Max-Planck-Institut für Biochemie

Abteilung Strukturforschung

Biologische NMR-Arbeitsgruppe

Biochemical and structural investigations of the retinoblastoma protein, its binding partners, and the BRG1 protein - a subunit of human SWI/SNF remodeling complexes

Mahavir Singh

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Dr. Adelbert Bacher
Prüfer der Dissertation:	1. apl. Prof. Dr. Dr. h. c. Robert Huber
	2. Univ.-Prof. Dr. Johannes Buchner

Die Dissertation wurde am 19.12.2005 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 14.03.2006 angenommen.

To my family and sister....

Acknowledgement

It has been a pleasure experience to work in the department of 'Strukturfoschung, NMR Gruppe'. I am indebted to all the people who helped me during this time.

First of all, I would like to thank my supervisor Dr. Tad A. Holak for his constant support, patience and encouragement. I am especially thankful to him for listening to all my ideas and for his guidance through out this time.

I am grateful to Professor Robert Huber for giving me the opportunity to work in his department and for being my 'Doktorvater'.

I would like to thank all the labmates: Ulli Rothweiler, Sudipta Majumdar, Aleksandra Mikolajka (Ola), Marcin Krajewaski, Przemyslaw Ozdowy, Tomasz Sitar , Grzegorz Popowicz, Loyola D'Silva, Joma Kanikadu Joy, and Kinga Brongel for creating a friendly atmosphere in the lab. My thanks also goes to previous labmates: Dr. Pawel Smialowski, Dr. Madhumita Ghosh, Dr. Sreejesh Shankar, Dr. Ashwini Thakur, and Dr. Igor Siwanowicz for their friendship.

I would specially like to thank, Ulli for his friendship, criticism, help in all matters, and coffee time; Igor for giving tips on muscle development in the gym, Gregorze for being patient with my questions on crystallography and Ola for helping me with many computer programmes. I would like to thank Pawel for the nice trip of Poland in September 2003.

I am grateful to Ms. Weyher-Stingl Elisabeth (Lissy) for introducing to me the techniques of ITC, CD, fluorescence etc. Many experiments in this thesis would not have been possible without her help.

Beside all these people, I would like to acknowledge the friendship of all the Indian friends in or arround the institute for providing the atmosphere of home. Especially, I would like to thank the Indian community for organizing many joyous events and cricket matches.

At last, all my gratitude and thanks goes to the constant support and encouragement of my family. Their love and sacrifices can never be repaid.

Publications

Parts of this thesis have been or will be published in due course:

Pawel Smialowski, **Mahavir Singh,** Aleksandra Mikolajka, Sudipta Majumdar, Joma K. Joy, Narashimha Nalabothula, Marcin Krajewski, Roland Degenkolbe, Hans-Ulrich Bernard and Tad A. Holak
NMR and mass spectrometry studies of putative interactions of cell cycle proteins pRb and CDK6 with cell differentiation proteins MyoD and Id-2
Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, 2005, *1750*:48-60

Mahavir Singh, Marcin Krajewski, Aleksandra Mikolajka and Tad A. Holak
Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences
Journal of Biological Chemistry, 2005, *280*:37868-37876

Loyola D'Silva, Przemyslaw Ozdowy, Marcin Krajewski, Ulli Rothweiler, **Mahavir Singh** and Tad A. Holak
Monitoring the Effects of Antagonists on Protein-Protein Interactions with NMR Spectroscopy
Journal of the American Chemical Society, 2005, *127*:13220-13226

Mahavir Singh, Loyola D'Silva and Tad A. Holak
DNA binding properties of the recombinant high-mobility-group-like AT hook-containing region from human BRG1 protein
Manuscript submitted

Mahavir Singh, Gregorze M Popowicz and Tad A. Holak
1.5 Å structure of bromodomain from human BRG1 and ramification for acetylated lysine recognition
Manuscript under preparation

Mahavir Singh and Tad A. Holak
Interaction of retinoblastoma protein with cellular targets. *Review*
Manuscript under preparation

Table of contents

1 Introduction

Eukaryotic cells have evolved a complex network of regulatory proteins to control a cellular event. One of the principle cell event is the cell cycle division. Understanding how various events of the cell cycle are controlled and coordinated is important, because their deregulation results in uncontrolled cell proliferation or 'tumorgenesis'. Studying various interactions at their molecular level would help the understanding of tumorgenesis and would aid in designing the therapy more rationally. In this thesis we studied some of the molecular players of two very important processes, cell cycle regulation and chromatin remodeling. Protein-protein interactions (direct or indirect as in complexes) and protein-DNA interactions (again direct or indirect) are the fundamental interactions through which a process is executed or controlled. These interactions are in turn regulated by the covalent modifications of proteins (e.g. phosphorylation, acetylation, SUMOylation, ubiqitination) or DNA (e.g. methylation).

The cell cycle is a complex interplay of various protein-protein interactions. The retinoblastoma protein (pRb) is called the master regulator of the cell cycle as it controls the cell proliferation by arresting the cell at G1 phase. Inactivation of pRb by either mutations or viral oncoproteins, or deregulation of its upstream regulators, causes tumorigenesis. pRb performs its function by binding to a plethora of proteins. In this thesis we set out to check some of these interactions at the molecular level, using purified proteins and a multimethod approach, which included NMR spectroscopy, ITC, mass spectrometry and other techniques. Section 5.1 documents the results and discussion of some of the interactions between pRb and other proteins with this approach.

Chromatin remodeling is a process by which cell makes its highly compact DNA in nucleosome accessible for the routine cellular processes like replication, transcription, and DNA repair. For these processes to occur the remodeling complex should be able to bind to the proteins and/or DNA components of the nucleosome. Mutations or deregulation of various components of chromatin remodeling complexes have been implicated in various tumors. BRG1 is the central ATPase of the human SWI/SNF chromatin remodeling complex. The AT hook motif of BRG1 is the DNA binding motif and bromodomain is the acetyl-lysine residue binding domain. Section 5.2 presents the results and discussion on the DNA binding properties of the AT hook motif of BRG1, and the structure of bromodomain at 1.5 Å and its interaction the acetylated histones.

2 Biological background

2.1 The cell cycle

The cell cycle is the process by which cells duplicate themselves, grow, and prepare to divide again. It is the basis for the reproduction and sustained growth of all living organisms. Its control in higher eukaryotes is important for preventing tumorgenesis, because cancer can occur when the regulation of cell cycle fails. In recent years the basic mechanisms underlying the cell cycle control have been unraveled and shown to be common to living organisms from yeast to the humans. The basic cell cycle is divided into four phases. During two of these phases cells execute the two basic events in cell division: generation of a single and faithful copy of its genetic material (the synthetic or S phase) and partitioning of all the cellular components between two identical daughter cells (mitosis or M phase). The two other phases of the cycle - G1 and G2 - represent 'gap' periods, during which cells prepare themselves for the successful completion of the S and M phases, respectively. When cells cease proliferation, either due to specific antimitogenic signals or to the absence of proper mitogenic signaling, they exit the cycle and enter a non-dividing, quiescent state known as G0. However, many G0 cells are anything but quiescent. They are busy carrying out their functions in the organism, e.g. secretion, attacking pathogens. Often G0 cells are terminally differentiated - they will never reenter the cell cycle but instead will carry out their functions in the organism until they die. For other cells, G0 can be followed by re-entry into the cell cycle. Most of the lymphocytes in human blood are in G0. However, with proper stimulation, such as encountering an appropriate antigen, they can be stimulated to re-enter the cell cycle (at G1) and proceed on to new rounds of alternating S phases and mitosis. G0 represents not simply the absence of signals for mitosis but an active repression of the genes needed for mitosis. Cancer cells cannot enter G0 and are destined to repeat the cell cycle indefinitely (Voet and Voet, 1995; Bruce et al. 2002).

To ensure proper progression through the cell cycle, cells have developed a series of checkpoints that prevent them from entering into a new phase until they have successfully completed the previous one. It is likely that newly divided or quiescent cells must also pass certain checkpoints before they can enter the cycle. For instance, cells must make sure that they have reached their homeostatic size; otherwise cells will become smaller with each round of division. Metazoans must also control the number of cells in every organ, which - coupled with cellular size - determines the size of the organism. The genetic elements that control these

parameters remain largely unknown. However, they can be influenced by external elements, such as the amount of available nutrients or the intensity of the mitogenic information that cells receive at any given time. Indeed, these two parameters are key determinants for cell division. For instance, if cells have mitogenic requirements that are too stringent, they might not be able to proliferate at crucial times, such as during wound healing or during an infection. Conversely, loosening these controls might lead to unscheduled proliferation and possibly neoplastic growth. In culture, cells undergo a period of mitogen dependence before they enter the cycle. This transition from mitogen dependence to mitogen independence called - the restriction point (R)- represents a point of no return that commits cells to DNA replication and a new round of cell division. Understanding the molecular mechanisms that allow cells to enter the cycle should provide important clues as to the determinants of normal versus abnormal proliferation in multicellular organisms. The cell responds to the external stimuli and communicates it through a cascade of intracellular phosphorylations cascade, by upregulating expression of the cyclins which associate with the cyclin-dependent kinases (CDKs) which in turn are controlled by regulatory proteins. CDKs have an active kinase subunit and a regulatory subunit, cyclin. CDK/cyclin complex is subjected to several kinds of regulation, both positive and negative, for instance, by reversible phosphorylation by CDK activator kinases and specific CDK phosphatases. In the CDK/cyclin holoenzyme, cyclin accounts for both substrate specificity and cell phase specificity of the holoenzyme. Their level fluctuates during the cell cycle, some being more abundant in specific cell phases than other (Table 2.1). Cyclin level is primarily regulated by gene expression (transcriptional regulation) and protein degradation (proteolytic degradation, ubiquitination). Cyclins respond to mitogenic signals, and unscheduled expression leads to uncontrolled proliferation, implicated in different cancers.

Table 2.1. *Various cyclins and associated CDKs involved during different phases of mammalian cell cycle*.

Cyclin	***Associated CDK***	***Function***
A	CDK1, CDK2	S phase entry and transition Anchorage-dependent growth
B1, B2	CDK1	G2 exit, mitosis
C	CDK8	Transcriptional regulation, G0-to-S–phase transition
D1, D2, D3	CDK4, CDK6	G0-to-S–phase transition
E	CDK2	G1-to-S–phase transition
F	?	G2-to-M–phase transition
G1, G2	CDK5	DNA damage response
H	CDK7	CDK activation, transcriptional, regulation, DNA repair
I	?	?
K	?	Transcriptional regulation, CDK activation
T1, T2	CDK9	Transcriptional regulation

There are protein inhibitors of CDKs, which negatively regulate the activity of CDKs. These inhibitors belong to two families: the Cip/Kip family ($p27^{Cip/Kip}$, $p21^{Cip/Kip}$ and $p57^{Cip/Kip}$) and INK4a ($p15^{INK4a}$, $p16^{INK4a}$, $p18^{INK4a}$ and $p19^{INK4a}$) family, INK4a family which are potent inhibitor of cyclin D-CDK4/6 complex.

There are proteins called 'tumor suppressor' which act as effectors of cell cycle clock. Some of them are extrinsic factors in that they only act if the cell is damaged. The most important are the retinoblastoma protein (pRb) and the transcription factor p53. The main function of pRb is to connect the cell cycle machinery to the transcriptional machinery (intrinsic mechanism). The pRb protein has a critical role in regulating G1 progression through the restriction point and will be discussed later in detail. p53 is a DNA binding protein that regulates the expression of genes involved in cell cycle arrest. It senses DNA damage and tells the cell to either stop proliferating (until damage is repaired) or to kill itself by apoptosis, a typical extrinsic mechanism. pRb and p53 pathways are now believed to be disregulated in almost all human cancers (reviews; Sandal, 2002; Sherr, 1996; Sherr, 2000; Sherr, 2002; Vogelstein and Kinzler, 2004) .

2.2 The retinoblastoma protein

The retinoblastoma gene was the first tumor suppressor to be cloned (Friends et al. 1986). The retinoblastoma protein (pRb), a 928 amino acid nucleophosphoprotein has been implicated in many cellular processes, such as regulation of the cell cycle, DNA-damage responses, DNA repair, DNA replication, protection against apoptosis, and differentiation, all of which could contribute to its function as a tumor suppressor (Knudsen et al., 1998; Wang et al., 1994). The pRb gene is closely related to two genes in mice and humans; p107 and 130, which, interestingly are not commonly mutated in tumors. It is now evident that pRb, p107, and p130, together known as the 'pocket proteins,' are central participants in a gene regulatory network that governs the cellular response to antimitogenic signals, and whose deregulation constitutes one of the hallmarks of cancer (Hanahan and Weinberg, 2000; Cobrinik, 2005). Biochemical experiments have provided information regarding the nature of various protein interactions with pRb. There are more than 130 reported pRb binding proteins till today, hence much detailed structure/function mapping is required to clarify the biological relevance of all these interactions. The mechanism behind the role of pRb as a tumor suppressor is not clearly understood, however it is clear that pRb controls the passage of a cell through G1-S transition point. pRb is the target of oncoproteins that are expressed by DNA tumor viruses indicated that pRb is a regulator of cell proliferation. The discovery that pRb overexpression caused cells to undergo arrest in the G1

phase of the cell cycle, whereas cells that are deficient in pRb gene show an accelerated G1 transition, also provided support for the idea of pRb as a cell-proliferation inhibitor.

The identification of viral oncoproteins that interact with and presumably interrupt the function of pRb also raised the possibility that analogous cellular proteins could exist. Subsequent experimentation led to the identification of the E2F1 transcription factor as the first cellular target of pRb.

An active 'hypophosphorylated' form of pRb arrests the cell in the early G1 phase whereas a inactive 'hyperphosphorylated' form of pRb allows the cell to pass through the G1-S restriction point. This process is highly regulated under normal cellular conditions, however, disregulalation leads to uncontrolled proliferation leading to cancer. How pRb regulates the G1-S transition point will be discussed later in the Section 2.2.3.

2.2.1 Domain organization of pRb

pRb contains four structural domains out of which three are protease resistant *in-vitro*. The N-domain (beginning at amino acid 8) is 40 kDa in size and constitutes most of the amino terminal half of pRb. Continued proteolysis of the N-domain generates two subfragments of 30 kDa (beginning at amino acid 8) and 10 kDa (beginning at amino acid 263). The rest of the pRb comprises domains A, B and C. The A (beginning at amino acid 379) and B (beginning at amino acid 642) domains have apparent molecular mass of approximately 24 and 18 kDa, respectively, and map to the carboxy terminal half of the protein. The A and B structural domains correspond essentially to functional domains that have been mapped genetically. Collectively, the A domain, linker region, and B domain are called the 'small pocket' (Fig. 2.1).

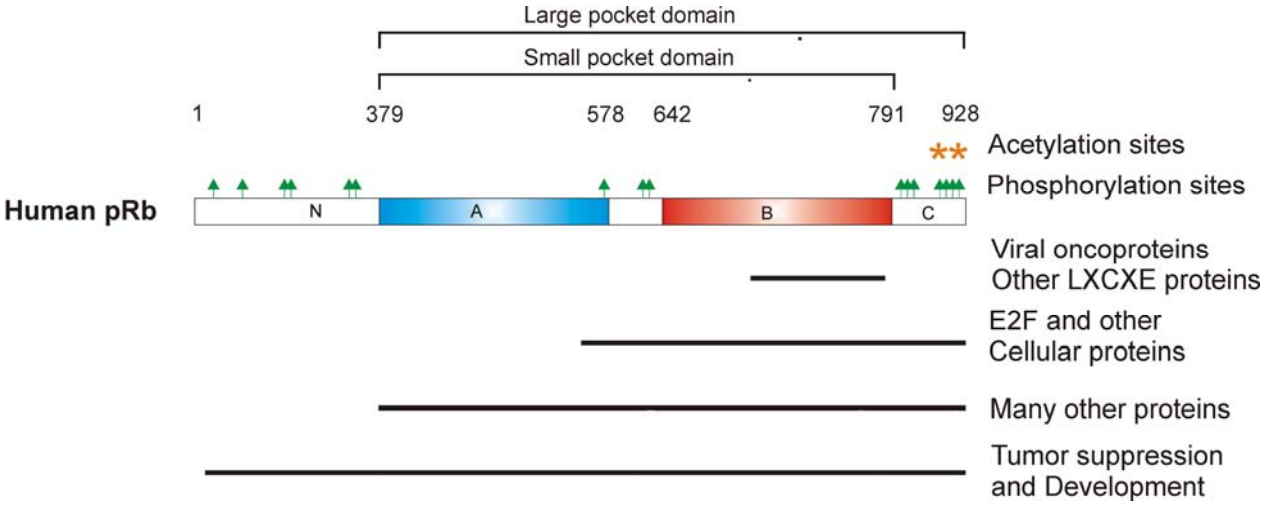

Figure 2.1. *Domain organization of the human pRb.*

The importance of the small pocket for pRb-mediated tumor suppression is underscored by the fact that many naturally occurring RB mutations disrupt its integrity (Chow and Dean, 1996; Lee et al., 1998). The A and B domains are also the most highly conserved regions of pRb throughout evolution (Fig. 2.1). Disruption of the small pocket compromises pRb activity in most *in-vitro* functional assays. Phosphorylation sites are also noticeably absent within the A/B domains suggesting that modification of this structure is not tolerated in the active protein. It is likely therefore that the integrity of the small pocket is essential for normal pRb function. In addition to the small pocket, sequences carboxy terminal to the B domain are critically important for pRb function. Amino acids 773–928 constitute the C-domain. Unlike the N-, A-, and B-domains, the C-domain does not exist as a detectable, protease resistant structure. Yet, the C domain is sufficient to mediate interactions with a subset of pRb binding partners including CDK2, PP1a, UBF, MDM2, and c-Abl. The C-domain together with the small pocket is called the 'large pocket'. With some exceptions, the large pocket is sufficient to mediate interaction with the numerous pRb binding partners, including E2Fs. The large pocket is also necessary and sufficient for pRb-mediated *in-vitro* assays such as regulation of the cell cycle (Hu et al., 1990; Huang et al., 1990; Kaelin et al., 1990; Edwards et al., 1992).

2.2.2 Pocket proteins

The 'pocket proteins' (pRb, p107 and p130) are best known for their roles in restraining the G1-S transition through regulation of E2F-responsive genes. The gene location for pRb is at chromosome 13q14. The gene for p107 (also called RBL1) is located at 20q11.2. The gene for p130 (also called RBL2) is located at 16q12.2 (Baldi et al., 1996; Claudio et al., 2002; Classon and Dyson, 2001; Ewen et al., 1991; Hong et al., 1989).

Similarities and differences among pocket proteins

Similarity at the gene level

The gene structure for all the three proteins is similar to the other house-keeping genes, including the lack of canonical TATA boxes or CAAT boxes found in the promoters of most of the differentially expressed genes, presence of a GC rich zone immediately surrounding the main transcription initiation site, the presence of multiple consensus sequences for binding of SP1 transcription factor, and the presence of multiple transcription sites.

Primary amino acid sequence similarities

The primary sequences have the homology beyond the pocket region. However p107 and p130 have more sequence homology compared to the pRb (Fig. 2.2A). A cyclin A/E binding region is present in the linker region between A and B domain, only in p107 and p130.

Ability to bind to viral oncoproteins

All three proteins interact with viral oncoproteins, such as, HPV E7, SV 40 large T antigen and adenovirus E1A. The viral proteins posses a LXCXE-like motif and interact with the B subdomain on the pocket proteins.

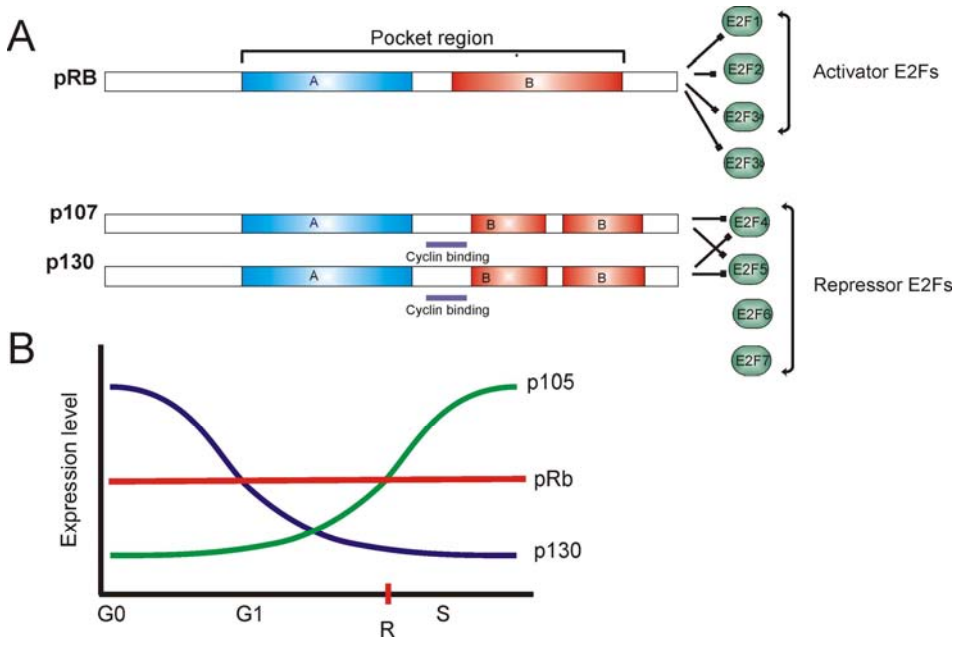

Figure 2.2. *Pocket proteins. (A) Genomic structure of pocket protein and their interaction with various E2Fs. (B) Expression pattern of pocket proteins during cell cycle.*

Ability to bind to E2F transcription factors

Each member of the pocket protein family associate with members of the E2F transcription factor family, and this common activity may provide the molecular basis for many of their overlapping functional properties. However, there is selectivity for E2Fs; whereas pRb associates *in-vivo* with E2F-1, -2, and -3, p107 and p130 associate almost exclusively with E2F-4 and E2F-5. The basis for this specificity is unclear but in general terms p107 and p130 interact with E2Fs that are primarily considered to be corepressor molecules, whereas pRb regulates E2Fs that are strong activators (Dimova and Dyson, 2005; Crobinik, 2005).

Differences in the expression pattern during cell cycle

p130 is highly expressed in quiescent and differentiated cells, and its levels drop rapidly when quiescent cells are stimulated to enter the cell cycle (Fig. 2.2B). In contrast, p107 levels are generally quite low in terminally differentiated cells and its levels rise when quiescent cells are stimulated to proliferate. Moderate levels of pRb can be found in most cell types as well as in

both quiescent and cycling cells (DeCaprio et al., 1989). At present, it is unclear which activities of these proteins are affected by these dynamic fluctuations in pocket protein levels.

Ability to form stable complexes with the cyclin A/E-dependent kinases

The p107 and p130 linker regions contain homologous sequences and include a high-affinity binding site for cyclin A/CDK2 and cyclin E/CDK2. This motif, which is not present in the pRb sequence, allows p107 and p130 to form stable complexes with these CDKs.

Mouse knockout phenotypes

Heterozygous gene deletion of p107 and p130 do not result in tumor in mice and cell remains unchanged, only observed phenotype is increases adipogenesis, whereas heterogyous gene deletion of pRb is lethal and mice die at E13-15 stage due to thyroid and pituitary tumor (reviewed in Whyte, 1995; Clauodio et al., 2002; Classan and Dyson, 2001; Cobrinik, 2005; Classon and Harlow, 2002).

2.2.3 pRb and G1-S control

The G1-S checkpoint, originally proposed by A. Pardee (1974, 1989) and called the restriction point, defines a time point in G1 at which cells are committed to enter S phase, even in the absence of growth factors (reviewed in Herwig and Strauss, 1997; Weinberg, 1995). Prior to this point, however, the cell may take alternative routes such as differentiation, senescence, or cell death, depending on the external signals that are sensed by key regulatory proteins. Comprehensive studies on the G1 restriction point have now identified the D-type cyclins, CDK4/CDK6, pRb and certain CDK-inhibitors (CKIs) as key regulators of a common pathway controlling the commitment to enter S phase. G1 progression requires sustained cyclin D expression, persisting as long as there are mitogen signals. If the mitogens are removed, the level of cyclin D rapidly decreases and the cell is arrested in G1 phase. A similar block in cyclin D-kinase activity and subsequent arrest in G1 phase is achieved by the INK4a family of CKIs. Four members of this family ($p15^{INK4a}$, $p16^{INK4a}$, $p18^{INK4a}$ and $p19^{INK4a}$) are known to bind and inhibit CDK4 and CDK6, without affecting other CDKs. In contrast Cip/Kip family of CKIs ($p27^{Cip/Kip}$, $p21^{Cip/Kip}$ and $p57^{Cip/Kip}$), play a positive role by stabilizing the CDK/cyclin complex.

The ultimate substrate of cyclin D/CDK4/6 in this pathway is the pRb (Fig. 2.3) and can therefore be considered to be the key regulator that holds proliferation in check (Goodrich and Lee, 1993; Hatakeyama et al., 1994; Lukas et al., 1994a; Mittnacht et al., 1994; Sherr, 1994). It does so by sequestering and thereby inactivating a set of regulatory proteins (of which the best characterized are E2Fs) that favors cell proliferation (Weinberg, 1995). This growth inhibitory function of pRb only applies to its hypophosphorylated ‘active’ form which predominates in

quiescence and early G1 phase, and this form of pRb binds to the E2F family of transcription factors. However as cell passes through G1 phase, pRb gets phosphorylated by cyclin D/CDK4/6 complexes, which became active via different signaling pathways, and the resulting hyperphosphorylated 'inactive' form of pRb can no longer bind to E2Fs (Buchkovich et al., 1989, Chen et al., 1989, DeCaprio et al., 1989, Ludlow et al., 1990; Mihara et al., 1989) (Fig. 2.3). E2F proteins act as transcription factors for many genes whose products are necessary for DNA synthesis and S phase progression. E2F triggers expression of proteins like dihydrofolate reductase (DHFR), thymidine kinase, different DNA polymerases and the late G1 cyclin E. Expression of cyclin E establishes a positive feedback loop of pRb phosphorylation, since cyclin E activates CDK2 and the cyclin E/CDK2 complex will continue to phosphorylate Rb, contributing to an irreversible transition into the S phase and cell cycle progression, even in the absence of growth factors. There is one more mechanism to activate CDK2. A cyclin D/CDK4/6 complex can bind to WAF1/KIP inhibitors without losing their kinase activity. This interaction titrates these inhibitors away from cyclin E/CDK2 complexes that activate CDK2 activities. Cells do not rely exclusively on the CDK4/6 activity to fully activate CDK2 kinase. Indeed, Kip1 is phosphorylated by CDK2 as well as by other, as yet unknown kinase that triggers its proteasome mediated degradation. Decrease in the CDK2 activity at the G1-S transition is mainly controlled by degradation of cyclin E, for which a specific ubiquitin ligase has been identified (Malumbres and Barbacid, 2001). Also cyclin D/CDK4 complexes phosphorylate Smad3, negatively regulating the functions of transcriptional complexes that mediate cell growth inhibition by proteins of the TGF-β family (Matsuura et al. 2004). Thus, the cyclin D-dependent kinases now appear to cancel the activities of at least two families of proteins (E2F family via phosphorylation of pRb and Smad3) that negatively regulate cell cycle dependent gene expression.

The repressor activity of pRb stems from its two types of activities. First, it binds directly to the transactiation domain of E2F in its hypophoyphorylated form, thereby blocking its activity, as described above. Second, it can actively represse transcription while it is tethered to promoters by E2F through recruitment of chromatin remodeling enzymes. These chromatin remodeling enzymes fall into to classes: histone deacetylases (HDACs) and ATP-dependent SWI/SNF complexes (Fig. 2.3). Direct recruitment of HDACs and SWI/SNF complexes to pRb has been reported, however, these interactions seems to be complicated, as other reports suggested indirect regulation of pRb. These complexes mediate chromatin condensation and subsequent inhibition of transcription (Harbour and Dean, 2000; Muchardt and Yaniv, 1999; Muchardt and Yaniv, 2001).

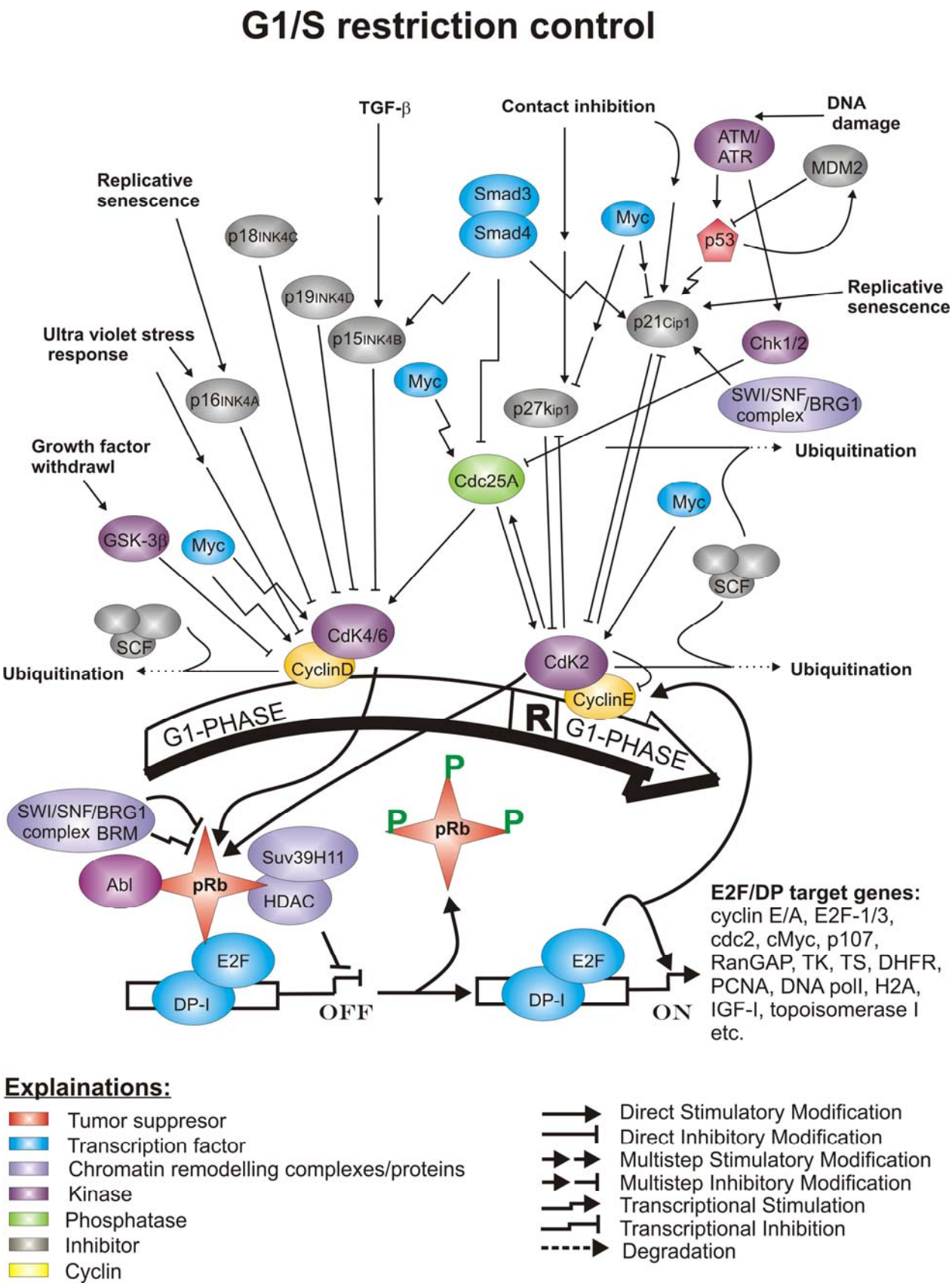

Figure 2.3. *Interplay of various proteins which play role in the G1-S transition in a mammalian cell. The cyclin/CDKs complexes are the censors for various mitogens or GFs which in turn regulate the activity of pRb by changing its phosphorylation status. pRb arrests, cells in the G1 phase by sequestering the E2F transcription factor. Phosphorylation of pRb leads to release of E2F, which in turn helps in transcription of various genes necessary for G1-S transition and S phase progression.*

Recent discoveries

Until now the 'classical view' for the entry into, progression through, and exit from the G1 phase of the mammalian cell cycle in response to extracellular mitogenic cues are presumed to be governed by cyclin-dependent kinases (CDKs) regulated by the D-type and E-type cyclins as discussed above. Studies performed over more than a decade have supported the view that these holoenzymes are important, if not required, for these processes. However, recent experiments in which the genes encoding all three D-type cyclins, the two E-type cyclins, cyclin D dependent CDK4 and CDK6, or cyclin E-dependent CDK2 have been disrupted in the mouse germ line, have revealed that much of fetal development occurs normally in their absence. Thus, none of these genes is strictly essential for cell cycle progression. To what extent is the prevailing dogma incorrect and how can the recent findings be reconciled with past work remains to be seen (Reviewed by Sherr and Roberts, 2004).

2.2.4 Structure of the pRb 'small-pocket'

The 3D structure of the pRb small pocket domain has been solved by X-ray crystallography. The structure of the A domain came first (Kim et al., 1997) and showed that it is helical with a cyclin-like fold. Later the structure of the complete small pocket, i.e. pRb-AB was solved in complex with peptides from HPV E7, E2F1, E2F2, and the N-terminal LXCXE containing domain from the SV40 large T antigen (Lee et al., 1998; Lee et al., 2002; Xiao et al., 2003; Kim et al., 2001). The structure of the small pocket showed that in addition to 10 helices of the two cyclin folds, it contains eight other helices, a β-hairpin, and an extended tail. The cyclin fold, which forms the basis of the A- and B- domain structures, consists of a three-helix bundle with two additional helices packing on two sides, giving a five-helix core motif (α3, α4, α6, α7, α8 for the A domain, and α11, α12, α14, α15, α17 for the B domain). The additional eight helices, β-hairpin, and extended tail of the pRb pocket serve either at the A-B interface (α9, α10, α13), or the LXCXE binding site (α16, βhairpin, α18), or cover surface regions of the A and B cyclin folds (α1, α2, α5, tail) (Fig. 2.4A, B). The interface between the A- and B-domain involve extensive hydrophobic interactions and constitutes the integral part of the overall pocket structure. This interface is continuous with that of the B domain, suggesting that the B domain needs the A-B interface for its stable folding. The A–B interface is formed through the packing of portions of seven helices, three from the A domain (α8, α9 and α10) and four from the B domain (α11, α12, α13, α14), and has a buried surface area of 2,013 $Å^2$. The interactions between the A and B domains are mediated by a compact hydrophobic core of about 20 side chains and networks of backbone and

side-chain hydrogen bonds. The importance of the conserved A-B interface can be understood by the fact that many missenese tumorigenic mutations are located in this region.

The LXCXE motif from the HPV E7 and SV40 large T antigen binds at a shallow hydrophobic surface surrounded by positively charged lysines residues on the B domain (Fig. 2.4A). The LXCXE peptide adopts a β-strand like extended conformation (Lee et al., 1998). For three major interactions of L, C, and E in the minimal LXCXE, the side chains of the conserved leucine and cysteine fit tightly to the hydrophobic groove on pRb-AB, whereas the carboxylate group of glutamic acid of the terminal LXCXE makes two hydrogen bonds with two backbone amide groups of helix-15 of pRb-AB. The interaction between pRb-AB and N-terminal domain of the SV40 large T antigen revealed that two-third of the total interaction occurs via the LXCXE motif. These studies showed that the LXCXE motif is the primary site of interaction for HPV E7 and the large T antigen (Kim et al., 2001).

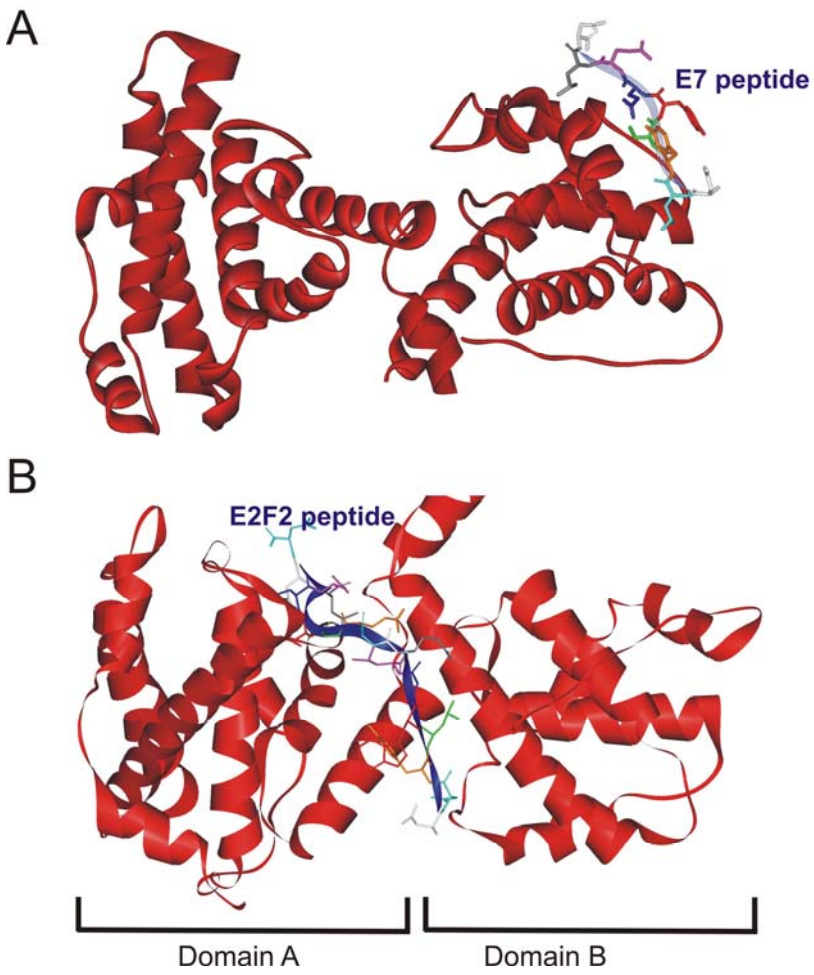

Figure 2.4. *X-ray structure of the pRb small pocket (pRb-AB). (A) The [pRb-AB]-[E7 LXCXE peptide] structure (adapted from pdb file 1GUX). (B)[pRb-AB]-[E2F2 transactivation domain peptide] structure (adapted from pdb file 1N4M).*

In contrast, the E2F peptide from its transactivation domain binds at the interface of A and B domains (Lee C et al. 2002; Xiao et al. 2003). The E2F peptide binding site is separated by ~30 Å from the E7 peptide binding site (Fig. 2.4B). In structures of pRb-AB/E2F1 and pRb-AB/E2F2, the peptides bind in a similar manner. N-terminal segment of the peptides adopt extended β-strand-like conformations and the six-residue C-terminal segments form a helical shape containing a 3_{10} helix. The two segments are separated by a bend formed from the five amino acid residues in the middle of the peptide. The conserved residues in the N-segment and C-segment of the E2F peptides that are biochemically known to affect binding of E2F to pRb make a major contribution to the buried surface area upon complex formation.

2.3 The retinoblastoma protein and cancer

Retinoblastoma as name suggests is the cancer of retina, first realized in 1951 in child patients. It was found to occur sporadically in some patients, but to be inherited in others (Neel and Falls, 1951). Later, pRb was found to be mutated in many additional cancers, such as small-cell lung carcinoma and osteosarcoma, cervical carcinomas, prostate carcinomas, breast carcinomas, and some forms of leukemia (Horowitz et al., 1989; Onadim, 1992; Sellers and Kaelin, 1997; Yandell, 1989). It is now believed that the pRb pathway together with the p53 tumor-suppressor gene pathway is inactivated in most of the cancers. Many regulators in the pRb pathways are also mutated in human cancers. The alterations include overexpression of cyclins (mainly D1 and E1), CDKs (mainly CDK4 and CDK6) and E2F1, as well as loss of CDK inhibitors (p16/INK4 and KIP1) and pRb expression. Hence the frequent loss of G1 regulation in human cancers has revealed targets for possible therapeutic intervention. Indeed restoring the restriction point (R point) control to cancer cells might allow them to return to quiescent state. Mainly CDKs are being targeted to find the drugs. However the best strategies to accomplish these goals have not been determined, and will still require proper experimentations, especially in animal models (Malumbres and Barbacid, 2001).

2.4 pRb binding partners

Over the years pRb has been shown to interact with more than 130 proteins (cellular as well as viral oncoproteins). These proteins belong to diverse class and family of proteins like for example, kinases, phosphatases, kinase-regulators, transcriptional regulators etc. which includes many other proteins that have not yet been shown to directly regulate transcription (Morris and

Dyson, 2001). This plethora of pRb binding protein raises several questions like, how many functions does pRb possess, which of these functions are important for development, and which contribute to tumor suppression? And above all which interactions are genuine. Researchers have mainly, used 'secondary techniques' like, Western-blot, co-immunopprecipitation, pull down assays for detecting interaction to pRb. The strength of these techniques is their sensitivity. However these techniques are prone to give false positive and are not suitable to study weak interactions. Also in many cases, these experiments are performed with whole cell lysate and the interactions are interpreted for the individual protein-protein interactions. We currently understand very little about how individual domains of pRb are regulated *in-vivo* and how these domains cooperate to modulate pRb functions. Small changes in pRb structure could translate into large functional outcomes in cellular function. To add further complexity, very little is known about how the pool of pRb is divided among such a large number of potential targets. Given the possibility that pRb might simultaneously bind to several distinct proteins it will be crucial to know what events regulate the assembly of different multisubunit complexes. These issues are very difficult to address, especially when one considers that the binding sites of many pRb-binding proteins have not been mapped on pRb with any degree of precision. Most worryingly, it is unclear how many of these interactions between pRb and binding proteins are even biologically relevant. Surely more work is required to find the genuine pRb binding partners. We studied some of these interactions as part of this thesis.

2.4.1 E2F transcription factors

E2F is the collective activity of a heterogeneous family of E2F/DP heterodimers (Cobrinik, 1996). Of the eight E2F genes cloned to date, four gene products (E2F1, E2F2, E2F3a, E2F3b) have been found in complexes with hypophosphorylated pRb (Shan et al., 1992; Kaelin et al., 1992; Helin et al., 1992; Ivey-Hoyle et al., 1993; Lees et al., 1993; Krek et al., 1993; Cress et al., 1993; Moberg et al., 1996; He et al., 2000; Leone et al., 2000). These interactions have been found both *in-vitro* and *in-vivo* (with endogenous and overexpressed proteins) in multiple cell types. A small pRb binding domain has been identified in E2F-1 that is conserved in five E2F genes (E2F1-5) and missing in E2F-6. This domain is embedded in the transactivation domain of E2F proteins, suggesting that pRb sterically prevents transcriptional activation. E2F binds with low affinity to the small pocket domain of pRb, but with high affinity to the large pocket. pRb-associated E2F complexes can contain either DP-1 or DP-2 which potentiate the pRb binding activity of E2F polypeptides (Girling et al., 1993; Bandara et al., 1994; Wu et al., 1995). The C-

terminal region of DP-1 has been shown to bind pRb *in-vitro* in a pocket-independent fashion (reviewed in Dyson, 1998; Dimova and Dyson, 2005; Crobinik, 2005)

2.4.2 MyoD and Id proteins

A role of pRb was suggested in the muscle differentiation. pRb cooperates with MyoD (an activator of muscle specific gene expression and myogenic conversion) to promote muscle differentiation when they are both expressed in Saos-2 cells after transfection (Gu et al., 1993). Moreover, MyoD induces pRb expression (Martelli et al., 1994) and pRb expression is required for proper muscle gene expression (Novitch et al., 1996). In this regards pRb is actually acting as a transcriptional activator. In addition, when myoblasts are formed from *pRb-/-* cells *in-vivo*, the cell nuclei of differentiated cells are able to re-enter the cell cycle in response to serum stimulation, suggesting that pRb is required for muscle cells to remain quiescent (Schneider et al., 1994). It was also shown that pRb is required for myoblasts to exit the cell cycle, for the maintenance of myoblasts, and for the formation of myofibers from myotubes. Gu and coworkers (1993) found MyoD/pRb and myogenin/pRb complexes in extracts of differentiated C2 myotube extracts and demonstrated that *in-vitro* association between these proteins was mediated by the basic helix-loop-helix region (bHLH) of MyoD and the C-terminal half (amino acids 605-928) of pRb. pRb/MyoD complexes were competed by a T Antigen peptide. Other myogenic bHLH factors, including myogenin, Myf-5, and MRF4 were reported to bind pRb similarly (Gu et al., 1993). However, several other *in-vivo* studies could find no evidence for such binary interactions suggesting that if there is any direct interaction it could be weak, making the significance of the *in-vitro* MyoD-pRb interaction questionable (Guo et al., 2003; Puri et al., 2001; Zhang et al., 1999a; Zhang et al., 1999b). Using the two-hybrid system Zhang et al. (1999a) could not detect direct MyoD or myogenin-pRb interactions.

Id proteins (Id-1 to Id-4) are another class of bHLH-leucine zipper transcription factors that inhibit cellular differentiation in several systems. Id-2 was initially reported to reduce the growth suppression activities of pRb on Saos-2 cells without affecting the phosphorylation status of pRb (Iavarone et al., 1994; Pagliuca et al., 1995). These proteins promote cell proliferation by binding to transcription factors from the bHLH subclass of proteins and inhibit their ability to bind DNA (Norton et al., 1998). Id-2 and MyoD have high degree of sequence homology. Only Id-2, and not other Id proteins, was reported to disrupt antiproliferative effects of pRb proteins via direct interactions (Lasorella et al. 1996). The bHLH domain of Id-2 was shown to interact with hypophosphorylated form of pRb (small pocket domain), p107, and p130.

We could not detect any binding between pRb-AB and MyoD or Id2 *in-vitro*. We postulated that indirect interactions, through additional binding partners in multiprotein complexes or modulation of gene expression levels of these proteins, are therefore their probable mode of action.

2.4.3 LXCXE sequence containing proteins

Since many viral pRb-binding proteins contain a conserved LXCXE pRb-binding motif, several studies have used the presence of an LXCXE-motif in a candidate cellular pRb-binding protein as evidence that it is a genuine pRb interactor. These types of arguments should be treated cautiously. For example, the β-chain of human insulin contains an LXCXE motif and was claimed to interact with the pRb (Radulescu and Wendtner, 1992), yet a biologically relevant interaction with pRb is clearly unexpected as one protein is extracellular and other resides in the nucleus. While the residues that contact viral LXCXE peptides have been conserved in pRb homologs from plants to animals, it is unclear which of the cellular proteins that might interact with this site are important for pRb function. Indeed, recent studies have shown that mutagenesis of the LXCXE-binding cleft prevents pRb from interacting with viral oncoproteins but results in a pRb-molecule that retains many of its cell cycle arrest and growth suppression properties (Dick et al., 2000; Chen and Wang, 2000; Dahiya et al., 2000). Thus, even for the LXCXE-containing proteins, a careful structure function analysis is needed before the significance of this interaction is clear. During this thesis we have shown that amino acid residues other than conserved L, C and E are also important for the interaction making it complex in nature. A few case scenarios could be imagined regarding the interaction between pRb and LXCXE sequences containing proteins. First, since the LXCXE motif is relatively small, it is not surprising to find this in the primary sequence of a protein, therefore if there is an interaction, this could still be biologically not relevant. Second, if the LXCXE motif is the part of a secondary structure like α-helix or β-sheet, this may not be in proper orientation to bind to pRb. In such a case, if during the experiment protein get denatured, the secondary structure can be lost and now this inactive protein can interact with the pRb, which can still be biologically irrelevant. Third, as we have shown, other residues beside L, C, and E are important for the interaction, as the surface of pRb, where LXCXE sequence binds is unique and surrounded by positively charged residues, therefore positive residues in or surrounding the LXCXE sequences would have repelling effect on the interaction. More in this subject is discussed in the section 5.1.7 of this thesis.

2.4.3.1 Viral oncoproteins

Adenovirus, human papilloma virus and simian virus 40, are the three viruses, whose interaction with pRb pathway has been studied in great detail. All these viruses rely heavily on the host cell for the synthesis of their own genome. They have developed ways to undermine the regulation of the normal cellular growth of their hosts and to stimulate progression from G1 to S phase so that the viruses can efficiently use the cellular DNA replicating machinery. One important way that viruses subvert normal cell growth is by encoding and expressing proteins during the early stages of viral infection that target key regulators of cellular proliferation and block the cellular transcription system. These viral proteins provide a strong stimulatory force that induces infected cells to assemble all of the machinery needed for DNA synthesis. This induction of DNA synthesis and subversion of cell-cycle regulation is achieved by the E1A in adenoviruses, the large T antigen in the case of SV40 and related polyoma viruses and the E7 gene product of human papilloma viruses. Although these proteins have only a weak similarity in their overall primary structures, they share a limited number of conserved regions and have similar biological properties (Moris and Dyson, 2001).

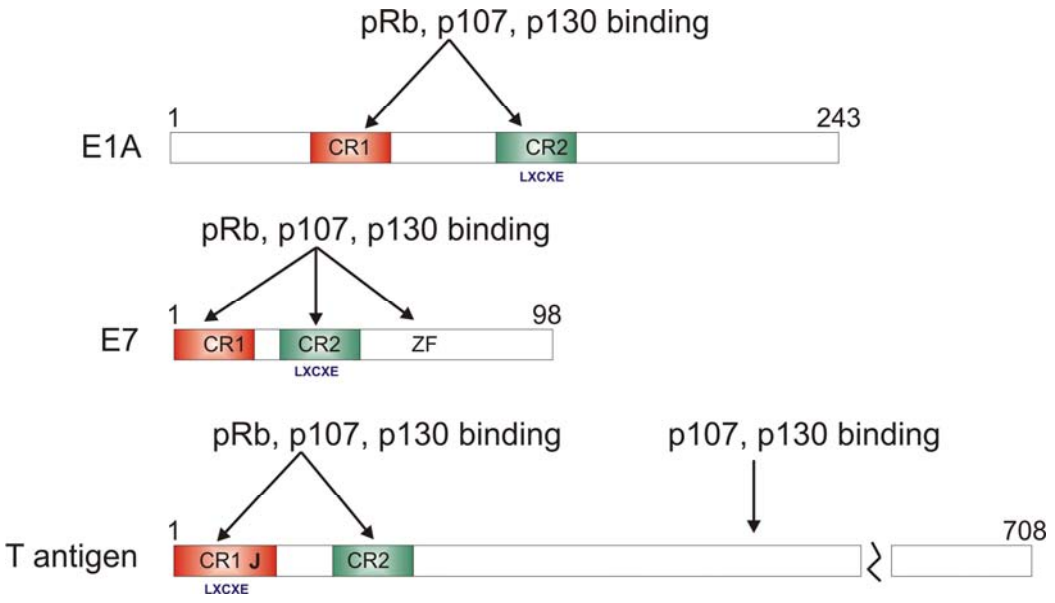

Figure 2.5. *Domain organization of E1A, E7 and T antigen viral oncoproteins and their interaction with pocket proteins. Conserved domains are highlighted in colors.*

E1A, E7 and T antigen

All of these viral oncoproteins have two conserved regions in their sequences called CR1 and CR2, which are the regions of interaction with the pRb (Fig. 2.5). E1A was the first protein found to be associated with pRb and other pocket proteins. The 'LXCXE' motif in case of E1A is located within the CR2 of the protein, which serve as the main binding site for pRb, CR1 binds to

the pRb with low affinity. HPV E7 was found to be expressed in human cervical cancers. Three regions of the E7 make association with the pRb. In the N-terminal half of E7 are CR1 and CR2 that share sequence homology with E1A CR1 and CR2 respectively. The third region is unusually large zinc finger in the C-terminal half of E7, which does not share homology with E1A. Like E1A, E7 possesses an LXCXE motif in CR2. In some studies it was also proposed that the E7 acts by degrading pRb and effectively reduces its concentration. The natural host of SV40 is the *Rhesus macaque*, however, SV40 tumor T antigen sequences and protein have been detected in human cancers and found to be associated with pRb. The SV40 large T antigen shares sequence homology with CR1 and CR2 of E1A and E7. The LXCXE sequence is located in the CR1 region of the protein. All three proteins bind to the hypophosphorylated form of pRb (Helt and Galloway, 2003; Lee and Cho, 2002).

2.4.3.2 Histone deacytylases (HDACs)

There are two classes of HDAC gene products: class I consists of HDAC1, 2, 3 and 8, whereas class II proteins include HDAC4, 5, 6, and 7. HDAC1 and 2 have been found to associate with two large histone-regulating complexes: a) the Sin3a/HDAC complex, which also contains RbAp46 and 48, as well as Sin3a, SAP18, and SAP30 (Zhang et al., 1997), and b) the NuRD complex, which also contains both RbAp46 and -48, as well as Mi2, p70, and p32 (Zhang et al., 1999). pRb binding has been detected with a number of these proteins suggesting that pRb function might occur within a large multisubunit regulatory network. In 1998, three simultaneous reports demonstrated that HDAC1 physically interacts with pRb to cooperatively repress E2F-mediated gene expression (Brehm et al., 1998; Luo et al., 1998; Magnaghi- Jaulin et al., 1998). The 'small pocket' was found to be necessary and sufficient for HDAC1 binding and transcriptional repression. Further analysis of pRb/HDAC interactions demonstrated that endogenous HDAC1, 2, and 3 bound pRb *in-vitro* and coimmunoprecipitated with the endogenous, hypophosphorylated pRb from H1299 cell extracts *in-vivo* (Lai et al., 1999; Nicolas et al., 2000). Moreover, a HDAC1 competing peptide, containing the LXCXE binding site, prevented pRb-precipitated polynucleosomal deacetylase activity in Jurkat extracts, and HDAC interactions with pRb could not be detected in 293T cells, which express E1A and T antigen suggesting common mode of interaction. p107 and p130 have also been shown to bind HDAC1 (Ferreira et al., 1998). Interestingly, although HDAC1 and 2 contain IXCXE motifs, HDAC3 does not, yet all three have been reported to be associated with pRb. Recent data suggest that HDACs might be tethered to pRb by other linker proteins, such as RBP1 (Lai et al., 1999). During this thesis we found that the IXCXE motif from HDAC1 could bind to the pRb-AB,

however, with the weaker affinity compared to the LXCXE motif from viral oncoproteins. We discuss the implications of the weak binding and provide explanations for many observations seen in published work in this area (section 5.1.7.2).

2.4.3.3 BRG1 (ATPase of SWI/SNF remodeling complex)

BRG1 is one of the central ATPases of SWI/SNF chromatin remodeling complexes. BRG1, and its homologous protein BRM1 contain LXCXE motifs, and were shown to associate with pRb *in-vitro*, and antibodies raised against either protein have been shown to coprecipitate pRb from cell lysate (Dunaief et al., 1994). In addition, all three pocket-proteins have been reported to interact with both hBRM1 and BRG1 in yeast two-hybrid assays (Strober et al., 1996). Dunaief and colleagues (1994) found that expression of hBRG1 caused growth arrest in SW-13 cells, which was overcome by the coexpression of E1A. Mutants of hBRG1 that were defective for the pRb-interaction had greatly reduced growth arrest activity but it is notable that these mutants remove a substantial portion of the wild-type protein. These experiments suggest that pRb functionally synergizes with hBRM1 and hBRG1, potentially through a direct physical interaction. In agreement with this, hBRM1 and hBRG1 were found to cooperate with pRb to repress E2F1 and *c-fos* transcription, respectively (Trouche et al., 1997; Murphy et al., 1999). hBRM1/pRb/E2F1 appear to form complexes both *in-vitro* and *in-vivo* (Trouche et al., 1997). Recently, pRb/BRG1/HDAC1 trimolecular complexes have been found in transiently-transfected C33A cells, and it has been suggested that a specific subset of E2F-regulated genes might be regulated by pRb/BRG1 complexes (Zhang et al., 2000). Currently it is unclear how SWI/SNF complexes and pRb co-operate in transcriptional repression. Moreover, some doubts have been raised regarding the interaction of BRG1 with pRb via the LXCXE motif. One report suggested that BRG1 and HDAC1 interact separately with pRb via differing modes of interaction (Dahiya et al., 2000). Now, if both HDAC1 and BRG1 are proposed to bind via the LXCXE motifs in their sequences, these results seems to be contradictory. Moreover, recently, Kang et al. have shown that the direct physical interaction of BRG1 with the pRb is not required for the BRG1 ability to arrest growth and flat cell formation but rather it controls the activity of pRb via regulation of $p21^{CIP/WAF1/SDI}$, an upstream regulator of pRb (Kang et al., 2004). All these observations have made the BRG1 LXCXE-pRb interaction questionable.

2.3.3.4 Gankyrin

Gankyrin was initially purified and characterized by Tanaka and coworkers (Hori et al., 1998) as a component of the regulatory subunit of the 26S proteasome, which is an ATP-dependent protease responsible for the degradation of proteins. Since gankyrin has a LXCXE motif it was

proposed to interact with pRb via this motif (Higashitsuji et al., 2000). However, there are reports which suggest that probably gankyrin affects the pRb pathway indirectly. Firstly, gankyrin was found to be associated only with pRb and not with the other members of the family, p107 and p130. If gankyrin interact via the LXCXE motif in its sequence, it would be expected to interact with all the pocket proteins with conserved LXCXE binding motif. When we looked at the LXCXE motif in the structure of gankyrin, it is located in a α-helix, considering that the LXCXE motif from T antigen and E7 binds in an extended conformation, it is difficult for gankyrin to bind via its LXCXE motif (Krzywda et al., 2004). The possibility that gankyrin unfolds before binding to the pRb, is unlikely, but cannot be ruled out. Gankyrin facilitates the phosphorylation and degradation of pRb, suggesting that increased expression of gankyrin promotes tumorigenicity by targeting pRb to the proteasome. Also gankyrin disrupts pRb function by other means as well. Gankyrin also binds CDK4 resulting in a gankyrin-CDK4-cyclinD ternary complex (Li and Tsai, 2002). In so doing, gankyrin competes with INK4A, an inhibitor of cyclin kinases, for binding to CDK4. Based upon these findings, gankyrin appears to indirectly activate CDK4, resulting in the hyperphosphorylation of pRb and concomitant deregulation of E2F1-mediated transcription and cell cycle progression. Taken together, these studies suggest that gankyrin deactivates the pRb tumor suppressor pathway at multiple levels, including direct binding to pRb or by ensuring its inactivation through the maintenance of CDK4 kinase activity (Lozano and Zambetti, 2005).

2.3.3.5 Plasminogen activator inhibitor (PAI) 2

Plasminogen activator inhibitor 2 (PAI2) is well documented as an inhibitor of the extracellular serine proteinase urokinase-type plasminogen activator (uPA). It was shown that PAI2 has a novel intracellular function as a retinoblastoma protein binding protein. PAI-2 co localized with pRb in the nucleus and inhibited the turnover of pRb, which led to increases in pRb protein levels and pRb-mediated activities. Although PAI2 contains an LXCXE motif, pRb binding was primarily mediated by the C-D interhelical region of PAI2, which was found to bind to the C pocket of pRb. The C-D interhelical region of PAI2 was shown to contain a novel pRb-binding motif, termed the PENF homology motif, which is shared by many cellular and viral pRb-binding proteins. PAI2 expression also protected pRb from the accelerated degradation mediated by human papillomavirus (HPV) E7, leading to recovery of pRb and inhibition of E6/E7 mRNA expression (Darnell et al., 2003).

2.5 Chromatin remodeling

A single human genome is about 2 m long and must be compacted to fit into a nucleus with a diameter that is about 200,000-fold smaller. Eukaryotic cells have solved this packaging problem by folding their DNA into a highly compacted chromatin structure. The fundamental structure of chromatin is the nucleosome, which consists of 146 base pairs of DNA wrapped around an octamer of two molecules each of the histones H2A, H2B, H3 and H4. Adjacent nucleosomes are connected by linker DNA and progressive coiling of nucleosomes leads to compact, higher-order chromatin structures. DNA that is contained within highly compacted chromatin, termed heterochromatin, is inaccessible to the transcription machinery. Therefore, chromatin needs to be restructured, in order to make the DNA accessible for processes such as gene expression, DNA replication, repair or recombination. Appropriate regulation of gene expression requires interplay of complexes that re-model chromatin structure and thereby regulate the accessibility of individual genes to sequence-specific transcription factors and basal transcription machinery. Dynamic modulation of chromatin structure and topology is now appreciated as a key component of the regulated expression of individual genes. Chromatin remodeling is performed by multi-subunit complexes that can be grouped into two broad classes: (A) ATP-dependent complexes, which use the energy of ATP hydrolysis to locally disrupt or alter the association of histones with DNA, and (B) histone acetyltransferase (HAT) and histone deacetylase (HDAC) complexes, which regulate the transcriptional activity of genes by determining the level of acetylation of the amino terminal domains of nucleosomal histones associated with them (reviewed in Flaus and Owen-Hughes, 2004; Simone, 2005; Roberts and Orkin, 2004).

All ATP-dependent chromatin-remodeling factors identified so far are multisubunit complexes that contain an ATPase subunit, which belongs to the Swi2/Snf2 ATPase superfamily. There are four different classes of ATPase subunit — SWI/SNF, ISWI, CHD (Mi-2) and INO80 (Fig. 2.6) and based on these ATPase subunits the ATP-dependent chromatin-remodeling factors are subdivided into four classes (Tsukiyama, 2002; Vignali et al., 2000).

SWI/SNF class:

First identified in the yeast *Saccharomyces cerevisiae*, SWI/SNF is a 2 MDa multisubunit assembly that is highly conserved in eukaryotes. SWI/SNF remodeling complexes alter the path of DNA around the nucleosomal histone core in an ATP-dependent manner, resulting in nucleosome mobilization. SWI/SNF class will be discussed further in section 2.5.1.

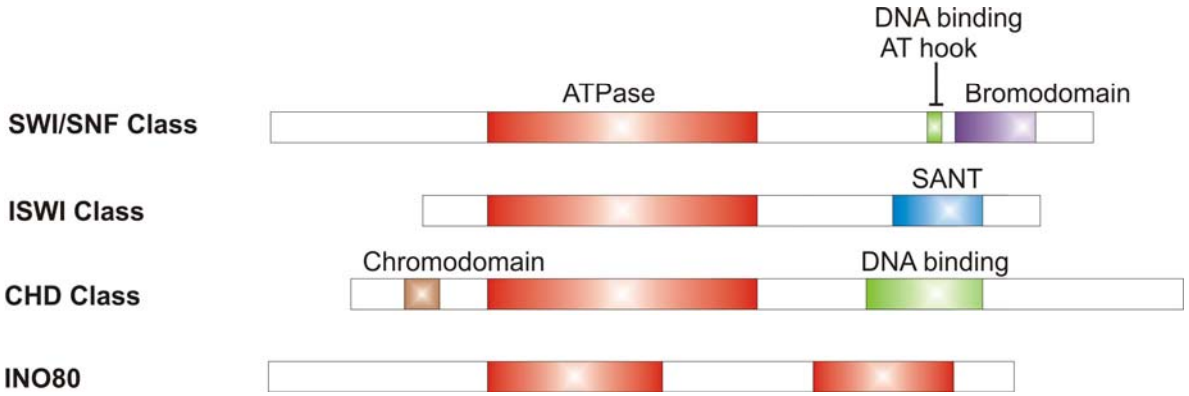

Figure 2.6. *Structure of four classes of ATPase in ATP-dependent chromatin-remodeling complexes. ATPase domains are shown in red. Signature motifs in each class-the bromodomain (violet), the SANT (SWI3, ADA2, NCoR, TFIIIB) domain (blue), the chromodomain (brown) and a putative DNA-binding domain (green) — are shown (adapted from Tsukiyama T, 2002).*

ISWI class

The founding member of the ISWI (imitation switch) class, *Drosophila* ISWI, was originally identified on the basis of its homology in its ATPase domain to that found in BRM1. Although *Drosophila* has only one member of this class, the ISWI protein forms at least three distinct complexes *in-vitro* — NURF (nucleosome remodeling factor), ACF (ATP-utilizing chromatin assembly and remodeling factor) and CHRAC (chromatin-accessibility complex). Each complex has a biochemical activity that alters the structure of chromatin.

CHD class

This class of factors is generally believed to repress transcription, partly because Mi-2 - a member of the CHD family in higher eukaryotes-forms large protein complexes that contain HDAC subunits *in-vivo*. The Mi-2 complexes in mammalian cells and *Xenopus laevis* egg also contain putative methylated-DNA-binding proteins. This indicates a function for the Mi-2 complex in transcriptional repression that is mediated by DNA methylation.

INO80 class

INO80 was originally found as a gene that is required for transcriptional activation of the *INO1* gene, a gene that is induced in the absence of inositol. INO80 has not been extensively studied yet, however, this protein is believed to represent a new class of ATPase, because it has a unique ATPase domain that is split into two subdomains. Purification from budding yeast yielded a 1.0 - 1.5 MDa complex with an ATP-dependent ability to alter chromatin structure *in-vitro*.

2.5.1 The SWI/SNF remodeling complex

The founding member of the SWI/SNF class of remodeling complexes is yeast SWI/SNF complex, which contains Swi2/Snf2 ATPase (Wang et al., 1996). Many subunits of the SWI/SNF complex were isolated in genetic screens that looked for factors that are required for mating-type switching, as well as for the use of sucrose as a carbon source. The isolated genes were therefore named *SWI*, for switching and *SNF*, for sucrose non-fermenting. Identification of the genes that are mutated in *SWI* and *SNF* mutants showed that many of them are the same genes, and this led to the conclusion that the same protein complex is responsible for both the *SWI* and *SNF* phenotypes (Stern et al. 1984; Neigeborn and Carlson, 1984). The *SWI* and *SNF* phenotypes are caused mainly by a lack of transcriptional activation of the homothallic switching (*HO*) endonuclease and sucrose hydrolyzing enzyme (*SUC2*) genes, respectively, and this shows that the SWI/SNF complex is required for the inducible transcription of genes. Consistent with this, SWI/SNF is necessary for the formation, *in-vivo* of an open chromatin structure at the promoter of the *SUC2* gene, where the DNA is accessible to structural probes such as endonucleases (Hirschhorn et al. 1992). The SWI/SNF group of remodelers can be subdivided further into two distinct, highly conserved subclasses. One subfamily comprises yeast SWI/SNF (ySWI/SNF), fly BAP and mammalian BAF complexes; whereas the second subfamily includes yeast RSC, fly PBAP and mammalian PBAF complexes. Mammalian SWI/SNF is a multiprotein chromatin-remodeling complex composed by at least 10 elements (Kadam et al., 2000). Two distinct SWI/SNF complexes were described, each characterized by the presence of a unique subunit: BAF250 or BAF180, defining BAF and PBAF, respectively (Table 2.2). In addition, BAF can contain either BRG1 or BRM1 as the core motor subunit, whereas PBAF only contains BRG1. All subunits of SWI/SNF are well conserved from yeast to humans, and structural analysis of their protein domains suggests specific functional properties (Fig. 2.7). The central core subunits BRG1 and BRM1 contain an ATPase domain and a bromodomain. Bromodomain is a recognition motif for acetylated lysines in histone tails or in other proteins. Further structural domains that are involved in protein-protein interaction are present in BAF250, BAF180, BAF170, BAF155, BAF60 (a, b, c1/2), BAF53 (a, b), BAF47, and G-actin, whereas BAF250, BAF180, BAF170, BAF155, and BAF57 contain different sequence-dependent and sequence-independent DNA binding domains (Simone, 2005; Klochendler-Yeivin et al., 2001).

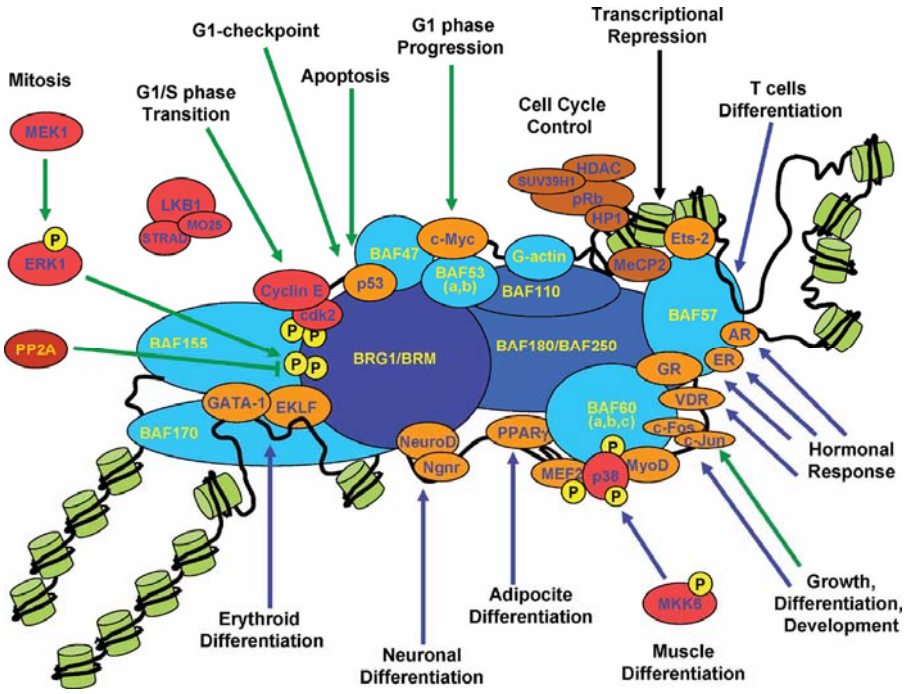

***Figure 2.7**: Mammalian SWI/SNF complex. A ~2 MDa complex contains BRG1 or BRM1 as the central ATPase and a number of other factors or proteins. Mammalian SWI/SNF complex is targeted by several signaling pathways. Adapted from Simone C, 2005.*

***Table 2.2.** Relationship between SWI/SNF-class of chromatin remodeling complexes from yeast, Drosophila and mammals. The conserved groups are grouped horizontally, with the ATPase subunit listed in the first row (adapted from Klochendler-Yeivin et al. 2001).*

Yeast		**Drosophila**		**Human**	
SWI/SNF	**RSC**	**BAP**	**PBAP**	**BAF**	**PBAF**
Swi2/Snf2	Sth1	Brahma	Brahma	BRG1 or BRM1	BRG1
		OSA		BAF250	
	Rsc1, Rsc2, Rsc4		Polybromo		Polybromo/BAF 180
	Rsc9		BAF170		
Swi3	Rsc8	Moria	Moria	BAF170 & BAF155	BAF170 & BAF155
		BAP111	BAP111	BAF57	BAF57
Swp73	Rsc6	BAP60	BAP60	BAF60a	BAF60a or BAF60b
Swp61/Arp7 Swp59	Rsc11/Arp7 Rsc12/Arp9	BAP55	BAP55	BAF53	BAF53
		actin	actin	actin	actin
Snf5	Sth1	Snr1	Snr1	hSNF5/INI1	hSNF5/INI1
	Rsc5, 7, 10, 13-15 Rsc3, Rsc30				
Swp82 Swp29/Tfg3/TAF30/Anc1 Snf6, 11					

2.5.2 BRG1 and BRM1 ATPases

A link between the chromatin structure and the capacity of the SWI/SNF complex to regulate transcription arose during screens for suppressors of SWI and SNF mutations. These mutations were found to exist within histones and other chromatin components and provided the first link between the SWI/SNF complex and chromatin. Since then, it has been shown that SWI/SNF is recruited to chromatin, hydrolyses ATP, and uses this energy to remodel nucleosomes. Several different assays have been used to detect this activity for example, monitoring for altered accesses for restriction enzymes and DNAse I, extrusion of DNA, and movement of nucleosomes along the DNA helix. Notably, although the complex normally consists of 9-12 subunits, only four of these-BRG1 or BRM1 (mutually exclusive ATPase subunits that are human homologues of the yeast Swi2/Snf2 ATPase subunit), SNF5 (also known as INI1), BAF155 and BAF170-are required *in-vitro* to remodel nucleosomes at a rate that is comparable to the entire complex (Phelan et al., 1999). The role of the other subunits remains ill defined, although it seems likely that they assist in directing the specificity of the complex through protein–protein interactions. Mammalian Swi2/Snf2-like ATPases - BRG1 and BRM1 form distinct complexes that have different biochemical activities to alter chromatin structure *in-vitro*, and mouse knockout studies show clear differences in their biological roles *in-vivo*. Mice that lack the *BRM1* gene develop normally; expression of BRG1 is significantly upregulated in these mice, which indicates a mechanism of functional compensation. Interestingly, *BRM1*-mutant mice are ~15% heavier than their wild-type littermates, and show increased cell proliferation-fibroblast cells that are isolated from the *BRM1*-mutant mice fail to arrest their cell cycles after cell confluence or DNA damage, which shows abnormal cell-cycle control (Reyes et al., 1998). It remains to be determined whether these phenotypes are due to the lack of BRM1, increased levels of BRG1, or both. The *BRG1* gene, by contrast, is essential for mouse development, and *BRG1*-mutant embryos die at the peri-implantaion stage suggesting a connection between the mammalian SWI/SNF complex and cell-cycle regulation (Bultman et al., 2000). BRG1 affect the cell cycle by interacting with pRb pathways directly or indirectly and was shown to physically associates with BRCA1 *in-vivo*. The BRG1 gene is mutated in various human tumor cell (Wong et al., 2000). The mammalian SWI/SNF-like complexes affect cell-cycle regulation and have roles in tissue differentiation. There are tissue- and cell-type-specific SWI/SNF-like complexes that regulate tissue-specific gene expression (Roberts and Orkin, 2004).

The human BRG1 protein is a 1647 amino acid long. There are defined domains and motif present in the protein, which are homologous to the domains found in many chromatin related proteins. Human BRG1 has an N-terminal proline rich domain, followed by a highly

charged domain (rich in Lys and Arg), central ATPase domain, followed by HMGA (high-mobility-group A) like KR/AT hook region and bromodomain (Fig. 2.8) (Khavari et al., 1993).

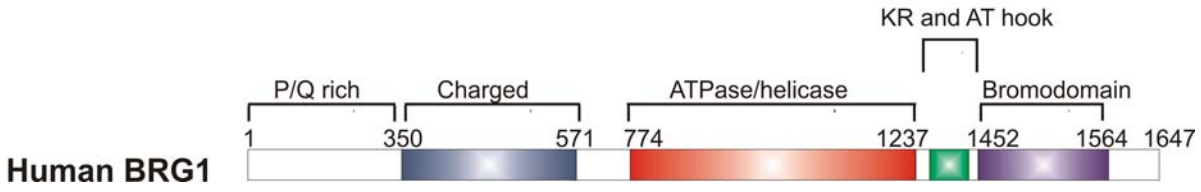

Figure 2.8. *Domain organization of human BRG1 protein.*

2.5.3 Functional domains or motif in hBRG1

2.5.3.1 ATPase or helicase

The central functional catalytic domain in the BRG1 and BRM1 proteins is the DNA dependent ATPase/helicase domain. This is the common domain found in all the ATP- dependent chromatin remodeling complexes. Khavari et al. cloned the cDNA of human BRG1 in 1993, and showed that has extensive homology to *SWI2* and *Drosophila* brahma proteins. The ATPase activity is essential for chromatin remodeling. Yeast and the *Drosophila* genomes encode single SWI2/SNF2 homologues that possess ATPase activity-Brahma (BRM1). Mammals, however, contain two such homologues-BRM1 and BRG1. The importance of the ATPase domain can be realized by the fact that a mutation in the ATP binding site of the ATPase/helicase-like domain form of BRG1 and BRM1 (ATPmut) strongly impairs the ability of the protein to revert the phenotype of ras-transformed fibroblasts (Khavari et al., 1993; Richmond and Peterson, 1996).

Recently, Dürr et al. (2005) solved the structure of the SWI2/SNF2 ATPase domain with the DNA substrate from *Sulfolobus solfataricus* bacteria (protein called SsoRad54cd). SsoRad54cd consists of two domains (domains 1 and 2). Domain 1 and 2 each contain a 'RecA' type α/β subdomain (denoted 1A and 2A, respectively), with a central β-sheet, flanked by α-helices. The seven classical sequence motifs that are implicated in ATP hydrolysis and/or DNA binding of helicases are located in loop regions of subdomains 1A (motifs I, Ia, II, III) and 2A (motifs IV, V, VI). The core RecA-like architecture of SsoRad54cd is related to that found in helicase structures (Singleton et al., 2001 and Subramanya et al., 1996), suggesting that SWI2/SNF2 enzymes and DExx box helicases share the basic ATP hydrolysis mechanism. The DNA duplex binds alongside the entrance of the active site cleft in a position where ATP-driven

changes between domain 1 and 2 could be directly linked to the translocase activity (Durr et al., 2005).

2.5.3.2 HMGA (high-mobility-group A) like DNA binding AT-hook motif

The SWI/SNF complex clearly binds DNA and nucleosomes with high affinity (Côté et al. 1994; Lorch et al., 1998; Quinn et al., 1996)**.** The ySWI/SNF complex is able to bind naked DNA in an ATP-independent manner with a K_D in the nanomolar range (Quinn et al., 1996). It is likely that this binding occurs through minor-groove interactions, since the complex can be displaced from DNA by distamycin A or chromomycin A3, two minor-groove-binding drugs (Quinn et al., 1996; Côté et al., 1998).

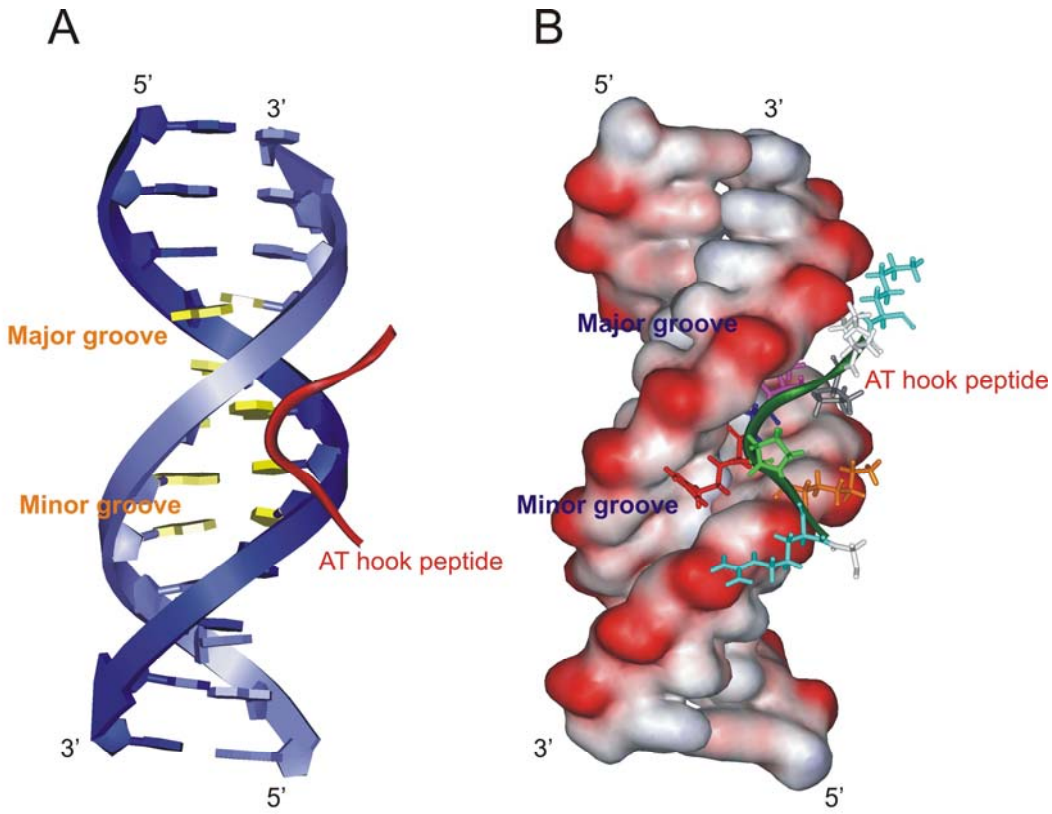

Figure 2.9*. Recognition of AT rich minor groove of B-DNA by AT-hook motif from HMGA1 protein (adapted from pdb file 2EZD). (A) The ladder form of DNA with the AT hook peptide. The AT base pairs in the minor groove are shown in yellow color. (B) The surface fill model of B-DNA with bound peptide.*

The DNA-binding properties of the ySWI/SNF complex are similar to those of HMG box proteins, which bind nonspecifically to DNA in a length-dependent manner with a preference for four-way junctions and cruciforms. It has been proposed that the complex might bind DNA via one of the two subunits that contain an HMG box. The first one is INI1/hSNF5L1, which prefers

to bind supercoiled DNA over relaxed circular DNA (Morozov et al., 1998**)**. The second subunit is BAF57, which can bind four-way junction DNA (Wang et al., 1998). Later, the BRM1 subunit of SWI/SNF complex was shown to have the DNA binding activity (Bourchot, 1999). A region in the C-terminus of the protein has an AT hook-like motif of HMGA1 proteins (Reeves, 2001). This region of the protein is rich in positively charged residues (Lys and Arg hence called KR region) and was shown to act as a nuclear localization signal (NLS) and harbor the DNA binding activity. AT hook motif was first identified in the HMGA1 protein and binds to the double stranded DNA in the AT rich sequences in the minor grove (Fig. 2.9). There is a similar homologous region in the C terminus of hBRG1 protein. Thus, it is possible that the SWI/SNF complex recognizes particular structural features in chromatin, such as crossovers in the linker DNA.

2.5.3.3 Bromodomain

The distinguishing feature of the SWI/SNF class of chromatin remodelers is the presence of a bromodomain in the ATPase subunit, which is absent in ISWI, CHD/Mi-2, and INO80 type of remodelers (Eberharter and Becker, 2004). Bromodomains are ~110 amino acid long modules first identified by sequence alignment in 1992 as a ~60 residue motif conserved among *Drosophila* Brahma and female-sterile homeotic (fsh) and four other potential transcription regulators (Haynes SR et al., 1992). Subsequently, this domain was found to be present in many transcription and chromatin regulators, suggesting its role in regulating chromatin structure and transcription. Bromodomains recognize acetylated lysines in histone tails and other proteins (Dhalluin et al., 1999; Hassan et al., 2002; Hudson et al., 2000; Jacobson et al., 2000; Marmorstein and Berger, 2001; Owen et al., 2000; Zeng and Zhou, 2002). This finding suggests a novel mechanism for regulating protein-protein interactions via lysine acetylation, which has broad implications for the mechanisms underlying a wide variety of cellular events, like transcriptional activation, and signal transduction apart from chromatin remodeling. In particular, the bromodomain of BRG1 has been implicated in translating the histone code. The histone code hypothesis states that different histone modifications such as acetylation, phosphorylation and methylation, are recognized in a specific manner by different effector proteins, thereby translating histone code into action (Strahl and Allis, 2000 and Kouzarides, 2002). For example, it has been shown that during activation of IFN-β gene only a subset of lysines in histone H4 and H3 are acetylated by GCN5 acetyltransferase. These acetylated lysines in histones H3 and H4 in turn recruit different promoter containing transcription complexes. Acetylation of histone H4 K8 mediates recruitments of the SWI/SNF complexes; whereas acetylation of K9 and K14 in histone

H3 is critical for the recruitment of TFIID. In general it was shown *in-vivo* that the bromodomain of BRG1 in SWI/SNF complex was capable of interacting with the K8 of histone H4 and K9 and K14 of histone H3 (Agalioti et al., 2002).

In terms of the structure, bromodomain is a helical domain (Dhalluin et al., 1999; Hudson et al., 2000; Jacobson et al., 2000; Owen et al., 2000). The three-dimensional structure of a prototypical bromodomain from the transcriptional co-activator p300/CBP-associated factor (P/CAF) determined by using NMR spectroscopy showed that the bromodomain adopts an atypical left-handed four-helix bundle (helices αZ, αA, αB and αC). A long intervening loop between helices αZ and αA (termed the ZA loop) is packed against the loop connecting helices αB and αC (named the BC loop) to form a surface-accessible hydrophobic pocket, which is located at one end of the four-helix bundle, opposite the amino and carboxy termini of the protein. This unique structural fold is highly conserved in the bromodomain family, as supported by several more structures of bromodomains from human GCN5 and *S. cerevisiae* GCN5p as well as the double bromodomain module of human $TAF_{II}250$ (Hudson et al., 2000; Jacobson et al., 2000 and Owen et al., 2000). The structural similarity amongst these bromodomains is very high for the four helices with pairwise root-mean-square deviations of 0.7–1.8 Å for the backbone C_α atoms. The majority of structural deviations are localized in the loop regions, particularly in the ZA and BC loops. This observation is in an agreement with the relatively high sequence variations in these loops. The modular structure supports the notion that bromodomains act as a functional unit for protein interactions, and multiple bromodomain modules can be placed sequentially in a protein to serve similar or distinct functions (Jeanmougin et al. 1997; Dhalluin et al., 1999).

Since residues in bromodomain, important for acetyl-lysine recognition are largely conserved, binding of lysine-acetylated proteins is likely a general biochemical function for bromodomains. However, differences in ligand selectivity may be attributed to a few but important differences in bromodomain sequences, which include; (i) variations in the ZA loops, which have relatively low sequence conservation and amino acid deletion or insertion in different bromodomains, and (ii) differences in bromodomain residues that directly interact with residues surrounding acetyl-lysine in a target protein. It is now an accepted theme that bromodomains, similar to Src homology 2 (SH2) and phosphotyrosine binding (PTB) domains (which specifically recognize tyrosine-phosphorylated proteins in signal transduction) (Shoelson, 1997; Kuriyan and Cowburn, 1993; Zhou and Fesik, 1993; Forman-Kay and Pawson, 1993) can also bind with high selectivity to lysine-acetylated proteins through specific interactions with amino acid residues flanking the acetyl-lysine (Owen et al., 2000; Mujtaba et al. 2004). New knowledge

of ligand specificity will facilitate our understanding of molecular mechanisms underlying specific biological functions of bromodomains in cellular processes including chromatin remodeling and transcriptional activation. These structural studies further emphasize the concept that functional diversity of a conserved protein modular structure is achieved by evolutionary changes of amino acid sequences in the ligand binding site. Tracking such an evolutionary trail on the functional divergence of bromodomains (particularly involving the ZA loop) will require detailed knowledge of protein structure–function relationships, which can be only acquired from three-dimensional structures of new bromodomains and in particular bromodomain/biological-ligand complexes.

2.6 SWI/SNF complexes and cancer

It has been established that inactivation of SWI/SNF can play a critical causal role in the development of human cancers. A number of subunits of the SWI/SNF complex are found mutated in many human cancers. Most compellingly, the human *SNF5* gene (SNF5 is a universal SWI/SNF component (Table 2.2) was found to be deleted or mutated in atypical teratoid and malignant rhabdoid tumors (ATRTs and MRTs) (Roberts et al., 2000), very aggressive cancers of early childhood (reviewed in Klochendler-Yeivin, 2002; Roberts and Orkin, 2004 and Mohrmann and Verrijzer, 2005). Human *SNF5* mutations were also found associated with some other neoplasms, including chronic myeloid leukemia, choroids plexus carcinoma, medulloblastoma and central primitive neurectodermal tumors (Klochendler-Yeivin, 2000). SNF5 is a direct target for several distinct DNA-binding regulators, stimulates the *in-vitro* remodeling activity of BRG1, and is required during post-recruitment remodeling *in-vivo*. When hSNF5 is re-expressed in the MRT cells it causes an accumulation of cells in G0/G1, cellular senescence and apoptosis through direct transcriptional activation of the tumor suppressor $p16^{INK4a}$. Strikingly, it was shown that hSNF5 is critical for the recruitment of BRG1 to the p16INK4a promoter and transcriptional activation (Oruetxebarria et al., 2004). It has been established that the induction of cellular senescence by hSNF5 was strictly dependent upon the $p16^{INK4a}$/pRb tumor suppressor pathway.

BRG1 was shown to be mutated in multiple human cell lines and in some primary tumors. BRG1 homozygous knock-outs mice die during peri-implantation stages, showing that it is essentials for very early developments (Wong et al., 2000; Bultman et. al., 2000). Similar to BRG1 homozygous knock-outs, mice deficient for *SNF5* die during peri-implantation. In non-small-cell-lung cancers, the loss of BRG1 expression correlates with a poorer prognosis. BAF and PBAF are implicated further in tumorgenesis by its association with proteins with a well-

established role in human cancers, including pRb, BRCA1, MLL, and h-catenin. However SNF5 and BRG1 are not only anti-proliferation factor. Some cells require BRG1 function for cell growth or differentiation rather than cell cycle arrest. Thus, the SWI/SNF complexes act in a cell-type specific manner, whereas some cells cannot survive without its activity, in others it activates the senescence program. In conclusion, inactivation of distinct SWI/SNF subunits results in very different phenotypes. Loss of mammalian BRM1 function does not affect viability or induce cancer, but these mice are bigger than their wild-type littermates. However, inactivation of either BRG1 or SNF5 gives rise to different types of neoplasias. Furthermore, many normal cell types require BRG1 or SNF5 for viability under physiological conditions. Taken together, these results re-enforce the notion that the different subunits in SWI/SNF complexes can contribute unique functional properties (Klochendler-Yeivin, 2002; Roberts and Orkin, 2004; Muchardt and Yaniv, 1999).

3 Goals of the study

The objective of this thesis work was to carry out the functional investigations of two proteins, the retinoblastoma protein (pRb) and human BRG1. Since these proteins play central role in cell cycle and chromatin remodeling processes respectively, the structural and functional details on these protein would be valuable in understanding these processes at the molecular level.

pRb coordinate a number of processes in the cell including cell cycle control, transcriptional repression, differentiation and apoptosis to name a few. One of the ways pRb could control these many processes is by interacting with other proteins. Indeed more than 130 proteins have been reported to associate with pRb. However, the data is ambiguous and is not quantitative. We sought to investigate the interaction of pRb with several targets using purified proteins and biochemical and biophysical methods. The target proteins included MyoD, Id-2, HPV E7, SV40 large T antigen, E2F, HDAC1, PAI2, gankyrin and BRG1. The objective was to check for the binding of these proteins with pRb and define them in terms of K_D values.

Human BRG1 consists of the ATPase/helicase domain, an AT hook motif and a bromodomain. The ATPase/helicase domain provides the catalytic activity, the AT hook motif is the DNA recognizing and bromodomain is acetyl lysine binding domain implicated in protein-protein interactions. Biophysical characterization of interactions between the various domains and targets would help in understanding how this protein acts. Characterization of DNA binding properties of the AT hook motif would allow us to understand the interaction of BRG1 and nucleosome. The structure of the bromodomain would help in elucidating the reasons for specificity of various bromodomains with different acetylated lysine residues in histone and other proteins. Also the comparison of bromodomain structures from different proteins would allow us to find the reason for the evolutionary adaptation of these domains. These finding would ultimately enhance our understanding about role of hBRG1 protein in chromatin remodeling.

4 Materials and laboratory methods

4.1 Materials

In this section of the thesis all the materials used in this work are listed together with the names corresponding producer companies. In cases of the most common chemicals, it should be assumed that they were purchased from Sigma or Merck Biosciences unless noted. All crystallization screen and chemicals were purchased from the Hampton Research.

4.1.1 *E. coli* strains and plasmids

Different *E. coli* strains and vectors used in this thesis are listed in Table 4.1 and Table 4.2, respectively.

Table 4.1. *List and properties of E. coli strains used in this work*

Strain	Genotype	Aim	Provider
One Shot TOP10	F¯ *mcr*A Δ(*mrr*¯*hsd*RMS¯*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Cloning	Invitrogen (Holland)
Novablue	*endA1 hsdR17*($r_{K12}^-m_{K12}^+$) *supE44 thi-1 recA1 gyrA96 relA1 lac*F'[*proA*$^+$*B*$^+$ *lacI qZΔM15*::Tn*10*] (TetR)	Cloning	Novagen (Canada)
BL21 Star	F¯ *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm*	Expression	Novagen (Canada)
BL21 Star (DE3)	F¯ *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Expression	Novagen (Canada)
BL21 Star (DE3) pLysS	F¯ *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)pLysS (CamR)	Expression	Novagen (Canada)
B834 (DE3)	B834(DE3) F¯ *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm met* (DE3)	Expression	Novagen (Canada)

Table 4.2*. List and properties of vectors used in this study*

Vector	Antibiotic^R	N-Terminal purification tag	Protease to cut N terminal tag	Provider
pET30 LIC/Xa	Kanamycin	6X His	Factor Xa	Novagen (Canada)
pET41 LIC/EK	Ampicillin	GST	Enterokinase	Novagen (Canada)
pET46 LIC/EK	Ampicillin	6X His	Enterokinase	Novagen (Canada)
pGEX 4T2	Ampicillin	GST	Thrombin	Amersham-Pharmacia (Sweden)

4.1.2 Cell growth media and stocks solutions

Luria Bertani (LB)

For 1 liter LB medium:
- 10 g bacto tryptone
- 5 g bacto yeast extract
- 10 g sodium chloride

pH was adjusted to 7.0 with NaOH. For the preparation of agar plates the medium was supplemented with 15 g agar.

Minimal medium (MM) for uniform enrichment with ^{15}N:

For 1 liter MM:
- 0.5 g NaCl
- 1.3 ml trace elements solution
- 1 g citric acid monohydrate
- 36 mg ferrous citrate
- 4.02 g KH_2PO_4
- 7.82 g K_2HPO_4 x $3H_2O$
- 1 ml Zn-EDTA solution
- 1 g NH_4Cl or $^{15}NH_4Cl$

pH was adjusted to 7.0 with NaOH, the mixture was autoclaved, upon cooling separately sterilized solutions were added: 25 ml glucose, 560 µl thiamin, antibiotics, 2 ml $MgSO_4$ stock.

Defined medium for selective ^{15}N labeling of proteins (also for Sel-Met labeling)

For 1 litre of medium:
- 400 mg Ala, Gln, Glu, Arg, Gly
- 255 mg Asp, Met
- 125 mg cytosine, guanosine, uracil
- 100 mg Asn, Leu, His, Lys, Pro, Thr,
- 100 mg Try
- 400 mg Ile, Val
- 50 mg Phe, thymine, thymidine
- 1.6 g Ser
- 10 mg $CaCl_2$
- 2 g NaAc
- 10 g K_2HPO_4
- 1 g citric acid
- 1.3 ml trace element solution

36 mg ferrous citrate
1 ml Zn-EDTA solution
1 g NH_4Cl

pH was adjusted to 7.0 with NaOH, the mixture was autoclaved. To the cooled medium, separately sterilized solutions were added: 25 ml glucose, 560 µl thiamin, antibiotics, 2 ml 1 M $MgSO_4$, sterile filtered and:

50 mg Cys, Trp, nicotinic acid
0.1 mg biotin
X mg ^{15}N-amino acid

Another portion of the ^{15}N-amino acid is added at the time of induction as well (same amount as added before, 0.22 µm filtered).

Stock solutions

Glucose: 20% (w/v) in deionised H_2O, autoclaved.

Ampicillin: 100 mg/ml of ampicillin in deionised H_2O, sterilized by filtration, stored in aliquots at -20°C until used. Working concentration: 200 µg/ml.

Chloramphenicol: 50 mg/ml of chloramphenicol in ethanol, stored in aliquots at -20°C until used. Working concentration: 50 µg/ml.

Kanamycin: 100 mg/ml of kanamycin in deionised H_2O, sterile filtrated and stored in aliquots at -20 °C until used. Working concentration: 100 µg/ml.

IPTG: A sterile filtered 1 M stock of IPTG in distilled water was prepared and stored in aliquots at -20 °C until used.

Thiamin, 1%, in deionised H_2O, sterilized by 0.22 µ filtration.

$MgSO_4$, 1 M, in deionised H_2O, sterilized by 0.22 µ filtration.

Zn-EDTA solution:
5 mg/ml EDTA
8.4 mg/ml $Zn(Ac)_2$

Trace elements solution:
2.5 g/l H_3BO_3
2.0 g/l $CoCl_2 \times H_2O$
1.13 g/l $CuCl_2 \times H_2O$
9.8 g/l $MnCl_2 \times 2H_2O$
2.0 g/l $Na_2MoO_4 \times 2H_2O$
pH lowered with citric acid or HCl.

4.1.3 Solutions for making chemically competent *E. coli* cells

Solutions for making competent cells

$CaCl_2.2H_2O$	1 M Filter sterilized (0.22 μ)
$MgCl_2$-$CaCl_2$ solution	80 mM $MgCl_2$ and 20 mM $CaCl_2$ Filter sterilized (0.22 μ)

4.1.4 Buffer for DNA agarose gel electrophoresis

50X TAE buffer (for 1 L)

40 mM Tris-acetate	242 g of Tris base
1 mM EDTA	100 ml of 0.5 M EDTA (pH 8.0)
Glacial acetic acid	57.1 ml

4.1.5 Protein purification buffers

Ion exchange and gel filtration chromatography buffers

Buffer P(0)	50 mM Na_2HPO_4 0.05% NaN_3 10 mM β-mercaptoethanol pH 7.4
Buffer P(1000)	50 mM Na_2HPO_4 1 M NaCl 10 mM β-mercaptoethanol 0.05% NaN_3 pH 7.4
PBS	140 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4 0.05% NaN_3 pH 7.3

Buffers for Immobilized metal-chelate chromatography (IMAC) under native conditions

Binding buffer (buffer N0)	50 mM NaH_2PO_4

	300 mM NaCl
	10 mM β-mercaptoethanol
	pH 8.0
Binding buffer (buffer N10)	50 mM NaH_2PO_4
	300 mM NaCl
	10 mM imidazole
	10 mM β-mercaptoethanol
	pH 8.0
Wash buffer (buffer N20)	50 mM NaH_2PO_4
	300 mM NaCl
	20 mM imidazole
	10 mM β-mercaptoethanol
	pH 8.0
Elution buffer (buffer N250)	50 mM NaH_2PO_4
	300 mM NaCl
	250 mM imidazole
	10 mM β-mercaptoethanol
	pH 8.0
Buffers for IMAC under denaturing conditions	
Buffer A	6 M guanidinium chloride
	100 mM NaH_2PO_4
	10 mM Tris
	10 mM β-mercaptoethanol
	pH 8.0
Buffer B	6 M guanidinium chloride
	100 mM NaH_2PO_4
	10 mM Tris
	10 mM β-mercaptoethanol
	pH 6.5
Buffer C	6 M guanidinium chloride

	100 mM Na acetate
	10 mM β-mercaptoethanol
	pH 4.0

Protease buffers

Buffer X (Factor Xa cleavage buffer)	100 mM NaCl
	2 mM $CaCl_2$
	20 mM Tris
	0.01% NaN_3
	pH 7.5
Buffer EK (Enterokinase cleavage buffer)	100 mM NaCl
	2 mM $CaCl_2$
	20 mM Tris
	0.01% NaN_3
	pH 7.5

4.1.6 Buffers for crystallization, NMR, ITC and CD spectroscopy

pRb-AB crysatallization buffer	20 mM Tris
	150 mM NaCl
	5 mM DTT
	0.05% NaN_3
	pH 7.5
Exbromodomain crystallization buffer	20 mM HEPES
	200 mM NaCl
	0.05% NaN_3
	2 mM DTT
	pH 7.4
NMR and ITC buffer for pRb-AB	50 mM Na_2HPO_4
	10 mM 2-mercaptoethanol
	0.05% NaN_3
	pH 7.4

NMR buffer for AT hook and Bromodomain	50 mM Na_2HPO_4 100 mM NaCl 0.05% NaN_3 pH 6.5
ITC buffer for AT hook and KR BRG1 domains	10 mM KH_2PO_4 100 mM KCl 0.05% NaN_3 pH 7.0

4.1.7 Reagents and buffers for SDS-PAGE and western blots

Buffers for the SDS-PAGE

Trcine SDS-PAGE

Anode buffer (+):	200 mM Tris pH 8.9
Cathode buffer (-):	100 mM Tris pH 8.25 100 mM tricine 0.1% SDS
Separation buffer:	1 M Tris pH 8.8 0.3% SDS
Stacking buffer:	1 M Tris pH 6.8 0.3% SDS

Pouring polyacrylamide gels

Separation gel	1.675 ml H_2O 2.5 ml separation buffer 2.5 ml 30 % acrylamide mix 0.8 ml glycerol 25 µl APS 2.5 µl TEMED

Intermediate gel	1.725 ml H_2O
	1.25 ml separation buffer
	0.75 ml 30 % acrylamide mix
	12.5 µl APS
	1.25 µl TEMED
Stacking gel	2.575 ml H_2O
	0.475 ml stacking buffer
	0.625 ml 30 % acrylamide mix
	12.5 µl 0.5 M EDTA, pH 8.0
	37.5 µl APS
	1.9 µl TEMED

Glycine SDS-PAGE

5X SDS tris glycine running buffer (for 1 L)

25 mM Tris-Cl	15.1 g of Tris base
250 mM glycine	94 g of glycine
0.1% SDS	50 ml of 10% SDS

Pouring polyacrylamide gels

Separating Gel 10 ml	(12%)	(15%)	
	3.3 ml	2.3 ml	H_2O
	4.0 ml	5.0 ml	30% acryalamide mix
	2.5 ml	2.5 ml	1.5 M Tris (pH 8.8)
	0.1 ml	0.1 ml	10% SDS
	0.1 ml	0.1 ml	10% APS
	10.0 µl	12.0 µl	TEMED

Stacking gel 5 ml	(5%)	
	3.40 ml	H_2O
	0.83 ml	30% acryalamide mix
	0.63 ml	0.5 M Tris (pH 6.5)
	0.05 ml	10% SDS
	0.05 ml	10% APS
	10.0 µl	TEMED

Protein visualization

Coomassie-blue solution:	0.25% W/V 45% ethanol 10% acetic acid
Destaining solution:	5% ethanol 10% acetic acid

Reagents for the Western blots

Transfer buffer	25 mM Tris pH 8.3 192 mM glycine To make the final working solution mix 80 ml of buffer with 20ml of Methanol
Alkaline phosphatase buffer	100 mM Tris pH 9.5 100 mM NaCl 5 mM $MgCl_2$
Wash buffer	10 mM Tris pH 8.0 150 mM NaCl 0.05% Tween20
1st antibody solution	1:1000 diluted in the alkaline phosphatase buffer with 5% non-fat milk
2nd antibody solution (linked to alkaline phosphatase)	1:2000 diluted in alkaline phosphatase buffer 5% milk in alkaline phosphatase buffer
Substrate for alkaline phosphatase	BCIP (Sigma) Dissolve 1 tablet in 10 ml of water (as directed by manufacturer)

4.1.8 Enzymes and other proteins

BSA	New England BioLabs (USA)

CIP	New England BioLabs (USA)
BamH I	New England BioLabs (USA)
Dpn I	Stratagene (USA)
EcoR I	New England BioLabs (USA)
Hind III	New England BioLabs (USA)
Nde I	New England BioLabs (USA)
Xho I	New England BioLabs (USA)
Vent^{R} polymerase	New England BioLabs (USA)
Pfu turbo DNA Polymerase	Stratagene (USA)
Thermopol	Qiagen (Germany)
Factor Xa	Novagen (Canada)
Enterokinase	Novagen (Canada)
Thrombin	Sigma (USA)
TAGzyme	Qiagen (Germany)
Anti His antibodies (mouse)	Santa Cruz biotech (USA)
Anti pRb mouse antibodies	Santa Cruz biotech (USA)
Goat anti mouse antibodies	Santa Cruz biotech (USA)

4.1.9 Kits and reagents

QIAquick PCR Purification Kit	Qiagen (Germany)
QIAprep Spin Miniprep Kit	Qiagen (Germany)
QIAGEN Plasmid Midi Kit	Qiagen (Germany)
QIAGEN Plasmid Maxi Kit	Qiagen (Germany)
QuikChange Site-Directed Mutagenesis Kit	Stratagene (USA)
Pre-Crystallization Test (PCT)	Hampton Research (USA)
Quick Ligation Kit	Qiagen (Germany)
Complete Protease Inhibitor Cocktail	Roche (Germany)
pET LIC cloning Kits	Novagen (Canada)

4.1.10 Protein and nucleic acids markers

Prestained Protein Marker	New England BioLabs (USA)
100 BP DNA marker	New England BioLabs (USA)
1Kb DNA marker	New England BioLabs (USA)
Broad Range (6-175 kDa) 1 kb DNA-Leiter	Peqlab (Germany)

4.1.11 Chromatography equipment, columns and media

ÄKTA explorer 10	Amersham Pharmacia (Sweden)
Peristaltic pump P-1	Amersham Pharmacia (Sweden)
Fraction collector RediFrac	Amersham Pharmacia (Sweden)
Recorder REC-1	Amersham Pharmacia (Sweden)
UV flow through detector UV-1	Amersham Pharmacia (Sweden)
BioLogic LP System	Biorad (USA)
HiLoad 26/60 Superdex S75pg	Amersham Pharmacia (Sweden)
HiLoad 10/30 Superdex S75pg	Amersham Pharmacia (Sweden)
Mono Q HR 5/5, 10/10	Amersham Pharmacia (Sweden)
Mono S HR 5/5, 10/10	Amersham Pharmacia (Sweden)
Ni-NTA-agarose	Qiagen (Germany)
SP sepharose	Amersham Pharmacia (Sweden)
GST Sepharose FF	Amersham Pharmacia (Sweden)

4.2 Laboratory methods and principles

The most common protocols for procedure or formulation of a solution is taken directly or modified from the 'Molecular Cloning' book by Sambrook and Russell (2001).

4.2.1 Construct design

Optimizing proper protein constructs is very important for NMR and X-ray crystallography. Unstructured portions of a protein or loop regions give rise to low signal dispersion of NMR signals of NMR spectra and are detrimental for crystallization. The idea behind designing protein domain constructs is to have as defined a domain as possible, and at the same time to have it functional. Leads for rational behind designing constructs mainly come from previously published literature, characterisation of constructs in lab using various techniques and in some case predicting the secondary structure with the help of bioinformatics tools.

The retinoblastoma protein (pRb) constructs were designed mainly on the basis of the previously published literature. We also employed NMR for defining the boundaries of my constructs. For example, the pRb-ABC construct had a NMR spectrum of partially folded protein, therefore, I deleted the C terminus part and the linker region between the A and B domain to finally get the folded pRb-AB construct. The main construct of pRb used in this thesis is pRb-AB, which comprises the 'small pocket' of pRb. Constructs of BRG1 were also designed mainly

on the basis of previously published reports or in some case guidance from the predicted secondary structure was taken into consideration.

In all the cases, final functionality of the protein construct was checked by functional assays, for example binding of a protein construct with its ligand (protein, peptide, DNA or chemical compound). Folding of a protein was checked by 1D-^{1}H NMR spectroscopy or CD spectroscopy.

4.2.2 Choice of expressions system

The objective of heterologous expression of a protein is to express it in high quantity. A large number of prokaryotic expression vectors currently available, possessing various restriction sites and encoding tags facilitating fusion protein purification (Makrides, 1996), makes a choice of a suitable vector an easy task. We frequently used pET or pGEX series of vectors for the expression of proteins in *E. coli* (pET system manual, Novagen, 2003; Amersham Pharmacia, 2003).

Removal of the fusion tag is often required for functional studies of a recombinant protein for example, crystallization. This is usually achieved with an aid of a specific restriction protease. The enzymes most commonly used are: Factor Xa (cleaves after R in an IEGRX sequence, X≠P), thrombin (cleaves after R in a LVPRGS sequence), and enterokinase (cleaves after K in a DDDDK sequence). Vectors with specific protease sites are commercially available otherwise they could be inserted using PCR. There are also available the exopeptidases, like dipeptidase (TAGzyme, Qiagen), which cleave dipeptides from the N-terminus of a fusion protein, and stops before Arg, Lys or 1 or 2 residues before Pro. The enzyme is active in the presence of PMSF and EDTA, making it ideal for work with degradation susceptible proteins. However, this enzyme is used for removal of only short N-terminal His-tags.

In choosing the restriction protease, enzyme – target protein compatibility has to be considered: the optimal cleavage conditions for the enzyme may not be appropriate for the stability of the protein in question. For instance, Factor Xa requires the presence of 1-4 mM Ca^{2+} in the buffer and oxidizing conditions, and will not work in the presence of chelating agents, phosphate buffer, reducing agents or high salt. Presence of even small amounts of detergents may hamper its selectivity. PreScission protease on the other hand requires strong reducing conditions.

I cloned most of the protein constructs used in thesis in pET30 LIC vectors with the 6xHis or GST purification tag. All the constructs have either factor Xa or thrombin cleavage site however, for MyoD constructs TAGZyme was also employed to remove an N-terminal His-tag.

4.2.3 DNA techniques

4.2.3.1 Preparation of plasmid DNA

The isolation of plasmid DNA from *E. coli* was carried out using dedicated plasmid purification kits from Qiagen. The kits employ a standard alkaline lysis of the precipitated bacteria in the presence of RNAse and strong ionic detergent, SDS, followed by neutralization/DNA renaturation with acetate. 10 ml of LB media with appropriate antibiotic was inoculated with a single colony and grown overnight at 37° C. For purification, a crude cell lysate is loaded onto a silica gel column, washed with an ethanol-containing buffer, and eluted in a small volume, yielding up to 20 μg of the plasmid DNA. This modified protocol gave more DNA and pure enough for further transformation and sequencing purposes.

4.2.3.2 PCR

A polymerase chain reaction was employed to amplify desired DNA fragments and genes, introduce restriction sites, and STOP codons and sequences encoding restriction protease cleavage sites. The primers were prepared according to standardized principles regarding the length, GC-content, melting temperature and occurrence of secondary structures of the hairpin type. All primers used for cloning and mutagenesis are listed in Table 4.3. Three different kinds of recombinant thermostable DNA polymerases were used, each operating at slightly different conditions:

	Melting temp.	Annealing temp.	Synthesis temp.
Themropol	94 °C	55 °C	72 °C
Pfu Turbo	95 °C	55 °C	68 °C
Vent^{R}	95 °C	55 °C	72 °C

Table 4.3. *List of primers used for cloning and mutagenesis in this work*

S.No	***Name***	***Sequence (5'-3')***
For pRb constructs		
1	nabc_forward	**GGTATTGAGGGTCGC**CAGAACAGGAGTGCACGGATAGCA
2	abc_forward380	**GGTATTGAGGGTCGC**ATGAACACTATCCAACAATTAATG
3	abc_reverse935	**AGAGGAGAGTTAGAGCC**TTAAAGGTCCTGAGATCCTCAT
4	abc_shortHis_sense	GCACCATCATCATCATCATATTGAGGGTCGCATGAACACTATCC
5	abc_shortHis_antisense	GGATAGTGTTCATGCGACCCTCAATATGATGATGATGATGGTGC
6	abc-c_sense	CTCACATTCCTCGAAGCCCTTACAAGTGAGGATCTCAGGACCTTTAAGGCTCT
7	abc-c_antisense	AGAGCCTAAAAGGTCCTGAGATCCTCACTTGTAAGGGCTTCGAGGAATGTGAG
8	linkerDel_sense	GATCTTATTAAACAATCAAAGGACTTGAAATCTACCTCTCTTTCACTG
9	linkerDel_antisense	CAGTGAAAGAGAGGTAGATTTCAAGTCCTTTGATTGTTTAATAAGATC
10	abc64_reverse	**AGAGGAGAGTTAGAGCCT**TATTCAGCACTTCTTTTGAGCACACG

S.No	***Name***	***Sequence (5'-3')***
11	n_forward	**GGTATTGAGGGTCGC**CAGAACAGGAGTGCACGGATAGCAAAACAAC
12	n_reverse	**AGAGGAGAGTTAGAGCCTTA**AACAGTCCTAACTGGAGTGTGTGGAG
13	na_reverse	**AGAGGAGAGTTAGAGCCTTA**GTCCTTTGATTGTTTAATAAGATC
14	Δ480_sense	GCAAACTTCTGAATGACATTTTTCATATGTCTTTATTGGCG
15	Δ480_antisense	CGCCAATAAAGACATATGAAAAATGTCATTCAGAAGTTTGC
16	S567A_sense	CTTGCATGGCTCTCAGAT***gCA***CCTTTATTTGATC
17	S567A_antisense	GATCAAATAAAGG***TGc***ATCTGAGAGCCATGCAAG
18	R661W_sense	GTATCGGCTAGCCTATCTC***tGG***CTAAATACACTTTGTGAAC
19	R661W_antisense	GTTCACAAAGTGTATTTAG***CCa***GAGATAGGCTAGCCGATAC
20	C712R_sense	GATGTGTTCCATGTATGGCATA***cGC***AAAGTGAAGAATATAGACC
21	C712R_antisense	GGTCTATATTCTTCACTTT***GCg***TATGCCATACATGGAACACATC
For BRG1 constructs		
22	IPPP_forward	**GGTATTGAGGGTCGC**ATCCCCACCCAGGGGCCTGGA
23	IPPP_reverse	**AGAGGAGAGTTAGAGCCTTA**AGGAGGAGGCTGGCCTGGCA
24	MV_forward	GGTATTGAGGGTCGCATGGTGCCACTGCACCAGAAGCAGA
25	IG_reverse	AGAGGAGAGTTAGAGCCTTACCCAATGGCAGGCGTCTGTCCT
26	LNTV_forward	**GGTATTGAGGGTCGC**CTGAACGGCATCCTGGCCGAC
27	LNTV_reverse	**AGAGGAGAGTTAGAGCCTTA**GACGGTGCAGAGGCGGAGCAC
28	LNQV_reverse	**AGAGGAGAGTTAGAGCCTTA**CACCTGCGTCAGCTTGCAGTG
29	PTST_forward	**GACGACGACAAGAT**GCCCACCATCTTCAAGAGCTGC
30	PTST_reverse	**GAGGAGAAGCCCGGTTTA**CGTGCTGCAGTGTCTGCT
31	LQKE_forward	**GGTATTGAGGGTCGC**CTGCAGTCCTACTATGCCGTG
32	KE_reverse	**AGAGGAGAGTTAGAGCCTTA**CTCCTTGCGGTGGCGGGA
33	NVKE_forward	**GGTATTGAGGTTCGC**AACGTGGACCAGAAGGTGATC
34	DLNA_forward	**GGTATTGAGGGTCGC**GATCTGTTCATGCGCATGGACCTGGAC
35	DLNA_reverse	**AGAGGAGAGTTAGAGCCTTA**TGCGTTCTGGCACAGGAGCATGACGT
36	LPES_forward	**GACGACGACAAGAT**GCTCCCCTCGTGGATCATCAAGGAC
37	KR domain_forward	**GACGACGACAAGAT**GGAGGTCCGGCAGAAGAAATCATCAC
38	AThook_forward	**GACGACGACAAGAT**GAAGGACGACGAGAGCAAGAAGCAG
39	AThook_reverse	**GAGGAGAAGCCCGGT**TTACTCCTTCTCGATTTTCTGCCGCAC
40	AThook-his_sense	ACTTTAAGAAGGAGATATACCATGAAGGACGACGAGAGCAAGAAG
41	AThook-his_antisense	CTTCTTGCTCTCGTCGTCCTTCATGGTATATCTCCTTCTTAAAGT
42	Bromo_forward	**GGTATTGAGGGTCGC**TCCCCTAACCCACCCAACCTC
43	Bromo_reverse	**AGAGGAGAGTTAGAGCCTTA**CTCGATTTTCTGGGCCACGCTG
44	Bromo-his_sense	CTTTAAGAAGGAGATATACATATGTCCCCTAACCCACCCAACCTC
45	Bromo-his_antisense	GAGGTTGGGTGGGTTAGGGGACATATGTATATCTCCTTCTTAAAG
46	Exbromo_forward	**GACGACGACAAGAT**GGCCGAGAAACTCTCCCCTAACCCA
47	Exbromo_reverse	**GAGGAGAAGCCCGGTTTA**ACTCTCCTCGCCTTCACTGTC

- The primers used for normal PCR are designated as 'forward' and 'reverse' whereas the primers used for site-directed mutagenesis are designated as 'sense' and 'antisense'.
- Bold/underlined sequences represent the adaptor sequences for the pET LIC vectors.
- Bold/italic shows the codon changed in mutagenesis, wherein the lowercase base was changed.

Stock solution of the primer was always 1 µg/µl. The working solution was 100 ng/µl. Usually 2 µl of the working solution was used per PCR reaction for each primer

4.2.3.3 Digestion with restriction enzymes

Usually, 1-2 units of each restriction enzyme were used per 1 μg of plasmid DNA to be digested. The digestion was performed in a buffer specified by the manufacturer at the optimal temperature (usually 37 °C) for 5-16 h. The fragments ends that occurred after digestion were cohesive or blunt. To eliminate possibility of plasmid recirculation (possible when double-digestion does not occur with 100% efficiency), 5'-ends of a vector were dephosphorylated using calf intestine phosphatase (CIP). CIP treatment was performed with 1 unit of enzyme per 3 μg of plasmid DNA, at 37 °C for 1 h.

4.2.3.4 Purification of PCR and restriction digestion products

DNA obtained from restriction digestion, phosphatase treatment or PCR was purified from primers, nucleotides, enzymes, buffering substances, salts, agarose, ethidium bromide, and other impurities, using a silica-gel column (QIAquick PCR Purification Kit, Qiagen). The QIAquick system uses a simple bind-wash-elute procedure. A binding buffer was added directly to the PCR sample or other enzymatic reaction, and the mixture was applied to the spin column. Nucleic acids adsorbed to the silica-gel membrane in the high-salt conditions provided by the buffer. Impurities and short fragments of single or double-stranded DNAs were washed away and pure DNA was eluted with a small volume of 10 mM Tris pH 8.0 or water.

4.2.3.5 Ligation independent cloning

The ligation independent cloning requires the PCR amplification of the gene of interest with the compatible overhang at both ends (LIC cloning kit). Following are the sequences of the overhangs:

<u>For pET30 LIC/Xa</u>

Forward primer 5' – GGTATTGAGGGTCGC – 3'

Reverse primer 5' – AGAGGAGAGTTAGAGCCTTA – 3'

<u>For pET41 or 46 LIC/Ek</u>

Forward primer 5' – GACGACGACAAGAT – 3'

Reverse primer 5' – GAGGAGAAGCCCGGT – 3'

Rest of the procedure was followed as per the instructions of the manufacturer (pET system: manual, Novagen, 2003).

4.2.3.6 Site directed mutagenesis

I used the following modified procedure to make the single site directed or long deletion mutations (Quick change site directed mutagenesis kit: instruction manual, 2003, Stratagene). Using this procedure up to 300 bp were also successfully deleted. The first step is to assemble the following reaction in separate tubes, separate for forward and reverse primer:

Component	**Volume**
10X Reaction Buffer	2.50 µl
S_primer/AS_primer (100 ng/ul)	1.25 µl
DNTP Mix 25mM	0.50 µl
Template DNA (75 ng/ul)	0.50 µl
Pfu Polymerase(5 u/µl)	0.50 µl
ADDW	20.0 µl
Total	**25.0 µl**

Both the reactions were mixed well with the pipette tips and following PCR reaction was made for each reaction separately:

Temperature	**Time**
95 °C	60 s
95 °C	45 s
62 °C	60 s (can vary with the primer to primer)
68 °C	11 min (2 min per 1 kb of DNA template)
Go to 2 and repeat 1 time	
4 °C	Hold

After this reaction, both the tubes were combined and following reaction were run:

Temperature	**Time**
95 °C	60 s
95 °C	45 s
62 °C	60 s (can vary with the primer to primer)
68 °C	11 min (2 min per 1 kb of DNA template)
Go to 2 and repeat 12 times	
4 °C	Hold

After the PCR cycling, the reaction mixture was subjected to *Dpn* I digestion. 1 µl of *Dpn* I restriction enzyme was added to the reaction mixture and incubated at 37 °C for 1 h. 2µl of the reaction mixture was transformed into the Novablue or Top10 cells

4.2.3.7 Agarose gel electrophoresis of DNA

The success of the PCR cycling or the restriction digestion was checked by running the samples of the PCR reaction or the restriction digestion on the agrose gel. Usually 1% agrose gel was prepared in the 1X TAE buffer. DNA samples were run along with the 100bp and/or 1kb DNA ladder (NEB or pEQ lab) at 100-120 V DC. Ethidium bromide stained DNA gel was visualized by UV illumination

4.2.4 Transformation of *E. coli*

4.2.4.1 Making chemically competent cells

A single colony of overnight grown bacteria from a LB agar plate was inoculated into 100 ml of LB media in a 500 ml flask. Culture was incubated at 37 °C with vigorous agitation, monitoring the growth of cells. Cells were grown till the OD_{600} reach ~0.4. The bacterial culture was transferred to sterile, disposable, ice-cold 50 ml polypropylene tubes and cooled down to 4 °C on ice for 10 min. Cells were recovered by centrifugation at 2700g for 10 min at 4 °C. Supernatant media was decanted and tubes were kept in an inverted position on a pad of paper towel for 1 min to allow the last traces of media to drain away. Pellets were resuspended by gentle vortexing in 30 ml of ice-cold $MgCl_2$-$CaCl_2$ solution. Again, cells were recovered by centrifugation at 2700g for 10 min at 4 °C. Supernatant solution was decanted and tubes were kept in an inverted position on a pad of a paper towel for 1 min to allow the last traces of solution to drain away. Pellet of the cells was recovered by gentle vortexing in 2 ml of ice-cold 0.1 M $CaCl_2$ containing 20% glycerol, for each 50 ml of original culture. After this cells were dispensed into aliquots of 50 µl, flash frozen in liquid nitrogen and stored at –70 °C.

4.2.4.2 Transformation of chemically competent cells

1 µl of DNA was mixed with freshly thawed chemically competent cells and the mixture was incubated on ice for 30 min. The mixture was then subjected to heat-shock by incubating it at 42 °C for 45 s. After 1 min incubation on ice 500 µl of SOC medium (Sambrook and Russell, 2001) was added to the mixture and incubated at 37 °C, 200 rpm for 1 h. After that 100 µl of

mixture was plated on the LB-agar plate with desired antibiotic and the plate was incubated at 37 °C for overnight.

4.2.5 Protein chemistry methods & techniques

4.2.5.1 Protein expression

It is important to optimize the conditions for expressing a protein. The aim is to get the maximum amount of a protein in the soluble fraction of the lysate. Often the recombinant protein goes to the insoluble inclusion bodies then the protein should be refolded back to the native form by employing various refolding strategies. A number of parameters were checked to get the maximum yields of the protein, which include optimization of a type of culture media, temperature, induction duration, induction OD, concentration of inducer (IPTG), and cell type. In the beginning, the knowledge about the prior published purification strategies of the same or similar protein can provide help; otherwise the best condition should be determined empirically for each protein. In this work all the proteins were cloned into pET vectors, which are IPTG inducible. The cell pellet after the expression was used directly for the protein purification or stored at -80 °C for longer time or -20 °C if to be used after 1-2 days.

4.2.5.2 Sonication

Sonication is a simple method used for the disruption of cells by ultrasounds. The high frequency sound wave induces formation of cavitations in the solution, the collapse of which produces sheer forces able to break cells. The advantage of this method for cell disruption over French press, freezing and thawing, or lysozyme digestion is comparatively low viscosity of the lysate due to the nucleic acid fragmentation. Pulsed mode of operation was applied (output control 8, 60% duty cycle) and sonication was carried out on ice, in 4 steps of 3 min each, with 5 min intervals between steps, to avoid overheating of the sample.

4.2.5.3 Protein refolding

There are no set rules for achieving best conditions for refolding proteins from inclusion bodies (IBs). During this work I have employed refolding by both rapid dilution and dialysis. Many variables can be tried, like arginine, glycine, reduced/oxidized glutathione ratio, protein concentration, urea, temperature, salt concentration, protease inhibitor, glycerol etc. For each protein the best condition should be found empirically.

4.2.5.4 SDS polyacrylamide gel electrophoresis (SDS PAGE)

The SDS-PAGE was run at various steps to check the purity, and identity of proteins. For smaller proteins Tris-tricine was used (Schagger and von Jagow, 1987) and for proteins > 10 kDa size normal tris glycine SDS PAGE was used (Sambrook and Russell, 2001).

Protein samples with high concentrations of guanidinium HCl were prepared as following: 20 µl of the protein solution in a denaturing buffer was diluted with 400 µl 20% trichloroacetic acid (TCA). The sample was incubated for 15 min at 4 °C followed by centrifugation for 5 min at 20000g. Supernatant was discarded by suction; the precipitated protein pellet was washed once by vortexing with 400 µl ethanol. After centrifugation and ethanol removal, the protein pellet was resuspended in 20 µl of 2x SB and the sample was boiled for 3 min.

For visualization of the protein bands, the gels were stained in a Coomassie-blue solution. Background was cleared by incubation of the gel in a destaining solution. Both processes were greatly accelerated by brief heating with microwaves of the gel submerged in an appropriate solution.

4.2.5.5 Western blot

Western blot is a functional assay to check the identity of proteins or identify the protein out of a number of proteins. In our lab we run the semi-dry Western blot. The Western blot assay starts with running the desired sample on the SDS-PAGE. The nitrocellulose membrane and six Watmann paper of the size of the SDS-PAGE gel were cut and soaked in the transfer buffer. Watmann paper, SDS-PAGE (gel), and nitrocellulose membrane were arranged in the following order over the electroblot: three wet Watmann paper, wet nitrocellulose paper, and SDS-PAGE gel followed by three Watmann wet papers. The apparatus was closed and run (transfer) at constant voltage of 100 V for 1-1:30 h. After the transfer, the nitrocellulose membrane is taken and kept in the blocking solution for 1 h with constant shaking. The SDS-PAGE is stained by Commassie blue solution to check the success of transfer. After blocking, membrane is washed three times with the wash buffer and incubated for 1 h at room temperature with the 1st antibody solution (the procedure can be stopped at this point by keeping the blot in the 1st antibody solution at 4 °C for overnight). Membrane was with the wash buffer and incubated in the 2nd antibody solution for 1 h at room temperature. After this the membrane was washed three times with the wash buffer and develop blot by incubating it in the substrate (BCIP) solution for 10 min.

4.2.5.6 Determination of protein concentration

The concentration of proteins in solution was estimated by means of the Bradford colorimetric assay. 10 µl of the protein sample was added to 1 ml (10x diluted stock) Bradford reagent (BioRad) in a plastic cuvette. After gentle mixing, A_{595} was measured and converted to the protein concentration on the basis of a calibration curve prepared for known concentrations of BSA.

A precise protein concentration determination was performed spectrophotometrically. Absorption at 280 nm was measured and converted to a protein concentration on the basis of theoretical extinction coefficients. It has been shown that it is possible to estimate the molar extinction coefficient E λ(Prot) of a protein from knowledge of its amino acid composition (Gill and Hippel, 1989). From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine residues do not absorb appreciably at wavelengths >260 nm, while cystine does) at a given wavelength λ the extinction coefficient of a protein can be computed using the equation:

Eλ(Prot) = Numb(Tyr)xExtλ(Tyr) + Numb(Trp)xExtλ(Trp) + Numb(cystine)xExtλ(cystine)

The absorbance (A, optical density) can be calculated using the following formula:

Aλ (Prot) = Eλ(Prot) / Molecular weight.

4.2.5.7 Protein purification strategy

To finally get a protein in a homogenous form, in large quantities, requires not only the experimental skills, but planning and patience. The required purification level of a protein depends on the use of the protein for later work. However, for crystallization or NMR experiments the highest level of a purified protein is imperative. Protein purification varies from simple one-step precipitation procedures to large scale multiple-steps processes. The key to successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements and combine them in a logical way to maximize yield and minimize the number of steps required. Most purification schemes involve some form of chromatography. Different chromatography techniques with different selectivities can form powerful combinations for the purification of any biomolecule. Biomolecules including proteins are purified using purification techniques that separate according to differences in specific properties (Table 4.4).

Invariable strategy used to get the purified proteins is shown in Figure 4.1. In this thesis, I have used three types of chromatography techniques extensively: Affinity chromatography, gel filtration size exclusion chromatography, and ion exchange chromatography (The QIAexpressionist™, 2003; Amersham-Pharmacia manual, 2003).

Table 4.4. *Types of chromatographic techniques based on the biochemical properties of a protein*

Property	Chromatography
Biorecognition (ligand specificity)	Affinity chromatography
Charge	Ion exchange
Size	Gel filtration (or size exclusion)
Hydrophobicity	Hydrophobic interaction
Hydrophobicity	Reversed phase

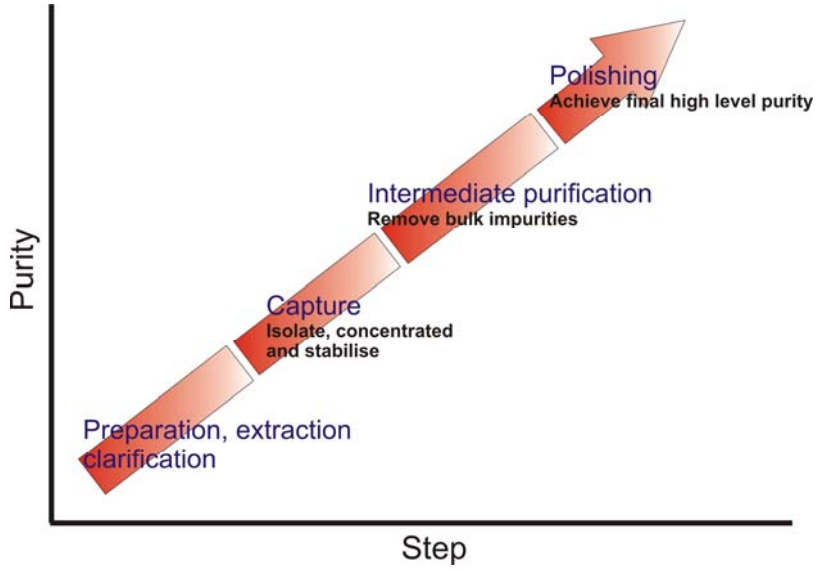

Figure 4.1. *A general optimized strategy to achieve high level purity (adapted from Amersham-Pharmacia manual, 2003).*

4.2.6 Other protein analytical methods

4.2.6.1 Pull down assays

The 'bait' proteins with the His–tag or GST–tag were bound to respective columns. 'Prey' protein without purification tag (His-tag or GST) was put over column saturated with the bait. Columns were washed with a suitable wash buffer and complexes were eluted with 250 mM imidazole or 10 mM reduced glutathione. Fractions obtained were visualized on SDS-PAGE.

4.2.6.2 Analytical gel filtration chromatography

A mixture of two proteins or a protein and peptide was subjected to separation on the Sephadex S75 analytical column. In case of protein-protein interactions, the shift in the position of peaks of the complex was compared with individual proteins and fractions containing the complex were checked on SDS-PAGE and analyzed by mass-spectrometry.

The oligomeric states of the proteins were also analyzed by the analytical gel filtration. The elution volume of the proteins (which corresponds with the relative molecular weight) was compared with the elution profile of a known standard.

4.2.6.3 Mass spectrometry

Mass spectrometry was carried out on an ESI–MS API 165 Perkin–Elmer Sciex (Langen) coupled with HPLC (column: Macherey–Nagel EC 125/2 Nucleosil 300–5 C4 MPN; pump system: Microgradient System 140B/C Perkin Elmer (solvent A: water, 0.05 % TFA, B: MeCN, 0.05% TFA; gradient 10–95% B); photodiode array Agilent HP1100PDA; software: Masschrom, Biomultiview).

4.2.7 NMR spectroscopy

All NMR spectra were carried out on the 600 MHz or 900 MHz spectrometer. 1D-^{1}H spectra require protein concentrations as low as 0.05 mM (preferably in the phosphate buffer) whereas 2D or higher dimensional spectra require protein to be labeled and require higher concentrations of protein (0.2 mM to 1 mM). Typically a 450 µl of the protein sample and 10 µl of $^{2}H_2O$ was mixed and added to the NMR tube. All 1D-^{1}H NMR spectra were recorded with a time domain of 32 K complex points and a sweep-width of 10,000 Hz. The 2D ^{1}H-^{15}N-HSQC spectra were recorded with a time domain of 1 K complex data points with 128 complex increments with a sweep width of 8 kHz in the ^{1}H dimension and 2 kHz in the ^{15}N dimension. All the spectra were analyzed with the XWINMR or/and Sparky software.

4.2.7.1 Perdeuteration of pRb-AB

The strategy for making all the reagents was same as for normal minimal media for uniform labeling, except that 99% D_2O was used in place of H_2O. Minimal media were prepared with 30% D_2O (100 ml), 60% D_2O (100 ml), 75% D_2O (100 ml), 90% D_2O (100ml), 99% D_2O (1000 ml). A single colony from a freshly transformed LB-agar plate was inoculated in 50 ml of 30% D_2O MM and adapted to the D_2O containing media with gradual increase of % D_2O. Rest of the procedure is described in Table 4.5:

Table 4.5. *Summary of the steps for culturing E. coli for perdeuterated pRb-AB*

Step	***Time (h)***	***OD_{600}***
Fresh colony from LB-agar plate in 50 ml of 30% D_2O minimal medium	32	1.5

Step	***Time (h)***	***OD_{600}***
25 µl of 30% in 50 ml of 60% D_2O minimal medium	12	0.667
1 ml of 60% in 49 ml of 75% D_2O minimal medium	12	0.658
1 ml of 75% in 49 ml of 90% D_2O minimal medium	12	0.956
5 ml of 90% in 95 ml of 99% D_2O minimal medium	12	0.456

After this, 100 ml of 99% of the last step was inoculated into 1000 ml of 99% D_2O minimal medium. Cells were grown at 37 °C for 11 h, till the cells reached OD_{600} of ~0.765. The temperature was reduced to 20 °C and induced with 1 mM IPTG for next 19 h. Cells were pelleted down and the protein was purified as per normal procedure.

4.2.8 Protein crystallization

I employed both sitting drop and hanging drop vapor diffusion techniques for crystallization. Crystallization was set up in the following manner: more then 95% pure protein sample was loaded on the preparative grade gel filtration column (S75 or S200) in (preferably) Tris or HEPES buffer. Initially the buffers with low salt (>100 mM salt) were tried, however in pRb-AB crystallization 150 mM NaCl was used whereas BRG1 bromodomain constructs were tried in different salt concentrations but the final crystal grew from 20 mM HEPES; pH 7.4, 200 mM NaCl. The protein was concentrated to ~15-20 mg/ml and centrifuged at high rpm briefly. 2 µl of protein was mixed with 2 µl of reservoir buffer in the well filled with 400 µl of buffer. Crystallization plates were set up at two temperatures 4 °C and room temperature.

For pRb-AB, crystals appear in two conditions in the initial screen (Hampton CSI No. 4 and Hampton Cyro No. 9). Optimizing this condition further produced the crystals. For the exbromodomain, crystals appear in three conditions, Hampton PEG/Ion No. 29, Hampton Index No. 45 and Hampton Index No. 62.

The process of data recording and structure calculations is described in the 'results and discussion' chapter 5 with the pRb-AB and bromodomain crystals analysis.

4.2.9 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a thermodynamic technique for monitoring any chemical reaction initiated by the addition of a binding component, and has become the method of choice for characterizing biomolecular interactions. When substances bind, heat is either generated or absorbed. Measurement of this heat allows accurate determination of dissociation constants (K_D), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS), thereby providing a complete thermodynamic profile of the molecular interaction in a single experiment. In ITC experiment, a

syringe containing a 'ligand' solution is titrated into a cell containing a solution of the 'macromolecule' at constant temperature. When ligand is injected into the cell, the two materials interact, and heat is released or absorbed in direct proportion to the amount of binding. As the macromolecule in the cell becomes saturated with ligand, the heat signal diminishes until only background heat of dilution is observed.

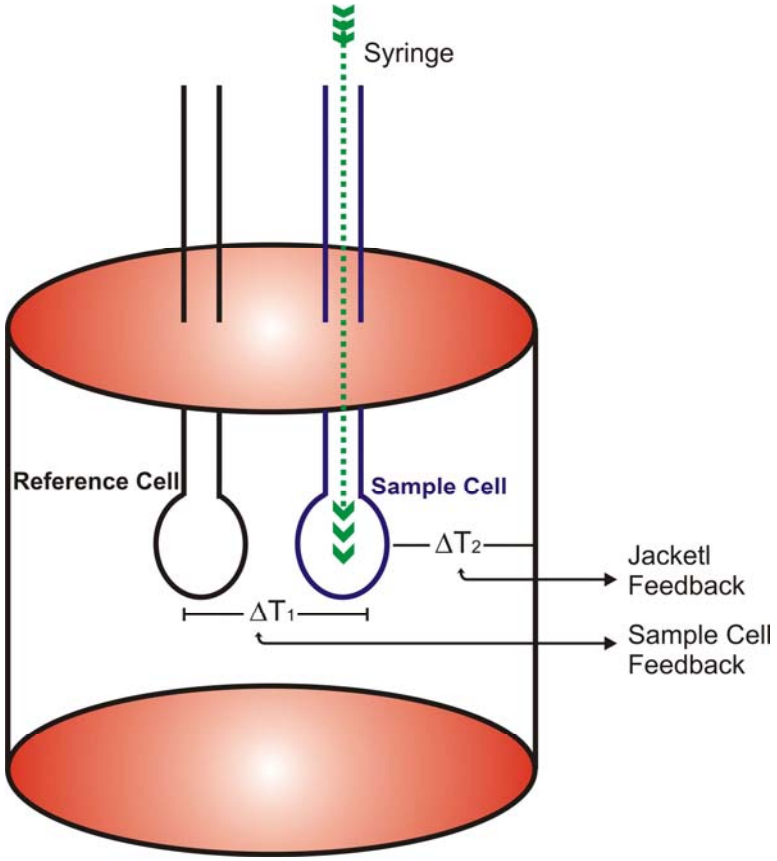

Figure 4.2. *Schematic drawing of the ITC microcalorimeter. A pair of identical coin shaped cells is enclosed in an adiabatic Outer Shield (Jacket). Access stems travel from the top exterior of the instrument to the cells. Both the coin shaped cells and the access stems are totally filled with liquid during operation. Temperature differences between the reference cell and the sample cell are measured and calibrated to power units. This generated signal is sometimes referred to as the 'feedback' power used to maintain temperature equilibrium.*

In the binding studies presented here, the VP-ITC system (MicroCal) was used. It is a sensitive isothermal titration calorimeter that uses a cell feedback network (CFB) to differentially measure heat produced or absorbed between the sample and reference cell. Twin coin-shaped cells are mounted in a cylindrical adiabatic environment, and connect to the outside through narrow access tubes (Fig. 4.2). A thermoelectric device measures the temperature difference between the two cells and a second device measures the temperature difference between the cells and the jacket. As chemical reactions occur in the sample cell, heat is generated or absorbed. The temperature difference between the sample and reference cells (ΔT_1) is kept at zero by the

addition of heat to the sample or reference cell, as appropriate, using the CFB system. The integral of the power required to maintain $\Delta T_1 = 0$ over time is a measure of total heat resulting from the process being studied.

I employed ITC to characterize the binding of pRb-AB to LXCXE peptides and of the KR/AT-hook domain and DNA. A typical ITC experiment was carried out as following: component 1 (ligand) in the syringe is titrated into the second component 2. Usually the ligand concentration is typically 10 to 20 times the concentration of the component in the cell. The volume of the component 1 in the cell was 1430 μl and component 2 in the syringe 600 μl. All solutions were degassed prior to measurements. Heat generated by protein dilution was determined in separate experiments by injecting protein solution into the PBS filled sample chamber. All data were corrected for the heat of protein dilution. The details of the experimental and injection parameters are described below (for parameters see Table 4.6).

In case of the pRb-AB and LXCXE peptide titrations, protein pRb-AB was the component 2 in the cell, whereas ligands were LXCXE peptide. The buffer system comprised: 50 mM Na_2HPO_4, 10 mM βME, pH 7.8, 0.01% NaN_3. Protein was extensively dialyzed against the buffer, whereas the peptides were dissolved in the final dialysis buffer, pH of peptide solutions were carefully checked and adjusted in some cases. Protein concentration was 20-50 μM whereas the peptide concentration was 200-400 μM. Experiments were carried out at 20-22 °C.

In case of the AT hook domain and DNA titrations, protein was the ligand (component 1) and component 2 in the cell was DNA solution. The buffer system comprised of 10 mM KH_2PO_4, 100 mM NaCl, pH 7.1, 0.01% NaN_3. Both protein and DNA were dialyzed against the same buffer extensively. Protein concentration was ~200-500 μM and DNA concentration in the cell was 20-50 μM. Experiments were performed at various temperatures ranging from 5-30 °C.

Table 4.6. *Parameters set up for ITC experiments*

Parameter	***pRb-AB vs LXCXE peptides***	***DNA vs AT hook domain or KR domain***
Total number of injections	45	28
Volume of a single injection [μl]	5	10
Duration of an injection [sec]	10	30
Intervals between injections [sec]	600	400
Filter period [sec]	2	2
Equilibrium cell temperature [°C]	20-22	5-30
Initial delay [sec]	60	60
Reference power [μCal/sec]	15	15

Parameter	***pRb-AB vs LXCXE peptides***	***DNA vs AT hook domain or KR domain***
Stirring speed [RPM]:	270	270

All data were corrected for the heat of protein dilution. Data were fitted using χ^2 minimization on a model assuming a single set of sites to calculate the binding affinity K_D. All steps of the data analysis were performed using ORIGIN (V5.0) software provided by the manufacturer.

4.2.10 CD spectroscopy

CD measures the interaction of circularly polarized light with molecules. Circularly polarized light comes in left- and right-handed forms. These forms interact equally with non-chiral chromophores, but differently when the chromophore has a right- or left-handed form. The result is a small difference in the extinction coefficients for left- and right-polarized light, which varies with wavelength. In proteins, the chiral arrangement of peptide bonds in secondary structures such as α-helix and β-sheet leads to characteristic CD spectra at UV wavelengths. Alteration of the relative orientations, due to conformational or structural variations, causes changes in the CD spectrum. Other molecules such as nucleic acids also have CD spectra at wavelengths where they absorb light.

Secondary structure can be determined by CD spectroscopy in the 'far-uv' spectral region (190-250 nm). Alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectrum (Fig. 4.3). The approximate fraction of each secondary structure type that is present in any protein can thus be determined by analyzing its far-uv CD spectrum as a sum of fractional multiples of such reference spectra for each structural type. Like all spectroscopic techniques, the CD signal reflects an average of the entire molecular population. Far-uv CD spectra require 20 to 200 µl of solution containing 1 mg/ml to 50 µg/ml protein, in any buffer which does not have a high absorbance in this region of the spectrum. (high concentrations of DTT, histidine, or imidazole, for example, cannot be used in the far-uv region). Thermal stability is assessed using CD by following changes in the spectrum with increasing temperature. In some cases the entire spectrum in the far- or near- UV CD region can be followed at a number of temperatures. Alternatively, a single wavelength can be chosen which monitors some specific feature of the protein structure, and the signal at that wavelength is then recorded continuously as the temperature is raised. CD is often used to assess the degree to which solution pH, buffers, and additives such as sugars, amino acids or salts alter the thermal stability, which is usually recorded as transition temperature (Tm). If the melting of a protein is reversible, the

thermodynamics of protein folding can be extracted from the data. The introduction presented here was adapted from Fasmann (1999) and Moss (2001).

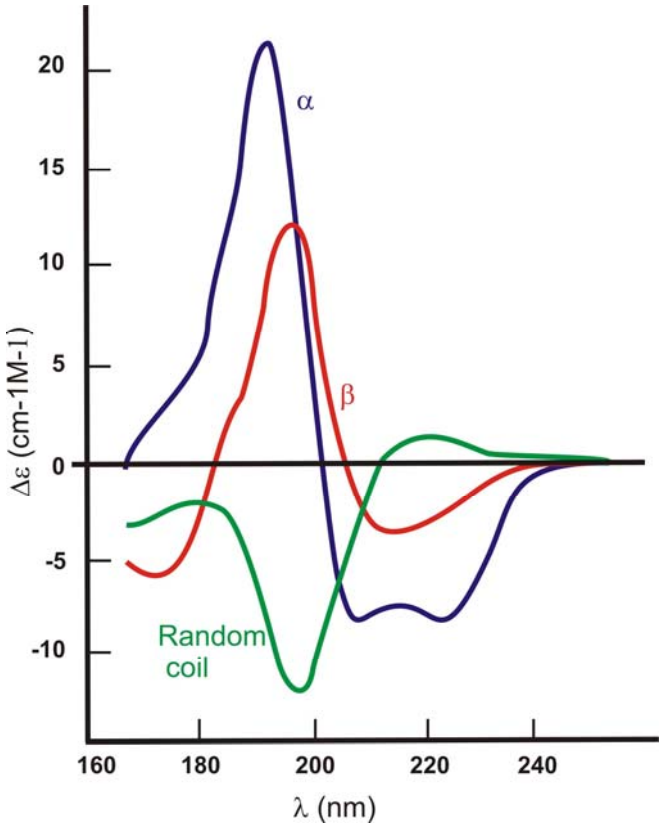

Figure 4.3. *Far UV CD spectra of various types of secondary structure. Blue curve represents ,100% α-helix; red, curve represents 100 % β-sheet and green curve represents 100% random coil conformation.*

All the CD spectra were recorded on JASCO J-715 spectrometer. Protein or DNA solution at concentration 5-10 μM in a suitable buffer was used in a 0.1 mm cuvette. To reduce random error and noise, each spectrum was recorded as an average of four scans. Only buffer spectra was recorded and subtracted to isolate the conformational change solely from protein or DNA. The following parameters were used for data acquisition (Table 4.7):

Table 4.7. *Parameters setup for a typical CD spectrum acquisition.*

Parameter	**Value**
response	2 sec
scanning speed	20 nm/min
bandwidth	1.0 nm
sensitivity	5 mdeg
step resolution	0.1 nm

The melting curve for a protein was recorded using CD spectroscopy by recording the CD signal at 222 nm (one of the characteristic minima for α helix) or 198 nm (the characteristic maxima for β sheet) with the increasing temperature (usually from 4 to 80 °C).

5 Results and discussion

This part of the thesis lists the results of experiments performed, and it sums up the protocol of protein constructs from cloning to their functional studies. Protein cloning, expression, purification and functional studies were performed in-house. Most of the compounds, peptides or oligonucleotides were purchased form outside companies or were provided by in-house service of the institute. Functional analysis of two proteins were carried out, the retinoblastoma protein (pRb) and the human BRG1 protein. Section 5.1 and its subsections document the results and discussion on pRb and its interaction with other proteins. Section 5.2 and its subsections document the results and discussion on various domains of the hBRG1 protein including the structure determination of the bromodomain from hBRG1.

5.1 The retinoblastoma protein and interaction partners

The objective of obtaining pRb domains was to study their interactions with the various binding partners with the ultimate aim of characterizing those interactions using various biochemical and biophysical techniques. These techniques included ITC, mass-spectrometry, X-ray crystallography, and NMR spectroscopy. Interaction of pRb with a number of proteins was checked, which included viral oncoproteins (HPV16 E7 and SV40 large T antigen), the E2F1 protein, MyoD, Id-2, HDAC1, PAI2, gankyrin and hBRG1. Attempts to crystallize pRb-AB and assign the 2D ^{15}N-^{1}H HSQC spectra of pRb-AB were also tried.

5.1.1 Cloning and expression

This section lists the results of cloning and expression of various protein constructs used. Only a limited number of proteins made into the next level i.e. the purification stage as in some cases expression was very low or none, and other times proteins refused to refold from the inclusion bodies. Results of various pRb constructs are summarized in Table 5.1.

5.1.1.1 pRb constructs

I started with the cloning of pRb-ABC construct, which expressed with moderate levels in *E. coli* in soluble fraction. Various other constructs were further cloned or made using site directed mutagenesis, taking the pRb full length protein or the pRb-ABC as a template (Fig. 5.1). pRb-NABC, pRb-NA, and pRB-ABC64 did not express under various conditions in *E. coli*. Mutant

pRb-AB (C706F) and pRb-AB (K713A) expressed well, whereas rest of the mutants did not express in *E. coli* (Table 5.1).

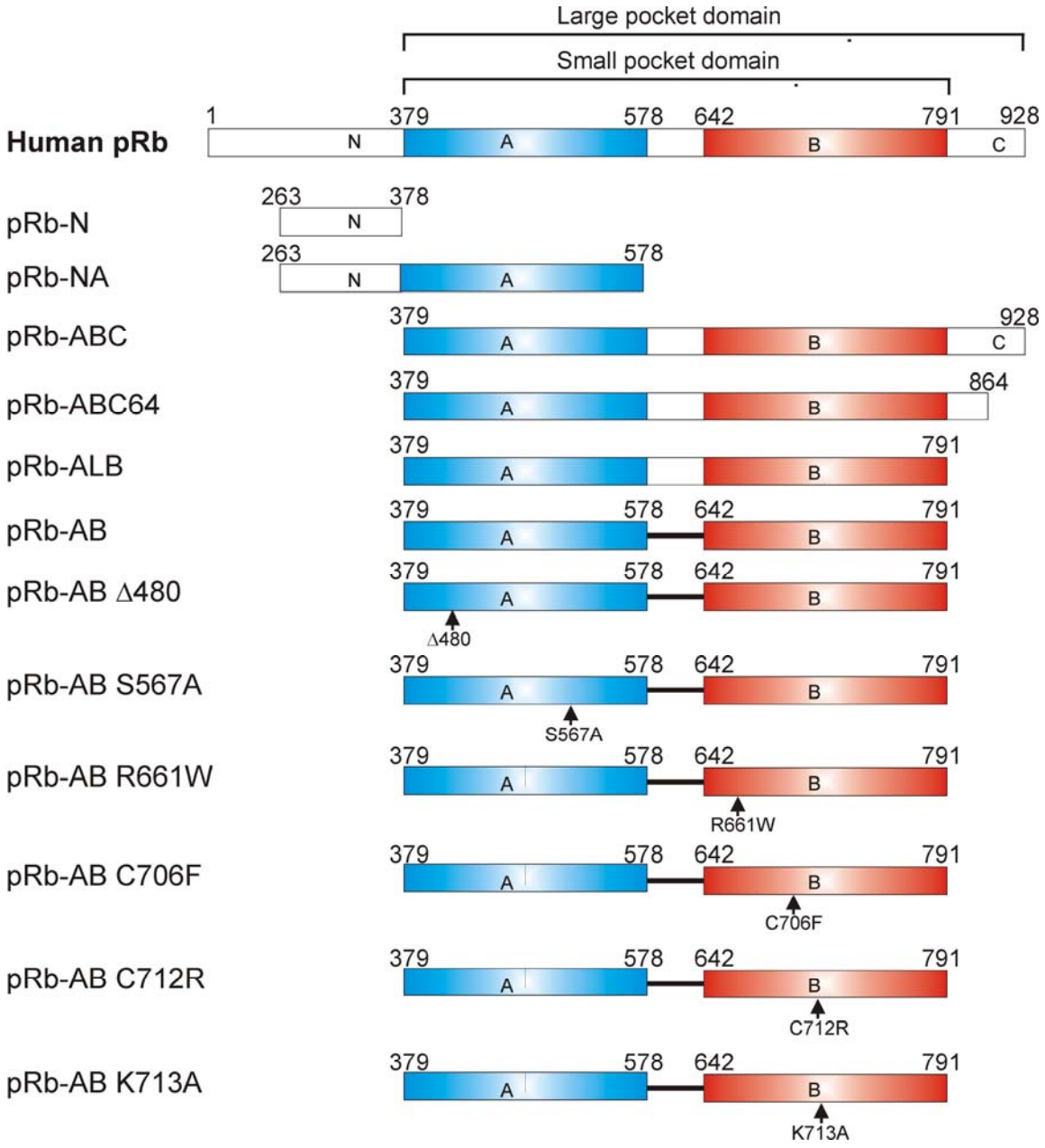

Figure 5.1. *Graphical representation of various pRb constructs cloned and used in this study. All the constructs are compared with the full-length protein.*

These results were unexpected 'though' interesting. The idea behind expressing these mutants was to check their physical properties in comparison with the wild type protein. The X-ray structure of pRb-AB shows that these residues are important for the structural integrity of the AB pocket thus these mutations probably make the protein unstable. Only pRb-ABC, pRb-AB and pRb-N were further used. For the large scale expression of pRb-ABC, pRb-AB, and pRb-N,

1 l of LB medium was inoculated with 10 ml of the overnight grown primary culture. Cells were induced with 1 mM IPTG, pRb-AB cultures were additionally supplemented with 100 mM $MgCl_2$. Induction temperature for pRb-AB was 18 °C and for pRb-N 26 °C. Induction was carried out for overnight (~14-16 h). Cells were pelleted down by centrifugation and stored at –20 °C until purification.

Table 5.1. *Results of cloning and expression of various pRb constructs produced during this thesis*

No.	***Construct Name***	***Primer no.****	***Vector***	***Expression in E. coli***	***1D-^{1}H NMR diagnosis***
1	pRb-NABC	1, 3	pET30 LIC/Xa	No	ND
2a	pRb-ABC	2, 3	pET30 LIC/Xa	Moderate	Partially folded
2b	pRb-ABC♣	4, 5	SDM of pRb-ABC	Moderate	Partially folded
3	pRb-ALB	6, 7	SDM of pRb-ABC	High	Partially folded
4	pRb-AB	8, 9	SDM of pRb- ALB	High	Folded
5	pRb-ABC64	2, 10	SDM of pRb-ABC	No	ND
6	pRb-N	11, 12	pET30 LIC/Xa	High	Unstructured
7	pRb-NA	11, 13	pET30 LIC/Xa	No	NA
8	pRb-AB(Δ420)	14, 15	SDM of pRb- AB	No	NA
9	pRb-AB(S567A)	16, 17	SDM of pRb- AB	No	NA
10	pRb-AB(R661W)	18, 19	SDM of pRb- AB	No	NA
11	pRb-AB(C706F)	(Made by Ola)	SDM of pRb- AB	Moderate	Folded
12	pRb-AB(C712R)	20, 21	SDM of pRb- AB	No	NA
13	pRb-AB(K713A)	(Made by Ola)	SDM of pRb- AB	Moderate	Folded

Abbreviations: LIC, ligation independent cloning; SDM, site directed mutagenesis; ND, not done; NA, not applicable.

* Please refer to Table 4.3 for the primer name, no. and the sequence in (section 4.2.3.2).

♣ This construct has short His-tag compared to pRb-ABC.

5.1.1.2 MyoD, Id-2 and HPV E7

The MyoD construct used in this work consisted of the basic helix-loop-helix (bHLH) region from amino acids 106 to 173 and was named MCNC. The Id-2 constructs comprised the full length protein. MyoD and Id-2 constructs were cloned into the pET30 LIC/Xa vector and expressed in the *E. coli* BL21 (DE3) cells. Cells were grown at 37 °C until OD_{600} ~0.7. Cells were induced with 1 mM IPTG for 4 h at 30 °C. Both MyoD and Id-2 were expressed in inclusion bodies.

Two constructs of the E7 protein were used in this work. The HPV16 E7 full-length construct was cloned into the pET8 vector and had a C-terminal 6xHis-tag. The HPV16 E7-short was cloned into the pET30 Xa/LIC vector and spanned residues 21-40. Protein expression was carried in a manner similar to the MyoD expression.

5.1.2 Protein purification strategies

5.1.2.1 pRb-N, pRb-AB, and pRb-ABC

The purification strategy for pRb-N, pRb-AB and pRb-ABC is shown in the form of a flow-chart (Fig. 5.2).

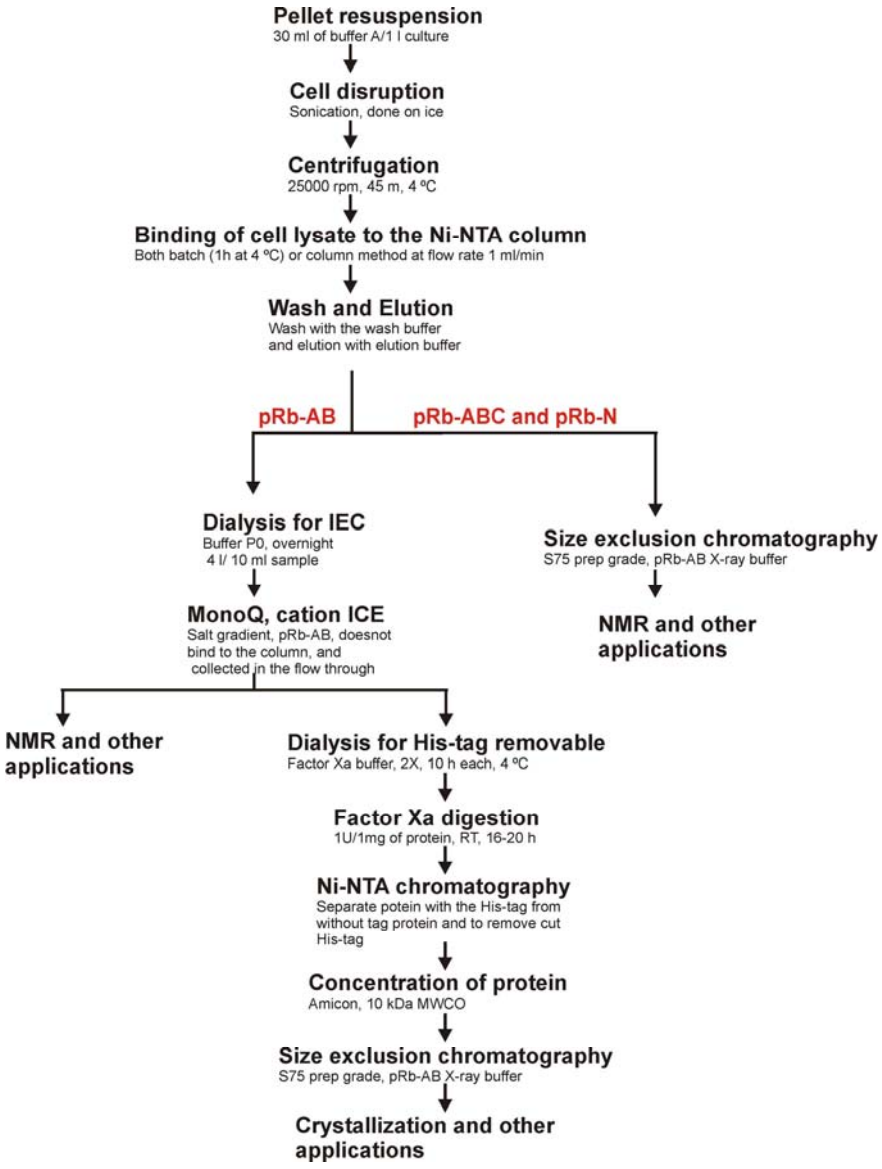

Figure 5.2*. Flow chart of the purification scheme for pRb-ABC, pRb-AB, and pRb-N constructs.*

Theoretical isoelectric points (pI), molecular masses (MW), and extinction coefficients (E_{280}) were deduced from the amino acids sequence of the studied protein in order to aid further purification. We frequently use ProtParam on www.expasy.ch for determining these parameters. Examination of already existing methods of purification of similar or homologous proteins was also used in designing the purification protocols.

***Table 5.2**. Physicochemical properties of pRb-ABC, pRb-AB, and pRb-N*

No.	Name	No of AA	MW(Da)*	E280 (M^{-1} cm^{-1})**	pI
1	pRb-ABC (with his-tag)	561	64566.4	43150	9.21
2.1	pRb-AB (with his-tag)	361	42461.4	40205	8.77
2.2	pRb-AB (without tag)	350	41051.8	40205	8.77
3	pRb-N (with his-tag)	169	19200.5	3960	6.19

*Mol. weight is calculated for the proteins with oxidized cysteines.

**E(Prot) = Numb(Tyr) x Ext(Tyr) + Numb(Trp) x Ext(Trp) + Numb(Cystine) x Ext(Cystine)

Closer examination of the proteins amino acid sequence may also reveal potential protease cleavage/degradation sites. Table 5.2 shows the basic physicochemical properties of pRb-N, pRb-AB and pRb-ABC. The theoretical pI informs of the net charge of a protein placed in solvent of a given pH. Thus a proper combination of buffer and ion-exchanger (cation- or anion-exchanger) medium can be chosen. Purification of pRb-AB employed Ni-NTA chromatography (or IMAC) followed by MonoQ anion exchange chromatography (Fig. 5.3 and 5.4). For MonoQ chromatography a buffer of pH 7.4-7.6 was used (buffer P0). At this pH, pRb-AB should be protonated i.e. positively charged, as this pH is below its pI value. Hence pRb-AB did not bind the column at this pH, eluted in the flow through while impurities were bound, giving almost a >95% pure protein.

5.1.2.2 MyoD, Id-2, and HPV E7

MyoD and Id-2 proteins were purified using Ni-NTA chromatography under denaturing conditions. Inclusion bodies were resolubilized using buffer: 6 M GuaHCl, 50 mM Tris pH 8. Buffer for the purification consisted of 6 M GuaHCl, 50 mM Tris (lysis pH 8, wash pH 6.0, and pH 5.2 elution pH 4.5). Proteins were refolded using dialysis. Proteins at low concentration (~1 mg/ml) were dialyzed to 50 mM NaH_2PO_4, 300 mM NaCl, pH 8, 10 mM βME (buffer N0).

The E7 full-length and E7 short proteins were purified using Ni-NTA chromatography under native condition followed by S75 gel filtration chromatography using the buffer: 50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 10 mM βME (buffer N0).

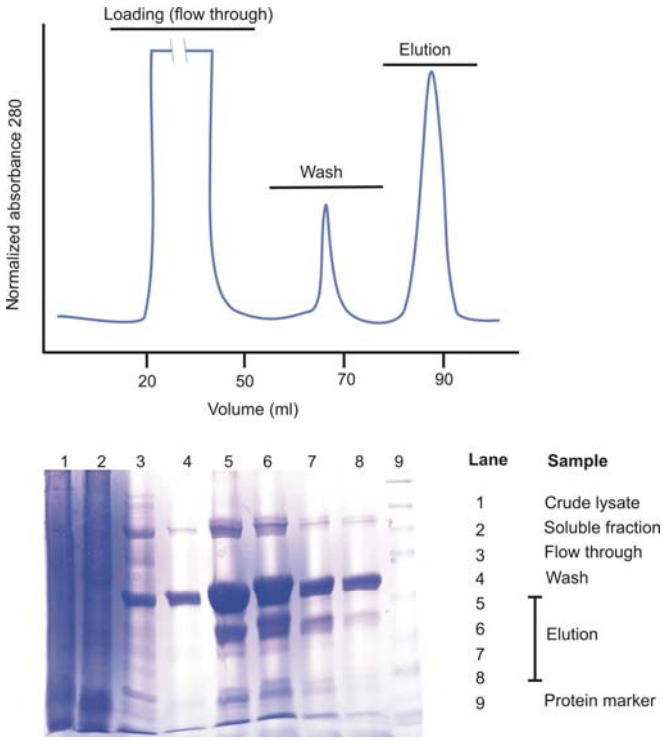

Figure 5.3. *A typical chromatogram and SDS-PAGE of the pRb-AB purification after Ni-NTA chromatography.*

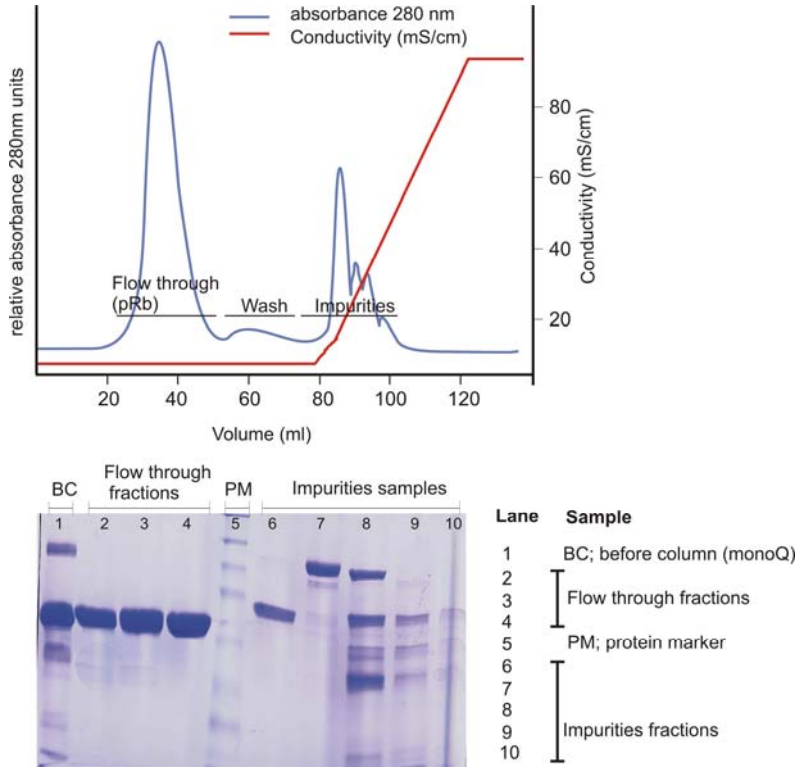

Figure 5.4. *A typical chromatogram and the SDS-PAGE analysis of the pRb-AB after MonoQ ion exchange chromatography. Note that pRb-AB did not bind to the column and eluted in the flow-through.*

5.1.3 Functional and structural studies of the retinoblastoma protein

5.1.3.1 Characterization of pRb-ABC, pRb-AB, and pRb-N

The recombinant proteins were checked for the purity and identity by SDS-PAGE analysis. Apart from this we characterized the recombinant proteins using analytical gel filtration chromatography, Western blot assay, N-terminal sequencing and mass spectrometry (Fig. 5.5). The folding of the proteins was checked by NMR spectroscopy and CD spectroscopy (Fig. 5.6).

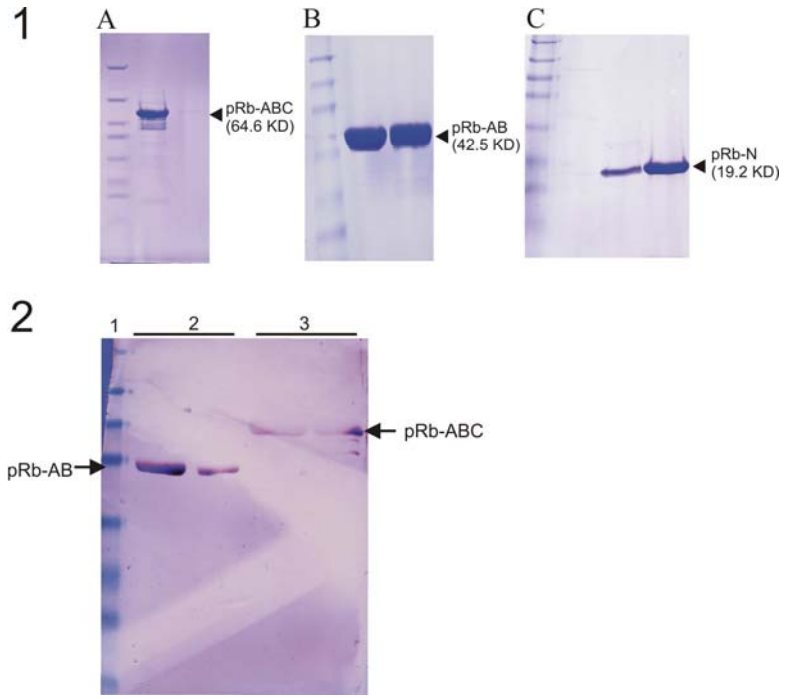

***Figure 5.5**. Characterization of the proteins using SDS-PAGE (panel 1) and Western-blot (panel 2). Panel 1, SDS-PAGE of (A) pRb-ABC, (B) pRb-AB, (C) pRb-N. First lane always shows the protein molecular weight marker (NEB, prestained). Usually proteins with this purity level were used for further work. Panel 2, A typical Western-blot of pRb-ABC and pRb-AB proteins using anti-His primary antibodies.*

pRb-N, pRb-AB, and pRb-ABC eluted as monomers in gel filtration. Internal cleavage at the C-terminus of pRb-ABC was observed over a period of time. The 1D-^{1}H spectra of the protein showed pRb-N as unfolded, pRb-ABC is partially unfolded, and pRb-AB as a well folded protein (Fig. 5.6)

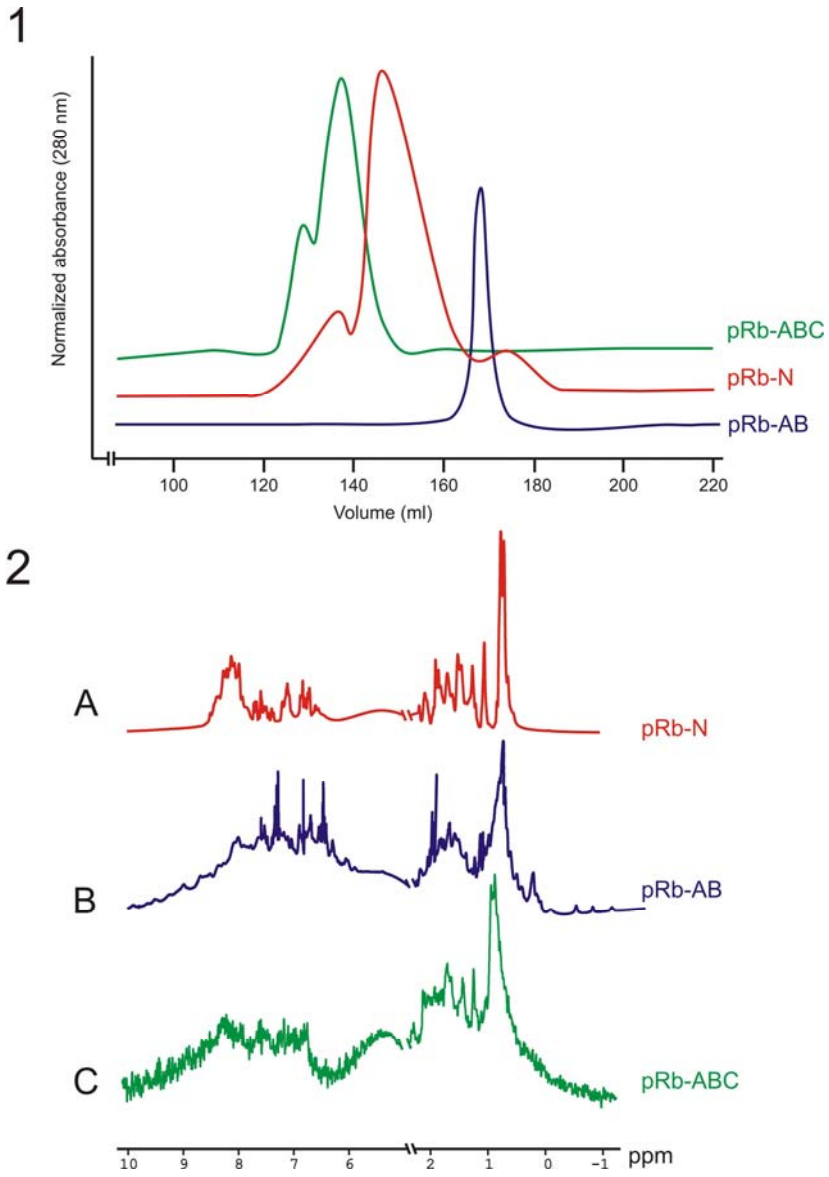

Figure 5.6. *1. Gel filtration profile of various pRb constructs under native conditions. A Superdex S75 prep grade column was used in this case. 2. 1D-^{1}H spectra of various pRb constructs further used in this work, showing the folding characteristics of these protein domains. (A) pRb-N. (B) pRb-AB. (C) pRb-ABC.*

5.1.3.2 The interaction between pRb and HPV E7, MyoD and Id-2 proteins

In this section we have checked for putative interactions between pRb-AB and MyoD and Id-2. Functionalities of the proteins were checked by protein-protein or protein-DNA interaction assays. NMR spectroscopy of the ^{15}N uniformly labeled MyoD sample showed that MyoD (MCNC) was able to bind target DNA (sequence 5'-TCAACAGCTGTTGA-3') proving the functionality of the protein. Additionally MyoD (MCNC) was checked for formation of a complex with Id-2 using the affinity chromatography pull-down assay and gel filtration chromatography. Gel filtration chromatography of 1:1 molar mixture of Id-2 and MyoD (MCNC) showed co-elution of both proteins in a single peak (Fig. 5.7A).

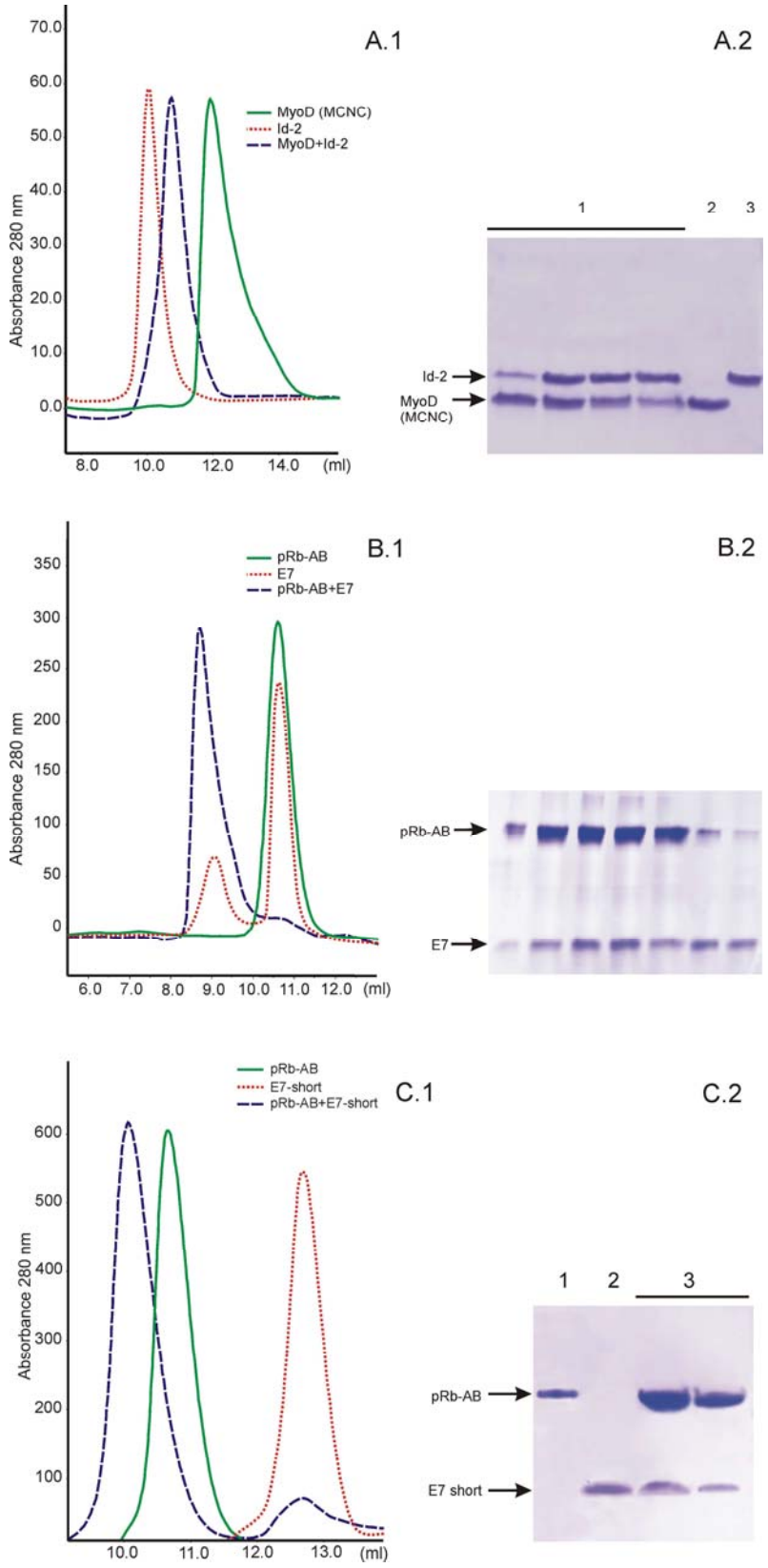

Figure 5.7. *Monitoring the formation of protein-protein complexes by gel filtration chromatography. Panels from A.1 to C.2 show chromatograms and panels A.2 to E.2 SDS-PAGE analysis of consecutive chromatography fractions. Panel A.1 shows MyoD (MCNC) – Id-2 complex (dash line), Id-2 alone as a dotted line (tetramer), MyoD (MCNC) alone as a straight line (tetramer); Panel A.2 SDS-PAGE analysis of chromatography fractions - Part 1 MyoD (MCNC) – Id-2 complex (consecutive fractions), Part 2 MyoD (MCNC) alone, Part 3 Id-2 alone. Panel B.1 shows the pRb-AB - E7 (full-length) complex (dash line), E7*

(full-length) alone (dimers and oligomers) as a dotted line, pRb-AB alone (straight line); Panel B.2 SDS-PAGE analysis of chromatography fractions - all lanes pRb-AB - E7 complex. Panel C.1 shows the pRb-AB - E7 (short) complex (dash line), the E7 (short) alone (dot line), pRb-AB alone (straight line); Panel C.2 SDS-PAGE analysis of chromatography fractions - part 1 pRb-AB, part 2 E7 (short), part 3 the E7 (short) – pRb-AB complex.

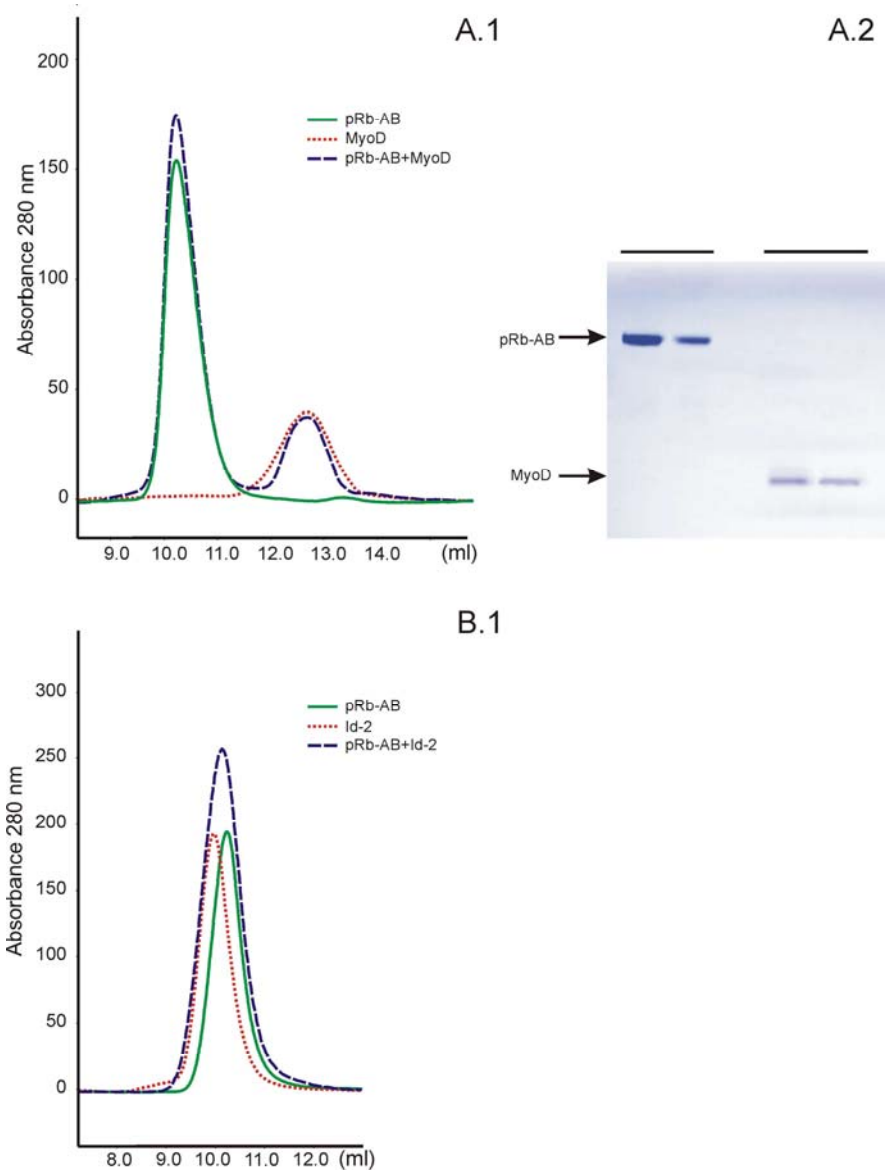

Figure 5.8. *Monitoring the formation of protein-protein complexes by gel filtration chromatography. Panel A.1 illustrates that no complex was formed between MyoD (MCNC) and pRb-AB (dash line), pRb-AB alone (straight line), MyoD (MCNC) alone (dotted line). Panel E.2 shows the results of SDS-PAGE analysis of chromatography fractions - Part 1 pRb-AB, Part 2 MyoD (MCNC). Panel B.1 demonstrates that no complex was formed between Id-2 and pRb-AB (dashed line). Id-2 alone is shown as a dotted line (tetramer), pRb-AB alone (straight line). No SDS-PAGE analysis was performed because both proteins elute at the same column volume. The Superdex 75 HR 10/30 gel filtration chromatography column was used in all experiments except experiment D where column HiLoad 26/60 Superdex 75 prep grade was used. If more than one lane corresponds to a given part of SDS-PAGE then samples for those lanes were taken from consecutive chromatography fractions.*

The pRb was functional as proven by its interaction with the HPV16 E7 full-length, the E7-short, E7 peptide and the SV40 large T antigen peptide. Gel filtration of an equimolar mixture of E7 (full-length), or E7-short, with pRb-AB both showed complex formation. In both cases proteins were eluted in a single peak (Fig. 5.7B and C). Binding between pRb and the large T antigen was demonstrated by gel filtration chromatography coupled with mass spectrometry (Fig. 5.10A). Next we checked the interaction between the small pocket of pRb (constructs pRb-ALB) and the bHLH domain of MyoD (MCNC) *in-vitro*. We observed no interaction *in-vitro* between the pRb-AB and the bHLH domain of MyoD (MCNC) using both gel filtration chromatography (Fig. 5.8A) and affinity chromatography pull-down assays (Fig. 5.9A)

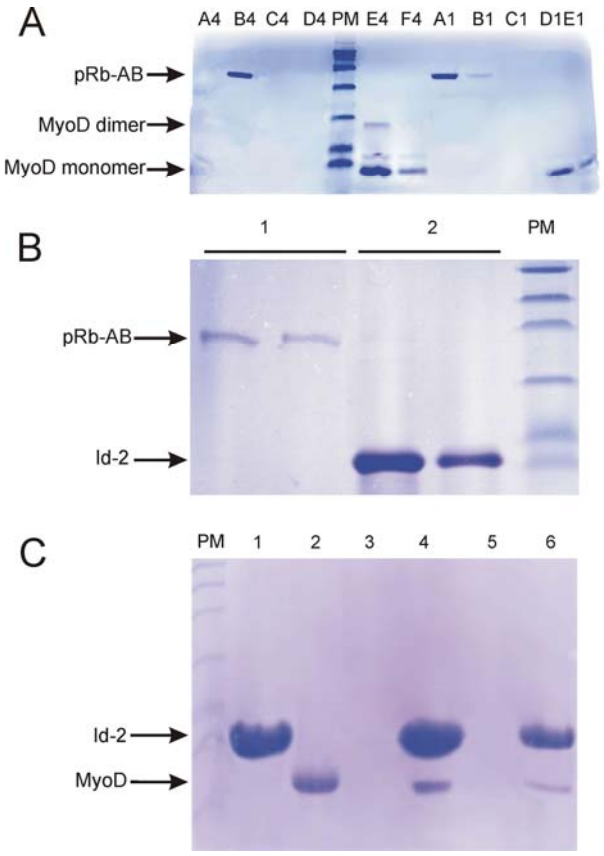

Figure 5.9. *SDS-PAGE of affinity chromatography pull-down assays. (A) the His-tagged MyoD (MCNC) (the bait) was bound to the Ni-NTA column and then pRb-AB (the prey) was passed through the column. A.4, MyoD flow-through; B4 and C4, pRb flow-through; D4, E4 and, F4, various fractions eluted with 250 mM imidazole. A.1, B.1 and C.1, pRb-AB flow-through (amount of MyoD decreased four times); D.1 and, E.1, 250 mM imidazole elutions. (B) His-tagged Id-2 (the bait) was bound to the Ni-NTA column and then pRb-AB (the prey) was passed through the column. 1- pRb-AB flow-through; 2 - 250 mM imidazole elution fractions. (C) His-tagged Id-2 was bound to the NiNTA column and then MyoD (MCNC) ('prey') was passed through the column. 1 - Id-2 alone, 2 - MyoD alone, 3,5 - MyoD flow-through, 4 - fraction eluted with 250 mM imidazole, 6 - fraction eluted with 250 mM imidazole (amount of Id-2 decreased four times), PM - is the protein marker.*

Mass spectrometry of concentrated fractions corresponding to the pRb peak in gel filtration chromatogram showed no MyoD (Fig. 5.10B). The same method clearly showed the presence of

the SV40 large T antigen peptide bound to pRb after the sample from gel filtration chromatography was analyzed by mass spectrometry (Fig. 5.10A).

An affinity chromatography pull-down assay in which MyoD (His-tagged MCNC) was the bait and pRb-AB the prey showed no binding (Fig. 5.9A). A reverse experiment gave the same results.

We observed no interaction between Id-2 and pRb using gel filtration chromatography (Fig. 5.8B). The lack of the interaction was further confirmed by the affinity chromatography pull-down assay, where Id-2 was the bait and pRb-AB the prey (Fig. 5.9B). Using the same affinity chromatography pull-down assay method we were able to show the binding between Id-2 and MyoD (MCNC) (Fig. 5.9C).

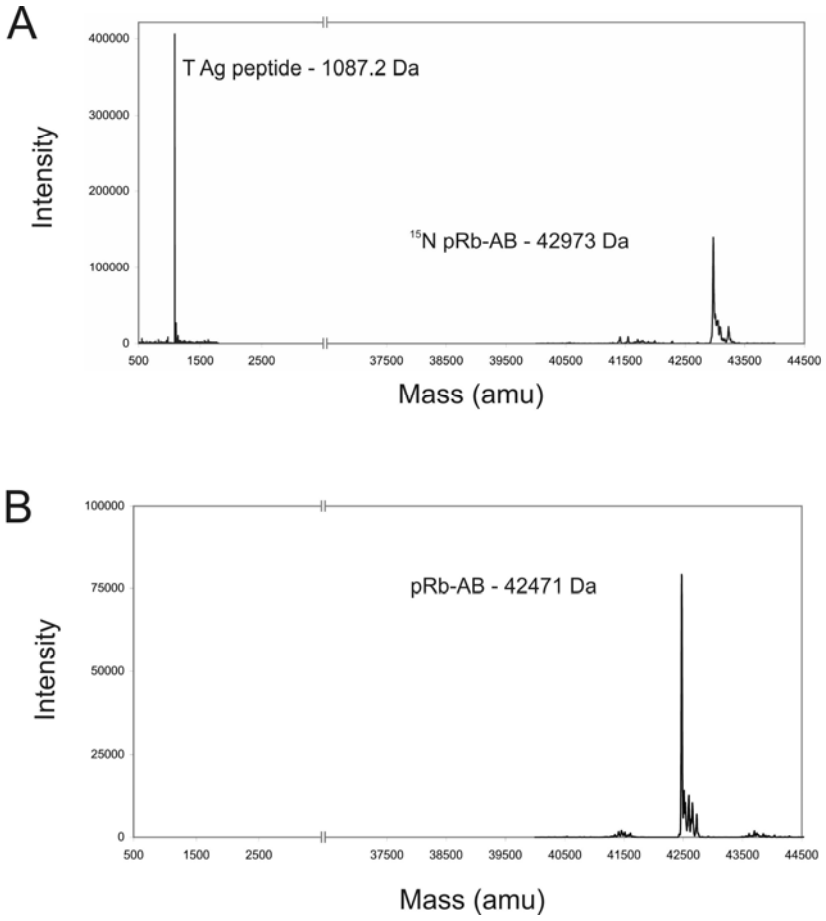

Figure 5.10. *Mass spectrometry results of: (A) the gel filtration chromatography purified sample of mixture 1:1 ^{15}N pRb-AB with the large T antigen peptide - both the protein and the peptide are detectable, indicating physical interaction between the two; (B) the sample of mixture 1:1 pRb-AB with the MyoD (MCNC) purified by gel filtration chromatography - the detectable protein corresponds to pRb-AB.*

It has been reported that MyoD, Id-2, and viral HPV16 E7 bind to the same region of pRb (Gu et al., 1993; Lasorella et al., 1996). We tested this interaction using 'selectively' ^{15}N labeled pRb-AB labeled with ^{15}N-Lys or ^{15}N-Leu. Addition of MyoD or Id-2 did not cause any changes

in ^{15}N-HSQC spectra, proving no *in-vitro* interactions (Fig 5.11A and C). In a positive control we were able to see clearly the interaction of the ^{15}N labeled pRb with HPV E7 peptides (Fig. 5.11B and D).

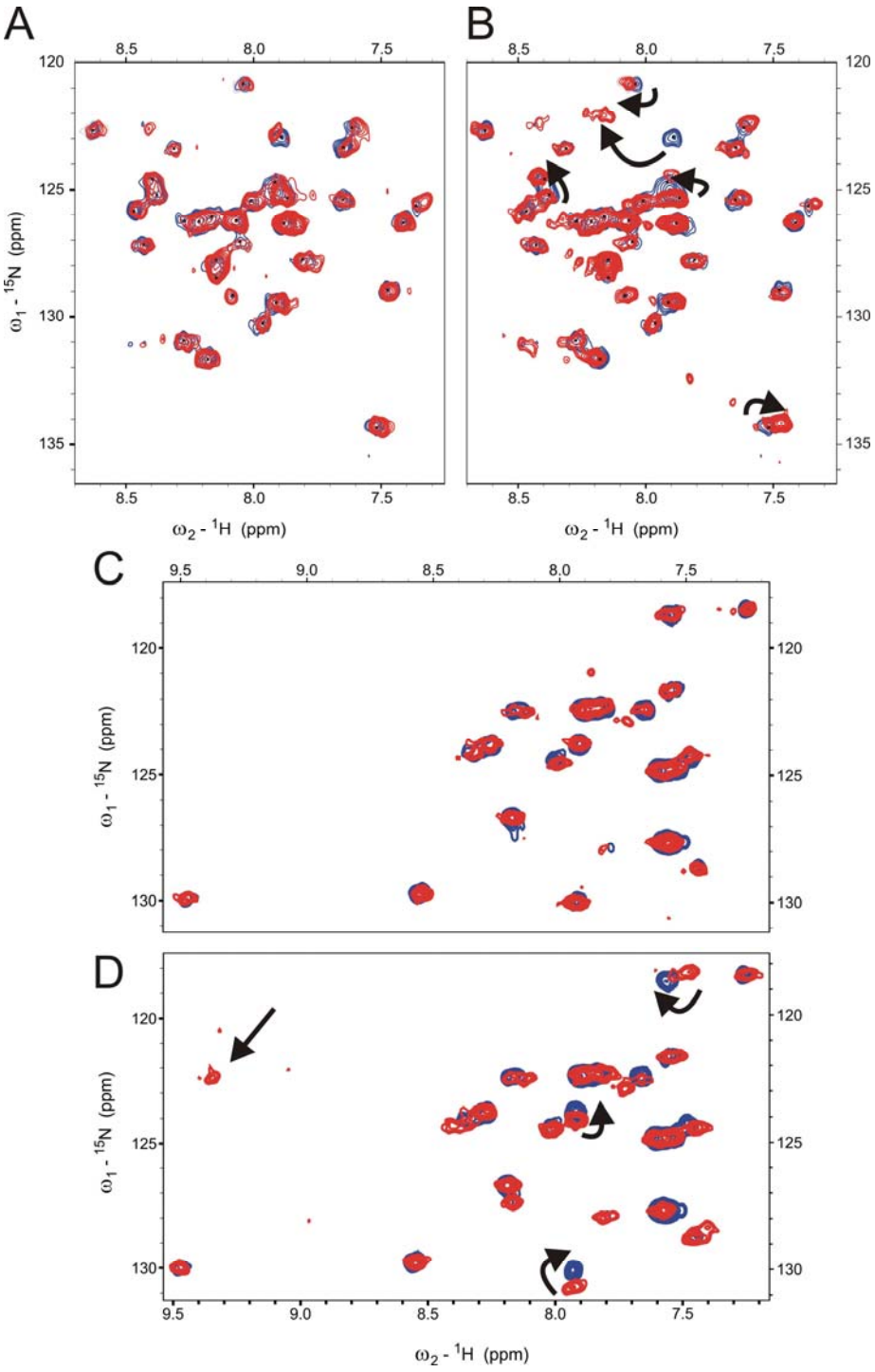

Figure 5.11. *Titration of the ^{15}N-labeled pRb-AB with Id-2 and MyoD (MCNC). The ^{15}N-HSQC spectrum in panel A of the selectively ^{15}N-Leu labeled pRb-AB mixed in equimolar ratio with Id-2 (red); in panel B ^{15}N-Leu labeled pRb mixed in equimolar ratio with the E7-1 peptide (red); in panel C ^{15}N-Lys labeled pRb-AB mixed in equimolar ratio with MyoD (MCNC) (red); in panel D ^{15}N-Lys labeled pRb-AB mixed in equimolar ratio with E7 peptide (red); pRb alone is shown in green in all panels. Arrows show cross-peaks that move. All spectra were recorded at pH 7.8.*

5.1.4 Interaction between pRb and LXCXE sequences

In this part of the thesis I document the results of the interaction of pRb-AB with LXCXE sequences from a number of proteins, which included viral oncoproteins HPV E7 and SV40 large T antigen, HDAC1, PAI2, gankyrin, BRG1, and pRb-N (the N-terminal part of pRb having a LXCXE like sequence). Table 5.3 lists the various LXCXE peptides or protein domains used in this work.

Table 5.3 *ITC and NMR titration results data for small pocket of pRb and various peptides/protein domains. K_D values were determined by NMR in case of fast exchange or weak binders. First column shows the various peptides or protein domains used in this study and second column shows the LXCXE sequence in the peptides or protein. The Leu (Ile), Cys, and Glu residues in the LXCXE motif are highlighted with red color and residues in blue were mutated*

Interaction of pRb-AB with:	***Peptide sequence Position P*** ------- +++++ ...54321012345...	***ITC K_D (μM)***	***NMR K_D (μM)***
HPV E7 peptide	DLYCYEQLN	0.19 ± 0.07 (0.11) [a]	tight binding
HPV E7 L(P-2)I peptide	DIYCYEQLN	0.32 ± 0.07	-[b]
HPV E7 L(P+4)F peptide	DLYCYEQFN	0.14 ± 0.03	-
HPV E7 D(P-3)R peptide	RLYCYEQLN	1.05 ±0.04	tight binding
HPV E7 D(P-3)R,Y(P+1)E peptide	RLYCEEQLN	4.79 ±0.20	-
T antigen peptide	NLFCSEEMD	0.44 ± 0.06	tight binding
HDAC1 peptide (17mer)	DKRIACEEEFSDSEEEG	10.0 ± 3.0	40.0[c]
HDAC1 large peptide (20mer)	SSDKRIACEEEFSDSEEEGE	10.0 ± 3.0	-
HDAC1 short peptide	RIACEEEFS	20.0 ± 4.0	50.0 [c]
HDAC1 F(P+4)L peptide	RIACEEELS	no binding	50.0 [c]
HDAC1 R(P-3)D peptide	DIACEEEFS	10.0 ± 5.0	8.0 [c]
HDAC1 R(P-3)D,E(P+1)Y peptide	DIACYEEFS	3.27 ± 0.07	no binding
E2F-1 large peptide	LDYHFGLEEGEGIRDLFD	1.24 ± 0.20 (0.49) [a]	tight binding
E2F-1 short peptide	GLEEGEGIR	no binding	no binding
PAI2 peptide	FLECAEEARKKINSWVK	no binding	no binding
pRb-N domain	..IEVLCKEHE..[d]	no binding	-
Gankyrin	..LHLACDEERVEEA..[d]	no binding	no binding
BRG1 LXCXE domain	..ERLTCEEEEEK..[d]	no binding	-
pRb-AB (K713A)[e] + E7 peptide	DLYCYEQLN	0.21 ± 0.04	tight binding
pRb-AB (C706F) [e] + E7 peptide	DLYCYEQLN	0.044 ± 0.02	-

[a] The data from Lee *et al*, (1998) are in parentheses.
[b] -, not determined.
[c] Specific values refer to the complexes in fast exchange between binding components, the estimated error is 10%.
[d] Protein domains having LXCXE sequences rather than peptides are used in these experiments.
[e] pRb-AB mutants are used in these experiments.

5.1.4.1 pRb and L/IXCXE peptides from HPV E7, SV40 large T antigen, and HDAC

The LXCXE sequences of HPV E7 and the SV40 large T antigen bind tightly to the small pocket of pRb, while the IXCXE sequence of HDAC1 associates transiently. We used the following peptides in the first round of our study: the LXCXE peptides from HPV E7, SV40 large T antigen, and PAI2 proteins; and the IXCXE peptides from HDAC1 and a peptide derived from the transactivation domain of E2F-1. Also LXCXE-containing fragments of BRG1 (NVKE construct) and the N-terminal of pRb were used (Table 5.3). We first measured the binding of these various peptides to pRb-AB using ITC. The equilibrium dissociation constant, K_D, for the binding between the pRb pocket and IXCXE peptide of HDAC1 (17mer) was difficult to measure but could be estimated to be ca. 10 μM (Fig. 5.12A and Table 5.3), which is weaker compared to the binding of LXCXE containing peptides from HPV E7 or SV 40 large T antigen (Fig. 5.12 B and C) and the peptide derived from the transactivation domain of E2F-1 (Table 5.3). Three HDAC1 peptides were tested: the HDAC1 peptide of Table 5.3 (17mer), a shorter version, RIACEEEFS, and a peptide extended by two N- and one C-terminal residues, SS and E, respectively (20mer). ITC showed the pRb-AB binding for all three peptides but for the later two quantatification of the data was difficult to obtain.

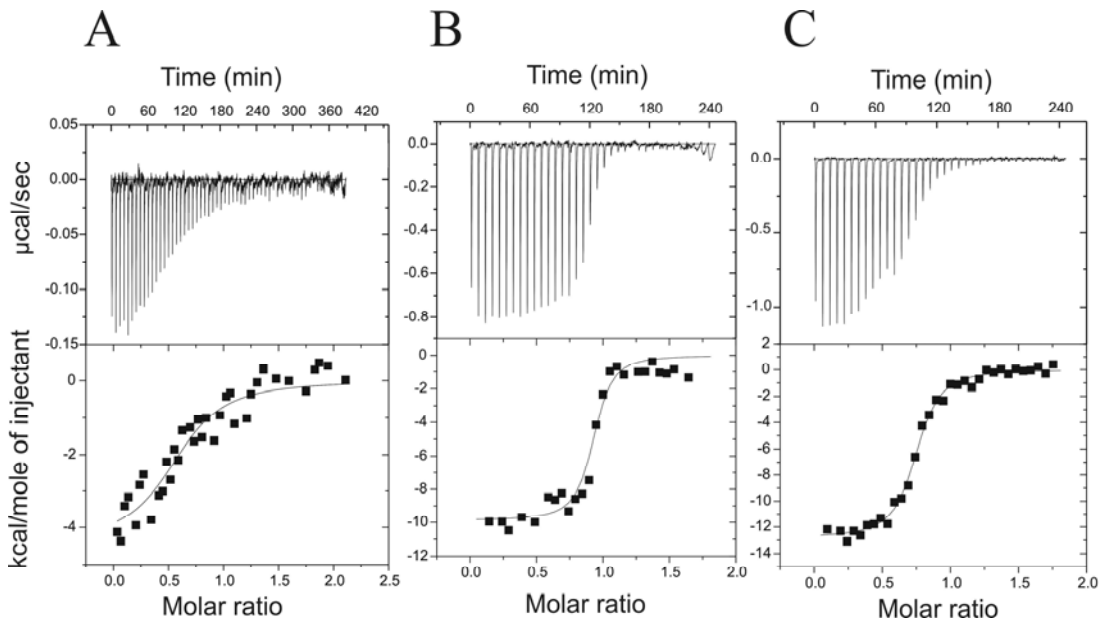

Figure 5.12. *ITC titration curve for various LXCXE peptides with the pRb-AB. (Upper) The raw data of ITC experiment each performed at 20-22 °C. (Lower) The integrated heat changes, corrected for the heat of dilution, and fitted curve based on a single site model. (A) pRb-AB titrated with HDAC1 peptide (17mer). (B) pRb-AB titrated with HPV E7 peptide. (C) pRb-AB titrated with Large T antigen peptide.*

We also checked the strength of this binding using mass spectrometry. The pRb-AB and the IXCXE HDAC1 peptide were mixed in 1:3 molar ratios and passed through a S75 analytical

Superdex column. A mass corresponding only to pRb-AB was detected in the eluent. The same procedure was used for the LXCXE peptide of large T antigen and in this case the masses of both the protein and the peptide were observed (Fig. 5.10A; section 5.1.3.2 and Table 5.4).

We next used NMR spectroscopy to check the binding. NMR measurements consisted of monitoring changes in chemical shifts and line widths of the backbone amide resonances of ^{15}N-enriched pRb samples upon addition of unlabeled peptides (Meyer and Peters, 2003; Pellecchia et al., 2002). As the assignment of the NMR spectra was not possible up to now (and not intended), NMR resonances influenced by binding to known binding partners (E7 peptide, E2F peptide, T Ag peptide) were identified (only K713 has been assigned using an ^{15}N lysine labeled sample of the pRb-AB K713S mutant) (Fig. 5.13).

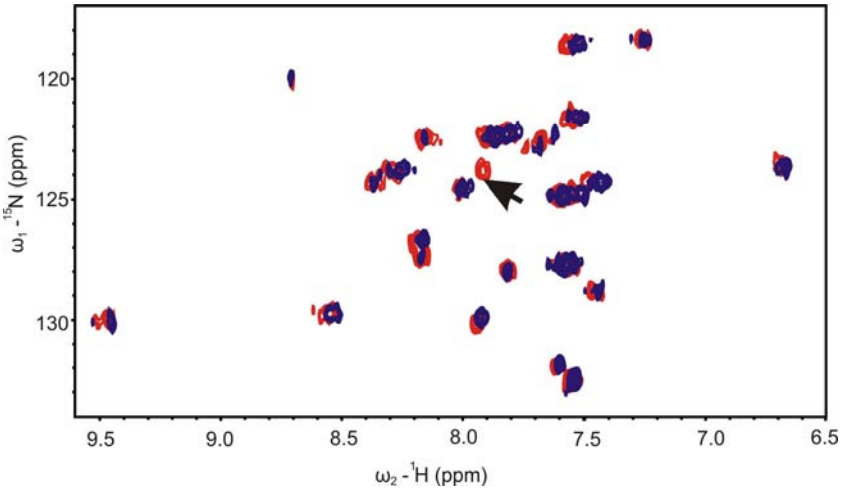

Figure 5.13. *Assignment of Lys713. Overlay of ^{15}N HSQC spectra pRb-AB (red) and pRb-AB (K713S) mutant (blue), all cross peaks except one are present in both spectra. The missing resonance peak corresponds to lysine 713.*

From the published crystal structure of the complex between the pRb small pocket and the HPV E7 peptide, and from a number of biochemical and biophysical assays, it has been shown that several lysine residues of pRb participate in the LXCXE-pRb interface (Kim et al., 2001; Lee et al., 1998; Chan et al., 2001; Dahiya et al., 2000; Chen and Wang, 2000; Dick and Dyson, 2002; Kennedy et al., 2001). In particular, the crystal structure of the pRb/E7 peptide shows that there are four lysine residues, which are close to and make contact with the E7 peptide backbone, viz. K713, K720, K722 and K765; beside these lysines there are two more lysine residues near to the LXCXE binding site namely, K729 and K740. All these lysine residues are situated in the B domain of pRb (Dick and Dyson, 2002; Brown and Gallie, 2002) (Fig. 5.14). We have therefore prepared an ^{15}N lysine labeled sample of pRb-AB and carried out titration with this sample (the HSQC spectrum of the ^{15}N uniformly labeled pRb-AB contained resonances from 361 residues

and was too crowded to be of practical use in these type of experiments (discussed in section 5.1.5). The ^{1}H-^{15}N HSQC spectrum of ^{15}N lysine labeled pRb-AB exhibited 27 cross peaks corresponding to the total of 27 lysines present in pRb-AB (Fig. 5.15).

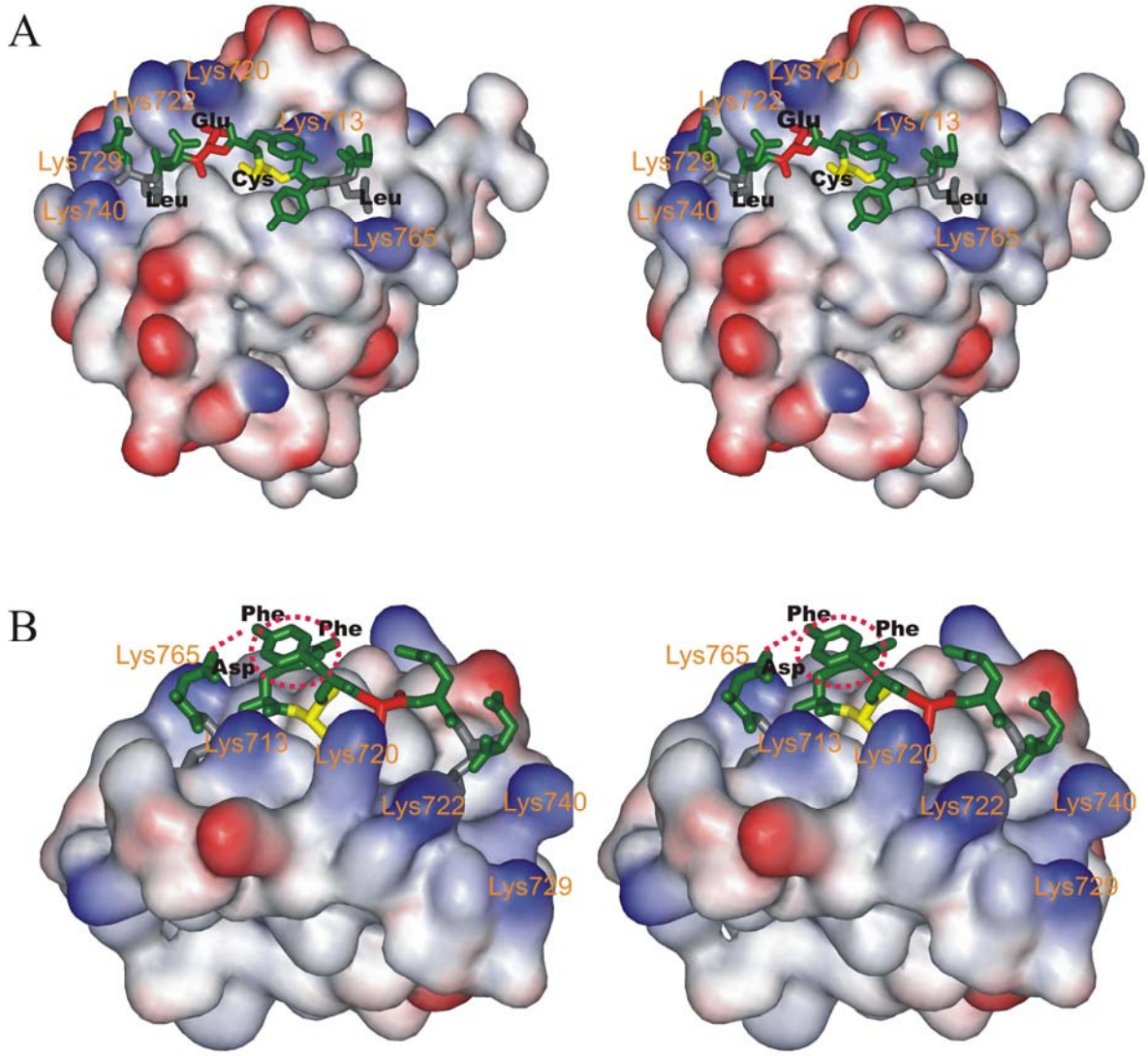

Figure 5.14. *(A) Surface of the B domain of pRb bound to the LXCXE peptide from HPV E7 (taken from the PDB file 1GUX (32). Blue color indicates positive charge and red color indicates negative charge. Lysine residues in the binding site and the residues of the LXCXE E7 peptide are highlighted. (B) Intramolecular interactions among the amino acid side chains. Two tyrosine aromatic rings stack with each other (dashed circle), while the hydroxyl group of tyrosine (P+1) makes a long distance hydrogen bond with the aspartic acid side chain at position (P-3)(dashed line).*

The positive control experiment constituted titration of the ^{15}N-Lys pRb-AB with the unlabeled E7 peptide. As can be seen in Figure 5.15A, five lysines exhibited changes in position and intensities of peaks upon complex formation with the E7 peptide. These induced shifts are proportional to the strength of the intermolecular interaction to a first approximation and therefore these lysine residues form the major E7-binding region of pRb-AB (Meyer and Peters, 2003; Pellecchia et al., 2002; Schnuchel et al., 1993). NMR spectra showed that the

E7 peptide/pRb-AB complex was long-lived on the NMR chemical shift time scale (Wüthrich, 1986, see also Methods) as two separate sets of ^{1}H-^{15}N HSQC resonances were observed in the intermediate stages of titration, one corresponding to free pRb-AB and the other to the pRb-AB bound to the E7 peptide (Fig. 5.15A and B). This is in agreement with the ITC data that showed a K_D of 0.22 μM. A similar result was obtained with the LXCXE peptide derived from the SV40 large T antigen. In the control experiment for no binding, we used a truncated version of the E2F-1 short peptide, Table 5.3, GLEEGEGIR, for which ITC showed no binding. Addition of this peptide did not cause any changes in positions of cross peaks of the ^{1}H-^{15}N HSQC spectrum within the NMR chemical shift differences of 0.01 ppm and 0.05 ppm in ^{1}H and ^{15}N dimensions, respectively.

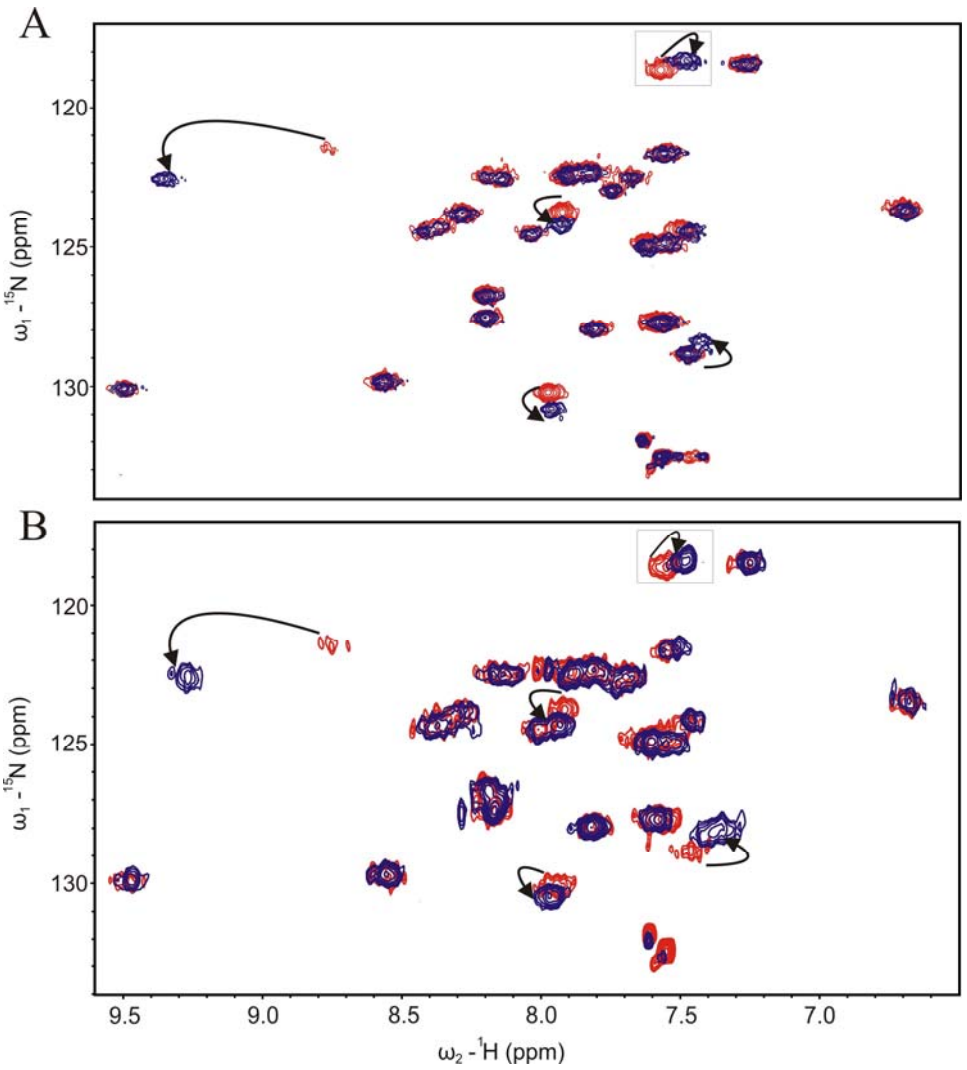

Figure 5.15. *(A) The ^{15}N lysine labeled pRb-AB was titrated with increasing concentrations of the E7 peptide and ^{1}H-^{15}N HSQC spectra were recorded at each step of ligand addition. Only two steps of titrations are shown: the starting (red, untitrated) and the final step (blue , at an E7 peptide/pRb-AB molar ratio of 1.60). Upon addition of E7 peptide five resonances were perturbed (arrows), indicating that these lysines are involved in the binding to pRb-AB, including the resonance peak from Lys713 (at the*

chemical shifts for ^{1}H 7.91 ppm, and for ^{15}N 123.9 ppm; the assignment of this peak is described in the S Figure 5.13). (B) A fresh sample of the ^{15}N lysine specifically labeled of pRb-AB was titrated with the IXCXE peptide from HDAC1. Only two steps of titrations are shown, red the starting (untitrated) and blue the final (at HDAC1 peptide/pRb-AB molar ratio of 3.76). Features of the cross peak (marked box, at 118.75 ppm, ω_1-^{15}N; 7.56 ppm, ω_2-^{15}H chemical shift values) in (A) and (B) are described in detail in Figure 5.16.

Figures 5.15B and 5.16B show titration of the ^{15}N-Lys pRb-AB with the HDAC1 peptide. Again five resonances are influenced and the changes were seen at the same pRb residues that were affected by the E7 peptide binding (Fig. 5.15A), indicating that both peptides bind to the same site on pRb-AB. However, the binding dynamics is different: the HDAC1 peptide/pRb-AB complex showed a continuous movement of five NMR peaks upon addition of increasing amounts of the peptide (Fig. 5.16B). This indicates that binding of the peptide to the pRb-AB is in fast exchange, which indicates weak binding. We calculated the K_D from the NMR titrations to be 40 ±10 µM at 27 ^{o}C, which agrees well with an approximate value estimated from ITC (10 ± 3 µM at 20 ^{o}C, see Methods). For the smaller version of the HDAC1 peptide, the changes in the chemical shift values were similar but smaller.

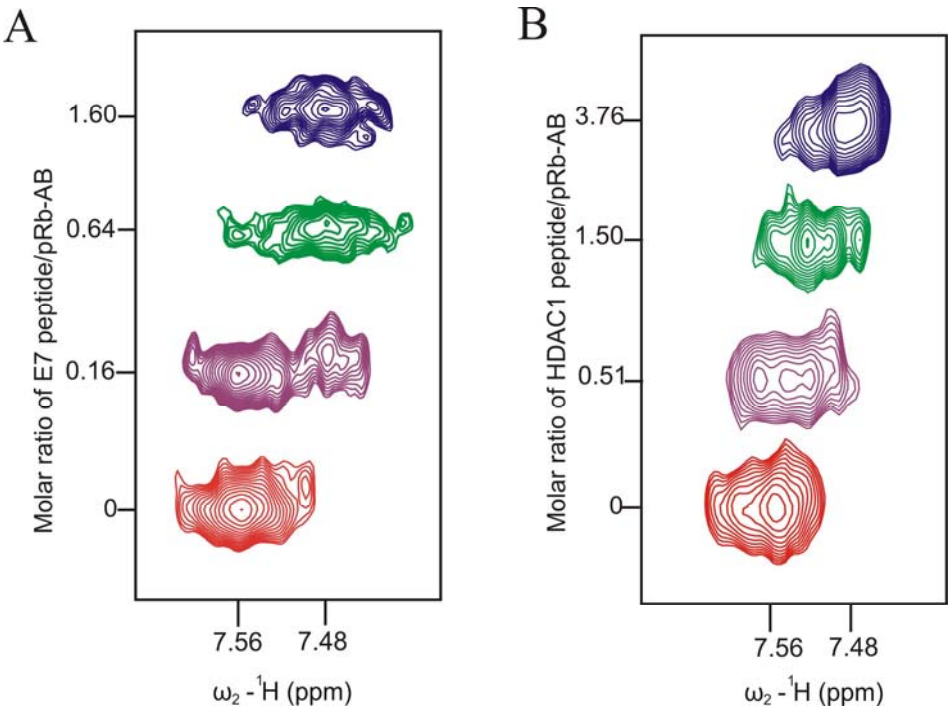

Figure 5.16. *The cross peak (at 118.75 ppm ω_1-^{15}N; 7.56 ppm ω_2-^{1}H, Figure 5.15) was chosen to demonstrate the difference between tight (A) (E7 peptide/pRb-AB) and weak (B) (HDAC1 peptide/pRb-AB) binding. Four steps of titrations are shown here at the specified peptide/pRb-AB molar ratio*

5.1.4.2 The molecular basis of the LXCXE interaction with pRb-AB

We then address the question why the IXCXE sequence from HDAC1 binds weaker to the pRb-AB compared to the LXCXE sequences from the E7 or T antigen. In order to determine which residues are important for bind to the LXCXE motif, we systematically designed various mutants of the basic sequences of E7 and HDAC1 (Table 5.3). We introduce here a nomenclature in which the cysteine residue of the LXCXE motif is denoted as the P0 site and residues on the target peptide N-terminal to this are denoted P-1, P-2 etc., and C-terminal P+1, P+2 etc (Table 5.3).

The first small but clear difference between LXCXE-like sequences of HDAC1 and HPV E7 (or SV40 large T antigen) is the presence of isoleucine at the P-2 position instead of conserved leucine. ITC titrations showed that the E7 peptide with Leu to Ile mutations also binds to the pRb-AB with approximately same affinity (K_D 0.32 ± 0.05, affinity lowered by only 1.68 times) (Table 5.3 and Fig. 5.17A). This result shows that this change is not significant for the binding.

Structures of complexes of the [pRb-AB]-[E7 LXCXE peptide] or the [pRb-AB]-[domain of SV40 large T antigen] reveal that the side chain of residues outside the minimal LXCXE sequence - leucine in position P+4 (DLYCYEQ*L*N) in HPV E7 and methionine (NLFCSEE*M*P) in SV40 large T antigen - also make strong hydrophobic interactions to pRb-AB (Kim et al., 2001; Lee et al., 1998). There is a larger phenylalanine residue in the HDAC1 at this position (RIACEEE*F*S). In order to check whether this substitution significantly affects binding, we used variants DLYCYEQ*F*N from E7 [HPV E7 L(P+4)F] and RIACEEE*L*S of HDAC1 [HDAC1 F(P+4)L], in which the E7 wild type Leu at position P+4 was changed to Phe and the HDAC1 wild type Phe at position P+4 was changed to Leu, respectively. ITC titration showed that the HPV E7 L(P+4)F peptide still binds strongly (Table 5.3 and Fig. 5.17B). The HDAC1 F(P+4)L peptide did not show any binding in ITC and NMR titration showed that the HDAC1 F(P+4)L peptide binds weakly to the pRb-AB (Table 5.3). These results show that the change of Leu to Phe is not a determinant for the binding, although we believe that the residue at this position should be hydrophobic, based on the crystal structures.

For three major interactions of L, C, and E in the minimal LXCXE, the side chains of the conserved leucine and cysteine fit tightly to the hydrophobic groove on pRb-AB, whereas the carboxylate group of glutamic acid of the terminal LXCXE makes two hydrogen bonds with two backbone amide groups of helix-15 of pRb-AB. As mentioned earlier, a lysine patch surrounds the LXCXE binding site groove on the B half of the AB pocket (Fig. 5.14A and B). Six lysine residues that surround this binding cleft are K713, K720, K722, and K729. K740 and K765. We

speculated that this highly positive surface on the LXCXE binding cleft would repel any positive residue in the LXCXE sequences or the flanking residues. The first confirmation of this hypothesis came from the observation that the longer peptide from HDAC1 bound better then the short peptide (Table 5.3). The longer peptide differs from the short one by having a number of extra negatively charged residues in the C terminal segment (Table 5.3). In viral oncoproteins such as HPV E7 or SV 40 large T antigen, the LXCXE motif is followed by a stretch of negatively charged residues. It has been proposed that these acidic residues may interact with the lysine patch. However, the [pRb-AB]-[domain of SV40 large T antigen] structure showed that this stretch is flexible, with relatively few interactions that were further proposed to be driven by the phosphorylation of pRb (Kim et al., 2001). The second substantiation of the hypothesis of repulsion via positive flanking residues came from the observation that the peptide from the PAI2 protein did not show any binding to pRb-AB (Table 5.3). The PAI2 sequence has positively charged residues at the C terminus of the peptide. We also confirmed the lack of interaction between pRb-AB and PAI2 protein or pRB-large pocket (amino acids 379-928) and PAI2 protein using gel filtration chromatography.

We designed two single mutant peptides, *R*LYCYEQLN, D(P-3)R peptide (Asp at P-3 changed to Arg), from HPV E7 and *D*IACEEEFS, R(P-3)D peptide (Arg at P-3 changed to Asp) from HDAC1 (Table 5.3). The ITC results showed that the HPV D(P-3)R peptide binds to the pRb-AB with 5.5 fold reduced affinity (Fig. 5.17C), and the HDAC1 R(P-3)D peptide also showed better binding compared to the normal peptide. ITC and NMR suggested K_D values of ~10 μM and 8 μM, respectively, the lowest values among all HDAC1 peptides so far. However, complex formation between pRb and the peptide was still in the fast exchange regime (Table 5.3). These results clearly show that any positive residue in or around the LXCXE sequences reduces the binding affinity to the pRb-AB.

The crystal structure shows that the E7 peptide binds to the pRb-AB in an extended conformation. There are interactions among the side chains of the peptide residues. In the E7 peptide *D*L*Y*C*Y*EQLN, the side chain aromatic ring of Tyr at P+1 makes a stacking interaction with the side chain aromatic ring of Tyr at P-1, and also makes a long distance hydrogen bond with the side chain of Asp at P+3 position (Fig. 5.14B)(Lee at al., 1998). Since the residues at the corresponding positions in HDAC1 differ significantly (*R*I*A*C*E*EEFS), we decided to design two peptides, each with two substitutions compared to wild type sequences: *R*LYC*E*EQLN [from HDAC1 with substitutions D(P-3)R and Y(P+1)E] from HPV E7 and *D*IAC*Y*EEFS, [from HPV E7 with substitutions R(P-3)D and E(P+1)Y], see Table 5.3. ITC measurements showed that the HPV E7 D(P-3)R,Y(P+1)E peptide binds to pRb-AB with a 25 fold reduced affinity (K_D = 4.79

±0.20 μM) compared to wild type E7 peptide and that the HDAC1 R(P-3)D,E(P+1)Y peptide also binds with relatively high affinity (K_D = 3.27 ± 0.07 μM) (Table 5.3 and Fig. 5.17D and E). We also checked these interactions using gel filtration chromatography followed by mass spectrometry. In case of the pRb-AB/HPV E7 D(P-3)R,Y(P+1)E peptide interaction, a mass corresponding only to pRb-AB was detected, but for the pRb-AB/ HDAC1 R(P-3)D,E(P+1)Y peptide interaction, both protein and peptide masses were detected (Table 5.4). These results demonstrate clearly the importance of intramolecular interactions between the peptide amino acid side chains and pRb.

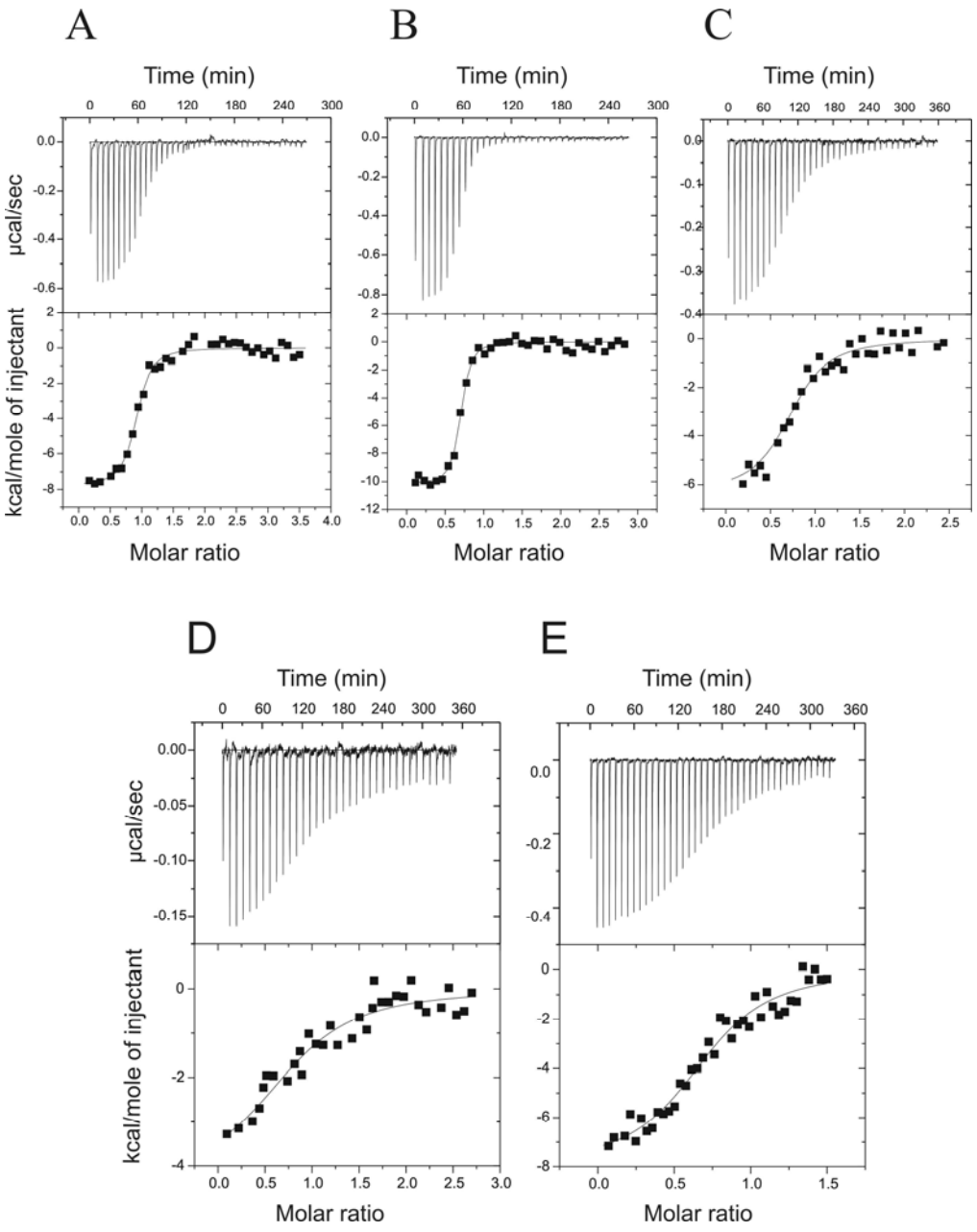

Figure 5.17. *ITC titration curve for the variant LXCXE peptides and pRb-AB. (A) pRb-AB titrated with E7 L(P-2)I peptide. (B) pRb-AB titrated E7 L(P+4)F peptide. (C) pRb-AB titrated with E7 D(P-3)R peptide.*

(D) pRb-AB titrated with HPV E7 D(P-3)R,Y(P+1)E peptide. (E) pRb-AB titrated with HDAC1 R(P-3)D,E(P+1)Y peptide.

Table 5.4. *Gel filtration followed by mass-spectrometry results. pRb-AB and peptides were mixed in 1:3 to 5 molar ratio, respectively. These complexes were then run on the Superdex S75 analytical gel filtration column. Eluent peak corresponding to the complex was collected and analyzed by mass spectrometry*

Interaction of pRb-AB with	pRb-AB (Da)	Peptide (Da)
Large T antigen peptide	42973.0[a]	1087.3
HPV E7 D(P-3)R,Y(P+1)E peptide	42467.0	-[b]
HDAC-1 peptide (17mer)	42467.0	-
HDAC-1 short peptide	42469.0	-
HDAC-1 R(P-3)D,E(P+1)Y peptide	42467.0	1076.6

[a] Mass detected corresponds to the uniformly ^{15}N labeled pRb-AB used in this case.
[b]- No mass corresponding to the peptide mass was detected.

5.1.4.3 Interaction between pRb and gankyrin, pRb-N and BRG1 LXCXE domain

The BRG1 protein, which was reported to interact with the pRb-AB, has an LXCXE motif in its sequence (…EVER*L*T*C*E*E*EEEK…) (Dunaief et al., 1994; Strober., 1996; Brown and Gallie, 2002). There is an LXCXE motif (...RIIE*V*L*C*K*E*HE…) present in the N terminal domain of pRb, however, no data in the literature could be found about binding of the N-terminal pRb to the pRb-AB small pocket. Gankyrin protein was shown to associate with the pRb-AB by the LXCXE motif in its sequences (…LH*L*A*C*D*E*ERVEEAK…) (Higashitsuji et al., 2000; Lozano and Zambetti, 2005)

All these proteins or domains have positive residues in or flanking the LXCXE sequence. In gel filtration chromatography under native conditions, we detected no complex formation between the BRG1 LXCXE domain (construct NVEK) and pRb-AB or the N-terminal pRb and pRb-AB or gankyrin and pRb-AB as all the protein eluted separately (Fig. 5.18A, B and C)). ITC titrations also showed no binding (Table 5.3). These results supplement the data presented in the previous sections, indicating that any positive charge around the LXCXE sequence are detrimental to the binding with the pRb. Gankyrin protein is helical with ankyrin repeats, whereas other two protein domain we checked was unfolded in solution. Folding of these proteins was checked by 1D-^{1}H NMR spectra (Fig. 5.19).

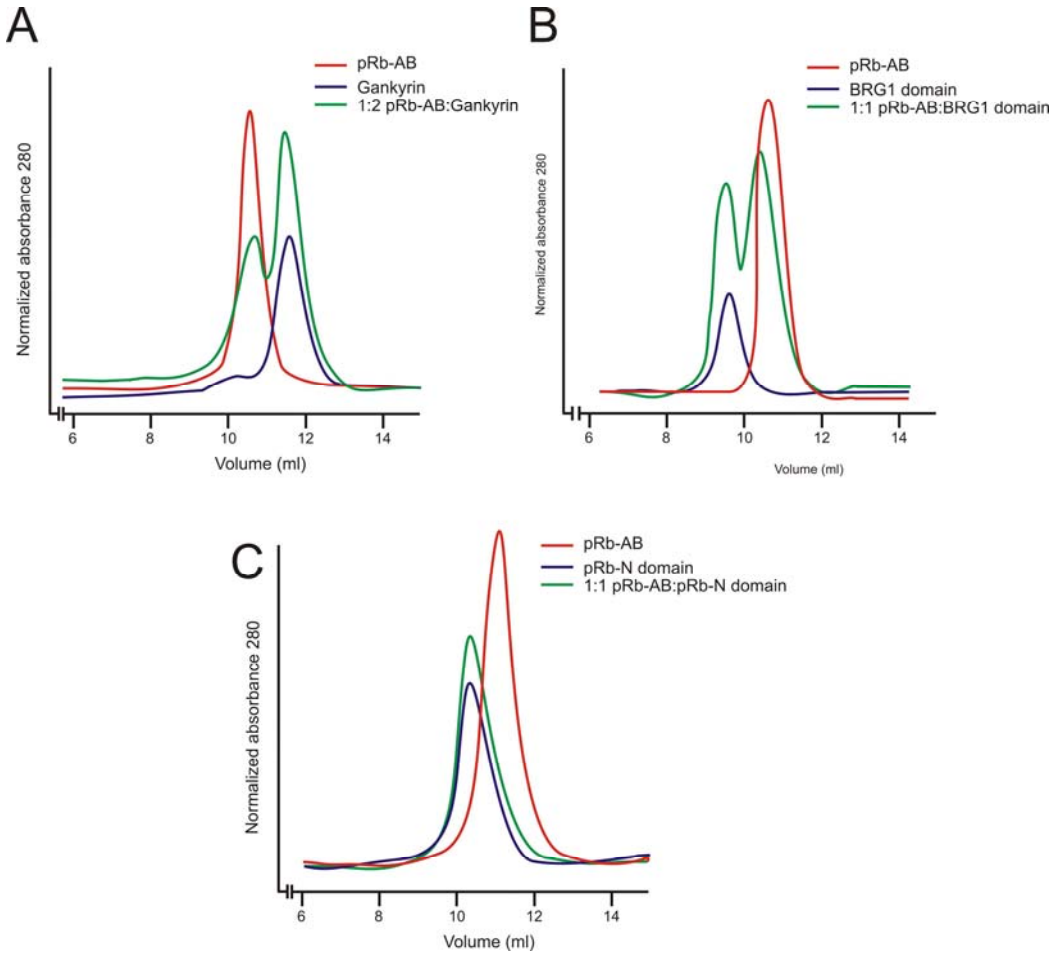

Figure 5.18. *Monitoring the formation of protein-protein complexes by gel filtration chromatography. (A) pRb-AB alone (red line), gankyrin alone (blue line), 1:2 molar ratio of pRb-AB: Gankyrin (green). No change in the shifts of the peak of either protein observed, as both proteins eluted separately. (B) pRb-AB alone (red line), A LXCXE containing domain from BRG1 (named NVKE domain) alone (as dimer; blue line), 1:1 molar ratio of pRb-AB: A LXCXE containing domain from BRG1 (green). No change in the shifts of the peak of either protein observed, as both proteins eluted separately. (C) pRb-AB alone (red line), pRb-N alone (as dimer; blue line), 1:1 molar ratio of pRb-AB: pRb-N (green). No change in the shifts of the peak of either protein observed, as both proteins eluted separately.*

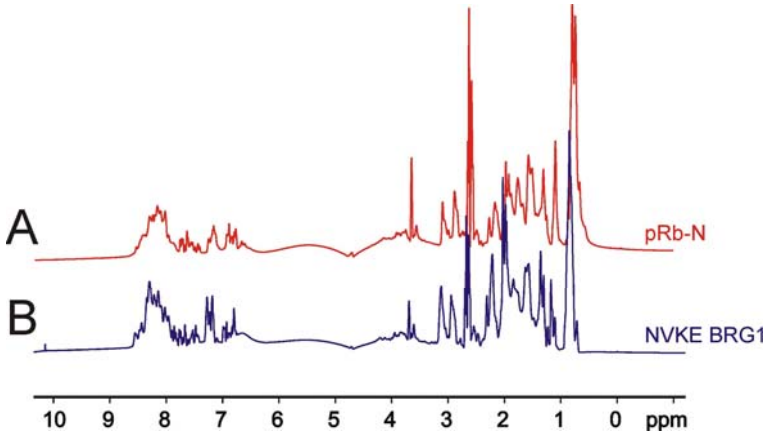

Figure 5.19. *1D-^{1}H spectra of (A) pRb-N and (B) NVKE BRG1 construct, showing the folding.*

5.1.4.4 Binding of the pRb-AB mutants to LXCXE sequences

We tested the binding of two pRb-AB mutants with the E7 LXCXE peptide. It has been reported that a pRb-AB mutant C706F is defective in its ability to bind to the LXCXE sequence containing viral proteins (Chan et al., 2001; Dahiya et al., 2000; Brown and Gallie, 2002; Dick et al., 2000; Kratzke et al., 1992). Another mutant, K713S, was also shown not to bind LXCXE sequences in one report (Chan et al., 2001), whereas other reports suggested the opposite (Chan et al., 2001; Brown and Gallie, 2002; Dick et al., 2000). Our binding experiments showed that both of these mutants bind tightly to the LXCXE peptide of HPV E7 (Table 5.3 and Fig. 5.20A and B). We have used NMR to check the structural integrity of the mutants (Pellecchia et al., 2002; Rehm et al., 2002). The NMR showed that these mutations do not disrupt the structural integrity of the small pocket as no noticeable differences were observed compared to the wild type pRb-AB pocket in the NMR spectra.

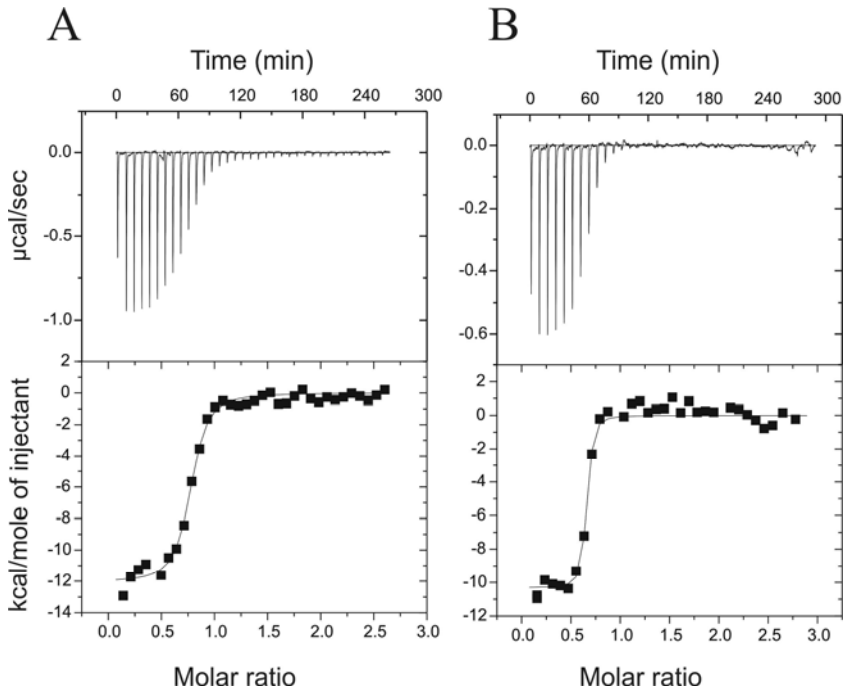

Figure 5.20. *ITC curves for pRb-AB mutants and E7 peptide. (A) pRb-AB (K713A) and E7 peptide. (B) pRb-AB (C706F) and E7 peptide.*

5.1.5 Perdeuteration of pRb-AB

Perdeuteration of pRb-AB was carried out with the objective of the backbone assignment of the protein. The protein sample was made as described in the Section 5.2.8. The level of perdeuteration was checked by mass spectrometry. The deduced mass of pRb-AB (from the protein sequence) ($^{2}H + ^{15}N$) is 45991.4 Da (we assume this mass for 100% perdeuteration). The mass obtained after the mass spectrometry analysis of perdeuterated pRb-AB is 45071.0 Da.

Therefore, the % of perdeuteration would be 98%. However we should keep in mind that some of the constituents of the minimal media were not deuterated, like for example, glucose which could account for ≠ 100% perdeuteration.

We recorded ^{15}N-^{1}H HSQC spectra of the perdeuterated pRb-AB on a 900 MHz spectrometer. There was a significant improvement of the HSQC spectrum but still the assignment was not attempted, although the peaks were sharper compared with the ^{15}N-^{1}H HSQC spectrum of protonated pRb-AB (Fig. 5.21A and B).

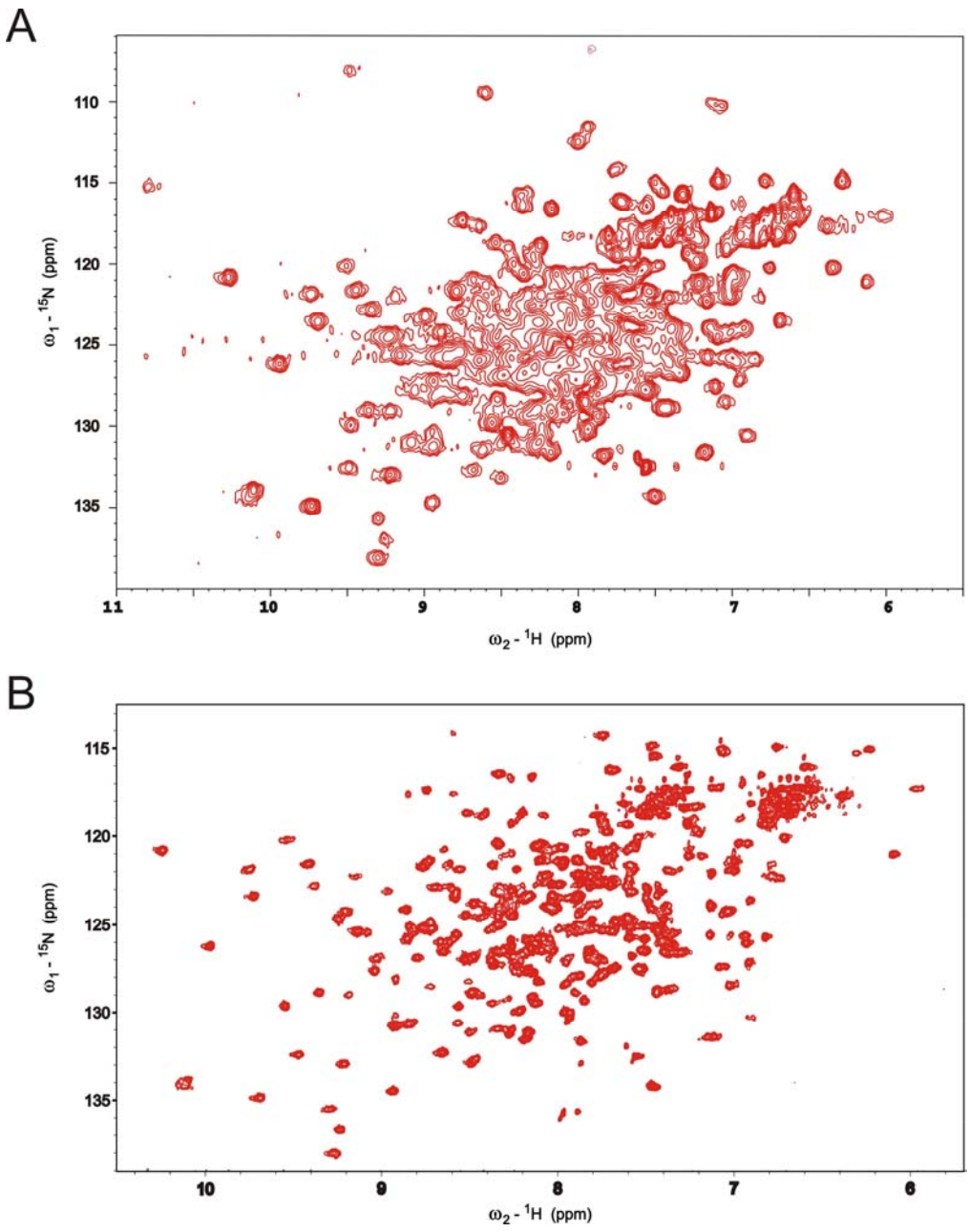

Figure 5.21. *^{15}N-^{1}H HSQC spectra of pRb-AB. (A) ^{15}N-^{1}H HSQC spectrum of protonated pRb. (B) ^{15}N-^{1}H HSQC spectrum of perdeuterated pRb-AB.*

5.1.5.2 Titration of perdeuterated pRb-AB and HPV E7

We titrated the perdeuterated pRb-AB with the E7 peptide, however the ^{15}N-HSQC spectra was too crowded to detect binding, although a number of peaks were perturbed. For large proteins of ca. 30 kDa, the HSQC spectra are normally too crowded to be of practical use in these types of experiments. The HSQC spectrum of the ^{15}N/^{2}H uniformly labeled pRb-AB contains NH resonances from 361 residues (except prolines) and is too crowded for detecting interactions with the pRb binding peptides. A remedy would be to prepare a ^{15}N selectively labeled protein with one amino acid type and obtain a less crowded NMR 'subspectrum'. Therefore we carried out our titration on selectively ^{15}N lysine labeled pRb-AB (discussed in the section 5.1.4)

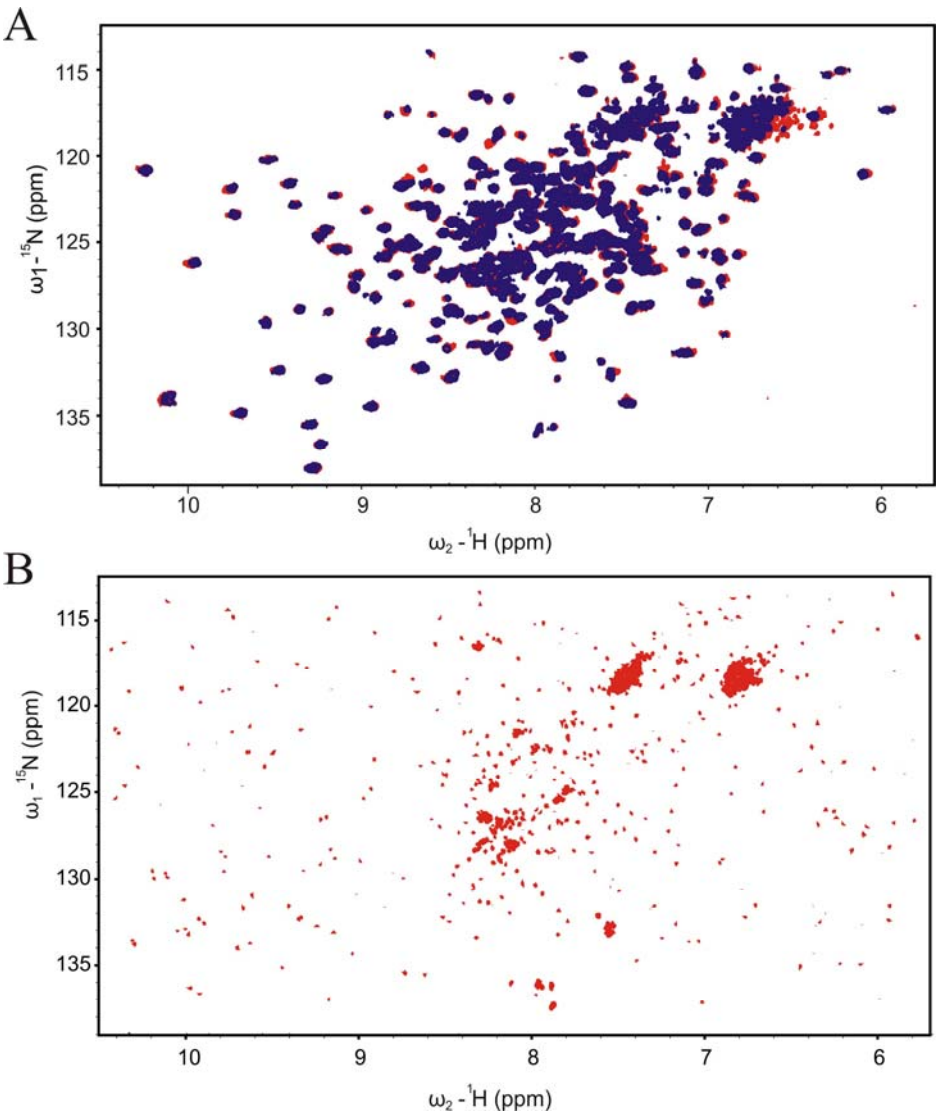

Figure 5.22. *(A) ^{15}N-HSQC spectra of the small pocket domain of [^{15}N, ^{2}H]-pRb. The reference spectrum is shown in red, and the one in blue is after titration with the E7 peptide. (B) On addition of the dimeric full-length E7 protein (104 residues for a monomer), the pRb spectrum disappears due to the formation of the complex.*

Figure 5.22A shows the spectra of ^{15}N-labeled pRb small pocket domain titrated with the HPV E7 peptide of 9 amino acids. Figure 5.22B shows that the complex formation E7-pRb can easily be detected with NMR if a larger fragment of the E7 protein is used. This happen because when the smaller protein binds to a larger one, the observed $1/T_2$ transverse relaxation rates of the bound protein in the complex increase significantly and broadening of NMR resonances results in the disappearance of most of the cross-peaks in the HSQC spectrum.

5.1.6 Crystallization of pRb-AB

I have undertaken the crystallization of pRb-AB with the objective to solve its structure and compare it with the previously published structures of pRb-AB with bound HVE E7 and E2F1/2 peptides. In addition I wanted to have the structure of the pRb small pocket domain with the IXCXE peptide from HDAC1. My attempts to co-crystallize pRb-AB with the IXCXE peptide from HDAC1 were not successful. However, I was able to crystallize pRb-AB in its unliganded form. Crystals grew in two conditions of Hampton screens: the Hampton CSI No. 4 and Hampton Cryo No. 9 (Fig. 5.23A-D). The composition of Hampton CSI No. 4 is 0.1 M HEPES; pH 8.5, 2.0 M ammonium sulfate, and the composition of Hampton Cyro No. 9 is 0.17 M ammonium sulfate, 0.085 M tri-sodium citrate dihydrate; pH 5.6, 25% w/v PEG 4000, 15% glycerol.

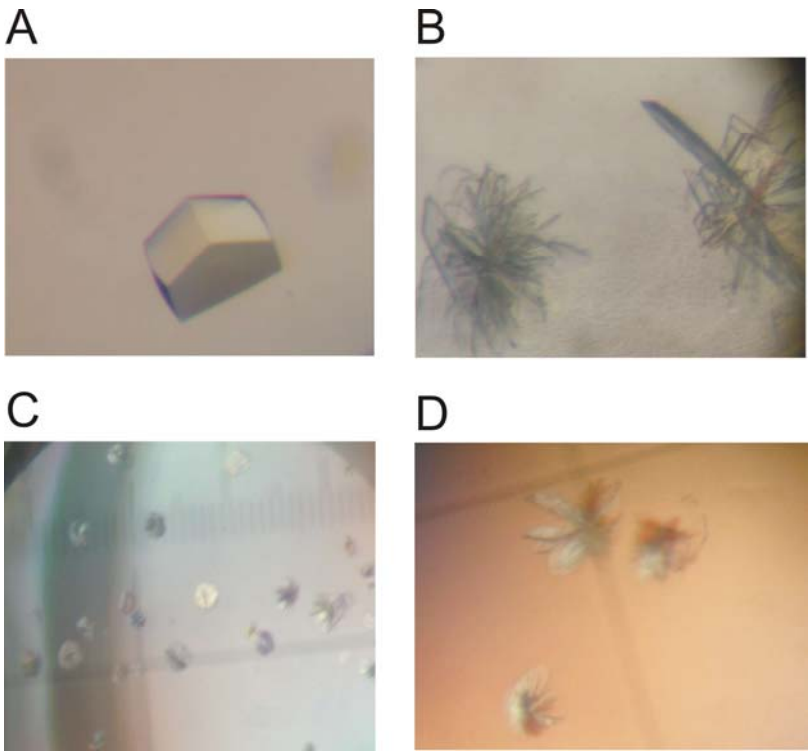

Figure 5.23. *Crystals of pRb-AB grown under different conditions. (A) Crystal grown in the Hampton CS1 No. 4 condition. (B – D) Different forms of crystals grown under the Hampton Cryo screen No. 9 and further optimized conditions of Hampton Cryo screen No. 9.*

One of the crystals out of the Hampton Screen I No. 4 diffracted up to 3.5 Å, however, unfortunately I could not repeat this crystallization condition (Fig. 5.24A). All attempts to optimize this condition were unsuccessful. Also, microseeding did not produce any crystals. The diffraction pattern for this form of the crystal is shown in Figure 5.24A and the preliminary data in Table 5.5.

Intertwined-like plate crystals were obtained from the Hampton screen Cyro No. 9. However, these crystals were very thin. Further optimization of the condition yielded crystals of a better shape and bigger sizes. We measured them on a synchrotron source in DESY Hamburg and they diffracted up to 3.5 Å, as well. I also soaked these crystals with the IXCXE peptide from HDAC1, however the crystals diffracted worse up to 4.0 Å.

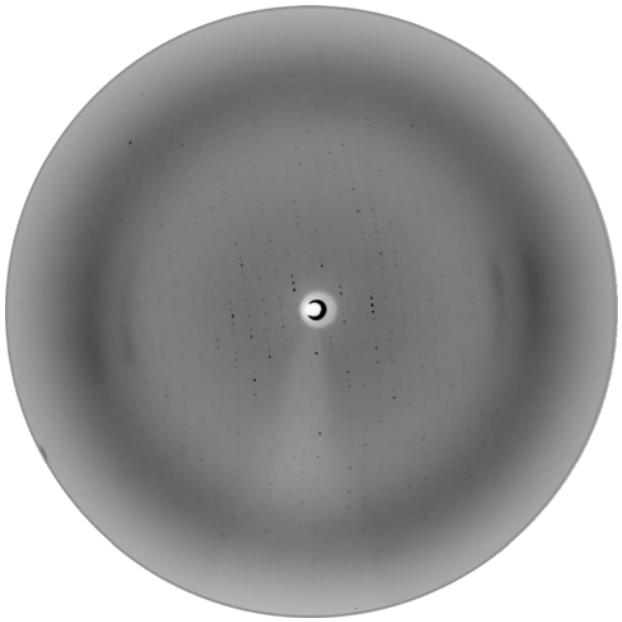

Figure 5.24. *Diffraction pattern of pRb-AB crystal. A frame from the MAR CCD165 image plate (BW6, DESY, Hamburg). Resolution at the edge of the detector is ca. 3.0 Å.*

Table 5.5. *Data collection the native pRb-AB crystal*

Data collection			
Space group	$P6_122$		
Unit cell dimensions (Å)	a= 80.590	b= 80.590	c= 257.040
(°)	α= 90.000	β= 90.000	γ=120.000
Resolution range (Å)	20-3.5		
Wavelength (Å)	1.05		

5.1.7 Discussion

5.1.7 pRb interacts with the viral oncoproteins but there is no interaction between pRb and MyoD/Id-2 proteins *in-vitro*

The bHLH class of proteins contains MyoD, myogenin, E12, E47, E2-5/ITF1, NeuroD/BETA2, and TAL (Massari and Murre, 2000). MyoD plays a key role in differentiation of all skeletal muscle lineages (Weintraub et al., 1991). Forced expression of MyoD in a variety of different cell backgrounds, including normal, transformed and tumor cell lines, was shown to induce growth arrest. This feature was correlated with the presence of the MyoD bHLH domain (Crescenzi et al., 1990). *In-vitro* protein binding and immunoprecipitation studies indicated that both MyoD and myogenin bind to the pocket domain of underphosphorylated pRb directly through the bHLH domain (Gu et a., 1993). Several other *in-vivo* studies however could find no evidence for such binary interactions suggesting that if there is any direct interaction it appears to be weak, making the significance of the *in-vitro* MyoD-pRb interaction questionable (Guo et al., 2003; Puri et al., 2001; Zhang et al., 1999a; Zhang et al, 1999b). Using the two-hybrid system Zhang et al. (1999a) could not detect direct MyoD or myogenin-pRb interactions.

Id-1 to Id-4 are another category of mammalian HLH proteins (Pagliuca et al., 19995). These proteins promote cell proliferation by binding to transcription factors from the bHLH subclass of proteins and inhibit their ability to bind DNA (Norton et al., 1998). Id-2 and MyoD have a high degree of sequence homology. Only Id-2, and not other Id proteins, was reported to disrupt the antiproliferative effects of pRb proteins via direct interactions (Lasorella et al., 1996). The HLH domain of Id-2 was shown to interact with pRb (small pocket domain), p107, and p130 (Lasorella et al., 1996; Iavarone et al., 1994).

To check whether these are direct *in-vitro* protein-protein interactions we used NMR spectroscopy, gel filtration chromatography coupled with mass spectrometry and affinity chromatography pull-down assay. We did not use the yeast two-hybrid system or immunoprecipitation, as these methods are known to give false positive results (Gietz et al., 1997; Toby and Golemis, 2001) we also excluded autoradiographic detection methods for the same reasons. The yeast two-hybrid system and immunoprecipitation are usually used to obtain a number of possible interaction partners and they form a good starting point for further experiment.

We could detect the interactions between MyoD and DNA or MyoD and Id-2 but there was no direct protein-protein interaction between MyoD (residues 99-173) or Id-2 (residues 2-132) and the small pocket domain of pRb (residues 379-791) *in-vitro.* We next used NMR to check for very weak (i.e. millimolar) interactions (Rehm et al., 2002). Standard molecular biology methods for

protein binding used in our study and also methods, such as for example ELISA, RIA, or Biacore fail to detect ligands with weak (i.e. millimolar) affinities. The ability of NMR to detect ligands that bind very weakly to target proteins has made NMR increasingly important in drug discovery and structural proteomics (Diercks et al., 2001; Meyer and Peters, 2003; Pellecchia et al., 2002; Stoll et al., 2001). NMR showed no association between these proteins.

Our results contradict earlier publications (Gu et al., 1993; Iavarone et al., 1994) but are in agreement with reports which show that MyoD influences the cell cycle as a transcription factor (Zhang et al., 1999a; Cenciarelli et al., 1999; Halevy et al., 1995; Magenta et al., 2003; Polesskaya et al., 2001). MyoD is widely accepted to be a transcription factor that can interact with p300 and CBP proteins, which are both histone acetyltransferases (Polesskaya et al., 2001). These interactions facilitate binding to a promoter sequence CRE. MyoD, together with CBP and CREB proteins, trigger transcriptions of pRb and p21 - both promoters of cell differentiation (Cenciarelli et al., 1999; Halevy et al., 1995; Magenta et al., 2003). It was also shown that augmented levels of p57, p27, and probably p21 cause an increased stability of MyoD (Reynaud et al., 1999; Zabludoff et al., 1998). These provide a positive feedback, important at the onset of myoblast differentiation. Since there is interplay between myogenesis and the cell cycle, it seems that the primary function of MyoD is in modulation of gene expression levels connected with both myogenesis and the cell cycle.

Genetic interactions between pRb and Id-2 during development have been reported (Lasorella et al., 2000). It is possible that Id-2 influences directly or indirectly expression levels of the pRb gene without the immediate protein-protein interaction with pRb. It has been shown that the bHLH protein E2A has the ability to block the G1 to S phase transition at least through transcriptional activation of the p21Cip1 gene. The Id proteins are able to inhibit the E2A mediated expression of the p21Cip1 gene (Prabhu et al., 1997). The p21Cip1 protein binds and inactivates CDKs and enhances the activity of pRb, thereby blocking the G1 to S phase transition. It is therefore possible that Id proteins mediate the G1 to S phase transition by down regulating the p21Cip1 gene expression.

5.2.2 Molecular determinants for the complex formation between the retionblastoma protein and LXCXE sequences

The small pocket of pRb binds to the LXCXE-like sequence containing proteins (Lee and Cho, 2002) and includes also a primary binding site for E2Fs (Flemington et al., 1993; Helin et al., 1992; Helin et al., 1993; Hiebert et a., 1992; Kaelin et al., 1992; Mundle and Saberwal, 2003). The large pocket region of pRb (amino acids 379-928) is known to have additional binding to

E2Fs and is necessary for binding to other cellular proteins, for example, MDM2 (Xiao et a., 1995). About 120 proteins have been reported to physically interact with pRb, mostly through the pRb pocket domain (Morris and Dyson, 2001; Cobrinik, 2005).

DNA tumor virus oncoproteins, such as adenovirus E1A, SV40 large T antigen, HPV E7, (Figge et al., 1988; Phelps et al., 1988; Stabel et al., 1985; Hu et al., 1990) and about 20 cellular proteins, such as for example, HDAC1, HDAC2, BRG1, cyclin D1, BRAC1, gankyrin and PAI2, possess an LXCXE-like motif in their sequences (Brehm et al., 1998; Magnaghi-Jaulin et al., 1998; Luo et al., 1998; Lai et al., 1999; Dunaief et al., 1994; Strober et al., 1996; Dowdy et al., 1993; Ewen et al., 1993; Fan et al., 2001; Darnell et al., 2003; Higashitsuji et al., 2000). The viral oncoproteins inactivate the pocket proteins by direct association through their LXCXE sequences. As viral oncoproteins utilize the LXCXE motif, it was proposed that cellular proteins would also interact with the pRb through this sequence. However, this model of interaction would not account for binding of protein like HDAC3 to pRb as it lacks any LXCXE-like motif (Chan et al., 2001; Dahiya et al., 2000). There are also reports that suggest that HDAC1 and HDAC2 do not utilize their LXCXE-like sequences for interaction with the small pocket of pRb (Chen at al., 2000; Dick et al., 2002; Kennedy et al., 2001); instead, recruitment of HDACs to pRb occurs via an intermediary, LXCXE containing protein, RBP-1 (Dunaief et al., 1994). How pRb recruits the class-1 HDACs to the repressor complex at the promoter remains a subject of controversy (Kennedy et al., 2001). Similar conflicting reports abound in the literature regarding the interaction of other proteins, for example BRG1, with the pRb (Dunaief et al., 1994; Strober et al., 1996). Some reports suggest direct interaction between BRG1 and pRb-AB, whereas other reports contradict this notion (Dahiya et al., 2000; Chen et al., 2000; Kang et al., 2004; Zhang et al., 2004). One report suggested that BRG1 and HDAC1 interact separately with pRb via differing modes of interaction (Dahiya et al., 2000).

Biochemical and structural studies showed that the E2F peptide binding site is separated by ~30 Å from the LXCXE peptide binding site (Lee et al., 2002; Xiao et al., 2003). There are other contact points on viral oncoproteins for pRb, but the major interaction is through the LXCXE motif. The crystal structure of the [pRb-AB]-[domain of SV40 large T antigen] shows that two-thirds of the total surface interaction between the two proteins is via the LXCXE motif (Kim et al. 2001). This motif interacts with the pRb-AB on the B subdomain in an extended conformation exactly like the LXCXE peptide from HPV E7 (Kim et al., 2001; Lee et al., 1998).

Since the assays reported in the literature were all based on inhomogeneous protein preparations, we decided to use homogenously purified proteins and check for pRb-AB/protein or peptide associations using direct *in-vitro* experiments. Most of the studies on the pRb/LXCXE

interaction have relied on mutagenesis whereby the LXCXE binding site on the pRb pocket was mutated and tested for binding with possible target proteins in cell lysate coupled with immunoprecipitation and Western blot assays. Our approach was different from these approaches in that we use direct binding methods *in-vitro*. Since our titration experiments use purified proteins, the results are less ambiguous and direct protein-protein interactions can easily be checked. In order to check for any interaction that could be weak, we have carried out our binding study also with NMR spectroscopy. Standard molecular biology methods for protein binding, for example ELISA or RIA, and Biacore fail to detect ligands with weak (i.e. millimolar) affinities. The ability of NMR to detect such ligands has made NMR increasingly important in drug discovery and structural proteomics (Meyer and Peters, 2003; Pellecchia et al., 2002).

Amino acids that form the LXCXE binding site in pRb are highly conserved across pRb homologs of different species (Ach et al., 1997; Lu and Horvitz, 1998). Many cellular pRb binding proteins contain LXCXE-like sequences. The simplest model would predict that pRb uses the LXCXE binding site to interact with target proteins, and viruses evolved an LXCXE sequence to mimic this interaction (Dick et al., 2000; Helt and Galloway, 2003). The model for the pRb-HDAC interaction has been tested by making the LXCXE binding site mutants of pRb (Chan et al., 2001; Dahiya et al., 2000; Chen and Wang, 2000; Dick and Dyson, 2002). pRb alleles deficient in LXCXE binding retain at least some ability to repress transcription in all the reports, but different conclusions were reached regarding the HDAC interaction. In two reports the authors concluded that the LXCXE mutant pRb retained the ability to bind HDAC1 (Dick and Dyson, 2002; Kennedy et al., 2001), whereas other reports showed reduced or no binding (Chan et al., 2001; Dahiya et al., 2000; Chen and Wang, 2000). From our data we conclude that although the E7 and HDAC1 LXCXE-like sequences bind to the same site on pRb, binding between the small pocket of pRb and E7 LXCXE peptide is tight whereas the HDAC1 IXCXE peptide binding is weak and the complex formation between the two is transient.

There have also been conflicting reports regarding the interaction of BRG1 with pRb-AB. One model suggests that BRG1 interacts with pRb through the LXCXE motif (Dunaief et al., 1994). However, using the LXCXE binding site mutant alleles of pRb, others reported that LXCXE motif may not be required for binding of BRG1 to pRb-AB (Dahiya et al., 2000; Chen and Wang, 2000), and moreover Zhang et. al. showed that HDAC1 and BRG1 interact with separate sites of pRb (Zhang et al., 2000). Recently, Kang et al. have shown that the direct physical interaction of BRG1 with the pRb is not required for the BRG1 ability to arrest growth and flat cell formation but rather it controls the activity of pRb via regulation of $p21^{CIP/WAF1/SDI}$, an upstream regulator of pRb (Kang et al., 2004). Our results showed no binding between pRb-

AB and a LXCXE containing BRG1 domain. Moreover no binding was detected between pRb-AB and the LXCXE containing peptide from the PAI2 protein and the N-terminal domain of pRb. We would like to note that these results do not rule out other interactions among the full length or larger domains of these proteins.

We also examined a series of E7 and HDAC1 derived peptides with single or double amino acid substitutions. Whereas substitutions Ile to Leu at the position P-2 and Leu to Phe at P+4 did not influence the binding, any positively charged residue in the peptides had significant effect on binding. In agreement with this trend, the peptide from the PAI2 protein, which has a series of positive residues at the C-terminus of LXCXE, did not show any binding to pRb-AB. An Asp to Arg mutation at P-3 in HPV E7 reduced binding to pRb-AB ~6 fold, a reverse Arg to Asp change at P-3 in HDAC1 showed tighter binding (with the K_D ~8-10 μM), showing that any positive charge in the XLXCXEXXX sequence can be detrimental for the interaction. In fact a frequently used mutation in biochemical assays Glu(107)Lys at P+1 position in SV40 large T antigen leads to its dissociation from pRb-AB (Zalvide et al., 1998). The double mutant, (an Asp to Arg substitution at P-3 coupled with Tyr to Glu at P+1), in HPV E7 reduced the binding ~25 fold, whereas an Arg to Asp mutation at P-3 coupled with Glu to Tyr at P+1 showed binding of 3.27 ± 0.07 μM, the tightest among all the HDAC1 derived peptides, indicating that the Tyr to Glu change is also important for this interaction. These results are in agreement with the crystal structure of pRb-AB/E7, which shows that the E7 amino acids at these positions make contact with each other, which may help keeping the LXCXE sequence in a correct, extended configuration. We can thus expect that a LXCXE motif that is a part of a regular secondary structure element, such as an α-helix or β-sheet, would present a conformation that is not optimal for the interaction with pRb.

Our results also showed that single point mutations C706F and K713S in pRb are not able to disrupt the structure of pRb-AB and these mutants were still able to bind HPV E7 LXCXE peptide strongly. We concluded therefore that a single amino acid substitution may not be sufficient for inactivation of the pRb binding to LXCXE sequences.

In conclusion, our data show that the LXCXE-like sequences with only the determinant triplet Leu, Cys and Glu in are not sufficient for tight complex formation with pRb-AB. These residues might provide specificity, but high affinity requires other flanking residue interactions. Specifically, a sequence XLXCXEXXX, where X should not be a positively charged amino acid (e.g. Lys or Arg), and X should preferably be a hydrophobic residue, should bind tightly to pRb-AB. Positively charged amino acids do not abolish binding, but weaken it to differing extents such that fast exchange XLXCXEXXX/pRb complexes form. Moreover, it is worth noting that

the Leu, Cys and Glu amino acids in the IXCXE motif of HDACs homologues across different species are not conserved, whereas the amino acid residues N-terminal to the IXCXE motif are highly conserved (Fig. 5.25). In fact the yeast homologue of class I HDACs, Rpd3, does not have any sequence homology in the IXCXE region of the HDAC1. It has been shown that in yeast, the HDAC dependent transcriptional repression by pRb does not require the pRb LXCXE binding cleft; rather, Msilp(RbAp48) was proposed to act as a bridging protein between pRb/Rpd3p. Similarly in some reports RBP1, an LXCXE containing protein, was shown to act as a bridging protein between pRb/HDAC (Lai et al., 1999; Kennedy et al., 2001). We envisage that class I HDACs (HDAC1, HDAC2 and HDAC3) have some functions common with their deactylase activity, whereas HDAC1 and HDAC2, which have IXCXE motifs, have other functions unique to this interaction. All these observations, combined with our results, suggest that the IXCXE motif in HDAC1 and HDAC2 evolved to serve a function unique to higher organisms. Thus for example, the HDAC1 RIACEEF/pRb interaction may not be important in some *in-vivo* situations, but this interaction does exist and its weak feature should be taken into consideration when considering models of the pRb/HDAC interaction. It should be mentioned here that a very different outcome if full-length proteins were used would be unlikely. This is because the primary interaction sites are located in the fragments we have studied.

				I	X	C	X	E		
Human	D	K	R	I	A	C	E	E	E	F
Mouse	D	K	R	I	A	C	E	E	E	F
Chicken	D	K	R	I	S	C	D	E	E	F
Sea Urchin	D	K	R	I	Q	R	D	D	E	F
C. Elegans	D	K	R	L	P	P	Q	I	T	D
Drosophila	D	K	R	I	V	P	E	N	E	Y

Figure 5.25. *Sequence alignment of L/IXCXE-motif of various HDACs from different organisms. Amino acids in the red indicate the position of I, C and E or corresponding residues. Conserved residues N-terminus to the IXCXE motif are highlighted by a box.*

Multiple functions of pRb (e.g. proliferation, apoptosis, cell differentiation, cellular senescence, control of developmental processes in extra-embryonic tissues, and maintenance of tissue homeostasis - to name a few) require the interaction with many cellular partners directly or indirectly (through in multiprotein complexes). These functions can be achieved not only by adding interacting partners to the system, but also by modulating the strength of the interaction by, for example, posttranslational modifications, such as phosphorylation and acetylation, or

simply (without any modifications) by the primary amino acid sequence present in binding sequences.

Viral proteins associate with pRb by displacing cellular proteins, such as E2F and HDAC1, with greatest viral advantage presumably when the interactions are strong. Interaction between the cellular proteins and pRb is highly regulated, so that and the regulatory interactions need to be reversible. For example, the phosphorylated C-terminus of pRb is suggested to interact with the positively charged lysine patch and displace HDAC1. This can only be achieved if the interaction between HDAC1 and the pRb B domain is transient (Harbour et al., 1999).

Another example of the complexity of the LXCXE/pRb interactions is provided by a recent study of the SUMO modification of pRb (Ledl et al., 2005). This study showed that pRb is SUMOylated at lysine 720 and the level of SUMOylation is controlled by the interaction of pRb with LXCXE-motif proteins. The viral E1A, E7 and the cellular EID-1, E1A-like inhibitor of differentiation proteins, completely abolished SUMO modification of pRb, while HDAC1 showed reduced SUMOylation of pRb, but failed to completely block it. Our data can provide an explanation for these observations: the SUMO inhibitory potential of these LXCXE proteins most likely reflects their different binding affinities for pRb.

5.3 Functional and structural studies of hBRG1 domains

5.2.1 Cloning and expression of hBRG1 constructs

Various constructs were designed to express functional domains of the BRG1 protein (Fig. 5.26). Constructs from the N-terminal part of the protein did not express well in *E. coli*, whereas constructs covering the central ATPase domain of the protein expressed into the inclusion bodies. Various refolding conditions were tried but refolding was not possible in most of these cases. Both rapid dilution and dialysis methods were employed but without much success. Refolding was possible only for the LNQV construct. NMR and CD spectroscopy results showed that this domain is unstructured in solution. The MVIG and NVKE constructs were expressed to a moderate level in the soluble fraction of *E. coli* however, both of these constructs were unstructured in solution. The NVEK construct has the LXCXE motif and was used for its binding studies with pRb-AB (Section 5.1.4.3). However it did not bind to pRb-AB.

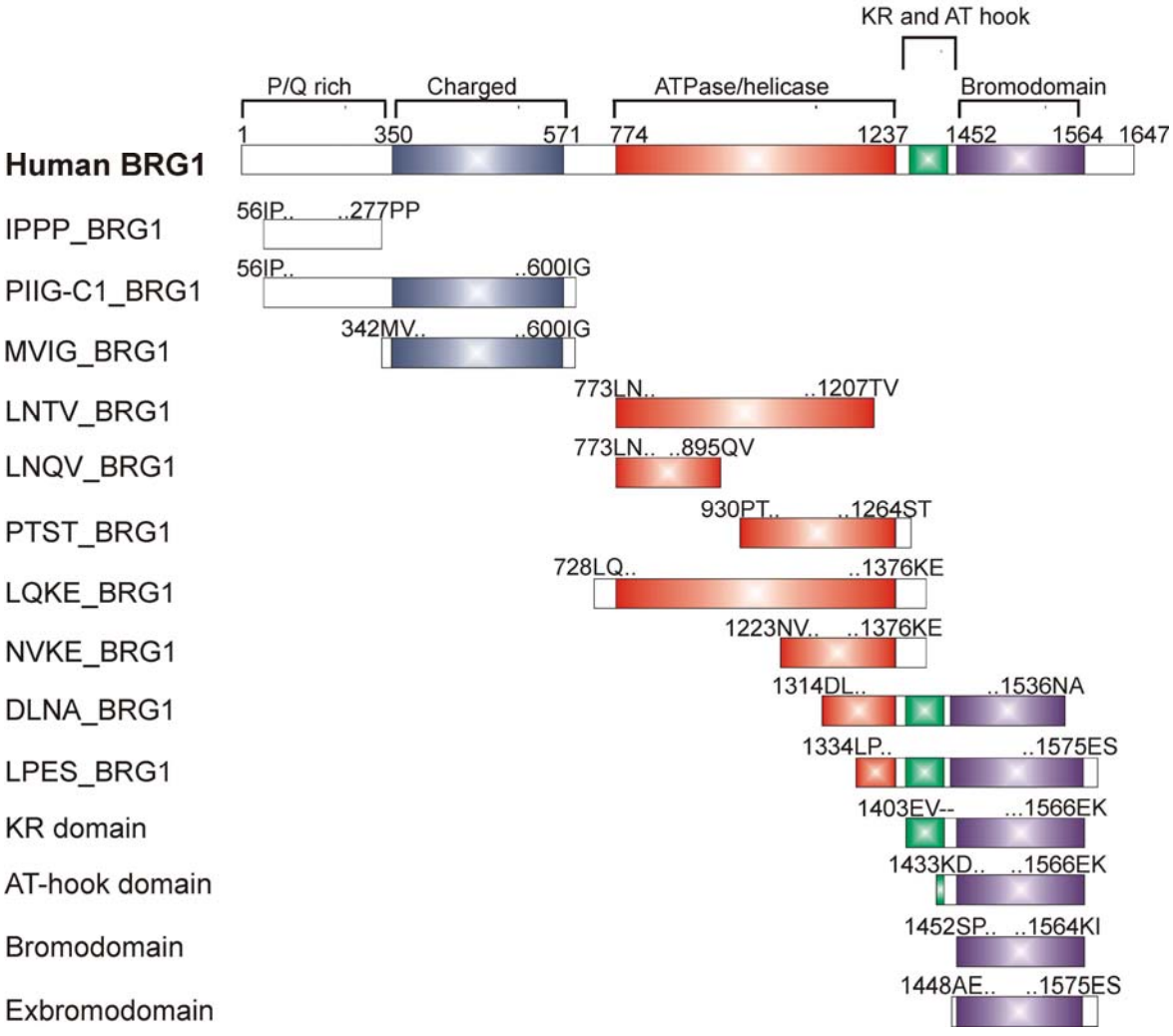

Figure 5.26. *Graphical representation of various hBRG1 constructs cloned and used in this study. All the constructs are compared with the full-length protein.*

Constructs from the C terminal part of the BRG1 protein, encompassing the bromodomain, were expressed in high level in *E. coli* and were folded as checked by NMR and CD spectroscopy. All of these constructs were expressed in a similar manner. For the large-scale expression of protein, 1 l of LB media or MM media was inoculated with 10 ml of an overnight grown primary culture. Cells were grown at 37 °C until the OD of ~0.6. Cells were induced with 1 mM IPTG for overnight time (~14-16 h) at 28 °C. Cells were pelleted down by centrifugation and stored at –20 °C. Table 5.6 summarizes the results of cloning and expression of various BRG1 constructs.

Table 5.6. *Results of cloning and expression of various BRG1 domains*

No.	Construct Name	Primer no.*	Vector	Expression in E. coli	1D-1H NMR diagnosis
1	IPPP_BRG1	22, 23	pET30 LIC/Xa	No	NA
2	PIIG-C1_BRG1	22, 25	pET30 LIC/Xa	No	NA
3	MVIG_BRG1	24, 25	pET30 LIC/Xa	Moderate	Unstructured
4	LNTV_BRG1	26, 27	pET30 LIC/Xa	High/IB	ND
5	LNQV_BRG1	26, 28	pET30 LIC/Xa	High/IB	Unstructured
6	LQKE_BRG1	31, 32	pET30 LIC/Xa	High/IB	ND
7	NVKE_BRG1	33, 32	pET30 LIC/Xa	Moderate	Unstructured
8	PTST_BRG1	29, 30	pET30 LIC/Xa	Low	Unstructured
9	DLNA_BRG1	34, 35	pET30 LIC/Xa	Moderate	Unstructured
10	LPES_BRG1	36, 47	pET46 LIC/Xa	High	Partially folded
11	KR-domain	37, 39	pET46 LIC/Xa	High	Folded
12a	AT-hook domain	38, 39	pET46 LIC/Xa	High	Folded
12b	AT-hookdomain♣	40, 41	SDM of AT-hook domain	High	Folded
13a	Bromodomain	42, 43	pET30 LIC/Xa	High	Folded
13b	Bromodomain♣	44, 45	SDM of Bromodomain	High	Folded
14	exbromodomain	46, 47	pET46 LIC/Xa	High	Folded

Abbreviations: LIC, ligation independent cloning; SDM, site directed mutagenesis; ND, not done; NA, not applicable.
* Please refer to Table 4.3 for the primer name, no. and sequence (section 4.2.3.2).
♣ These constructs have no His-tag.

5.2.2 Protein purification and characterization strategies

5.1.2.1Purification of the exbromo-, bromo-, AT-hook- and KR- domains

All recombinant domains from C terminal part of the protein were purified using the same strategy. The proteins were purified over a Ni-NTA Agarose column in the first step followed by N-terminal His-tag removal and gel filtration chromatography. However in a few cases the bromodomain and the AT-hook domain were purified without the use of a N-terminal His-tag. In this strategy cell lysate was directly applied to the SP Sepharose anion exchange column (~8 ml). Proteins were eluted using a salt gradient. This step was followed by gel filtration (Fig. 5.27). Both strategies gave >95% pure protein in the end (Fig. 5.28A-D).

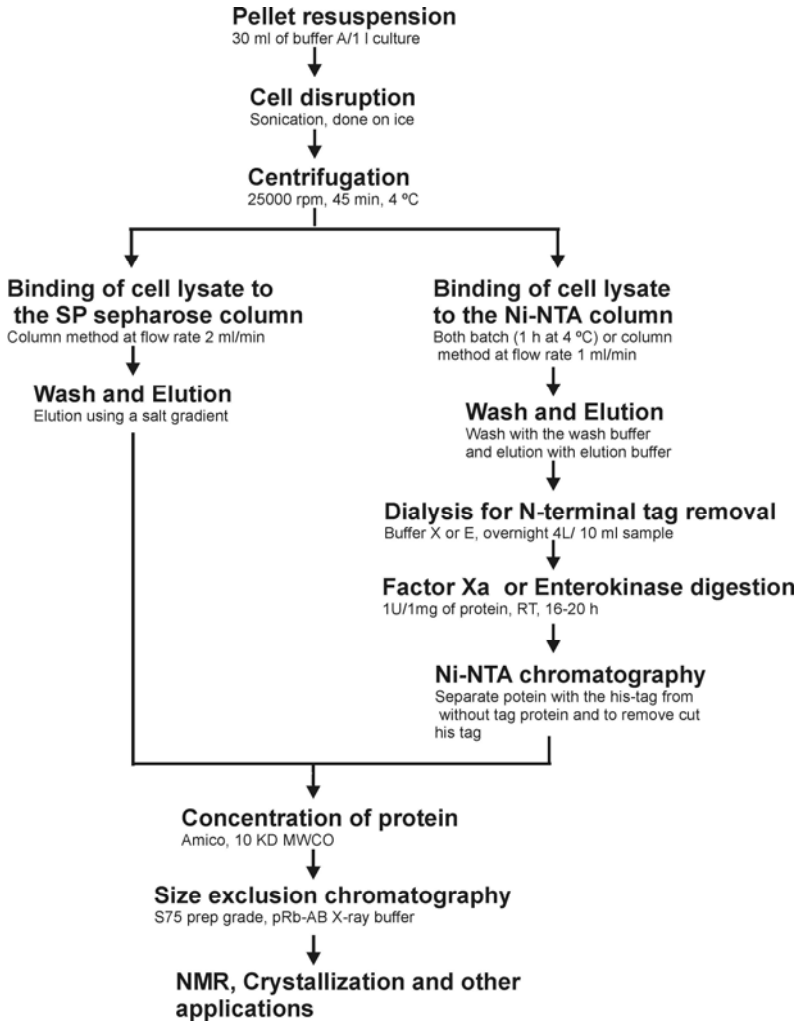

Figure 5.27. *Flow chart of the purification scheme for the exbromodomain, the bromodomain, the AT-hook domain and the KR-domain.*

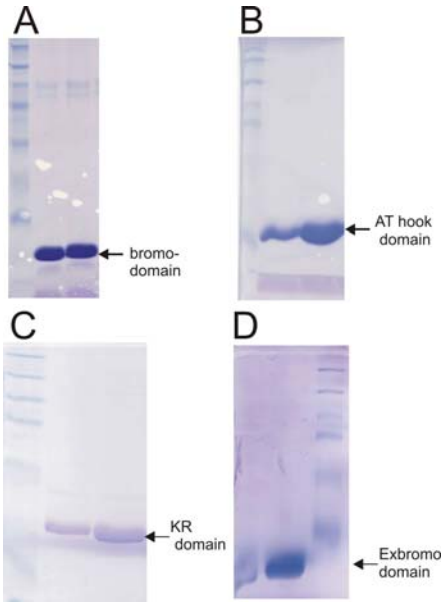

Figure 5.28. *SDS-PAGE of various BRG1 constructs after purification. (A) Bromodomain. (B) AT-hook domain. (C) KR-domain (D) Exbromodomain*

5.1.2.2 Exbromo-, bromo-, AT-hook- and KR- domains characterization

A purified protein was characterized using a variety of methods. The folding of a recombinant protein was checked using NMR spectroscopy and CD spectroscopy analysis. Figure 5.29A and B shows the gel filtration chromatography and 1D-^{1}H NMR spectra of bromodomain, exbromodomain, AT-hook domain and KR-domain.

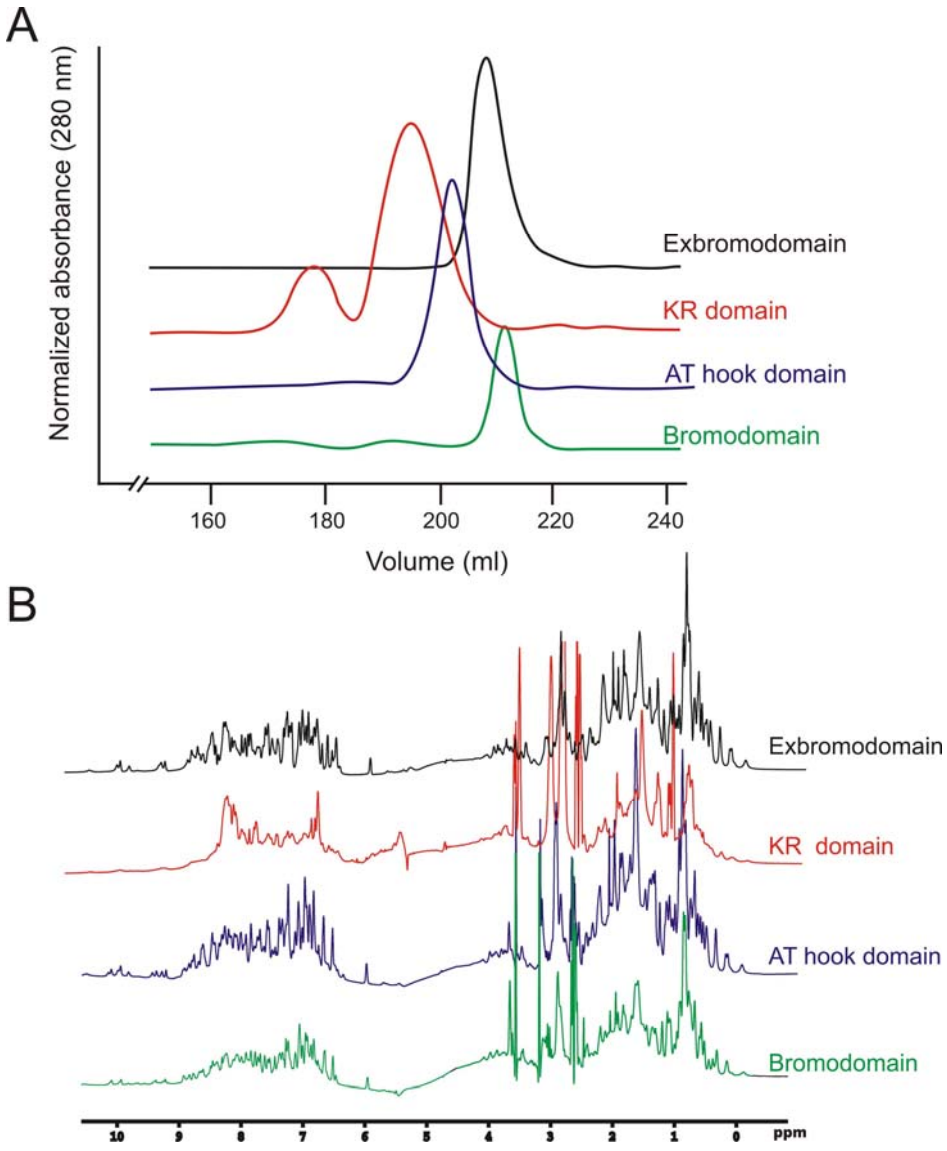

Figure 5.29. *Characterization of the KR-, AT-hook-, exbromo- and bromodomains. (A) S75 prep grade profile of various domains. Bromodomain, exbromodomain, and AT-hook domain eluted at approximate size of monomer however, the KR-domain eluted at as approximate dimer size with a small part at a tetramer size as well. (B) 1D-^{1}H NMR spectra of various constructs. The bromodomain, the exbromodomain and the AT-hook domain gave well dispersed spectra of a folded protein, however in case of the KR-domain the line-width of the signal was broadened, corroborating with the gel-filtration chromatography results, showing that the protein is in a dimeric form in solution.*

5.2.3 Peptides and oligonucleotides

The peptides were synthesized using solid phase synthesis, purified with the C8 reverse phase chromatography and checked by mass spectrometry. Peptides were dissolved in the buffer that was used for NMR titration of the proteins. Table 5.7 lists the peptides used in this work.

Table 5.7. *Peptides used in this work*

S. No.	Name of the peptide	Sequence (N- to C- terminus)
1	acetylated-K8 peptide histone H4	GKGGK(Ac)GLGKG
2	acetylated K9/14 peptide histone H3	TKQTARK(Ac)STGGK(Ac)APR

Oligonucleotides were ordered from Metabion GmbH (Germany) and were used without further purification. To make the double stranded DNA, complementary strands were dissolved in water and an equimolar ratio mixed with each other at 95 °C for 10 min and allowed to cool slowly at room temperature for overnight. Table 5.8 lists the oligos used in this study. The 4H DNA was also prepared by mixing equimolar ratios of each strand (Binachi, 1988).

Table 5.8. *Oligonucleotides used in this work*

No.	DNA	Sequence (5'-3')
1	DNA1	Sense GGGAAATTCCGC
		Antisense GCGGAATTTCCC
2	DNA2	Sense CGCGAATTCGCG
		Antisense CGCGAATTCGCG
3	4H	Strand1 CCCTATAACCCCTGCATTGAATTCCAGTCTGATAA
		Starnd2 GTAGTCGTGATAGGTGCAGGGGTTATAGGG
		Starnd3 AACAGTAGCTCTTATTCGAGCTCGCGCCCTATCACGACTA
		Starnd4 TTTATCAGACTGGAATTCAAGCGCGAGCTCGAATAAGAGCTACTGT

5.2.4 Functional investigation on the bromodomain, the AT-hook and the KR-domains

5.2.4.1 Interaction of the bromodomain with the acetylated lysine peptides from histones

It has been suggested that the bromodomain of BRG1 should bind to the acetylated lysine of histone H3 and H4. Our NMR titration experiments show no binding between the bromodomain and the acetylated K8 peptide from histone H4, whereas we detected the change in NMR chemical shifts of a number of amino acids upon addition of a doubly acetylated K9/14 peptide from histone H3 (Fig 5.30A,B). However the binding seems to be in fast exchange, i.e. weak binding. Weak binding of the bromodomain to an acetylated histone peptide is expected since it is known that binding between bromodomains and acetylated peptides is weak, the dissociation constants being in the higher micromolar range (Dhallium et al., 1999; Hudson et al., 2000;

Jacobson et al., 2000). Similar results were obtained when the AT-hook domain was titrated with the doubly acetylated K9/14 peptide from histone H3 (Fig. 5.30C). Nevertheless these results show that the recombinant protein domains in our studies are functional and active.

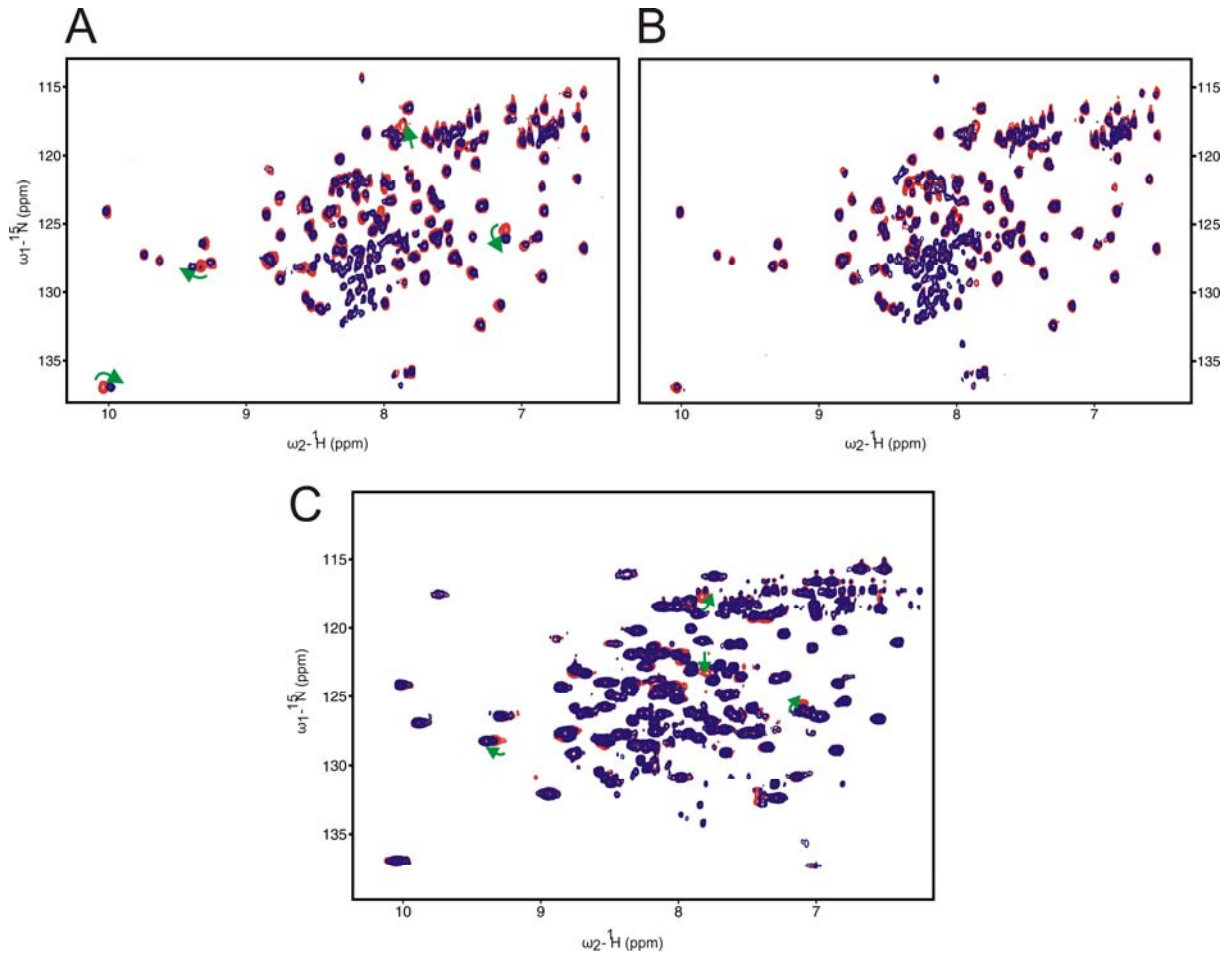

Figure 5.30. *Monitoring the binding of N-acetyl-lysine peptides to the bromodomain and the AT-hook domain using NMR spectroscopy. 2D-HSQC spectra at two steps of titrations are shown here at the final protein to peptide ratio of 1:2 (blue, the starting and red, the final step) (A) ^{15}N labeled sample of bromodomain was titrated with a N-acetylated K9 and K19 peptide from H3, a numbers of peaks were perturbed upon addition of the peptide (green arrows) showing that the two component interact. (B) The ^{15}N labeled sample of bromodomain was titrated with an acetylated K8 peptide from histone H4, no peaks were perturbed, showing that there is no interaction. (C) The ^{15}N labeled sample of the AT-hook domain was titrated with an N-acetylated K9 and K19 peptide from H3, again a number of peaks were perturbed upon addition of the peptide (green arrows) showing that the two components interact.*

5.2.4.2 Interaction of the AT-hook and the KR-domain with linear DNA

We checked the binding of the AT-hook domain from BRG1 to the linear DNA. Table 5.7 lists the DNA oligonucleotides that were used in our experiments. We used two double stranded DNAs in this work, designated DNA1 and DNA2. DNA1 corresponds to a part of PRDII, a point in the enhancer of the human interferon-β gene (nIFN-β), a naturally occurring binding site for the HMGA protein (Thanos and Maniatis, 1992). This sequence of DNA has been used for the

structural and functional investigation of the HMGA1 AT-hook-DNA interaction (Huth et al., 1997; and Dragan et al., 2003). DNA2 is a palindromic decamer that was used by Murphy et al. (1999) for solving the structure of HMG1 from Drosophila. We used NMR chemical shift perturbation method to monitor binding of ligands to the proteins (Rehm et al., 2002; and Peters and Mayer, 2003). An ^{15}N labeled AT-hook domain was titrated with DNA1, resulting in shifts or perturbations of some residues, indicating binding between the two components (Fig. 5.31A). A closer look at the spectra indicates that the residues that were perturbed or shifted showed the characteristics of slow exchange, indicating tight binding. Addition of DNA2 also showed binding to the AT-hook domain, however, weak compared to DNA1 (Fig. 5.31B).

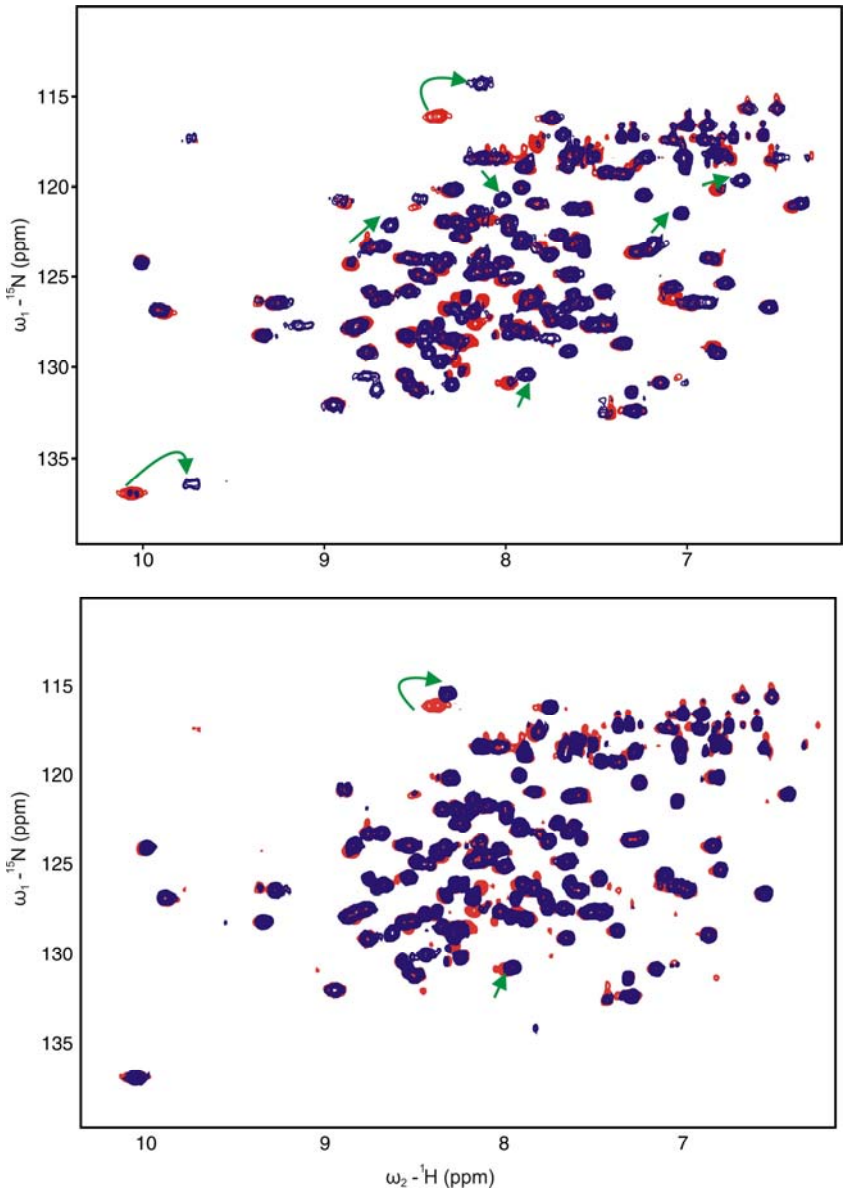

Figure 5.31. *Titration of the AT-hook domain with linear DNA. An ^{15}N labeled AT-hook domain was titrated with the duplex DNA. The figure shows the overlay of 2D-HSQC spectra of the AT-hook domain*

only (red) and the AT-hook domain with DNA at 1:1 molar ratio (blue). Many peaks were either shifted or perturbed (green arrows), which is the hallmark of binding. (A) AT-hook domain titrated with DNA1. (B) AT-hook domain titrated with DNA2.

We next checked the binding of the AT-hook and KR-domains with DNA1 using ITC. Results of a typical ITC experiment on binding of the AT-hook to DNA1 in 100 mM KCl at 5 °C and 30 °C are shown in Figures 5.32A and B. Binding at these temperatures proceeds with heat effect of opposite sign: positive at low temperature and negative at high temperature. Linear regression analysis of the interaction fits well to the stoichiometry of 1:1. We also measured the binding of the KR-domain with DNA1, the stoichiometry of 1:2 was observed in this case (Fig. 5.32C). No binding was detected between the bromodomain and DNA1 (Table 5.8). Binding constants (K_D) are given in Table 5.9.

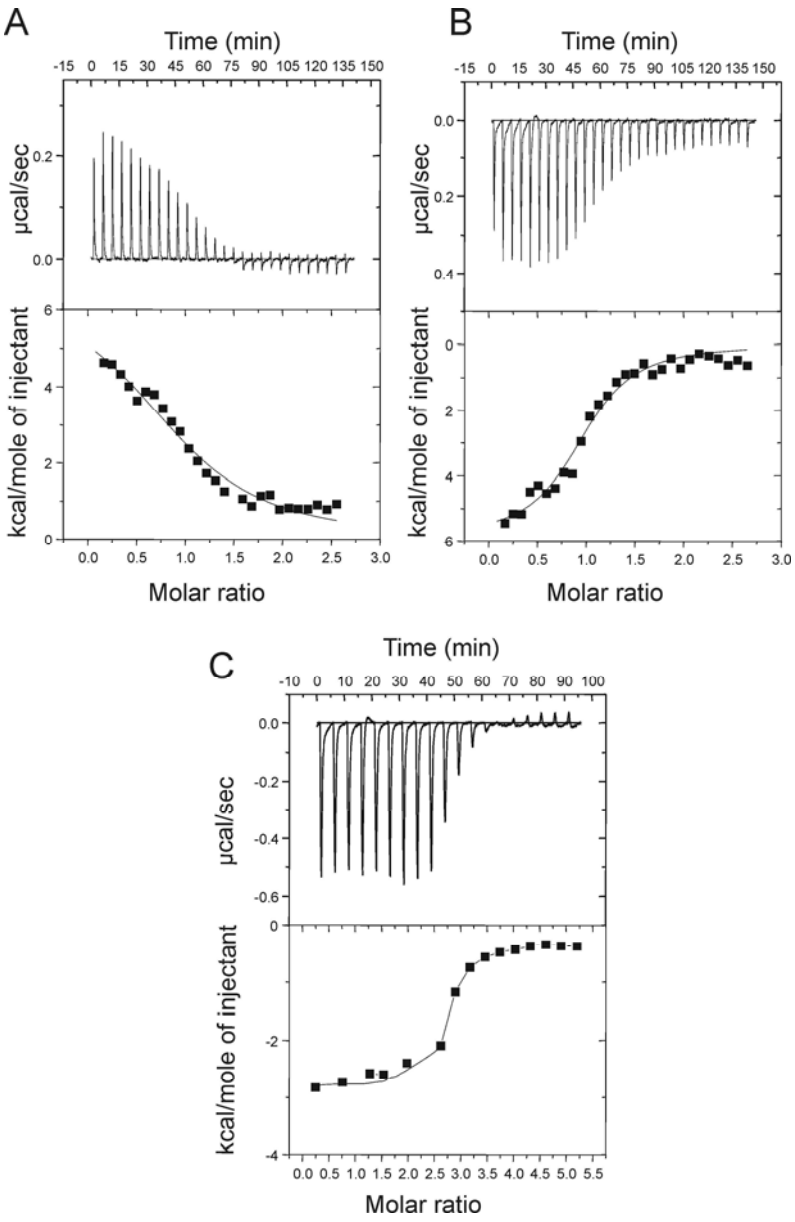

Figure 5.32. *ITC data for binding of the AT-hook domain to DNA1 (A) titration carried out at 30 °C and (B) at 5 °C. (C) Binding of the KR-domain with DNA1 carried out at 30 °C*

Table 5.9. *Dissociation constants for various domains of BRG1 and DNA determined by ITC tirations. The stochiometry ratio is give as 'n'*

Interaction of DNA with:	**Temperature (°C)**	**K_D (μM)**	***n***
Bromodomain	30	no binding	-
AT-hook domain	5	6.94 ± 0.17	1.05
AT-hook domain	30	1.80 ± 0.03	0.99
KR-domain	30	1.09 ± 0.50	2.80

5.2.4.3 The interaction of the AT-hook domain with 4H DNA

We checked the interaction of the AT-hook domain and the 4H DNA using NMR spectroscopy. When we titrated the ^{15}N labeled AT-hook with 4H DNA, a number of peaks were perturbed showing that the two components interact (Fig. 5.33A).

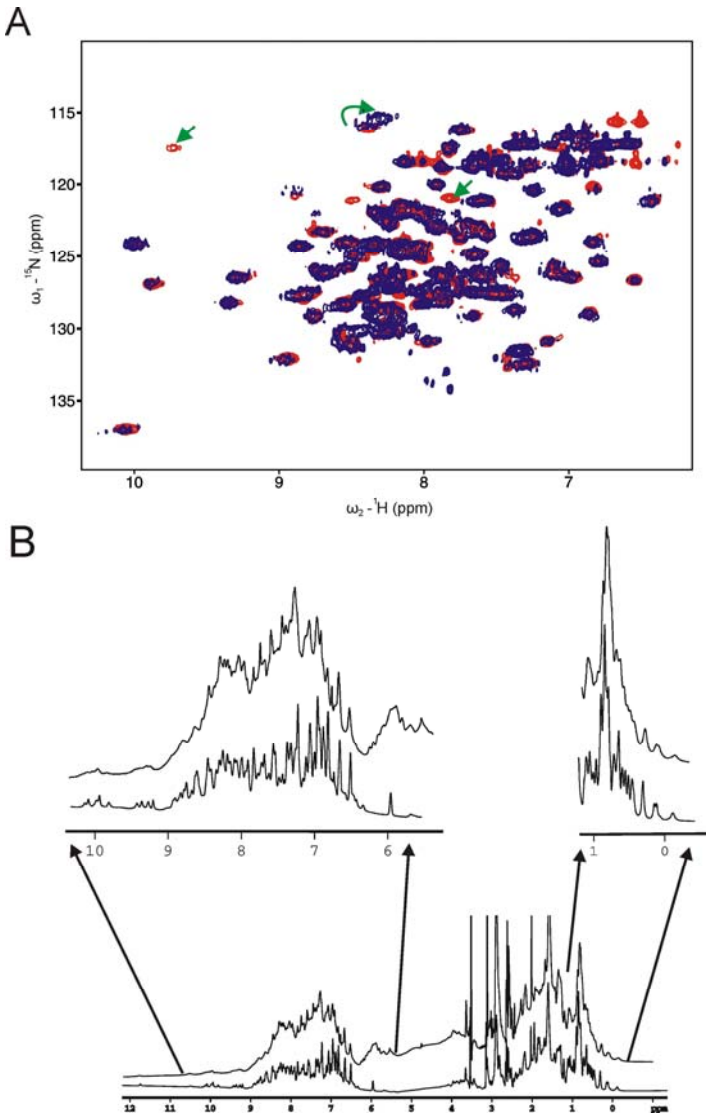

Figure 5.33. *Interaction of the AT-hook region with 4H DNA. (A) An ^{15}N labeled AT-hook domain was titrated with 4H DNA. The figure shows the overlay of 2D-HSQC spectra of the AT-hook domain (red) and*

the AT kook domain with 4H DNA at 1:0.3 molar ratio (blue), respectively. Many peaks were either shifted or perturbed or disappeared, which indicates binding. Overall broadening of peaks was also observed (B) The 1D proton spectra of the AT-hook domain before and after titration with the 4H DNA clearly shows the broadening of the line widths.

However, an interesting observation was made, the line width of the signals were broadened significantly. The line width of NMR signals is correlated to the size of the molecule and to exchange processes taking place (Wuthrich, 1986; Abragam, 1961). The broadening of the line width was also clear in 1D proton spectra after the titration (Fig. 5.33B). This shows that the AT-hook domain-4H complex behaves like a big complex. Because of the faster relaxation mechanisms, the NMR signal from the protein, a big complex now, decays faster compared to the AT-hook domain alone, or AT-hook domain-DNA1 complex. This in turn produces broader lines for the resonances.

5.2.3.4 Conformational properties of the AT-hook domain-DNA complex

We used CD spectroscopy to monitor conformational changes in the linear DNA1 or 4H DNA after binding to the AT-hook domain. In the region between 230 and 300 nm, the CD spectra of DNA are not significantly affected by the spectral contribution of the protein, and the recorded spectra therefore show structural changes of DNA only. The CD spectrum of linear DNA1 is characterized by a strong negative signal at 247 nm and a strong positive signal at 280 nm. Upon addition of increasing amount of the AT-hook domain, the negative signal at 265 nm is shifted in the positive direction, whereas the positive maximum at 280 nm was reduced (Fig. 5.34A). These results clearly demonstrate the conformational change in DNA upon interaction with the protein. Such changes especially decrease in the intensity at ~280 nm, suggests that the number of base pairs per turn decreases after addition of the AT-hook domain, which may indicate the change in the conformation of DNA from B-form to A-form (Zhang et al., 2003).

Next, we checked the effect of the AT-hook domain on the 4H DNA. The CD spectrum of 4H DNA is characterized by a strong negative signal at 247 nm and a strong broad positive signal at 275 nm. No significant spectral changes were obtained upon addition of increasing amounts of the AT-hook domain implying that no conformational changes in 4H DNA occur when it binds to the AT-hook domain (Fig. 5.34B).

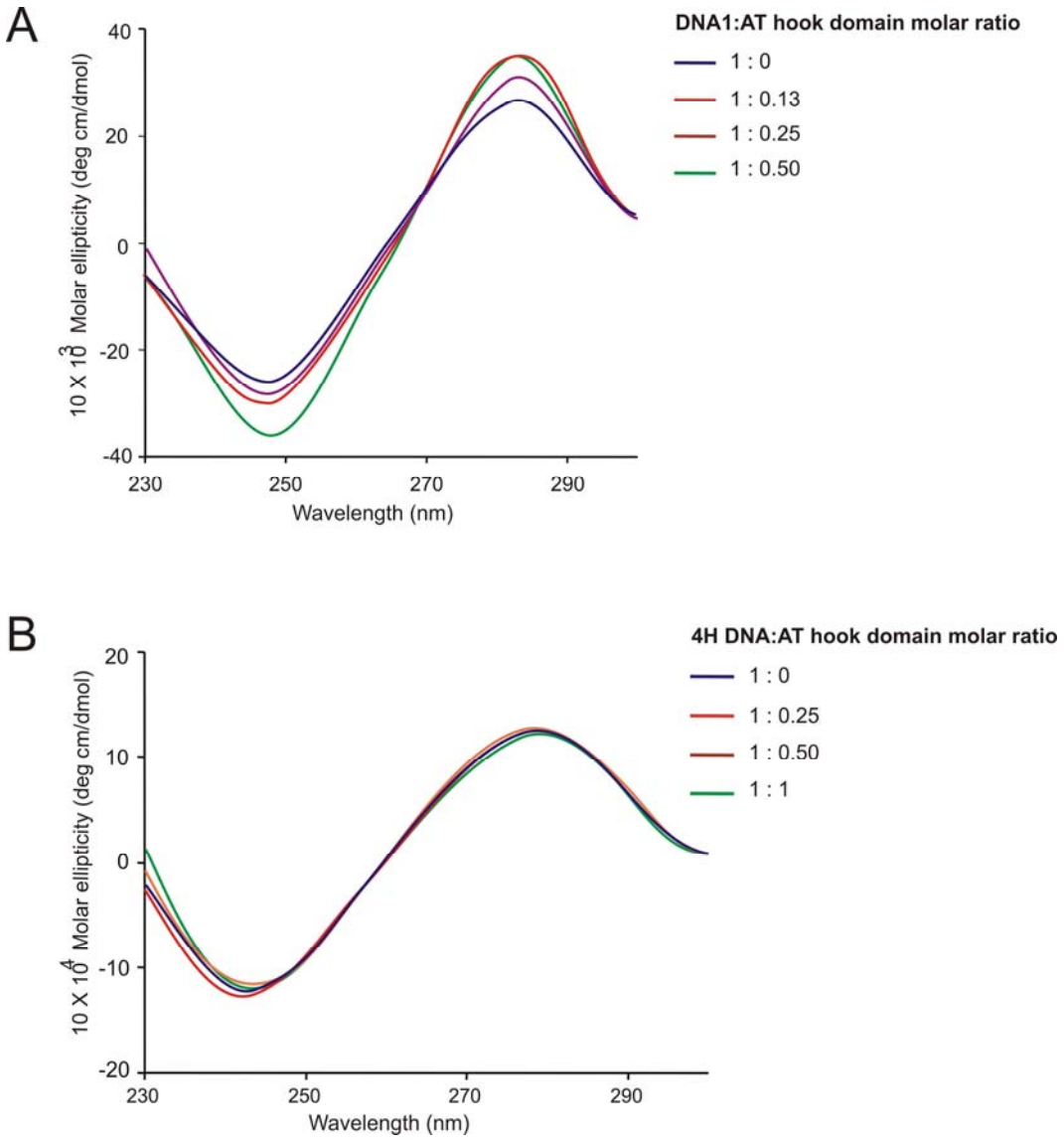

Figure 5.34. *CD spectra monitoring conformational changes in DNA upon interaction with the AT-hook domain of hBRG1. (A) Spectral changes upon binding of the AT-hook domain to linear DNA, DNA1. Molar ratios of DNA1 and AT-hook domain at each step are indicated. (B) Spectral changes upon binding of the AT-hook domain to four-way-junction (4H) DNA. Molar ratio of 4H DNA and AT-hook domain at each step are indicated.*

5.2.5 Crystallization of the bromodomain from hBRG1

1D-^{1}H spectra of bromodomain, AT-hook and KR- domains showed that they were well structured in solution, however all my attempts to crystallize them were unsuccessful. I tried different crystallization screens, which included the Hampton screens: 1, 2, Lite, PEG/Ion, Index and SaltRx, and two temperature 4 °C and room temperature. I therefore decided to make another construct named the exbromodomain. This construct crystallized quickly under three conditions: the Hampton screen PEG/Ion No. 29, Index No. 45, and Index No. 62 (Table 5.10).

Table 5.10 . *List of various conditions under which the bromodomain crystallized.*

S. No.	Condition	Composition
1	Hampton PEG/Ion screen No. 29	0.2 M potassium acetate; 20% w/v PEG 3350; pH 7.8
2	Hampton Index screen No. 45	0.1 M Tris pH 8.5; 25% w/v PEG 3350
3	Hampton Index screen No. 62	0.2 M Trimethylamine N-oxide dihydrate; 0.1 M Tris pH 8.5; 20% PEG 2000
4	Self made Index no. 45 pH 8.5	0.1 M Tris pH 8.5; 25% w/v PEG 3350
5	Self made Index no. 45 pH 8.8	0.1 M Tris pH 8.8; 25% w/v PEG 3350

5.2.5.1 Optimizing the construct of the bromodomain for crystallization

A highly charged protein behaves as a polarized molecule at near physiological pH value. The pIs of the bromodomain, the AT-hook domain and the KR-domain are close to ~10 (Table 5.11). This is because of the uneven balance of positive and negative charged amino acids in their sequences (positive > negative, Table 5.11). The solubility of the AT-hook domain and the KR-domain at pH 7.2 was high (>30 mg/ml), whereas the solubility of the bromodomain was very low (<5 mg/ml) and degradation of the protein was also observed at concentration above 5 mg/ml. The bromodomain and the AT-hook domain eluted as monomers after gel-filtration, whereas the KR-domain eluted approximately as a dimer (Fig. 5.29, Section 5.2.2.2). I decided to balance the ratio of charged amino acids and designed the exbromodomain in a way that the ratio of positive/negative amino acids is close to 1.0 (Fig. 5.26 and Table 5.6; Section 5.2.1). This brings theoretical pI value of the exbromodomain down to ~7.7 (Table 5.11). The exbromodomain expresses well in *E. coli* in the soluble fraction. The solubility of exbromodomain was around 20 mg/ml at pH 7.3 and eluted as a monomer after analytical gel-filtration and was stable at room temperature.

Table 5.11. *Physical properties of bromo-, AT-hook, KR and exbromodomain*

Construct	Length (AA)	Arg + Lys residues	Glu + Asp residues	Theoretical pI	Tm (°C)	Oligomeric state	Solubility (mg/ml)
Bromodomain	114	21	15	9.46	32.85	Monomer	< 5
AT-hook domain	136	30	20	9.66	47.00	Monomer	~30
KR-domain	165	40	23	10.00	46.55	Dimer	~30
Exbromodomain	129	23	22	7.76	57.06	Monomer	~20

AA, amino acid; pI, isoelectric point; Tm, melting temperature

5.2.5.2 Thermostability of various constructs

Since the behavior of the, the bromodomain, the AT-hook domain and the KR-domain was significantly different, I decided to check their thermostability by thermal denaturation employing CD spectroscopy. All these recombinant constructs encompass the helical bromodomain; therefore they gave typical CD spectra of a helical protein.

The CD signal from the 222 nm wavelength (a characteristic minimum for α-helix) was monitored to see the effect of increasing temperature on the protein structure. The data were analyzed by the program provided by the supplier and the melting temperatures calculated are shown in Figure 5.35 and Table 5.11. In all cases thermal denaturation of proteins was irreversible.

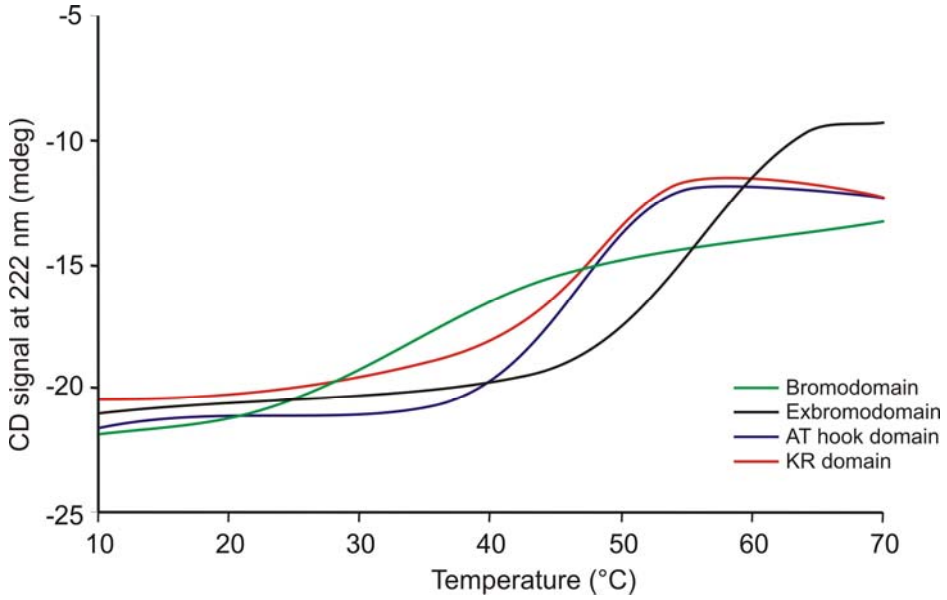

Figure 5.35*. Thermostability of the bromodomain, the exbromodomain, the AT-hook domain, and the KR-domain determined by CD spectroscopy. The curve for respective construct is shown by color indicated.*

The results were strikingly different. The bromodomain was the least stable whereas the exbromodomain was the most stable. The stabilities of the AT-hook domain and the KR-domain were between the bromodomain and exbromodomain (Table 5.11). These results also explain the reasons for degrading nature of the bromodomain.

5.2.5.3 Crystallization of the exbromodomain

The crystals for the exbromodomain appeared in both commercial Hampton screens, as well as self made solutions. Best crystals appeared with the protein concentration of ~8 mg/ml at room temperature (Fig. 5.36A-C). Since the molecular replacement (MR) and multiple isomorphous replacement (MIR) method for finding phases were not successful, I made a Sel-Met labeled

protein for multiwavelength anomalous diffraction (MAD). The crystal of the Sel-Met labeled protein appeared in the same conditions as for the native protein (Fig 5.36D). The 'exbromodomain' will be referred as the 'bromodomain' from now on to have the clarity in terms of comparing the structure with those of other bromodomains.

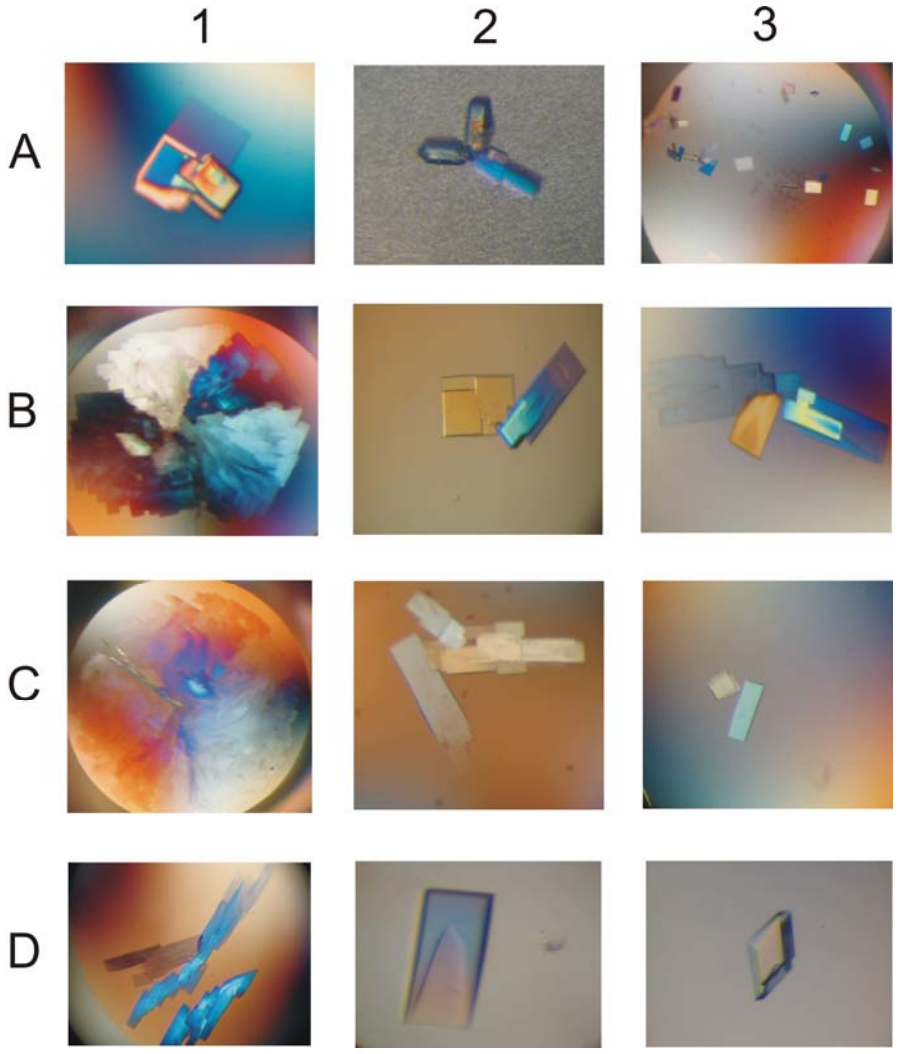

Figure 5.36. *Crystals of the exbromodomain grown under different conditions. Row A, crystals grown in the Hampton screen PEG/Ion solution No. 29 (1 from Hampton screen, 2, 3 from self made solutions). Row B, crystal grown in the Hampton screen Index No. 45 (1 from Hampton screen, 2, 3 from self made solutions). Row C, crystal grown in the Hampton screen Index. No. 45. Row D, crystals of the Sel-Met labeled protein grown under the Hampton screen Index solution No. 45 (1, 2) and no. 62 (3).*

5.2.5.4 Data collection and multiwavelength anomalous diffraction

Crystals were plunge-frozen after 30 s in the cryoprotectant solution containing 30% of MPD in the mother liquor. The diffraction data was measured on the MPG/GBF beamline BW6 at DESY (Fig. 5.37). Data were indexed, integrated and scaled with the XDS package (Kabsch, 1999). The native dataset was collected at 1.05 Å. The selenomethionine derivative was measured at 0.9793

Å for the peak, 0.9796 Å for the inflection point and 0.9770 Å for the high remote datasets. Datasets were of high quality and showed a strong anomalous signal. Two selenium atom sites (out of three) were found using the SHELXD software (Schneider and Sheldrick, 2002). Initial atom positions were refined using the autoSHARP software package (De La Fortelle and Bricogne, 1997). Resulting phases were improved by the DM program (CCP4, 1994) and used for an automated model building with the Arp/Warp software (Perrakis et al., 1999). The resulting model of about 90% completeness was inspected and finished manually with the Xfit program (McRee, 1999). Restrained refinement enforced by the Refmac5 software was then performed using the native data and followed by addition of water molecules by Arp/Warp (Lamzin and Wilson, 1993). Data collection, phasing, and refinement statistics are presented in Table 5.12. Most of the model has a clear and well interpretable electron density with the exception for the first two N-terminal residues (including one of the methionines) and last six C-terminal residues of the bromodomain construct. Additionally, some solvent exposed side chains had no interpretable electron density. The above parts were omitted in the final model. The R-factor of the presented structure is 25.0% and R-free 26.5%. The structure factors and coordinates will be deposited in the PDB.

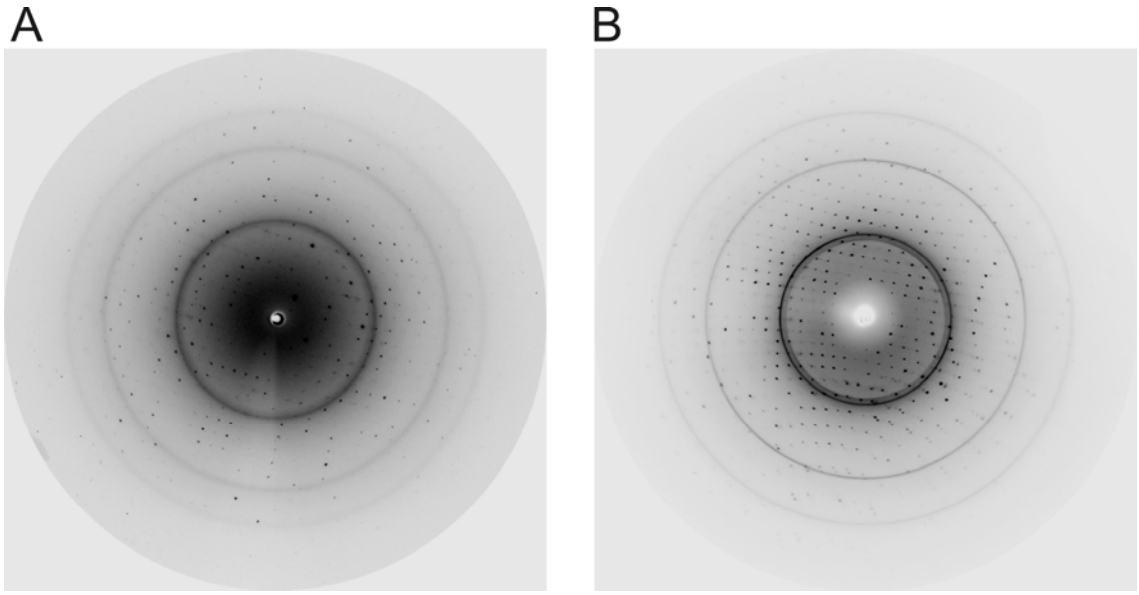

Figure 5.37. *Diffraction patterns of crystals of the bromodomain.* ***A****: A frame from the MAR CCD165 (BW6, DESY, Hamburg). The crystal of the native bromodomain (space group $P2_1$) was rotated 2°. The edge of the image is about 1 Å. (B) A frame from the MAR CCD165 (BW6, DESY, Hamburg). The crystal of native SelMet labeled bromodomain (space group $P2_1$) was rotated 2°. The edge of the image is about 1 Å.*

***Table 5.12**. Data collection and refinement statistics for the exbromodomain*

Data collection	**Native**	**Se-Met:peak**	**inflection**	**high remote**
Space group	$P2_1$	$P2_1$		
Cell constants (Å)				
a	29.86	29.81		
b	30.33	30.46		
c	66.82	66.66		
β	90.27	90.44		
Resolution range (Å)	20-1.5	20-1.8		
Wavelength (Å)	1.05	0.9793	0.9796	0.9770
Observed reflections	83320	151738	63867	64508
Unique reflections	19371	21729	21747	21773
Whole resolution range:				
Completeness (%)	99.1	99.5	99.5	99.3
R_{merge}	2.0	3.0	4.0	4.2
I/σ(I)	40	26.5	19.6	20.5
Last resolution shell:				
Resolution range (Å)	1.5-1.6	1.8-1.9	1.8-1.9	1.8-1.9
Completeness (%)	84.8	77.6	70.9	74.4
R_{merge}	8.3	10	11.1	10.8
I/σ(I)	11	9.7	8.7	8.9
Phasing				
Number of sites found/present	2/3			
Phasing power	1.7			
FOM	0.63			
Refinement				
No. of reflections	18402			
Resolution (Å)	20 – 1.5			
R-factor (%)	25.0			
R_{free} (%)	26.5			
Average B ($Å^2$)	17.7			
R.m.s bond length(Å)	0.007			
R.m.s. angles (°)	1.04			
Content of asymmetric unit				
No. of protein molecules	1			
No. of protein residues/atoms	120/1166			
No. of solvent atoms	165			
No. of Se atoms identified	2			

5.2.5.4 Molecular structure of the bromodomain at 1.5 Å

The crystal structure of the human BRG1 bromodomain was solved using phasing from multiple anomalous diffraction (MAD) and final structure was refined to 1.5 Å using native data. The structure is largely helical (61%) with a small antiparallel β sheet. The helical part of the BRG1 bromodomain is similar to the other bromodomain structures solved, however, the antiparallel β sheet is unique to the BRG1 bromodomain (Fig. 5.38A,B).

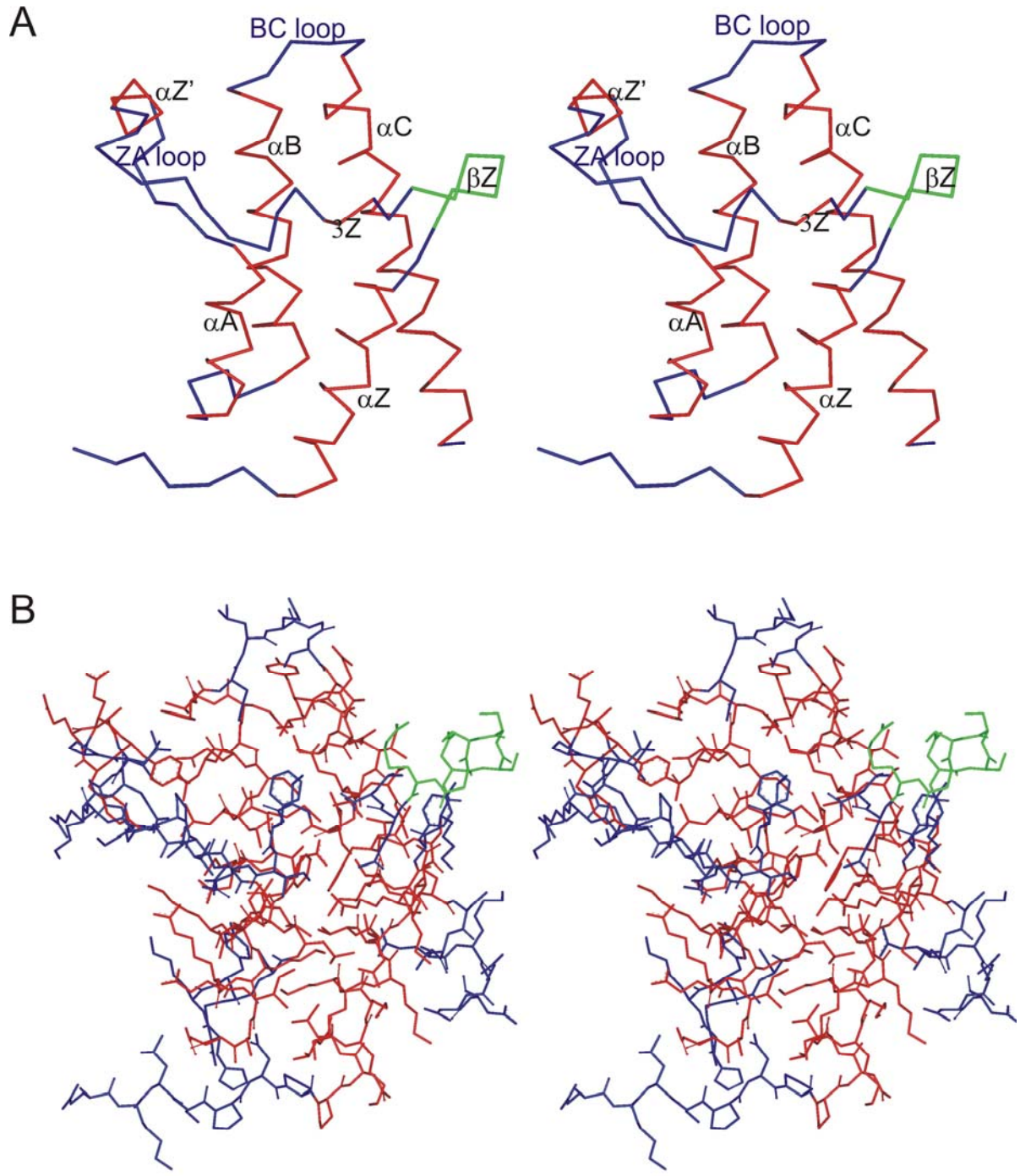

Figure 5.38. *The stereo picture of the BRG1 bromodomain. The helical part is shown in red and novel β-sheet in green. The loops are shown in blue. (A) The backbone representation with the position of secondary structure elements and the loops are marked. (B) A side-chain representation.*

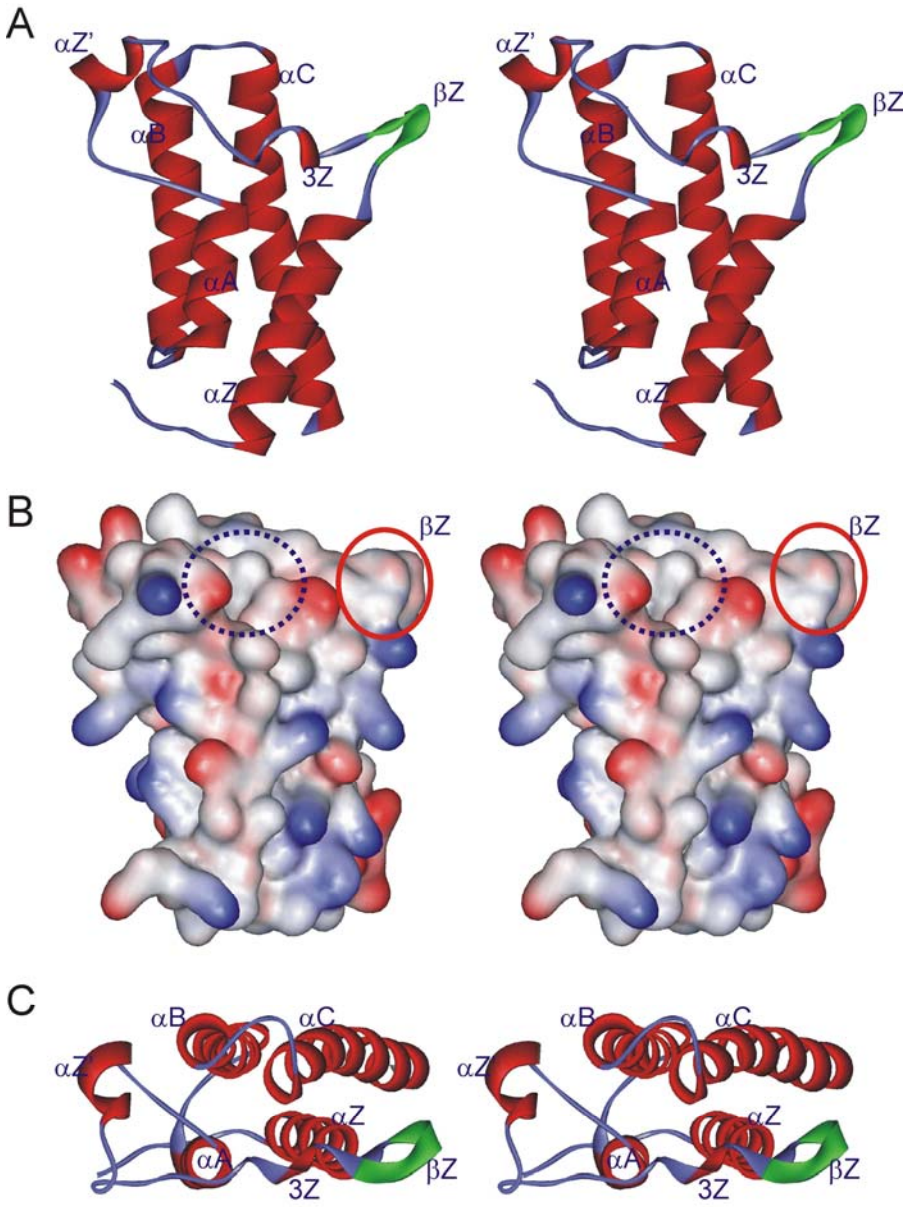

Figure 5.39. *Stereo structures of the BRG1 bromodomain. (A) A ribbon form representation. (B)A surface representation, the position of the β-sheet is marked with a red circle. The acetyl-lysine binding site is marked with a dashed blue circle. (C) Top view representation. All the secondary structure elements are marked.*

The RMSD value for superimposition of the BRG1 bromodomain with the other bromodomains range from 1-1.5 Å for backbone atoms. The helical part consists of five α-helices and a small 3_{10}-helix which are, made up of αZ (residues 1456-1471), 3Z (residues 1482-1483), αZ' (residues 1496-1500), αA (residues 1507-1516), αB (residues 1522-1539), αC (residues 1545-1568) (the helix nomenclature is derived from early predictions of the bromodomains secondary structures and later from the 3D structures by X-ray crystallography as well as NMR spectroscopy {Haynes 1992; Jeanmougin, 1997, Dhallium 1999; Hudson 2000; Owen, 2000}).

The unique antiparallel β-sheet covers residues 1474-1479 (Fig. 5.39A-C). The acetyl-lysine binding site is created by the ZA and the BC loops (Fig. 5.39B)

5.2.5.5 The Novel β sheet in the ZA loop region

The antiparallel βZ-sheet is located between the αZ and 3_{10} (3Z) helices. It is positioned at an angle of approximately 146° in comparison with αZ and approximately 180° in comparison to the αZ' helix. The βZ-sheet and αZ' helix are separated approximately by a distance of 30 Å. The approximate angle of the αZ helix-βZ sheet-αZ' helix is 136°. The βZ sheet is not coincidental as it involves an intricate network of intra-sheet hydrogen bonds and also makes a network of hydrogen bonding with the main chain and solvent (Fig. 5.40 and Table 5.13). A characteristic inter chain H-bond is formed between carbonyl oxygen and amide hydrogen of Asp1474 and Gly1478. The side chain of Arg1479 makes two hydrogen bonds with both the side chain oxygen of Asp1549 (contacting the βZ sheet with the αC helix). The amide hydrogen of Tyr1472 makes a hydrogen bond with the carbonyl group of Lys1471 in αZ helix also helping in stabilizing the sheet. Many other intra-sheet hydrogen bonds that help in stabilizing this structure are listed in Table 5.13.

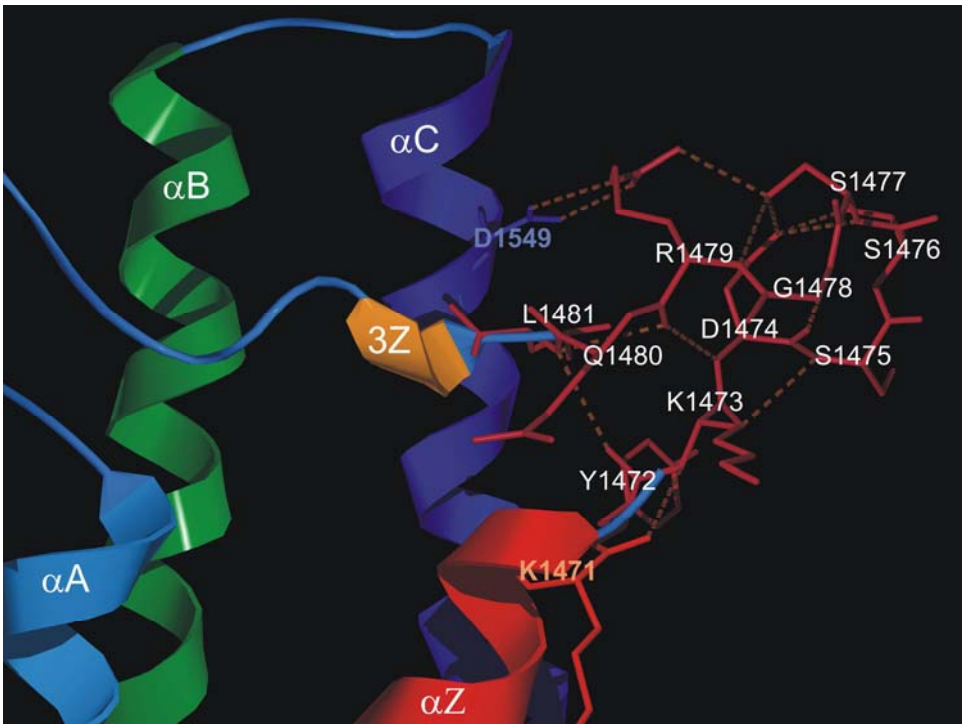

Figure 5.40. *The intricate network of hydrogen bonds involved in stabilizing the βZ-sheet. The amino acids in theβZ-sheet are shown as sticks and marked, hydrogen bonds are shown with dashed lines. The residue Asp1549 in the αC helix and Lys1471 in the αZ-helix are also marked. These residues make hydrogen bonds with the residues in the βZ-sheet. The helices are marked and shown in different colors.*

Table 5.13. *List of hydrogen bonds made by the amino acids in the βZ sheet*

-Group of	Amino acid 1	To -Group of	Amino acid 2	H bond distance (Å)
-NH	Asp1474	-CO	Arg1479	2.90
-CO	Asp1474	-NH	Gly1478	2.69
-δO2	Asp1474	-γO	Ser1476	2.55
-δO2	Asp1474	-NH	Ser1447	2.89
-γO2	Asp1474	-γO	Ser1447	2.67
-NH	Ser1475	-CO	Lys1473	3.28
-γO	Ser1476	-δO2	Asp1474	2.55
-γO	Ser1476	-NH	Ser1477	2.75
-NH	Ser1477	-δO2	Asp1474	2.89
-NH	Ser1477	-γO	Ser1476	2.74
-CO	Ser1477	-NH	Arg1479	3.24
-γO	Ser1477	-δO2	Asp1474	2.67
-γO	Ser1477	-NH	Arg1479	3.22
-γO	Ser1477	-ηNH	Arg1479	3.26
-γO	Ser1477	-NH	Gly1478	3.14
-NH	Gly1478	-CO	Asp1474	2.69
-NH	Gly1478	-γO	Ser1477	3.14
-NH	Arg1479	-γO	Ser1477	3.22
-NH	Arg1479	-CO	Ser1477	3.24
-CO	Arg1479	-NH	Asp1474	2.90
-CO	Arg1479	-NH	Leu1481	3.28
-εN	Arg1479	-δO2	Asp1549	2.97
-γN1	Arg1479	-γO	Ser1477	3.26
-γN2	Arg1479	-δO1	Asp1549	2.79
-γN2	Arg1479	-δO2	Asp1549	3.29

5.2.6 Discussion

5.2.6.1 Functions of the BRG1 bromodomain and DNA binding activities of hBRG1

Bromodomain is a conserved acetyl lysine recognizing and binding domain found in many chromatin associated proteins and in almost all histone acetyltransferases (HATs) (Jeanmougin et al., 1997) and it plays a crucial role in chromatin remodeling and signal transduction (Yang, 2004; Zeng, 2002). The bromodomain of BRG1 has been implicated in translating the histone code. The histone code hypothesis states that different histone modifications such as acetylation, phosphorylation and methylation, are recognized in a specific manner by different effector proteins, thereby translating histone code into action (Strahl and Allis, 2000; Kouzarides, 2002). For example, it has been shown that during activation of the IFN-β gene only a subset of lysines in histone H4 and H3 are acetylated by GCN5 acetyltransferase. These acetylated lysines in turn recruit different promoter containing transcription complexes. Acetylation of the histone H4 Lys8 mediates recruitments of the SWI/SNF complexes; whereas acetylation of Lys9 and Lys14 in histone H3 is critical for the recruitment of TFIID. *In-vivo* the bromodomain of BRG1 in

SWI/SNF complex was capable of interacting with Lys8 of histone H4 and Lys9 and Lys14 of histone H3 (Agalioti at al., 2002). The structure of the BRG1 bromodomain would shed light on its role in these processes.

To remodel chromatin, remodeling complexes have to be able to recognise and bind their substrates. Different mechanisms have been proposed for targeting the SWI/SNF complex to a specific chromosomal locus. In some, the complex is recruited to the promoter through the interaction with sequence specific transcription factors, such as the glucocorticoid receptor or the estrogen receptor (Yoshinaga et al., 1992; Direnzo et al., 2000) or through stable association with the RNA polymerase II holoenzyme (Wilson et a., 1996). However, in other models the recruitment was proposed through the intrinsic DNA binding activities of the SWI/SNF complex. The DNA binding activities have been shown for both yeast and human SWI/SNF remodeling complexes (Côté et al., 1994; Peterson and Herskowitz, 1992). The DNA binding properties of the SWI/SNF complex were suggested to be similar to high-mobility-group (HMG) proteins (Quinn et al., 1996; Vignali et al., 2000). The HMG group of 'architectural' proteins belongs to three families (HMGA, HMGB, and HMGN). These HMGs have different structures, nevertheless, broadly similar functions. They all organize the chromatin by binding, bending, and plasticizing DNA (Reeves and Beckerbauer, 2001; Bianchi and Agresti, 2005). HMGA1 proteins contain AT-hooks; 9-10 amino acid peptide segments that are unstructured in solution, bind to the AT-rich DNA in the minor groove (Reeves, 2001). Upon binding to DNA, an unstructured-to-structured transition in the AT-hook motif occurs as it adopts a crescent like conformation (Huth et al., 1997; Reeves and Nissen, 1990). HMGB proteins contain HMG boxes that bind to a minor groove of DNA with low sequence specificity. HMGN proteins bind inside the nucleosomes (Bianchi and Agresti, 2005). Two individual subunits from the SWI/SNF complex have been shown to have HMG-like DNA binding properties. BAF57 has been shown to contain a HMG1 domain and recombinant BAF57 subunit was shown to interact with the four-way-junction (4H) DNA, with high affinity, and to a linear DNA with low affinity, characteristics for HMGB proteins (Wang et al., 1998). The conformation of 4H DNA is thought to mimic the topology of DNA as it enters or exits the nucleosome and it also mimics the intermediates during the recombination events. The BRM1 subunit of the SWI/SNF complex has been shown to have the DNA binding activity (Bourchot et al., 1999). A region in the C-terminus of the protein has an AT-hook-like motif of HMGA1 proteins. This region of the protein is rich in positively charged residues (Lys and Arg, hence called also the KR region). It acts as a nuclear localization signal (NLS) and harbours the DNA binding activity. The KR region of the BRM1 protein binds to linear DNA with high affinity and to 4H DNA with low affinity, characteristics of HMGA

proteins. This region of BRM1 protein was required for cooperation of BRM1 with the glucocorticoid receptor (Brourchot et al., 1999).

Human BRG1 protein also has a KR region similar to the HMGA-like AT-hook motif. In the BRG1 protein this motif is sandwiched between the ATPase/helicase domain and the bromodomain (Aravind and Landsman, 1998). This raises an important question as to how the AT-hook motif recognizes the DNA in such proteins, as compared with HMGA proteins. The AT-hook motif seems to be an auxiliary motif necessary for cooperation with other DNA-binding activities in the same or different proteins. In the chromatin remodeling complex, such as SWI/SNF, wherein the main function is to remodel the chromatin, it is possible that this site harbours many types of DNA binding activities.

5.2.6.2 AT-hook motif of BRG1 and BRM1

We have compared the AT-hook-containing region from hBRG1 and hBRM1 with the AT-hooks from the HMGA1 protein. The HMGA1 protein has three AT-hook sequences, termed DNA binding domain 1 (DBD1), DBD2, and DBD3. The core of each contains a central Arg-Gly-Arg (RGR) tripeptide and has proline residues at their N- and C-ends. The resulting PRGRP palindrome tends to be flanked at both the ends by positively charged residues. The AT-hooks have been classified into two types, type I and type II. The type I AT-hooks have additional residues at the C terminus in addition to the central PRGRP palindrome, with the flanking charged R and K residues. These residues make an extensive network of hydrophilic and hydrophobic contacts with the sugar phosphate backbone of the DNA, which increases the surface contact area and hence the stability of the complexes, DBD2 of HMGA1 belongs to this category. The types II AT-hooks do not have the central PRGRP palindrome with the flanking charged R and K residues. DBD1 and DBD3 belong to this category (Dragan et al., 2003). The NMR structure of the HMGA1 DBD2/DNA complex revealed that the central 'RGR' core deeply penetrates the minor groove of AT base pair with two arginine residues forming extensive electrostatic and hydrophobic contacts with the floor of the minor groove. The two proline residues on each side of the 'RGR' core direct the protein away from the floor of the minor groove and position the positively charged arginine or lysine near the negatively charged phosphate backbone, making further contacts. The AT-hooks from BRG1 and BRM1 seem to have the characteristics of both types (Fig. 5.41). However, the core 'RGR' in hBRG1 is flanked by a proline residue on one side. A lysine residue at the position of the conserved lysine residue (Lys65) of DBD2 of HMGA1 is present in both, BRG1 and BRM11 AT-hooks. *In-vivo* acetylation of Lys65 in HMGA1 (by the transcriptional coactivator/HAT CBP) was found to turn

off the hIFN-β gene expression. It was postulated that acetylation of Lys65 causes major changes in its interaction with the DNA, resulting in destabilization of the hIFN-β enhanceosome (Munshi et al., 1998). However, using the acetylated peptide (acetylated K65), Dragan et al. (2003) showed that the peptide could still interact with the DNA. Whether the Lys residue at this position in the AT-hooks of BRG1 and BRM1 have any role in *in-vivo* remains to be seen. Based on these observations, the BRG1 and BRM1 AT-hooks resemble DBD2 of HMGA1. These similarities with some differences could show the evolutionary adaptation of AT-hook in these proteins.

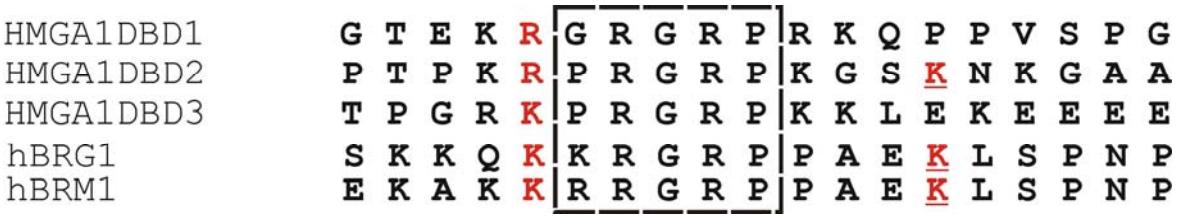

Fig. 5.41. *Comparison of the AT-hook motifs from the HMGA1 protein with the AT-hook motifs of BRG1 and BRM1. The marked box highlights the central region of the AT-hook motif. Homologous positive residues are shown in red, the underlined lysine residue in the HMGA1 DBD2 is acetylated in-vivo, and corresponding lysine residues in the AT-hooks of BRG1 and BRM1 are also underlined.*

5.2.6.3 Interaction of the AT-hook domain from hBRG1 with DNA

The interaction between the SWI/SNF complex and DNA or nucleosomes is not well characterized at the level of SWI/SNF subunits. The direct binding of the SWI/SNF complex to DNA was reported and later interactions of its individual subunits were investigated for binding with the DNA (Cote et al., 1994; Quinn et al., 1996). The BAF57 and BRM1 subunit have been shown to bind to the linear DNA as well as to the 4H DNA (Wang et al., 1998; Brouchot et al., 1999). We have characterized the binding of the HMGA-like region from the hBRG1 subunit of the SWI/SNF complex with the linear and to the 4H DNA. The hBRG1 bromodomain alone could interact with the acetylated lysine peptide from H3 whereas it did not show any binding with the DNA, however, the construct that included the AT-hook region and the bromodomain showed the binding with DNA. Since the segment N-terminal to the AT-hook-like region of hBRG1 has a number of positively charged residues, we made two constructs of protein, the first, named 'AT-hook domain', has the core of the AT-hook motif and the bromodomain, and second, named 'KR-domain' includes the KR rich region and core AT-hook motif plus the bromodomain (Fig. 5.26). The KR-domain binds linear DNA with approximately a two fold lower K_D than the binding of the AT-hook domain to the linear DNA. We thus conclude that the core DNA binding motif is situated in the AT-hook domain, with positive residues around this motif having additive

affect on the interaction. Using NMR spectroscopy we showed that many residues were affected upon binding with DNA. The AT-hook domain showed interaction with both, linear DNA and 4H DNA. A few peaks that were either shifted or perturbed are the same in two cases, suggesting that some common residues are involved in binding with linear and 4H DNA. We speculate that the AT-hook of BRG1 targets the AT rich regions in both types of DNA with extra non-specific interactions in case of AT-hook–4 H DNA interactions. A larger complex of protein-DNA, as expected, in the case of the AT-hook domain-4H DNA complex, produced as expected significant broadening of signals in NMR.

It is known that the HMG group of protein induces bending, unwinding and straightening of DNA. The HMGA protein also changes the topological DNA conformations and induces supercoils into relaxed DNA *in-vitro*. To check if the AT-hook region of hBRG1 has similar activity, we studied the conformational changes in linear DNA and 4H DNA. In the case of linear DNA, the AT-hook domain induces conformational changes by unwinding the DNA. For the HMGA protein, such changes in the topology of DNA have many biological implications. For example, HMGA proteins regulate gene activation by acting as architectural cofactors to facilitate formation of enhanceosomes on the promoter and/or enhancer regions. Human HMGA protein recruits transcription factors by inducing conformational changes within the IFN-β enhancer (Falvo et a., 1995). Many implications can be envisaged for the DNA unwinding properties of the hBRG1 protein. Such activity can have direct role in the chromatin remodeling or play a role in recruitments of transcription factors to promoter or enhancer regions. Indeed, the BRG1 containing SWI/SNF complexes are shown to be recruited to the human immunodeficiency virus promoter and estrogen responsive promoters (Henderson et al, 2004; Direnzo et al., 2000). Fan et al. (2003) have shown that BRG1 makes nucleosomal DNA accessible for regulatory factors by creating DNA loops on the surface of a nucleosome. Our results corroborates with this finding showing that, at least partially, the AT-hook region of BRG1 can be responsible for such activity.

Interestingly, there was no change in the conformation of 4H DNA upon its interaction with the AT-hook domain of BRG1. The rigidity of the open form of 4H DNA may explain the lack of conformational change. Similar kinds of observations were made by Zhang et al. (2003) for the wheat HMGA and 4H DNA interactions.

In summary, we showed that the AT-hook-containing region of hBRG1 harbors the DNA binding activity similar to the HMGA group of proteins. Further structural-functional investigations on individual subunits of the SWI/SNF remodeling complex should help in understanding the mechanism of its interaction with DNA in nucleosomes in details. The

differences in the core of the AT-hook motif in different proteins reflect its evolutionarily adapted role in the interaction with DNA.

5.2.6.4 The BRG1 bromodomain structure vs HAT bromodomains

I have compared the 3D structures of my BRG1 bromodomain with those of the yeast GCN5, human GCN5, TAFII250, P/CAF, and CBP which all belong to the histone acetyl-transferase (HAT) class of proteins. The structures have similar fold of 5 helices, except for the βZ sheet (Fig. 5.42). The RMSD value for the superposition of the BRG1 bromodomain and other bromodomain is given in the Table 5.14. The RMSD value for the helical part is low and major deviations are located in loop regions of these proteins.

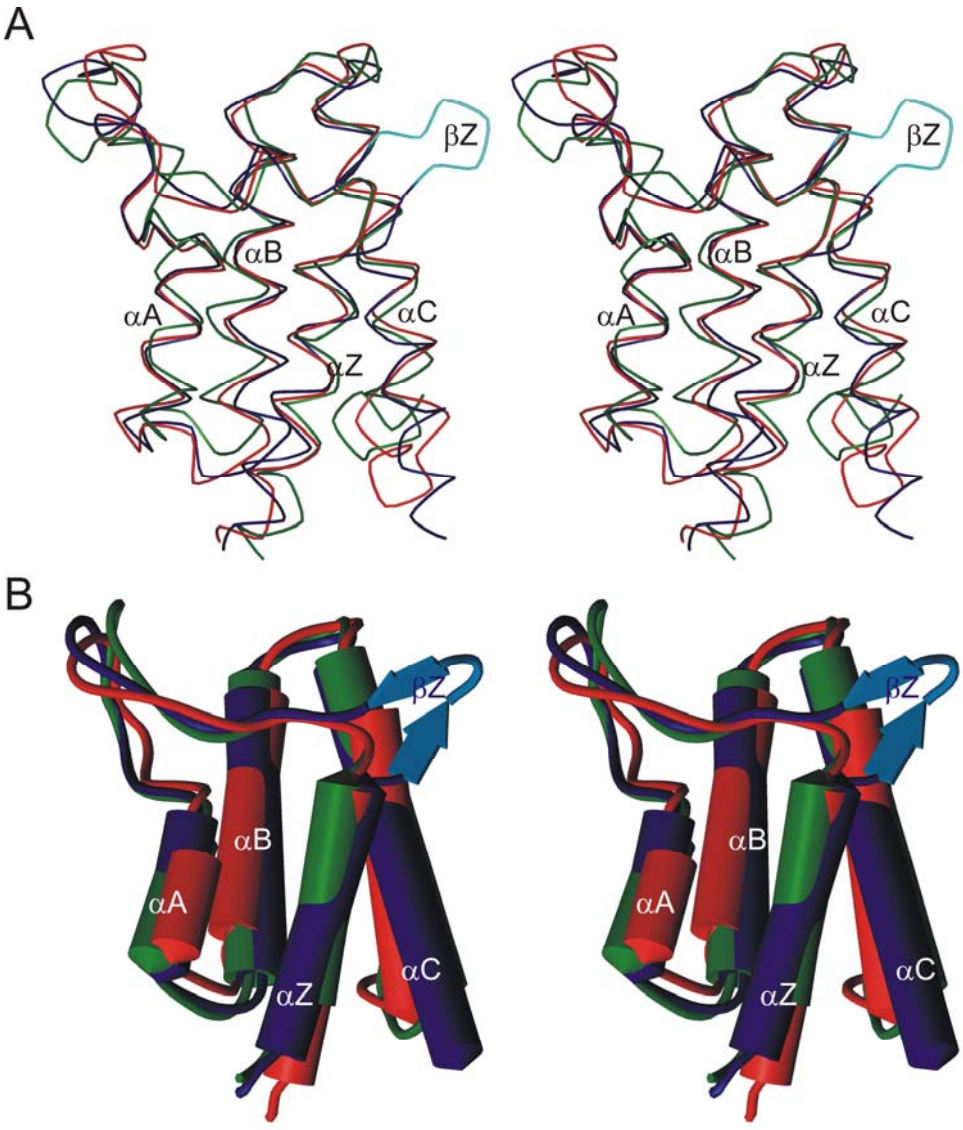

Figure 5.42. *Comparison of the BRG1 bromodomain structure with the yeast GCN5 (X-ray), and P/CAF (NMR) bromodomain structures. The BRG1 bromodomain is shown in blue, the yeast GCN5*

bromodomain in red and the P/CAF bromodomain in green. The novel βZ sheet is colored in light blue. (A) The tube form representation. (B) Schematic representation showing the major helices and βZ-sheet.

Table 5.14. *RMSD of BRG1 bromodomain structure from other published bromodomains*

Comparisons of BRG1 bromodomain (X-ray) with:	***RMSD value (Å)***
Yeast bromodomain (X-ray)	1.09
Human GCN5 bromodomain (NMR)	1.17
CBP (NMR)	1.54
P/CAF (NMR)	1.30
TAFII250 (NMR)	1.50 for both domains

Sequence alignment of all these bromodomains clearly shows that the βZ-sheet is unique to the BRG1 protein. A further interesting observation is that the βZ-sheet seems to be even longer in the case of BRG1 homologue protein, BRM1 (Fig. 5.43). Both BRG1 and BRM1 are the mutually exclusive ATPases of the human SWI/SNF complex and share >80% sequence homology. The insertion point for the βZ sheet seems to be clearly defined by conserved residues. The variable βZ-sheet is guarded by a conserved Ser and Phe residues at N- and C-terminus respectively (Fig. 5.43).

Sequence alignment of various bromodomains also showed a number of highly conserved residues. In the BRG1 bromodomain, these residues are mainly located in ZA and BC loops and are involved in creating an acetyl-lysine binding cleft, in stabilizing the two helixes by hydrophobic interactions, or providing the helix capping activities to nearby helices (Fig. 5.44). These activities show their universally conserved role in all these functions (Table 5.15).

Table 5.15. *List and position of all the conserved residues in the BRG1 bromodomain.*

Conserved residue	**Location in the structure**
Met1462	αZ
Val 1466	αZ
Ser1475	βZ
Phe1485	ZA loop
Pro1489	ZA loop
Tyr1497	αA’
Tyr1498	αA’
Glu1499	αA’
Leu1500	ZA loop
Ile1501	ZA loop
Pro1504	ZA loop
Val1505	ZA loop
Asp1506	αA
Phe1507	αA
Ile1514	αA
Tyr1519	AB loop

Conserved residue	Location in the structure
Leu1532	αB
Asn1535	αB
Phe1539	αB
Asn1540	BC loop
Leu1553	αC

Conserved residues in the helices are involved in stabilizing the two helices by providing hydrophobic interactions. For example, the side chains of conserved Met1462 and Ile1514 make hydrophobic interactions and help in stacking the αZ and αA helices. The amide group of conserved Phe1485 makes a hydrogen bond with the carbonyl group of Ser1482, helping in stabilizing the 3Z helix, also the carbonyl group of Phe1485 makes a hydrogen bond with the amide group of Phe1507 closing the neck of the ZA loop. Conserved Pro1489, Tyr1497, Tyr1498, Glu1499, Leu1500, and Ile1501, make up the acetyl-lysine binding site. The carbonyl group of conserved Pro1489, makes a hydrogen bond with the side chain hydroxyl of conserved Tyr1498 providing the helix capping interaction for αZ' helix.

The carbonyl group of the conserved Tyr1497 makes hydrogen bonds with amide hydrogens of conserved Leu1500 and Ile1501, essentially stabilizing the helix αZ'. The carbonyl group of conserved Pro1504 makes a long distance helix-capping hydrogen bond with the side-chain NH of Lys1509. The carbonyl group of conserved Asp1506 is making three helix-capping hydrogen bonds with the amide hydrogens of Lys1509, Ile1510 and Lys1508, capping the αA helix which is further stabilized by a hydrogen bond between the side chain oxygen of Asp1506 and amide nitrogen of Lys1509 of the αA helix. The side chain oxygen of Asp1506 also makes a hydrogen bond with the Leu1488 in the ZA loop. The amide group of conserved Phe1507 makes a hydrogen bond with the carbonyl group of Phe1485 again linking the neck of the ZA loop. The side chain hydroxyl group of conserved Tyr1519 makes a hydrogen bond with the side chain oxygen of Asp1528 helping in keeping the αB helix in position. The conserved Leu1532 and Asn1535 residues are also involved in making the acetyl-lysine binding sited. The side chain NH of Asn1535 makes a hydrogen bond with the carbonyl group of Lys1503 closing the acetyl-lysine binding cleft from one side. The conserved residues Asn1535, Phe1539 and Asn140 in the BC loop are also involved in creating the acetyl-lysine binding site. The side chain NH of Asn1540 makes a hydrogen bond with the carbonyl group of Ala1536 in the αB helix providing the helix capping interaction (Fig. 5.44).

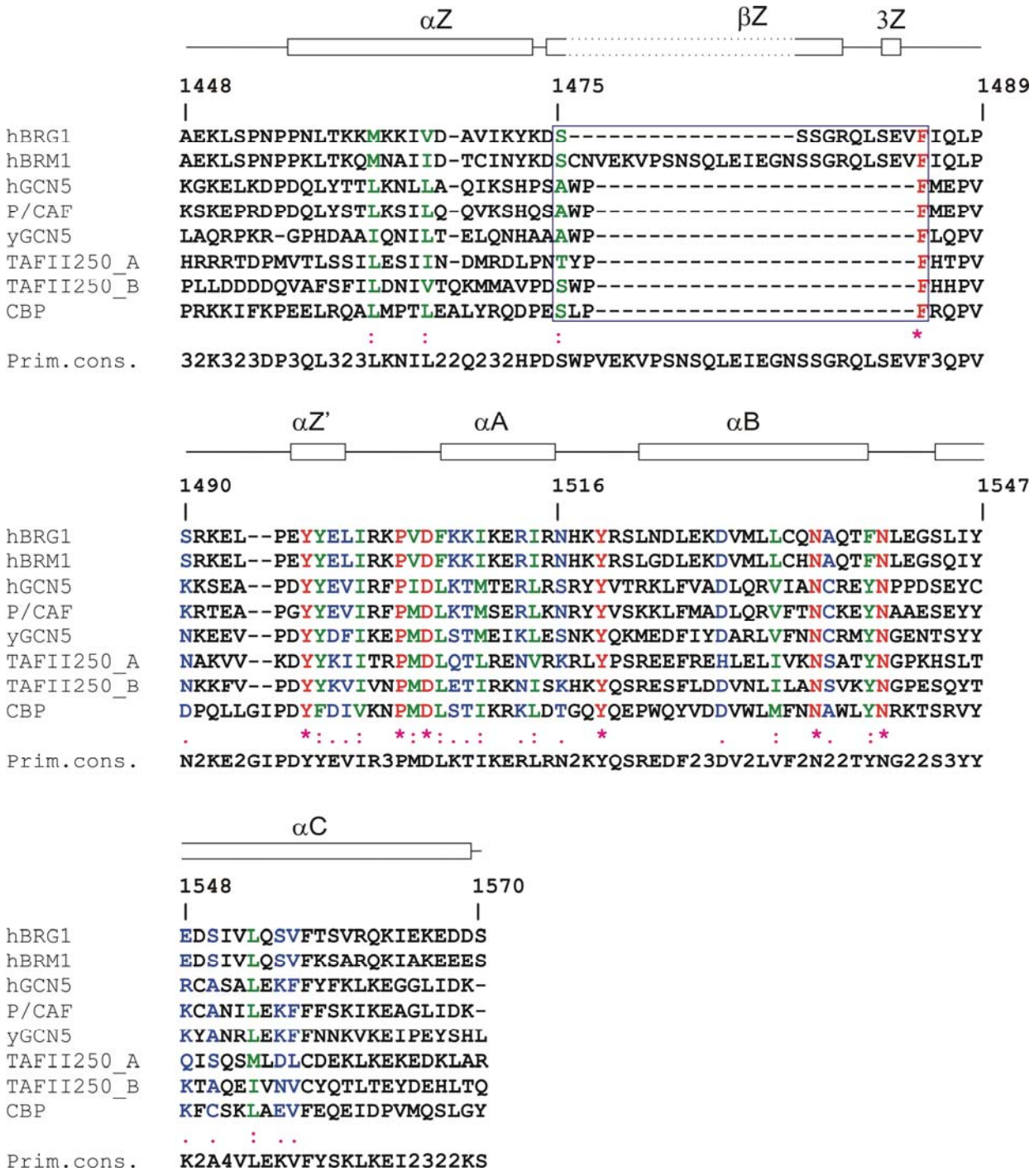

Figure 5.43. *A structurally guided sequence alignment of bromodomains from different proteins obtained using the CLUSTALW sequence alignment tool (Combet et al., 2000). Sequences are compared with the hBRG1 sequence. The amino acid numbering belongs to hBRG1 protein. Secondary structure representation if shown in the top belongs to the hBRG1 bromodomain. Amino acid residues which are identical in all proteins are shown in red and **, strongly similar residues in green and **:**, and similar residues are shown in blue and **.**. Primary consensus sequence (at the bottom) shows the most common residue or the number which shows the set of residues most prominent in all the sequences. TAFII250 has two bromodomains, each bromodomain subunit is aligned separately termed as TAFII250_A and TAFII250_B.*

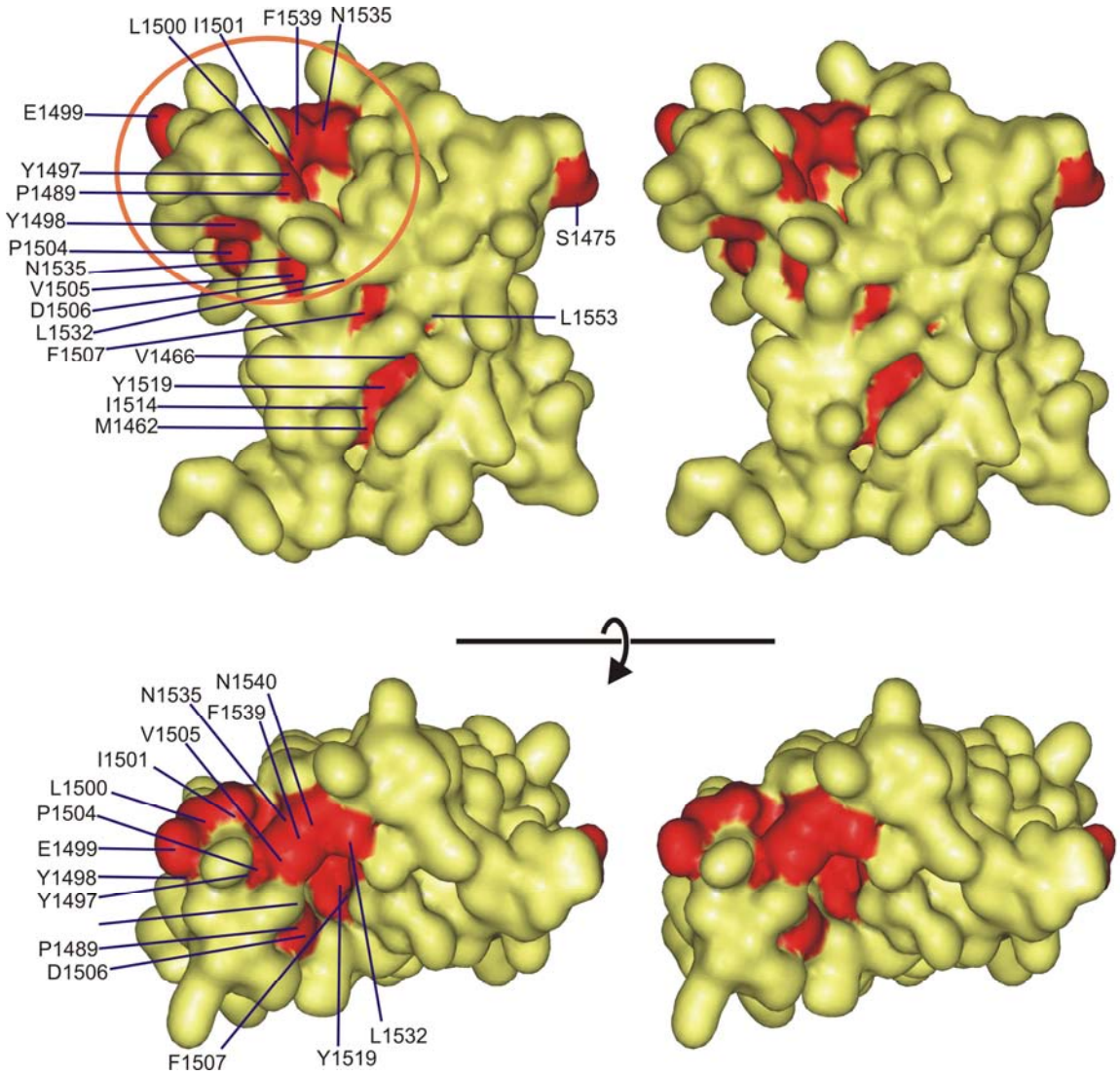

Figure 5.44. *Positions of highly conserved amino acids in the BRG1 bromodomain. Conserved residues are shown in red and marked. (A) A front stereo view of the structure with the conserved residues marked. The circle shows the acetyl-lysine binding site in the structure. (B) A 90° flip of the front view.*

5.2.6.5 The N-acetyl-lysine peptide binding site in the BRG1 bromodomain

A crystal structure of the bromodomain from yeast GCN5, P/CAF, and CBP have been solved with the bound acetylated peptides (Owen et al., 2000; Mujtaba et al., 2002; Mujtaba et al., 2004). The [N-acetylated lysine H4 peptide]-[yGCN5 bromodomain] was solved by X-ray crystallography, whereas [N-acetylated lysine p53 peptide]-[P/CAF bromodomain] and [N-acetylated lysine HIV tat]-[CBP bromodomain] complex structures were deduced by NMR spectroscopy. The peptide binding site is created by residues in the ZA and BC loops in all these structures. Acetylated peptides in yGCN5 and P/CAF binds in a largely extended conformations whereas the acetylated HIV-tat peptide in the CBP bromodomain was shown to adopt a β turn-like conformation with the acetylated lysine positioned at the beginning of the turn. In all structures however, the side chain of the acetylated lysine intercalates into a deep hydrophobic

cavity created by conserved residues of ZA and BC loops. Table 5.16 lists amino acid residues involved in binding an acetylated peptide and the corresponding residues in the BRG1 bromodomain. It is clear from this compilation that there a number of common residues that are involved in binding all acetylated peptides (Table 5.16, color highlighted residues).

Table 5.16. *List of amino acids involved in the acetylated peptide binding. Corresponding residue in the BRG1 bromodomain are listed in the next column. Residues which are common in all three are highlighted with pink and residues which are common in two cases are highlighted in light blue. Three complexes of [bromodomain]-[acetylated peptide] have been solve so far, both by X-ray and NMR methods. Residues N-terminal to the Ac-K are denoted as +1, +2, etc. whereas residues C-terminal to ac-K are denoted as -1, -2, etc.*

[yGCN5]-[histone H4 peptide]		
Residue of H4 peptide (AK(Ac)RHRKILRNSIQGI9	***Residue affected in yGCN5***	***Corresponding residue in BRG1***
Acetyl-lysine	Pro351	--
Acetyl-lysine	Val361	Leu1494
Acetyl-lysine	Tyr364	Tyr1497
Acetyl-lysine	Tyr413	Ile1546/Tyr1547
Acetyl-lysine	Asn407	Asn1540
Acetyl-lysine	Phe352	Phe1485
Acetyl-lysine	Val356	Pro1489
His (Ac-K+2)	Tyr406	Phe1539
His (Ac-K+2)	Phe367	Leu1500
Arg (Ac-K+3)	Arg404	Gln1537
Arg (Ac-K+3)	Asn407	Asn1540
[P/CAF]-[HIV-tat peptide] NMR structure		
Residue of HIV-tat peptide (SYGRK(Ac)KRRQR)	***Residue affected in P/CAF***	***Corresponding residue in BRG1***
Acetyl-lysine	Phe748	Phe1485
Acetyl-lysine	Val752	Pro1489
Acetyl-lysine	Tyr760	Tyr1497
Acetyl-lysine	Ile764	Ile1501
Acetyl-lysine	Tyr802	Phe1539
Acetyl-lysine	Tyr809	Ile1546/Tyr1547
Tyr (Ac-K-3)	Val763	Leu1500
Gln (Ac-K+4)	Glu756	Glu1493
[CBP]-[p53 peptide] NMR structure		
Residue of p53 peptide (SHLKSKKGQSTSRHKK(Ac)LMFK	***Residue affected in CBP***	***Corresponding residue in BRG1***
Acetyl-lysine	Val1115	Pro1489
Acetyl-lysine	Leu1120	Leu1490
Acetyl-lysine	Tyr1125	Tyr1497
Acetyl-lysine	Ile1128	Leu1500
Acetyl-lysine	Tyr1167	Phe1539
Acetyl-lysine	Val1174	Ile1546
Acetyl-lysine	Phe1177	Asp1549
Lys (Ac-K-1)	Tyr1167	Phe1497
Leu (Ac-K+1)	Val1115	Pro1489
Leu (Ac-K+1)	Leu1120	Leu1494
Leu (Ac-K+1)	Ile1122	--

Met (Ac-K+2)	Tyr1167	Phe1539
Met (Ac-K+2)	Val1174	Ile1546

Figure 5.45 shows the positions of all these amino acids in the structure of the BRG1 bromodomain. It is clear that all these residues are creating the groove for acetyl-lysine binding.

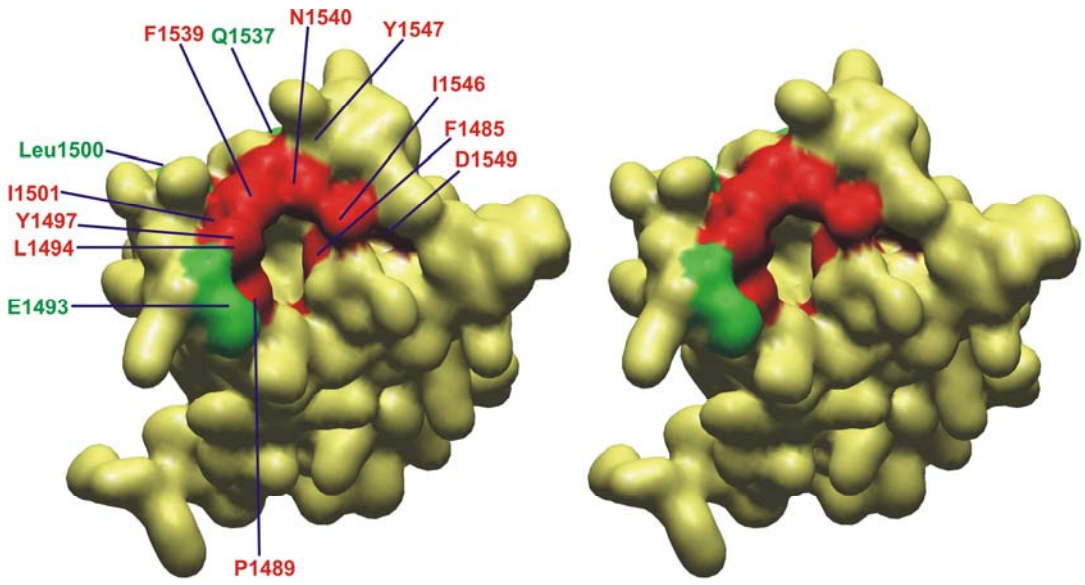

Figure 5.45. *Stereoview of the BRG1 bromodomain showing the residues postulated to be important for the acetylated-lysine peptide binding. The marked residues correspond to amino acids shown to be involved in the acetylated-lysine binding in yGCN5, P/CAF, and CBP. Residues in red are directly involved in binding of an acetyl-lysine residue, and residues in green correspond to the amino acids involved in binding other residues surrounding the acetyl-lysine in peptides.*

Next, I chose the [histone H4 peptide]-[yGCN5] complex to further probe the acetyl-lysine binding site in the BRG1 bromodomain. This was done for two reasons: (A) this structure has been determined by X-ray diffraction method similar to the BRG1 bromodomain and (B) the RMSD value for this structure and BRG1 bromodomain structure is smallest in comparison to other structures (Table 5.14). We modeled the acetylated H4 peptide using the pdb file 1E6I {[yGCN5-acetylated lysine H4 peptide]-[yGCN5 bromodomain] complex}. The modeling shows that the peptide would bind to the protein in the hydrophobic groove in a manner similar to the yGCN5 protein (Fig. 5.46).

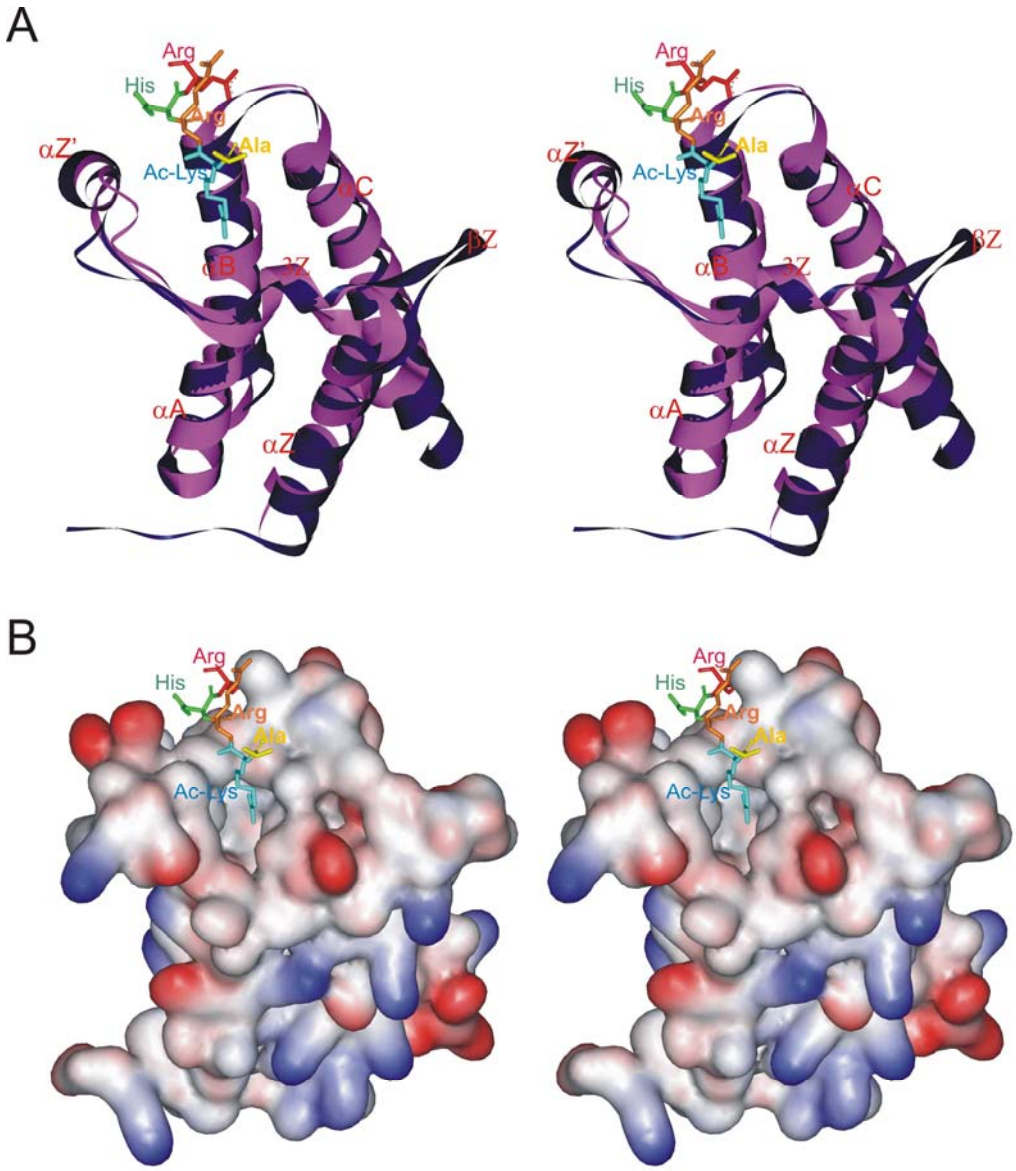

Figure 5.46. *Model of an acetylated-lysine H4 peptide bound to the BRG1 bromodomain, modeling was done using the PDB file 1E6I for GCN5 and BRG1 bromodomains. (A) A ribbon form of superimposed structures of [N-acetylated lysine H4 peptide]-[yGCN5 bromodomain] violet, and BRG1 bromodomain blue. (B) The surface representation of the BRG1 bromodomain with the modeled N-acetylated lysine H4 peptide.*

In the yGCN5 structure the acetylated-lysine H4 peptide was shown to bind to the bromodomain in an extended conformation with the major determinant being the acetylated lysine (Owen et al., 2000). This residue sits in the deep cleft, which is accessible to solvent on one side. The wall of the cleft is formed by the Val361, Tyr364, Pro351, and Tyr413. This accounts for the high hydrophobic potential of the groove. The interaction that plays the major role in orientating the N-acetyl group, such that its carbonyl points towards the hydrophobic interior of the protein, is a hydrogen bond formed between the amide nitrogen of Asn407 and oxygen of the acetyl carbonyl group (Fig. 5.46A and Fig. 5.47A). Also the N-acetyl group is kept

in a fixed position by a network of its hydrogen bonding among the solvent and other carbonyl groups and side chains of the protein, in particular the carbonyl group of Pro351 and hydroxyl group of Tyr364 (Fig. 5.47A). The methyl of the N-acetyl group is surrounded at the base of the pocket by the exposed hydrophobic surfaces of Phe352 and Val356. Apart from the acetyl-lysine residue in the H4 peptide, the amino acids at positions Ac-K+2 and Ac-K+3, His and Arg respectively also make interactions with the protein. The histidine residue sits in a shallow hydrophobic pocket formed by Tyr406 and Phe367. The Arg residue makes two hydrogen bonds to the backbone carbonyl group of Arg404 and Asn407 (Fig. 5.47A).

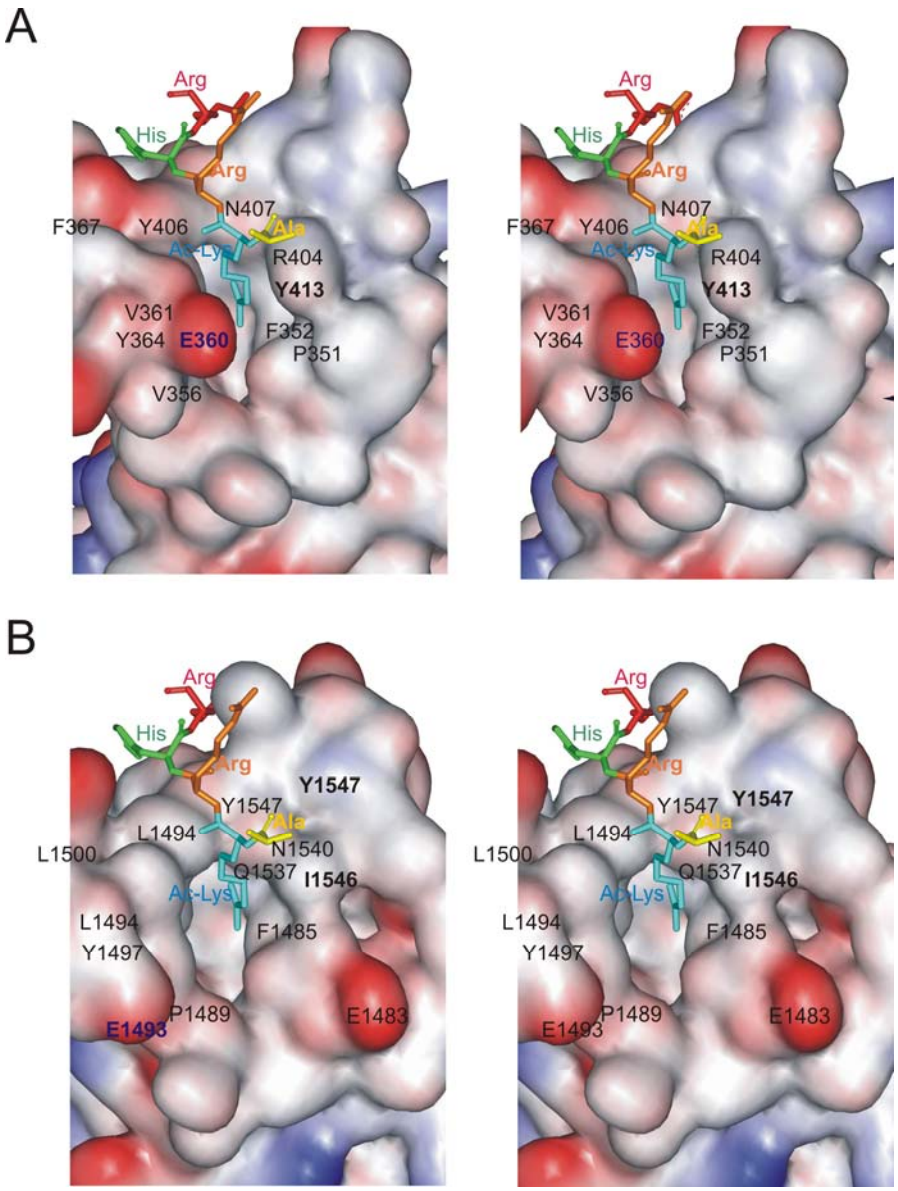

Figure 5.47. *Enlarged view of acetyl-lysine binding site in (A) GCN5 bromodomain and (B) BRG1 bromodomain. The bound peptide is modeled using WebLabView program. Residues of the bromodomain involved in binding are shown. Residues of the peptide are marked with three letter code.*

In the model of [BRG1 bromodomain]-[Acetylated-lysine H4 peptide] complex, I postulate that the N-acetyl-lysine would sit in the hydrophobic cavity formed by residues Leu1494, Tyr1494, Ile154, and Tyr1547 (Fig. 5.47B). These are the residues corresponding to the Val361, Tyr364, Pro351, and Tyr413 of the yGCN5 protein. Also in my modeled peptide the oxygen of the acetyl carbonyl group of acetylated-lysine is at the hydrogen bonding distance from Asn1540 residue, in a way very similar to yGCN5 protein (Fig. 5.47A). I postulate further that this interaction would also primarily help in orienting the acetyl-lysine side chain towards the hydrophobic interior of the protein in a manner similar to the yGCN5 bromodomain.

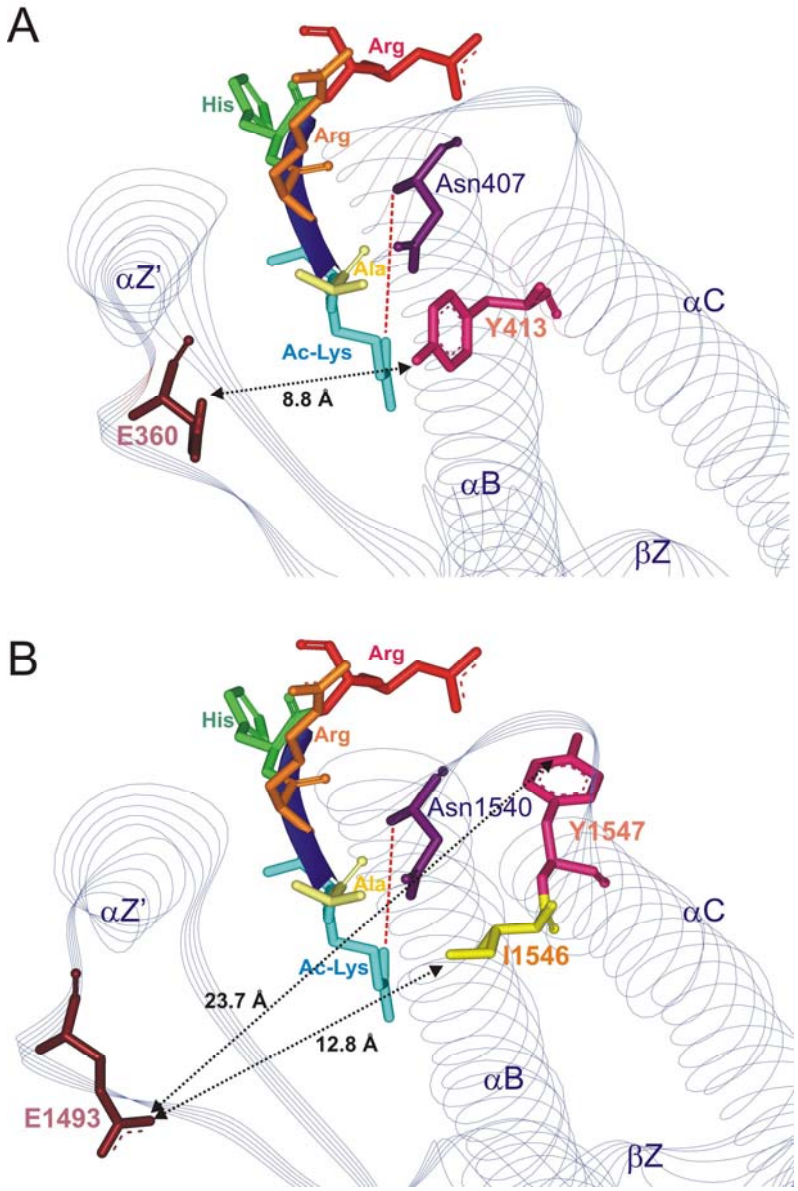

Figure 5.48. *Position of N-acetylated H4 peptide in comparison with the bromodomain residues involved in keeping it in a proper orientation. The amino acids of the peptide are shown with three letter codes and the amino acids of the bromodomains are shown with single letter code and number. Dotted lines in black with arrows show the approximate distances between the two residues. The dotted lines in red represent*

the hydrogen bond (A) Positions of Glu360, Tyr413 and Asn407 in comparison with the N-acetylated H4 peptide in the yGCN bromodomain. (B) Positions of Glu1493, Tyr1547, Ile1546, and Asn1540 in comparison with the N-acetylated H4 peptide from H4 peptide modeled with the BRG1 bromodomain.

However, some differences are clear. In the yGCN5 bromodomain the hydrophobic cleft is complementary to the acetyl-lysine moiety. This is because the opening of this cleft is closed by Glu360 (at one end) and Tyr413 (at the other end). The distance between these amino acids is ~8.8 Å as a result these two residues primarily would make this cavity smaller (Fig. 5.48A). This complementary association between the acetylated lysine and the protein would make this interaction a slow-exchange process by decreasing the off-rate. In the BRG1 bromodomain the amino acid corresponding to Glu360 is Glu1493 and amino acid corresponding to Tyr413 is Ile1546 however there is Tyr1547 next to Ile1546. These amino acids (Glu1493, Ile1546, and Tyr1547) are oriented in different ways (Fig. 5.48B). The distance between Glu1493 and Ile1546 is 12.8 Å and between Glu1493 and Tyr1547 is 23.7 Å. This make the acetyl lysine cleft open and more solvent accessible in comparison with the yGCN5. We postulate that this would make the [N-acetyl-lysine]-[BRG1 bromodomain] interaction to be in fast-exchange time scale by increasing the off-rates of the reaction. However, this would still allow the acetyl-lysine binding albeit weakly.

Overall we showed that the BRG1 bromodomain has significant differences in comparison to other bromodomain structures. This is the first bromodomain structure from any ATP dependent chromatin remodeling subunit. Previously published structures are from the HAT group of proteins. These differences are clearly seen in the βZ-sheet and in the N-acetyl-lysine binding site. Acetylated-lysine binding by bromodomains is an accepted theme now, however, bromodomains from different proteins have been shown to have different preferential affinities for the acetylated-lysine in target proteins. The differences in the structure could account for an evolutionarily adapted role of bromodomains in recognizing the acetylated target proteins preferentially.

6 Summary

Two projects were undertaken in this thesis. The first project involved functional investigations on the retinoblastoma protein (pRb) and its interaction with binding partners. The second project involved the characterization of functional domains of the human BRG1 protein including the 3D structure determination of BRG1 bromodomain.

The retinoblastoma tumor suppressor protein (pRb) is a key negative regulator of cell proliferation that is frequently deregulated in human cancer. More than 130 proteins have been reported to bind to pRb directly. Many of these interactions were reported with the pRb small pocket, which is the major focus of tumorigenic mutations in pRb, and comprises the A and B cyclin - like domains (amino acids 379-578 and 641-791, respectively). However, results found in the literature are ambiguous and contradictory. We have investigated some of these interactions using purified proteins and *in-vitro* biophysical and biochemical methods, which included, nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry, isothermal titration calorimetry (ITC), the affinity chromatography pull-down assays, and gel filtration chromatography. The interactions of pRb with the following proteins were checked using this approach: MyoD, Id-2, E2F1, HDAC1, PAI2, HPV E7, SV40 large T antigen, BRG1, and gankyrin.

Myogenic differentiation promoting, MyoD and inhibiting, Id-2, factors were reported to interact with the pRb pocket domain to coordinate the cell growth and differentiation during the process of myogenesis. There have been contradictory reports in the literature regarding the direct interaction between pRb and MyoD, pRb and Id-2 proteins. We probed these interactions using our multimethod *in-vitro* approach. Using this approach and we were able to document interactions between pRb and HPV-E7, pRb and SV40 large T antigen, MyoD and DNA, and MyoD and Id-2. However using the same approach, we could unambiguously show that there is no direct protein-protein interaction *in-vitro* between the small pocket domain of pRb and the bHLH domain of MyoD and between the small pocket domain of pRb and Id-2. We propose that the indirect interactions, through additional binding partners in multiprotein complexes or modulation of gene expression levels of these proteins, are therefore their probable mode of action.

Many viral oncoproteins (for example, HPV E7, E1A) are known to bind to the pRb pocket domain via a LXCXE binding motif. There are also some 20 cellular proteins that contain a LXCXE motif and have been reported to associate with the pocket domain of pRb. Using the *in-vitro* methods including NMR spectroscopy and isothermal calorimetry titration, I showed that

LXCXE peptides of viral oncoproteins bind strongly to the pocket domain of pRb. Additionally, I showed that the LXCXE-like peptides of histone deacetylase 1 (HDAC1) bind to the same site on pRb, with however a weak (micromolar) and transient association. Systematic substitution of residues other than conserved L, C, and E show that the residues flanking the LXCXE are important for the binding. Positively charged amino acids in the XLXCXEXXX sequence significantly weaken the interaction. Also I could show that the interactions among the peptide residues are important for orienting the LXCXE motif in a proper conformation. I discuss the ramification of the weak binding of IXCXE sequences of HDAC1 to pRb.

The BRG1 protein is the central conserved ATPase of the human SWI/SNF class of the ATP dependent chromatin remodeling complex. Human BRG1 protein has an ATPase domain, an AT-hook containing region, and a bromodomain. An AT-hook motif, first described in the high-mobility-group of non-histone chromosomal proteins HMGA1/2 (formally called HMGI/Y), is a DNA binding motif. Bromodomain is an acetyl-lysine recognizing domain found in many chromatin-associated proteins.

The AT-hook motif binds to the AT rich DNA sequences in the minor groove of B-DNA in a non-sequence specific manner. In this study I cloned and purified a fragment of hBRG1 which encompasses the AT hook region and the bromodomain. NMR and CD spectroscopy show that this entire recombinant domain is well-structured in solution. Functionalities of subdomains were checked by their interactions with N-acetylated peptides from histones. I also characterized the interaction of the AT-hook region with DNA. The ITC data shows that the primary interaction is through the AT-hook motif. The AT-hook region interacts with the linear DNA by unwinding it. These properties resemble the characteristics of the HMGA1/2 proteins and their interaction with DNA.

BRG1 and BRM1 ATPase subunits of the SWI/SNF complex also contain a bromodomain at the C-terminal end of the protein. Here I report the crystal structure of human BRG1 bromodomain at a 1.5 Å resolution. This is a first structure of a bromodomain from a subunit of the ATP dependent chromatin remodeling complex. The four helices in the structure are similar to the helices from other published bromodomains from histone acetyltransferase (HAT) class of proteins (GCN5, P/CAF, CBP, and double bromodomain module of TAFII250), however, a novel small anti-parallel β-sheet is present in the ZA loop region (termed βZ-sheet) of the BRG1 bromodomain. Sequence alignment shows that this βZ-sheet is present only in hBRG1 and hBRM1 proteins and between these two it is longer in the case of hBRM1 protein. *In-silico* modeling shows that the N-acetyl-lysine peptide would bind in a manner similar to other bromodomain. However, we postulate that in BRG1 bromodomain the N-acetyl lysine binding

groove is more open, and this would make [N-acetylated-lysine peptide]-[BRG1 bromodomain] interaction weaker. These changes in the acetyl-lysine binding site could account for the evolutionarily adapted functions of bromodomain in different proteins.

Overall the results of this work should help in understanding and delineating the functions pRb and BRG1 in the important processes of cell cycle regulation and chromatin remodeling.

7 Zusammenfassung

In dieser Doktorarbeit wurden zwei Projekte durchgeführt. Das erste behandelt Untersuchungen der Funktion des Retinoblastoma Proteins und seiner Wechselwirkung mit Bindungspartnern. Das zweite Projekt beschreibt die Charakterisierung von funktionalen Domänen des humanen BRG1 Proteins und die Bestimmung der 3D Struktur der BRG1 Bromodomäne.

Das Retinoblastoma Tumor Suppressor Protein (pRb) ist ein negativer Schlüsselregulator der Zellproliferation, der häufig in menschlichen Tumoren dereguliert ist. Von mehr als 130 Proteine wurde eine direkte Bindung an pRb nachgesagt. Viele dieser Wechselwirkungen sind mit der kleinen Pocket Domäne von pRb, welche die A und B Cyclin-ähnlichen-Domänen (AS 379-578 und 641-791) umfasst, in der bevorzugt Mutationen auftreten, die bei der Krebsentstehung beobachtet wurden. Jedoch sind die Ergebnisse in der Literatur teilweise zweideutig und widersprüchlich. Wir haben einige dieser Interaktionen an gereinigten Proteinen mit verschiedenen Methoden, darunter *in-vitro* biophysikalische und biochemische Methoden wie Kern Spin Resonanz Spektroskopie (NMR), Massenspektrometrie, Isotermale Titrations Calorimetrie (ITC), Affinitäts-Chromatographie *pull-down* Experimente und Gelfiltrations Chromatographie getestet. Mit Hilfe dieses Ansatz wurden folgende Interaktionspartner von pRb untersucht: MyoD, Id-2, E2F1, HDAC1, PAI2, HPV E7, SV40 *large T* Antigen, hBRG1, und Gankyrin.

Die Muskelzelldifferenzierung fördernde Faktoren wie MyoD und inhibierende Faktoren wie Id-2, wurden als Interaktionspartner der kleinen Pocket Domäne von pRb beschrieben, die durch direkte Wechselwirkung mit pRb das Zellwachstum und die Zelldifferenzierung während der Myogenese steuern. Es gib aber widersprüchliche Angaben zu einer direkten Wechselwirkung zwischen pRb einerseitz und MyoD bzw. Id-2 andererseits. Wir überprüften die Interaktion mit dem Multimethoden Ansatz *in vitro*. Mit diesem Ansatz konnte wir zeigen, dass es direkte Interaktionen zwischen pRb und HPV E7, pRb und SV40 *large T* Antigen, MyoD und DNA sowie MyoD und Id-2 gibt. Mit dem gleichen Ansatz konnten wir zweifelsfrei zeigen, dass es keine direkte Wechselwirkung zwischen der kleinen Pocket Domäne von pRb und der bHLH Domäne von MyoD sowie zwischen der kleinen Pocket Domäne von pRb und Id-2 gibt. Indirekte Wechselwirkung über weitere Bindungspartner in einem Multiproteinkomplex oder die Modulation der Genexpression sind daher die wahrscheinlicheren Aktionen.

Viele virale Onkoproteine (wie z.B.: HPV E7, E1A) binden an die kleine Pocket Domäne von pRb über ihr LXCXE Bindungsmotiv. Es wurde ebenfalls von ca. 20 zelluläre Proteine, die ein LXCXE Motiv haben, berichtet, dass sie mit der kleinen Pocket Domäne von pRB assoziiert

sind. Mit Hilfe von *in vitro* Methoden wie NMR und ITC konnte ich zeigen, dass die LXCXE Peptide von viralen Onkoproteinen stark an die kleine Pocket Domäne von pRb binden. Zusätzlich habe ich gezeigt, dass das LXCXE ähnliche Peptid der Histon Deacetylase 1 (HDAC1) an die gleiche Bindungsstelle an pRb bindet, jedoch mit einer schwächeren (im micromolaren Bereich) und transienten Bindung. Systematischer Ersatz von Aminosäureresten außer den konservierten Resten L, C und E zeigte, dass die flankierende Reste des LXCXE Motivs wichtig für die Bindung an pRb sind. Positiv geladene Reste in der XLXCXEXXX Sequenz schwächen die Bindung signifikant. Ich konnte ebenfalls zeigen, dass die intramolekularen Wechselwirkungen in den LXCXE Peptiden wichtig sind um die Aminosäuren für eine optimale Orientierung für die Bindung an pRb auszurichten.

Das BRG1 Protein ist die zentrale, konservierte ATPase der humanen SWF/SNF Klasse von ATP abhängigen Chromatin-Remodulierenden Komplexen. Das humane BRG1 Protein hat eine ATPase Domäne, ein AT-*hook* Region und eine Bromodomäne. Ein AT-*hook* Motiv, ein DNA Bindungsmotiv, wurde erstmals in der *high-mobility-group* des Nicht-Histon chromosomalen Proteins HMGA 1/2 (auch HMGI/Y genannt) gefunden. Die Bromodomäne ist eine Acetyllysin erkennende Domäne, die in vielen Chromatin assoziierten Proteinen vorkommt.

Das AT-*hook* Motiv bindet unspezifisch an die kleine Furche in AT reiche DNA Sequenzen von B-DNA. In dieser Arbeit habe ich ein Fragment von hBRG1, welche die AT-*hook* Region und die Bromodomäne umfassten, kloniert und gereinigt. NMR und CD Spektroskopie zeigten, dass die rekombinante Domäne gefaltet in Lösung vorliegt. Die Funktion der Subdomäne wurde durch Interaktionsstudien mit N-Acetylierten Peptiden von Histonen überprüft. Ebenso habe ich die AT-*hook* Region auf ihre DNA Bindung charakterisiert und ITC Daten zeigten, dass die Hauptinteraktion mit DNA über das AT-*hook* Motiv geht. Das AT-*hook* Motiv interagiert mit linearer DNA indem es sie aufwindet. Diese Eigenschaften gleichen den Charakteristika des HMGA 1/2 Proteins und dessen Interaktion mit DNA.

Die BRG1 und BRM1 ATPase Untereinheiten des SWI/SNF Komplex beinhalten auch eine Bromodomäne am C-terminalen Ende des Proteins. Ich stelle im letzten Abschnitt meiner Doktorarbeit die Kristallstruktur der hBRG1 Bromodomäne bei einer Auflösung von 1.5 Å vor. Dies ist die erste Struktur einer Bromodomäne einer Untereinheit des ATP abhängigen Chromatin-Remodulierenden Komplex. Die vier Helixes in der Struktur sind ähnlich wie die Helixes in den anderen publizierten Bromodomänen der Histone Acetyltransferase (HAT) Klasse von Proteinen (GCN5, P/CAF, CBP, und dem doppel Bromodomäne Module von TAFII250), angeordnet, jedoch gibt es ein neues kleines antiparalleles β-Faltblatt (βZ-Faltblatt genannt) in der ZA *loop* Region der hBRG1 Bromodomäne. Sequenzvergleiche zeigten, dass dieses

β-Faltblatt ausschließlich in hBRG1 und hBRM1 Proteinen, wobei im Falle von hBRM1 der Bereich größer ist, vorkommt. *In-silico* Modelle zeigten, dass das N-Acetyllysin Peptid in gleicher Weise an die Bromodomäne von BRG1 bindet wie in den Fällen der anderen Bromodomänen andere Proteine. Jedoch ist die Bindungsfurche der BRG1 Bromodomäne größer, was eine schwächer Bindung zwischen N-Acetyllysin Peptiden und der hBRG1 Bromodomäne vermuten lässt. Diese Unterschiede in der Acetyllysin Bindungsstelle könnten das Ergebnis einer evolutionären Anpassung an die unterschiedlichen Funktionen der Bromodomänen in den verschiedenen Proteinen sein.

Die Ergebnisse dieser Arbeit geben ein besseres Verständnis und Bild über die Funktionen von pRb und hBRG1 in den wichtigen Prozessen der Zellzyklus Regulation und der Chromatin-Remodulierung.

8 Appendix

8.1 Full Length retinoblastoma protein sequence

```
     1          11         21         31         41         51
     |          |          |          |          |          |
   1 MPPKTPRKTA ATAAAAAAEP PAPPPPPPPE EDPEQDSGPE DLPLVRLEFE ETEEPDFTAL    60
  61 CQKLKIPDHV RERAWLTWEK VSSVDGVLGG YIQKKKELWG ICIFIAAVDL DEMSFTFTEL   120
 121 QKNIEISVHK FFNLLKEIDT STKVDNAMSR LLKKYDVLFA LFSKLERTCE LIYLTQPSSS   180
 181 ISTEINSALV LKVSWITFLL AKGEVLQMED DLVISFQLML CVLDYFIKLS PPMLLKEPYK   240
 241 TAVIPINGSP RTPRRGQNRS ARIAKQLEND TRIIEVLCKE HECNIDEVKN VYFKNFIPFM   300
 301 NSLGLVTSNG LPEVENLSKR YEEIYLKNKD LDARLFLDHD KTLQTDSIDS FETQRTPRKS   360
 361 NLDEEVNVIP PHTPVRTVMN TIQQLMMILN SASDQPSENL ISYFNNCTVN PKESILKRVK   420
 421 DIGYIFKEKF AKAVGQGCVE IGSQRYKLGV RLYYRVMESM LKSEEERLSI QNFSKLLNDN   480
 481 IFHMSLLACA LEVVMATYSR STSQNLDSGT DLSFPWILNV LNLKAFDFYK VIESFIKAEG   540
 541 NLTREMIKHL ERCEHRIMES LAWLSDSPLF DLIKQSKDRE GPTDHLESAC PLNLPLQNNH   600
 601 TAADMYLSPV RSPKKKGSTT RVNSTANAET QATSAFQTQK PLKSTSLSLF YKKVYRLAYL   660
 661 RLNTLCERLL SEHPELEHII WTLFQHTLQN EYELMRDRHL DQIMMCSMYG ICKVKNIDLK   720
 721 FKIIVTAYKD LPHAVQETFK RVLIKEEEYD SIIVFYNSVF MQRLKTNILQ YASTRPPTLS   780
 781 PIPHIPRSPY KFPSSPLRIP GGNIYISPLK SPYKISEGLP TPTKMTPRSR ILVSIGESFG   840
 841 TSEKFQKINQ MVCNSDRVLK RSAEGSNPPK PLKKLRFDIE GSDEADGSKH LPGESKFQQK   900
 901 LAEMTSTRTR MQKQKMNDSM DTSNKEEK
```

Most frequently used pRb constructs composition

Construct name	Amino acid (from-to) in sequence
pRb-N	257 to 378
pRb-ABC	379 to 928
pRb-ALB	379 to 791
pRb-AB	379 to 578 and 642 to 791

8.2 Full length human BRG1 protein sequence

```
     1          11         21         31         41         51
     |          |          |          |          |          |
   1 MSTPDPPLGG TPRPGPSPGP GPSPGAMLGP SPGPSPGSAH SMMGPSPGPP SAGHPIPTQG    60
  61 PGGYPQDNMH QMHKPMESMH EKGMSDDPRY NQMKGMGMRS GGHAGMGPPP SPMDQHSQGY   120
 121 PSPLGGSEHA SSPVPASGPS SGPQMSSGPG GAPLDGADPQ ALGQQNRGPT PFNQNQLHQL   180
 181 RAQIMAYKML ARGQPLPDHL QMAVQGKRPM PGMQQQMPTL PPPSVSATGP GPGPGPGPGP   240
 241 GPGPAPPNYS RPHGMGGPNM PPPGPSGVPP GMPGQPPGGP PKPWPEGPMA NAAAPTSTPQ   300
 301 KLIPPQPTGR PSPAPPAVPP AASPVMPPQT QSPGQPAQPA PMVPLHQKQS RITPIQKPRG   360
 361 LDPVEILQER EYRLQARIAH RIQELENLPG SLAGDLRTKA TIELKALRLL NFQRQLRQEV   420
 421 VVCMRRDTAL ETALNAKAYK RSKRQSLREA RITEKLEKQQ KIEQERKRRQ KHQEYLNSIL   480
 481 QHAKDFKEYH RSVTGKIQKL TKAVATYHAN TEREQKKENE RIEKERMRRL MAEDEEGYRK   540
 541 LIDQKKDKRL AYLLQQTDEY VANLTELVRQ HKAAQVAKEK KKKKKKKKAE NAEGQTPAIG   600
 601 PDGEPLDETS QMSDLPVKVI HVESGKILTG TDAPKAGQLE AWLEMNPGYE VAPRSDSEES   660
 661 GSEEEEEEEE EEQPQAAQPP TLPVEEKKKI PDPDSDDVSE VDARHIIENA KQDVDDEYGV   720
 721 SQALARGLQS YYAVAHAVTE RVDKQSALMV NGVLKQYQIK GLEWLVSLYN NNLNGILADE   780
 781 MGLGKTIQTI ALITYLMEHK RINGPFLIIV PLSTLSNWAY EFDKWAPSVV KVSYKGSPAA   840
 841 RRAFVPQLRS GKFNVLLTTY EYIIKDKHIL AKIRWKYMIV DEGHRMKNHH CKLTQVLNTH   900
 901 YVAPRRLLLT GTPLQNKLPE LWALLNFLLP TIFKSCSTFE QWFNAPFAMT GEKVDLNEEE   960
 961 TILIIRRLHK VLRPFLLRRL KKEVEAQLPE KVEYVIKCDM SALQRVLYRH MQAKGVLLTD  1020
1021 GSEKDKKGKG GTKTLMNTIM QLRKICNHPY MFQHIEESFS EHLGFTGGIV QGLDLYRASG  1080
1081 KFELLDRILP KLRATNHKVL LFCQMTSLMT IMEDYFAYRG FKYLRLDGTT KAEDRGMLLK  1140
```

```
1141 TFNEPGSEYF IFLLSTRAGG LGLNLQSADT VIIFDSDWNP HQDLQAQDRA HRIGQQNEVR  1200
1201 VLRLCTVNSV EEKILAAAKY KLNVDQKVIQ AGMFDQKSSS HERRAFLQAI LEHEEQDESR  1260
1261 HCSTGSGSAS FAHTAPPPAG VNPDLEEPPL KEEDEVPDDE TVNQMIARHE EEFDLFMRMD  1320
1321 LDRRREEARN PKRKPRLMEE DELPSWIIKD DAEVERLTCE EEEEKMFGRG SRHRKEVDYS  1380
1381 DSLTEKQWLK AIEEGTLEEI EEEVRQKKSS RKRKRDSDAG SSTPTTSTRS RDKDDESKKQ  1440
1441 KKRGRPPAEK LSPNPPNLTK KMKKIVDAVI KYKDSSSGRQ LSEVFIQLPS RKELPEYYEL  1500
1501 IRKPVDFKKI KERIRNHKYR SLNDLEKDVM LLCQNAQTFN LEGSLIYEDS IVLQSVFTSV  1560
1561 RQKIEKEDDS EGEESEEEEE GEEEGSESES RSVKVKIKLG RKEKAQDRLK GGRRRPSRGS  1620
1621 RAKPVVSDDD SEEEQEEDRS GSGSEED
```

Most frequently used BRG1 constructs composition

Construct name	**Amino acid (from-to) in sequence**
NVKE_BRG1	1223 to 1376
KR domain	1403 to 1566
AT hook domain	1452 to 1564
Bromodomain	1448 to 1564
Exbromodomain	1448 to 1575

9 Abbreviations

- 1D one-dimensional
- 2D two-dimensional
- Å Ångstrøm (10^{-10} m)
- aa amino acid
- APS ammonium peroxodisulfate
- bHLH basic helix-loop-helix
- bp base pair
- BRG1 brahma related gene 1
- BSA bovine serum albumin
- CBP CREB binding protein
- cDNA complimentary DNA
- CD circular dichroism
- COSY correlation spectroscopy
- Da Dalton (g mol-1)
- DMSO dimethylsulfoxide
- DNA deoxyribonucleic acid
- Dnase I deoxyribonuclease I
- EDTA ethylenediamine tetraacetic acid
- ELISA enzyme-linked immunosorbant assay
- g gravity (9.81 m s^{-2})
- GCN5 general control of amino acid synthesis 5
- GH growth hormone
- GSH reduced glutathione
- GSSG oxidized glutathione
- GST glutathione S-transferase
- HAT histone acetyltransferase
- HDAC histone deacetylase
- HIV human immunodeficiency virus
- HSQC heteronuclear single quantum coherence
- Hz Hertz
- IB inclusion bodies
- IMAC immobilized metal affinity chromatography

- IPTG isopropyl-β-thiogalactopyranoside
- ITC isothermal titration calorimetry
- K_D dissociation constant
- kDa kilo dalton
- LB Luria-Broth medium
- MAD multiwavelength anomalous diffraction
- MAP mitogen-activated protein kinase
- MIR multiple isomorphous replacement
- MM minimal medium
- MW molecular weight
- MWCO molecular weight cut off
- NA not applicable
- ND not done
- Ni-NTA nickel-nitrilotriacetic acid
- NLS nuclear localization signal
- NMR nuclear magnetic resonance
- NOE nuclear Overhauser effect
- NOESY nuclear Overhauser enhancement spectroscopy
- OD optical density
- PAGE polyacrylamide gel electrophoresis
- PAI plasminogen activator inhibitor
- P/CAF p300/CBP-associated factor
- PBS phosphate-buffered saline
- ppm parts per million
- pRb retinoblastoma protein
- RMSD root mean square deviation
- SDM site directed mutagenesis
- SAR structure-activity relationship
- SDS sodium dodecyl sulfate
- SWI switching
- SNF sucrose non fermenting
- TAFII250 transcription initiation factor II 250 kDa
- TEMED N,N,N',N'-tetramethylethylendiamine

Amino acids and nucleotides are abbreviated according to either one or three letter IUPAC code.

10 References

Abragam, A. (1961) *Principles of Nuclear Magnetism, International Series of Monographs on Physics* **No. 32** Oxford University Press, New York.

Ach, R.A., Durfee, T., Miller, A.B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P.C., and Gruissem, W. (1997) RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Mol. Cell Biol.* **17,** 5077-5086.

Agalioti, T., Chen, G., and Thanos, D. (2002) Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111,** 381-392.

Aravind, L., and Landsman, D. (1998) AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* **26,** 4413-4421.

Baldi, A., Boccia, V., Claudio, P.P., De Luca, A., and Giordano, A. (1996) Genomic structure of the human retinoblastoma-related Rb2/p130 gene. *Proc Natl Acad Sci U.S.A.* **93,** 4629-4632.

Bandara, L.R., Lam, E.W., Sorensen, T.S., Zamanian, M., Girling, R., and La Thangue, N.B. (1994) DP-1: a cell cycle-regulated and phosphorylated component of transcription factor DRTF1/E2F which is functionally important for recognition by pRb and the adenovirus E4 orf 6/7 protein. *EMBO J.* **13,** 3104-3114.

Bianchi, M.E. (1988) Interaction of a protein from rat liver nuclei with cruciform DNA. *EMBO J.* **7,** 843-849.

Bianchi, M.E., and Agresti, A. (2005) HMG proteins: dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev.* **15,** 496-506.

Bourachot, B., Yaniv, M., and Muchardt, C. (1999) The activity of mammalian brm/SNF2alpha is dependent on a high-mobility-group protein I/Y-like DNA binding domain. *Mol. Cell. Biol.* **19,** 3931-3939.

Brehm, A.., Miska, E.A., McCance, D.J., Reid, J.L., Bannister, A.J., and Kouzarides, T. (1998) Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* **391,** 597-601.

Brown, V.D., and Gallie, B.L. (2002) The B-domain lysine patch of pRB is required for binding to large T antigen and release of E2F by phosphorylation. *Mol. Cell. Biol.* **22,** 1390-1401.

Bruce, A., Alexander, J., Julian, L., Martin, R., Keith, R., and Peter, W. (2002) *Molecular Biology of the Cell.* **4th edition**. New York: Garland Publishing.

Buchkovich, K., Duffy, L.A., and Harlow, E. (1989) The retinoblastoma protein is phosphorylated during specific phases of the cell cycle. *Cell* **58,** 1097-1105.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G., and Magnuson, T. (2000) A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* **6,** 1287-1295.

CCP4, Collaborative computational project, Number 4. (1994) *Acta. Crystallogr. D.* **50,** 760-763.

Chan, H.M., Smith, L., and La Thangue, N.B. (2003) Role of LXCXE motif-dependent interactions in the activity of the retinoblastoma protein. *Oncogene* **20,** 6152-6163.

Chen, P.L., Scully, P., Shew, J.Y., Wang, J.Y., and Lee, W.H. (1989) Phosphorylation of the retinoblastoma gene product is modulated during the cell cycle and cellular differentiation. *Cell* **58,** 1193-1198.

Chen, T.T., and Wang, J.Y. (2000) Establishment of irreversible growth arrest in myogenic differentiation requires the RB LXCXE-binding function. *Mol. Cell. Biol.* **20,** 5571-5580.

Chow, K.N., and Dean, D.C. (1996) Domains A and B in the Rb pocket interact to form a transcriptional repressor motif. *Mol. Cell. Biol.* **16,** 4862-4868.

Classon, M., and Harlow, E. (2002) The retinoblastoma tumour suppressor in development and cancer. *Nat. Rev. Cancer* **2,** 910-917.

Classon, M., and Dyson, N. (2001) p107 and p130: versatile proteins with interesting pockets. *Exp. Cell. Res.* **264,** 135-147.

Claudio, P.P., Tonini, T., and Giordano, A. (2002) The retinoblastoma family: twins or distant cousins? *Genome Biol.* **3,** 3012.

Cobrinik, D. (1996) Regulatory interactions among E2Fs and cell cycle control proteins. *Curr. Top. Microbiol. Immunol.* **208,** 31-61.

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

Combet, C., Blanchet, C., Geourjon, C., and Deléage G. (2000) NPS@: Network Protein Sequence Analysis. *T.I.B.S.* **25,** 147-150.

Côté, J., Peterson, C.L., and Workman, J.L. (1998) Perturbation of nucleosome core structure by the SWI/SNF complex persists after its detachment, enhancing subsequent transcription factor binding. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 4947-4952.

Côté, J., Quinn, J., Workman, L., and Peterson, C.L. (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265,** 53-60.

Cress, W.D., Johnson, D.G., and Nevins, J.R. (1993) A genetic analysis of the E2F1 gene distinguishes regulation by Rb, p107, and adenovirus E4. *Mol. Cell. Biol.* **13,** 6314-6325.

Dahiya, A., Gavin, M.R., Luo, R.X., and Dean, D.C. (2000) Role of the LXCXE binding site in Rb function. *Mol. Cell. Biol.* **20,** 6799-6805.

Darnell, G.A., Antalis, T.M., Johnstone, R.W., Stringer, B.W., Ogbourne, S.M., Harrich, D., and Suhrbier, A. (2003) Inhibition of retinoblastoma protein degradation by interaction with the serpin plasminogen activator inhibitor 2 via a novel consensus motif. *Mol. Cell. Biol.* **23,** 6520-6532.

DeCaprio, J.A., Ludlow, J.W., Lynch, D., Furukawa, Y., Griffin, J., Piwnica-Worms, H., Huang, C.M., and Livingston, D.M. (1989) The product of the retinoblastoma susceptibility gene has properties of a cellcycle regulatory element. *Cell* **58,** 1085-1095.

De La Fortelle, E., and Bricogne, G. (1997). Maximum-Likelihood Heavy-Atom Parameter Refinement in the MIR and MAD Methods. *Methods Enzymol.* **276,** 472-494.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491-496.

Dick, F.A., and Dyson, N.J. (2002) Three regions of the pRB pocket domain affect its inactivation by human papillomavirus E7 proteins. *J. Virol.* **76,** 6224-6234.

Dick, F.A., and Dyson N. (2003) pRB contains an E2F1-specific binding domain that allows E2F1-induced apoptosis to be regulated separately from other E2F activities. *Mol. Cell.* **12,** 639-649.

Dick, F.A., Sailhamer, E., and Dyson, N.J. (2000) Mutagenesis of the pRB pocket reveals that cell cycle arrest functions are separable from binding to viral oncoproteins. *Mol. Cell. Biol.* **20,** 3715-3727.

Dimova, D.K., and Dyson, N.J. (2005) The E2F transcriptional network: old acquaintances with new faces. *Oncogene* **24,** 2810-2826.

DiRenzo, J., Shang, Y., Phelan, M., Sif, S., Myers, M., Kingston, R., and Brown, M. (2000) BRG-1 is recruited to estrogen-responsive promoters and cooperates with factors involved in histone acetylation. *Mol. Cell. Biol.* **20,** 7541-7549.

Dowdy, S.F., Hinds, P.W., Louie, K., Reed, S.I., Arnold, A.., and Weinberg, R.A. (1993) Physical interaction of the retinoblastoma protein with human D cyclins. *Cell* **73,** 499-511.

Dragan, A.I., Liggins, J.R., Crane-Robinson, C., and Privalov, P.L. (2003) The energetics of specific binding of AT-hooks from HMGA1 to target DNA. *J. Mol. Biol.* **327,** 393-411.

Dunaief, J.L., Strober, B.E., Guha, S., Khavari, P.A., Alin, K., Luban, J., Begemann, M., Crabtree, G.R., and Goff SP. (1994) The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell* **79,** 119-130.

Durr, H., Korner, C., Muller, M., Hickmann, V., and Hopfner, K.P. (2005) X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121,** 363-373.

Dyson, N. (1998) The regulation of E2F by pRB-family proteins. *Genes Dev.* **12,** 2245-2262.

Eberharter, A., and Becker, P.B. (2004) ATP-dependent nucleosome remodeling: factors and functions. *J. Cell Sci.* **117,** 3707-3711.

Edwards, G.M., Huber, H.E., DeFeo-Jones, D., Vuocolo, G., Goodhart, P.J., Maigetter, R.Z., Sanyal, G., Oliff, A.., Heimbrook, D.C. (1992) Purification and characterization of a functionally homogeneous 60-kDa species of the retinoblastoma gene product. *J. Biol. Chem.* **267,** 7971-7974. Erratum in: *J. Biol. Chem.* **267,** 13780.

Ewen, M.E., Sluss, H.K., Sherr, C.J., Matsushime, H., Kato, J., and Livingston, D.M. (1993) Functional interactions of the retinoblastoma protein with mammalian D-type Cyclins. *Cell* **73,** 487-497.

Ewen, M.E., Xing, Y.G., Lawrence, J.B., and Livingston, D.M. (1991) Molecular cloning, chromosomal mapping, and expression of the cDNA for p107, a retinoblastoma gene product-related protein. *Cell* **66,** 1155-1164.

Falvo, J.V., Thanos, D.T., and Maniatis, T. (1995) Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). *Cell* **83,** 1101-1111.

Fasman, G.D. (1996) *Circular Dichroism and the Conformational Analysis of Biomolecules.* Plenum Press: New York.

Fan, H.Y., He, X., Kingston, R.E. and Narlikar, G.J. (2003) Distinct strategies to make nucleosomal DNA accessible. *Mol. Cell* **11,** 1311-1322.

Fan, S., Yuan, R., Ma, Y.X., Xiong, J., Meng, Q., Erdos, M., Zhao, J.N., Goldberg, I.D., Pestell, R.G., and Rosen, E.M. (2001) Disruption of BRCA1 LXCXE motif alters BRCA1 functional activity and regulation of RB family but not RB protein binding. *Oncogene* **20,** 4827-4841.

Ferreira, R., Magnaghi-Jaulin, L., Robin, P., Harel-Bellan, A., and Trouche, D. (1998) The three members of the pocket proteins family share the ability to repress E2F activity through recruitment of a histone deacetylase. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 10493-10498.

Figge, J., Webster, T., Smith, T.F., and Paucha, E. (1988) Prediction of similar transforming regions in simian virus 40 large T, adenovirus E1A, and myc oncoproteins. *J. Virol.* **62,** 1814-1818.

Flaus, A., and Owen-Hughes, T. (2004) Mechanisms for ATP-dependent chromatin remodeling: farewell to the tuna-cancoctamer? *Curr. Opin. Genet. Dev.* **2,** 165-173.

Flemington, E.K., Speck, S.H., and Kaelin, W.G. Jr. (1993) E2F-1-mediated transactivation is inhibited by complex formation with the retinoblastoma susceptibility gene product. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 6914-6918.

Forman-Kay, J.D., and Pawson, T. (1999) Diversity in protein recognition by PTB domains. *Curr. Opin. Struct. Biol.* **9,** 690–695.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., Dryja. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323,** 643-646.

Gill, S.C., and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182,** 319-326.

Girling, R., Partridge, J.F, Bandara, L.R., Burden, N., Totty, N.F., Hsuan, J.J., and La Thangue, N.B. (1993) A new component of the transcription factor DRTF1/E2F. *Nature* **365,** 468.

Goodrich DW, Lee WH. (1993)Molecular characterization of the retinoblastoma susceptibility gene. *Biochim. Biophys. Acta.* **1155,** 43-61.

Gu, W., Schneider, J.W., Condorelli, G., Kaushal, S., Mahdavi, V., and Nadal-Ginard, B. (1993) Interaction of myogenic factors and the retinoblastoma protein mediates muscle cell commitment and differentiation. *Cell* **72,** 309-324.

Guo, C.S., Degnin, C., Fiddler, T.A., Stauffer, D., and Thayer, M.J. Regulation of MyoD activity and muscle cell differentiation by MDM2, pRb, and Sp1. *J. Biol. Chem.* 278, 22615–22622.

Hanahan, D., and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell* **100,** 57-70.

Harbour, J.W., and Dean, D.C. (2000) Rb function in cell-cycle regulation and apoptosis. *Nat. Cell. Biol.* **2,** E65-67.

Harbour, J.W., Luo R.X., Dei Santi, A., Postigo, A.A., and Dean, D.C. (1999) Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell* **98,** 859-69.

Hassan, A.H., Prochasson, P., Neely, K.E., Galasinski, S.C., Chandy, M., Carrozza, M.J., Workman, J.L. (2002) Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell* **111,** 369-379.

Hatakeyama, M., Brill, J.A., Fink, G.R., and Weinberg, R.A. (1994) Collaboration of G1 cyclins in the functional inactivation of the retinoblastoma protein. *Genes Dev.* **8,** 1759-1771.

Haynes, S.R., Dollard, C., Winston, F., Beck, S., Trowsdale, J., and Dawid, I.B. (1992) The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res.* **20,** 2603.

He, Y., Armanious, M.K., Thomas, M.J., and Cress, W.D. (2000) Identification of E2F-3B, an alternative form of E2F-3 lacking a conserved N-terminal region. *Oncogene* **19,** 3422-3433.

Helin, K., Harlow, E., and Fattaey, A. (1993) Inhibition of E2F-1 transactivation by direct binding of the retinoblastoma protein. *Mol. Cell. Biol.* **13,** 6501-6508.

Helin, K., Lees, J.A., Vidal, M., Dyson, N., Harlow, E., and Fattaey, A. (1992) A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell* **70,** 337-350.

Helt, A.M., and Galloway, D.A. (2003) Mechanisms by which DNA tumor virus oncoproteins target the Rb family of pocket proteins. *Carcinogenesis* **24,**159-169.

Henderson, A.., Holloway, A.., Reeves, R., and Tremethick, D.J. (2004) Recruitment of SWI/SNF to the human immunodeficiency virus type 1 promoter. *Mol. Cell. Biol.* **24,** 389-397

Herwig, S., and Strauss, M. (1997) The retinoblastoma protein: a master regulator of cell cycle, differentiation and apoptosis. *Eur. J. Biochem.* **246,** 581-601.

Hiebert, S.W., Chellappan, S.P., Horowitz, J.M., and Nevins, J.R. (1992) The interaction of RB with E2F coincides with an inhibition of the transcriptional activity of E2F. *Genes Dev.* **6,** 177-185.

Higashitsuji, H., Itoh, K., Nagao, T., Dawson, S., Nonoguchi, K., Kido, T., Mayer, R.J., Arii, S., and Fujita, J. (2000) Reduced stability of retinoblastoma protein by gankyrin, an oncogenic ankyrin-repeat protein overexpressed in hepatomas. *Nat. Med.* **6,** 96-99.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* **6,** 2288-2298.

Hong, F.D., Huang, H.J., To, H., Young, L.J., Oro, A., Bookstein, R., Lee, E.Y., and Lee, W.H. (1989) Structure of the human retinoblastoma gene. *Proc. Natl. Acad. Sci. U.S.A.* **86,** 5502-5506.

Hori, T., Kato, S., Saeki, M., DeMartino, G.N., Slaughter, C.A., Takeuchi, Toh-e, A.., and Tanaka, K. (1998) cDNA cloning and functional analysis of p28 (Nas6p) and p40.5 (Nas7p), two novel regulatory subunits of the 26S proteasome. *Gene* **216,** 113–122.

Horowitz, J.M., Yandell, D.W., Park, S.H., Canning, S., Whyte, P., Buchkovich, K., Harlow, E., Weinberg, R.A., and Dryja, T.P. (1989) Point mutational inactivation of the retinoblastoma antioncogene. *Science* **243,** 937-940.

Hu, Q.J., Dyson, N., and Harlow, E. (1990) The regions of the retinoblastoma protein needed for binding to adenovirus E1A or SV40 large T antigen are common sites for mutations. *EMBO J.* **9,** 1147-1155.

Huang, S., Wang, N.P., Tseng, B.Y., Lee, W.H., and Le, E.H. (1990) Two distinct and frequently mutated regions of retinoblastoma protein are required for binding to SV40 T antigen. *EMBO J.* **9,** 1815-1822.

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J., and Wright, P.E. (2000) Solution structure and acetyl-lysine binding activity of the GCN5 bromodomain. *J. Mol. Biol.* **304,** 355-370.

Huth, J.R., Bewley, C.A., Nissen, M.S., Evans, J.N., Reeves, R., Gronenborn, A.M., and Clore, G.M. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* **4,** 657-665.

Iavarone, A., Garg, P., Lasorella, A., Hsu, J., and Israel, M.A. (1994) The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. *Genes Dev.* **8,** 1270-1284.

Ivey-Hoyle, M., Conroy, R., Huber, H.E., Goodhart, P.J., Oliff, A., and Heimbrook, D.C. (1993) Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F. *Mol. Cell. Biol.* **13,** 7802-7812.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000) Structure and function of a human TAFII250 double bromodomain module. *Science* **288,** 1422-1425.

Jeanmougin, F., Wurtz, J.M., Douarin, B.L., Chambon, P., and Losson., R. (1997) The bromodomain revisited . *Trends Biochem. Sci.* **22,** 151–153.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26,** 795-800.

Kadam, S., McAlpine, G.S., Phelan, M.L., Kingston, R.E., Jones, K.A., and Emerson, B.M. (2000) Functional selectivity of recombinant mammalian SWI/SNF subunits. *Genes Dev.* **14,** 2441-2451.

Kaelin, W.G. Jr., Ewen, M.E., and Livingston, D.M. (1990) Definition of the minimal simian virus 40 large T antigen- and adenovirusE1A-binding domain in the retinoblastoma gene product. *Mol. Cell. Biol.* **10,** 3761-37619.

Kaelin, W.G. Jr., Krek, W., Sellers, W.R., DeCaprio, J.A., Ajchenbaum, F., Fuchs, C.S., Chittenden, T., Li, Y., Farnham, P.J., and Blanar MA, et al. (1992) Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. *Cell* **70,** 351-364.

Kang, H., Cui, K., and Zhao, K. (2004) BRG1 controls the activity of the retinoblastoma protein via regulation of p21CIP1/WAF1/SDI. *Mol. Cell. Biol.* **24,** 1188-1199.

Kennedy, B.K., Liu, O.W., Dick, F.A., Dyson, N., Harlow, E., and Vidal, M. (2001) Histone deacetylase-dependent transcriptional repression by pRB in yeast occurs independently of interaction through the LXCXE binding cleft. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 8720-8725.

Khavari, P.A., Peterson, C.L., Tamkun, J.W., Mendel, D.B., and Crabtree, G.R. (1993) BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366,** 170-174.

Kim, H.Y., Ahn, B.Y., and Cho, Y. (2001) Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *EMBO J.* **20,** 295-304.

Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., and Yaniv, M. (2000) The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression, *EMBO Rep.* **1,** 500-506.

Klochendler-Yeivin, A., Muchardt, C., and Yaniv, M. (2002) SWI/SNF chromatin remodeling and cancer. *Curr. Opin. Genet. Dev.* **12,** 73-79.

Klochendler-Yeivin, A., and Yaniv, M. (2001) Chromatin modifiers and tumor suppression. *Biochim Biophys Acta.* **1551,** M1-10.

Knudsen, E.S., Buckmaster, C., Chen, T.T., Feramisco, J.R., and Wang, J.Y. (1998) Inhibition of DNA synthesis by RB: effects on G1/S transition and S-phase Progression. *Genes Dev.* **12,** 2278-2292.

Kouzarides, T. (2002) Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **12,** 198-209.

Kratzke, R.A., Otterson, G.A., Lin, A.Y., Shimizu, E., Alexandrova, N., Zajac-Kaye, M., Horowitz, J.M., and Kaye, F.J. (1992) Functional analysis at the Cys706 residue of the retinoblastoma protein. *J. Biol. Chem.* **267,** 25998-26003.

Krek, W., Livingston, D.M., and Shirodkar, S. (1993) Binding to DNA and the retinoblastoma gene product promoted by complex formation of different E2F family members. *Science* **262,** 1557-1560.

Krzywda, S., Brzozowski, A.M., Higashitsuji, H., Fujita, J., Welchman, R., Dawson, S., Mayer, R.J., and Wilkinson, A.J. The crystal structure of gankyrin, an oncoprotein found in complexes with cyclin-dependent kinase 4, a 19 S proteasomal ATPase regulator, and the tumor suppressors Rb and p53. *J. Biol. Chem.* **279,** 1541-1545.

Kuriyan, J., and Cowburn, D. (1993) Structures of SH2 and SH3 domains. *Curr. Opin. Str. Biol.* **3,** 828–837.

Lai, A., Lee, J.M., Yang, W.M., DeCaprio, J.A., Kaelin, W.G. Jr., Seto, E., and Branton, P.E. (1999) RBP1 recruits both histone deacetylase-dependent and -independent repression activities to retinoblastoma family proteins. *Mol. Cell. Biol.* **19,** 6632-6641.

Lamzin, V. S., and Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallogr. sect. D* **49,** 129-149.

Lasorella, A., Iavarone, A., and M.A. Israel. (1996) Id2 specifically alters regulation of the cell cycle by tumor suppressor proteins. *Mol. Cell. Biol.* **16,** 2570–2578.

Lasorella, A., Noseda, M., Beyna, M., Yokota, Y., and Iavarone, A. (2000) Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins.*Nature* **407,** 592-8. Erratum in: (*2000*) *Nature* **408,** 498.

Ledl, A.., Schmidt, D., and Muller, S. (2005) Viral oncoproteins E1A and E7 and cellular LxCxE proteins repress SUMO modification of the retinoblastoma tumor suppressor. *Oncogene* **24,** 3810-3188.

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

Lee, C., Chang, J.H., Lee, H.S., and Cho, Y. (2002) Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor. *Genes Dev.* **16,** 3199-3212.

Lee, J.O., Russo, A.A., and Pavletich, N.P. (1998) Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature* **391,** 859-865.

Leone, G., Nuckolls, F., Ishida, S., Adams, M., Sears, R., Jakoi, L., Miron, A., and Nevins, J.R. (2000) Identification of a novel E2F3 product suggests a mechanism for determining specificity of repression by Rb proteins. *Mol. Cell. Biol.* **20,** 3626-3632.

Li, J., and Tsai, M.D. (2002) Novel insights into the INK4-CDK4/6-Rb pathway: counter action of gankyrin against INK4 proteins regulates the CDK4-mediated phosphorylation of Rb. *Biochemistry* **41,** 3977–3983.

Lorch, Y., Cairns, B.R., Zhang, M., and Kornberg, R.D. (1998) Activated RSC-nucleosome complex and persistently altered form of the nucleosome. *Cell* **94,** 29-34.

Lu, X., and Horvitz, H.R. (1998) lin-35 and lin-53, two genes that antagonize a C. elegans Ras pathway, encode proteins similar to Rb and its binding protein RbAp48. *Cell* **95,** 981-991.

Lukas, J., Pagano, M., Staskova, Z., Draetta, G., and Bartek, J. (1994) Cyclin D1 protein oscillates and is essential for cell cycle progression in human tumour cell lines. *Oncogene* **9,** 707-718.

Lees, J.A., Saito, M., Vidal, M., Valentine, M., Look, T., Harlow, E., Dyson, N., and Helin, K. (1993) The retinoblastoma protein binds to a family of E2F transcription factors. *Mol. Cell. Biol.* **12,** 7813-7825.

Lozano, G., and Zambetti, G.P. (2005) Gankyrin: an intriguing name for a novel regulator of p53 and RB. *Cancer Cell* **8,** 3-4.

Ludlow, J.W., Shon, J., Pipas, J.M., Livingston, D.M., and DeCaprio, J.A. (1990) The retinoblastoma susceptibility gene product undergoes cell cycle-dependent dephosphorylation and binding to and release from SV40 large T. *Cell* **60,** 387-396.

Luo, R.X., Postigo, A.A., and Dean, D.C. (1998) Rb interacts with histone deacetylase to repress transcription. *Cell* **92,** 463-473.

Magnaghi-Jaulin, L., Groisman, R., Naguibneva, I., Robin, P., Lorain, S., Le Villain, J.P., Troalen, F., Trouche, D., and Harel-Bellan, A. (1998) Retinoblastoma protein represses transcription by recruiting a histone deacetylase. *Nature* **391,** 601-605.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol. Rev.* **60,** 512-538.

Malumbres, M., and Barbacid, M. (2001) To cycle or not to cycle: a critical decision in cancer. *Nat. Rev. Cancer* **1,** 222-231.

Martelli, F., Cenciarelli, C., Santarelli, G., Polikar, B., Felsani, A., and Caruso, M. (1994) MyoD induces retinoblastoma gene expression during myogenic differentiation. *Oncogene* **9,** 3579-3590.

Marmorstein, R., and Berger, S.L. (2001) Structure and function of bromodomains in chromatin-regulating complexes. *Gene* **272,** 1-9.

Matsuura, I., and Denissova, N.G., Wang, G., He, D., Long, J., and Liu, F. (2004) Cyclin-dependent kinases regulate the antiproliferative function of Smads. *Nature* **430,** 226-231.

McRee, D. E. (1999). XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. *J. Struc. Biol.* **125,** 156-165.

Meyer, B., and Peters, T. (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int. Ed. Engl.* **42,** 864-890.

Mihara, K., Cao, X.R., Yen, A., Chandler, S., Driscoll, B., Murphree, A.L., T'Ang, A., Fung, and Y.K. (1989) Cell cycle-dependent regulation of phosphorylation of the human retinoblastoma gene product. *Science* **246,** 1300-1303.

Mittnacht S, Lees JA, Desai D, Harlow E, Morgan DO, Weinberg RA. (1994) Distinct sub-populations of the retinoblastoma protein show a distinct pattern of phosphorylation. *EMBO J.* **13,** 118-127.

Moberg, K., Starz, M.A., and Lees, J.A. (1996) E2F-4 switches from p130 to p107 and pRB in response to cell cycle reentry. *Mol. Cell. Biol.* **16,** 1436-1449.

Mohrmann, L., and Verrijzer, C.P. (2005) Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochim. Biophys. Acta.* **1681,** 59-73.

Morris, E.J., and Dyson, N.J. (2001) Retinoblastoma protein partners. *Adv. Cancer Res.* **82,** 1-54.

Morozov, A., Yung, E., and Kalpana, G.V. (1998) Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 1120-1125.

Moss, T. (2001) DNA-Protein Interactions, Principles and protocols. *Methods in Molecular biology*, Volume **148,** Human press.

Muchardt, C., and Yaniv, M. (1999) The mammalian SWI/SNF complex and the control of cell growth. *Semin. Cell Dev. Biol.* **10,** 189-195.

Muchardt, C., and Yaniv, M. (2001) When the SWI/SNF complex remodels...the cell cycle. *Oncogene* **20,** 3067-3075.

Mujtaba, S., He, Y., Zeng, L., Farooq, A; Carlson, J.E., Ott, M., Verdin, E., Zhou M M. (2002) Structural basis of lysine-acetylated HIV-1 tat recognition by PCAF bromodomain. *Mol. Cell* **9,** 575-586.

Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, Sanchez, R., Zeleznik-Le, N.J., Ronai, Z., and Zhou M M. (2004) Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol. Cell* **13,** 251-263.

Mundle, S.D., and Saberwal, G. (2003) Evolving intricacies and implications of E2F1 regulation. *FASEB J.* **17,** 569-574.

Munshi, N., Merika, M., Yie, J., Senger, K., Chen, G., and Thanos, D. (1998) Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome. *Mol. Cell* **2,** 457-667.

Murphy, F.V 4th., Sweet, R.M., and Churchill, M.E. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J.* **18,** 6610-6618.

Neel, J.V., and Falls, H.F. (1951) The rate of mutation of the gene responsible for retinoblastoma in man. *Science* **114,** 419-422.

Neigeborn, L., and Carlson, M. (1984) Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. *Genetics* **108,** 845-858.

Nicolas, E., Morales, V., Magnaghi-Jaulin, L., Harel-Bellan, A., Richard-Foy, H., and Trouche, D. (2000) RbAp48 belongs to the histone deacetylase complex that associates with the retinoblastoma protein. *J. Biol. Chem.* **275,** 9797-9804.

Norton, J.D., Deed, R.W., Craggs, G., and Sablitzky, F. (1998) Id helix–loop–helix proteins in cell growth and differentiation. *Trends Cell Biol.* **8,** 58–65.

Novitch, B.G., Mulligan, G.J., Jacks, T., and Lassar, A.B. (1996) Skeletal muscle cells lacking the retinoblastoma protein display defects in muscle gene expression and accumulate in S and G2 phases of the cell cycle. *J. Cell Biol.* **135,** 441-456.

Onadim, Z., Hogg, A., Baird, P.N., and Cowell, J.K. (1992) Oncogenic point mutations in exon 20 of the RB1 gene in families showing incomplete penetrance and mild expression of the retinoblastoma phenotype. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 6177-6181.

Oruetxebarria, I., Venturini, F., Kekarainen, T., Houweling, A., Zuijderduijn, L.M., Mohd-Sarip, A., Vries, R.G., Hoeben, R.C., and Verrijzer, C.P. (2004) P16INK4a is required for hSNF5 chromatin remodeler-induced cellular senescence in malignant rhabdoid tumor cells. *J. Biol. Chem.* **279,** 3807–3816.

Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000) The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J.* **19,** 6141-6149.

Pagliuca, A., Bartoli, P.C., Saccone, S., Della Valle, G., and Lania, N. (1995) Molecular cloning of ID4, a novel dominant negative helix–loop–helix human gene on chromosome 6p21.3–p22. *Genomics* **27,** 200–203.

Pardee, A.B. (1974) A restriction point for control of normal animal cell proliferation. *Proc. Natl. Acad. Sci. U.S.A.* **71.** 1286-1290.

Pardee, A.B. (1989) G1 events and regulation of cell proliferation. *Science* **246,** 603-608.

Pellecchia, M., Sem, D.S., and Wuthrich K. (2002) NMR in drug discovery. *Nat. Rev. Drug. Discov.* **1,** 211-219.

Perrakis, A., Morris, R. and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6,** 458-463.

pET system: manual. **10** th edition. (2003) Novagen.

Peterson, C.L., and Herskowitz, I. (1992) Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell* **68,** 573-583.

Phelan, M.L., Sif, S., Narlikar, G.J., and Kingston, R.E. (1999) Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Mol. Cell* **3,** 247-253.

Phelps, W.C., Yee, C.L., Munger, K., and Howley, P.M. (1988) The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. *Cell* **53,** 539-47.

Protein purification: handbook collection. (2003) Amersham-pharmacia.

Puri, P.L., Iezzi, S., Stiegler, P., Chen, T.T., Schiltz, R.L., Muscat, R.L., Giordano, R., Kedes, R., Wang, J.Y., and Sartorelli, V. (2001) Class I histone deacetylases sequentially interact with MyoD and pRb during skeletal myogenesis. *Mol. Cell* **8,** 885–897.

Quick change site directed mutagenesis kit: instruction manual. (2003) Stratagene.

Quinn, J., Fyrberg, A.M., Ganster, R.W., Schmidt, M.C., and Peterson, C.L. (1996) DNA-binding properties of the yeast SWI/SNF complex. *Nature* **379,** 844-847.

Radulescu, R.T., and Wendtner, C.M. (1992) Proposed interaction between insulin and retinoblastoma protein. *J. Mol. Recognit.* **5,** 132-137.

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

Reeves, R. (2001) Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* **277,** 63-81.

Reeves, R., and Beckerbauer, L. (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim. Biophys. Acta.* **1519,** 13-29.

Reeves R & Nissen MS (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *J. Biol. Chem.* **265,** 8573-8582.

Reyes, J.C., Barra, J., Muchardt, C., Camus, A., Babinet, C., and Yaniv M. (1998) Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha). *EMBO J.* **17,** 6979-6991.

Richmond, E., and Peterson, C.L. (1996) Functional analysis of the DNA-stimulated ATPase domain of yeast SWI2/SNF2. *Nucleic Acids Res.* **24,** 3685-92.

Roberts, C.W., and Orkin, S.H. (2004) The SWI/SNF complex--chromatin and cancer. *Nat. Rev. Cancer.* **4,** 133-142.

Roberts, C.W., Galusha, S.A., McMenamin, M.E., Fletcher, C.D., and Orkin, S.H. (2000) Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 13796-13800.

Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sandal, T. (2002) Molecular aspects of the mammalian cell cycle and cancer. *Oncologist* **7,** 73-81.

Schagger, H., von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166,** 368-379.

Schneider, J.W., Gu, W., Zhu, L., Mahdavi, V., and Nadal-Ginard, B. (1994) Reversal of terminal differentiation mediated by p107 in Rb-/- muscle cells. *Science* **264,** 1467-1471.

Schneider, T. R., and Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. sect. D.* **58,** 1772-1779.

Schnuchel, A., Wiltscheck, R., Czisch, M., Herrler, M., Willimsky, G., Graumann, P., Marahiel, M.A., and Holak TA. (1993) Structure in solution of the major cold-shock protein from Bacillus subtilis. *Nature* **364,** 169-171.

Sellers, W.R., and Kaelin, W.G Jr. (1997) Role of the retinoblastoma protein in the pathogenesis of human cancer. *J. Clin. Oncol.* **15,** 3301-3312.

Shoelson, S.E. (1997) SH2 and PTB domain interactions in tyrosine kinase signal transduction. *Curr. Opin. Chem. Biol.* **2,** 227-234.

Shan, B., Zhu, X., Chen, P.L., Durfee, T., Yang, Y., Sharp, D., and Lee W.H. (1992) Molecular cloning of cellular genes encoding retinoblastoma-associated proteins: identification of a gene with properties of the transcription factorE2F. *Mol. Cell. Biol.* **12,** 5620-5631.

Sherr, C.J., and McCormick, F. (2002) The RB and p53 pathways in cancer. *Cancer Cell* **2,** 103-112.

Sherr, C.J., and Roberts, J.M. (2004) Living with or without cyclins and cyclin-dependent kinases. *Genes Dev.* **18,** 2699-2711.

Sherr, C.J. (1996) Cancer cell cycles. *Science* **274,** 1672-1677.

Sherr, C.J. (2000- 2001) Cell cycle control and cancer. *Harvey Lect.* **96,** 73-92.

Sherr, C.J. (2004) Principles of tumor suppression. *Cell* **116,** 235-246.

Simone C. (2005) SWI/SNF: The crossroads where extracellular signaling pathways meet chromatin. *J Cell Physiol.* 2005 Sep **9,** (published online).

Stabel, S., Argos, P., and Philipson, L. (1985) The release of growth arrest by microinjection of adenovirus E1A DNA. *EMBO J.* **4,** 2329-2336.

Strahl, B.D., and Allis, C.D. (2000) The language of covalent histone modifications. *Nature* **403,** 41-45.

Stern, M., and Jensen, R., and Herskowitz, I. (1984) Five SWI genes are required for expression of the HO gene in yeast. *J. Mol. Biol.* **178,** 853-868.

Strober, B.E., Dunaief, J.L., Guha, and Goff, S.P. (1996) Functional interactions between the hBRM/hBRG1 transcriptional activators and the pRB family of proteins. *Mol. Cell. Biol.* **16,** 1576-1583.

Thanos, D., and Maniatis, T. (1992) The high mobility group protein HMGI(Y) is required for NF-κβ-dependent virus induction of the human IFN-β gene. *Cell* **71,** 777–789.

The QIAexpressionist™: *A handbook for high-level expression and purification of 6xHis-tagged proteins*. (2003) Qiagen.

Trouche, D., Le Chalony, C., Muchardt, C., Yaniv, M., and Kouzarides, T. (1997) RB and hbrm cooperate to repress the activation functions of E2F1. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 11268-11273.

Tsukiyama, T. (2002) The in vivo functions of ATP-dependent chromatin-remodeling factors. *Nat. Rev. Mol. Cell Biol.* **3,** 422-429.

Vignali, M., Hassan, A.H., Neely, K.E., and Workman, J.L. (2000) ATP-dependent chromatin-remodeling complexes. *Mol. Cell. Biol.* **20,** 1899-1910.

Voet, D., and Voet, J.G. (1995) *Biochemistry. John Wiley & Sons,* New York.

Vogelstein, B., and Kinzler, K.W. (2004) Cancer genes and the pathways they control. *Nat. Med.* **10,** 789-799.

Wang, J.Y., Knudsen, E.S., and Welch, P.J. (1994) The retinoblastoma tumor suppressor protein. *Adv. Cancer Res.* **64,** 25-85.

Wang, W., Chi, T., Xue, Y., Zhou, S., Kuo, A., and Crabtree, G.R. (1998) Architectural DNA binding by a high-mobility-group/kinesin-like subunit in mammalian SWI/SNF-related complexes. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 492-498.

Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R., Muchardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L., and Crabtree, G.R. (1996) Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J.* **15,** 5370-5382.

Weinberg, R.A. (1995) The retinoblastoma protein and cell cycle control. *Cell* **81,** 323-330.

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

Wilson, C.J., Chao, D.M., Imbalzano, A.N., Schnitzler, G.R., Kingston, R.E., Young, R.A. (1996) RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84,** 235-244.

Wong, A.K., Shanahan, F., Chen, Y., Lian, L., Ha, P., Hendricks, K., Ghaffari, S., Iliev, D., Penn, B., Woodland, A.M., Smith, R., Salada, G., Carillo, A., Laity, K., Gupte, J., Swedlund, B., Tavtigian, S.V., Teng, D.H., and Lees E. (2000) BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines. *Cancer Res.* **60,** 6171-6177.

Wu, C.L., Zukerberg, L.R., Ngwu, C., Harlow, E., and Lees, J.A. (1995) In vivo association of E2F and DP family proteins. *Mol. Cell Biol.* **15,** 2536-2546.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.

Xiao, B., Spencer, J., Clements, A., Ali-Khan, N., Mittnacht, S., Broceno, C., Burghammer, M., Perrakis, A., Marmorstein, R., and Gamblin, S.J. (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.

Xiao, Z.X., Chen, J., Levine, A.J., Modjtahedi, N., Xing, J., Sellers, W.R., and Livingston, D.M. (1995) Interaction between the retinoblastoma protein and the oncoprotein MDM2. *Nature* **375,** 694-698.

Yandell, D.W., Campbell, T.A., Dayton, S.H., Petersen, R., Walton, D., Little, J.B., McConkie-Rosell, A., Buckley, E.G., and Dryja, T.P. (1989) Oncogenic point mutations in the human retinoblastoma gene: their application to genetic counseling. *N. Engl. J. Med.* **321,** 1689-1695.

Yang, X.J. (2004) Lysine acetylation and the bromodomain: a new partnership for signaling. *Bioessays* **26,** 1076-1087.

Yoshinaga, S.K., Peterson, C.L., Herskowitz, I., Yamamoto, K.R. (1992) Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors. *Science* **258,** 1598-1604.

Zalvide, J., Stubdal, H., and DeCaprio J.A. (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.

Zeng, L., and Zhou, M.M. (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS Lett.* **513,** 124-128.

Zhang, H.S., Gavin, M., Dahiya, A., Postigo, A.A., Ma, D., Luo, R.X., Harbour, J.W., and Dean, D.C. (2000) Exit from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-Rb-hSWI/SNF and Rb-hSWI/SNF. *Cell* **101,** 79-89.

Zhang, J.M., Wei, Q., Zhao, W., and Paterson, B.M. (1999) Coupling of the cell cycle and myogenesis through the cyclin D1-dependent interaction of MyoD with cdk4. *EMBO J.* **18,** 926–933.

Zhang, J.M., Zhao, X., Wei, Q., and Paterson, B.M. (1999) Direct inhibition of G(1) cdk kinase activity by MyoD promotes myoblast cell cycle withdrawal and terminal differentiation. *EMBO J.* **18,** 6983–6993.

Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (1997) Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89,** 357-364.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* **13,** 1924-1935.

Zhang, W., Wu, Q., Pwee, K.H., Jois, S.D., and Kini, R.M. (2003) Characterization of the interaction of wheat HMGa with linear and four-way junction DNAs. *Biochemistry* **42,** 6596-6607.

Zhou, M.M., and Fesik, S.W. (1995) Structure and function of the phosphotyrosine binding (PTB) domain. Prog. Biophys. *Mol. Biol*. **64,** 221–235.