UNIVERSITY OF CALIFORNIA
SANTA CRUZ

STRUCTURAL CONSERVATION AND E2F BINDING SPECIFICITY WITHIN THE RETINOBLASTOMA POCKET PROTEIN FAMILY

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Tyler James Liban

December 2016

The Dissertation of Tyler J. Liban is approved:

_____
Professor Seth Rubin, Chair

_____
Professor Michael Stone

_____
Professor Doug Kellogg

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

Page

List of Figures

List of Tables

Abstract

Tyler Liban

Structural Conservation and E2F Binding Specificity Within the Retinoblastoma Pocket Protein Family

The regulation of cell division is critical for ensuring the outcome of two viable daughter cells. Improper regulation of the cell cycle can lead to aberrant proliferation and tumor formation. The human pocket proteins retinoblastoma (Rb), p107, and p130 are critical regulators of the cell cycle and contribute to tumor suppression. Their growth suppressive function is achieved through the regulation of E2F transcription factors, but within the pocket protein family, there is specificity for the particular E2F transcription factors each negatively regulates. The molecular mechanisms governing this specificity remain unknown, as does how E2F specificity contributes to the different functions observed among pocket proteins. In this study, I explore the molecular determinants governing the interactions between pocket proteins and E2Fs, how this E2F specificity may influence the function of pocket proteins, as well as the evolution of tumor suppression characteristics in the protein Rb. Using structural biology, biochemistry and biophysical techniques, we identified key regions of Rb that evolved to regulate a specific subset of E2Fs, giving rise to the unique functional activities associated with Rb.

Acknowledgements

I would like to gratefully acknowledge the support my mentor, Dr. Seth Rubin, has provided me through the course of my thesis research. I acknowledge the scientists at the Advanced Photon Source Beamline 23-IDB (Argonne National Laboratory) and the Advanced Light Source Beamlines 5.0.1 and 8.3.1 (Lawrence Berkeley National Laboratories). I acknowledge Dr. Sarvind Tripathi for his expertise in structural biology, without which this research would not have been possible. This work was supported by a QB3 fellowship, the US National Institute of Health National Cancer Institute (CA132685), and the Department of Defense Breast Cancer Research Program Breakthrough Award (W81XWH-14-1-0329). Finally, I would like to acknowledge my friends and family for their unending encouragement and support.

Introduction

1.1 Rb and cancer

Regulation of cell proliferation is essential to ensuring that in response to growth factors and environment cues, the cell proceeds through a single round of division and produces two healthy daughter cells. Proper development and overall health is linked to the ability to divide in response to growth factors, as well as inhibiting division until proliferative signals once again accumulate. The Retinoblastoma tumor suppressor protein (Rb) is one of the most intensively studied proteins that functions as a tumor suppressor and negative regulator of the cell cycle. Rb represents the first tumor suppressor gene cloned, and has changed the discussion on how tumorigenesis arises. The original hypothesis relating to cancer progression and aberrant cellular proliferation was that overexpression of mutant growth factors or constitutively active growth factors, termed oncogenes, were required to drive cell proliferation. In 1971, Alfred Knudson proposed a new “two-hit hypothesis”, where individuals displaying the retinoblastoma disease inherited a mutant copy of a tumor suppressor allele, the first “hit”. The second “hit” comes from a sporadic mutation in the remaining copy of this allele, ultimately causing retinoblastoma, and further predisposes the individual to additional cancers in a host of tissues. The search for these tumor suppressor genes ultimately led to the discovery of Rb, a protein that plays a central role in regulation of the cell cycle, and is thought to be deregulated or absent in over 90% of human cancers (Hanahan and Weinberg, 2011).

Following the discovery of Rb, attempts to characterize how a tumor suppressor might function allowed for the discovery of several pieces of information key to our understanding of Rb function and its role in regulation of the cell cycle. First, the initial discovery of Rb found it to be a phosphoprotein located in the nucleus of dividing cells (Lee et al. 1987). Second, the timing of Rb phosphorylation occurred near the transition of the cell cycle from the G1 growth phase to S phase, a critical point in regulation termed the

restriction point (Buchkovich et al. 1989; Chen et al. 1989; DeCaprio et al. 1989; Mihara et al. 1989). Third, Rb was discovered to associate with a family of transcription factors named E2F in a phosphorylation dependent manner, and that E2F controlled the expression of many genes required for S-phase (Bagchi et al. 1991; Bandara et al. 1991; Chellappan et al. 1991; Nevins 1992). Lastly, Rb-E2F complexes could be disrupted by the addition of the adenovirus E1A protein, releasing free E2F and promoting the transition into S phase (Bandara et al. 1991; Chittenden et al. 1991). The structure of Rb bound to E1A demonstrates the molecular mechanism by which E1A can inhibit the tumor suppressor properties of Rb and cause release of E2F (Liu and Marmorstein, 2007). These findings highlight the role of the Rb-E2F interaction towards achieving negative regulation of the cell cycle, and that this negative regulation is lost when Rb is missing, hyperphosphorylated, or inactivated by viral oncoproteins, and releases E2F.

1.2 Pocket proteins p107 and p130

The discovery of two additional pocket proteins displaying Rb-like characteristics was met with excitement. Initial studies found that p107 and p130 also interacted with members of the E2F family, had the ability to arrest cell growth, and associated with viral proteins E1A and the SV40 T-antigen similar to Rb (Yee and Branton, 1985; Harlow et al. 1986; Dyson et al. 1989; Classon et al. 2001, Cobrinik, 2005). Additionally, p107 and p130 were found to be substrates of CDK activity and when hyperphosphorylated released E2Fs leading to cell cycle progression (Ahazawa et al. 2001; Calbo et al. 2002; Farkas et al. 2002). Given the functional overlap shared with Rb, p107 and p130 were assumed to also have tumor suppressor capabilities. While p107 and p130 possess some tumor suppression activity, mutations of these proteins are less frequently associated in cancer, and loss of either p107 or p130 alone does not result in tumor formation in mice (Burkhart

et al. 2008; Classon et al. 2001; Classon et al. 2002; Dannenberg et al. 2004; Robanus-Maandag et al. 1998).

Immunoprecipitation experiments with anti-p107 antibodies found that like Rb, p107 associated with 62-65 kDa proteins resembling E2F1. However, immunological and biochemical analyses with antibodies specific for different members of the E2F family found that p107 associated with an E2F that was distinct from E2F1 (Dyson et al. 1993). Additional studies confirmed that p107 and p130 associate exclusively with E2F4 and E2F5 (Dyson et al. 1993; Sardet et al. 1997; Trimarchi and Lees, 2002). Our current understanding of this pocket protein-E2F relationship has shown that while Rb is capable of associating with E2F1-5 in cells, p107 and p130 only associate with E2F4 and E2F5. This preference for distinct subsets of E2Fs within the pocket protein family may also give rise to some of the diverse phenotypes that are experienced in the absence of different pocket proteins.

1.3 Mouse genetics experiments reveal differences in pocket protein function

To identify the roles of Rb, p107, and p130 in normal mammalian development, genetically modified mice specifically lacking different members of the pocket protein family were generated. The results from these experiments showed a clear difference between pocket protein functions, and further pointed towards the role of Rb as a potent tumor suppressor. Mice that are nullzygous for Rb do not develop past embryonic week 13, and mice that are heterozygous for Rb display a strong tumor phenotype, with tumors appearing in the pituitary and thyroid tissues with high penetrance (Clarke et al. 1992; Lee et al. 1992; Jacks et al 1992).

In contrast, mice that are nullzygous for p107 or p130 can develop with no major defects, and do not display any tumor phenotype. (Cobrinik et al. 1996; Jacks et al. 1992; Lee et al. 1996). Dual knockout of p107 and p130 results in death shortly after birth,

suggesting that p107 or p130 can compensate for the loss of the other pocket protein, but at least one copy of p107 or p130 is necessary for development and viability (Cobrinik et al. 1996). While individual knockout of p107 has no tumor phenotype, p107 appears to protect against tumor formation to some degree. Rb heterozygous mice develop pituitary and thyroid tumors, but deletion of p107 in addition to removing a copy of Rb results in retinoblastoma and an increased tumor phenotype (Wirt and Sage, 2010).

It has been hypothesized that this drastic difference in developmental phenotypes and tumor progression is that deregulation of E2Fs is affected differently between the knockouts. Because both p107 and p130 regulate E2F4 and E2F5, one could reason that the lack of phenotype associated with a p107 or p130 knockout has to do with the remaining pocket protein compensating for loss of the other. If p107 is mutated, p130 is available to regulate E2F4 and E2F5. However, in the absence of Rb, the activator E2Fs (1,2,3) are now free from pocket protein regulation, as p107 and p130 cannot compensate. This deregulation of activator E2Fs also explains the prevalence of Rb mutations associated with human cancers, while mutations in p107 and p130 are quite rare (Burkhart and Sage, 2008). Additionally, reduction of activator E2F levels can reduce unscheduled proliferation of several cell types and reduces expression of oncogenes in small cell lung carcinoma, further pointing towards a role for activator E2Fs in promoting tumorigenesis (Liu et al. 2015; Chen et al. 2007; Landman et al. 2013; Saavedra et al. 2002; Coe et al. 2013). Clearly, regulation of the activator E2Fs is crucial in preventing aberrant proliferation and tumorigenesis. However, a molecular explanation detailing how Rb, but not p107 or p130, can regulate this class of E2Fs is lacking. This question of pocket protein molecular architecture and its influence on E2F binding specificity is addressed in Chapter 2 and Chapter 3.

.

1.4 p107 and p130 in DREAM

Another major difference between members of the pocket protein family is the ability to participate in formation of the DREAM (DP, Rb-like, E2F, and MuvB) complex, a negative regulator of hundreds of cycle genes (Litovchick et al. 2007). In quiescence, cells are temporarily removed from the cell cycle, with formation of the repressive DREAM complex as a major component of preventing transcription of cell cycle genes and maintaining quiescence (Sadasivam and DeCaprio, 2013). The human DREAM complex is a multi-protein complex, consisting of what is termed the "MuvB core" group of proteins LIN9, LIN37, LIN52, LIN54, and RbAp48, joined by either p107 and p130, as well as E2F4/5 and DP. This MuvB group of proteins are stably associated throughout the cell cycle and can participate in both repressive complexes, such as DREAM, and activating complexes when associated with the protein BMYB (Litovchick et al. 2007; Schmit et al. 2007). The functions of the proteins comprising MuvB are not well characterized. A small fragment of LIN52 is required for the formation of DREAM, LIN54 contains a DNA binding domain which represents a very small portion of the protein, LIN9 contains a putative Tudor domain, which typically bind histone tails, and RbAp48 is thought to be a chromatin assembly factor (Litovchick et al. 2007, Litovchick et al. 2011, Guiley et al. 2015, Marceau et al. 2016). While some functions of this complex have been elucidated, the mechanism by which it can function both as an activator and repressor remains unknown.

The expression profile of p130 is highest in G0, and therefore is the predominant pocket protein found in the DREAM complex (Grana et al. 1998; Mulligan and Jacks 1998). In cells lacking p130, p107 expression increases and substitutes for p130 in a functional DREAM complex (Sadasivam and DeCaprio, 2013). Concomitant knockdown of p107 and p130 results in the disassembly of DREAM and is associated with an increase in E2F target gene expression (Litovchick et al. 2007; Litovchick et al. 2011). While p107 or p130 can assemble into DREAM, *in vivo* studies have not found Rb capable of interacting with the

MuvB core group of proteins. The molecular mechanisms detailing the ability of p107 or p130 to form DREAM, but not Rb, remains a mystery. Rb is capable of regulating the same group of E2Fs as p107 and p130, so the mechanism behind the specificity of p107 and p130 for DREAM complex formation must be independent of E2F regulation. In Chapter 5 of this study, we provide insight into the molecular differences between p107/p130 and Rb that exclude Rb from participating in DREAM complex formation.

1.5 Domain structure of pocket proteins and E2Fs

The overall structure of pocket proteins is well conserved, and can be divided into three separate domains; the N-terminus, the pocket domain, and the C-terminus (Fig. 7). The Rb N-terminus is well structured, and though it does not directly interact with E2F, it plays an important role in the dissociation of E2F in response to CDK phosphorylation, as well as weakening the affinity for proteins binding via the LxCxE motif (Burke et al. 2013; Lamber et al; 2013). Separated by a short unstructured region from the N-terminus is the pocket domain, the most conserved domain amongst pocket proteins. The pocket domain can be sectioned in to two subdomains, RbA and RbB, separated by a loop in the primary structure (Balog et al. 2011; Burke et al. 2010; Burke et al. 2012; Burke et al 2013; Lee et al. 1998; Lee et al. 2002; Rubin et al. 2005). The pocket domain contains two distinct binding regions that contribute to its growth arrest potential; the E2F binding pocket and the LxCxE cleft. Experiments in which the Rb pocket domain was fused to E2F1 determined the pocket to be necessary, but not sufficient, for inhibiting E2F1 activity and establishing a G1/S block (Sellers et al. 1995). The C-terminal domain of Rb (RbC) is an unstructured region of the protein that also serves a role in regulation of E2Fs and is necessary for cell cycle arrest (Rubin et al. 2005, Whitaker et al. 1998).

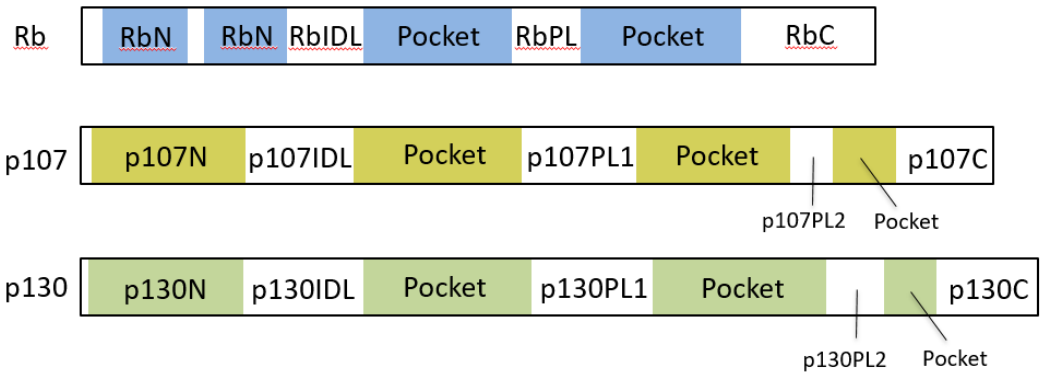


Figure 1: Structural conservation among pocket proteins. Each protein contains an N-terminal domain (RbN and p107N), a pocket domain, and disordered sequences including the C-terminus (RbC and p107C), the interdomain linker(RbIDL and p107IDL) and the pocket loops (RbPL, p10PL, and p107PL2). Domains with known structures are colored, although in the case of p107 the coloring and boundaries are based on structure prediction.

The five canonical E2Fs show strong conservation in their structural organization (Fig. 8) The DNA binding domain (DBD), which functions as a heterodimer with DP to associate with E2F target promoters containing the TTTSSCGC sequence, is similar among E2F1-5. The noncanonical E2F7 and E2F8 contain two DNA binding domains and binds DNA without the DP DBD. Structural comparison of E2F4 and E2F8 show that both use a similar motif to bind E2F promoter sequences and show preference for the same promoter sequences (Morgunova et al. 2015; Zheng et al. 1999). Two additional dimerization domains, the coiled-coil (CC) and marked-box region (MB), associate with DP and together serve as a docking interface for pocket C-termini (Rubin et al. 2005). The most C-terminal domain, the transactivation domain, is well conserved and is responsible for interacting with the pocket domain, and recruiting transcriptional machinery to drive gene expression (Xiao et al. 2002; Lee et al. 2002).

The canonical E2Fs commit a large portion of their structures to interacting with their obligate heterodimerization partner, DP. Though there are two copies of DP in humans, DP1 and DP2, dimerization of either DP appears to achieve the same function.

The E2F family, however, can be separated in to groups of activators and repressors based on their transactivation potential at E2F target genes. E2F1, 2, and 3 belong to the activator family, and are thought to be responsible for driving the cell cycle after release by Rb. E2F4 and E2F5 seem to function mostly as repressors of genes required for S phase. ChIP data looking at promoter occupancy of pocket protein-E2F complexes find E2F4 to be located on chromatin in quiescence and early G1, accompanied by p107 or p130 (Chicas et al. 2010). In late G1, these repressive complexes are replaced by E2F1-3 that have dissociated from pocket proteins (Araki et al. 2003). This change in E2F promoter occupancy also coincides with increased levels of expression at these sites and a commitment to the cell cycle.

The inability of E2F4 and E2F5 to transactivate genes stems from two nuclear export signals located in the N-termini of these proteins. E2F1-3 contain nuclear import signals, allowing them to stay in the nucleus when E2F4 and E2F5 are localized to the cytoplasm (Dimova et al. 2005; Muller et al. 1997). Overexpression of the nuclear export protein CRM1 is sufficient to reduce repression of E2F targets by exporting E2F4/5 to the cytoplasm (Gaubatz et al. 2001). Additionally, inclusion of the E2F1 nuclear import sequence on E2F4 allows for E2F4 to remain in the nucleus following the release by pocket proteins, and can transactivate target genes (Muller et al. 1997). From these studies it seems clear that while E2F4 and E2F5 contain the necessary domains to recruit transcriptional machinery, their localization in a cell-cycle dependent manner dictates that the main function of E2F4 and E2F5 is to repress target genes during the cell cycle. However, more recent studies have provided evidence that E2F4 can function outside its canonical role as a repressor of cell cycle genes. In mice, E2F4 can upregulate the expression of cilia in multiciliate cells, and loss of E2F4 results in the absence of ciliated cells (Hsu and Sage, 2016). Knockdown of E2F4 in human intestinal cells leads to downregulation of E2F target genes, as well as a decrease in proliferation (Hsu and Sage,

2016). Interestingly, E2F4 overexpression in the epidermis of mice promotes the formation of skin tumors (Hsu and Sage, 2016). While E2F4 clearly plays a role in the repression of cell cycle genes, there is growing evidence that E2F4 can function as an activator in a context- or tissue-dependent manner.

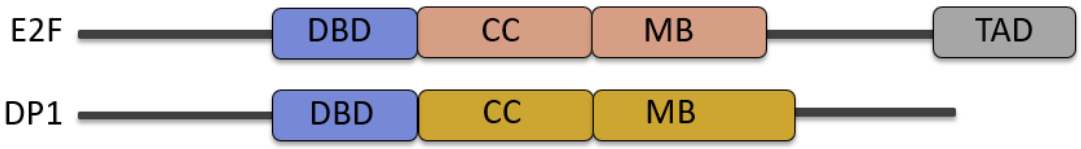


Figure 2: Structural schematic of E2F and DP1 domain organization.
The structured regions of E2F and DP1 are boxed and colored, with the unstructured regions depicted as lines. The DNA binding domain(DBD), coiled-coil (CC) and marked box (MB) domains are contained by E2F and DP1, and function as heterodimers. The transactivation domain (TAD) does not require dimerization with DP1. E2F1-E2F5 all contain the domains depicted above.

1.6 Interaction of Rb with E2Fs

The regulation of the E2F family of transcription is the best characterized activity of Rb. In the early stages of the cell cycle, Rb-E2F complexes are found within the nucleus. As G1 progresses and the expression of cyclins leads to active Cyclin-dependent kinase (CDK) activity, Rb becomes hyperphosphorylated, dissociates from E2F and is exported out of the nucleus (Jiao et al. 2006). This release of E2F seems to be sufficient for bypassing the restriction point and committing to the cell cycle, as reintroduction of growth suppression factors has no effect after this point (Pardee et al. 1989).

Structural work with the Rb-E2F complex have identified two different interaction surfaces that allow for suppression of E2F driven transactivation (Lee et al. 1998; Lee et al. 2002; Rubin et al. 2005; Xiao et al. 2003). First, the best conserved region of pocket proteins, the well-structured pocket domain, associates with the transactivation domain (TAD) of E2Fs. This region of E2Fs has been thought to be responsible for recruiting transcription machinery when not in complex with pocket proteins, and sequestration of the

TAD prevents recruitment of these factors to E2F genes. In a second interaction, the predicted unstructured region of Rb associates with the coiled-coil and marked box (CM) regions of E2F and DP. There are conflicting results regarding whether this is an association that is specific for Rb and E2F1, or is a general mechanism employed by all pocket proteins to bind E2F1-5 (Cecchini et al. 2014; Cecchini et al. 2014; Dick and Dyson 2003). *In vivo* work suggesting the specificity of this interaction pointed toward the presence of valine 276 in the E2F1 CM sequence. Only E2F1 contains a valine at this position, and when mutated to a proline to match the sequence present in other E2Fs, the association with Rb is lost. However, work by Rubin *et a*l using purified components, as well as the work presented in this thesis, suggests that the interaction of the pocket protein C-terminus with the E2F-DP CM domain is a general interaction shared by all pocket proteins.

Contained within the pocket domain is a well-conserved region known as the LxCxE binding cleft. In fact, this site is the most conserved region of the pocket protein family, and may represent the original function of these proteins. Through the LxCxE cleft, it is thought that pocket proteins can localize chromatin modifiers, such as histone deacetylases and histone methyltransferases, which promote heterochromatin formation and acts to reduce E2F target gene expression (Luo et al, 1998; Nielsen et al. 2001; Shao et al. 1995). However, biochemical data detailing the interactions of pocket proteins with chromatin modifiers, or indeed any other proteins involved in cell cycle regulation, is quite limited. The best characterized interaction of LxCxE containing proteins interacting in this cleft is that of viral oncoproteins. Structures of Rb and p107 have detailed that the E7 effector protein of the human papilloma virus binds to both proteins using a similar mechanism. Binding data studying this interaction has shown that p130 also interacts with E7 and has a similar affinity for E7 when compared to Rb or p107.

1.7 CDK activity regulating pocket protein function

Initial studies looking at the function of Rb found that it was phosphorylated in a cell cycle-dependent manner and allowed for the transition from G1 to S phase (Buchkovich et al. 1989; Chen et al. 1989; DeCaprio et al. 1989; Mihara et al. 1989). Work conducted before these studies had shown that in yeast the cell division control protein 2 (cdc2) was critical for the G1 to S phase transition as well (Nurse et al. 1976). Hypothesizing that the homolog in humans (CDK1) might also function as a regulator of this transition through phosphorylation of Rb, trypsin-digested fragments of Rb were incubated with CDK1 and were found to be phosphorylated, identifying the CDK family of proteins as potential regulators of Rb activity (Lin et al. 1991). When SaOS-2 cells, which lack Rb, express Rb ectopically, Rb is capable of halting proliferation. However, if Rb is ectopically expressed with either cyclin A or cyclin E, essential components of CDKs, the Rb-induced block on proliferation is removed (Dowdy et al. 2003). This result, combined with previous data, identified the CDK family of kinases as negative regulators of Rb activity.

The molecular mechanisms by which CDK can inhibit the activity of Rb have been well characterized structurally. Two CDK sites located in the flexible linker regions of Rb, S608 and S795, directly compete with the E2F transactivation binding the pocket domain when phosphorylated (Burke et al. 2010, Burke et al. 2013). Phosphorylation of S788/S795 also appears to inhibit binding between the RbC and E2F-DP CM domain (Rubin et al. 2005). Phosphorylation of T385, located in the RbIDL, induces an intramolecular association of the RbN with the Rb pocket domain. This association causes a rotation in the Rb pocket such that the E2F binding groove between the A and B subdomains widens, and weakens the affinity for E2Fs (Burke et al. 2012; Lamber et al. 2013). Located in the RbC, residues T821 and T826 inhibit the E2F-DP CM domain from interacting with the RbC, and also associates with the pocket domain such that binding to the LxCxE cleft is

occluded (Rubin et al. 2005). Clearly, Rb has evolved multiple mechanisms that regulate the many interaction surfaces used in binding E2Fs and LxCxE containing proteins.

p107 and p130 are also subject to CDK activity, and they share many of the CDK sites that have known functions in Rb. p107 and p130 are sensitive to expression of cyclin D and cyclin E, resulting in a decrease in pocket protein activity and a release of E2F (Calbo et al. 2002). Farkas et al (2002) mutated select CDK sites in p107 and p130, and demonstrated that these mutants showed an increase in resistance to CDK activity, and a higher percentage of cells in G1 compared to wild type p107 or p130. Phosphorylation of p107 or p130 by CDKs is also thought to be the mechanism that dissociates the DREAM complex and leads to cell cycle re-entry (Guiley et al. 2015). In addition to the phosphorylation sites that p107 and p130 share with Rb, each contains CDK consensus sites not present in Rb. Structural information regarding the mechanisms by which p107 and p130 are inhibiting following CDK activity is lacking, so in this study we set out to elucidate how CDK activity affects the function of p107 and p130 at the molecular level.

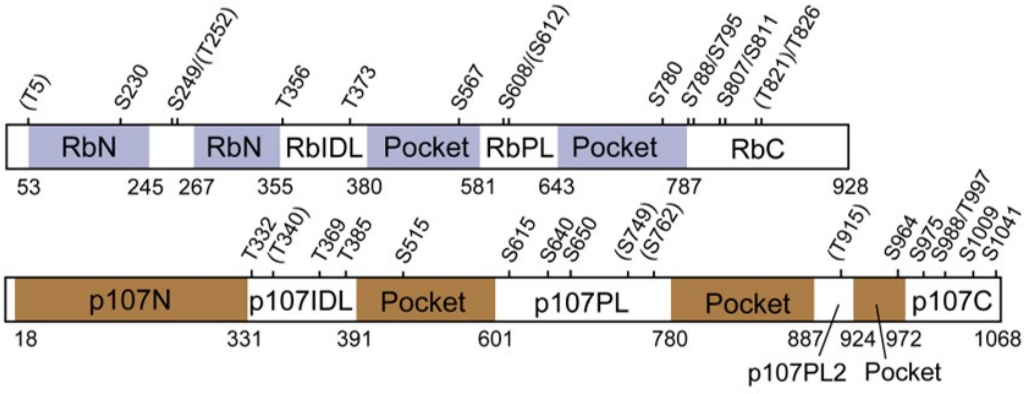


Figure 3: Overall architecture of pocket proteins Rb and p107 and location of CDK phosphorylation sites. Consensus CDK phosphorylation sites (SP and TP) are indicated, and those in parentheses are not even conserved among vertebrates.

1.2 Approach

The primary goal of this research has been to identify the molecular determinants that allow for E2F transcription factor binding specificity within the human pocket protein family, with the overall goal of relating this specificity back to the distinct functions observed between different pocket proteins. Decades of research spent studying Rb have demonstrated that Rb contains potent tumor suppressor activity, a function not shared by p107 or p130, and this function is potentially related to the regulation of a specific subgroup of E2Fs. Additionally, this research will focus on determining the mechanisms by which CDKs can inhibit pocket protein activity. In an attempt to answer these questions, we use a variety of biochemical approaches. To identity molecular determinants of specificity, functional truncates of whole proteins are used in X-ray crystallographic studies to determine differences in protein architecture at the atomic level. We monitor the binding affinities of different pocket protein mutants or pocket protein chimeras to verify we have identified molecular determinants of specificity using isothermal titration calorimetry (ITC). Lastly, we can quantitatively monitor the addition of phosphates at specific residues using LC-MS, and test the effect of phosphorylation at that residue using ITC. The fact that Rb is more commonly mutated in cancers, while p107 and p130 are typically functional, suggests the interesting therapeutic potential of tuning the E2F selectivity of p107 and/or p130 such that they regulate E2Fs in an Rb-like manner, and restore regulation of the cell cycle. Understanding the molecular determinants governing the E2F specificity within the pocket protein family is a first step towards achieving this potential.

Chapter 2: Comparison of the Rb and p107 pocket domain structures identifies molecular determinants of E2F transactivation domain binding specificity

2.1 Introduction

The retinoblastoma tumor suppressor protein (Rb) and its paralogs p107 and p130 negatively regulate proliferation through controlling the cell cycle, differentiation, senescence, and apoptosis (Burkhart et al. 2008; Classon et al. 2001; Classon et al. 2002; Cobrinik et al. 2005; Dick et al. 2013; Claudio et al. 2002).. Consistent with their sequence homology, the three members of the "pocket protein" family share a number of cellular functions. They all arrest cells when expressed, primarily due to their ability to associate with E2F transcription factor family members and repress E2F-mediated gene expression. Pocket proteins regulate transcription both through their direct inhibition of E2F and by recruiting transcriptional co-regulators that modify histones and chromatin structure. Similarly, all are inactivated by Cyclin-dependent kinase (CDK) activity and by proteins from oncogenic tumor viruses such as the E7 protein from human papilloma virus.

In contrast, a number of observations point to important differences in pocket proteins. Rb knockout is embryonic lethal in mice, whereas knockout of p107 or p130 results in a viable phenotype (Cobrinik et al. 1996; Jacks et al. 1992; Lee et al. 1996). Mice lacking different combinations of pocket protein genes suggest that p107 and p130 have an overlapping role in development that is distinct from Rb (Cobrinik et al. 1996). Importantly, the tumor suppressor properties of the Rb gene are thought to be stronger than p107 and p130. Mutations within the p107 and p130 genes are less common in cancer compared with Rb (Burkhart et al. 2008; Classon et al. 2001; Classon et al. 2002). Heterozygous mutant Rb mice spontaneously develop tumors, while p107 and p130 mutant mice do not (Cobrinik et al. 1996; Jacks et al. 1992; Lee et al. 1996; Mulligan et al. 1998). However, there is evidence that p107 and p130 have some tumor suppression function and can

compensate for Rb loss in certain context (Dannenberg et al. 2004). Mutations that lead to general pocket protein inactivation through kinase upregulation are much more common in tumors than loss of Rb, suggesting that broad pocket protein inactivation is necessary in many tissues (Burkhart et al. 2008; Classon et al. 2002; Dannerberg et al. 2004; Malumbres et al. 2001). p130 levels are inversely correlated with tumor grade in many cases, suggesting p130 can protect against tumor progression by stabilizing cell cycle exit (Claudio et al. 2002). Finally, p107 protects against Rb loss in several mouse cancer model (Wirt and Sage, 2010). The potential for p107 and p130 to replace Rb and each other suggests the intriguing therapeutic strategy of manipulating pocket protein functions so they can functionally compensate for each other where needed. To realize this potential, a better understanding of the molecular origins of their functional differences is needed.

Reflecting their genetic disparities, several distinct cellular functions for pocket proteins have been observed. Pocket proteins control distinct E2F target genes and arrest cells in different cell cycle phases (Hurford et al. 1997; Jiang et al. 2000). In a striking example, a genome-wide screen of gene repression in fibroblast cells revealed a unique role for Rb in promoting senescence (Chicas et al. 2010). Other additional activities have been attributed specifically to Rb, including its ability to stabilize the CDK inhibitor p27, promote apoptosis, and maintain genomic stability (Dick and Rubin, 2013). Conversely, p107 and p130 exclusively act as CDK inhibitors and promote quiescence through their assembly into the DREAM complex (Litovchick et al. 2007).

These similar and divergent pocket protein functions most likely reflect overlapping and distinct sets of protein interactions made by Rb and p107 or p130 (p107/p130). The pocket domain contains a cleft that binds an ‘LxCxE’ motif originally identified in viral oncoproteins (Lee et al. 1998). Histone deacetylases, chromatin remodeling complexes, Cdh1, and a number of other cellular proteins bind Rb in an ‘LxCxE’ cleft-dependent manner (Binne et al. 2007; Brehm et al. 1999; Morris and Dysonm, 2001), although detailed

structural analysis of these associations is limited (Singh et al. 2005). The best-characterized pocket protein interactions are with transcription factors. The pocket domain binds the E2F transactivation domain ($E2F^{TD}$), and a second binding interface exists between a section of the pocket protein C-terminal domain and the marked box (MB) domains of E2F and its obligate heterodimer partner DP (Fig. 7) (Lee et al. 2002; Rubin et al. 2005; Xiao et al. 2003). p107 and p130 also bind viral and chromatin factors through their 'LxCxE' cleft, and they use this cleft to bind LIN52 in the DREAM complex (Mulligan et al. 1998; Guiley et al. 2015). It has been observed in cells that p107/p130 binds E2F4 and E2F5 but not E2F1, E2F2, and E2F3, while Rb is found bound to E2F1 through E2F5 (Trimarchi and Lees, 2002). Several other protein interactions are specific for Rb or p107/p130. Most notably, Rb exclusively binds Cdh1 and PP1 (Binne et al. 2007; Hirschi et al. 2010) while p107/p130 bind CDKs with higher affinity, the LIN52 component of the MuvB complex to form DREAM, and protein phosphatase 2A (Mulligan et al. 1998; Litovchick et al. 2007; Guiley et al. 2015; Jayadeva et al. 2010).

CDK phosphorylation inhibits the ability of pocket proteins to negatively regulate the cell cycle, to bind E2Fs and other factors, and to repress E2F-activated transcription (Ashizawa et al. 2001; Buchkovich et al. 1989; Chen et al. 1989; Farkas et al. 2002; Garriga et al; 1998; Hansen et al. 2001; Lees et al. 1991; Rubin 2013). Analysis of the effects of CDK phosphorylation on Rb structure has revealed that specific phosphorylation events mediate distinct conformational changes in Rb that disrupt interactions with E2F and potentially 'LxCxE'-binding partners (Rubin 2013). Like Rb, p107/p130 contains a large number of CDK phosphorylation sites that regulate their activity and protein interactions (Ashizawa et al. 2001; Farkas et al. 2002; Garriga et al; 1998; Hansen et al. 2001), however the structural effects of p107/p130 phosphorylation have not been explored.

We have investigated here the structural conservation of p107/p130, its interactions with E2F, and how these interactions are inhibited by specific phosphorylation

events. The goal is a comparative analysis of pocket protein structure, biochemical interactions, and regulation such that genetic and functional differences can be understood in the context of their structural origins. We use the recently determined crystal structure of the p107 pocket domain to identify the molecular determinants of E2F specificity among pocket proteins and to demonstrate that binding specificity can be tuned both *in vitro* and in cells.

2.2 Results

2.2.1 Structural conservation of the pocket domain

Our understanding of pocket protein structure primarily comes from detailed analysis of Rb. Rb contains a structured N-terminal domain (RbN), the pocket domain, and an intrinsically disordered C-terminal domain (RbC) (Fig. 7a) (Dick and Rubin 2013). There are several other unstructured sequences, including an interdomain linker between RbN and the pocket (RbIDL) and a ~60 amino acid loop in the pocket domain (RbPL). p107 and p130 share 54% sequence identity with each other, share 30% identity with Rb, and appear to have similar domain structure as Rb. For example, p107 has a pocket domain (Guiley et al. 2015), a predicted structured N-terminal domain (p107N) and disordered C-terminus (p107C), an interdomain linker (p107IDL), and two large internal loops within the pocket domain (p107PL and p107PL2). Notably, while CDK sites are distributed throughout the pocket proteins, their localization to sequences within the disordered IDL, PL, and C regions is similar (Fig. 7a).

The crystal structure of the p107 pocket domain was recently determined in complexes with peptides that bind the 'LxCxE' site (Guiley et al. 2015). We performed a detailed comparison of the p107 and Rb structures in order to understand differences that may account for separate functions, particularly differences in E2F binding specificity. The

overall similarity between the p107 and Rb pocket domains is high with an RMSD in C_a position of 1.1 Å (PDB codes 1GUX and 4YOZ). Moreover, most of the secondary structural elements compare with the exception of two extra small helices in p107 (α4' and α10').

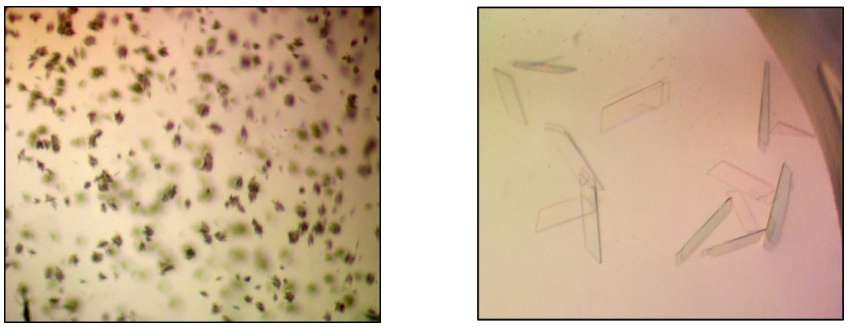


Figure 4: Crystals of the p107 pocket domain in complex with the HPV E7 peptide
Initial screening of 384 conditions resulted in 3 crystal hits, one of which was optimized to achieve diffraction quality crystals. The initial crystal hit (left) produced small clusters of needles that grew after 1-2 weeks. After several rounds of optimization, these crystals adopted a plate-like morphology(right), and although fragile, diffracted well and allowed us to determine the first structure of the p107 pocket domain.

To identify the most highly conserved pocket structural features, we generated an alignment of 51 pocket domain sequences from 31 metazoan organisms (Fig. 5 and 6). The alignment produces 30 amino acids that are identical in 90% of the sequences and can be divided into four clusters (Fig. 7b). Two highly conserved clusters form the structural core of the A subdomain (in human p107: E507, F519, P520, F533, I537, L547, H554, L555) and B subdomain (Y796, L806, L810, W821, R838, D841, Q842, I864, Y868). The third group contains residues critical for the A-B interface (K535, E538, E559, W568, K794, R802, H839), and the fourth group forms directly or structurally supports the LxCxE binding site (F861, F933, Y934, N935). The conservation of N935 is particularly notable, as it plays no obvious role supporting pocket domain structure. N935 faces outward and contacts proteins that bind at the cleft. In the crystal structures of p107 bound to either the LIN52 or the E7 peptide, the N935 amide sidechain makes bidentate hydrogen bonds with the 'LxSxE' or 'LxCxE' backbone (Guiley et al. 2015), and the same interaction is observed in

the Rb-E7 complex (Lee et al. 1998). The structural conservation is consistent with the similar affinity of the E7 peptide for both Rb and p107 pocket domains (Guiley et al. 2015; Lee et al. 1998).

	p107-E7
Data collection	
Space group	$C2_1$
Cell dimensions	
a, b, c	99.7, 76.6, 74.7
α, β, γ	90, 120.3, 90
Resolution (Å)	57.2 – 2.2
R_{pim}	5.8% (25.2%)
Total reflections	48201
Unique reflections	21552
I/σ	8.6 (2.1)
Completeness	93.2 % (94.4%)
Redundancy	2.2 (2.2)
Refinement	
Resolution	57.2 – 2.2
Number of reflections	21552 (2171)
R_{work}/R_{free}	19.0 / 25.1
Number of atoms	2949
Protein	2801
Water	128
RMS deviations	
Bond lengths	0.008
Bond angles	1.06
Average B factor	31.60
Ramachandran analysis	
Favored	96.0 %
Outliers	0.3 %

Table 1: X-ray crystallography collection and refinement data statistics for the p107 pocket domain in complex with the HPV E7 peptide
Crystallographic data was collected at the APS beamline 23-IDB.

Figure 5: Pocket domain sequence alignment. Sequences are shown for 51 pocket protein sequences from 31 organisms. Degree of conservation is shown in the colored bars.

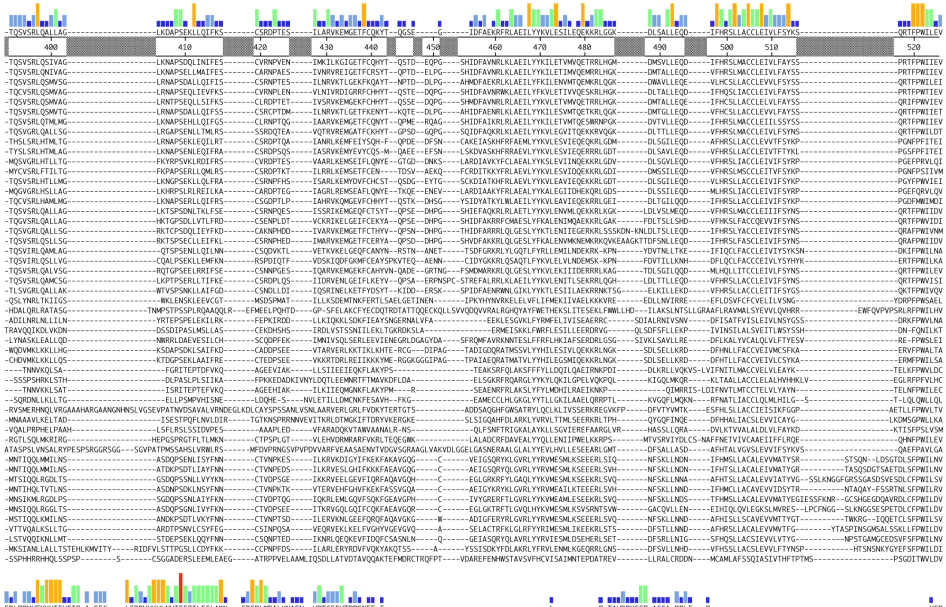
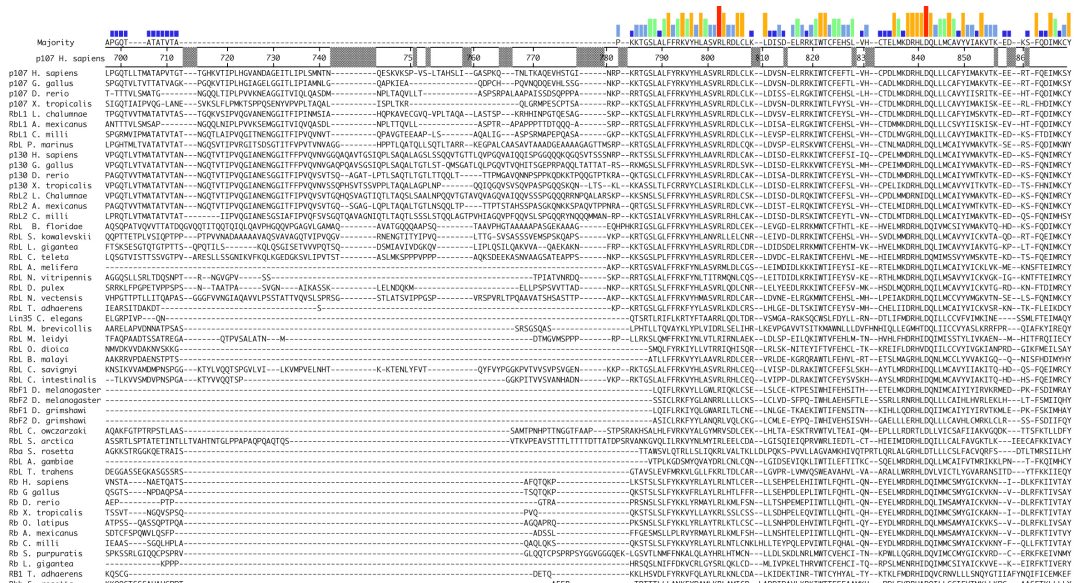


Figure 6: Pocket domain sequence alignment. Sequences are shown for 51 pocket protein sequences from 31 organisms. Degree of conservation is shown in the colored bars.



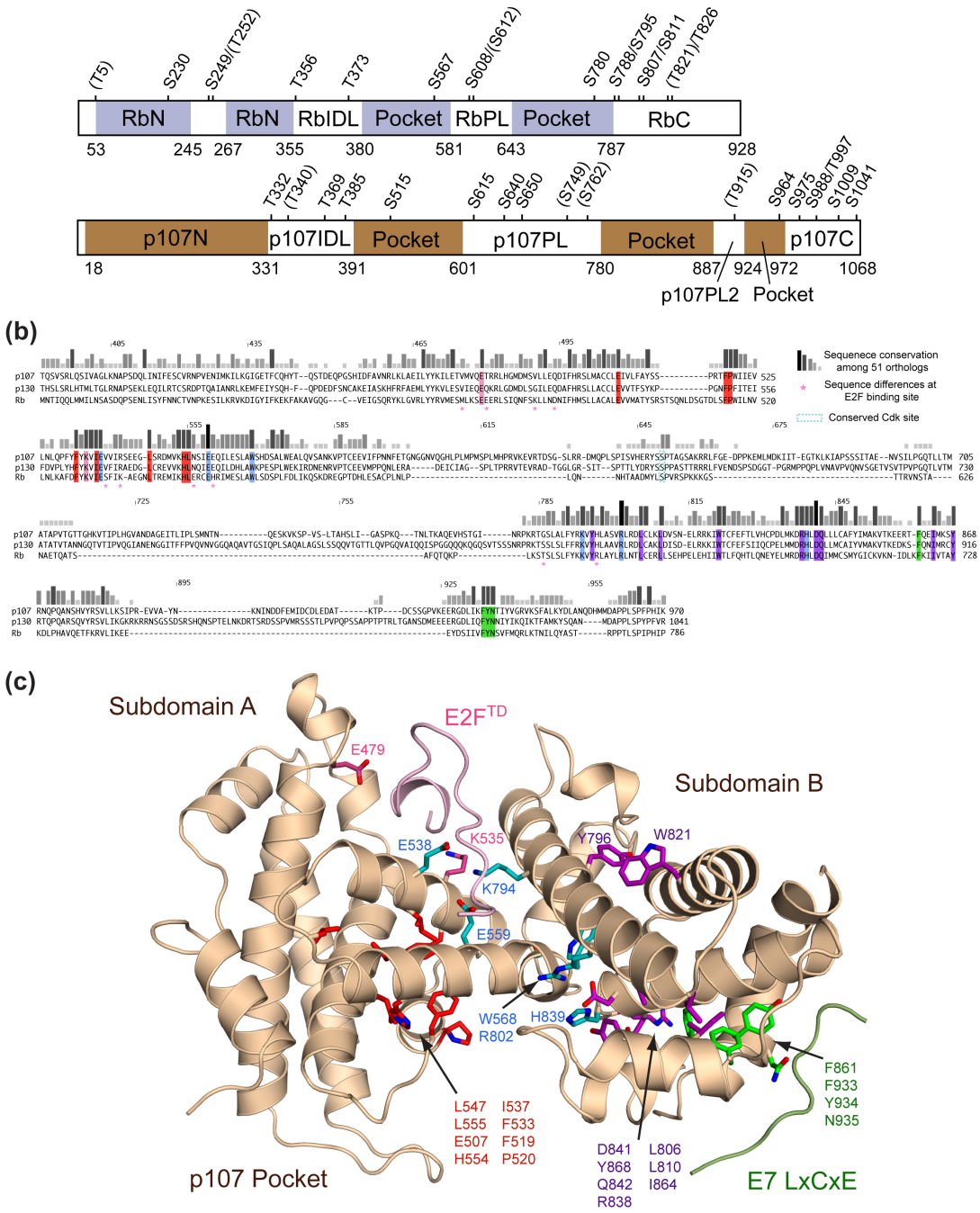


Figure 7: Structural conservation among pocket domains. (a) Overall architecture of Rb and p107. Each protein contains an N-terminal domain (RbN and p107N), a pocket domain, and disordered sequences including the C-terminus (RbC and p107C), interdomain linker (RbIDL and p107IDL), and the pocket loops (RbPL, p107PL, and p107PL2). Domains with known structure are colored, although in the case of p107N, the

coloring and boundaries are based on structure prediction. Consensus CDK phosphorylation sites (SP or TP) are indicated, and those in parentheses are not even conserved among vertebrates. **(**b) Pocket domain sequence alignment. Sequences are shown for human Rb, p107, and p130. Relative conservation among 51 sequences from 31 organisms is represented by the bars (full alignment shown in Fig. 5 and 6). Residues that are conserved among >90% of the sequences are highlighted in colors and divided into clusters by location in their structure. These residues are either in the structural cores of subdomain A (red) and subdomain B (magenta), at the A-B interface (blue), or at the E2F (pink) and 'LxCxE' (green) binding sites. Residues that line the $E2F^{TD}$ binding site and differ between Rb and p107 are noted with a pink asterisk as is the conserved CDK phosphorylation site in the pocket loop (blue dashed box). (c) Structure of the human p107 pocket domain (PDB Code: 4YOZ). $E2F2^{TD}$ is modeled in a bound conformation by aligning the p107 structure with the Rb-$E2F2^{TD}$ structure (PDB Code: 1N4M). Residues with >90% conservation are colored as in (b).

We further compared the p107 pocket domain with structures of the Rb-$E2F^{TD}$ complex to examine conservation of the $E2F^{TD}$ binding site and identify residues that may explain differences in the binding of human pocket proteins to different E2Fs. $E2F^{TD}$ binds the A-B pocket interface by making two sets of contacts (Lee et al. 2005; Xiao et al. 2003). At its N-terminal end, $E2F^{TD}$ binds in an extended conformation and makes interactions with α8 and α9 in the A-subdomain and α11 in the B-subdomain. The C-terminal portion binds as a short helix to α4, α5, α6 and α8 in the A-subdomain. With the exception of E479 and K535, there is much less conservation within the E2F binding site of pocket protein orthologs (Fig. 7). A few other conserved sidechains (E538, E559, and K794) contact E2F but also clearly play a role in stabilizing the A-B interface. By aligning the structures of human p107 and Rb pockets, we identified ten amino acids near the pocket-E2F interface that differ between the two human paralogs (Fig. 7), and we examine the effect of these residues on binding specificity below.

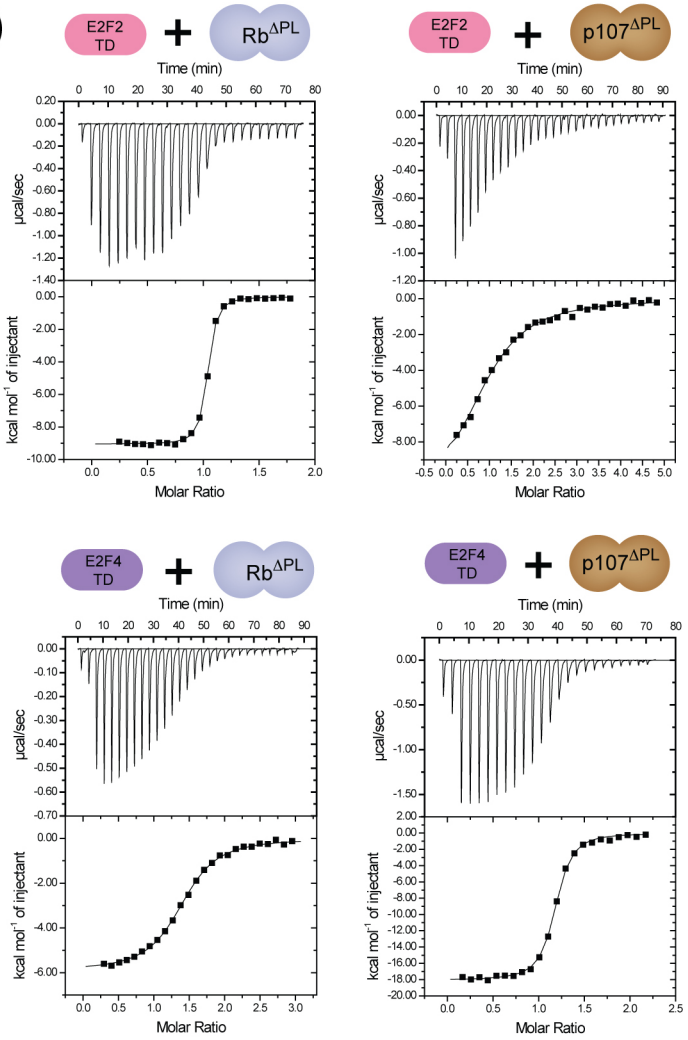
2.2.2 Structural basis for $E2F^{TD}$-pocket specificity determinants

We applied sequence comparison of E2F family members and the structures of the Rb and p107 pocket domains to understand the structural origins of binding specificity among E2F and pocket protein family members. The E2F transcription factor family is commonly divided into activating (E2F1-E2F3) and repressive (E2F4-E2F8) subgroups based on their respective roles in regulating transcription. While E2F1-E2F5 all contain pocket-binding domains, based on co-immunoprecipitation experiments from cell extracts, it has been understood that Rb binds all E2Fs whereas p107/p130 have specificity for the repressive E2Fs (E2F4 and E2F5) [27].

We tested whether the observed specificity in cells correlates with differences in pocket protein-E2F transactivation domain binding affinities. We purified recombinant Rb and p107 pocket domains that lack their large internal loops ($Rb^{\Delta PL}$: residues 380-787 with 581-642 deleted, $p107^{\Delta PL}$: residues 391-972 with 601-779 and 888-923 deleted). We measured affinity for the E2F1, E2F2, E2F3, and E2F4 transactivation domains using isothermal titration calorimetry (Fig. 8). Consistent with previous measurements (Lee et al. 2002; Burke et al. 2014), we found that $Rb^{\Delta PL}$ has higher affinity for $E2F1^{TD}$ and $E2F2^{TD}$ than for $E2F3^{TD}$ and $E2F4^{TD}$. In contrast, we found here that $p107^{\Delta PL}$ binds $E2F4^{TD}$ with higher affinity than $E2F1^{TD}$, $E2F2^{TD}$, and $E2F3^{TD}$.

Figure 8: Specificity in pocket-$E2F^{TD}$ affinities. (a) Isothermal titration calorimetry data for the indicated binding reactions and (b) table of ITC measured affinities of the Rb and p107 pocket domains for each $E2F^{TD}$.

(a)



(b)

	$Rb^{\Delta PL}$	$p107^{\Delta PL}$
	K_d (µM)	
E2F1 TD	0.07 ± 0.01	1.06 ± 0.04
E2F2 TD	0.04 ± 0.01	3.1 ± 0.3
E2F3 TD	1.2 ± 0.1	7 ± 1
E2F4 TD	0.5 ± 0.2	0.15 ± 0.01

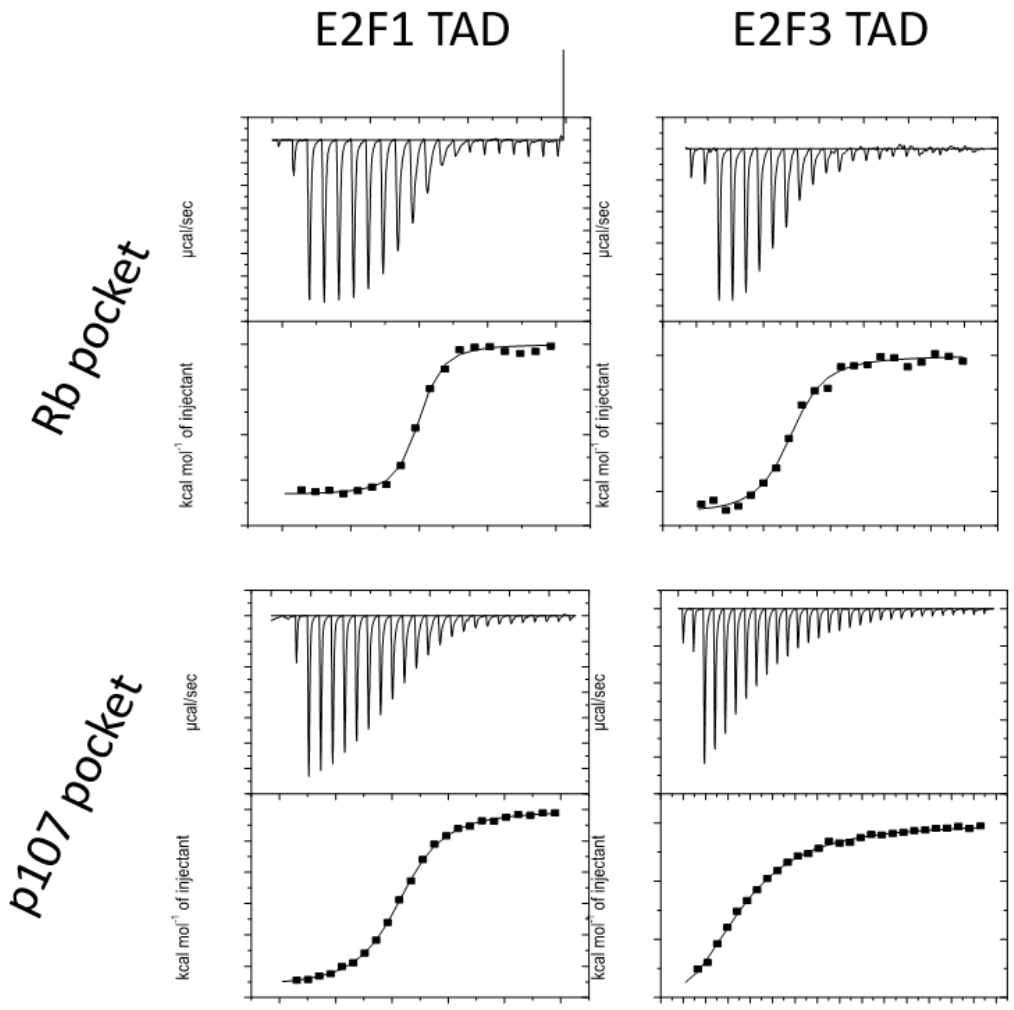


Figure 9: Representative ITC binding data for Rb and p107 pocket domains binding E2F1 and E2F3 as presented in Figure 5.

We first sought an understanding of the higher affinity of Rb for E2F1 and E2F2 compared to the other E2Fs. We looked at an E2F sequence alignment for residues that

are conserved in E2F1 and E2F2 but distinct in the other E2Fs (Fig. 10a). The glycine at position 414 in E2F1 (415 in E2F2) fits this criterion. In both the Rb-E2F1^{TD} and Rb-E2F2^{TD} structures, this glycine packs against helix α11 of the pocket B subdomain such that K652 in Rb hydrogen bonds to the glycine carbonyl oxygen (Fig. 10b). K652 also makes a sidechain hydrogen bond to E416/E417 in E2F1/E2F2, which is not conserved as a glutamate in E2F3. The close proximity of G414/G415 to α11 suggests that the bulkier sidechains present in the other E2Fs may clash with the helix and disrupt the ideal position of K652.

We mutated G415 in E2F2 to resemble E2F4, and we found that the G415N mutant E2F2^{TD} binds Rb$^{\Delta PL}$ with 6-fold weaker affinity (Fig. 10c). We also found that an N395G mutation in E2F4 results in a five-fold increase in affinity of E2F4^{TD} for Rb$^{\Delta PL}$ such that its affinity is similar to E2F1^{TD} and E2F2^{TD}. These data point to G414/G415 as a key residue for mediating the high affinity interaction of Rb for E2F1 and E2F2. The importance of this glycine for E2F^{TD} binding to p107$^{\Delta PL}$ is less evident, although we do observe a weaker affinity of E2F2^{TD} G415N compared to wild-type (Fig. 10c). p107 has R793 at the analogous position to K652 in Rb and may also clash with N395 in E2F4. We propose that other interactions between E2F4 and p107 must stabilize E2F4 binding to p107 relative to E2F2 as observed in Fig. 8.

We performed a similar analysis to attempt to understand why p107 binds E2F4^{TD} with higher affinity than the activating E2Fs. Examination of the E2F^{TD} sequences identifies E2F4 V402 and C403 as candidate specificity-determining residues that are similar in E2F4 and E2F5 but different in E2F1-E2F3. We found that a double I422V and S423C mutation in E2F2^{TD} has little effect on affinity for p107$^{\Delta PL}$. However, unexpectedly, a double V402I and C403S mutation increases E2F4^{TD} affinity slightly.

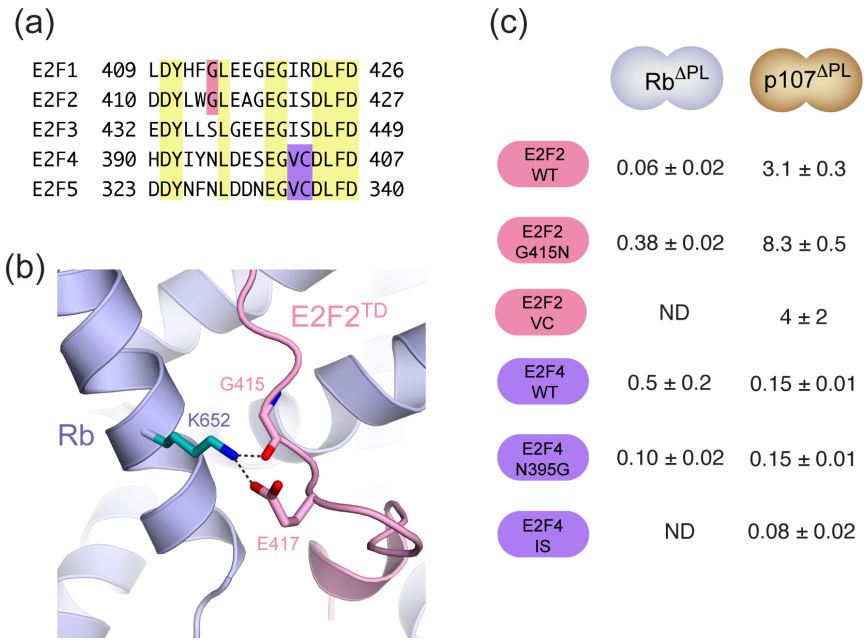


Figure 10: Preference of Rb for E2F1/E2F2. (a) Sequence alignment of E2F transactivation domains. Conserved residues are highlighted yellow, G414/G415 are highlighted pink, and residues that are only similar in E2F4/E2F5 are highlighted purple. (b) Structure of $Rb^{\Delta PL}$-$E2F2^{TD}$ complex shows proximity of G415 to K652 in helix α11. The structure predicts that glycine substitution would lead to potential steric clash. (c) ITC measurements of $Rb^{\Delta PL}$ and $p107^{\Delta PL}$ for the indicated E2F construct. VC is an I422V and S423C mutation in $E2F2^{TD}$, and IS is a V402I and C403S mutation in $E2F4^{TD}$.

2.2.3: Structural differences explain activating E2F preference for Rb

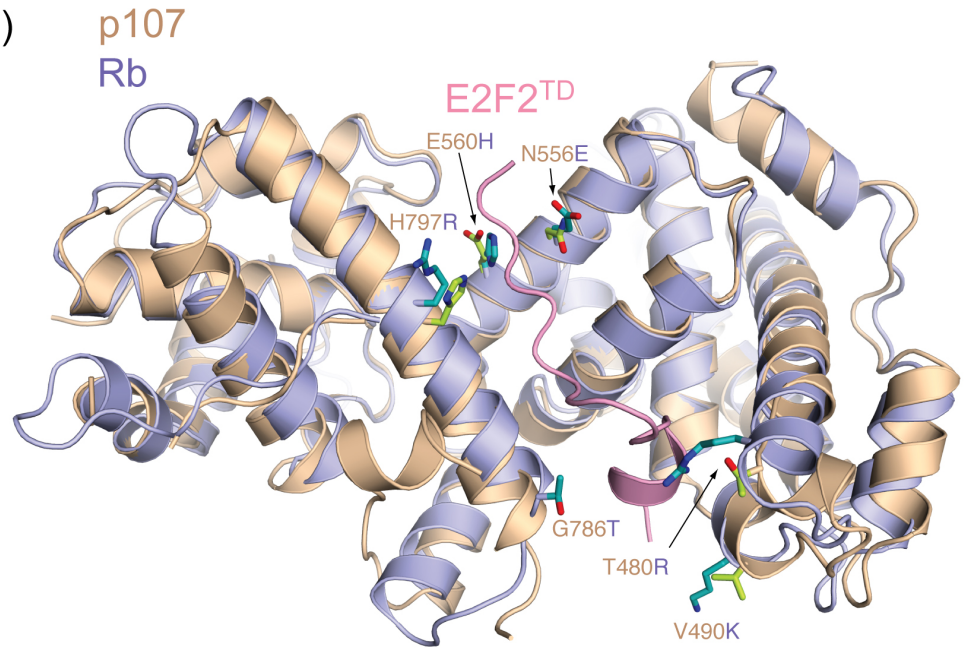
We next compared and manipulated the pocket domain structures to determine the basis for the higher affinity of the activating E2Fs for Rb compared to p107. There are ten residues that line the E2F binding cleft in the pocket and are distinct between p107 and Rb (Fig. 7). We predicted that six mutations in p107 (T480R, V490K, G786T, N556E, E560H, H797R) would influence E2F binding based on conservation and their observed interactions with the transactivation domain in the Rb-$E2F^{TD}$ structures (Fig. 11a). We mutated these six residues in $p107^{\Delta PL}$ to create a construct ($p107^{6x}$) that has an Rb-like E2F binding cleft. We then tested by ITC and found that $p107^{6x}$ bound $E2F1^{TD}$ and $E2F2^{TD}$ with three-fold and twenty-four fold higher affinity than $p107^{\Delta PL}$ respectively but bound

E2F4^{TD} with similar affinity (Fig. 11b). Notably, the affinity of p107^{6x} for E2F2^{TD} is nearly similar to the affinity of Rb for the activator E2Fs.

We then mutated these six residues individually in p107$^{\Delta PL}$ and tested binding to E2F2^{TD} (Fig. 11c). We found that E560H, H797R, and V490K each bound with approximately 3-fold increases in affinity, while the effects of the other three mutations were more modest. The structure of the Rb pocket domain bound to E2F2^{TD} provides rationale for why E560H, H797R, and V490K increase affinity of p107 for E2F2^{TD} (Fig. 11d). E560H likely creates a favorable electrostatic interaction with nearby D411 and D410. While D410 is further from the H555/E560 position in the Rb-E2F2^{TD} structure, it is notable that E2F1 and E2F4 have leucine and histidine respectively at the analogous 410 position (Fig. 10a), which possibly explains why the increase in affinity of the 6x-mutation is most prominent for E2F2. To support this idea, we mutated the E2F2 aspartates to histidines in order to create favorable electrostatic interactions with the E560 in wild-type p107$^{\Delta PL}$. We found that a D410H mutation increases the affinity of E2F2^{TD} for p107$^{\Delta PL}$ about 3-fold (Fig 11c). In contrast, making the histidine mutation at D411, which is conserved across all E2Fs, does not influence the affinity. The H797R mutation creates a favorable hydrophobic interaction with L413, as the imidazole group is replaced with a methylene in the arginine sidechain. The V490K mutation creates a favorable electrostatic interaction with the sidechain of D427, which is conserved in all E2F^{TD}s.

Figure 11: Preference of activating E2Fs for Rb over p107. (a) Overlay of Rb$^{\Delta PL}$ and p107$^{\Delta PL}$ structures. Sidechains are shown for six residues that are near the E2F binding site and differ between the two pocket domains. (b) ITC measurements E2F peptides binding to wild-type p107$^{\Delta PL}$ and p107$^{\Delta PL}$ containing six mutations that are changes of p107 to the Rb sequence (p107^{6x}). (c) ITC measurements of individual mutations as indicated. (d) Comparison of the Rb$^{\Delta PL}$ and p107$^{\Delta PL}$ structures explains the preference of Rb for E2F2^{TD}.

(a)



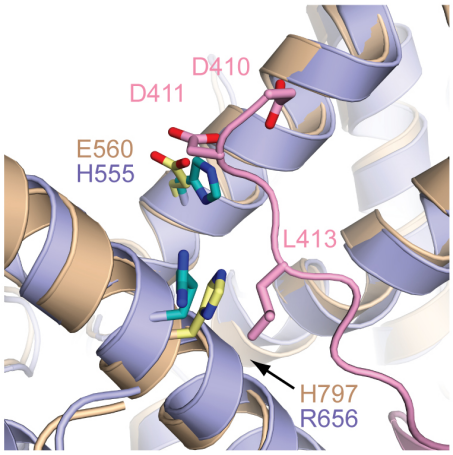
(b)

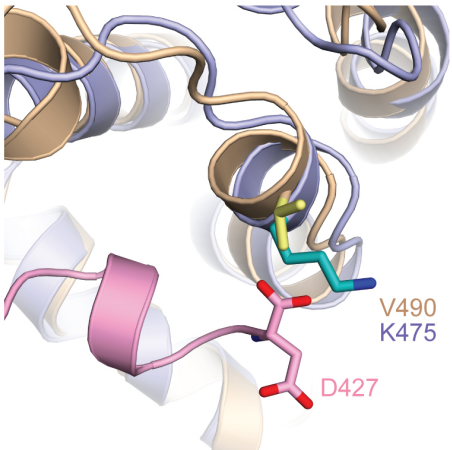
	p107ΔPL	p107 6x
	Kd (μM)	
E2F1 WT	1.06 ± 0.04	0.37 ± 0.02
E2F2 WT	3.1 ± 0.3	0.13 ± 0.02
E2F4 WT	0.15 ± 0.01	0.14 ± 0.02

(c)

p107ΔPL mutant	E2F2 WT Kd (μM)
T480R	2.6 ± 0.1
V490K	1.00 ± 0.01
N556E	2.2 ± 0.2
E560H	1.1 ± 0.5
G786T	2.1 ± 0.2
H797R	0.9 ± 0.4

(d)





2.2.4 The $p107^{6X\text{-}FL}$ protein associates with and regulates the activity of E2F2

We tested whether the mutations that increase the affinity of the p107 pocket domain for $E2F2^{TD}$ lead p107 to display activity in cells that resembles a gain in E2F2 regulation. We engineered the six E2F-binding mutations into full-length p107 ($p107^{6x\text{-}FL}$) and transfected the construct into C33A cells. Consistent with our calorimetry data using the isolated domains, we find that $p107^{6x\text{-}FL}$ co-precipitates more transfected E2F2 than wild-type p107 (Fig. 12a). We used a luciferase reporter assay to measure E2F2 driven gene expression and tested the effects of $p107^{6x\text{-}FL}$ in repressing E2F2. This experiment reveals that $p107^{6x\text{-}FL}$ more efficiently represses transcription by transfected E2F2 than $p107^{WT}$ (Fig. 12b). These data support the conclusion that structural differences between pocket protein-$E2F2^{TD}$ interactions can alter E2F activity in cells.

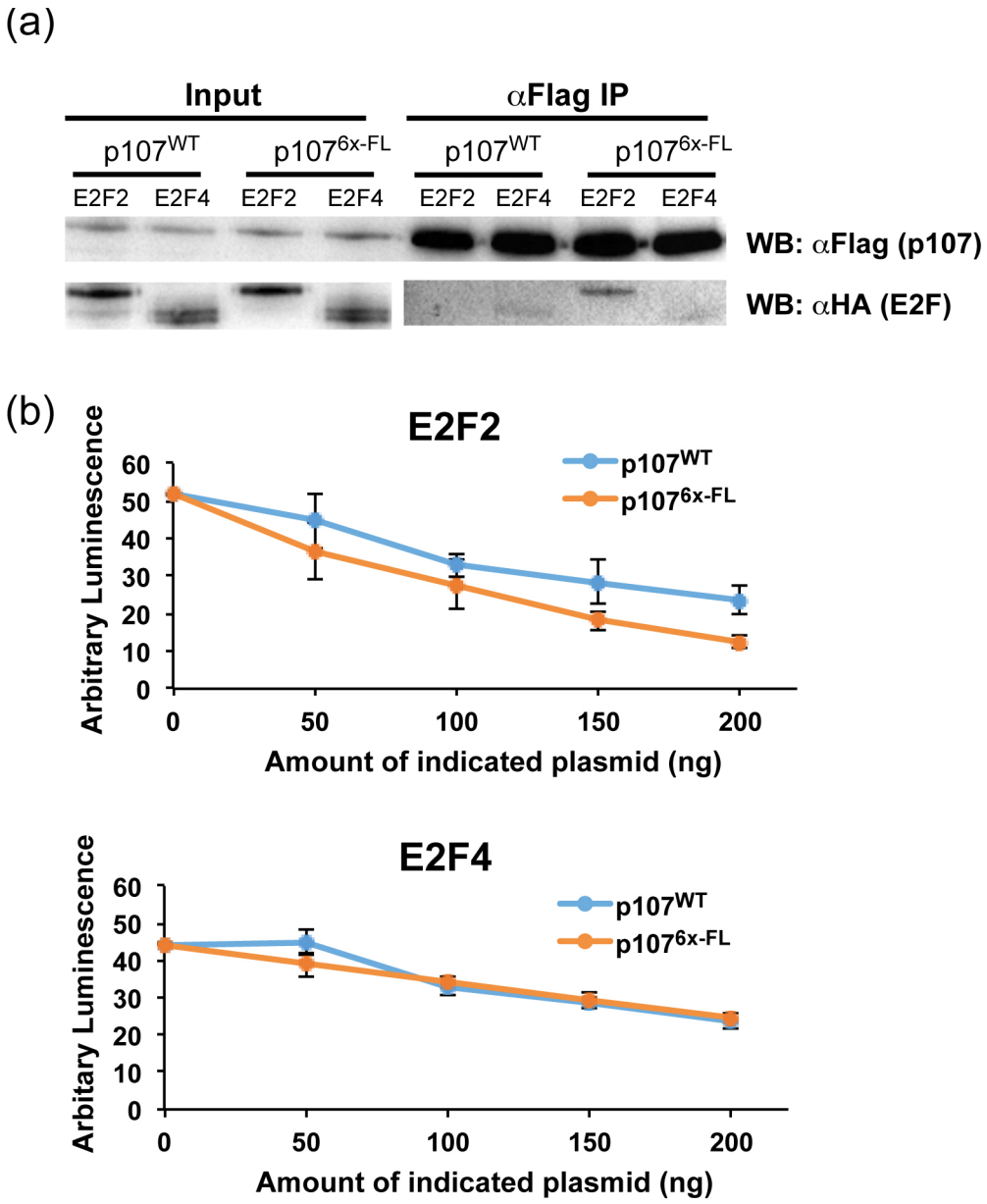


Figure 12: Mutation of the E2F-binding cleft in p107 enhances affinity for and regulation of E2F2. (a) C33A cells were co-transfected with Flag-tagged $p107^{WT}$ or $p107^{6x\text{-}FL}$, HA-tagged E2F2 or E2F4, and DP1 plasmids. Cell extracts were immunoprecipitated with anti-Flag antibodies and western blots of extracts and precipitated proteins were probed with the indicated antibodies. (b) Saos2 cells were transfected with an HA-E2F2 expression plasmid, DP1, and a p107 mutant or wild-type construct. Transfections also included an E2F responsive luciferase reporter plasmid and transcriptional activity was determined by luciferase activity and normalized to a non-p107 transfected control. Error bars indicate one standard deviation from the mean (n=9). Means were compared by a t-test and significant differences are indicated by an asterisk (P<0.05).

2.3: Discussion

Two broad questions about the pocket protein family remain unaddressed: what are the different cellular functions of Rb, p107, and p130 that account for their different tumor suppressor properties and what are the distinguishing molecular features that account for these differences? We have explored here the idea that functional differences may arise from different interactions with E2F family members and demonstrate that the pocket protein-E2F pairings observed in cells result from sequence differences that impart different binding affinities. We have found that Rb possesses higher affinity for the transactivation domains of E2F1 and E2F2, whereas p107 has preference for E2F4. We have accounted for these preferences by examining sequence and structural features that differ among pocket protein and E2F family members, and importantly, we have shown that affinities and preferences can be altered. For example, we have observed that the $p107^{6x\text{-}FL}$ mutant is capable of co-precipitating E2F2 and represses more efficiently E2F2-dependent transcription in cells. We propose using this reagent to examine whether the capacity to bind activating E2Fs confers other Rb-specific functions to p107 that may compensate for Rb deficiency in cells.

We examined pocket protein orthologous sequences from thirty metazoan species to identify the most conserved features. The highest conserved residues support the pocket domain structure, either by stabilizing the A and B subdomain folds or by stabilizing the interface between the two subdomains. Residues in the B subdomain at the LxCxE cleft are also highly conserved, suggesting that the best conserved and perhaps most ancient function of pocket proteins is binding protein targets at the LxCxE site. In contrast to pocket structure and 'LxCxE'-mediated protein interactions, the detailed mechanism of E2F-binding is less conserved. The expansion of the Rb family that occurs in early vertebrates is paralleled by an expansion in the E2F family. Subtle sequence changes that subsequently occurred have led to specificity in the interactions between family members.

Several of our binding measurements suggest that details of p107-E2F4^{TD} binding may be different than anticipated from the Rb-E2F1^{TD} and Rb-E2F2^{TD} structures. For example, we have not been able to understand why p107 has higher affinity for E2F4^{TD} compared to the activating E2F^{TD}s (Fig. 10). It is also interesting that the 6x-mutation, which changes p107 to mimic Rb, increases p107 affinity for E2F2^{TD} but does not weaken affinity for E2F4^{TD} in the ITC experiment. The structure of the p107 pocket domain has not yet been determined bound to an E2F^{TD}, and we expect that it may provide some unexpected details that explain the unique preference of p107 and p130 for the repressive E2Fs.

2.4: Materials and Methods

2.4.1 Protein expression, purification, and phosphorylation

All Rb, p107, and E2F constructs were expressed in *Escherichia coli*, using PET-derived vectors, as fusion proteins containing an N-terminal glutathione S-transferase (GST) tag. Transformed BL21(DE3) cells were grown to an OD_{600} of 0.6-0.8 and induced with 1 mM IPTG. Protein expression took place overnight at 22°C. Cells were resuspended in lysis buffer containing 25 mM Tris-HCl, 250 mM NaCl, 5 mM DTT, 5% glycerol, and 1 mM PMSF (pH 8). Cells were lysed by passing them three times through a cell homogenizer, and following centrifugation the resulting soluble fraction was purified using glutathione sepharose affinity chromatography. The subsequent eluate was further purified by anion-exchange chromatography and cleaved with GST-TEV overnight at 4°C. To remove cleaved GST, proteins were again passed over glutathione sepharose resin, and finally subjected to size exclusion chromatography to achieve a pure sample.

2.4.2 Isothermal Titration Calorimetry

Proteins were prepared for ITC by dialyzing overnight at 4°C in a buffer containing 100 mM NaCl, 20 mM Tris-HCl, and 1mM beta-mercaptoethanol (pH 8.0). Using a Micro-Cal VP-ITC calorimeter, typical binding experiments involved injecting 0.5-1 mM peptide into a 20-40 μM solution of p107 or Rb at 25°C. Binding constants were generated by fitting the data to a one-site binding model using Origin software. The error associated with the reported dissociation constants reflect the standard deviation calculated from 2-4 separate binding experiments.

2.4.3 Immunoprecipitation and Western Blotting

Immunoprecipitation was carried out as previously described (Cecchini and Dick, 2011). To generate extracts C33A cells were plated at $6x10^6$ cells per 15 cm plate and transfected with 40μg of either CMV-FLAG-$p107^{WT}$ or CMV-FLAG-$p107^{6x\text{-}FL}$, 20μg of either CMV-HA-E2F2 or CMV-HA-E2F4 and 20μg CMV-HA-DP1. Extracts were then normalized for transfection efficiency and immunoprecipitated using anti-FLAG M2 (Sigma). Immunoblotting was carried out using anti-FLAG M2 (Sigma) and anti-HA 3F10 (Roche).

2.4.4 Luciferase Reporter Assays

SAOS2 cells were plated at $7.5x10^5$ cells per well in a six well plate and transfected 24h later. Cells were transfected in triplicate with Fugene HD according to manufacturer's instructions (Fugene HD + 100μl DMEM). All transfections included the following: 100ng of pE2F4B-Luc reporter plasmid, 200ng CMV-βGal, 15 ng of CMV-HA-E2F2, and 15ng of CMV-HA-DP1. E2F-repression assays also included either 0, 50 ng, 100 ng, 150 ng, or 200 ng of CMV-FLAG-$p107^{WT}$ or CMV-FLAG-$p107^{6x\text{-}FL}$. CMV-CD20 was added to normalize p107 and CD20 plasmids to 200ng. Luciferase and βGal assays were performed as previously described (Dick et al. 2000), and luciferase activity was normalized to βGal from the same transfected extract.

Chapter 3: Conservation and Divergence of C-terminal Domain Structure in the Retinoblastoma Protein Family

3.1 Introduction

E2F transcription factors regulate the mammalian cell cycle by controlling expression of genes required for DNA synthesis and cell division (Lammens et al. 2009; Trimarchi and Lees, 2002). E2F activity is regulated by the retinoblastoma (Rb) "pocket" protein family members Rb, p107, and p130, which bind and inhibit E2F and recruit repressive factors to E2F-driven promoters (Burkhart and Sage, 2008; Classon and Dyson, 2001; Classon and Harlow, 2002; Cobrinik 2005; Dick and Rubin, 2013). These pocket protein-E2F complexes are the focal point of signaling pathways that trigger diverse cellular

processes including proliferation, differentiation, and apoptosis. Improper inactivation of pocket proteins is a common mechanism by which cancerous cells maintain aberrant proliferation (Burkhart and Sage, 2008; Classon and Harlow, 2002; Sherr 1996; Malumbres and Barbacid, 2001; Dyson 2016). Pocket protein-E2F dissociation and subsequent E2F activation is induced by cyclin-dependent kinase phosphorylation (Ashizawa et al. 2001; Buchkovitch et al. 1989; Chen et al. 1989; Farkas et al. 2002; Hansen et al. 2001; Rubin 2013)) or binding of viral oncoproteins such as the SV40 T-antigen (DeCaprio 2009; Felsani et al. 2006; Nevins 1994).

The E2F family contains eight members, five of which (E2F1-E2F5) form complexes with pocket proteins ((Lammens et al. 2009; Trimarchi and Lees, 2002). E2F1-E2F3 associate exclusively with Rb and are potent activators of transcription during the G1 and S phases of the cell cycle (Helin et al. 1992; Lees et al. 1993). E2F4 is found in complexes with all three pocket proteins and typically occupies promoters of repressed genes. E2F4 is thought of primarily as a repressor because it is exported from the nucleus upon release from pocket proteins (Dyson et al. 1993; Moberg et al. 1996; Muller et al. 1997; Verona et al. 1997). E2F4 and E2F5 are also found in the DREAM complex, which contains p130 and represses cell-cycle genes during quiescence (Litovchick et al. 2007). In contrast, several studies of E2F4 function during development suggest that E2F4 may stimulate proliferation in certain contexts, acting through association with other transcription factors (Hsu and Sage, 2016; Lee et al. 2011; Ma et al. 2014; Wang et al; 2000). Better characterization of how E2F4 and E2F5 associate with pocket proteins and other factors is needed to understand their different functions and how they are regulated.

While all three pocket proteins similarly inhibit the cell cycle and proliferation, genetic observations suggest important distinct functions. For example, only Rb deletion is embryonic lethal in the mouse (Cobrinik et al. 1996; Jacks et al. 1992; Lee et al. 1996). Rb is a more potent tumor suppressor in mouse cancer models (Cobrinik et al. 1996; Jacks et

al. 1992; Lee et al. 1996; Mulligan and Jacks, 1998), and mutations are more commonly observed in human cancers (Burkhart and Sage, 2008; Classon and Harlow, 2002), although mutations in p107 and p130 are also found. One proposed explanation for these observations is that Rb forms unique complexes with the activating E2Fs (E2F1-E2F3), although other pocket protein-specific binding interactions may confer distinct functions (Binne et al. 2007; Hirschi et al. 2010; Jayadeva et al. 2010).

The five canonical E2Fs each contain a DNA binding domain (DBD), a transactivation domain (TAD), and a coiled-coil and marked-box domain (CM) (Fig. 14a). The DBDs are homologous and bind similar DNA sequences as heterodimers with one of three DP proteins (Zheng et al. 1999). E2F7 and E2F8 contain two DBDs each, do not dimerize with DP, and recognize promoter sites that are similar to other E2Fs (Margunova et al. 2015). The CM domain of E2F also heterodimerizes with a similar domain in DP (21), and the CM heterodimer binds other transcription factors as a proposed mechanism for specific E2F family members to activate distinct genes (Giangrande et al. 2004; Hallstrom et al. 2003). The Rb family pocket domains bind the E2F transactivation domain and bind other cellular and viral proteins using a distinct surface called the LxCxE-cleft (Binne et al. 2007; Morris and Dyson, 2001; Guiley et al. 2015; Lee et al. 2002; Lee et al. 1998; Xiao et al. 2003) (Fig. 14a). Each pocket protein also contains a C-terminal domain (CTD) that is required for growth suppression and E2F inhibition and has a role in protein stability (Hiebert 1993; Qin et al. 1992; Sengupta et al. 2015; Zhu et al. 1995). A crystal structure demonstrates that the Rb CTD (RbC) binds the E2F1-DP1 CM domains (Rubin et al. 2005), but several studies suggest that this particular association may be specific to Rb and E2F1 (Cecchini and Dick, 2011; Cecchini et al. 2014; Dick and Dyson, 2003).

To better understand how the Rb proteins regulate E2F function, we have characterized the association of pocket protein CTDs with the E2F CM domain. We determined crystal structures of the E2F4-DP1 CM domain (E2F4-DP1^{CM}) and E2F5-DP1^{CM}

in complex with the p107 CTD (p107C). The structure of the ternary complex clarifies the generality of this domain association among all family members and reveals molecular details that explain the respective preferences of activating E2Fs for Rb and repressive E2Fs for p107 and p130 (p107/p130). We conclude that Rb evolved sequences that make it uniquely suited to bind and regulate the activating E2Fs. Our combination of structural and biochemical data with phylogenetic analyses provides novel insights into the co-evolution of a protein-protein interaction critical for control of cell proliferation.

3.2 Results

3.2.1 Distinct E2F-binding properties of RbC and p107C

We first tested whether the binding preferences of Rb pocket proteins for different E2F family members result from different affinities between the pocket protein CTDs and E2F CM domains. We expressed and purified E2F4-DP1^{CM} and E2F5-DP1^{CM} complexes that were identified based on homology to E2F1-DP1^{CM} (Fig. 14b). We used a co-precipitation assay to identify a minimal fragment of p107C (residues 994-1031, called p107C$^{994-1031}$) that is suitable for structural studies and is sufficient to bind the CM domains (Fig. 13). We measured the affinity of this minimal fragment for E2F4-DP1^{CM} using isothermal titration calorimetry (Fig. 14c). We found that p107C$^{994-1031}$ binds with similar affinity as that previously reported for full-length p107C (residues 949-1068) (Rubin et al. 2005). Both p107C$^{994-1031}$ and p107C$^{949-1068}$ bind E2F1-DP1^{CM} with lower affinity than they bind E2F4-DP1^{CM}.

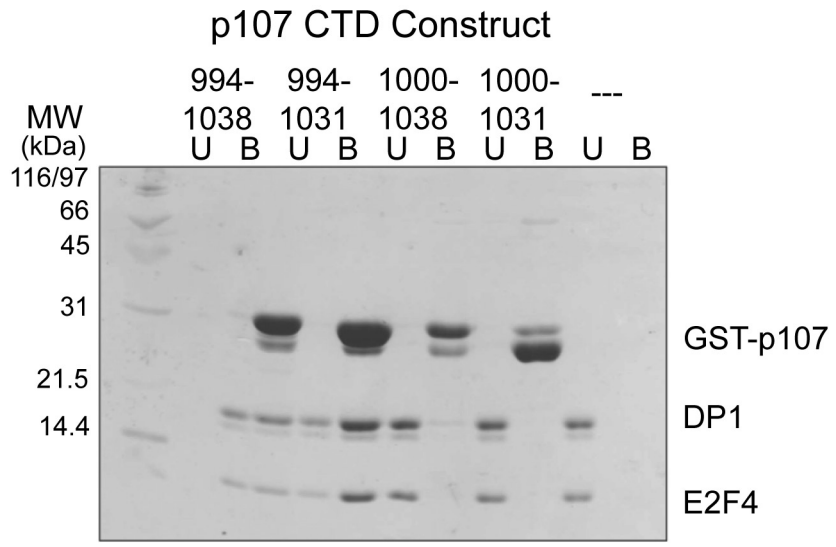


Figure 13: Co-precipitation assay identifying a minimal p107 CTD fragment that binds E2F4-DP1^{CM}.
100 µg of the indicated GST-p107 fragment and 100 µg of purified E2F4-DP1^{CM} were incubated on ice for 1 hour in a reaction volume of 200 µL containing 100 mM NaCl, 25 mM Tris, and 5 mM DTT (pH 8.0). Proteins were affinity precipitated with Glutathione Sepharose resin, washed, and eluted in the binding buffer plus 10 mM glutathione. For each binding reaction, the eluate containing bound proteins (B) was loaded onto an SDS-PAGE gel along with a sample of the unbound (U) reaction.

Comparing these measurements with previous measurements of RbC reveals several differences between how RbC and p107C bind to E2F-DP^{CM} domains (Fig. 14C) (Rubin et al. 2005). First, the affinity of the full RbC sequence (residues 771-928) is four-fold tighter than full p107C for E2F4-DP1 and fifty-fold tighter for E2F1-DP1. Second, while RbC makes a bipartite association with contributions from residues 786-801 (RbC^{nter}) and residues 829-864 (RbC^{core}) (Fig. 14B) (Rubin et al. 2005), all the interactions made by p107C are contained within p107C$^{994-1031}$. Third, while RbC^{core} has similar affinity for E2F1 and E2F4 (Rubin et al. 2005), p107^{core} has higher affinity for E2F4 than E2F1. We next determined the structural basis for these affinity differences.

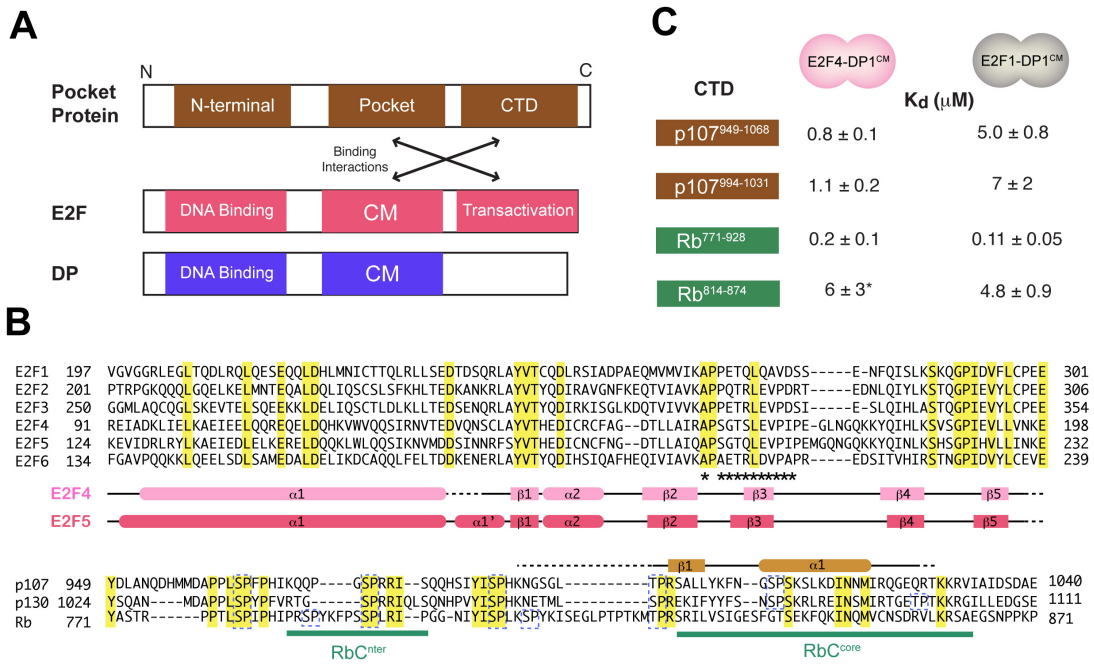


Figure 14: Interactions between pocket protein C-terminal domains (CTDs) and the E2F coiled-coil and marked-box (CM) domains. (A) Domain architecture of pocket proteins, E2F, and DP. Two associations stabilize the pocket protein complex with E2F-DP. The E2F transactivation domain binds the pocket domain and the E2F-DP CM domain binds the CTD. (B) Sequence alignment of the human E2F CM domains and pocket protein CTDs. The residues that are strictly conserved among paralogs are highlighted yellow. The secondary structure elements present in our crystal structures of E2F4-DP1^{CM} and p107C-E2F5-DP1^{CM} are indicated along with residues in E2F5 (*) that contact the p107 CTD (p107C). RbC^{nter} and RbC^{core} are the two sequences in the Rb CTD (RbC) that were previously identified as critical for E2F1-DP1^{CM} binding. CDK phosphorylation sites are indicated with blue dashed lines. (C) Isothermal titration calorimetry (ITC) measurements of RbC and p107C constructs for different E2F-DP CM domains. The values for p107$^{994-1031}$ binding were determined here, whereas the other measurements in the table were previously reported (21). The asterisk indicates that the previous measurement was made with E2F4-DP2^{CM}.

3.2.2 Conservation of E2F-DP^{CM} structures

We grew crystals of E2F4-DP1^{CM} alone and E2F5-DP1^{CM} (Fig. 15) bound with p107C$^{994-1031}$ and determined structures with resolution of 2.3 Å and 2.9 Å respectively (Table 1 and Fig. 16). In both structures, the E2F and DP polypeptides have similar secondary structure topology, and the chains entwine to create an extensive interface (Fig. 14B and 16A). The structure consists mainly of a heterodimeric coiled-coil subdomain and a heterodimeric β-sandwich subdomain that are bridged by two small helices and two small

strands. The coiled coil contains two long parallel helices, one from each protein. The hydrophobic and salt bridge interactions forming the coiled-coil interface resemble those seen in a canonical leucine zipper, though not strictly throughout the length of the helices. The sandwich domain has an immunoglobulin fold, consisting of two four-stranded β-sheets, with a hydrophobic core generated from both E2F and DP sidechains. The hydrophobic residues that hold the coiled-coil and sandwich domains together also originate from both proteins. The intertwined structure and dependence on DP to complete the hydrophobic core explain why heterodimerization is necessary for E2F stability, DNA binding, and transcriptional activity (Trimarchi and Lees, 2002; Rubin et al. 2005; Zheng et al. 1999).

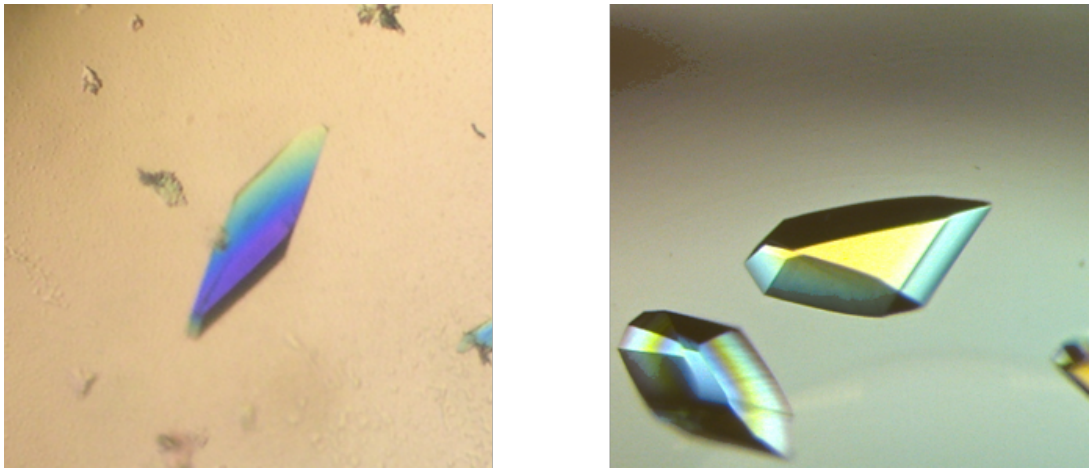


Figure 15: Crystals of E2F4-DP1 and E2F5-DP1-p107C
Crystals of the E2F4-DP1 dimer (left) readily crystallized, forming diffraction quality crystals that diffracted to 2.3A in the JCSG commercial screen. Obtaining suitable crystals of the E2F5-DP1-p107C ternary complex (right) was incredibly difficult. After setting over 10,000 protein drops, we were finally able to grow crystals that diffracted to 2.9A.

We considered sequence and structural conservation among E2F paralogs and identified regions in the coiled-coil and marked-box domains that may be involved in shared or distinct functions. Several aspects of the E2F4-DP1^{CM} and E2F5-DP1^{CM} structures are similar to the previously determined structure of E2F1-DP1^{CM}, including the topology of the secondary structure elements (Trimarchi and Lees, 2002; Rubin et al. 2005; Zheng et al.

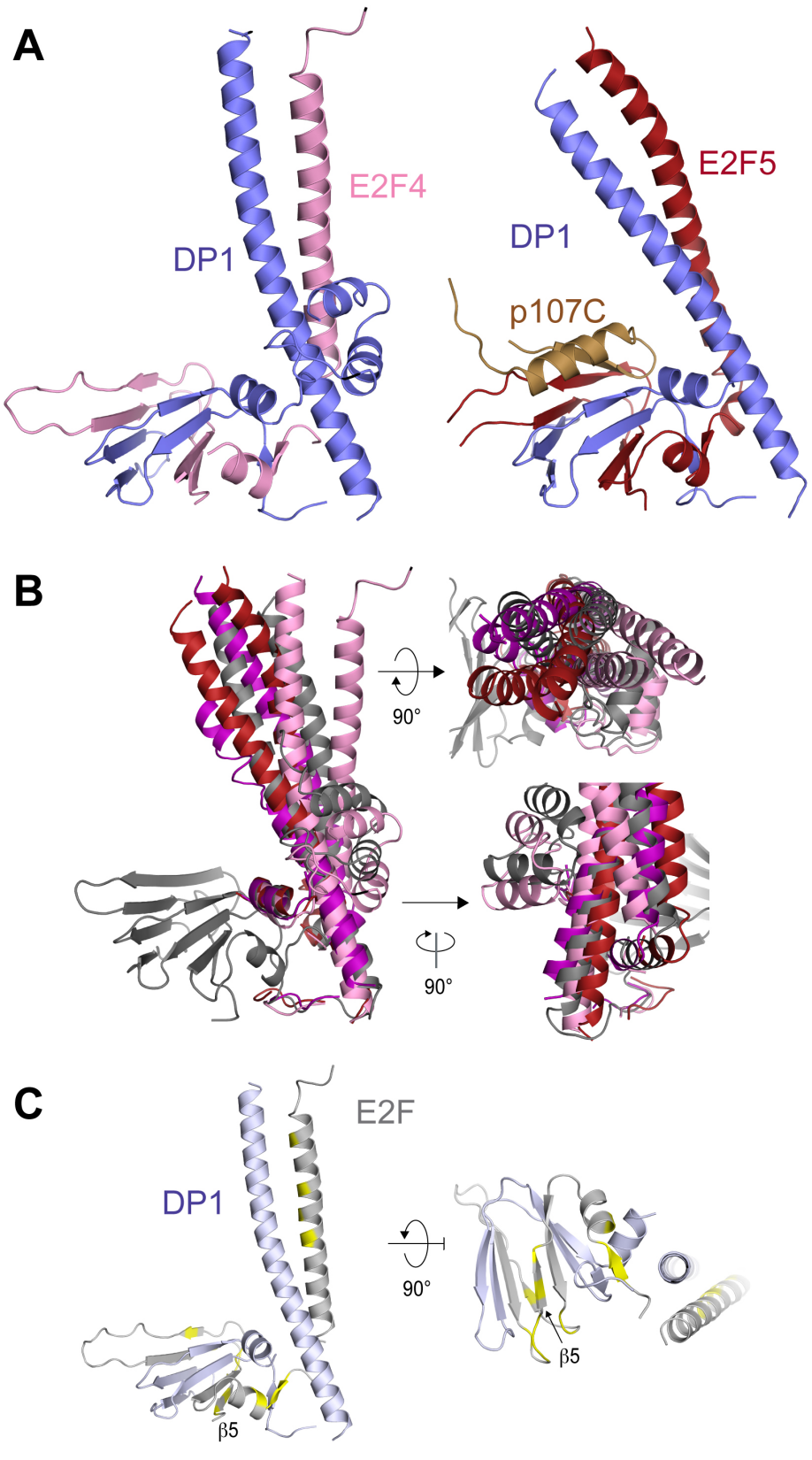
1999). Pairwise alignment of Cα atoms from the different E2F-DP β-sandwich domains show root mean square deviations of only 0.8-1.4 Å (Fig. 17). One notable variation among the structures is the orientation of the coiled-coil domain relative to the β-sandwich domain (Fig. 16B). Alignment of the overall structures suggests that the coiled-coil domain can pivot about a fixed contact point made with the α2 helix in DP1. Considering that the E2F-DP DNA binding domains are N-terminal to the start of the coiled-coil domain, we suggest that this flexibility may be important for bridging the interaction with DNA and interactions with other transcription factors that potentially bind the marked-box domain or C-terminal regions in E2F (Giangrande et al. 2004, Hallstrom and Nevins, 2003).

Table 2: Data Collection and Refinement statistics (Values in parentheses are for higher resolution shell)
Molecular replacement was used to generate an initial model for the E2F4-DP1 data set. A single anomalous dispersion (SAD) (middle column) data set was collected at the selenium peak to generate an initial electron density map. The initial model from SAD phasing was later refined against a higher resolution native data set (right column). Data collection was performed at ALS beamline 5.0.1 for E2F4-DP1 and APS beamline 23-IDB for E2F5-DP1-p107C.

	E2F4-DP1	E2F5-DP1-P107 (SAD)	E2F5-DP1-P107
Data collection	APS	APS	APS
Space group	I2	$P2_1$	$P2_1$
Unit cell dimensions (a,b,c) (Å), (⟨, ®, © (º)	73.57, 37.54, 109.9 90, 103.4, 90	62.47, 56.17, 99.30 90, 93.7, 90	60.98, 57.34, 99.20 90, 96.4, 90
Resolution Range (Å) (highest shell)	53.96-2.25 (2.32-2.25)	99.09-4.09(4.58-4.09)	60.57-2.90 (3.08-2.90)
Wavelength (Å)	1.02	1.02	1.02
Total observations	46394(2904)	146148(41730)	71791(11049)
Unique reflections (highest shell)	13397(1302)	5606(1578)	15301(1529)
Completeness (%) (highest shell)	95.2(93.6)	99.9(99.7)	99.0(99.2)
R_{merge} (highest shell)	5.6(30.8)	11.9(21.7)	10.4(41.3)
<I/ ∫> (highest shell)	20.7(2.9)	22.6(14.9)	10.6(5.5)
CC1/2	0.997(0.679)	0.999(0.998	0.992(0.921)
Redundancy (highest shell)	3.5 (2.4)	26.1(26.4)	4.7(4.5)
Refinement			
R_{work} %/ R_{free} %	19.68/23.77		21.67/26.38
Number of atoms	1976		3539
Protein	1928		3539
Ligand/ion	12		
Water	36		
B-factor (Wilson)	44.60		60.88
Unmodeled residues	5		148
RMSD Bond length (Å)	0.010		.003
RMSD Bond angle	1.21		0.92

Sequence comparison of the human E2F1-E2F6 proteins reveals that twenty residues are identical within the CM domain (Fig. 14B). The identical amino acids map primarily to three regions of the structure (Fig. 16C). There are conserved residues that form the coiled-coil interface and conserved residues that contribute to the structural core that bridges the β-sandwich and coiled-coil domains. It is clear that these amino acids contribute to the overall stability of the E2F-DP heterodimer. The third highly conserved structural element contains residues in the last E2F strand (β5) and the preceding loop. Sidechains from these residues form a surface on the edge of the sandwich opposite the edge that binds pocket proteins (Fig. 16A and 16C). The conservation here suggests that the exposed hydrophobic cleft along this sandwich edge is a potential protein interaction surface common to all E2F family members. The most notable region of the structure that is distinct among paralogs (but conserved among orthologs of each family member) is the end of E2F β3 and the loop between β3 and β4. This region contains a number of residues that contact p107C in the ternary complex, and we explore below the idea that sequence divergence in this region accounts for differences in specificity for different pocket. We also propose that other potential E2F-DP interacting proteins, particularly those that have specificity for particular E2F family members (Giangrande et al. 2004, Hallstrom and Nevins, 2003), bind at this site.

Figure 16: Crystal structures of the E2F4-DP1 and E2F5-DP1 CM domains. (A) Overall structures show similar topologies with the E2F and DP chains forming an extensive interface. Both the coiled-coil and b-sandwich subdomains contain secondary structure elements from both proteins. (B) Overlay of the CM domain structures determined for E2F1-DP1 (grey, PDB: 2AZE), E2F4-DP1 (pink), and E2F5-DP1 (red and purple correspond to the two different molecules in the asymmetric unit). Structures were overlaid by alignment of β-sandwich domain Cα atoms (Fig. 17). For clarity, only the β-sandwich domain of E2F1-DP1 is shown. (C) Structure of the E2F4-DP1 domain shown with residues that are conserved among E2F paralogs highlighted in yellow.


A
E2F4
DP1
E2F5
DP1
p107C
B
90°
90°
C
E2F
DP1
90°
β5
β5

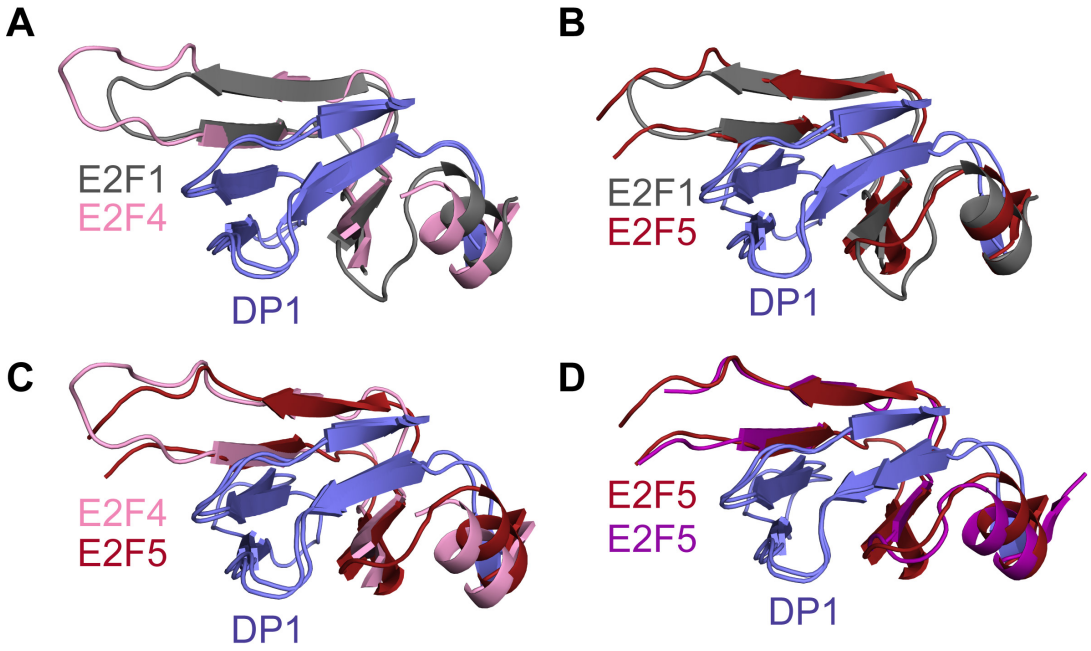


Figure 17: Comparison of CM β-sandwich domains reveals similar structures. A) Alignment of E2F4-DP1^{CM} and E2F1-DP1^{CM} structures. The root mean square deviation (rmsd) of C**a** position is 0.77 Å. B) Alignment of E2F5-DP1^{CM} and E2F1-DP1^{CM} structures. Rmsd is 1.24 Å. C) Alignment of E2F5-DP1^{CM} and E2F4-DP1^{CM} structures. Rmsd is 1.35 Å. D) Alignment of two different E2F5-DP1^{CM} heterodimers in the asymmetric unit of the E2F5-DP1-p107c crystals. Rmsd is 1.14 Å.

3.2.3 Structural investigation of the p107C-E2F5-DP1 interface

p107C binds the E2F5-DP1 marked-box domain using a strand-loop-helix motif (Fig. 18). The strand adds on in an anti-parallel direction to the β-sheet in the immunoglobulin sandwich domain that is distal to the coiled coil (Fig. 18A). Residues L1002, Y1004, and F1006 in p107C make a series of backbone hydrogen bonds with G195, T196, L198, and V200 in E2F5. There are additional interactions between sidechains from the p107C strand. For example, Y1004 and F1006 pack into the hydrophobic core of the β-sandwich domain, contacting sidechains in both E2F5 and DP1.

The amphipathic p107C helix covers the core of the β-sandwich domain. The hydrophobic sidechains of L1014, I1017, M1020, and I1021 from p107C pack into the core. They make van der Waals contacts with L198, V200, I202, and P203 from E2F5 and I262, T290, F291, I293, and D295 from DP1. N1018 in the p107C helix makes hydrogen bonds with backbone atoms in the DP1 β-strand at the edge of the β-sandwich.

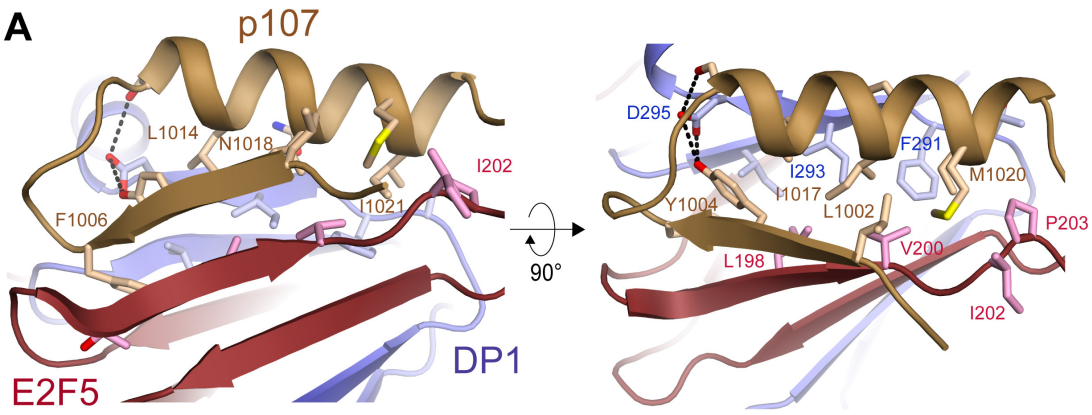


Figure 18: Structure of the p107C-E2F5-DP1 interface. (A) p107C binds the E2F5-DP1 β-sandwich domain using a strand-loop-helix motif and forms a hydrophobic interface with residues from both E2F and DP.

We used the Cancer Genome Atlas (cancergenome.nih.gov) to identify cancer-associated mutations in p107 and p130 that are localized to the CTD, and we mapped these mutations onto the crystal structure (Fig. 19B). Several missense mutations have been found at I1021 in p107 (I1092 in p130), which is found on the p107C helix and is buried into the E2F5-DP1^{CM} sandwich core (Fig. 18A and 19B).

We found that an I1021A mutation weakens affinity of p107C for E2F4-DP1^{CM} thirty-fold, while an I1021M mutation found in colorectal cancer has a relatively modest two-fold effect (Fig. 19C). Most of the other mutations map to the surface of the p107C strand

and helix that does not form the interface with E2F5-DP1^{CM}. We did find that two mutations found in uterine cancer, K1012E in p107 and R1093C in p130 (R1022C in p107), and one mutation found in both colorectal and prostate cancer, R1093H in p130 (R1022H in p107), have two to three-fold effects on the complex affinity. We conclude that these mutations may slightly impair the ability of p107/p130 to bind E2F. However, an interesting alternative possibility, given the localization of these residues to the exposed face of the p107C helix, is that these mutations interfere with other protein interactions. Notably, this region of p107/p130 has been implicated in regulation of protein stability, and mutation of K774 in Drosophila (K1012 in human p107) leads to a developmental defect from loss of regulation of protein levels (Sengupta et al. 2015; Zhang et al. 2014). In addition to E2F-binding, the C-terminal helix may mediate protein interactions that control p107/p130 degradation.

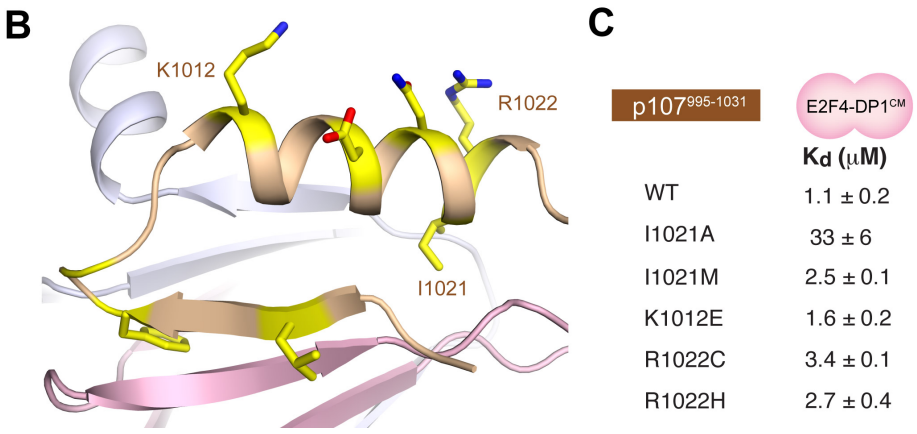


Figure 19: Location of residues in p107 associated with various forms of cancer. (B) Residues in p107 or p130 that are found mutated in human tumor samples (Cancer Genome Atlas) are highlighted yellow. (C) ITC measurements of E2F4-DP1^{CM} binding to p107C constructs containing example cancer mutations.

3.2.4 Specificity in Rb and p107 interactions with E2F-DP^{CM}

We compared our structure of the p107C-E2F5-DP1 complex with the structure of the RbC-E2F1-DP1 complex to understand the binding preferences revealed by our affinity measurements. First we addressed the question of why E2F4-DP1 has higher affinity for p107C^{core} than RbC^{core} (Fig. 14C). In general, the mode of RbC binding to the E2F1-DP1 marked-box domain is similar to the observed for p107C (Fig. 20A). RbC uses a similar strand-loop-helix motif, and several residues from the RbC helix (I848, M851, V852) that insert into the core are identical or similar in the p107C helix (I1017, M1020, I1021). However, the contacts between hydrophobic residues near the N-terminus of the helix and C-terminus of the strand are distinct with V833, I835, T841, and F845 in Rb replaced with Y1004, F1006, and L1014 in p107 (Fig. 20A). We suggest that tighter packing of this interface stabilizes p107C binding relative to RbC.

A second observed binding specificity is the higher affinity of p107C for E2F4 compared to E2F1. To understand this preference, we identified residues at the CM interface with p107C that are different between E2F5 and E2F1. The most striking difference is towards the C-terminus of strand β3, in which the sequence motif in E2F4 and E2F5 is VPIP, while the sequence in E2F1 is AVDS (Fig. 14B and 20B). The bulkier V200 in E2F5 (V167 in E2F4) can interact better with I1017 and M1020 in p107 compared to the smaller A275 in E2F1 (Fig. 20B). In addition, P201 in E2F5 (P168 in E2F4) causes the strand to bulge such that P203 (P170 in E2F4) is in position to contact I1021. D277 in E2F1 is at the same position as P203 in E2F5 and likely makes weaker interactions.

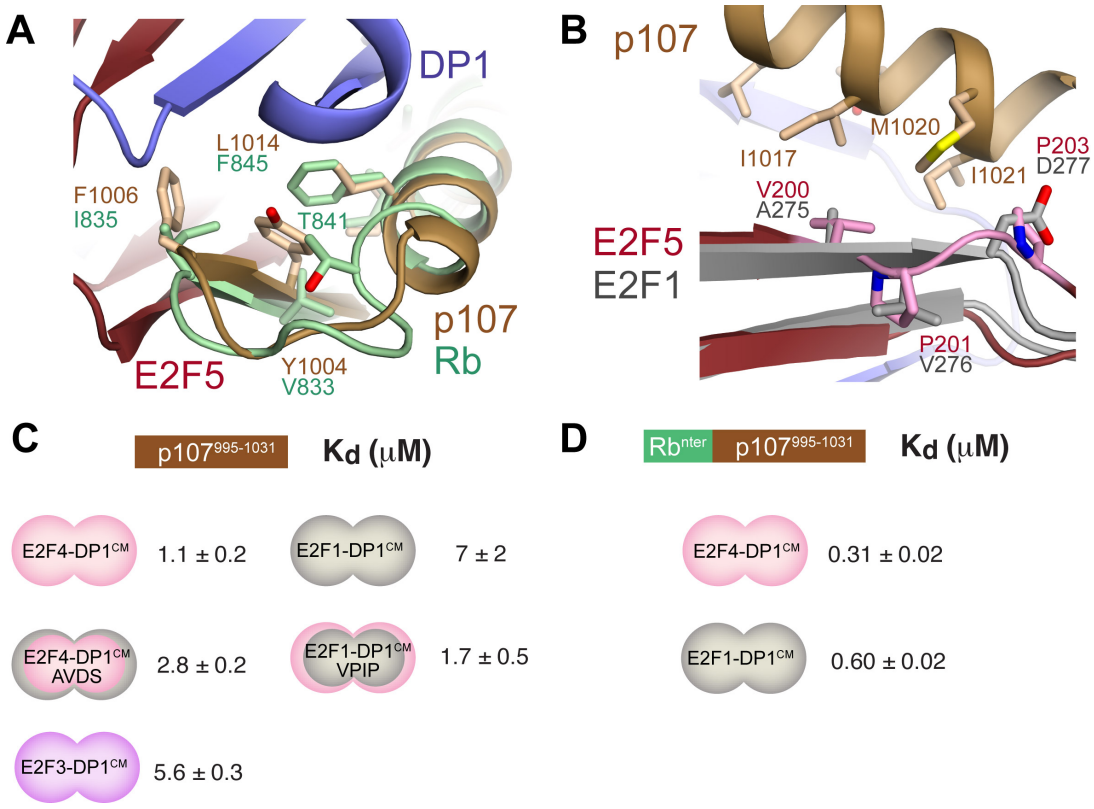


Figure 20: Comparison of p107C-E2F5 and RbC-E2F1 binding interfaces. Structures of p107C-E2F5-DP1 and RbC-E2F1-DP1 (PDB: 2AZE) were overlaid by alignment of β-sandwich domain Cα atoms (Fig. 17). (A) Comparison of how RbC and p107C would bind E2F5. (B) Comparison of how p107C would bind E2F1 and E2F5. (C) ITC binding affinity measurements for the p107C construct present in the crystal structure and the indicated E2F-DP^{CM} domains. The E2F4-DP1^{CM} AVDS mutant has the E2F4 VPIP sequence (residues 167-170) mutated to AVDS, while the E2F1-DP1^{CM} mutant has the E2F1 AVDS sequence (residues 275-278) mutated to VPIP. (D) Affinity measurements of an RbC-p107C hybrid protein containing residues 771-794 of Rb (RbC^{nter}) fused to residues 975-1031 of p107.

We used the calorimetry assay to test the importance of the E2F4/E2F5-conserved VPIP motif for p107C$^{994-1031}$ affinity (Fig. 20C). We found that changing the VPIP sequence in E2F4 to the AVDS sequence in E2F1 yields a mutant E2F4-DP1^{CM} heterodimer that binds p107C nearly three-fold weaker than wild-type. Conversely, mutation of the E2F1 AVDS sequence to VPIP increases the affinity of E2F1-DP1^{CM} for p107C$^{994-1031}$ four-fold.

We also found that p107C$^{994\text{-}1031}$ binds E2F3-DP1^{CM} more weakly than E2F4-DP1^{CM} and more similar to E2F1-DP1^{CM} (Fig. 20C). This observation is interesting because E2F3 has VPDS as its sequence in this critical region (Fig. 14B), and although E2F3 has the first proline to induce the bulge in the strand, the S331 at the position of the second proline in E2F4/E2F5 is suboptimal for contacting I1021 (like D277 in E2F1). Together these data demonstrate that the sequence in β3 strand is a critical determinant for p107 binding repressive E2Fs with higher affinity than activating E2Fs.

Unlike p107, RbC binds E2F1-DP1^{CM} and E2F4-DP1^{CM} with similar affinity (Fig. 14C) (Rubin et al. 2005). Rb contains a valine (V852) at the analogous position as I1021 in in the p107C helix. Alignment of the p107C and RbC ternary complex structures suggests the smaller Rb sidechain would not contact P203 in E2F5 (P170 in E2F4), and we observe loss of affinity due to substitution of the I1021 sidechain with a smaller hydrophobic group (Fig. 19C). The structural comparison suggests the explanation that Rb is less sensitive to the differences in E2F1 and E2F4/E2F5 at this binding interface because of the weaker interactions between V852 and the E2F β3 strand.

Additional interactions involving the RbC^{nter} sequence enhance RbC binding to both E2F1-DP1^{CM} and E2F4-DP1^{CM} (Rubin et al. 2005). In contrast, our measurements here suggest that the sequence in p107 N-terminal to the core binding region in the crystal structure does not make these stabilizing interactions (Fig. 14C). We found that replacing the p107C N-terminal sequence (residues 949-974) with the RbC^{nter} sequence (residues 771-822) results in a hybrid p107C construct that binds E2F1-DP1^{CM} and E2F4-DP1^{CM} with increased affinity. This observation demonstrates that the RbC^{nter} sequence enables RbC to bind both activator and repressive E2F proteins with higher affinity than p107C.

While Rb is found in cells in complexes with both activating and repressive E2Fs, it has been proposed that the RbC association is specific to E2F1 (Lees et al. 1993; Cecchini and Dick, 2011; Cecchini et al. 2014; Dick and Dyson, 2003). One observation supporting

this specificity is that a mutation in the pocket domain (R467E/K548E) at the transactivation-binding site inhibits Rb complex formation in cell extracts with E2F3 but not E2F1 (Cecchini and Dick, 2011; Cecchini et al. 2014). Our quantitative data however suggest that at least in the RbC^{core} region, interactions are similar with all E2Fs (Fig. 14B) (Rubin et al. 2005). We considered that the Rb-E2F1 specificity observed in cells may be due to differences in other interactions outside the C-terminus. For example, the E2F1 and E2F2 TADs bind the Rb pocket domain with significantly higher affinity than other E2Fs (Liban et al. 2016). We measured the affinity of the R467E/K548E Rb mutant for both the E2F1 and E2F3 TADs by ITC. We find that while the mutations abolish E2F3 binding completely, we still observe some weak affinity for E2F1 (Fig. 21). We suggest that such differences in transactivation domain binding and not E2F-DP^{CM} binding explain the preference of Rb for different E2Fs in cells when probed under various conditions. In contrast, differences in affinity for both the transactivation domain (Liban et al. 2016) and the CM domain (Fig. 14C) contribute to the preferences of p107 for different E2Fs.

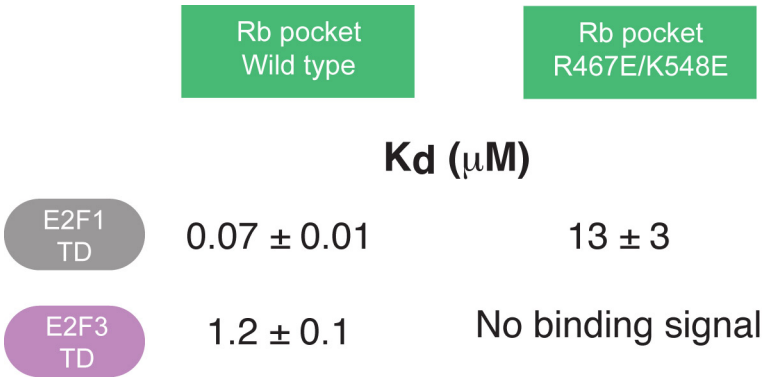


Figure 21: Mutations to the E2F transactivation domain (TD) binding site in Rb do not completely abolish E2F1 TD binding. ITC measurements of Rb pocket domain (residues 380–787) for E2F1 TD (residues 409-426) and E2F3 (residues 432-449). Wild-type measurements were previously reported (58).

3.2.5 Rb sequence elements that confer E2F binding affinity co-evolved with E2F1 and E2F2

Our data support the conclusion that Rb is unique among pocket proteins in its ability to bind E2F1 with high affinity. To test the hypothesis that this property of Rb co-evolved with E2F1, we examined the evolutionary history of pocket proteins and E2Fs along the metazoan lineage from a subset of 52 genomes. Our phylogenetic analysis reveals a number of gene duplication events that resulted in the expansion of the pocket protein and E2F families (Fig. 22, Fig. S4, Fig. S5, and Table S2)(Refer to publication for additional supplemental material). In agreement with previous work (Cao et al. 2010), we find that the divergence of Rb and RbL (the p107/p130 ancestor) from their common ancestor (aRb) precedes the emergence of Eumetazoa, possibly after the divergence of Choanoflagellata (*S. rosetta*) and before the emergence of the Placozoa lineage (*T. adhaerens*). This emergence of Rb appears to coincide with the emergence of two E2F proteins, one that is the ancestor of E2F4 and E2F5 (E2F45) and one that is the ancestor of E2F1, E2F2, E2F3 and E2F6 (E2F1236). Additional gene duplication events occurred at the base of the Craniata lineage after the divergence of the jawless fish lineage (*P. marinus*, “lamprey”), when RbL2 (p130) and RbL1 (p107) emerged from RbL, E2F4 and E2F5 emerged from E2F45, and E2F1, E2F2, E2F3, and E2F6 emerged from E2F1236.

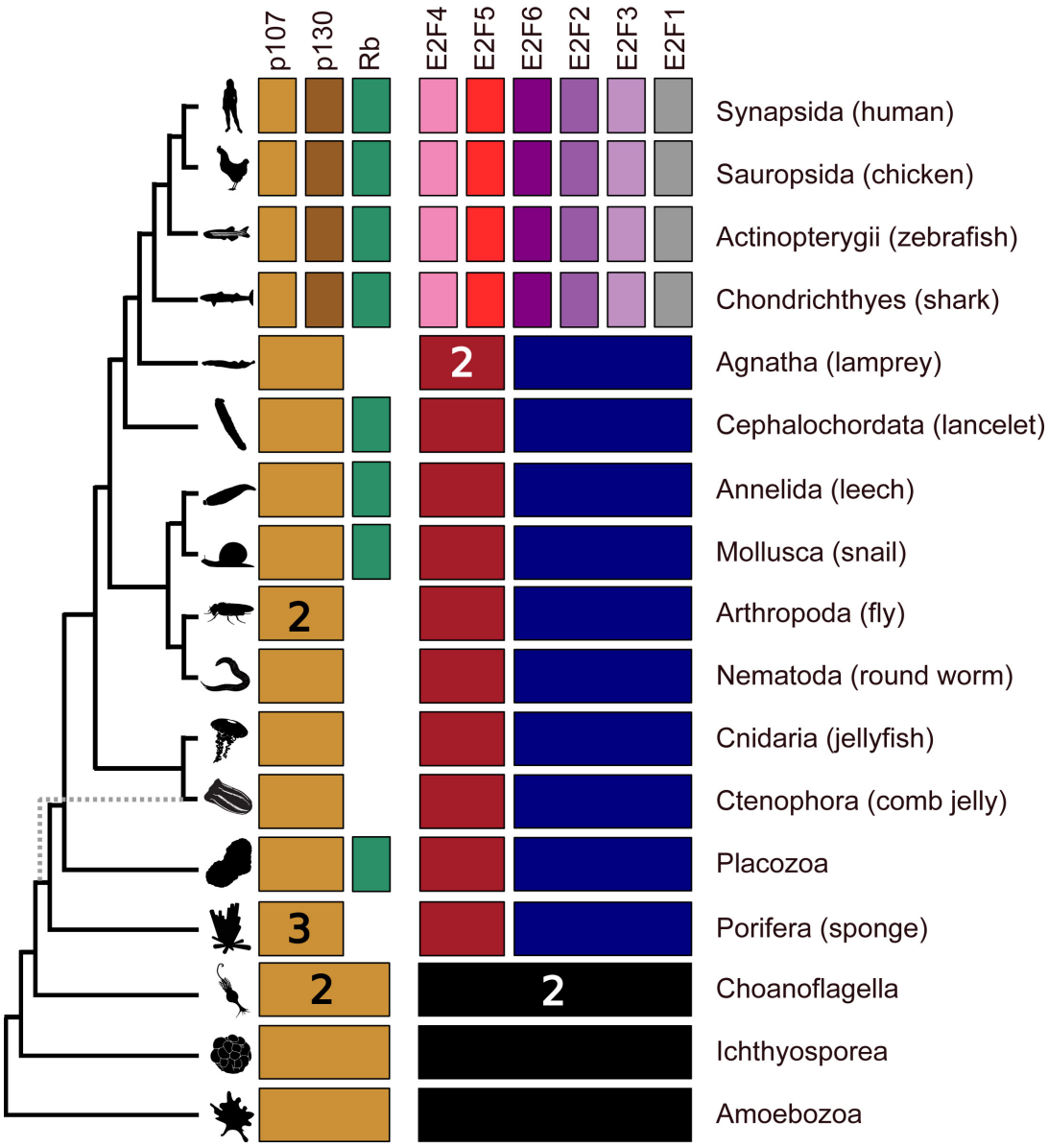


Figure 22: Phylogenetic distribution of pocket protein and E2F sub-families. Detected pocket proteins and E2Fs from each genome were classified into sub-families based on phylogenetic analyses (see Methods). The resulting distribution of homologs was condensed and ordered depicts the current consensus on the phylogenetic relationships between major animal lineages (Telford et al. 2015), including alternative locations of the Ctenophora (dashed grey line) and with Amoebozoa as an outgroup. Common name of exemplary species in parenthesis. For a list of taxa depicted here refer to Table S2. Taxa silhouttes obtained from http://www.phylopic.org.

We focused on the evolution of structures that play a role in determining pocket protein-E2F binding specificity. There is considerable conservation in the pocket protein CTD helix (in human p107 residues 1011-1023), which plays a prominent role in stabilizing the interface with the E2F CM domain (Fig. 18, Fig. 19 and Fig. 20). For example, the helix residues along the interface are hydrophobic in all the sequences dating back to the early metazoa, and several positions (I1017, N1018, and M1020 in p107) are nearly strictly conserved (Fig. 23 and Fig. 24). Two positions that give rise to differences in how Rb binds the E2Fs--L1014 (F845 in human Rb) and I1021 (V852)--emerge in Rb in sharks (Fig. 23 and Fig. 24). This emergence is coincident with the expansion of the protein families at the base of the Craniata lineage (Fig. 22).

We also examined the sequence corresponding to the end of E2F β3 strand (V200-P203 in human E2F5), which in our affinity measurements is a key source of binding preferences between the E2F and pocket proteins (Fig. 20). The ancestor E2F at the base of the phylogenetic tree and the two E2Fs in early metazoa (E2F45 and E2F1236) all contain the VP*P motif at these positions (Fig. 23 and Fig. 24). This motif is kept in the E2F4 and E2F5 lineages through humans. As seen in our E2F5 structure, the VP*P sequence places the hydrophobic V200 and P203 sidechains at the interface with p107C, and these residues are critical for high affinity binding of p107C to human E2F4-DP1^{CM} (Fig. 20). The VP*P motif is eventually lost in all three activator E2F lineages after the gene expansion at the base of Craniata. It is lost first in E2F1 and E2F2, which emerge in sharks and lack the first and second proline respectively (Fig. 23). The E2F3 lineage maintains the VP*P motif until more recently diverged animals, in which it lacks the second proline. Notably, we observe that both human E2F1 and E2F3 have weaker affinity for p107C than E2F4 (Fig. 20C).

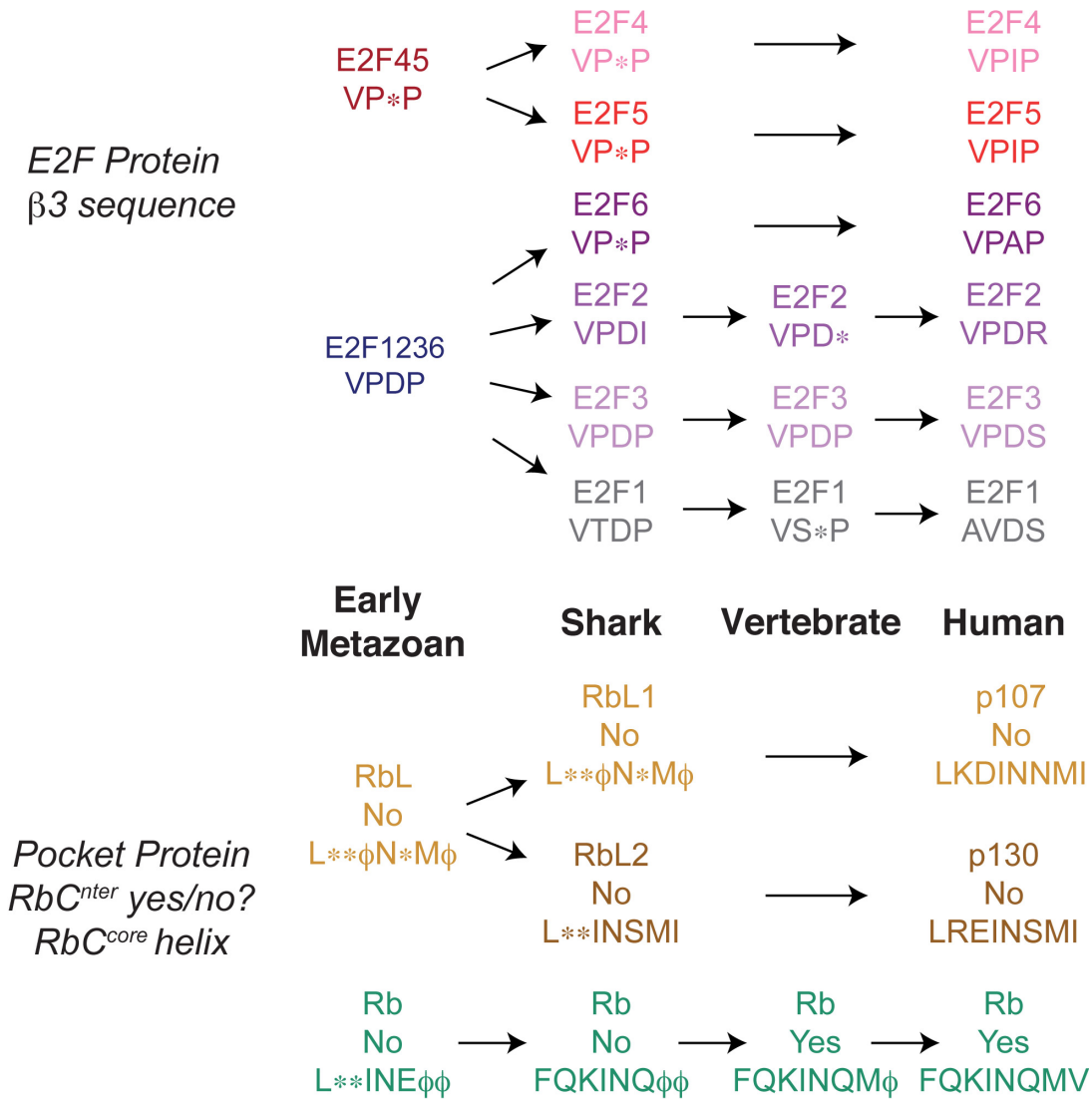


Figure 23: Evolutionary model of sequences involved in E2F^{CM}-pocket protein association. The VP*P motif in the β3 strand of the E2F marked–box domain, which is necessary for affinity binding to p107C, is lost in activating E2Fs. Hydrophobic residues in RbC^{core} that are critical for E2F binding are conserved in pocket proteins, with subtle changes emerging in Rb in sharks. The RbC^{nter} sequence, which increases Rb affinity for E2F proteins, is first observed in Rb in vertebrates and is not present in p107, p130 and the Rb sequences from deeper nodes in our tree.

Our phylogenetic analyses support the hypothesis that the VP*P motif is a key distinguishing feature between activator and repressive E2F structure and function. Examination of the pocket protein CTD sequences indicates that the RbC^{nter} sequence emerges at the same point as E2F1 and E2F2, despite the divergence of Rb from RbL much earlier in the tree. For example, *T. adhaerens* harbors an Rb-like protein, however, *T. adharens* Rb resembles RbL around the sequence location of RbC^{nter} (Fig. 22 and Fig. 24). The Rb proteins from *C. mil* and other cartilaginous fish (*L. eri* and *S. can*) have sequences that more resembles RbC^{nter}, but they lack the S788 (human Rb numbering) CDK phosphorylation site (Fig. 24). The first complete RbC^{nter} appears in vertebrates (e.g. *D. rer*).

Our analysis suggests that the appearance of the E2F1-E2F3 lineages and loss of the VP*P-motif in the β3 strand are coincident with the appearance of the RbC^{nter} sequence. It is remarkable that the loss of affinity between RbL (p107/p130) and E2F1 and E2F2 is compensated by the gain in affinity of Rb due to the emergent RbC^{nter} sequence. We could recapitulate this hypothesized adaptation by adding RbC^{nter} to p107C (Fig. 20D). The hybrid protein has enhanced affinity for E2F1 compared to p107C. We see a similar trend in pocket residues that confer higher affinity of Rb for activator E2F transactivation domains. For example, H555 and K475 (human Rb numbering), both of which increase affinity of the Rb pocket domain for $E2F2^{TD}$ relative to the affinity of the p107 pocket domain (58), emerge in sharks around the E2F2 emergence (Fig. 25). We propose that Rb has maintained sequence variations from its RbL ancestor that contribute to its high–affinity interactions with activator E2Fs, which are necessary for transcription factor inhibition.

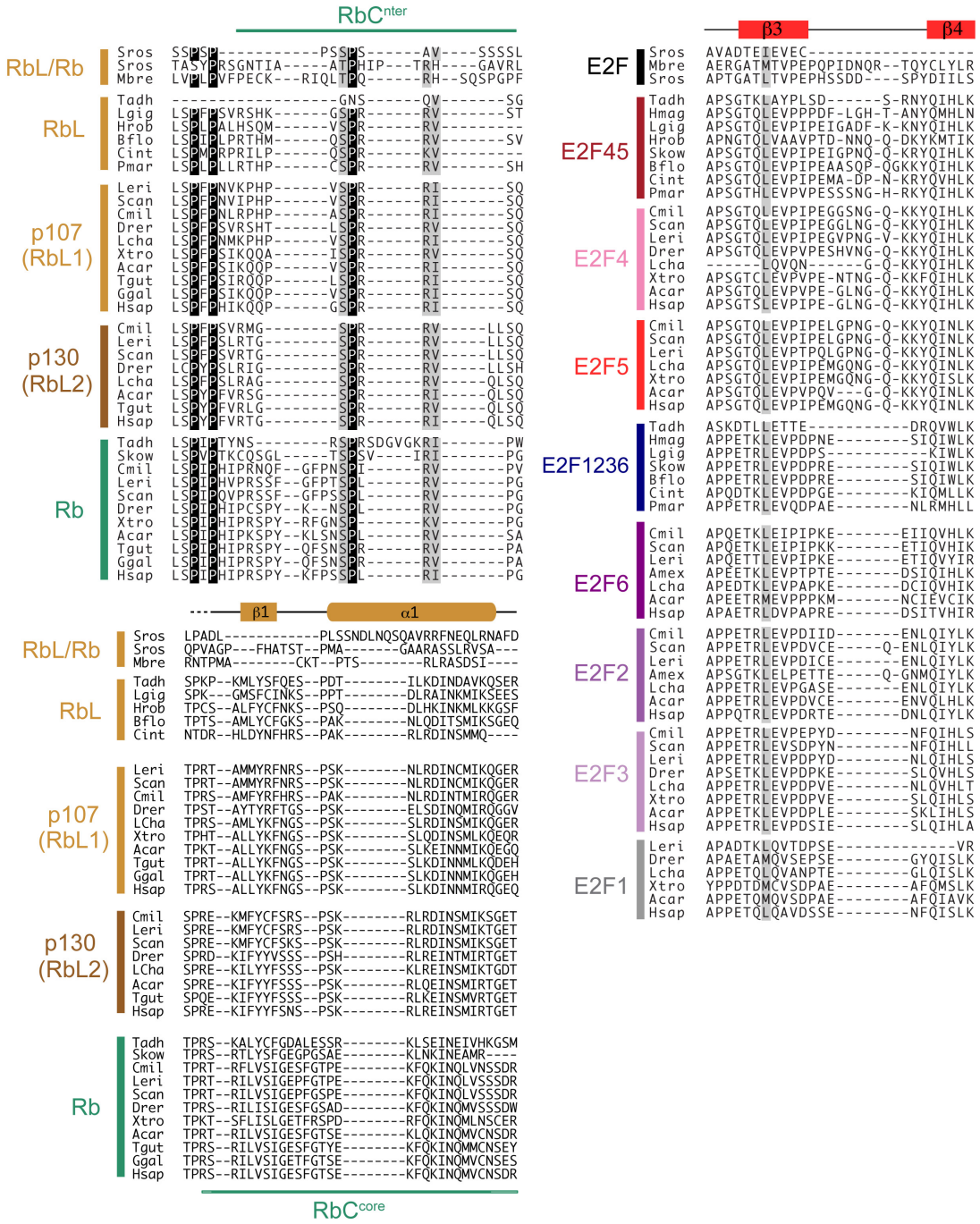


Figure 24: Alignment of pocket protein and E2F sequences. Sequences in the pocket protein C-terminus around the structured core and RbC^{nter} are shown. E2F sequences containing the β3 strand in the E2F^{CM} domain are shown. Full organism names can be found in Table S2.

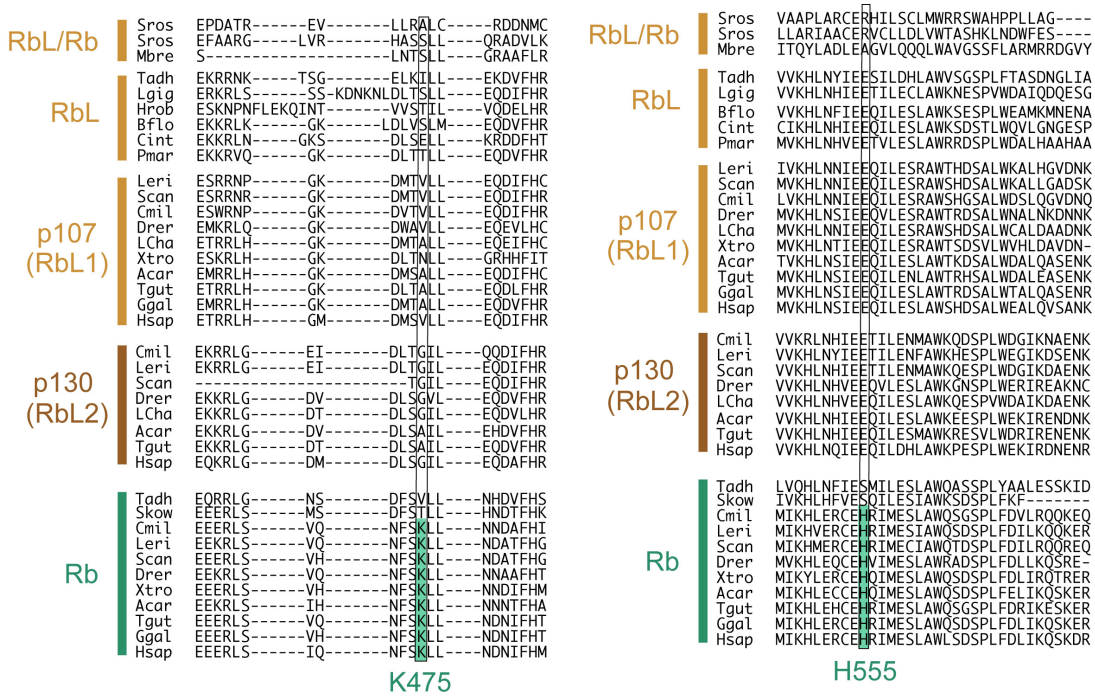


Figure 25: Sequence alignment of pocket protein sequences that show emergence of K475 and H555. It has been demonstrated that K475 and H555 (human Rb numbering) contribute to the higher affinity of Rb for activator E2F transactivation domains compared to p107 (Liban et al. 2016). Alignment of sequences in the pocket protein pocket domains demonstrate that K475 and H555 first appear in the cartilaginous fish (Cmil, Leri, Scan) but are maintained through the human sequence. Full organism names can be found in Table S2.

3.3 Discussion

Our structural and biochemical data support the conclusion that the association between the pocket protein CTD and E2F-DP^{CM} domain is a general binding mechanism shared by all pocket proteins and canonical E2F family members. Functional studies clearly point to a unique role for Rb in regulating E2F1 activity, and it had been proposed that RbC has unique affinity for E2F1^{CM} (Lees et al. 1993; Cecchini and Dick, 2011; Cecchini et al, 2014; Dick and Dyson, 2003). We have demonstrated that RbC associates similarly with different E2F CM domain, and we suggest that the unique Rb-E2F1 binding observed in cells is conferred by other specific features of the Rb-E2F1 association such as the stronger pocket-TAD association and additional affinity conferred by association of RbC^{nter} (Rubin et al. 2005; Liban et al. 2016). In contrast to RbC, we find here that p107C binds E2F1^{CM} and E2F3^{CM} with weaker affinity than E2F4^{CM}, and the crystal structure of the p107C complex offers a clear explanation for the preference for E2F4. Notably, differences in the sequence in strand β3 result in weaker binding of activator E2Fs for p107C compared to the repressor E2Fs.

Our phylogenetic analyses and examination of specific sequence motifs involved in pocket protein-E2F interactions also point to a special relationship between Rb and the activating E2Fs, especially E2F1 and E2F2. We find that the emergence of E2F1 and E2F2 is coincident with Rb-specific sequences that result in higher E2F affinity. These sequences nevertheless result in tighter Rb binding to other E2Fs as well. For example, RbC^{nter} increases affinity of Rb and p107 for both E2F4 and E2F1 (Fig 20). We propose that coincident with the divergence of E2F1 and E2F2 from their ancestor and the accumulation of changes that weakened p107/p130 binding, Rb underwent adaptive changes that resulted in increased E2F biding affinity. While these adaptive changes do not result in higher affinity specifically for E2F1 and E2F2, the foregoing poor affinity of p107 and p130 for activator E2Fs rendered Rb the only pocket protein with high affinity for

activating E2Fs. We find it interesting that the RbC^{nter} sequence includes a CDK phosphorylation site that is known to weaken RbC-E2F affinity (Rubin et al. 2005). We suggest that the additional E2F-binding motif co-evolved with a regulatory phosphorylation mechanism such that Rb-E2F complexes can be dissociated for cell-cycle progression.

Despite decades of research, the critical cellular functions of Rb that give rise to tumor suppression remain uncertain (Burkhart and Sage, 2008; Dick and Rubin, 2013; Malumbres and Barbacid, 2001). In the canonical model, Rb negatively regulates proliferation through inhibition of E2F activity and repression of cell cycle gene expression. Considering that p107 and p130 can also inhibit E2F and the cell cycle, it has been argued that other Rb functions must underlie its unique potency as a tumor suppressor (Burkhart and Sage, 2008; Dick and Rubin, 2013; Malumbres and Barbacid, 2001). Indeed, other Rb functions have been identified including regulation of heterochromatin formation, chromosome stability, DNA repair, and apoptosis. But to what extent do these functions entail E2F regulation as a mechanism and to what extent do they involve interactions with other cellular proteins? The answer is not always clear. With the exception of the E2F association, the relevant protein interactions and whether they are Rb-specific are often poorly characterized. Here we demonstrate that there are unique Rb structural features that underlie its exclusive ability to regulate activator E2Fs. These observations complement and explain previous studies that implicate E2F1, E2F2, and E2F3 activity as the cause of aberrant phenotypes in Rb knockout cells (Liu et al. 2015; Chen et al. 2007; Landman et al. 2013; Saavedra et al. 2002). Rb regulation of activator E2Fs cannot be complemented by p107/p130, because they fail to bind with sufficient affinity. Our data support the model that the unique role of Rb in development and tumor suppression arises from its unique capacity to regulate the activator E2Fs. At the same time, our data indicate that p107/p130 have some weak affinity for activating E2Fs that may be exploited. It has been observed that p107 represses E2F1 upon overexpression (Cress et al. 1993), and

endogenous p107/p130 complexes with E2F1 or E2F3 can be detected in mouse fibroblasts that lack E2F4 and thus have higher free p107/p130 concentrations (Lee et al. 2002). Increasing p107/p130 association with the activator E2Fs may be a viable therapeutic strategy towards harnessing their activity to compensate for Rb loss.

3.4 Materials and Methods

3.4.1 Protein expression and purification

Human DP1 (residues 199-350) and E2F4 (residues 91-198) or E2F5 (124-232) proteins were co-expressed from pET-derived vectors containing N-terminal glutathione S-transferase and N-terminal 6xHis tags, respectively. *Eschericia coli* BL21(DE3) cells co-transformed with two both vectors were grown to an OD_{600} of 0.6-0.8 and induced with 500mM IPTG, with protein expression taking place overnight at 18°C. Cells were pelleted at 4,000 rpm for 10 minutes and resuspended in a lysis buffer containing 25 mM Hepes, 100 mM NaCl, 5 mM DTT, 5mM EDTA, 10% glycerol, and 1 mM PMSF (pH 8). A cell homogenizer was used to lyse the cells, and following centrifugation, the resulting soluble fraction was purified using glutathione sepharose affinity chromatography followed immediately by Ni-sepharose affinity chromatography. Following the affinity chromatography columns, the eluate was further purified by anion-exchange chromatography and cleaved with GST-TEV overnight at 4°C. To remove cleaved GST and His tags, proteins were again passed over glutathione and nickel sepharose resin, and finally the heterodimer was subjected to size exclusion chromatography to achieve the final pure sample. All p107C and RbC constructs were expressed and purified similarly to as described (Rubin et al. 2005). .

3.4.2 Protein Crystallization

Prior to crystallization trials, the DP1-E2F4 heterodimer was purified over a SD75 Superdex column in 20 mM Tris 8.0, 100 mM NaCl, and 4 mM TCEP. Crystal trays were set at 13mg/ml, using a 1:1 protein to buffer ratio at 18°C. Diffraction quality crystals appeared within 3-4 days in 100 mM MES, 100 mM NaCl, and 16% PEG 6000 (pH 6.3). Crystals were harvested, incubated in the above condition plus 25% ethylene glycol as a cryoprotectant, and flash frozen in liquid nitrogen prior to data collection.

To crystalize the trimeric protein complex, p107C was mixed in a 2:1 molar ratio with purified DP1-E2F5. The complex was isolated using a SD75 superdex column equilibrated in 20 mM Hepes 7.0, 100 mM NaCl, and 4 mM TCEP. Crystal trays were set at 15 mg/ml and incubated at 18 °C. Diffracting crystals were grown in 100 mM Hepes 7.0, 7% PEG 5000, 5% 1-propanol, and 2% 2-propanol. Crystals were flash frozen in liquid nitrogen in a buffer containing the crystallization condition supplemented with 25% glycerol. Both native and selenomethionine containing crystals were grown using this process. All crystals in this study were grown using the sitting drop vapor diffusion method.

To incorporate selenomethionine into DP1-E2F5 protein complexes, co-transformed BL21 cells were grown in M9 minimal media to an OD_{600} of 0.6. The methionine pathway was inhibited through addition of lysine, phenylalanine, and threonine at 100mgs per liter, isoleucine and valine at 50mgs per liter, and supplemented with 100mgs of selenomethionine per liter. Twenty minutes after addition of the amino acids, cells were induced with 500mM IPTG. DP1-E2F5 containing selenomethionine was otherwise expressed and purified similar to unlabeled protein.

3.4.3 X-ray Structure Determination

Data were collected from single crystals under cryogenic condition (100K) at the Advanced Light Source (Beamline 8.3.1) and Advanced Photon Source (Beamline 23IDB).

Mosflm (Leslie 2006) and SCALA (Bailey 1994), implemented in CCP4 suite, were used for indexing and scaling. For the E2F4-DP1, a molecular replacement solution was found with Phaser (Mccoy et al. 2007) using PDB code 2AZE as a search model. We could not find a molecular replacement solution for the p107C-E2F5-DP1 native diffraction data, perhaps because of the different angles between the coiled-coil and marked-box domains (Fig. 16B). Phases were alternatively calculated from a SAD experiment using a selenomethionine derivative. SAD data were collected at the selenium peak energy. Six selenium sites and an initial structural model were generated by using the program Autosol (Terwilliger et al. 2009) from within the Phenix interface (Adams et al. 2010). While we expected incorporation of 12 selenium atoms (6 per molecule in the asymmetric unit), only 8 methionine residues were visible in the structure. High-resolution native data and the initial model from Autosol were then used for further model building and refinement by using COOT (Emsley and Cowtan, 2004) and Phenix (Adams et al. 2010) respectively. Coordinates and structure factors for the E2F4-DP1^{CM} and p107C-E2F5-DP1^{CM} structures were deposited to the Protein Data Bank under accession codes 5TUU and 5TUV respectively.

3.4.4 Isothermal Titration Calorimetry

Preceding ITC experiments, protein samples were dialyzed overnight at 4°C in 20 mM Tris, 100 mM NaCl, and 1 mM beta-mercaptoethanol. Binding experiments involving p107C were conducted at pH 8.0, and experiments involving RbC were performed at pH 7.0. Typically, 0.5-1 mM peptide was injected into 20-40 μM DP1-E2F^{CM} using a Micro-Cal VP-ITC. Origin software was used to calculate binding constants by fitting the data to a one-site binding model. The error associated with the reported binding constants is the standard deviation calculated from 2-4 independent binding experiments.

3.4.5 Homolog detection and phylogenetic analyses

We used the HMMER 3 package (Eddy 2011) and profile-Hidden Markov Models (profile-HMMs) built previously for eukaryotic-wide phylogenetic analyses (Medina et al. 2016) to retrieve homologs (E-value threshold of 1E-10) of the E2F-DP and pocket protein family from 52 genomes with particular focus on the Metazoa. Some detected homologs were discarded because they were incomplete or too divergent to be included in our phylogenies. All reliable homologs were aligned using MAFFT-L-INS-i (-maxiterate 1000) (Katoh and Standley, 2013). Resulting alignments were masked using probabilistic alignment masking with ZORRO (Wu et al. 2012). ProtTest 3 was used to determine the empirical amino-acid evolutionary model that best fit each of our protein datasets using several criteria: Akaike Information Criterion, corrected Akaike Information Criterion, Bayesian Information Criterion and Decision Theory (Darriba et al. 2011) (E2F-DP: JTT+I+G+F & Rb: LG+I+G+F). Last, for each dataset (E2FDP_alignment.fasta, Rb_alignment.fasta) and its best-fitting model, we ran different phylogenetic programs that use maximum-likelihood methods with different algorithmic approximations (RAxML and PhyML) to reconstruct the phylogenetic relationships between proteins. For RAxML analyses, the best likelihood tree was obtained from five independent maximum likelihood runs started from randomized parsimony trees using the empirical evolutionary model provided by ProtTest. We assessed branch support via rapid bootstrapping (RBS) with 100 pseudo-replicates. PhyML 3.0 phylogenetic trees were obtained from five independent randomized starting neighbor-joining trees (RAND) using the best topology from both NNI and SPR moves. Non-parametric Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRTs) and parametric *à la Bayes* aLRTs (aBayes) were calculated to determine branch support from two independent PhyML 3.0 runs. The Rb phylogeny (Fig. S4) and

E2F phylogeny (Fig. S5) were then used to classify orthologs and paralogs to create Table S2.

Chapter 4: Phosphorylation of p107 by CDK reduces the affinity for E2F4 using multiple mechanisms

4.1 Introduction

The phosphorylation of pocket proteins by CDKs has shown to be an important step in the inactivation of pocket proteins and progression into S phase. The presence of hyperphosphorylated pocket proteins is concomitant with the dissociation of E2Fs from their pocket protein interaction partners. Rb and p107 contain 16 and 17 consensus CDK sites, respectively, and p130 contains 24 (Fig. 1). The general distribution for these CDK sites is conserved, often appearing in groups and localized to the unstructured linker regions between N-terminus and pocket domain, and along the C-terminal domain. p130 is unique in that it contains additional phosphorylation sites at the very N-terminus and in the loop between the two B-subdomains in the pocket, which are not shared by Rb or p107. The function of these sites however is unknown.

The mechanism by which phosphorylation triggers the release of E2F has been well characterized in Rb. Phosphorylation of T373 in Rb causes an interdomain association between the N-terminus and the pocket domain (Burke et al. 2012). The association of the N-terminus with the pocket has two effects. First, the docking of the N-terminus results in a rotation of the pocket domain which widens the E2F binding groove, resulting in a reduced affinity for E2Fs. Second, the affinity of peptides known to bind the pocket domain using the LxCxE is found to be reduced. SAXS data also confirms the association of RbN with the pocket domain, displaying a more compact structure when T373 is phosphorylated versus the phosphorylated protein (Burke et al. 2012; Lamber et al. 2013).

Within the loop connecting the two halves of the Rb pocket, phosphorylation of S608 also reduces the affinity of the E2F TAD for the pocket domain(Burke et al. 2010). While the presence of the Rb loop does not compete for E2F binding in the unphosphorylated state, phosphorylation of S608 causes the pocket loop to flip into the E2F biding groove and directly compete with E2F for access to the pocket domain. The structure of a S608 phosphomimetic mutant reveals that the loop adopts a confirmation similar to how E2F binds (Burke et al. 2010). S608E makes several hydrogen bond contacts with the Rb pocket, as well as Y606. The analogous serine and tyrosine residues are well conserved in p107 and p130, suggesting that pocket loop phosphorylation is a general strategy used to dissociate E2Fs from pocket proteins.

Lastly, two pairs of CDK sites in the Rb C-terminus have an effect on E2F binding. The first pair of sites, S788 and S795, regulates the interaction between the RbC and the E2F1-DP1 CM domains (Rubin et al. 2005). The mechanism used to reduce the affinity for E2F following phosphorylation of S788/S795 is less clear, as structural information regarding this interaction is lacking. However, it is clear that part of the interaction between RbC and E2F1-DP1 is lost following phosphorylation. The second set of phosphorylation sites, T821 and T826, do not seem to have a direct effect in reducing the affinity for E2F1-DP1, but instead causes the RbC to form an intramolecular association with the pocket domain, which ultimately destabilizes the interaction the E2F1-DP1 indirectly (Rubin et al. 2005).

Many of the CDK sites shown to have an effect in dissociating the Rb-E2F interaction are conserved in p107 and p130. p107 and p130 also contain CDK sites conserved with one another, but not conserved in Rb. To address whether phosphorylation of p107 causes release of E2F using similar mechanisms as demonstrated in Rb, we attempted to repeat the studies performed in Rb using p107, as well as test additional sites in p107 that are not conserved in Rb to assay the function, if any, of these CDK sites.

4.2 Results

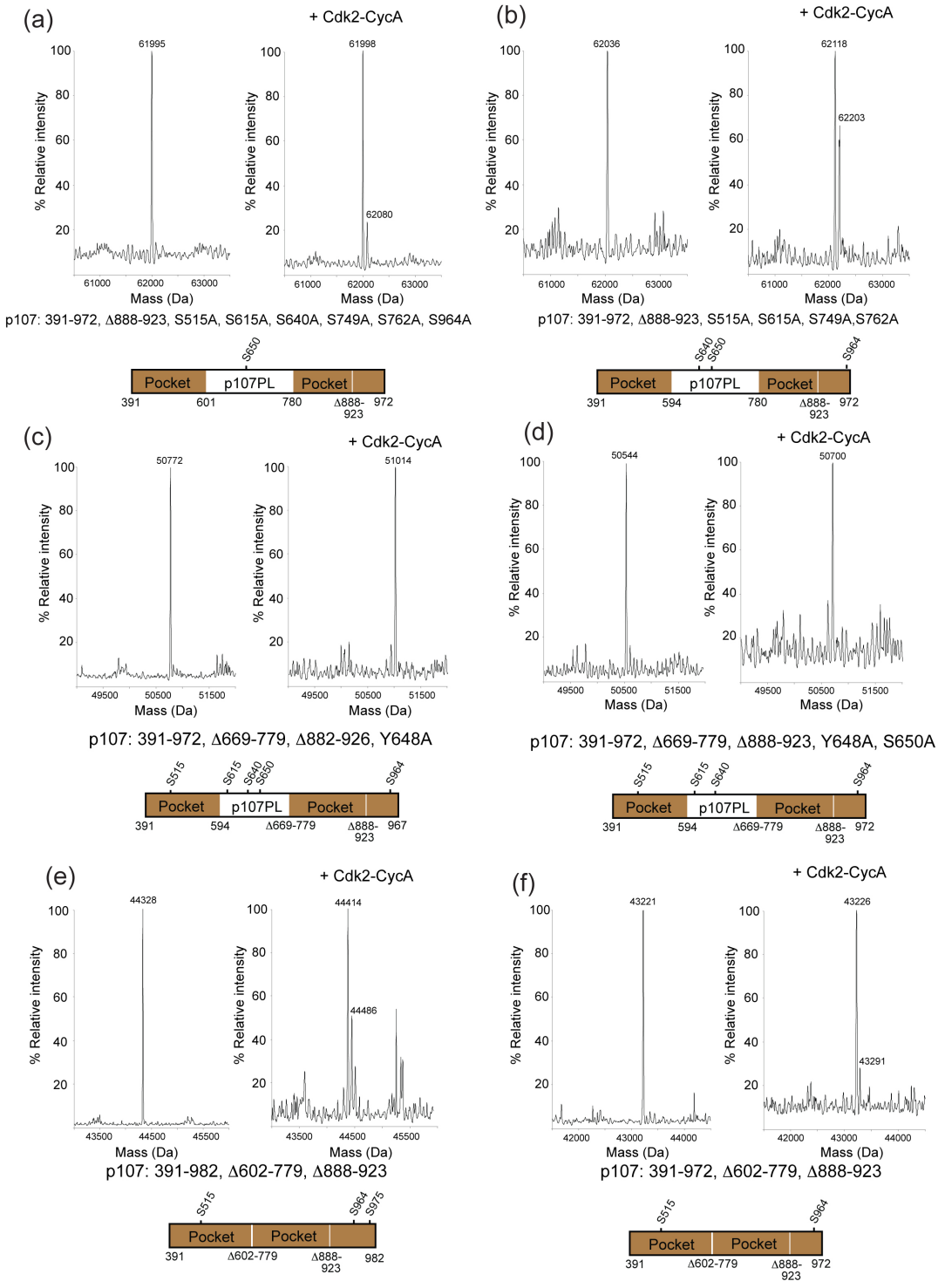
4.2.1. Phosphorylation of S650 in p107 reduces affinity of the p107 pocket domain for the E2F4 TAD

In the first loop connecting the A and B subdomains of the p107 pocket domain there are several CDK sites that have the potential to regulate the affinity of p107 for E2F4 (Fig. 1). The p107 pocket loop contains consensus sites S615, S640, S650, S749, and S762. Based on previously published results demonstrating the ability of the RbPL to actively compete with E2F for binding the pocket groove, we wondered whether the same mechanisms employed by Rb were conserved in p107 and p130 as well. To test this idea, we purified the p107 pocket domain containing the first pocket loop but lacking the second pocket loop connecting the two B subdomains. We then measured the affinity of this unphosphorylated construct for E2F4 using ITC (Fig. 27). Interestingly, we found that the presence of PL1 in this construct caused a 10-fold decrease in affinity. In Rb, the presence of the pocket loop has no effect on the affinity of Rb for different E2Fs.

We then tested this construct against several other E2Fs, and found that for all binding measurements there was roughly a 10-fold decrease in affinity compared to the pocket domain without loops tested against the same E2F (data not shown). This led us to hypothesize that even in the unphosphorylated state the PL1 could be competing with E2Fs for access to the pocket domain. Using the structure of Rb (PDB code 4ELL) and sequence conservation as a guide, we mutated Y648 to alanine and tested binding again to E2F4. In the Rb structure, Y606 is buried in the E2F binding pocket and forms a hydrogen bonding network with several sidechains of Rb along the E2F binding pocket (Burke et al. 2010). If PL1 was indeed occupying the E2F binding site, mutation of this tyrosine should reduce the affinity of PL1 for the pocket domain and result in an increase in affinity for

E2F4. When we tested the Y648A mutant for binding against E2F4, we found a significant increase in affinity, with a binding constant only two-fold above that for the loopless pocket domain, indicating that PL1 can compete in the unphosphorylated form with E2F4 (Fig. 27b). We then continued with this Y648A to determine the effect of phosphorylation of residues along PL1. We were unsuccessful in achieving an S650 fully phosphorylated population in the wild type construct. We reasoned that the S650 site is not accessible given the nature of this loop to be buried within the p107 pocket. Phosphorylation of the Y648A construct allowed us to achieve a homogenously phosphorylated population with incorporation of three phosphates (S615, S640, S650) (Fig. 26c). Testing this phosphorylated protein against E2F4 resulted in a decrease in affinity for E2F4, suggesting that upon phosphorylation the PL1 once again directly competes with E2F4 (Fig. 27b). Repeating this binding experiment with a phosphorylated S650A mutant also resulted in a decrease in affinity for E2F4 compared to wild-type, but bound tighter than the construct containing a phosphorylated S650. This results suggests that phosphorylation of S650 in the p107PL1 is important for regulating the p107-E2F4 interaction. It is possible that the other CDK sites (S615 and S640), also have an effect, but our binding data, as well as conservation among pocket proteins, suggest S650 as the primary phosphorylation site regulating this interaction.

Figure 26: Electrospray ionization mass spectrometry of p107 pocket domain constructs. Mass spectra were acquired for each construct without phosphorylation and with CDK2-CycA phosphorylation. Approximately 0.5-1 mg of purified protein was desalted using Sep-Pak cartridges with C8 reverse phase resin (Waters). Protein was eluted in 3.5 mL of acetonitrile and injected into the mass spectrometer. (a) Pocket domain construct that lacks p107PL2 and is absent of all consensus CDK sites except S650 is resistant to phosphorylation. (b) Same construct as in (a) except S640, S650, and S964 are intact phosphoacceptor sites. The kinase adds mostly one or two phosphates. (c) p107 pocket construct that has a shortened PL1, no PL2, and a Y648A mutation is phosphorylated on three sites. (d) Same construct as in (c) but with a S650A mutation is phosphorylated on two sites. (e) Pocket construct without both PL1 and PL2 and extended to residue 982 is phosphorylated on one site. (f) The same construct as (e) but ends at 972 is resistant to phosphorylation. Together we conclude the following from these data: S615 and S640 are readily phosphorylated, while S515 and S964 are resistant to CDK phosphorylation. S650 is not phosphorylated except when Y648 is mutated.


(a)
+ Cdk2-CycA
61995
61998
62080
% Relative intensity
Mass (Da)
p107: 391-972, Δ888-923, S515A, S615A, S640A, S749A, S762A, S964A
S650
Pocket
p107PL
Pocket
391
601
780
Δ888-923
972
(b)
+ Cdk2-CycA
62036
62118
62203
p107: 391-972, Δ888-923, S515A, S615A, S749A,S762A
S640
S650
S964
391
594
780
Δ888-923
972
(c)
+ Cdk2-CycA
50772
51014
p107: 391-972, Δ669-779, Δ882-926, Y648A
S515
S615
S640
S650
S964
391
594
Δ669-779
Δ888-923
967
(d)
+ Cdk2-CycA
50544
50700
p107: 391-972, Δ669-779, Δ888-923, Y648A, S650A
S515
S615
S640
S964
391
594
Δ669-779
Δ888-923
972
(e)
+ Cdk2-CycA
44328
44414
44486
p107: 391-982, Δ602-779, Δ888-923
S515
S964
S975
391
Δ602-779
Δ888-923
982
(f)
+ Cdk2-CycA
43221
43226
43291
p107: 391-972, Δ602-779, Δ888-923
S515
S964
391
Δ602-779
Δ888-923
972

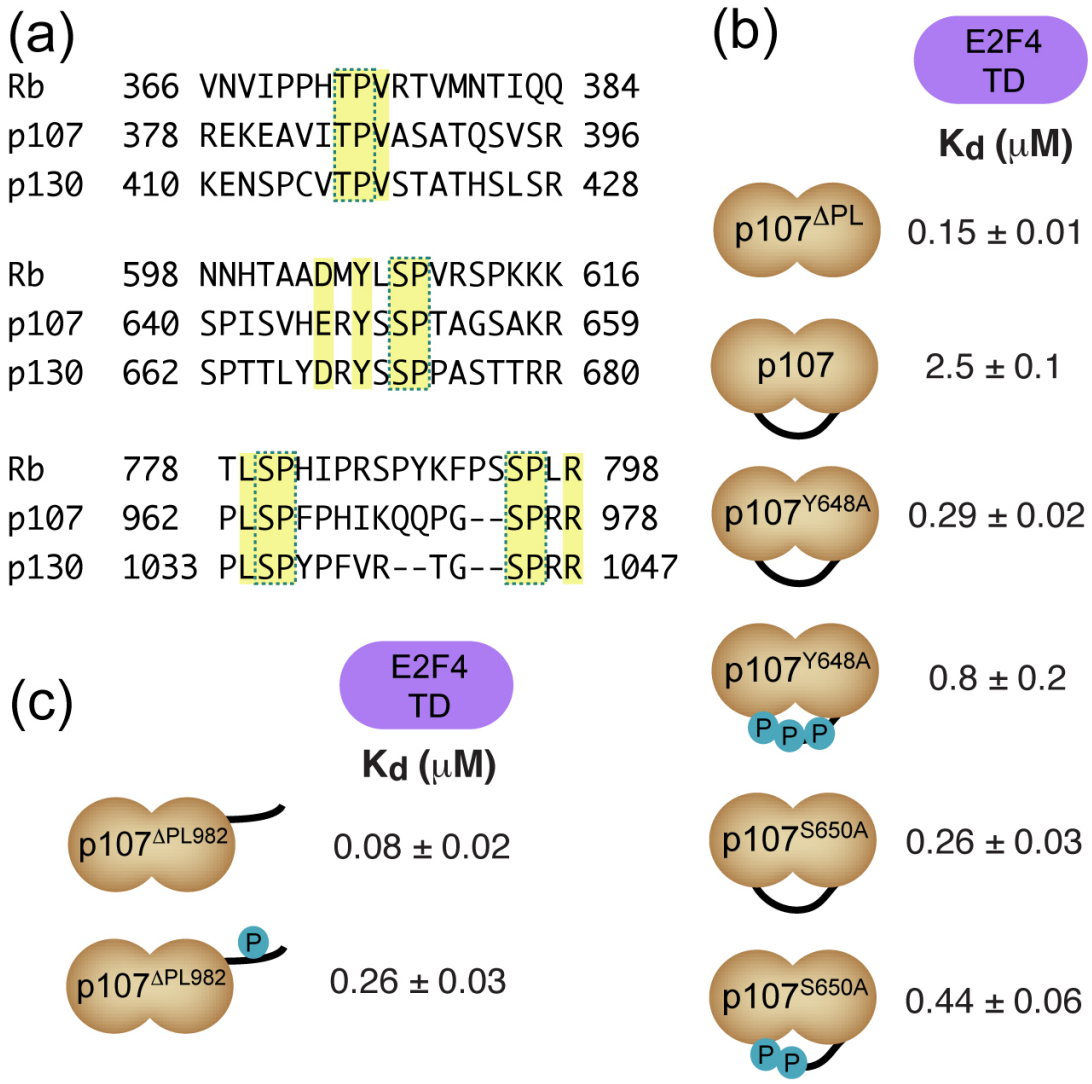


Figure 27: Conservation of CDK inhibitory mechanisms in pocket proteins. (a) Sequence comparison of human Rb, p107, and p130 shows conservation of CDK phosphorylation sites that are known to inhibit $E2F^{TD}$ binding to Rb. (b) and (c) ITC measurements of unphosphorylated and phosphorylated p107 constructs for $E2F4^{TD}$.

4.2.2 Phosphorylated S975 directly competes with E2F4^{TAD} and weakens binding affinity

The p107 pocket domain contains CDK two sites, (S964 and S975) at the far C-terminus of the B subdomain (Fig. 1). Though this region extends about twenty amino acids beyond the structured region of the B subdomain, several structures of Rb and p107 pocket domains show this region to pack against the surfaces of Rb and p107 in an unstructured manner, generally termed the "bar" (Burke et al. 2010; Burke et al 2013; Lee et al. 2002; Guiley et al. 2015). Both Rb and p107 contain a pair of CDK consensus sites in this region at amino acids S788/S795 and S964/S975, respectively (Fig. 1). In both Rb and p107 pocket domain crystal structures, the first phosphorylation site of each pair is visible in the electron density and interacts with the B subdomain, slightly distal from the E2F binding interface (Burke et al. 2010; Burke et al 2013; Lee et al. 2002; Guiley et al. 2015). The second CDK site of each pair was either not included in the crystallographic construct, or is not visible in the electron density. For the latter, this suggests that this region of the C-terminus is no longer associated with the pocket domain and is disordered in solution. While S788 and S964 can't directly compete with E2F binding given their location in the structure, the S795 and S975 sites have the potential to reach the E2F binding groove and directly compete with E2F for access.

We purified two p107 pocket domain constructs, one including only S964 and ending at Q972, and the other that included both S964 and S975, ending at Q982. Both constructs bound E2F4 with similar affinity in the unphosphorylated state. When we tried to phosphorylate the shorter construct that contained only S964 and quantified phosphate incorporation using mass spec, it was determined that the majority of the protein population was unphosphorylated, with only ~10% containing a single phosphate addition (Fig. 26f). Similar to S650 in PL1, we reasoned that the location of this site along the B subdomain occludes it from CDK and results in a low percentage of it getting phosphorylated. The longer p107 construct, containing both S964 and S975, showed ~90% of the population

phosphorylated at one site, and 10% phosphorylated at two sites (Fig. 26e). The unstructured nature of S975 likely allows for it to be readily phosphorylated by CDK, while S964 remains buried in the pocket domain and is resistant to CDK activity. The affinity of this longer phosphorylated construct was four-fold weaker than wild-type, suggesting that S975 phosphorylation results in a decrease in E2F4 affinity (Fig. 27b). While we weren't able to test the effect of S964 phosphorylation alone, our mass spec data combined with the location of this residue in the structures of Rb and p107 suggests that phosphorylation of this residue would not reduce E2F4 binding.

4.2.3 The affinity of p107C for E2F4-DP1^{CM} is diminished following phosphorylation of p107C by CDK at residues T997 and S1009

We next explored the question of how CDK phosphorylation of p107 weakens the p107C- E2F-DP^{CM} association. p107 contains two phosphorylation sites (T997 and S1009) within the C-terminal fragment that is sufficient for E2F-DP^{CM} binding (Fig. 14B). We phosphorylated these sites in the p107$^{994-1031}$ construct using purified CDK2-CycA and found by ITC that the affinity of the phosphorylated peptide was eleven-fold weaker than the unphosphorylated peptide (Fig. 28A). We then made T997A and S1009A mutations in two separate constructs and found that phosphorylation at the remaining site in each construct still leads to a weakened affinity. These measurements demonstrate that both phosphorylation events in p107C inhibit binding to E2F-DP and that their effects are additive.

In the crystal structure of the ternary complex, S1009 is visible in the loop between the p107C strand and helix (Fig. 28B). The loop folds back towards the secondary structure elements, and the S1009 sidechain makes a hydrogen bond with S1013, which is in the p107C helix. Phosphorylation of S1009 likely weakens affinity by destabilizing this

bound conformation. In the construct used for crystallization, electron density for T997 is not visible, suggesting that T997 is disordered. It is less clear then why phosphorylation of T997 inhibits the association.

The phosphorylation pattern within the CTD of p107 and p130 is distinct from the pattern in Rb. In Rb, there are two threonine CDK sites (T821 and T826), but they are both N-terminal to the CTD strand, and their phosphorylation does not directly inhibit binding of RbC^{core} to E2F1-DP1 (Rubin et al. 2005) (Fig. 1). Instead, phosphorylation of these Rb sites induces an interdomain association between phosphorylated RbC and the pocket domain, which competes with RbC^{core} binding to E2F-DP^{CM}. We found that phosphorylation of T997 and S1009 directly inhibits E2F-DP^{CM} binding (Fig. 28a). We could not detect binding of phosphorylated p107C to the p107 pocket domain (data not shown), which is consistent with a report that RbC contains a sequence that can uniquely inhibit T-antigen binding to the pocket domain of pocket proteins (Knudsen and Wang, 1998).

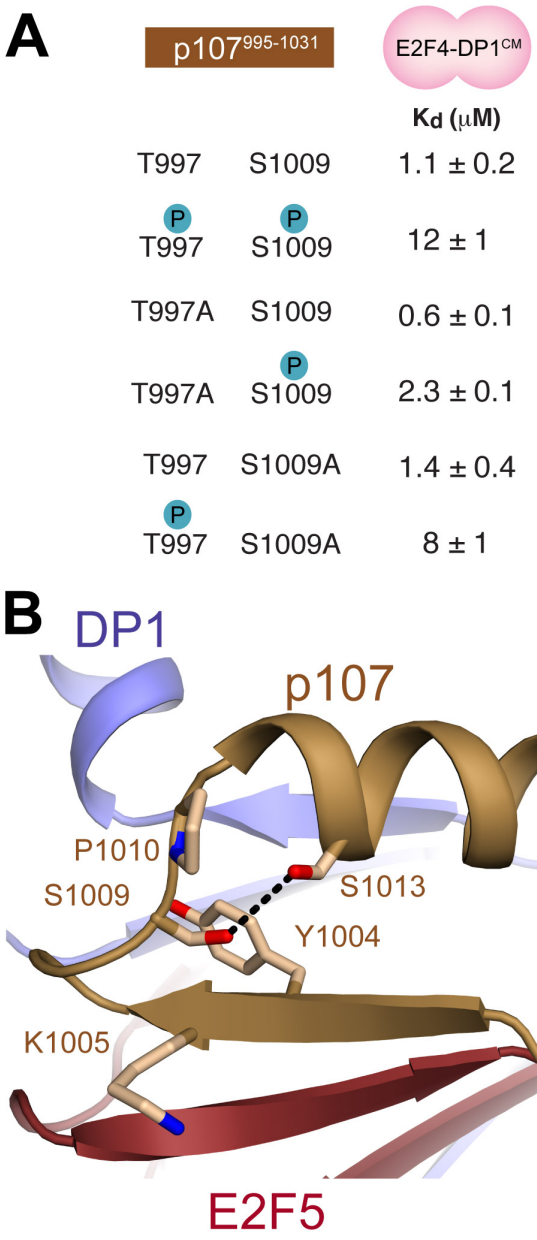


Figure 28: Phosphorylation of CDK sites in p107C directly inhibits E2F binding. (A) ITC measurements of p107C peptides phosphorylated at the indicated sites. (B) Structure of the p107C-E2F5-DP1 interface

4.2.4 p107 and p130 CDK sites located in the linker region between the NTD and the pocket domain are resistant to phosphorylation.

In Rb, phosphorylation of T373 causes the NTD to associate with the pocket Domain (Burke et al. 2012; Lamber at et. 2013). This association of the NTD results in a rotation of the pocket domain that enlarges the E2F binding pocket, reducing affinity for E2Fs. Additionally, access to the LxCxE binding cleft in the pocket is also occluded following phosphorylation of T373 (Burke et al. 2012). Both p107 and p130 contain analogous CDK sites in their interdomain linkers at residues T385 and T417, respectively (Fig. 1). We wondered whether this mechanism was conserved across all pocket proteins, and set out to phosphorylate p107 and p130 constructs that contained both the N-terminus and pocket domain. Unfortunately, we were not able to resolve our mass spec data and quantitatively determine whether we had incorporated phosphates. We conducted binding assays with p107/p130 and E2F4 or LIN52 individually, before and after phosphorylation, and found no significance difference in affinity for either binding partner. However, as we were unable to measure the incorporation of phosphates, we were not able to draw any conclusions from these binding experiments. Both p107 and p130 are known to contain CDK inhibitory regions in the constructs that we used, so we reasoned it was possible that the lack of difference in binding affinity is due to inhibition of CDK activity by its substrate. To qualitatively determine if our constructs were inhibiting CDK activity, we used radioactive phosphate in our kinase experiments, and performed autoradiography to assess CDK activity. The p32 assay clearly shows that the activity of CDK is not inhibited by p107 or p130, as both appear to be readily phosphorylated (Fig. 29). However, each of these constructs contains multiple phosphorylation sites, and without the ability to perform mass spec, determining if all of our sites are phosphorylated is challenging.

p107 construct	available phos sites
1) 391-972, Δ669-779, Δ888-923, Y648A/S650A	S515, S615, S640, S964
2) 391-982, Δ602-779, Δ888-923	S515, S964, S975
3) 1-335	T332
4) 1-972, Δ602-779, Δ888-923, T332A/T340A	T369, ***T385***, S515, S964
5) 18-972, Δ602-779, Δ888-923, T332A/T340A/T385A/S515A/S964A	T369

Figure 29: Labeling of p107 substrates with p32 to assay phosphate incorporation
Lanes 1-5 represent different p107 constructs used for our p32 assay with CDK2-CyclinA. P107 constructs for each lane are listed below, with the available phosphorylation sites shown at the right. Residues highlighted in red are the predicted sites that are incorporation radiolabeled phosphates, based on our mass spec data. T835 is the residue we were hoping to phosphorylate.

4.3 Discussion

In our attempt to compare the mechanisms relating CDK activity and E2F affinity between pocket proteins, we discovered both conserved and unique mechanisms by which phosphorylation triggers the release of E2F (Fig. 27 and Fig. 28). We do observe here phosphorylation events in human p107 that weaken E2F binding similar to previous observations for Rb (Fig. 27). However, these sites (S650 and S975 in p107 and S608 and S795 in Rb) are only conserved in vertebrate pocket proteins, suggesting that E2F regulation has evolved more recently. In general, we have found here that the inhibitory effects of p107 phosphorylation on E2F affinity are more modest than the effects of Rb phosphorylation (Burke et al. 2013), and CDK cannot quantitatively phosphorylate p107 at a key inhibitory site S650 *in vitro* (Fig. 26). These observations are consistent with the prevalence of p107-E2F4 complexes in S phase and may be critical for how p107 and Rb differentially regulate E2F in the cell cycle (Cao et al. 1992, Devoto et al. 1992; Moberg et al. 1996; Shirodkar et al. 1992).

The p107C sites T997 and S1009, when phosphorylated, represent a regulatory mechanism that is distinct from Rb (Fig. 28). S1009, which is not conserved in Rb but is conserved in p130, forms a hydrogen bond with S1013 (Fig. 28), and likely assists in the formation of the sheet-turn-helix motif displayed by p107C. When phosphorylated, this hydrogen bonding network is likely to be disrupted, resulting in a decreased affinity of p107C for E2F4-DP1 (Fig. 28). S1009 phosphorylation represents a mechanism by which CDK activity can directly reduce the affinity of p107 for E2F4-DP1 that is not conserved in Rb. Interpreting the result that T997 phosphorylation causes a decrease in affinity of p107C for E2F4-DP1 is not as straightforward. T997 is included in our crystallography construct, but for both copies in the asymmetric unit there is no density N-terminal of S1000, suggesting that this region is unstructured. Interestingly, T997 phosphorylation

appears to be responsible for most of the binding defect observed when both sites are phosphorylated (Fig. 28).

Our attempts to phosphorylate either p107 or p130 NTD-pocket constructs, and observe the level of phosphorylation quantitatively, was immediately met with resistance. Despite numerous attempts and multiple strategies employed, we were unable to use mass spectrometry to assay the level of phosphorylation. We could resolve spectra for untreated p107 and p130 constructs, but following CDK treatment, it appeared that the protein was either crashing out on the desalting resin, or refusing to elute. Regardless, we took CDK treated proteins and assayed binding to E2F4, LIN52, or E7, and found no significant difference in affinity. We initially reasoned that perhaps the CDK-inhibitory region present in p107 and p130 was acting antagonistically to our phosphorylation efforts. However, kinase assays using radiolabeled ATP showed that our kinase is active and capable of phosphorylating a variety of p107 and p130 constructs, including those containing the proposed CDK inhibitory region (Fig. 29). Our radiolabeled ATP assay demonstrates that our kinase is active, but does not quantitatively show the number of phosphates incorporated per molecule. It could be that T385 in p107, similar to S650 S964, is buried and is not accessible to CDK, or that T385 is just a poor CDK substrate.

Currently, it is difficult to make any claims towards whether T385 in p107 has any effect on E2F or LxCxE containing protein affinity when phosphorylated. Structural conservation with Rb suggest that T385 phosphorylation would function in a manner similar to Rb, but our results suggest that perhaps this mechanism is not conserved across the pocket protein family, even though the CDK site itself is conserved among higher order species. Additional work will need to be performed to definitively answer whether phosphorylation of T385 in p107 has any effect on the binding affinities of proteins occupying the E2F or LxCxE binding clefts.

4.4 Materials and Methods

4.4.1 Protein expression and purification

All p107 constructs were expressed in *Escherichia coli*, using PET-derived vectors, as fusion proteins containing an N-terminal glutathione S-transferase (GST) tag. Transformed BL21(DE3) cells were grown to an OD_{600} of 0.6-0.8 and induced with 1 mM IPTG. Protein expression took place overnight at 22°C. Cells were resuspended in lysis buffer containing 25 mM Tris-HCl, 250 mM NaCl, 5 mM DTT, 5% glycerol, and 1 mM PMSF (pH 8). Cells were lysed by passing them three times through a cell homogenizer, and following centrifugation the resulting soluble fraction was purified using glutathione sepharose affinity chromatography. The subsequent eluate was further purified by anion-exchange chromatography and cleaved with GST-TEV overnight at 4°C. To remove cleaved GST, proteins were again passed over glutathione sepharose resin, and finally subjected to size exclusion chromatography to achieve a pure sample. Purification of DP1-E2F protein complexes was performed as previously described.

4.4.2 Protein phosphorylation

Phosphorylation of p107 constructs was achieved by adding 10% by mass CDK2-CyclinA to the p107 substrate in 25mM Tris, 150mM NaCl, 5mM DTT, 10mM $MgCl_2$, and 1mM ATP. The kinase was purified away from phosphorylated p107 using size exclusion chromatography. Incorporation of phosphate was determined quantitatively using electrospray mass spectrometry. Samples were desalted and directly injected into the spectrometer.

4.4.3 Isothermal Titration Calorimetry

Preceding ITC experiments, protein samples were dialyzed overnight at 4°C in 20 mM Tris, 100 mM NaCl, and 1 mM beta-mercaptoethanol, pH 8.0. Typically, 0.5-1 mM

peptide was injected into 20-40 μM protein using a Micro-Cal VP-ITC. Origin software was used to calculate binding constants by fitting the data to a one-site binding model. The error associated with the reported binding constants is the standard deviation calculated from 2-4 independent binding experiments.

Chapter 5: Structural Mechanisms of DREAM Complex Assembly and Regulation

The data presented in the following chapter is the result of work from many talented individuals. My contributions to this paper include solving the first structure of p107, bound to the human papillomavirus (HPV) E7 peptide. The conditions used to solve this structure proved to be essential for solving additional p107 structures in this paper, as p107-LIN52 crystals were non-existent in commercial screens and only appeared in screens previously optimized for p107-E7. Additional work included ITC binding

experiments between the p107 pocket domain, both wild-type and mutant, and the HPV E7 peptide.

*Supplemental data referred to, but not presented in this thesis, can be found in the original publication

5.1 Introduction

Cell cycle exit is a critical process for differentiation and tumor prevention, and cancer cells often have lesions in pathways that control temporary (quiescence) or permanent (senescence) exit. Understanding fundamental mechanisms of cell cycle exit is critical for understanding development and for ultimately designing therapeutic strategies that manipulate exit pathways for halting tumor proliferation. Quiescence has recently been shown to be dependent on the highly conserved protein complex known as DREAM (Litovchick et al. 2007; Litovchick et al. 2011; Sadasivam and DeCaprio 2013). Genetics experiments in model organisms reveal an essential role for DREAM components in differentiation, cell proliferation, and tumor suppression (Korenjak et al. 2004; Lewis et al. 2004; Harrison et al. 2006; Litovchick et al. 2007; van den Heuvel and Dyson 2008; Reichert et al. 2010). DREAM deficient mice die shortly after birth with bone developmental defects that result from aberrant chondrocyte differentiation (Forristal et al. 2014). In human cell culture, DREAM is assembled upon serum starvation in an experimental state of quiescence, and disruption of the DREAM complex drives cells back into the cell cycle despite environmental cues to arrest (Pilkinton et al. 2007; Schmit et al. 2007; Litovchick et al. 2011).

Chromatin immunoprecipitation (ChIP) has located DREAM proteins at a majority of E2F and cell cycle homology regions (CHR) promoters in human cells, and gene expression analysis has implicated DREAM as a repressor of cell cycle genes (Litovchick

et al. 2007; Schmit et al. 2007; Muller et al. 2012). The mechanism of how DREAM regulates transcription has not been elucidated, but it is clear that its central components scaffold a number of key cell cycle transcription factors including E2F4/5, B-Myb, FoxM1, and the Rb tumor suppressor family paralogs p107 and p130 (Korenjak et al. 2004; Lewis et al. 2004; Litovchick et al. 2007; Schmit et al. 2007; Sadasivam et al. 2012). DREAM was originally isolated through biochemical purification in flies and worms, and in each case the complex contained multi-vulval class B protein homologs LIN9, LIN37, LIN52, LIN54, and RBAP48 (mammalian protein names), which together are called the MuvB subcomplex (Korenjak et al. 2004; Lewis et al. 2004; Harrison et al. 2006; van den Heuvel and Dyson 2008). The biochemical functions of these proteins are unknown, with the exception of RBAP48 and LIN54, which bind histones and DNA respectively (van den Heuvel and Dyson 2008; Sadasivam and DeCaprio 2013).

In mammalian cells, MuvB associates with p130-E2F4/5 to form DREAM in G_0 and early G_1, and it associates with B-Myb during S phase to form the Myb-MuvB (MMB) complex (Pilkinton et al. 2007; Schmit et al. 2007; Litovchick et al. 2011; Sadasivam et al. 2012). While DREAM represses gene expression, MMB activates late cell cycle genes during S phase and G2 both with and without FoxM1 (Georlette et al. 2007; Litovchick et al. 2007; Pilkinton et al. 2007; Schmit et al. 2007; Wen et al. 2008; Sadasivam et al. 2012). MuvB thus has an intriguing role in unifying cell cycle dependent gene expression, however the mechanisms by which DREAM and MMB-FoxM1 are assembled and regulated to carry out their specific functions are not well understood. DREAM formation requires the phosphorylation of MuvB protein LIN52 at serine residue 28 (phosS28) by the DYRK1A kinase (Litovchick et al. 2011). “Delta DREAM” mice that lack p130 and carry a p107 mutant allele incapable of binding MuvB display an identical phenotype to p107/p130 double knockout mice, suggesting an intimate relationship between the pocket proteins and MuvB proteins during cell cycle arrest (Forristal et al. 2014). While there are some reports

of MuvB binding Rb (Gagrica et al. 2004; Korenjak et al. 2004), other evidence indicates that the DREAM complex only assembles with either p107 or p130 (Litovchick et al. 2007; Pilkinton et al. 2007; Schmit et al. 2007; Forristal et al. 2014). p130 expression levels are high during quiescence, and p130 is an established biomarker at this stage of the cell cycle in which DREAM is a repressor of gene expression (Smith et al. 1996; Henley and Dick 2012). Loss of p130 results in upregulated expression of p107 and subsequent recruitment of p107 to MuvB (Litovchick et al. 2007; Forristal et al. 2014). MuvB dissociation from pocket proteins coincides with Cyclin-dependent kinase (CDK) activity (Pilkinton et al. 2007), suggesting that CDK phosphorylation of p107 and p130 may inhibit DREAM formation in cycling cells. Knowledge of the molecular architecture of the MuvB complex, how MuvB binds different transcription factors, and how these associations are manipulated throughout the cell cycle is important for understanding such fundamental processes as entry into quiescence and mitotic cell cycle control.

Here we describe the structural mechanisms of DREAM assembly following DYRK1A phosphorylation and DREAM inhibition by either oncogenic viral proteins or CDK phosphorylation. We find that phosphorylated LIN52 binds the pocket domain of p107 or p130 using a short sequence that is a tunable variation on the canonical 'LxCxExL' Rb-binding sequence. A crystal structure of a LIN52-p107 complex explains the requirement for phosphorylation, why the association is specific for p107/p130 and not Rb, and how DREAM is inhibited by viral proteins. The structural observations lead us to uncover a novel model for cell cycle dependent regulation of DREAM assembly in which CDK phosphorylation of pocket proteins induces a conformation that is incompatible with LIN52 binding.

5.2 Results

5.2.1 p107 and p130 directly associate with LIN52

It has previously been shown that p107 and p130 co-immunoprecipitate with MuvB proteins if S28 in LIN52 is phosphorylated by DYRK1A (Litovchick et al. 2011). To determine if LIN52 directly associates with p130, we performed a series of binding experiments with recombinant, purified proteins. We first purified full length GST-LIN52-HIS and GST-p130 both expressed in Sf9 cells and found co-precipitation using Ni^{2+} affinity resin (Fig. 30a). Electrospray mass spectrometry indicated that the LIN52 purified from Sf9 cells is phosphorylated (Figure 30A and 30B). Given the observation that the binding appears substoichiometric, we postulated that p130 is heterogeneously phosphorylated in Sf9 cells, as previously observed for Rb (Burke et al. 2010), and that phosphorylation weakens LIN52 association. We treated p130 with lambda phosphatase during purification and observed a band shift on a Phos-tag SDS-PAGE gel (Supplemental Figure 1C*). After treatment with and subsequent separation from phosphatase, the p130-LIN52 complex stability appears increased (Fig. 30a).

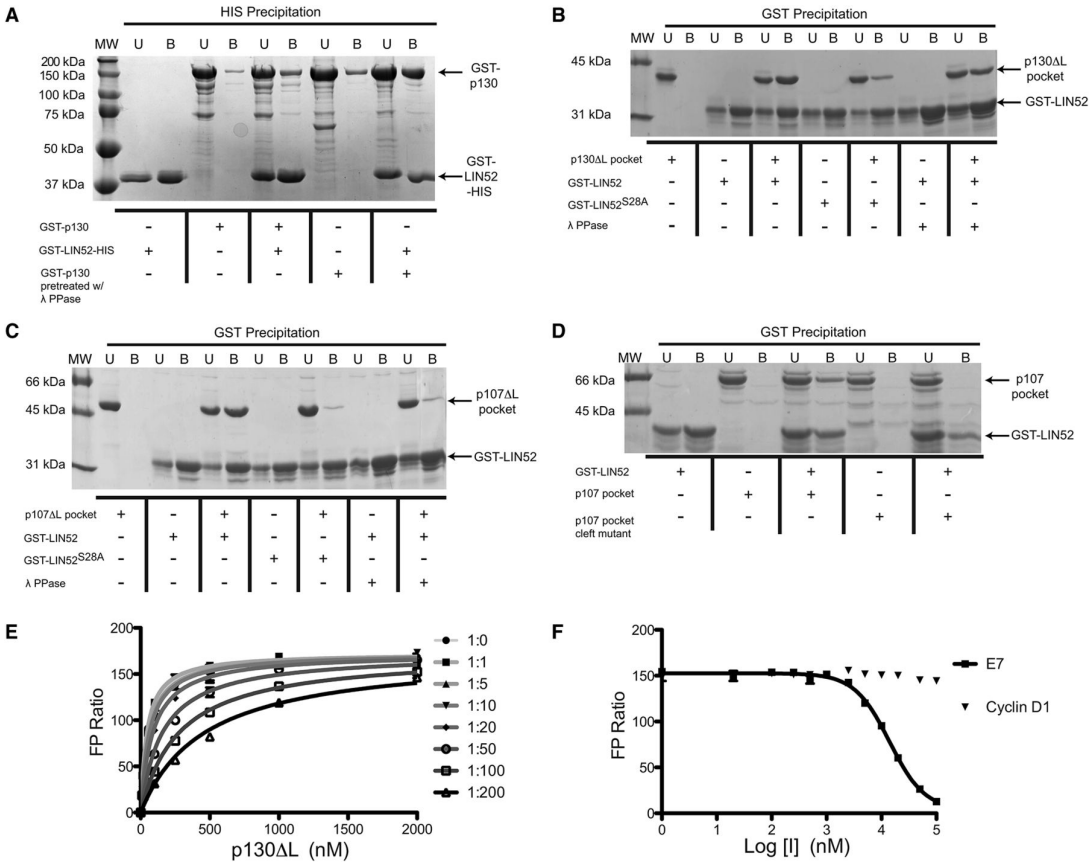


Figure 30: Direct association between LIN52 and p107/p130. Co-precipitation experiments performed with recombinant purified proteins. The indicated proteins were mixed, precipitated with the appropriate resin, and both unbound (U) and bound (B) fractions were analyzed using polyacrylamide gel electrophoresis and Coomassie staining. (A) Co-precipitation of the full length GST-p130 and GST-LIN52-HIS expressed in Sf9 cells using Ni^{2+}-NTA resin. Treatment of the p130 with lambda phosphatase during purification increases the amount of p130 appearing in the bound fraction. (B) Co-precipitation of the untagged p130DL pocket domain expressed in Sf9 cells with GST-$LIN52^{13-45}$ using GS4B sepharose resin. Mutation of S28 in LIN52 or its treatment with lambda phosphatase weakens the p130DL association. (C) Same as (B) but with p107DL pocket domain. (D) Similar experiment to (B) but using p107 pocket domain (loops intact) and an LxCxE cleft (I931A, N935A, and V939A) mutant. (E) A fluorescence polarization assay demonstrating inhibition of the LIN52-p130 association by an HPV E7 peptide. 10 nM TMR-$LIN52^{12-34;phosS28}$ was mixed with increasing concentration of p130DL. Different saturation curves are at the indicated molar ratio of TMR-$LIN52^{12-34;phosS28}$ to $E7^{21-29}$ peptide. The affinity of TMR-$LIN52^{12-34;phosS28}$ in the absence of $E7^{21-29}$ is K_d = 53 ± 2 nM. (F) FP measurements made of 10 nM TMR-$LIN52^{12-34;phosS28}$ in the presence of saturating (2000 nM) p130DL and increasing concentrations of $E7^{21-29}$ peptide (squares) as the inhibitor (I). As a negative control, a peptide corresponding to the N-terminus of Cyclin D1 (residues 1-17) was used (triangles). This peptide does not associate with p130DL (Supplementary Figure 5*).

We next tested whether phosphorylation of S28 in LIN52 is necessary for its direct interaction with pocket proteins. p130 and p107 pocket domains were expressed and purified from *E. coli* without posttranslational modifications. These constructs (p107DL and p130DL) each lack two internal loops and are similar to crystallized Rb constructs that maintain E2F and viral protein binding activity (Lee et al. 1998; Balog et al. 2011). We also purified LIN52 constructs expressed in Sf9 cells containing the N-terminal domain (residues 13-45; $LIN52^{13-45}$). We found that $LIN52^{13-45}$ is sufficient for direct association with the p130 and p107 pocket domains. An S28A mutation in $LIN52^{13-45}$ or lambda phosphatase treatment reduces binding to p130 and p107 (Fig. 30b and 30c), demonstrating the importance of S28 phosphorylation for formation of the complex.

The pocket domain contains two protein interaction interfaces--the LxCxE cleft and the E2F transactivation domain ($E2F^{TD}$) binding site (Lee et al. 1998; Lee et al. 2002; Xiao et al. 2003; Dick and Rubin 2013). The LxCxE cleft binds viral and endogenous proteins containing an "LxCxExf" sequence motif (x is any and f is a hydrophobic amino acid) (Jones et al. 1990; Lee et al. 1998; Singh et al. 2005). A p107 cleft mutant fails to assemble DREAM *in vivo* (Forristal et al. 2014), and we therefore hypothesized that LIN52 directly binds to the LxCxE cleft rather than the $E2F^{TD}$ site. In support of this hypothesis, a purified p107 pocket domain cleft mutant (I931A, N935A, and V939A) fails to bind LIN52 (Fig. 30d). The DREAM complex is perturbed in cervical cancer cells infected with human papillomavirus type 16 (HPV16), and knockdown of the oncoprotein HPV16 E7 rescues DREAM assembly and promotes cell cycle arrest (Nor Rashid et al. 2011). To determine if HPV16 E7 can inhibit the direct LIN52 association with the p130 pocket domain, we used a fluorescence polarization (FP) competition assay. A $LIN52^{12-34;phosS28}$ peptide was labeled with a tetramethylrhodamine (TMR) dye and mixed with increasing amounts of p130DL to obtain FP saturation curves. Addition of $E7^{21-29}$ peptide to the binding reaction shifts the curves such that p130DL appears to bind with weaker affinity (Fig. 30e). We then

quantified the E7 peptide inhibition (K_i = 170 ± 50 nM) by displacing labeled LIN52$^{12\text{-}34;phosS28}$ peptide in the FP assay (Fig. 30e). E7$^{21\text{-}29}$ contains an LxCxExL motif and inhibits LIN52 from interacting with p130, indicating that E7 disrupts DREAM by competitive inhibition at the LxCxE cleft of p130.

5.2.2 Crystal structures of LIN52-p107 and E7-p107 complexes

To identify the molecular interactions involved in DREAM assembly and oncoprotein-facilitated disruption, we solved the crystal structures of the p107DL pocket domain bound to a phosphorylated synthetic LIN52 peptide (LIN52$^{12\text{-}34;phosS28}$) at 2.3 Å and bound to an E7 (E7$^{21\text{-}29}$) peptide at 2.2 Å (Fig. 31, Table 3). The structures reveal that the overall fold of the p107 pocket domain is similar to that of Rb, consisting of two helical subdomains called the A and B boxes. Each subdomain contains a five-helix Cyclin fold, flanked by additional helices that form the A-B interface, the LxCxE interface, or cover hydrophobic surfaces. As with previous Rb pocket domain structures (Lee et al. 1998; Lee et al. 2002; Xiao et al. 2003; Balog et al. 2011), most residues in the central loop between the two subdomains were left out of the crystallized construct, and the residues that remain are not visible in the electron density.

Structural alignment of the p107 and Rb pocket domains (PDB: 3POM) shows a root-mean square deviation in Cα position of only 1.2 Å (Fig. 32a). While most secondary structural elements compare, relative to Rb, the p107 A box has two additional small helices (α4' and α10'). These helices consist of sequences for which the corresponding electron density in Rb is not observable.

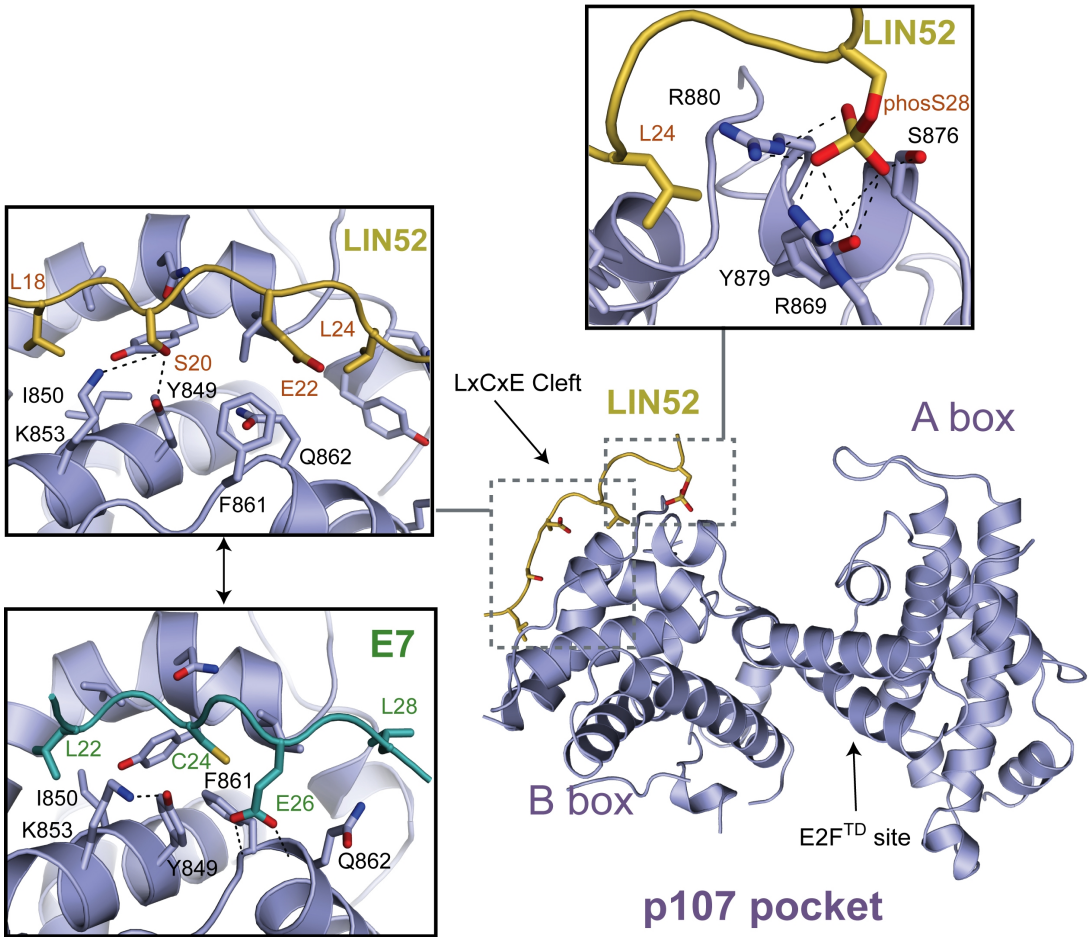


Figure 31: Crystal structures of the p107 pocket domain in complex with LIN52$^{12\text{-}34;phosS28}$ and E7$^{21\text{-}29}$. The LIN52 and E7 peptides both bind at the 'LxCxE' cleft (insets, left). The LIN52 association is mediated by the S28 phosphate, which contacts a p107/p130-specific binding pocket (inset top).

	$E7^{21-29}$	$LIN52^{12-34;phosS28}$	$LIN52^{12-34;P29A;\ phosS28}$
Data collection			
Space group	$C2_1$	$C222_1$	$C222_1$
Cell dimensions			
a, b, c	99.7, 76.6, 74.7	75.4, 101.1, 140.7	74.8, 100.5, 142.8
α, β, γ	90, 120.3, 90	90, 90, 90	90, 90, 90
Resolution	57.2 Å–2.2 Å	60.5 Å–2.3 Å	71.4 Å–2.4 Å
Total reflections	48,201	159,897	112,012
Unique reflections	21,552	24,252	21,397
I/σ	8.6 (2.1)	7.1 (2.2)	6.6 (1.8)
Completeness	93.2% (94.4%)	99.5% (97.9%)	99.3% (97.9%)
Redundancy	2.2 (2.2)	6.6 (6.4)	5.2 (5.4)
R_{pim}	5.8% (25.2%)	7.0% (39.2%)	8.3% (44.2%)
$CC^{1/2}$	0.992 (0.759)	0.991 (0.556)	0.987 (0.472)
Refinement			
Resolution	57.2 Å–2.2 Å	60.5 Å–2.3 Å	71.4 Å–2.4 Å
Number of reflections	21,552 (2171)	24,232 (2297)	21,369 (2084)
R_{work}	18.4% (21.9%)	19.9% (27.4%)	19.5% (26.9%)
R_{free}	24.0 (26.9%)	24.2% (30.0%)	24.2% (33.6%)
Number of atoms	3008	2970	3082
Protein	2877	2849	2934
Water	111	95	128
RMSDs			
Bond lengths	0.005	0.005	0.007
Bond angles	0.87°	0.93°	1.23°
Average B factor			
Overall	36.00 $Å^2$	61.30 $Å^2$	52.70 $Å^2$
p107ΔL	34.53 $Å^2$	59.10 $Å^2$	50.25 $Å^2$
Peptide	66.10 $Å^2$	129.12 $Å^2$	92.95 $Å^2$
Ramachandran analysis			
Favored	97.4%	96.0%	96.0%
Outliers	0.0%	0.3%	0.3%

Values in parenthesis are for the high-resolution shell.

Table 3: Xray crystallographic data collection and refinement statistics for crystals containing p107ΔL and the indicated peptide. Values in parenthesis are for the high-resolution shell.

10' occurs at the C-terminus of the A box and creates additional A-B interface contacts through packing against α11. The E7 peptide binds to the LxCxE cleft in the B-box of p107 as previously observed in the Rb-E7 complex structure (Lee et al. 1998) (Fig. 31 and Fig. 32b). The similar specific contacts made are consistent with the high sequence homology between Rb and p107 within the LxCxE cleft.

In the p107DL-$LIN52^{12-34;phosS28}$ structure, LIN52 binds at the LxCxE cleft consistent with our co-precipitation experiments. The LIN52 peptide binds with an LxSxExL (residues 18-24) motif and makes a similar set of interactions as the E7 peptide (Fig. 31, Fig. 33A). However, several striking differences arise from the presence of the more polar hydroxyl group in the LIN52 serine sidechain. S20 occupies the analogous place in the sequence as the cysteine in the canonical E7 peptide LxCxExf motif. C24 in the E7 peptide fits in a

hydrophobic pocket formed by F861 and Y849. In contrast, S20 in LIN52 makes hydrogen bonds with the K853 and Y849 sidechains. The position of both the F861 and Y849 sidechains are changed in the LIN52 complex such that the F861 phenyl ring is flipped out towards solvent. Whereas the E7 peptide E26 sidechain makes hydrogen bonds with the backbone of F861 and Q862, the LIN52 E22 sidechain is occluded from the analogous position by the flipped out F861 sidechain. These structural differences suggest that the LIN52 serine-containing motif binds with weaker affinity than the canonical E7 motif.

The S28 phosphoserine sidechain in LIN52 binds a positively charged pocket in p107 consisting of R869, S876, Y879, and R880. These p107 sidechains are highly conserved in p107 and p130 paralogs but not in Rb (Fig. 33a). Rb notably has F739 at the Y879 position in p107 and lacks the phenolic hydroxyl group that hydrogen bonds with the LIN52 phosphate. The lack of the phosphate-binding pocket in Rb supports and explains previous observations that the DREAM complex does not assemble with Rb (Litovchick et al. 2007; Pilkinton et al. 2007; Schmit et al. 2007; Forristal et al. 2014). The intervening amino acids between the 'LxSxExL' sequence and phosS28 loop out from the p107 domain and do not make any direct contacts (Fig. 31). We also solved a crystal structure of p107DL in complex with a mutant peptide ($LIN52^{12\text{-}34;P29A;phosS28}$) that binds with higher affinity. The electron density corresponding to the LIN52 peptide is stronger in this structure and allows for observation of additional interactions between W32 in LIN52 and E863, M865, R869, and Y879 in p107.

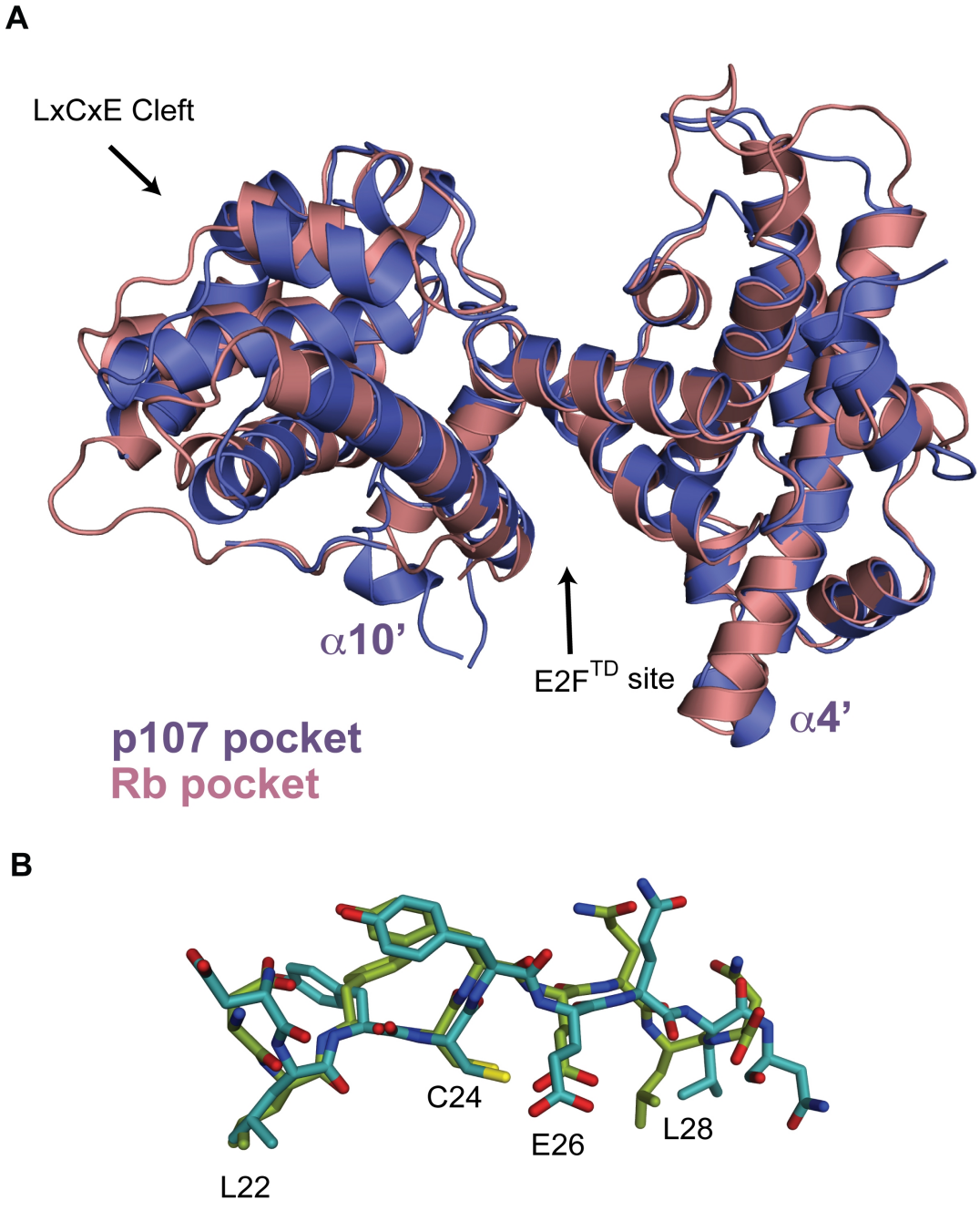


Figure 32: Comparison of p107 and Rb structures. (A) Overlay of the pocket domains constructed by aligning Cα positions of the p107-E7 structure and the unliganded Rb structure (PDB ID: 3POM). (B) Overlay of the E7 'LxCxExL' peptide bound to p107 (teal) and Rb (green). The image was constructed by aligning the pocket domains of the p107-E7 structure and the Rb-E7 structure (PDB ID: 1GUX).

While p107 and p130 loss in genetic models does not readily lead to tumors, there is evidence that p107 and p130 have tumor suppression function and can compensate for Rb loss in certain contexts (Dannenberg et al. 2004; Wirt and Sage 2010). The cBioPortal for Cancer Genomics catalogs a number of missense and nonsense mutations in both p107 and p130 that occur in human cancer samples (Gao et al. 2013). We found that 31 of the currently reported missense mutations for p107 are within the structured pocket domain. Based on the location of these residues in our crystal structure, we predict that 21 of these 31 mutations would destabilize the protein and likely result in functional loss. Three other mutations map to the E2F binding surface based on the analogous Rb-E2F structure (Lee et al. 2002; Xiao et al. 2003). Notably, two mutations (R880I and Y934C) and an analogous p130 mutation map to the LIN52 interface, and we found that these p107 mutations weaken the LIN52$^{12\text{-}34;phosS28}$ peptide affinity. These observations support the idea that p107/p130-dependent growth control, both through E2F inhibition and DREAM function, plays an important role in tumor suppression.

5.2.3 LIN52 S28 phosphorylation increases 'LxSxExL' affinity

Our structural data suggest that the 'LxSxExL' sequence in LIN52 binds to the pocket domain with lower affinity than a canonical 'LxCxExL' sequence and that phosphorylation increases affinity specifically in p107 and p130 to make a more stable complex. We tested these ideas by quantitatively measuring the affinity of different LIN52 peptides using isothermal titration calorimetry. LIN52$^{12\text{-}34;phosS28}$ binds p107DL with an affinity of K_d = 1.4 ± 0.9 μM and the p130ΔL with a similar affinity of K_d = 1.0 ± 0.1 μM. An unphosphorylated LIN52$^{12\text{-}34}$ peptide binds p107ΔL with weaker affinity (K_d = 5.9 ± 0.9 μM). Mutation of S20 to a cysteine increases affinity for p107ΔL (LIN52$^{12\text{-}34;S20C}$ K_d = 0.120 ± 0.007 μM). The sensitivity of the canonical LxCxExL motif to a serine substitution has previously been

observed in the case of the E7 peptide and Rb (Jones et al. 1990). We find that the LIN52$^{12\text{-}34;S20C}$ affinity for p107ΔL is similar to the affinity of the HPV16 E7$^{21\text{-}29}$ peptide for p107ΔL (K_d = 0.16 ± 0.04 μM). S28 phosphorylation in the context of the S20C mutation still increases affinity (LIN52$^{12\text{-}34;S20C;phosS28}$ K_d = 0.021 ± 0.004 μM) such that it is tighter than the wild-type phosphorylated LIN52 sequence.

No binding was observable upon mixing of either LIN52$^{12\text{-}34}$ or LIN52$^{12\text{-}34;phosS28}$ peptide with Rb. The LIN52$^{12\text{-}34;S20C}$ peptide (K_d = 0.3 ± 0.1 μM) binds Rb with similar affinity as p107, however phosphorylation results in no additional increase in affinity (LIN52$^{12\text{-}34;S20C;phosS28}$ K_d = 0.3 ± 0.1 μM). This result corroborates the lack of a phosphate-binding pocket in Rb. We also mutated the residues in p107ΔL that contact LIN52 phosS28 in the crystal structure such that they resemble the Rb sequence (R869K, S876Q, Y879F, and R880K). LIN52 and LIN52$^{12\text{-}34;S20C}$ bind the Rb-like mutant (p107^{mutant}) with weaker affinity compared to p107ΔL, and phosphorylation does not enhance affinity (Fig. 33b). The E7$^{21\text{-}29}$ peptide binds p107^{mutant} and p107ΔL with similar affinity, which is consistent with a lack of interaction between E7 and the phosphate pocket.

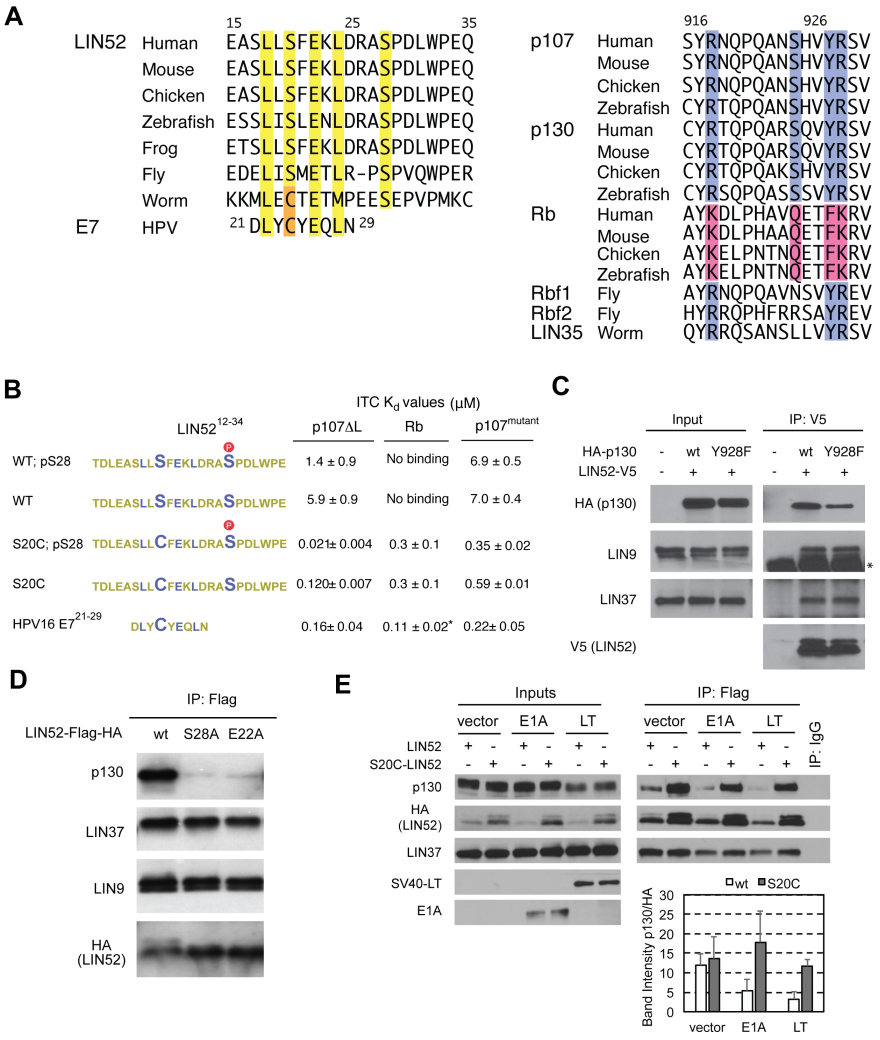


Figure 33: Role of the 'LxSxExL' motif and S28 phosphorylation on LIN52-p107/p130 association. (A) Alignment of LIN52 (left) and pocket protein (right) sequences. The LIN52 'LxSxExLxxxpS' motif and S28 phosphate binding residues are highlighted. (B) ITC affinity measurements of $LIN52^{12-34}$ variants and $E7^{21-29}$ binding to pocket proteins. *Value previously reported (Lee et al. 1998). (C) T98G cells were co-transfected with LIN52-V5 and the indicated HA-p130, extracts were immunoprecipitated with an anti-V5 antibody, and proteins were detected with the appropriate antibodies in a Western blot. The band in the LIN9 blot marked with an asterisk is nonspecific. (D) T98G cells stably expressing the wild-type or mutant LIN52-Flag-HA alleles were serum starved for 48 hours, and the co-immunoprecipitation was performed as in (C) except with an anti-Flag antibody. (E) As in (C), wild-type (WT) and S20C mutant LIN52-Flag-HA and viral proteins SV40 Large T antigen (LT) and adenovirus E1A were transiently expressed in cycling T98G cells as indicated. Bands intensities for the immunoprecipitated HA-LIN52 and p130 were quantified, and the p130/LIN52 ratio is plotted. Error bars are standard deviations for two different replicates of the experiment.

Peptide	p107ΔL affinity
$LIN52^{12–34;phosS28}$	1.4 μM ± 0.8 μM
$LIN52^{12–34;L18A;phosS28}$	14.9 μM ± 0.1 μM
$LIN52^{12–34;L24A;phosS28}$	7.4 μM ± 0.9 μM
$LIN52^{12–34;E22A;phosS28}$	5.5 μM ± 0.3 μM
$LIN52^{12–34;+A28;phosS29}$	1.7 μM ± 0.3 μM
$LIN52^{12–34;–A27;phosS27}$	0.126 μM ± 0.002 μM
$LIN52^{12–34;P29A;phosS28}$	0.035 μM ± 0.003 μM

Table 4: Affinity for LIN52 peptides for the p107 pocket domain

The fact that $LIN52^{12-34}$ and $LIN52^{12-34;phosS28}$ bind $p107^{mutant}$ with greater affinity than Rb suggests that p107-specific interactions other than the phosphate-binding pocket stabilize LIN52 association. One possible explanation observed in the crystal structure is the presence of I850 in p107. I850 contacts Y849 and stabilizes the Y849 position for hydrogen bonding to S20 in LIN52 (Fig. 32). In contrast, G710 at the corresponding position in Rb cannot play the same stabilizing role and is likely a source of the lower Rb affinity for the noncanonical 'LxSxExL' motif relative to p107/p130.

We tested the significance of the p107/p130 specific phosS28-binding site for DREAM assembly in cells (Fig. 33c). T98G cells were co-transfected with V5-tagged LIN52 and either wild-type HA-p130 or HA-p130 with a Y928F mutation (equivalent to Y879F in p107). LIN52-V5 was precipitated with an anti-V5 antibody, and we examined the amount of precipitated p130 by Western blot. LIN52 co-precipitated less Y928F mutant p130 protein than wild-type, supporting the requirement of an intact phosphate-binding pocket as observed in the crystal structure for DREAM assembly.

5.2.4 the 'LxSxExL' sequence motif is critical for DREAM assembly

To explore further the important elements of the LIN52 'LxSxExL' binding motif, we measured the affinity of different $LIN52^{12-34}$ peptides for the p107 pocket domain using the

ITC assay. Both the L18A (K_d = 14.9 ± 0.1 μM) and L24A (K_d = 7.4 ± 0.9 μM) consensus mutations in LIN52 reduce affinity. Although E22 in LIN52 does not make any interactions with p107 in the crystal structure (Fig. 31), we also find that an E22A mutation (K_d = 5.5 ± 0.1 μM) reduces affinity 4-fold compared to wild-type (K_d = 1.4 ± 0.8 μM). We suggest that in solution there is an additional alternate conformation in equilibrium that resembles the E7 peptide conformation but is less stable because of the buried S20 hydroxyl group.

We also measured the effects of moving the position of phosS28 relative to the 'LxSxExL' motif. Addition of an alanine to the intervening three residues results in similar affinity ($LIN52^{12\text{-}34;+A28;phosS29}$; K_d = 1.7 ± 0.3 μM), while the peptide with A27 deleted binds with 10-fold tighter affinity ($LIN52^{12\text{-}34;-A27;phosS27}$; K_d = 0.126 ± 0.002 μM). This observation is consistent with a lack of interactions between the intervening sequence and pocket domain (Fig. 31). The tighter affinity of the shorter sequence may result from a reduced entropic penalty for binding. Many kinases phosphorylate substrates bearing an (S/T)P motif, which include DYRK1A and the CDKs (Ubersax and Ferrell 2007). We tested a LIN52 phosS28 P29A mutant to determine if p107 binding is also dependent upon this motif. The mutant has a higher affinity ($LIN52^{12\text{-}34;P29A;phosS28}$; K_d = 0.035 ± 0.003 μM), than wild-type, suggesting that the proline is required for kinase recognition of S28 but is not preferred at this position for pocket binding.

To determine if the p130 pocket association with the LIN52 'LxSxExL' motif is required for DREAM assembly in cells, we generated LIN52-HA-FLAG mutants stably expressed in T98G cells. We then immunoprecipitated the FLAG sequence and examined the co-precipitation of p130 by Western blot. When either E22 or S28 is mutated in LIN52, p130 binding is diminished while binding of the MuvB proteins LIN37 and LIN9 is preserved (Fig. 33d). We also tested the stabilizing effect of an S20C mutant on the LIN52-p130 complex by co-precipitation (Fig. 33e). The LIN52-HA-FLAG S20C mutant expresses at a higher level than wild-type but does not co-precipitate a relative greater amount of p130. However,

when co-expressed with either E1A or large T-antigen viral proteins, both of which contain 'LxCxExf' sequences, LIN52 S20C does co-precipitate a relative greater amount of p130 than wild-type. We conclude that the wild-type sequence is sufficient to co-precipitate p130 but when challenged with high affinity viral proteins, the increase in affinity from the S20C mutation allows LIN52 to better compete with E1A and large T-antigen for p130 binding. This result emphasizes how the weaker 'LxSxExL' motif in LIN52, even with the addition of S28 phosphorylation, enables DREAM disassembly by competitor viral oncoproteins.

5.2.5 B-Myb requires LIN52 to bind MuvB but does not compete with p130

We next investigated the determinants of B-Myb binding to MuvB in order to compare the mechanisms of DREAM and MMB assembly. The evolutionarily conserved MuvB proteins in the *Drosophila melanogaster* DREAM complex were found to interact with the C-terminus of B-Myb (Andrejka et al. 2011). To test whether this association is direct and conserved in the human proteins, we performed a series of precipitation experiments with recombinant B-Myb and MuvB proteins from Sf9 cells (Fig. 34). We reconstituted the core MuvB subcomplex by co-expressing GST-LIN9$^{94-542}$, which lacks its poorly conserved and putative unstructured N-terminus, along with full-length GST-LIN37, Strep-RBAP48, and GST-LIN52 as indicated. The purified complexes were then co-precipitated in the presence of the C-terminus of B-Myb (GST-B-Myb$^{654-700}$). We find that GST-B-Myb$^{654-700}$ binds the MuvB subcomplex only in the presence of LIN52 (Fig. 34a). However, full-length GST-LIN52-HIS, which is sufficient to co-precipitates p130 (Fig. 30a), is not sufficient to co-precipitate GST-B-Myb$^{654-700}$ (Fig. 34B). We conclude that LIN52 is necessary but not sufficient for B-Myb binding. The requirement for LIN52 could result from direct B-Myb-LIN52 contacts or because LIN52 is necessary for a properly structured MuvB core.

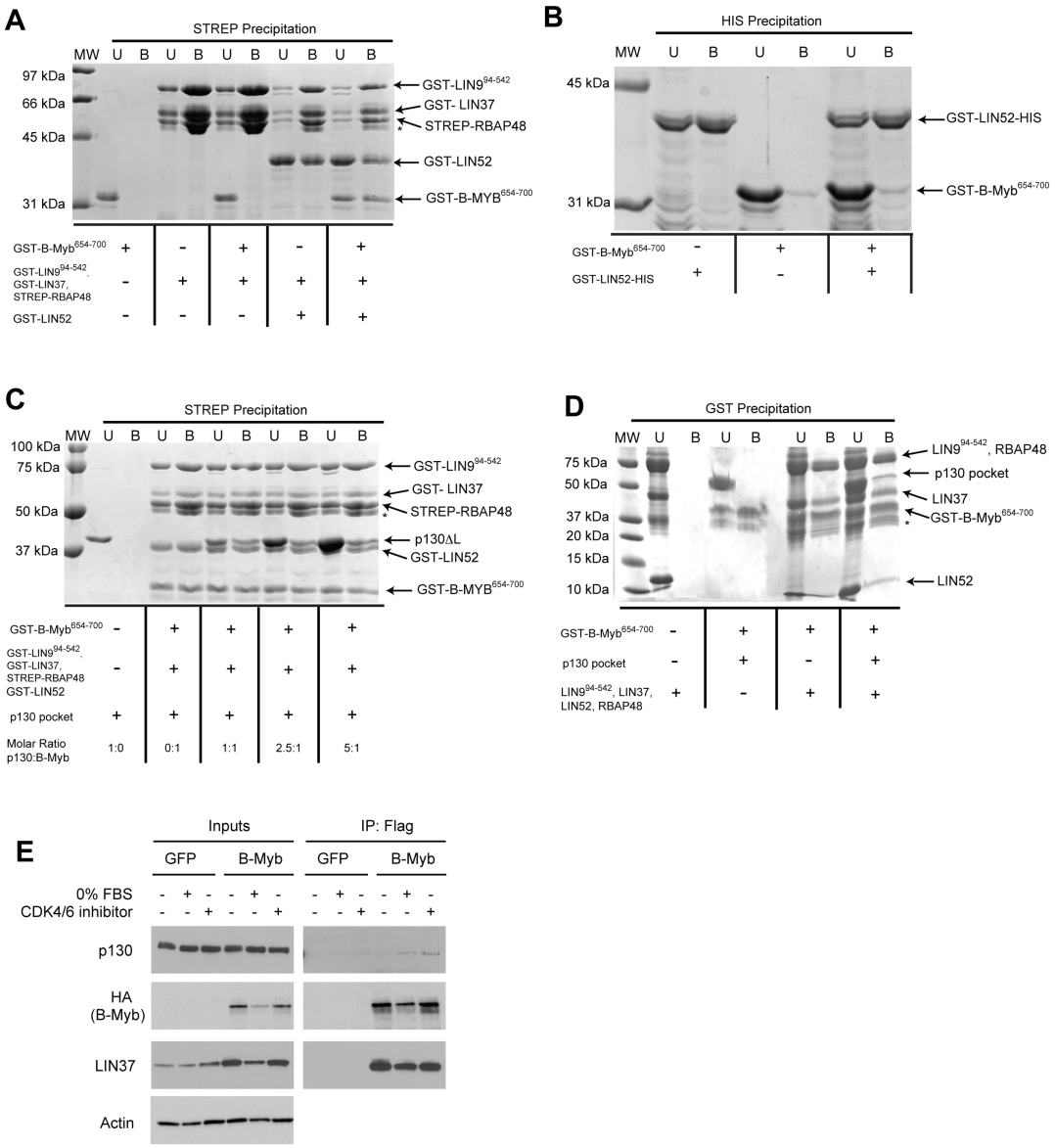


Figure 34: B-Myb and p130 bind distinct surfaces of MuvB. Co-precipitation experiments were conducted with recombinant proteins as in Fig. 30. (A) Streptactin co-precipitation of Step-tagged RbAP48, GST-LIN9$^{94-542}$, GST-LIN37 with GST-B-Myb$^{654-700}$ in the absence and presence of GST-LIN52 demonstrates the requirement of LIN52 for B-Myb-MuvB binding. The band marked with an asterisk is a GST-LIN9$^{94-542}$ degradation product. (B) Ni^{2+}-NTA fails to co-precipitate GST-B-Myb$^{654-700}$ with GST-LIN52-HIS alone. (C) Streptactin co-precipitation of Step-tagged RbAP48, GST-LIN9$^{94-542}$, GST-LIN37, GST-LIN52 with GST-B-Myb$^{654-700}$ in the presence of increasing amounts of p130DL. The amount of co-precipitated GST-B-Myb$^{654-700}$ does not change, demonstrating that the p130 pocket domain does not compete with B-Myb for MuvB binding. The band marked with an

asterisk is a GST-LIN9$^{94-542}$ degradation product. (D) GST-B-Myb$^{654-700}$ co-precipitates the untagged p130 pocket domain only in the presence of the untagged MuvB components, indicating both B-Myb and p130 can simultaneously bind MuvB. The band marked with an asterisk are degradation products of GST-B-Myb$^{654-700}$. (E) Association of p130 and MuvB with ectopically expressed B-Myb. BJ-hTERT fibroblasts stably expressing HA-Flag tagged GFP (control) or B-Myb were incubated for 24h in complete medium, in the medium without FBS, or in medium containing CDK4/6 inhibitor. Proteins of interest were detected by Western blot in the cell extracts (Inputs) and in the anti-Flag pull-downs (IP:FLAG).

We next tested whether p130 and B-Myb compete for access to MuvB or whether they can both simultaneously bind. We find that Strep-tagged RbAP48, together with the other components of the MuvB subcomplex, co-precipitates both GST-B-Myb$^{654-700}$ and GST-p130 pocket when all components are mixed in the same solution (Fig. 34c). Increasing the amount of GST-p130 does not change the amount of GST-B-Myb$^{654-700}$ that co-precipitates, consistent with a lack of competition (Fig. 34c). We also find that GST-B-Myb$^{654-700}$ co-precipitates p130 pocket if and only if the components of MuvB are present, demonstrating the presence of a complex that contains B-Myb$^{654-700}$, p130 pocket, and MuvB all together (Fig. 34d). Based on these observations that p130 and B-Myb can simultaneously bind MuvB and that they have different requirements for binding, we conclude that they have non-overlapping binding sites on the MuvB core. In order to demonstrate that full-length p130 and B-Myb associate together with MuvB in cells, we precipitated HA-Flag-B-Myb that is stably expressed in BJ-hTERT fibroblasts (Fig. 34e). We observe co-precipitation of endogenous LIN37 and p130 under conditions that support DREAM assembly, including serum starvation and CDK4/6 inhibition with palbociclib (PD-0332991, Pfizer, Inc.). Although we cannot rule out a direct association between full-length p130 and B-Myb in this experiment, the data are consistent with B-Myb co-precipitating p130 through a common MuvB core.

5.2.6 CDK phosphorylation promotes DREAM disassembly

Components of the MuvB complex fail to associate with p107 and p130 in cycling cells, however the mechanism of DREAM inhibition upon exit from quiescence has not been elucidated. Considering that LIN52 is found phosphorylated and unphosphorylated at S28 when bound to B-Myb during S phase (Litovchick et al. 2011), dephosphorylation of LIN52 S28 is unlikely the primary mechanism promoting DREAM disassembly. Our result that B-Myb and p130 bind MuvB simultaneously also argues against a competitive mechanism for DREAM disassembly whereby upon expression, B-Myb displaces p130 from MuvB. Noting the coincident timing of CDK activation and DREAM disassembly (Pilkinton et al. 2007), we tested directly whether CDK activity correlated with p130 binding to MuvB. We find that the MuvB component LIN37 can co-precipitate a greater amount of p130 from the extracts of cycling T98G cells in the presence of the CDK4/6 inhibitor (Fig. 35A). To observe the effect of CDK inhibition on DREAM disassembly upon cell cycle entry, we released T98G cells from serum starvation and found that LIN37 co-precipitation of p130 persists to a greater extent in the presence of the CDK4/6 inhibitor (Fig. 35b). In the absence of inhibitor, the loss of p130 co-precipitation, beginning at 10 hours, correlates with the disappearance of the hypophosphorylated p130 form (lower band in gel). These results demonstrate that CDK activity in cells inhibits p130-MuvB association.

We asked whether Cyclin D1 could inhibit DREAM assembly by directly competing with the LIN52 interaction at the p130 LxCxE cleft. Cyclin D1 was previously proposed to bind the LxCxE cleft in the Rb pocket domain (Dowdy et al. 1993). However, we found in our FP assay (Fig. 30f) that a Cyclin D1 peptide does not displace LIN52 form p130DL. We also titrated the Cyclin D1 peptide bearing its 'LxCxExE' motif into p107DL and p130DL in the ITC assay and found no observable binding, suggesting the Cyclin D1 sequence does not directly associate with the pocket domain. These results are supported by our ITC data

and previous reports (Singh et al. 2005) that a hydrophobic residue is necessary at the downstream glutamic acid position in Cyclin D1.

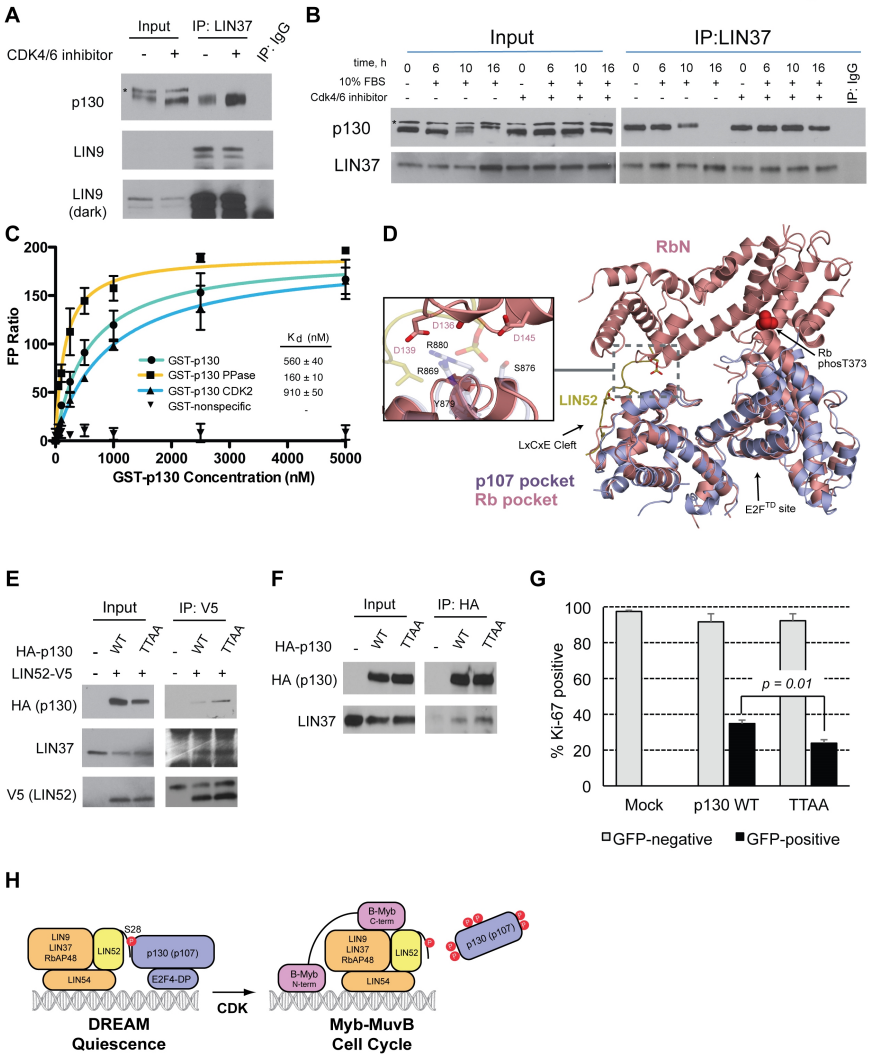


Figure 35: p130 phosphorylation weakens LIN52 binding and DREAM activity
(A) T98G cells were incubated for 24h in the presence or absence of CDK4/6 inhibitor. Binding of p130 to LIN37 was determined by co-immunoprecipitation with an anti-LIN37 antibody and detection of proteins in a Western blot. The band marked with an asterisk in the p130 input blot is a nonspecific band. (B)T98G cells were serum starved for 72 hours

and then released with addition of 10% FBS in the absence or presence of the CDK4/6 inhibitor palbociclib. After harvesting cells at the indicated time points, extracts were immunoprecipitated with an anti-LIN37 antibody and probed for LIN37 and p130. The band marked with an asterisk in the p130 input blot is a nonspecific band. (C) Fluorescence polarization binding assay of TMR-labeled LIN52$^{12\text{-}34;phosS28}$ to GST-p130. The GST-p130 is purified from sf9 cells and used untreated (green, circles), treated with lambda phosphatase (yellow, squares), and treated with CDK2 kinase (blue, triangles). Purified GST alone was used as a negative control (inverted triangles). The expected effect of enzyme treatment is corroborated with a Phos-tag gel (Supplemental Figure 1C*). (D) Overlay of the LIN52$^{12\text{-}34;phosS28}$-p107 structure and phosphorylated Rb$^{55\text{-}787}$ structure (PDB: 4ELJ). Residues in p107 that bind the S28 phosphate in LIN52 (rendered as transparent blue sticks) are analogous to residues in the Rb pocket domain that bind the docked RbN domain (pink) when Rb is phosphorylated. (E) U2-OS cells were transfected with LIN52-V5 and the wild-type or T401A/T417A (TTAA) mutant HA-p130. Extracts were immunoprecipitated with anti-V5 antibody and the indicated proteins were detected by Western blot. (F) Same as (E) except an anti-HA antibody was used for immunoprecipitation. (G) U2OS cells were co-transfected with the wild-type or T410A/T417A (TTAA) mutant of p130 and GFP as tracer. The expression of Ki-67 was determined by indirect immunofluorescence cell staining at 48 h post-transfection. The graph shows the average values and standard deviations (error bars) of 3 replicate experiments in which at least 100 cells were counted per condition. The p-value evaluating statistical significance was calculated using a two-tailed student's t-test. (H) Model for DREAM complex disassembly and Myb-MuvB complex assembly upon CDK activation.

We hypothesized that CDK phosphorylation of p130/p107 inhibits the LIN52 interaction with the p130/p107 pocket domain. In support of this mechanism, we found that lambda phosphatase treatment of the partially phosphorylated p130 purified from Sf9 cells increased LIN52-p130 complex stability in a qualitative co-precipitation assay (Fig. 30a). We quantified the effect of p130 phosphorylation on LIN52 binding using the fluorescence polarization (FP) assay described in Fig. 1E. The TMR-LIN52$^{12\text{-}34;phosS28}$ peptide was mixed with increasing amounts of GST-p130 purified from Sf9 cells without treatment, with phosphatase treatment, and with kinase treatment. While phosphatase treatment increases the affinity 3.5-fold, CDK2 phosphorylation decreases the affinity nearly 2-fold, demonstrating that p130 phosphorylation weakens LIN52 binding (Fig. 35c). The observation of a change in affinity upon treatment with either enzyme reflects the partial phosphorylation of GST-p130 purified from Sf9 cells.

Our structural observation that phosphorylated LIN52 binds p107/p130 at the 'LxCxE' cleft suggests that CDK phosphorylation events resulting in occlusion of the 'LxCxE'-site would inhibit DREAM assembly. Studies of inactive Rb have revealed phosphorylation-induced structural changes that block the 'LxCxE' cleft (Rubin et al. 2005; Burke et al. 2012; Rubin 2013). We considered one mechanism in which phosphorylation of conserved sites in the linker between the N-terminal (RbN) and pocket domains results in RbN-pocket docking (Burke et al. 2012). Alignment of the phosphorylated Rb structure with the p107-LIN52 complex suggests that an analogous closed conformation in p107/p130 would inhibit LIN52 binding (Fig. 35d). In the Rb structure, residues in RbN form hydrogen bonds and salt bridge interactions with Rb pocket residues; the homologous residues in the p107 pocket domain contact the LIN52 S28 phosphate in the structure presented here.

To determine whether CDK phosphorylation disrupts the DREAM complex through phosphorylation of sites in the putative p107/p130 interdomain linker, we examined the effect of p130 T401A and T417A mutation on MuvB binding in cycling U2OS cells. We expressed LIN52 together with p130 wild-type and phosphosite mutant and found that LIN52 co-precipitates a greater amount of the T401A/T417A double mutant (Fig. 35e). We also found that the p130 double mutant can co-precipitate a great amount of endogenous LIN37 than wild-type (Fig. 35f). In order to test the functional consequences of mutating CDK sites, we transfected U2OS cells with wild-type p130 and the T401A/T417A phosphosite mutant and examined proliferation of cells using S-phase-specific Ki-67 staining similar to as previously described (Canhoto et al. 2000). We find that the T401A/T417A mutant induces a more potent growth arrest than wild-type, which is consistent with stabilization of the DREAM complex (Fig. 35g). T401 and T417 phosphorylation has also been shown to be relevant for regulation of E2F4 activity (Farkas et al. 2002). Mutation of these sites may also influence p130-E2F4 affinity and contribute

to the observed growth arrest by stabilizing E2F binding.

5.3 Discussion

Our structural and binding data support a model in which the DREAM complex is mediated by the direct association of LIN52 with the p107/p130 pocket (Fig. 35h). Assembly is promoted by DYRK1A phosphorylation of S28 in LIN52, while disassembly is induced by CDK phosphorylation of sites in p107/p130 that promote structural changes to occlude the LIN52 binding surface. B-Myb binding is dependent on LIN52 and other members of the MuvB core and can occur simultaneously with p107/p130 binding in our assays with recombinant protein and in cells under conditions in which B-Myb is ectopically expressed and p130 is known to be hypophosphorylated. The fact that B-Myb and p130 are not found together in complexes purified from mammalian cells is likely due to the coincidence of B-Myb expression and p130 phosphorylation in those cells (Litovchick et al. 2007). Complexes have been purified from Drosophila cells containing both Myb and Rbf and may point to a distinct mechanism of regulation in lower metazoans. The *D. melanogaster* Rbf sequences notably lack conserved phosphorylation sites equivalent to T401 and T417 in p130, which further points to their importance in disassociating p130 from MuvB.

LIN52 accesses the 'LxCxE' cleft of the pocket domain using a non-canonical and non-optimal 'LxSxExL' motif in combination with phosphorylation at S28. The weak affinity of the 'LxSxExL' sequence sensitizes the association to regulatory phosphorylation by DYRK1A. Kinase phosphorylation acts as a switch, because it increases LIN52 binding to submicromolar affinity. We note that in contrast to higher orthologs, *C. elegans* LIN52 maintains the canonical, tight-binding LxCxE motif and lacks the equivalent DYRK1A consensus surrounding the S28 residue. We predict that DREAM in *C. elegans* is not regulated equivalently by phosphorylation, and LIN52 and LIN35 constitutively form a

stable complex. The LIN52 sequence in *D. melanogaster* resembles our LIN52$^{12\text{-}34;DA27;phosS27}$ peptide, suggesting Drosophila Rbf proteins and LIN52 also form a highly stable, phosphorylation-dependent interaction.

The requirement for LIN52 phosphorylation is an important source of specificity for p107 and p130 in DREAM assembly. Unlike Rb, the p107 and p130 pocket domains contain the proper phosphate-binding site. Phylogenetic sequence analysis of pocket proteins suggests that Rb appears more recently than p107 and p130, which are closer in sequence to pocket proteins such as LIN35 and Rbf in lower metazoans (van den Heuvel and Dyson 2008; Sadasivam and DeCaprio 2013). It is not clear then why Rb evolved such that it does not bind MuvB, although one possibility is that the LxCxE cleft in Rb must be reserved for some alternate unique function.

Despite their diverse cellular roles, the universal biochemical function of pocket proteins is to assemble protein complexes (Cobrinik 2005; Dick and Rubin 2013). It is estimated that several hundred proteins interact with Rb and its pocket protein paralogs, and the 'LxCxE' cleft has been described as the key interface for these complexes (Morris and Dyson 2001). The crystal structure presented here of the p107-LIN52 complex provides the first structural insights into how cellular proteins contact the cleft. It implicates the 'LxSxExLxxxpS' sequence as a novel binding motif that can be regulated by phosphorylation. We searched the Scansite 3 server (Obenauer et al. 2003) to identify other potential p107/p130 binding partners that use this motif and the motifs identified in our binding studies that bind the novel phosphate-binding pocket. We identified 203 proteins that matched our motifs, although only 13 are known to be phosphorylated on the consensus serine (Supplemental Table 1*).

CDK phosphorylation has been well characterized as a mechanism for inactivating pocket proteins, and both structural and functional studies have implicated the effects of phosphorylation on inhibiting pocket protein interactions (Cobrinik 2005; Rubin 2013). Our

results here further support phosphorylation of p107/p130 as a mechanism of inactivating the DREAM complex by dissociating pocket proteins from the MuvB core. Mutation of conserved CDK sites T401 and T417 enhances the growth arrest properties of p130 and stabilizes the DREAM complex (Fig. 35). Structural and sequence comparison between p107/p130 and Rb suggests that a conserved closed conformation that occludes the 'LxCxE' cleft is induced by phosphorylation. We note that additional CDK sites in p107/p130 may contribute to inhibition of LIN52 binding and that phosphorylation may also contribute to DREAM dissociation by inhibiting p107/p130 association with E2F proteins.

It has been suspected that cancer cells can in certain contexts avert the effect of cytotoxic treatments by entering quiescence. As an important mediator of cell cycle exit, DREAM may be an important additional target during chemotherapy. For example, cells from gastrointestinal stromal tumors enter quiescence following treatment with imatinib, and inhibition of DREAM assembly through genetic knockdown alternatively induces apoptosis (Boichuk et al. 2013). Our structural characterization of LIN52 association with p107 suggests that inhibitors of protein interactions at the LxCxE cleft in pocket proteins would likely block DREAM function *in vivo*. It is encouraging that only a short peptide corresponding to the E7 sequence is sufficient for inhibition of LIN52 association *in vitro* (Fig. 30f), and small molecule inhibitors of the E7 LxCxE-Rb interaction have been reported (Fera et al. 2012). Discovery of more potent chemical inhibitors of LxCxE-pocket binding may prove to be a viable strategy for preventing cancer cells from escaping to quiescence as a mechanism for surviving therapeutic treatment.

5.4 Materials and Methods

5.4.1 Protein expression and peptides

Human Rb pocket with its loop deleted (residues N380-G581; K643-R787), p107 pocket (T391-Q972), p107DL (T391-T599; N780-I887; K924-Q972) and p130 DL (H424-

D632; N828-K935; E999-Q1049) were expressed and purified from *E. coli* BL21 cells as GST fusion proteins. Cells were induced with 1 mM IPTG and grown overnight at 20°C. Lysates were first purified by GS4B affinity chromatography. The elution fraction was then subjected to TEV protease cleavage and dialyzed overnight in 25 mM Tris, 200mM NaCl, 1 mM DTT, 0.5 mM EDTA pH 8.0. The protein was then passed over GS4B affinity resin again to remove free GST, concentrated, and stored in a buffer containing 20 mM Tris, 200mM NaCl, 1 mM DTT, and 20% glycerol pH 8.0.

LIN52, LIN37, RBBP48, LIN9, B-Myb and full-length p130 were expressed and purified from Sf9 cells (Expression Systems, Davis CA) using baculovirus vectors. Proteins were purified as described above, and the GST-free samples were passed over a Superdex-200 column that was equilibrated in the dialysis buffer following TEV cleavage. p130 was phosphatase treated with 1% lambda phosphatase by mass using the manufacturer protocol (New England Biolabs). CDK2 treatment was performed as described (Burke et al. 2010), except the CDK2 activator Speedy was used instead of a Cyclin activator. Following treatment, GST-p130 was separated from the enzymes using a Superdex 200 column.

The LIN52$^{12\text{-}34;\text{phosS28}}$ and E7$^{21\text{-}29}$ peptides used for crystallization were synthesized by BioPeptide LLC., San Diego CA. The LIN52$^{12\text{-}34;\text{P29A;phosS28}}$ peptide and all other peptides used for ITC and the FP assay were synthesized by GenScript Inc. Piscataway NJ.

5.4.2 Crystallization, data collection, structure determination, and model refinement

p107DL was prepared for crystallization by elution from a Superdex 75 (GE Healthcare) column in a buffer containing 25 mM Tris, 500 mM NaCl, and 5 mM DTT, pH 8.0. LIN52$^{12\text{-}34;\text{phosS28}}$ was added in 3-fold molar excess to 14 mg/mL p107DL, and E7$^{21\text{-}29}$ was added in 2-fold molar excess to 12 mg/mL p107ΔL. After incubation on ice for 30 minutes, both complexes were crystallized by sitting-drop vapor diffusion at 4°C. Plates formed after two weeks in 100 mM MES pH 6.5, 1.6 M $(NH_4)_2SO_4$, and 4% PEG 400 for the

LIN52$^{12\text{-}34;phosS28}$ complex and in 100 mM HEPES pH 7.5, 1.7 M $(NH_4)_2SO_4$, and 4% PEG 400 for the E7$^{21\text{-}29}$ complex. Crystals were frozen in the proper well buffer with 20% ethylene glycol. Some crystals containing the LIN52$^{12\text{-}34;phosS28}$ peptide were soaked with 20-fold molar excess of the P29A mutant peptide before freezing.

Data were collected at the Advanced Photon Source, Argonne National Laboratory at Beamline 23-IDB. Diffraction spots were integrated using MOSFLM (Leslie 2006), and data were merged and scaled using Scala (Bailey 1994). Phases were first solved for the E7$^{21\text{-}29}$ complex by molecular replacement using PHASER (Mccoy et al. 2007). A homology model of p107ΔL, which was constructed using the Rb pocket domain (PDB code: 3POM), was used as a search model. p107ΔL complexes with LIN52 peptides were then solved using the E7-p107 complex as a search model as the crystal form was slightly different. Peptides were built with Coot (Emsley and Cowtan 2004), and the models were refined with Phenix (Adams et al. 2010). Coordinates and structure factors have been deposited in the PDB for the structures with the E7$^{21\text{-}29}$, LIN52$^{12\text{-}34;phosS28}$, and LIN52$^{12\text{-}34;P29A;phosS28}$ peptides under accession codes 4YOZ, 4YOS, and 4YOO, respectively.

5.4.3 Co-precipitation assays

For assays with recombinant protein, 50 mg of each indicated protein was incubated with 15 µL of either GS4B resin, Streptactin resin, or Nickel Sepharose Excel resin (all from GE Healthcare) in a buffer containing 25 mM Tris, 150 mM NaCl, 1 mM DTT, pH 8.0 on ice for 30 minutes. The resin was washed three times using the same buffer or with the addition of 20 mM imidazole, and eluted using 30 µL of elution buffer containing 25 mM Tris, 200 mM NaCl, 1 mM DTT, pH 8.0 with either 20 mM glutathione, 400 mM imidazole, or 10 mM desthiobiotin as needed. Under these conditions, we do not find that GST dimerization influences our results. For example, in Fig. 34a, GST-B-Myb$^{654\text{-}700}$ only

co-precipitates with other GST-tagged components of the MuvB core when LIN52 is present. Samples were analyzed by SDS-PAGE with Coomassie staining.

For assays in cells, stable cell lines were generated using a pMSCV-FLAG-HA retroviral vector as described (Litovchick et al. 2007). Alternatively, cells were transiently transfected with pcDNA3.1 vectors encoding HA tagged wild-type or mutant p130 alleles, or pBabe vectors encoding SV40 large T antigen or adenovirus E1A. Mutant LIN52 and p130 alleles were obtained using site-directed mutagenesis (QuikChange, Agilent). T98G and BJ-hTERT cells were chosen for assaying DREAM assembly because they have intact Rb family genes, they have significant G1 populations in asynchronously growing cells, and they arrest in G0/G1 readily (Litovchick et al. 2007). U2OS cells are highly proliferative and sensitive to overexpression of Rb family proteins; therefore, they were used for characterizing p130 phosphorylation site mutants as described (Farkas et al. 2002). Cells were extracted using EBC lysis buffer (50 mM Tris, 150 mM NaCl, 0.5% NP-40, pH 7.4), immunoprecipitated and subjected to western blot analysis as described previously (Litovchick et al. 2007). Band intensities were quantified using Image J software.

5.4.4 Proliferation assay

U2OS cells were seeded onto six-well plates at 2 x 10^5 cells per well and transfected with pcDNA3.1 vectors encoding p130 constructs and GFP. The p130 vector was transfected at a 4:1 molar ratio relative to the GFP vector. 48 hours post-transfection, cells were trypsinized and seeded in triplicate onto 6 well plates containing glass coverslips and allowed to attach for 24 hours. Proliferating GFP-positive and GFP-negative cells were detected by indirect immunofluorescence staining using anti-Ki67 (Millipore) antibodies as described in (Canhoto et al. 2000). To determine the fraction of proliferating cells in each sample, the number of Ki67-positive cells was determined in at least 100 cells per condition using EVOS fluorescent microscope (AMG advanced microscope group) and 20X

objective. U2OS cells were chosen for this assay because they have intact Rb family members and because they are highly proliferative and thus sensitive to increases in the DREAM arrest activity.

5.4.5 Calorimetry

Equilibrium dissociation constants for p107, p130, Rb, LIN52, and E7 binding were obtained using isothermal titration calorimetry with the Micro Cal VP-ITC system (GE Healthcare). Peptides and proteins were dialyzed overnight and titrated into a buffer containing 25 mM Tris, 200 mM NaCl, and 5 mM BME, pH 8.0 at 15°C. Because of the difficulty in accurately determining peptide concentration, the peptide concentration was adjusted in the fitting such that all the stoichiometry values were equal to 1.0.

5.4.6 Fluorescence polarization assay

TMR-LIN52$^{12\text{-}34;phosS28}$ was mixed at 10 nM with p130DL GST-p130 in a buffer containing 40 mM Tris, 150 mM NaCl, 1 mM DTT, 0.1% Tween, pH 8.0. 20 µL of the reaction was used for the measurement in a 384-well plate well. FP measurements were made in triplicate using a Perkin-Elmer EnVision plate reader, and reported FP values were determined using instrument software. Error bars in the plots show standard deviations for the three measurements of each point, while the reported errors in K_d and K_I are errors derived from curve fits. We note that absolute K_d measurements were typically tighter using the FP assay compared to ITC and suggest that this difference may be due to the presence of the dye in the peptide or other subtle differences in experimental conditions.

6.1 References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221.

Andrejka, L., Wen, H., Ashton, J., Grant, M., Iori, K., Wang, A., . . . Lipsick, J. S. (2011). Animal-specific C-terminal domain links myeloblastosis oncoprotein (Myb) to an ancient repressor complex. *Proc Natl Acad Sci U S A, 108*(42), 17438-17443.

Araki, K., Nakajima, Y., Eto, K., & Ikeda, M. A. (2003). Distinct recruitment of E2F family members to specific E2F-binding sites mediates activation and repression of the E2F1 promoter. *Oncogene, 22*(48), 7632-7641.

Ashizawa, S., Nishizawa, H., Yamada, M., Higashi, H., Kondo, T., Ozawa, H., . . . Hatakeyama, M. (2001). Collective inhibition of pRB family proteins by phosphorylation in cells with p16INK4a loss or cyclin E overexpression. *J Biol Chem, 276*(14), 11362-11370.

Ashizawa, S., Nishizawa, H., Yamada, M., Higashi, H., Kondo, T., Ozawa, H., . . . Hatakeyama, M. (2001). Collective inhibition of pRB family proteins by phosphorylation in cells with p16INK4a loss or cyclin E overexpression. *J Biol Chem, 276*(14), 11362-11370.

Bagchi, S., Weinmann, R., & Raychaudhuri, P. (1991). The retinoblastoma protein copurifies with E2F-I, an E1A-regulated inhibitor of the transcription factor E2F. *Cell, 65*(6), 1063-1072.

Bailey, S. (1994). The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography, 50*, 760-763.

Balog, E. R., Burke, J. R., Hura, G. L., & Rubin, S. M. (2011). Crystal structure of the unliganded retinoblastoma protein pocket domain. *Proteins, 79*(6), 2010-2014.

Bandara, L. R., & La Thangue, N. B. (1991). Adenovirus E1a prevents the retinoblastoma gene product from complexing with a cellular transcription factor. *Nature, 351*(6326), 494-497.

Binne, U. K., Classon, M. K., Dick, F. A., Wei, W., Rape, M., Kaelin, W. G., Jr., . . . Dyson, N. J. (2007). Retinoblastoma protein and anaphase-promoting complex physically interact and functionally cooperate during cell-cycle exit. *Nat Cell Biol, 9*(2), 225-232.

Boichuk, S., Parry, J. A., Makielski, K. R., Litovchick, L., Baron, J. L., Zewe, J. P., . . . Duensing, A. (2013). The DREAM complex mediates GIST cell quiescence and is a novel therapeutic target to enhance imatinib-induced apoptosis. *Cancer Res, 73*(16), 5120-5129.

Brehm, A., & Kouzarides, T. (1999). Retinoblastoma protein meets chromatin. *Trends Biochem Sci, 24*(4), 142-145.

Buchkovich, K., Duffy, L. A., & Harlow, E. (1989). The retinoblastoma protein is

phosphorylated during specific phases of the cell cycle. *Cell, 58*(6), 1097-1105.

Burke, J. R., Deshong, A. J., Pelton, J. G., & Rubin, S. M. (2010). Phosphorylation-induced conformational changes in the retinoblastoma protein inhibit E2F transactivation domain binding. *J Biol Chem, 285*(21), 16286-16293.

Burke, J. R., Hura, G. L., & Rubin, S. M. (2012). Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control. *Genes Dev, 26*(11), 1156-1166.

Burke, J. R., Liban, T. J., Restrepo, T., Lee, H. W., & Rubin, S. M. (2014). Multiple mechanisms for E2F binding inhibition by phosphorylation of the retinoblastoma protein C-terminal domain. *J Mol Biol, 426*(1), 245-255.

Burkhart, D. L., & Sage, J. (2008). Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer, 8*(9), 671-682.

Calbo, J., Parreno, M., Sotillo, E., Yong, T., Mazo, A., Garriga, J., & Grana, X. (2002). G1 cyclin/cyclin-dependent kinase-coordinated phosphorylation of endogenous pocket proteins differentially regulates their interactions with E2F4 and E2F1 and gene expression. *J Biol Chem, 277*(52), 50263-50274.

Canhoto, A. J., Chestukhin, A., Litovchick, L., & DeCaprio, J. A. (2000). Phosphorylation of the retinoblastoma-related protein p130 in growth-arrested cells. *Oncogene, 19*(44), 5116-5122

Cao, L., Faha, B., Dembski, M., Tsai, L. H., Harlow, E., & Dyson, N. (1992). Independent binding of the retinoblastoma protein and p107 to the transcription factor E2F. *Nature, 355*(6356), 176-179.

Cao, L., Peng, B., Yao, L., Zhang, X., Sun, K., Yang, X., & Yu, L. (2010). The ancient function of RB-E2F pathway: insights from its evolutionary history. *Biol Direct, 5*, 55.

Cecchini, M. J., & Dick, F. A. (2011). The biochemical basis of CDK phosphorylation-independent regulation of E2F1 by the retinoblastoma protein. *Biochem J, 434*(2), 297-308.

Cecchini, M. J., Thwaites, M. J., Talluri, S., MacDonald, J. I., Passos, D. T., Chong, J. L., . . . Dick, F. A. (2014). A retinoblastoma allele that is mutated at its common E2F interaction site inhibits cell proliferation in gene-targeted mice. *Mol Cell Biol, 34*(11), 2029-2045.

Chellappan, S. P., Hiebert, S., Mudryj, M., Horowitz, J. M., & Nevins, J. R. (1991). The E2F transcription factor is a cellular target for the RB protein. *Cell, 65*(6), 1053-1061.

Chen, D., Opavsky, R., Pacal, M., Tanimoto, N., Wenzel, P., Seeliger, M. W., . . . Bremner, R. (2007). Rb-mediated neuronal differentiation through cell-cycle-independent regulation of E2f3a. *PLoS Biol, 5*(7), e179.

Chen, H. Z., Tsai, S. Y., & Leone, G. (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer, 9*(11), 785-797.

Chen, P. L., Scully, P., Shew, J. Y., Wang, J. Y., & Lee, W. H. (1989). Phosphorylation of the retinoblastoma gene product is modulated during the cell cycle and cellular differentiation. *Cell, 58*(6), 1193-1198.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., . . . Lowe, S. W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. *Cancer Cell, 17*(4), 376-387.

Chittenden, T., Livingston, D. M., & Kaelin, W. G., Jr. (1991). The T/E1A-binding domain of the retinoblastoma product can interact selectively with a sequence-specific DNA-binding protein. *Cell, 65*(6), 1073-1082.

Clarke, A. R., Maandag, E. R., van Roon, M., van der Lugt, N. M., van der Valk, M., Hooper, M. L., . . . te Riele, H. (1992). Requirement for a functional Rb-1 gene in murine development. *Nature, 359*(6393), 328-330.

Classon, M., & Dyson, N. (2001). p107 and p130: versatile proteins with interesting pockets. *Exp Cell Res, 264*(1), 135-147.

Classon, M., & Harlow, E. (2002). The retinoblastoma tumour suppressor in development and cancer. *Nat Rev Cancer, 2*(12), 910-917.

Claudio, P. P., Tonini, T., & Giordano, A. (2002). The retinoblastoma family: twins or distant cousins? *Genome Biol, 3*(9), reviews3012.

Cobrinik, D. (2005). Pocket proteins and cell cycle control. *Oncogene, 24*(17), 2796-2809.

Cobrinik, D., Lee, M. H., Hannon, G., Mulligan, G., Bronson, R. T., Dyson, N., . . . Jacks, T. (1996). Shared role of the pRB-related p130 and p107 proteins in limb development. *Genes Dev, 10*(13), 1633-1644.

Coe, B. P., Thu, K. L., Aviel-Ronen, S., Vucic, E. A., Gazdar, A. F., Lam, S., . . . Lam, W. L. (2013). Genomic deregulation of the E2F/Rb pathway leads to activation of the oncogene EZH2 in small cell lung cancer. *PLoS One, 8*(8), e71670.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr, 50*(Pt 5), 760-763.

Cress, W. D., Johnson, D. G., & Nevins, J. R. (1993). A genetic analysis of the E2F1 gene distinguishes regulation by Rb, p107, and adenovirus E4. *Mol Cell Biol, 13*(10), 6314-6325.

Dannenberg, J. H., Schuijff, L., Dekker, M., van der Valk, M., & te Riele, H. (2004). Tissue-specific tumor suppressor activity of retinoblastoma gene homologs p107 and p130. *Genes Dev, 18*(23), 2952-2962.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics, 27*(8), 1164-1165.

DeCaprio, J. A. (2009). How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. *Virology, 384*(2), 274-284.

DeCaprio, J. A., Ludlow, J. W., Lynch, D., Furukawa, Y., Griffin, J., Piwnica-Worms, H., . . . Livingston, D. M. (1989). The product of the retinoblastoma susceptibility gene has properties of a cell cycle regulatory element. *Cell, 58*(6), 1085-1095.

Devoto, S. H., Mudryj, M., Pines, J., Hunter, T., & Nevins, J. R. (1992). A cyclin A-protein kinase complex possesses sequence-specific DNA binding activity: p33cdk2 is a component of the E2F-cyclin A complex. *Cell, 68*(1), 167-176.

Dick, F. A., & 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*(3), 639-649.

Dick, F. A., & Rubin, S. M. (2013). Molecular mechanisms underlying RB protein function. *Nat Rev Mol Cell Biol, 14*(5), 297-306.

Dick, F. A., Sailhamer, E., & 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*(10), 3715-3727.

Dimova, D. K., & Dyson, N. J. (2005). The E2F transcriptional network: old acquaintances with new faces. *Oncogene, 24*(17), 2810-2826.

Dowdy, S. F., Hinds, P. W., Louie, K., Reed, S. I., Arnold, A., & Weinberg, R. A. (1993). Physical interaction of the retinoblastoma protein with human D cyclins. *Cell, 73*(3), 499-511.

Dyson, N., Buchkovich, K., Whyte, P., & Harlow, E. (1989). The cellular 107K protein that binds to adenovirus E1A also associates with the large T antigens of SV40 and JC virus. *Cell, 58*(2), 249-255.

Dyson, N., Dembski, M., Fattaey, A., Ngwu, C., Ewen, M., & Helin, K. (1993). Analysis of p107-associated proteins: p107 associates with a form of E2F that differs from pRB-associated E2F-1. *J Virol, 67*(12), 7641-7647.

Dyson, N. J. (2016). RB1: a prototype tumor suppressor and an enigma. *Genes Dev, 30*(13), 1492-1502.

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Comput Biol, 7*(10), e1002195.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2126-2132.

Farkas, T., Hansen, K., Holm, K., Lukas, J., & Bartek, J. (2002). Distinct phosphorylation events regulate p130- and p107-mediated repression of E2F-4. *J Biol Chem, 277*(30), 26741-26752.

Felsani, A., Mileo, A. M., & Paggi, M. G. (2006). Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins. *Oncogene, 25*(38), 5277-5285.

Fera, D., Schultz, D. C., Hodawadekar, S., Reichman, M., Donover, P. S., Melvin, J., . . .

Marmorstein, R. (2012). Identification and characterization of small molecule antagonists of pRb inactivation by viral oncoproteins. *Chem Biol, 19*(4), 518-528.

Forristal, C., Henley, S. A., MacDonald, J. I., Bush, J. R., Ort, C., Passos, D. T., . . . Dick, F. A. (2014). Loss of the Mammalian DREAM Complex Deregulates Chondrocyte Proliferation. *Mol Cell Biol, 34*(12), 2221-2234.

Gagrica, S., Hauser, S., Kolfschoten, I., Osterloh, L., Agami, R., & Gaubatz, S. (2004). Inhibition of oncogenic transformation by mammalian Lin-9, a pRB-associated protein. *EMBO J, 23*(23), 4627-4638.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal, 6*(269).

Garriga, J., Limon, A., Mayol, X., Rane, S. G., Albrecht, J. H., Reddy, E. P., . . . Grana, X. (1998). Differential regulation of the retinoblastoma family of proteins during cell proliferation and differentiation. *Biochem J, 333 (Pt 3)*, 645-654.

Gaubatz, S., Lees, J. A., Lindeman, G. J., & Livingston, D. M. (2001). E2F4 is exported from the nucleus in a CRM1-dependent manner. *Mol Cell Biol, 21*(4), 1384-1392.

Georlette, D., Ahn, S., MacAlpine, D. M., Cheung, E., Lewis, P. W., Beall, E. L., . . . Botchan, M. R. (2007). Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. *Genes Dev, 21*(22), 2880-2896.

Giangrande, P. H., Zhu, W., Rempel, R. E., Laakso, N., & Nevins, J. R. (2004). Combinatorial gene control involving E2F and E Box family members. *EMBO J, 23*(6), 1336-1347.

Grana, X., Garriga, J., & Mayol, X. (1998). Role of the retinoblastoma protein family, pRB, p107 and p130 in the negative control of cell growth. *Oncogene, 17*(25), 3365-3383.

Guiley, K. Z., Liban, T. J., Felthousen, J. G., Ramanan, P., Litovchick, L., & Rubin, S. M. (2015). Structural mechanisms of DREAM complex assembly and regulation. *Genes Dev, 29*(9), 961-974.

Hallstrom, T. C., & Nevins, J. R. (2003). Specificity in the activation and control of transcription factor E2F-dependent apoptosis. *Proc Natl Acad Sci U S A, 100*(19), 10848-10853.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674.

Hansen, K., Farkas, T., Lukas, J., Holm, K., Ronnstrand, L., & Bartek, J. (2001). Phosphorylation-dependent and -independent functions of p130 cooperate to evoke a sustained G1 block. *EMBO J, 20*(3), 422-432.

Harlow, E., Whyte, P., Franza, B. R., Jr., & Schley, C. (1986). Association of adenovirus early-region 1A proteins with cellular polypeptides. *Mol Cell Biol, 6*(5), 1579-1589.

Harrison, M. M., Ceol, C. J., Lu, X., & Horvitz, H. R. (2006). Some C. elegans class B synthetic multivulva proteins encode a conserved LIN-35 Rb-containing complex distinct from a NuRD-like complex. *Proc Natl Acad Sci U S A, 103*(45), 16782-16787.

Helin, K., Lees, J. A., Vidal, M., Dyson, N., Harlow, E., & Fattaey, A. (1992). A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell, 70*(2), 337-350.

Henley, S. A., & Dick, F. A. (2012). The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle. *Cell Div, 7*(1), 10.

Hiebert, S. W. (1993). Regions of the retinoblastoma gene product required for its interaction with the E2F transcription factor are necessary for E2 promoter repression and pRb-mediated growth suppression. *Mol Cell Biol, 13*(6), 3384-3391.

Hirschi, A., Cecchini, M., Steinhardt, R. C., Schamber, M. R., Dick, F. A., & Rubin, S. M. (2010). An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. *Nat Struct Mol Biol, 17*(9), 1051-1057.

Hsu, J., & Sage, J. (2016). Novel functions for the transcription factor E2F4 in development and disease. *Cell Cycle*, 1-8.

Hurford, R. K., Jr., Cobrinik, D., Lee, M. H., & Dyson, N. (1997). pRB and p107/p130 are required for the regulated expression of different sets of E2F responsive genes. *Genes Dev, 11*(11), 1447-1463.

Jacks, T., Fazeli, A., Schmitt, E. M., Bronson, R. T., Goodell, M. A., & Weinberg, R. A. (1992). Effects of an Rb mutation in the mouse. *Nature, 359*(6393), 295-300.

Jayadeva, G., Kurimchak, A., Garriga, J., Sotillo, E., Davis, A. J., Haines, D. S., . . . Grana, X. (2010). B55alpha PP2A holoenzymes modulate the phosphorylation status of the retinoblastoma-related protein p107 and its activation. *J Biol Chem, 285*(39), 29863-29873.

Ji, P., Jiang, H., Rekhtman, K., Bloom, J., Ichetovkin, M., Pagano, M., & Zhu, L. (2004). An Rb-Skp2-p27 pathway mediates acute cell cycle inhibition by Rb and is retained in a partial-penetrance Rb mutant. *Mol Cell, 16*(1), 47-58.

Jiang, H., Karnezis, A. N., Tao, M., Guida, P. M., & Zhu, L. (2000). pRB and p107 have distinct effects when expressed in pRB-deficient tumor cells at physiologically relevant levels. *Oncogene, 19*(34), 3878-3887.

Jiao, W., Datta, J., Lin, H. M., Dundr, M., & Rane, S. G. (2006). Nucleocytoplasmic shuttling of the retinoblastoma tumor suppressor protein via Cdk phosphorylation-dependent nuclear export. *J Biol Chem, 281*(49), 38098-38108.

Jones, R. E., Wegrzyn, R. J., Patrick, D. R., Balishin, N. L., Vuocolo, G. A., Riemen, M. W., . . . Oliff, A. (1990). Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. *J Biol Chem, 265*(22), 12782-12785.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780.

Knudsen, E. S., & Wang, J. Y. (1998). Hyperphosphorylated p107 and p130 bind to T-antigen: identification of a critical regulatory sequence present in RB but not in p107/p130. *Oncogene, 16*(13), 1655-1663.

Korenjak, M., Taylor-Harding, B., Binne, U. K., Satterlee, J. S., Stevaux, O., Aasland, R., . . . Brehm, A. (2004). Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell, 119*(2), 181-193.

Lamber, E. P., Beuron, F., Morris, E. P., Svergun, D. I., & Mittnacht, S. (2013). Structural insights into the mechanism of phosphoregulation of the retinoblastoma protein. *PLoS One, 8*(3), e58463.

Lammens, T., Li, J., Leone, G., & De Veylder, L. (2009). Atypical E2Fs: new players in the E2F transcription factor family. *Trends Cell Biol, 19*(3), 111-118.

Landman, A. S., Danielian, P. S., & Lees, J. A. (2013). Loss of pRB and p107 disrupts cartilage development and promotes enchondroma formation. *Oncogene, 32*(40), 4798-4805.

Lee, B. K., Bhinge, A. A., & Iyer, V. R. (2011). Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis. *Nucleic Acids Res, 39*(9), 3558-3573.

Lee, C., Chang, J. H., Lee, H. S., & Cho, Y. (2002). Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor. *Genes Dev, 16*(24), 3199-3212.

Lee, E. Y., Cam, H., Ziebold, U., Rayman, J. B., Lees, J. A., & Dynlacht, B. D. (2002). E2F4 loss suppresses tumorigenesis in Rb mutant mice. *Cancer Cell, 2*(6), 463-472.

Lee, E. Y., Chang, C. Y., Hu, N., Wang, Y. C., Lai, C. C., Herrup, K., . . . Bradley, A. (1992). Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. *Nature, 359*(6393), 288-294.

Lee, J. O., Russo, A. A., & Pavletich, N. P. (1998). Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature, 391*(6670), 859-865.

Lee, M. H., Williams, B. O., Mulligan, G., Mukai, S., Bronson, R. T., Dyson, N., . . . Jacks, T. (1996). Targeted disruption of p107: functional overlap between p107 and Rb. *Genes Dev, 10*(13), 1621-1632.

Lee, W. H., Bookstein, R., Hong, F., Young, L. J., Shew, J. Y., & Lee, E. Y. (1987). Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science, 235*(4794), 1394-1399.

Lees, J. A., Buchkovich, K. J., Marshak, D. R., Anderson, C. W., & Harlow, E. (1991). The

retinoblastoma protein is phosphorylated on multiple sites by human cdc2. *EMBO J, 10*(13), 4279-4290.

Lees, J. A., Saito, M., Vidal, M., Valentine, M., Look, T., Harlow, E., . . . Helin, K. (1993). The retinoblastoma protein binds to a family of E2F transcription factors. *Mol Cell Biol, 13*(12), 7813-7825.

Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 1), 48-57.

Lewis, P. W., Beall, E. L., Fleischer, T. C., Georlette, D., Link, A. J., & Botchan, M. R. (2004). Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev, 18*(23), 2929-2940.

Liban, T. J., Thwaites, M. J., Dick, F. A., & Rubin, S. M. (2016). Structural Conservation and E2F Binding Specificity within the Retinoblastoma Pocket Protein Family. *J Mol Biol, 428*(20), 3960-3971.

Litovchick, L., Florens, L. A., Swanson, S. K., Washburn, M. P., & DeCaprio, J. A. (2011). DYRK1A protein kinase promotes quiescence and senescence through DREAM complex assembly. *Genes Dev, 25*(8), 801-813.

Litovchick, L., Sadasivam, S., Florens, L., Zhu, X., Swanson, S. K., Velmurugan, S., . . . DeCaprio, J. A. (2007). Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol Cell, 26*(4), 539-551.

Liu, H., Tang, X., Srivastava, A., Pecot, T., Daniel, P., Hemmelgarn, B., . . . Leone, G. (2015). Redeployment of Myc and E2f1-3 drives Rb-deficient cell cycles. *Nat Cell Biol, 17*(8), 1036-1048.

Liu, X., & Marmorstein, R. (2007). Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes Dev, 21*(21), 2711-2716.

Ma, L., Quigley, I., Omran, H., & Kintner, C. (2014). Multicilin drives centriole biogenesis via E2f proteins. *Genes Dev, 28*(13), 1461-1471.

Malumbres, M., & Barbacid, M. (2001). To cycle or not to cycle: a critical decision in cancer. *Nat Rev Cancer, 1*(3), 222-231.

Marceau, A. H., Felthousen, J. G., Goetsch, P. D., Iness, A. N., Lee, H. W., Tripathi, S. M., . . . Rubin, S. M. (2016). Structural basis for LIN54 recognition of CHR elements in cell cycle-regulated promoters. *Nat Commun, 7*, 12301.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-674.

Medina, E. M., Turner, J. J., Gordan, R., Skotheim, J. M., & Buchler, N. E. (2016). Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi. *Elife, 5*.

Mihara, K., Cao, X. R., Yen, A., Chandler, S., Driscoll, B., Murphree, A. L., . . . Fung, Y. K. (1989). Cell cycle-dependent regulation of phosphorylation of the human retinoblastoma gene product. *Science, 246*(4935), 1300-1303.

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

Morgunova, E., Yin, Y., Jolma, A., Dave, K., Schmierer, B., Popov, A., . . . Taipale, J. (2015). Structural insights into the DNA-binding specificity of E2F family transcription factors. *Nat Commun, 6*, 10050.

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

Muller, G. A., Quaas, M., Schumann, M., Krause, E., Padi, M., Fischer, M., . . . Engeland, K. (2012). The CHR promoter element controls cell cycle-dependent gene transcription and binds the DREAM and MMB complexes. *Nucleic Acids Res, 40*(4), 1561-1578.

Muller, G. A., Wintsche, A., Stangner, K., Prohaska, S. J., Stadler, P. F., & Engeland, K. (2014). The CHR site: definition and genome-wide identification of a cell cycle transcriptional element. *Nucleic Acids Res*.

Muller, H., Moroni, M. C., Vigo, E., Petersen, B. O., Bartek, J., & Helin, K. (1997). Induction of S-phase entry by E2F transcription factors depends on their nuclear localization. *Mol Cell Biol, 17*(9), 5508-5520.

Mulligan, G., & Jacks, T. (1998). The retinoblastoma gene family: cousins with overlapping interests. *Trends Genet, 14*(6), 223-229.

Nevins, J. R. (1992). E2F: a link between the Rb tumor suppressor protein and viral oncoproteins. *Science, 258*(5081), 424-429.

Nevins, J. R. (1994). Cell cycle targets of the DNA tumor viruses. *Curr Opin Genet Dev, 4*(1), 130-134.

Nor Rashid, N., Yusof, R., & Watson, R. J. (2011). Disruption of repressive p130-DREAM complexes by human papillomavirus 16 E6/E7 oncoproteins is required for cell-cycle progression in cervical cancer cells. *J Gen Virol, 92*(Pt 11), 2620-2627.

Nurse, P., Thuriaux, P., & Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *Mol Gen Genet, 146*(2), 167-178.

Obenauer, J. C., Cantley, L. C., & Yaffe, M. B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res, 31*(13), 3635-3641.

Pardee, A. B. (1989). G1 events and regulation of cell proliferation. *Science, 246*(4930), 603-608.

Pilkinton, M., Sandoval, R., & Colamonici, O. R. (2007). Mammalian Mip/LIN-9 interacts

with either the p107, p130/E2F4 repressor complex or B-Myb in a cell cycle-phase-dependent context distinct from the Drosophila dREAM complex. *Oncogene, 26*(54), 7535-7543.

Qin, X. Q., Chittenden, T., Livingston, D. M., & Kaelin, W. G., Jr. (1992). Identification of a growth suppression domain within the retinoblastoma gene product. *Genes Dev, 6*(6), 953-964.

Reichert, N., Wurster, S., Ulrich, T., Schmitt, K., Hauser, S., Probst, L., . . . Gaubatz, S. (2010). Lin9, a subunit of the mammalian DREAM complex, is essential for embryonic development, for survival of adult mice, and for tumor suppression. *Mol Cell Biol, 30*(12), 2896-2908.

Robanus-Maandag, E., Dekker, M., van der Valk, M., Carrozza, M. L., Jeanny, J. C., Dannenberg, J. H., . . . te Riele, H. (1998). p107 is a suppressor of retinoblastoma development in pRb-deficient mice. *Genes Dev, 12*(11), 1599-1609.

Rubin, S. M. (2013). Deciphering the retinoblastoma protein phosphorylation code. *Trends Biochem Sci, 38*(1), 12-19.

Rubin, S. M., Gall, A. L., Zheng, N., & Pavletich, N. P. (2005). Structure of the Rb C-terminal domain bound to E2F1-DP1: a mechanism for phosphorylation-induced E2F release. *Cell, 123*(6), 1093-1106.

Sadasivam, S., & DeCaprio, J. A. (2013). The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat Rev Cancer, 13*(8), 585-595.

Sadasivam, S., Duan, S., & DeCaprio, J. A. (2012). The MuvB complex sequentially recruits B-Myb and FoxM1 to promote mitotic gene expression. *Genes Dev, 26*(5), 474-489.

Sandoval, R., Pilkinton, M., & Colamonici, O. R. (2009). Deletion of the p107/p130-binding domain of Mip130/LIN-9 bypasses the requirement for CDK4 activity for the dissociation of Mip130/LIN-9 from p107/p130-E2F4 complex. *Exp Cell Res, 315*(17), 2914-2920.

Schmit, F., Korenjak, M., Mannefeld, M., Schmitt, K., Franke, C., von Eyss, B., . . . Gaubatz, S. (2007). LINC, a human complex that is related to pRB-containing complexes in invertebrates regulates the expression of G2/M genes. *Cell Cycle, 6*(15), 1903-1913.

Sengupta, S., Lingnurkar, R., Carey, T. S., Pomaville, M., Kar, P., Feig, M., . . . Henry, R. W. (2015). The Evolutionarily Conserved C-terminal Domains in the Mammalian Retinoblastoma Tumor Suppressor Family Serve as Dual Regulators of Protein Stability and Transcriptional Potency. *J Biol Chem, 290*(23), 14462-14475.

Sherr, C. J. (1996). Cancer cell cycles. *Science, 274*(5293), 1672-1677.

Shirodkar, S., Ewen, M., DeCaprio, J. A., Morgan, J., Livingston, D. M., & Chittenden, T. (1992). The transcription factor E2F interacts with the retinoblastoma product and a p107-cyclin A complex in a cell cycle-regulated manner. *Cell, 68*(1), 157-166.

Singh, M., Krajewski, M., Mikolajka, A., & Holak, T. A. (2005). Molecular determinants for the complex formation between the retinoblastoma protein and LXCXE sequences. *J Biol Chem, 280*(45), 37868-37876.

Smith, E. J., Leone, G., DeGregori, J., Jakoi, L., & Nevins, J. R. (1996). The accumulation of an E2F-p130 transcriptional repressor distinguishes a G0 cell state from a G1 cell state. *Mol Cell Biol, 16*(12), 6965-6976.

Telford, M. J., Budd, G. E., & Philippe, H. (2015). Phylogenomic Insights into Animal Evolution. *Curr Biol, 25*(19), R876-887.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., . . . Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr, 65*(Pt 6), 582-601.

Trimarchi, J. M., & Lees, J. A. (2002). Sibling rivalry in the E2F family. *Nat Rev Mol Cell Biol, 3*(1), 11-20.

Ubersax, J. A., & Ferrell, J. E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol, 8*(7), 530-541.

van den Heuvel, S., & Dyson, N. J. (2008). Conserved functions of the pRB and E2F families. *Nat Rev Mol Cell Biol, 9*(9), 713-724.

Verona, R., Moberg, K., Estes, S., Starz, M., Vernon, J. P., & Lees, J. A. (1997). E2F activity is regulated by cell cycle-dependent changes in subcellular localization. *Mol Cell Biol, 17*(12), 7268-7282.

Wang, D., Russell, J. L., & Johnson, D. G. (2000). E2F4 and E2F1 have similar proliferative properties but different apoptotic and oncogenic properties in vivo. *Mol Cell Biol, 20*(10), 3417-3424.

Wen, H., Andrejka, L., Ashton, J., Karess, R., & Lipsick, J. S. (2008). Epigenetic regulation of gene expression by Drosophila Myb and E2F2-RBF via the Myb-MuvB/dREAM complex. *Genes Dev, 22*(5), 601-614.

Whitaker, L. L., Su, H., Baskaran, R., Knudsen, E. S., & Wang, J. Y. (1998). Growth suppression by an E2F-binding-defective retinoblastoma protein (RB): contribution from the RB C pocket. *Mol Cell Biol, 18*(7), 4032-4042.

Wirt, S. E., & Sage, J. (2010). p107 in the public eye: an Rb understudy and more. *Cell Div, 5*, 9.

Wu, M., Chatterji, S., & Eisen, J. A. (2012). Accounting for alignment uncertainty in phylogenomics. *PLoS One, 7*(1), e30288.

Xiao, B., Spencer, J., Clements, A., Ali-Khan, N., Mittnacht, S., Broceno, C., . . . 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*(5), 2363-2368.

Yee, S. P., & Branton, P. E. (1985). Detection of cellular proteins associated with human adenovirus type 5 early region 1A polypeptides. *Virology, 147*(1), 142-153.

Zhang, L., Wei, Y., Pushel, I., Heinze, K., Elenbaas, J., Henry, R. W., & Arnosti, D. N. (2014). Integrated stability and activity control of the Drosophila Rbf1 retinoblastoma protein. *J Biol Chem, 289*(36), 24863-24873.

Zheng, N., Fraenkel, E., Pabo, C. O., & Pavletich, N. P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev, 13*(6), 666-674.

Zhu, L., Enders, G., Lees, J. A., Beijersbergen, R. L., Bernards, R., & Harlow, E. (1995). The pRB-related protein p107 contains two growth suppression domains: independent interactions with E2F and cyclin/cdk complexes. *EMBO J, 14*(9), 1904-1913.