# Conservation and divergence of C-terminal domain structure in the retinoblastoma protein family

Tyler J. Liban[a], Edgar M. Medina[b,c], Sarvind Tripathi[a], Satyaki Sengupta[d], R. William Henry[d], Nicolas E. Buchler[b,c], and Seth M. Rubin[a,1]

[a]Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064; [b]Department of Biology, Duke University, Durham, NC 27708; [c]Center for Genomic and Computational Biology, Duke University, Durham, NC 27708; and [d]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824

Edited by Nicholas J. Dyson, Massachusetts General Hospital Cancer Center, Charlestown, MA, and accepted by Editorial Board Member Carol Prives March 29, 2017 (received for review November 21, 2016)

**The retinoblastoma protein (Rb) and the homologous pocket proteins p107 and p130 negatively regulate cell proliferation by binding and inhibiting members of the E2F transcription factor family. The structural features that distinguish Rb from other pocket proteins have been unclear but are critical for understanding their functional diversity and determining why Rb has unique tumor suppressor activities. We describe here important differences in how the Rb and p107 C-terminal domains (CTDs) associate with the coiled-coil and marked-box domains (CMs) of E2Fs. We find that although CTD–CM binding is conserved across protein families, Rb and p107 CTDs show clear preferences for different E2Fs. A crystal structure of the p107 CTD bound to E2F5 and its dimer partner DP1 reveals the molecular basis for pocket protein–E2F binding specificity and how cyclin-dependent kinases differentially regulate pocket proteins through CTD phosphorylation. Our structural and biochemical data together with phylogenetic analyses of Rb and E2F proteins support the conclusion that Rb evolved specific structural motifs that confer its unique capacity to bind with high affinity those E2Fs that are the most potent activators of the cell cycle.**

cell cycle | tumor suppressor protein | E2F | protein–protein interactions | evolution

E2F transcription factors regulate the mammalian cell cycle by controlling expression of genes required for DNA synthesis and cell division (1). 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 (2–5). Beyond cell-cycle regulation, these pocket protein–E2F complexes are the focal point of signaling pathways that trigger diverse cellular processes including proliferation, differentiation, apoptosis, and the stress response. Improper inactivation of pocket proteins is a common mechanism by which cancerous cells maintain aberrant proliferation (1, 5–7). Pocket protein–E2F dissociation and subsequent E2F activation is induced by cyclin-dependent kinase phosphorylation (3–5, 8) or binding of viral oncoproteins such as the SV40 T-antigen (9).

The E2F family contains eight members, five of which (E2F1 to E2F5) form complexes with pocket proteins (1). E2F1 to E2F3 associate exclusively with Rb and are potent activators of transcription during the G1 and S phases of the cell cycle (10, 11). These "activating" E2Fs also specifically induce apoptosis (12). E2F4 is found in complexes with all three pocket proteins and is thought of primarily as a repressor, because it typically occupies promoters of repressed genes and is exported from the nucleus upon release from pocket proteins (1, 13, 14). 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 (14). 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.

Although 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 (3, 15). Rb is a more potent tumor suppressor in mouse cancer models (3, 15), and mutations are more commonly observed in human cancers (6, 16). One proposed explanation for these observations is that Rb forms unique complexes with the activating E2Fs (E2F1 to E2F3), although other pocket protein-specific binding interactions may confer distinct functions (4, 17). For example, through unique protein interactions, Rb functions in processes outside of cell-cycle control including apoptosis, chromosome stability, transcriptional silencing, and metabolic regulation (5, 18).

The five canonical E2Fs each contain a DNA-binding domain (DBD), transactivation domain, and coiled-coil and marked-box domain (CM) (Fig. 1*A*). The DBDs are homologous and bind similar DNA sequences as heterodimers with one of three DP proteins (1). The CM domain of E2F also heterodimerizes with a similar domain in DP (19), and the CM heterodimer binds other transcription factors as a proposed mechanism for how specific E2F family members activate distinct genes (20, 21). The Rb family pocket domains bind the E2F transactivation domain and other cellular and viral proteins using a distinct surface called the LxCxE cleft (4, 17) (Fig. 1*A*). 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 (22–24). A crystal structure demonstrates that the Rb CTD (RbC) binds the

## Significance

**The retinoblastoma (Rb) pocket protein and E2F transcription factor families regulate cell division and are commonly deregulated in proliferating cancer cells. An important question has been what distinguishing molecular features of Rb and its interaction with E2F result in its unique potency as a tumor suppressor relative to its homologous proteins p107 and p130. Here we identify structures in Rb, p107, and E2Fs that determine the specificity in their association. We explain binding preferences with an X-ray crystal structure of a p107–E2F5–DP1 complex, and present phylogenetic analyses that implicate coevolving protein interactions between family members as a key determinant of their evolution.**

Author contributions: T.J.L., E.M.M., S.S., R.W.H., N.E.B., and S.M.R. designed research; T.J.L., E.M.M., S.T., and S.S. performed research; T.J.L., E.M.M., S.T., R.W.H., N.E.B., and S.M.R. analyzed data; and T.J.L., N.E.B., and S.M.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. N.J.D. is a guest editor invited by the Editorial Board.

Data deposition: The crystallography, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5TUU and 5TUV).

[1]To whom correspondence should be addressed. Email: srubin@ucsc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619170114/-/DCSupplemental.

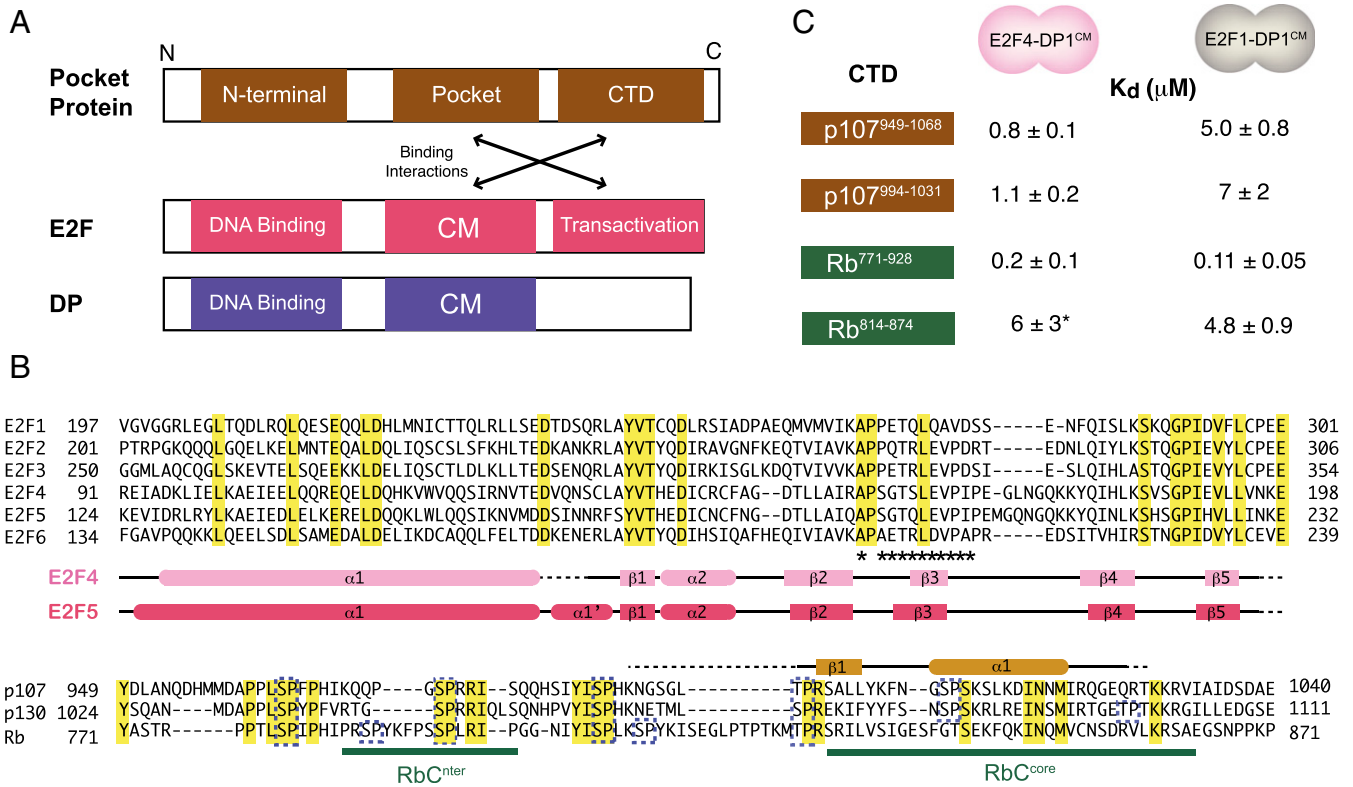


**Fig. 1.** Interactions between pocket protein C-terminal domains and the E2F coiled-coil and marked-box domains. (*A*) Domain architecture of pocket proteins, E2F, and DP. The E2F transactivation domain binds the pocket domain, and the E2F–DP CM domain binds the CTD. (*B*) Sequence alignment of the human E2F CM domains and pocket protein CTDs. The secondary structure elements present in our crystal structures are indicated along with residues in E2F5 (*) that contact the p107 CTD. Cdk phosphorylation sites are indicated with blue dashed outlines. (*C*) Calorimetry measurements of RbC and p107C binding to E2F–DP CM domains. $p107^{994–1031}$ binding was measured here, whereas the other values were previously reported (19). The asterisk indicates that the measurement was made with $E2F4–DP2^{CM}$.

E2F1–DP1 CM domains (19), but several studies suggest that this particular association may be specific to Rb and E2F1 (25, 26).

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 domains ($E2F4–DP1^{CM}$) 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 coevolution of a protein–protein interaction critical for control of cell proliferation.

## Results

**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 (Fig. 1 *B* and *C*). We used a coprecipitation assay to identify a minimal fragment of p107C (residues 994 to 1031; $p107C^{994–1031}$) that is suitable for structural studies and sufficient to bind $E2F4–DP1^{CM}$ (Fig. S1). Using isothermal titration calorimetry (ITC) (Fig. 1*C*), we found that $p107C^{994–1031}$ binds with similar affinity as that previously reported for full-length p107C (residues 949 to 1068) (19). Both $p107C^{994–1031}$ and $p107C^{949–1068}$ bind $E2F1–DP1^{CM}$ with lower affinity than they bind $E2F4–DP1^{CM}$. These affinity measurements performed with purified protein fragments are consistent with previous observations of interaction specificities among Rb and E2F family proteins in cells, and suggest that these specificities arise at least in part from intrinsic structural differences (1, 10–13).

Comparing these measurements with previous measurements of RbC reveals several differences between how RbC and p107C bind to $E2F–DP^{CM}$ domains (Fig. 1*C*) (19). First, the affinity of the full RbC sequence (residues 771 to 928) is 4-fold tighter than full p107C for E2F4–DP1 and 50-fold tighter for E2F1–DP1. Second, although RbC makes a bipartite association with contributions from residues 786 to 801 ($RbC^{nter}$) and residues 829 to 864 ($RbC^{core}$) (Fig. 1*B*) (19), all of the interactions made by p107C are contained within $p107C^{994–1031}$ ($p107C^{core}$). Third, whereas $RbC^{core}$ has similar affinity for E2F1 and E2F4 (19), $p107C^{core}$ has higher affinity for E2F4 than E2F1. We next determined the structural basis for these affinity differences.

**Conservation of E2F–DP CM Structures.** We grew crystals of $E2F4–DP1^{CM}$ alone and $E2F5–DP1^{CM}$ bound with $p107C^{994–1031}$ and determined structures with a resolution of 2.3 and 2.9 Å, respectively (Fig. 2 and Table S1). In both structures, the E2F and DP polypeptides have similar secondary structure topology, and the chains entwine to create an extensive interface (Figs. 1*B* and 2*A*). The CM 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 intertwined structure and dependence on DP to complete the hydrophobic core explain why heterodimerization is necessary for E2F stability, DNA binding, and transcriptional activity (1, 19).

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

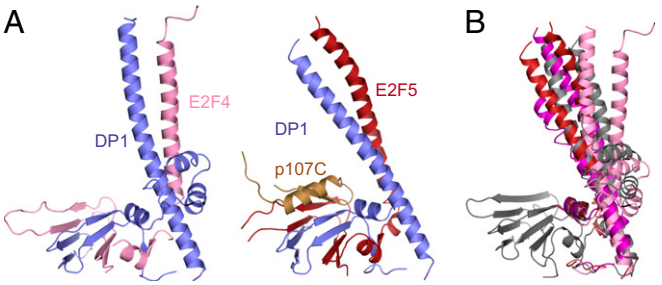


**Fig. 2.** Crystal structures of the E2F4–DP1 and E2F5–DP1 CM domains. (*A*) Overall structures show similar topologies with the E2F and DP chains forming an extensive interface. (*B*) Overlay of the CM domain structures determined for E2F1–DP1 [gray; Protein Data Bank (PDB) ID code 2AZE], E2F4–DP1 (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. S2) so that the different positions of the coiled-coil domains reflect their different orientations relative to the β-sandwich domain. Only the β-sandwich domain of E2F1–DP1 is shown.

are similar to the previously determined structure of E2F1–DP1$^{CM}$ (19), including the topology and structures of the β-sandwich domains (Fig. S2). One notable variation among the structures is the orientation of the coiled-coil domain relative to the β-sandwich domain (Fig. 2*B*). Alignment of the overall structures with the β-domain fixed 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 (20, 21).

Sequence comparison of the human E2Fs reveals that 20 residues are identical within the CM domain (Fig. 1*B*). They map primarily to the coiled-coil interface and the structural core that bridges the β-sandwich and coiled-coil domains (Fig. S2*E*). These amino acids contribute to the overall stability of the E2F–DP heterodimer. The most notable region of the structure that is distinct among paralogs is the end of E2F β3 and the loop between β3 and β4. We explore below the idea that sequence divergence in this region accounts for differences in specificity for different pocket proteins.

**Specificity in Rb and p107 Interactions with E2F–DP$^{CM}$.** p107C binds the E2F5–DP1 marked-box domain using a strand-loop-helix motif (Figs. 2 and 3*A*). The strand adds on in an antiparallel direction to the β-sheet in the sandwich domain that is distal to the coiled coil. The amphipathic p107C helix covers the core of the β-sandwich domain (Fig. 3*A*). The hydrophobic side chains 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. These residues in E2F5 are all conserved in E2F4 (Fig. 1*B*), and we anticipate that E2F4 binds p107 through identical interactions.

We used the Cancer Genome Atlas (https://cancergenome.nih.gov) to identify cancer-associated mutations in p107 and p130 that are localized to the CTD. We mapped these mutations onto the p107C–E2F5–DP1 crystal structure and tested their effects on binding with ITC (Fig. S3). We conclude that most of these cancer-associated mutations map to the exposed surface of the CTD helix and only slightly impair the ability of p107 to bind E2F.

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. 1*C*). In general, the mode of RbC binding to the E2F1–DP1 marked-box domain resembles p107C binding to E2F5–DP1 (Fig. 3) (19). 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. 3*B*). 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 and E2F5 compared with E2F1 (Fig. 3*D*). To understand this preference, we considered residues toward the C terminus of strand β3 in the E2F5 structure (residues 200 to 203). In addition to L198, which is conserved among all E2Fs, these residues contain the only E2F side chains that directly contact p107C, and are different between E2F5 and E2F1. The sequence in E2F5 and E2F4 is VPIP, whereas the sequence in E2F1 is AVDS (Figs. 1*B* and 3*C*). The bulkier V200 in E2F5 (V167 in E2F4) can interact better with I1017 and M1020 in p107 compared with the smaller A275 in E2F1 (Fig. 3*C*). 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.

We used the calorimetry assay to test the importance of the E2F4/E2F5-conserved VPIP motif for p107C$^{994–1031}$ affinity (Fig. 3*D*). We primarily used E2F4 in our binding measurements because E2F4 is more abundant in cells and expresses well as a recombinant protein. E2F4 and E2F5 are highly conserved in the β3-strand that binds p107C (Fig. 1*B*), and they both bind wild-type p107C with similar affinity (Fig. 3*D*). We found that changing the

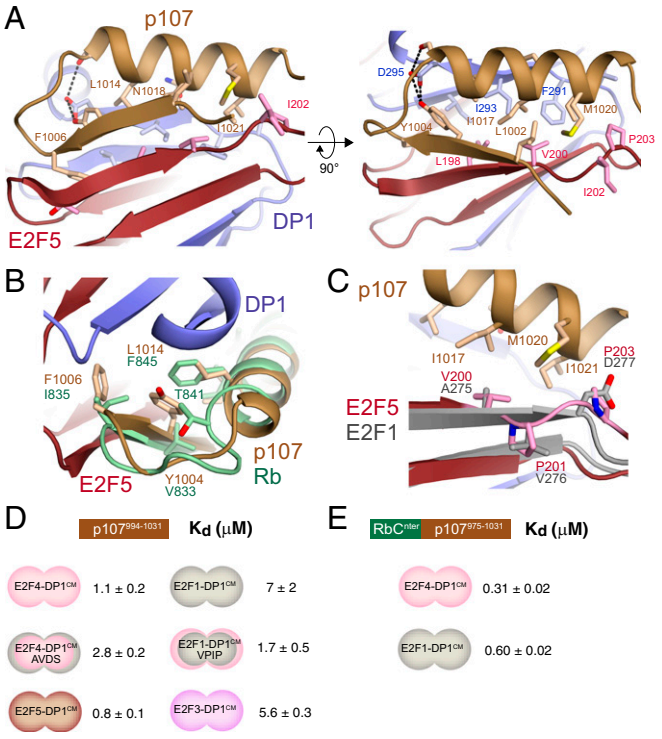


**Fig. 3.** Comparison of p107C–E2F5 and RbC–E2F1 binding interfaces. (*A*) p107C binds the E2F5–DP1 β-sandwich domain using a strand-loop-helix motif and forms a hydrophobic interface with residues from both E2F and DP. (*B*) Overlay of RbC (taken from the RbC–E2F1 structure) with p107C models how RbC would bind E2F5. (*C*) Overlay of E2F1 and E2F5 models how p107C would bind E2F1. (*D*) Affinity measurements for p107C$^{994–1031}$ binding to the indicated E2F–DP$^{CM}$ domains. The E2F4–DP1$^{CM}$ AVDS mutant has the E2F4 VPIP sequence (residues 167 to 170) mutated to AVDS, whereas the E2F1–DP1$^{CM}$ mutant has the E2F1 AVDS sequence (residues 275 to 278) mutated to VPIP. (*E*) Affinity measurements of an RbC–p107C hybrid protein containing residues 771 to 794 of Rb (RbC$^{nter}$) fused to residues 975 to 1031 of p107.

VPIP sequence in E2F4 to the AVDS sequence in E2F1 yields a mutant E2F4–DP1$^{CM}$ heterodimer that binds p107C nearly threefold more weakly than wild type. Conversely, mutation of the E2F1 AVDS sequence to VPIP increases the affinity of E2F1–DP1$^{CM}$ for p107C$^{994–1031}$ fourfold. We also found that p107C$^{994–1031}$ binds E2F3–DP1$^{CM}$ more weakly than it binds E2F4–DP1$^{CM}$ and E2F5–DP1$^{CM}$ and more similar to how it binds E2F1–DP1$^{CM}$ (Fig. 3*D*). Although E2F3 has the first proline to induce the bulge in the strand (Fig. 1*B*), 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 the 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. 1*C*) (19). Rb contains a valine (V852) at the analogous position as I1021 in the p107C helix. Structural alignment suggests that the smaller Rb side chain would not contact P203 in E2F5 (P170 in E2F4), and we observe loss of affinity due to substitution of the I1021 side chain with a smaller hydrophobic group (Fig. S3). 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$^{CM}$ (19). 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. 1*C*). We found that replacing the p107C N-terminal sequence (residues 949 to 974) with the RbC$^{nter}$ sequence (residues 771 to 822) results in a hybrid p107C construct that binds E2F1–DP1$^{CM}$ and E2F4–DP1$^{CM}$ with increased affinity compared with p107C$^{994–1031}$ and p107C$^{949–1068}$ (Fig. 3*E*). This observation demonstrates that the RbC$^{nter}$ sequence enables RbC to bind both activator and repressive E2F proteins with higher affinity than p107C.

Although Rb is in complexes with both activating and repressive E2Fs, it has been proposed that the RbC association is specific to E2F1 (11, 25, 26). We find here that RbC binds different E2F–DP$^{CM}$ domains with similar affinity (Fig. 1*C* and Fig. S4). As we discuss further in Fig. S4, this apparent discrepancy arises from differences in the affinity of different E2F transactivation domains for the Rb pocket domain. In contrast to Rb, differences in affinity for both the transactivation domain (27) and the CM domain (Fig. 1*C*) contribute to the preference of p107 for different E2Fs.

**T997 and S1009 Phosphorylation Regulates p107C Binding to E2F–DP$^{CM}$.** We next explored the question of how Cdk phosphorylation of p107 weakens the p107C–E2F–DP$^{CM}$ association. We phosphorylated the two Cdk sites in p107$^{994–1031}$ (T997 and S1009; Fig. 1*B*) with purified Cdk2–CycA and found by ITC that the affinity of the phosphorylated peptide was 11-fold weaker than the unphosphorylated peptide (Fig. 4*A*). We then made T997A and S1009A mutations in two separate constructs and found that phosphorylation at the remaining site in each construct still weakens affinity. These measurements demonstrate that both phosphorylation events in p107C inhibit binding to E2F-DP$^{CM}$ 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. 4*B*). The loop folds back toward the secondary structure elements, and the S1009 side chain makes a hydrogen bond with S1013, which is in the p107C helix. Phosphorylation of S1009 likely weakens affinity by destabilizing this bound conformation. Electron density for T997 is not visible, suggesting that T997 is disordered. It is less clear, then, why T997 phosphorylation inhibits the association.

The phosphorylation pattern within the CTD of p107 and p130 is distinct from the pattern in Rb (Fig. 1*B*). 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

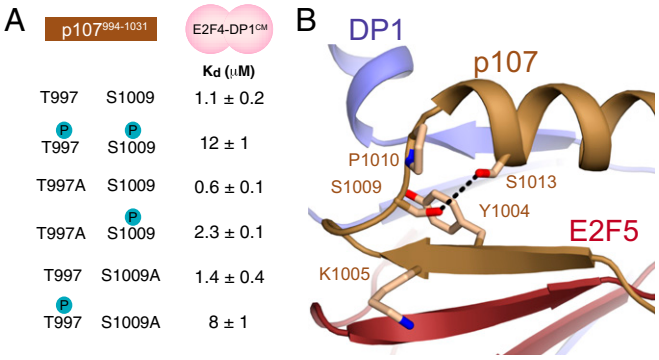


**Fig. 4.** Phosphorylation of Cdk sites in p107C directly inhibits E2F binding. (*A*) ITC measurements of p107C peptides phosphorylated at the indicated sites. (*B*) Structure of the p107C–E2F5–DP1 interface.

inhibit binding of RbC$^{core}$ to E2F1–DP1 (19). 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 p107C T997 and S1009 directly inhibits E2F–DP$^{CM}$ binding, and we could not detect binding of phosphorylated p107C to the p107 pocket domain.

**Rb Sequence Elements That Confer E2F Binding Affinity Coevolved 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 coevolved with E2F1, we examined the evolutionary history of pocket proteins and E2Fs along the metazoan lineage from a subset of 52 genomes (Datasets S1 and S2). Our phylogenetic analysis reveals a number of gene-duplication events that resulted in the expansion of the pocket protein and E2F families (Fig. 5 and Figs. S5, S6, and S7). In agreement with previous work (28), 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 and before the emergence of the Placozoa lineage. 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 Agnantha lineage ("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.

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 to 1023), which plays a prominent role in binding E2F–DP$^{CM}$ (Fig. 3). For example, the helix residues along the interface are hydrophobic in all of the sequences dating back to the early Metazoa, and several positions are nearly strictly conserved (Fig. 6 and Fig. S8). 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. 6 and Fig. S8). This emergence is coincident with the expansion of the protein families at the base of the Craniata lineage (Figs. 5 and 6).

We also examined the sequence corresponding to the end of the E2F β3-strand (V200 to P203 in human E2F5), which our data implicate as a key source of binding preferences between the E2F and pocket proteins (Fig. 3). The ancestral 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. 6 and Fig. S8). This motif is kept in the E2F4 and E2F5 lineages through humans. As seen in our E2F5 structure, the VP*P sequence places

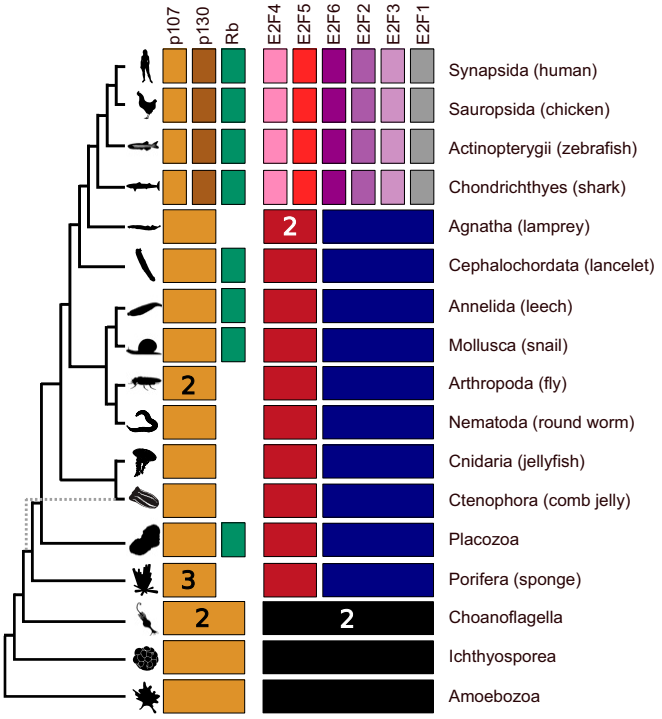


**Fig. 5.** Phylogenetic distribution of pocket protein and E2F subfamilies. Detected pocket proteins (Dataset S1) and E2Fs (Dataset S2) from each genome are classified into subfamilies based on phylogenetic analysis (*SI Materials and Methods*). The resulting distribution of homologs was condensed and ordered according to the current consensus on phylogenetic relationships between major animal lineages (36), including alternative locations of the Ctenophora (dashed gray line) and with Amoebozoa as an outgroup. Common names of exemplary species are in parentheses.

the hydrophobic V200 and P203 side chains at the interface with p107C, and these residues are critical for high-affinity binding of p107C to human E2F4–DP1$^{CM}$ (Fig. 3). The VP*P motif is eventually lost in all three activator E2F lineages. It is lost first in E2F1 and E2F2, which emerge in sharks and lack the first and second proline, respectively (Fig. 6). 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 and E2F5 (Fig. 3*D*). Therefore, the unbiased approach that groups E2Fs based on phylogenetic distance over their entire sequence (Figs. 5 and 6) matches a grouping based on their affinity for p107C (Fig. 3*D*), which arises in large part from sequence variation in just the four motif residues.

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 (Fig. 6 and Fig. S8). The Rb proteins from *Callorhinchus milii* and other cartilaginous fish (*Leucoraja erinacea* and *Scyliorhinus canicula*) have sequences that resemble RbC$^{nter}$, although they lack the S788 (human Rb numbering) Cdk phosphorylation site (Fig. S8). The first complete RbC$^{nter}$ appears in vertebrates (e.g., *Danio rerio*).

Our analysis suggests that the appearance of the E2F1 to E2F3 lineages and loss of the VP*P motif are coincident with the appearance of the RbC$^{nter}$ sequence. 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. 3*E*). The hybrid protein has enhanced affinity for E2F1 over 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 (27), emerge in sharks around the E2F1 and E2F2 emergence (Fig. S9). We propose that Rb has maintained sequence variations from its ancestor that contribute to its high-affinity interactions with and ability to inhibit activator E2Fs.

## 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 has been proposed that RbC has unique affinity for E2F1$^{CM}$ (11, 25, 26). We demonstrate here that RbC associates similarly with different E2F CM domains, and we suggest that RbC–E2F1 binding was thought to be specific because previous Rb–E2F1 coimmunoprecipitation experiments were detecting the stronger pocket–E2F1 transactivation domain association (19, 25–27) (Fig. S4). In contrast to RbC, we find here that p107C binds E2F1$^{CM}$ and E2F3$^{CM}$ with weaker affinity than E2F4$^{CM}$ and E2F5$^{CM}$, and the crystal structure of the p107C complex offers a clear explanation for the preference for the repressor E2Fs (Fig. 3). These data are consistent with previous observations that the p107C domain is required for its growth-suppressive function and that p107 associates exclusively with repressor E2Fs except when it is overexpressed (13, 24, 29). We find that p107 has some capacity to bind all E2Fs, but its preference is tuned by the molecular details of the p107C–E2F$^{CM}$ association.

Our phylogenetic analyses and examination of specific sequence motifs involved in pocket protein–E2F interactions point to a special relationship between Rb and the activating E2Fs. The emergence of E2F1 and E2F2 is accompanied by the emergence

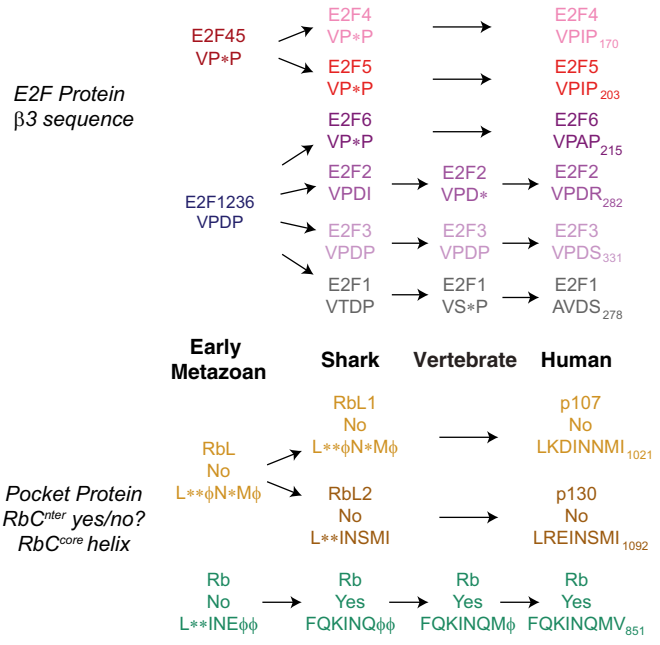


**Fig. 6.** Evolutionary model of sequences involved in E2F$^{CM}$–pocket protein association. Family members are arranged according to their phylogenetic distances over their entire sequences. The VP*P motif in the E2F β3-strand, hydrophobic residues in RbC$^{core}$, and whether the RbC$^{nter}$ sequence is present are indicated. The amino acid number of the last residue in each human sequence is shown.

of Rb-specific sequences that result in higher E2F affinity. 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 binding affinity. Although 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 (19). We suggest that the additional E2F-binding motif coevolved with a regulatory mechanism such that Rb–E2F complexes can be dissociated.

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 (30–33). Rb regulation of activator E2Fs cannot be complemented by p107/p130, because they fail to bind with sufficient affinity. In more recently derived animal lineages with multiple pocket proteins and E2Fs, exclusive relationships between family members may allow for independent regulation of different processes. For example, specific Rb inhibition of activating E2Fs is likely relevant to other functions beyond the cell cycle such as apoptosis or response to DNA damage (1, 12). 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 results 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 (29), 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 (34). Increasing p107/p130 association with the activator E2Fs may be a viable therapeutic strategy toward harnessing their activity to compensate for Rb loss.

## Materials and Methods

Proteins were expressed and purified using standard methods. Crystallization was performed using vapor diffusion in sitting drops, and X-ray diffraction data were collected at the Advanced Photon Source. The dimer and trimer structures were solved by molecular replacement and single anomalous diffraction methods. ITC experiments were performed with a MicroCal VP-ITC instrument, and the reported errors are the SDs from two to four measurements. Phosphorylation of p107C was performed as previously described (27). Profile-hidden Markov models were built and used to retrieve E2F and pocket protein homologs, as described in ref. 35. Following sequence alignment, phylogenetic analysis was performed using maximum-likelihood methods. Details of all experimental procedures can be found in *SI Materials and Methods*.

**ACKNOWLEDGMENTS.** This research used resources of the Advanced Photon Source, beamline 23-ID-B, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. This work was supported by grants from the National Institutes of Health (to S.M.R.; R01CA132685) and the Burroughs Wellcome Fund CASI Award (to N.E.B.; BWF1005769.01).

1. Trimarchi JM, Lees JA (2002) Sibling rivalry in the E2F family. *Nat Rev Mol Cell Biol* 3:11–20.
2. Classon M, Dyson N (2001) p107 and p130: Versatile proteins with interesting pockets. *Exp Cell Res* 264:135–147.
3. Cobrinik D (2005) Pocket proteins and cell cycle control. *Oncogene* 24:2796–2809.
4. Dick FA, Rubin SM (2013) Molecular mechanisms underlying RB protein function. *Nat Rev Mol Cell Biol* 14:297–306.
5. Dyson NJ (2016) RB1: A prototype tumor suppressor and an enigma. *Genes Dev* 30:1492–1502.
6. Classon M, Harlow E (2002) The retinoblastoma tumour suppressor in development and cancer. *Nat Rev Cancer* 2:910–917.
7. Sherr CJ (1996) Cancer cell cycles. *Science* 274:1672–1677.
8. Rubin SM (2013) Deciphering the retinoblastoma protein phosphorylation code. *Trends Biochem Sci* 38:12–19.
9. DeCaprio JA (2009) How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40. *Virology* 384:274–284.
10. Helin K, et al. (1992) A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. *Cell* 70:337–350.
11. Lees JA, et al. (1993) The retinoblastoma protein binds to a family of E2F transcription factors. *Mol Cell Biol* 13:7813–7825.
12. Attwooll C, Lazzerini Denchi E, Helin K (2004) The E2F family: Specific functions and overlapping interests. *EMBO J* 23:4709–4716.
13. Dyson N, et al. (1993) Analysis of p107-associated proteins: p107 associates with a form of E2F that differs from pRB-associated E2F-1. *J Virol* 67:7641–7647.
14. Hsu J, Sage J (2016) Novel functions for the transcription factor E2F4 in development and disease. *Cell Cycle* 15:3183–3190.
15. Mulligan G, Jacks T (1998) The retinoblastoma gene family: Cousins with overlapping interests. *Trends Genet* 14:223–229.
16. Burkhart DL, Sage J (2008) Cellular mechanisms of tumour suppression by the retinoblastoma gene. *Nat Rev Cancer* 8:671–682.
17. Morris EJ, Dyson NJ (2001) Retinoblastoma protein partners. *Adv Cancer Res* 82:1–54.
18. Ishak CA, et al. (2016) An RB-EZH2 complex mediates silencing of repetitive DNA sequences. *Mol Cell* 64:1074–1087.
19. Rubin SM, Gall AL, Zheng N, Pavletich NP (2005) Structure of the Rb C-terminal domain bound to E2F1-DP1: A mechanism for phosphorylation-induced E2F release. *Cell* 123:1093–1106.
20. Giangrande PH, Zhu W, Rempel RE, Laakso N, Nevins JR (2004) Combinatorial gene control involving E2F and E box family members. *EMBO J* 23:1336–1347.
21. Hallstrom TC, Nevins JR (2003) Specificity in the activation and control of transcription factor E2F-dependent apoptosis. *Proc Natl Acad Sci USA* 100:10848–10853.
22. Qin XQ, Chittenden T, Livingston DM, Kaelin WG, Jr (1992) Identification of a growth suppression domain within the retinoblastoma gene product. *Genes Dev* 6:953–964.
23. Sengupta S, et al. (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* 290:14462–14475.
24. Zhu L, et al. (1995) The pRB-related protein p107 contains two growth suppression domains: Independent interactions with E2F and cyclin/cdk complexes. *EMBO J* 14:1904–1913.
25. Cecchini MJ, Dick FA (2011) The biochemical basis of CDK phosphorylation-independent regulation of E2F1 by the retinoblastoma protein. *Biochem J* 434:297–308.
26. Cecchini MJ, et al. (2014) A retinoblastoma allele that is mutated at its common E2F interaction site inhibits cell proliferation in gene-targeted mice. *Mol Cell Biol* 34:2029–2045.
27. Liban TJ, Thwaites MJ, Dick FA, Rubin SM (2016) Structural conservation and E2F binding specificity within the retinoblastoma pocket protein family. *J Mol Biol* 428:3960–3971.
28. Cao L, et al. (2010) The ancient function of RB-E2F pathway: Insights from its evolutionary history. *Biol Direct* 5:55.
29. Cress WD, Johnson DG, Nevins JR (1993) A genetic analysis of the E2F1 gene distinguishes regulation by Rb, p107, and adenovirus E4. *Mol Cell Biol* 13:6314–6325.
30. Liu H, et al. (2015) Redeployment of Myc and E2f1–3 drives Rb-deficient cell cycles. *Nat Cell Biol* 17:1036–1048.
31. Chen D, et al. (2007) Rb-mediated neuronal differentiation through cell-cycle-independent regulation of E2f3a. *PLoS Biol* 5:e179.
32. Landman AS, Danielian PS, Lees JA (2013) Loss of pRB and p107 disrupts cartilage development and promotes enchondroma formation. *Oncogene* 32:4798–4805.
33. Saavedra HI, et al. (2002) Specificity of E2F1, E2F2, and E2F3 in mediating phenotypes induced by loss of Rb. *Cell Growth Differ* 13:215–225.
34. Lee EY, et al. (2002) E2F4 loss suppresses tumorigenesis in Rb mutant mice. *Cancer Cell* 2:463–472.
35. Medina EM, Turner JJ, Gordân R, Skotheim JM, Buchler NE (2016) Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi. *Elife* 5:e09492.
36. Telford MJ, Budd GE, Philippe H (2015) Phylogenomic insights into animal evolution. *Curr Biol* 25:R876–R887.
37. Mccoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
38. Terwilliger TC, et al. (2009) Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65:582–601.
39. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
40. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
41. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195.
42. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780.
43. Wu M, Chatterji S, Eisen JA (2012) Accounting for alignment uncertainty in phylogenomics. *PLoS One* 7:e30288.
44. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.
45. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
46. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321.
47. Bergsten J (2005) A review of long-branch attraction. *Cladistics* 21:163–193.