Protein Engineering, Design & Selection, 2018, pp. 1–9
doi: 10.1093/protein/gzx068
Review

OXFORD

Review

Short linear motif core and flanking regions modulate retinoblastoma protein binding affinity and specificity

Nicolás Palopoli[1,2,†], Nicolás S. González Foutel[3,†], Toby J. Gibson[4], and Lucía B. Chemes[3,5,6,*]

[1]Department of Science and Technology, Universidad Nacional de Quilmes, CONICET. Roque Sáenz Peña 352. CP (B1876BXD), Bernal, Buenos Aires, Argentina, [2]Structural Bioinformatics Unit, Fundación Instituto Leloir and IIBBA-CONICET. Av. Patricias Argentinas 435 CP 1405, Buenos Aires, Argentina, [3]Protein Structure Function and Engineering Laboratory, Fundación Instituto Leloir and IIBBA-CONICET. Av. Patricias Argentinas 435 CP 1405, Buenos Aires, Argentina, [4]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117, Heidelberg, Germany, [5]Instituto de Investigaciones Biotecnológicas IIB-INTECH, Universidad Nacional de San Martín. Av. 25 de Mayo y Francia CP 1650, San Martín, Buenos Aires, Argentina, and [6]Departamento de Fisiología y Biología Molecular y Celular (DFBMC), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160, Ciudad Universitaria, C1428EGA, Argentina

[*]To whom correspondence should be addressed. E-mail:lchemes@iibintech.com.ar
[†]These authors contributed equally to this work.
Paper Edited By: Huan-Xiang Zhou, Board Member for PEDS

Received 21 October 2017; Revised 12 December 2017; Editorial Decision 13 December 2017; Accepted 10 January 2018

Abstract

Pocket proteins retinoblastoma (pRb), p107 and p130 are negative regulators of cellular proliferation and multifunctional proteins regulating development, differentiation and chromatin structure. The retinoblastoma protein is a potent tumor suppressor mutated in a wide range of human cancers, and oncogenic viruses often interfere with cell cycle regulation by inactivating pRb. The LxCxE and pRb AB groove short linear motifs (SLiMs) are key to many pocket protein mediated interactions including host and viral partners. A review of available experimental evidence reveals that several core residues composing each motif instance are determinants for binding. In the LxCxE motif, a fourth hydrophobic position that might allow variable spacing is required for binding. In both motifs, flanking regions including charged stretches and phosphorylation sites can fine-tune the binding affinity and specificity of pocket protein SLiM-mediated interactions. Flanking regions can modulate pocket protein binding specificity, or tune the high affinity interactions of viral proteins that hijack the pRb network. The location of SLiMs within intrinsically disordered regions allows faster evolutionary rates that enable viruses to acquire a functional variant of the core motif by convergent evolution, and subsequently test numerous combinations of flanking regions towards maximizing interaction specificity and affinity. This knowledge can guide future efforts directed at the design of peptide-based compounds that can target pocket proteins to regulate the G1/S cell cycle checkpoint or impair viral mediated pRb inactivation.

Key words: cell cycle checkpoint, flanking regions, LxCxE motif, retinoblastoma, SLiM

Introduction

The human pocket proteins retinoblastoma (pRb), p107 and p130 are central negative regulators of the eukaryotic cell cycle. Pocket proteins act as potent tumor suppressors, especially pRb, which is inactivated in a wide range of human cancers (Burkhart and Sage, 2008). pRb prevents cell proliferation by inactivating E2F transcription factor family members and repressing the expression of S-phase genes. While pocket proteins have partially overlapping functions, distinct pocket protein activities might be mediated by their different E2F transcription factor binding partners (Liban *et al.*, 2016). Mitogenic signaling leads to Cyclin/CDK-mediated phosphorylation of pRb, which causes its dissociation from E2F transcription factors, enabling cell cycle progression. In turn, multiple oncogenic viruses have evolved to inactivate pocket proteins, leading to cell cycle reentry and transformation (de Souza *et al.*, 2009; Davey *et al.* 2011). Despite their central role in cell cycle control, pocket proteins can be best understood as a multifunctional protein family (Dick and Rubin, 2013; Dyson, 2016). In addition to repressing E2F transcription family factors, pRb also regulates cell cycle progression through E2F-independent mechanisms including the pRb–SKP2–p27 pathway, which leads to p27 stabilization, cyclin-dependent kinase inhibition and cell cycle arrest. Pocket proteins regulate development and differentiation across the eukaryotic lineage as well as cell cycle withdrawal, for example through the assembly of the DREAM complex involving p107 and p130 together with MuvB proteins (Dick and Rubin, 2013). Recent studies have uncovered a novel role for pRb in regulating genomic stability, with pocket protein-defective cells presenting abnormal chromosome structure including chromosome missegregation, tangling and aneuploidy. Additional roles include the regulation of apoptosis and the DNA damage response. However, despite their key role in cell cycle regulation and cancer, knowledge of the molecular mechanisms underlying pocket protein functions and specificity remain limited. Most of these functions are mediated through the interaction of pocket proteins with a wide array of protein partners (Morris and Dyson, 2001; Dick, 2007).

Role of SLiMs in pocket protein family function

Short Linear Motifs (SLiMs) are protein sub-sequences usually located within intrinsically disordered (IDP) regions, which makes them accessible for binding (Gouw *et al.*, 2017). SLiMs play a central role in cellular signaling by regulating protein–protein interactions, intracellular targeting, post-translational modifications and degradation (Tompa *et al.*, 2014). Several SLIMs present within pRb, p107 and p130 regulate their function, including Cyclin docking (Lowe *et al.*, 2002), phosphorylation by CDKs (Lees *et al.*, 1991) and other kinases (Inoue *et al.*, 2007; Delston *et al.*, 2011), dephosphorylation by protein phosphatase 1 (Hirschi *et al.*, 2010), binding to regulators of the DNA damage response (Carr *et al.*, 2014), Caspase cleavage (Fattman *et al.*, 2001), nuclear targeting (Fontes *et al.*, 2003) and degradation (Tedesco *et al.*, 2002).

Two SLiMs, the LxCxE and the pRb AB groove motifs, mediate the interaction of multiple target proteins with the central AB domain of pocket proteins. The LxCxE motif binds to the conserved LxCxE-binding cleft in the pRb B domain (Lee *et al.*, 1998) (Fig. 1A). The presence of an LxCxE motif has been reported in at least 30 out of more than 100 reported cellular pRb interactors (Morris and Dyson, 2001; Dick, 2007) and in multiple viral proteins that hijack the pRb network (Gouw *et al.*, 2017). However, an in depth assessment of the available evidence indicates that in many cases the reported LxCxE motifs are known or predicted to be located within folded domains as in the case of Ubiquitin hydrolase USP4 (Luna-Vargas *et al.*, 2011), Chromobox protein CBX5 (Kaustov *et al.*, 2011), and the RBBP9 hydrolase (Vorobiev *et al.*, 2009), which might affect their accessibility and functionality (Gouw *et al.*, 2017). In other cases, including transcription factor UBF-1, DNA repair endonuclease CtIP (RBBP8), heat shock protein HSP75 (TRAP-1), Aryl hydrocarbon receptor AhR, Replication Factor C (RFC-1), and transcription factor ELF1, the biochemical evidence supporting the involvement of the LxCxE motifs is limited. An updated list of proteins proposed to use the LxCxE motif for pRb interaction can be found at the ELM resource (http://elm.eu.org) (Gouw *et al.*, 2017). To date, strong biochemical evidence demonstrating the involvement of LxCxE motifs in pRb binding is available for 18 proteins from mammalian and plant cells, and for 14 proteins from animal and plant viruses, listed in Table I. These motifs have been validated using several independent biochemical techniques, and are known or predicted to be located in unstructured, accessible protein domains or regions. These motifs are found in proteins involved in chromatin regulation and histone modification including histone deacetylases HDAC1/2 and ARI4A, histone demethylase KDM5A, histone methyltransferase PRDM2, the chromatin remodeling complex containing BRG1 and BRM (SMCA 2/4) proteins, transcriptional regulators such as HBP1 and EID1, and plant and animal Cyclin D proteins, among others. Undoubtedly, this list will continue to grow as we learn more about LxCxE-mediated interactions. For example, recent reports show that the interaction between pRb and the anaphase promoting complex APC/C^{CDH1} (Binne *et al.*, 2007) and between pRb and the Condensin II complex (Longworth *et al.*, 2008) are dependent on the LxCxE cleft in pRb, but the specific interactions mediated by LxCxE motifs in each case remain to be identified.

The second pRb binding SLiM is the pRb AB groove motif, found so far only in the E2F transactivation domain (E2F-TD) of transcription factor family members E2F 1–5, as well as in the early E1A protein from Adenovirus (AdE1A) (Table I). The pRb AB groove motif binds to a conserved groove joining the pRb A and B domains (Lee *et al.*, 2002) (Fig. 1C and D) and mediates the E2F-repressive function of pRb family members. This SLiM is mimicked by viral proteins to displace E2F from the pRb–E2F complexes and induce cell cycle entry, and the same surface is also targeted by an intramolecular helical mimic of the pAB groove motif that involves binding of the phosphorylated (S608) pRbAB loop region to the AB cleft (Burke *et al.*, 2012).

Affinity determinants of core LxCxE sequences

Available data indicates that whereas viral LxCxE motifs have pRb affinities in the nanomolar range, cellular LxCxE motifs reveal generally lower affinities, ranging from submicromolar to micromolar (Table II). The first quantitative reports for LxCxE:pRb binding were made by Jones *et al.* for the Human Papillomavirus E7 (HPV16E7) and Adenovirus E1A (AdE1A) proteins (Jones *et al.*, 1990). In this work the authors defined residues 21–29 in HPV16E7 as the minimal pRb-binding module, reporting mid-nanomolar affinities for HPV16E7, and a somewhat lower affinity of the AdE1A LxCxE motif. Later studies using more sensitive techniques yielded higher affinities (dissociation constant K_D ~5 nM) for comparable HPV16 E7 LxCxE sequences (Lee *et al.*, 1998; Singh *et al.*, 2005; Chemes *et al.*, 2010, 2011) and for the related Large T Antigens from Simian virus 40 (SV40) and Merkel Cell Polyomavirus

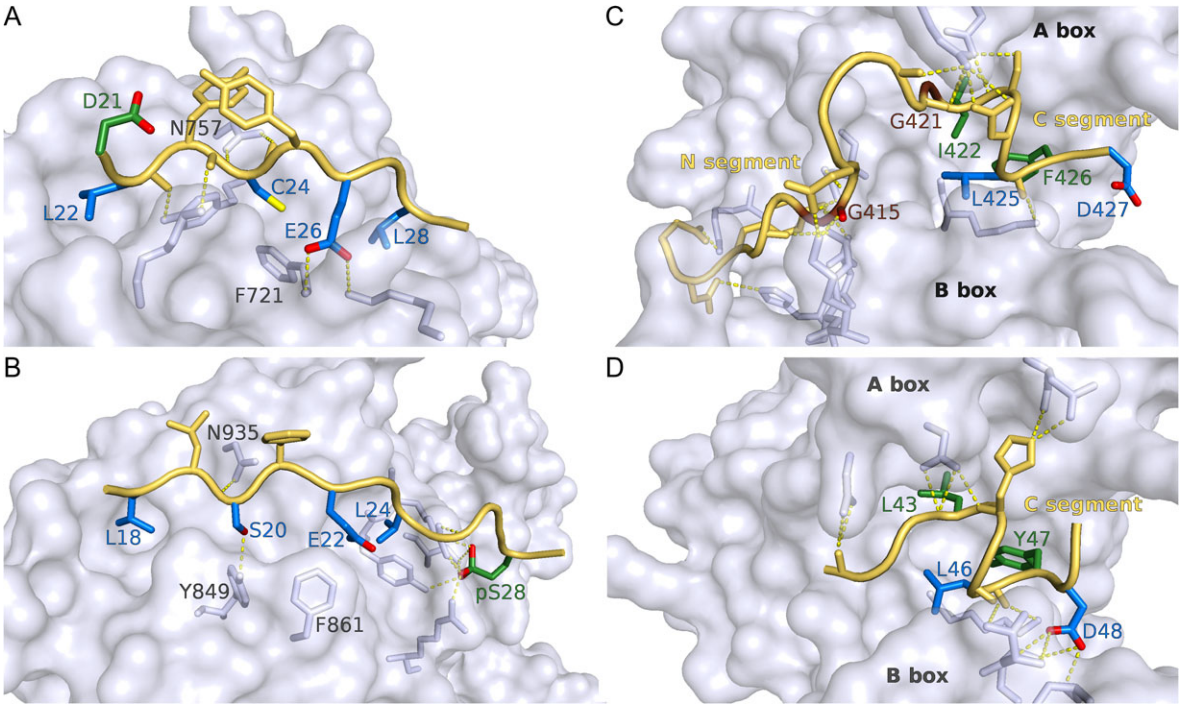

Fig. 1 Representative structures of pocket protein binding SLiMs. (**A**) Structure of the HPV16E7 LxCxE motif bound to the pRb pocket domain-binding cleft (PDB ID: 1GUX). Core residues and L22, C24, E26 and L28 are marked as blue sticks. Other relevant side chains such as D21 are marked as green sticks. Side chains of other motif residues [Y23, Y25] are marked as yellow sticks. (**B**) Structure of the phosphorylated LIN52 LxSxE motif bound to the p107 pocket domain-binding cleft (PDB ID: 4YOS). Core residues L18, S20, E22 and L24 are marked as blue sticks. Other relevant side chains such as phospho-S28 are marked as green sticks. Side chains of other core residues [L19, Y21] are marked as yellow sticks. (**C**) Structure of the E2F2 pRb AB groove motif bound to the pRb AB cleft (PDB ID: 1N4M). The pRb A-Box and B-box sides, the extended N-Segment and the helical C-segment of the E2F2 peptide are marked for reference. Core motif residues I422, L425, F426 and D427 are depicted as blue and green sticks. G415 and G421 in the N-Segment are colored red. G421 allows a kink in the peptide orientation. Side chains from the N-Segment that establish additional intermolecular bonds are marked as yellow sticks. (**D**) Structure of the Adenovirus E1A (AdE1A) pRb AB groove motif bound to the pRb AB cleft (PDB ID: 2R7G). The short segment containing the helical motif is depicted, with core motif residues L43, L46, Y47 and D48 marked as blue and green sticks. Additional peptide residues forming intermolecular hydrogen bonds are depicted as yellow sticks. In all panels, the peptide backbone is shown in cartoon representation, with the residues most relevant for the interaction marked as sticks, and the pRb pocket protein surface displayed in gray tint. Polar, intermolecular hydrogen bonds are depicted as yellow dashed lines connecting to pRb residues shown as gray sticks. Images were rendered with PyMol [The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC]. A color version of the figure is available online.

(MCPyV), although with weaker binding affinities compared to HPV16 E7 (Borchert *et al.*, 2014) (Table II). Among the cellular LxCxE sequences, Histone Deacetylase 1 (HDAC-1) demonstrates a weaker (K_D ~20 μM) and more transient association with the pRb pocket domain (Singh *et al.*, 2005) while the IRCDE motif from the Retinoblastoma protein-interacting zinc finger protein (RIZ) was reported to bind with high-nanomolar affinity to pRb (Sun *et al.*, 2015).

In HPV16 E7, the core motif residues L22 and C24 bind tightly into hydrophobic pockets of the LxCxE groove, while E26 establishes hydrogen bonds to the side chain of residues F721 and K722 in the pRb LxCxE binding cleft (Fig. 1A). Hydrogen bonding of K713 and N757 in pRb to the LxCxE main chain holds the peptide backbone in a fixed position (Fig. 1A). The core L, C, E residues are strong determinants for binding, as shown by 5- to 10-fold decreases in binding affinity upon alanine mutagenesis (Jones *et al.*, 1990; Guiley *et al.*, 2015) (Table II). Both structural and mutagenesis data indicate that a fourth hydrophobic residue located in position +2 to the C-terminal E residue in the LxCxE motif (L28 in HPV16 E7) plays a central role in LxCxE motif binding to pRb (Fig. 2A and B). This hydrophobic residue is conserved in many LxCxE motifs (Fig. 2A) including HPV16E7, SV40LT, HDAC and LIN52 (Table II). Inspection of the LxCxE-pRb structures reveals that the fourth hydrophobic residue binds into a broad binding groove formed by pRb residues V725, F739, I752 and I753 (Lee *et al.*, 1998; Kim *et al.*, 2001) (Fig. 2D), and a similar binding pattern is also observed in the p107/LIN52 LxSxE complex (Guiley *et al.*, 2015). Mutation of the hydrophobic L24 residue in LIN52 is as destabilizing as mutation of core residues L18, C20 and E22 (Table II). The broader configuration of the binding groove might allow for an optimal spacing at +2 with variations that contribute to tuning binding affinity, as in the adenovirus E1A LxCxE sequence, where the hydrophobic residue is at +3 (DLTCHEAGFP) (Fattaey *et al.*, 1993). In support of this human Cyclin D1, which lacks a hydrophobic at +2 and has a hydrophobic residue in position +4 (QLLCCEVETIR) binds very weakly to pRb (Dowdy *et al.*, 1993) and the isolated peptide is unable to bind to p130 or compete for LIN52 binding (Guiley *et al.*, 2015). In comparison Cyclin D2, which has a hydrophobic residue in position +3 (ELLCHEVDPV), binds to pRb more strongly (Ewen *et al.*, 1993). However, further experiments are required to test the specific requirements for the spacing and identity of the fourth hydrophobic position, as natural sequence variants harbor several other substitutions that could explain the changes in binding affinity. Due to the requirement for the fourth hydrophobic position, it has been suggested that the motif be renamed LxCxExφ (Guiley *et al.*, 2015). Because the original name is so widely used, it retrieves

Table I. Experimentally validated LxCxE and pRb AB groove binding motifs

Protein	Uniprot name	Organism	PDB entry	PMID
Cellular LxCxE motifs				
Lysine-specific Demethylase 5A	KDM5A_HUMAN	*Homo sapiens*		1857421
AT-rich interacting domain protein 4A	ARI4A_HUMAN	*Homo sapiens*		1857421; 8414517
Histone Deacetylase 1	HDAC1_HUMAN	*Homo sapiens*		9468139; 16118215
Histone Deacetylase 2	HDAC2_HUMAN	*Homo sapiens*		10496602; 9468139
Kinetochore protein NDC80 homolog	NDC80_HUMAN	*Homo sapiens*		10779342; 18455984
HMG-box containing protein 1	HBP1_RAT	*Rattus novergicus*		9030690; 9178770
EP300-interacting inhibitor of differentiation 1	EID1_HUMAN	*Homo sapiens*		11073990; 11073889
PR domain zinc finger protein 2 (RIZ)	PRDM2_HUMAN	*Homo sapiens*		7538672; 25640033
Protein phosphatase 1 regulatory subunit 26	PPR26_HUMAN	*Homo sapiens*		26442585
Transcription activator SNF2L2 (Brm)	SMCA2_HUMAN	*Homo sapiens*		8657132; 9326598
Transcription activator BRG1	SMCA4_HUMAN	*Homo sapiens*		8657132
Brg1 protein	Q63928_9MURI	*Mus musculus*		7923370
G1/S specific Cyclin D1	CCND1_HUMAN	*Homo sapiens*		19237565; 8490963
G1/S specific Cyclin D2	CCND2_HUMAN	*Homo sapiens*		8343202; 8449399
G1/S specific Cyclin D3	CCND3_HUMAN	*Homo sapiens*		8343202; 8449399
Cyclin D1-1	CCD11_ARATH	*A. Thaliana*		769881
Cyclin D2-1	CCD21_ARATH	*A. Thaliana*		769881
Cyclin D3-1	CCD31_ARATH	*A. Thaliana*		769881
Viral LxCxE motifs				
Large T antigen	LT_SV40	*Simian Virus 40*	1GH6[a]	1122619; 2839300
Large T antigen	B8ZX42_9POLY	*Merkel cell polyomavirus*		24371076; 21413015
E7 oncoprotein	VE7_HPV16	*Human papillomavirus 16*	1GUX[a]; 4YOZ[b]	1312637; 9495340
E1A oncoprotein	E1A_ADE05	*Human adenovirus 5*		8084002; 1534854
UL97 Serine-Threonine Kinase	GCVK_HCMVA	*Human cytomegalovirus (HHV-5)*		18321963
Wsv069	Q77J89_WSSVS	*Shrimp white spot syndrome Virus*		24027329
Wsv056	Q77J94_WSSVS	*Shrimp white spot syndrome Virus*		24027329
MC007 *	Q98178_MCV1	*Molluscum contagiosum virus*		18701596
RepA protein	REPA_BEYDV	*Bean yellow dwarf virus*		10191192
RepA protein	REPA_WDVS	*Wheat dwarf virus*		7664747
RepA protein	REPA_MSVS	*Maize streak virus*		15722542
Clink protein	CLINK_FBNY1	*Faba bean necrotic virus*		10708410
UNK protein	Q9WKM8_BBTV	*Banana bunchy top virus*		10640570
Cellular LxSxE motifs (p107 binding)				
LIN-52	LIN52_HUMAN	*Homo sapiens*	4YOS[b]	25917549
Cellular pRb AB groove motifs				
E2F1 protein	E2F1_HUMAN	*Homo sapiens*	1O9K[a]	8346196; 8413249
E2F2 protein	E2F2_HUMAN	*Homo sapiens*	1N4M[a]	12502741
E2F3 protein	E2F3_HUMAN	*Homo sapiens*		27567532
E2F4 protein	E2F4_HUMAN	*Homo sapiens*		27567532
E2F5 protein	E2F5_HUMAN	*Homo sapiens*		12502741
pRb protein AB Loop	RB_HUMAN	*Homo sapiens*	4ELL[a]	22569856; 20223825
Viral pRb AB groove motifs				
E1A protein	E1A_ADE05	*Human adenovirus 5*	2R7G[a]	17974914; 1534854

All motifs reported in Table I are supported by several sources of biochemical evidence (IP-Western, mutagenesis, *in vitro* binding) and are known or predicted to be located in unstructured regions. A comprehensive curation of LxCxE and pAB groove motifs can be found at http://elm.eu.org.

[a]PDB ID for interaction with Human pRb.

[b]PDB ID for interaction with Human p107.

* The Protein has two LxCxE motifs.

a large literature spanning nearly three decades, and because the final position is likely to allow variable spacing, we are still cautious as to whether renaming the motif would be beneficial. Conservative substitutions in the two hydrophobic core positions (L22I and L28F in HPV16E7) do not alter pRb binding affinity, indicating that different hydrophobic residues might be allowed at these positions (Singh *et al.*, 2005). Mutation of side chains such as Y25 in the HPV16E7 peptide that do not contact pRb but establish intramolecular contacts produce significant decreases in binding affinity, suggesting that preformed conformations present in the free peptide might modulate pRb binding (Singh *et al.*, 2005).

Regions flanking the LxCxE core motif modulate pocket protein binding affinity

Available data suggests that residues outside the core motif modulate pRb binding affinity. For example, Singh *et al.* have shown that the D21R mutation in HPV16E7 produces an eight-fold decrease in pRb-binding affinity. Conversely, the mutation R413D in HDAC-1 increases the affinity for pRb two-fold (Table II), indicating that the presence of an acidic residue in position (–1) with respect to the LxCxE motif favors pRb-binding (Fig. 1A and B). While the acidic residue at (–1) does not have a defined pattern of interaction with

Table II. Affinities reported for Retinoblastoma protein family binding motifs

Protein	Partner	Sequence	Technique[a]	K_D value (M)	Reference
Viral LxCxE sequences					
HPV16 E7 (16–31)	pRb	QPETTDLYCYEQLNDS	Stopped Flow	$(2.10 \pm 0.03) \cdot 10^{-9}$	Chemes *et al.* (2011)
HPV16 E7 (16–40)	pRb	QPETTDLYCYEQLNDSSEEEDEIDG	Stopped Flow	$(3.10 \pm 0.05) \cdot 10^{-9}$	Chemes *et al.* (2011)
HPV16 E7 (16–40) pS31pS32	pRb	QPETTDLYCYEQLNDSSEEEDEIDG	Stopped Flow	$(0.36 \pm 0.08) \cdot 10^{-9}$	Chemes *et al.* (2011)
HPV16 E7 (21–29)	pRb	DLYCYEQLN	F. Spectroscopy	$(4.7 \pm 1.7) \cdot 10^{-9}$	Chemes *et al.* (2010)
HPV16 E7 (16–31)	pRb	QPETTDLYCYEQLNDS	F. Spectroscopy	$(5.1 \pm 1.3) \cdot 10^{-9}$	Chemes *et al.* (2010)
HPV16 E7 (16–40)	pRb	QPETTDLYCYEQLNDSSEEEDEIDG	F. Spectroscopy	$(6.5 \pm 1.0) \cdot 10^{-9}$	Chemes *et al.* (2010)
HPV16 E7 (16–40) pS31pS32	pRb	QPETTDLYCYEQLNDSSEEEDEIDG	F. Spectroscopy	$(1.8 \pm 0.4) \cdot 10^{-9}$	Chemes *et al.* (2010)
HPV16 E7 (21–29)	pRb	DLYCYEQLN	ITC	$(0.11 \pm 0.03) \cdot 10^{-6}$	Lee *et al.* (1998)
HPV16 E7 (21–29)	pRb	DLYCYEQLN	ITC	$(0.19 \pm 0.07) \cdot 10^{-6}$	Singh *et al.* (2005)
HPV16 E7 (21–29) L22I	pRb	DIYCYEQLN	ITC	$(0.32 \pm 0.07) \cdot 10^{-6}$	Singh *et al.* (2005)
HPV16 E7 (21–29) L28F	pRb	DLYCYEQFN	ITC	$(0.14 \pm 0.03) \cdot 10^{-6}$	Singh *et al.* (2005)
HPV16 E7 (21–29) D21R	pRb	RLYCYEQLN	ITC	$(1.05 \pm 0.04) \cdot 10^{-6}$	Singh *et al.* (2005)
HPV16 E7 (21–29) D21R/Y25E	pRb	RLYCEEQLN	ITC	$(4.79 \pm 0.20) \cdot 10^{-6}$	Singh *et al.* (2005)
SV40 LT (102–110)	pRb	NLFCSEEMD	ITC	$(0.44 \pm 0.06) \cdot 10^{-6}$	Singh *et al.* (2005)
SV40 LT (7–117)	pRb	…NLFCSEEMPSSDDEAT	MT	$(1.95 \pm 0.37) \cdot 10^{-6}$	Borchert *et al.* (2014)
MCPyV LT (1–244)	pRb	…EDLFCDESLSSPEPPSSSEEPEEPPSSRSSPRQP	MT	$(5.13 \pm 3.81) \cdot 10^{-8}$	Borchert *et al.* (2014)
Cellular LxCxE sequences					
HDAC (409–428)	pRb	SSDKRIACEEEFSDSEEEGE	ITC	$(10.0 \pm 3.0) \cdot 10^{-6}$	Singh *et al.* (2005)
HDAC1 (413–421)	pRb	RIACEEEFS	ITC	$(20.0 \pm 4.0) \cdot 10^{-6}$	Singh *et al.* (2005)
HDAC1 (413–421) F420L	pRb	RIACEEELS	ITC	No binding	Singh *et al.* (2005)
HDAC1 (413–421) R413D	pRb	DIACEEEFS	ITC	$(10.0 \pm 5.0) \cdot 10^{-6}$	Singh *et al.* (2005)
HDAC1 (413–421) R413D/E417Y	pRb	DIACYEEFS	ITC	$(3.27 \pm 0.07) \cdot 10^{-6}$	Singh *et al.* (2005)
RIZ AR (297–341)	pRb	**…**EIRCDEKPEDLLEEPKTTSEETLEDCSECTPAM	ITC	$(6.40 \pm 0.70) \cdot 10^{-7}$	Sun *et al.* (2015)
RIZ AR (297–341)	pRb	**…**EIRCDEKPEDLLEEPKTTSEETLEDCSECTPAM	F. Spectroscopy	$1.10 \cdot 10^{-7}$	Sun *et al.* (2015)
RIZ (309–319)	pRb	EIRCDEKPEDL	F. Spectroscopy	$3.60 \cdot 10^{-7}$	Sun *et al.* (2015)
LIN52 (12–34)	pRb	TDLEASLLSFEKLDRASPDLWPE	ITC	No binding	Guiley *et al.* (2015)
LIN52 (12–34) pS28	pRb	TDLEASLLSFEKLDRASPDLWPE	ITC	No binding	Guiley *et al.* (2015)
LIN52 (12–34)	p107	TDLEASLLSFEKLDRASPDLWPE	ITC	$(5.9 \pm 0.9) \cdot 10^{-6}$	Guiley *et al.* (2015)
LIN52 (12–34) pS28	p107	TDLEASLLSFEKLDRASPDLWPE	ITC	$(1.4 \pm 0.9) \cdot 10^{-6}$	Guiley *et al.* (2015)
LIN52 (12–34) pS28 L18A	p107	TDLEASALSFEKLDRASPDLWPE	ITC	$(14.9 \pm 0.1) \cdot 10^{-6}$	Guiley *et al.* (2015)
LIN52 (12–34) pS28 E22A	p107	TDLEASLLSFAKLDRASPDLWPE	ITC	$(5.5 \pm 0.3) \cdot 10^{-6}$	Guiley *et al.* (2015)
LIN52 (12–34) pS28 L24A	p107	TDLEASLLSFEKADRASPDLWPE	ITC	$(7.4 \pm 0.9) \cdot 10^{-6}$	Guiley *et al.* (2015)
LIN52 (12–34) pS28	p130	TDLEASLLSFEKLDRASPDLWPE	ITC	$(1.0 \pm 0.1) \cdot 10^{-6}$	Guiley *et al.* (2015)
Viral pRb AB groove sequences					
E1A–CR1-linker (37–121)	pRb	…HFEPPTLHELYDL…	ITC	$(0.9 \pm 0.3) \cdot 10^{-6}$	Liu
Cellular pRb AB groove sequences					
E2F-2 (410–427)	pRb	DDYLWGLEAGEGISDLFD	ITC	$(0.19 \pm 0.04) \cdot 10^{-6}$	Lee *et al.* (2002)
E2F-2 (410–427) E420A	pRb	DDYLWGLEAGAGISDLFD	ITC	$(0.32 \pm 0.04) \cdot 10^{-6}$	Lee *et al.* (2002)
E2F-2 (410–427) G421S	pRb	DDYLWGLEAGESISDLFD	ITC	$(10.2 \pm 1.4) \cdot 10^{-6}$	Lee *et al.* (2002)
E2F-5 (323–346)	pRb	DDYNFNLDDNEGVCDLFD	ITC	$(0.69 \pm 0.1) \cdot 10^{-6}$	Lee *et al.* (2002)
E2F-1 (409–426)	pRb	LDYHFGLEEGEGIRDLFD	ITC	$(0.34 \pm 0.02) \cdot 10^{-6}$	Xiao *et al.* (2003)
E2F-1 (380–437)	pRb	…LDYHFGLEEGEGIRDLFD…	ITC	$(0.16 \pm 0.01) \cdot 10^{-6}$	Xiao *et al.* (2003)

(*Continued*)

Table II. Continued

Protein	Partner	Sequence	Technique[a]	K_D value (M)	Reference
E2F-1 (409–426)	pRb	LDYHFGLEEGEGIRDLFD	ITC	$(0.07 \pm 0.01) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-1 (409–426)	p107	LDYHFGLEEGEGIRDLFD	ITC	$(1.06 \pm 0.04) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-2 (410–427)	pRb	DDYLWGLEAGEGISDLFD	ITC	$(0.04 \pm 0.01) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-2 (410–427) G415N	pRb	DDYLWNLEAGEGISDLFD	ITC	$(0.38 \pm 0.02) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-2 (410–427)	p107	DDYLWGLEAGEGISDLFD	ITC	$(3.1 \pm 0.3) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-2 (410–427) G415N	p107	DDYLWNLEAGEGISDLFD	ITC	$(8.3 \pm 0.5) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-3 (432–449)	pRb	EDYLLSLGEEGISDLFD	ITC	$(1.2 \pm 0.1) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-3 (432–449)	p107	EDYLLSLGEEGISDLFD	ITC	$(7 \pm 1) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-4 (390–407)	pRb	HDYIYNLDESEGVCDLFD	ITC	$(0.5 \pm 0.2) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-4 (390–407) N395G	pRb	HDYIYGLDESEGVCDLFD	ITC	$(0.10 \pm 0.02) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-4 (390–407)	p107	HDYIYNLDESEGVCDLFD	ITC	$(0.15 \pm 0.01) \cdot 10^{-6}$	Liban *et al.* (2016)
E2F-4 (390–407) N395G	p107	HDYIYGLDESEGVCDLFD	ITC	$(0.15 \pm 0.01) \cdot 10^{-6}$	Liban *et al.* (2016)

[a]FI spectroscopy, fluorescence intensity or anisotropy; ITC, isothermal titration calorimetry; MT, microscale thermophoresis.

pRb, it is likely to contribute to binding or the correct positioning of the motif by charge complementarity to basic residues K713 and K765 located in the rim of the LxCxE cleft (Fig. 2C). In addition, several viral oncoproteins and cellular proteins such as HPVE7, SV40 LT and HDAC-1, present a negatively charged stretch following the LxCxE motif (Fig. 2A) that establishes electrostatic interactions with a complementary positively charged surface that surrounds the LxCxE binding cleft in pocket proteins (Chemes *et al.*, 2011) (Fig. 2C). This tenet is supported by experimental data showing that LxCxE peptides from HPV16E7, HDAC-1 and RIZ that include the acidic segments show higher binding affinities (Table II) (Singh *et al.*, 2005; Chemes *et al.*, 2011; Sun *et al.*, 2015). In addition, the interaction between pRb and HPV16E7 or Polyomavirus Large T antigens is weakened upon charge screening at high salt concentration, indicating that electrostatic interactions stabilize the pRb–LxCxE interaction (Chemes *et al.*, 2011; Borchert *et al.*, 2014).

Viral transforming proteins HPV16 E7, MPCyV LT, SV40 LT, AdE1A and also cellular HDAC-1 present serine-phosphorylation sites within the acidic stretch following the LxCxE motif (Chemes *et al.*, 2010). It has been reported that in vitro phosphorylation of these sites increases the affinity for pRb (Table II). Chemes *et al.* have shown that phosphorylation in S31 and S32 increases HPV16 E7 affinity for pRb pocket domain four-fold compared to the unphosphorylated HPV16 E7 fragment (Chemes *et al.*, 2010, 2011). A similar behavior is also observed when S132 from the AdE1A LxCxE motif is phosphorylated (Whalen *et al.*, 1996). Schrama et al. (2016) have reported that the S220A mutation in MCPyV-LT weakens pRb binding, and phosphorylation proximal to a non-canonical LxSxE motif in the LIN52 protein enhances binding to p107 (Guiley *et al.*, 2015). The functional relevance of these phosphorylation sites is underscored by their requirement for viral mediated cell transformation (Whalen *et al.*, 1996; Schrama *et al.*, 2016), viral protein nuclear shuttling (Rihs *et al.*, 1991) and HDAC-1 histone deacetylase activity (Pflum *et al.*, 2001).

Specificity determinants for LxCxE motif–pocket protein family interactions

Recent structural and biochemical work has shed light on determinants of pocket protein binding specificity by LxCxE motifs. Guiley *et al.* have shown that a non-canonical LxSxE motif in the LIN52 protein mediates interactions with the pocket domains of p107 and p130, as part of the highly conserved DREAM complex that regulates differentiation, cell proliferation and tumor suppression (Guiley *et al.*, 2015). This LxSxE motif differs from the LxCxE motif due to the presence of the more polar hydroxyl group in the serine side chain replacing the cysteine, which changes the backbone and side chain hydrogen bonding pattern, disrupting hydrogen bonds between N935 and the peptide backbone in p107, and between the side chains of the LIN52 E22 residue and F861 in p107 yielding a suboptimal binding affinity (Fig. 1B). However, LIN52 phosphorylation at S28, following the core motif, increases the binding affinity through an additional stabilizing hydrogen bond network with the p107 and p130 pocket domains at a site adjacent to the LxCxE cleft (Table II and Fig. 1B) (Guiley *et al.*, 2015). Differences in the residues lining this adjacent site in p130/p107 and pRb lead to the phosphorylated LIN52 LxSxE motif being able to bind to p130/p107 but not to pRb, determining the pocket protein binding specificity of LxCxE and LxSxE motifs, and providing a

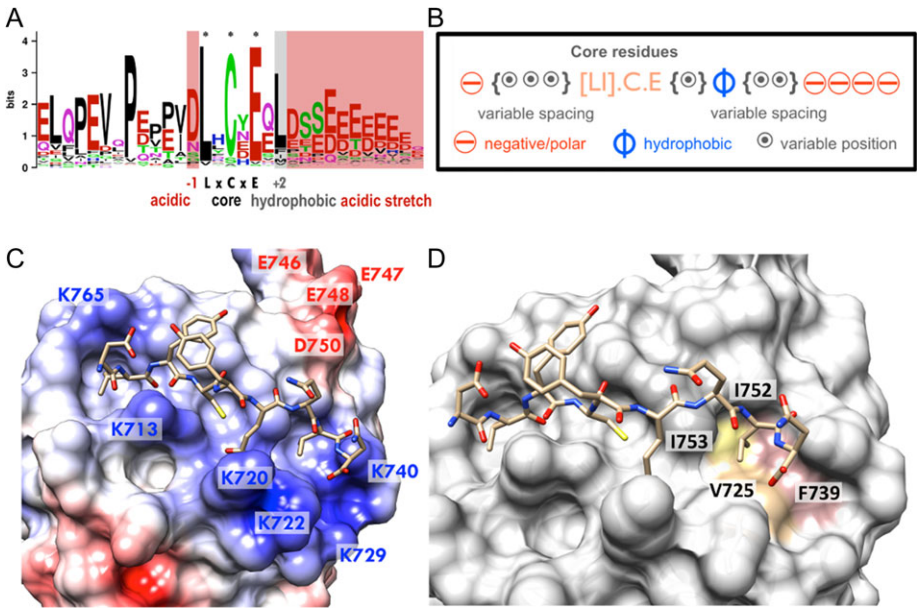

Fig. 2 Core and flanking regions of the LxCxE motif. (**A**) Sequence logo of the Papillomavirus E7 oncoprotein showing strong conservation of the L, C, E core residues as well as of the flanking regions including an acidic residue in position (–1) to the core motif, a fourth hydrophobic residue in position (+2) to the core motif, and an acidic stretch with phosphorylatable serines C-terminal to the core motif. Adapted from Chemes *et al.* (2012a). (**B**) Schematic representation of the LxCxE motif core and its modulatory flanking regions. Many of the flanking elements can be at variable distances from the core motif, and the acidic stretch can contribute variable negative charge. (**C**) Representation of the LxCxE motif (sticks) bound to the pRb LxCxE cleft [PDB ID: 1GUX]. The rendering highlights the surface electrostatic potential (blue = positive charge; red = negative charge). A rim of positively charged Lysine residues (labeled and depicted in blue) surrounds the LxCxE binding groove. The image was rendered using Coulombic surface coloring in the UCSF Chimera package (Pettersen *et al.*, 2004). (**D**) Representation of the LxCxE motif (sticks) bound to the pRb LxCxE cleft [PDB ID: 1GUX]. The rendering highlights the position and identity of hydrophobic residues from the retinoblastoma B domain that form the binding pocket for the hydrophobic L28 residue in HPV16E7 (V725: pale orange; F739: pink; I752: yellow and I753: dark gray). A color version of the figure is available online.

mechanism through which phosphorylation regulates LxSxE–pocket protein interactions.

Binding specificity of the pRb AB groove motif and its flanking regions

Pocket protein binding to E2F transcription factors regulates cell-cycle dependent gene expression. The pRb AB groove motif is a short amphipathic helical stretch that binds at the interface of the A and B cyclin folds and is present in E1A from Adenovirus and the members of eukaryotic E2F transcription factors family (E2F1 to E2F5). To date, a single quantitative report for viral E1A-pRb binding affinity that includes the minimal pRb AB groove motif plus 70C-terminal residues reveals a micromolar K_D (Liu and Marmorstein, 2007) whereas reported eukaryotic E2F family transcription factor K_{Ds} have tighter binding affinities, in the range of nanomolar to submicromolar (Lee *et al.*, 2002; Xiao *et al.*, 2003; Liban *et al.*, 2016) (Table II). In E2F transcription factors, the Rb binding surface is bipartite, and involves an extended N-terminal segment followed by the helical C-terminal pRb AB groove motif (Fig. 1C)

The motif core is formed by three hydrophobic residues in viral E1A (L43, L46 and Y47) and the E2F2 transcription factor (I422, L425 and F426), which form one face of the amphipathic helix that binds to a conserved hydrophobic surface in pRb, establishing van der Waals interactions with conserved residues L476, F482 and K530 in pRb (Lee *et al.*, 2002) (Fig. 2C and D). Residues L43 and Y47 in E1A are critical for binding to pRb and for displacement of E2F by E1A upon infection (Liu and Marmorstein, 2007). Additional determinants in E1A include the conserved D48 acidic residue that establishes hydrogen bond interactions with residues in the pRb B domain. The E1A-specific H44 residue establishes a hydrogen bond network with the main chain and side chain of pRb residues E464 and N472 (Fig. 1D) causing significant rearrangements of N472 (Liu and Marmorstein, 2007). The pattern of extensive main chain and side chain bonds seems to extend beyond the core motif residues both in E1A and the E2F family (Fig. 2C and D). Despite having a lower affinity pRb AB groove motif, additional pRb binding sites in the disordered E1A protein include the LxCxE motif, enabling a cooperative and high affinity bipartite interaction that allows effective displacement of E2F, leading to viral-mediated cell cycle deregulation (Fattaey *et al.*, 1993).

Additional binding elements in E2F include an extended N-terminal segment exclusive to the E2F transcription factors that extends the minimal helical motif, increasing binding affinity by establishing additional hydrophobic and hydrogen bonding interactions with residues in both the A-box and B-box sections of the pRb groove (Lee *et al.*, 2002; Xiao *et al.*, 2003) (Table II and Fig. 2C). The E2F2 G421 residue does not make intermolecular contacts with pRb but produces a sharp kink that orients the N-terminal and C-terminal segnemts. Correspondingly, mutation G421S in E2F2 disrupts the flexibility of the hinge region joining the N-terminal segment and the pRb AB groove motif, decreasing pRb binding affinity (Lee *et al.*, 2002) (Table II and Fig. 2C). Pocket proteins exhibit specific preferences for binding to different E2F transcription factors, so whereas pRb has a preference for E2F1 and E2F2, p107 and p130 bind almost exclusively to E2F4 and E2F5. Structural and biochemical studies have shown that while some of the selectivity for E2F family members can be explained by differences in the pocket protein binding cleft that binds to the helical C-terminal segment, natural substitutions in the sequence of the N-terminal segment act as

additional specificity determinants for pocket protein binding, such as positions G415 in E2F2 and N395 in E2F4 (Liban *et al.*, 2016) (Table II and Fig. 2C). Additional pocket protein specificity determinants map to the interaction of the coiled-coil and marked box domain (CM) of E2F/DP with the pRb C terminal domain (CTD). The CM–CTD interactions contribute to E2F partner specificity and are differentially regulated by CTD phosphorylation patterns (Liban *et al.*, 2017). In this scenario, pRb seems to have 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.

Interestingly, the AB groove can also be bound by pRb itself as part of the mechanism of Cyclin/CDK-mediated Rb inactivation. The flexible loop between the pRb A and B cyclin folds carries a CDK phosphorylation site (S608) that can be recognized by Cyclin-dependent kinase 4 (Inoue *et al.*, 2007). Upon phosporylation of S608, this loop gains order and binds the pRb pocket domain, mimicking the E2F helical motif and establishing additional hydrogen bonds that block the binding site for E2F (Burke *et al.*, 2012), resembling the E1A–CR1 interaction (Liu and Marmorstein, 2007) and leading to E2F displacement and activation.

Role of flanking regions in SLiM-mediated interactions

Domain–SLiM interactions have generally been studied by looking at the binding preferences between positions in the motif and the residues occupying the complementary surface in the globular partner. The Eukaryotic Linear Motif database (ELM) (Dinkel *et al.*, 2016) is the most comprehensive, curated catalog of known SLiMs in eukaryotes and their pathogens. It lists almost 300 classes of SLiMs, each populated by several motif instances that most likely differ in sequence but share physicochemical properties at key positions. This empirical pattern of amino acid variability, represented as a regular expression of conserved and degenerate positions, defines the motif core that is mostly responsible for the binding affinity.

The experimental evidence presented above for the binding preferences of two retinoblastoma pocket protein binding SLiMs adds to a growing body of evidence indicating that besides the motif core, flanking positions that are in contact with the partner can be relevant for defining specificity and functionality of the motif (Van Roey *et al.*, 2014). These SLiM modulatory regions may resemble conservation patterns typical of core positions, and they can also comprise regions of low complexity and variable length, located within structurally disordered segments (Davey *et al.*, 2012). The modulatory regions may also carry residues that are targets for phosphorylation, acting as regulatory switches of activity (Akiva *et al.*, 2012).

Five amino acids in the canonical LxCxE expression represent the core of the pRb-binding motif. As reviewed above, its binding affinity seems to be finely tuned by variable stretches of negatively charged amino acids at both sides, plus a hydrophobic residue over-represented immediately on the C-terminal side (Chemes *et al.*, 2015) (Fig. 2B). Visual inspection of the available LxCxE peptide structures bound to pRb or p107 (in which a serine substitutes the core cysteine), reveals a positively charged rim surrounding the LxCxE binding cleft that complements the acidic stretch following the LxCxE motif (Fig. 2C) (Lee *et al.*, 1998; Guiley *et al.*, 2015). Differences in binding affinity of up to three orders of magnitude between LxCxE instances (Table II) may thus be explained by the combination of a disordered core LxCxE motif, necessary but likely insufficient for high affinity interaction, with subtle or defining contributions from modulatory contacts in flanking regions that include a fourth hydrophobic residue and acidic stretches on both sides of the motif (Fig. 2B). Fine-tuning of these modulatory elements might contribute to establishing the specificity of LxCxE-mediated interactions and the high interaction affinities of viral LxCxE containing proteins.

Flanking sequences modulate the affinity of other well-studied motif-mediated interactions. An analysis of experimental data on peptides binding yeast Src homology-3 (SH3) domains revealed that even for core motif instances that perfectly comply with the canonical definition, the binding specificities are up or down modulated by their flanking positions (Gorelik and Davidson, 2012; Kelil *et al.*, 2016), usually enriched with a positively charged residue (Teyra *et al.*, 2012). Neighboring positions affect SH2 binding capacities in a similar manner (Liu *et al.*, 2010). Human PDZ domain interactions are also strongly affected by the context of the core binding motif (Luck *et al.*, 2012). A comparative study using phage display peptide libraries showed that up to seven residues upstream of the core motif may contribute to binding energy, with different effects on affinity and specificity for each family member (Zhang *et al.*, 2006). These results agree with a large-scale computational assessment on binding energy contributions in many SLiMs of known structure, which suggested that the core motif is needed to secure the interaction but the modulator regions are essential for specificity (Stein and Aloy, 2008).

Disorder and evolution of flanking regions

Most SLiMs occur in natively disordered polypeptide regions, which favors the *de novo* acquisition of linear motifs as the sequence can be resampled by point mutations without affecting a folded structure (Davey *et al.*, 2015). SLiMs often show a higher level of evolutionary conservation compared with their immediate sequence context (Davey *et al.*, 2009). However, the region immediately surrounding the motif can also carry a rich conservation signal that may reflect structural constraints even after changes in sequence (Stein and Aloy, 2008; Davey *et al.*, 2009). It has been shown that flanking regions share similar local sequence and structural propensities of core motifs, which points to a concerted evolution among both regions (Chemes *et al.*, 2012a, 2012b), possibly for improved regulation (Chica *et al.*, 2009). As SLiMs are key elements of signaling pathways, viruses frequently take control of the host cell machinery by mimicking the natural motif (Davey, 2011). Faster evolutionary rates allow viruses not only to acquire a functional variant of the core motif by convergent evolution, but also to easily test numerous combinations of flanking regions towards maximizing interaction specificity (Chemes *et al.*, 2015). It is therefore possible that viruses and other pathogens can enhance the binding affinity by extending the number of flanking positions that favor the interaction, even when individual contributions towards binding energy are small.

Concluding remarks

A comprehensive review of available structural and biochemical data on pocket protein binding SLiMs including the LxCxE and pRb AB groove motifs found in host and viral proteins reveals the presence of a well-defined motif core that determines binding, which is modulated by flanking regions that contribute to fine tuning of binding affinity and specificity. This adds to a growing body of evidence showing that the sequence context of a given SLiM might have strong effects on binding preferences and functionality. This

knowledge can guide future efforts directed at the design of peptide-based compounds that can target pocket proteins to regulate the G1/S cell cycle checkpoint or impair viral mediated pRb inactivation.

Acknowledgements

L.B.C. and N.P. hold Research Positions from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). N.G.F. is a CONICET predoctoral fellow.

Funding

This work was supported by Ministerio de Ciencia, Tecnología e Innovación Productiva and Deutscher Akademischer Austauschdienst (MinCyT-DAAD) [Grant CyCMotif DA/16/05 to L.B.C. and T.J.G.]; Agencia Nacional de Promocion Científica y Tecnológica (ANPCyT) [Grants PICT 2013/1985 to L.B.C. and PICT 2015/3367 to N.P.]; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) [Grant PIP 2014/2016 11220130100558CO to L.B.C.].

References

Akiva,E., Friedlander,G., Itzhaki,Z. and Margalit,H. (2012) *PLoS Comput. Biol.*, **8**, e1002341.

Binne,U.K., Classon,M.K., Dick,F.A., Wei,W., Rape,M., Kaelin, W.G., Jr, Naar,A.M. and Dyson,N.J. (2007) *Nat. Cell. Biol.*, **9**, 225–232.

Borchert,S., Czech-Sioli,M., Neumann,F., Schmidt,C., Wimmer,P., Dobner, T., Grundhoff,A. and Fischer,N. (2014) *J. Virol.*, **88**, 3144–3160.

Burke,J.R., Hura,G.L. and Rubin,S.M. (2012) *Genes Dev.*, **26**, 1156–1166.

Burkhart,D.L. and Sage,J. (2008) *Nat. Rev. Cancer*, **8**, 671–682.

Carr,S.M., Munro,S., Zalmas,L.P. *et al.* (2014) *Proc. Natl. Acad. Sci. USA*, **111**, 11341–11346.

Chemes,L.B., de Prat-Gay,G. and Sanchez,I.E. (2015) *Curr. Opin. Struct. Biol.*, **32**, 91–101.

Chemes,L.B., Glavina,J., Alonso,L.G., Marino-Buslje,C., de Prat-Gay,G. and Sanchez,I.E. (2012a) *PLoS ONE*, **7**, e47661.

Chemes,L.B., Glavina,J., Faivovich,J., de Prat-Gay,G. and Sanchez,I.E. (2012b) *J. Mol. Biol.*, **422**, 336–346.

Chemes,L.B., Sanchez,I.E. and de Prat-Gay,G. (2011) *J. Mol. Biol.*, **412**, 267–284.

Chemes,L.B., Sanchez,I.E., Smal,C. and de Prat-Gay,G. (2010) *FEBS J.*, **277**, 973–988.

Chica,C., Diella,F. and Gibson,T.J. (2009) *PLoS ONE*, **4**, e6052.

Davey,N.E., Cyert,M.S. and Moses,A.M. (2015) *Cell. Commun. Signal.*, **13**, 43.

Davey,N.E., Shields,D.C. and Edwards,R.J. (2009) *Bioinformatics*, **25**, 443–450.

Davey,N.E., Trave,G. and Gibson,T.J. (2011) *Trends Biochem. Sci.*, **36**, 159–169.

Davey,N.E., Van Roey,K., Weatheritt,R.J. *et al.* (2012) *Mol. Biosyst.*, **8**, 268–281.

de Souza,R.F., Iyer,L.M. and Aravind,L. (2009) *Biochim. Biophys. Acta*, **1799**, 302–318.

Delston,R.B., Matatall,K.A., Sun,Y., Onken,M.D. and Harbour,J.W. (2011) *Oncogene*, **30**, 588–599.

Dick,F.A. (2007) *Cell. Div.*, **2**, 26.

Dick,F.A. and Rubin,S.M. (2013) *Nat. Rev. Mol. Cell. Biol.*, **14**, 297–306.

Dinkel,H., Van Roey,K., Michael,S. *et al.* (2016) *Nucleic Acids Res.*, **44**, D294–D300.

Dowdy,S.F., Hinds,P.W., Louie,K., Reed,S.I., Arnold,A. and Weinberg,R.A. (1993) *Cell*, **73**, 499–511.

Dyson,N.J. (2016) *Genes Dev.*, **30**, 1492–1502.

Ewen,M.E., Sluss,H.K., Sherr,C.J., Matsushime,H., Kato,J. and Livingston,D. M. (1993) *Cell*, **73**, 487–497.

Fattaey,A.R., Harlow,E. and Helin,K. (1993) *Mol. Cell. Biol.*, **13**, 7267–7277.

Fattman,C.L., Delach,S.M., Dou,Q.P. and Johnson,D.E. (2001) *Oncogene*, **20**, 2918–2926.

Fontes,M.R., Teh,T., Jans,D., Brinkworth,R.I. and Kobe,B. (2003) *J. Biol. Chem.*, **278**, 27981–27987.

Gorelik,M. and Davidson,A.R. (2012) *J. Biol. Chem.*, **287**, 9168–9177.

Gouw,M., Michael,S., Samano-Sanchez,H. *et al.* (2017) *Nucleic Acids Res.* 10.1093/nar/gkx1077.

Guiley,K.Z., Liban,T.J., Felthousen,J.G., Ramanan,P., Litovchick,L. and Rubin,S.M. (2015) *Genes Dev.*, **29**, 961–974.

Hirschi,A., Cecchini,M., Steinhardt,R.C., Schamber,M.R., Dick,F.A. and Rubin,S.M. (2010) *Nat. Struct. Mol. Biol.*, **17**, 1051–1057.

Inoue,Y., Kitagawa,M. and Taya,Y. (2007) *EMBO J.*, **26**, 2083–2093.

Jones,R.E., Wegrzyn,R.J., Patrick,D.R. *et al.* (1990) *J. Biol. Chem.*, **265**, 12782–12785.

Kaustov,L., Ouyang,H., Amaya,M. *et al.* (2011) *J. Biol. Chem.*, **286**, 521–529.

Kelil,A., Levy,E.D. and Michnick,S.W. (2016) *Proc. Natl. Acad. Sci. USA*, **113**, E3862–E3871.

Kim,H.Y., Ahn,B.Y. and Cho,Y. (2001) *EMBO J.*, **20**, 295–304.

Lee,C., Chang,J.H., Lee,H.S. and Cho,Y. (2002) *Genes Dev.*, **16**, 3199–3212.

Lee,J.O., Russo,A.A. and Pavletich,N.P. (1998) *Nature*, **391**, 859–865.

Lees,J.A., Buchkovich,K.J., Marshak,D.R., Anderson,C.W. and Harlow,E. (1991) *EMBO J.*, **10**, 4279–cdc4290.

Liban,T.J., Medina,E.M., Tripathi,S., Sengupta,S., Henry,R.W., Buchler,N.E. and Rubin,S.M. (2017) *Proc. Natl. Acad. Sci. USA*, **114**, 4942–4947.

Liban,T.J., Thwaites,M.J., Dick,F.A. and Rubin,S.M. (2016) *J. Mol. Biol.*, **428**, 3960–3971.

Liu,B.A., Jablonowski,K., Shah,E.E., Engelmann,B.W., Jones,R.B. and Nash, P.D. (2010) *Mol. Cell. Proteomics*, **9**, 2391–2404.

Liu,X. and Marmorstein,R. (2007) *Genes Dev.*, **21**, 2711–2716.

Longworth,M.S., Herr,A., Ji,J.Y. and Dyson,N.J. (2008) *Genes Dev.*, **22**, 1011–1024.

Lowe,E.D., Tews,I., Cheng,K.Y., Brown,N.R., Gul,S., Noble,M.E., Gamblin, S.J. and Johnson,L.N. (2002) *Biochemistry*, **41**, 15625–15634.

Luck,K., Charbonnier,S. and Trave,G. (2012) *FEBS Lett.*, **586**, 2648–2661.

Luna-Vargas,M.P., Faesen,A.C., van Dijk,W.J., Rape,M., Fish,A. and Sixma, T.K. (2011) *EMBO Rep.*, **12**, 365–372.

Morris,E.J. and Dyson,N.J. (2001) *Adv. Cancer Res.*, **82**, 1–54.

Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) *J. Comput. Chem.*, **25**, 1605–1612.

Pflum,M.K., Tong,J.K., Lane,W.S. and Schreiber,S.L. (2001) *J. Biol. Chem.*, **276**, 47733–47741.

Rihs,H.P., Jans,D.A., Fan,H. and Peters,R. (1991) *EMBO J.*, **10**, 633–639.

Schrama,D., Hesbacher,S., Angermeyer,S. *et al.* (2016) *Int. J. Cancer*, **138**, 1153–1162.

Singh,M., Krajewski,M., Mikolajka,A. and Holak,T.A. (2005) *J. Biol. Chem.*, **280**, 37868–37876.

Stein,A. and Aloy,P. (2008) *PLoS ONE*, **3**, e2524.

Sun,Y., Stine,J.M., Atwater,D.Z., Sharmin,A., Ross,J.B. and Briknarova,K. (2015) *Biochemistry*, **54**, 1390–1400.

Tedesco,D., Lukas,J. and Reed,S.I. (2002) *Genes Dev.*, **16**, 2946–2957.

Teyra,J., Sidhu,S.S. and Kim,P.M. (2012) *FEBS Lett.*, **586**, 2631–2637.

Tompa,P., Davey,N.E., Gibson,T.J. and Babu,M.M. (2014) *Mol. Cell.*, **55**, 161–169.

Van Roey,K., Uyar,B., Weatheritt,R.J., Dinkel,H., Seiler,M., Budd,A., Gibson,T.J. and Davey,N.E. (2014) *Chem. Rev.*, **114**, 6733–6778.

Vorobiev,S.M., Su,M., Seetharaman,J. *et al.* (2009) *Proteins*, **74**, 526–529.

Whalen,S.G., Marcellus,R.C., Barbeau,D. and Branton,P.E. (1996) *J. Virol.*, **70**, 5373–5383.

Xiao,B., Spencer,J., Clements,A. *et al.* (2003) *Proc. Natl. Acad Sci. USA*, **100**, 2363–2368.

Zhang,Y., Yeh,S., Appleton,B.A. *et al.* (2006) *J. Biol. Chem.*, **281**, 22299–22311.