



HAL
open science

Conformational buffering underlies functional selection in intrinsically disordered protein regions

Nicolás González-Foutel, Juliana Glavina, Wade Borchers, Matías Safranchik, Susana Barrera-Vilarmau, Amin Sagar, Alejandro Estaña, Amelie Barozet, Nicolás Garrone, Gregorio Fernandez-Ballester, et al.

► To cite this version:

Nicolás González-Foutel, Juliana Glavina, Wade Borchers, Matías Safranchik, Susana Barrera-Vilarmau, et al.. Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nature Structural and Molecular Biology*, 2022, 29 (8), pp.781-790. 10.1038/s41594-022-00811-w . hal-03767337

HAL Id: hal-03767337

<https://laas.hal.science/hal-03767337>

Submitted on 1 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Conformational buffering underlies functional selection in intrinsically**
2 **disordered protein regions**

3
4 Nicolas S. Gonzalez-Foutel^{1,2†}, Juliana Glavina^{1,3†}, Wade M. Borchers⁴, Matías Safranchik¹,
5 Susana Barrera-Vilarmau^{5,6}, Amin Sagar⁷, Alejandro Estaña^{7,8}, Amelie Barozet⁸, Nicolás A.
6 Garrone¹, Gregorio Fernandez-Ballester⁹, Clara Blanes-Mira⁹, Ignacio E Sánchez³, Gonzalo de
7 Prat-Gay², Juan Cortés⁸, Pau Bernadó⁷, Rohit V. Pappu^{5*}, Alex S. Holehouse^{5,10*}
8 Gary W. Daughdrill^{4*} and Lucía B. Chemes^{1,2*}

9
10 ¹ Instituto de Investigaciones Biotecnológicas (IIBiO-CONICET), Universidad Nacional de San Martín, Av. 25 de
11 Mayo y Francia, CP1650 Buenos Aires, Argentina

12 ² Fundación Instituto Leloir e Instituto de Investigaciones Bioquímicas (IIB-CONICET), Av. Patricias Argentinas 435,
13 CP1405 Ciudad Autónoma de Buenos Aires, Argentina

14 ³ Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN-CONICET), Universidad
15 de Buenos Aires, Ciudad Universitaria CP1428 Buenos Aires Argentina

16 ⁴ Department of Cell Biology, Microbiology, and Molecular Biology and, University of South Florida. Tampa, Florida

17 ⁵ Department of Biomedical Engineering, Center for Science & Engineering of Living Systems, Washington University
18 in St. Louis, St. Louis, MO 63130, USA

19 ⁶ Instituto de Química Avanzada de Cataluña (IQAC-CSIC), Jordi Girona, 18-26, 08034, Barcelona, Spain

20 ⁷ Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, 34090 Montpellier, France

21 ⁸ LAAS-CNRS, Université de Toulouse, CNRS, 31400 Toulouse, France

22 ⁹ Instituto de Investigación, Desarrollo e Innovación en Biotecnología Sanitaria de Elche (IDIBE), Universidad Miguel
23 Hernández, Elche, 03202, Alicante, Spain

24 ¹⁰ Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO
25 63110

26 † These authors contributed equally to this work

27 * correspondence should be addressed to: lchemes@iib.unsam.edu.ar, pappu@wustl.edu,

28 alex.holehouse@wustl.edu, gdaughdrill@usf.edu

29

30 **ABSTRACT**

31 **Many disordered proteins conserve essential functions in the face of extensive sequence**
32 **variation, making it challenging to identify the mechanisms responsible for functional**
33 **selection. Here, we identify the molecular mechanism of functional selection for the**
34 **disordered adenovirus early gene 1A (E1A) protein. E1A competes with host factors to**
35 **bind the retinoblastoma (Rb) protein, subverting cell cycle regulation. We show that two**
36 **binding motifs tethered by a hypervariable disordered linker drive picomolar affinity Rb**
37 **binding and host factor displacement. Compensatory changes in amino acid sequence**
38 **composition and sequence length lead to conservation of optimal tethering across a large**
39 **family of E1A linkers. We refer to this compensatory mechanism as conformational**
40 **buffering. We also detect co-evolution of the motifs and linker, which can preserve or**
41 **eliminate the tethering mechanism. Conformational buffering and motif-linker coevolution**
42 **explain robust functional encoding within hypervariable disordered linkers and could**
43 **underlie functional selection of many disordered protein regions.**

44

45

46

47

48 **Running title:** Entropic tethering underlies viral hijack in a minimal viral model system

49

50

51

52

53

54 INTRODUCTION

55

56 Intrinsically disordered proteins and protein regions (IDRs) [1,2] use short linear motifs
57 (SLiMs) to bind cellular partners. These conserved interaction modules play essential roles in cell
58 biology [3]. In contrast, the regions connecting SLiMs often have lower sequence conservation
59 and a high frequency of insertions and deletions [4]. Under the classical structure-function
60 paradigm, these features indicate weak evolutionary restraints, leading to the view that these
61 IDRs might play the roles of passive “spacers”, stringing together ordered domains and disordered
62 SLiMs. However, recent progress in the quantitative description of sequence-ensemble
63 relationships (SERs) in IDR conformations [5] indicates that specific features in these less
64 conserved regions are required for function [6,7,8,9]. The fact that IDRs with different sequence
65 characteristics have conserved SERs that are responsible for function [10], suggests that SERs
66 are under natural selection. There is growing evidence that IDRs which function as flexible tethers
67 that physically join ordered domains and/or disordered SLiMs fall into this category [11,12,13].

68 Tethering is essential for kinase signaling [8,14,15], gene silencing [7], enzyme catalysis [16],
69 transcriptional regulation [13,17,18] and the formation of biomolecular condensates [9,19]. Tethering
70 allows intra- or intermolecular coupling between ordered domains and/or SLiMs [20]. This coupling
71 can increase the effective concentrations of interacting partners [21], and relatively simple polymer
72 models such as the Worm Like Chain (WLC) [22,23,24] can estimate the affinity enhancement from
73 tethering [22,25,26,12]. An emerging hypothesis is that SERs that encode IDR dimensions -as
74 determined by sequence length, composition and patterning- [27,28,29,30,5] play an important role in
75 tethered interactions by determining the effective concentrations of binding modules around
76 binding partners either in *cis* or in *trans* [12,13,31,32,16,33]. This leads to the expectation that
77 evolutionary pressure will preserve these dimensions in spite of large-scale sequence variation.

78 However, the lack of a well-defined model system in which molecular function is unambiguously
79 conserved in the face of a hypervariable tether has hampered the ability to test this hypothesis.

80 In order to establish a model system for quantitatively understanding tethering
81 mechanisms and their evolution, we chose the intrinsically disordered adenovirus early region 1A
82 (E1A) protein. Viruses are under constant selection pressure from a changing environment, and
83 many viral proteins utilize protein disorder to acquire novel traits [^{34,35,36,37,38}]. This makes them
84 robust as model systems to investigate functional selection of IDRs. E1A is a multifunctional
85 signaling hub that employs multiple SLiMs [^{36,37,38}] tethered by disordered linkers to hijack cell
86 signaling [³⁹]. Here we test the central hypothesis that conserved SERs drive functional selection
87 of the disordered E1A protein. Our results demonstrate that IDRs with dramatic changes in the
88 linear sequence have a conserved tethering function. We also found evidence for compensatory
89 co-evolution between disordered tethers and SLiMs. Taken together, our findings have broad
90 implications for understanding IDR function and evolution.

91

92

93 **MAIN TEXT**

94

95 ***Tethering enhances Rb binding and promotes E2F displacement***

96 The subversion of cell cycle regulation by E1A involves essential interactions with the
97 retinoblastoma (Rb) tumor suppressor, which displaces E2F transcription factors, triggering S-
98 phase entry and viral genome replication (**Fig. 1a, b**). To identify the molecular mechanisms
99 responsible for E2F displacement, we performed detailed structural and biophysical binding
100 studies using the central RbAB domain of Rb (hereafter referred to as Rb) and the minimal Rb
101 binding region from the adenovirus E1A protein (hereafter referred to as E1A_{WT}). Rb contains the
102 binding sites for the E2F and LxCxE SLiMs [40,41] and E1A_{WT} harbors the E1A_{E2F} and E1A_{LxCxE}
103 SLiMs [42] connected by a 71-residue disordered linker [43]. This linker contains additional SLiMs
104 for the CREB binding protein (CBP) TAZ2 domain [43] and the BS69 transcriptional repressor
105 MYND domain [39], which mediate the formation of ternary complexes [44] (**Fig.1b**).

106 To assess the affinity of E1A_{WT} and the relative contributions of the two motifs, we also
107 tested E1A constructs comprising the individual SLiMs or fragments where the E2F (E1A_{ΔE}) or
108 LxCxE (E1A_{ΔL}) motifs were mutated to poly-alanine (**Extended Data Fig. 1 and Fig. 1b**) and the
109 E2F SLiM (E2F2) taken from the host transcription factor E2F2 (**Fig. 1b**). Isothermal titration
110 calorimetry (ITC) (**Extended Data Fig. 2 and Supplementary Data Table 1**) and size exclusion
111 chromatography coupled to static light scattering (SEC-SLS) experiments (**Supplementary Data**
112 **Table 2**) confirmed that all E1A constructs bound to Rb with 1:1 stoichiometry. To quantify binding
113 affinities, we performed fluorescence polarization measurements using fluoresceine 5-
114 isothiocyanate (FITC)-labelled constructs (**Extended Data Fig. 3 and Supplementary Data**
115 **Tables 1 and 3**). While the host-derived E2F2 SLiM bound to Rb with high affinity ($K_D = 1$ nM),
116 the E1A_{E2F} SLiM had a $K_D = 119$ nM, suggesting it would be a weak competitor of E2F2 (**Fig. 1c**).
117 Based on a previous study showing the E2F and LxCxE motifs bind simultaneously to Rb [45] we
118 suspected that a protein containing both SLiMs and the linker (E1A_{WT}) would increase the binding

119 affinity of the individual SLiMs by enhancing local concentrations. In support of this, we found that
120 E1A_{WT} had picomolar binding affinity ($K_D = 24$ pM), conferring a 4000-fold enhancement compared
121 to the individual E1A SLiMs and a 40-fold enhancement compared to E2F2, consistent with a role
122 of tethering in affinity enhancement (**Fig. 1c**).

123 To further test the role of tethering in E2F displacement, we carried out competition
124 assays. Synthetic peptides corresponding to the E1A_{LXCxE} and E1A_{E2F} motifs as well as the E1A_{ΔL}
125 mutant were unable to effectively displace E2F from Rb (**Fig. 1d**). However, E1A_{WT} was a strong
126 competitor, disrupting the [E2F2:Rb] complex at low nanomolar concentration (**Fig. 1d**). The
127 agreement among ITC, direct titration and competition experiments confirmed that tethering was
128 required for high affinity Rb binding and E2F displacement (**Fig. 1e, Supplementary Data Table**
129 **1**).

130 We anticipated that tethering the two SLiMs would play a prominent role in the affinity
131 enhancement between the independent and linked SLiMs of E1A by increasing the effective
132 concentration (C_{eff}) of the second motif once a primary interaction is established (**Fig. 1f, Model**
133 **A**). However, alternative mechanisms that are not mutually exclusive with tethering could also
134 contribute to the stability of the complex. The E1A linker could enhance affinity by establishing
135 additional stabilizing interactions with Rb (**Fig. 1f, Model B**). Alternatively, a primary interaction
136 by the E1A_{E2F} or E1A_{LXCxE} SLiMs could induce an allosteric change in Rb that enables the
137 complementary motif to bind with higher affinity (**Fig. 1f, Model C**). We tested each of these
138 mechanisms using a combination of structural biophysics and thermodynamic analysis.

139

140 ***Linker-mediated interactions do not stabilize binding to Rb***

141 We first sought to confirm the disordered nature of E1A_{WT} using nuclear magnetic
142 resonance (NMR) spectroscopy. The transverse optimized relaxation (TROSY) spectrum of ¹⁵N-
143 labeled E1A_{WT} revealed narrow chemical shift dispersion in the ¹H-dimension. This is a

144 characteristic signature of disordered regions and is consistent with previous work on E1A
145 fragments (**Fig. 2a**) [43,46,47]. Further, the $^{13}\text{C}_\alpha$ secondary chemical shifts ($\Delta\delta\text{C}_\alpha$) showed minimal
146 deviation from random coil values obtained from disordered proteins (**Fig. 2b I**) and negative ^1H -
147 ^{15}N nuclear overhauser effect (NHNOE) values observed for E1A_{WT} indicated fast backbone
148 dynamics (**Fig. 2b II**). Finally, sequence analysis also predicted that E1A_{WT} is globally disordered
149 (**Fig. 2b IV**). These results confirmed that the conformational ensemble of E1A_{WT} is characterized
150 by high heterogeneity (disorder) and with fast interconversion between distinct conformations on
151 the nanosecond to picosecond timescale (flexibility).

152 Next, we used NMR spectroscopy to determine the structural basis for E1A_{WT} binding to
153 Rb. For this, we dissected the relative contributions of the SLiMs, their flanking regions, and the
154 linker. Previous NMR work mapped binding of E1A fragments containing individual Rb binding
155 motifs [43] but did not examine the structural details of E1A bivalently tethered to Rb. The TROSY
156 spectrum of labeled E1A_{WT} in complex with unlabeled Rb (MW=54.6 kDa, **Supplementary Data**
157 **Table 2**) reveals a complete loss of peaks for residues corresponding to the E2F and LxCxE
158 SLiMs (L43 to Y47 and L122 to E126) (**Fig. 2a, 2b III and Extended Data Fig. 4**) consistent with
159 slow exchange expected from the high affinities of the untethered motifs and the tethered complex
160 [48]. The E1A_{ΔL} and E1A_{ΔE} constructs retain binding of the wild type motif (**Fig. 2b III**), consistent
161 with independent binding of each motif to Rb. Based on previous reports [49], we anticipated that
162 the regions flanking the canonical E1A_{E2F} or E1A_{LxCxE} motifs contribute stabilizing interactions to
163 the complex. In agreement with this expectation, the peaks corresponding to the flanking residues
164 (E39 to T52 and V119 to E135) disappeared upon binding, yielding near-zero I/I_0 ratios (**Fig. 2b**
165 **III**).

166 Binding experiments using fluorescence polarization and ITC confirmed the stabilizing role
167 of the flanking regions: the affinity of the E1A_{LxCxE} motif increased incrementally upon adding the
168 acidic stretch following the motif (E1A_{LxCxE-AC}), and with Ser132 phosphorylation (E1A_{LxCxE-ACP})

169 (Fig. 1b and Supplementary Data Table 1) leading to an overall 5-fold increase in binding affinity.
170 ITC revealed different origins for thermodynamic stability in each core motif and the flanking
171 regions (Supplementary Data Table 1). While binding of E1A_{E2F} is entropically driven, suggesting
172 complex stabilization is dominated by the desolvation of apolar surfaces, binding of E1A_{LxCxE} is
173 enthalpically driven, likely due to the contribution from hydrogen bonds between the LxCxE motif
174 and Rb.

175 The N-terminal linker region (residues 50-85) encompassing the TAZ2 binding motif is
176 highly conserved and has a lower disorder propensity due to its hydrophobic nature (Fig. 2b IV-
177 V). This region showed a decrease in peak intensities (Fig. 2b III) that a previous report suggested
178 was due to weak interactions with Rb [43]. Consistent with this observation, the N-terminal linker
179 region does not show increased chemical shift dispersion or large chemical shift changes upon
180 binding Rb (Fig. 2a and Extended Data Fig. 4a-d). The C-terminal linker region (residues 86-
181 120) shows no changes in chemical shifts or resonance intensities, indicating this region remains
182 globally disordered and flexible when bound to Rb. These interpretations are supported by the
183 lack of change in secondary structure upon binding Rb, as measured by circular dichroism (CD)
184 (Fig. 2c).

185 Additional ITC studies using an isolated fragment from the N-terminal linker region that
186 showed the largest decrease in peak intensities (E1A₆₀₋₈₃) did not show any detectable association
187 to Rb (Extended Data Fig. 2i). Further, E1A constructs that include the linker did not show higher
188 binding affinities when compared to isolated E1A motifs (Fig. 2d, Supplementary Data Table 1).
189 Taken together, these data rule out the presence of a high affinity binding site. To test for weak
190 interactions that depend on bivalent tethering, we also designed a construct where the TAZ2
191 binding region (₇₁MLAVQEGID₇₉), which showed the largest reduction in I/I_0 , was replaced by a
192 GS stretch (E1A_{WTΔHyd}). The binding affinity of this mutant actually increased by 1.5-fold compared
193 to E1A_{WT} in fluorescence competition experiments (Supplementary Data Table 4), revealing a
194 weak destabilizing effect of the TAZ2 binding site.

195 In order to identify thermodynamic contributions of the linker binding to Rb, temperature
196 dependent ΔH measurements were used to infer changes in accessible surface area (ΔASA_T)
197 and the number of residues (X_{res}) that fold upon binding to Rb (**Extended Data Fig. 5** and
198 **Supplementary Data Tables 5 and 6**) (See **Methods**). ΔASA_T values calculated using
199 conventional and IDP-specific models [^{50,51}] failed to reveal an increase in ΔASA_T of the motif-
200 linker construct (E1A_{ΔL}) compared to the individual motif (E1A_{E2F}) (**Fig. 2d** and **Supplementary**
201 **Data Table 6**), demonstrating that the linker did not contribute to additional surface desolvation.
202 The IDP-specific method yielded $X_{res} = 33$ residues for the 16-mer E1A_{E2F} binding to Rb, indicating
203 a similar number of Rb residues fold at the E1A-binding interface. However, X_{res} did not increase
204 for E1A_{ΔL} compared to E1A_{E2F}, suggesting that no additional linker residues were involved in
205 coupled folding and binding.

206 Collectively, these results demonstrate that the linker does not contribute to the
207 thermodynamics of complex formation through coupled folding and binding or through persistent
208 molecular interactions with Rb. While the hydrophobic TAZ2-binding region may establish
209 transient, weak interactions with Rb that have a minor destabilizing effect and result in the
210 resonance intensity reductions we observe, our results do not support a model that invokes linker-
211 mediated interactions (**Fig. 1f, Model B**) as a source for affinity enhancement.

212

213 ***Allosteric coupling in Rb does not increase E1A-Rb affinity***

214 To assess whether allosteric coupling between the E2F and LxCxE binding sites in Rb
215 play a role in affinity enhancement (**Fig. 1f, Model C**), we saturated Rb with the E1A_{E2F} or E1A_{LxCxE}
216 motifs and performed ITC titrations with the complementary motif (**Extended Data Fig. 5**). If a
217 positive allosteric effect is at play, E1A_{LxCxE} should bind more tightly to Rb when E1A_{E2F} is already
218 bound, and vice versa. This was measured as the change in Gibbs free energy $\Delta\Delta G =$
219 $\Delta G_{SATURATED} - \Delta G_{UNSATURATED}$, where a negative value for $\Delta\Delta G$ indicates positive cooperativity. For
220 both motifs, the values of $\Delta\Delta G$ were in the range +/- 0.25 kcal/mol (**Supplementary Data Table**

221 7). In E1A_{LXCXE} binding assays, saturation with E1A_{ΔL} instead of E1A_{E2F} did not change the
222 outcome, indicating that neither the motif nor the motif + linker arrangement behaved as an
223 allosteric effector on the complementary site. Therefore, our results suggest that allosteric
224 coupling in Rb (**Fig. 1f, Model C**) does not make a major contribution to affinity enhancement.

225

226 ***Entropic tethering optimizes affinity of E1A for Rb***

227 Our results indicate the positive cooperativity of the tethered E1A_{E2F} and E1A_{LXCXE} motifs
228 binding to Rb results from an increase in the effective concentration (C_{eff}) of one motif once the
229 other motif is bound [⁴⁵] (**Fig. 1f, Model A**). It is well established that this form of cooperativity can
230 be described using a simple Worm Like Chain (WLC) model [^{23,24,22, 33}] that treats the linker as an
231 entropic tether (**Fig. 3a,b**) wherein the dimensions of the linker will determine the degree of the
232 affinity enhancement. A short linker would be unable to straddle the distance between the two
233 binding sites and lead to low affinity enhancement (**Fig. 3a,b I**), an optimal linker would maximize
234 C_{eff} , leading to maximal positive cooperativity (**Fig. 3a,b II**), and a longer than optimal linker would
235 decrease C_{eff} (**Fig. 3a,b III**). Application of the WLC model to the E1A linker predicts a C_{eff} value
236 of 0.92 mM, which is close to the optimal value (**Fig. 3b**) and within a factor of two of the C_{eff} (0.52
237 ± 0.09 mM) obtained from the affinities of E1A_{WT} and the isolated motifs (**Supplementary Data**
238 **Table 1**). For E1A_{WTΔHyd}, where the destabilizing effect of the linker region is removed, the
239 agreement with the WLC model improves ($C_{eff} = 0.78 \pm 0.24$ mM), indicating that this mutated
240 linker behaves more like an entropic tether optimized to bind Rb with near-maximal affinity.

241 To further test the tethering model, we performed Small Angle X-ray Scattering (SAXS)
242 on Rb, E1A_{WT}, and the [E1A_{WT}:Rb] complex (**Fig. 3c, Extended Data Fig. 6**). The experimental
243 SAXS profile of the Rb domain could be fit to the theoretical SAXS profile derived from its crystal
244 structure ($\chi^2 = 1.3$) and further refined (RMSD = 1.7 Å) using a SAXS-driven modelling approach
245 ($\chi^2 = 0.82$) (**Fig. 3c, Extended Data Fig. 6a**), indicating that Rb in solution retained its folded
246 structure. Alternatively, the Kratky plots of E1A_{WT} were characteristic of an IDP. Fitting of the

247 SAXS profiles using the Ensemble Optimization Method (EOM) [52] indicated that E1A_{WT} adopts
248 highly expanded conformations (**Extended Data Fig. 6b**). To analyze the conformation of the
249 linker in the [E1A_{WT}:Rb] complex, we applied a sampling method [53] to generate a pool of 10250
250 realistic conformations [54] and computed theoretical SAXS profiles that were selected using EOM
251 analysis. The SAXS profile of the complex was best described by sub-ensembles where the linker
252 sampled expanded conformations (**Fig. 3c-e, Extended Data Fig. 6c**) with hydrodynamic radius
253 (R_h) values ($R_{h,EOM} = 3.36$ nm) in good agreement with those obtained from SEC-SLS experiments
254 ($R_{h,SEC} = 3.20 \pm 0.12$ nm) (**Fig. 3f-g, Extended Data Fig. 6d and Supplementary Data Table 2**)
255 and R_g/R_h ratios consistent with bivalent tethering (**Supplementary Data Table 2**).

256 Our structural and thermodynamic dissection establishes E1A as a quantitative model
257 system for entropic tethering, demonstrating that other mechanisms have a negligible contribution
258 to affinity enhancement (with linker interactions having $\Delta G \sim +200$ cal.mol⁻¹ over a total $\Delta G_{BINDING}$
259 = -14240 cal.mol⁻¹). Our ability to isolate tethering as the key determinant of binding affinity is
260 unparalleled and provides us with a unique opportunity to test how tethering operates in biological
261 systems. A longstanding question is whether the sequences of regions that encode tethering have
262 any unique relationships with the conformational ensembles they form. If they do not it is difficult
263 to imagine that a family of linkers with extensive variation in sequence and length could function
264 in the same way. The model we present below is a comprehensive assessment of how this is
265 possible.

266

267 ***Hypervariable E1A linkers have a conserved functional length***

268 Inspection of selected linker sequences representative of mastadenoviruses that infect a
269 wide range of mammalian hosts (**Fig. 4a**) revealed that while the N- and C-terminal acidic
270 extensions and the aromatic/hydrophobic TAZ2 binding region were highly conserved, the linker
271 lengths and compositions vary considerably within the central region enriched predominantly with
272 polar, hydrophobic and proline residues (**Fig. 4a and 2b V**). To understand how function is

273 conserved in the face of these extensive differences in linker length and sequences, we performed
274 all atom simulations [9] and generated conformational ensembles of 27 E1A linker sequences with
275 linker lengths from 27 to 75 residues (**Fig. 4a**). While the shortest linkers from Bovine/Ovine E1A
276 proteins had smaller end-to-end distances, the average end-to-end distance of linkers 41 to 75
277 residues long remained roughly constant despite almost doubling the length (**Fig. 4b**). This
278 suggested that the linkers have a conserved functional length [55] that is determined by a joint
279 contribution of sequence length, amino acid composition, and sequence patterning as
280 determinants of end-to-end distances. To test the feasibility of this hypothesis, we performed
281 simulations for 140 random synthetic sequences of variable length that matched the amino acid
282 composition of one of the shortest linkers (HF_HAdV40). In sharp contrast to natural sequences,
283 the synthetic sequences showed the expected monotonic increase in end-to-end distance with
284 chain length ($R_{\text{natural}} = 0.37$, $R_{\text{synt}} = 0.99$, **Extended Data Fig. 7a**). To examine the sequence-
285 encoded origins of this compensation we analyzed various statistical properties (**Extended Data**
286 **Fig. 7 b,c**). Net charge per residue (NCPR) had the strongest positive correlation with normalized
287 end-to-end distance, with more expanded chains having a higher NCPR (**Extended Data Fig.**
288 **7b**). This is in agreement with previous findings that net charge and patterning are major
289 determinants of IDR dimensions in natural [55,13,30,29,11] and synthetic [30,27,32] sequences. Longer
290 chains also tend to have higher proline content with fewer hydrophobic and charged residues
291 (**Extended Data Fig. 7c**).

292 The results of the simulations suggest that the functional length of the linkers is conserved
293 and the linker dimensions are key to providing optimal affinity enhancement by tethering. Based
294 on these results, we hypothesize that the end-to-end distances of disordered linkers are under
295 functional selection through compensatory covariations in sequence length and composition, an
296 adaptive mechanism that we term conformational buffering.

297

298 ***Conformational buffering preserves optimal tethering***

299 The conformational buffering mechanism predicts that linker dimensions and optimal
300 tethering will be conserved across E1A proteins with very different linker sequences. To test this
301 prediction, we constructed a series of E1A chimeras by grafting different linker sequences with
302 the E1A_{E2F} and E1A_{LxCxE} motifs (**Fig. 4c**) and determined Rb binding affinity using the competition
303 assay of **Fig.1d (Supplementary Data Table 4 and Extended Data Figure 8)**. We selected
304 linkers from E1A types infecting a wide range of mammalian hosts (**Fig.4a,c**). These sequences
305 cover a wide range of linker lengths (27-75), amino acid composition, and sequence patterning
306 (**Fig. 4a and Extended Data Fig. 7c**). The E1A variants were expressed as MBP fusion proteins
307 (**Extended Data Fig. 1e**) and we verified that MBP-E1A_{WT} had the same binding affinity as
308 cleaved E1A_{WT} (**Supplementary Data Table 4**).

309 The sequence and structure of the Rb domain that binds to E1A is highly conserved across
310 the host range covered in our experiments (> 95% sequence identity and RMSD < 1.2 Å). The
311 residues that make up the E2F and LxCxE binding clefts and the spacing between the sites are
312 also highly conserved suggesting that functional length is under selection (**Extended Data Fig. 9**
313 and **Supplementary Fig. 1**). This conservation implies that human Rb is an excellent proxy for
314 the mammalian Rb proteins.

315 We predict that optimal tethering depends mainly on the linker dimensions and variants
316 with conserved end-to-end distances will confer similar affinity to E1A_{WT}. In accordance with this
317 prediction, human and simian E1A linkers ranging from 41 to 75 residues have similar binding
318 affinities, with $K_D/K_{D,E1A_{WT}}$ ratios between 0.4 and 1.2 (**Fig. 4c,d**). E1A linkers with the highest
319 affinities (Hum-2 and Sim-1) had polar residues interrupting the weakly destabilizing hydrophobic
320 interactions in the TAZ2 SLiM (**Fig. 4a**). To directly assess linker dimensions we used SEC to
321 measure R_h for selected E1A variants after MBP cleavage (**Extended Data Fig. 1f,g and 7d**).
322 Both K_D and R_h agreed closely with those predicted from the atomistic simulations (**Fig. 4d,**
323 **Supplementary Data Table 4 and Extended Data Fig. 7d**). We also created a tandem repeat

324 where the Hum-2 linker was duplicated (Hum-2-2x, **Fig. 4c,d**). $K_{D,Hum-2-2x}/K_{D,E1AWT}$ was 1.1,
325 suggesting its dimensions are still optimized.

326 E1A linkers appear to be under strong functional selection to preserve optimal tethering
327 using a mechanism that requires compensatory covariations in sequence length and composition
328 (i.e. conformational buffering). These results underscore the functional implications of preserving
329 sequence-ensemble-relationships (SERs), which in the case of E1A is achieved by preserving
330 the dimensions of the disordered linkers, which is necessary for hijacking the eukaryotic cell cycle.

331

332 ***Linker-motif coevolution modulates conformational buffering***

333 The shorter bovine linkers (**Fig. 4a**) had $K_D/K_{D,E1AWT}$ ratios between 13 and 20 (**Fig. 4c,d**).
334 This weaker than predicted affinity (**Fig. 4d**) was not due to the linkers being less expanded since
335 the predicted and experimental R_h values for Bov-1 were similar (**Extended Data Fig. 7d**).
336 Instead, it suggests that a minimal sequence length, not predicted by WLC, is necessary to
337 overcome entropic effects required for proper orientation of the SLiMs to bind Rb. This highlights
338 a limitation of the WLC model which is not unexpected since this simplified homopolymer model
339 does not include excluded volume or local changes in the chain stiffness.

340 We expect that the E1A linkers and SLiMs are co-evolving in a way that may not be
341 represented in the chimeras. For instance, the canine and bat chimeras had similar predicted and
342 experimental R_h values (**Extended Data Fig. 7d**) but showed reduced binding affinity, with
343 $K_D/K_{D,E1AWT}$ ratios between 5 and 6 (**Fig. 4c,d**) even though FoldX predicted the bat SLiMs to have
344 higher affinity for Rb than E1A_{WT} motifs. This reduction in binding could be due to additional
345 destabilizing interactions of the bat/canine linkers with Rb, implying the possibility of
346 compensatory changes that optimize but do not maximize binding affinity. To test this hypothesis,
347 we measured the affinity of a variant containing the endogenous SLiMs and linker from bat (Bat-
348 ED), which recovered high affinity binding with Rb ($K_D/K_{D,E1AWT} = 1.5$). This is a clear signature of

349 coevolution whereby linker mutations that weaken affinity enhancement by tethering are
350 compensated by SLiM mutations that directly increase Rb affinity.

351 The Bovine linker is predicted to have a smaller than optimal end-to-end distance
352 compared with other E1A linkers (**Fig. 4b**) and the bovine SLiMs are predicted by FoldX to bind
353 Rb with lower affinity (**Fig. 5a**), suggesting that the SLiMs and linker for Bovine are suboptimal.
354 To test this prediction, we measured the affinity of a variant with the endogenous Bov-1 SLiMs
355 and linker (Bov-1-ED). Bov-1-ED was unable to displace E2F in our competition experiments
356 ($K_D/K_{D,E1AWT} > 20,000$) (**Fig. 4 c,d** and **Extended Data Fig. 8**). Taken together, these results
357 suggest that Bovine E1A cannot displace E2F to hijack the host cell cycle.

358

359 ***Evolutionary conservation of E1A tethering***

360 Our results suggest that conformational buffering is a selection mechanism that conserves
361 end-to-end distances and affinity enhancement by tethering for E1A, and that motifs and linkers
362 co-evolve. To test these hypotheses on a larger family of sequences⁵⁶, we predicted global
363 binding affinities for 110 distinct E1A SLiMs and linkers. We used FoldX to predict SLiM affinities
364 and we predicted C_{eff} using either the WLC model with a single persistence length (E1A WLC), or
365 the sequence-specific persistence length (E1A Lp-Sim) from the simulations of the 27 linkers in
366 **Fig. 4a** (**Extended Data Fig. 10**). The results are shown in **Figure 5a** together with the measured
367 affinities from the grafting experiments (E1A Graft) or from endogenous variants (E1A ED).

368 The conservation of affinity enhancement by tethering is predicted across E1A proteins
369 from adenoviruses infecting human (HA-G), simian (SA/B/F), canine (CA), bat (BtA/B) and equine
370 (EA) species. These results are in agreement with our binding affinity measurements for human,
371 simian and bat E1A proteins. The structural conservation of the p107/p130 paralogs that harbor
372 the same SLiM binding sites (**Extended Data Fig. 9**) suggests that E1A uses the same
373 mechanism to displace E2F factors bound to all Rb paralogs.

374 In contrast, in a divergent branch of E1A proteins infecting rodents (MA/B/C), treeshrew
375 (TSA) and artiodactyls (including bovine, sheep and pig OA/BA/PA), binding to Rb seems
376 impaired or lost completely due to the presence of short linkers coupled to low affinity (PC/OA/BA)
377 or missing SLiMs. For instance, E1A proteins from rodents retain the LxCxE motif but lose the
378 E2F motif. These E1A proteins could interfere with host factors binding to the LxCxE cleft but
379 would be unable to displace E2F. These results suggest that the SLiMs and the linker are under
380 co-evolutionary selection, such that either the SLiMs and linker are jointly optimized, or selection
381 pressure is lost on both elements, leading to a loss of E2F displacement and possibly a loss of
382 E1A's ability to hijack the eukaryotic cell cycle (**Fig. 5a**). This branch of divergent adenoviruses is
383 likely to employ alternative mechanisms to induce host cell proliferation.

384 In summary, we demonstrate that tethering is the main mechanism that allows E1A to bind
385 Rb with picomolar affinity and displace E2F transcription factors. We show that the functional
386 length of the linkers is conserved and fine-tuned through conformational buffering to enable
387 maximal affinity enhancement in the face of extensive changes in sequence composition and
388 length. We also uncover a previously unknown linkage between the evolution of linkers and their
389 tethered motifs. This study shows that strong functional selection can operate both on the motifs
390 and on the physical properties of an IDR linker, providing important insights regarding the
391 evolution of sequence features and tethering functions in IDRs.

392

393 **DISCUSSION**

394 Here, we demonstrate how E1A hijacks the eukaryotic cell cycle using two SLiMs tethered
395 by a flexible linker with conserved dimensions [12,26]. The proposed docking and displacement
396 mechanism is conserved across divergent E1A proteins by conformational buffering and
397 coevolution of the SLiMs and tether. Conformational buffering promotes robust encoding of a core
398 function (**Fig. 5b, upper**) while supporting the extensive sequence variation necessary to rewire
399 the E1A interactome (**Fig.5b, lower**) and adapt to different hosts by gaining or losing additional

400 SLiMs [56-58], as we show for several SLiMs in **Fig.5a** [39,43,56,59]. Our work challenges the view
401 that IDRs with extensive sequence variation evolve neutrally. We also demonstrate that
402 conserved SERs that encode for IDR dimensions -as determined by sequence length,
403 composition and patterning- can be detected with atomistic simulations even if they are obscured
404 by naïve sequence alignments.

405 Conformational buffering results in the conservation of tethering for any sequence solution
406 that preserves the functional length. Our experimental validation using a collection of E1A linkers
407 largely supports this hypothesis, demonstrating that linkers with a broad range of sequence
408 compositions and lengths are functionally equivalent (**Fig. 4 a,d**). However, the molecular
409 evolution of tethered systems will be constrained by competing evolutionary pressures [6,2]. In the
410 case of the E1A linkers, correlated changes in NCPR and proline content (**Fig. 4 a,b**) maintain
411 linker extension and prevent folding. The linker can contain additional SLiMs that mediate the
412 formation of higher order complexes [43] and impose restrictions on sequence variation that could
413 prevent optimal tethering, analogous to the frustrated energy landscapes in protein folding [60]. By
414 dissecting the contribution of linker versus SLiMs, we found signatures of these competing forces:
415 E1A_{WT} confers optimal tethering and harbors a hydrophobic SLiM (TAZ2 motif) that plays a minor
416 destabilizing role, while the Bat E1A linkers evolved stronger destabilizing interactions with Rb
417 that are compensated by mutations that restore optimal tethering by increasing the affinity of the
418 SLiMs for Rb. Thus it appears that competition between linker tethering and SLiM binding
419 constrains IDR evolution due to different contributions from conformational buffering and coupled
420 folding and binding. This can result in linker sequence conservation patterns ranging from highly
421 variable [this work,^{11,13}] to highly conserved [12]. Other systems such as the intrinsically disordered
422 Notch RAM region show similar mixed contributions from optimal tethering and sequence-specific
423 effects [61]

424 Our work establishes E1A as an example of optimal tethering. The low picomolar affinity
425 of E1A for Rb and the 4000-fold affinity enhancement enforced by the E1A linker is amongst the

426 highest reported positive cooperativity produced by tethering in a natural system, similar to the
427 POU domain activator Oct-1 binding to DNA ($K_D = 71$ pM and 2100-fold enhancement) [22]. E1A
428 is the first adenoviral gene expressed, and the picomolar affinity is likely required for E1A to bind
429 stably to Rb and efficiently displace E2F and hijack the cell cycle at low expression levels during
430 early infection [62]. As a comparison, intramolecular MdmX inhibition, exhibits optimal tethering
431 with a 400-fold enhancement and $K_{INTRA} = 250$ [12] and other bivalently tethered systems show
432 variable degrees of enhancement with affinities in the nanomolar range [63,64,65]. At the opposite
433 extreme, multiple low affinity interactions tethered by short and/or non-optimal linkers might
434 promote the dynamic binding required for multivalent binding or liquid-liquid phase separation
435 [66,17]. Our work suggests that conformational buffering can tune the functional length of linkers to
436 produce an optimal functional output. Thus, conformational buffering may be a widespread
437 mechanism driving dimensional compensation among IDRs.

438

439

440

441

442

443

444

445 **Acknowledgements**

446 This work was supported by: Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT)
447 Grants PICT #2013-1895 and #2017-1924 (LBC), #2012-2550 and #2015-1213 (IES) and
448 #2016-4605 (GPG). US National Institutes of Health #GM115556 and #CA141244 (GWD) and
449 #5R01NS056114 (RVP), FLDOH #20B17 (GWD), US National Science Foundation #MCB-
450 1614766 (RVP). Travel award from the USF Nexus Initiative and a Creative Scholarship Grant
451 from the USF College of Arts and Sciences (GWD and LBC). Labex EpiGenMed
452 «Investissements d'avenir» program #ANR-10-LABX-12-01 (PB), French National Research
453 Agency #ANR-10-INBS-04-01 and #ANR-10-INBS-05 (PB). Spanish Ministerio de Ciencia y
454 Universidades MICYU-FEDER #RTI2018-097189-C2-1 (GFB). Consejo Nacional de
455 Investigaciones Científicas y Técnicas (CONICET, Argentina) doctoral fellowship (NGF, MS and
456 NAG), postdoctoral fellowship (JG), and permanent researcher (LBC, GdPG, IES). Fulbright
457 Visiting Scholar Program (NSGF). Ministerio de Ciencia e Innovación, España #BES-2013-
458 063991 and #EEBB-I-16-11670 (SBV). Longer Life Foundation: A RGA/Washington University
459 Collaboration (ASH). HPC resources of the CALMIP supercomputing center #2016-P16032
460 (GFB) and Cluster of Scientific Computing (<http://ccc.umh.es/>) of the Miguel Hernández University
461 (UMH) (GFB). The synchrotron SAXS data was collected at beamline P12 operated by EMBL
462 Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We thank Kathryn Perez at
463 the Protein Expression and Purification Core Facility at EMBL (Heidelberg) for critical help with
464 ITC experiments and Pedro Aramendia for providing critical access to fluorescence spectrometry
465 equipment at Centro de Investigaciones en Bionanociencias (CIBION, Argentina).

466

467 **Author Contribution Statement**

468 LBC, GWD, ASH and RVP designed research and conceived the study. NSGF WMB and MS
469 produced reagents. NSGF and WMB performed FP, ITC and NMR experiments and WB, NSGF,
470 GWD and LBC analyzed data. JG designed and conducted bioinformatic analyses of E1A variants

471 and Rb proteins. MS purified E1A protein variants and NAG performed SEC experiments. AS
472 and PB performed and analyzed SAXS experiments. AE, AB and JC produced and analyzed E1A
473 conformational ensembles. SBV and ASH performed and analyzed all atom simulations of E1A
474 linkers. GFB, CBM and IES computed and analyzed FOLDX matrices. NSGF, JG, AS and ASH
475 produced figures. LBC, GWD, PB, JC, GdPG, IES, ASH and RVP Supervised research. LBC,
476 NSGF, JG, RVP, ASH and GWD wrote the paper with critical feedback from all authors.

477

478 **Competing Interests Statement**

479 A.S.H. is a scientific consultant with Dewpoint Therapeutics Inc. and R.V.P. is a member of the
480 scientific advisory board of Dewpoint Therapeutics Inc. This work has not been influenced by the
481 affiliation with Dewpoint. The rest of the authors have no competing interests.

482

483

484

485

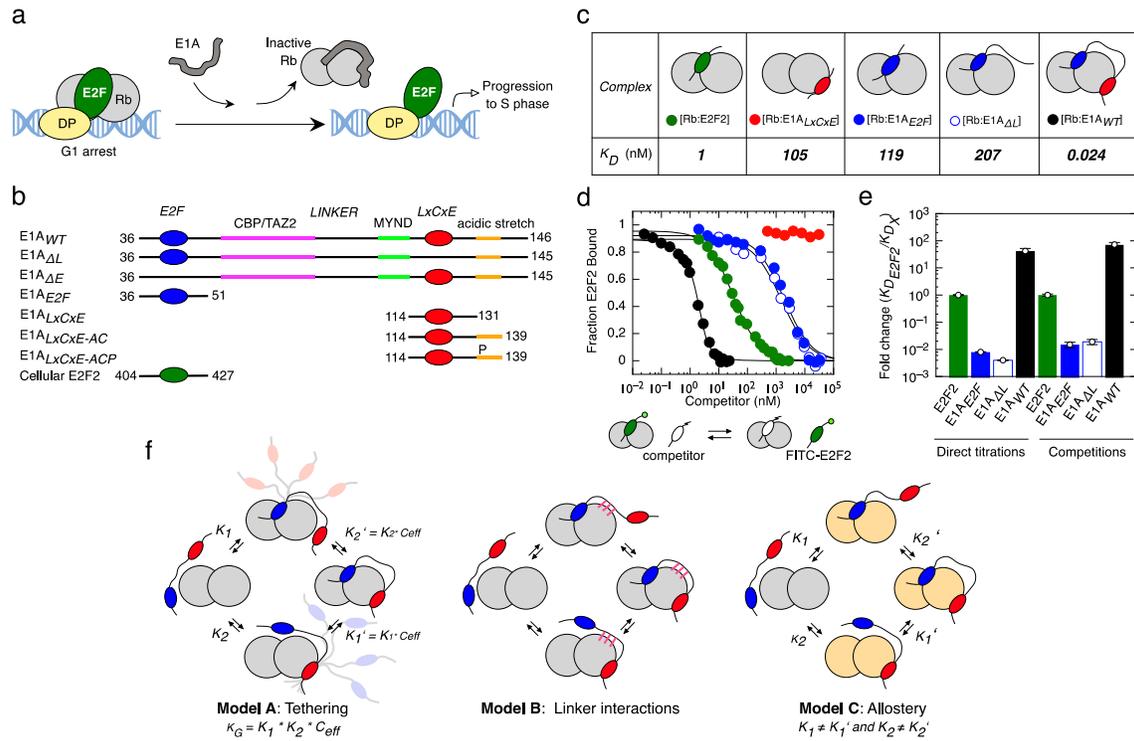
486

487

488

489

490 **FIGURE LEGENDS**



491

492 **FIGURE 1. Tethering is required for high affinity Rb binding and E2F displacement by E1A.**

493 **a)** Model for disruption of the repressive Rb-E2F complex by E1A. **b)** Schematic representation

494 of E1A and E2F2 constructs used in this study. Color coding for the E2F, LxCxE, TAZ2 and MYND

495 SLiMs, the acidic stretch and S132 phosphorylation are maintained throughout figures. **c)**

496 Representative interactions tested using fluorescence spectroscopy (**Extended Data Fig. 3 and**

497 **Supplementary Data Tables 1 and 3**). **d)** E2F competition titrations. Color code is as in panel c.

498 **e)** Comparison of the fold-change in binding affinity from direct titrations versus competition

499 assays. The height of the bar is obtained by dividing the K_D of E2F2 by each K_D ($n=1$), and values

500 higher than unity indicate an increase in binding affinity with respect to E2F2. For direct titrations,

501 each K_D value was obtained by averaging (global fitting) over several independent binding

502 isotherms (E2F2: $n=5$, E1A^{E2F}: $n=3$, E1A^{ΔL}: $n=3$, E1A^{WT}: $n=3$) containing 16-22 points each (see

503 **Source Data**). For competition experiments, each K_D was obtained by fitting of a single binding

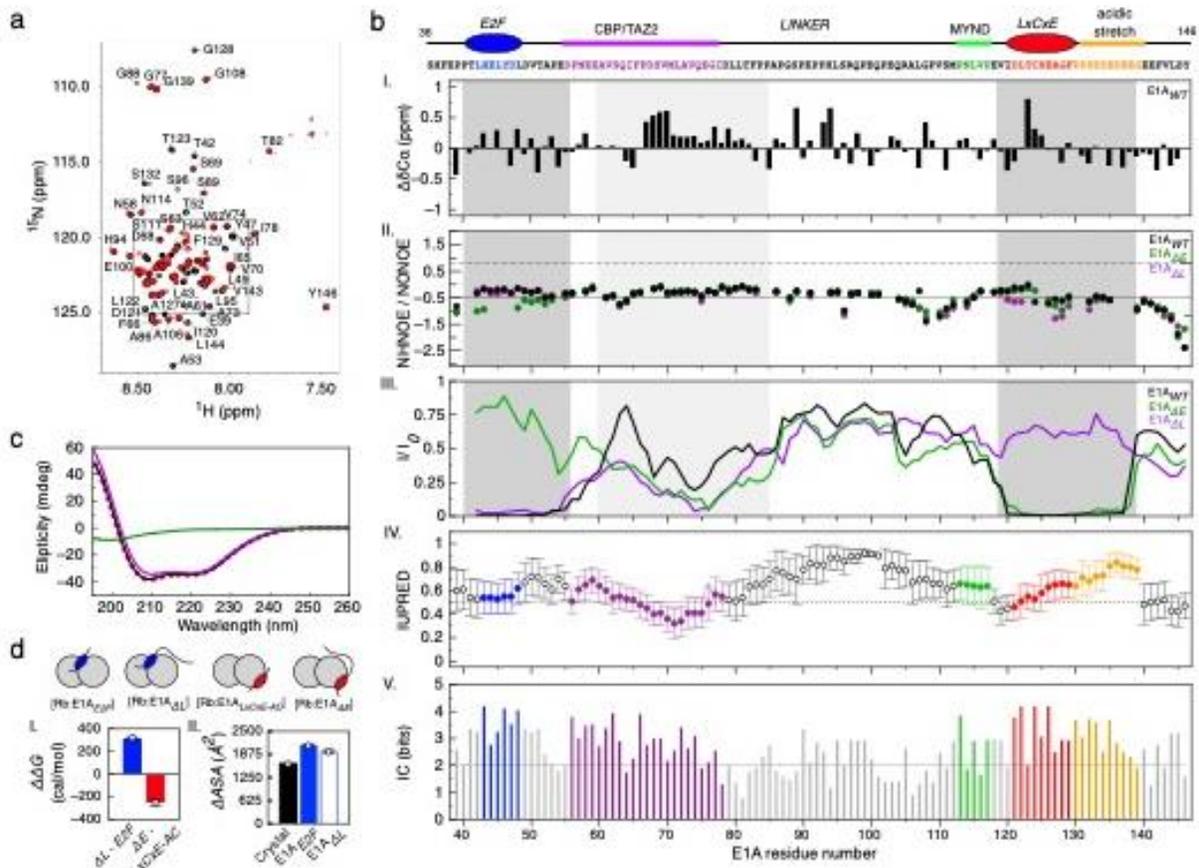
504 isotherm ($n=1$). Error bars correspond to the propagated standard deviation of the averaged K_D

505 values. **f)** Three models that account for affinity enhancement in the Motif-Linker-Motif E1A

506 arrangement (See main text for details).

507

508



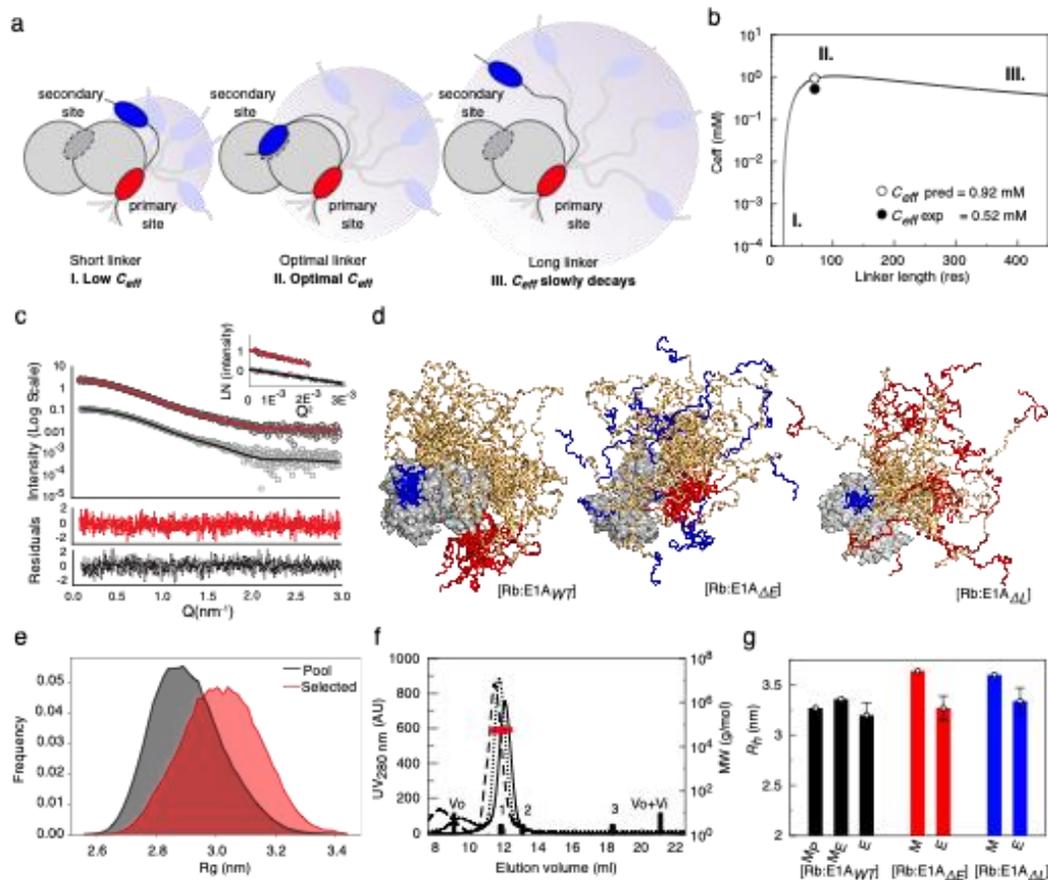
509

510 **FIGURE 2. NMR and ITC analysis of the [E1A_{WT}:Rb] complex.** **a)** ¹H-¹⁵N TROSY spectra of free
 511 ¹⁵N-E1A_{WT} (black) and ¹⁵N-E1A_{WT} bound to unlabeled Rb (red). ¹⁵N-E1A_{WT} peak assignments for
 512 the inset are shown in **Extended Data Fig. 4.** **b)** I. ¹³Cα secondary chemical shift (ΔδCα) of ¹⁵N-
 513 E1A_{WT}. II. NHNOE/NONOE ratio for ¹⁵N-E1A_{WT}. Dashed line: reference value for rigid backbone.
 514 III. Intensity ratio plots of bound state (*I*) with respect to the free state (*I*₀) for E1A_{WT}, E1A_{ΔL} and
 515 E1A_{ΔE}. Dark gray: E2F/LxCxE SLiMs and flanking regions; Light gray: N-terminal linker region.
 516 **IV-V.** Disorder propensity and residue conservation (information content: IC) were predicted from
 517 an alignment of E1A sequences (n=110) (**Supplementary Data File 1**). For disorder prediction,
 518 data points represent the mean IUPred value at each position and error bars represent the
 519 standard deviation of the mean. The number of residues averaged at each position is variable
 520 depending on the number of gaps in the alignment. For the conservation plot, the height of each
 521 bar represents the IC value at each position. **c)** Far-UV CD spectra for E1A_{WT} (green line), Rb
 522 (violet line), the [E1A_{WT}:Rb] complex (black line) and the arithmetic sum of the Rb and E1A_{WT}
 523 spectra (red dashed line). The latter CD spectra largely overlap. While it is possible the low salt
 524 concentration of the CD experiments might mask hydrophobic interactions occurring at the higher
 525 salt concentration used for NMR and other binding experiments, such effects are unlikely to

526 prevail for the types of monovalent salts used in our binding experiments. **d)** Left: Plot of the
527 change in free energy of binding ($\Delta\Delta G$) for E1A fragments containing or lacking the linker region,
528 measured by ITC. The bar height results from the subtraction between mean ΔG values obtained
529 by averaging several independent binding experiments: ΔG E1A $_{\Delta L}$ (n=3), ΔG E1A $_{E2F}$ (n=1), ΔG
530 E1A $_{\Delta E}$ (n=3) and ΔG E1A $_{LXCxE-AC}$ (n=3) (**Supplementary Data Table 1**). Right: Plot of the change
531 in ΔASA for E1A fragments containing or lacking the linker region. The height of the bar represents
532 the ΔASA value from PDB structure 2R7G (n=1, black bar) or that derived from ITC experiments
533 using parameters from Murphy & Freire for [E1A $_{E2F}$:Rb] (n=1, blue bar) and [E1A $_{\Delta L}$:Rb] (n=1,
534 empty blue bar) (**Supplementary Data Table 6**). ΔASA was calculated by ITC measurements at
535 several temperatures (n=4 [E1A $_{E2F}$:Rb], n=3 [E1A $_{\Delta L}$:Rb]). Error bars correspond to the propagated
536 mean standard errors of the ΔASA value.

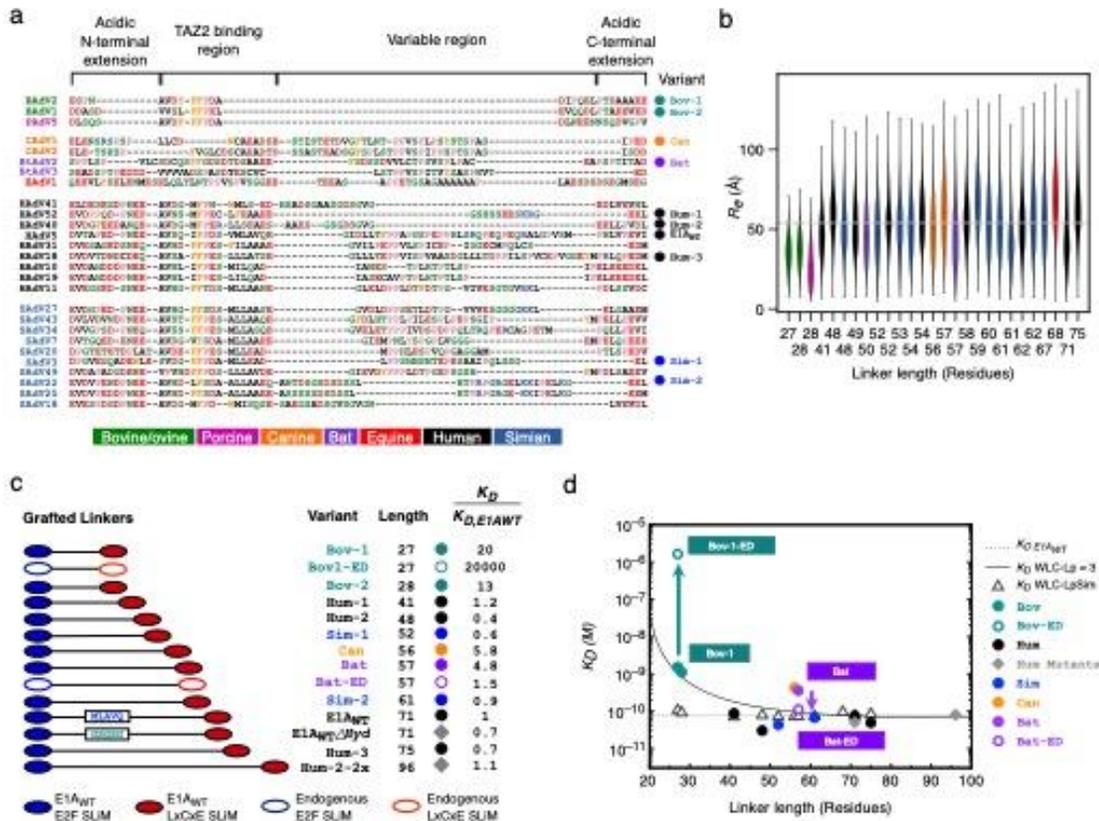
537

538



539

540 **FIGURE 3: The E1A linker behaves as an entropic tether. a)** Schematic representation of how
 541 C_{eff} depends on linker length. **b)** C_{eff} curve from the WLC model. The scenarios depicted in a) are
 542 shown as regions (I, II, III). **c)** SAXS intensity profile of: Rb (gray squares) with best fit to the
 543 theoretical profile derived from the Rb crystal structure (RbAB domain, black line); and the
 544 $[E1A_{WT}:Rb]$ complex (black circles) with best fit from the EOM method (red line). Inset: Guinier
 545 plots for Rb and $[E1A_{WT}:Rb]$. **d)** SAXS-selected $[E1A_{WT}:Rb]$ EOM ensemble (both motifs bound)
 546 and simulated ensembles for $[E1A_{\Delta E}:Rb]$ and $[E1A_{\Delta L}:Rb]$ (one motif bound). **e)** R_g distribution of
 547 the ensemble pool for $[E1A_{WT}:Rb]$ (black) and the EOM ensemble (red). The linker samples
 548 conformations more extended than the random-coil model of the pool. **f)** SEC-SLS of $[E1A_{WT}:Rb]$
 549 (solid line), $[E1A_{\Delta E}:Rb]$ (dotted line) and $[E1A_{\Delta L}:Rb]$ (dashed line). Black bars: BSA 66 kDa (1),
 550 MBP 45 kDa (2) and Lysozyme 14.3 kDa (3). Black line: SEC profile, Red line: MW value (g/mol).
 551 **g)** Comparison between the hydrodynamic radius (R_h) of modeled (M_P = pool, M_E = EOM) and
 552 experimental (E) ensembles for $[E1A_{WT}:Rb]$ (black bars), $[E1A_{\Delta E}:Rb]$ (red bars) and $[E1A_{\Delta L}:Rb]$
 553 (blue bars). The height of each bar represents the R_h value. Modeled R_h values ($n=1$) have no
 554 associated error. For Experimental R_h values ($n=1$) error bars represent the propagated error
 555 obtained from estimation of the R_h parameter (see **Methods**).



556

557 **FIGURE 4. Conformational buffering leads to conserved functionality of E1A proteins. a)**

558 Global alignment of 27 selected E1A linker sequences. *Mastadenovirus* types are indicated on

559 the left and the color coding (bottom panel) indicates the host range. The variants used for the

560 design of chimeras are shown to the right, with three letter codes indicating the host range. Amino

561 acids color code: acidic (red), basic (blue), polar (green), hydrophobic (black), aromatic (orange)

562 and proline (pink). **b)** End-to-end distance calculated from all-atom simulations using the set of

563 E1A linkers from panel a. Violin plots are colored by host range as in panel a. For each sequence,

564 n=15 independent simulations were run (see **Methods**). The horizontal line within each violin plot

565 represents the median end-to-end distance (R_e) value and the ends of the whiskers indicate the

566 maximal and minimal values. Horizontal dotted line: mean R_e value (53.39 Å) obtained by

567 averaging the median R_e values of all sequences excluding Bov-1, Bov-2 and Porcine. **c)** Motif-

568 Linker-Motif constructs used in the E1A linker grafting experiment. Filled circles: grafting of linkers

569 into the HAdV5 E1A_{E2F} and E1A_{LxCxE} motifs. Diamonds: Mutant where the hydrophobic

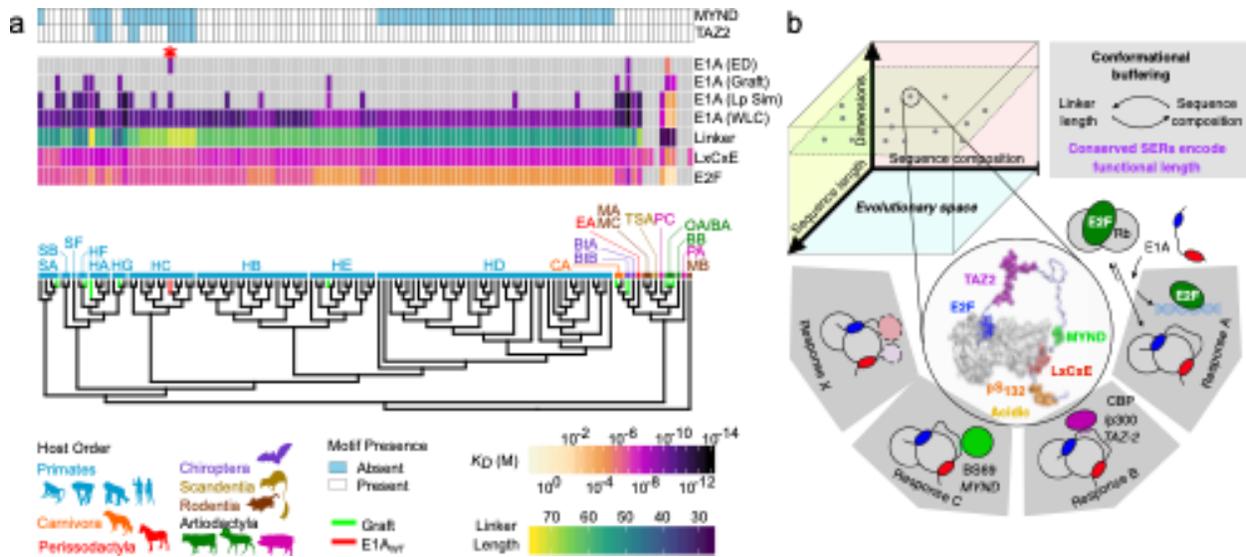
570 MLAVQEGID region was replaced by a GS stretch (E1A_{WTΔHyd}) or where the HAdV40 linker

571 sequence was duplicated (Hum2-2x). Empty circles: Variants harboring endogenous linker and

572 motifs (ED). **d)** Global K_D as a function of linker length for the Motif-Linker-Motif constructs. K_D for

573 each variant was measured using an E2F displacement experiment (symbols as in c) or predicted

574 using the WLC model. The K_D values \pm errors for all measurements are reported in
575 **Supplementary Data Table 4**. The predicted value of the K_D for the grafted linkers was calculated
576 as $K_D = (K_{D,E2F} * K_{D,LxCxE}) / C_{eff}$ (see **Methods**) using the known affinity of the E1A_{E2F} and E1A_{LxCxE}
577 motifs from E1A_{WT} (**Supplementary Data Table 1**) and C_{eff} values obtained using a sequence
578 independent (Straight line: WLC-Lp=3) or sequence-dependent (Empty triangles: WLC-LpSim)
579 persistence length (L_p) parameter (see **Extended Data Fig. 10 and Methods**). Dotted line:
580 Experimental K_D value of the E1A_{WT} construct (75 ± 17 pM). Under the sequence-independent
581 WLC model the K_D is expected to increase gradually with decreasing linker length, while LpSim
582 predicts the K_D to remain constant in the 41-75 linker length range. Experimental K_D values are in
583 good agreement with both models for longer linker lengths, but are closer to LpSim for shorter
584 linker lengths (41, 48 and 52).
585
586



587

588 **FIGURE 5. Evolutionary conservation of tethering by E1A proteins.** **a)** Phylogenetic tree of
 589 mastadenovirus E1A proteins with species denoted by two letter codes. The affinity of the
 590 E2F/LxCxE SLiMs and E1A_{WT}, and linker length are indicated by color scales. E1A (WLC): Global
 591 K_D for E1A proteins predicted by the WLC model with standard L_p values ($L_p = 3$); E1A (LpSim):
 592 K_D for E1A proteins predicted by the WLC model with sequence-dependent L_p values; E1A (Graft):
 593 Experimental K_D measured for the grafted linkers of **Fig. 4d**; E1A (ED): Experimental K_D
 594 measured for the variants harboring endogenous linker and motifs of **Fig. 4d**. Gray box: absent
 595 motif/linker. Light/blue box: present TAZ2/MYND SLiMs. The E1A_{WT} protein is marked as a red
 596 asterisk and as a red terminal branch in the tree and all other sequences used in the experiments
 597 are marked as green terminal branches in the tree. **b)** Upper: E1A sequences evolved a
 598 multiplicity of solutions in the sequence length-composition space to achieve conserved SERs
 599 through conformational buffering. Lower: The model represents one pose of the conformational
 600 [E1A_{WT}:Rb] ensemble with E2F/LxCxE SLiMs bound to Rb. The evolvable E1A interaction
 601 platform performs highly conserved functions (E2F activation) while allowing adaptive changes in
 602 functionality (TAZ2, MYND and other protein binding).

603

604

605 **REFERENCES:**

- 606
- 607
- 608 1. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein
609 structure-function paradigm. *J Mol Biol* **293**, 321–331 (1999).
- 610 2. van der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem*
611 *Rev* **114**, 6589–6631 (2014).
- 612 3. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the
613 molecular biologist. *Mol Cell* **55**, 161–169 (2014).
- 614 4. Brown, C. J., Johnson, A. K., Dunker, A. K. & Daughdrill, G. W. Evolution and disorder.
615 *Curr Opin Struct Biol* **21**, 441–446 (2011).
- 616 5. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form
617 and function of intrinsically disordered proteins. *Curr Opin Struct Biol* **32**, 102–112 (2015).
- 618 6. Daughdrill, G. W., Narayanaswami, P., Gilmore, S. H., Belczyk, A. & Brown, C. J.
619 Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of
620 negligible amino acid sequence conservation. *J Mol Evol* **65**, 277–288 (2007).
- 621 7. Beh, L. Y., Colwell, L. J. & Francis, N. J. A core subunit of Polycomb repressive complex
622 1 is broadly conserved in function but not primary sequence. *Proc Natl Acad Sci U S A*
623 **109**, E1063-71 (2012).
- 624 8. Das, R. K., Huang, Y., Phillips, A. H., Kriwacki, R. W. & Pappu, R. V. Cryptic sequence
625 features within the disordered protein p27Kip1 regulate cell cycle signaling. *Proc Natl*
626 *Acad Sci U S A* **113**, 5616–5621 (2016).
- 627 9. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase
628 behavior of prion-like domains. *Science* **367**, 694–699 (2020).
- 629 10. Zarin, T. *et al.* Proteome-wide signatures of function in highly diverged intrinsically
630 disordered regions. *Elife* **8**, (2019).
- 631 11. Buske, P. J., Mittal, A., Pappu, R. V & Levin, P. A. An intrinsically disordered linker plays
632 a critical role in bacterial cell division. *Semin Cell Dev Biol* **37**, 3–10 (2015).

- 633 12. Borchers, W. *et al.* Optimal Affinity Enhancement by a Conserved Flexible Linker
634 Controls p53 Mimicry in MdmX. *Biophys J* **112**, 2038–2042 (2017).
- 635 13. Sherry, K. P., Das, R. K., Pappu, R. V & Barrick, D. Control of transcriptional activity by
636 design of charge patterning in the intrinsically disordered RAM region of the Notch
637 receptor. *Proc Natl Acad Sci U S A* **114**, E9243–E9252 (2017).
- 638 14. Hantschel, O. *et al.* A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* **112**, 845–
639 857 (2003).
- 640 15. Ayrapetov, M. K. *et al.* Conformational basis for SH2-Tyr(P)527 binding in Src
641 inactivation. *J Biol Chem* **281**, 23776–23784 (2006).
- 642 16. Dyla, M. & Kjaergaard, M. Intrinsically disordered linkers control tethered kinases via
643 effective concentration. *Proc Natl Acad Sci U S A* **117**, 21413–21419 (2020).
- 644 17. Cordeiro, T. N. *et al.* Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer
645 and Its Corepressor Regulates Gene Expression. *Structure* **27**, 1270-1285.e6 (2019).
- 646 18. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo
647 Binding Specificity. *Mol Cell* **79**, 459-471.e4 (2020).
- 648 19. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered
649 linkers determine the interplay between phase separation and gelation in multivalent
650 proteins. *Elife* **6**, (2017).
- 651 20. Huang, Q., Li, M., Lai, L. & Liu, Z. Allostery of multidomain proteins with disordered
652 linkers. *Curr Opin Struct Biol* **62**, 175–182 (2020).
- 653 21. Jencks, W. P. On the attribution and additivity of binding energies. *Proc Natl Acad Sci U*
654 *S A* **78**, 4046–4050 (1981).
- 655 22. Zhou, H. X. The affinity-enhancing roles of flexible linkers in two-domain DNA-binding
656 proteins. *Biochemistry* **40**, 15069–15073 (2001).
- 657 23. Zhou, H. X. Polymer models of protein stability, folding, and interactions. *Biochemistry* **43**,
658 2141–2154 (2004).

- 659 24. Morrison, G. & Thirumalai, D. Semiflexible chains in confined spaces. *Phys Rev E Stat*
660 *Nonlin Soft Matter Phys* **79**, 11924 (2009).
- 661 25. van Dongen, E. M. W. M. *et al.* Variation of linker length in ratiometric fluorescent sensor
662 proteins allows rational tuning of Zn(II) affinity in the picomolar to femtomolar range. *J*
663 *Am Chem Soc* **129**, 3494–3495 (2007).
- 664 26. Bertagna, A., Topygin, D., Brand, L. & Barrick, D. The effects of conformational
665 heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case
666 of biologically tuned disorder. *Biochem Soc Trans* **36**, 157–166 (2008).
- 667 27. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue
668 modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad*
669 *Sci U S A* **107**, 8183–8188 (2010).
- 670 28. Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically
671 disordered proteins. *Biophys J* **98**, 2383–2390 (2010).
- 672 29. Müller-Spätth, S. *et al.* From the Cover: Charge interactions can dominate the dimensions
673 of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **107**, 14609–14614 (2010).
- 674 30. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are
675 influenced by linear sequence distributions of oppositely charged residues. *Proc Natl*
676 *Acad Sci U S A* **110**, 13392–13397 (2013).
- 677 31. Van Rosmalen, M., Krom, M. & Merkx, M. Tuning the Flexibility of Glycine-Serine Linkers
678 to Allow Rational Design of Multidomain Proteins. *Biochemistry* **56**, 6565–6574 (2017).
- 679 32. Sorensen, C. S. & Kjaergaard, M. Effective concentrations enforced by intrinsically
680 disordered linkers are governed by polymer physics. *Proc Natl Acad Sci U S A* **116**,
681 23124–23131 (2019).
- 682 33. Kjaergaard, M., Glavina, J. & Chemes, L. B. Predicting the effect of disordered linkers on
683 effective concentrations and avidity with the “C(eff) calculator” app. *Methods Enzymol*
684 **647**, 145–171 (2021).

- 685 34. Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N. & Tawfik, D. S. Do viral
686 proteins possess unique biophysical features? *Trends Biochem Sci* **34**, 53–59 (2009).
- 687 35. Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M. & Andino, R. Rapid evolution of virus
688 sequences in intrinsically disordered protein regions. *PLoS Pathog* **10**, e1004529 (2014).
- 689 36. Hagai, T., Azia, A., Babu, M. M. & Andino, R. Use of host-like peptide motifs in viral
690 proteins is a prevalent strategy in host-virus interactions. *Cell Rep* **7**, 1729–1739 (2014).
- 691 37. Davey, N. E., Trave, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends*
692 *Biochem Sci* **36**, 159–169 (2011).
- 693 38. Chemes, L. B., de Prat-Gay, G. & Sanchez, I. E. Convergent evolution and mimicry of
694 protein linear motifs in host-pathogen interactions. *Curr Opin Struct Biol* **32**, 91–101
695 (2015).
- 696 39. King, C. R., Zhang, A., Tessier, T. M., Gameiro, S. F. & Mymryk, J. S. Hacking the Cell:
697 Network Intrusion and Exploitation by Adenovirus E1A. *MBio* **9**, (2018).
- 698 40. Liu, X. & Marmorstein, R. Structure of the retinoblastoma protein bound to adenovirus
699 E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor.
700 *Genes Dev* **21**, 2711–2716 (2007).
- 701 41. Lee, J. O., Russo, A. A. & Pavletich, N. P. Structure of the retinoblastoma tumour-
702 suppressor pocket domain bound to a peptide from HPV E7. *Nature* **391**, 859–865
703 (1998).
- 704 42. Dyson, N., Guida, P., McCall, C. & Harlow, E. Adenovirus E1A makes two distinct
705 contacts with the retinoblastoma protein. *J Virol* **66**, 4606–4611 (1992).
- 706 43. Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Structural basis for
707 subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc Natl*
708 *Acad Sci U S A* **106**, 13260–13265 (2009).
- 709 44. Ferreon, A. C., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by
710 protein intrinsic disorder. *Nature* **498**, 390–394 (2013).

- 711 45. Fattaey, A. R., Harlow, E. & Helin, K. Independent regions of adenovirus E1A are
712 required for binding to and dissociation of E2F-protein complexes. *Mol Cell Biol* **13**,
713 7267–7277 (1993).
- 714 46. Hosek, T. *et al.* Structural and Dynamic Characterization of the Molecular Hub Early
715 Region 1A (E1A) from Human Adenovirus. *Chemistry (Easton)* **22**, 13010–13013 (2016).
- 716 47. Haberz, P., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the
717 interactions of adenoviral E1A proteins with the p160 nuclear receptor coactivator binding
718 domain of CBP. *Protein Sci* **25**, 2256–2267 (2016).
- 719 48. Zuiderweg, E. R. P. Mapping protein-protein interactions in solution by NMR
720 spectroscopy. *Biochemistry* **41**, 1–7 (2002).
- 721 49. Palopoli, N., Gonzalez Foutel, N. S., Gibson, T. J. & Chemes, L. B. Short linear motif core
722 and flanking regions modulate retinoblastoma protein binding affinity and specificity.
723 *Protein Eng Des Sel* **31**, 69–77 (2018).
- 724 50. Perozzo, R., Folkers, G. & Scapozza, L. Thermodynamics of protein-ligand interactions:
725 history, presence, and future aspects. *J Recept Signal Transduct Res* **24**, 1–52 (2004).
- 726 51. Theisen, F. F. *et al.* Quantification of Conformational Entropy Unravels Effect of
727 Disordered Flanking Region in Coupled Folding and Binding. *J Am Chem Soc* **143**,
728 14540–14550 (2021).
- 729 52. Bernado, P., Mylonas, E., Petoukhov, M. V, Blackledge, M. & Svergun, D. I. Structural
730 characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*
731 **129**, 5656–5664 (2007).
- 732 53. Estaña, A. *et al.* Realistic Ensemble Models of Intrinsically Disordered Proteins Using
733 a Structure-Encoding Coil Database. *Structure* **27**, 381-391.e2 (2019).
- 734 54. Cortes, J., Simeon, T., Remaud-Simeon, M. & Tran, V. Geometric algorithms for the
735 conformational analysis of long protein loops. *J Comput Chem* **25**, 956–967 (2004).

- 736 55. Cohan, M. C., Eddelbuettel, A. M. P., Levin, P. A. & Pappu, R. V. Dissecting the
737 Functional Contributions of the Intrinsically Disordered C-terminal Tail of *Bacillus subtilis*
738 FtsZ. *J Mol Biol* **432**, 3205–3221 (2020).
- 739 56. Glavina, J., Rodriguez de la Vega, R., Risso, V.A., Leonetti, C.O., Chemes, L.B.,
740 Sánchez, I.E. Host diversification is concurrent with linear motif evolution in a
741 mastadenovirus hub protein. *Journal of Molecular Biology* (2022).
- 742 57. Hoppe, E. *et al.* Multiple Cross-Species Transmission Events of Human Adenoviruses
743 (HAdV) during Hominine Evolution. *Mol Biol Evol* **32**, 2072–2084 (2015).
- 744 58. Glavina, J. *et al.* Interplay between sequence, structure and linear motifs in the
745 adenovirus E1A hub protein. *Virology* **525**, 117–131 (2018).
- 746 59. Lau, L., Gray, E. E., Brunette, R. L. & Stetson, D. B. DNA tumor virus oncogenes
747 antagonize the cGAS-STING DNA-sensing pathway. *Science* **350**, 568–571 (2015).
- 748 60. Ferreiro, D. U., Komives, E. A. & Wolynes, P. G. Frustration in biomolecules. *Q Rev*
749 *Biophys* **47**, 285–363 (2014).
- 750 61. Sherry, K. P., Johnson, S. E., Hatem, C. L., Majumdar, A. & Barrick, D. Effects of Linker
751 Length and Transient Secondary Structure Elements in the Intrinsically Disordered Notch
752 RAM Region on Notch Signaling. *J Mol Biol* **427**, 3587–3597 (2015).
- 753 62. Crisostomo, L., Soriano, A. M., Mendez, M., Graves, D. & Pelka, P. Temporal dynamics
754 of adenovirus 5 gene expression in normal human cells. *PLoS One* **14**, e0211192 (2019).
- 755 63. Ramirez, J. *et al.* Targeting the Two Oncogenic Functional Sites of the HPV E6
756 Oncoprotein with a High-Affinity Bivalent Ligand. *Angew Chem Int Ed Engl* **54**, 7958–
757 7962 (2015).
- 758 64. Cheng, J. *et al.* Stabilized recombinant suppressors of RNA silencing: functional effects
759 of linking monomers of Carnation Italian Ringspot virus p19. *Biochim Biophys Acta* **1774**,
760 1528–1535 (2007).

- 761 65. Travers, T. *et al.* Combinatorial diversity of Syk recruitment driven by its multivalent
762 engagement with FcεR1γ. *Mol Biol Cell* **30**, 2331–2347 (2019).
- 763 66. Milles, S. *et al.* Plasticity of an ultrafast interaction between nucleoporins and nuclear
764 transport receptors. *Cell* **163**, 734–745 (2015).
- 765
- 766

767 **METHODS**

768

769 ***Protein purification and peptide synthesis and labeling***

770 *Protein expression and purification*: The human Retinoblastoma protein (Uniprot
771 ID: P06400) AB domain with a stabilizing loop deletion (372-787Δ582-642), named Rb,
772 was recombinantly expressed from a pRSET-A vector in *E. coli* BL21(DE3). Briefly, Rb
773 cultures were induced with 1 mM IPTG and grown at 28 °C overnight. Rb was purified
774 from the soluble fraction using a Ni²⁺-nitrilotriacetic acid immobilized metal affinity
775 chromatography resin, followed by a purification with a sulfate cation exchange (SP-
776 Sepharose) resin and size exclusion (Superdex 75) chromatography [67]. The
777 adenovirus serotype 5 (HAdV5) Early 1A protein fragment (36-146) (Uniprot ID:
778 P03255), named E1A_{WT}, was subcloned into *Bam*HI/*Hind*III sites of a modified pMalC2x
779 vector (NewEnglandBioLabs, Hitchin, UK). E1A_{ΔE} (43-LHELY-47Δ43-AAAA-46) and
780 E1A_{ΔL} (122-LTCHE-126Δ122-AAAA-125) variants were obtained by site-directed
781 mutagenesis of the wild type vector. E1A proteins were expressed as MBP fusion
782 products in *E. coli* BL21(DE3). Unlabeled and single (¹⁵N) and double (¹⁵N/¹³C) labeled
783 samples were obtained from 2TY medium and M9-minimal medium supplemented with
784 ¹⁵NH₄Cl and ¹³C-glucose respectively. Cultures were induced with 0.8 mM IPTG at 0.7
785 OD₆₀₀ and grown at 37 °C overnight in 2TY medium or for 5 h after induction in M9-
786 minimal medium. Harvested cells were lysed by sonication and proteins isolated
787 performing amylose affinity chromatography of the soluble fraction, followed by Q-
788 HyperD Ion exchange and size exclusion (Superdex 75) chromatography. The MBP tag
789 was cleaved with Thrombin (Sigma-Aldrich, USA) at 0.4 unit per mg of protein. Synthetic
790 MBP-E1A fusion constructs (construct sequences available in the **Source File for**
791 **Figure 4**) subcloned into the pMalC4x vector (GenScript, USA) were expressed in *E.*
792 *coli* BL21(DE3) followed by Amylose purification and Superdex 75 chromatography as

793 described above. All E1A protein stocks were stored at -80 °C in buffer containing 20mM
794 Sodium Phosphate pH 7.0, 200mM NaCl, 20mM DTT and 2mM PMSF. Protein purity (>
795 90%) and conformation were assessed by SDS-PAGE, SEC-SLS and circular dichroism
796 analysis (**Extended Data Fig. 1**).

797 Peptide synthesis: Peptides corresponding to individual E1A or E2F2 binding
798 motifs were synthesized by FMoc chemistry at >95% purity (GenScript, USA) and
799 quantified by Absorbance at 280 nm or by quantitation of peptide bonds at 220 nm in
800 HCl -when Tryptophan or Tyrosine residues were absent. The peptide sequences are:

801

802	E1A _{E2F}	36-SHFEPPTLHELYDLDV-51
803	E1A _{LxCxE}	116-VPEVIDLTCHEAGFPP-131
804	E1A _{LxCxE-AC}	116-VPEVIDLTCHEAGFPPSDDDEDEEG-139
805	E1A _{LxCxE-ACP}	116-VPEVIDLTCHEAGFPP _p SDDDEDEEG-139
806	Human E2F2	404-SPSLDQDDYLWGLEAGEGISDLFD-427

807 FITC labeling: Proteins and peptides were labeled at their N-terminus with
808 Fluorescein 5-Isothiocyanate (FITC, Sigma), purified and quantified following a
809 described protocol [67]. F/P (FITC/Protein) ratio was above 0.8 in all cases.

810 **Circular Dichroism (CD)**

811 Far-UV CD spectra were measured on a Jasco J-810 (Jasco, Japan)
812 spectropolarimeter equipped with a Peltier thermostat using 0.1 or 0.2 cm path-length
813 quartz cuvettes (Hellma, USA). Five CD scans were averaged from 195 to 200 nm at
814 100nm/min scan speed, and buffer spectra were subtracted from all measurements. All
815 spectra were measured in 10mM Sodium Phosphate buffer pH 7.0 and 2mM DTT at 20
816 ± 1 °C and 5 µM protein concentration.

817

818 **Size Exclusion Chromatography, Hydrodynamic radii calculations and Light**
819 **Scattering Experiments**

820 Analytical size exclusion chromatography (SEC) was performed on a Superdex
821 75 column (GE Healthcare) calibrated with globular standards: BSA (66 kDa), MBP (45
822 kDa) and Lysozyme (14.3 kDa). All runs were performed by injecting 100 μ l protein
823 sample (E1A_{WT} and E1A_{ΔL} at 270 μ M and E1A_{ΔE} at 540 μ M) in 20 mM Sodium
824 Phosphate buffer pH 7.0, 200 mM NaCl, 2 mM DTT. For each protein or complex a
825 partition coefficient (K_{av}) was calculated and apparent molecular weights were
826 interpolated from the $-\log MW$ vs K_{av} calibration curve. Experimental hydrodynamic radii
827 (R_h) were calculated following empirical formulations developed by Uversky and col. [68]:

$$828 \log R_h = -0.204 + 0.357 \log MW \quad (1)$$

829 Where MW is the apparent molecular weight derived from SEC experiments.
830 The predicted R_h for E1A_{WT} was calculated following the formulation developed by Marsh
831 and Forman-Kay [3].

832 The exponent ν was calculated from $R_h = R_o \cdot N^\nu$ using the experimental R_h values,
833 with $R_o = 2.49$ nm for E1A_{WT} and $R_o = 4.92$ nm for Rb, following [28]. For E1A_{WT}, ν was
834 calculated from $R_g = R_o \cdot N^\nu$ using R_g obtained from SAXS measurements and $R_o = 2.1$
835 nm, following [69]. In both cases, N is the number of residues in the chain
836 (**Supplementary Data Table 2**).

837 Static Light Scattering (SLS) coupled to SEC was carried out to determine the
838 average molecular weight of individual protein peaks and the stoichiometry of [E1A:Rb]
839 complexes using a PD2010 detector (Precision Detectors Inc, China) coupled in tandem
840 to an HPLC system and an LKB 2142 differential refractometer. The 90° light scattering
841 (LS) and refractive index (RI) signals of the eluting material were analyzed with
842 Discovery32 software (Precision Detectors).

843 Dynamic Light Scattering (DLS) was used to measure the hydrodynamic size
844 distribution of E1A, using a Wyatt Dynapro Spectrometer (Wyatt Technologies,
845 USA). Data was fitted using Dynamics 6.1 software. All measurements were performed
846 in 20 mM Sodium Phosphate buffer pH 7.0, 200 mM NaCl, 1 mM DTT at 2 mg/ml.
847 Samples were filtered by 0.22 μ M filters (Millipore) and placed into a 96 Well glass
848 bottom black plate (In Vitro Scientific P96-1.5H-N) covered by a high performance cover
849 glass (0.17+/-0.005mm) before measurements were taken.

850

851 ***Fluorescence Spectroscopy Experiments***

852 Measurements were performed in a Jasco FP-6200 (Nikota, Japan)
853 spectropolarimeter assembled in L geometry coupled to a Peltier thermostat. Excitation
854 and emission wavelengths were 495 nm and 520 nm respectively, with a 4 nm
855 bandwidth. All measurements were performed in 20 mM Sodium Phosphate buffer pH
856 7.0, 200 mM NaCl, 2 mM DTT and 0.1% Tween-20 at 20 ± 1 °C.

857 For direct titrations, a fixed concentration of FITC-labeled protein/peptide was
858 titrated with increasing amounts of Rb until saturation was reached. Maximal dilution
859 was 20% and samples equilibrated for 2 min ensuring steady state. Titrations performed
860 at concentrations 10 times higher than the equilibrium dissociation constant (K_D) allowed
861 estimation of the stoichiometry of each reaction. Binding titrations performed at sub-
862 stoichiometric concentrations allowed an estimation of K_D , by fitting the titration curves
863 to a bimolecular association model:

$$864 \quad Y = Y_F + \frac{(Y_B - Y_F)}{P_0} * \frac{(x + P_0 + K_D) + \sqrt{(x - P_0 + K_D)^2 - (4 * P_0 * x)}}{2} + C * x \quad (2)$$

865 Where Y is the measured anisotropy signal, Y_F and Y_B are the free and bound
866 labeled peptide signals, P_0 is the total labeled peptide concentration, x is Rb
867 concentration, and K_D is the equilibrium dissociation constant in Molar units. The $[C * x]$

868 linear term accounts for slight bleaching or aggregation. Data was fitted using the Profit
869 7.0 software (Quantumsoft, Switzerland), yielding a value for each parameter and its
870 corresponding standard deviation. Titrations for each complex were performed in
871 triplicate at least at three different concentrations of FITC-labeled sample, and
872 parameters were obtained from fitting individual titrations or by global fitting of the K_D
873 parameter using normalized titration curves at different concentrations, obtaining an
874 excellent agreement between individual and global fits (**Supplementary Data Table 3**
875 **and Extended Data Fig. 3**).

876 Competition experiments were carried out by titrating the pre-assembled
877 complex [Rb:FITC-E2F2] (1:1 molar ratio, 5 nM) with increasing amounts of unlabeled
878 competitors and following the decrease in the anisotropy signal until the value
879 corresponding to free FITC-E2F2 was reached. IC50 values were estimated directly
880 from the curves as the concentration where the competitor produced a decrease in 50%
881 of the maximal anisotropy value. K_D values were calculated by fitting the data
882 considering the binding equilibrium of the labeled peptide and the unlabeled competitors
883 according to [70], obtaining $K_{D(comp)}$ values that differed only slightly (2 to 3-fold) from
884 those obtained from direct titrations. K_D and $K_{D(comp)}$ values also displayed similar fold
885 changes in binding affinity relative to E2F2 within each method (**Supplementary Data**
886 **Table 1**). The agreement between the K_D values obtained from fluorescence and ITC
887 titrations (**Supplementary Data Table 1**) confirmed that FITC moiety did not cause
888 significant changes in Rb binding affinity. MBP-E1A fusion protein affinities
889 (**Supplementary Data Table 4 and Extended Data Fig. 8**) were determined by
890 performing competition experiments assembling a [Rb:FITC-E2F2] complex at 10nM
891 concentration, after verifying that MBP-E1A_{WT} and E1A_{WT} (cleaved and uncleaved
892 HAdV5 proteins) had the same binding affinity (Extended Data Table 7). Measurements
893 were performed on a PTI Quantamaster QM40 spectrofluorimeter (Horiba, Japan)

894 equipped with polymer film polarizers and coupled to a Peltier thermostat with excitation
895 parameters as described above.

896

897 ***ITC Experiments***

898 Direct titrations. ITC experiments were performed on MicroCal VP-ITC and
899 MicroCal PEAQ-ITC equipment (Malvern Panalytical) in 20 mM Sodium Phosphate pH
900 7.0, 200 mM NaCl, 5mM 2-mercapto ethanol at 20.0 ± 0.1 °C, unless stated otherwise.
901 Prior to titrations, cell and titrating samples were co-dialyzed in the aforementioned
902 buffer for 48 h at 4 ± 1 °C and then de-gassed. Measurements performed in the MicroCal
903 VP-ITC used 28 10- μ l injections at a flow rate of 0.5 μ l/s and those performed in the
904 MicroCal PEAQ-ITC used 13 3- μ l injections. The concentration range of cell and titrating
905 samples are detailed in **Extended Data Figs. 2 and 5**. Data were analyzed using the
906 Origin software.

907 Allosteric coupling experiments. First, a pre-assembled [Rb:E1A_{LxCxE}] complex
908 (1:1 molar ratio, 30 μ M) was titrated with E1A_{E2F} or E1A _{Δ L} to assess whether binding of
909 the LxCxE motif modified the binding affinity for the E2F site. Conversely, pre-
910 assembled [Rb: E1A_{E2F}] or [Rb: E1A _{Δ L}] complexes were titrated with E1A_{LxCxE} to assess
911 whether binding of the E2F motif modified the binding affinity for the LxCxE site
912 (**Supplementary Data Table 7**).

913 Calculation of ΔC_p and ΔASA_T parameters from ITC data. A series of titrations
914 were carried out at different temperatures (10.0, 15.0, 20.0 and 30.0 ± 0.1 °C) and the
915 change in binding heat capacity (ΔC_p) was obtained from the slope of the linear
916 regression analysis of the plot of ΔH vs temperature (**Extended Data Fig. 5**). The
917 changes in accessible surface area (ΔASA_T) and the number of residues that fold upon
918 binding (X_{res}) were estimated by solving semi-empirical equations from protein folding
919 studies applied to protein-ligand binding [⁵⁰] and from models that use parameters

920 derived for intrinsically disordered proteins (IDP-specific model) [51] (Details on the
 921 model-specific parameter values are provided in **Supplementary Data Tables 5 and**
 922 **6**). First we calculated $\Delta H_{int(T)}$ from:

$$923 \quad \Delta H_{int(T)} = \Delta H_{int(T_H)} + \Delta C_p (T - T_H) \quad (3)$$

924 where $\Delta H_{int(T)}$ is the change in enthalpy measured at experimental temperatures
 925 (K), ΔC_p is the change in heat capacity, $\Delta H_{int(T_H)}$ is the change in enthalpy at the
 926 temperature of enthalpic convergence and T_H is the temperature of enthalpic
 927 convergence at which the apolar contribution is assumed to be zero (295.15 K). Then,
 928 ΔASA_T values were calculated as the sum of the contribution of changes in polar
 929 (ΔASA_p) and non-polar (ΔASA_{np}) accessible surface areas, by solving the following set
 930 of equations:

$$931 \quad \Delta H_{int(T_H)} = \Delta h_{np} \Delta ASA_{np} + \Delta h_p \Delta ASA_p \quad (4)$$

$$932 \quad \Delta C_p = \Delta c_{np} \Delta ASA_{np} + \Delta c_p \Delta ASA_p \quad (5)$$

933 where Δh_{np} , Δh_p , Δc_{np} and Δc_p are constants that assume different values
 934 according to the model used [50,51] (See **Supplementary Data Table 6** for specific
 935 values). Finally, X_{res} was calculated from:

$$936 \quad X_{res} = \Delta S_{config} / \Delta S_{residue} \quad (6)$$

937 Where $\Delta S_{residue}$ is the change in configurational entropy per residue and the
 938 change in configurational entropy (ΔS_{config}) was calculated as the sum of changes in
 939 rotation-translation (ΔS_{rt}) and solvation (ΔS_{solv}) entropy:

$$940 \quad \Delta S_{config} = \Delta S_{rt} + \Delta S_{solv} \quad (7)$$

941 with ΔS_{solv} defined as:

$$942 \quad \Delta S_{solv} = C_1 * \Delta C_p * \ln (T/T_S) \quad (8)$$

943 where T is the experimental temperature (K) and T_S is the temperature for
944 entropic convergence (385 K). ΔS_{rt} , $\Delta S_{residue}$ and the constant C_1 , which depend on the
945 relationship of apolar to polar surface area, assume different values depending on the
946 model used [50,51] (see **Supplementary Data Table 6** for model-specific values).

947

948 ***NMR Experiments***

949 NMR experiments were carried out using a Varian VNMRs 800 MHz
950 spectrometer equipped with triple resonance pulse field Z-axis gradient cold probe. A
951 series of two-dimensional sensitivity-enhanced ^1H - ^{15}N HSQC and three-dimensional
952 HNCACB, HNCO and CBCA(CO)NH experiments [71,72] were performed for backbone
953 resonance assignments on uniformly ^{13}C - ^{15}N -labeled samples of E1A_{WT}, E1A_{ΔE} and
954 E1A_{ΔL} at 700 μM, 975 μM and 850 μM respectively. All measurements were performed
955 in 10 % D2O, 20 mM Sodium Phosphate pH 7.0, 200 mM NaCl, 2 mM DTT at 25 °C.
956 The HSQC used 9689.9 Hz and 1024 increments for the t1 dimension and 2106.4 Hz
957 with 128 increments for the t2. The HNCACB used 9689.9, 14075.1, and 2106.4 Hz,
958 with 1024, 128, and 32 increments for the t1, t2, and t3 dimensions, respectively. The
959 HNCO used 9689.9, 2010.4 Hz, and 2106.4 Hz with 1024, 64, and 32 increments for
960 the t1, t2, and t3 dimensions, respectively. The CBCA(CO)NH used 9689.9, 14072.6,
961 and 2106.4 Hz, with 1024, 128, and 32 increments for the t1, t2, and t3 dimensions,
962 respectively. For E1A_{WT} 88% of non-proline backbone ^1H and ^{15}N nuclei, 75% of $^{13}\text{C}'$
963 nuclei and 90% of $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ of E1A nuclei were assigned (**Supplementary Data**
964 **File 2**). For E1A_{ΔE} and E1A_{ΔL} 85% of non-proline backbone ^1H and ^{15}N nuclei, 72% of
965 $^{13}\text{C}'$ nuclei and 87% of $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ E1A nuclei were assigned.

966 NMRPipe and NMRViewJ software packages were used to process and analyze
967 all the NMR spectra [73]. Residue-specific random coil chemical shifts were generated

968 for the three sequences using the neighbor-corrected IDP chemical shift library [74].
969 Secondary chemical shifts ($\Delta \delta$), were calculated by subtracting random coil chemical
970 shifts from the experimentally obtained chemical shifts.

971 Two-dimensional ^1H - ^{15}N TROSY experiments were performed on single ^{15}N -
972 labeled samples of free E1A_{WT}, E1A _{Δ E} and E1A _{Δ L} and on each E1A protein bound
973 stoichiometrically to Rb (1:1 molar ratio) at 525 μM (E1A_{WT}), 300 μM (E1A _{Δ E}) and 315
974 μM (E1A _{Δ L}). The ratio between the peak intensity in the bound state (I) and the peak
975 intensity in the free state (I_0) was calculated, allowing interacting residues to be
976 determined together with additional data.

977

978 ***Molecular modelling of Rb:E1A conformational ensembles***

979 Conformations of E1A_{WT} bound to Rb were modeled using an extended version
980 of a recently proposed method to generate realistic conformational ensembles of IDPs
981 [53]. Conformational ensemble models of [E1A_{WT}:Rb] were generated using a stochastic
982 sampling algorithm implemented in the MoMA software suite (<https://moma.laas.fr>). This
983 method exploits local, sequence-dependent structural information encoded in a
984 database of three-residue fragments and builds conformations incrementally sampling
985 dihedral angles values from the database, while avoiding steric clashes. In order to
986 model the double-bound [Rb:E1A_{WT}] complex, the E2F and LxCxE motifs were
987 considered to be static, preserving the conformations extracted from experimentally
988 determined structures (2R7G and 1GUX). The 71-residue fragment between these two
989 motifs was considered as a long protein loop that adapts its conformation in order to
990 maintain the two ends rigidly positioned. Conformational sampling considering such
991 loop-closure constraints was performed using a robotics-inspired method [54] adapted
992 to use dihedral angle values from the aforementioned database. For each feasible
993 conformation of the central fragment, geometrically compatible conformations of the

994 short N- and C-terminal tails were sampled using the basic strategy explained in [53]. For
995 singly bound models [E1A_{ΔL}:Rb] and [E1A_{ΔE}:Rb], only one of the two motifs were
996 considered to be statically bound to Rb and the other motif behaved as the flexible linker.
997 The loop sampling method used to model the linker between the two binding motifs can
998 be used via a web server (<https://moma.laas.fr/applications/LoopSampler/>). Binaries
999 can be provided upon request.

1000

1001 **SAXS Experiments**

1002 SAXS experiments for Rb and [E1A_{WT}:Rb] were carried out at the European
1003 Molecular Biology Laboratory beamline P12 of PETRAIII storage ring, using the X-ray
1004 wavelengths of 1.24 Å and a sample-to-detector distance of 3.0 m [75]. The scattering
1005 profiles measured covered a momentum transfer range of $0.0026 < s < 0.73 \text{ \AA}^{-1}$. SAXS
1006 data for E1A was collected at the at the SWING beamline at the SOLEIL synchrotron,
1007 France, on an Eiger 4M detector with a sample-to-detector distance of 2.0 m. SAXS
1008 data were measured for Rb, E1A_{WT} and the [E1A_{WT}:Rb] complex at 10° C.
1009 Concentrations used for E1A_{WT} were 7.0, 5.6 and 4.2 mg/ml, for Rb were 4.0, 2.0, 1.0
1010 mg/ml, and for and [E1A_{WT}:Rb] were 2.7, 1.4, and 0.7 mg/ml, in 20 mM Sodium
1011 Phosphate pH 7.0, 200 mM NaCl, 1mM DTT. The scattering patterns of the buffer
1012 solution were recorded before and after the measurement of each sample. Multiple
1013 repetitive measurements were performed to detect and correct for radiation damage.
1014 The initial data processing steps including masking and azimuthal averaging were
1015 performed using the SASFLOW version 3.0. pipeline for Rb and [E1A_{WT}:Rb] and the
1016 program FOXTROT version 3.5.2. [76] for E1A. Final curves at each concentration were
1017 derived after the averaged buffer scattering patterns were subtracted from the protein
1018 sample patterns. No sign of aggregation was observed in any of the curves. Final SAXS

1019 profiles for the systems were obtained by merging curves for the lowest and highest
1020 concentrations to correct small attractive interparticle effects observed. The SAXS
1021 profiles were analyzed using the ATSAS suite of programs version 2.8.4 [77]. The
1022 forward scattering intensity, $I(0)$, and the radius of gyration, R_g , were evaluated using
1023 Guinier's approximation [78], assuming that at very small angles ($s < 1.3/R_g$, the intensity
1024 can be well represented as $I(s) = I(0) \exp(-(sR_g)^2/3)$). The $P(r)$ distribution functions
1025 were calculated by indirect Fourier Transform using GNOM [79] applying a momentum
1026 transfer range of $0.01 < s < 0.33 \text{ \AA}^{-1}$ and $0.013 < s < 0.27 \text{ \AA}^{-1}$ for Rb and [Rb:E1A],
1027 respectively. For E1A_{WT} a SEC-SAXS experiment was also performed which was
1028 processed using the program CHROMIX [80] which is a part of ATSAS 2.8.4. to obtain
1029 the SAXS profile from a highly monodisperse sample. This profile overlaid perfectly with
1030 the E1A_{WT} merged curve from the three batch experiments, discarding aggregation
1031 problems.

1032 The fitting of the crystallographic structure of Rb (PDB: 3POM [81]) to the
1033 experimental SAXS curve was performed with FOXS [82,83]. An optimal fit ($\chi^2=0.86$) was
1034 obtained after modelling the missing parts (loops, N- and C-termini) and a subsequent
1035 refinement with the program AllosMod-FoXS [84]. SAXS data measured for [Rb:E1A]
1036 were analyzed with the Ensemble Optimization Method (EOM) [52,85]. Briefly, theoretical
1037 SAXS profiles of the 10250 structures of the complex were computed with CRY SOL [86].
1038 200 different sub-ensembles of 20 or 50 conformations collectively describing the
1039 experimental curve were collected with EOM and analyzed in terms of R_g distributions.
1040 The experimental SAXS data of [E1A_{WT}:Rb] complex is compatible with three distinct
1041 scenarios: a 100% doubly-bound ensemble where the linker is highly expanded, a 100%
1042 singly-bound ensemble where the linker is highly compact and thirdly, an ensemble with

1043 a combination of 76% doubly bound and :24% singly-bound species, which resulted
1044 from the linear combination of a curve representing the ensemble average of all singly-
1045 and all doubly-bound conformations. However, thermodynamic (K_D for $E1A_{WT}$) data
1046 strongly argue against the last two scenarios as it indicates an extremely low expected
1047 population of the singly-bound forms at any concentration of the complex used in the
1048 SAXS experiments.

1049

1050 ***Hydrodynamic radii for generated conformations***

1051 Hydrodynamic radii were calculated using the program HydroPro (version 10)
1052 [87,88]. HydroPro was run on 1000 models selected by EOM for the doubly-bound
1053 conformations and 1000 randomly selected conformations of N- and C-terminal bound
1054 conformations. The calculations were done at temperatures of 20 and 25 °C with
1055 corresponding solvent viscosities of 0.01 and 0.009 poise, respectively. The values of
1056 atomic element radius (AER), Molecular Weight, Partial Specific Volume and Solvent
1057 Density were set to 2.9 Å, 54590 Da, 0.702 cm³/g and 1.0 g/cm³, respectively. These
1058 values have no associated error.

1059

1060 ***All-atom simulations of E1A Linker sequences***

1061 All-atom simulations were run using the CAMPARI simulation engine (V2)
1062 Version 2.0 (<http://campari.sourceforge.net>) and ABSINTH implicit solvent model ABS-
1063 OPLS3.2 [89,90]. All simulations were run at 320 K; while this is a slightly elevated
1064 temperature compared to the experimental temperature, none of the terms the
1065 Hamiltonian lacks temperature dependence such that this slightly high temperature
1066 serves to improve sampling quality in a uniform way across all simulations. This
1067 approach has been leveraged to great effect in previous studies and is especially
1068 convenient in the case of simulating many different sequences that span a range of

1069 sequence properties and lengths [8]. A collection of Monte Carlo moves was used to
1070 fully sample conformational space as previously described [91,92,13].

1071 For all simulations of natural sequences, 15 independent simulations were run
1072 per sequence for a total of 90K conformations per sequence across 27 different
1073 sequences (405 independent simulations, 5.25×10^8 Monte Carlo steps per sequence).
1074 Simulations were performed in 15 mM NaCl in a simulation droplet size sufficiently large
1075 for each sequence, calibrated in a length dependent manner. Simulations were analyzed
1076 using the MDTraj package version 1.9.5 [93] and SOURSOP version 0.1.3
1077 (<https://soursop.readthedocs.io/>). Sequence analysis was performed using the local
1078 CIDER software package [94] with all parameters reported in the **Source Data for**
1079 **Extended Data Figure 7**. Normalized end-to-end distance was calculated as the
1080 absolute end-to-end distance divided by the end-to-end distance expected for an
1081 equivalently long Gaussian chain. Motif-linker-motif simulations were performed in a
1082 manner analogous to the linker-only motifs. Each independent simulation was run for 86
1083 $\times 10^6$ steps, with 6×10^6 steps discarded as equilibration and conformations saved every
1084 50,000 steps. Over 10 independent replicas, this approach generates ensembles of
1085 16,000 conformations per sequence. To calculate the hydrodynamic radius we used the
1086 approach of Nygaard et al. to convert the radius of gyration into the hydrodynamic radius
1087 [95]. Code for this conversion is provided in the supporting GitHub repository.

1088

1089 ***Length titration Simulations***

1090 The linker from HF_HAdV40 was used to determine the overall amino acid
1091 composition and generate random sequences across a range of lengths that
1092 recapitulated this composition. Specifically, for each length (45, 50, 55, 60, 65, 70, 75)
1093 twenty random sequences were generated for a total of 140 randomly generated

1094 sequences. Each sequence was simulated under equivalent simulation conditions for
1095 35×10^9 simulation steps, with the goal of elucidating the general relationship between
1096 sequence length and end-to-end distance for an arbitrary sequence of the composition
1097 associated with HF_HAdV40. The mean end-to-end distance for the collection of
1098 sequences at a given length was determined, such that the mean value is a double
1099 average over both conformational space and sequence space.

1100

1101 **WLC modelling**

1102 The worm like chain (WLC) model: A worm like chain (WLC) model [22] was used
1103 to describe the end-to-end probability density distribution function of the E1A linker and
1104 estimate the effective concentration term (C_{eff}) used in the tethering model (**Fig. 1,**
1105 **Model A and Fig. 3**). In this model, the disordered linker behaves as a random polymer
1106 chain whose dimensions depend on the persistence length (L_p), which represents the
1107 chain stiffness, or the length it takes for the chain motions to become uncorrelated and
1108 on the contour length (L_c), which is the total length of the chain. For long peptides, L_p
1109 assumes a standard value of 3\AA and L_c is $L_c = N_{res} * b$, where N_{res} is the number of linker
1110 residues and b is the average unit size of one amino acid (3.8\AA) [23]. Under this model,
1111 the probability density function $p(r)$ is defined by:

$$1112 \quad p(r) = 4\pi r^2 \left(\frac{3}{4\pi L_p L_c} \right)^{\frac{3}{2}} \exp \left(\frac{-3r^2}{4L_p L_c} \right) \zeta(r, L_p, L_c) \quad (3)$$

1113 Where $p(r)$ is a function of distance r and depends on L_p and L_c . The last term in
1114 the equation is expanded in [22, 23]. The end-to-end probability density function can be
1115 related to the effective concentration in the bound state when the linker is restrained to
1116 a fixed distance between binding sites, r_o [22]. In this case, the effective concentration
1117 C_{eff} is defined by:

1118
$$C_{\text{eff}} = \frac{p(r_0) 10^{27} \text{\AA}^3 1^{-1}}{4\pi r^2 N_A} \quad (4)$$

1119 Where N_A is Avogadro's number and (r_0) is the distance separating the binding
 1120 sites obtained from the X-ray structure of the complex (49 Å calculated from PDB: 2R7G
 1121 [40] and 1GUX [41]). Multiplying Eq. (4) by 10^3 yields C_{eff} in millimolar units.

1122

1123 **Calculation of experimental and predicted C_{eff} values for the E1A_{WT}:Rb interaction**

1124 Experimental C_{eff} values: In Model A the global dissociation constant is calculated as:
 1125 $K_G = K_1 * K_2 * C_{\text{eff}}$ (**Fig. 1f**) where K_G , K_1 and K_2 are equilibrium association constants ($K =$
 1126 $1/K_D$). Therefore, this relationship can be expressed equivalently as $K_D =$
 1127 $K_{D,E2F} * K_{D,LxCxE} * C_{\text{eff}}^{-1}$. Here, $K_{D,E2F}$ and $K_{D,LxCxE}$ are the equilibrium dissociation constants
 1128 of the E1A_{E2F} and E1A_{LxCxE} motifs respectively (reported in **Supplementary Data Table**
 1129 **1**) and is the equilibrium dissociation constant for E1A_{WT} (reported in **Supplementary**
 1130 **Data Table 1**). The condition $K_1 = K_1'$ and $K_2 = K_2'$ (no allosteric coupling between sites)
 1131 was met (**Extended Data Fig. 5 and Supplementary Data Table 7**). Therefore, the
 1132 experimentally derived C_{eff} was calculated from the measured binding constants as: C_{eff}
 1133 $= (K_{D,E2F} * K_{D,LxCxE}) / K_D$ (**Fig. 3a**).

1134 Predicted C_{eff} values: The C_{eff} value predicted from the WLC model (**Fig. 3a**) was
 1135 obtained by applying Eq. [4] with the designated L_p parameter (standard model $L_p = 3\text{\AA}$
 1136 and $b = 3.8\text{\AA}$), using a linker length of 71 residues for HAdV5 E1A. The separation
 1137 between binding sites, r_0 , was 49 Å (from PDB:1GUX and PDB:2R7G).

1138

1139 **Calculation of predicted global binding affinity for grafted E1A linkers**

1140 We predicted the K_D values expected for each of the grafted linker variants of
 1141 **Figure 4c,d** under a sequence-independent WLC model or accounting for sequence-

1142 dependent changes in the persistence length. We calculated K_D for E1A_{WT} as $K_D =$
 1143 $(K_{D,E2F} * K_{D,LxCxE}) / C_{eff}$, where $K_{D,E2F}$ and $K_{D,LxCxE}$ are the K_D values for the E1A_{E2F} and
 1144 E1A_{LxCxE} motifs of E1A_{WT} (reported in **Supplementary Data Table 1**) and C_{eff} was
 1145 calculated using the WLC model as described below. Competition experiments for
 1146 E1A_{WT} typically yield K_D values ~3-fold higher than those obtained by direct titration
 1147 (**Supplementary Data Tables 1 vs 4**). To correct for this effect, the predicted K_D values
 1148 were corrected by a factor of three. Since the only element changing for each grafted
 1149 variant in the grafting experiment is the linker (i.e. the C_{eff} value), we calculated $K_{D,VARIANT}$
 1150 as $K_{D,VARIANT} = K_{DE1AWT} / (C_{eff}Ratio)$, where $C_{eff}Ratio = C_{eff.VARIANT} / C_{eff.E1AWT}$. All predicted
 1151 C_{eff} , L_{pSim} and K_D values for the grafting experiment are reported in the **Source Data**
 1152 **Files for Figure 5 and Extended Data Figure 10**. The C_{eff} values were calculated as
 1153 follows:

1154 WLC-Lp=3 model: For the standard assumption of a sequence-independent model
 1155 (WLC-Lp=3) we calculated the C_{eff} function as a function of linker length ($C_{eff(L)}$) using
 1156 equations (3) and (4) with a standard value for the persistence length parameter ($L_p =$
 1157 3). We calculated the expected K_D as a function of linker length as $K_D =$
 1158 $(K_{D,E2F} * K_{D,LxCxE}) / C_{eff(L)}$ (Straight line, **Fig. 4d**).

1159 WLC-LpSim model: For the sequence-dependent model (WLC-LpSim) we calculated
 1160 C_{eff} for each linker using equations (3) and (4) applying the specific number of residues
 1161 (N_{res}) of each linker and an individual sequence-dependent L_p value for each linker
 1162 (L_{pSim}), which was obtained from the simulations. L_{pSim} values were calculated from
 1163 the average end-to-end distance of each simulated ensemble using the equation $\langle r^2 \rangle =$
 1164 $2 * L_p * L_c$, where $L_c = N_{res} * b$ and b takes the value 3.8 Å. This equation is an approximation
 1165 for the value of $\langle r^2 \rangle$ for a worm like chain in the case where the contour length of the
 1166 chain is much larger than its persistence length ($L_c \gg L_p$) [23].

1167

1168 ***Calculation of predicted global binding affinity for a large family of E1A linkers***

1169 The WLC model was used to estimate the C_{eff} values and global Rb binding affinities of
1170 a collection of 110 natural linker sequences of different length changing the length value
1171 for each linker and keeping other parameters constant. All values are reported in **the**
1172 **Source Data File for Figure 5 and Extended Data Figure 10.**

1173 Dataset: A previously reported alignment and phylogenetic tree of 116 Mastadenovirus
1174 E1A sequences [56,58] was used to identify the E2F and LxCxE motifs as described 58,
1175 collecting 110 sequences in which both motifs were present (**Supplementary Data File**
1176 **1**). For all sequences, the length of the linker region between both motifs was recorded.
1177 Individual motif binding affinities, C_{eff} values and E1A global affinity ($K_{D,E1A}$) were
1178 calculated as explained below (**Source Data File for Figure 5 and Extended Data**
1179 **Figure 10**).

1180 Calculation of E1A binding affinity: The global binding affinity $K_{D,E1A}$ (**Extended Data**
1181 **Figure 10**) was calculated as $K_{D,E1A} = (K_{D,E2F} * K_{D,LxCxE}) / C_{eff}$, where C_{eff} is the C_{eff} value
1182 predicted under a naïve or sequence-dependent assumption (see details below) and
1183 $K_{D,E2F}$ and $K_{D,LxCxE}$ are the predicted binding affinities of each motif calculated using
1184 FoldX:

1185 Prediction of Motif binding affinities using FoldX: To estimate the binding affinity of
1186 individual E2F and LxCxE motifs ($K_{D,E2F}$ and $K_{D,LxCxE}$) present in each sequence, FoldX
1187 v5.0 [96] was used to build substitution matrices for all 20 amino acids at each position
1188 (**Source Data File for Extended Data Figure 10**). Briefly, given a structural complex
1189 the FoldX algorithm assesses the change in binding free energy produced by mutating
1190 each position of the motif for each one of the 20 amino acids. For the E2F matrix, the
1191 structure of the HAdV5 E1A_{E2F} motif in complex with Rb (PDB: 2R7G) was used as input.

1192 For the LxCxE matrix, the structure used as input was a model of the HAdV5 E1A_{LxCxE}
 1193 motif in complex with Rb (**Supplementary Data File 3**), built using FlexPepDock [⁹⁷]
 1194 and the structure of the HPV E7 LxCxE motif bound to Rb (PDB: 1GUX). The total
 1195 change in binding free energy with respect to the wild type sequence ($\Delta\Delta G_{\text{FoldX}}$) was
 1196 calculated by adding up the free energy terms for each residue at each matrix position
 1197 (**Source Data File for Extended Data Figure 10**). The predicted equilibrium
 1198 dissociation constant of the E2F and LxCxE motifs for each sequence ($K_{D \text{ SEQ}}$) was
 1199 calculated as:

$$1200 \quad \Delta\Delta G_{\text{FoldX}} = \Delta G_{\text{SEQ}} - \Delta G_{\text{WT}} = RT \ln(K_{D \text{ SEQ}}) - RT \ln(K_{D \text{ WT}}) \quad (5)$$

$$1201 \quad K_{D \text{ SEQ}} = \frac{K_{D \text{ WT}}}{e^{(-\Delta\Delta G_{\text{FoldX}}/RT)}} \quad (6)$$

1202 Where $\Delta\Delta G_{\text{FoldX}}$ is the total predicted change in binding energy calculated using
 1203 FoldX, RT is 0.582 kcal mol⁻¹, $K_{D \text{ WT}}$ is the experimentally measured binding affinity of
 1204 the sequence (HAdV5 E1A) present in the model structure ($K_{D,E2F}$ and $K_{D,LxCxE}$ measured
 1205 in this work, **Supplementary Data Table 1**).

1206 Prediction of C_{eff} values under the naïve WLC model: The C_{eff} value was calculated for
 1207 the collection of 110 natural E1A linkers using Equations (3) and (4) with $L_p = 3 \text{ \AA}$ (L_p
 1208 WLC, **Fig. 5a**) and the specific length (number of residues) of each linker, which defines
 1209 L_C .

1210 Prediction of C_{eff} values using a WLC model with sequence-dependent L_p parameters:

1211 For the subset of 27 natural E1A linkers used in all-atom simulations (**Fig. 4a**) we
 1212 calculated sequence-specific L_p values from all atom simulations (L_p^{Sim}) in order to
 1213 represent sequence-dependent changes in chain expansion. The details of these
 1214 calculations are explained under the Methods section “**Calculation of predicted global**
 1215 **binding affinity for grafted E1A linkers**”. New C_{eff} values were derived using the same

1216 parameters described above, but replacing the standard L_p value by the $L_{p\ Sim}$ value. The
1217 $L_p\ Sim$ values are reported in the **Source Data File for Extended Data Figure 10**.

1218 Statistical analysis. We used bootstrapping [98] to generate 99% confidence
1219 intervals (CI) for $K_{D,E2F}$, $K_{D,LxCxE}$ and $K_{D,E1A}$ average values, and compared the lower and
1220 upper end points against the value of $K_{D,E2F2}$ ($1\ 10^{-9}\ M$). The lower bound of the 99% CI
1221 for $K_{D,E2F}$ and $K_{D,LxCxE}$ is higher than $K_{D,E2F2}$ and the upper bound of the 99% CI for all
1222 $K_{D,E1A}$ are lower than $K_{D,E2F2}$. We also used permutation tests [98] to assess the null
1223 hypothesis that the C_{eff} , L_p and average K_D average values did not differ between all
1224 pairs of groups. In order to control for the false discovery rate, the p-values were
1225 corrected using the Benjamini-Hochberg [99] correction for multiple comparisons.

1226 Calculations of disorder propensity and conservation: All calculations were
1227 performed on the dataset from **Supplementary Data File 1**, using the methods
1228 described in [58]. For disorder propensity we recorded the mean IUPRED value \pm SD per
1229 position using IUPRED 2a [100] and for residue conservation we recorded the information
1230 content (IC) per position.

1231
1232 **Sequence conservation and Evolutionary Scores**
1233 We collected 77 mammalian orthologous sequences of the Retinoblastoma protein with no
1234 unassigned residues within the pocket domain from the Ensembl Database [101].
1235 Sequences were aligned using MUSCLE v3.8.1551 and manually curated according to
1236 structural information. The evolutionary conservation scores were calculated with the
1237 ConSurf 2016 webserver [102] using the E7-Rb complex (PDB: 1GUX) as the structural
1238 model. The sequence analyses and alignment graphics were performed using Jalview
1239 v2.11 [103]. The alignment was coloured according to residue identity and conservation
1240 scores calculated according to [104]. We analyzed the conservation of residues making up

1241 the E2F and LxCxE motif binding sites according to previously reported contacts [40,41].
1242 Results are presented in **Supplementary Fig. 1**.

1243

1244 **Structural Modeling**

1245 The structures of the human Rb (RbAB pocket domain) bound to E1A (PDB: 2R7G) and
1246 E7 (PDB: 1GUX), and the structure from the human paralogue p107 pocket domain (PDB:
1247 4YOZ) were collected from the protein data bank. Structural modeling of the human
1248 paralogue p130, and the retinoblastoma pocket domains from Macaque (*Macaca mulatta*),
1249 Chimpanzee (*Pan troglodytes*), Dog (*Canis lupus familiaris*), Microbat (*Myotis lucifugus*),
1250 Sheep (*Ovis aries*), Pig (*Sus scrofa*), Cow (*Bos taurus*), Horse (*Equus caballus*) and Tree
1251 shrew (*Tupaia belangeri*) were obtained by using AlphaFold v2.0 [105] implemented in
1252 ColabFold v1.0 [106]. The template multiple sequence alignments were generated using
1253 MMseqs2 [107] implemented within ColabFold v1.0. Template information and the predicted
1254 structure relaxation using amber force fields, were included. The distance between the E2F
1255 and LxCxE binding sites were measured between the alpha carbons of the C-terminal
1256 anchor site of the E2F cleft and the N-terminal anchor site of the LxCxE cleft. Molecular
1257 graphics and analyses were performed with UCSF Chimera v1.5 [108]. Results are
1258 presented in **Extended Data Fig. 9**.

1259

1260 **Data availability**

1261 SAXS raw data for Rb, E1A_{WT} and the [E1A_{WT}:Rb] complex has been deposited in SASDB
1262 (<https://www.sasbdb.org>) with codes SASDNK6 (Rb 1mg/ml), SASDNL6 (Rb 2mg/ml),
1263 SASDNM6 (Rb 4mg/ml), SASDNN6 (E1A_{WT} 4.2mg/ml), SASDNP6 (E1A_{WT} 5.6mg/ml),
1264 SASDNQ6 (E1A_{WT} 7.0mg/ml), SASDNR6 ([E1A_{WT}:Rb] 0.7mg/ml), SASDNS6 ([E1A_{WT}:Rb]
1265 1.4mg/ml), SASDNT6 ([E1A_{WT}:Rb] 2.7mg/ml), SASDNU6 ([E1A_{WT}:Rb] merged data),
1266 SASDNV6 (E1A_{WT}, SEC-SAXS). Refined conformational ensemble models for E1A_{WT} and

1267 [E1A_{WT}:Rb] have been deposited in the Protein Ensemble Database
1268 (<https://proteinensemble.org/P03255>) with codes PED00175 (E1A_{WT}) and PED00174
1269 ([E1A_{WT}:Rb]). Unfiltered conformational ensembles for the [E1A_{WT}:Rb], [E1A_{ΔL}:Rb] and
1270 [E1A_{ΔE}:Rb] complexes are available at (<https://moma.laas.fr/data/>) under the description
1271 “Conformational ensemble models of the IDP E1A bound to Rb protein”. NMR assignments
1272 of backbone resonances for E1A_{WT}, E1A_{ΔE} and E1A_{ΔL} are provided in Supplementary Data
1273 File 2. Trajectories for all E1A linker ensembles are provided at: Zenodo
1274 (<https://zenodo.org/record/6332925>), and trajectory analysis results are provided at:
1275 <https://github.com/holehouse->
1276 [lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021](https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021). PDB codes used in data
1277 analysis and prediction are: PDB:1GUX, PDB:3POM, PDB:2R7G, PDB:4YOZ. Raw data
1278 underlying Main Figures 1-5 and Extended Data Figures 1, 4, 7 and 10 are available as
1279 Source Data Files.

1280

1281 **Code Availability**

1282 The loop sampling method used to model the linker between the two binding motifs can be
1283 used via a web server (<https://moma.laas.fr/applications/LoopSampler/>), and binaries can
1284 be provided upon request. All code used to analyze the E1A linker trajectories are provided
1285 at [https://github.com/holehouse-](https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021)
1286 [lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021](https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021).

1287

1288 **REFERENCES**

1289

- 1290 67. Chemes, L. B., Noval, M. G., Sanchez, I. E. & de Prat-Gay, G. Folding of a cyclin box:
1291 linking multitarget binding to marginal stability, oligomerization, and aggregation of the
1292 retinoblastoma tumor suppressor AB pocket domain. *J Biol Chem* **288**, 18923–18938
1293 (2013).
- 1294 68. Uversky, V. N. What does it mean to be natively unfolded? *Eur J Biochem* **269**, 2–12
1295 (2002).
- 1296 69. Hofmann, H. *et al.* Polymer scaling laws of unfolded and intrinsically disordered proteins
1297 quantified with single-molecule spectroscopy. *Proc Natl Acad Sci U S A* **109**, 16155–
1298 16160 (2012).
- 1299 70. Kuzmic, P., Moss, M. L., Kofron, J. L. & Rich, D. H. Fluorescence displacement method
1300 for the determination of receptor-ligand binding constants. *Anal Biochem* **205**, 65–69
1301 (1992).
- 1302 71. Muhandiram, D. R. & Kay, L. E. Gradient-Enhanced Triple-Resonance Three-
1303 Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic*
1304 *Resonance, Series B* **103**, 203–216 (1994).
- 1305 72. Wittekind, M. & Mueller, L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate
1306 Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in
1307 Proteins. *Journal of Magnetic Resonance, Series B* **101**, 201–205 (1993).
- 1308 73. Johnson R.A., B. A. R. ; B. NMRView: a computer program for the visualization and
1309 analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
- 1310 74. Tamiola, K., Acar, B. & Mulder, F. A. Sequence-specific random coil chemical shifts of
1311 intrinsically disordered proteins. *J Am Chem Soc* **132**, 18000–18003 (2010).

- 1312 75. Blanchet, C. E. *et al.* Versatile sample environments and automation for biological
1313 solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl*
1314 *Crystallogr* **48**, 431–443 (2015).
- 1315 76. Girardot, R., Viguier, G., Pérez, J. & Ounsy, M. M. FOXTROT: A JAVA-based application
1316 to reduce and analyse SAXS and WAXS piles of 2D data at synchrotron SOLEIL,
1317 Synchrotron Soleil, Saint-Aubin, France, canSAS-VIII, Apr. 14–16, J-PARC, Tokai,
1318 Japan. (2015).
- 1319 77. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle
1320 scattering from macromolecular solutions. *J Appl Crystallogr* **50**, 1212–1225 (2017).
- 1321 78. Guinier, A. Diffraction of x-rays of very small angles-application to the study of
1322 ultramicroscopic phenomenon. *Ann. Phys* **12**, 161–237 (1939).
- 1323 79. Svergun Semenyuk, A. V.; Feigin, L. A., D. I. Small-angle-scattering-data treatment by the
1324 regularization method. *Acta Crystallogr. Sect. A Found. Crystallogr.* **44**, 244–250 (1988).
- 1325 80. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of
1326 chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–
1327 1946 (2018).
- 1328 81. Balog, E. R., Burke, J. R., Hura, G. L. & Rubin, S. M. Crystal structure of the unliganded
1329 retinoblastoma protein pocket domain. *Proteins* **79**, 2010–2014 (2011).
- 1330 82. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile
1331 computation and its assessment by contrast variation experiments. *Biophys J* **105**, 962–
1332 974 (2013).
- 1333 83. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and
1334 MultiFoXS: Single-state and multi-state structural modeling of proteins and their
1335 complexes based on SAXS profiles. *Nucleic Acids Res* **44**, W424-9 (2016).
- 1336 84. Weinkam, P., Pons, J. & Sali, A. Structure-based model of allostery predicts coupling
1337 between distant sites. *Proc Natl Acad Sci U S A* **109**, 4875–4880 (2012).

- 1338 85. Tria, G., Mertens, H. D., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of
1339 flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).
- 1340 86. Svergun Barberato, C.; Koch, M. H. J., D. CRY SOL – a Program to Evaluate X-ray Solution
1341 Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr* **28**,
1342 768–773 (1995).
- 1343 87. Garcia De La Torre, J., Huertas, M. L. & Carrasco, B. Calculation of hydrodynamic
1344 properties of globular proteins from their atomic-level structure. *Biophys J* **78**, 719–730
1345 (2000).
- 1346 88. Ortega, A., Amoros, D. & Garcia de la Torre, J. Prediction of hydrodynamic and other
1347 solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**,
1348 892–898 (2011).
- 1349 89. Vitalis, A. & Pappu, R. V. ABSINTH: a new continuum solvation model for simulations of
1350 polypeptides in aqueous solutions. *J Comput Chem* **30**, 673–699 (2009).
- 1351 90. Vitalis, A. & Pappu, R. V. Methods for Monte Carlo simulations of biomacromolecules.
1352 *Annu Rep Comput Chem* **5**, 49–76 (2009).
- 1353 91. Kozlov, A. G. *et al.* Intrinsically disordered C-terminal tails of E. coli single-stranded DNA
1354 binding protein regulate cooperative binding to single-stranded DNA. *J Mol Biol* **427**, 763–
1355 774 (2015).
- 1356 92. Metskas, L. A. & Rhoades, E. Conformation and Dynamics of the Troponin I C-Terminal
1357 Domain: Combining Single-Molecule and Computational Approaches for a Disordered
1358 Protein Region. *J Am Chem Soc* **137**, 11962–11969 (2015).
- 1359 93. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular
1360 Dynamics Trajectories. *Biophys J* **109**, 1528–1532 (2015).
- 1361 94. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. & Pappu, R. V. CIDER:
1362 Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered
1363 Proteins. *Biophys J* **112**, 16–21 (2017).

- 1364 95. Nygaard, M., Kragelund, B. B., Papaleo, E. & Lindorff-Larsen, K. An Efficient Method for
1365 Estimating the Hydrodynamic Radius of Disordered Protein Conformations. *Biophys J*
1366 **113**, 550–557 (2017).
- 1367 96. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res* **33**,
1368 W382-8 (2005).
- 1369 97. London, N., Raveh, B., Cohen, E., Fathi, G. & Schueler-Furman, O. Rosetta
1370 FlexPepDock web server--high resolution modeling of peptide-protein interactions.
1371 *Nucleic Acids Res* **39**, W249-53 (2011).
- 1372 98. Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. (Springer-Verlag
1373 New York, 2005). doi:10.1007/b138696.
- 1374 99. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: a Practical and
1375 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
1376 *(Methodological)* **57**, 289–300 (1995).
- 1377 100. Mészáros, B., Erdos, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of
1378 protein disorder as a function of redox state and protein binding. *Nucleic Acids Res* **46**,
1379 W329–W337 (2018).
- 1380 101. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res* **49**, D884–D891 (2021).
- 1381 102. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize
1382 evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-50 (2016).
- 1383 103. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview
1384 Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*
1385 **25**, 1189–1191 (2009).
- 1386 104. Livingstone, C. D. & Barton, G. J. Protein sequence alignments: a strategy for the
1387 hierarchical analysis of residue conservation. *Comput Appl Biosci* **9**, 745–756 (1993).
- 1388 105. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
1389 583–589 (2021).

- 1390 106. Mirdita, M. *et al.* ColabFold - Making protein folding accessible to all. *bioRxiv*
1391 2021.08.15.456425 (2021) doi:10.1101/2021.08.15.456425.
- 1392 107. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for
1393 fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
- 1394 108. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and
1395 analysis. *J Comput Chem* **25**, 1605–1612 (2004).
- 1396