nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-022-00811-w

Check for updates

Conformational buffering underlies functional selection in intrinsically disordered protein regions

Nicolás S. González-Foutel[1,2,11], Juliana Glavina[1,3,11], Wade M. Borcherds[4], Matías Safranchik[1], Susana Barrera-Vilarmau[5,6], Amin Sagar[7], Alejandro Estaña[7,8], Amelie Barozet[8], Nicolás A. Garrone[1], Gregorio Fernandez-Ballester[9], Clara Blanes-Mira[9], Ignacio E. Sánchez[3], Gonzalo de Prat-Gay[2], Juan Cortés[8], Pau Bernadó[7], Rohit V. Pappu[5] ✉, Alex S. Holehouse[5,10] ✉, Gary W. Daughdrill[4] ✉ and Lucía B. Chemes[1,2] ✉

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

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

Tethering is essential for kinase signaling[8,14,15], gene silencing[7], enzyme catalysis[16], transcriptional regulation[13,17,18] and the formation of biomolecular condensates[9,19]. Tethering allows intra- or intermolecular coupling between ordered domains and/or SLiMs[20]. This coupling can increase the effective concentrations of interacting partners[21], and relatively simple polymer models, such as the worm-like chain (WLC)[22–24], can estimate the affinity enhancement from tethering[12,22,25,26]. An emerging hypothesis is that SERs that encode IDR dimensions—as determined by sequence length, composition and patterning[5,27–30]—play an important role in tethered interactions by determining the effective concentrations of binding modules around binding partners either in *cis* or in *trans*[12,13,16,31–33]. This leads to the expectation that evolutionary pressure will preserve these dimensions in spite of large-scale sequence variation. However, the lack of a well-defined model system in which molecular function is unambiguously conserved in the face of a hypervariable tether has hampered the ability to test this hypothesis.

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

[1]Instituto de Investigaciones Biotecnológicas (IIBiO-CONICET), Universidad Nacional de San Martín, Buenos Aires, Argentina. [2]Fundación Instituto Leloir e Instituto de Investigaciones Bioquímicas (IIB-CONICET), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina. [3]Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN-CONICET), Universidad de Buenos Aires, Buenos Aires, Argentina. [4]Department of Cell Biology, Microbiology, and Molecular Biology and, University of South Florida, Tampa, FL, USA. [5]Department of Biomedical Engineering, Center for Science & Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO, USA. [6]Instituto de Química Avanzada de Cataluña (IQAC-CSIC), Barcelona, Spain. [7]Centre de Biologie Structurale (CBS), Université de Montpellier, INSERM, CNRS, Montpellier, France. [8]LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France. [9]Instituto de Investigación, Desarrollo e Innovación en Biotecnología Sanitaria de Elche (IDiBE), Universidad Miguel Hernández, Elche, Alicante, Spain. [10]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, USA. [11]These authors contributed equally: Nicolas S. Gonzalez-Foutel, Juliana Glavina. ✉e-mail: pappu@wustl.edu; alex.holehouse@wustl.edu; gdaughdrill@usf.edu; lchemes@iib.unsam.edu.ar

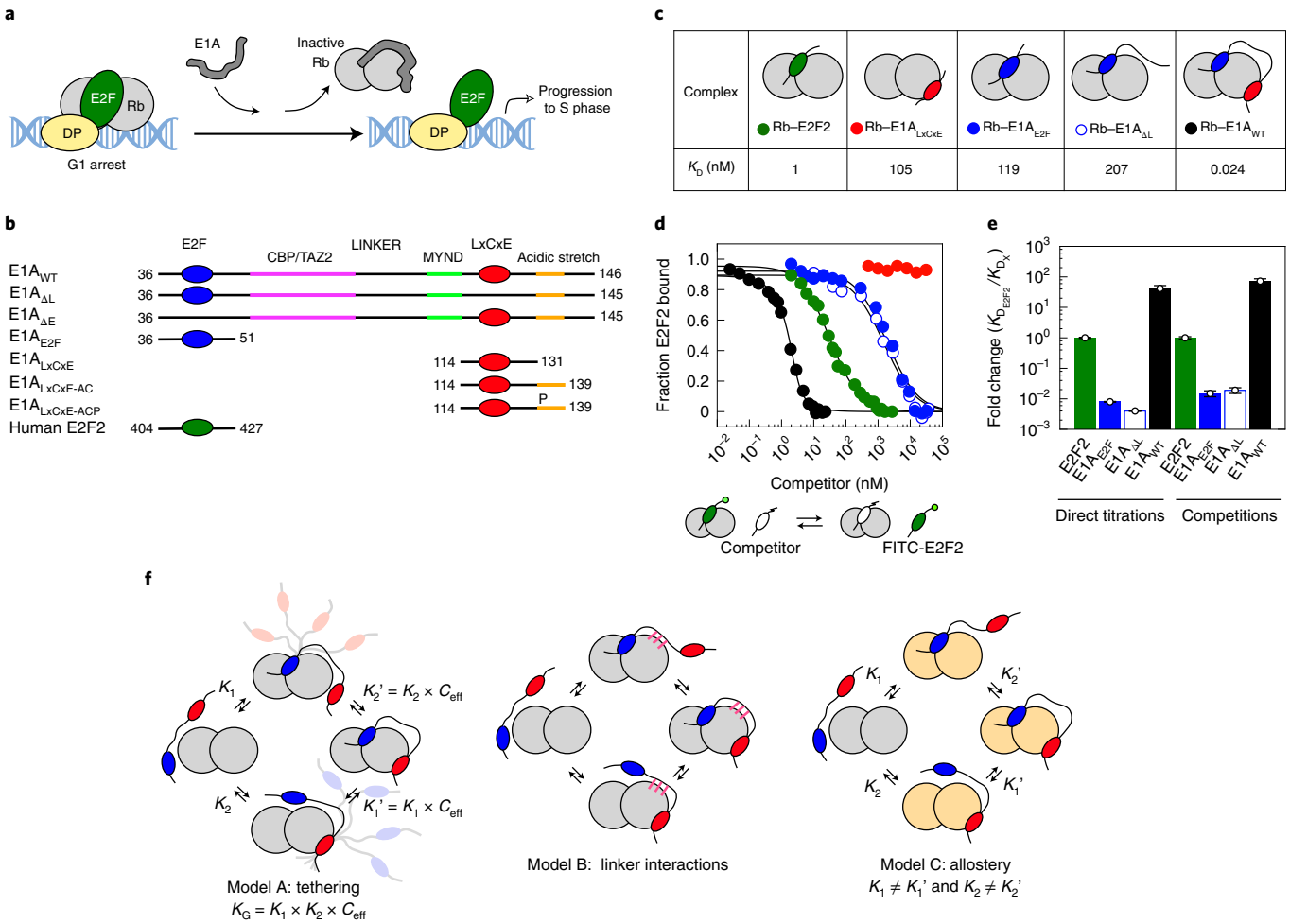


Fig. 1 | Tethering is required for high-affinity Rb binding and E2F displacement by E1A. a, Model for disruption of the repressive Rb–E2F complex by E1A. The functional E2F/DP transcriptional complex is depicted. E2F/DP is a heterodimer composed of one E2F transcripion factor and one DP transcription factor. **b**, Schematic representation of the E1A and E2F2 constructs used in this study. Color coding for the E2F, LxCxE, TAZ2 and MYND SLiMs, the acidic stretch and S132 phosphorylation are maintained throughout the figures. **c**, Representative interactions tested using fluorescence spectroscopy (Extended Data Fig. 3 and Supplementary Tables 1 and 3). Representative interactions tested using fluorescence spectroscopy. The KD values were measured by global fitting of several binding isotherms: (Rb-E2F2: $n = 5$, Rb-E1A$_{LxCxE}$: $n = 5$, Rb-E1A$_{E2F}$: $n = 6$, Rb-E1A$_{\Delta L}$: $n = 3$, Rb-E1A$_{WT}$: $n = 11$). **d**, E2F competition titrations. Color code is as in **c**. **e**, Comparison of the fold-change in binding affinity from direct titrations versus competition assays. The height of the bar is obtained by dividing the K_D of E2F2 by each K_D ($n=1$), and values higher than one ($10^0 = 1$) indicate an increase in binding affinity with respect to E2F2. For direct titrations, each K_D value was obtained by averaging (global fitting) over several independent binding isotherms (E2F2: $n=5$, E1A$_{E2F}$: $n=3$, E1A$_{\Delta L}$: $n=3$, E1A$_{WT}$: $n=3$) containing 16–22 points each (see source data). For competition experiments, each K_D was obtained by fitting of a single binding isotherm ($n=1$). Error bars correspond to the propagated s.d. of the averaged K_D values. **f**, Three models that account for affinity enhancement in the motif–linker–motif E1A arrangement (see main text for details). K_G = global equilibrium association constant for the formation of the E1A$_{WT}$-Rb complex, K_1 = equilibrium association constant for the formation of the E1A$_{E2F}$-Rb complex, K_2 = equilibrium association constant for the formation of the E1A$_{LxCxE}$-Rb complex, K_1' = equilibrium association constant for the formation of the E1A$_{E2F}$-Rb complex once the E1A$_{LxCxE}$ motif is bound, K_2' = equilibrium association constant for the formation of the E1A$_{LxCxE}$-Rb complex once the E1A$_{E2F}$ motif is bound.

Tethering enhances Rb binding and promotes E2F displacement

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

To assess the affinity of E1A$_{WT}$ and the relative contributions of the two motifs, we also tested E1A constructs comprising the individual SLiMs or fragments in which the E2F (E1A$_{\Delta E}$) or LxCxE (E1A$_{\Delta L}$) motifs were mutated to poly-alanine (Extended Data Fig. 1 and Fig. 1b) and the E2F SLiM (E2F2) taken from the host transcription factor E2F2 (Fig. 1b). Isothermal titration calorimetry (ITC) (Extended Data Fig. 2 and Supplementary Table 1) and size-exclusion chromatography coupled to static light scattering (SEC–SLS) experiments (Supplementary Table 2) confirmed that all E1A constructs bound to Rb with 1:1 stoichiometry. To quantify

binding affinities, we performed fluorescence polarization measurements using fluorescein 5-isothiocyanate (FITC)-labeled constructs (Extended Data Fig. 3 and Supplementary Tables 1 and 3). Although the host-derived E2F2 SLiM bound to Rb with high affinity (dissociation constant (K_D) = 1 nM), the $E1A_{E2F}$ SLiM had a K_D of 119 nM, suggesting that it would be a weak competitor of E2F2 (Fig. 1c). On the basis of a previous study showing that the E2F and LxCxE motifs bind simultaneously to Rb[45], we suspected that a protein containing both SLIMs and the linker ($E1A_{WT}$) would increase the binding affinity of the individual SLiMs by enhancing local concentrations. In support of this, we found that $E1A_{WT}$ had picomolar binding affinity (K_D = 24 pM), conferring a 4,000-fold enhancement compared with the individual E1A SLiMs and a 40-fold enhancement compared with E2F2, consistent with a role of tethering in affinity enhancement (Fig. 1c).

To further test the role of tethering in E2F displacement, we carried out competition assays. Synthetic peptides corresponding to the $E1A_{LxCxE}$ and $E1A_{E2F}$ motifs as well as the $E1A_{\Delta L}$ mutant, where the LxCxE motif is mutated to poly-alanine, were unable to effectively displace E2F from Rb (Fig. 1d). However, $E1A_{WT}$ was a strong competitor, disrupting the E2F2–Rb complex at a low nanomolar concentration (Fig. 1d). The agreement among results from ITC, direct titration and competition experiments confirmed that tethering was required for high-affinity Rb binding and E2F displacement (Fig. 1e and Supplementary Table 1).

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

Linker-mediated interactions do not stabilize binding to Rb

We first sought to confirm the disordered nature of $E1A_{WT}$ using nuclear magnetic resonance (NMR) spectroscopy. The transverse optimized relaxation (TROSY) spectrum of $[^{15}N]E1A_{WT}$ revealed narrow chemical-shift dispersion in the ^1H dimension. This is a characteristic signature of disordered regions and is consistent with previous work on E1A fragments (Fig. 2a)[43,46,47]. Further, the $^{13}C_\alpha$ secondary chemical shifts ($\Delta\delta C_\alpha$) showed minimal deviation from random-coil values obtained from disordered proteins (Fig. 2b(i)), and negative ^1H–^{15}N nuclear Overhauser effect (NHNOE) values observed for $E1A_{WT}$ were indicative of fast backbone dynamics (Fig. 2b(ii)). Finally, sequence analysis predicted that $E1A_{WT}$ is globally disordered (Fig. 2b(iv)). These results confirmed that the conformational ensemble of $E1A_{WT}$ is characterized by high heterogeneity (disorder), with fast interconversion between distinct conformations on the nanosecond to picosecond timescale (flexibility).

Next, we used NMR spectroscopy to determine the structural basis for $E1A_{WT}$ binding to Rb. For this, we dissected the relative contributions of the SLiMs, their flanking regions and the linker. Previous NMR work has mapped binding of E1A fragments containing individual Rb binding motifs[43], but it did not examine the structural details of E1A bivalently tethered to Rb. The TROSY spectrum of labeled $E1A_{WT}$ in complex with unlabeled Rb (molecular weight (MW) = 54.6 kDa, Supplementary Table 2) revealed a complete loss of peaks for residues corresponding to the E2F and LxCxE SLiMs (L43 to Y47 and L122 to E126) (Figs. 2a,b(iii) and Extended Data Fig. 4), consistent with the slow exchange expected from the high affinities of the untethered motifs and the tethered complex[48]. The $E1A_{\Delta L}$ and $E1A_{\Delta E}$ constructs containing a single wild type motif retained binding to Rb, consistent with independent binding of each motif to Rb. On the basis of previous reports[49], we anticipated that the regions flanking the canonical $E1A_{E2F}$ or $E1A_{LxCxE}$ motifs contribute stabilizing interactions to the complex. In agreement with this expectation, the peaks corresponding to the flanking residues (E39 to T52 and V119 to E135) disappeared upon binding, yielding near-zero resonance-intensity ratios.

Binding experiments using fluorescence polarization and ITC confirmed the stabilizing role of the flanking regions: the affinity of the $E1A_{LxCxE}$ motif increased incrementally upon addition of the acidic stretch following the motif ($E1A_{LxCxE\text{-}AC}$) and with Ser132 phosphorylation ($E1A_{LxCxE\text{-}ACP}$) (Fig. 1b and Supplementary Table 1), leading to an overall fivefold increase in binding affinity. ITC revealed different origins for thermodynamic stability in each core motif and the flanking regions (Supplementary Table 1). Although binding of $E1A_{E2F}$ is entropically driven, suggesting that complex stabilization is dominated by the desolvation of apolar surfaces, binding of $E1A_{LxCxE}$ is enthalpically driven, likely owing to the contribution from hydrogen bonds between the LxCxE motif and Rb.

The amino-terminal (N-terminal) linker region (residues 50–85) encompassing the TAZ2-binding motif is highly conserved and has a lower propensity towards disorder as assessed by the disorder predictor IUPred, likely owing to its hydrophobic nature (Fig. 2b(iv,v)). This region showed a decrease in peak intensities (Fig. 2b(iii)) that a previous report suggested was due to weak interactions with Rb[43]. Consistent with this observation, the N-terminal linker region does not show increased chemical-shift dispersion or large chemical-shift changes upon binding to Rb (Fig. 2a and Extended Data Fig. 4a–d). The carboxy-terminal linker region (residues 86–120) showed no changes in chemical shifts or resonance intensities, indicating that this region remains globally disordered and flexible when bound to Rb. These interpretations are supported by the lack of change in secondary structure upon binding Rb, as measured by circular dichroism (CD) (Fig. 2c).

Additional ITC studies using an isolated fragment from the N-terminal linker region that encompasses the region with the largest decrease in peak intensities ($E1A_{60-83}$) did not show any detectable association to Rb (Extended Data Fig. 2(i)). Further, E1A constructs that include the linker did not show higher binding affinities than those of isolated E1A motifs (Fig. 2d and Supplementary Table 1). Taken together, these data rule out the presence of a high-affinity binding site. To test for weak interactions that depend on bivalent tethering, we designed a construct in which the TAZ2-binding region ($_{71}$MLAVQEGID$_{79}$), which showed the largest reduction in I/I_0, was replaced by a glycine-serine stretch ($E1A_{WT\Delta Hyd}$). The binding affinity of this mutant actually increased by 1.5-fold compared with that of $E1A_{WT}$ in fluorescence competition experiments (Supplementary Table 4), revealing a weak destabilizing effect of the TAZ2-binding site.

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

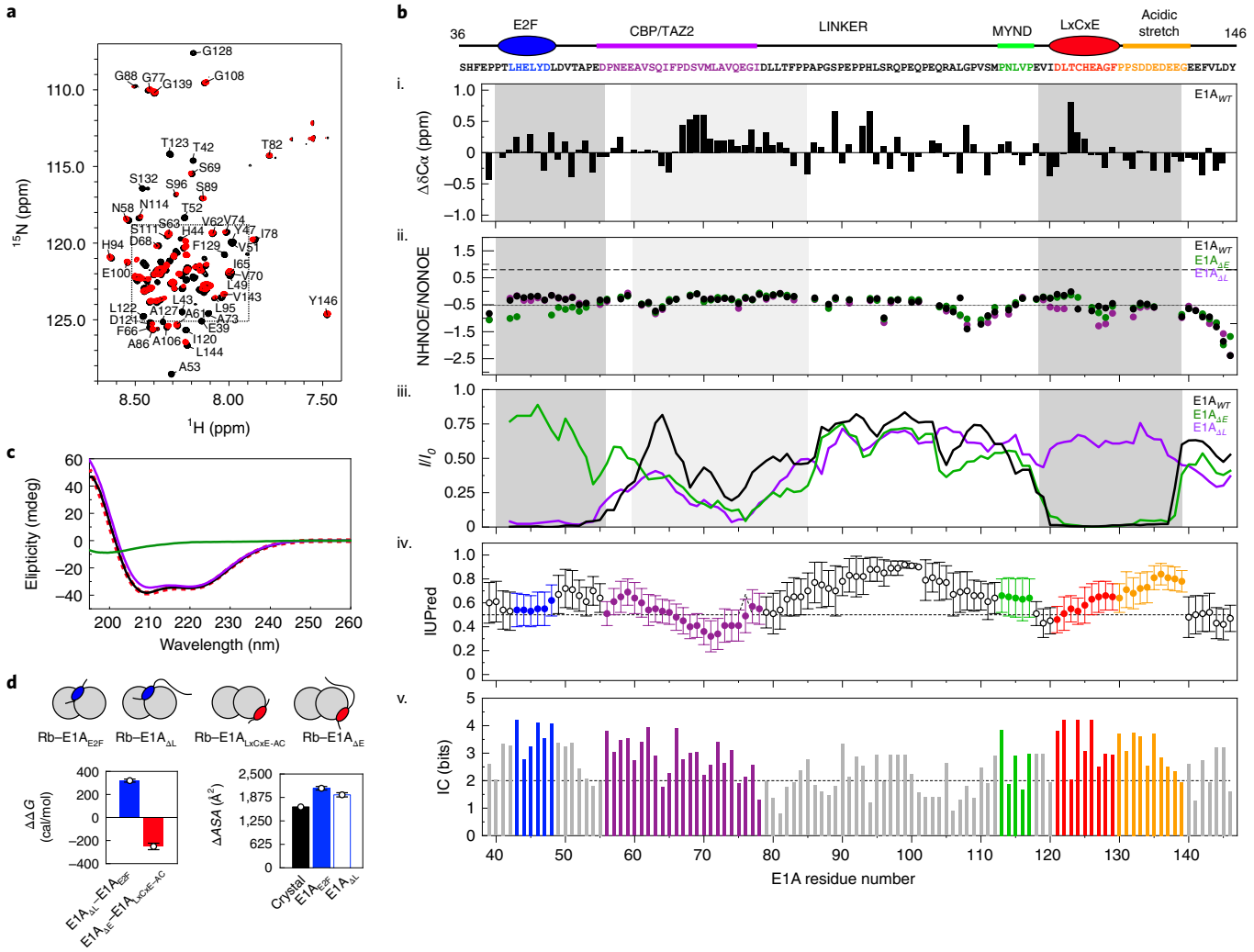


Fig. 2 | NMR and ITC analysis of the E1A$_{WT}$–Rb complex. a, ^{1}H-^{15}N TROSY spectra of free [^{15}N]E1A$_{WT}$ (black) and [^{15}N]E1A$_{WT}$ bound to unlabeled Rb (red). [^{15}N]E1A$_{WT}$ peak assignments for the inset are shown in Extended Data Fig. 4. **b**, i, ^{13}Cα secondary chemical shift (ΔδCα) of [^{15}N]E1A$_{WT}$. ii, ^{1}H-^{15}N nuclear Overhauser effect (NHNOE) was measured as the NHNOE/NONOE ratio for [^{15}N]E1A$_{WT}$, [^{15}N]E1A$_{ΔE}$ and [^{15}N]E1A$_{ΔL}$, where NONOE is a control experiment with no nuclear Overhauser effect. Dashed line: reference value for a rigid backbone. iii, Intensity ratio plots of bound state (I) with respect to the free state (I_0) for E1A$_{WT}$, E1A$_{ΔL}$ and E1A$_{ΔE}$. Dark gray: E2F and LxCxE SLiMs and flanking regions; light gray: N-terminal linker region. iv,v, Disorder propensity (iv) and residue conservation (information content, IC) (v) were predicted from an alignment of E1A sequences (n = 110) (Supplementary Data 1). For disorder prediction, data points represent the mean IUPred value at each position and error bars represent the standard deviation of the mean. The number of residues averaged at each position is variable depending on the number of gaps in the alignment. For the conservation plot, the height of each bar represents the IC value at each position. **c**, Far-UV CD spectra for E1A$_{WT}$ (green line), Rb (violet line), the E1A$_{WT}$–Rb complex (black line) and the arithmetic sum of the Rb and E1A$_{WT}$ spectra (red dashed line). The latter CD spectra largely overlap. Although it is possible the low salt concentration of the CD experiments might mask hydrophobic interactions occurring at the higher salt concentration used for NMR and other binding experiments, such effects are unlikely to prevail for the types of monovalent salts used in our binding experiments. **d**, Left, plot of the change in free energy of binding (ΔΔG) for E1A fragments containing or lacking the linker region, measured by ITC. The bar height results from the subtraction between mean ΔG values obtained by averaging several independent binding experiments: ΔG E1A$_{ΔL}$ (n = 3), ΔG E1A$_{E2F}$ (n = 1), ΔG E1A$_{ΔE}$ (n = 3) and ΔG E1A$_{LxCxE-AC}$ (n = 3) (Supplementary Table 1). Right, plot of the change in ΔASA for E1A fragments containing or lacking the linker region. The height of the bar represents the ΔASA value from PDB structure 2R7G (n = 1, black bar) or that derived from ITC experiments for E1A$_{E2F}$–Rb (n = 1, blue bar) and E1A$_{ΔL}$–Rb (n = 1, empty blue bar) (Supplementary Table 6). ΔASA was calculated by ITC measurements at several temperatures (n = 4 E1A$_{E2F}$–Rb; n = 3 E1A$_{ΔL}$–Rb). Error bars correspond to the propagated mean s.e. of the ΔASA value.

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

Allosteric coupling in Rb does not increase E1A–Rb affinity

To assess whether allosteric coupling between the E2F and LxCxE binding sites in Rb play a role in affinity enhancement (Fig. 1f, Model C), we saturated Rb with the E1A$_{E2F}$ or E1A$_{LxCxE}$ motifs and performed titrations with the complementary motif to determine the change in Gibbs free energy of binding of the saturated complex ($ΔG_{saturated}$) and compared this value to the change in Gibbs free energy of binding of the unsaturated complex ($ΔG_{unsaturated}$) (Extended Data Fig. 5). If a positive allosteric effect is at play, E1A$_{LxCxE}$ should bind more

tightly to Rb when $E1A_{E2F}$ is already bound, and vice versa. This was measured as the difference between the ΔG value of the saturated and unsaturated complexes: $\Delta\Delta G = \Delta G_{saturated} - \Delta G_{unsaturated}$, where a negative value for $\Delta\Delta G$ indicates positive cooperativity. For both motifs, the values of $\Delta\Delta G$ were in the range of ±0.25 kcal/mol (Supplementary Table 7). In $E1A_{LxCxE}$ binding assays, saturation with $E1A_{\Delta L}$ instead of $E1A_{E2F}$ did not change the outcome, indicating that neither the motif nor the motif plus linker arrangement behaved as an allosteric effector on the complementary site. Therefore, our results suggest that allosteric coupling in Rb (Fig. 1f, Model C) does not make a major contribution to affinity enhancement.

Entropic tethering optimizes affinity of E1A for Rb

Our results indicate that the positive cooperativity of the tethered $E1A_{E2F}$ and $E1A_{LXCXE}$ motifs binding to Rb results from an increase in C_{eff} of one motif once the other motif is bound[45] (Fig. 1f, Model A). It is well established that this form of cooperativity can be described using a simple WLC model[22–24,33] that treats the linker as an entropic tether (Fig. 3a,b) wherein the dimensions of the linker will determine the degree of the affinity enhancement. A short linker would be unable to straddle the distance between the two binding sites and would lead to low affinity enhancement (Fig. 3a,b(i)); an optimal linker would maximize C_{eff}, leading to maximal positive cooperativity (Fig. 3a,b(ii)), and a longer than optimal linker would decrease C_{eff} (Fig. 3a,b(iii)). Application of the WLC model to the E1A linker predicts a C_{eff} value of 0.92 mM, which is close to the optimal value (Fig. 3b) and within a factor of 2 of the C_{eff} (0.52 ± 0.09 mM), obtained from the affinities of $E1A_{WT}$ and the isolated motifs (Supplementary Table 1). For $E1A_{WT\Delta Hyd}$, where the destabilizing effect of the linker region is removed, the agreement with the WLC model improves (C_{eff} = 0.78 ± 0.24 mM), indicating that this mutated linker behaves more like an entropic tether optimized to bind Rb with near-maximal affinity.

To further test the tethering model, we performed small angle X-ray scattering (SAXS) on Rb, $E1A_{WT}$ and the $E1A_{WT}$–Rb complex (Fig. 3c and Extended Data Fig. 6). The experimental SAXS profile of the Rb domain could be fit to the theoretical SAXS profile derived from its crystal structure (Pearson's chi-squared value $\chi_i^2 = 1.3$) and further refined (root mean square deviation (RMSD) = 1.7 Å) using a SAXS-driven modeling approach ($\chi_i^2 = 0.82$) (Fig. 3c and Extended Data Fig. 6a), indicating that Rb in solution retained its folded structure. Alternatively, the Kratky plots of $E1A_{WT}$ were characteristic of an IDP. Fitting of the SAXS profiles using the ensemble optimization method (EOM)[52] indicated that $E1A_{WT}$ adopts highly expanded conformations (Extended Data Fig. 6b). To analyze the conformation of the linker in the $E1A_{WT}$–Rb complex, we applied a sampling method[53] to generate a pool of 10,250 realistic conformations[54] and computed theoretical SAXS profiles that were selected using EOM analysis. The SAXS profile of the complex was best described by sub-ensembles in which the linker sampled expanded conformations (Fig. 3c–e and Extended Data Fig. 6c) with hydrodynamic radius (R_h) values ($R_{h\ EOM}$ = 3.36 nm) in good agreement with those obtained from SEC–SLS experiments ($R_{h\ SEC}$ = 3.20 ± 0.12 nm) (Fig. 3f,g, Extended Data Fig. 6d and Supplementary Data Table 2) and radius of gyration to hydrodynamic radius (R_g/R_h) ratios consistent with bivalent tethering (Supplementary Table 2).

Our structural and thermodynamic dissection establishes E1A as a quantitative model system for entropic tethering, demonstrating that other mechanisms have a negligible contribution to affinity enhancement (with linker interactions having ΔG = ~+200 cal/mol over a total $\Delta G_{binding}$ = −14,240 cal/mol). Our ability to isolate tethering as the key determinant of binding affinity is unparalleled and provides us with a unique opportunity to test how tethering operates in biological systems. A longstanding question is whether the sequences of protein regions that encode tethering have any unique relationships with the conformational ensembles that they form. If they do not, it is difficult to imagine that the members of a family of linkers with extensive variation in sequence and length could function in the same way. The model we present below is a comprehensive assessment of how this is possible.

Hypervariable E1A linkers have a conserved functional length

Inspection of selected linker sequences that are representative of mastadenoviruses that infect a wide range of mammalian hosts (Fig. 4a) revealed that, although the N- and C-terminal acidic extensions and the aromatic/hydrophobic TAZ2-binding region were highly conserved, the linker lengths and compositions vary considerably within the central region enriched predominantly with polar, hydrophobic and proline residues (Figs. 4a and 2b(v)). To understand how function is conserved in the face of these extensive differences in linker length and sequences, we performed all-atom simulations[9] and generated conformational ensembles of 27 E1A linker sequences, with linker lengths from 27 to 75 residues (Fig. 4a). While the shortest linkers from bovine/ovine E1A proteins had smaller end-to-end distances, the average end-to-end distance of linkers 41–75 residues long remained roughly constant despite an almost doubled length (Fig. 4b). This suggested that the linkers have a conserved functional length[55] that is determined by a joint contribution of sequence length, amino acid composition and sequence patterning as determinants of end-to-end distances. To test the feasibility of this hypothesis, we performed simulations for 140 random synthetic sequences of variable length that matched the amino acid composition of one of the shortest linkers (HF_HAdV40). In sharp contrast to natural sequences, the synthetic sequences showed the expected monotonic increase in end-to-end distance with chain length ($R_{natural}$ = 0.37, R_{synt} = 0.99; Extended Data Fig. 7a). To examine the sequence features that underlie the conservation of the functional length, we analyzed various statistical properties (Extended Data Fig. 7b,c). Net charge per residue (NCPR) had the strongest positive correlation with normalized end-to-end distance, with more expanded chains having a higher NCPR (Extended Data Fig. 7b). This is in agreement with previous findings that net charge and patterning are major determinants of IDR dimensions in natural[11,13,29,30,55] and synthetic[27,30,32] sequences. Longer chains also tend to have higher proline content, with fewer hydrophobic and charged residues (Extended Data Fig. 7c).

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

Conformational buffering preserves optimal tethering

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

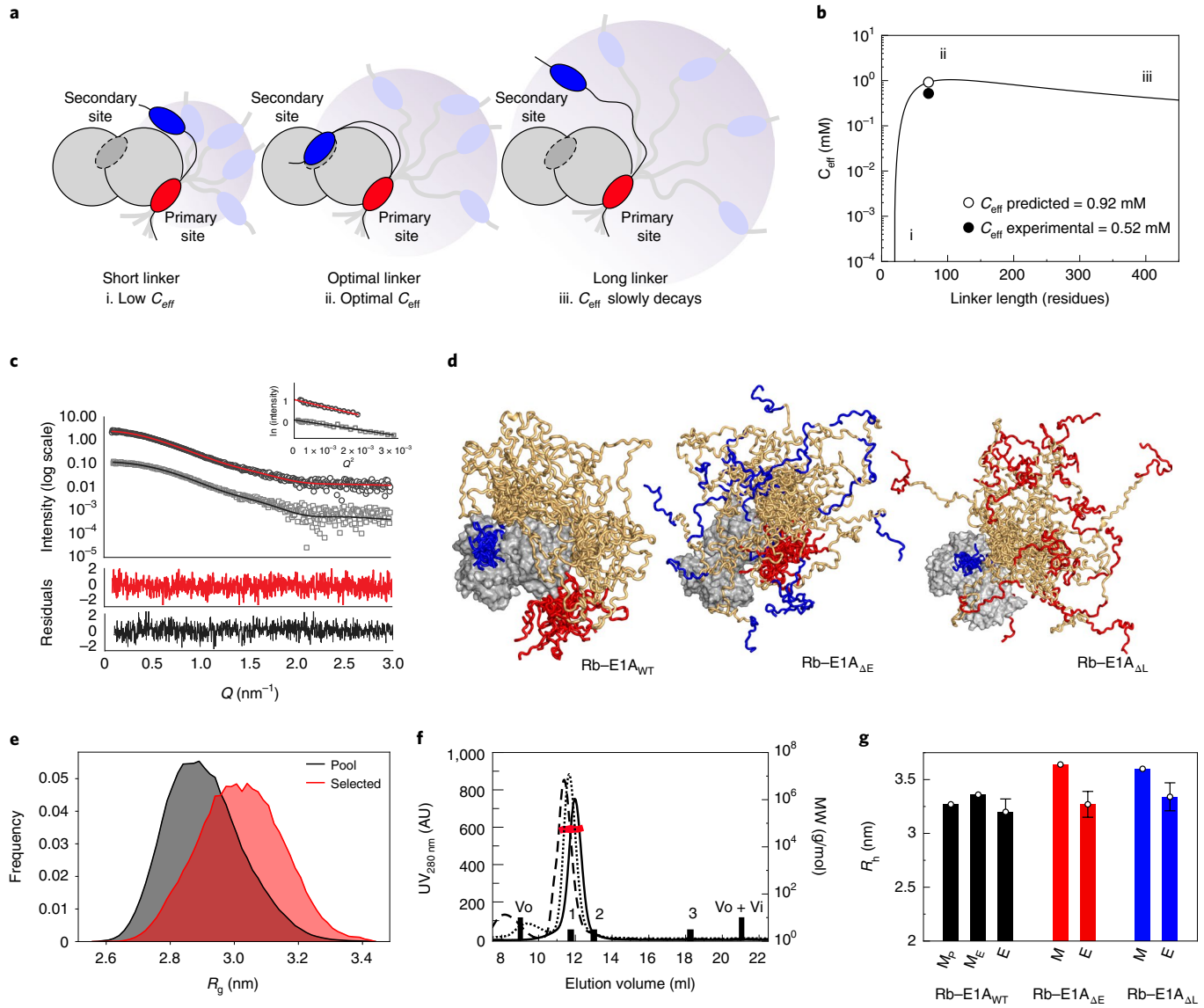


Fig. 3 | The E1A linker behaves as an entropic tether. a, Schematic representation of how C_{eff} depends on linker length. **b**, C_{eff} curve from the WLC model. The scenarios depicted in **a** are shown as regions (i, ii, iii). **c**, SAXS intensity profile of: Rb (gray squares) with best fit to the theoretical profile derived from the Rb crystal structure (RbAB domain, black line); and the E1A$_{WT}$–Rb complex (black circles) with best fit from the EOM method (red line). Inset: Guinier plots for Rb and E1A$_{WT}$–Rb. Q^2 is the square of momentum transfer (in nm^{-2}). **d**, SAXS-selected E1A$_{WT}$–Rb EOM ensemble (both motifs bound) and simulated ensembles for Rb–E1A$_{\Delta E}$ and Rb–E1A$_{\Delta L}$ (one motif bound). **e**, R_g distribution of the ensemble pool for E1A$_{WT}$–Rb (black) and the EOM ensemble (red). The linker samples conformations more expanded than the random-coil model of the pool. **f**, SEC–SLS of E1A$_{WT}$–Rb (solid line), E1A$_{\Delta E}$–Rb (dotted line) and E1A$_{\Delta L}$–Rb (dashed line). Black bars indicate the Vo and Vo+Vi, and the numbered bars indicate the elution volume of standards: Bovine serum albumin (BSA) 66 kDa (1), MBP 45 kDa (2) and lysozyme 14.3 kDa (3). Black line: SEC profile, Red line: MW value (g/mol). **g**, Comparison between the hydrodynamic radius (R_h) of modeled (M_P = pool, M_E = EOM) and experimental (E) ensembles for E1A$_{WT}$–Rb (black bars), E1A$_{\Delta E}$–Rb (red bars) and E1A$_{\Delta L}$–Rb (blue bars). The height of each bar represents the R_h value. Modeled R_h values (n = 1) have no associated error. For experimental R_h values (n = 1), error bars represent the propagated error obtained from estimation of the R_h parameter (see Methods).

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

We predict that optimal tethering depends mainly on the linker dimensions and variants with conserved end-to-end distances will confer similar affinity to E1A$_{WT}$. In accordance with this prediction, human and simian E1A linkers ranging from 41 to 75 residues have similar binding affinities, with $K_D/K_{D,E1AWT}$ ratios between 0.4 and 1.2 (Fig. 4c,d). E1A linkers with the highest affinities (Hum-2 and Sim-1) had polar residues interrupting the weakly destabilizing hydrophobic interactions in the TAZ2 SLiM (Fig. 4a). To directly assess linker dimensions, we used SEC to measure R_h for selected E1A variants after MBP cleavage (Extended Data Figs. 1f,g and 7d). Both K_D and R_h agreed closely with those predicted from the atomistic simulations (Fig. 4d, Supplementary Table 4 and Extended Data Fig. 7d). We also created a tandem repeat in which the Hum-2 linker was duplicated (Hum-2-2x, Fig. 4c,d).

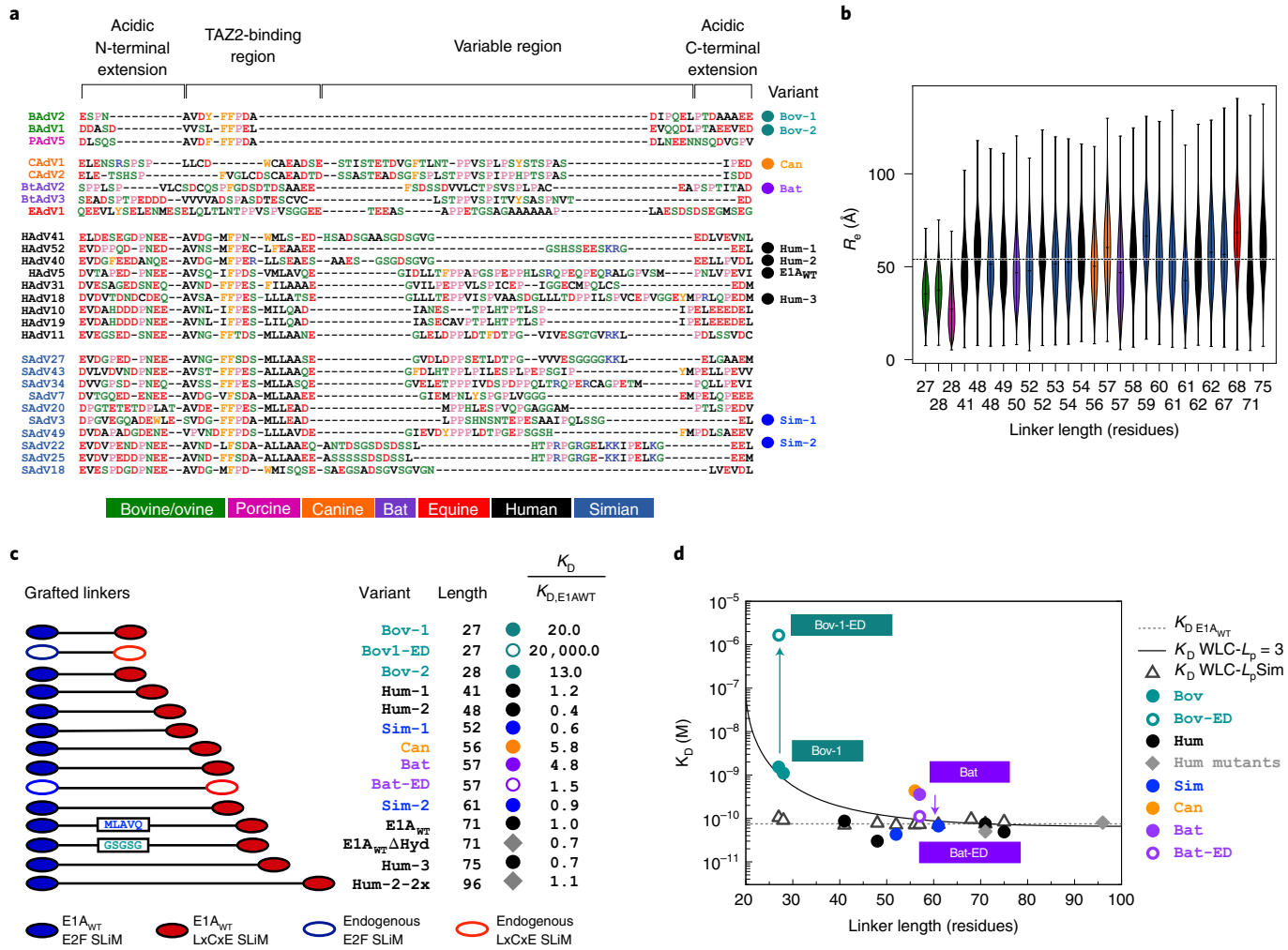


Fig. 4 | Conformational buffering leads to conserved functionality of E1A proteins. a, Global alignment of 27 selected E1A linker sequences. *Mastadenovirus* types are indicated on the left and the color coding (bottom) indicates the host range. The variants used for the design of chimeras are shown to the right, with three letter codes indicating the host range. Amino acids color code: acidic (red), basic (blue), polar (green), hydrophobic (black), aromatic (orange) and proline (pink). **b**, End-to-end distance calculated from all-atom simulations using the set of E1A linkers from **a**. Violin plots are colored by host range, as in **a**. For each sequence, $n = 15$ independent simulations were run (see Methods). The horizontal line within each violin plot represents the median end-to-end distance (R_e) value, and the ends of the whiskers indicate the maximal and minimal values. Horizontal dotted line: mean R_e value (53.39 Å) obtained by averaging the median R_e values of all sequences excluding Bov-1, Bov-2 and Porcine. **c**, Motif–linker–motif constructs used in the E1A linker grafting experiment. Filled circles: grafting of linkers into the HAdV5 E1A$_{E2F}$ and E1A$_{LxCxE}$ motifs. Diamonds: mutant in which the hydrophobic MLAVQEGID region was replaced by a GS stretch (E1A$_{WT\Delta Hyd}$) or in which the HAdV40 linker sequence was duplicated (Hum-2-2x). Empty circles: variants harboring endogenous linker and motifs (ED). **d**, Global K_D as a function of linker length for the motif–linker–motif constructs. K_D for each variant was measured using an E2F displacement experiment (symbols as in **c**) or predicted using the WLC model. The K_D values ± errors for all measurements are reported in Supplementary Table 4. The predicted value of the K_D for the grafted linkers was calculated as $K_D = (K_{D,E2F} \times K_{D,LxCxE}) / C_{eff}$ (see Methods) using the known affinity of the E1A$_{E2F}$ and E1A$_{LxCxE}$ motifs from E1A$_{WT}$ (Supplementary Table 1) and C_{eff} values obtained using a sequence-independent (straight line: WLC-$L_p = 3$) or sequence-dependent (empty triangles: WLC-L_pSim) persistence length (L_p) parameter (see Extended Data Fig. 10 and Methods). Dotted line, Experimental K_D value of the E1A$_{WT}$ construct (75 ± 17 pM). Under the sequence-independent WLC model, the K_D is expected to increase gradually with decreasing linker length, while L_pSim predicts the K_D to remain constant, in the 41–75 linker length range. Experimental K_D values are in good agreement with both models for longer linker lengths, but are closer to L_pSim for shorter linker lengths (41, 48 and 52).

$K_{D,Hum\text{-}2\text{-}2x}/K_{D,E1AWT}$ was 1.1, suggesting that its dimensions are still optimized.

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

Linker–motif coevolution modulates conformational buffering

The shorter bovine linkers (Fig. 4a) had $K_D/K_{D,E1AWT}$ ratios between 13 and 20 (Fig. 4c,d). This weaker-than-predicted affinity (Fig. 4d) was not due to the linkers being less expanded, because the predicted and experimental R_h values for Bov-1 were similar (Extended Data Fig. 7d). Instead, it suggests that a minimal sequence length, not predicted by the WLC, is necessary to overcome entropic effects required for proper orientation of the SLiMs to bind Rb. This highlights a limitation of the WLC model, which was not unexpected

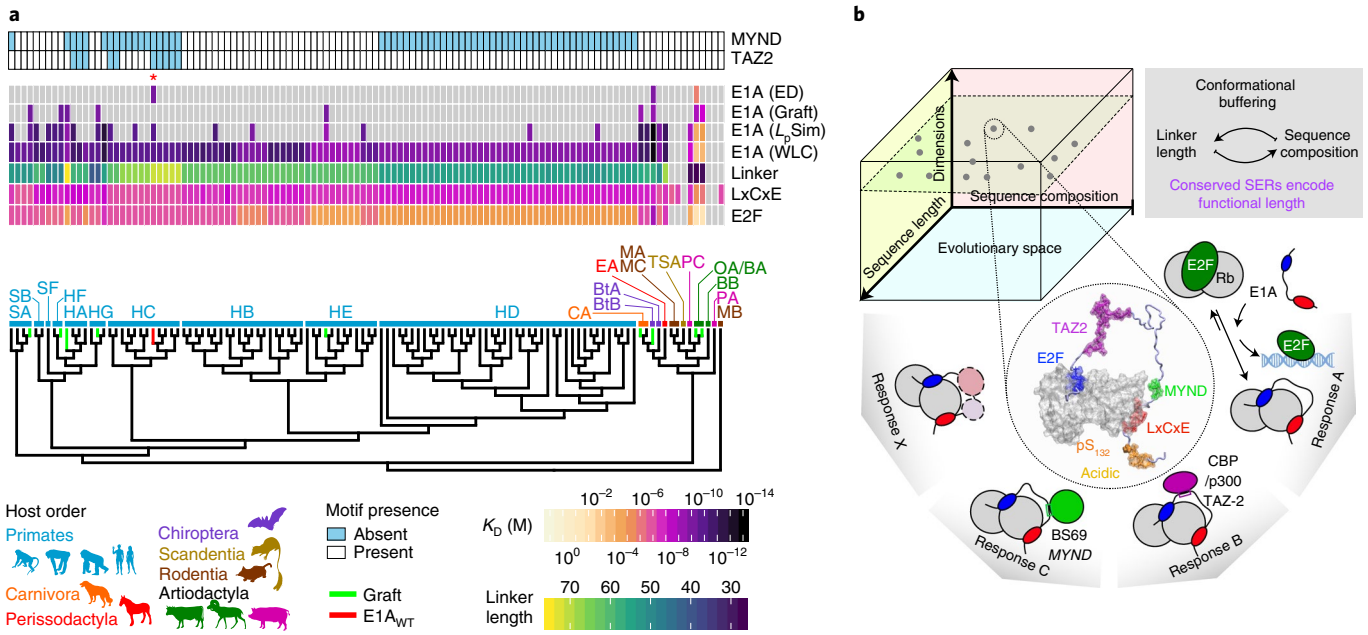


Fig. 5 | Evolutionary conservation of tethering by E1A proteins. a, Phylogenetic tree of mastadenovirus E1A proteins with species denoted by two-letter codes. The affinity of the E2F and LxCxE SLiMs and E1A$_{WT}$, and linker length are indicated by color scales (unit, number of amino acids or number of residues). E1A (WLC): global K_D for E1A proteins predicted by the WLC model, with standard L_p values (L_p = 3); E1A (L_pSim): K_D for E1A proteins predicted by the WLC model with sequence-dependent L_p values; E1A (Graft): experimental K_D measured for the grafted linkers in Fig. 4d; E1A (ED): experimental K_D measured for the variants harboring endogenous linkers and motifs in Fig. 4d. Gray box: absent motif/linker. Light/blue box: present TAZ2/MYND SLiMs. The E1A$_{WT}$ protein is marked as a red asterisk and as a red terminal branch in the tree and all other sequences used in the experiments are marked as green terminal branches in the tree. **b**, Top, E1A sequences evolved a multiplicity of solutions in the sequence length-composition space to achieve conserved SERs through conformational buffering. Bottom, the model represents one pose of the conformational E1A$_{WT}$–Rb ensemble with E2F/LxCxE SLiMs bound to Rb. The evolvable E1A interaction platform performs highly conserved functions (E2F activation) while allowing adaptive changes in functionality (TAZ2, MYND and other protein binding).

since this simplified homopolymer model does not include excluded volume or local changes in chain stiffness.

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

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

Evolutionary conservation of E1A tethering

Our results suggest that conformational buffering is a selection mechanism that conserves end-to-end distances and affinity enhancement by tethering for E1A, and that motifs and linkers co-evolve. To test these hypotheses on a larger family of sequences[56], we predicted global binding affinities for 110 distinct E1A SLiMs and linkers. We used FoldX to predict SLiM affinities, and we predicted C_{eff} using either the WLC model with a single persistence length (E1A WLC), or the sequence-specific persistence length (E1A L_pSim) from the simulations of the 27 linkers in Fig. 4a (Extended Data Fig. 10). The results are shown in Fig. 5a together with the measured affinities from the grafting experiments (E1A graft) or from endogenous variants (E1A ED).

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

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

adenoviruses is likely to employ alternative mechanisms to induce host cell proliferation.

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

Discussion

Here we demonstrate how E1A hijacks the eukaryotic cell cycle using two SLiMs tethered by a flexible linker with conserved dimensions[12,26]. The proposed docking and displacement mechanism is conserved across divergent E1A proteins by conformational buffering and coevolution of the SLiMs and tether. Conformational buffering promotes robust encoding of a core function (Fig. 5b, top) while supporting the extensive sequence variation necessary to rewire the E1A interactome (Fig. 5b, bottom) and adapt to different hosts by gaining or losing additional SLiMs[56–58], as we show for several SLiMs in Fig. 5a[39,43,56,59]. Our work challenges the view that IDRs with extensive sequence variation evolve neutrally. We also demonstrate that conserved SERs that encode for IDR dimensions—as determined by sequence length, composition and patterning—can be detected with atomistic simulations, even if they are obscured by naive sequence alignments.

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

Our work establishes E1A as an example of optimal tethering. The low picomolar affinity of E1A for Rb and the 4,000-fold affinity enhancement enforced by the E1A linker are among the highest reported positive cooperativities produced by tethering in a natural system, similar to the POU domain activator Oct-1 binding to DNA (K_D = 71 pM and 2,100-fold enhancement)[22]. E1A is the first adenoviral gene expressed, and the picomolar affinity is likely required for E1A to bind stably to Rb and efficiently displace E2F and hijack the cell cycle at low expression levels during early infection[62]. As a comparison, intramolecular MdmX inhibition exhibits optimal tethering with a 400-fold enhancement and K_{intra} = 250 (ref. [12]), and other bivalently tethered systems show variable degrees of enhancement with affinities in the nanomolar range[63–65]. At the opposite extreme, multiple low-affinity interactions tethered by short and/or non-optimal linkers might promote the dynamic binding required for multivalent binding or liquid–liquid phase separation[17,66]. Our work suggests that conformational buffering can tune the functional length of linkers to produce an optimal functional output. Thus, conformational buffering may be a widespread mechanism driving dimensional compensation among IDRs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-022-00811-w.

Received: 22 June 2021; Accepted: 23 June 2022;
Published online: 10 August 2022

References

1. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
2. van der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
3. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Mol. Cell* **55**, 161–169 (2014).
4. Brown, C. J., Johnson, A. K., Dunker, A. K. & Daughdrill, G. W. Evolution and disorder. *Curr. Opin. Struct. Biol.* **21**, 441–446 (2011).
5. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).
6. Daughdrill, G. W., Narayanaswami, P., Gilmore, S. H., Belczyk, A. & Brown, C. J. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J. Mol. Evol.* **65**, 277–288 (2007).
7. Beh, L. Y., Colwell, L. J. & Francis, N. J. A core subunit of polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proc. Natl Acad. Sci. USA* **109**, E1063–E1071 (2012).
8. Das, R. K., Huang, Y., Phillips, A. H., Kriwacki, R. W. & Pappu, R. V. Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. *Proc. Natl Acad. Sci. USA* **113**, 5616–5621 (2016).
9. Martin, E. W. et al. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
10. Zarin, T. et al. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *eLife* **8**, e46883 (2019).
11. Buske, P. J., Mittal, A., Pappu, R. V. & Levin, P. A. An intrinsically disordered linker plays a critical role in bacterial cell division. *Semin. Cell Dev. Biol.* **37**, 3–10 (2015).
12. Borcherds, W. et al. Optimal affinity enhancement by a conserved flexible linker controls p53 mimicry in MdmX. *Biophys. J.* **112**, 2038–2042 (2017).
13. Sherry, K. P., Das, R. K., Pappu, R. V. & Barrick, D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proc. Natl Acad. Sci. USA* **114**, E9243–E9252 (2017).
14. Hantschel, O. et al. A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* **112**, 845–857 (2003).
15. Ayrapetov, M. K. et al. Conformational basis for SH2-Tyr(P)527 binding in Src inactivation. *J. Biol. Chem.* **281**, 23776–23784 (2006).
16. Dyla, M. & Kjaergaard, M. Intrinsically disordered linkers control tethered kinases via effective concentration. *Proc. Natl Acad. Sci. USA* **117**, 21413–21419 (2020).
17. Cordeiro, T. N. et al. Interplay of protein disorder in retinoic acid receptor heterodimer and its corepressor regulates gene expression. *Structure* **27**, 1270–1285 (2019).
18. Brodsky, S. et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* **79**, 459–471 (2020).
19. Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
20. Huang, Q., Li, M., Lai, L. & Liu, Z. Allostery of multidomain proteins with disordered linkers. *Curr. Opin. Struct. Biol.* **62**, 175–182 (2020).

21. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl Acad. Sci. USA* **78**, 4046–4050 (1981).
22. Zhou, H. X. The affinity-enhancing roles of flexible linkers in two-domain DNA-binding proteins. *Biochemistry* **40**, 15069–15073 (2001).
23. Zhou, H. X. Polymer models of protein stability, folding, and interactions. *Biochemistry* **43**, 2141–2154 (2004).
24. Morrison, G. & Thirumalai, D. Semiflexible chains in confined spaces. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **79**, 11924 (2009).
25. van Dongen, E. M. W. M. et al. Variation of linker length in ratiometric fluorescent sensor proteins allows rational tuning of Zn(II) affinity in the picomolar to femtomolar range. *J. Am. Chem. Soc.* **129**, 3494–3495 (2007).
26. Bertagna, A., Toptygin, D., Brand, L. & Barrick, D. The effects of conformational heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case of biologically tuned disorder. *Biochem. Soc. Trans.* **36**, 157–166 (2008).
27. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl Acad. Sci. USA* **107**, 8183–8188 (2010).
28. Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* **98**, 2383–2390 (2010).
29. Müller-Späth, S. et al. From the cover: charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl Acad. Sci. USA* **107**, 14609–14614 (2010).
30. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl Acad. Sci. USA* **110**, 13392–13397 (2013).
31. Van Rosmalen, M., Krom, M. & Merkx, M. Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins. *Biochemistry* **56**, 6565–6574 (2017).
32. Sorensen, C. S. & Kjaergaard, M. Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proc. Natl Acad. Sci. USA* **116**, 23124–23131 (2019).
33. Kjaergaard, M., Glavina, J. & Chemes, L. B. Predicting the effect of disordered linkers on effective concentrations and avidity with the "C_{eff} calculator" app. *Methods Enzymol.* **647**, 145–171 (2021).
34. Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N. & Tawfik, D. S. Do viral proteins possess unique biophysical features? *Trends Biochem. Sci* **34**, 53–59 (2009).
35. Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M. & Andino, R. Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog.* **10**, e1004529 (2014).
36. Hagai, T., Azia, A., Babu, M. M. & Andino, R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep.* **7**, 1729–1739 (2014).
37. Davey, N. E., Trave, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **36**, 159–169 (2011).
38. Chemes, L. B., de Prat-Gay, G. & Sanchez, I. E. Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions. *Curr. Opin. Struct. Biol.* **32**, 91–101 (2015).
39. King, C. R., Zhang, A., Tessier, T. M., Gameiro, S. F. & Mymryk, J. S. Hacking the cell: network intrusion and exploitation by adenovirus E1A. *MBio* **9**, e00390-18 (2018).
40. Liu, X. & Marmorstein, R. Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes Dev.* **21**, 2711–2716 (2007).
41. Lee, J. O., Russo, A. A. & Pavletich, N. P. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature* **391**, 859–865 (1998).
42. Dyson, N., Guida, P., McCall, C. & Harlow, E. Adenovirus E1A makes two distinct contacts with the retinoblastoma protein. *J. Virol.* **66**, 4606–4611 (1992).
43. Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc. Natl Acad. Sci. USA* **106**, 13260–13265 (2009).
44. Ferreon, A. C., Ferreon, J. C., Wright, P. E. & Deniz, A. A. Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390–394 (2013).
45. Fattaey, A. R., Harlow, E. & Helin, K. Independent regions of adenovirus E1A are required for binding to and dissociation of E2F-protein complexes. *Mol. Cell. Biol.* **13**, 7267–7277 (1993).
46. Hosek, T. et al. Structural and dynamic characterization of the molecular hub early region 1A (E1A) froM Human Adenovirus. *Chemistry* **22**, 13010–13013 (2016).
47. Haberz, P., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of adenoviral E1A proteins with the p160 nuclear receptor coactivator binding domain of CBP. *Protein Sci.* **25**, 2256–2267 (2016).
48. Zuiderweg, E. R. P. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1–7 (2002).
49. Palopoli, N., Gonzalez Foutel, N. S., Gibson, T. J. & Chemes, L. B. Short linear motif core and flanking regions modulate retinoblastoma protein binding affinity and specificity. *Protein Eng. Des. Sel.* **31**, 69–77 (2018).
50. Perozzo, R., Folkers, G. & Scapozza, L. Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J. Recept. Signal Transduct. Res.* **24**, 1–52 (2004).
51. Theisen, F. F. et al. Quantification of conformational entropy unravels effect of disordered flanking region in coupled folding and binding. *J. Am. Chem. Soc.* **143**, 14540–14550 (2021).
52. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 (2007).
53. Estaña, A. et al. Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. *Structure* **27**, 381–391 (2019).
54. Cortes, J., Simeon, T., Remaud-Simeon, M. & Tran, V. Geometric algorithms for the conformational analysis of long protein loops. *J. Comput. Chem.* **25**, 956–967 (2004).
55. Cohan, M. C., Eddelbuettel, A. M. P., Levin, P. A. & Pappu, R. V. Dissecting the functional contributions of the intrinsically disordered C-terminal tail of *Bacillus subtilis* FtsZ. *J. Mol. Biol.* **432**, 3205–3221 (2020).
56. Glavina, J. et al. Host diversification is concurrent with linear motif evolution in a mastadenovirus hub protein. *J. Mol. Biol.* **434**, 167563 (2022).
57. Hoppe, E. et al. Multiple cross-species transmission events of human adenoviruses (HAdV) during hominine evolution. *Mol. Biol. Evol.* **32**, 2072–2084 (2015).
58. Glavina, J. et al. Interplay between sequence, structure and linear motifs in the adenovirus E1A hub protein. *Virology* **525**, 117–131 (2018).
59. Lau, L., Gray, E. E., Brunette, R. L. & Stetson, D. B. DNA tumor virus oncogenes antagonize the cGAS–STING DNA-sensing pathway. *Science* **350**, 568–571 (2015).
60. Ferreiro, D. U., Komives, E. A. & Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **47**, 285–363 (2014).
61. Sherry, K. P., Johnson, S. E., Hatem, C. L., Majumdar, A. & Barrick, D. Effects of linker length and transient secondary structure elements in the intrinsically disordered notch RAM region on notch signaling. *J. Mol. Biol.* **427**, 3587–3597 (2015).
62. Crisostomo, L., Soriano, A. M., Mendez, M., Graves, D. & Pelka, P. Temporal dynamics of adenovirus 5 gene expression in normal human cells. *PLoS ONE* **14**, e0211192 (2019).
63. Ramirez, J. et al. Targeting the two oncogenic functional sites of the HPV E6 oncoprotein with a high-affinity bivalent ligand. *Angew. Chem. Int. Ed. Engl.* **54**, 7958–7962 (2015).
64. Cheng, J. et al. Stabilized recombinant suppressors of RNA silencing: functional effects of linking monomers of Carnation Italian Ringspot virus p19. *Biochim. Biophys. Acta* **1774**, 1528–1535 (2007).
65. Travers, T. et al. Combinatorial diversity of Syk recruitment driven by its multivalent engagement with FcεRIγ. *Mol. Biol. Cell* **30**, 2331–2347 (2019).
66. Milles, S. et al. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734–745 (2015).

Methods

Protein purification and peptide synthesis and labeling. *Protein expression and purification.* The human retinoblastoma protein (Uniprot ID: P06400) AB domain (372-787 aa) with a stabilizing loop deletion (Δ582–642), named Rb, was recombinantly expressed from a pRSET-A vector in *Escherichia coli* Bl21(DE3). Briefly, Rb cultures were induced with 1 mM IPTG and grown at 28 °C overnight. Rb was purified from the soluble fraction using a Ni^{2+}-nitrilotriacetic acid immobilized metal affinity chromatography resin, followed by a purification with a sulfate cation exchange (SP-sepharose) resin and size-exclusion (Superdex 75) chromatography[67]. The adenovirus serotype 5 (HAdV5) Early 1A protein fragment (36–146) (Uniprot ID: P03255), named $E1A_{WT}$, was subcloned into BamHI and HindIII sites of a modified pMalC2x vector (New England Biolabs). $E1A_{\Delta E}$ (43-LHELY-47Δ43-AAAA-46) and $E1A_{\Delta L}$ (122-LTCHE-126Δ122-AAAA-125) variants were obtained by site-directed mutagenesis of the wild-type vector. E1A proteins were expressed as MBP fusion products in *E. coli* BL21(DE3). Unlabeled and single (^{15}N) and double ($^{15}N/^{13}C$) labeled samples were obtained from 2TY medium and M9-minimal medium supplemented with $^{15}NH_4Cl$ and [^{13}C]glucose, respectively. Cultures were induced with 0.8 mM IPTG at 0.7 OD_{600} and grown at 37 °C overnight in 2TY medium or for 5 hours after induction in M9-minimal medium. Collected cells were lysed by sonication, and proteins were isolated performing amylose affinity chromatography of the soluble fraction, followed by Q-HyperD Ion exchange and size-exclusion (Superdex 75) chromatography. The MBP tag was cleaved with Thrombin (Sigma-Aldrich, USA) at 0.4 unit per mg of protein. Synthetic MBP-E1A fusion constructs (construct sequences available in the Source File for Fig. 4) subcloned into the pMalC4x vector (GenScript) were expressed in *E. coli* Bl21(DE3), followed by Amylose purification and Superdex 75 chromatography as described above. All E1A protein stocks were stored at −80 °C in buffer containing 20 mM sodium phosphate pH 7.0, 200 mM NaCl, 20 mM DTT and 2 mM PMSF. Protein purity (>90%) and conformation were assessed by SDS–PAGE, SEC–SLS and circular dichroism analysis (Extended Data Fig. 1).

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

$E1A_{E2F}$ 36-SHFEPPTLHELYDLDV-51
$E1A_{LxCxE}$116-VPEVIDLTCHEAGFPP-131
$E1A_{LxCxE\text{-}AC}$116-VPEVIDLTCHEAGFPPSDDEDEEG-139
$E1A_{LxCxE\text{-}ACp}$116-VPEVIDLTCHEAGFPP*p*SDDEDEEG-139
Human E2F2404-SPSLDQDDYLWGLEAGEGISDLFD-427

FITC labeling. Proteins and peptides were labeled at their N terminus with fluorescein 5-isothiocyanate (FITC, Sigma), purified and quantified following a described protocol[67]. The FITC/protein ratio was above 0.8 in all cases.

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

Size-exclusion chromatography, hydrodynamic radii calculations and light scattering experiments. Analytical SEC was performed on a Superdex 75 column (GE Healthcare) calibrated with globular standards: BSA (66 kDa), MBP (45 kDa) and lysozyme (14.3 kDa). All runs were performed by injecting 100 μl protein sample ($E1A_{WT}$ and $E1A_{\Delta L}$ at 270 μM and $E1A_{\Delta E}$ at 540 μM) in 20 mM sodium phosphate buffer pH 7.0, 200 mM NaCl and 2 mM DTT. For each protein or complex, a partition coefficient (K_{av}) was calculated, and apparent molecular weights were interpolated from the –log (MW) versus K_{av} calibration curve. Experimental hydrodynamic radii (R_h) were calculated following empirical formulations developed by Uversky et al.[68]:

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

where MW is the apparent molecular weight derived from SEC experiments. The predicted R_h for $E1A_{WT}$ was calculated following the formulation developed by Marsh and Forman-Kay[28].

The scaling exponent ν was calculated from the hydrodynamic radius $R_h = R_0 \times N^{\nu}$ using the experimental R_h values, with the constant that depends on the persistent length of the polymer $R_0 = 2.49$ nm for $E1A_{WT}$ and $R_0 = 4.92$ nm for Rb, following ref. [28]. For $E1A_{WT}$, the scaling exponent ν was calculated from $R_g = R_0 \times N^{\nu}$ using R_g obtained from SAXS measurements and $R_0 = 2.1$ nm, following ref. [69]. In both cases, N^{ν} corresponds to the scaling factor, and N is the number of residues in the chain (Supplementary Table 2).

SLS–SEC was carried out to determine the average molecular weight of individual protein peaks and the stoichiometry of E1A–Rb complexes using a PD2010 detector (Precision Detectors), coupled in tandem to an HPLC system and an LKB 2142 differential refractometer. The 90° light scattering (LS) and refractive index (RI) signals of the eluting material were analyzed with Discovery32 software (Precision Detectors).

Dynamic light scattering was used to measure the hydrodynamic size distribution of E1A, using a Wyatt Dynapro Spectrometer (Wyatt Technologies). Data were fitted using Dynamics 6.1 software. All measurements were performed in 20 mM sodium phosphate buffer pH 7.0, 200 mM NaCl, 1 mM DTT at 2 mg/ml. Samples were filtered by 0.22-μM filters (Millipore) and placed into a 96-well glass-bottom black plate (In Vitro Scientific, P96-1.5H-N) covered by a high-performance cover glass (0.17 ± −0.005 mm) before measurements were taken.

Fluorescence spectroscopy experiments. Measurements were performed in a Jasco FP-6200 (Nikota) spectropolarimeter assembled in L geometry, coupled to a Peltier thermostat. Excitation and emission wavelengths were 495 nm and 520 nm, respectively, with a 4-nm bandwidth. All measurements were performed in 20 mM sodium phosphate buffer pH 7.0, 200 mM NaCl, 2 mM DTT and 0.1% tween-20 at 20 ± 1 °C.

For direct titrations, a fixed concentration of FITC-labeled protein/peptide was titrated with increasing amounts of Rb until saturation was reached. Maximal dilution was 20%, and samples were equilibrated for 2 minutes to ensure steady state. Titrations were performed at concentrations ten times higher than the equilibrium K_D, allowing estimation of the stoichiometry of each reaction. Binding titrations performed at sub-stoichiometric concentrations allowed an estimation of K_D, by fitting the titration curves to a bimolecular association model:

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

where Y is the measured anisotropy signal, Y_F and Y_B are the free and bound labeled peptide signals, P_0 is the total labeled peptide concentration, x is Rb concentration, and K_D is the equilibrium dissociation constant in molar units. The $C \times x$ linear term accounts for slight bleaching or aggregation. Data were fitted using the Profit 7.0 software (Quantumsoft), yielding a value for each parameter and its corresponding standard deviation. Titrations for each complex were performed in triplicate for at least at three different concentrations of FITC-labeled sample, and parameters were obtained from fitting individual titrations or by global fitting of the K_D parameter using normalized titration curves at different concentrations, obtaining an excellent agreement between individual and global fits (Supplementary Table 3 and Extended Data Fig. 3).

Competition experiments were carried out by titrating the pre-assembled complex Rb–FITC-E2F2 (1:1 molar ratio, 5 nM), with increasing amounts of unlabeled competitors and following the decrease in the anisotropy signal, until the value corresponding to free FITC-E2F2 was reached. Half-maximal inhibitory concentration values were estimated directly from the curves as the concentration where the competitor produced a decrease in 50% of the maximal anisotropy value. K_D values were calculated by fitting the data considering the binding equilibrium of the labeled peptide and the unlabeled competitors, according to ref. [70], obtaining $K_{D(comp)}$ values that differed only slightly (two- to threefold) from those obtained from direct titrations. K_D and $K_{D(comp)}$ values also displayed similar fold changes in binding affinity relative to E2F2 within each method (Supplementary Table 1). The agreement between the K_D values obtained from fluorescence and ITC titrations (Supplementary Table 1) confirmed that FITC moiety did not cause substantial changes in Rb binding affinity. MBP-E1A fusion protein affinities (Supplementary Data Table 4 and Extended Data Fig. 8) were determined by performing competition experiments assembling a Rb–FITC-E2F2 complex at 10 nM concentration, after verifying that MBP-$E1A_{WT}$ and $E1A_{WT}$ (cleaved and uncleaved HAdV5 proteins) had the same binding affinity (Supplementary Table 7). Measurements were performed on a PTI Quantamaster QM40 spectrofluorometer (Horiba) equipped with polymer film polarizers and coupled to a Peltier thermostat with excitation parameters as described above.

Isothermal titration calorimetry experiments. *Direct titrations.* ITC experiments were performed on MicroCal VP-ITC and MicroCal PEAQ-ITC equipment (Malvern Panalytical) in 20 mM sodium phosphate pH 7.0, 200 mM NaCl, 5 mM 2-mercaptoethanol at 20.0 ± 0.1 °C, unless stated otherwise. Prior to titrations, cell and titrating samples were co-dialyzed in the aforementioned buffer for 48 hours at 4 ± 1 °C and then de-gassed. Measurements performed in the MicroCal VP-ITC used 28 10-μl injections at a flow rate of 0.5 μl/second, and those performed in the MicroCal PEAQ-ITC used 13 3-μl injections. The concentration ranges of cell and titrating samples are detailed in Extended Data Figs. 2 and 5. Data were analyzed using the Origin software.

Allosteric coupling experiments. First, a pre-assembled Rb–$E1A_{LxCxE}$ complex (1:1 molar ratio, 30 μM) was titrated with $E1A_{E2F}$ or $E1A_{\Delta L}$ to assess whether binding of the LxCxE motif modified the binding affinity for the E2F site. Conversely, pre-assembled Rb–$E1A_{E2F}$ or Rb–$E1A_{\Delta L}$ complexes were titrated with $E1A_{LxCxE}$ to assess whether binding of the E2F motif modified the binding affinity for the LxCxE site (Supplementary Table 7).

Calculation of ΔC_p and ΔASA parameters from ITC data. A series of titrations were carried out at different temperatures (10.0, 15.0, 20.0 and 30.0 ± 0.1 °C) and the change in binding heat capacity (ΔC_p) was obtained from the slope of the linear regression analysis of the plot of ΔH versus temperature (Extended Data Fig. 5). The changes in accessible surface area (ΔASA_T) and X_{res} were estimated by solving semi-empirical equations from protein-folding studies applied to protein–ligand binding[50] and from models that use parameters derived for IDPs (IDP-specific model)[51] (details on the model-specific parameter values are provided in Supplementary Tables 5 and 6). First we calculated $\Delta H_{int(TH)}$ from:

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

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

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

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

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

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

where $\Delta S_{residue}$ is the change in configurational entropy per residue. The change in configurational entropy (ΔS_{config}) was calculated as the sum of changes in rotation–translation (ΔS_{rt}) and solvation (ΔS_{solv}) entropy:

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

with ΔS_{solv} defined as:

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

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

Nuclear magnetic resonance experiments. NMR experiments were carried out using a Varian VNMRS 800 MHz spectrometer equipped with a triple resonance pulse field *z*-axis gradient cold probe. A series of two-dimensional sensitivity-enhanced ^{1}H–^{15}N HSQC and three-dimensional HNCACB, HNCO and CBCA(CO)NH experiments[71,72] were performed for backbone resonance assignments on uniformly ^{13}C–^{15}N-labeled samples of E1A$_{WT}$, E1A$_{\Delta E}$ and E1A$_{\Delta L}$ at 700 µM, 975 µM and 850 µM, respectively. All measurements were performed in 10% D_2O, 20 mM sodium phosphate pH 7.0, 200 mM NaCl and 2 mM DTT at 25 °C. The HSQC used 9,689.9 Hz and 1,024 increments for the t_1 dimension and 2,106.4 Hz with 128 increments for the t_2 dimension. The HNCACB experiment used 9,689.9, 14,075.1 and 2,106.4 Hz with 1,024, 128 and 32 increments for the t_1, t_2 and t_3 dimensions, respectively. The HNCO experiment used 9,689.9, 2,010.4 Hz and 2,106.4 Hz with 1,024, 64 and 32 increments for the t_1, t_2 and t_3 dimensions, respectively. The CBCA(CO)NH experiment used 9,689.9, 14,072.6 and 2,106.4 Hz with 1,024, 128, and 32 increments for the t_1, t_2 and t_3 dimensions, respectively. For E1A$_{WT}$, 88% of non-proline backbone ^{1}H and ^{15}N nuclei, 75% of ^{13}C′ nuclei and 90% of $^{13}C_\alpha$ and $^{13}C_\beta$ of E1A nuclei were assigned (Supplementary Data 2). For E1A$_{\Delta E}$ and E1A$_{\Delta L}$, 85% of non-proline backbone ^{1}H and ^{15}N nuclei, 72% of ^{13}C′ nuclei and 87% of $^{13}C_\alpha$ and $^{13}C_\beta$ E1A nuclei were assigned.

NMRPipe and NMRViewJ software packages were used to process and analyze all the NMR spectra[73]. Residue-specific random-coil chemical shifts were generated for the three sequences using the neighbor-corrected IDP chemical shift library[74]. Secondary chemical shifts ($\Delta\delta$) were calculated by subtracting random-coil chemical shifts from the experimentally obtained chemical shifts.

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

Molecular modeling of Rb–E1A conformational ensembles. Conformations of E1A$_{WT}$ bound to Rb were modeled using an extended version of a recently proposed method to generate realistic conformational ensembles of IDPs[53]. Conformational ensemble models of E1A$_{WT}$–Rb were generated using a stochastic sampling algorithm implemented in the MoMA software suite (https://moma.laas.fr). This method exploits local, sequence-dependent structural information encoded in a database of three-residue fragments and builds conformations incrementally sampling dihedral angles values from the database, while avoiding steric clashes. In order to model the double-bound Rb–E1A$_{WT}$ complex, the E2F and LxCxE motifs were considered to be static, preserving the conformations extracted from experimentally determined structures (2R7G and 1GUX). The 71-residue fragment between these two motifs was considered as a long protein loop that adapts its conformation in order to maintain the two ends rigidly positioned. Conformational sampling considering such loop-closure constraints was performed using a robotics-inspired method[54] adapted to use dihedral angle values from the aforementioned database. For each feasible conformation of the central fragment, geometrically compatible conformations of the short N- and C-terminal tails were sampled using the basic strategy explained in ref. [53]. For singly-bound models E1A$_{\Delta L}$–Rb and E1A$_{\Delta E}$–Rb, only one of the two motifs was considered to be statically bound to Rb, and the other motif behaved as the flexible linker. The loop sampling method used to model the linker between the two binding motifs can be used via a web server (https://moma.laas.fr/applications/LoopSampler/). Binaries can be provided upon request.

Small angle X-ray scattering experiments. SAXS experiments for Rb and E1A$_{WT}$–Rb were carried out at the European Molecular Biology Laboratory beamline P12 of PETRA-III storage ring, using the X-ray wavelengths of 1.24 Å and a sample-to-detector distance of 3.0 m[75]. The scattering profiles measured covered a momentum transfer range of $0.0026 < s < 0.73$ Å^{-1}. SAXS data for E1A were collected at the at the SWING beamline at the SOLEIL synchrotron on an Eiger 4M detector with a sample-to-detector distance of 2.0 m. SAXS data were measured for Rb, E1A$_{WT}$ and the E1A$_{WT}$–Rb complex at 10 °C. Concentrations used for E1A$_{WT}$ were 7.0, 5.6 and 4.2 mg/ml, for Rb were 4.0, 2.0, 1.0 mg/ml and for and E1A$_{WT}$–Rb were 2.7, 1.4, and 0.7 mg/ml, in 20 mM sodium phosphate pH 7.0, 200 mM NaCl and 1 mM DTT. The scattering patterns of the buffer solution were recorded before and after the measurement of each sample. Multiple repetitive measurements were performed to detect and correct for radiation damage. The initial data-processing steps, including masking and azimuthal averaging, were performed using the SASFLOW version 3.0. pipeline for Rb and E1A$_{WT}$–Rb, and the program FOXTROT version 3.5.2 (ref. [76]). for E1A. Final curves at each concentration were derived after the averaged buffer scattering patterns were subtracted from the protein sample patterns. No sign of aggregation was observed in any of the curves. Final SAXS profiles for the systems were obtained by merging curves for the lowest and highest concentrations to correct small attractive interparticle effects observed. The SAXS profiles were analyzed using the ATSAS suite of programs version 2.8.4 (ref. [77]). The forward scattering intensity, $I(0)$, and the radius of gyration, R_g, were evaluated using Guinier's approximation[78], assuming that at very small angles $s < 1.3 / R_g$, the intensity can be well represented as $I(s) = I(0)\ e^{-(sRg)2/3}$. The $P(r)$ distribution functions were calculated by indirect Fourier Transform using GNOM[79], applying a momentum transfer range of $0.01 < s < 0.33$ Å^{-1} and $0.013 < s < 0.27$ Å^{-1} for Rb and Rb–E1A, respectively. For E1A$_{WT}$ a SEC–SAXS experiment was also performed which was processed using the program CHROMIX[80] which is a part of ATSAS 2.8.4. to obtain the SAXS profile from a highly monodisperse sample. This profile overlaid perfectly with the E1A$_{WT}$ merged curve from the three batch experiments, discarding aggregation problems.

The fitting of the crystallographic structure of Rb (PDB: 3POM)[81] to the experimental SAXS curve was performed with FOXS[82,83]. An optimal fit ($\chi^2 = 0.86$) was obtained after modeling the missing parts (loops, N and C termini) and a subsequent refinement with the program AllosMod-FoXS[84]. SAXS data measured for Rb–E1A were analyzed with the Ensemble Optimization Method (EOM)[52,85]. Briefly, theoretical SAXS profiles of the 10,250 structures of the complex were computed with CRYSOL[86]. Two hundred sub-ensembles of 20 or 50 conformations collectively describing the experimental curve were collected with EOM and analyzed in terms of R_g distributions. The experimental SAXS data for the E1A$_{WT}$–Rb complex are compatible with three distinct scenarios: a 100% doubly-bound ensemble where the linker is highly expanded, a 100% singly-bound ensemble where the linker is highly compact, or an ensemble with a combination of 76% doubly-bound and 24% singly-bound species, resulting from the linear combination of a curve representing the ensemble average of all singly- and all doubly-bound conformations. However, thermodynamic (K_D for E1A$_{WT}$) data, which indicate an extremely low expected population of the singly-bound forms at any concentration of the complex used in the SAXS experiments, strongly argue against the last two scenarios.

Hydrodynamic radii for generated conformations. Hydrodynamic radii were calculated using the program HydroPro (version 10)[87,88]. HydroPro was run on 1,000 models selected by EOM for the doubly-bound conformations and 1,000 randomly selected conformations of N- and C-terminal-bound conformations. The calculations were done at temperatures of 20 °C and 25 °C with corresponding solvent viscosities of 0.01 and 0.009 poise, respectively. The values of atomic element radius, molecular weight, partial specific volume and solvent density were set to 2.9 Å, 54,590 Da, 0.702 cm^3/g and 1.0 g/cm^3, respectively. These values have no associated error.

All-atom simulations of E1A linker sequences. All-atom simulations were run using the CAMPARI simulation engine (V2) version 2.0 (http://campari.sourceforge.net) and ABSINTH implicit solvent model ABS-OPLS3.2 (refs. [89,90]). All simulations were performed at a temperature of 320 K; although this is a slightly elevated temperature compared with the experimental temperature, none of the terms the Hamiltonian lacks temperature dependence such that this slightly high temperature serves to uniformly improve sampling quality across all simulations. This approach has been leveraged to great effect in previous studies and is especially convenient in the case of simulating many sequences spanning a range of sequence properties and lengths[8]. A collection of Monte Carlo moves was used to fully sample conformational space, as previously described[13,91,92].

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

Length titration simulations. The linker from HF_HAdV40 was used to determine the overall amino acid composition and to generate random sequences across a range of lengths that recapitulated this composition. Specifically, for each length (45, 50, 55, 60, 65, 70, 75) 20 random sequences were generated for a total of 140 randomly generated sequences. Each sequence was simulated under equivalent simulation conditions for 35×10^9 simulation steps, with the goal of elucidating the general relationship between sequence length and end-to-end distance for an arbitrary sequence of the composition associated with HF_HAdV40. The mean end-to-end distance for the collection of sequences at a given length was determined, such that the mean value is a double average over both conformational space and sequence space.

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

$$p(r) = 4\pi r^2 \left(\frac{3}{4\pi L_p L_c}\right)^{\frac{3}{2}} \exp\left(\frac{-3r^2}{4L_p L_c}\right) \zeta(r, L_p, L_c) \quad (9)$$

where p(r) is a function of distance r and depends on L_p and L_c. The last term in the equation is expanded on in refs. [22,23]. The end-to-end probability density function can be related to the effective concentration in the bound state when the linker is restrained to a fixed distance between binding sites, r_0 (ref. [22]). In this case, C_{eff} is defined by:

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

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

Calculation of experimental and predicted C_{eff} values for the E1A$_{WT}$-Rb interaction. Experimental C_{eff} values. In Model A, the global association constant is calculated as: $K_G = K_1 \times K_2 \times C_{eff}$ (Fig. 1f) where K_G, K_1 and K_2 are equilibrium association constants ($K = 1 / K_D$). Therefore, this relationship can be expressed equivalently as $K_D = K_{D,E2F} \times K_{D,LxCxE} \times C_{eff}^{-1}$. Here, $K_{D,E2F}$ and $K_{D,LxCxE}$ are the equilibrium dissociation constants of the E1A$_{E2F}$ and E1A$_{LxCxE}$ motifs, respectively (reported in Supplementary Table 1), and K_D is the global equilibrium dissociation constant for E1A$_{WT}$ binding to Rb (reported in Supplementary Table 1). The condition $K_1 = K_1'$ and $K_2 = K_2'$ (no allosteric coupling between sites) was met (Extended Data Fig. 5 and Supplementary Data Table 7). Therefore, the experimentally derived C_{eff} was calculated from the measured binding constants as: $C_{eff} = (K_{D,E2F} \times K_{D,LxCxE}) / K_D$ (Fig. 3a).

Predicted C_{eff} values. The C_{eff} value predicted from the WLC model (Fig. 3a) was obtained by applying Eq. 10 with the designated L_p parameter (standard model $L_p = 3$ Å and $b = 3.8$ Å), using a linker length of 71 residues for HAdV5 E1A. The separation between binding sites, r_0, was 49 Å (from PDB: 1GUX and PDB: 2R7G).

Calculation of predicted global binding affinity for grafted E1A linkers. We predicted the K_D values expected for each of the grafted linker variants of Fig. 4c,d under a sequence-independent WLC model or accounting for sequence-dependent changes in the persistence length. We calculated K_D for E1A$_{WT}$ as $K_D = (K_{D,E2F} \times K_{D,LxCxE}) / C_{eff}$, where $K_{D,E2F}$ and $K_{D,LxCxE}$ are the K_D values for the E1A$_{E2F}$ and E1A$_{LxCxE}$ motifs of E1A$_{WT}$ (reported in Supplementary Table 1) and C_{eff} was calculated using the WLC model, as described below. Competition experiments for E1A$_{WT}$ typically yield K_D values approximately threefold higher than those obtained by direct titration (Supplementary Tables 1 versus 4). To correct for this effect, the predicted K_D values were corrected by a factor of three. Because the only element changing for each grafted variant in the grafting experiment is the linker (that is, the C_{eff} value), we calculated $K_{D,VARIANT}$ as $K_{D,VARIANT} = K_{D,E1AWT} / (C_{eff}\text{ratio})$, where $C_{eff}\text{ratio} = C_{eff,VARIANT} / C_{eff,E1AWT}$. All predicted C_{eff}, L_pSim and K_D values for the grafting experiment are reported in the source data for Fig. 5 and Extended Data Fig. 10. The C_{eff} values were calculated as follows.

WLC-L_p = 3 model. For the standard assumption of a sequence-independent model (WLC-L_p = 3), we calculated the C_{eff} function as a function of linker length ($C_{eff(L)}$) using Eqs. 9 and 10 with a standard value for the persistence length parameter (L_p = 3). We calculated the expected K_D as a function of linker length as $K_D = (K_{D,E2F} \times K_{D,LxCxE}) / C_{eff(L)}$ (straight line, Fig. 4d).

WLC-L_pSim model. For the sequence-dependent model (WLC-L_pSim), we calculated C_{eff} for each linker using Eqs. 9 and 10, applying the specific number of residues (N_{res}) of each linker and an individual sequence-dependent L_p value for each linker (L_pSim), which was obtained from the simulations. L_pSim values were calculated from the average end-to-end distance of each simulated ensemble (<r²>) using the equation $\langle r^2 \rangle = 2 \times L_p \times L_c$, where $L_c = N_{res} \times b$, and b takes the value 3.8 Å. This equation is an approximation for the value of $\langle r^2 \rangle$ for a WLC in the case where the contour length of the chain is much larger than its persistence length ($L_c \gg L_p$)[23].

Calculation of predicted global binding affinity for a large family of E1A linkers. The WLC model was used to estimate the C_{eff} values and global Rb binding affinities of a collection of 110 natural linker sequences of different lengths, changing the length value for each linker and keeping other parameters constant. All values are reported in the source data for Fig. 5 and Extended Data Fig. 10.

Dataset. A previously reported alignment and phylogenetic tree of 116 mastadenovirus E1A sequences[56,58] was used to identify the E2F and LxCxE motifs, as described[58]; we collected 110 sequences in which both motifs were present (Supplementary File 1). For all sequences, the length of the linker region between both motifs was recorded. Individual motif-binding affinities, C_{eff} values and E1A global affinity ($K_{D,E1A}$) were calculated as explained below (source data for Fig. 5 and Extended Data Fig. 10).

Calculation of E1A binding affinity. The global binding affinity $K_{D,E1A}$ (Extended Data Fig. 10) was calculated as $K_{D,E1A} = (K_{D,E2F} \times K_{D,LxCxE}) / C_{eff}$, where C_{eff} is the C_{eff} value predicted under a naive or sequence-dependent assumption (see details below), and $K_{D,E2F}$ and $K_{D,LxCxE}$ are the predicted binding affinities of each motif calculated using FoldX:

Prediction of motif-binding affinities using FoldX. To estimate the binding affinity of individual E2F and LxCxE motifs ($K_{D,E2F}$ and $K_{D,LxCxE}$) present in each sequence, FoldX v5.0 (ref. [96]) was used to build substitution matrices for all 20 amino acids at each position (source data for Extended Data Fig. 10). Briefly, given a structural complex, the FoldX algorithm assesses the change in binding free energy produced by mutating each position of the motif for each of the 20 amino acids. For the E2F matrix, the structure of the HAdV5 E1A$_{E2F}$ motif in complex with Rb (PDB: 2R7G) was used as input. For the LxCxE matrix, the structure used as input was a model of the HAdV5 E1A$_{LxCxE}$ motif in complex with Rb (Supplementary Data File 3), built using FlexPepDock[97], and the structure of the HPV E7 LxCxE motif bound to Rb (PDB: 1GUX). The total change in binding free energy with respect to the wild-type sequence ($\Delta\Delta G_{FoldX}$) was calculated by adding the free-energy terms for each residue at each matrix position (source data for Extended Data Fig. 10). The predicted equilibrium dissociation constant of the E2F and LxCxE motifs for each sequence ($K_{D\,SEQ}$) was calculated as:

$$\Delta\Delta G_{FoldX} = \Delta G_{SEQ} - \Delta G_{WT} = RT\ln(K_{D\,SEQ}) - RT\ln(K_{D\,WT}) \quad (11)$$

$$K_{D\,SEQ} = \frac{K_{D\,WT}}{e^{(-\Delta\Delta G_{FoldX}/RT)}} \quad (12)$$

where $\Delta\Delta G_{FoldX}$ is the total predicted change in binding energy calculated using FoldX, RT is 0.582 kcal mol^{-1}, $K_{D\,WT}$ is the experimentally measured binding affinity of the sequence (HAdV5 E1A) present in the model structure ($K_{D,E2F}$ and $K_{D,LxCxE}$ measured in this work, Supplementary Table 1).

Prediction of C_{eff} values under the naive worm-like chain model. The C_{eff} value was calculated for the collection of 110 natural E1A linkers using Eqs. 9 and 10, with L_p = 3 Å (L_p WLC, Fig. 5a) and the specific length (number of residues) of each linker, which defines L_C.

Prediction of C_{eff} values using a worm-like chain model with sequence-dependent L_p parameters. For the subset of 27 natural E1A linkers used in all-atom simulations (Fig. 4a) we calculated sequence-specific L_p values from all-atom simulations (L_pSim) in order to represent sequence-dependent changes in chain expansion. The details of these calculations are explained in 'Calculation of predicted global binding affinity for grafted E1A linkers'. New C_{eff} values were derived using the same parameters described above, but replacing the standard L_p value by the $L_{p\,Sim}$ value. The L_pSim values are reported in the source data for Extended Data Fig. 10.

Statistical analysis. We used bootstrapping[98] to generate 99% confidence intervals (CI) for $K_{D,E2P}$, $K_{D,LxCxE}$ and $K_{D,E1A}$ average values, and compared the lower and upper end points against the value of $K_{D,E2F2}$ (1×10^{-9} M). The lower bound of the 99% CI for $K_{D,E2F}$ and $K_{D,LxCxE}$ is higher than $K_{D,E2F2}$ and the upper bound of the 99% CI for all $K_{D,E1A}$ are lower than $K_{D,E2F2}$. We also used permutation tests[98] to assess the null hypothesis that the C_{eff}, L_p and K_D average values did not differ between all pairs of groups. In order to control for the false-discovery rate, the P values were corrected using the Benjamini–Hochberg[99] correction for multiple comparisons.

Calculations of disorder propensity and conservation. All calculations were performed on the dataset from Supplementary Data 1, using the methods described in ref. [58]. For disorder propensity, we recorded the mean IUPred value ± s.d. per position using IUPred 2a[100], and for residue conservation, we recorded the IC per position.

Sequence conservation and evolutionary scores. We collected 77 mammalian orthologous sequences of the retinoblastoma protein with no unassigned residues within the pocket domain from the Ensembl Database[101]. Sequences were aligned using MUSCLE v3.8.1551 and manually curated according to structural information. The evolutionary conservation scores were calculated with the Consurf 2016 web server[102] using the E7–Rb complex (PDB: 1GUX) as the structural model. The sequence analyses and alignment graphics were performed using Jalview v2.11 (ref. [103]). The alignment was colored according to residue identity, and conservation scores were calculated according to ref. [104]. We analyzed the conservation of residues making up the E2F and LxCxE motif-binding sites according to previously reported contacts[40,41]. Results are presented in Supplementary Fig. 1.

Structural modeling. The structures of the human Rb (RbAB pocket domain) bound to E1A (PDB: 2R7G) and E7 (PDB: 1GUX), and the structure from the human paralogue p107 pocket domain (PDB: 4YOZ) were collected from the protein data bank. Structural modeling of the human paralogue p130, and the retinoblastoma pocket domains from macaque (*Macaca mulatta*), chimpanzee (*Pan troglodytes*), dog (*Canis lupus familiaris*), microbat (*Myotis lucifugus*), sheep (*Ovis aries*), pig (*Sus scrofa*), cow (*Bos taurus*), horse (*Equus caballus*) and tree shrew (*Tupaia belangeri*) were obtained by using Alphafold v2.0 (ref. [105]) implemented in ColabFold v1.0 (ref. [106]). The template multiple sequence alignments were generated using MMseqs2 (ref. [107]) implemented within ColabFold v1.0. Template information and the predicted structure relaxation using amber force fields were included. The distances between the E2F and LxCxE binding sites were measured between the alpha carbons of the C-terminal anchor site of the E2F cleft and the N-terminal anchor site of the LxCxE cleft. Molecular graphics and analyses were performed with UCSF Chimera v1.5 (ref. [108]). Results are presented in Extended Data Fig. 9.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

SAXS raw data for Rb, E1A$_{WT}$ and the E1A$_{WT}$–Rb complex have been deposited in SASDB (https://www.sasbdb.org) with codes SASDNK6 (Rb 1 mg/ml), SASDNL6 (Rb 2 mg/ml), SASDNM6 (Rb 4 mg/ml), SASDNN6 (E1A$_{WT}$ 4.2 mg/ml), SASDNP6 (E1A$_{WT}$ 5.6 mg/ml), SASDNQ6 (E1A$_{WT}$ 7.0 mg/ml), SASDNR6 (E1A$_{WT}$–Rb 0.7 mg/ml), SASDNS6 (E1A$_{WT}$–Rb 1.4 mg/ml), SASDNT6 (E1A$_{WT}$–Rb 2.7 mg/ml), SASDNU6 (E1A$_{WT}$–Rb merged data) and SASDNV6 (E1A$_{WT}$, SEC–SAXS). Refined conformational ensemble models for E1A$_{WT}$ and E1A$_{WT}$–Rb have been deposited in the Protein Ensemble Database (https://proteinensemble.org) with codes PED00175 (E1A$_{WT}$) and PED00174 (E1A$_{WT}$–Rb). Unfiltered conformational ensembles for the E1A$_{WT}$–Rb, E1A$_{ΔL}$–Rb and E1A$_{ΔE}$–Rb complexes are available at (https://moma.laas.fr/data/) under the description 'Conformational ensemble models of the IDP E1A bound to Rb protein.' NMR assignments of backbone resonances for E1A$_{WT}$, E1A$_{ΔE}$ and E1A$_{ΔL}$ are provided in Supplementary Data 2. Trajectories for all E1A linker ensembles are provided at Zenodo (https://zenodo.org/record/6332925), and trajectory analysis results are provided at https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021. PDB codes used in data analysis and prediction are: 1GUX, 3POM, 2R7G and 4YOZ. Source data are provided with this paper.

Code availability

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

References

67. Chemes, L. B., Noval, M. G., Sanchez, I. E. & de Prat-Gay, G. Folding of a cyclin box: linking multitarget binding to marginal stability, oligomerization, and aggregation of the retinoblastoma tumor suppressor AB pocket domain. *J. Biol. Chem.* **288**, 18923–18938 (2013).
68. Uversky, V. N. What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12 (2002).
69. Hofmann, H. et al. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl Acad. Sci. USA* **109**, 16155–16160 (2012).
70. Kuzmic, P., Moss, M. L., Kofron, J. L. & Rich, D. H. Fluorescence displacement method for the determination of receptor-ligand binding constants. *Anal. Biochem.* **205**, 65–69 (1992).
71. Muhandiram, D. R. & Kay, L. E. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B* **103**, 203–216 (1994).
72. Wittekind, M. & Mueller, L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson. B* **101**, 201–205 (1993).
73. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
74. Tamiola, K., Acar, B. & Mulder, F. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 18000–18003 (2010).
75. Blanchet, C. E. et al. Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443 (2015).
76. Girardot, R., Viguier, G., Pérez, J. & Ounsy, M. M. FOXTROT: A JAVA-based application to reduce and analyse SAXS and WAXS piles of 2D data at synchrotron SOLEIL, Synchrotron Soleil. In *Proc. 8th canSAS Meeting, Tokai, Japan* (2015).
77. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
78. Guinier, A. Diffraction of X-rays of very small angles-application to the study of ultramicroscopic phenomenon. *Ann. Phys.* **12**, 161–237 (1939).
79. Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. Small-angle-scattering-data treatment by the regularization method. *Acta Crystallogr. A* **44**, 244–250 (1988).
80. Panjkovich, A. & Svergun, D. I. CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
81. Balog, E. R., Burke, J. R., Hura, G. L. & Rubin, S. M. Crystal structure of the unliganded retinoblastoma protein pocket domain. *Proteins* **79**, 2010–2014 (2011).
82. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
83. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
84. Weinkam, P., Pons, J. & Sali, A. Structure-based model of allostery predicts coupling between distant sites. *Proc. Natl Acad. Sci. USA* **109**, 4875–4880 (2012).
85. Tria, G., Mertens, H. D., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207–217 (2015).
86. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

87. Garcia De La Torre, J., Huertas, M. L. & Carrasco, B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730 (2000).
88. Ortega, A., Amoros, D. & Garcia de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **101**, 892–898 (2011).
89. Vitalis, A. & Pappu, R. V. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* **30**, 673–699 (2009).
90. Vitalis, A. & Pappu, R. V. Methods for Monte Carlo simulations of biomacromolecules. *Annu. Rep. Comput. Chem.* **5**, 49–76 (2009).
91. Kozlov, A. G. et al. Intrinsically disordered C-terminal tails of *E. coli* single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA. *J. Mol. Biol.* **427**, 763–774 (2015).
92. Metskas, L. A. & Rhoades, E. Conformation and dynamics of the troponin I C-terminal domain: combining single-molecule and computational approaches for a disordered protein region. *J. Am. Chem. Soc.* **137**, 11962–11969 (2015).
93. McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
94. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. & Pappu, R. V. CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys. J.* **112**, 16–21 (2017).
95. Nygaard, M., Kragelund, B. B., Papaleo, E. & Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys. J.* **113**, 550–557 (2017).
96. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
97. London, N., Raveh, B., Cohen, E., Fathi, G. & Schueler-Furman, O. Rosetta FlexPepDock web server—high resolution modeling of peptide-protein interactions. *Nucleic Acids Res.* **39**, W249–W253 (2011).
98. Good, P. *Permutation, Parametric, and Bootstrap Tests of Hypotheses* (Springer-Verlag, 2005).
99. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Ser. B* **57**, 289–300 (1995).
100. Mészáros, B., Erdos, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
101. Howe, K. L. et al. Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
102. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
103. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
104. Livingstone, C. D. & Barton, G. J. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Appl. Biosci.* **9**, 745–756 (1993).
105. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
106. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
107. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
108. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

This work was supported by: Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) Grants PICT no. 2013-1895 and no. 2017-1924 (L. B. C.), no. 2012-2550 and no. 2015-1213 (I. E. S.) and no. 2016-4605 (G. P. G.); US National Institutes of Health no. GM115556 and no. CA141244 (G. W. D.) and no. 5R01NS056114 (R. V. P.); Florida Department of Health (FLDOH) no. 20B17 (G. W. D.); US National Science Foundation no. MCB-1614766 (R. V. P.); a travel award from the USF Nexus Initiative and a Creative Scholarship Grant from the USF College of Arts and Sciences (G. W. D. and L. B. C.); Labex EpiGenMed (Investissements d'avenir) program no. ANR-10-LABX-12-01 (P. B.); French National Research Agency no. ANR-10-INBS-04-01 and no. ANR-10-INBS-05 (P. B.); Spanish Ministerio de Ciencia y Universidades MICYU-FEDER no. RTI2018-097189-C2-1 (G. F.-B.); Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) doctoral fellowship (N. S. G.-F., M. S. and N. A. G.), postdoctoral fellowship (J. G.), and permanent researcher (L. B. C., G. d. P.-G. and I. E. S.); Fulbright Visiting Scholar Program (N. S. G.-F.); Ministerio de Ciencia e Innovación, España, no. BES-2013-063991 and no. EEBB-I-16-11670 (S. B.-V.); Longer Life Foundation: A RGA/Washington University Collaboration (A. S. H.); HPC resources of the CALMIP supercomputing center no. 2016-P16032 (G. F.-B.); and Cluster of Scientific Computing (http://ccc.umh.es/) of the Miguel Hernández University (G. F.-B.). The synchrotron SAXS data were collected at beamline P12, operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). We thank K. Perez at the Protein Expression and Purification Core Facility at EMBL (Heidelberg) for critical help with ITC experiments, and P. Aramendia for providing critical access to fluorescence spectrometry equipment at Centro de Investigaciones en Bionanociencias (CIBION, Argentina).

Author contributions

L. B. C., G. W. D., A. S. H. and R. V. P. designed research and conceived the study. N. S. G.-F., W. M. B. and M. S. produced reagents. N. S. G.-F. and W. M. B. performed FP, ITC and NMR experiments, and W. B., N. S. G.-F., G. W. D. and L. B. C. analyzed data. J. G. designed and conducted bioinformatic analyses of E1A variants and Rb proteins. M. S. purified E1A protein variants and N. A. G. performed SEC experiments. A. S. and P. B. performed and analyzed SAXS experiments. A. E., A. B. and J. C. produced and analyzed E1A conformational ensembles. S. B.-V. and A. S. H. performed and analyzed all-atom simulations of E1A linkers. G. F.-B., C. B.-M. and IES computed and analyzed FOLDX matrices. N. S. G.-F., J. G., A. S. and A. S. H. produced figures. L. B. C., G. W. D., P. B., J. C., G. d. P.-G., I. E. S., A. S. H. and R. V. P. supervised research. L. B. C., N. S. G.-F., J. G., R. V. P., A. S. H. and G. W. D. wrote the paper, with critical feedback from all authors.

Competing interests

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

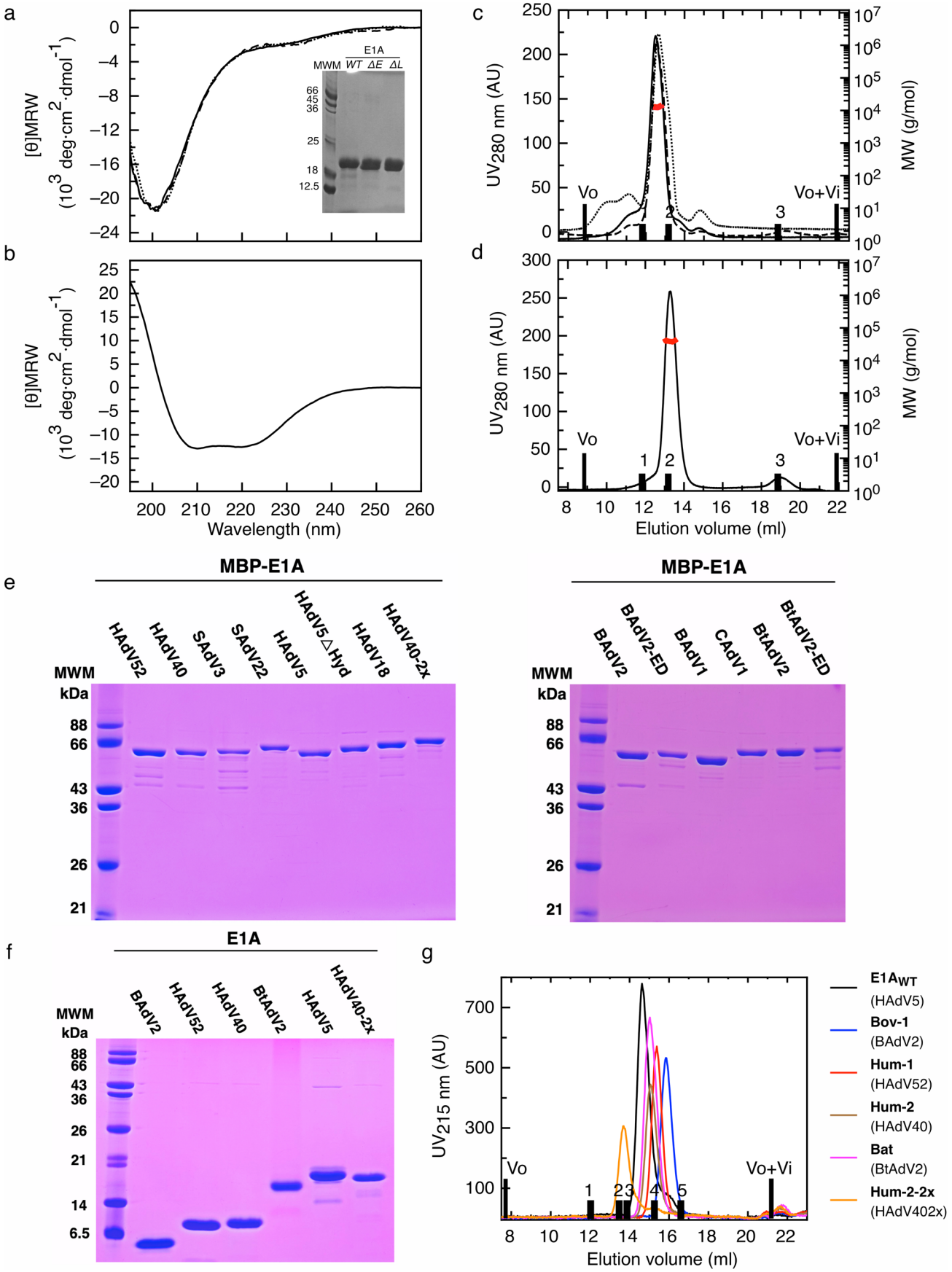
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-022-00811-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-022-00811-w.

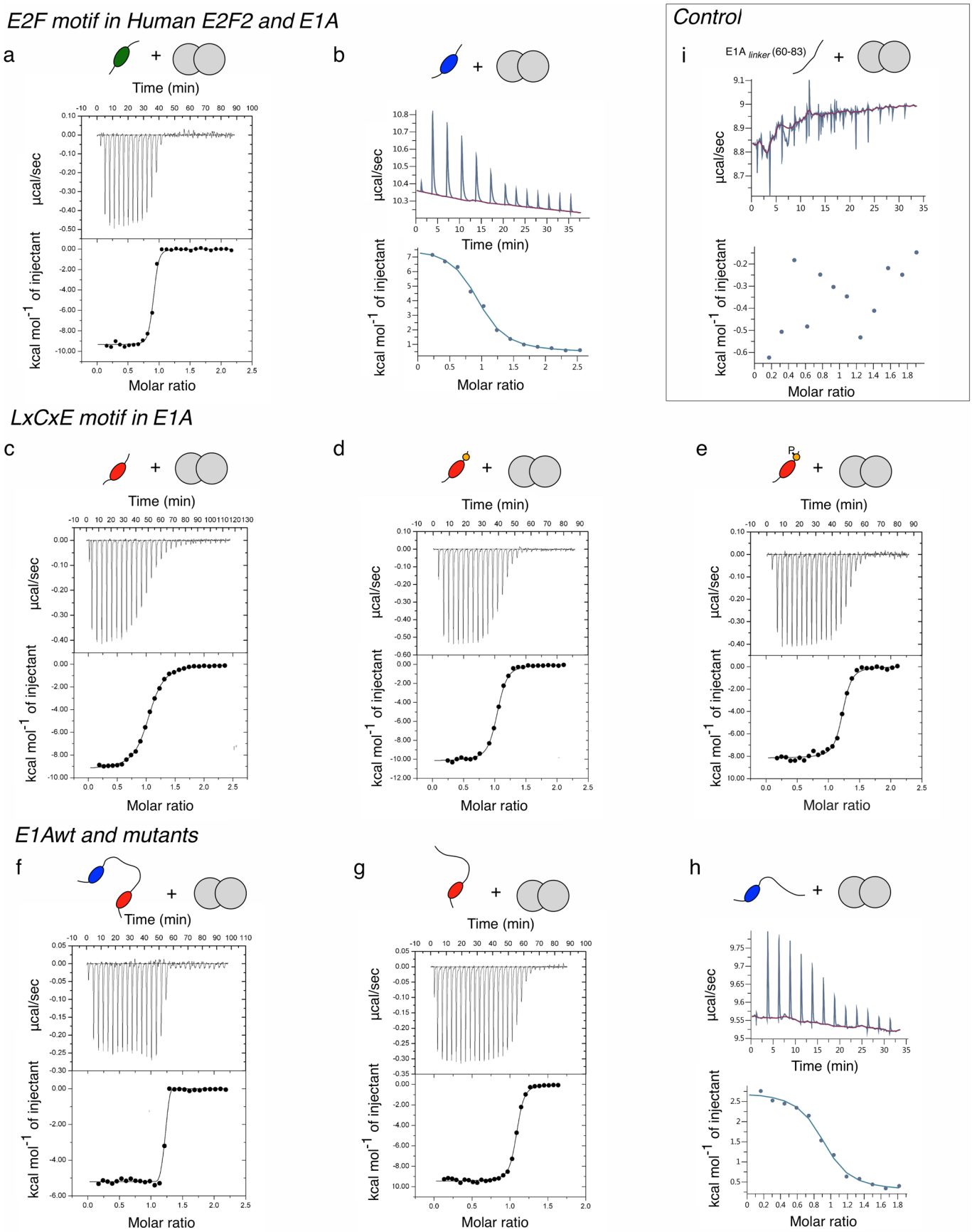
Correspondence and requests for materials should be addressed to Rohit V. Pappu, Alex S. Holehouse, Gary W. Daughdrill or Lucía B. Chemes.

Peer review information *Nature Structural and Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling editors: Anke Sparmann and Florian Ullrich, in collaboration with the *Nature Structural & Molecular Biology* team. Peer reviewer reports are available.



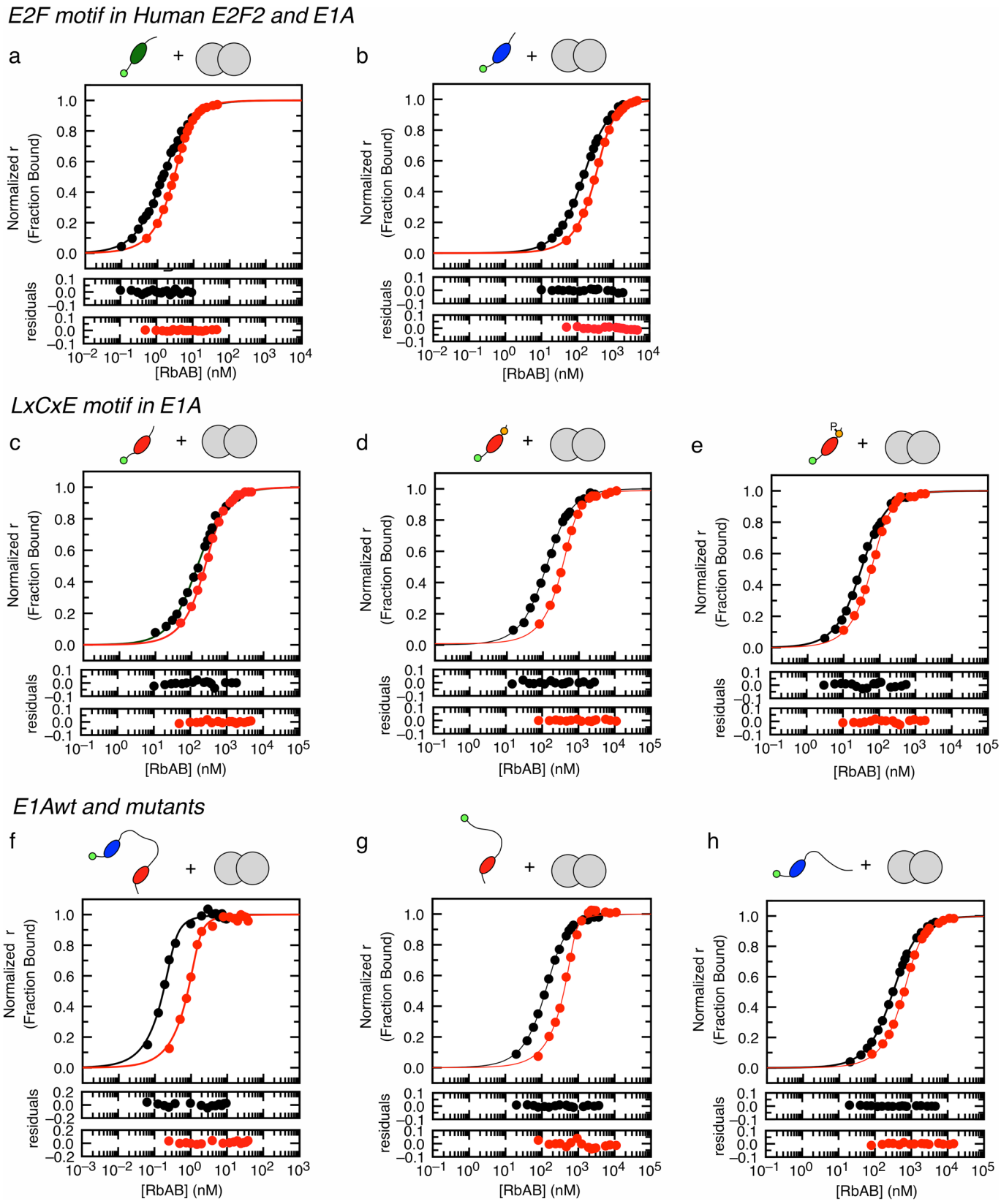
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Biophysical characterization of recombinant Rb and E1A proteins. a, Far UV-CD spectra of $E1A_{WT}$ (solid line), $E1A_{\Delta E}$ (dotted line), $E1A_{\Delta L}$ (dashed line). Inset: 15% SDS-PAGE gel of purified recombinant E1A proteins (purity > 90%). **b**, Far UV-CD spectrum of the Rb (RbAB domain). **c**, SEC-SLS experiments of $E1A_{WT}$ (solid line), $E1A_{\Delta E}$ (dotted line) and $E1A_{\Delta L}$ (dashed line). **d**, SEC-SLS experiment of Rb. For c) and d), black bars correspond to the elution volume of globular protein markers: BSA 66 kDa (1), MBP 45 kDa (2) and Lysozyme 14.3 kDa (3). Black line: SEC profile, red line: measurement of the molecular weight. **e**, 12.5% SDS-PAGE of MBP-E1A fusion protein variants. Gel1: Grafting of selected linkers from Human and Simian E1A proteins into the $E1A_{WT}$ construct containing the HAdV5 motifs. Types are: HAdV52, HAdV40, SAdV3, SAdV22, HAdV5, HAdV5*ΔHyd*, HAdV18, HAdV40-2x. Gel 2: Grafting of linkers from Bovine, Canine and Bat E1A proteins into the $E1A_{WT}$ sequence and endogenous variants carrying the cognate motifs for each species: BAdV2, BAdV2-ED, BAdV1, CAdV1, BtAdV2 and BtAdV2-ED. **f**, 17% SDS-PAGE of cleaved E1A protein variants: BAdV2, HAdV52, HAdV40, BtAdV2, HAdV5 and HAdV40-2x. **g**, Size exclusion chromatography experiment performed on a Superdex 200 column to determine R_h of cleaved E1A variants. Black bars correspond to V_o and $V_o + V_i$, and to the elution volume of globular protein markers: Gamma Globulin 150 kDa (1), Transferrin 80 kDa (2), BSA 66 kDa (3) MBP 45 kDa (4) and Trypsin Inhibitor 21 kDa (5). The E1A types are referenced to the names used in Fig. 4d.



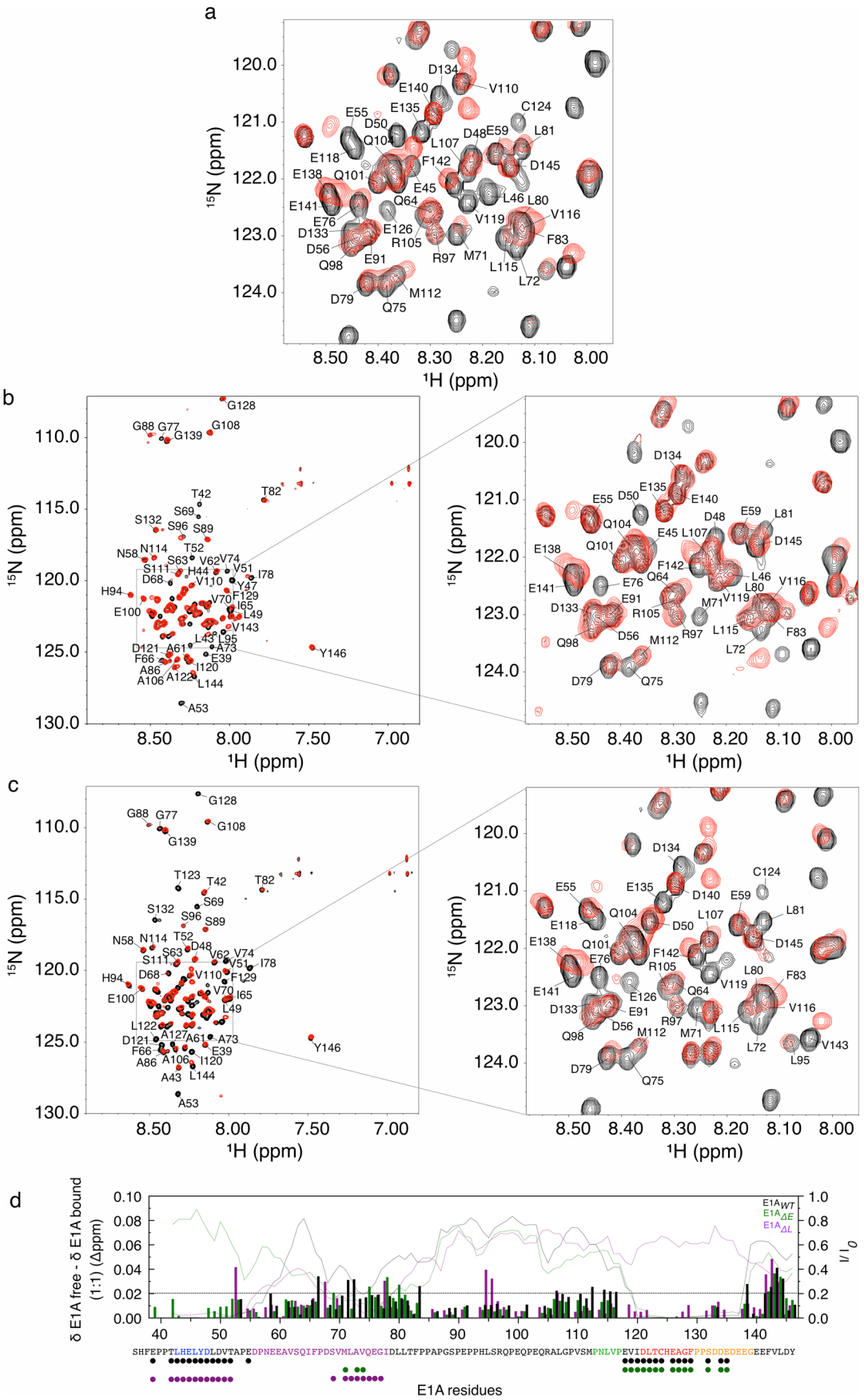
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Representative ITC binding isotherms for Rb:peptide/protein complexes. Measurements were performed loading the cell with Rb solution and the syringe with the different peptides or proteins as titrants. Panels show heat exchanged as a function of time (upper panel), and the enthalpy per mole of injectant plotted as a function of [peptide/protein]/[Rb] molar ratio (lower panel, black circles) and the corresponding fit using a single site binding model (lower panel, black lines). Binding traces here represented correspond to: **a**, Rb (5 μM) and Human E2F2 (50 μM); **b**, Rb (30 μM) and $E1A_{E2F}$ (300 μM); **c**, Rb (15 μM) and $E1A_{LxCxE}$ (150 μM); **d**, Rb (15 μM) and $E1A_{LxCxE\text{-}AC}$ (150 μM); **e**, Rb (15 μM) and $E1A_{LxCxE\text{-}ACP}$ (150 μM); **f**, Rb (15 μM) and $E1A_{WT}$ (150 μM); **g**, Rb (15 μM) and $E1A_{\Delta E}$ (150 μM); **h**, Rb (30 μM) and $E1A_{\Delta L}$ (300 μM). Thermodynamic parameters derived from the fitting are shown in Supplementary Table 1. Exothermic binding to Rb was observed for the Human E2F2 peptide and E1A peptides and protein fragments harboring the LxCxE motif, while $E1A_{E2F}$ and $E1A_{\Delta L}$ harboring only the E1A E2F motif clearly showed an endothermic behavior. **i)** ITC curve of a peptide corresponding to the TAZ2 region in the E1A linker (63-80) that showed intensity decreases in the NMR experiments (Fig. 2) binding to Rb. The titration was performed at 30 μM Rb and 300 μM E1A linker peptide at 20 °C. A schematic representation of each interacting pair is shown above the ITC traces: Rb (grey double circle) and each peptide/protein, where binding motifs are represented as follows: Human-E2F2 (green oval), E2F motif (blue oval), LxCxE motif (red oval), LxCxE acidic stretch (orange circle), phosphorylation (letter P).



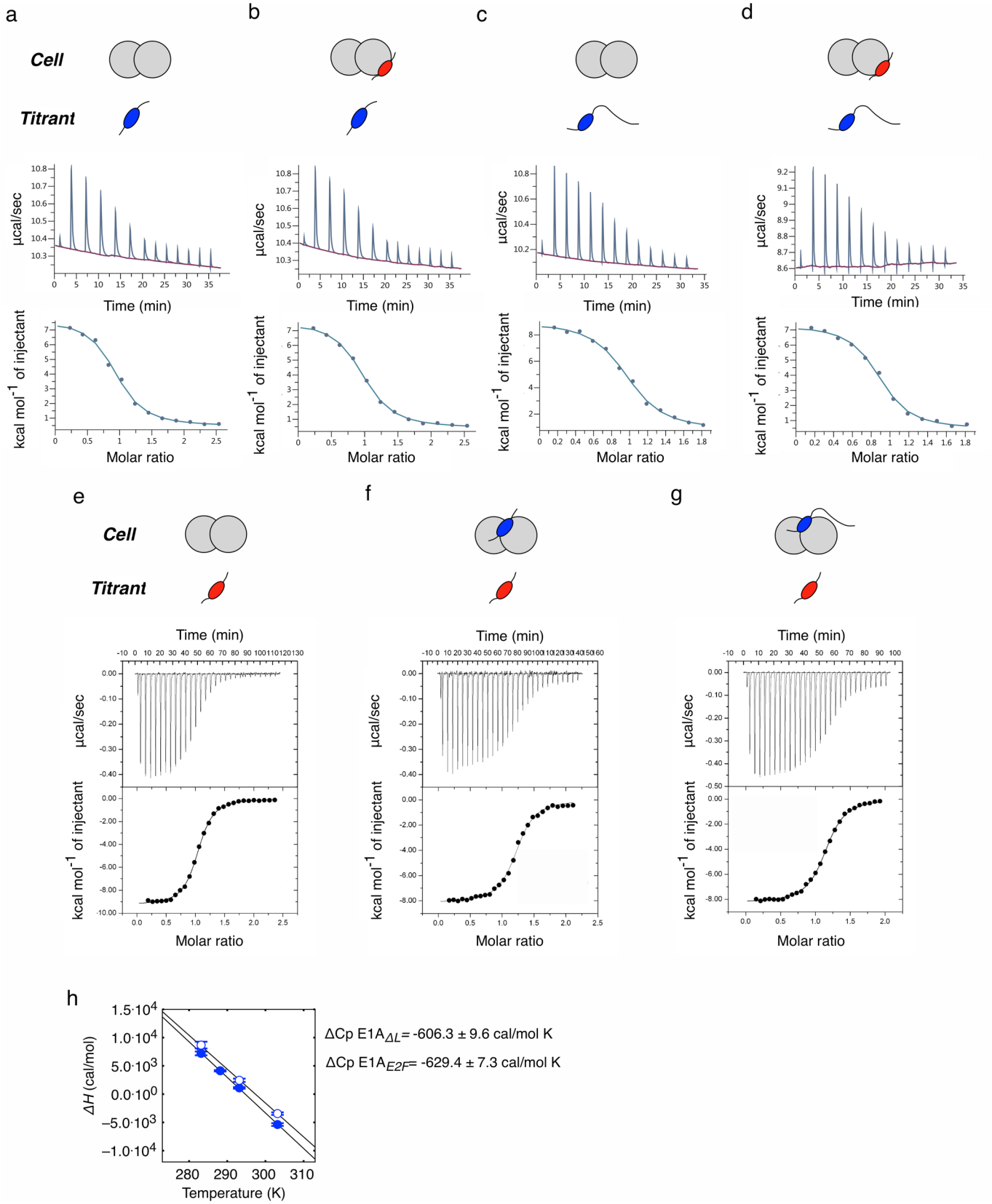
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Fluorescence Spectroscopy titration experiments of E1A-Rb and E2F-Rb interactions. Representative titration binding curves at equilibrium for each FITC-labeled peptide/protein-Rb interaction tested in this work. Normalized anisotropy signals (circles) are shown, along with the global fit to a 1:1 binding model (lines) that yielded the K_D value. The residuals for the fit are shown in the lower panels. Binding traces here represented correspond to two probe (FITC-labeled peptide/protein) concentrations: **a**, Human E2F2: 1 nM (black) and 5 nM (red); **b**, $E1A_{E2F}$: 100 nM (black) and 500 nM (red); **c**, $E1A_{LxCxE}$: 100 nM (black) and 500 nM (red); **d**, $E1A_{LxCxE\text{-}AC}$: 130 nM (black) and 700 nM (red); **e**, $E1A_{LxCxEACP}$: 30 nM (black) and 100 nM (red); **f**, $E1A_{WT}$: 0.5 nM (black) and 2 nM (red); **g**, $E1A_{\Delta E}$: 200 nM (black) and 800 nM (red); **h**, $E1A_{\Delta L}$: 200 nM (black) and 800 nM (red). The K_D values obtained by global fitting to a 1:1 model (Supplementary Data Table 1) were in excellent agreement with those obtained when fitting individual binding curves using non-normalized anisotropy or fluorescence data (Supplementary Table 2). A schematic representation of each interacting pair is shown above the binding traces: Rb (grey double circle); FITC-moiety at the N-terminus of the sequence (light green circle). Binding motifs are represented as follows: Human-E2F2 (green oval), E2F motif (blue oval), LxCxE motif (red oval), acidic stretch (orange circle), phosphorylation (letter P). The linker is represented by a black line.



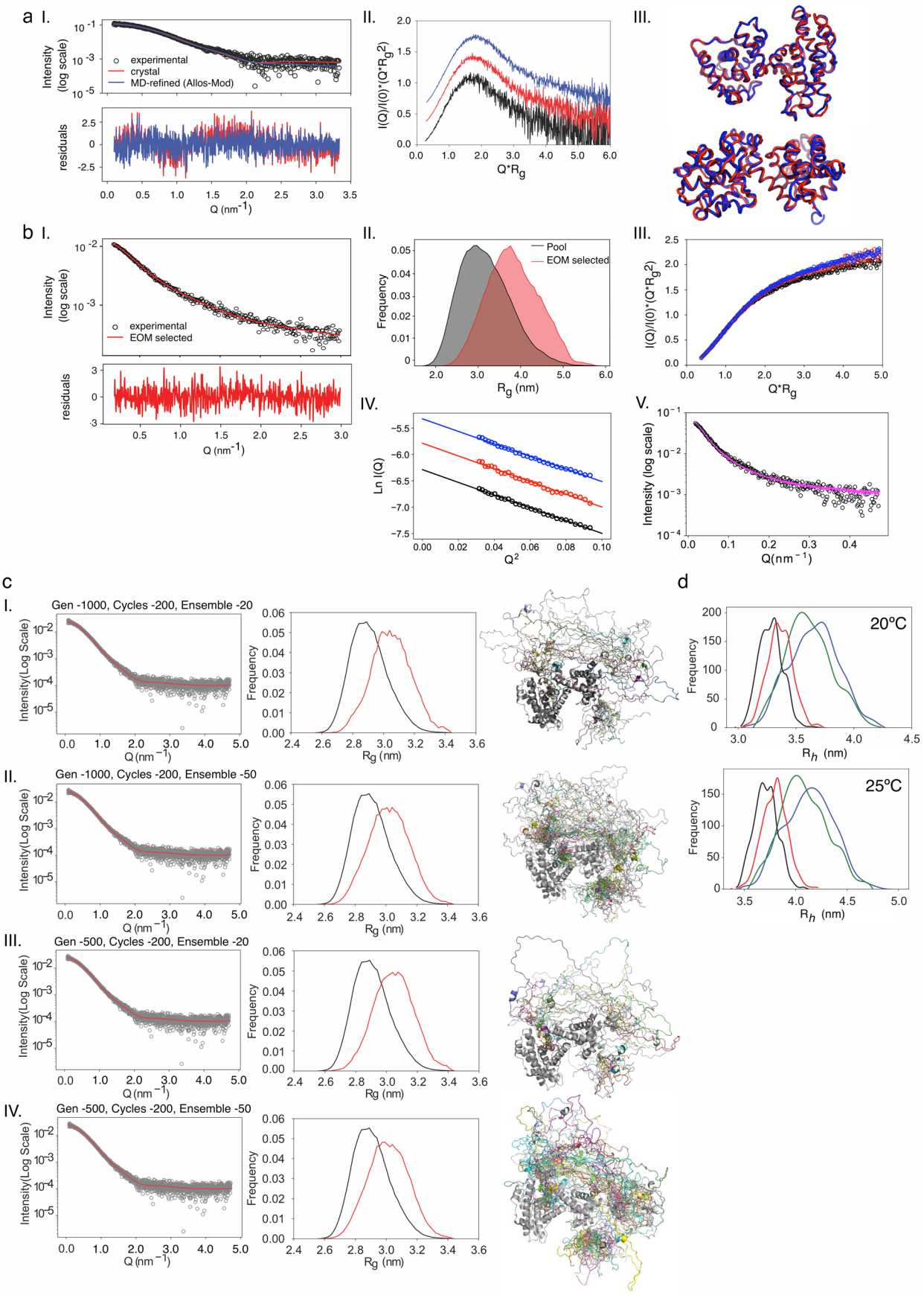
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | NMR experiments of [Rb:E1A] complexes. a, Central region of ^{1}H-^{15}N TROSY spectra of free ^{15}N-labeled E1A (black) and a 1:1 molar ratio complex of ^{15}N-labeled E1A and unlabeled Rb (red) at 525 μM, with assigned peaks of the free form indicated. The full spectrum of this complex is shown in Fig. 2a. **b**, Left panel: Overlay of the ^{1}H-^{15}N TROSY spectra of free ^{15}N-labeled E1A$_{\Delta L}$ (black) and a 1:1 molar ratio complex of ^{15}N -labeled E1A$_{\Delta L}$ and unlabeled Rb (red) at 315 μM. Right panel: central region of the spectra with assigned peaks of the free form indicated **c**, Left panel: Overlay of the ^{1}H-^{15}N TROSY spectra of free ^{15}N-labeled E1A$_{\Delta E}$ (black) and a 1:1 molar ratio complex of ^{15}N-labeled E1A$_{\Delta E}$ and unlabeled Rb (red) at 315 μM. Right panel: central region of the spectra with assigned peaks of the free form indicated. The low chemical shift dispersions in the ^{1}H dimension for E1A$_{\Delta L}$ and E1A$_{\Delta E}$ denote their disordered nature, like that seen in E1A. There is no change in peak dispersion upon binding with Rb, indicating that linker regions of the E1A$_{\Delta L}$ and E1A$_{\Delta E}$ mutants remain largely disordered in the [E1A$_{\Delta L}$:Rb] and [E1A$_{\Delta E}$:Rb] complexes. **d**, Plot of chemical shift changes upon binding as a function of residue number for E1A$_{WT}$, E1A$_{\Delta L}$ and E1A$_{\Delta E}$. Dashed line at 0.2 ppm corresponds to the digital resolution of the experiment. The small chemical shift changes for almost all of the linker residues suggest very little if no interaction with Rb. I/I_0 ratio is overlaid for comparison (colored lines). Dots on the bottom correspond to the residues of each variant whose ^{1}H-^{15}N intensities in the bound state is = 0, so the chemical shift changes could not be measured.



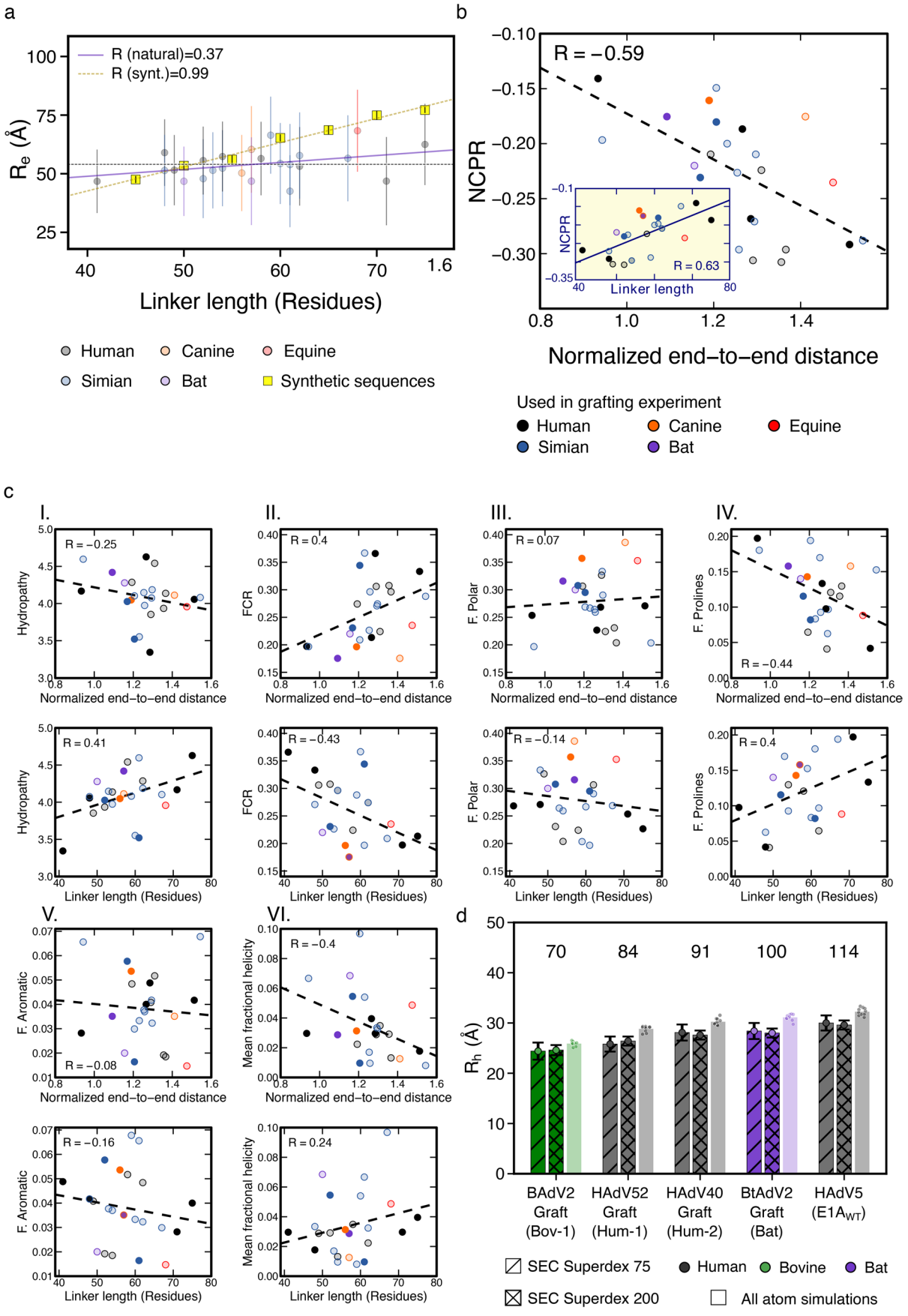
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Analysis of allosteric effects in the formation of the Rb-E1A complex. Measurements were performed by loading the cell with Rb or with a pre-assembled complex of Rb with peptide/proteins containing one of the interacting motifs and titrating with peptide/proteins containing the complementary motif loaded into the syringe. Panels show heat exchanged as a function of time, (upper panel) and the enthalpy per mole of injectant plotted as a function of [peptide or protein]/[Rb] molar ratio (Lower panel, black circles) along with the corresponding fit using a single site binding model (Lower panel, black lines). Binding traces correspond to: **a**, Rb (30 μM, cell) titrated with $E1A_{E2F}$ (300 μM, syringe) at 10 °C; **b**, [$E1A_{LxCxE}$:Rb] (30 μM, cell) titrated with $E1A_{E2F}$ (300 μM, syringe) at 10 °C; **c**, Rb (30 μM, cell) titrated with $E1A_{\Delta L}$(300 μM, syringe) at 10 °C; **d**, [$E1A_{LxCxE}$:Rb] (30 μM, cell) titrated with $E1A_{\Delta L}$ (300 μM, syringe) at 10 °C; **e**, Rb (15 μM, cell) titrated with $E1A_{LxCxE}$ (150 μM, syringe) at 20 °C; **f**, [$E1A_{E2F}$:Rb] (15 μM, cell) titrated with $E1A_{LxCxE}$ (150 μM, syringe) at 20 °C; **g**, [$E1A_{\Delta L}$:Rb] (15 μM, cell) titrated with $E1A_{LxCxE}$ (150 μM, syringe) at 20 °C. Thermodynamic parameters derived from the fitting are shown in Supplementary Table 1. A schematic representation of each titration design is shown above the ITC traces: Rb: grey double circle, E2F motif: blue oval, LxCxE motif: red oval. The E1A linker is depicted as a black line. **h**, ITC measurements of $E1A_{E2F}$ and $E1A_{\Delta L}$ at different temperatures. The heat capacity change (ΔCp) was calculated from the slope of the plot of ΔH vs temperature. $E1A_{E2F}$: filled blue circles; $E1A_{\Delta L}$: open blue circles. Thermodynamic parameters are reported in Supplementary Data Table 5.



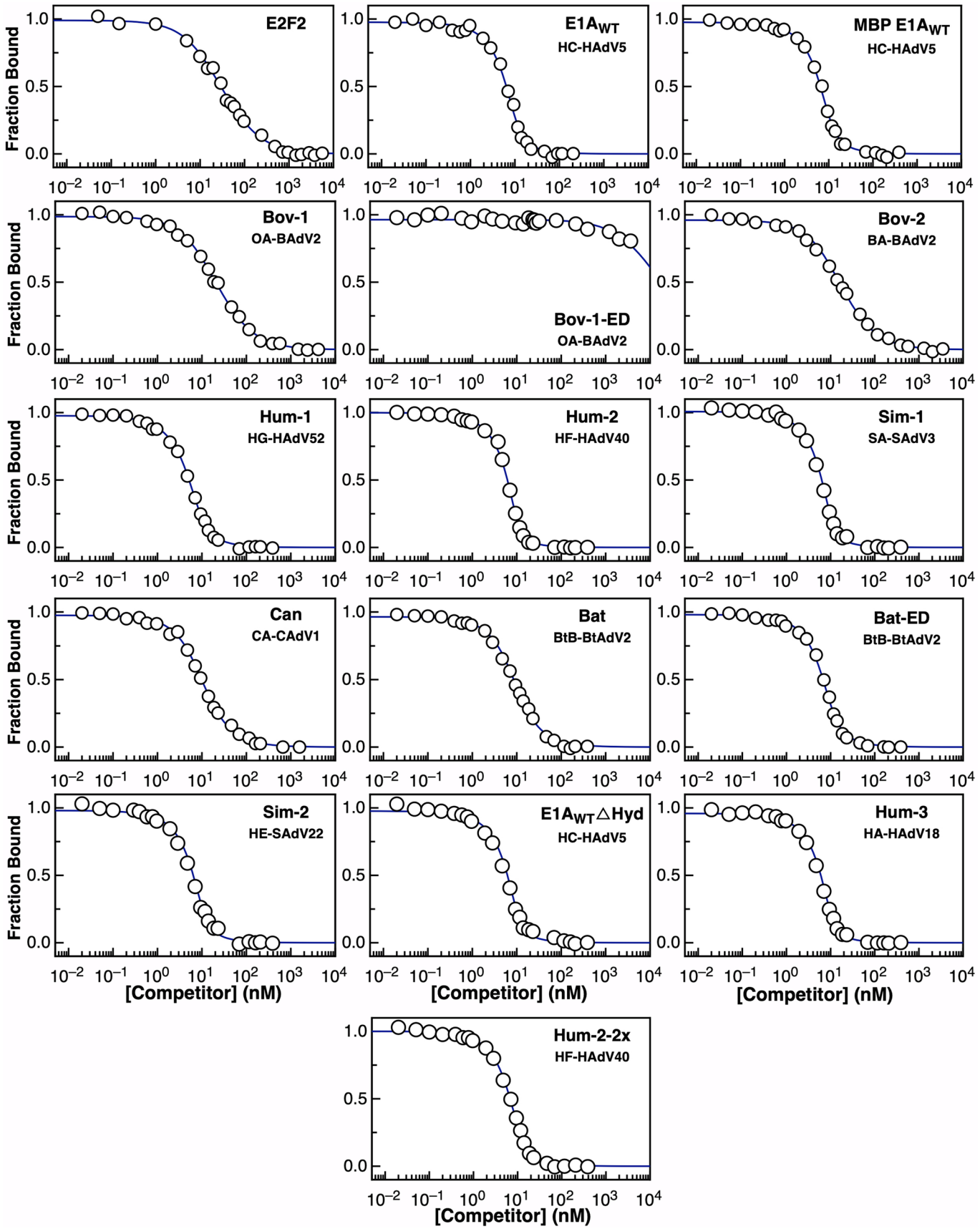
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | SAXS analysis of Rb, E1A and the [E1A$_{WT}$:Rb] complex. a, I. Experimental SAXS intensity profile (black empty circles) versus theoretical profiles obtained from the crystal structure of the unliganded RbAB domain (PDB ID: 3POM) (red line) or a refined model where flexible loops were added (Allos-Mod-FoXS, blue line). Residuals are shown below the fits. II. Kratky plots of Rb at 4.0 mg/ml (blue line), 2.0 mg/ml (red line) and 1.0 mg/ml (black line). III. Orthogonal views of the RbAB crystal structure (red) and optimized model (blue) (RMSD = 1.7 Å). **b**, I. SAXS intensity profile of E1A$_{WT}$ (black circles) and the best fit from the EOM method (red line). Below, residual of the fit. II. R_g distribution of the E1A$_{WT}$ ensemble pool (black area) and EOM-selected ensemble (red area). III-IV. Kratky plots (III) or Guinier plots (IV) of E1A$_{WT}$ at 7.0 mg/ml (blue empty circles), 5.6 mg/ml (red empty circles) and 4.2 mg/ml (black empty circles). V. Overlay of SEC-SAXS profile of E1A$_{WT}$ (blue empty circles) and the merged curve from SAXS experiments at three concentrations (pink line). **c**, Theoretical SAXS profiles computed for a pool of 10250 [E1A$_{WT}$:Rb] structures compared to experimental SAXS profiles and EOM fitting. Four fitting conditions are shown: I. 1000 generations with ensemble size N = 20, II. 1000 generations with N = 50, III. 500 generations with N = 20 and IV. 500 generations with N = 50. Left: experimental SAXS intensity profiles (grey circles) and EOM fitting (red lines). Middle: R_g distributions of pool ensembles (black line) and EOM-selected sub-ensembles (red line). Right: EOM-selected sub-ensembles. Fitting condition II is presented in Fig. 3. **d**, Calculated R_h for [E1A$_{WT}$:Rb] (black) [E1A$_{\Delta E}$:Rb] (green) and [E1A$_{\Delta L}$:Rb] (blue) pool ensembles and the EOM-selected [E1A$_{WT}$:Rb] sub-ensemble (red).



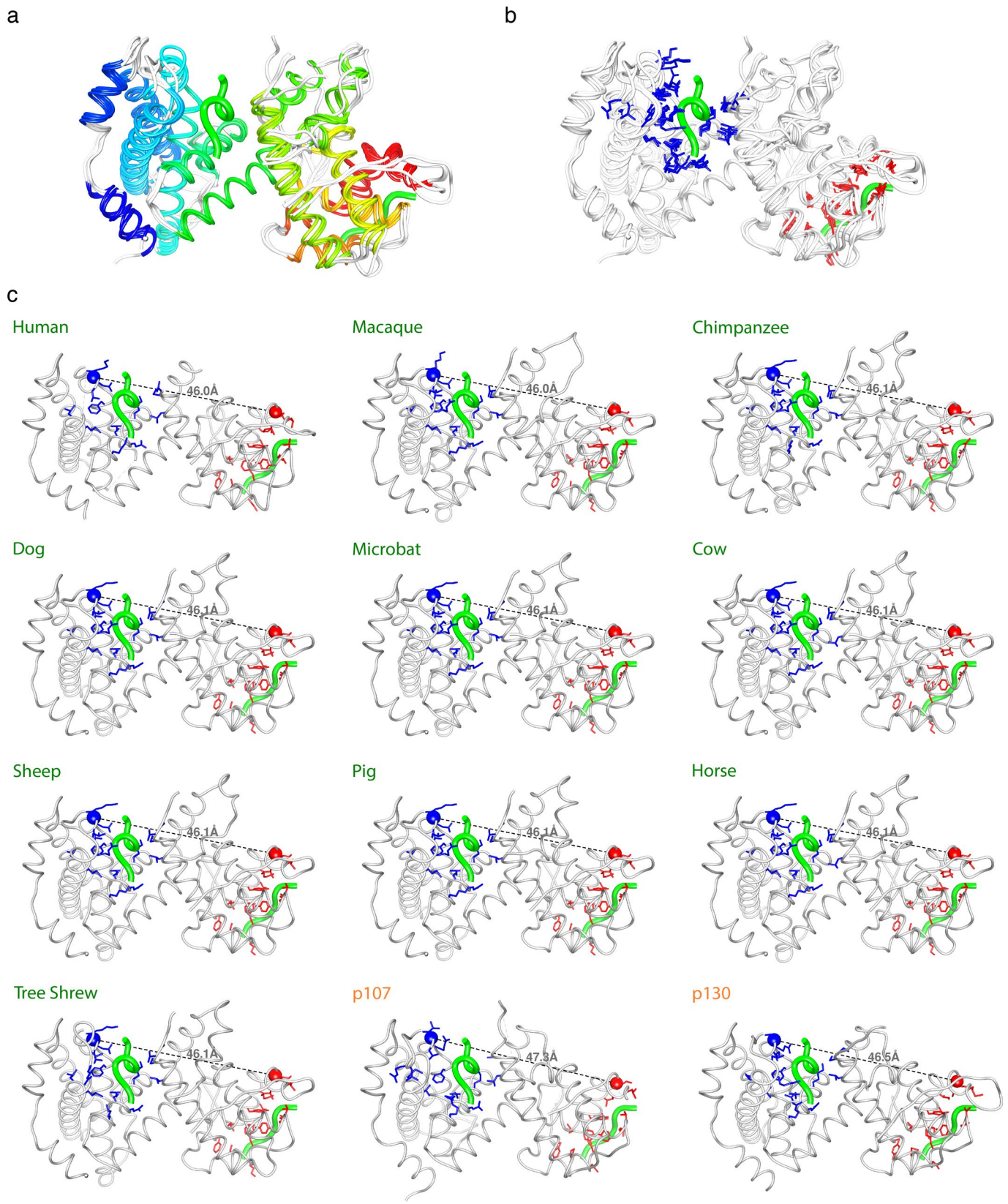
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Correlation of E1A linker dimensions with sequence-encoded features. a, Linker length control titration experiment. End-to-end distance (R_e) of natural sequences (colored circles) compared to synthetic sequences of varying length and constant sequence composition matching the HF_HAdV40 linker (yellow squares). Natural sequences: n = 15 independent simulations were run for each sequence, points represent the mean R_e value and error bars represent the standard deviation over the population obtained from the total ensemble from 15 simulations. Synthetic sequences: n = 20 random permutations were generated for each length and simulated under equivalent conditions. The mean R_e value (yellow square) is a double average over both conformational space and sequence space. Lines within the yellow squares represent the standard error of the mean across all simulations of a given length, shown to confirm that all random permutations have very similar R_e values. **b**, Net-charge per residue (NCPR) as a function of normalized end-to-end distance for the 27 linkers of Fig. 4a. Inset: NCPR as a function of linker length. Sequences used in the grafting experiment are shown as solid circles and the rest as transparent circles. R = Pearson's correlation coefficient. **c**, Correlation between distinct sequence parameters and normalized end-to-end distance (upper panels) or linker length (lower panels) (Supplementary Text 1). R = Pearson's correlation coefficient. Most R values are < 0.3 with several exceptions. **d**, Hydrodynamic radius (R_h) for motif-linker-motif constructs of five cleaved E1A variants (shown in Extended Data Fig. 1f,g). The length of each construct is indicated above each bar. R_h was determined from size exclusion chromatography run on Superdex 75 (n = 1, striped colored bars) or Superdex 200 (n = 1, cross-hatched colored bars). The height of each bar indicates the estimated R_h value and the error bars represent the standard deviation obtained from interpolation in the –logMW vs K_{av} calibration curve (see **Methods**). R_h was also predicted from all-atom simulations (colored bars). The height of each bar represents the mean R_h value from ten independent simulations of each construct (n = 10), while each individual marker is the mean of each independent simulation.



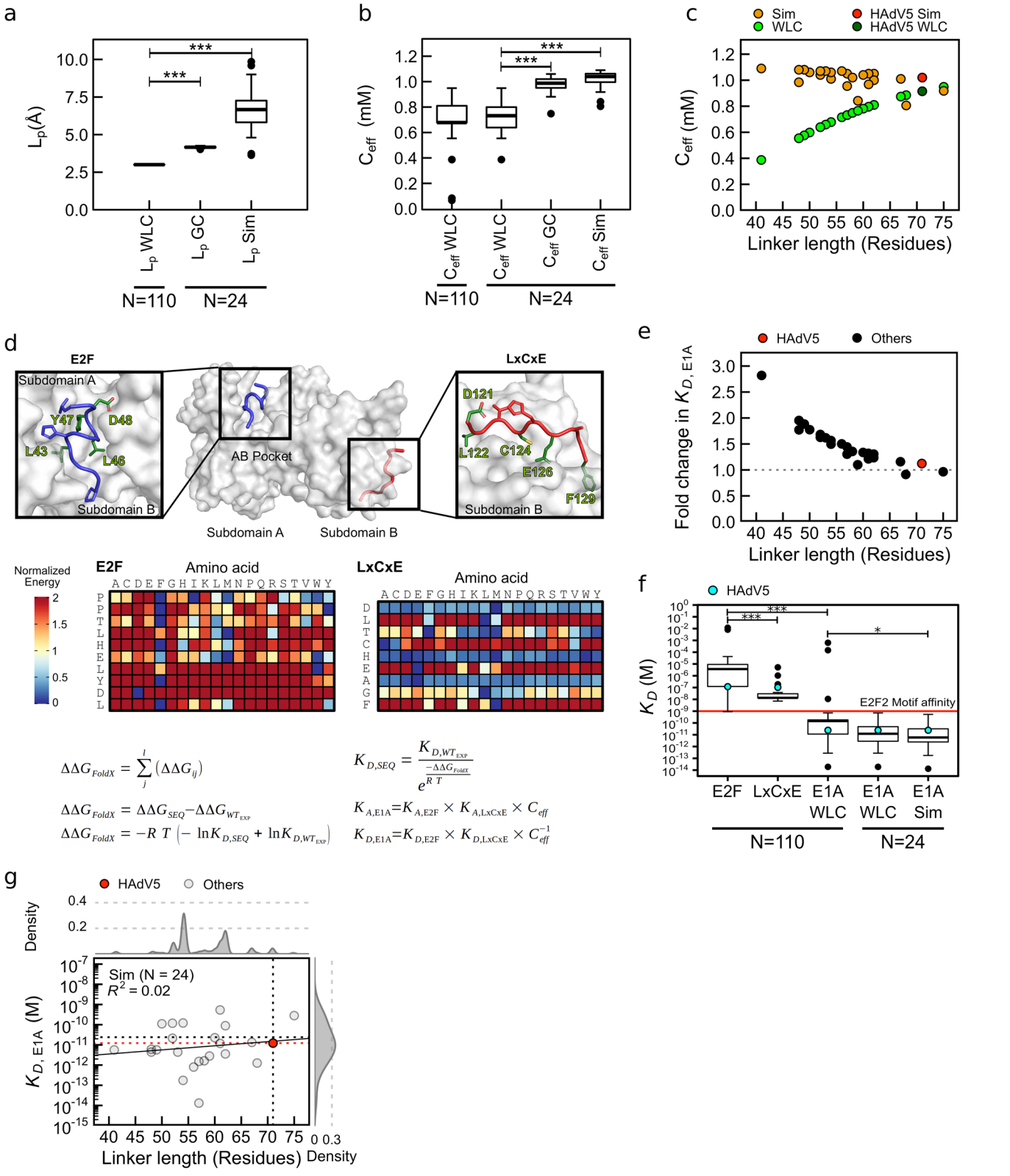
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | E2F displacement ability and Rb-binding affinity of E1A variants. Competition displacement curves were performed by competing a preassembled equimolar [FITC-E2F2:Rb] complex at 10 nM concentration with increasing concentrations of each variant. One representative example is shown for each variant reported on Supplementary Table 7. The displacement reaction was followed by recording the fluorescence anisotropy of the FITC moiety, with excitation at 490 nm and emission at 520 nm. In every case except for Bov-1-ED, the E1A variants were able to displace FITC-E2F2 from binding to Rb. The anisotropy value of free FITC-E2F2 was 0.042 ± 0.002 and the anisotropy value of the [FITC-E2F2:Rb] complex was 0.14 ± 0.01. In every case, the anisotropy value obtained at the end of the titration was equal to the anisotropy value of the free FITC-E2F2 peptide, confirming the complete displacement of FITC-E2F2. The anisotropy values were normalized to calculate the fraction of Rb-bound FITC-E2F2 and fitted to estimate the K_D value for the [Variant:Rb] complex.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Conservation of pocket domain structure and linear motif binding sites across mammalian pocket proteins. **a**, Structural conservation of the pocket domain across mammalian pocket proteins. The human Rb pocket domain (PDB:1GUX) is shown aligned with 9 structural models of Rb pocket domains from representative mammalian species plus the human paralogs p107 (PDB:4YOZ) and p130. The models of the Rb pocket domains and p130 were obtained by using Alphafold2 implemented in ColabFold (See Methods). Secondary structure is depicted in rainbow colors. The E2F (left) and LxCxE (right) motifs are depicted as green ribbons (PDB 2R7G and 1GUX respectively). **b**, Structural conservation of the E2F and LxCxE clefts in pocket proteins. Structural alignment shown in panel A with the residues that mediate binding to the E2F and LxCxE motifs (marked as asterisks in Supplementary Fig. 1) depicted as blue and red sticks respectively. **c**, The distance between the E2F and LxCxE binding sites is highly conserved across mammalian pocket proteins. The spacing was measured between the C-terminal anchor site of the E2F cleft (blue sphere) and the N-terminal anchor site of the LxCxE cleft (red sphere). Distances are: 46.0 Å (human, macaque and chicken), 46.1 Å (chimpanzee, dog, microbat, cow, sheep, pig, horse and tree shrew), 47.3 Å (p107) and 46.5 Å (p130). These distances are slightly shorter than the distance between binding sites used in the C_{eff} calculations ($r_0 = 49$ Å), which was measured between the C-terminal residue of the E2F motif and the N-terminal residue of the LxCxE motif using the structures of the motifs bound to Rb (PDB: 2R7G and 1GUX).



$$\Delta\Delta G_{FoldX} = \sum_{j}^{l} (\Delta\Delta G_{ij})$$

$$\Delta\Delta G_{FoldX} = \Delta\Delta G_{SEQ} - \Delta\Delta G_{WT_{EXP}}$$

$$\Delta\Delta G_{FoldX} = -R\,T\,(-\ln K_{D,SEQ} + \ln K_{D,WT_{EXP}})$$

$$K_{D,SEQ} = \frac{K_{D,WT_{EXP}}}{e^{\frac{-\Delta\Delta G_{FoldX}}{R\,T}}}$$

$$K_{A,E1A} = K_{A,E2F} \times K_{A,LxCxE} \times C_{eff}$$

$$K_{D,E1A} = K_{D,E2F} \times K_{D,LxCxE} \times C_{eff}^{-1}$$

Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Global prediction of E1A-Rb binding affinity. a,**b**, L_p and C_{eff} values for E1A linkers. Boxplots: center line represents the median, lower and upper bounds represent the first and third quartiles and upper and lower whiskers extend from the top and bottom of the box by 1.4 the interquartile range. Black dots: outliers. p-values were calculated using a two-sided permutation test (10000 permutations) and the Benjamini-Hochberg correction for multiple comparisons to control the false discovery rate. ***p-value < 0.001 (detection limit of the test). N = 110: All E1A linkers, N = 24: Simulated linkers. **c**, C_{eff} as a function of linker length for 24 linkers calculated using the WLC model (L_p = 3 Å) (green dots), or L_p values from all atom simulations (L_p Sim, orange dots). Dark green/red dots: $E1A_{WT}$. **d**, Upper panel: E2F (blue) and LxCxE (red) motifs From E1A bound to Rb. Green sticks: core residues, blue/red sticks: variable residues. Lower panel: FoldX energy matrices with energy normalized in the range 0-2 kcal/mol. **e**, Fold-change in affinity ($K_{D,E1A}$ (L_p = 3 Å) / $K_{D,E1A}$ (L_p Sim)) using naïve versus simulated L_p. Red dot: $E1A_{WT}$. **f**, Predicted K_D for the $E1A_{E2F}$ and $E1A_{LXCXE}$ SLiMs and for the motif-linker-motif construct for 110 sequences (E1A WLC) and for 24 simulated sequences using L_p = 3 Å (K_D WLC) or sequence-specific L_p from the simulations (E1A Sim). Boxplot elements and p-values are defined as in panel a. Cyan dots: experimental value for $E1A_{WT}$. Red line: E2F2 motif affinity. **g**, Global Rb binding affinity ($K_{D,E1A}$) as a function of linker length for 24 sequences using the L_pSim values. $K_{D,E1A} = K_{D,E2F} \cdot K_{D,LxCxE} \cdot C_{eff}^{-1}$. The low R^2 value indicates that $K_{D,E1A}$ is uncorrelated to linker length. Upper panel: density plot of linker length for 107 E1A linkers (three short linkers were excluded). Right panel: density plot of $K_{D,E1A}$. Red dot/line: Predicted $K_{D,E1A}$ for HAdV5 ($E1A_{WT}$). Grey cross line: experimental $K_{D,E1A}$ for $E1A_{WT}$.

nature portfolio

Corresponding author(s): Chemes Lucia B., Daughdrill Gary W., Holehouse Alex S. and Pappu Rohit V.

Last updated by author(s): Mar 10, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Conformational ensemble models of E1AWT:Rb were generated using algorithms implemented in the MoMA software suite (no version number) (https://moma.laas.fr). The loop sampling method used to model the linker between the two binding motifs can be used via a web server (https://moma.laas.fr/applications/LoopSampler/), and binaries are available upon request. Unfiltered conformational ensembles for the [E1AWT:Rb], [E1AΔL:Rb] and [E1AΔE:Rb] complexes are available at (https://moma.laas.fr/data/) under the description "Conformational ensemble models of the IDP E1A bound to Rb protein". E1A linker simulations were generated using the CAMPARI simulation engine Version 2.0 (http://campari.sourceforge.net) and ABSINTH implicit solvent model ABS-OPLS3.2. All code used to generate the E1A linker simulations is provided at (https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021) and trajectories are deposited on Zenodo (https://zenodo.org/record/6332925). All published methods used in data generation are referenced in the Materials and Methods section

Data analysis

Profit 7.0 software (Quantumsoft, Switzerland) was used for data fitting (https://quansoft.com). DLS data was analyzed with the Dynamics 6.1 software. NMR data was analyzed with NMRPipe (https://www.ibbr.umd.edu/nmrpipe/) and NMRViewJ (https://nmrfx.org/nmrfx/nmrviewj). FOXS (no version number) (https://modbase.compbio.ucsf.edu/foxs/), AllosMod-FoXS (no version number) (https://modbase.compbio.ucsf.edu/allosmod-foxs/), EOM (version 2.1), Crysol (version 2.8.4) HydroPro v10, SASFLOW 3.0, FOXTROT 3.5.2, ATSAS suite program version 2.8.4 and CHROMIX (part of ATSAS v2.8.4) were used for the analyses of SAXS-derived data. E1A linker simulations were analyzed using the MDTraj package version 1.9.5 and SOURSOP v0.1.3 (https://soursop.readthedocs.io/) and sequences were analyzed using the localCIDER 0.1.19 software package. All code used to analyze the E1A linker simulations and scripts used to analyze trajectories are provided at (https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021). Sequence analysis and structure-based predictions were performed with the FlexPepDock webserver (http://flexpepdock.furmanlab.cs.huji.ac.il/index.php), FoldX v5.0, Jalview v2.11, UCSF Chimera v1.5, IUPRED 2a, MUSCLE v3.8.1551, CONSURF 2016, Alphafold v2.0, MMseqs2, and ColabFold v1.0. PDB codes used for data analysis and structural predictions are: PDB: 3POM, PDB: 1GUX, PDB: 2R7G, PDB: 4YOZ. All published methods used in data analysis are referenced in the Materials and Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

SAXS raw data for Rb, E1AWT and the [E1AWT:Rb] complex has been deposited in SASDB (https://www.sasbdb.org) with codes SASDNK6 (Rb 1mg/ml), SASDNL6 (Rb 2mg/ml), SASDNM6 (Rb 4mg/ml), SASDNN6 (E1AWT 4.2mg/ml), SASDNP6 (E1AWT 5.6mg/ml), SASDNQ6 (E1AWT 7.0mg/ml), SASDNR6 ([E1AWT:Rb] 0.7mg/ml), SASDNS6 ([E1AWT:Rb] 1.4mg/ml), SASDNT6 ([E1AWT:Rb] 2.7mg/ml), SASDNU6 ([E1AWT:Rb] merged data), SASDNV6 (E1AWT, SEC-SAXS). Refined conformational ensemble models for E1AWT and [E1AWT:Rb] have been deposited in the Protein Ensemble Database (https://proteinensemble.org/P03255) with codes PED00175 (E1AWT) and PED00174 ([E1AWT:Rb]). Unfiltered conformational ensembles for the [E1AWT:Rb], [E1AΔL:Rb] and [E1AΔE:Rb] complexes are available at (https://moma.laas.fr/data/) under the description "Conformational ensemble models of the IDP E1A bound to Rb protein". NMR assignments of backbone resonances for E1AWT, E1AΔE and E1AΔL are provided in Supplementary Data File 2. Trajectories for all E1A linker ensembles are provided at: Zenodo (https://zenodo.org/record/6332925), and trajectory analysis results are provided at: https://github.com/holehouse-lab/supportingdata/tree/master/2021/Gonzalez_Foutel_2021. PDB codes used in data analysis and prediction are: PDB:1GUX, PDB:3POM, PDB:2R7G, PDB:4YOZ. Raw data underlying Main Figures 1-5 and Extended Data Figures 1, 4, 7 and 10 are available as Source Data Files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Our biological samples are recombinant proteins. We use recombinantly expressed and purified (>95% purity) proteins which do not have high sample-to-sample variability under our stringent criteria for analysis of stock purity and stability (FarUV-CD, SEC, SDS-PAGE). For this reason, we use different protein stocks across the work but don't perform controlled replications, as this is not a large source of variance. Therefore, in all our experiments the sample size equals the number of independent titrations/NMR experiments/SAXS measurements performed, i.e. our replicates test for the reliability of the instrument and the experimental error in the performance of the different methods. However, we note that for both Kd estimation and Rg/Rh estimation, we used several orthogonal techniques to derive the parameters, obtaining a very good agreement across techniques, which supports the robustness of the parameter estimation.

For affinity data (fluorescence polarization and ITC), we performed three to five independent titrations for each construct, except for very few cases which are reported in the Supplementary Tables and Figure legends. Each titration is composed of 13-22 data points, each one of which is an individual binding reaction. The number of points in each titration determines the precision of the measurements and is chosen such that the estimation of Kd parameter has an average SD value <15% (See Supplementary Tables). The number of independent titrations performed for ITC and fluorescence titrations was determined to allow estimating the average Kd values with SD < 20%. This sample size is also chosen, as a standard in the field, such that it allows to distinguish a 2-fold (100%) change in the Kd value reliably. For DLS experiments three measurements were performed and allowed to estimate average Rh values with SD < 5%. SEC experiments were performed once since they were time consuming and involved calibration with 4-6 protein standards, but the Rh values derived from SEC are in very good agreement across the orthogonal DLS technique (SD < 15%). For NMR experiments, each protein was measured from 15N labeled and as 15N-13C labeled samples, and the assignments were performed from the 15N-13C labeled samples after verifying the 15N and 15N/13C spectra were similar. Provided the single and double labeled samples have similar spectra, the measurements are taken from one sample, following the standard in the field. For SAXS experiments, three independent runs were performed at three concentrations and the final SAXS profile was built from the lowest and highest concentrations. This method allows to test the replicability of the derived parameters, which was very good. EOM fitting was performed for 200 independent sub-ensembles containing 20 or 50 conformations taken from a pool data composed of 10250 structures. This number of sub-ensembles ensured the robustness of the results over multiple sub-ensembles of the data.

	For E1A linker simulations 15 (E1A linkers) or 10 (motif-linker-motif constructs) or 20 (length titration) independent simulations were performed (see methods for details), and parameters (Re, Rh) were obtained for each simulation and then averaged over all the simulations. The large number of independent replicates ensures an extensive sampling of the conformational space.
Data exclusions	There were no data exclusions.
Replication	Fluorescence polarization (FP) titrations were replicated 3-5 times. ITC titrations were replicated 3 times (except if noted differently in the Supplementary Tables). For ITC 28 or 13 injections were used for VP-ITC and PEAQ-ITC respectively and three independent binding isotherms were fit individually and averaged to obtain the mean Kd value ± SD. For fluorescence titrations 16-22 individual data points per titration were used and three to five binding curves were fit globally to obtain the estimated Kd ± SD. Far UV-CD experiments are reported from a single measurement, but they could be replicated with very low (<10%) variance. DLS experiments were replicated three times to obtain Rh values. SEC-SLS measurements were performed once and Rh was derived from a single run on a size exclusion column calibrated with 4-6 protein standards. In several cases a second sample was run, and the measurements agreed with a variance <15%. NMR experiments were performed once but the agreement of 15N and 15N/13C spectra from two independent samples was confirmed for all E1A proteins. SAXS experiments were performed three times at different sample concentrations for all proteins/complexes. For linker simulations 15 (E1A linker), 20 (length titration) or 10 (motif-linker-motif) independent simulations were performed. All attempts at replication of the experiments were successful, and the Rh/Rg and Kd parameters obtained by at least two complementary techniques were in very good agreement, indicating the robustness of the calculated parameters.
Randomization	Randomization is not applicable to our studies, since we did not perform any clinical or preclinical studies involving human or animal subjects or any studies of efficacy of a treatment or drug that required assignment of subjects to groups. All our work involves purified recombinant protein samples.
Blinding	Blinding is not applicable to our studies, since we did not perform any clinical or preclinical studies involving human or animal subjects or any studies of efficacy of a treatment or drug that required blinding of the subjects undergoing the treatment, or of the researchers analyzing the data. All our work involves purified recombinant protein samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging