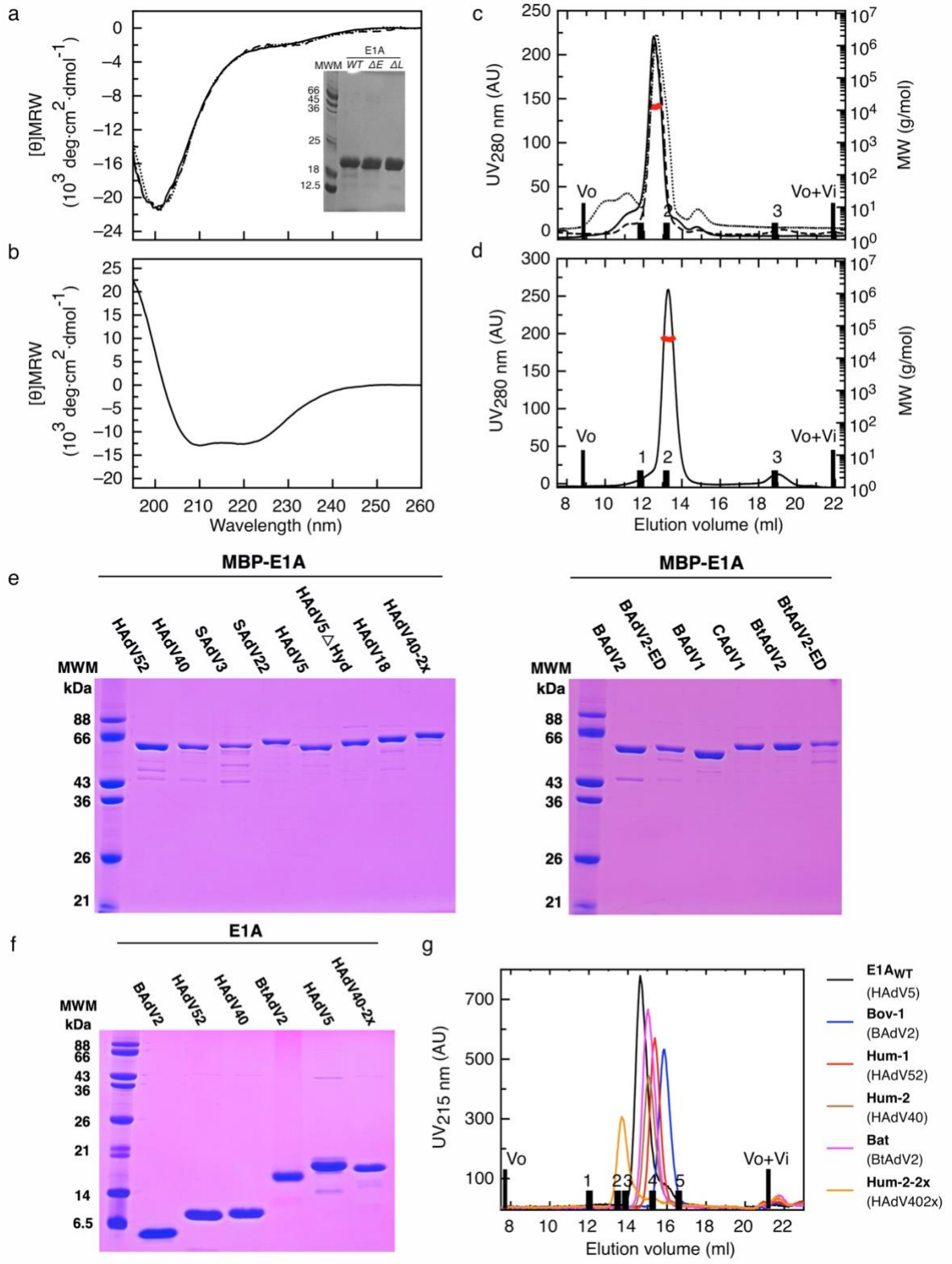
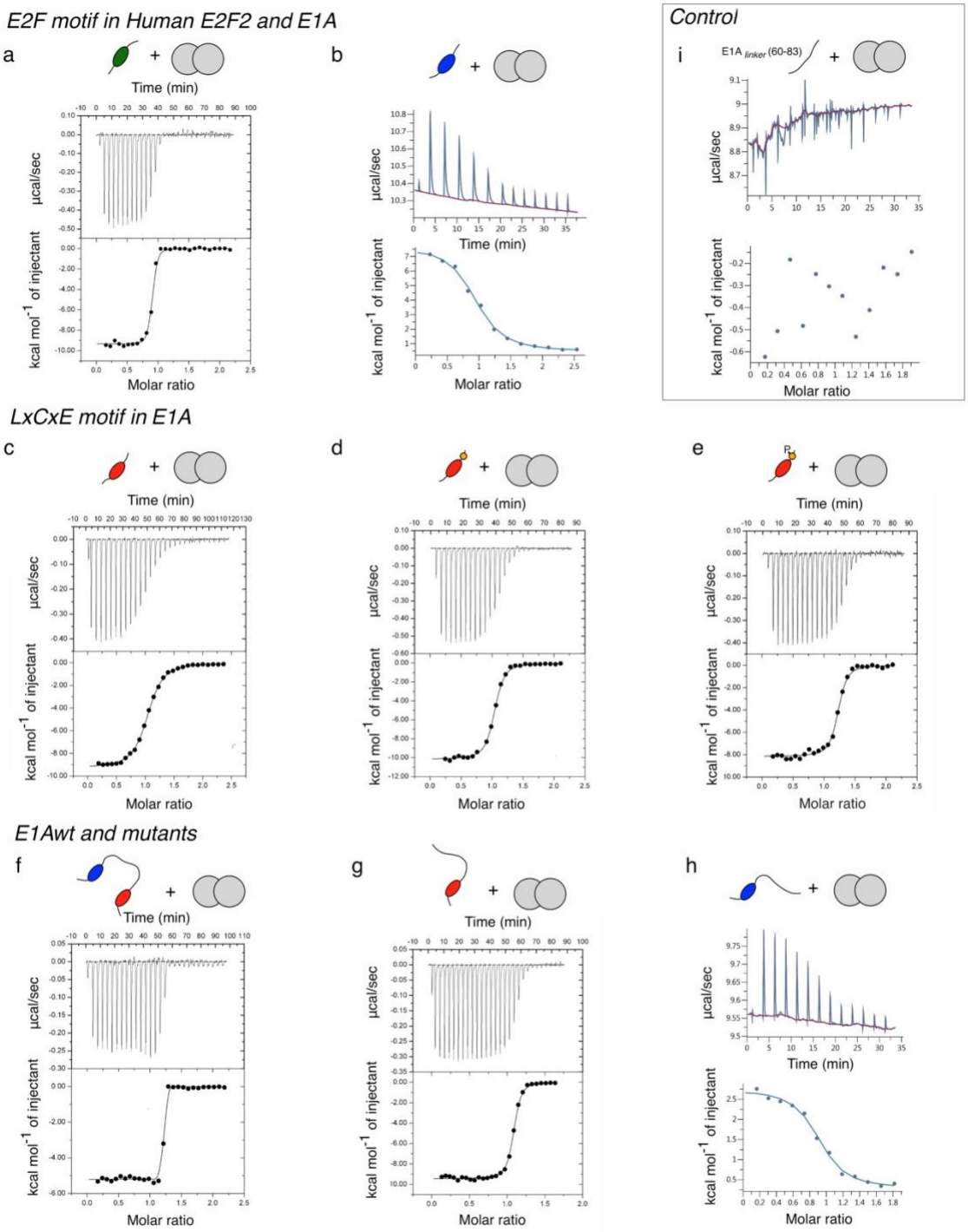
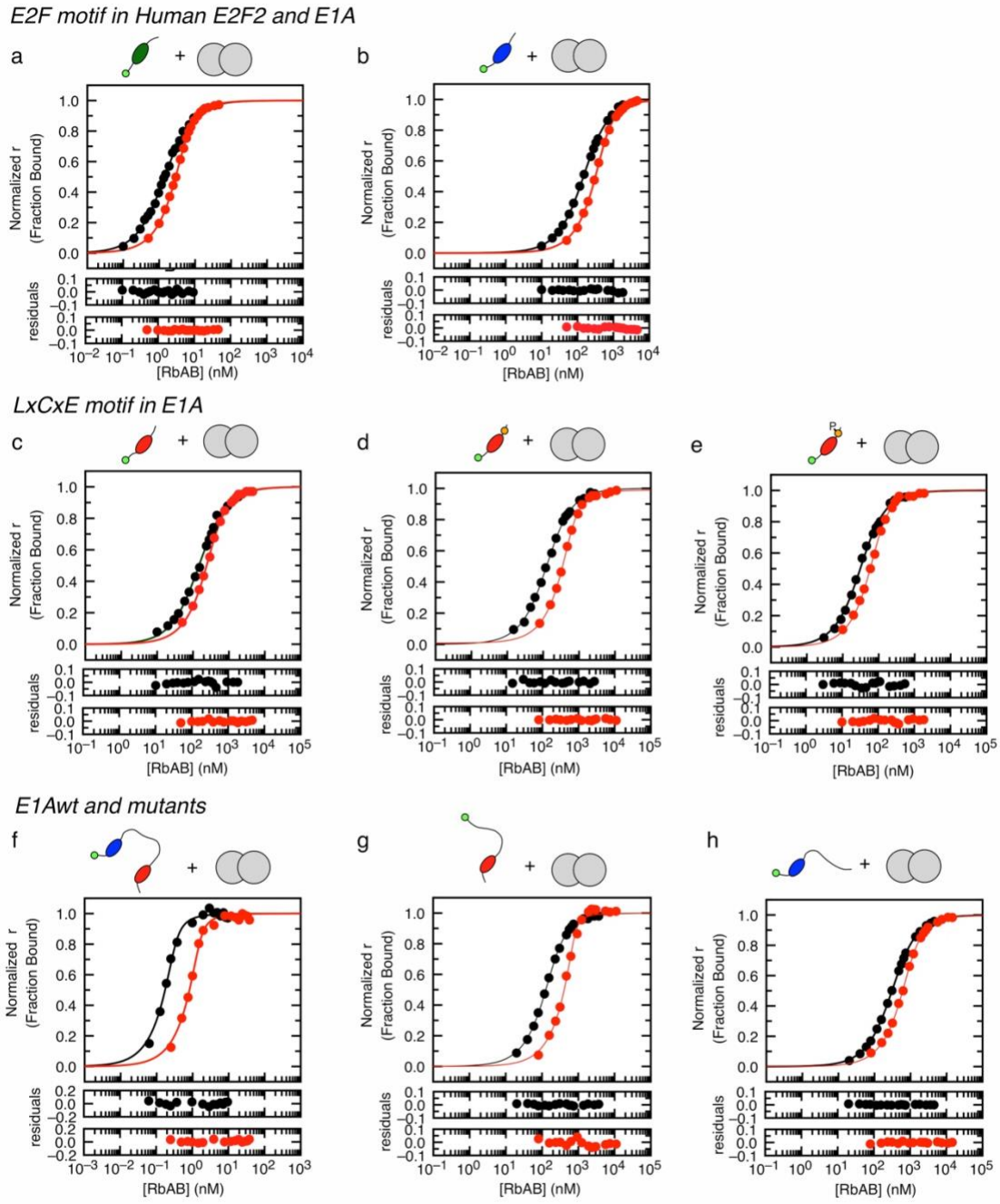
EXTENDED DATA FIGURE LEGENDS



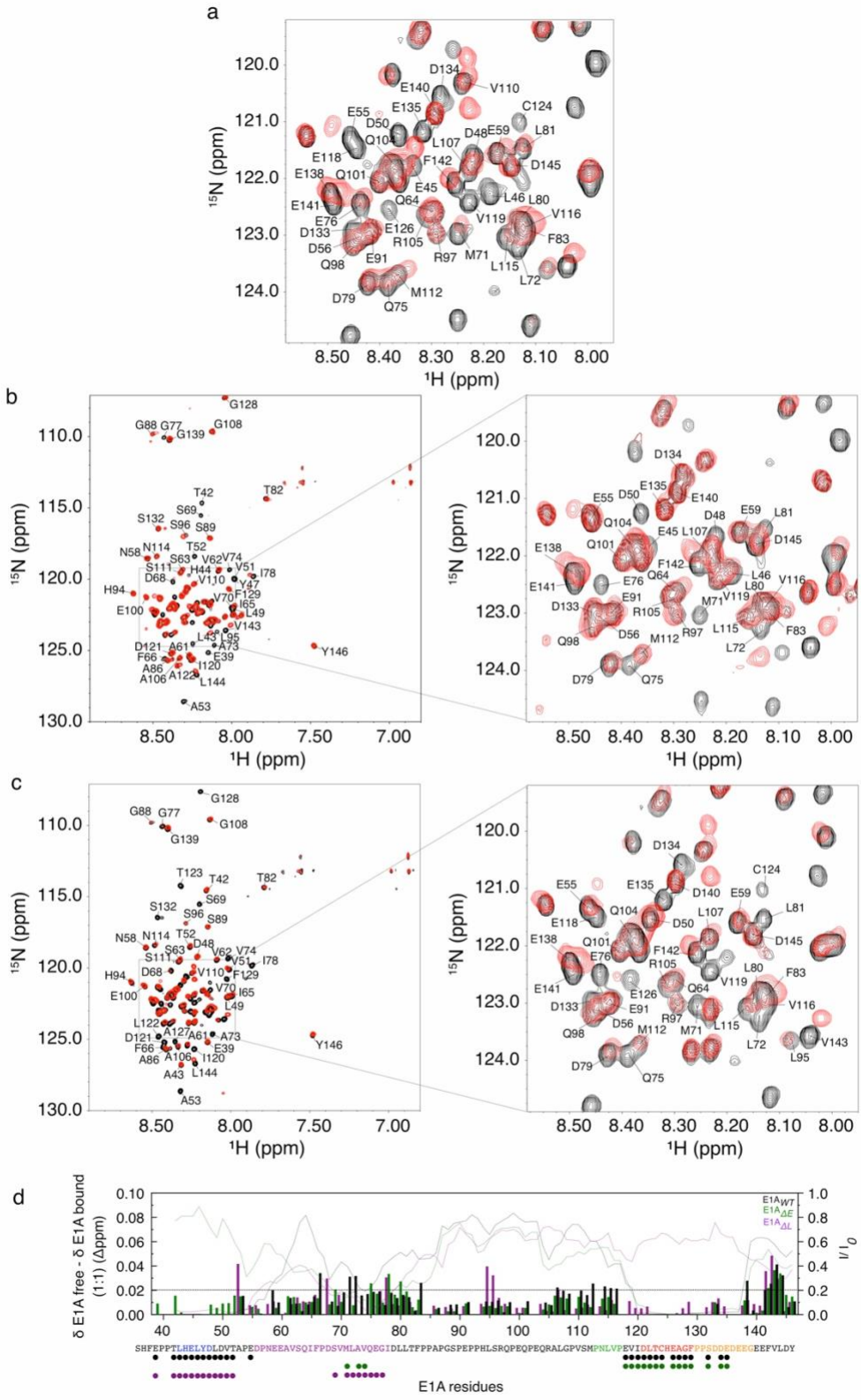
EXTENDED DATA FIGURE 1: Biophysical characterization of recombinant Rb and E1A proteins. a) Far UV-CD spectra of $E1A_{WT}$ (solid line), $E1A_{\Delta E}$ (dotted line), $E1A_{\Delta L}$ (dashed line). Inset: 15% SDS-PAGE gel of purified recombinant E1A proteins (purity > 90%). **b)** Far UV-CD spectrum 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 b) and c), 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**.


E2F motif in Human E2F2 and E1A
Control
a
b
i
E1A linker (60-83)
LxCxE motif in E1A
c
d
e
E1Awt and mutants
f
g
h
Time (min)
µcal/sec
kcal mol-1 of injectant
Molar ratio

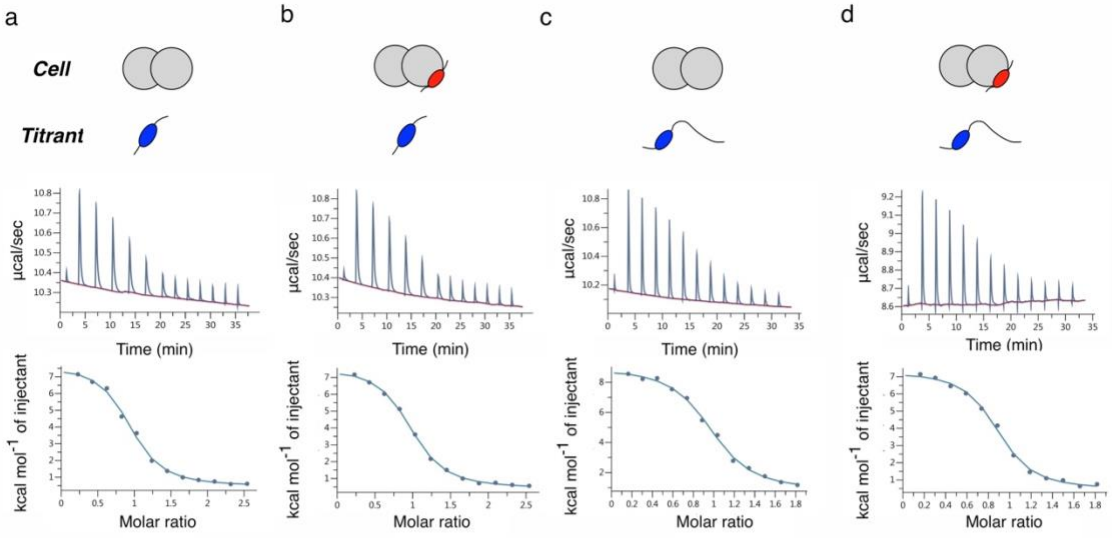
EXTENDED DATA FIGURE 2: Representative ITC binding isotherms for Rb:peptide/protein complexes. Measurements were performed loading the cell with Rb solution and the syringe with the different peptides or proteins as titrants. Panels show heat exchanged as a function of time (upper panel), and the enthalpy 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 Data Table 1.** Exothermic binding to Rb was observed for the Cellular 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 ºC. 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).

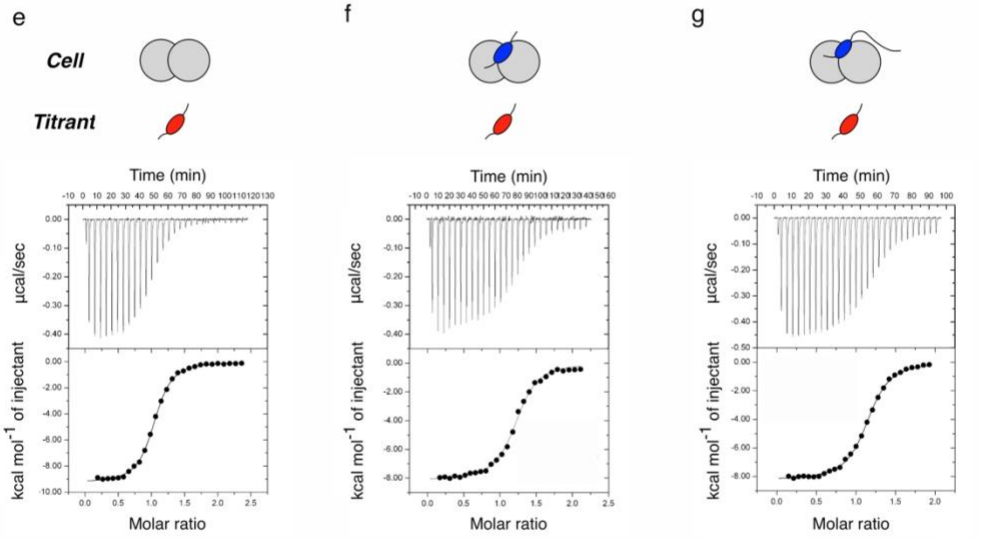

E2F motif in Human E2F2 and E1A
a
b
Normalized r
(Fraction Bound)
residuals
[RbAB] (nM)
LxCxE motif in E1A
c
d
e
P
E1Awt and mutants
f
g
h

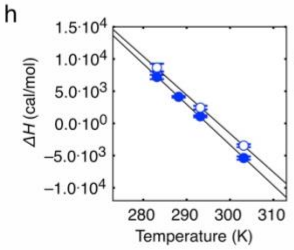
EXTENDED DATA FIGURE 3: Fluorescence Spectroscopy titration experiments of E1A-Rb and E2F-Rb interactions. Representative titration binding curves at equilibrium for each FITC-labeled peptide/protein-Rb interaction tested in this work. Normalized anisotropy signals (circles) are shown, along with the 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)** Cellular 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 Data 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.


a
b
c
d
^{15}N (ppm)
^{1}H (ppm)
δ E1A free - δ E1A bound (1:1) (Δppm)
I/ I$_0$
E1A$_{WT}$
E1A$_{ΔE}$
E1A$_{ΔL}$
E1A residues

EXTENDED DATA FIGURE 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. 2 a, 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.



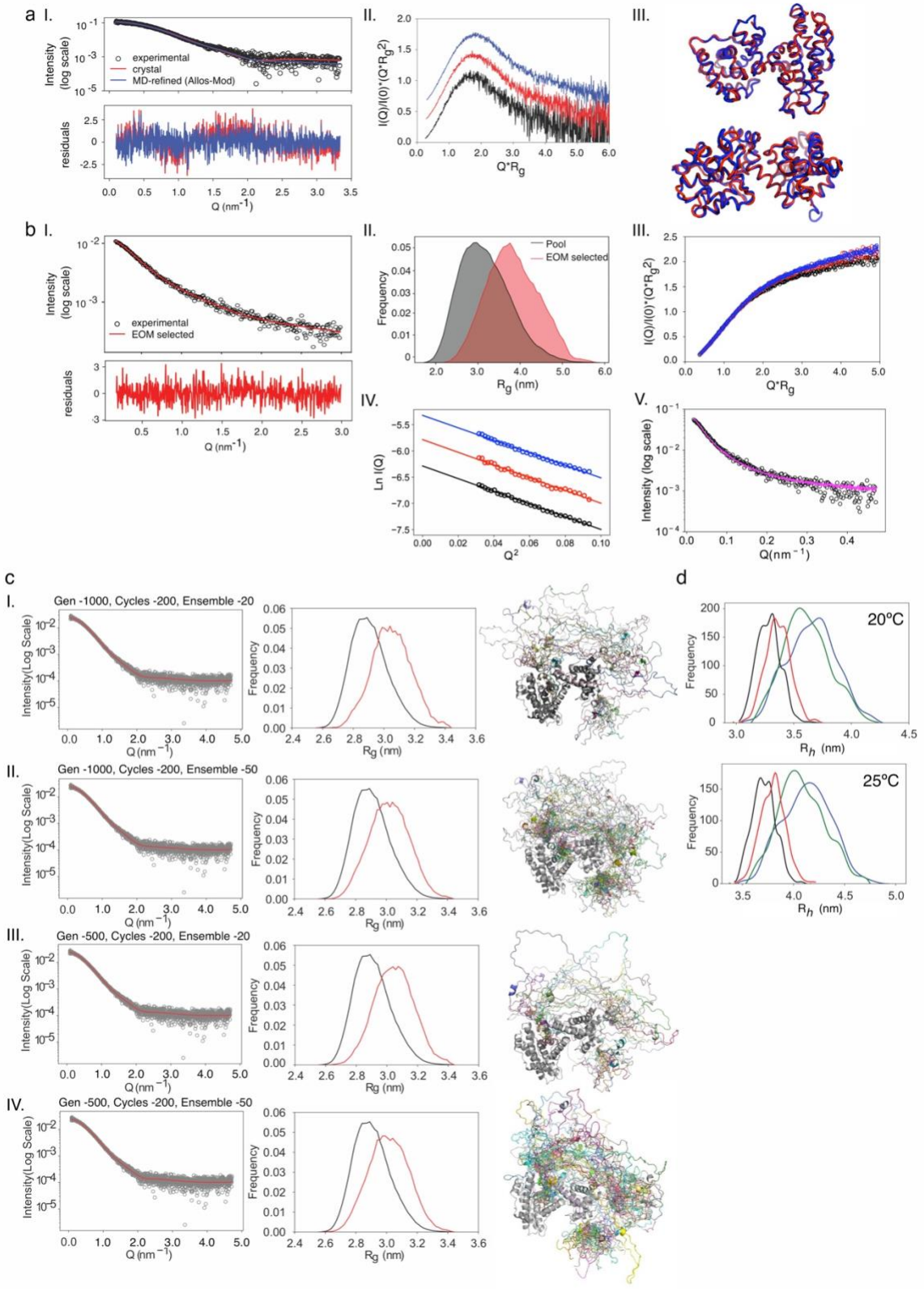




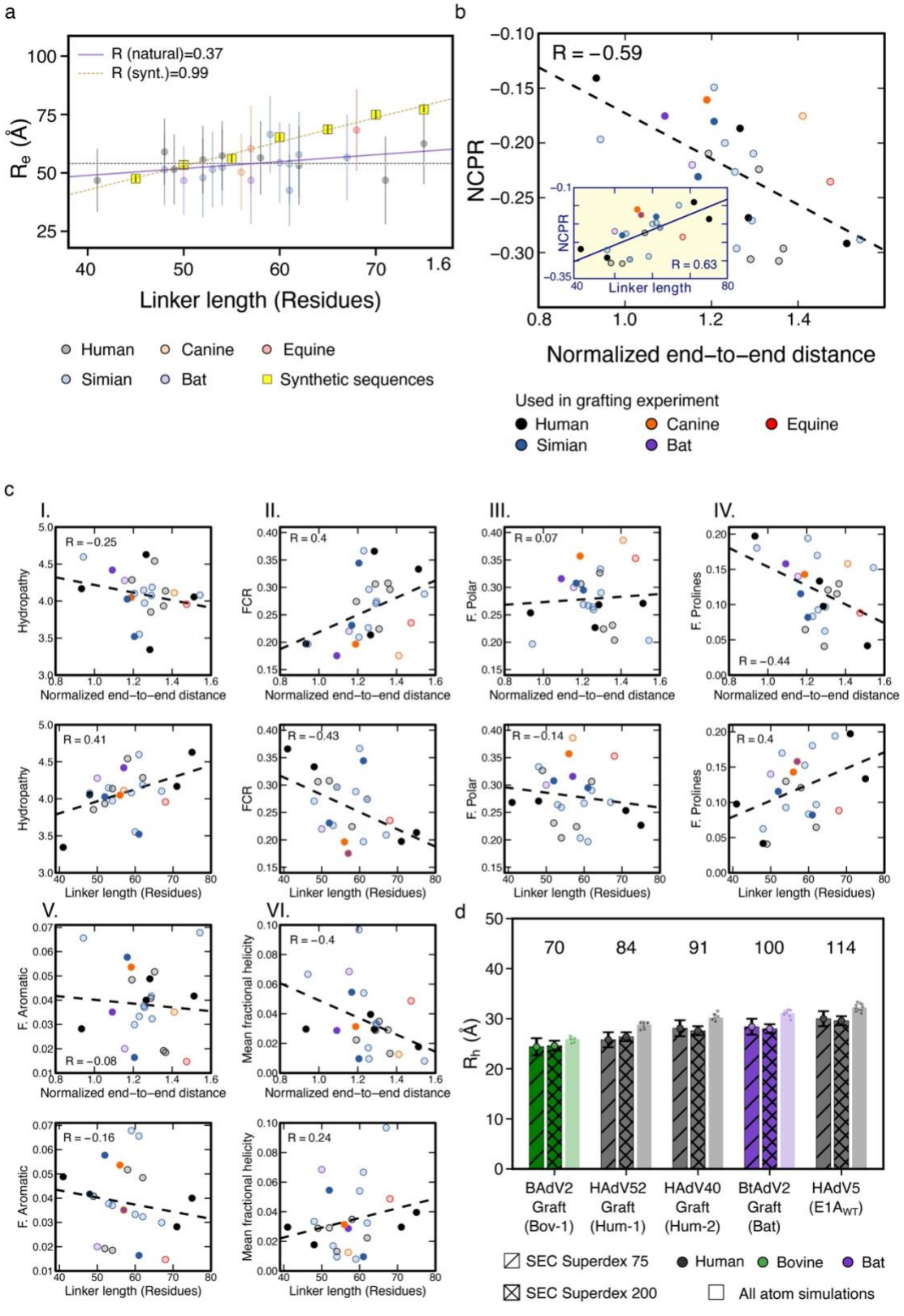
ΔCp E1A$_{\Delta L}$ = -606.3 ± 9.6 cal/mol K

ΔCp E1A$_{E2F}$ = -629.4 ± 7.3 cal/mol K

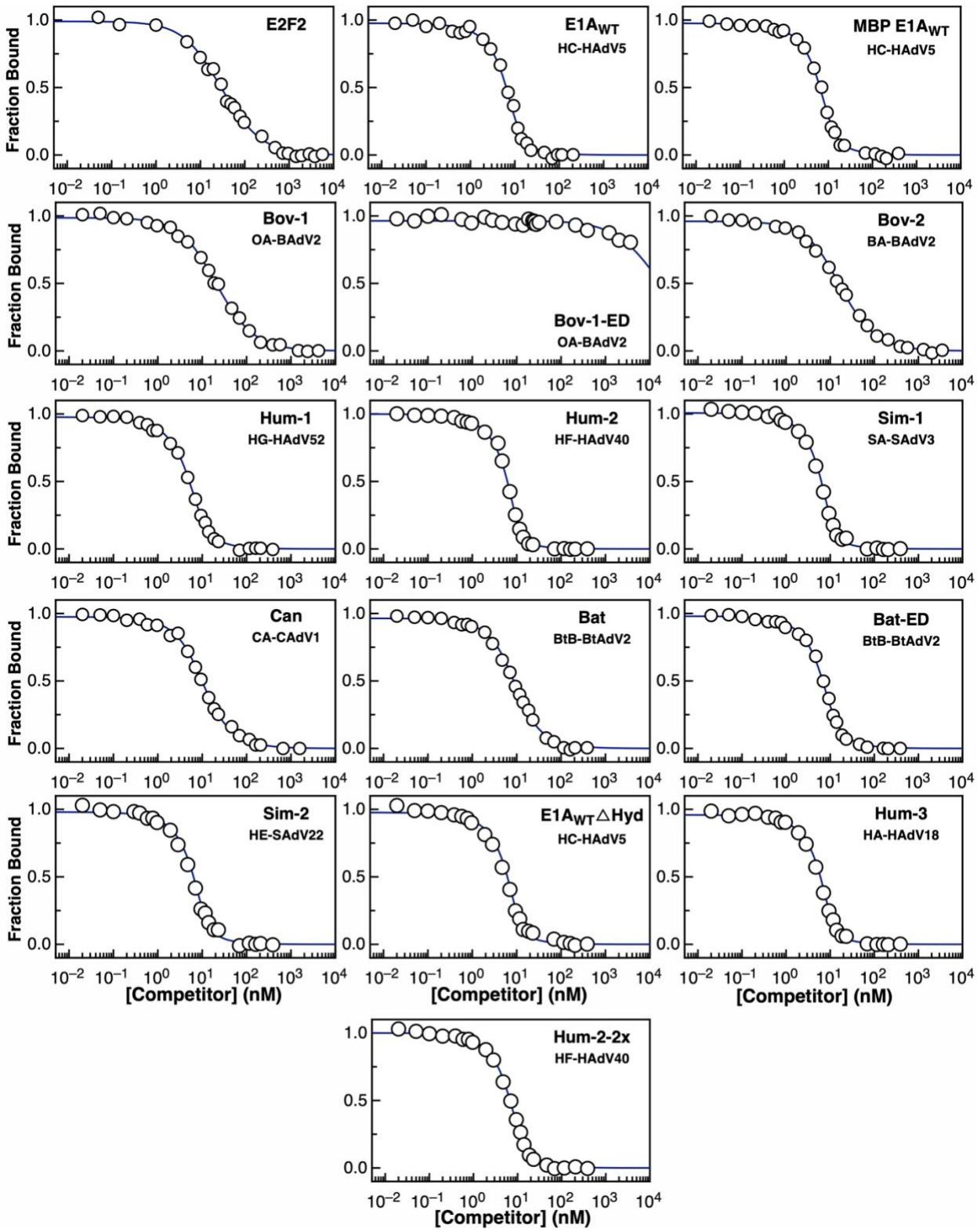
EXTENDED DATA FIGURE 5: Analysis of allosteric effects in the formation of the Rb-E1A complex. Measurements were performed by loading the cell with Rb or with a pre-assembled complex of Rb with peptide/proteins containing one of the interacting motifs and titrating with peptide/proteins containing the 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 ºC; **b)** [$E1A_{LxCxE}$:Rb] (30 μM, cell) titrated with $E1A_{E2F}$ (300 μM, syringe) at 10 ºC; **c)** Rb (30 μM, cell) titrated with $E1A_{\Delta L}$(300 μM, syringe) at 10 ºC; **d)** [$E1A_{LxCxE}$:Rb] (30 μM, cell) titrated with $E1A_{\Delta L}$ (300 μM, syringe) at 10 ºC; **e)** Rb (15 μM, cell) titrated with $E1A_{LxCxE}$ (150 μM, syringe) at 20 ºC; **f)** [$E1A_{E2F}$:Rb] (15μM, cell) titrated with $E1A_{LxCxE}$ (150 μM, syringe) at 20 ºC; **g)** [$E1A_{\Delta L}$:Rb] (15 μM, cell) titrated with $E1A_{LxCxE}$ (150 μM, syringe) at 20 ºC. Thermodynamic parameters derived from the fitting are shown in **Supplementary Data 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 bars; $E1A_{\Delta L}$: open blue bars. Thermodynamic parameters are reported in **Supplementary Data Table 5.**


a
I.
Intensity (log scale)
experimental
crystal
MD-refined (Allos-Mod)
residuals
Q (nm-1)
II.
I(Q)/I(0)*(Q*Rg)2)
Q*Rg
III.
b
I.
Intensity (log scale)
experimental
EOM selected
residuals
Q (nm-1)
II.
Pool
EOM selected
Frequency
Rg (nm)
III.
I(Q)/I(0)*(Q*Rg)2)
Q*Rg
IV.
Ln I(Q)
Q2
V.
Intensity (log scale)
Q(nm-1)
c
I.
Gen -1000, Cycles -200, Ensemble -20
II.
Gen -1000, Cycles -200, Ensemble -50
III.
Gen -500, Cycles -200, Ensemble -20
IV.
Gen -500, Cycles -200, Ensemble -50
Intensity(Log Scale)
Q (nm-1)
Frequency
Rg (nm)
d
20°C
25°C
Frequency
Rh (nm)

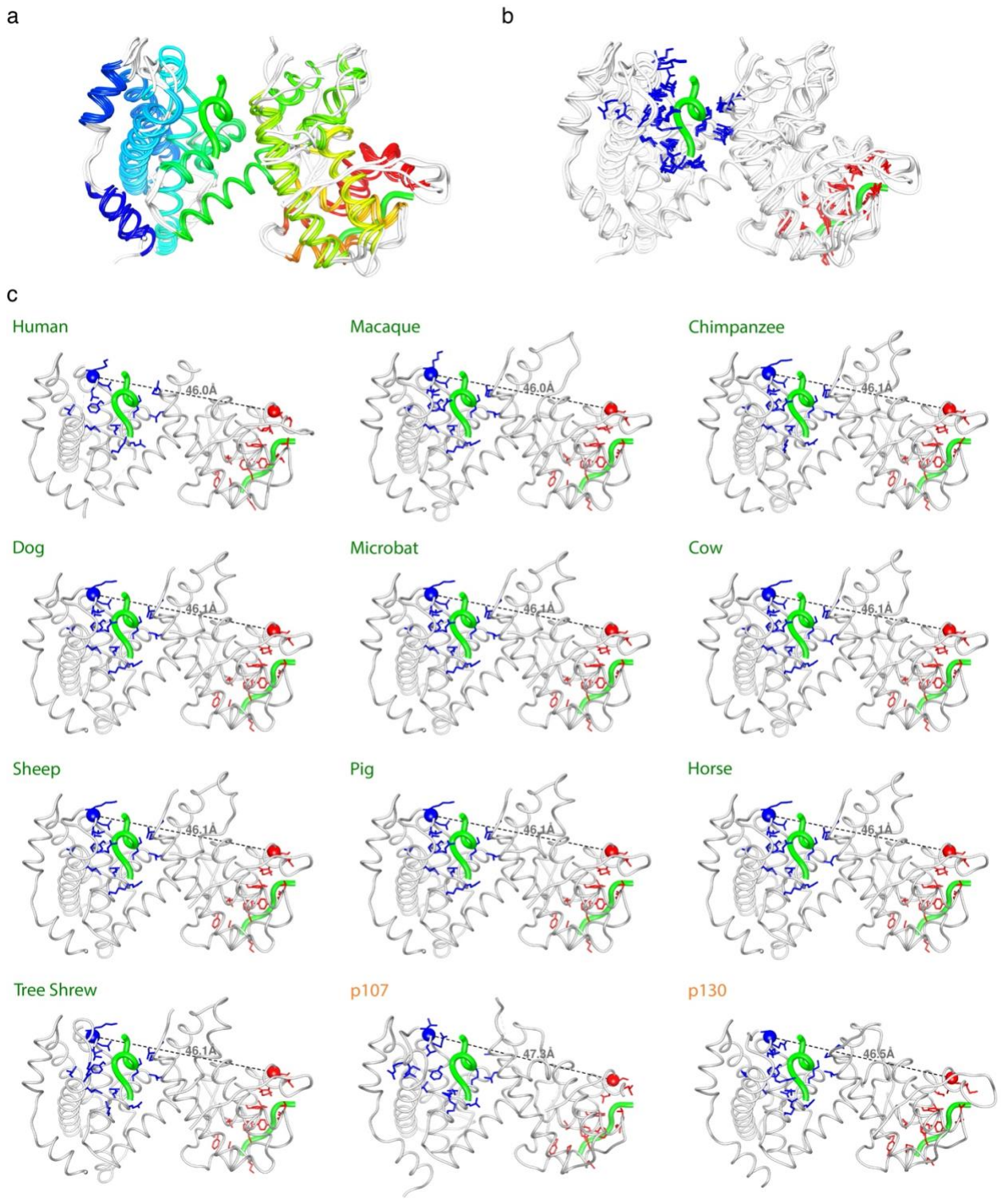
EXTENDED DATA FIGURE 6: SAXS analysis of Rb, E1A and the [E1A$_{WT}$:Rb] complex. a) I. Experimental SAXS intensity profile (black empty circles) versus theoretical profiles obtained from the crystal structure of the unliganded RbAB domain (PDB ID: 3POM) (red line) or a refined model where flexible loops 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. Rg 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).


a
R (natural)=0.37
R (synt.)=0.99
R_e (Å)
Linker length (Residues)
Human
Canine
Equine
Simian
Bat
Synthetic sequences
b
R = −0.59
NCPR
Normalized end−to−end distance
R = 0.63
Linker length
Used in grafting experiment
Human
Canine
Equine
Simian
Bat
c
I.
R = −0.25
Hydropathy
II.
R = 0.4
FCR
III.
R = 0.07
F. Polar
IV.
R = −0.44
F. Prolines
Normalized end−to−end distance
R = 0.41
R = −0.43
R = −0.14
R = 0.4
Linker length (Residues)
V.
R = −0.08
F. Aromatic
VI.
R = −0.4
Mean fractional helicity
R = −0.16
R = 0.24
d
70
84
91
100
114
R_h (Å)
BAdV2 Graft (Bov-1)
HAdV52 Graft (Hum-1)
HAdV40 Graft (Hum-2)
BtAdV2 Graft (Bat)
HAdV5 ($E1A_{WT}$)
SEC Superdex 75
SEC Superdex 200
Human
Bovine
Bat
All atom simulations

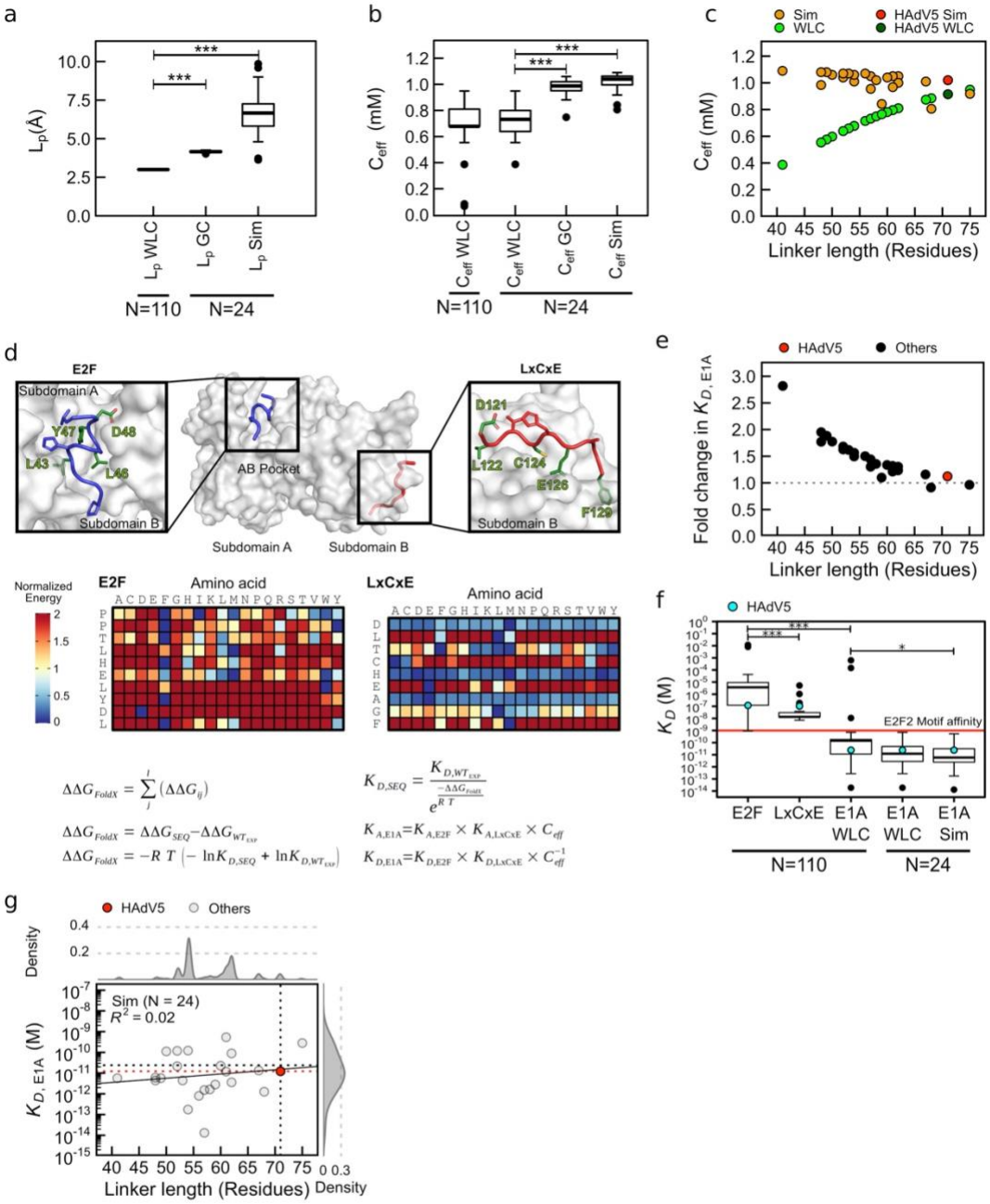
EXTENDED DATA FIGURE 7: Correlation of E1A linker dimensions with sequence-encoded features. a) Linker length control titration experiment. End-to-end distance (R_e) of natural sequences (colored circles) compared to synthetic sequences of varying length and constant sequence composition matching the 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. 1 f, 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 Kav 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.


E2F2
E1A$_{WT}$
HC-HAdV5
MBP E1A$_{WT}$
HC-HAdV5
Bov-1
OA-BAdV2
Bov-1-ED
OA-BAdV2
Bov-2
BA-BAdV2
Hum-1
HG-HAdV52
Hum-2
HF-HAdV40
Sim-1
SA-SAdV3
Can
CA-CAdV1
Bat
BtB-BtAdV2
Bat-ED
BtB-BtAdV2
Sim-2
HE-SAdV22
E1A$_{WT}$△Hyd
HC-HAdV5
Hum-3
HA-HAdV18
Hum-2-2x
HF-HAdV40
Fraction Bound
[Competitor] (nM)

EXTENDED DATA FIGURE 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 10nM concentration with increasing concentrations of each variant. One representative example is shown for each variant reported on **Supplementary Data Table 7**. The displacement reaction was followed by recording the fluorescence anisotropy of the FITC moiety, with excitation at 490nm and emission at 520nm. 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.


a
b
c
Human
46.0Å
Macaque
46.0Å
Chimpanzee
46.1Å
Dog
46.1Å
Microbat
46.1Å
Cow
46.1Å
Sheep
46.1Å
Pig
46.1Å
Horse
46.1Å
Tree Shrew
46.1Å
p107
47.3Å
p130
46.5Å

EXTENDED DATA FIGURE 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 = 49Å), 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).



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

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

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

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

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

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

EXTENDED DATA FIGURE 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 = 3Å (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 $E1A_{WT}$.