LETTER

Modulation of allostery by protein intrinsic disorder

Allan Chris M. Ferreon[1]*, Josephine C. Ferreon[1]*, Peter E. Wright[1,2] & Ashok A. Deniz[1]

Allostery is an intrinsic property of many globular proteins and enzymes that is indispensable for cellular regulatory and feedback mechanisms. Recent theoretical[1] and empirical[2] observations indicate that allostery is also manifest in intrinsically disordered proteins, which account for a substantial proportion of the proteome[3,4]. Many intrinsically disordered proteins are promiscuous binders that interact with multiple partners and frequently function as molecular hubs in protein interaction networks. The adenovirus early region 1A (E1A) oncoprotein is a prime example of a molecular hub intrinsically disordered protein[5]. E1A can induce marked epigenetic reprogramming of the cell within hours after infection, through interactions with a diverse set of partners that include key host regulators such as the general transcriptional coactivator CREB binding protein (CBP), its paralogue p300, and the retinoblastoma protein (pRb; also called RB1)[6,7]. Little is known about the allosteric effects at play in E1A–CBP–pRb interactions, or more generally in hub intrinsically disordered protein interaction networks. Here we used single-molecule fluorescence resonance energy transfer (smFRET) to study coupled binding and folding processes in the ternary E1A system. The low concentrations used in these high-sensitivity experiments proved to be essential for these studies, which are challenging owing to a combination of E1A aggregation propensity and high-affinity binding interactions. Our data revealed that E1A–CBP–pRb interactions have either positive or negative cooperativity, depending on the available E1A interaction sites. This striking cooperativity switch enables fine-tuning of the thermodynamic accessibility of the ternary versus binary E1A complexes, and may permit a context-specific tuning of associated downstream signalling outputs. Such a modulation of allosteric interactions is probably a common mechanism in molecular hub intrinsically disordered protein function.

Binding promiscuity is a hallmark of most hub proteins involved in signalling networks (for example, p53 and BRCA1)[8]. The inherent flexibility and structural adaptability of intrinsically disordered proteins makes them ideal hub proteins for binding to diverse partners. Not surprisingly, viruses widely use intrinsically disordered linear motifs to orchestrate subversion of the host cellular interactome[9].

The intrinsically disordered adenoviral protein E1A uses its amino-terminal region and conserved regions CR1 (residues 42–83) and CR2 (residues 121–139) in a cooperative manner to recruit numerous cellular regulatory proteins, thereby subverting signalling pathways in the infected cell[10]. The TAZ2 domain of CBP/p300 and the pocket domain of pRb each bind to two non-contiguous and largely non-overlapping regions of E1A to form binary complexes (E1A–pRb and E1A–TAZ2) and a ternary complex (pRb–E1A–TAZ2) (Fig. 1a)[11]. The major interaction site of CBP/p300 TAZ2 is within the E1A CR1 region, with a secondary binding site in the N-terminal region of E1A (Fig. 1b)[11]. pRb binds the characteristic LXCXE motif (residues 122–126) within the E1A CR2 region as well as a second binding site within CR1 (residues 42–49), in a region immediately preceding the TAZ2 binding site[12]. The E1A interaction sites occupy different regions of the CBP TAZ2 or the pRb surface[11,12]. The TAZ2 domain does not bind directly to the pocket domain of pRb, but rather associates with pRb only within

ternary complexes formed by binding of both proteins to E1A[11]. To identify potential allosteric effects that fine-tune the interactions between the three proteins and assess the energetic contributions of each E1A interaction motif (N terminus, CR1 and CR2) to binding, several truncated E1A constructs were generated and studied (Fig. 1b, Methods).

Previous attempts at measuring dissociation constants (K_d) for E1A complexes with CBP by isothermal titration calorimetry (ITC) and nuclear
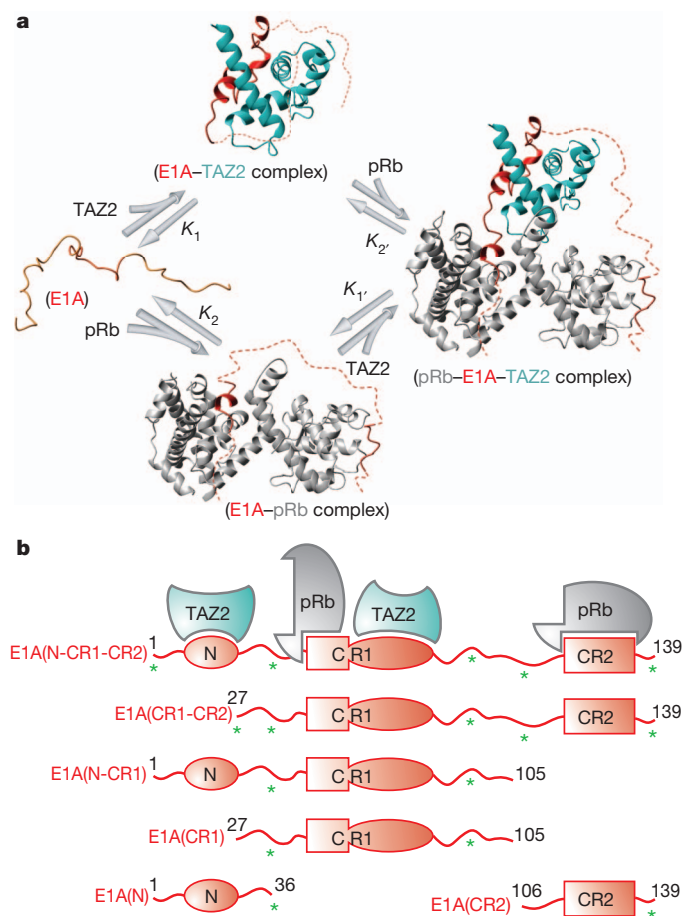
Figure 1 | Folding of the intrinsically disordered protein E1A induced by binding to pRb and the TAZ2 domain of CBP/p300. a, E1A binding and folding equilibria, showing formation of the ternary complex from the unbound intrinsically disordered protein state by way of two binary intermediate complexes. **b**, E1A constructs used to study the contributions of the N-terminal, CR1 and CR2 regions to formation of the binary and ternary complexes. Asterisks indicate the locations of single- or dual-site dye labelling for fluorescence measurements (that is, residue positions −3, 36, 88, 111 and 137, where residues 1–139 comprise the E1A sequence and positions −4 to −1 are the residues GSHM).

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [2]Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.
*These authors contributed equally to this work.

magnetic resonance (NMR) failed because E1A is highly aggregation-prone[11]. Even at concentrations as low as 10 μM, E1A(N-CR1-CR2) (residues 1–139 comprising the N-terminal region, CR1 and CR2) forms visible precipitates upon binding to CBP TAZ2. Hence, the previous NMR experiments, performed at micromolar concentrations, could demonstrate ternary complex formation between pRb, E1A and the CBP TAZ2 domain only for the short E1A CR1 region (E1A(CR1); residues 27–91)[11]. To overcome these problems, ensemble fluorescence anisotropy measurements were first attempted to measure binding affinities for longer E1A constructs that include more, or all, of the CBP, TAZ2 and pRb binding sites, and under more physiological concentrations (nanomolar to micromolar).

Formation of binary and ternary complexes of Ad2 E1A with the TAZ2 domain of mouse CBP and the human pRb pocket domain was monitored by fluorescence anisotropy titrations (Fig. 2a, b and Supplementary Fig. 1). From these ensemble fluorescence measurements we were able to obtain accurate dissociation constants for binary complexes with K_d greater than 25 nM (Supplementary Table 1 and Supplementary Fig. 2) that were in agreement with published values for pRb[12]. However, two issues impeded quantitative analysis of the binding data. First, most of the affinities are very high ($K_d < 25$ nM), and outside the reliable detection limit of fluorescence anisotropy measurements (Methods, Supplementary Table 1 and Supplementary Fig. 2). Second, in the presence of the N-terminal region, and especially for E1A(N-CR1-CR2) (residues 1–139), aggregation of the E1A constructs occurred at the relatively high concentrations required for competition fluorescence anisotropy assays.

To overcome the aforementioned problems, we used single-molecule fluorescence resonance energy transfer (smFRET). Owing to its high detection sensitivity, smFRET is an ideal method for investigating aggregation-prone and high-affinity systems, using low concentrations of fluorescently labelled protein (that is, ≤100 pM). In addition, the absence of ensemble averaging enables direct observation of free and bound populations, allowing for straightforward K_d measurements (Methods).

Binding affinities were measured by monitoring changes in intramolecular FRET that accompany folding of E1A upon binding to CBP TAZ2 or pRb. smFRET measurements were performed using freely diffusing E1A dual-labelled with donor and acceptor dyes (Fig. 1b). Free and bound populations in the resulting smFRET histograms (Fig. 3a–d and Supplementary Figs 3–6) have characteristic FRET efficiencies (E_{FRET}) (Supplementary Table 3) that are related to inter-dye distances[13,14]. In its free unbound state, E1A (labelled at multiple donor-acceptor sites) exhibited relatively low E_{FRET} (0.2–0.5), indicating extended structures. In contrast, E1A exhibited higher E_{FRET} (0.4–0.9) for both binary and ternary complexes, consistent with formation of more compact structures due to folding upon binding. The distances estimated from E_{FRET} values for bound E1A(CR1) are in agreement with those estimated from NMR and X-ray structures of E1A bound to TAZ2 and pRb pocket domains[11,12,15].

Using E1A(CR1(27–105/36C88C)), with fluorescent labels attached at Cys introduced at residue positions 36 and 88 (Fig. 1b), titrations with increasing concentrations of CBP TAZ2 resulted in a gradual disappearance of the free E1A peak (E_{FRET} ~0.46), concurrent with a gradual appearance of the binary E1A–CBP TAZ2 peak (E_{FRET} ~0.9; Fig. 3a). The increased E_{FRET} of the latter peak is due to folding of E1A upon binding to TAZ2, forming a more compact E1A structure as observed by NMR[11]. The smFRET data can be fitted to a one-site binding model with a $K_d = 11.7 \pm 0.4$ nM (Methods and Supplementary Fig. 3). In the presence of ≥1 μM pRb, TAZ2 binds to the pRb-bound E1A(CR1(27–105/36C88C)) with $K_d = 37 \pm 6$ nM to form a ternary complex (Fig. 3b and Supplementary Fig. 3).

Next, we performed similar experiments using the more aggregation-prone E1A construct containing both the N-terminal region (residues 1–26) and the CR1 region (E1A(N-CR1(1–105/36C88C))). The observed E_{FRET} values for free and bound E1A(N-CR1) were very similar to those for E1A(CR1), suggesting that the CR1 region, located between the fluorescent probes, adopts similar conformations in both complexes, unperturbed by the presence of the N terminus. The N-terminal region of E1A seems to interact weakly with TAZ2, as the binding affinity increases ~4-fold when it is present (Supplementary Table 2). When the N-terminal region of E1A is free to participate in the binding interactions, $K_d = 3.2 \pm 0.5$ nM for TAZ2 binding to E1A alone, whereas $K_d = 1.5 \pm 0.3$ nM for binding of TAZ2 to E1A in the presence of 1 μM pRb (Fig. 3c, d and Supplementary Fig. 3). Binding of TAZ2 to the binary E1A(N-CR1)–pRb complex is much stronger than to the E1A(CR1)–pRb complex that lacks the E1A N terminus ($K_d = 1.5$ versus 37 nM), showing that the N-terminal region of E1A makes interactions that stabilize the ternary complex. Similar results were obtained for E1A constructs containing the CR2 motif (E1A(CR1-CR2(27–139/36C88C)) and E1A(N-CR1-CR2(1–139/36C88C))), where the E1A N terminus enhances the binding affinity for TAZ2 ($K_d = 1.6$ versus 7.5 nM for the shorter construct) but has no effect on binding to pRb (Supplementary Fig. 4 and Supplementary Table 2). This increase in affinity is attributed to additional binding interactions mediated by the E1A N terminus, which binds dynamically to a surface of TAZ2 opposite the CR1 binding site and causes exchange broadening of NMR resonances[11].

We next used the affinity data to generate protein phase diagrams (Fig. 3e–h), which provide graphical representations of folding and binding linkage equilibria[16]. These diagrams provide population information for different E1A species (free, binary and ternary species) versus concentrations of CBP TAZ2 and pRb. Each phase separation line (for example, black line, Fig. 3e–h) represents ligand concentrations where a corresponding state (for example, unbound E1A) is 50% populated relative to all other states. In cells, concentrations of signalling proteins range from nanomolar to micromolar and can be as high as millimolar with co-localization[17]. Therefore, the concentration ranges shown for the phase diagrams are well within physiological ranges. Asymmetry in the central white areas (where the population of none of the states exceeds 50%) reflects cooperative binding. Thus, for E1A(CR1), the decrease in TAZ2–E1A binding affinity in the presence of pRb and corresponding positive slope in the white area in Fig. 3e demonstrate negative cooperativity between pRb and CBP TAZ2, with the formation of the binary E1A complexes favoured over the ternary complex at lower concentrations of pRb and TAZ2. Notably, and in contrast with E1A(CR1), the E1A(N-CR1) binding phase diagram (Fig. 3f) reveals positive cooperativity for the interactions between CBP TAZ2 and pRb, clearly reflected in a negative slope of the white area. Therefore, the availability of the E1A N-terminal region can modulate the sign of the cooperativity of CBP TAZ2 and pRb binding to E1A(CR1). In the cell, this situation might have a key role when a binding partner sequesters the E1A CR2 region. Our observations are also directly relevant to the interactions of cellular proteins with CR2-deleted E1A produced by oncolytic adenovirus mutants that are in clinical trials for cancer therapy[18]. We note that truncated versions of E1A, lacking the N
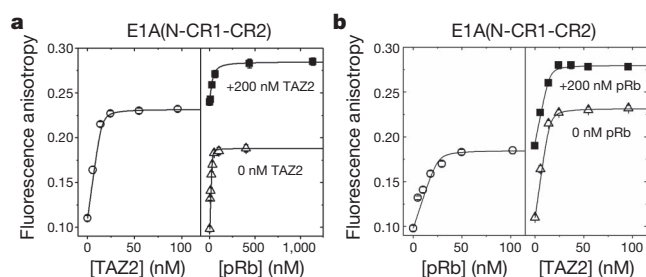

Figure 2 | E1A–TAZ2–pRb ternary complex formation detected by ensemble fluorescence anisotropy. a, b, TAZ2/pRb titration of free (open symbols) and TAZ2- or pRb-bound (solid symbols) Alexa Fluor 594-labelled E1A(N-CR1-CR2(1–139/S88C)).
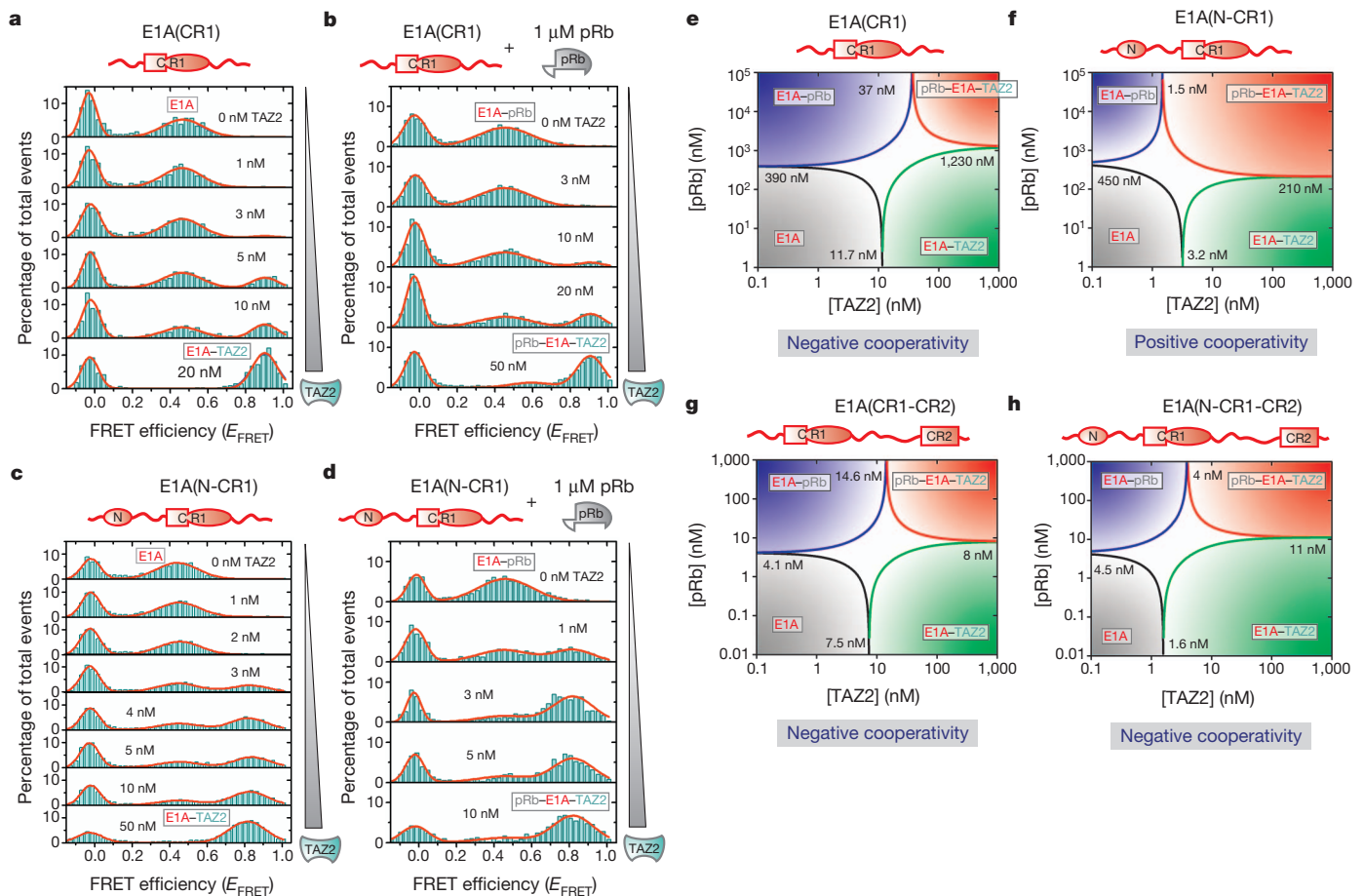
Figure 3 | E1A–TAZ2–pRb allosteric interactions probed using single-molecule fluorescence resonance energy transfer. a–d, smFRET histograms for the TAZ2 titration of Alexa Fluor 488- and 594-labelled E1A(CR1(27–105/36C88C)) (**a**, **b**) and E1A(N-CR1(1–105/36C88C)) (**c**, **d**) constructs in the absence (**a**, **c**) and presence (**b**, **d**) of 1 μM pRb. **e–h**, pRb–TAZ2 phase diagrams for FRET-labelled E1A(CR1) (**e**), E1A(N-CR1) (**f**), E1A(CR1-CR2(27–139/36C88C)) (**g**) and E1A(N-CR1-CR2(1–139/36C88C)) (**h**) constructed using the K_d values derived from ensemble and single-molecule fluorescence measurements (Supplementary Tables 1 and 2). The K_d values for the binding of E1A with pRb in the presence of CBP TAZ2 ($K_{2'}$) cannot be determined experimentally owing to overlap of E_{FRET} signals but can be calculated from a thermodynamic cycle analysis (Fig. 1a); $K_{1'}/K_1 = K_{2'}/K_2$. These values correspond to 1,230, 210, 8 and 11 nM in **e–h**, respectively.

terminus or other interaction domains, are commonly used to study the cellular response to E1A (ref. 5).

Previous NMR data suggest a plausible molecular basis for the observed negative cooperativity (Fig. 3e, g, h). Chemical shift titrations[11] indicate that binding of pRb disrupts a small subset of the intermolecular interactions that exist in the binary E1A(CR1)–CBP TAZ2 complex, indicating that negative allostery may be associated with partial overlap between the pRb and CBP TAZ2 binding sites in the E1A CR1 region. The molecular origin of the positive cooperativity observed for E1A (N-CR1) (Fig. 3f) is less obvious. However, allosteric coupling between sites in intrinsically disordered proteins can be either positive or negative and does not require mechanical linkage but can arise through perturbations of the energetic balance by binding events at individual sites[19].

The phase diagrams also provide a direct visualization of how cooperativity affects E1A population distributions and thereby their functional outcomes. Negative cooperativity (Fig. 3e, g, h) results in the ternary complex occupying a smaller area relative to the binary complexes. Conversely, positive cooperativity (Fig. 3f) results in the ternary complex occupying broader concentration ranges. Together, the phase diagrams demonstrate how multiple layers of regulation can be imposed on the E1A hub, depending on which domains of E1A are available for interaction with CBP/p300 and pRb, permitting the cooperativity of the system to be fine-tuned over a broad concentration range.

So far, there are relatively few examples of negatively cooperative biological systems[20]. Positive cooperativity is a common mechanism

for increasing the binding potential. Positive cooperativity in ternary complex formation would enhance a critical function of E1A: the CBP/p300-mediated acetylation of pRb to force permanent exit from the cell cycle and promote differentiation of the host cell[21–23]. However, for a promiscuous molecular hub intrinsically disordered protein such as E1A (Fig. 4a), negative cooperativity has an equally important role because it broadens the stimulus range[24], increasing the population of intermediate binding states (binary complexes) and facilitating their interactions with other partners (Fig. 4b). This would permit a context-dependent modulation of different molecular species that contribute to the potency of viral E1A in subverting host cellular mechanisms[6,7].

Our results indicate that intrinsically disordered protein systems can be tuned to optimize population distributions and cellular outcome by changing the available binding sites. This could occur by competition between different molecular partners for the same E1A binding sites (Fig. 4a), resulting in allosteric modulation of the interaction and signalling networks involving CBP/p300 and pRb (Fig. 4b). E1A exhibits multiple activities in infected cells, mediating CBP/p300-dependent pathways that are independent of pRb (transcriptional activation or repression; green quadrant in Fig. 4b), pRb-dependent pathways that are independent of CBP/p300 (cell cycle progression, blue quadrant in Fig. 4b), and pathways that are dependent on both CBP/p300 and pRb (differentiation-specific functions of E1A; red quadrant in Fig. 4b)[10,21–23,25,26]. Our allosteric interaction modulation model provides an important mechanistic paradigm for understanding
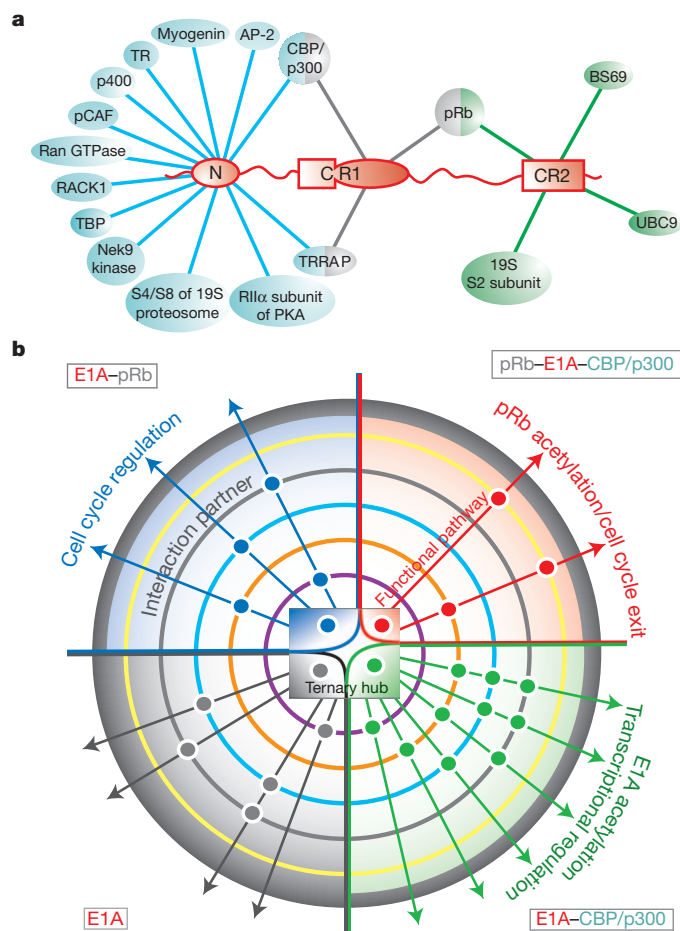
Figure 4 | E1A functional complexity achieved through binding promiscuity. a, Interactions of the N-terminal, CR1 and CR2 motifs of E1A with cellular proteins. Interactions mediated by the CR3 and CR4 regions of E1A[5] are not shown. **b**, Allosteric modulation of signalling pathways by interactions of the E1A–CBP/p300–pRb ternary hub. This hub, represented by a central phase diagram, has four E1A states: free E1A, E1A–pRb, E1A–CBP/p300, and ternary complex (grey, blue, green and red quadrants, respectively). Coloured concentric circles surrounding the hub represent additional protein partners with different interaction propensities for individual hub states. Each positive interaction is represented by a dot, coloured by hub state, and positioned based on the interaction partner. These ternary hub interactions with different sets of partners result in multiple functional pathways, the control of which may be achieved by modulating the central E1A–CBP/p300–pRb hub equilibria.

regulation of such varying signalling outputs, although other parameters are also likely to be important in a cellular context. A recent theoretical study[27] showed that allosteric ensembles associated with intrinsic protein disorder can upregulate or downregulate activity in response to different physiological stimuli, a feature of E1A that both activates and represses gene expression[10]. The capacity to expand protein functionality through modulation of accessible interaction sites has some similarities to alternative splicing, where the different protein isoforms generated increase the functional complexity of the genome[28]. Given the small size of the viral genome, it would be advantageous for adenovirus to amplify functional complexity using only a small number of proteins while maintaining the potential for maximum cellular control. A way to accomplish this is through a hub intrinsically disordered protein such as E1A that initially interacts with a small number of major binding partners (for example, pRb and CBP TAZ2) to form a series of hub interaction complexes (Fig. 4b). Additional binding partners

then interact with the hub complex, with varied interaction preferences against different molecular forms, resulting in altered signalling outputs. Thus, modulation of allostery using intrinsically disordered protein regions that can bind to diverse partners may be a mechanism by which a promiscuous molecular hub intrinsically disordered protein can manage its functional complexity.

METHODS SUMMARY

K_d measurements using ensemble fluorescence anisotropy for the different E1A samples were performed using direct and competition titration methods[29].

The smFRET confocal instrumentation, and single-molecule data collection and analyses, were previously described[30]. For each smFRET measurement carried out at specific experimental conditions, an average of ~5,000 events were collected. For the whole series of smFRET titrations, data collection was in excess of 700,000 events. The E_{FRET} histograms were fitted to Gaussian functions using OriginPro 8 (OriginLab Corp.). K_d values were determined by using a one-site binding model and by analysing the ligand concentration dependence of the peak areas of free and bound populations. See Methods and Supplementary Fig. 7 for more details.

Full Methods and any associated references are available in the online version of the paper.

1. Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl Acad. Sci. USA* **104,** 8311–8315 (2007).
2. Garcia-Pino, A. *et al.* Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142,** 101–111 (2010).
3. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol.* **6,** 197–208 (2005).
4. Xie, H. *et al.* Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* **6,** 1882–1898 (2007).
5. Pelka, P., Ablack, J. N., Fonseca, G. J., Yousef, A. F. & Mymryk, J. S. Intrinsic structural disorder in adenovirus E1A: a viral molecular hub linking multiple diverse processes. *J. Virol.* **82,** 7252–7263 (2008).
6. Ferrari, R. *et al.* Epigenetic reprogramming by adenovirus e1a. *Science* **321,** 1086–1088 (2008).
7. Horwitz, G. A. *et al.* Adenovirus small e1a alters global patterns of histone modification. *Science* **321,** 1084–1085 (2008).
8. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D² concept. *Annu. Rev. Biophys.* **37,** 215–246 (2008).
9. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends Biochem. Sci.* **36,** 159–169 (2011).
10. Berk, A. J. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene* **24,** 7673–7685 (2005).
11. 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).
12. 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).
13. Ferreon, A. C. M., Moran, C. R., Gambin, Y. & Deniz, A. A. Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol.* **472,** 179–204 (2010).
14. Deniz, A. A., Mukhopadhyay, S. & Lemke, E. A. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface* **5,** 15–45 (2008).
15. 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).
16. Ferreon, A. C. M., Ferreon, J. C., Bolen, D. W. & Rösgen, J. Protein phase diagrams II: nonideal behavior of biochemical reactions in the presence of osmolytes. *Biophys. J.* **92,** 245–256 (2007).
17. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–990 (2007).
18. Heise, C. *et al.* An adenovirus E1A mutant that demonstrates potent and selective systemic anti-tumoral efficacy. *Nature Med.* **6,** 1134–1139 (2000).
19. Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl Acad. Sci. USA* **104,** 8311–8315 (2007).
20. Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Sci.* **17,** 1295–1307 (2008).
21. Wang, H.-G. H., Moran, E. & Yaciuk, P. E1A promotes association between p300 and pRB in multimeric complexes required for normal biological activity. *J. Virol.* **69,** 7917–7924 (1995).
22. Chan, H. M., Krstic-Demonacos, M., Smith, L., Demonacos, C. & La Thangue, N. B. Acetylation control of the retinoblastoma tumour-suppressor protein. *Nature Cell Biol.* **3,** 667–674 (2001).

23. Nguyen, D. X., Baglia, L. A., Huang, S.-M., Baker, C. M. & McCance, D. J. Acetylation regulates the differentiation-specific functions of the retinoblastoma protein. *EMBO J.* **23,** 1609–1618 (2004).

24. Koshland, D. E. Jr. The structural basis of negative cooperativity: receptors and enzymes. *Curr. Opin. Struct. Biol.* **6,** 757–761 (1996).

25. Sang, N., Avantaggiati, M. L. & Giordano, A. Roles of p300, pocket proteins, and hTBP in E1A-mediated transcriptional regulation and inhibition of p53 transactivation activity. *J. Cell. Biochem.* **66,** 277–285 (1997).

26. Green, M., Panesar, N. K. & Loewenstein, P. M. The transcription-repression domain of the adenovirus E1A oncoprotein targets p300 at the promoter. *Oncogene* **27,** 4446–4455 (2008).

27. Motlagh, H. N. & Hilser, V. J. Agonism/antagonism switching in allosteric ensembles. *Proc. Natl Acad. Sci. USA* **109,** 4134–4139 (2012).

28. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30,** 13–19 (2002).

29. Lee, C. W., Ferreon, J. C., Ferreon, A. C. M., Arai, M. A. & Wright, P. E. Graded enhancement of p53 binding to CBP/p300 by multisite phosphorylation. *Proc. Natl Acad. Sci. USA* **107,** 19290–19295 (2010).

30. Ferreon, A. C. M., Gambin, Y., Lemke, E. A. & Deniz, A. A. Interplay of α-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proc. Natl Acad. Sci. USA* **106,** 5645–5650 (2009).

METHODS

Sample preparation. Protein expression and purification were carried out as described[11]. The Ad2 E1A short constructs (E1A(CR1(27–105)), E1A(N-CR1(1–105)) and E1A(CR2(106–139))) were obtained via thrombin digestion of the longer E1A constructs (E1A(CR1-CR2(27–139)) or E1A(N-CR1-CR2(1–139)))[11]. All E1A Cys mutants used for ensemble fluorescence[29] or single-molecule fluorescence resonance energy transfer (smFRET)[30] experiments have the additional mutations C6S and C124S, which replace two natural Cys residues that are respectively located in CBP TAZ2 and pRb binding regions. Although C124 is in a conserved pRb LXCXE binding motif, it has been shown that a Cys to Ser mutation in the site exhibits marginal effects on E1A binding[31]. Alexa Fluor 488 and 594 (Molecular Probes) fluorescent dyes were attached at sites that are unlikely to cause structural perturbations or affect E1A binding to CBP TAZ2 or pRb (residue positions −3, 36, 88, 111 and 137, where residues 1–139 comprise the E1A sequence and positions −4 to −1 are the residues GSHM).

For direct E1A titrations against CBP TAZ2 and/or pRb monitored by ensemble fluorescence anisotropy, E1A constructs with single Cys (S36C for E1A(N(1–36)); E137C for E1A(CR2(106–139)); otherwise, S88C) were used to attach Alexa Fluor 594 probes. For ensemble competition experiments, the competing E1A ligands (E1A(N-CR1-CR2(1–139)), E1A(CR1-CR2(27–139)), E1A(N(1–36)), and E1A(CR2(106–139))) have the wild-type E1A sequence except for the G139 residue that was mutated to Trp for more accurate protein concentration determination by ultraviolet spectroscopy. For E1A(CR2(106–139)), protein with the wild-type sequence was used in direct titration measurements, with the E1A protein N-terminally labelled with Dylight594 (Thermo Scientific) NHS ester probe. To investigate the role of the E1A N terminus in the protein's binding properties, four sets of pair-labelled E1A constructs were used for smFRET studies: E1A(CR1(27–105)) and E1A(N-CR1(1–105)), 36C88C; and, E1A(CR1-CR2(27–139)) and E1A(N-CR1-CR2(1–139)), 36C88C, (−3)C111C, 36C137C.

All E1A single Cys mutants were labelled in 50 mM Tris, 6 M guanidine HCl, pH 7.2 using ~3–5 fold molar excess of maleimide dye. For E1A double Cys mutants, approximately 5 nmol of E1A were incubated with 1:3 concentration ratio of Alexa Fluor 488: Alexa Fluor 594 dye. Labelling reactions were run for 2 h at room temperature. All dye-labelled E1A samples were purified using an analytical C18 reverse-phase HPLC column, and were checked for correct mass and for incorporation of the Alexa dyes by MALDI–TOF mass spectrometry.

Ensemble fluorescence spectroscopy. Isothermal titrations in 20 mM Tris, 50 mM NaCl, 1 mM dithiothreitol (DTT), pH 7.0 at 21 °C were performed by monitoring ensemble fluorescence anisotropy. Two titration methods were used: direct protein–ligand titration and competition binding measurements[29]. Direct titrations were carried out by detecting fluorescence anisotropy changes in solutions containing 25 nM of dye-labelled E1A macromolecule (M) as a function of ligand (L) concentration (CBP TAZ2 or pRb). Dissociation constants (K_d) were determined using OriginPro 8.0 (OriginLab Corp.) by nonlinear least-squares (NLS) fitting of the data to a one-site binding model (see equation (2) below). To determine the goodness of fit and test the validity of the simplified model, simulations were performed using the fitted parameters and compared to the data on the basis of a more complete binding model that considers the macromolecule concentration (see the model defined below and described by equation (3)). Application of the more exact model is not feasible for the analysis of the ensemble fluorescence anisotropy data due to the number of fitting parameters. For cases of low-affinity binding, where $K_d \gg M$ (such as with the titration of E1A(CR1(27–105)) and E1A(CR2(106–139)) against pRb), the assumptions of the simplified one-site binding model become valid, as can be shown by simulations using the derived parameters as applied to the second model. For cases of high-affinity binding, where $K_d \ll M$, the estimates for K_d using the first binding model are not accurate. In such cases, an upper bound for the K_d was used (Supplementary Table 1). For the competition method, 25 nM labelled E1A were initially bound with 500–1,000 nM pRb or 350–500 nM CBP TAZ2, and competed with the unlabelled E1A counterpart to see the effect of the probes (for example, E1A(N-CR1(1–105)), 88C-Alexa Fluor594, competed against wild-type sequence E1A(N-CR1(1–105))). An estimate of the K_d from the direct titration was necessary to fit the K_d of the competing ligand.

Single-molecule spectroscopy. Single-molecule smFRET experiments were carried out as described previously[30] using a home-built laser confocal microscope system that uses an Axiovert 200 microscope (Zeiss). Excitation was achieved by focusing the 488-nm line of a 543-AP-A01 tunable argon-ion laser (Melles Griot) into the sample solution, 30 μm above a glass coverslip surface, using a water immersion objective (1.2 NA, ×63; Zeiss). The fluorescence emission was collected using the same objective, separated from the excitation light using a dichroic mirror (Q495LP; Chroma Tech. Corp.), spatially filtered using a 100 μm-pinhole then split into donor and acceptor components using a second dichroic mirror (560 DCXR; Chroma). The donor and acceptor signals were further filtered using an

HQ 525/50M band-pass filter (donor; Chroma) and a 590 LPV2 long-pass filter (acceptor; Chroma), then detected using SPCM-AQR-14 avalanche photodiode (APD) photon counting modules (Perkin-Elmer Optoelectronics). Photon counts were recorded using a photon counting card (PCI 6602; National Instruments) interfaced with a computer.

FRET efficiency (E_{FRET}) histograms were generated by using a two-channel data collection mode to simultaneously record donor and acceptor signals as a function of time, with a binning time of 500 μs. The donor–acceptor solutions used were ~100 pM in fluorophore concentration (that is, ~100 pM FRET-labelled E1A), ensuring that virtually all of the detected signals were from single molecules. The background counts, the leakage of donor emission into the acceptor channel (~8%) and the acceptor emission due to direct excitation (~5%) were determined in separate experiments, and used to correct the signals before FRET analysis. A threshold of 50 counts (the sum of signals from the two channels) was used to separate background noise from fluorescence signals, and E_{FRET} values were calculated for each accepted event using equation (1) and plotted in the form of histograms:

$$E_{FRET} = I_A/(I_A + \gamma I_D) \qquad (1)$$

I_D and I_A are the corrected donor and acceptor fluorescence intensities, respectively, and γ is a correction factor that is dependent on the donor and acceptor fluorescence quantum yields, and donor channel and acceptor channel detection efficiencies. Using the same experimental setup and FRET dye-pair, we previously measured γ to be approximately equal to unity[30]. Although the accuracy of the determined γ value is critical for measurement of inter-dye distances, γ does not have a part in the calculation of population distributions[30].

Direct detection of binding events using smFRET. Binding of unlabelled CBP TAZ2 and/or pRb to different constructs of E1A labelled with Alexa Fluor 488 (donor) and 594 (acceptor) (see 'Sample preparation' section above and Fig. 1b) was detected using smFRET at room temperature (~21 °C). The same solution conditions were used as for the ensemble fluorescence measurements (20 mM Tris, 50 mM NaCl, 1 mM DTT, pH 7.0). An average of ~5,000 single-molecule events was collected for each smFRET histogram measurement. In total, the complete set of smFRET titration data reported here comprise in excess of 700,000 events. Representative smFRET histograms are shown in Fig. 3a–d and Supplementary Figs 3–6.

E_{FRET} histograms were fitted to Gaussian functions by using OriginPro 8.0 with the peak positions, areas and widths at half height used as fitting parameters. For experimental conditions where E1A predominantly adopts a single binding state (that is, free, CBP TAZ2-bound, pRb-bound, or in ternary complex with CBP TAZ2 and pRb), smFRET histograms showed two peaks: one corresponding to the protein signal, and another to the 'zero peak', which is present in all histograms due to molecules with photo-bleached, missing, or non-fluorescent acceptor probe. These histograms of single populations or 'pure states' were used as references in determining (via independent NLS Gaussian fits) the characteristic E_{FRET} signatures of the different E1A binding states (see Supplementary Table 3). These precisely determined E_{FRET} values were then used as fixed parameter inputs in the analyses of smFRET histograms exhibiting resolved multiple 'protein peaks' that correspond to different E1A conformations (for example, unbound and CBP TAZ2-bound states). The areas under each protein peak determined by NLS Gaussian fitting were then used to calculate fractional populations (for example, fraction unbound and fraction CBP TAZ2-bound), which were analysed further as a function of ligand concentration (for example, [CBP TAZ2]) to determine binding constants (see Fig. 1a and Supplementary Table 2, and K_d determination method discussion below). The smFRET data presented in Supplementary Figs 3a–f, 4a–f, 5a–c and 6a–d were all analysed independently via NLS Gaussian fitting.

In some cases, smFRET histograms acquired under different solution conditions were analysed simultaneously, sharing fitting parameters that correspond to the same protein states. This global analysis was especially useful in cases where protein peaks were not resolved well or when E_{FRET} values cannot be satisfactorily determined independently using just the histograms of pure states. Fractional populations were calculated using the area parameters derived from global fitting and analysed further for K_d determination (see below). The smFRET data presented in Supplementary Fig. 5d–f were analysed via global NLS Gaussian fitting.

K_d determination by smFRET. Detection of macromolecular interactions in solution at single-molecule resolution holds a number of important advantages over ensemble methods, including the direct measurement of population distributions, the ability to experimentally handle aggregation-prone systems, and improved resolution in the study of high-affinity interactions.

Applying smFRET to derive K_d values for the binding of a ligand L to a macromolecule M, assuming that the bound and unbound populations of the FRET-labelled

macromolecule exhibit distinct FRET efficiencies, is straightforward and can be performed empirically, without the need for model fitting. K_d, or the ligand concentration at which the bound (ML) and unbound (M) macromolecule populations are equal, can be determined simply by titrating M with increasing concentration of L until the measured smFRET histogram shows approximately equal areas for the peaks associated with the two binding states. The process can then be repeated several times to achieve the desired precision.

Alternatively, population distributions can be similarly measured, then used as an experimental variable that depends on $[L]$ and analysed using a binding model. The model described by equation (2) (see below) assumes that the total ligand concentration $[L_T]$ is approximately equal to the concentration of free ligand, that is, $[M] \ll K_d$, a requirement that is easily achieved using smFRET, in which measurements are usually performed using 100 pM labelled molecules (or less).

Binding constants for the E1A–CBP TAZ2–pRb ternary system (Fig. 1a) were determined as a function of ligand concentration (that is, [CBP TAZ2] or [pRb]) using the fractional populations (that is, fraction bound and unbound) directly derived from the analyses of smFRET histogram data (described above). Fraction populations plotted against total ligand concentration (expressed in terms of $-\log[\text{ligand}]_{\text{total}}$ or pL_T) were analysed graphically using the general binding model, $ML \leftrightarrow M + L$, and fitted with OriginPro 8.0 using equation (2):

$$Y = (\alpha Y_M + Y_{ML})/(1 + \alpha) \qquad (2)$$

M represents a macromolecule binding to a ligand L, Y is the experimental observable (that is, fraction bound or unbound), Y_M and Y_{ML} are the binding transition baselines (that is, the constants 0 and 1, respectively, if using fraction bound as Y; otherwise, 1 and 0), $\alpha = 10^\zeta$, $\zeta = pL_T - pK_d$, $[\text{ligand}]_{\text{total}}$ represents both bound and unbound forms of L, $pK_d = -\log[K_d]$, and K_d is the dissociation constant. The model assumes that the concentration of unbound ligand is approximately equal to $[\text{ligand}]_{\text{total}}$, that is, the total concentration of the macromolecule E1A ($\sim$100 pM) is significantly less than the K_d values being measured, which in the case here are in the 1–50 nM range (see Supplementary Table 2). In addition, Y can be any observable/signal that is able to distinguish the different binding states, for example, E_{FRET} (see Supplementary Fig. 5c).

A more general expression describing the same model (that is, $ML \leftrightarrow M + L$) is given by equation (3):

$$Y = ([ML]/[M_T])(Y_{ML} - Y_M) + Y_M \qquad (3)$$

M_T is the total M concentration independent of ligation state, Y is the observable, Y_M and Y_{ML} are the binding transition baselines, and $[ML] = (-b - (b^2 - 4ac)^{0.5})/2a$, with $a = 1$, $b = -K_d - [M_T] - [L_T]$, and $c = [M_T][L_T]$. Presented in Supplementary Fig. 7 are simulations for ligand binding at different M_T, highlighting the advantage of single-molecule detection in resolving binding constants of high-affinity interactions.

Protein phase diagrams. Using the binding constants derived from ensemble and single-molecule measurements (see above and Supplementary Tables 1 and 2), phase diagrams were generated to visualize the ligand concentration dependence of E1A interaction with its binding partners CBP TAZ2 and pRb (Figs 3e–h). Detailed descriptions of the general properties of protein phase diagrams, and their construction and interpretation, are provided elsewhere[16,32].

Here, we use the reaction mechanism depicted in Fig. 1a to describe the coupled folding and binding of E1A with CBP TAZ2 and pRb. K_1 and $K_{1'}$, and K_2 and $K_{2'}$ are equilibrium constants for E1A binding to CBP TAZ2 in the absence and presence of pRb, and to pRb in the absence and presence of CBP TAZ2. Because the reaction scheme constitutes a complete thermodynamic cycle, it can be shown that $K_1/K_{1'} = K_2/K_{2'}$. 50% phase separation lines were constructed as previously described[32], using partition functions (Q_i) that describe each of the four binding states (that is, unfolded and unbound (U), folded and CBP TAZ2-bound (FL$_1$), folded and pRb-bound (FL$_2$) and ternary (FL$_1$L$_2$) states). For example, 50% phase separation lines between the U state and the three other binding states are calculated by equating Q_U with the sum of the remaining partition functions Q_{FL1}, Q_{FL2} and Q_{FL1L2}.

31. Kadri, Z. *et al.* Direct binding of pRb/E2F-2 to GATA-1 regulates maturation and terminal cell division during erythropoiesis. *PLoS Biol.* **7,** 1–15 (2009).
32. Rösgen, J. & Hinz, H. J. Phase diagrams: a graphical representation of linkage relations. *J. Mol. Biol.* **328,** 255–271 (2003).