






Allostery through DNA drives phenotype switching

Gabriel Rosenblum ¹✉, Nadav Elad², Haim Rozenberg ¹, Felix Wiggers ¹, Jakub Jungwirth ³ & Hagen Hofmann ¹✉

Allostery is a pervasive principle to regulate protein function. Growing evidence suggests that also DNA is capable of transmitting allosteric signals. Yet, whether and how DNA-mediated allostery plays a regulatory role in gene expression remained unclear. Here, we show that DNA indeed transmits allosteric signals over long distances to boost the binding cooperativity of transcription factors. Phenotype switching in *Bacillus subtilis* requires an all-or-none promoter binding of multiple ComK proteins. We use single-molecule FRET to demonstrate that ComK-binding at one promoter site increases affinity at a distant site. Cryo-EM structures of the complex between ComK and its promoter demonstrate that this coupling is due to mechanical forces that alter DNA curvature. Modifications of the spacer between sites tune cooperativity and show how to control allostery, which allows a fine-tuning of the dynamic properties of genetic circuits.

¹Department of Chemical and Structural Biology, Rehovot, Israel. ²Department of Chemical Research Support, Rehovot, Israel. ³Department of Chemical and Biological Physics, Weizmann Institute of Science, Rehovot, Israel. ✉email: gabriel.rosenblum@weizmann.ac.il; hagen.hofmann@weizmann.ac.il

Allostery is the structural coupling between ligand sites in biomolecules. Binding of a ligand to one site facilitates or hampers the binding of a second ligand to a distant site^{1–5}. The resulting cooperativity regulates the activity of many proteins and molecular machines^{6–8}, but it is also key for the behaviour of genetic circuits with binary⁹, oscillatory¹⁰, excitable^{11,12}, or pulsing¹³ dynamics. The past decades have seen growing evidence that allostery is also an inherent property of DNA^{14–27}, which has far reaching consequences for our understanding of promoter sequences. Yet, most insights on DNA-mediated allostery upon transcription factor (TF) binding were either based on artificial promoters^{14,28} or found to be short-ranged^{19,20,25}. Whether natural promoters evolved to efficiently transmit allosteric signals across many nanometres remained largely unclear. Here, we show that *Bacillus subtilis* bacteria utilize long-range allostery in a stochastic and reversible phenotype switch (Fig. 1a). In the competent phenotype, *B. subtilis* can take up DNA from the medium^{11,29}. The master regulator of the switch from vegetative to competent cells is the TF ComK^{11,12}. The model postulates a positive feedback regulation of *comK* gene expression once ComK levels stochastically surpass a critical threshold (Fig. 1b)³⁰. The critical threshold acts like an analogue-to-digital converter: the ComK target promoter is inactive at low concentrations, but it switches cooperatively to an active state within a narrow ComK concentration range (Fig. 1c)^{11,12}. Yet, how the promoter mediates cooperative ComK binding is unknown. Using single-molecule Förster resonance energy transfer (smFRET) and cryo-electron microscopy (cryo-EM) we show how ComK binding at one site enhances binding to a distant site via allosteric changes in DNA.

Results

Distant binding sites communicate. ComK target promoters consist of two elements (box 1 and box 2, hereafter) separated by spacer sequences of variable length (Fig. 1d)³¹. Three spacers of 8, 18, and 31 base pairs (bp) are known in the ComK response genes *addAB*, *comG*, and *comK*, respectively³¹. Each box in a promoter contains an adenine–thymine (AT)-rich sequence (Fig. 1d). Such A-tracts are ubiquitous throughout all kingdoms of life,

presumably due to their potential to curve DNA^{32,33}. To assess the structural properties of the promoter, we engineered a set of *comG* promoters (18 bp spacer) that we site-specifically labelled with donor and acceptor fluorophores for confocal smFRET measurements (Supplementary Fig. 2 and Supplementary Table 1). When we compared the experimental FRET efficiencies with theoretical values calculated for extended B-type DNA³⁴, we noted substantial deviations across one box, suggesting an overall curved topology (Fig. 2a). Given that DNA curvature scales with DNA stiffness³⁵, which is sensitive to the protonation state of the phosphate backbone, we repeated the experiment under more acidic conditions (pH 4.0, Fig. 2b). Indeed, we obtained reduced FRET values that showed a better apparent agreement with the calculated B-DNA profile. This result indicated that the *comG* promoter is curved at neutral pH. Importantly, the correlation between donor fluorescence lifetimes and FRET values is close to the expectation from Förster theory (see ‘Methods’), which indicates that the structure of the *comG* promoter does not fluctuate substantially at the length scale probed by FRET (Fig. 2a, b inset). Yet, small-amplitude motions cannot be ruled out and require a more sophisticated analysis to be identified³⁶.

To probe ComK binding cooperativity, we labelled the *comG* promoter with a FRET-pair either at each box or within the spacer (Fig. 2c). In the absence of ComK, we find FRET values between 0.4 and 0.5. Upon addition of ComK, the FRET peaks shift to lower values for box 1 and box 2, equivalent to a distance increase, whereas the FRET efficiency increased for the spacer region, corresponding to a slight distance reduction (Fig. 2c). Whereas the measured changes in FRET efficiencies were small, we were able to accurately derive relative populations of free and ComK-bound *comG* molecules by fitting our data to superpositions of two-state Gaussian peaks (Fig. 2c, Supplementary Fig. 2, and ‘Methods’). We found that the fraction of ComK–*comG* complexes indeed increased in a sigmoidal fashion with ComK concentrations (Fig. 2d), in agreement with the cooperative mode of binding required by mathematical models of the phenotype switch^{11,12,37}. Fits with the Hill equation (Fig. 2d)³⁸ result in Hill exponents of $n = 3.6 \pm 0.1$ for box 1 and $n = 3.4 \pm 0.3$ for box 2. Independent controls confirm that ComK is monomeric in solution

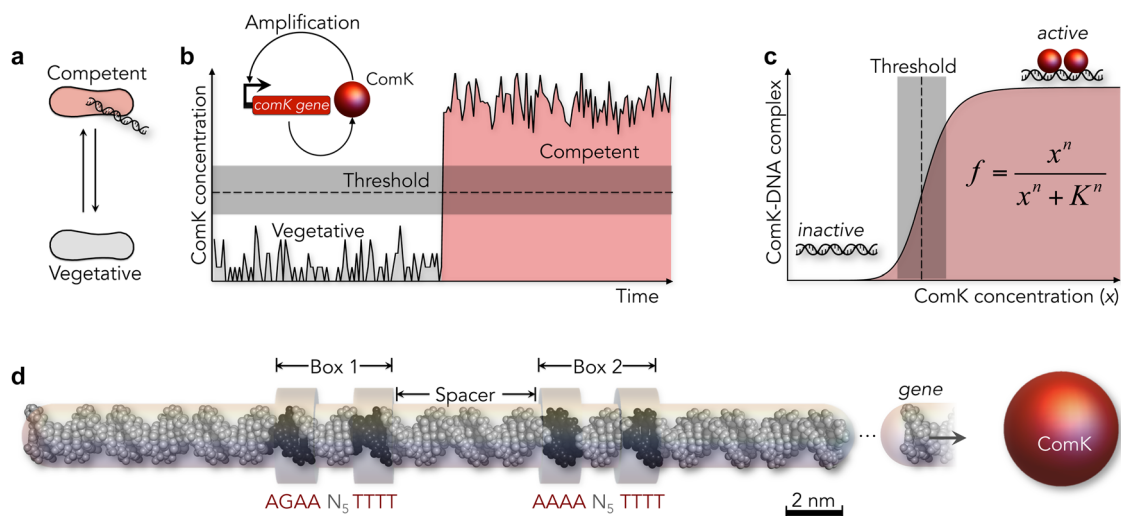


Fig. 1 Schematics describing the model for phenotype switching in *B. subtilis*. **a** *B. subtilis* can reversibly switch to a competent phenotype (red). **b** Switching is triggered by copy number fluctuations of ComK (grey) that, once stochastically exceeding a threshold (dashed line), cause its auto-amplification to high concentrations (red). Inset: scheme of the auto-amplification of ComK. **c** The threshold (grey) requires a cooperative DNA binding of ComK, modelled by the Hill equation with cooperativity n , affinity K , and fraction of complexes f . **d** ComK promoters consist of two boxes separated by a spacer. Here, the *comG* promoter with an 18 bp spacer and an imperfect A-tract in box 1 is shown. For size comparison, a ComK monomer (Stokes radius: 2.07 ± 0.03 nm, Supplementary Fig. 1) is shown as red sphere.

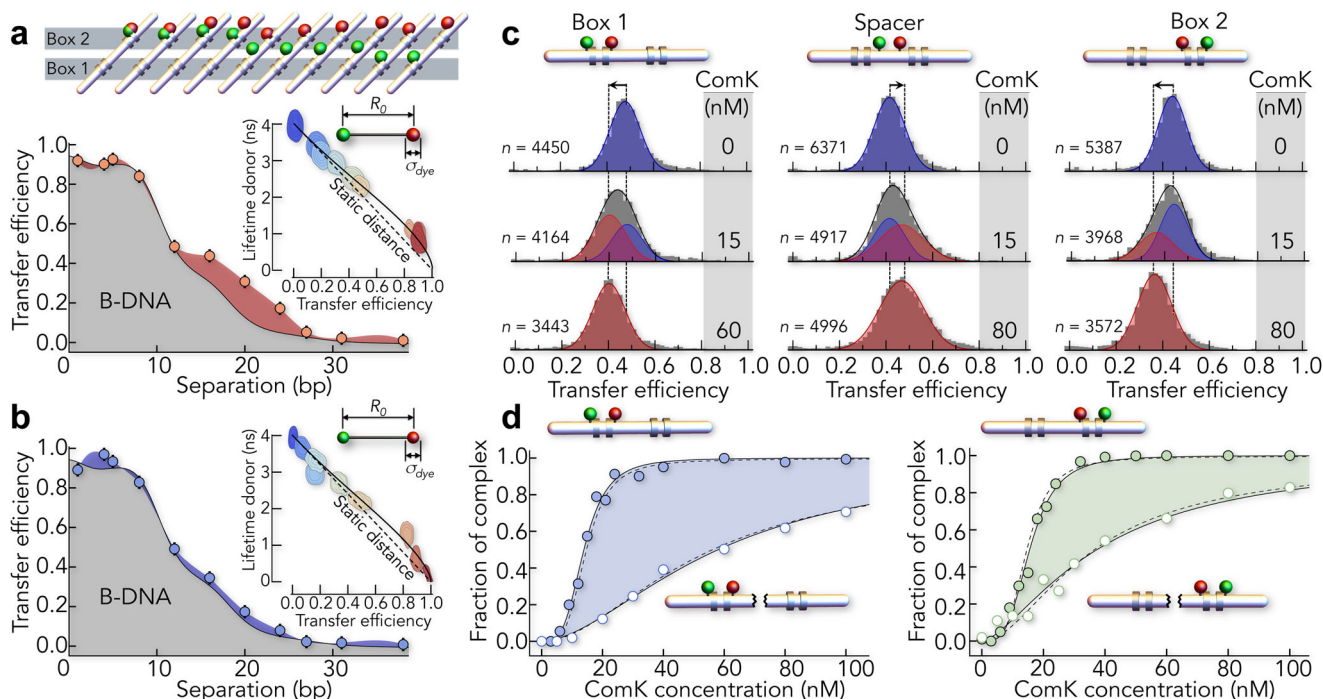


Fig. 2 DNA-binding cooperativity of ComK probed with smFRET. **a, b** Mapping of distances in box 2 with FRET at higher (**a** pH 7) and lower (**b** pH 4) persistence length. Grey region indicates the expected dependence for B-DNA³⁴. Insets: fluorescence lifetimes of the donor as function of the transfer efficiency. Dashed line is the expectation for a single distance and solid line is a fit with a Gaussian distance distribution to account for dye flexibility with widths of 0.76 and 0.66 nm for pH 7 and pH 4, respectively (see ‘Methods’). The dimension of the dye clouds is schematically indicated compared to the Förster distance $R_0 = 5.4$ nm. **c** smFRET histograms of the *comG* promoter (93 bp) with an 18 bp spacer indicate a change in DNA on ComK binding. FRET is probed in box 1 (left), in the spacer (middle), and in box 2 (right). Solid lines (black) are fits with superpositions of two Gaussian peaks for the free (blue) and ComK-bound (red) promoter. **d** Fraction of the ComK-DNA complexes probed in box 1 (left) and box 2 (right) in the presence (filled circles) and absence (empty circles) of the respective other box. The single-box constructs are 43 bp in length. Lines are fits with the Hill equation (solid) and with a mechanistic binding model (dashed, ‘Methods’ Eqs. (20) and (21)).

(Supplementary Fig. 1 and Supplementary Table 3). To interrogate coupling between box 1 and box 2, we created shorter single-box constructs (Fig. 2d and Supplementary Table 2). Remarkably, we obtained $n = 1.7 \pm 0.1$ for box 1 and $n = 1.5 \pm 0.1$ for box 2. Since Hill exponents provide a lower limit of the number of binding sites, these values agree with earlier reports of a 2:1 (protein:DNA) stoichiometry per box³¹ at least. The twofold higher Hill exponents in the natural promoter show that the FRET probes on one box also report on ComK binding to the other box, which requires an allosteric communication between the boxes.

Structure of the complex. One classical mechanism to achieve allostery is DNA looping that allows contacts between proteins bound to different boxes^{31,39}. Alternatively, ComK may bridge the boxes by interacting with the spacer, which could explain the distance decrease in the spacer upon binding of ComK (Fig. 2c, middle). To distinguish between these options, we used single-particle cryo-EM that overcomes the strong aggregation tendency and low stability of ComK (Supplementary Fig. 3). We determined 2D class averages of ComK bound to promoters with spacers of 8 bp (*addAB*: 57 bp, 35.2 kDa), 18 bp (*comG*: 93 bp, 57.5 kDa), and 31 bp (*comK*: 93 bp, 57.5 kDa). Isolated DNA complexes of the expected size were readily identified in the cryo-EM micrographs and 2D class averages revealed two major populations (Fig. 3a).

While the first population was pure promoter DNA, the second population contained DNA with extra density. Three observations are notable: (i) the extra density bound to DNA is confined to two locations that agree with the position of box 1 and box 2, thus

assigning this density to ComK; (ii) the DNA in these complexes is not looped, and (iii) only a few particles with ComK density on only one box are found for promoters with 8 and 31 bp spacers (Fig. 3b). The first observation demonstrates that ComK binds specifically to box 1 and box 2, but not to the spacer between the boxes. Second, DNA looping, as suggested in previous studies^{31,40}, can be excluded as an allosteric mechanism by the 2D class averages. Indeed, we did not find an increase in FRET efficiency upon ComK binding to promoters labelled at the 5' and 3' end (Supplementary Fig. 4). Third, the low abundance of particles with ComK on only one box even under the vitrified conditions used in cryo-EM (Fig. 3b) suggests an all-or-none binding of the protein, in accord with the high cooperativity found with smFRET at room temperature (Fig. 2d). Since Hill exponents can also be interpreted as variance of the distribution of bound ligands relative to that of the binomial distribution⁴¹, we computed the Hill exponents (n_{EM}) from the cryo-EM distributions (see ‘Methods’). These values agreed qualitatively with those obtained from smFRET (Fig. 3b and Supplementary Table 4). To determine the precise arrangement of ComK on the DNA, we reconstructed the 3D maps of the complexes with spacers of 8 and 18 bp (Fig. 3c and Supplementary Figs. 5–7). To overcome the severe preferred orientation of the complexes relative to the electron beam, cryo-EM images were collected while tilting the sample (see ‘Methods’). The 3D maps show that ComK molecules in the two boxes face the same side of the DNA. In addition, we find a pseudo C_2 -symmetry of the protein density in each box, indicating two molecules of ComK per box, i.e., four per promoter, in excellent accord with the Hill exponents (Fig. 2d) and a previous report³¹. We further noted that ComK densities at individual boxes were

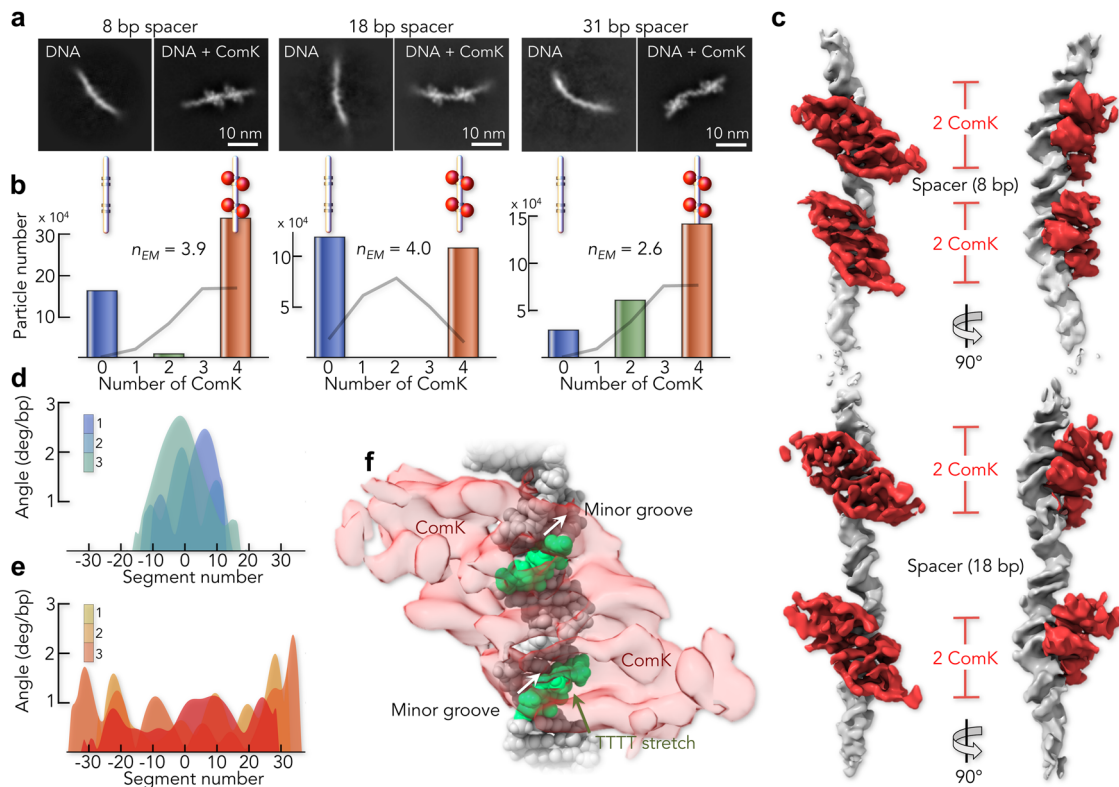


Fig. 3 Single-particle cryo-EM analysis on ComK-DNA complexes. **a** Class averages of free DNA and the ComK-DNA complex for the promoters with an 8 bp (left), 18 bp (middle), and 31 bp (right) spacer. The scale bar is indicated. **b** Distribution of particles with different numbers of bound ComK molecules for the promoters with spacers of length 8 bp (left), 18 bp (middle), and 31 bp (right). Grey line indicates the binomial distribution for the case of no cooperativity. The Hill exponent calculated from these distributions (n_{EM}) is indicated. **c** 3D reconstruction of ComK bound to the promoter with an 8 bp (top) and 18 bp (bottom) spacer (Supplementary Figs. 5–7). DNA (grey) and ComK (red) densities were segmented following fitting of an atomic DNA model (see ‘Methods’). **d, e** Curvature angles for free DNA (**d**) and the ComK-DNA complex (**e**). The 3D classes from which curvature angles were derived are shown in Supplementary Fig. 6. Colour code indicates the 3D classes. Angles between successive segments (segment length: 0.34 nm = 1 bp) for different classes were aligned with respect to the middle of the DNA. Averaged over all segments, the difference in curvature angles is 0.6°/bp between free and ComK-bound DNA. **f** Orientation of a ComK dimer (transparent red) relative to the A-tracts in a single box of the *addAB* promoter DNA (8 bp spacer). ComK faces the thymine bases (green) in the minor groove (indicated). The ComK density is shown relative to the fitted DNA model.

continuous, which hints towards significant protein–protein contacts within a box (Fig. 3c). By contrast, we detected no ComK interactions across the spacer region, which suggests DNA-mediated allostery as source for the high binding cooperativity. To determine the allosteric mechanism, we analysed the DNA curvature of free (Fig. 3d) and bound promoters (Fig. 3e), and found a significant reduction in the presence of ComK, which hints at a potential mechanism. ComK is known to interact with the minor groove of DNA³¹, which is significantly narrower in A-tracts compared to B-DNA⁴². Indeed, a fit of our 3D density with an atomistic representation of the *addAB* promoter (8 bp spacer) that agrees well with our smFRET values (Supplementary Fig. 8, and Supplementary Tables 5 and 6), suggests that ComK faces the minor groove by forming contacts with the thymine bases of the A-tracts (Fig. 3f). Widening the minor groove by ComK would lower DNA curvature around a box, thus affecting the width of the minor groove at the second box. Transmitting such curvature changes from one box to another likely involves structural alterations of the spacer, which might cause the spacer contraction observed with smFRET (Fig. 2c). Importantly, the model stipulates that DNA-mediated communication between boxes ought to be dependent on the spacer length.

Mechanical force transduction in DNA. To test this mechanism, we altered the spacer length and measured the effect on cooperativity using a Koshland-Nemethy-Filmer (KNF) binding

model⁵ that is derived from the cryo-EM structures (Fig. 4a). Each box has two binding sites for ComK. The first ComK binds a box with the association constant K . The second molecule binds the same box with σ -fold increased affinity due to protein–protein contacts between the ComK molecules. Once the first box is saturated with ComK, the altered DNA curvature increases the affinity at the second box by a factor J . The quantity $\Delta g_j = -k_B T \log J$ is the DNA-mediated coupling free energy between two boxes. Using this model, we examined the coupling between boxes using the three promoters *addAB*, *comG*, and *comK* with spacer lengths of 8, 18, and 31 bp, respectively (Fig. 4b). When analysing the binding isotherms (Fig. 4b), we found a significant decrease in the Hill exponent with increasing spacer length, which is also found in the ligand distribution derived from cryo-EM (Fig. 3b). Similarly, the coupling free energy ($-\Delta g_j$) drops from $5.8 \pm 0.9 k_B T$ (8 bp) over $4.5 \pm 1.2 k_B T$ (18 bp) to $1.9 \pm 0.1 k_B T$ (31 bp). The result shows that spacer length is key for allostery, in line with the idea that reducing promoter curvature is the main origin for allosteric signal transmission. Compared to allosteric effects found previously in artificial promoters without curvature^{14,23}, the coupling free energies are almost threefold larger, which converts to a 40-fold stronger effect on the association constants. Hence, changes in curvature, though not mandatory^{14,23}, greatly enhance DNA-mediated allostery. An elastic-coupling theory of DNA bending predicts a decrease of Δg_j with increasing spacer length¹⁸. Yet, this decay should be

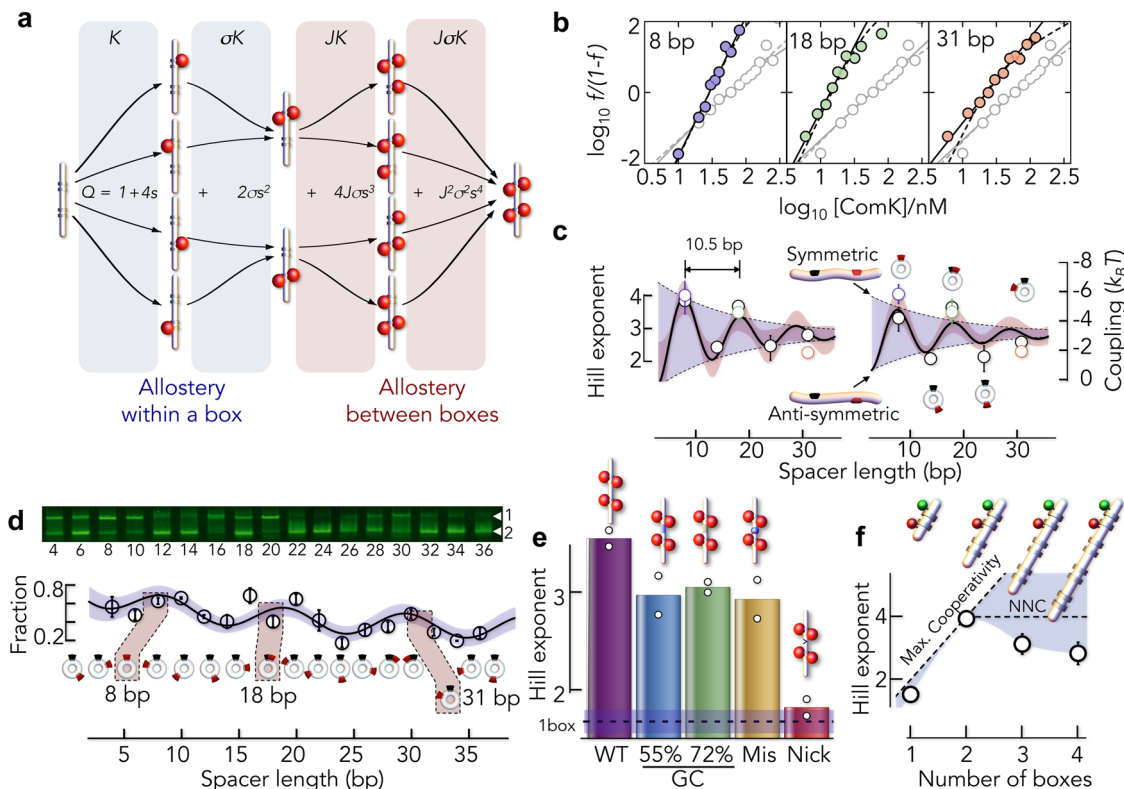


Fig. 4 Tuning allostery. **a** Schematics of the KNF binding model (‘Methods’ Eqs. (20) and (21)) with DNA (grey) and ComK (red). A single box is first occupied by two ComK molecules before the second box is bound. The model contains a microscopic association constant K and the allosteric parameters σ and J due to allostery within a box and between boxes, respectively. Binding constants for each step are identical for all paths. (Middle) Binding polynomial (Q) of the model with $s = K [\text{ComK}]$. **b** Hill plots of the binding isotherms for promoters with varying spacer length (indicated) in comparison to an isolated box (grey). Solid lines are fits with the Hill equation and dashed lines are fits with the model in **a**. **c** Hill exponents (left) and inter-box coupling free energies (right) as function of the spacer length for natural promoters (coloured symbols) and for designed promoters (Supplementary Table 2). Solid black lines are fits with the elastic-coupling model (‘Methods’ Eqs. (26) and (27)). Blue shaded areas indicate the spacer-dependent damping for symmetric and antisymmetric arrangements of the boxes as indicated schematically. Red shaded areas are the 90% confidence interval of the fit. Circles represent the mean and error bars indicate the minimal and maximal values from independent measurements ($n \geq 2$). **d** Gel assay to screen ComK binding to promoters with different spacer lengths with free DNA (lower band: 2) and ComK-bound DNA (upper band: 1) at a ComK concentration of 300 nM. (Bottom) Relative fraction of the ComK-DNA band as function of the spacer length. Promoters with 8, 18, and 31 bp are highlighted. Error bars indicate \pm SD of independent triplicates. Solid line is a fit with a sum of a cosine and exponential function. Shaded area is the 90% confidence band. **e** Hill exponents for *comG* promoters with modified spacer. White circles indicate the two independent measurements. The dashed line indicates the limit of an isolated box. **f** Hill exponents of artificial promoters (8 bp spacer) with multiple boxes (circles). The diagonal line indicates maximal cooperativity and the horizontal line indicates maximum nearest-neighbour cooperativity (NNC). Error bars of the Hill exponents represent the error of the fit with Eq. (22).

modulated by the helical periodicity of the DNA (see ‘Methods’). For instance, boxes in the natural promoters *addAB* and *comG* are one and two helical turns away from each other, respectively, thus leading to symmetric arrangements in which ComK always faces the same side of the DNA (Fig. 3c). We expect antisymmetric arrangements with non-integer turns to have significantly lower cooperativity (see ‘Methods’) ¹⁸. We therefore designed five promoters with 8, 14, 18, 24, and 31 bp spacers (Supplementary Table 2). Indeed, we find that non-integer turns show much lower cooperativity (Fig. 4c and Supplementary Table 4), resulting in a modulation of Hill exponents and coupling free energies. To determine the modulation period more precisely, we performed a qualitative gel assay that probes the fraction of ComK-DNA complexes and shows good agreement with the B-DNA periodicity of 10.5 bp (Fig. 4d). With this value, we used elastic-coupling theory ¹⁸ to fit our experimental Hill coefficients and coupling free energies (Fig. 4c, and see ‘Methods’ Eqs. (26) and (27)). The model predicts a dampening of the allosteric coupling with increasing spacer length. Dampening is characterized by the decay length $\xi = (k_B T l_p / f)^{1/2}$ with the persistence length $l_p = 40$ nm (ref. ⁴³) and a tension force f . We found $\xi = 14 \pm 8$ bp and

$f \sim 7$ pN. The model also provides the average change in bending $1.1 \pm 0.3^\circ/\text{bp}$ (see ‘Methods’), which is in good accord with the average change determined from the cryo-EM structures ($0.6^\circ/\text{bp}$; Fig. 3d, e). Yet, since bending is not uniform across the promoter, any comparison of average bending angles between cryo-EM and fits with the elastic-coupling theory remain qualitative.

Tuning allostery via DNA sequences. To understand the sequence determinants of allostery, we altered the spacer sequence in *comG* by increasing its GC content, thus increasing DNA stability. The GC content of the native spacer (40%) is close to the genome average ⁴⁴. We generated two random spacer sequences of higher GC content (55 and 72%, Supplementary Table 2). Indeed, increasing GC content lowered cooperativity (Fig. 4e). Nevertheless, the Hill exponents remained above the values found for the isolated boxes, suggesting that spacer sequences fine-tune allostery at best. Next, we tested whether allostery requires a correct base pairing by introducing a mismatch of four bases at the centre of the spacer. However, even though the Hill exponent is reduced compared to wild-type *comG*, it is comparable to those found for increased GC contents

(Fig. 4e). Remarkably, this result suggests that the allosteric communication between boxes does not stringently require a continuous base pair stacking in the spacer. These results point at the DNA backbone as major determinant for allosteric signal transmission. Indeed, when we introduce a nick in the spacer, cooperativity is abolished and we obtain a Hill exponent of $n = 1.8 \pm 0.1$, i.e., close to the value found for the isolated boxes (Fig. 4e). Clearly, nicked DNA has greater flexibility and a quantitative comparison with the effects found in double-stranded DNA has to be taken with care. Yet, given that neither a higher GC content nor a mismatch show comparable effects, we conclude that curvature-induced tension between the boxes propagates mainly via an intact DNA backbone, but it can be fine-tuned by the spacer sequence and length.

Finally, we tested promoters with multiple boxes. In fact, the *comG* promoter contains an A-tract located 10 bp upstream of the first box (Supplementary Table 2)⁴⁰ that is not occupied in our cryo-EM structure and controls indeed show that ComK is incapable of binding isolated A-tracts (Supplementary Fig. 9 and Supplementary Table 7). However, increasing the number of perfect boxes may be a simple strategy to further boost cooperativity. Yet, even the presence of four boxes with short spacing (8 bp) does not increase the Hill exponent beyond the value found for the promoter with two boxes. This implies that curvature changes do not propagate across multiple boxes, but rather mediate nearest-neighbour communication (Fig. 4f).

Discussion

Our results show that DNA-mediated allostery generates high cooperativity in TF binding via mechanical deformations of the DNA across distances larger than 6 nm (18 bp). Notably, this mechanism is a built-in feature of natural ComK promoters to filter copy number fluctuations of ComK (Fig. 1b, c): copy numbers below the midpoint of the Hill curve leave the promoter unoccupied, those above the midpoint cause near full occupation. With increasing cooperativity, this filter increases the variability (noise) of free and bound promoters, which would also affect the ratio of vegetative and competent cells in a population. However, even if genetic circuits function in the deterministic regime of high copy numbers, sigmoidal dose responses of gene expression activity with TF concentrations are key for their dynamics^{45–47}. Altering the steepness of this response, i.e., the molecular cooperativity, will unambiguously alter these dynamics^{10,48}. Hence, besides the architecture of wiring genes^{49,50}, designing dose-response curves via promoter sequences will provide a more precise level of engineering the dynamics of gene networks in the future.

Methods

DNA labelling and duplex preparation. Single-stranded DNA constructs (Supplementary Tables 1 and 2) were synthesized and HPLC-purified by the Keck Biotechnology Resource Laboratory (Yale School of Medicine). For labelling with the amino group-reactive dyes AlexaFluor 488 (donor) and AlexaFluor 594 (acceptor) NHS ester (Thermo Fisher), thymine or adenine bases at the labelling positions contained an amino modifier C6-dT/dA CE phosphoramidite (Glen Research; Supplementary Tables 1 and 2). In constructs used for smFRET ComK-binding experiments, the positions of the modified A- and T-bases were chosen such that nucleotides that had previously been shown to be protected by ComK binding³¹ were avoided. For labelling, we followed the procedure by Masoud et al.⁵¹ with the exception that the labelling reaction was incubated for 1 h at 45 °C. The donor and acceptor dyes were used to label the forward and reverse ssDNA strands, respectively. Reversed-phase HPLC was used to remove unreacted dye, and unlabelled ssDNA. To this end, a ZORBAX Eclipse Plus C18 (3.5 µm) column (Agilent) was used, equilibrated with TEAA buffer (0.1 M trimethylamine buffered with acetic acid glacial to pH 6.5 and supplemented with 5% acetonitrile). A gradient of 7–30% acetonitrile over 40 ml was used for the separation. Labelled fractions were dried overnight (SpeedVac) and dissolved in 50 µl ddH₂O. DNA and dye concentrations were quantified by UV-VIS photometry, using the extinction coefficients of the free dyes $\epsilon_{495} = 73,000 \text{ M}^{-1} \text{ cm}^{-1}$ (AlexaFluor 488) and $\epsilon_{590} = 92,000 \text{ M}^{-1} \text{ cm}^{-1}$

(AlexaFluor 594). Only fractions containing 1:1 ratio of dye:ssDNA were used for smFRET experiments. Annealing of ssDNA donor and acceptor strands for smFRET experiments was obtained by mixing 2 pmol donor-labelled ssDNA with 3 pmol acceptor-labelled ssDNA in 50 mM Tris pH 7.5, 0.1 M NaCl, and 3 mM Mg₂Cl. The samples were heated to 95 °C and cooled gradually over 1 h in a PCR cycler. Unlabelled DNA duplexes for cryo-EM and EMSAs (electrophoretic mobility shift assays) were obtained by mixing a 1:1 ratio of forward and reverse ssDNA strands, using the same annealing procedure.

Expression and purification of ComK. The sequence of ComK (Uniprot ID: P40396), codon-optimized for expression in *Escherichia coli* (Genscript; Supplementary Table 3), was inserted into the NcoI and HindIII sites of a pET28b vector, and the protein was expressed in inclusion bodies (IBs) in BL21(DE3). Bacteria were grown in 4 l LB medium containing 50 µg/ml kanamycin. Cultures were induced at OD₆₀₀ 0.7 with 0.3 mM IPTG. After 4 h at 37 °C, the cells were harvested by centrifugation and the cell pellets were re-suspended in lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 100 µM PMSF, 1.2 µg/ml leupeptin, and 1 µM pepstatin A). After sonication (70% power, pulse on 2 s, pulse off 10 s, 60 cycles, on ice) and centrifugation (48,200 × g, 30 min at 4 °C), the IB precipitate was mixed with 15 ml of 60 mM EDTA, 6% Triton, 1.5 M NaCl, homogenized, and incubated at 4 °C for 1 h to solubilize membrane debris. After centrifugation (48,200 × g, 30 min at 4 °C), the resulting IB precipitate was supplemented with 0.1 M Tris pH 8, 1 mM EDTA, homogenized, and centrifuged using the same conditions. Finally, the IB-pellet was re-solubilized (>2 h, 4 °C) in 50 mM Tris pH 8.0, 6 M GdmCl, 10 mM imidazole, containing 10 mM DTT. After complete resolubilization, the buffer was exchanged to 50 mM Tris pH 8.0, 6 M GdmCl, and 10 mM imidazole, using a HiPrep desalting column (26/10, GE) to remove DTT. The protein containing fractions were then loaded on a HisTrap Ni-NTA column (5 ml, GE) and the protein was eluted with 50 mM Tris pH 8.0, 6 M GdmCl, and 0.5 M imidazole, using a linear gradient of 3 ml/min for 10 min. The eluted protein (20 ml) was supplemented with 0.1 mg/ml HRV3C and 5 mM DTT and the solution was dialysed overnight (4 °C) against 300 volumes of cleavage buffer (50 mM Tris and 100 mM L-Arg at pH 7, 1 mM DTT). HRV3C protease refolds efficiently under these conditions. Due to the strong aggregation propensity of ComK, the protein partially precipitates during this procedure. The protein was fully precipitated by the addition of an equivalent volume of a saturated ammonium sulfate solution. The precipitate was collected by centrifugation (48,200 × g, for 20 min at 4 °C) and the protein pellet was dissolved in 50 mM Tris pH 8.0, 6 M GdmCl, and 10 mM imidazole. The cleaved His-tag and HRV3C protease were removed using Ni-NTA affinity chromatography (5 ml HisTrap column, GE). Cleaved, i.e., untagged protein was obtained in the flow-through fractions, while the purification tags and HRV3C protease were retained on the column. Finally, the buffer of the ComK-containing fractions was exchanged to 0.5 M L-Arg pH 7.3 using a HiTrap desalting column (5 ml, GE). The protein concentration was determined using an extinction coefficient of $\epsilon_{280} = 24,660 \text{ M}^{-1} \text{ cm}^{-1}$ computed from the amino acid sequence of ComK. Since ComK contains four cysteine residues, we added 10 mM DTT and stored the samples at –80 °C. For mass determination, the purified ComK in 0.5 M L-Arg was injected into a combined LC-MS system (Waters ACUITY UPLC class H) equipped with a C4 column (300 Å, 1.7 µm, 21 mm × 100 mm) and the mass was determined in positive ion mode using electrospray ionization using a desolvation temperature of 500 °C, with flow rate of 1000 l/h. The voltage was 0.69 kV for the capillary and 46 V for the cone. Spectra, de-convoluted with the MaxEnt1 software, resulted in a single peak with a mass of $22,652 \pm 10 \text{ Da}$, in good agreement with a calculated mass of 22,640 Da.

Labelling of ComK for 2fFCS experiments. Singly labelled ComK was used to determine the oligomeric state of ComK in solution using two-focus fluorescence correlation spectroscopy (2fFCS)⁵². Since ComK contains four intrinsic cysteine residues at positions 31, 39, 55, and 102 (Supplementary Table 3), our approach was to replace them by alanine. However, the resulting ΔCys variant was incapable of binding DNA at protein concentrations $\geq 100 \text{ nM}$ as demonstrated in smFRET experiments using doubly labelled *comG* (Supplementary Fig. 1d). The result surprises given the facts (i) that ComK is a cytosolic protein that acts in a reducing environment and (ii) that all our smFRET experiments with wild-type ComK are performed under strongly reducing conditions (20 mM DTT). The result shows that the reduced cysteine residues are either crucial for the structural integrity of ComK, e.g., as metal binding centres, and/or are key for the interaction with the DNA.

To obtain active and fluorescently labelled ComK for 2fFCS measurements, we therefore introduced an additional well-accessible cysteine at the N-terminus of ComK (N-Cys ComK) for labelling. We checked that the N-terminal cysteine is labelled preferentially (Supplementary Fig. 1a inset). To this end, wild-type ComK and the variant N-Cys ComK were diluted to 10 µM and 0.9 mol equivalents of AlexaFluor 488 C5 maleimide were added to the solution. Samples (50 µl) were drawn after different times and quenched by the addition of 100 mM DTT. The samples were subjected to C18-reversed-phase chromatography and the labelling ratio of the protein elution was determined from the absorbance at 280 and 493 nm. For 2fFCS sample preparation, we added 1.4 mol equivalents of AlexaFluor 488 C5 maleimide to 46 µM ComK in 0.5 M arginine-HCl, pH 7.3. After 30 min at room temperature, the reaction was quenched by the addition of 10 mM DTT and

the buffer was exchanged to 50 mM Tris-HCl pH 8.0, and 6 M GdmCl. Unlabelled ComK was separated from labelled ComK using reversed-phase chromatography (RP-C18 column) in 0.1% TFA in ddH₂O. The protein was eluted using a gradient of 25–50% acetonitrile at a flow rate of 1 ml/min for 40 min. The fraction containing labelled ComK was first lyophilized and then dissolved in 50 mM Tris-HCl pH 8.0, and 6 M guanidinium chloride (GdmCl) buffer. A UV-Vis quantification confirmed a 1:1 ratio of ComK to dye.

Determination of the dimerization state without labelling ComK. To obtain a labelling-independent confirmation that ComK is monomeric in solution, we performed quantitative size-exclusion chromatography experiments. Here, the exclusion volume of a Superdex 75 column (GE Healthcare) was first determined using dextran-blue and the total volume was determined by ascorbic acid. In a second step, we performed a calibration of the column using standard proteins (LMW, GE Healthcare) of known size in 20 mM Tris pH 7.0, 50 mM L-Arg, 5 mM MgCl₂, 150 mM KCl, 20 mM DTT, and 0.001% Tween-20 (Supplementary Fig. 1c). Finally, we injected 3 and 4 μM ComK. The elution volume of ComK was in accord with a mass of 20.3 ± 0.8 kDa, indicating that ComK is a monomer at these concentrations⁵³.

Optimizing solution conditions to prevent ComK aggregation. Our results show that ComK aggregates at concentrations as low as 2 μM (Supplementary Fig. 3c, d). High concentrations of L-Arg (0.5 M) effectively prevent this aggregation up to a concentration of >100 μM. However, our experiments also show that such high L-Arg concentrations suppress the binding of ComK to DNA. A quantification of the aggregation of ComK using UV/VIS photometry revealed that at least 50 mM L-Arg (Supplementary Fig. 3) is required to ensure soluble ComK at a concentration of 2 μM. Even though the L-Arg concentration exceeds the cytosol concentration of 0.57 mM (*E. coli*)⁵⁴, our control experiments show that the conformation of the DNA (*comG*) is unaffected by L-Arg up to a concentration of 100 mM (Supplementary Fig. 3d inset), and ComK is fully active in binding DNA (Supplementary Fig. 2b–m). We therefore used 50 mM L-Arg in all smFRET and cryo-EM experiments. Notably, L-Arg is known to stabilize proteins, thus preventing aggregation⁵⁵; however, it does not solubilize pre-formed aggregates in our hands.

SmFRET and 2fFCS experiments. All smFRET and 2fFCS experiments with ComK were performed at 20 mM Tris pH 7.0, 50 mM L-Arg, 5 mM MgCl₂, 150 mM KCl, 20 mM DTT, and 0.001% Tween-20. In contrast, the smFRET experiments, in which we map the curvature of the *comG* promoter (Fig. 2a, b) were performed in 20 mM sodium phosphate pH 7 or 4, 100 mM β-mercaptoethanol, and 0.001% Tween-20, which allowed an easier adjustment of the solution pH. The ionic strength was kept constant at 100 mM in these experiments by the addition of NaCl. The concentration of labelled DNA was 30 pM.

All smFRET experiments were performed with a MicroTime 200 confocal microscope (PicoQuant) equipped with an Olympus IX73 inverted microscope. Briefly, we used linearly polarized light from a 485 nm diode laser (LDH-D-C-485, PicoQuant) and unpolarized light at 594 nm from a supercontinuum light source (Solea, PicoQuant) to excite donor and acceptor alternately with a repetition rate of 20 MHz in an interleaved manner, thus resulting in a period length of 50 ns that contains one donor and one acceptor pulse. This pulsed-interleaved excitation (PIE) scheme⁵⁶ was used to distinguish molecules with only donor dye from those carrying both dyes, but with a large inter-dye distance (Supplementary Fig. 2a). The excitation beam was guided through a major dichroic mirror (ZT 470-491/594 rpc, Chroma) to a 60x, 1.2NA water objective (Olympus) that focuses the beam into the sample. The sample was placed in a home-made cuvette with a volume of 50 μl made of round quartz cover slips of 25 mm diameter (Esco Optics) and borosilicate glass 6 mm diameter cloning cylinder (Hilgenberg), using Norland 61 optical adhesive (Thorlabs). All measurements were performed at a laser power of 100 μW (485 nm) and 20 μW (594 nm) measured at the back aperture of the objective. Photons emitted from the sample were collected by the same objective and after passing the major dichroic mirror (ZT 470-491/594 rpc, Chroma), the residual excitation light was filtered by a long-pass filter (BLP01-488R, Semrock) and focused on a 100 μm pinhole. The sample fluorescence was detected either with two channels (Fig. 2a, b) or four channels (all other smFRET experiments). Donor and acceptor fluorescence was separated via a dichroic mirror (T585 LPXR, Chroma), and each colour was focused onto a single-photon avalanche diode (SPAD; Excelitas) with additional bandpass filters: FF03-525/50, (Semrock) for the donor SPAD and FF02-650/100 (Semrock) for the acceptor SPAD. The arrival time of every detected photon was recorded with a HydraHarp 400 M time-correlated single-photon counting module (PicoQuant) and stored with a resolution of 32 ps.

For 2fFCS⁵², the sample was excited with two orthogonally polarized lasers at 485 nm (LDH-D-C-485, PicoQuant). The beams of both lasers were combined and laterally shifted by a Nomarsky prism at the back aperture of the objective. Emitted light passed through a 150 μm pinhole. Afterwards, the emitted light was split using a polarizing beam splitter and focused on two separate SPAD detectors. The distance between the two resulting foci was experimentally determined using the known Stokes radii (R_S) of the calibration samples: the dye Oregon Green ($R_S = 0.6$ nm, 0.36 kDa) and the proteins aprotinin ($R_S = 1.29$ nm, 6.5 kDa), RNaseA ($R_S = 1.73$ nm, 13.7 kDa), carboxy anhydrase ($R_S = 2.19$ nm, 29 kDa), ovalbumin

($R_S = 2.65$ nm, 44 kDa), and conalbumin ($R_S = 3.1$ nm, 75 kDa; Supplementary Fig. 1a, b). For calibration, all proteins were labelled with AlexaFluor 488 NHS ester at primary amino groups of lysine residues. The autocorrelation functions of the signal of each focus and the two cross-correlation functions were fit globally with a model that contained one diffusion component and two exponential decay components to account for the triplet dynamics of the dyes (Supplementary Fig. 1a). The global fits were performed iteratively to minimize the squared difference between the calibration values and the values determined with 2fFCS. Subsequently, the Stokes radius of labelled ComK was determined in the absence and presence of increasing amounts of unlabelled ComK (Supplementary Figs. 1b and 3a). All measurements were performed at 23 °C.

Determination of the thermodynamic stability of ComK. To understand the strong aggregation tendency of ComK, we used 2fFCS to determine the stability of ComK under physiological conditions. To this end, we denatured labelled ComK at different concentration of the denaturant GdmCl. With increasing concentrations of GdmCl, the Stokes radius increases in a cooperative manner (Supplementary Fig. 3b). To obtain the stability, we fit the data using a two-state model of protein folding. Assuming a spherical shape of folded and unfolded ComK, the average Stokes radius (R) at any concentration of GdmCl is given by

$$\langle R \rangle = \sqrt[3]{\frac{3}{4\pi} [f v_f + (1-f) v_u]} \quad (1)$$

here, f is the fraction of folded molecules, v_f and v_u are the volumes of folded and unfolded molecules, respectively. Since it is known that the dimension of unfolded polypeptide chains responds sensitively to the concentration of denaturants^{57–59}, we assume that the volume of the unfolded state changes linearly with the concentration of GdmCl (x) according to

$$v_u = m_u x + n_u \quad (2)$$

with the slope m_u and the intercept n_u . The fraction f for a two-state system is given by

$$f = e^{-\Delta G/RT} / (1 + e^{-\Delta G/RT}) \quad (3)$$

here, R is the ideal gas constant, T is the temperature in Kelvin, and ΔG is the free energy difference between folded and unfolded molecules. We further assume a linear free energy relationship of the type

$$\Delta G = mx + \Delta G_0 \quad (4)$$

The data in Supplementary Fig. 3b were fit using a combination of Eqs. (1)–(4). The fit results in free energy of stabilization of $\Delta G_0 = -8.8 \pm 0.3$ kJ/mol.

SmFRET data analysis. Instrumental imperfections and differences in the brightness of donor and acceptor require a correction of the detected raw photon counts in the two channels (1 and 2)^{60,61}. These corrections are defined by five parameters: γ_1 and γ_2 that account for the different detection probabilities of photons from the two dyes, β_{21} and β_{12} , the leakage of donor photons into the acceptor channel 1 and the leakage of acceptor photons into the donor channel 2, respectively, and α , the probability to directly excite the acceptor dye at the wavelength specific for the donor. If n_1 and n_2 are the detected photons in the acceptor and donor channel, respectively, and b_1 and b_2 are the background rates in both channels, the corrected photon counts for acceptor and donor (n'_A and n'_D) are given by

$$\begin{pmatrix} n'_A \\ n'_D \end{pmatrix} = \begin{pmatrix} \gamma_1 - \beta_{DA} - \beta_{AD}\gamma_2 \\ \gamma_2 - \beta_{DA} - \beta_{AD}\gamma_1 \end{pmatrix} \begin{pmatrix} n_1 - b_1 T \\ n_2 - b_2 T \end{pmatrix} \text{ and } n'_A = n_A - \alpha(n'_D + n_A) \quad (5)$$

here, T is a time specifying the length of a burst. The correction parameters were determined⁶² with two separate samples of the dyes, in which their concentrations were adjusted such that both samples have an absorbance of 0.1 at the excitation wavelength of 485 nm. Setting arbitrarily $\gamma_1 = 1$, we obtain $\gamma_2 = 1.12 \pm 0.09$, $\beta_{DA} = 0.050 \pm 0.003$, and $\beta_{AD} = 0.0021 \pm 0.0004$ over 5 years with 21 measurements of these correction factors. We note that using these correction factors for different DNA samples assumes that the fluorescence quantum yields of donor and acceptor do not depend on the specific labelling position on the DNA. The probability of directly exciting the acceptor dye at 485 nm is given by $\alpha = \epsilon_A / (\epsilon_A + \epsilon_D)$, where ϵ_A and ϵ_D are the extinction coefficients of the dyes at the donor excitation wavelength of 485 nm. For our dye pair, $\alpha = 0.049$. Transfer efficiency histograms were computed with the fully corrected photon counts according to

$$E = n'_A / (n'_A + n'_D) \quad (6)$$

Importantly, the corrections (Eq. (5)) are already taken into account during burst identification. Bursts were identified from the measured photon traces following Eggleling et al.⁶⁰ and Hoffmann et al.⁶¹. If $\Delta t_i = t_i - t_{i-1}$ is the inter-photon time of the i th photon, the photon is retained if $\Delta t_i \leq \gamma_j(i) \Delta t_{\max}$ ($\Delta t_{\max} = 100$ μs) with $\gamma_j(i)$ being the correction factor of the i th photon detected in channel $j = [1, 2]$ (Eq. (5)). The algorithm then proceeds to the next photon $i + 1$, stops after n photons once $\Delta t_{i+n} > \gamma_j(i+n) \Delta t_{\max}$, and provides the total length of the burst by $T = t_{i+n} - t_{i-1}$. The resulting string of photons is now corrected via Eq. (5), using estimated background rates b_1 and b_2 . The initial guess of b_1 and b_2 is given by all detected photons in channels 1 and 2, respectively, divided by the total

measurement time. A burst is identified if $n'_A + n'_D > 80$. The photons belonging to this burst are removed from the photon trace, and a new guess for b_1 and b_2 is computed based on the remaining photons. Subsequently, the burst search is performed again with updated background rates. This procedure converges after three iterations to constant background rates and a constant number of identified bursts. Since the identified bursts also contain molecules for which the acceptor bleached during the transit through the confocal spot, thus masking the true transfer efficiency, we further cleaned the FRET histograms by these events⁶³. For a burst with n'_D donor photons with the arrival times $t_{D,1} \dots t_{D,n'_D}$ and n'_A acceptor photons with the arrival times $t_{A,1} \dots t_{A,n'_A}$, the average arrival times are given by $\langle t_D \rangle = n'^{-1}_D \sum_i t_{D,i}$ and $\langle t_A \rangle = n'^{-1}_A \sum_i t_{A,i}$. The burst asymmetry is defined by $\alpha_{DA} = \langle t_D \rangle - \langle t_A \rangle$. If the acceptor dye bleaches, we find $\alpha_{DA} > 0$. Taking shot noise into account, the distribution of α_{DA} has a standard deviation given by

$$\sigma_{DA} = \frac{T}{2\sqrt{3}} \left(\frac{1}{n'_D} + \frac{1}{n'_A} \right)^{1/2} \quad (7)$$

To eliminate molecules with a bleached acceptor, we excluded all molecules for which $|\alpha_{DA}| > \sigma_{DA}$.

In addition, only molecules containing active acceptor and donor dyes were included in the analysis (Supplementary Fig. 2a). To this end, we computed the donor-acceptor stoichiometry (S) for each burst according to

$$S = \frac{n'_{DD} + n'_{DA}}{n'_{DD} + n'_{DA} + \gamma_{PIE} n'_{AA}} \quad (8)$$

here, $\gamma_{PIE} \sim 2.5$ is a correction factor to account for the different excitation intensities for donor and acceptor. Furthermore, the first subscript indicates the emission and the second subscript indicates the excitation. Only molecules with $S < 0.8$ were used for constructing smFRET histograms.

Determination of the fraction of ComK-bound promoters from smFRET. FRET histograms for ComK-binding experiments were fitted with a combination of two Gaussian distributions⁶⁴, since all constructs used in these experiments had FRET values close to 0.5, thus showing no skewing. The width and position of these FRET peaks were fixed to the values obtained for the FRET distribution in the absence of ComK and in the presence of saturating amounts of ComK, which minimized the number of free fitting parameters. Only the amplitudes of the Gaussian peaks were free fitting parameters. The area under the histogram curve for each subpopulation was determined using numerical integration to obtain the fraction of ComK-bound promoter.

FRET histograms of the DNA constructs used to identify curvature in the *comG* promoter (Fig. 2a, b and Supplementary Fig. 2a) were fitted with Gaussian distributions for constructs with intermediate FRET efficiency, and with log-normal functions for histograms with mean FRET values close to 0 and 1.

Fluorescence lifetime analysis. The average lifetimes of the donor in the presence of acceptor τ_{DA} (Fig. 2a, b inset) were computed from the mean arrival time of the photons relative to the laser pulse exciting the donor. The expected value of the fluorescence lifetime of the donor in the presence of the acceptor τ_{DA} for a static inter-dye distance (dashed line in Fig. 2a, b insets) is given by $\tau_{DA} = \tau_D(1 - E)$, where $\tau_D = 3.6\text{--}4\text{ ns}$ is the lifetime of the donor in the absence of acceptor as determined from molecules lacking an active acceptor. However, if the *comG* promoter would fluctuate between different conformations, thus sampling a distance distribution $P(r)$, the simple linear relationship between the average donor fluorescence lifetime $\langle \tau_{DA} \rangle$ and the average FRET efficiency $\langle E \rangle$ will deviate from the linear Förster-dependence for a static distance^{65,66}. The resulting deviation can be used to determine the width of the distribution $P(r)$. We therefore fit the peak position of the average fluorescence lifetime $\langle \tau_{DA} \rangle$ and mean FRET efficiency $\langle E \rangle$, using a Gaussian distance distribution with mean μ and width σ . The analytical expressions for the average lifetime and FRET efficiency are given by

$$\langle \tau_{DA} \rangle = \int_0 \tau_{DA}(r)^2 P(r) dr / \int_0 \tau_{DA}(r) P(r) dr \quad (9)$$

and

$$\langle E \rangle = \int_0 E(r) P(r) dr / \int_0 P(r) dr \quad (10)$$

here, $\tau_{DA}(r) = \tau_D [1 + (R_0/r)^6]^{-1}$ and $E(r) = R_0^6 / (R_0^6 + r^6)$. The dependence between $\langle \tau_{DA} \rangle$ and $\langle E \rangle$ is implicit and depends on μ , which is a variable, and σ , which is the fitting parameter. We obtained $\sigma = 0.76\text{ nm}$ at pH 7 and $\sigma = 0.66\text{ nm}$ at pH 4. Assuming spherically symmetric and Gaussian distributed dye clouds for donor and acceptor due to the long C₆-alkyl linkers of our dyes, we can compute the distribution width σ_{dye} of these clouds. Assuming that donor and acceptor have the same spatial distribution (same σ_{dye}), we find $\sigma_{dye} = \sigma/\sqrt{2}$, which results in $\sigma_{dye} = 0.54\text{ nm}$ at pH 7 and $\sigma_{dye} = 0.47\text{ nm}$. These widths are shown as radius of the spheres in the schemes in Fig. 2a, b (inset) and are very similar to the positional

uncertainty (L) of the dye positions found previously by Wozniak et al.³⁴ (see Eq. (11) below).

Calculation of the FRET values for B-DNA. To determine the transfer efficiency profile of B-DNA (Fig. 2b, c), we followed the work of Clegg et al.⁶⁷ and Wozniak et al.³⁴ Keeping the nomenclature of Wozniak et al., the mean distances R_{DA} of donor (D)-acceptor (A) pairs on B-DNA is given by

$$R_{DA} = \sqrt{(L + \Delta_{bp} z_{bp})^2 + r_A^2 + r_D^2 - 2r_A r_D \cos(\alpha + \Delta_{bp} \beta_{bp})} \quad (11)$$

here, α is the offset angle between the fluorophores if they were both bound to the same base. L accounts for the fact that projections of the centres of D and A onto the helix axis do not necessarily coincide with the position of bases. The values r_A and r_D are the distances from the helical axis due to the dye linkers, z_{bp} is the increase along the axis per base pair, and β_{bp} is the increase in D-A angle per base pair. For dyes with a similar size of those used here, Wozniak et al. found $L = 0.617\text{ nm}$, $z_{bp} = 0.338\text{ nm}$, $r_A = 1.177\text{ nm}$, $r_D = 1.247\text{ nm}$, $\beta_{bp} = 36^\circ$, and $\alpha = 89.8^\circ$ for B-DNA, and we used the same parameters to compute the transfer efficiency via

$$E_{mp} = R_0^6 / (R_0^6 + R_{DA}^6) \quad (12)$$

with $R_0 = 5.4\text{ nm}$ ⁵⁷. Since our experimentally measured transfer efficiencies are an average over the positional distribution of the dyes due to the flexibility of their linkers, the E_{mp} values were converted to the mean transfer efficiencies $\langle E \rangle$ accessible in our experiments by solving the empirical implicit equation³⁴

$$E_{mp} = 0.008 + 0.679\langle E \rangle + 1.470\langle E \rangle^2 - 1.141\langle E \rangle^3 \quad (13)$$

These values are shown as black line in Fig. 2a, b. Notably, assuming that the *comG* promoter is in a B-DNA conformation at pH 4, we also used Eqs. (11)–(13) in a fit with R_0 as a free parameter and obtained a value of $5.6 \pm 0.1\text{ nm}$, i.e., very similar to the literature value of 5.4 nm used in Fig. 2a, b. Yet, this comparison between experimental smFRET values and the described model is based on two assumptions, (i) distance fluctuations of the DNA are marginal and (ii) dye motions caused by the flexible linkers are slower than the FRET transfer rate. Since we cannot unambiguously exclude violations of these assumptions, the described agreement has to be considered qualitative.

Thermodynamic ComK-DNA-binding model. To determine coupling free energies between the binding boxes of ComK, we fit the data using a mechanistic binding model (Fig. 4a). In this model, we denote the promoter as P and the four binding sites are denoted by subscripts $P_1 \dots P_4$. Here, the first two binding sites are in box 1 and the second two binding sites are in box 2. We further denote ComK as X such that the concentration of the promoter with one ComK bound to the first site is $[P_1X]$. The concentration of free ComK is x . The initial binding of a ComK to each site is given by four association reactions with identical association constant K defined by

$$K = \frac{[P_1X]}{[P]x} = \frac{[P_2X]}{[P]x} = \frac{[P_3X]}{[P]x} = \frac{[P_4X]}{[P]x} \quad (14)$$

The second ComK binds exclusively to a box that already contains a bound ComK. We introduced this constraint based on the determined cryo-EM structures that indicates substantial protein-protein contacts between ComK molecules in one box (Fig. 3c). Hence, binding the second ComK gives the relations

$$\sigma K = \frac{[P_{12}X_2]}{[P_1X]x} = \frac{[P_{12}X_2]}{[P_2X]x} = \frac{[P_{34}X_2]}{[P_3X]x} = \frac{[P_{34}X_2]}{[P_4X]x} \quad (15)$$

The factor σ accounts for the higher affinity of binding a second ComK to a box due to allostery via protein-protein contacts. The third ComK now binds to a different box, thus leading to

$$JK = \frac{[P_{134}X_3]}{[P_{34}X_2]x} = \frac{[P_{234}X_3]}{[P_{34}X_2]x} = \frac{[P_{123}X_3]}{[P_{12}X_2]x} = \frac{[P_{124}X_3]}{[P_{12}X_2]x} \quad (16)$$

where the factor J accounts for the inter-box cooperativity. Finally, the last ComK molecule binds to the remaining site that now includes intra- (σ) and inter- (J) box cooperativity

$$J\sigma K = \frac{[P_{1234}X_4]}{[P_{234}X_3]x} = \frac{[P_{1234}X_4]}{[P_{134}X_3]x} = \frac{[P_{1234}X_4]}{[P_{124}X_3]x} = \frac{[P_{1234}X_4]}{[P_{132}X_3]x} \quad (17)$$

Replacing $s = Kx$ for clarity and with $[P_0]$ as the total concentration of promoter DNA, we obtain the mass conservation equation

$$[P_0] = [P](1 + 4s + 2\sigma s^2 + 4Js^3 + J^2\sigma^2 s^4) \quad (18)$$

which gives the binding polynomial shown in Fig. 4a

$$Q = 1 + 4s + 2\sigma s^2 + 4Js^3 + J^2\sigma^2 s^4 \quad (19)$$

The fraction of fully bound promoter is therefore given by

$$f = \frac{J^2 \sigma^2 s^4}{1 + 4s + 2\sigma s^2 + 4J\sigma s^3 + J^2 \sigma^2 s^4} \quad (20)$$

With the same logic, we obtain for the isolated boxes

$$f = \frac{\sigma s^2}{1 + 2s + \sigma s^2} \quad (21)$$

Notably, these equations only hold under conditions at which the concentration of DNA is very low compared to K^{-1} , which, with $[P_0] = 30 \text{ pM}$ is given at all our experimental conditions. To analyse our binding data, we first fit the data of the isolated boxes 1 and 2 globally, using Eq. (21) (Fig. 2d). Here, the value $\sigma = e^{-\Delta g_\sigma}$ was a global parameter, whereas $K = e^{-\Delta g_K}$ was a local parameter that differed between box 1 and box 2. These fits provide the free energy changes Δg_σ and Δg_K in Supplementary Table 4. To reduce the number of parameters in Eq. (20), we kept the value of Δg_σ in all fits of promoters with two boxes and expressed the coupling between two boxes as $J = e^{-\Delta g_J}$ with the coupling free energy Δg_J (Supplementary Table 4). For promoters in which we mapped both boxes (*addAB* and *comG*), we fit the data using a global value for Δg_J (Figs. 2e and 4b). To allow an easier comparison of the cooperativity in the different constructs, all data were also fitted using the standard Hill equation

$$f = x^n / (K_{\text{Hill}}^n + x^n) \quad (22)$$

where n is the Hill exponent that is an empirical measure for cooperativity, and K_{Hill} is the effective dissociation constant.

Determination of Hill exponents from the distribution of bound ComK. If i is the number of ligands to a single particle then $\langle i \rangle = \sum_{i=0}^N ip(i)$ is the average number of bound ligands to an ensemble of particles and $\langle i^2 \rangle = \sum_{i=0}^N i^2 p(i)$ is the second moment of the distribution $p(i)$ of bound ligands, and N is the total number of possible binding sites. Following Wyman⁴¹ (combine Eqs. (8.2) and (9.1) therein), the Hill coefficient is given by

$$n = \frac{\langle i^2 \rangle - \langle i \rangle^2}{\langle i \rangle (1 - \langle i \rangle / N)} \quad (23)$$

The nominator of Eq. (23) is the variance of the observed distribution of bound ligands and the denominator is the variance of the corresponding binomial distribution. Since $p(i)$ is approximately accessible from the evaluation of particles in cryo-EM (Fig. 3b), the Hill coefficient was calculated from the cryo-EM statistics, using Eq. (23).

The elastic-coupling model. We described the obtained coupling free energies (Fig. 4c) using a model derived by Rudnick and Bruinsma¹⁸. The model starts from the Hamiltonian for a two-dimensional worm-like chain (WLC)

$$H[\theta(s)] = \int ds \left\{ k_B T \frac{l_p}{2} \left(\frac{d\theta}{ds} \right)^2 - f \cos\theta \right\} \quad (24)$$

Here, s is the arclength of the chain, $\theta(s)$ is the angle between the chain and one coordinate, l_p is the persistence length, and f is a tension applied to the rod. Solving Eq. (24) for small bending angles ($\cos\theta \sim \theta^2/2$) and introducing two discontinuities (kinks) separated by Δs , gives odd and even solutions that lead to an enthalpic contribution of the kinks

$$\Delta E = 2\alpha^2 \left(l_p / \xi \right) \left(1 \pm e^{-\Delta s / \xi} \right) \text{ in units of } k_B T \quad (25)$$

Here, Δs is the distance between the two kinks (protein binding sites), $\xi = \left(k_B T l_p / f \right)^{1/2}$ is the decay length of tension along the DNA, l_p is the persistence length, and α is the local bending angle at the binding site. Compared to the enthalpic effect, the entropic contribution is rather small¹⁸ and was therefore omitted. Eq. (25) describes the effective change in binding energy of two proteins for the case, in which they introduce bending in DNA. Interestingly, Eq. (25) has a non-zero asymptotic limit at long distances between the binding sites, which accounts for the energy required to bend DNA by each protein separately. The plus sign is for the symmetric case, i.e., the two sites couple repulsively, which hampers binding of both proteins, and the minus sign indicates the attractive antisymmetric case. Clearly, the opposite is the case for ComK. Rather than introducing bending, binding of ComK reduces bending, which we interpret to be equivalent to a switch in signs. The helicity of DNA is incorporated via a cosine modulation between the signs in Eq. (25), thus leading to

$$\Delta g_J = 2\alpha^2 \left(l_p / \xi \right) \left(1 + \cos \left[\frac{2\pi}{\lambda} (\Delta s + \Delta s_0) \right] e^{-\Delta s / \xi} \right) \quad (26)$$

for the coupling free energy. We used a similar expression to fit the change of Hill

exponents with increasing spacer length

$$n = a \left(b + \cos \left[\frac{2\pi}{\lambda} (\Delta s + \Delta s_0) \right] e^{-\Delta s / \xi} \right). \quad (27)$$

Here, λ is the periodicity of DNA (10.5 bp), Δs is the absolute spacer length, $\Delta s_0 = 8 \text{ bp}$ is the phase shift. The qualitative fractions of the EMSA were empirically fitted with a sum of a cosine and exponential function (solid line in Fig. 4d) with a periodicity of 10.5 bp, which resulted in a phase shift of $\Delta s_0 = 7.2 \pm 0.5 \text{ bp}$. When setting $\lambda = 10.5 \text{ bp}$ and $\Delta s_0 = 8 \text{ bp}$ based on these results (Fig. 4d), a fit of the Hill exponents with Eq. (27) resulted in $\xi = 14 \pm 8 \text{ bp}$ (Fig. 4c left). To obtain the local bending angle, which can be interpreted as the change in bending in a box upon ComK binding, we used Eq. (26) and fixed the decay length to $\xi = 14$, resulting in $\alpha = 24^\circ \pm 7^\circ$ per box. The average change in bending across the full *comG* promoter (44 bp from box 1 to box 2) then gives a value of $2\alpha/44 = 1.1^\circ \pm 0.3^\circ/\text{bp}$ that can be compared to the result from cryo-EM (Fig. 3d, e).

Gel retardation assay. The gel retardation assay, also termed EMSA, is capable of providing a qualitative estimate of ComK–DNA binding. The forward strand (Supplementary Table 7) and the reverse strand, at a final concentration of $10 \mu\text{M}$ were mixed in the following buffer: 50 mM Tris pH 7.5, 0.1 M NaCl, and 3 mM MgCl. The mixture was heated to 95°C and slowly cooled in a PCR cycler, as previously described. The reaction mixture contained 25 nM dsDNA and 300 nM ComK, in the following buffer: 20 mM Tris pH 7.0, 50 mM l-Arg, 5 mM MgCl₂, 150 mM KCl, 10 mM DTT, and 0.001% Tween-20. The reaction was incubated for 20 min and supplemented with glycerol to a final concentration of 10%. The reaction was loaded on a 5% native PAGE, composed of 4.2 ml of 30% polyacrylamide 29:1, 8.3 ml of 1.5 M Tris-HCl pH 8.8, 12.2 ml ddH₂O, 150 μl of 10% ammonium persulfate, and 150 μl TEMED. The gels were run in a Hoefer system at 100 V for 30 min, under TAE running buffer. The gels were then incubated in 100 ml ddH₂O and a single drop of 0.625 mg/ml ethidium bromide, for 10 min. The gels were then washed and imaged with a Typhoon FLA 9500 scanner (GE). The fractions were computed by numerically integrating the bands in each lane, using Mathematica 11.2 (Wolfram).

Sample preparation for cryo-EM. ComK at a concentration of $40 \mu\text{M}$ (500 μl) in 50 mM Tris pH 8.0, 6 M GdmCl, and 10 mM imidazole was injected on a HiTrap desalting column (5 ml, GE Healthcare) pre-equilibrated with 0.5 M arginine–HCl, pH 7.3. The protein containing fractions (32 μM) were then supplemented with 10 mM DTT to prevent the oxidation of the intrinsic cysteine residues of ComK. The ComK–DNA complexes were formed by quickly mixing 2.5 μl of ComK (32 μM) and 0.6 μl of DNA (20 μM) with 22 μl 20 mM Tris pH 7.0, 5 mM MgCl₂, 150 mM KCl, 3 mM DTT, and 0.001% Tween-20. Immediately, after mixing, 2.5 μl of this mixture was transferred to UltraAuFoil R 1.2/1.3 300 mesh grids (Quantifoil). The grids were glow discharged before applying the sample, blotted for 2.5 s at 4°C and 100% humidity, and plunge frozen in liquid ethane cooled by liquid nitrogen using a Vitrotron plunger (Thermo Fisher Scientific).

Cryo-EM data acquisition. Cryo-EM data sets were collected on a Titan Krios G3i transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV. ComK–DNA complexes featured preferred orientation, and therefore initial data sets collected at 0° tilt were used only to calculate binding statistics by 2D classification (Fig. 3a). To this end, 10,526, 3909, and 5108 0° tilt movies were recorded for the 8, 18, and 31 bp spacers, respectively. For 3D reconstructions, data were collected primarily at 40° tilt. The 8 bp spacer data set comprised 6537 40° tilt movies. The 18 bp spacer data set comprised 8024 40° tilt movies and 431 0° tilt movies. Movies were recorded in counting mode on a K3 direct detector (Gatan) at the end of BioQuantum energy filter (Gatan), using a slit of 20 or 15 eV at a nominal magnification of 105,000 \times , which corresponds to a physical pixel size of 0.86 \AA . The dose rate was set to 20.6 e⁻/pixel/s and the total exposure time was 2 s, resulting in an accumulated dose of 55.7 e⁻/ \AA^2 . Each movie was fractionated into 50 frames of 0.04 s. The nominal defocus range was -0.8 to $-1.3 \mu\text{m}$; however, the actual defocus range was larger because of tilting. Imaging was done using an automated low-dose procedure implemented in SerialEM⁶⁸, in which a single image was collected from the centre of each hole. Image shift was used to target an array of up to six neighbouring holes along the tilt axis (depending on hole array orientation), and stage shift to move between arrays. Beam tilt was adjusted to achieve coma-free alignment when applying image shift.

Single-particle cryo-EM image processing. Image processing was performed using CryoSPARC software⁶⁹. Movie frames were aligned using patch motion correction, followed patch CTF estimation. A total of 4024 micrographs of the 8 bp spacer data set, which featured a resolution better than 10.0 \AA , relative ice thickness lower than 1.4, and full frame motion correction smaller than 100 pixels were selected for further processing. The processing procedure is outlined in Supplementary Fig. 5. An initial particle data set was created by manual picking, followed by 2D classification and automated picking, using the newly generated 2D classes as templates. About 130,000 automatically picked particles were extracted, binned 4×4 (110-pixel box size, 3.44 $\text{\AA}/\text{pixel}$), and subjected to multiple rounds of 2D classification in order to clean the data set. Particles were separated into

ComK–DNA complexes and free DNA (100,695 and 9500 particles, respectively), followed by an *ab initio* 3D reconstruction of the ComK–DNA particles with two 3D classes. The higher-resolution 3D class (62,177 particles) was refined using the unbinned data (440 box size) to a final resolution of 6.6 Å.

For the 18 bp spacer, 7450 movies were retained using the same criteria as above (Supplementary Fig. 6). Automated picking resulted in ~1,361,000 particles, which were cleaned by 2D classification and separated into two data sets, ComK–DNA and free DNA, comprising 422,425 and 29,171 particles, respectively. *Ab initio* 3D reconstruction followed using 4 × 4 binned images (3.44 Å/pixel). Both the ComK–DNA and free DNA particles were classified in 3D based on particle curvature, resulting in three well-resolved classes in each data set. In addition, in an attempt to find a homogeneous subset that will yield the highest possible resolution, the ComK–DNA particles were classified in 3D sequentially, as depicted in Supplementary Fig. 6. All 3D classes were refined by homogenous refinement using the unbinned data and 440-pixel boxes. Final resolutions and angular distributions can be seen in Supplementary Fig. 7. 3D visualization was performed using UCSF Chimera⁷⁰ (Version 1.12) or ChimeraX⁷¹ (Version 0.92). The resulting 3D maps show a plane that connects the surface of a ComK dimer in the first box with that of the second box. In addition, the peripheral surface of the ComK dimers that faces away from the DNA appear distorted in the maps. This indicates that the ComK–DNA complexes are adsorbed to the water–air interface via this surface (Supplementary Fig. 10).

Binding stoichiometry of ComK to DNA was determined using the initial 0° tilt data sets, as described above. Movie frames were aligned using patch motion correction, followed by CTF estimation, and the highest quality micrographs were retained. A total of 10,526, 3909, and 4982 micrographs were retained in the 8, 18, and 31 bp spacer data sets, respectively. Initial 2D classes were generated by manual picking, followed by automated picking using the newly generated 2D templates. The number of complexes in each group was determined by 2D classification after discarding classes that did not resolve well. In total, the 8, 18, and 31 bp spacer data sets comprised 508,389, 22,6121, and 231,710 particles, respectively.

Calculation of curvature angles of free and ComK-bound comG. Curvature angles (Fig. 3d, e) were computed from a coarse-grained spline that was determined from the EM maps of free DNA and of the ComK–comG complex. The EM maps were first binned (four *z*-slices per bin for free DNA, three *z*-slices per bin for ComK–comG) and subsequently fit with a Bezier 3D-spline, using Mathematica 11.2 (Wolfram). The spline was subsequently divided into segments of $b = 0.34$ nm length. The resulting coordinates \mathbf{u}_i were used to compute the vector \mathbf{v}_i of each segment i via $\mathbf{v}_i = \mathbf{u}_{i+1} - \mathbf{u}_i$. The angle between neighbored segments was finally computed according to $\theta_i = \cos^{-1} \mathbf{v}_{i+1} \cdot \mathbf{v}_i / b^2$.

DNA model building and refinement. The structures of the two promoters *addAB* and *comG* in complex with ComK were built and refined, using the respective cryo-EM maps. Idealized B-DNA fibre-like double helix models were produced using Coot⁷² with the sequence of the promoters *addAB* (78 bp in total) with a spacer of 8 bp between boxes, and *comG* (93 bp in total) with a spacer of 18 bp. The resulting straight DNA duplexes were used as starting models for alignment to their respective density maps, using manual fitting and real-space rigid-body refinement with the whole duplex defined as one group, leading to two idealized models of opposite direction for both promoters. The directional uncertainty of the models along the long axis of the EM map was solved by accounting for the asymmetry of the sequences at the 5'- and 3'-ends relatively to the binding boxes. However, whereas the size of the EM envelope of the *addAB* promoter at reasonable sigma value (typically 3.5–5.0) corresponded well to the length and diameter of the corresponding idealized model (~260 and 20 Å respectively), the EM map of the longer promoter, *comG*, was too short by ~30 Å (10 bp) to accommodate a 93 bp double helix with an expected length of ~310 Å. Most probably the longer DNA-end of the promoter was more flexible than the rest of the duplex. As a result, 10 bp are missing in the model of the *comG* structure. The aligned models were then submitted to a second real-space rigid-body cycle. To this end, the number and size of the rigid-body groups were determined by identifying the base pairs for which the DNA duplex deviated from the straight initial model. These parts were then adapted to the EM map, using Phenix⁷³. This step led to correlation coefficients (CC) of $CC(\text{mask}) = 0.35$ and $CC(\text{box}) = 0.48$ (from 0.11 and 0.30, respectively) for the whole *addAB* duplex. Subsequent cycles of manual corrections with Coot⁷² and real-space refinement with Phenix⁷³ led to the final values of $CC(\text{mask}) = 0.74$ and $CC(\text{box}) = 0.66$. For the refinement of the *comG* model, successive cycles of manual corrections and real-space rigid-body refinement were entirely conducted with Coot⁷². The final real-space refinements of this incomplete model (84 bp out of 93 bp) using Phenix⁷³ led to the final values of $CC(\text{mask}) = 0.75$ and $CC(\text{box}) = 0.63$. For both structures, the refinement cycles with Phenix⁷³ included restraints on bond lengths and bond angles, on Watson–Crick base-pairing H-bonds, planarity, parallelism, and on stacking. Despite the low resolution, the atomic displacement parameters (ADPs or B-factors) were refined during the last cycle of refinement. Remarkably, in both structures, the lowest values were obtained for the nucleotides forming A-tracts targets of the ComK monomers. During the entire process, the progress was monitored by systematic inspection of the models in their respective EM map using Coot⁷² and validated using MolProbity⁷⁴. Assuming that two ComKs bind their respective box in a similar manner, we were able to position the asymmetrically bent DNA double

helices along the long axis of the map with a level of uncertainty of ±2 bp. The refinement statistics are given in the Supplementary Table 5.

Comparison of the DNA model with smFRET. Predictions of mean FRET values for the ComK–*addAB* complex were done by first simulating possible dye positions ($\mathbf{s}_i, \mathbf{t}_j$) along with their probabilities (p_i, q_j) for donor (i) and acceptor (j) fluorophores. The FRET efficiency was then evaluated as $E = \sum_i \sum_j \frac{R_0^6}{R_{ij}^6 + R_0^6} p_i q_j$, where $r_{ij} = |\mathbf{s}_i - \mathbf{t}_j|$ and $R_0 = 5.4$ nm is the Förster distance. The values for $\mathbf{s}_i, \mathbf{t}_j, p_i,$ and q_j were obtained by modelling the fluorophores as bulky spheres, which were attached to an adenine base in the following manner: a nitrogen atom was positioned at a distance of 132 pm from the C8-atom of the adenine base of the atomic model of the *addAB* promoter (see previous paragraph). The atom was placed in the adenine plane at a N–C8–N9 angle of 123°. The fluorophore was attached to this N-atom via a polyethylene chain of seven C-atoms. The conformational freedom of the linker was accounted for using the rotational isomeric state model⁷⁵: each bond along the chain samples the anti, gauche+ and gauche– rotamers. Chains that sterically clashed⁷⁶ with the complex were discarded. However, the linker was seriously sterically constrained by the tight packing of bases and the DNA backbone, which is unlikely to be precise in its details given the low resolution of our 3D-EM maps that does not allow us to correctly model tilt and buckling of the base pairs. In addition, it cannot be excluded that the dye attachment causes local distortions of the DNA that are not reflected in the atomic model. Given the resulting steric constraints, we performed a set of simulations in which we allow clashes of linker atoms with the DNA. Three sets were considered, clashes of only the first linker atom, clashes of the first four linker atoms, and clashes of the first seven linker atoms (Supplementary Fig. 8 and Supplementary Table 6). Geometric and energetic parameters of polyethylene were taken from Rehan et al.⁷⁵, and the fluorophore size was taken from Kalinin et al.⁷⁷. The excluded volume of ComK was taken into account. To this end, the EM map of ComK given on a 3D grid with 1.718 Å spacing was used. We performed calculations for two different contour levels of the EM map for ComK, leading to volumes of 19.6 and 47.8 nm³ for a ComK dimer. The volumes were filled with atoms of van der Waals radius of 0.859 Å. The results are summarized in Supplementary Table 6. For comparison, with an average amino acid volume⁷⁸ of 0.13 nm³ and a length of a ComK monomer of 195 amino acids, we expect a volume of 50.7 nm³ for a ComK dimer.

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

Data availability

Data supporting the findings of this manuscript are available from the corresponding authors upon reasonable request. A reporting summary for this article is available as a Supplementary Information file. Cryo-EM maps and atomic coordinates have been deposited in the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB). The accession codes are EMD-11022 and 6Z0S for the *comG*–ComK complex (18 bp spacer) and EMD-12260 and 7NBN for the *addAB*–ComK complex (8 bp spacer). The single-molecule raw data were deposited on the Zenodo data repository under the <https://doi.org/10.5281/zenodo.4468372>. Source data are provided with this paper.

Code availability

The single-molecule analysis code is accessible as a custom WSTP add-on for Mathematica (Wolfram Research) at <https://schuler.bioc.uzh.ch/programs/>.

Received: 7 July 2020; Accepted: 13 April 2021;

Published online: 20 May 2021

References

- Monod, J. & Jacob, J. General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).
- Monod, J., Changeux, J.-P. & Jacob, F. Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329 (1963).
- Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
- Koshland, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA* **44**, 98–104 (1958).
- Koshland, D. E., Némethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385 (1966).
- Thirumalai, D., Hyeon, C., Zhuravlev, P. I. & Lorimer, G. H. Symmetry, rigidity, and allosteric signaling: from monomeric proteins to molecular machines. *Chem. Rev.* **119**, 6788–6821 (2019).

7. Gruber, R. & Horowitz, A. Allosteric mechanisms in chaperonin machines. *Chem. Rev.* **116**, 6588–6606 (2016).
8. Yuan, Y., Tam, M. F., Simplaceanu, V. & Ho, C. New look at hemoglobin allostery. *Chem. Rev.* **115**, 1702–1724 (2015).
9. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
10. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
11. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).
12. Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317**, 526–529 (2007).
13. Locke, J. C. W., Young, J. W., Fontes, M., Hernández Jiménez, M. J. & Elowitz, M. B. Stochastic pulse regulation in bacterial stress response. *Science* **334**, 366–369 (2011).
14. Kim, S. et al. Probing allostery through DNA. *Science* **339**, 816–819 (2013).
15. Pohl, F. M., Jovin, T. M., Baehr, W. & Holbrook, J. J. Ethidium bromide as a cooperative effector of a DNA structure. *Proc. Natl Acad. Sci. USA* **69**, 3805–3809 (1972).
16. Hogan, M., Dattagupta, N. & Crothers, D. M. Transmission of allosteric effects in DNA. *Nature* **278**, 521–524 (1979).
17. Parekh, B. S. & Hatfield, G. W. Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. *Proc. Natl Acad. Sci. USA* **93**, 1173–1177 (1996).
18. Rudnick, J. & Bruinsma, R. DNA-protein cooperative binding through variable-range elastic coupling. *Biophys. J.* **76**, 1725–1733 (1999).
19. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111–1123 (2007).
20. Moretti, R. et al. Targeted chemical wedges reveal the role of allosteric DNA modulation in protein–DNA assembly. *ACS Chem. Biol.* **3**, 220–229 (2008).
21. Koslover, E. F. & Spakowitz, A. J. Twist- and tension-mediated elastic coupling between DNA-binding proteins. *Phys. Rev. Lett.* **102**, 178102 (2009).
22. García, H. G. et al. Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep.* **2**, 150–161 (2012).
23. Singh, J. & Purohit, P. K. Elasticity as the basis of allostery in DNA. *J. Phys. Chem. B* **123**, 21–28 (2019).
24. Dršata, T. et al. Mechanical model of DNA allostery. *J. Phys. Chem. Lett.* **5**, 3831–3835 (2014).
25. Hancock, S. P., Cascio, D. & Johnson, R. C. Cooperative DNA binding by proteins through DNA shape complementarity. *Nucleic Acids Res.* **47**, 8874–8887 (2019).
26. Morten, M. J., Lopez, S. G., Steinmark, I. E., Rafferty, A. & Magennis, S. W. Stacking-induced fluorescence increase reveals allosteric interactions through DNA. *Nucleic Acids Res.* **46**, 11618–11626 (2018).
27. Schurr, J. M., Delrow, J. J., Fujimoto, B. S. & Benight, A. S. The question of long-range allosteric transitions in DNA. *Biopolymers* **44**, 283–308 (1997).
28. Crothers, D. M. Biophysics. Fine tuning gene regulation. *Science* **339**, 766–767 (2013).
29. Smits, W. K., Kuipers, O. P. & Veening, J.-W. Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.* **4**, 259–271 (2006).
30. Turgay, K., Hahn, J., Burghoorn, J. & Dubnau, D. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J.* **17**, 6730–6738 (1998).
31. Hamoen, L. W., Van Werkhoven, A. F., Bijlsma, J. J., Dubnau, D. & Venema, G. The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev.* **12**, 1539–1550 (1998).
32. Haran, T. E. & Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **42**, 41–81 (2009).
33. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. & Shakked, Z. DNA bending by an adenine-thymine tract and its role in gene regulation. *Proc. Natl Acad. Sci. USA* **98**, 8490–8495 (2001).
34. Wozniak, A. K., Schröder, G. F., Grubmüller, H., Seidel, C. A. M. & Oesterhelt, F. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl Acad. Sci. USA* **105**, 18337–18342 (2008).
35. Stellwagen, E., Peters, J. P., Maher, L. J. & Stellwagen, N. C. DNA A-tracts are not curved in solutions containing high concentrations of monovalent cations. *Biochemistry* **52**, 4138–4148 (2013).
36. Ingargiola, A., Weiss, S. & Lerner, E. Monte Carlo diffusion-enhanced photon inference: distance distributions and conformational dynamics in single-molecule FRET. *J. Phys. Chem. B* **122**, 11598–11615 (2018).
37. Schultz, D., Ben Jacob, E., Onuchic, J. N. & Wolynes, P. G. Molecular level stochastic model for competence cycles in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **104**, 17582–17587 (2007).
38. Hill, A. V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* **40**, iv–vii (1910).
39. Hertel, K. J., Lynch, K. W. & Maniatis, T. Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* **9**, 350–357 (1997).
40. Susanna, K. A. et al. Mechanism of transcription activation at the comG promoter by the competence transcription factor ComK of *Bacillus subtilis*. *J. Bacteriol.* **186**, 1120–1128 (2004).
41. Wyman, J. Linked functions and reciprocal effects in hemoglobin: a second look. *Adv. Protein Chem.* **19**, 223–286 (1964).
42. Hud, N. V. & Plavec, J. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers* **69**, 144–158 (2003).
43. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).
44. Kunst, F. et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).
45. Santillan, M. On the use of the Hill functions in mathematical models of gene regulatory networks. *Math. Model. Nat. Phenom.* **3**, 85–97 (2008).
46. Gonze, D., Abou-Jaoudé, W., Ouattara, D. A. & Halloy, J. How molecular should your molecular model be? On the level of molecular detail required to simulate biological networks in systems and synthetic biology. *Meth. Enzymol.* **487**, 171–215 (2011).
47. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall/CRC, 2007).
48. Buse, O., Pérez, R. & Kuznetsov, A. Dynamical properties of the repressilator model. *Phys. Rev. E* **81**, 066206 (2010).
49. Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
50. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
51. Masoud, R. et al. Studying the structural dynamics of bipedal DNA motors with single-molecule fluorescence spectroscopy. *ACS Nano* **6**, 6272–6283 (2012).
52. Dertinger, T. et al. Two-focus fluorescence correlation spectroscopy: a new tool for accurate and absolute diffusion measurements. *Chemphyschem* **8**, 433–443 (2007).
53. Siegel, L. M. & Monty, K. J. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta* **112**, 346–362 (1966).
54. Bennett, B. D. et al. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599 (2009).
55. Lange, C. & Rudolph, R. Suppression of protein aggregation by L-arginine. *Curr. Pharm. Biotechnol.* **10**, 408–414 (2009).
56. Müller, B. K., Zaychikov, E., Bräuchle, C. & Lamb, D. C. Pulsed interleaved excitation. *Biophys. J.* **89**, 3508–3522 (2005).
57. Schuler, B., Lipman, E. & Eaton, W. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002).
58. 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).
59. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu. Rev. Biophys.* **45**, 207–231 (2016).
60. Eggeling, C. et al. Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163–180 (2001).
61. Hoffmann, A. et al. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl Acad. Sci. USA* **104**, 105–110 (2007).
62. Schuler, B. Application of single molecule Förster resonance energy transfer to protein folding. *Methods Mol. Biol.* **350**, 115–138 (2007).
63. Hillger, F. et al. Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angew. Chem. Int. Ed. Engl.* **47**, 6184–6188 (2008).
64. Benke, S., Nettels, D., Hofmann, H. & Schuler, B. Quantifying kinetics from time series of single-molecule Förster resonance energy transfer efficiency histograms. *Nanotechnology* **28**, 114002 (2017).
65. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu. Rev. Biophys.* **45**, 207–231 (2016).
66. Gopich, I. V. & Szabo, A. Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. *Proc. Natl Acad. Sci. USA* **109**, 7747–7752 (2012).
67. Clegg, R. M., Murchie, A. I., Zechel, A. & Lilley, D. M. Observing the helical geometry of double-stranded DNA in solution by fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA* **90**, 2994–2998 (1993).
68. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

69. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
70. Pettersen, E. F. et al. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
71. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
72. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
73. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
74. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
75. Rehahn, M., Mattice, W. L. & Suter U. W. *Rotational Isomeric State Models in Macromolecular Systems*, 1 edn. (Springer, 1997).
76. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
77. Kalinin, S. et al. A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9**, 1218–1225 (2012).
78. Zamyatnin, A. A. Amino acid, peptide, and protein volume in solution. *Annu. Rev. Biophys. Bioeng.* **13**, 145–165 (1984).

Acknowledgements

We enjoyed the critical discussions and helpful comments of many colleagues. Our thanks go to Deborah Fass, Gilad Haran, Amnon Horovitz, Benjamin Schuler, and Philipp Selenko. We also thank Harry Greenblatt for computational support, as well as Christian Dubiella and Ronen Gabizon for their help with the LC-MS. This research was supported by the Israel Science Foundation (grant no. 1549/15), the Benozio Fund for the Advancement of Science, the Carolito Foundation, the Gurwin Family Fund for Scientific Research, the Leir Charitable Foundation, and the Koshland family. The work was further supported by the Irving and Cherna Moskowitz Center for Nano and Bio-Nano Imaging at the Weizmann Institute of Science.

Author contributions

G.R. and H.H. designed research. G.R. and F.W. labelled DNA constructs. G.R. and H.H. performed and analysed single-molecule experiments. G.R. and N.E. performed cryo-EM

work and analysed these data. H.R. conducted the fitting of the 3D maps with atomic models of the DNA. J.J. performed the mean FRET predictions. H.H. wrote the paper with the help of all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23148-2>.

Correspondence and requests for materials should be addressed to G.R. or H.H.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021