7554–7565 *Nucleic Acids Research, 2018, Vol. 46, No. 15* *Published online 14 June 2018*
doi: 10.1093/nar/gky549

Allosterism and signal transfer in DNA

Alexandra Balaceanu[1], Alberto Pérez[2], Pablo D. Dans[1] and Modesto Orozco[1,3]

[1]Joint IRB-BSC Program on Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain, [2]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA and [3]Department of Biochemistry and Biomedicine, University of Barcelona, 08028 Barcelona, Spain

Received February 07, 2018; Revised May 11, 2018; Editorial Decision June 01, 2018; Accepted June 06, 2018

ABSTRACT

We analysed the basic mechanisms of signal transmission in DNA and the origins of the allostery exhibited by systems such as the ternary complex BAMHI–DNA–GRDBD. We found that perturbation information generated by a primary protein binding event travels as a wave to distant regions of DNA following a hopping mechanism. However, such a structural perturbation is transient and does not lead to permanent changes in the DNA geometry and interaction properties at the secondary binding site. The BAMHI–DNA–GRDBD allosteric mechanism does not occur through any traditional models: direct (protein-protein), indirect (reorganization of the secondary site) readout or solvent-release. On the contrary, it is generated by a subtle and less common entropy-mediated mechanism, which might have an important role to explain other DNA-mediated cooperative effects.

INTRODUCTION

Macromolecules are capable to transport information signals emerging from, for example ligand binding, to distant regions activating a variety of secondary effects. One of them is allostery, which implies changes in the binding of a second ligand due to prior interaction of the allosteric effector (1–5). Allostery has been deeply studied for proteins, where it has been characterized as one of the main mechanisms of control of their activities (5–7). Much less studied, but equally important, DNA-mediated allostery has been attributed a crucial role in the control of DNA–protein interactions (8–17). Most cases of cooperativity in DNA can be explained by a 'direct read-out' mechanism. Accordingly, a primary protein recognizes a DNA sequence by means of specific interactions with one of its domains, whilst other(s) domain(s) make(s) specific interactions with a secondary protein, positioning it near the second DNA sequence to be recognized (9–11). In addition to the 'direct readout' model, two other allosteric mechanisms have been suggested: the 'indirect read-out', where the primary protein distorts the structure of DNA improving the binding characteristics of the secondary site (12,13), and the 'solvent release mechanism' that assumes that primary binding induces changes in water or ion distribution reducing the desolvation cost required for the secondary binding (14). Recent experiments (15,18,19), and theoretical models (17,20) have shown cases of cooperativity in DNA–protein binding that apparently do not fit within any of the traditional paradigms, as binding sites are distant and binding-induced structural changes in DNA are absent or very mild. Although in proteins a model of 'allostery without conformational change' (21) has been described, very few studies have focused on similar processes in DNA. In fact, cooperative binding of BAMHI type II Endonuclease (22) and the glucocorticoid receptor DNA-binding domain (GRDBD) (23) to the DNA is (to our knowledge) the first described example of 'allostery without conformational changes' involving DNA (15,24). The difficulties in understanding the origins of the BAMHI–DNA–GRDBD allostery highlight our limited knowledge on the mechanisms in which information is transferred along DNA (18,25–27).

In this contribution, we propose a model of DNA allostery based on communication from site-to-site by entropy transfer using correlated motions to transmit information through the system. Moreover, we show that the source of allostery is the directionality of time-delayed correlations between the internal degrees of freedom of DNA, which accounts for causality and explains the thermodynamics of complex formation.

MATERIALS AND METHODS

Simulated systems

We explore allosteric effects in duplex DNA containing the canonical BAMHI binding site (d(GGATCC)) and the canonical GRDBD binding site (d(AGAACATGATGT TCT) separated by linkers of increasing length (4,7,11,15). In all cases four systems were simulated: the naked DNA, the BAMHI–DNA complex, the GRDBD–DNA complex and the BAMHI–DNA–GRDBD trimer (see Supplementary Figure S1). The 4 × 4 systems were created using

*To whom correspondence should be addressed.Tel: +34 93 4037156; Fax: +34 93 4037156; Email: modesto.orozco@irbbarcelona.org

Nucleic Acid Builder and standard B-DNA geometrical parameters (28,29), except for the binding region where the geometries were transferred from the respective crystal structures (PDB codes: 2BAM and 1R4R). Note that the GRDBD binding site used (the one coming from the X-ray structure), correspond to what authors from ref. (15) labelled as the 'reverse' GRDBD sequence.

System preparation

The systems were immersed in octahedral boxes of water, which were defined to guarantee no DNA atom was placed at less than 10 Å from any face of the periodic cell. Hydrated systems were then neutralized by adding Na^+ ions and extra 100 mM NaCl and subjected to a standard minimization/thermalization/pre-equilibration (30,31) procedure followed by a 50 ns equilibration prior to production runs. All the topologies and coordinate files were build with the *leap* program of AmberTools 15 (32) and/or with the utilities provided by GROMACS (33), and run with GROMACS machinery.

Production runs

Molecular dynamics (MD) trajectories were collected in the isothermal ($T = 298$ K), isobaric ($P = 1$ atm) ensemble, using a Langevin thermostat (34) and Andersen-Parrinello barostat (35,36). The parmbsc1 force-field was used for DNA (37), coupled with ff99SB-ILDN for proteins (38), Dang's parameters for ions (39) and TIP3P waters (40). Periodic boundary conditions and Particle Mesh Ewald (real space cutoff 12 Å and grid spacing 1.2 Å) were used to account for remote electrostatic interactions (41). Van der Waals contacts were truncated at the real space cutoff. All bonds containing hydrogen were constrained using LINCS (42), which allowed us to use an integration step of 2 fs. Trajectories were extended for 1 μs each, and system sizes range from 91 164 to 263 834 atoms. Additional trajectories where BAMHI was instantaneously bound in a 1 μs equilibrated naked DNA (linker size = 7) were collected to test the mechanism in which perturbation is transferred along DNA. As the direct contacts between the two proteins are possible for the 4-nt linker, most of the discussion is limited to the 7-, 11- and 15-nt linkers. All collected trajectories are available through our BigNasim database (43).

Analysis of the trajectories

Conformational analysis was performed with the use of Curves+ and Canal programs (44) to obtain the DNA internal coordinates at each time frame such as helical parameters and backbone dihedrals. Standard analysis on those helical parameters were performed using tools implemented in our NaFleX server (45), and further analysis was done with a combination of in-house tools.

Analysis of the structural response

Correlations between geometrical variables were evaluated taking into account the nature of the coordinates involved, either linear or circular (backbone torsions). Correlation between two directional variables was assessed with the use of Jammalamadaka formula (46). To compute time-delayed correlations we used the cross-correlation (47) between the time series of major groove widths at two positions on the DNA with a maximum lag of 5 ns (see the Supplementary Data for a more comprehensive description of these techniques). Correlations network analysis of backbone torsions was performed by computing all circular-circular correlation (edges) between backbone angles (nodes), and using them to build an interaction network represented as a descriptive network graph using the R package igraph (48). Classical molecular interaction potentials were computed (using our CMIP code (49)) to determine the changes in recognition properties from an enthalpic point of view (only considering Coulomb and VdW interactions), induced by BAMHI binding, on the region of DNA that binds GRDBD. The ionic strength and the reaction-field dielectric constant were set to 0.15 and 78.4 M, respectively, whilst the dielectric constant for DNA was set to 8 (50). A protonated methylamine was used as probe particle. Cation analysis was performed by determining the cation distribution in curvilinear cylindrical coordinates. The distribution of sodium cations around the DNA was determined from the last 200 ns of each MD trajectory, and analysed using Canion (51). The limits of the grooves were defined as reported elsewhere (52,53). In order to analyse the effect of protein interaction with DNA whilst reducing the thermal noise, we defined as protein 'sensing contacts' all pairs of amino-acid/nucleotide that when coming in close proximity to each other (distances between centres of mass below 7 Å and at least one atom pair distance below 3 Å) produce the most significant perturbations in the DNA. Selecting structures from the trajectories using these criteria yielded meta-trajectories consisting of at least 10 000 disperse frames that were analysed together.

Analysis of thermodynamical properties

Entropy calculations were performed using both the Schlitter and Andricioaei/Karplus methods (54,55) employing two separate alignment methods. Absolute entropies of the DNA heavy atoms in the secondary binding region (GRDBD recognition site) were calculated for naked DNA, and all complexes at increasing time windows (50 ns to 1 μs). From these time-dependent values, we used the Harris' extrapolation scheme (56) to obtain converged absolute entropies at infinite simulation time, and used them to calculate the entropy changes upon protein binding. The dihedral entropy for DNA backbone torsions was computed as a function of the Kullback–Leibler divergence (57) of real dihedral state populations from the assumed independent populations, as described by Cukier (58). The set of dihedrals was chosen at the interface between linker and secondary binding regions, and it includes torsions on both Watson–Crick strands (α, β, ε, γ and ζ). Using major groove widths fluctuations along the DNA we calculate the transfer entropy (TE) between 2 bp following the method proposed by Schreiber (59), computing the TE as a summation of Shannon entropy terms (57), which stems from the calculation of conditional entropies between time series separated in time by τ (chosen here to

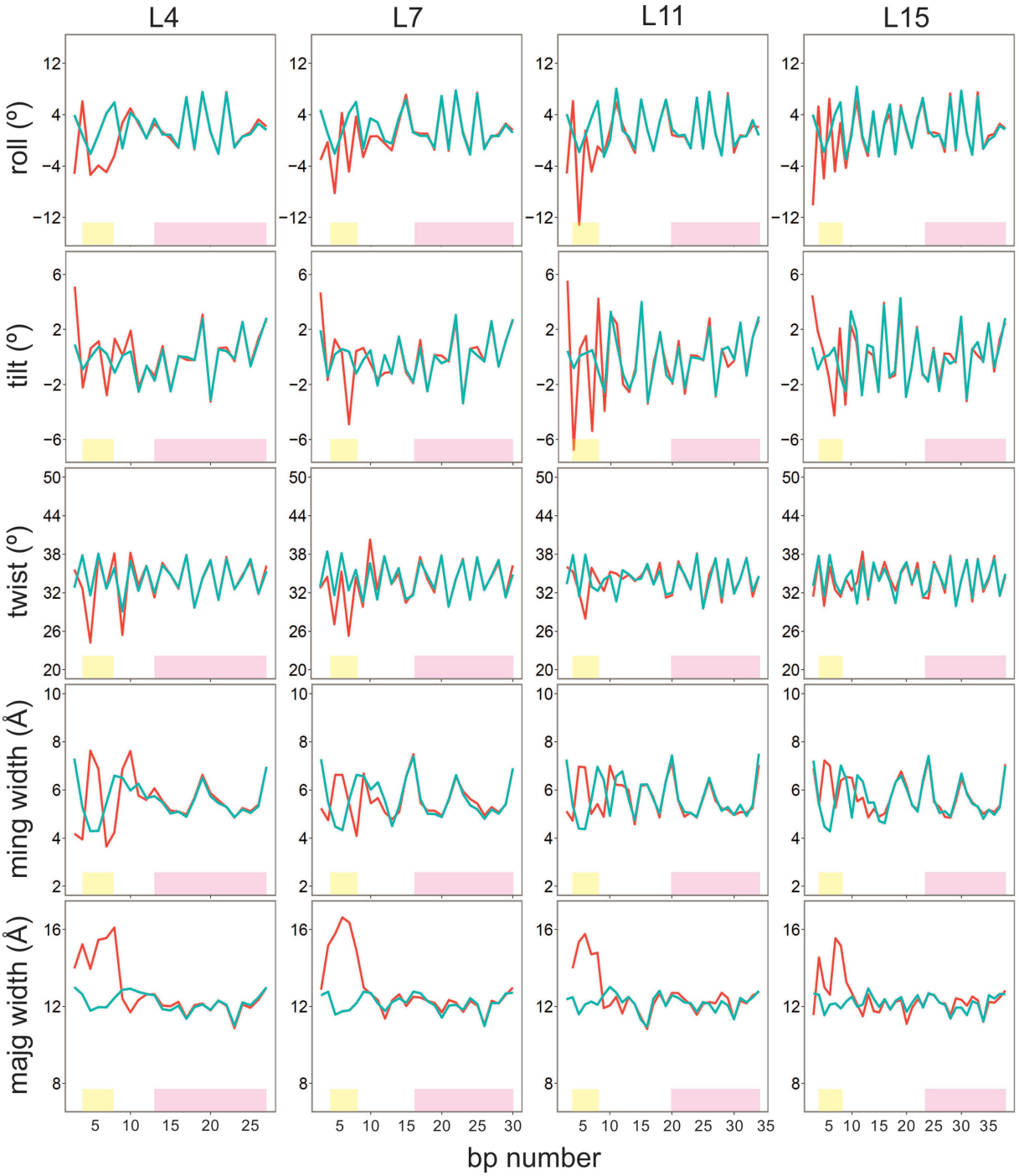

Figure 1. Average (0.2–1 μs) rotational inter base-pair parameters (in degrees) and groove widths (Major: majg and Minor: ming in Å) for the DNA duplexes with linkers of different lengths. Values for free DNA (cyan) versus those for BAMHI-bound DNA (red), where the binding region of BAMHI is highlighted in yellow and the recognition site for the GRDBD in magenta.

be 2 ns). A rough estimate of the difference in free energy of binding GRDBD to the free and protein bound DNA was obtained by following an adaptation of the confine-convert-release (CCR) method described by Roy *et al*. (60), based itself on previously described confinement methods (61,62). We calculate the energy of confining each structure (naked DNA, BAMHI–DNA, GRDBD–DNA and BAMHI–DNA–GRDBD complexes) to its energy minimum by thermodynamic integration with increasing restraints. The negative of this energy is the release term. In the convert step that completes the thermodynamic cycle, we calculate the energy difference of the DNA atoms between the two highly restrained complexes. Finally, the total binding free energy is calculated from the sum of these individual contributions.

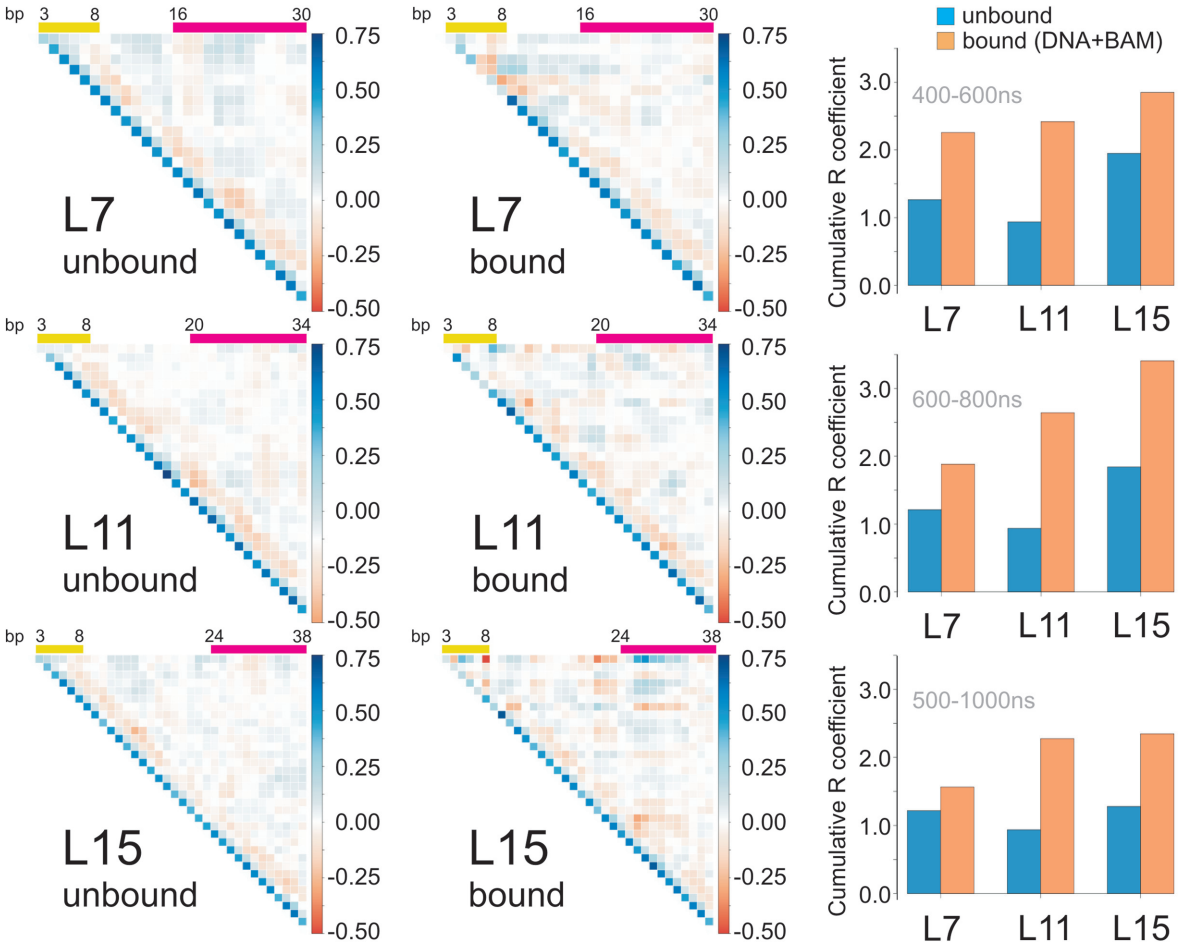

Figure 2. Left: upper diagonal matrix of major groove width correlation coefficients along the sequence for the unbound and BAMHI-bound DNA. The regions where the two proteins bind to the DNA are highlighted in yellow and magenta. Correlations have been calculated for the fully equilibrated last 200 ns of each trajectory. Right: cumulative R-coefficients between the nucleotides involved in BAMHI binding and the ones belonging to the GRDBD recognition site for different time windows of the simulation (P-value < 0.01 between replicas, see main text).

More details on all analyses and methods used are given in the Supplementary Data.

RESULTS AND DISCUSSION

The structural response

We explore here BAMHI–DNA–GRDB allostery (15) using a variety of theoretical approaches. We first investigated the possibility that direct protein–protein interactions can justify the observed cooperative binding of BAMHI and GRDBD to DNA (the list of simulated systems is shown in Supplementary Figure S1). To this end, we computed the protein–protein interaction energy during the last 100 ns of the 1 μs MD trajectories of DNA bound to BAMHI and GRDBD (10^4 snapshots for each case). Whilst for a short (4 bp) linker protein-protein interaction is sizeable (-17.7 ± 2.0 kcal/mol), for longer linkers it is negligible (<0.5 kcal/mol in all cases), precluding a 'direct readout' mechanism. Furthermore, analysis of helical parameters and groove dimensions (Figure 1 and Supplementary Figure S2) demonstrates that (when well equilibrated trajectories are used) the interaction of DNA with BAMHI does not significantly alter the helical geometry at the secondary (GRDBD) binding site, arguing against an 'indirect readout' model. Additionally, analysis of the MD trajectories clearly shows that BAMHI-induced changes in water and ion environment are restricted to the BAMHI region (see Supplementary Figure S3A), also arguing against the prevalence of an ion- or water-release mechanism. The ability of the DNA at the secondary binding region to recognize charged amino acids in the presence or absence of the effector protein (BAMHI) was assessed with our classical molecular interaction potential (CMIP (49)). We used a protonated methylamine probe to simulate the presence of a charged amino acid sidechain(see Supplementary Figure S3B) and found essentially no difference in the electrostatic or van der Waals terms between the bound and free DNA for any of the linker region sizes (see Supplementary Table S1). The cooperative effect detected experimentally should then be explained by a less common alternative mechanism, which implies a subtle flow of information between primary and secondary binding sites.

To investigate the mechanism of the information transfer along DNA we explored the correlation between the movements in primary and secondary binding sites for naked

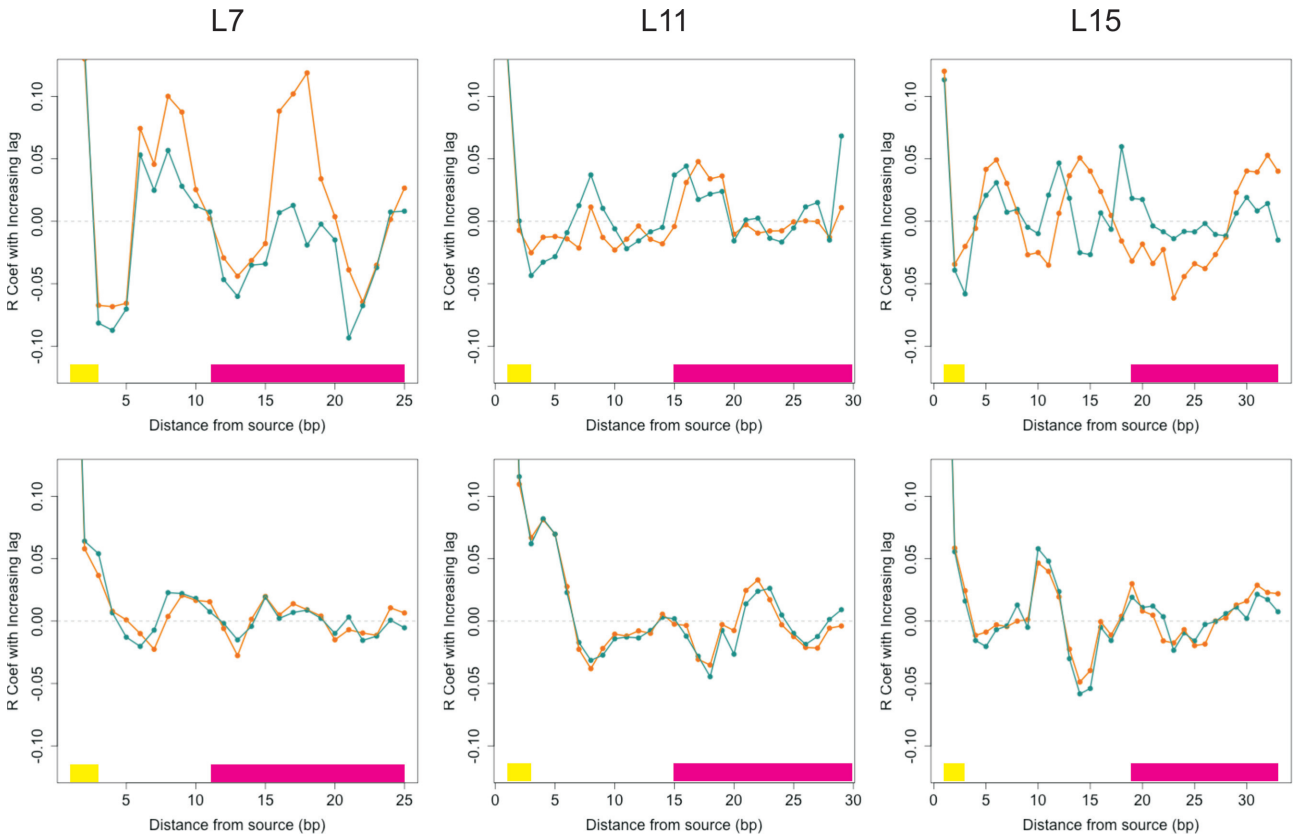

Figure 3. Time-delayed cross correlation of major groove widths between bp 5 belonging to the BAMHI binding regions and all subsequent bases in all linker size system. Top: bound complexes; Bottom: Corresponding naked DNAs. The abscissa denotes the distance in base pairs from the perturbation source (bp 5). 'Forward' correlations (5′ to 3′ on Watson strand, corresponding to cross-correlations from bp 5 to all other bases) are depicted in orange, whilst 'reverse' correlations (3′ to 5′ direction, referring to cross-correlations from each bp to bp 5) appear in blue. The curves are calculated as described in the Supplementary Data for the DNA+BAMHI complexes of each linker size.

and BAMHI bound DNAs. Early studies on this system suggested the existence of strong correlations in the movements at BAMHI and GRDBD binding sites (15,16), whilst more refined calculations showed that such correlations may emerge from equilibration artefacts (17). Indeed (as seen in Supplementary Figure S4), short simulations lead to an overestimation of the correlation between the two binding sites.But when long equilibration windows are considered, correlations between the two binding sites are still clearly larger for the DNA–BAMHI complex than for the naked DNA (Supplementary Figure S4 and Figure 2), suggesting that the GRDBD binding site feels in a dynamic way the presence of the BAMHI, even for the longest linker. Depending on the selected time window and the linker size, the differences in correlation strength between naked and bound DNA can vary,but they are always larger in the presence of the protein, most visibly so in the case of the 15 bp linker system, which also shows the higher cooperativity experimentally.

To establish if the structural correlations observed imply causation, we computed a time-delayed correlation between DNA residues (63), which account for the time lag that might appear as the signal travels from one binding site to the other (see Supplementary Data). The expected time lag at each position was calculated based on the linear progression of correlation maxima at the first 3 bp away from the source (bp 5), which show the least amount of noise. The assumption made here is that the signal travels at a constant speed through the sequence. Figure 3 shows correlation coefficients of the major groove widths between bp 5 and all subsequent base pairs with their corresponding time delay, either in the forward (5′→3′ on the Watson strand—from bp 5 to each other bp) or reverse direction (3′→5′—from all further base pairs to bp 5). The top half of Figure 3 depicts a comparison of such time-delayed major groove width correlations between BAMHI bound DNA of the different liker sizes.In the systems with cooperativity-favourable linker sizes (7 and 15 bp), the delayed correlation in the 5′→3′ direction (forward, orange line) is out of phase from the correlation in the 3′→5′ direction (reverse, blue line), so that at key positions, the correlation is significantly stronger in the forward direction. This is due to the fact that the forward correlation at such positions decays significantly slower than in the reverse case. These results indicate that the effect of the major groove fluctuations at BAMHI site (when this protein is bound), on later fluctuations at the secondary binding site, persist for longer times in the case of 7 and 15 bp linkers. In contrast, the 11 bp linker system has practically symmetric responses at these key positions. This suggest that the major groove width fluctuations in the BAMHI binding region drive the motions in the GRDBD binding region using the specific 5′-3′ direc-

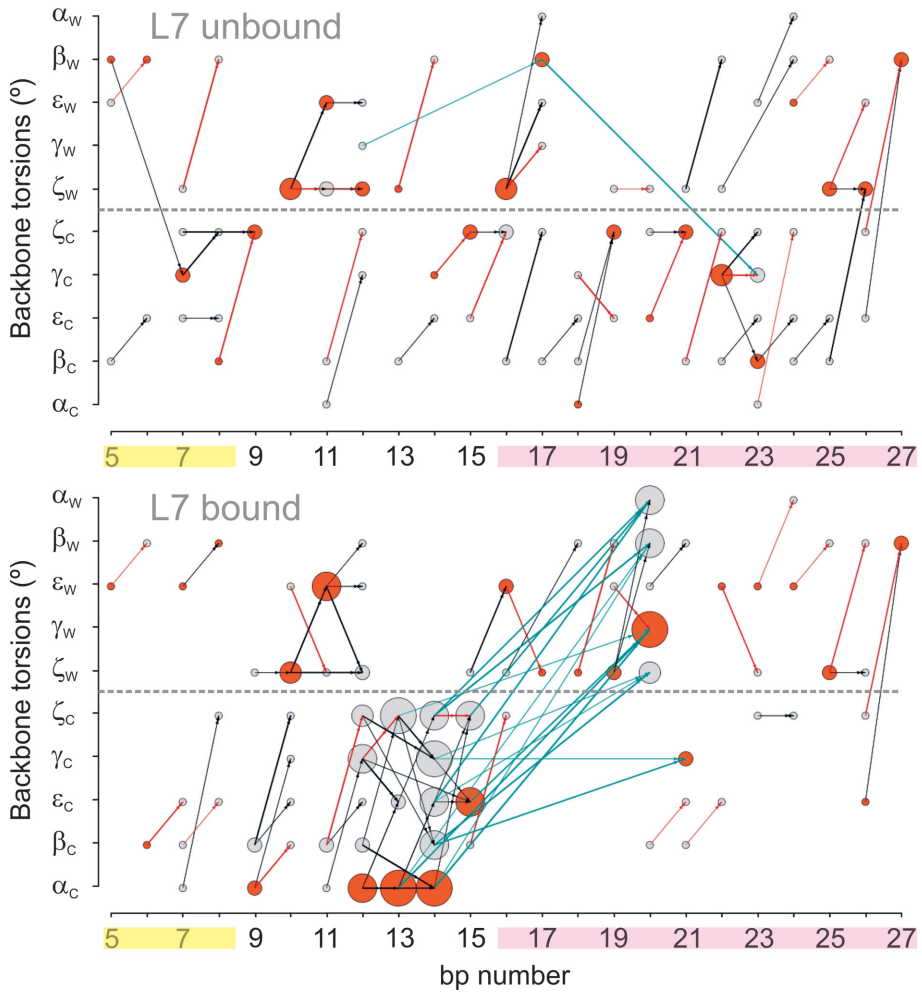

Figure 4. Correlations network analysis showing through space propagation of correlated motions in the DNA backbone (both the Watson (W) and Crick (C) strand torsions) with 0.5>|R|>0.4 for naked and BAMHI-bound DNA (case of 7 bp linker) from the centre of the BAMHI binding site along the sequence. Vertex size is proportional to their degree (the number of connections) and, for each level, the vertex with the highest degree is coloured in orange. Correlations between levels more than 4 bp apart are depicted as cyan arrows. From each level, the strongest correlation with a starting point on that level is shown as an orange arrow, unless it is already coloured in blue. Protein binding sites are highlighted in yellow and magenta.

tion for linkers 7 and 15, whereas the nature of the structural correlation between binding sites for the 11 bp linker system is small and non-directional. The bottom half of the figure shows the corresponding plots for the naked DNA of different linker sizes. In the absence of the perturbation induced by the binding of the effector protein, the phase shift does not appear. To further confirm that this behaviour is due to the introduction of a perturbation at the effector binding site, we ran a separate simulation where we introduced a gradual harmonic tear opening the major groove width at bp 5 (bound region) and looked on the effect of this perturbation to the forward and reverse delayed correlations (see Supplementary Figure S5). This result shows that a perturbation introduced in the binding region of BAMHI will indeed affect the symmetry of the cross-correlation. There is, additionally, significant difference (P-value < 0.01 for the cumulative r-square in the GRDBD binding site of favourable linker systems, obtained through pair-wise t-test) between correlation coefficients at the secondary binding region between free and bound DNA when taking the time delay into consideration. This suggests that correlations in the naked DNA might be intrinsic and determined by simply the shape of the double helix that synchronizes motions instantaneously, whereas in the protein-bound duplex the response to perturbation takes a certain amount of time to travel through the sequence.

The changes in the connection pattern between recognition sites in DNA due to BAMHI binding became even more evident when network analysis tools (see Supplementary Data) are used to find connectivity maps between the different backbone torsions (Figure 4; see results for a broader correlation range in Supplementary Figure S6). Clearly, the presence of BAMHI enriches the connectivity between the different torsional degrees of freedom. Interestingly, whilst such connections are mostly local and sequential for the naked DNA, BAMHI binding triggers crosstalk between the backbones of the two strands, through the space, and from the linker region to the secondary binding site, as expected from a 'hopping' information transfer mechanism (Figure 4 and Supplementary Figure S6), i.e.

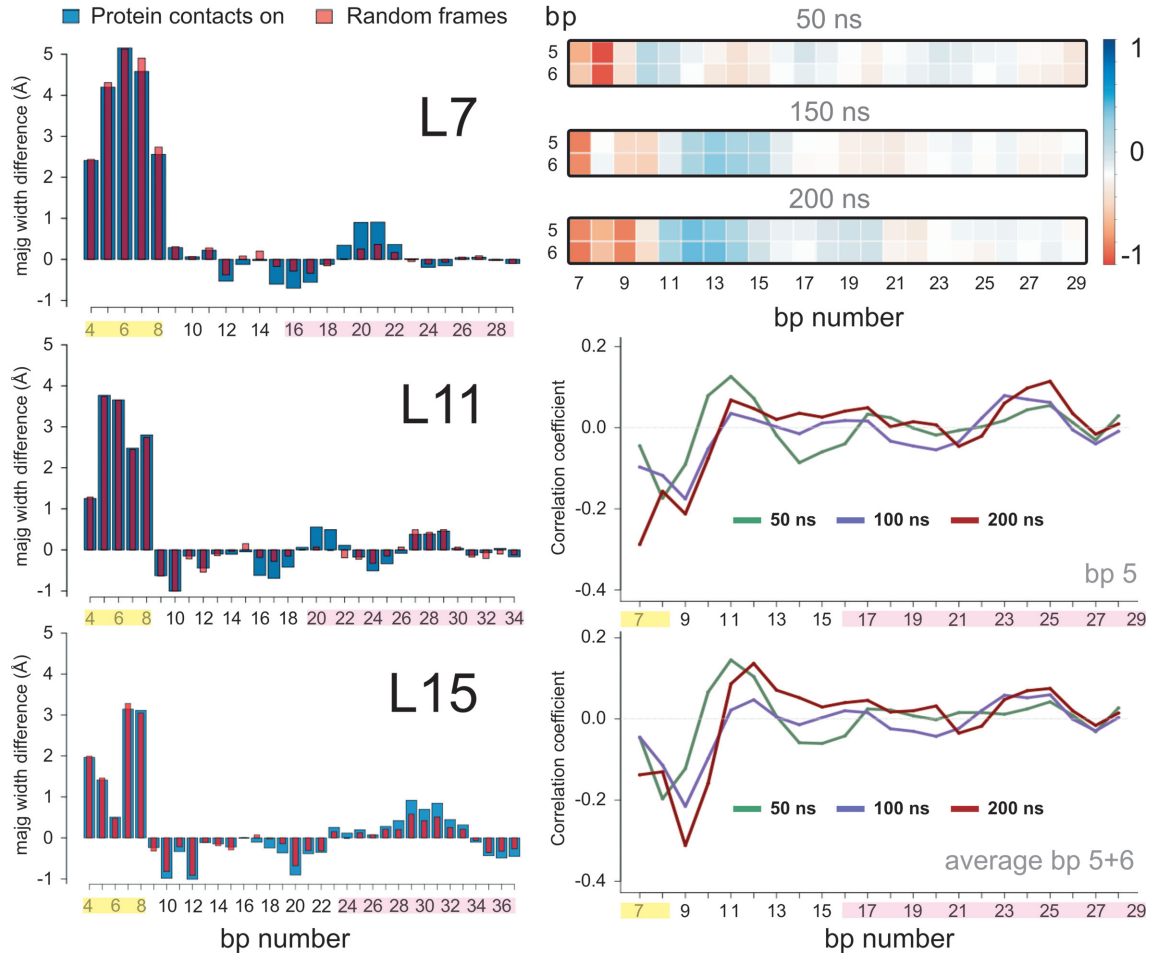

Figure 5. Left: differences in major groove width between naked and BAMHI-bound DNAs (7 bp linker) for those frames (around 10 000) where there are strong DNA–protein contacts (bars in blue). In each plot there is also an additional control obtained by choosing random snapshots of the same trajectories (i.e. ensembles not enriched in strong-contacts; bars in red). The protein binding sites are highlighted in yellow and magenta. Right: correlation coefficients between the groove width after 50, 100 and 200 ns of simulation time after the instantaneous insertion of BAMHI in its binding site in equilibrated DNAs for the 7-bp liker DNA. Average results over five replicas illustrate a large initial perturbation at the BAMHI binding site that over time is propagated and gains amplitude at the second binding site.

where the information flows by giving hops between the dihedrals backbone of non-sequential residues.

As noted above, the crosstalk between the two binding sites detected upon BAMHI binding does not lead in average to dramatic geometrical changes at distant regions (Figure 1 and Supplementary Figure S3), but generates pulses of distortion that can travel quite long distances generating non-negligible temporary geometrical distortions in the duplex. Although differences in major groove width between naked and BAMHI-bound DNAs are in average rather small out of the BAMHI binding site (Figure 5), they increase dramatically for those selected structures where we detected the strongest DNA–protein contacts (around 10^4 frames, see 'Materials and Methods' section). This suggests that 'protein-sensing' leads to a sizeable distortion pulse at quite long distances thanks to a 'hopping' mechanism with $\frac{1}{2}$ and full turn periodicities (Figure 5, left). During a protein-sensing event the signal can be transferred to remote regions of the DNA, but once the contact is released, the structure relaxes and the signal starts to dissipate, as shown by the evolution in time of major groove width correlations along the sequence during and after protein clenching (Supplementary Figure S7). Very interestingly, the distortion pattern introduced by 'protein sensing' is enhanced at the GRDBD binding site for linkers 7 and 15, i.e. those showing experimentally strong cooperativity, whilst for linker 11, where cooperativity is not experimentally detected, the peak of the perturbation wave is displaced with respect to the secondary recognition site. Therefore, the linker size modulates the impact of the distortion signal at the secondary binding site, suggesting that protein contacts affect not only the major groove width at the secondary binding site, but also the cross-talk between the two binding sites, as it can be observed from the time evolution of structural correlations during and after protein sensing (Supplementary Figure S7).

To further validate the idea that 'protein-sensing' generates a wave of distortion transmitting a pulse of information to distant regions of DNA, we analysed the correlation between the groove width after 50, 100 and 200 ns of the instantaneous insertion of BAMHI in its binding site in

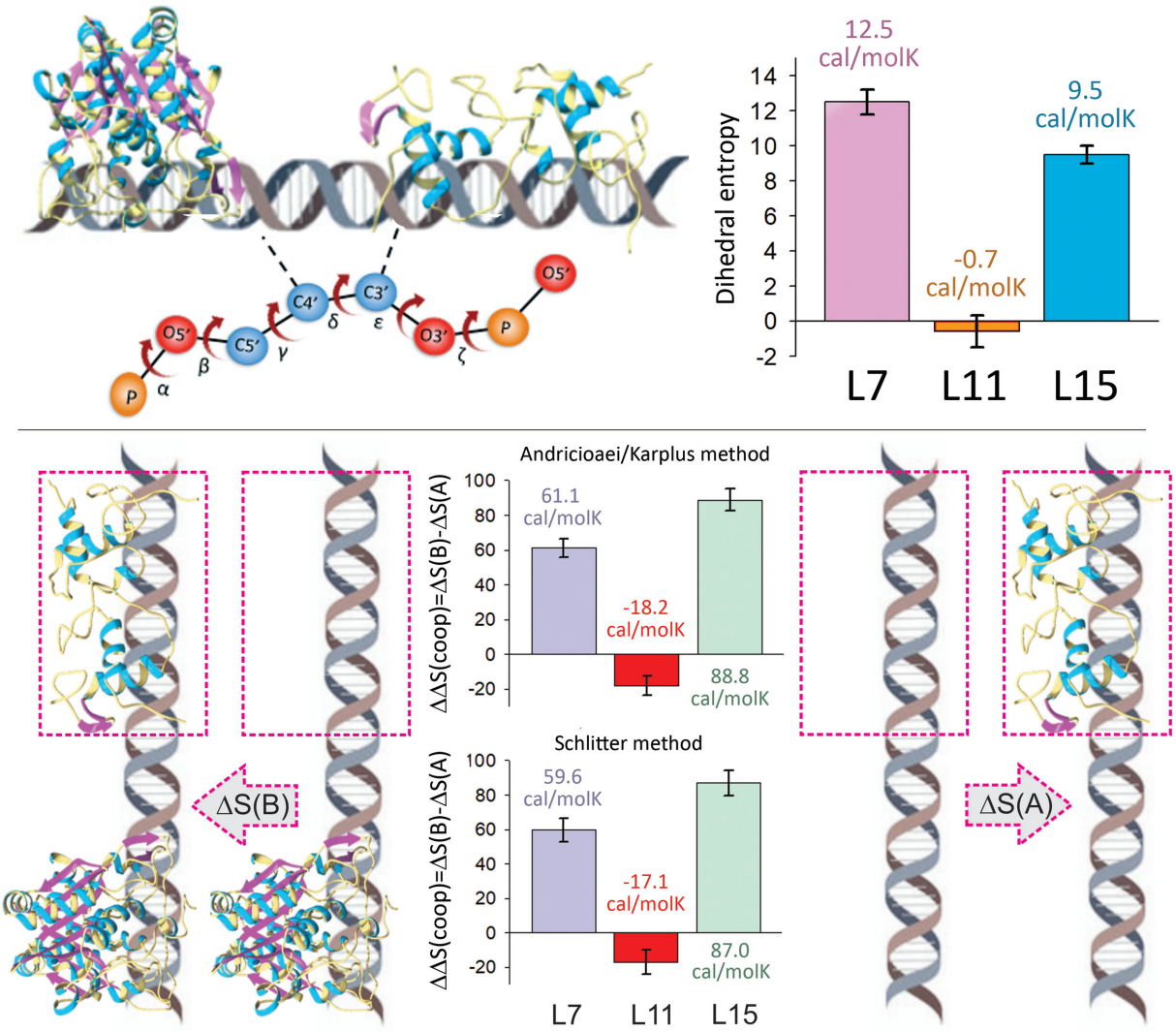

Figure 6. Top: dihedral entropy change at the interface between linker and secondary binding regions for linkers of 7, 11 and 15 bp. Bottom: DNA entropy variation ΔS(B)-ΔS(A) induced by the binding of GRDBD at the secondary binding region for DNA duplexes containing 7, 11 and 15 bp linkers. Positive values mean that entropy contributes favourably to the cooperative binding of BAMHI and GRDBD.

equilibrated DNAs. Average results in Figure 5 illustrate the generation of a perturbation wave by 'protein-sensing' that travels with a $\frac{1}{2}$ and 1 turn periodicity to reach the GRDBD binding site, confirming previous suggestions on the transient nature of the structural distortion.

The entropic origin of cooperativity

The analysis of the structural response presented above suggests that DNA acts as a wire transmitting pulses of information originated at the primary binding site that travel to distant regions. The existence of such a mechanism of information transfer is a necessary, but not sufficient condition for the appearance of cooperativity. So, the question is now how these changes in the dynamics of DNA affect binding thermodynamics. To answer this question, we first evaluate the impact that backbone correlations (shown in Figure 4) have in the DNA entropy. With this purpose we computed the dihedral entropy at the interface of the linker region and the secondary binding site for all linker sizes following the method described by R.I. Cukier (58), which measures the decrease of entropy arising from the dependence amongst the dihedrals (see Supplementary Data). The results (Figure 6, top) strongly suggest that the entropic change associated with the network of correlations at these positions depend on the linker size, in good qualitative agreement with experimental data (Figure 6, see Supplementary Data for details).

We further processed our equilibrated MD trajectories (see Supplementary Data) of the naked DNA, BAMHI–DNA, GRDBD–DNA and BAMHI–DNA–GRDBD complexes to determine the (DNA) entropy change arising from GRDBD binding in naked DNA and when DNA is already bound to BAMHI. From these values we can define the entropy cooperativity as $\Delta\Delta S(coop) = \Delta S(B)\text{-}\Delta S(A)$, where $\Delta S(A)$ and $\Delta S(B)$ are the entropic variation associated to the binding of the GRDBD protein to naked or BAMHI-bound DNA, respectively. Entropies were computed from the analysis of the mass-weighted covariance matrix as described by Andricioaiei–Karplus (55), and to

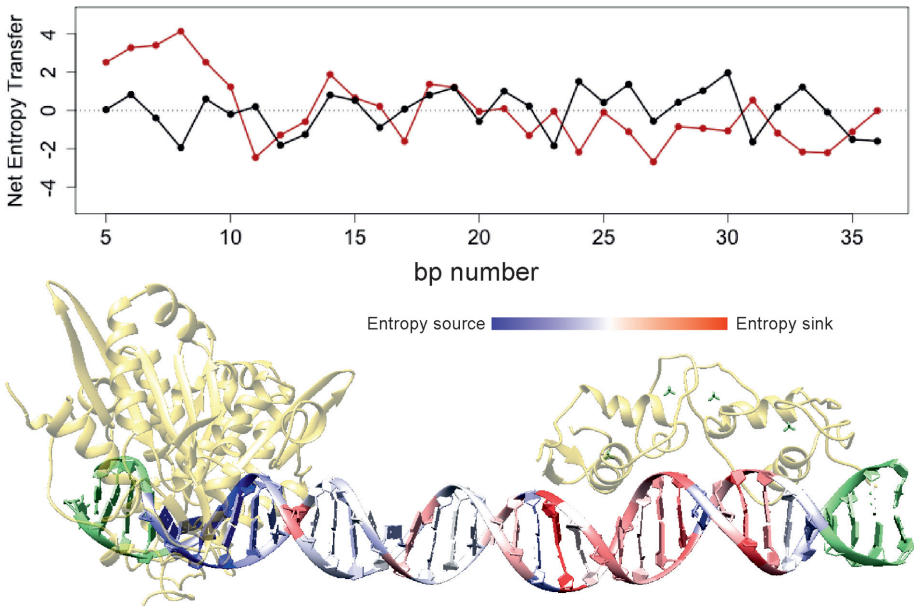

Figure 7. Top: net entropy transfer from each residue to the rest of the sequence, calculated for free (black line) and BAMHI–DNA (red line). Residues with positive values of net entropy transfer are entropy sources, whilst residues with negative values are entropy sinks. Bottom: values of net entropy transfer difference between bound and naked DNA are mapped onto the three-dimensional structure of the double helix (blue is for base pairs that are stronger entropy sources in the presence of BAMHI and red is for base pairs that behave more like entropy acceptors when the protein is bound compared to the free DNA).

have an additional estimate from Schlitter's formulation (54). To gain extra confidence on the robustness of the results two alignment methods were used to define the average structure, and estimates were obtained for different time-windows, which were then combined using Harris' extrapolation scheme (56) to obtain values extrapolated at infinite simulation time (see Supplementary Data for details). An example of the obtained results is summarized in Figure 6, bottom (the results are quite robust to the approach used to align the duplexes, to the procedure followed to transform oscillations into entropy measures, to the extension of the trajectory, or even through the simulation of replicas; see Supplementary Figure S8). Thus, as suggested by dihedral entropy measures above, the cooperativity studied here has an entropic origin. Very interestingly, for linkers of 7 and 15 bp, where large cooperative effects were detected (15), we observe that the entropy change associated to the binding of GRDBD is significantly reduced when the DNA is previously interacting with BAMHI (leading to a positive cooperative entropy term). On the contrary, for duplexes with an 11 bp linker no significant entropy differences are found when binding happens in naked or BAMHI-bound DNA, suggesting no sizeable cooperativity, in perfect qualitative agreement with experimental findings (15).

It is worth noting that the entropy-mediated mechanism of cooperative binding observed herein, is also supported by relative changes in the effective temperature computed from atomic oscillations in the presence/absence of the first protein (see Supplementary Figure S9). Additionally, the CCR calculations (61–62) (see Supplementary Data and Supplementary Figure S10) further confirm the expected free-energy change associated to cooperativity for linkers 7 and 15, in agreement with the relative k_{off} measured experimentally (15).

We went one step further and examined the information transfer landscape of the naked and BAMHI-bound DNA using Schreiber's formulation of entropy transfer (59). This approach allows us to find entropy sinks and sources upon the binding of the protein to DNA, and explains how given pairs of nucleotides from one binding site to the other communicate with each other using entropy transfer (64–67). Thus, based on the Shannon formulation of entropy (57), but taking into account the time delayed conditional probabilities of time series (59), we quantify the allosteric communication through the DNA (for details see Supplementary Data). Results are summarized in Figure 7 and Supplementary Figure S11 and show the entropy transfer landscape between residues of the DNA when BAMHI is bound (Supplementary Figure S11, right), and the quite uniform landscape of the naked DNA (Supplementary Figure S11, left). Without the protein, only few residues in the diagonal of the entropy map display net entropy ($T^{NET}_{i\rightarrow j}$) transfer (the communication is local in nature), whilst for most of the residues the information flowing in and out to the rest of the sequence is basically the same. The binding of BAMHI produces a drastic change, dominated by a sizeable net entropy transfer from the bp in the BAMHI bound region to the secondary binding region (thus, in the 5′→3′ direction), which involves several bp and is clearly non-local (Figure 7).The results show that in the presence of the effector protein (BAMHI), the base pairs of its binding site are major entropy sources for several base pairs along the sequence, whereas base pairs in the secondary binding region specifically change their entropy transfer characteristics, becoming notable acceptors of entropy (Figure 7). Analysing the

provenance of these changes as shown in the entropy transfer landscape of Supplementary Figure S11, the BAMHI binding region seems to be an exceptionally strong entropy source for the bases that bind to the GRDBD protein (Supplementary Figure S11), displaying directionality in the interactions of the two binding regions, which could be considered as an entropic switch that controls the binding of GRDBD.

CONCLUSION

Results reported here suggest that BAMHI binding to DNA generates a perturbation wave that travels to quite distant regions, and if the linker length is suitable, produces a change in structural correlations between residues in the secondary binding site. This change reduces the entropy cost associated to the second binding. We are pointing then to protein-induced changes in DNA-entropy as the origin of cooperativity in the explanation for BAMHI–DNA–GRDBD binding cooperativity. This type of entropy-mediated allostery was previously suggested for protein–protein interactions (21,68,69), and for the binding of small minor groove binders to DNA (27,56), but to our knowledge, it has not been previously described at the molecular level for DNA–protein binding. Our work also highlights the significant information transfer between base pairs in these systems. From the entropy transfer point of view, allosteric communication may be a general property of DNA that should be taken into consideration. Furthermore, we demonstrate that the knowledge of time delayed correlations and entropy transfer is needed to quantify allosteric cross-talk through the DNA, as an alternative to the established paradigms of cooperativity and allosterism. Time delayed events and causality analyses have only recently started to be viewed as crucial tools for studying allosteric communication in proteins. We now show that information transfer through DNA merits the same attention as a mechanism to explain cooperativity. We speculate that this entropy-mediated cooperativity can be quite general, considering that many proteins involved in DNA recognition are too small to make significant protein–protein contacts to account for the direct readout mechanism, that in many cases proteins do not introduce large remote structural distortions in DNA upon binding making the indirect readout also unlikely, and that the rearrangement of solvent molecules is usually quite local in nature precluding for most of the cases the solvent-release paradigm (70–72).

DATA AVAILABILITY

The trajectories and the associated analyses are accessible from the MuGBigNAsim portal: http://www.multiscalegenomics.eu/MuGVRE/modules/BigNASimMuG/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. R. Lavery for very useful scientific discussions, and Prof. J.L. Gelpi and Dr A. Hospital for help in the calculations of protein–protein interactions. M.O. is an ICREA (Institució Catalana de Recerca i Estudis Avançats) academia researcher. P.D.D. is a PEDECIBA (Programa de Desarrollo de las Ciencias Básicas) and SNI (Sistema Nacional de Investigadores, Agencia Nacional de Investigación e Innovación, Uruguay) researcher.

FUNDING

Spanish Ministry of Science [BFU2014-61670-EXP, BFU2014-52864-R]; Catalan SGR, the Instituto Nacional de Bioinformática; European Research Council (ERC SimDNA); European Union's Horizon 2020 research and innovation program [676556]; Biomolecular and Bioinformatics Resources Platform (ISCIII PT 13/0001/0030) co-funded by the Fondo Europeo de Desarrollo Regional (FEDER); MINECO Severo Ochoa Award of Excellence (Government of Spain) (to IRB Barcelona); PRACE 9th Call (Computer Resources). Funding for open access charge: European Research Council (ERC SimDNA) (to M.O.).
Conflict of interest statement. None declared.

REFERENCES

1. Changeux,J.-P. and Edelstein,S.J. (2005) Allosteric mechanisms of signal transduction. *Science*, **308**, 1424–1428.
2. Liu,J. and Nussinov,R. (2016) Allostery: an overview of its history, concepts, methods, and applications. *PLOS Comput. Biol.*, **12**, e1004966.
3. Cui,Q. and Karplus,M. (2008) Allostery and cooperativity revisited. *Protein Sci.*, **17**, 1295–1307.
4. Motlagh,H.N., Wrabl,J.O., Li,J. and Hilser,V.J. (2014) The ensemble nature of allostery. *Nature* **508**, 331–339.
5. Hilser,V.J., Wrabl,J.O. and Motlagh,H.N. (2012) Structural and energetic basis of allostery. *Annu. Rev. Biophys.*, **41**, 585–609.
6. Monod,J., Wyman,J. and Changeux,J.P (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.*, **12**, 88–118.
7. Koshland,D.E., Némethy,G. and Filmer,D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, **5**, 365–385.
8. Chaires,J.B. (2008) Allostery: DNA does it, too. *ACS Chem. Biol.*, **3**, 207–209.
9. Georges,A.B., Benayoun,B.A., Caburet,S. and Veitia,R.A. (2010) Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *FASEB J.*, **24**, 346–356.
10. Lelli,K.M., Slattery,M. and Mann,R.S. (2012) Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.*, **46**, 43–68.
11. Slattery,M., Riley,T., Liu,P., Abe,N., Gomez-Alcala,P., Dror,I., Zhou,T., Rohs,R., Honig,B., Bussemaker,H.J. *et al.* (2011) Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. *Cell*, **147**, 1270–1282.
12. Moretti,R., Donato,L.J., Brezinski,M.L., Stafford,R.L., Hoff,H., Thorson,J.S., Dervan,P.B. and Ansari,A.Z. (2008) Targeted chemical wedges reveal the role of allosteric DNA modulation in Protein−DNA assembly. *ACS Chem. Biol.*, **3**, 220–229.
13. Chenoweth,D.M. and Dervan,P.B. (2009) Allosteric modulation of DNA by small molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 13175–13179.
14. Harris,L.-A., Williams,L.D. and Koudelka,G.B. (2014) Specific minor groove solvation is a crucial determinant of DNA binding site recognition. *Nucleic Acids Res.*, **42**, 14053–14059.
15. Kim,S., Brostromer,E., Xing,D., Jin,J., Chong,S., Ge,H., Wang,S., Gu,C., Yang,L., Gao,Y.Q. *et al.* (2013) Probing allostery through DNA. *Science*, **339**, 816–819.
16. Xu,X., Ge,H., Gu,C., Gao,Y.Q., Wang,S.S., Thio,B.J., Hynes,J.T., Xie,X.S. and Cao,J. (2013) Modeling spatial correlation of DNA Deformation: DNA allostery in protein binding. *J. Phys. Chem. B*, **117**, 13378–13387.

17. Dršata,T., Zgarbová,M., Jurečka,P., Šponer,J. and Lankaš,F. (2016) On the use of molecular dynamics simulations for probing allostery through DNA. *Biophys. J.*, **110**, 874–876.
18. Lesne,A., Foray,N., Cathala,G., Forné,T., Wong,H. and Victor,J.M. (2015) Chromatin fiber allostery and the epigenetic code. *J. Phys. Condens Matter*, **27**, 64114.
19. Camunas-Soler,J., Alemany,A. and Ritort,F. (2017) Experimental measurement of binding energy, selectivity, and allostery using fluctuation theorems. *Science*, **355**, 412–415.
20. Xu,X., Ge,H., Gu,C., Gao,Y.Q., Wang,S.S., Thio,B.J., Hynes,J.T., Xie,X.S. and Cao,J. (2013) Modeling spatial correlation of DNA deformation: DNA allostery in protein binding. *J. Phys. Chem. B*, **117**, 13378–13387.
21. Cooper,A. and Dryden,D.T. (1984) Allostery without conformational change. A plausible model. *Eur. Biophys. J.*, **11**, 103–109.
22. Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656–663.
23. Luisi,B.F., Xu,W.X., Otwinowski,Z., Freedman,L.P., Yamamoto,K.R. and Sigler,P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, **352**, 497–505.
24. Narasimhan,K., Pillay,S., Huang,Y.-H., Jayabal,S., Udayasuryan,B., Veerapandian,V., Kolatkar,P., Cojocaru,V., Pervushin,K. and Jauch,R. (2015) DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. *Nucleic Acids Res.*, **43**, 1513–1528.
25. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in Protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
26. Panne,D. (2008) The enhanceosome. *Curr. Opin. Struct. Biol.*, **18**, 236–242.
27. Tevis,D.S., Kumar,A., Stephens,C.E., Boykin,D.W. and Wilson,W.D. (2009) Large, sequence-dependent effects on DNA conformation by minor groove binding compounds. *Nucleic Acids Res.*, **37**, 5550–5558.
28. Cheatham,T.E. 3rd, Brooks,B.R. and Kollman,P.A. (2001) Molecular modeling of nucleic acid structure. *Curr. Protoc. Nucleic Acid Chem.*, **6**, 7.5.1–7.5.12.
29. Neidle,S. (1999) *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press.
30. Dans,P.D., Danilāne,L., Ivani,I., Dršata,T., Lankaš,F., Hospital,A., Walther,J., Pujagut,R.I., Battistini,F. and Gelpí,J.L. (2016) Long-timescale dynamics of the Drew–Dickerson dodecamer. *Nucleic Acids Res.*, **44**, 4052–4066.
31. Pérez,A., Luque,F.J. and Orozco,M. (2007) Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.*, **129**, 14739–14745.
32. Case,D.A., Babin,V., Berryman,J., Betz,R.M., Cai,Q., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Gohlke,H. *et al.* (2014) *Amber 14*. University of California, San Francisco, CA
33. Lindahl,E., Hess,B. and van der Spoel,D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**, 306–317.
34. Loncharich,R.J., Brooks,B.R. and Pastor,R.W. (1992) Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N?-methylamide. *Biopolymers*, **32**, 523–535.
35. Andersen,H.C. (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, **72**, 2384–2393.
36. Parrinello,M. and Rahman,A. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, **52**, 7182–7190.
37. Ivani,I., Dans,P.D., Noy,A., Pérez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goñi,R., Balaceanu,A. *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.
38. Lindorff-Larsen,K., Piana,S., Palmo,K., Maragakis,P., Klepeis,J.L., Dror,R.O. and Shaw,D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.
39. Smith,D.E. and Dang,L.X. (1994) Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.*, **100**, 3757–3766.
40. Jorgensen,W.L., Chandrasekhar,J. and Madura,J.D. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
41. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
42. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
43. Hospital,A., Andrio,P., Cugnasco,C., Codo,L., Becerra,Y., Dans,P.D., Battistini,F., Torres,J., Goñi,R., Orozco,M. *et al.* (2016) BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res.*, **44**, D272–D278.
44. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
45. Hospital,A., Faustino,I., Collepardo-Guevara,R., González,C., Gelpí,J.L. and Orozco,M. (2013) NAFlex: a web server for the study of nucleic acid flexibility. *Nucleic Acids Res.*, **41**, W47–W55.
46. Jammalamadaka,S.R. and SenGupta,A. (1963) *Topics in Circular Statistics*. World Scientific, River Edge, New Jersey, 2001.
47. Bolinder,E.F. (1963) The Fourier integral and its applications. *Proc. IEEE*, **51**, 267.
48. Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research | BibSonomy. *InterJ. Complex Syst.*, **1695**, 1–9.
49. Gelpí,J.L., Kalko,S.G., Barril,X., Cirera,J., de La Cruz,X., Luque,F.J. and Orozco,M. (2001) Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. *Proteins*, **45**, 428–437.
50. Cuervo,A., Dans,P.D., Carrascosa,J.L., Orozco,M., Gomila,G. and Fumagalli,L. (2014) Direct measurement of the dielectric polarization properties of DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E3624–E3630.
51. Lavery,R., Maddocks,J.H., Pasi,M. and Zakrzewska,K. (2014) Analyzing ion distributions around DNA. *Nucleic Acids Res.*, **42**, 8138–8149.
52. Dans,P.D., Faustino,I., Battistini,F., Zakrzewska,K., Lavery,R. and Orozco,M. (2014) Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.*, **42**, 11304–11320.
53. Pasi,M., Maddocks,J.H. and Lavery,R. (2015) Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res.*, **43**, 2412–2423.
54. Schlitter,J. (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.*, **215**, 617–621.
55. Andricioaei,I. and Karplus,M. (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.*, **115**, 6289–6292.
56. Harris,S.A., Gavathiotis,E., Searle,M.S., Orozco,M. and Laughton,C.A. (2001) Cooperativity in drug-DNA recognition: a molecular dynamics study. *J. Am. Chem. Soc.*, **123**, 12658–12663.
57. Cover,T.M. and Thomas,J.A. (2015) *Elements of Information Theory*. Wiley-Interscience, New Jersey, 2006.
58. Cukier,R.I. (2015) Dihedral angle entropy measures for intrinsically disordered proteins. *J. Phys. Chem. B*, **119**, 3621–3634.
59. Schreiber,T. (2000) Measuring information transfer. *Phys. Rev. Lett.*, **85**, 461–464.
60. Roy,A., Perez,A., Dill,K.A. and MacCallum,J.L. (2014) Computing the relative stabilities and the Per-Residue components in protein conformational changes. *Structure*, **22**, 168–175.
61. Tyka,M.D., Clarke,A.R. and Sessions,R.B. (2006) An efficient, Path-independent method for free-energy calculations. *J. Phys. Chem. B*, **110**, 17212–17220.
62. Cecchini,M., Krivov S,V., Spichty,M. and Karplus,M. (2009) Calculation of free-energy differences by confinement simulations. Application to peptide conformers. *J. Phys. Chem. B*, **113**, 9728–9740.
63. Vatansever,S., Gümüş,Z.H. and Erman,B. (2016) Intrinsic K-Ras dynamics: a novel molecular dynamics data analysis method shows causality between residue pair motions. *Sci. Rep.*, **6**, 37012.
64. Gourévitch,B. and Eggermont,J.J. (2007) Evaluating information transfer between auditory cortical neurons. *J. Neurophysiol.*, **97**, 2533–2543.
65. Staniek,M. and Lehnertz,K. (2008) Symbolic transfer entropy. *Phys. Rev. Lett.*, **100**, 158101.

66. Hacisuleyman,A. and Erman,B. (2017) Entropy transfer between residue pairs and allostery in Proteins: Quantifying allosteric communication in Ubiquitin. *PLOS Comput. Biol.*, **13**, e1005319.
67. Kamberaj,H. and van der Vaart,A. (2009) Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys. J.*, **97**, 1747–1755.
68. Capdevila,D.A., Braymer,J.J., Edmonds,K.A., Wu,H. and Giedroc,D.P. (2017) Entropy redistribution controls allostery in a metalloregulatory protein. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 4424–4429.
69. Guo,J. and Zhou,H.-X. (2016) Protein allostery and conformational dynamics. *Chem. Rev.*, **116**, 6503–6515.
70. Remenyi,A., Lins,K., Nissen,L.J., Reinbold,R., Schöler,H.R. and Wilmanns,M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.*, **17**, 2048–2059.
71. Tahirov,T.H., Sato,K., Ichikawa-Iwata,E., Sasaki,M., Inoue-Bungo,T., Shiina,M., Kimura,K., Takata,S., Fujikawa,A., Morii,H. *et al.* (2002) Mechanism of c-Myb-C/EBP beta cooperation from separated sites on a promoter. *Cell*, **108**, 57–70.
72. Shiina,M., Hamada,K., Inoue-Bungo,T., Shimamura,M., Uchiyama,A., Baba,S., Sato,K., Yamamoto,M. and Ogata,K. (2015) A novel allosteric mechanism on Protein–DNA interactions underlying the Phosphorylation-Dependent regulation of Ets1 target gene expressions. *J. Mol. Biol.*, **427**, 1655–1669.