Mechanisms of DNA-Mediated Allostery

Midas Segers,[1] Aderik Voorspoels,[1] Takahiro Sakaue,[2] and Enrico Carlon[1]
[1]Soft Matter and Biophysics, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
[2]Department of Physical Sciences, Aoyama Gakuin University,
5-10-1 Fuchinobe, Chuo-ku, Sagamihara, Kanagawa 252-5258, Japan

(Received 23 May 2023; accepted 7 November 2023; published 7 December 2023)

Proteins often regulate their activities via allostery—or action at a distance—in which the binding of a ligand at one binding site influences the affinity for another ligand at a distal site. Although less studied than in proteins, allosteric effects have been observed in experiments with DNA as well. In these experiments two or more proteins bind at distinct DNA sites and interact indirectly with each other, via a mechanism mediated by the linker DNA molecule. We develop a mechanical model of DNA/protein interactions which predicts three distinct mechanisms of allostery. Two of these involve an enthalpy-mediated allostery, while a third mechanism is entropy driven. We analyze experiments of DNA allostery and highlight the distinctive signatures allowing one to identify which of the proposed mechanisms best fits the data.

DOI: 10.1103/PhysRevLett.131.238402

Introduction.—The term "allostery" indicates an action at a distance in biological macromolecules where the binding of a ligand at one site modifies the binding of another ligand at a distinct site. Many proteins regulate their activities via allostery [1], through mechanisms that are not fully understood and presently debated, see e.g., [2]. Although hitherto most of the focus has been on proteins, allosteric effects have been observed in DNA as well [3,4] and discussed in models and simulations [5–7]. In DNA allostery two or more proteins binding at distinct sites interact with each other through some signal carried by the linker DNA, see Fig. 1(a). Experiments show that the interaction is weakly dependent on the DNA sequence [3,4], suggesting that allostery may be described by a homogenous DNA model. The interaction is strongly attenuated if one of the two strands is cut (DNA nicking), which shows that allostery requires an intact DNA molecule. This and other experimental evidences [3,4] show that the interaction is transmitted through DNA, and not via direct (electrostatic) or solvent-mediated effects. A model of force-induced allostery was discussed in [8]. However, this mechanism does not apply to experiments in which DNA is not under tension [3,4].

In all generality, the total free energy for the system consisting of DNA and two bound proteins, separated by a linker sequence of m base pairs, is

$$F_{ab} = F_0 + \Delta F_a + \Delta F_b + \Delta\Delta F_{\rm int}(m) \qquad (1)$$

where F_0 is the bulk contribution from DNA in absence of bound proteins, ΔF_a and ΔF_b are the excess free energies when only one of the two proteins, either "a" or "b", is bound. The interaction term, $\Delta\Delta F_{\rm int}(m)$, is the excess free energy when both proteins are bound which vanishes as $m \to \infty$. If $\Delta\Delta F_{\rm int} < 0$ the simultaneous binding of the two proteins leads to a net decrease of the total free energy (cooperative binding). If $\Delta\Delta F_{\rm int} > 0$, the simultaneous binding is destabilized. We introduce a model which predicts three distinct mechanisms of allostery, corresponding to different forms of $\Delta\Delta F_{\rm int}$. We introduce collective variables X_n at each base pair position $0 \le n < N$, which are local reaction coordinates associated to DNA-protein binding. We define $\bar{l} \equiv \langle X_n \rangle$, the equilibrium value, and

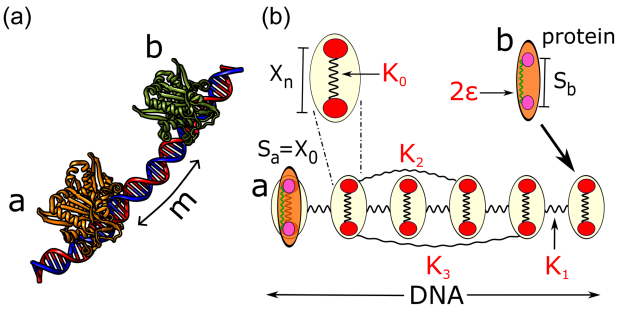

FIG. 1. (a) DNA-mediated interaction for two bound proteins separated by a linker molecule of m base pairs. (b) Model of DNA allostery. The DNA substrate (yellow) is described as a set of variables $X_n \equiv u_n + \bar{l}$ defined at each base-pair site, with $\langle X_n \rangle = \bar{l}$ the equilibrium value. These variables are characterized by a local stiffness (K_0) and distal couplings ($K_1, K_2, \ldots$), with energy given by (2). Upon binding, the protein (orange blob) modifies the local mechanical properties of the DNA substrate. The distal couplings carry the signal to distinct sites. The schematic plot in (b) shows a protein interacting with a single DNA site. The more realistic case of proteins binding to several DNA sites is also considered.

$u_n \equiv X_n - \bar{l}$. At base-pair level, DNA deformations are described by several coarse-grained coordinates like the 12 canonical ones (twist, roll, tilt, rise, …) of the rigid base model [9]. In our model u_n could be one of these coordinates, a combination thereof, or any other local deformation parameter. Experimental data, discussed further, put constraints on the properties of u_n which gives insights on candidate allostery-carrying variables. In the u_n the energy of free DNA is

$$H_0 = \frac{1}{2}\sum_{n=0}^{N-1}\left[K_0 u_n^2 + \sum_{p=1}^{L} K_p(u_n u_{n+p} + u_n u_{n-p})\right] \quad (2)$$

which is quadratic with local stiffness K_0 and distal couplings $K_p u_n u_{n+p}$ [Fig. 1(b)] here assumed to extend to a finite range L. Distal couplings naturally arise from the collective nature of u_n and are indeed observed in simulations [10–14]. They typically extend to a few flanking nucleotides [truncated at a distance L in (2)] and the strength and decay of the interactions depend on the coarse-grained variable considered [13]. Distal couplings are essential to generate allostery. The model (2) is coarse-grained with one degree of freedom per DNA site and it can be solved analytically. Using periodic boundary conditions ($u_N \equiv u_0$) we write (2) as

$$H_0 = \frac{1}{2N}\sum_q \tilde{K}_q |\mathcal{U}_q|^2 \quad (3)$$

where we introduced the discrete Fourier transforms

$$\mathcal{U}_q = \sum_{n=0}^{N-1} e^{-2\pi i n q/N} u_n, \qquad \tilde{K}_q = \sum_{n=-L}^{L} K_n \cos\left(\frac{2\pi n q}{N}\right) \quad (4)$$

with q an integer, $K_{-p} = K_p$, and $-N/2 < q < N/2$ [15]. In absence of distal couplings ($K_p = 0$ for $p \geq 1$) the q stiffness $\tilde{K}_q = K_0$ is constant, thus a q dependence of $\tilde{K}_q$ reflects the existence of couplings between distal sites. Note that the couplings K_p can take any value as long as $\tilde{K}_q > 0$ for real q (stability condition).

We consider first proteins interacting with a single DNA site. An unbound protein is thus described by a single collective variable S, with average $\langle S\rangle = \bar{s}$ and energy $H_p = \varepsilon(S - \bar{s})^2$. The binding to DNA (at site $n = 0$) forces the corresponding collective variables to assume the same value $S = X_0 = \bar{l} + u_0$ so that the total energy of DNA and protein ($H_0 + H_p$) takes the form

$$H = H_0 + \varepsilon[u_0^2 + 2(\bar{l} - \bar{s})u_0 + (\bar{l} - \bar{s})^2], \quad (5)$$

omitting a constant binding energy which does not influence $\Delta\Delta F_{\text{int}}$. Equation (5) shows that protein binding introduces perturbation "fields" proportional to u_0 and u_0^2. If the equilibrium value of the collective coordinates of DNA and protein coincide ($\bar{s} = \bar{l}$), the linear term vanishes and only the term proportional to u_0^2 "survives" in (5). Conversely, if $|\bar{l} - \bar{s}|$ is large one can neglect the quadratic term contribution [16]. In the following we compute $\Delta\Delta F_{\text{int}}$ for three different cases where the proteins induce a linear or a quadratic field.

A quantity of central interest is the propagator

$$S_m \equiv \frac{1}{N}\sum_q \frac{e^{2\pi i m q/N}}{\tilde{K}_q} = \beta\langle u_0 u_m\rangle_0, \quad (6)$$

where $\beta = 1/k_B T$ is the inverse temperature and $\langle\cdot\rangle_0$ indicates a thermal average with respect of H_0. Equation (6) follows from the equipartition theorem

$$\beta\langle\mathcal{U}_q\mathcal{U}_p\rangle_0 = N\tilde{K}_q^{-1}\delta_{q,-p}. \quad (7)$$

Transforming the sum in (6) into an integral ($N \to \infty$ limit), one obtains the asymptotic behavior of S_m from the leading pole, i.e., the solution of $\tilde{K}_q = 0$ with the smallest imaginary part. This equation cannot have solutions for real q as stability requires $\tilde{K}_q > 0$. In the most general case the leading pole has real and imaginary parts. Since $\tilde{K}_q$ is real for real q and symmetric in $\pm q$ [see (4)] there are at least four poles, one of which is

$$\frac{2\pi q_E}{N} \equiv \phi + \frac{i}{\xi_E} \quad (8)$$

and the others are $-q_E$, q_E^* and $-q_E^*$, where * denotes complex conjugation. The asymptotic behavior is governed by q_E

$$S_m \overset{m\gg 1}{\sim} \Gamma\cos(m\phi + \phi_0)e^{-m/\xi_E} \quad (9)$$

with ϕ_0 a phase shift and Γ a scale factor [17].

Enthalpic allostery.—We consider first

$$H^E = H_0 - h(u_0 + u_m) \quad (10)$$

with $h = -2\varepsilon(\bar{l} - \bar{s})$, following (5). We find [17]

$$\Delta\Delta F_{\text{int}}^E = -h^2 S_m \quad (11)$$

which, being temperature independent, describes an interaction of enthalpic origin [30]. The fields in 0 and m shift the equilibrium values of u_0 and u_m and the distal couplings $K_1, K_2, \ldots$ propagate this perturbation to flanking sites, leading to $\langle u_n\rangle \neq 0$. Asymptotically the interaction decays via damped oscillations, see (9). We refer to ξ_E in (9) as the enthalpic allosteric length. The interaction stabilizes or destabilizes the simultaneous protein binding depending on their distance m, see Fig. 2(a). The calculation can be

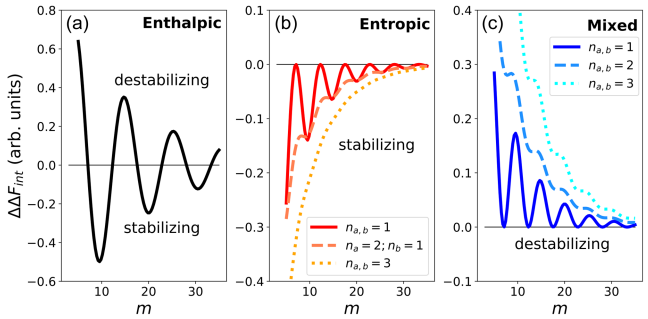

FIG. 2. Plots of $\Delta\Delta F_{\rm int}$ vs linker DNA length m for the three different mechanisms of DNA-mediated allostery proposed: (a) enthalpic, (b) entropic, and (c) mixed. An angular frequency $\phi = 2\pi/10.5$, corresponding to the periodicity of the DNA double helix and $\xi_E = 15$ bp were used. These parameters match those observed in experiments [see Fig. 3]. In the case (a) the interaction is stabilizing/destabilizing depending on the values of m and the asymptotic oscillating behavior is universal, i.e., also valid for interactions involving n_a, n_b sites for protein a and b, respectively. In the cases (b) and (c) the interaction is always stabilizing and destabilizing, respectively. In (b) and (c) the asymptotic decay is nonuniversal, being dependent on the number of interacting sites $n_{a,b}$ per protein.

generalized to protein-DNA couplings involving more than one site, i.e., of the type $\sum_{k=0}^{n_a-1} h_k u_k + \sum_{l=n_a-1}^{n_a+n_b-2} h_{m+l} u_{m+l}$ where n_a (n_b) are the number of sites to which first (second) protein binds and m is the number of base pairs separating the nearest edges of the two proteins. The asymptotic decay remains of the form (9) which is universal [17].

Entropic allostery.—We consider next

$$H^S = H_0 + \varepsilon\left(u_0^2 + u_m^2\right). \tag{12}$$

Differently from the enthalpic case, here $\langle u_n \rangle = 0$ for all sites. We find [17]

$$\Delta\Delta F^S_{\rm int} = \frac{k_B T}{2} \log\left[1 - \left(\frac{2\varepsilon S_m}{1 + 2\varepsilon S_0}\right)^2\right]. \tag{13}$$

The interaction is of entropic origin $\Delta\Delta F^S_{\rm int} = -T\Delta\Delta S \le 0$, implying a net increase in entropy when both proteins are bound (cooperative binding). This can be understood as follows. The local stiffening to $K_0 + 2\varepsilon$ at sites 0 and m induces an entropy reduction in two regions surrounding the two perturbed sites. When m is sufficiently small, the two regions overlap which leads to a net entropy gain, hence $\Delta\Delta S > 0$. This is reminiscent of entropic attractions observed in soft condensed matter systems, such as polymer-colloid mixtures [31]. In the limit $m \gg 1$, $\Delta\Delta F^S_{\rm int}$ vanishes as S_m^2. This implies [Eq. (9)] a decay length which is half of the enthalpic allosteric length $\xi_S = \frac{1}{2}\xi_E$, and an oscillating prefactor proportional to $\cos^2(m\phi + \phi_0)$, as shown in Fig. 2(b), red solid line. We extended the analysis of $\Delta\Delta F^S_{\rm int}$ for protein-DNA contacts at more than one site. We consider first $\varepsilon(u_0^2 + u_1^2 + u_{m+1}^2)$, which can be solved analytically ([17], Eq. (S43)) shown as dashed line in Fig. 2(b). Unlike (13), this extended binding case contains terms proportional to S_m^2, S_{m+1}^2, and $S_m S_{m+1}$, each oscillating but with different phases. Figure 2(b) (dotted) shows $\Delta\Delta F^S_{\rm int}$ for an interaction term $\varepsilon(\sum_{l=0}^{2} u_l^2 + \sum_{k=m}^{m+2} u_k^2)$ in which each protein couples to a block of $3u$'s. There is in this case a very weak modulation of the exponential decay. Summarizing, the asymptotic behavior for generic quadratic interactions is

$$\Delta\Delta F^S_{\rm int} \sim f(m) e^{-2m/\xi_E} \tag{14}$$

with a nonuniversal prefactor $f(m) \le 0$, which depends on details of the protein-DNA bindings (unlike the universal behavior of the enthalpic case).

Mixed allostery.—Finally, we consider the mixed case

$$H^M = H_0 - h u_0 + \varepsilon u_m^2 \tag{15}$$

for which we find [17]

$$\Delta\Delta F^M_{\rm int} = \frac{\varepsilon h^2 S_m^2}{1 + 2\varepsilon S_0} \tag{16}$$

which is positive and thus a destabilizing interaction term. It is temperature independent and thus of enthalpic nature. The term $-hu_0$ produces a $\langle u_n \rangle \neq 0$, which contributes to the enthalpic part, but we find no entropy change in this model. As $\Delta\Delta F^M_{\rm int}$ depends on S_m^2 the asymptotic is very similar to the entropic case, with decay length $\xi_M = \frac{1}{2}\xi_E$ and oscillations proportional to $\cos^2(m\phi + \phi_0)$. As in the entropic case, interactions to more than one site lead to a decay of the type (14), with $f(m) \ge 0$. We note that (11) and (13) [but not (16)] were also derived in a study of interactions of point defects in fluctuating membranes [32]. Their applicability is general and not limited to a one-dimensional chain.

Experiments.—In principle, one could distinguish the three scenarios in experiments from the sign of $\Delta\Delta F_{\rm int}$ (Fig. 2). Kim *et al.* analyzed the binding of several different pairs of proteins on DNA [3]. The binding free energy showed a decaying oscillating behavior with alternating sign which is consistent with an enthalpic allostery (10), in agreement with the analysis performed by other authors [3,5,6,33]. A fit to (10) with the asymtpotic expression for S_m (9) is shown in Fig. 3(a). A different system was analyzed by Rosenblum *et al.* [4] who found DNA-mediated allostery in the binding of bacterial transcription factors ComK. The experiments showed a cooperative binding ($\Delta\Delta F_{\rm int} < 0$) for varying spacer lengths, see Fig. 3(b). The negative $\Delta\Delta F_{\rm int}$ indicates an allostery of (predominantly) entropic type. The data are fitted (dashed line) against a model containing both linear and quadratic

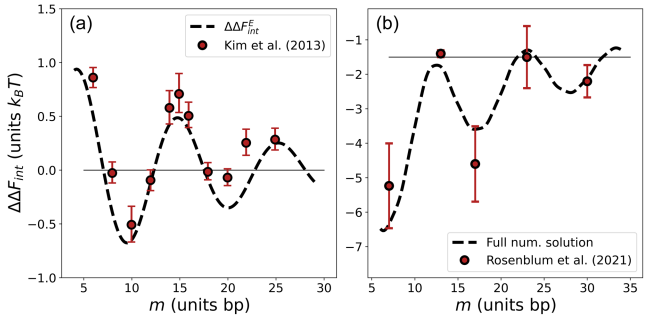

FIG. 3. (a) Symbols: experimental data of interaction free energies for BamHI-GRDBD [3]. Oscillating sign in $\Delta\Delta F_{\rm int}$ indicates an enthalpic type of allostery. Dashed line: fit to the asymtptotic expression (9) ($\phi = 2\pi/10.5$, $\xi_E = 15$ bp). (b) Symbols: experimental data for $\Delta\Delta F_{\rm int}$ for the ComK system [4]. The negative sign indicates an allosteric interaction which is of predominant entropic. Dashed line: full numerical solution of $\Delta\Delta F_{\rm int}(m)$ ($\phi = 2\pi/10.5$, $\xi_E = 22$ bp) for a global allostery model with extended perturbations [17]. As $\Delta\Delta F_{\rm int}$ does not seem to converge to zero for large m, we have added an asymptotic nonzero offset (dotted line).

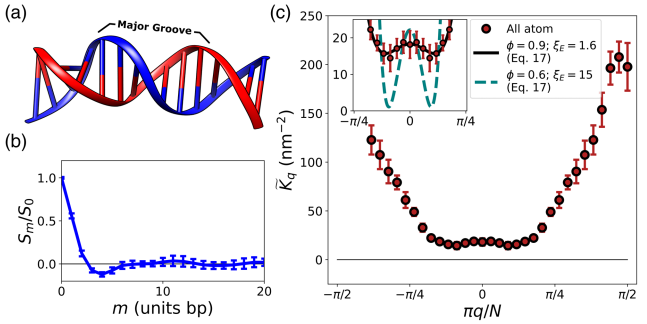

FIG. 4. (a) Schematic representation of the major groove width of DNA, for which the Curves+ [35] definition is used. (b) Normalized correlation function of the major groove width as obtained from all-atom simulations obtained using the Curves+ software [35]. (c) Plot of the momentum-space stiffness for the major groove width obtained from all-atom data. The inset coplots $\tilde{K}_q$ for the minimal model [Eq. (17)] for two sets of parameters. The solid line is a direct fit of the stiffness data. The dashed line uses ϕ and ξ_E from a fit of the experimental $\Delta\Delta F_{\rm int}$ data. The error bars in (b) and (c) indicate the standard deviation calculated over 21 time windows of 10 ns.

terms, using a Monte Carlo fitting procedure [17]. These fields act on several sites reflecting the extended contact regions of the ComK-DNA interaction. The oscillating component is due to the enthalpic part, which, as seen above, has a universal oscillatory decaying behavior. In the fit the same value of $\phi = 2\pi/10.5$ as Fig. 3(a) was used. Instead for the correlation length we used $\xi_E = 22$ bp, larger than the value used in (a), possibly indicating that allosteric coupling is carried over by different collective variables in the two cases. We note that the ComK have much larger $|\Delta\Delta F_{\rm int}|$ than the data shown in Fig. 3(a). It is possible that to quantitatively describe the ComK data one would need to go beyond linear elasticity (anharmonic effects) or to more complex multimodal models [34]. However, the harmonic model with linear and quadratic terms spanning several sites generates a $\Delta\Delta F_{\rm int}$ consistent with experimental data.

Simulations.—So far we have assumed a model with distal couplings generating allosteric interactions, using a generic u_n. In principle u_n could be one of the several DNA local deformation modes used in the rigid base model [9], or a combination thereof. However, experiments indicate that allosteric interactions decay via damped oscillations, which puts some constraints on u_n, as several variables do not have this property. For instance, we can exclude pure bending or twist modes as candidates for u_n, as the stiffness $\tilde{K}_q$ for these coordinates does not produce oscillations with the desired periodicity [17]. Prior work suggested the DNA major groove width as mediating allosteric interactions [3]. However, extensive 1 μs all-atom simulations of a 33-bp sequence showed no signature of periodicity in the groove width correlations [33]. We have performed simulations using two different 44-bp sequences for 100 ns, with major groove width calculated from the algorithm Curves+ [35] using the setup discussed in [13]. Our results for the normalized propagator $S_m/S_0 = \langle u_0 u_m\rangle_0/\langle u_0^2\rangle_0$ (with u_n the deviation of the major groove width from the equilibrium value) are given in Fig. 4(b) and are in close agreement with those reported in [33]. To extract parameters from simulations we have calculated the q stiffness $\tilde{K}_q$ from the equipartition relation (7), as recently done for twist and bending deformations [13,36]. A minimal model with just four poles $\pm q_E$, $\pm q_E^*$ (8) gives

$$\tilde{K}_q = A(q^2 - q_E^2)(q^2 - q_E^{*2}) = A(q^4 - \mu q^2 + \lambda^2) \quad (17)$$

with A a scale factor, $\mu = q_E^2 + q_E^{*2}$, and $\lambda = |q_E|^2$. We note that (17) is not of the form (4), but should be interpreted as the continuum long wavelength limit ($q \to 0$) of model (2). Figure 4(c) shows $\tilde{K}_q$ as obtained from simulation data averaged over two sequences (red circles). The double-well shaped curve is fitted, for small q, to (17) giving $\phi = 0.9$ and $\xi_E = 1.6$ (solid line in the inset of Fig. 4). The former parameter is close to the double helix periodicity $\phi = 2\pi/10.5 \approx 0.6$, while the latter appears to be quite small as compared to experimental data which predict $\xi_E \approx 15$ bp. We show on the same inset a plot of Eq. (17) with the latter values for ϕ and ξ_E. We conclude that simulations of the major-groove width qualitatively support the distal couplings model (2), but a quantitative matching remains an open challenge, as pointed out earlier [7,33]. See [17] for an extended discussion on possible origins of these discrepancies.

Conclusions.—We have studied a coarse-grained model which predicts three types of allosteric DNA-mediated interactions. One could distinguish between the three cases (or about the dominance of one of these) from the sign of the interaction free energy and on its dependence on the length of the DNA linker sequence separating the two protein-binding sites. Prior work [3,5,6] pointed to some examples of enthalpic DNA-mediated allostery characterized by a free energy of oscillating sign. We have argued here that recent ComK data [4] show an allostery which is of predominant entropic nature, as $\Delta\Delta F_{\rm int} < 0$. Entropic allostery (often referred as dynamic allostery) was discussed in the case of proteins [37,38], but it should manifest itself in DNA as well. The model introduced here predicts additionally a "mixed" allostery, obtained when coupling two different proteins in which one exerts a linear field and the other a quadratic one. This mixed allostery is of enthalpic nature and we are not aware that such interaction was discussed in the protein literature. Differently from the protein case, in DNA-mediated allostery one can vary the spacer sequence length, probing the decay of $\Delta\Delta F_{\rm int}$, therefore the type of allostery (enthalpic, entropic, or mixed) should be easier to identify. By varying the binding sites sequences one can bring in close vicinity proteins of different types and which couple differently to the DNA (e.g., predominantly via linear or quadratic fields), thereby probing the three scenarios predicted by the model.

Discussions with G. Bussi, H. Hofmann and G. Rosenblum are gratefully acknowledged. M. S. acknowledges financial support from Fonds Wetenschappelijk Onderzoek (FWO 11O2323N). E. C. thanks the Aoyama-Gakuin University (Japan) and Laboratoire C. Coulomb of the University of Montpellier (France), where part of this work was done, for the kind hospitality. He acknowledges financial support from the Japan Society for the Promotion of Science (FY2020 JSPS).

[1] K. Gunasekaran, B. Ma, and R. Nussinov, Proteins **57**, 433 (2004).
[2] S. J. Wodak *et al.*, Structure **27**, 566 (2019).
[3] S. Kim *et al.*, Science **339**, 816 (2013).
[4] G. Rosenblum *et al.*, Nat. Commun. **12**, 2967 (2021).
[5] T. Držata, M. Zgarbová, N. Špačková, P. Jurečka, J. Šponer, and F. Lankaš, J. Phys. Chem. Lett. **5**, 3831 (2014).
[6] J. Singh and P. K. Purohit, J. Phys. Chem. B **123**, 21 (2018).
[7] A. Balaceanu, A. Pérez, P. D. Dans, and M. Orozco, Nucleic Acids Res. **46**, 7554 (2018).
[8] J. Rudnick and R. Bruinsma, Biophys. J. **76**, 1725 (1999).
[9] W. K. Olson *et al.*, J. Mol. Biol. **313**, 229 (2001).
[10] F. Lankaš, J. Šponer, P. Hobza, and J. Langowski, J. Mol. Biol. **299**, 695 (2000).
[11] A. Noy and R. Golestanian, Phys. Rev. Lett. **109**, 228101 (2012).
[12] B. Eslami-Mossallam and M. Ejtehadi, J. Chem. Phys. **134**, 125106 (2011).
[13] E. Skoruppa, A. Voorspoels, J. Vreede, and E. Carlon, Phys. Rev. E **103**, 042408 (2021).
[14] Y. A. Gutierrez Fosado, F. Landuzzi, and T. Sakaue, Phys. Rev. Lett. **130**, 058402 (2023).
[15] Models as (2) were used in the literature to describe large scale packing of chromosomal DNA, see A. Amitai and D. Holcman, Phys. Rev. E **88**, 052604 (2013) and K. Polovnikov, S. Nechaev, and M. Tamm, Soft Matter **14**, 6561 (2018). Some of these models considered a slowly algebraic decay of the off-site couplings $K_p \sim p^{-a}$ (and $L \to \infty$), which gives a Fourier space stiffness $\tilde{K}_q$ nonanalytic in $q = 0$. We consider here couplings extending to a finite range L, hence such nonanalytic behavior will not occur.
[16] The linear term is dominant when $|l_0 - s_0|^2 \gg \langle u_0^2 \rangle_0$.
[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.238402, which contains references [18–29], for details of mathematical derivations, simulations, and additional information about the models used.
[18] F. Lankaš, O. Gonzalez, L. M. Heffler, G. Stoll, M. Moakher, and J. H. Maddocks, Phys. Chem. Chem. Phys. **11**, 10565 (2009).
[19] A. Voorspoels, J. Vreede, and E. Carlon, J. Chem. Theory Comput. **19**, 902 (2023).
[20] D. Petkevičiūtė, M. Pasi, O. Gonzalez, and J. Maddocks, Nucleic Acids Res. **42**, e153 (2014).
[21] M. Abrahams, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, SoftwareX **1–2**, 19 (2015).
[22] A. Perez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, Biophys. J. **92**, 3817 (2007).
[23] I. Ivani *et al.*, Nat. Methods **13**, 55 (2016).
[24] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[25] S. Li, W. Olson, and X.-J. Lu, Nucleic Acids Res. **47**, W26 (2019).
[26] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[27] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[28] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).
[29] K. Liebl and M. Zacharias, Biophys. J. **122**, 2841 (2023).
[30] The coarse-grained variables u_n derive from an underlying partial integration of microscopic degrees of freedom. This may lead to a temperature dependence of the couplings K_n. Here enthalpic and entropic interactions refer to this coarse-grained level of description. Our focus is on the sign of $\Delta\Delta F_{\rm int}$ rather than its temperature dependence.
[31] T. Witten and P. Pincus, *Structured Fluids: Polymers, Colloids, Surfactants* (OUP, Oxford, 2010).
[32] R. R. Netz, J. Phys. I (France) **7**, 833 (1997).
[33] T. Držata, M. Zgarbová, P. Jurečka, J. Šponer, and F. Lankaš, Biophys. J. **110**, 874 (2016).

[34] K. López-Güell, F. Battistini, and M. Orozco, Nucleic Acids Res. **51**, 2633 (2023).
[35] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska, Nucleic Acids Res. **37**, 5917 (2009).
[36] M. Segers, A. Voorspoels, T. Sakaue, and E. Carlon, J. Chem. Phys. **156**, 234105 (2022).
[37] R. J. Hawkins and T. C. B. McLeish, Phys. Rev. Lett. **93**, 098104 (2004).
[38] N. V. Dokholyan, Chem. Rev. **116**, 6463 (2016).