

Theoretical Framework for Calculating Sequence-Dependent Physical Properties of Nucleosomes

Enrico Skoruppa ^{1,2,*}, Manish Yadav ^{2,*}, Thor van Heesch ³, Jocelyne Vreede ³, and Helmut Schiessel ^{2,4}

¹*School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom*

²*Cluster of Excellence Physics of Life, TU Dresden, Arnoldstraße 18, 01307 Dresden, Germany*

³*Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

⁴*Institut für Theoretische Physik, TU Dresden, 01062 Dresden, Germany*



(Received 14 November 2025; accepted 20 April 2026; published 21 May 2026)

The positioning of nucleosomes, the most abundant nucleoprotein complex in eukaryotes, and their physical properties are strongly influenced by sequence-dependent geometric and elastic properties of the wrapped DNA. At present, the theoretical study of this system is largely limited to numerical computation, and access to entropic contributions has been challenging. In particular, the toolset for the systematic introduction of binding-site-mediated constraints that retains analytical tractability—beyond uniform superhelical wrapping—is currently lacking. Here, we present a mathematical framework for introducing such constraints in the context of the rigid-base-pair model that permits local DNA relaxation and yields closed-form expressions for elastic binding free energies. Importantly, the methodology enables the evaluation of individual contributions to the free energy, including entropy. Benchmarking against numerical solutions obtained via Monte Carlo sampling and validation with experimental data ranging from competitive nucleosome reconstitution to spontaneous nucleosome breathing and force-induced unwrapping demonstrates quantitative fidelity. The approach enables genome-wide evaluation of nucleosome binding affinities and can be readily extended to noncanonical and epigenetically modified DNA or histone cores. While the present study is based on the rigid-base-pair model, the same ideas extend to any rigid-body-based model, opening a route to a broad class of problems in biomolecular modeling.

DOI: [10.1103/rh7r-ysrf](https://doi.org/10.1103/rh7r-ysrf)

I. INTRODUCTION

Eukaryotic DNA is highly compacted within the nucleus of each cell. In human cells, for example, 46 DNA molecules, totaling roughly 2 m in length, are packed inside a micron-sized nucleus. This compaction is achieved through a hierarchical structure, with nucleosomes forming its lowest level. Given their central role in chromosomal organization, nucleosomes dictate a wide range of biological processes, including gene regulation, recombination, and replication. A nucleosome consists of 147 base pairs (bp) of DNA (about one persistence length) wrapped around a cylindrical protein complex composed of eight histone proteins [1,2]. Histone-DNA interactions primarily involve electrostatic interactions between negatively charged DNA phosphate groups and positively charged amino acid side chains at the surface of the octamer. There are 14 distinct binding locations where the minor groove of the DNA double helix faces the histone octamer. Together, these nearly equally spaced contacts wrap the nucleosomal DNA onto a superhelical path spanning roughly $1\frac{3}{4}$ turns.

Various experiments have shown that nucleosomes are highly dynamic structures. They temporarily expose portions of their wrapped DNA through spontaneous unspooling from either end via a process called site exposure or nucleosome breathing [3], they diffuse along DNA in a process called nucleosome sliding [4,5] (via single-bp twist defects [6–12] and 10-bp bulges [9,10,12–14]), and they show slow spontaneous gaping, which refers to the transient separation of adjacent superhelical turns of DNA around the histone core [3,15,16]. To understand some of these experimental observations, early theoretical studies modeled the nucleosomal DNA as a homogeneous polymer, typically a perfectly flexible or a wormlike chain, and the histone octamer as a sphere or a cylinder. These models were built to explain various phenomena, e.g., the overcharging of the nucleosome complex by the negatively charged DNA [17–23] or the surprising resistance of nucleosomes against external tension [24–27].

However, more recent experiments have demonstrated that nucleosome behavior is strongly governed by the sequence-dependent mechanical properties of the wrapped DNA. This is particularly evident from observations showing different behavior for the two DNA entry and exit sites. For example, nucleosomes have been shown to breathe asymmetrically [28,29], and external force-induced DNA dissociation was demonstrated to favor unwrapping from one end [30]. Given the nucleosome's twofold symmetry about the dyad axis, this asymmetric behavior may seem surprising at first. However, the involved DNA sequences are nonpalindromic, which, together with the sequence-dependent DNA elasticity and

*These authors contributed equally to this work.

†Contact author: esk.phys@gmail.com

shape, explains why these complexes exhibit asymmetric physical properties. It has been shown that such asymmetric nucleosomes present “polar barriers” for RNA polymerases, making transcription in one direction more efficient than in the other [31,32]. In short, sequence matters for many physical properties of nucleosomes [33].

In the meantime, the sequence-dependent properties of nucleosomes have been studied extensively with a variety of computational and theoretical approaches. At the highest resolution, atomistic simulations have been instrumental in unravelling a multitude of phenomena [34–44], such as the effect of histone tails and their epigenetic modifications on nucleosome dynamics and stability [45–54]; see also reviews in Refs. [55,56]. While offering high resolution, this approach is limited to simulation times of several microseconds—much shorter than the timescales relevant to many important chromatin phenomena. Larger timescales can be reached with coarse-grained nucleosome models, and many such models accounting for sequence-dependent DNA properties have been developed. The most frequently used models fall into three categories: those based on the rigid-base-pair model (RBP) [57,58], including Refs. [8,59–71], on the three-site-per-nucleotide model (3SPN) [72], including Refs. [9–12,73,74] and those based on cgDNA+ [75–78]. Importantly, these models are often sufficient to obtain qualitative and even quantitative insights into the role of bp sequence in nucleosome organization.

For analytical modeling and genome-wide analysis, the RBP formalism offers a good balance between structural detail and mathematical complexity. In this framework, each bp is treated as a rigid body, with an associated orientation and position. The intrinsic structure of a DNA molecule is determined by the sequence-specific ground state of relative rotations and translations between adjacent bps. Deformations away from this ground state are governed by sequence-dependent elastic energies. Together, these two properties provide a natural link between sequence and mechanical response. By embedding nucleosomal DNA conformations within the RBP framework, elastic contributions to nucleosome stability can be systematically quantified [64,65,67,68,71].

Some approaches based on the RBP “thread” DNA onto a rigid scaffold—either one taken directly from crystallographic data [59,60] or an idealized superhelix [60,62,68,79,80]. While such frameworks yield qualitative insights into, for example, the bp step preferences of nucleosomes [68], they largely overestimate elastic energies as they do not permit local relaxation of the DNA between the binding sites. Furthermore, such approaches are unable to account for entropic effects.

In contrast, allowing for local relaxation was shown to improve the agreement with a wide range of experiments. Ref. [64] introduced a coarse-grained nucleosome model in which histone-DNA interaction-mediated confinement is induced on the RBP through rigid constraints at the 28 DNA-histone contact points. In Ref. [64], using Monte Carlo simulations, it was shown that the model predicts the bp step preferences of nucleosomes [81,82] and the binding free energies in competitive reconstitution experiments [82–84]. Next, in Ref. [65], the model successfully predicted the

bp sequence-dependent response to external tension [30]. In Ref. [67], this framework was used to interpret experiments measuring the bp sequence-dependent accessibility of nucleosomes to restriction enzymes [28,85]. Moreover, the model was used to predict the amount of spontaneous DNA unwrapping in an ensemble of nucleosomes, as observed through small-angle x-ray scattering experiments [29]. These studies all relied on numerical computations and focused solely on enthalpic effects, while neglecting entropy. In addition, a Markov Chain Monte Carlo approach allowing for the determination of free energies was introduced [86]. However, this method requires a new set of extensive simulations with each change of the model (e.g., when going to a different set of DNA parameters), and it does not permit comparison of free energies for nucleosomes in different unwrapping states.

In the current work, we introduce an analytical framework for solving the RBP-based nucleosome model that allows for local relaxation [64] and the calculation of the full free energy, i.e., including entropy. The general idea of this procedure is to enforce the geometrical constraints imposed on the DNA by the DNA-histone contacts and then evaluate the free energy of the resulting restrained system by integrating over all accessible configurations according to the canonical measure. These constraints impose structural deformations—most prominently the required superhelical bending—which incur an enthalpic penalty, while simultaneously restricting the accessible phase space and thereby modifying the configurational entropy. Both effects are manifestly sequence-dependent and determine how the structural and elastic properties of the DNA influence nucleosome positioning. These elastic costs are unfavorable for nucleosome association and must be balanced by stabilizing components such as electrostatic contacts, counterion release, and hydration effects. In our description, these effects are collectively accounted for by site-specific—yet sequence-independent—adsorption energies calibrated from experimental data. Importantly, this approach assumes a separation of timescales, particularly when nucleosome breathing is considered; specifically, each wrapping state must be sufficiently long-lived to explore the accessible phase space. We argue that these conditions are met, as spontaneous unwrapping typically occurs at timescales of seconds [87], while configurations within a given bound state decorrelate on sub-microsecond timescales.

The inherent difficulty of analytically applying the nucleosome-imposed constraints on the wrapped DNA is overcome by leveraging the methodology developed in Ref. [88] to construct junctions spanning multiple bp steps. These junctions correspond to the effective dynamical variables on which the constraints are imposed. Reexpressing the system to include these degrees of freedom, therefore, allows for the convenient splitting of restrained and unrestrained degrees of freedom. As a result, the contributions from elastic energy and entropy can be expressed fully analytically.

To account for the free-energy difference between different wrapped states, we expand the model by incorporating the influence of histone-DNA contacts on DNA fluctuations. This extended model is parametrized based on molecular dynamics simulations and experimental nucleosome unzipping data [89].

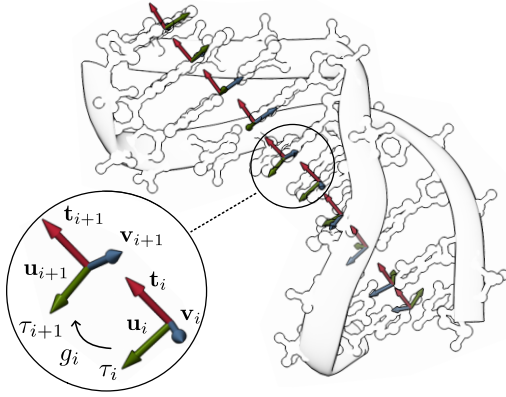


FIG. 1. Illustration of the RBP for a small 10-bp oligomer. Each base pair is represented by a right-handed triad that captures the orientation of the molecule and its local geometry. Deformations and, therefore, the energetics of the model are expressed in terms of the junctions spanning between adjacent base pairs.

We verify our analytical calculation through Monte Carlo simulations of the same RBP model. Specifically, we calculate the free-energy, enthalpy, and entropy landscapes along a stretch of the yeast genome and find excellent agreement between theory and simulations. Finally, we apply our analytical model to interpret data from three different types of nucleosome experiments, namely competitive nucleosome reconstitution, nucleosome breathing, and force-induced nucleosome unwrapping.

In summary, the main results of this work can be presented as follows: Our analytical nucleosome model is based on a coarse-grained DNA model, the RBP model Fig. 1, which is forced into a superhelix by histone-induced constraints Fig. 2. By introducing junction variables Fig. 2(a), we arrive at an analytical expression for the free energy [Eq. (40)] that shows excellent agreement with the results of Monte Carlo simulations Fig. 4. Furthermore, we improve this approach by incorporating soft constraints that account for the flexibility of the protein core and are parametrized by all-atom MD simulations Fig. 5. Finally, we compare our model with data from three different experiments: competitive nucleosome reconstitution Fig. 7, nucleosome breathing Fig. 8, and force-induced nucleosome unwrapping Fig. 9. The comparisons show quantitative agreement and suggest that these experiments can be largely interpreted on the basis of the bp sequence-dependent elasticity and geometry of the DNA double helix.

II. THEORY

A. Rigid-base-pair framework

Throughout this work, we employ the RBP model [58] for DNA, which represents bps as individual rigid bodies; see Fig. 1. These are fully captured by a right-handed reference frame (triad) $\mathcal{T} \in \text{SO}(3)$ and a position vector $\mathbf{r} \in \mathbb{R}^3$. Both components are conveniently expressed within a unified framework as elements of the special Euclidean group $\text{SE}(3)$, which in explicit representation can be written as 4×4 matrices,

$$\tau = \begin{bmatrix} \mathcal{T} & \mathbf{r} \\ \mathbf{0}^\top & 1 \end{bmatrix} \in \text{SE}(3). \quad (1)$$

To map atomistic bp coordinates onto rigid bodies, we employ the Curves+ definition for bp triad vectors [90], which reflect the local geometry of the double helix, such as the orientation of the quasiparallel bp, and the location of the backbone and the grooves. Note, however, that we define the three base vectors as the columns of the respective triad, $\mathcal{T} = [\hat{\mathbf{u}} \ \hat{\mathbf{v}} \ \hat{\mathbf{t}}]$ (as opposed to the definition given in the original Curves+ publication [90]). A configuration of an $N + 1$ bp molecule is thus captured by the set of triads $\{\tau_0, \tau_1, \dots, \tau_N\}$.

In this representation, a given DNA molecule is characterized by a sequence-dependent ground state structure and an elastic energy penalizing deformations away from this state. Both of these characteristics are expressed in terms of the junctions connecting adjacent bp's

$$g_i \equiv \tau_i^{-1} \tau_{i+1} = \begin{bmatrix} \mathcal{T}_i^\top \mathcal{T}_{i+1} & \mathcal{T}_i^\top (\mathbf{r}_{i+1} - \mathbf{r}_i) \\ \mathbf{0}^\top & 1 \end{bmatrix} = \begin{bmatrix} \mathcal{R}_i & \mathbf{w}_i \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (2)$$

which are likewise elements of $\text{SE}(3)$. Note that g_i is defined as a right-action transformation $\tau_{i+1} = \tau_i g_i$, i.e., it is a body-fixed rather than space-fixed transformation. As such, it reflects the orientation of the triads τ_{i+1} relative to the frame of τ_i and is invariant under global Euclidean transformations of the chain as a whole.

These junctions and, more broadly, elements of $\text{SE}(3)$ can be parameterized by six independent parameters: three rotational components that represent elements of $\text{SO}(3)$, and the three components of the translation vector $\mathbf{w}_i$. In the context of DNA, the three rotational components are commonly referred to as tilt, roll, and twist, while the three translational components are called shift, slide, and rise.

While the rotation matrices, $\mathcal{R}_i$, may be parametrized in various ways, we adopt the rotation vector parametrization. Each rotation matrix encodes a rotation of a certain angle around a certain axis. Both of these features are captured by vectors $\boldsymbol{\Omega}^\top = [\Omega_u \ \Omega_v \ \Omega_t]$ —also known as Euler vectors—that point along the rotation axis and have length equal to the rotation angle. Further details about the relationship between rotation matrices and Euler vectors and useful group properties of $\text{SO}(3)$ and the associated Lie algebra $\mathfrak{so}(3)$ are provided in Appendix A. As shorthand notation, we write the six junction components as a single six-vector

$$\mathbf{X}_i = \begin{bmatrix} \boldsymbol{\Omega}_i \\ \mathbf{w}_i \end{bmatrix} \equiv \mathcal{P}(g_i), \quad (3)$$

where $\mathcal{P}$ is a mapping between elements of $\text{SE}(3)$ and corresponding coordinates.

For the development of composite junctions, necessary to impose histone wrapping constraints on the DNA, we will employ the formalism developed in Ref. [88], which requires bp step junctions g_i to be split into static (s_i) and dynamic (d_i) components

$$g_i = s_i d_i. \quad (4)$$

Here, the static component s_i encodes the sequence-dependent equilibrium geometry, while the dynamic component d_i captures thermal deformations away from this state. This separation enables the expansion of dynamic components while

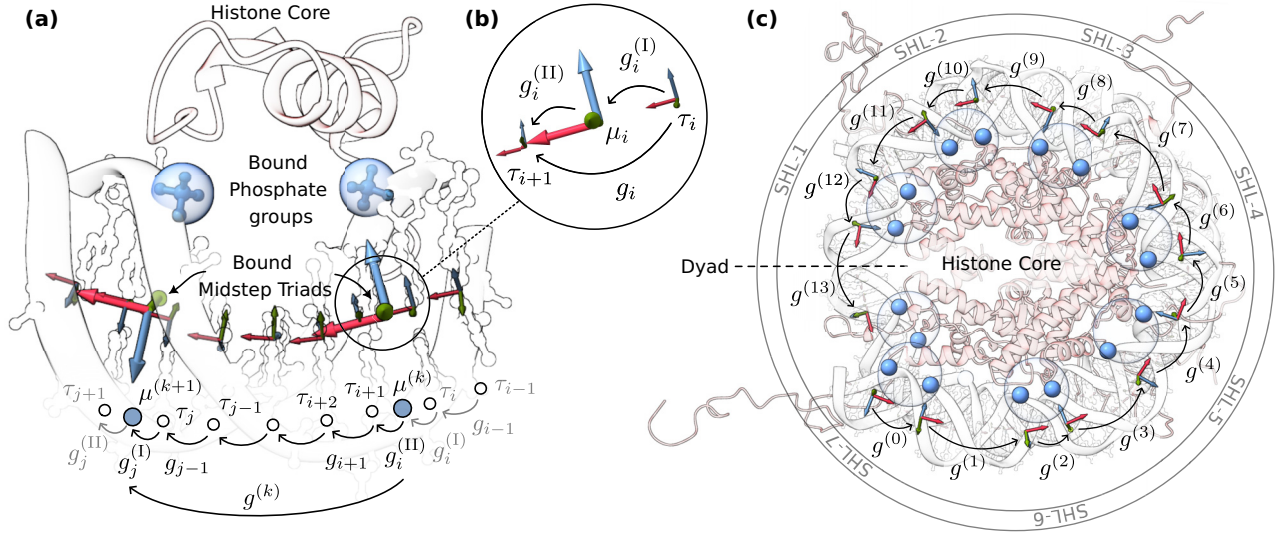


FIG. 2. Illustration of the mathematical description of the nucleosome in the context of the RBP. (a) A segment of nucleosomal DNA wrapped around the histone core, with two bound phosphate groups (blue spheres). Large reference frames denote corresponding midstep triads $\mu^{(k)}$. Junction connecting these triads $g^{(k)}$ spans across multiple bp steps g_i and comprises two half-step junctions (b), defined by $g_i = g_i^{(I)} g_i^{(II)}$ and $\mu_i = \tau_i g_i^{(I)}$ (see Appendix B 2 for full details). (c) Complete nucleosome core particle (PDB 1KX5 [2]) highlighting bound phosphate groups (blue spheres), associated midstep triads, and junctions $g^{(k)}$ spanning between them. Each superhelical location (SHL) contains two binding sites, here grouped by circles.

the typically larger static contributions can be retained without approximation.

The set of static components,

$$s_i = \begin{bmatrix} \mathcal{S}_i & \mathbf{s}_i \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (5)$$

parametrized by the six-vectors

$$\mathbf{Y}_{0,i} = \begin{bmatrix} \Phi_{0,i} \\ \mathbf{s}_i \end{bmatrix} \equiv \mathcal{P}(s_i), \quad (6)$$

determines the ground state of a given molecule and remains constant. Deformations away from that ground state are captured by the dynamic components,

$$d_i = s_i^{-1} g_i = \begin{bmatrix} \mathcal{D}_i & \mathbf{d}_i \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (7)$$

with corresponding coordinates

$$\mathbf{Y}_{\Delta,i} = \begin{bmatrix} \Phi_{\Delta,i} \\ \mathbf{d}_i \end{bmatrix} \equiv \mathcal{P}(d_i). \quad (8)$$

Following the common small fluctuation approach, we assume the elastic energy to be quadratic in these deformations. This elastic energy is conveniently expressed by collecting all dynamic junction elements into a single $6N$ -dimensional vector

$$\tilde{\mathbf{Y}}_{\Delta} \equiv \text{stack}\{\mathbf{Y}_{\Delta,0}, \dots, \mathbf{Y}_{\Delta,N-1}\} \equiv \begin{bmatrix} \mathbf{Y}_{\Delta,0} \\ \mathbf{Y}_{\Delta,1} \\ \vdots \\ \mathbf{Y}_{\Delta,N-1} \end{bmatrix}. \quad (9)$$

The most general quadratic elastic energy then takes the form

$$\beta E_{\text{free}} = \frac{1}{2} \tilde{\mathbf{Y}}_{\Delta}^\top \mathbf{M} \tilde{\mathbf{Y}}_{\Delta}, \quad (10)$$

where $\beta = 1/k_B T$ is the inverse temperature. Depending on the chosen parametrization the stiffness matrix $\mathbf{M} \in \mathbb{R}^{6N \times 6N}$ may be block diagonal [57,91] or may contain interactions spanning beyond nearest neighbors [88,92,93]. Note that the framework of splitting static and dynamic components at the rotation group (matrix) level deviates from usual conventions. However, this choice proves more convenient for calculating nucleosome free energies. For the conversion of more conventional frameworks to our framework and, in particular, the transformation of the stiffness matrix and static component vectors, we refer to Ref. [88]. Sequence-dependent stiffness matrices and ground states can be derived from different sources including crystallographic data [57] and atomistic molecular dynamics simulations [76,91,94].

B. Nucleosome model

The wrapping of nucleosomal DNA around the histone octamer is stabilized by DNA-histone contacts, which are primarily mediated by electrostatic interactions between positively charged amino acid residues and the phosphate groups along the DNA backbones. In particular, conserved arginine side chains form short, hydrogen-bonded salt bridges with the phosphate oxygens, providing specific, directional anchoring that both neutralizes and bends the DNA around the histone core [1,2].

There are 14 distinct binding sites—usually referred to as superhelical locations [SHLs; see Fig. 2(c)]—involving two phosphate groups each; one per backbone displaced by a few bp steps. Altogether, this yields 28 DNA-histone contact points. In Ref. [64], these 28 strongly bound phosphates were identified based on local minima in the crystallographic B-factor of the NCP147 structure [2].

To impose these phosphate-mediated constraints within the RBP description, we follow the approach first introduced by Eslami-Mossallam *et al.* [64]. Binding is assumed to constrain the midstep triad between the two bp's associated with the bound phosphate group. Midstep triads are defined as the half-translation and half-rotation between a pair of bp triads [64,90], τ_i and τ_{i+1} ,

$$\mu_i \equiv \begin{bmatrix} \tau_i \sqrt{\mathcal{R}_i} & \frac{1}{2}(\mathbf{r}_i + \mathbf{r}_{i+1}) \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (11)$$

with $\sqrt{\mathcal{R}}$ the rotation matrix defining rotation over the same axis as $\mathcal{R}$ but over half the angle; see Eq. (A9). The assumption of a fixed midstep triad in Ref. [64] was based on the observation in several nucleosome crystal structures that the pair of DNA phosphates between bp steps, whether bound to the octamer or not, remains largely stationary with respect to the midstep triad. This allows the RBP model to implicitly capture the effects of bound phosphates, even though they are not explicitly represented.

At first glance, constraining bp triads rather than midstep triads might seem the more natural choice. However, since backbone phosphates are located between adjacent bp planes, there is no single bp triad that unambiguously represents the constraint. The only symmetric alternative would be to constrain both flanking bp triads simultaneously, but this would be considerably more restrictive than our midstep constraint. It would entirely freeze the local bp step at the contact point, eliminating any possibility of relaxation at that junction. For a more detailed justification, we refer to the Supporting Information (Text S1 and Fig. S2) of Ref. [64].

In the canonical nucleosome structure—composed of 147 bp of DNA—the constrained midstep triads lie midway between the triads τ_{i_k} and τ_{i_k+1} with [64]

$$i_k \in \{2, 6, 14, 17, 24, 29, 34, 38, 45, 49, 55, 59, 65, 69, 76, 80, 86, 90, 96, 100, 107, 111, 116, 121, 128, 131, 139, 143\}. \quad (12)$$

When considering partially wrapped states, open binding sites should be excluded from this set. For convenience, we introduce the shorthand notation

$$\mu^{(k)} \equiv \mu_{i_k}. \quad (13)$$

At first, we will follow the approach of Refs. [64,67,71,95] and assume bound midstep triads to remain fixed,

$$\mu^{(k)} = \mu_0^{(k)}, \quad (14)$$

which we refer to as the hard constraint. Subsequently, we shall demonstrate that imposing these constraints precludes the computation of relative free energies between distinct wrapping states. This limitation will be alleviated by imposing a soft constraint that allows midstep triads to fluctuate around their preferred state $\mu_0^{(k)}$.

Both types of constraints act on the relative positions and orientations of neighboring bound midstep triads

$$g^{(k)} \equiv \mu^{(k)-1} \mu^{(k+1)}, \quad (15)$$

which is the composite junction connecting these midstep triads, i.e., it spans across multiple individual bp steps and

involves two half-step junctions; see Figs. 2(a) and 2(b) and Appendix B 2 for a more detailed discussion.

The general approach is to express the coordinates of these composite junctions in terms of the coordinates of the underlying bp steps and then apply a change of variables with the associated reformulation of the elastic Hamiltonian that includes the composite degrees of freedom as dynamic variables. On these new coordinates, it becomes straightforward to impose the discussed constraints.

To implement this procedure, we draw on the methodology of Ref. [88], which requires the bp step junctions to be split into static and dynamic components at the transformation level as introduced in Eq. (4). This results in composite junctions with likewise splitting

$$g^{(k)} = s^{(k)} d^{(k)}. \quad (16)$$

The ground state $s^{(k)}$ [exact definition given in Eq. (B70)] again reflects the shape of the DNA molecule in the absence of the histone-imposed constraint. For the dynamic components, we introduce the six-vector coordinates

$$\mathbf{Y}_\Delta^{(k)} = \mathcal{P}(d^{(k)}). \quad (17)$$

Thus, we aim to apply a change of variables to include the $\mathbf{Y}_\Delta^{(k)}$ as independent dynamic variables. This implies replacing one of the original bp step junction coordinates $\mathbf{Y}_{\Delta,l}$ for each such composite. The system will remain fully described as long as each replaced junction $\mathbf{Y}_{\Delta,l}$ is part of its respective composite $\mathbf{Y}_\Delta^{(k)}$.

Such a change of variables results in a set of composite junctions spanning between constrained midstep triads (indicated with the label M) and the set of remaining original junctions (label R). Consequently, for $N_M + 1$ constrained midstep triads there will be N_M composite six-vectors, which we collectively write as

$$\bar{\mathbf{Y}}_\Delta^M = \text{stack}\{\mathbf{Y}_\Delta^{(0)}, \dots, \mathbf{Y}_\Delta^{(N_M-1)}\}. \quad (18)$$

Likewise, we label the full vector of the remaining dynamic bp step coordinates as

$$\bar{\mathbf{Y}}_\Delta^R = \text{stack}\{\mathbf{Y}_{\Delta,0}, \dots, \mathbf{Y}_{\Delta,i_{k+1}-1}, \mathbf{Y}_{\Delta,i_{k+1}+1}, \dots\}, \quad (19)$$

which contains the same elements as Eq. (9) except for the replaced coordinates. Together, these vectors constitute an alternative representation of the system

$$\bar{\mathbf{E}}_\Delta = \begin{bmatrix} \bar{\mathbf{Y}}_\Delta^R \\ \bar{\mathbf{Y}}_\Delta^M \end{bmatrix}. \quad (20)$$

In Appendix B we show how the composites $\mathbf{Y}_\Delta^{(k)}$ —i.e., the components of $\bar{\mathbf{Y}}_\Delta^M$ —may be approximated as linear combinations of the original junctions, such that the change of coordinates can be expressed as a linear transformation characterized by a change of basis matrix A_Ξ , that transforms the old basis $\bar{\mathbf{Y}}_\Delta$ [Eq. (9)] into the new bases [Eq. (20)],

$$\bar{\mathbf{E}}_\Delta = A_\Xi \bar{\mathbf{Y}}_\Delta. \quad (21)$$

The Jacobian determinant associated with the change of variables is $(\det A_\Xi)^{-1}$, such that

$$d\bar{\mathbf{Y}}_\Delta = \frac{d\bar{\mathbf{E}}_\Delta}{\det A_\Xi}. \quad (22)$$

This transformation allows us to reexpress the energy Eq. (10) in terms of the new coordinates

$$\beta E = \frac{1}{2} \bar{\mathbf{E}}_{\Delta}^{\top} M_{\Xi} \bar{\mathbf{E}}_{\Delta}, \quad (23)$$

with M_{Ξ} defined by the congruence transformation,

$$M_{\Xi} = A_{\Xi}^{-\top} M A_{\Xi}^{-1}, \quad (24)$$

which preserves the positive-definite structure of the stiffness matrix.

We have deliberately constructed $\bar{\mathbf{E}}_{\Delta}$ to first list the remaining bp step coordinates and then the midstep coordinates such that the stiffness matrix M_{Ξ} may be decomposed into blocks corresponding to remaining and midstep coordinates,

$$M_{\Xi} = \begin{bmatrix} M_R & M_{RM} \\ M_{RM}^{\top} & M_M \end{bmatrix}, \quad (25)$$

where the dimensionality of the square matrices M_R and M_M matches those of the vectors $\bar{\mathbf{Y}}_{\Delta}^R$ and $\bar{\mathbf{Y}}_{\Delta}^M$, respectively.

To separate fluctuations of $\bar{\mathbf{Y}}_{\Delta}^R$ and $\bar{\mathbf{Y}}_{\Delta}^M$ the elastic energy Eq. (23) may be rewritten as

$$\beta E = \beta E_R(\bar{\mathbf{Y}}_{\Delta}^R) + \beta E_M(\bar{\mathbf{Y}}_{\Delta}^M), \quad (26)$$

with

$$\beta E_R(\bar{\mathbf{Y}}_{\Delta}^R) = \frac{1}{2} (\bar{\mathbf{Y}}_{\Delta}^R - \bar{\mathbf{Y}}^R)^{\top} M_R (\bar{\mathbf{Y}}_{\Delta}^R - \bar{\mathbf{Y}}^R) \quad (27)$$

and

$$\beta E_M(\bar{\mathbf{Y}}_{\Delta}^M) = \frac{1}{2} \bar{\mathbf{Y}}_{\Delta}^M{}^{\top} \tilde{M}_M \bar{\mathbf{Y}}_{\Delta}^M. \quad (28)$$

Here we have introduced the Schur complement [58,96,97]

$$\tilde{M}_M = M_M - M_{RM}^{\top} M_R^{-1} M_{RM} \quad (29)$$

and the $\bar{\mathbf{Y}}_{\Delta}^M$ -dependent shift

$$\bar{\mathbf{Y}}^R = -M_R^{-1} M_{RM} \bar{\mathbf{Y}}_{\Delta}^M. \quad (30)$$

Note that as long as the degrees of freedom $\bar{\mathbf{Y}}_{\Delta}^R$ remain unrestrained, which will be the case throughout this work, the contribution of βE_R to the free energy can be readily integrated and thus only contributes a constant term. The integral is a Gaussian with shifted mean, which is independent of the shift $\bar{\mathbf{Y}}^R$ (and hence independent of $\bar{\mathbf{Y}}_{\Delta}^M$). Therefore, all coupling effects between internal and collective modes are exactly absorbed into the Schur complement $\tilde{M}_M$, which represents the effective stiffness of collective modes after accounting for internal mode compliance. This relaxation of internal modes as a response to collective mode fluctuations leads to the effective softening of the latter. Integration of $\bar{\mathbf{Y}}_{\Delta}^R$ then leaves the energy in function of $\bar{\mathbf{Y}}_{\Delta}^M$, Eq. (28), which is exactly the portion of the elastic energy relevant for binding DNA to the histone core.

In this spirit, we can express the partition function as

$$\begin{aligned} Z &= \int \frac{d\bar{\mathbf{Y}}_{\Delta}^R d\bar{\mathbf{Y}}_{\Delta}^M}{\det A_{\Xi}} e^{-\beta E_R(\bar{\mathbf{Y}}_{\Delta}^R) - \beta E_M(\bar{\mathbf{Y}}_{\Delta}^M)} \\ &= \frac{(2\pi)^{3N_R}}{\det A_{\Xi} \sqrt{\det M_R}} \int d\bar{\mathbf{Y}}_{\Delta}^M e^{-\beta E_M(\bar{\mathbf{Y}}_{\Delta}^M)}, \end{aligned} \quad (31)$$

where $N_R = N - N_M$ (since $\bar{\mathbf{Y}}_{\Delta}^R \in \mathbb{R}^{6N_R}$). The remaining integral depends on the nature of the imposed constraints, which are discussed below.

C. Hard constraint

Fixing the bound midstep triads to $\mu_0^{(k)}$, relative to the nucleosome core, renders the composite junctions between them into constant transformations

$$g^{(k)} = \mu_0^{(k)-1} \mu_0^{(k+1)} = c^{(k)}, \quad (32)$$

such that corresponding dynamic coordinates assume constant values

$$\mathbf{C}_{\Delta}^{(k)} \equiv \mathcal{P}(s^{(k)-1} c^{(k)}). \quad (33)$$

As before, we collect all constant dynamic composites into the state vector

$$\bar{\mathbf{C}}_{\Delta}^M = \text{stack}\{\mathbf{C}_{\Delta}^{(0)}, \dots, \mathbf{C}_{\Delta}^{(N_M-1)}\}. \quad (34)$$

Setting $\bar{\mathbf{Y}}_{\Delta}^M$ to $\bar{\mathbf{C}}_{\Delta}^M$ in Eq. (28), the minimal-energy state under the histone constraint, hereafter referred to as the enthalpic contribution, is given by

$$\beta F_{\text{enth}}^{\text{hc}} = \frac{1}{2} \bar{\mathbf{C}}_{\Delta}^M{}^{\top} \tilde{M}_M \bar{\mathbf{C}}_{\Delta}^M. \quad (35)$$

Further, substituting $\bar{\mathbf{C}}_{\Delta}^M$ into Eq. (30) shows that the remaining degrees of freedom $\bar{\mathbf{Y}}_{\Delta}^R$ fluctuate around

$$\bar{\mathbf{C}}_{\Delta}^R = -M_R^{-1} M_{RM} \bar{\mathbf{C}}_{\Delta}^M. \quad (36)$$

In the $\bar{\mathbf{E}}_{\Delta}$ coordinates, the lowest-energy state in the bound state is then simply

$$\bar{\mathbf{C}}_{\Delta}^{\Xi} = \begin{bmatrix} \bar{\mathbf{C}}_{\Delta}^R \\ \bar{\mathbf{C}}_{\Delta}^M \end{bmatrix}, \quad (37)$$

which can be transformed back into the original bp step coordinates via

$$\bar{\mathbf{C}}_{\Delta} = A_{\Xi}^{-1} \bar{\mathbf{C}}_{\Delta}^{\Xi}. \quad (38)$$

This vector defines the compromise state structure of the nucleosome-bound DNA.

To derive the free energy, we first evaluate the partition function. Starting from Eq. (31) we find

$$\begin{aligned} Z^{\text{hc}} &= \frac{(2\pi)^{3N_R}}{\det A_{\Xi} \sqrt{\det M_R}} \int d\bar{\mathbf{Y}}_{\Delta}^M e^{-\beta E_M(\bar{\mathbf{Y}}_{\Delta}^M)} \delta(\bar{\mathbf{Y}}_{\Delta}^M - \bar{\mathbf{C}}_{\Delta}^M) \\ &= \frac{(2\pi)^{3N_R}}{\det A_{\Xi} \sqrt{\det M_R}} e^{-\beta F_{\text{enth}}^{\text{hc}}}. \end{aligned} \quad (39)$$

The free energy then follows as

$$\beta F^{\text{hc}} = -\log Z^{\text{hc}} = \beta F_{\text{fluc}}^{\text{hc}} + \beta F_{\text{enth}}^{\text{hc}}, \quad (40)$$

where the contribution stemming from fluctuations, which includes both thermal elastic energy ($\frac{1}{2}k_B T$ per degree of freedom) and entropy, is given by

$$\beta F_{\text{fluc}}^{\text{hc}} = -3N_R \log(2\pi) + \log \det A_{\Xi} + \frac{1}{2} \log \det M_R. \quad (41)$$

D. Soft constraint

Imposing the hard constraint reduces the number of fluctuating degrees of freedom in the system as the composite junctions are set to a fixed value. This leads to inconsistencies when comparing the entropy between different wrapped states, i.e., with varying numbers of constrained midstep

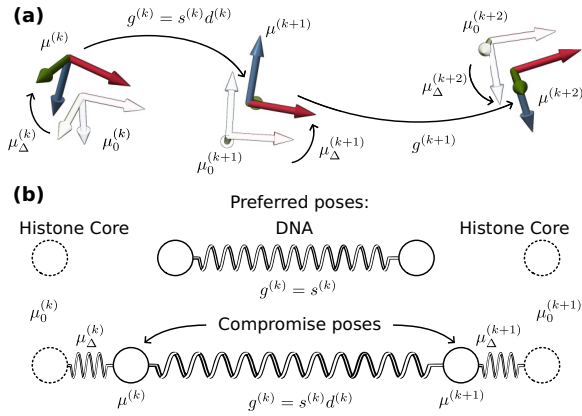


FIG. 3. Visualization of the soft-constraint nucleosome model. (a) Rather than fixing nucleosome-bound midstep triads $\mu^{(k)}$, they are assumed to have innately preferred poses $\mu_0^{(k)}$, which depend only on the histone core, but may fluctuate around this state in both position and orientation. The ability to fluctuate enables these triads to accommodate the shape of the associated DNA segments by assuming a frustrated state that minimizes the total energy, while satisfying neither the preferred pose of the DNA nor the histone core. This compromise is illustrated in panel (b).

triads. The reason for this can be understood by considering a continuous transition between the free system and a full holonomic constraint. One can, for example, subject all corresponding degrees of freedom to additional harmonic potentials with stiffness k . Letting k go to infinity will remove fluctuations, but leads inadvertently to the divergence of the entropy.

Therefore, rather than fully fixing the midstep triads $\mu^{(k)}$, we assume them to be constrained within a histone core-mediated harmonic potential, see Fig. 3. To achieve this, we split midstep triads into preferred poses $\mu_0^{(k)}$ —again expressed relative to the nucleosome core—and fluctuating poses $\mu_\Delta^{(k)}$,

$$\mu^{(k)} = \mu_0^{(k)} \mu_\Delta^{(k)}, \quad (42)$$

and assign the coordinates

$$\mathcal{P}(\mu_\Delta^{(k)}) \equiv \mathbf{Z}_\Delta^{(k)} = \begin{bmatrix} \boldsymbol{\Theta}_\Delta^{(k)} \\ \mathbf{z}_\Delta^{(k)} \end{bmatrix}. \quad (43)$$

Once again, consolidating all dynamic coordinates into a single-state vector

$$\bar{\mathbf{Z}}_\Delta = \text{stack}\{\mathbf{Z}_\Delta^{(0)}, \dots, \mathbf{Z}_\Delta^{(N_M)}\}, \quad (44)$$

the harmonic soft constraint can be expressed as

$$\beta E_Z = \frac{1}{2} \bar{\mathbf{Z}}_\Delta^\top K \bar{\mathbf{Z}}_\Delta, \quad (45)$$

where the stiffness matrix $K \in \mathbb{R}^{6(N_M+1) \times 6(N_M+1)}$ characterizes the histone core-mediated constraint on the midstep triads $\mu^{(k)}$. Note that K may capture allosteric interactions via off-diagonal couplings. We omit further specification of its structure and interaction strengths here but will return to this point in the results section.

Before calculating the free energy, we note that since

$$d^{(k)} = s^{(k)-1} (\mu_\Delta^{(k-1)} \mu_0^{(k-1)} \mu_0^{(k+1)} \mu_\Delta^{(k+1)}), \quad (46)$$

[see Eqs. (15), (16), and (42)] the coordinates $\bar{\mathbf{Y}}_\Delta^M$ of the mid-step junctions $g^{(k)}$, spanning between adjacent $\mu^{(k)}$, are fully determined by the coordinates $\bar{\mathbf{Z}}_\Delta$ of the midstep triad $\mu^{(k)}$ (provided the intrinsic poses $\mu_0^{(k)}$ are known). This suggests that $\bar{\mathbf{Y}}_\Delta^M$ may be approximated as a linear combination of $\bar{\mathbf{Z}}_\Delta$. Indeed, in Appendix D we show that

$$\bar{\mathbf{Y}}_\Delta^M = B \bar{\mathbf{Z}}_\Delta + \bar{\mathbf{b}}, \quad (47)$$

with B a linear transformation and $\bar{\mathbf{b}}$ a constant vector, both of which depend only on the static components $s^{(k)}$ and $\mu_0^{(k)}$; explicit definitions are given by Eq. (D19) and Eq. (D20), respectively. This linear transformation maps $(N_M + 1)$ -dimensional vectors into N_M -dimensional vectors and has a null space spanned by the nucleosome's overall translational and rotational degrees of freedom.

The constant offset $\bar{\mathbf{b}}$ arises whenever there is a mismatch between the shape of the composite segment $s^{(k)}$ spanning between adjacent midstep triads and the innate position of the bound midstep triads $\mu_0^{(k)}$, i.e., when the system cannot simultaneously satisfy $\bar{\mathbf{Y}}_\Delta^M = \mathbf{0}$ and $\bar{\mathbf{Z}}_\Delta = \mathbf{0}$. In that case, it assumes a compromise state that minimizes the total elastic energy, while some or all degrees of freedom remain excited—a phenomenon sometimes referred to as frustration [98].

The total energy of the bound midstep triad system, i.e., the sum of the energies corresponding to fluctuations of $\bar{\mathbf{Y}}_\Delta^M$ [Eq. (28)] and $\bar{\mathbf{Z}}_\Delta$ [Eq. (45)], is then given by

$$\begin{aligned} \beta E_{\text{comb}} &= \frac{1}{2} \bar{\mathbf{Z}}_\Delta^\top K \bar{\mathbf{Z}}_\Delta + \frac{1}{2} \bar{\mathbf{Y}}_\Delta^M \tilde{M}_M \bar{\mathbf{Y}}_\Delta^M \\ &= \frac{1}{2} (\bar{\mathbf{Z}}_\Delta - \bar{\boldsymbol{\alpha}})^\top K_{\text{comb}} (\bar{\mathbf{Z}}_\Delta - \bar{\boldsymbol{\alpha}}) + \beta F_{\text{enth}}^{\text{sc}}, \end{aligned} \quad (48)$$

where the effective stiffness

$$K_{\text{comb}} = K + B^\top \tilde{M}_M B \quad (49)$$

emerges as the combined effect of DNA-histone contacts and innate DNA stiffness. Furthermore, the enthalpy of the soft-constraint system is written as

$$\beta F_{\text{enth}}^{\text{sc}} = \frac{1}{2} \bar{\mathbf{b}}^\top (\tilde{M}_M - \tilde{M}_M B K_{\text{comb}}^{-1} B^\top \tilde{M}_M) \bar{\mathbf{b}}, \quad (50)$$

and the aforementioned compromise state for $\bar{\mathbf{Z}}_\Delta$ reads

$$\bar{\boldsymbol{\alpha}} = -K_{\text{comb}}^{-1} B^\top \tilde{M}_M \bar{\mathbf{b}}. \quad (51)$$

This implies that in the absence of fluctuations, the midstep triad $\mu^{(k)}$ will assume the compromise pose

$$\mu_\alpha^{(k)} = \mu_0^{(k)} \mathcal{P}^{-1}(\boldsymbol{\alpha}^{(k)}). \quad (52)$$

Further, inserting Eq. (51) into Eq. (47) one finds that the corresponding compromise state for $\bar{\mathbf{Y}}_\Delta^M$ is given by

$$\bar{\boldsymbol{\alpha}}^M = (\mathbb{1} - B K_{\text{comb}}^{-1} B^\top \tilde{M}_M) \bar{\mathbf{b}}. \quad (53)$$

The compromise state of the full system, i.e., including the remaining original junctions $\bar{\mathbf{Y}}_\Delta^R$, is found by following the same procedure as for the hard constraint [Eqs. (36)–(38) and identifying $\bar{\boldsymbol{\alpha}}^M$ with $\bar{\mathbf{C}}_\Delta^M$].

Proceeding to compute the partition function, starting from Eq. (31), one has

$$\begin{aligned} Z^{\text{sc}} &= \frac{(2\pi)^{3N_R}}{\det A_{\Xi} \sqrt{\det M_R}} \\ &\times \int d\bar{\mathbf{Y}}_{\Delta}^R d\bar{\mathbf{Z}}_{\Delta} e^{-\beta E_M - \beta E_Z} \delta(\bar{\mathbf{Y}}_{\Delta}^M - B\bar{\mathbf{Z}}_{\Delta} - \bar{\mathbf{b}}) \\ &= \frac{(2\pi)^{3(N+1)}}{\det A_{\Xi} \sqrt{\det M_R} \sqrt{\det K_{\text{comb}}}} e^{-\beta F_{\text{enth}}^{\text{sc}}}, \end{aligned} \quad (54)$$

with $F_{\text{enth}}^{\text{sc}}$ given by Eq. (50). The free energy is again split into an enthalpic and a fluctuation contribution

$$\beta F^{\text{sc}} = -\log Z^{\text{sc}} = \beta F_{\text{fluc}}^{\text{sc}} + \beta F_{\text{enth}}^{\text{sc}}, \quad (55)$$

where the fluctuating contribution reads

$$\begin{aligned} \beta F_{\text{fluc}}^{\text{sc}} &= -3(N+1)\log(2\pi) + \log \det A_{\Xi} \\ &+ \frac{1}{2} \log \det M_R + \frac{1}{2} \log \det K_{\text{comb}}. \end{aligned} \quad (56)$$

Finally, we will show how the stiffness matrix K can be calculated from a canonical ensemble of midstep triad pose coordinates $\bar{\mathbf{Z}}_{\Delta}$ and prior knowledge about the innate structure and elasticity of the associated DNA, i.e., the static junctions s_i and the innate stiffness matrix M are given—implying likewise access to $s^{(k)}$ and $\bar{M}_M$. Assuming the given $\bar{\mathbf{Z}}_{\Delta}$ to be Gaussian distributed, one notes that according to Eq. (48) the effective midstep triads stiffness K_{comb} can be deduced via covariance inversion

$$K_{\text{comb}} = \langle (\bar{\mathbf{Z}}_{\Delta} - \langle \bar{\mathbf{Z}}_{\Delta} \rangle) (\bar{\mathbf{Z}}_{\Delta} - \langle \bar{\mathbf{Z}}_{\Delta} \rangle)^{\top} \rangle^{-1}. \quad (57)$$

We note that given just an ensemble of $\mu^{(k)}$, the intrinsically preferred poses $\mu_0^{(k)}$ are a priori unknown. In particular, the ensemble means [99] will be related to the compromise poses $\mu_{\alpha}^{(k)}$ [see Eq. (52)] rather than the $\mu_0^{(k)}$. In Appendix E we show how the $\mu_0^{(k)}$ can still be deduced by invoking a self-consistency relation that directly leads to an efficient gradient-descent optimization procedure.

E. Correction for translations

The approach of linearly approximating composites as given by Eq. (24)—and worked out in more detail in Appendix B—turns out to be highly accurate for the rotational components but struggles with the translational degrees of freedom, exhibiting particularly large discrepancies in the rise. While rotations are independent of translations, the directionality of translations is set by the rotational state, such that the coordinates of translational deformations intimately depend on the rotational coordinates. The dependence of the composite junction coordinates on these couplings is amplified by lever-arm effects: small errors in the rotational state of the first junction are magnified as they propagate downstream through the composite. In particular, bending fluctuations skew the distribution of translational fluctuations, as already noted in Ref. [88]. This compounding effect constitutes a fundamental limitation of linearly approximating composites in SE(3) around states with nonzero rotational deformations.

This observation suggests that one can leverage the proficiency of the linear transformation in capturing rotational

deformations to construct a two-step predictor-corrector procedure that improves translational accuracy. On the first pass (predictor step), the nucleosome-bound compromise state is computed using the previously described formalism from which the rotational components are adapted; Eq. (38) for the hard constraint and analogous for the soft constraint. On the second pass (corrector step), these rotations are used to construct an improved expansion for translations. The fidelity of this approach critically depends on the quality of the rotational estimate provided by the predictor step and will significantly degrade in contexts of strong rotational deformation, where first-order contributions are no longer sufficient to capture the majority of the rotation.

The idea is fully worked out in Appendix C and provides a modified form of the transformation matrix A_{Ξ} while also introducing an additional constant shift

$$\bar{\Xi}_{\Delta} = A'_{\Xi} \bar{\mathbf{Y}}_{\Delta} + \bar{\mathbf{a}}_{\Xi}, \quad (58)$$

with

$$\bar{\mathbf{a}}_{\Xi} = \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{a}} \end{bmatrix}. \quad (59)$$

The net effect of this augmented transformation is that the matrix M_{Ξ} is changed to $M'_{\Xi} = A'^{\top} M A'^{-1}$ and that $\bar{\mathbf{Y}}_{\Delta}^M$ in Eqs. (28) and (30) is replaced by $\bar{\mathbf{Y}}_{\Delta}^M - \bar{\mathbf{a}}$. All other relations of the hard and soft-constraint systems are unaffected.

III. RESULTS

A. Monte Carlo verification

The analytical solution to the nucleosome model fundamentally relies on a series of approximations. In particular, the nucleosome-mediated constraints are imposed on composite degrees of freedom—spanning several bp step junctions—that are constructed as a linear expansion of the underlying degrees of freedom. In this section, we demonstrate the quantitative fidelity of the analytical framework via numerical validation. We performed independent Monte Carlo (MC) simulations of the nucleosome model based on the hard-constraint assumptions. In the following, for both the analytical treatment and the MC simulation, we use the same set of parameters. As in our earlier computational work, we use here the hybrid parametrization [100] in which intrinsic deformations are derived from protein-DNA crystals [57] and the stiffnesses from atomistic molecular dynamics (MD) simulations [91].

The simulations are numerical replicas of the hard-constraint theory, consisting of a chain of rigid bodies [see Fig. 4(a)] where relative deformations are penalized according to the elastic Hamiltonian Eq. (10) and where bound mid-step triads are left invariant [see Fig. 4(b)]. The evaluation of the wrapped structure—i.e., the deformations necessary to achieve the superhelical binding—and fluctuational free energy is achieved by employing two complementary MC protocols. Low-temperature simulated annealing (cooling geometrically from 300 K to 10^{-5} K) is utilized to identify the compromise ground state $\bar{\mathbf{C}}_{\Delta}$ [see Eq. (38)] and to determine the enthalpic contribution of the free energy. Moreover, relative free energies between different nucleosome positioning sequences are evaluated with thermodynamic integration (TI) [101,102]. At this stage, we exclusively consider the fully

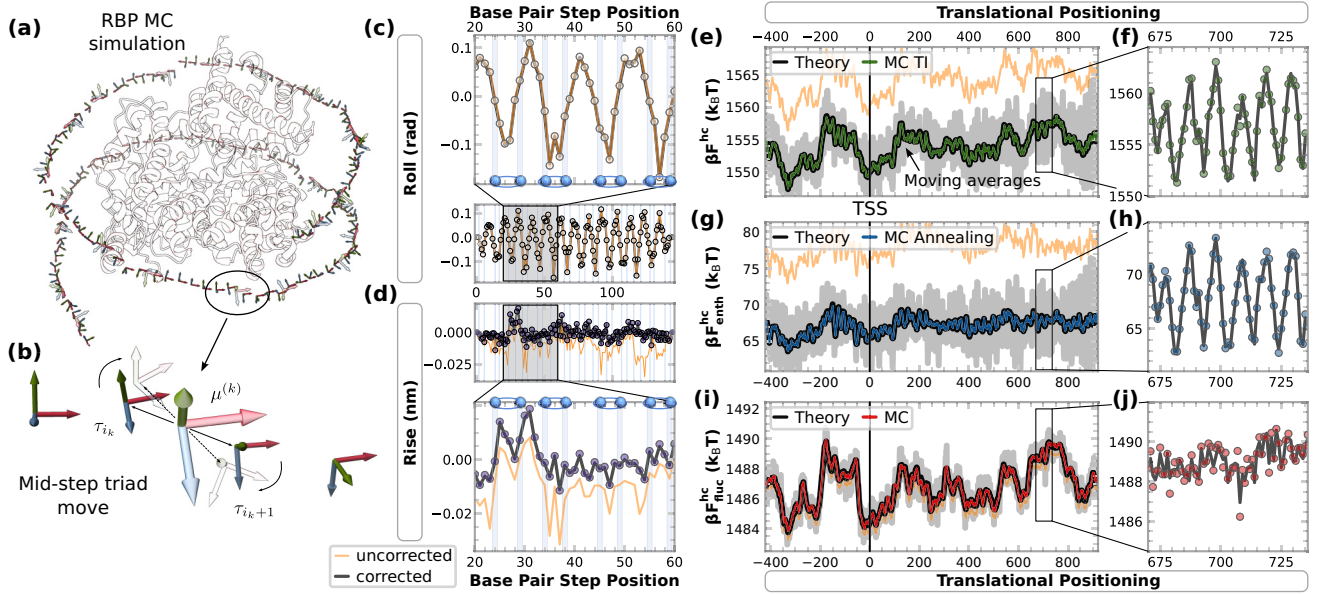


FIG. 4. Monte Carlo validation of the nucleosome theory. (a) Illustration of the representation used for Monte Carlo simulations, which consists of 147 reference frames. The nucleosome is considered implicitly via the hard constraint imposed on bound midstep triads. (b) Triads adjacent to binding sites are permuted via a custom midstep triads move that moves both associated triads in tandem and in such a way that the midstep triads remain unaltered. This is achieved by applying translations of equal distance but opposite direction and rotations over the same axis and angle, but in the opposite sense. (c), (d) Nucleosome-bound ground state excess deformation ($\bar{C}_{\Delta,i}$) for roll (c) and rise (d). Scatters show values determined with MC simulated annealing, orange lines correspond to the uncorrected hard-constraint theory [Eq. (38)], and black solid lines represent the translation correction discussed in Sec. II E. Blue circles and blue shaded vertical regions indicate the position of individual binding sites. (e)–(j) Comparison between theoretical and numerical nucleosome positioning free energies for all possible positions over a 1500 bp promoter region of the YAL002W gene in *Saccharomyces cerevisiae*: (e), (f) Full free energy, (g), (h) enthalpic component, and (i), (j) fluctuational component. Full range figures (e), (g), (i) display 11-bp moving averages for numerical results [(e) green, (g) blue, and (i) red], uncorrected theory [Eqs. (35), (40), and (41); orange line], and corrected theory (black solid line). Note that the numerical results are right on top of those obtained with the corrected theory. Results obtained with the corrected theory without application of the moving average are displayed as gray lines. Zoom-ins (f), (h), (j) display only the corrected theoretical values without averaging (black solid line).

wrapped state. Details about the MC protocols, used configurational permutation moves, and convergence are included in Appendix I.

Based on the Widom 601 sequence [103], we compare the analytic hard-constraint ground-state shape to the corresponding lowest-energy MC configuration. For both, we use hybrid elastic DNA parameters (structure from crystallographic data [57] and elasticity from MD [91]). Figures 4(c) and 4(d) depict the roll and rise components of $\bar{C}_{\Delta}$ as defined in Eq. (38), respectively. Analytical values obtained without the application of the translation correction (orange lines) show excellent agreement with MC sampled values (scatters) for the rotational degrees of freedom, as is exemplified by the roll component. However, discrepancies between analytical and numerical values are significantly more pronounced for the translational components, in particular, for rise; see Fig. 4(d). This highlights the necessity of the translation correction introduced in Sec. II E, which leverages the agreement for rotational components achieved in the original calculation. Application of the latter (black solid lines) eliminates the discrepancies almost entirely.

We next benchmark the analytical free-energy landscape against its numerical counterpart by sliding the position of the 147 bound bp's along the 1500 bp promoter region of the YAL002W gene in *Saccharomyces cerevisiae* (−500 bp to +1000 bp around TSS), at single bp increments. For each

position, we determine the total free energy—using MC TI in the simulations and Eq. (40) for the analytical solution [Figs. 4(e) and 4(f)]—and the enthalpic contribution—using low-temperature MC annealing and Eq. (35), respectively [Figs. 4(g) and 4(h)]—and infer the fluctuational term from the difference of the two [Figs. 4(i) and 4(j)]. Uncorrected values—using composite coordinates obtained by Eq. (21)—for the free energy (orange line) are found to systematically deviate from the MC-derived counterpart, overestimating the latter by about $10 k_B T$ on average. These deviations originate predominantly from the enthalpic contribution, which highlights the inherent deficiency of the uncorrected theory to predict the correct ground state. Invoking the translation correction lifts these discrepancies virtually completely, reproducing larger trends in the positioning free energy—as highlighted by the 11-bp moving averages of the free energy and its two components [thick solid lines in Figs. 4(e), 4(g), and 4(h); note that colored solid curves are right on top of the black solid curves in Figs. 4(e), 4(g), and 4(i)]—and local rotational positioning—as shown in the zoom-ins given in Figs. 4(f), 4(h), and 4(j). These zoom-ins correspond to the raw data traces shown in light gray in Figs. 4(e), 4(g), and 4(i). Over the entire range of 1354 evaluated nucleosome positions, we find RMSDs between corrected theory and MC-sampled values of $0.32 k_B T$, $0.18 k_B T$, and $0.29 k_B T$, for the F , F_{enth} , and F_{fluc} , respectively.

Comparing Figs. 4(f) and 4(h), we observe that the rotational positioning of nucleosomes is driven by the enthalpic part of the free energy. Moreover, it is known that the 10 bp undulations, which coincide with the DNA helical pitch, are caused by the local intrinsic curvature of the involved DNA stretch. For a comparison to experimental data, it is thus sufficient to focus on the enthalpic part. This analysis was performed in Ref. [64] using the computational model, and it was shown that the enthalpic landscape successfully predicts rotational positioning.

B. Parametrization of K matrix

Applying the soft-constraint model (Sec. IID) requires the stiffness matrix K , which characterizes the restriction of DNA fluctuations due to DNA-histone interactions at the binding sites. This matrix is essential for quantifying both the binding-mediated entropy loss and the extent to which each constrained midstep triad can adjust its pose to accommodate the shape of adjacent DNA segments. K is assumed to be sequence-independent as it characterizes the interactions between phosphate groups and the histone core. This assumption may not hold fully as midstep triads are used as proxies for phosphate groups, and the relative pose and flexibility of phosphate groups relative to midsteps may exhibit mild sequence-dependence. While this warrants further investigation in the future, the present assumption represents a natural leading-order approximation, as the dominant sequence-dependence in the model enters through the intrinsic shape and stiffness of the DNA itself rather than through the histone-DNA contacts. Improving on this without determining a separate K for each considered sequence would require a description that explicitly incorporates phosphate groups as distinct entities, such as the rigid body model cgNA+ [76].

We estimated K for canonical histones via atomistic MD simulations of a nucleosome built from the crystal structure NCP147 (PDB: 1KX5) [2]. Complementary simulations for free DNA were conducted to characterize the intrinsic stiffness and structure of the simulated bound DNA. This is necessary to decompose the effective stiffness K_{comb} into contributions stemming from innate DNA properties and DNA-histone interactions K [see Eq. (49)]. Trajectories were processed and mapped onto the RBP representation using MDTraj [104] and MDNA [105]. Unrestrained DNA stiffness M and structure s_i were then deduced via standard equipartition analysis [88,91] and nucleosome-bound midstep poses $\mu^{(k)}$ were calculated according to Eq. (11). Further details on the simulations and analysis are provided in Appendix F.

Our analysis reveals the $\mu^{(k)}$ to exhibit considerable fluctuations relative to the histone core, both in their orientations [see Fig. 5(a)] and positions [Fig. 5(b)]. To validate the assumption of Gaussianity, we calculated mean poses using the method described in Ref. [99] (once rotational averages are determined, the translational average is trivial) and parametrized the fluctuations around these poses in terms of the usual six coordinates. Resulting distributions confirm harmonicity [see Fig. 5(c)] as stipulated by the model (see Sec. IID). Systematic calculation of K was then performed via the procedure outlined in Appendix E.

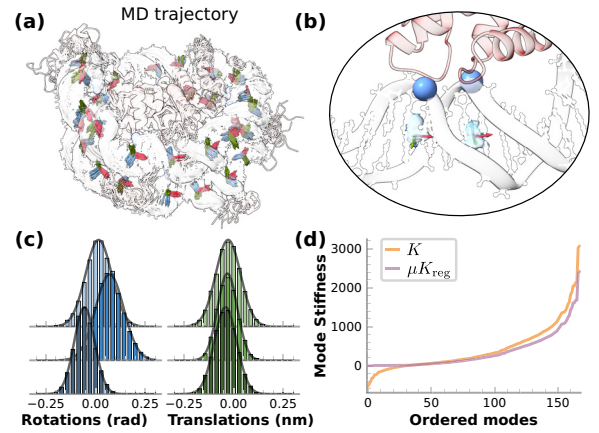


FIG. 5. Estimation of stiffness K via MD simulations. (a) Superimposed snapshots from a nucleosome MD simulation. (b) Snapshot of two constrained midstep triads showing the typical translational volume of midstep triad fluctuations (light blue region). (c) Distributions of the three rotations (top to bottom show analogues to: tilt, roll, twist) and the three translations (top to bottom show analogues to: shift, slide, rise) parametrizing fluctuations around a single averaged midstep triad. (d) Ordered eigenvalue spectra of the original (orange) and clipped-and-rescaled K matrix (purple).

The resulting K proved indefinite, i.e., it features negative eigenvalues, so it cannot serve as a proper stiffness matrix. These negative eigenvalues can have several causes. First, the high dimensionality of the free-energy landscape may require a large sample size to capture all relevant components of the system's configurational space. The statistics obtained with our MD simulations may be insufficient to capture all relevant modes. Second, our model treats the DNA's mechanical properties as identical in the histone-bound and free states—assuming histones exert only external forces—while in reality, histone-DNA interactions may alter the molecule's preferred conformation and soften specific modes. Subtracting the unsoftened (free DNA) modes from the combined (innate and histone-DNA contact) stiffness K_{comb} [see Eq. (49)] may then yield the observed negative eigenvalues for K .

To regularize K for the soft-constraint model, we perform spectral clipping by replacing all negative eigenvalues with small positive values (10^{-6}). Since this procedure inadvertently stiffens the combined system (histone contacts plus junctions), we rescale the regularized matrix K_{reg} such that the free energy corresponding to the regularized system matches that of the original system. Specifically, we introduce a scaling factor μ satisfying

$$\det(\mu K_{\text{reg}} + B^T \tilde{M}_M B) = \det K_{\text{comb}}. \quad (60)$$

A comparison between the original and clipped-and-rescaled spectra is shown in Fig. 5(d).

Finally, after regularization, we symmetrize the matrix with respect to the DNA-wrapping sense to reflect the dyad symmetry of the histone core. This is required to obtain truly symmetric wrapping energies for palindromic positioning sequences. Details about the symmetrization are provided in Appendix E 3.

We note that for canonical nucleosomes and under the assumption of sequence-independence, K needs to be sym-

metric to reflect the structure of the histone core. However, asymmetric K may arise in cases where nucleosomes feature asymmetric histone cores, as observed in Refs. [106,107].

C. Site-specific adsorption energies

The theory described thus far allows for the calculation of the free-energy cost of wrapping nucleosomal DNA around the histone core. To stabilize binding, this energetic penalty has to be counterbalanced by adsorption energies at the superhelical binding sites. Assuming sequence-independent adsorption energies, the relative elastic free energies of the full wrapped state of different sequences give meaningful results. However, when comparing different states of wrapping, i.e., cases where the number of engaged binding sites changes, such as for nucleosome breathing, site-specific adsorption energies need to be known. We assume all sequence-independent contributions, such as histone conformational entropy, counterion-release and solvent effects to be absorbed into these effective adsorption energies.

A first experimental attempt to determine these binding strengths was based on the micromanipulation of nucleosome arrays [108]. The data suggested unreasonably large adsorption energies. However, subsequently it was shown that the energetic barrier observed during pulling on a nucleosome reflects the bending energy of the in- and outgoing DNA during the nucleosome unwrapping event [24], as explained in more detail in Sec. III D. To avoid this unwanted barrier, a new experimental setup was devised, where histone-DNA contacts are disengaged via unzipping [89]. In this setup, recorded dwell times of the unzipping fork preceding subsequent dissociation yield insights about the locations of the binding sites and associated energy barriers. In particular, the resolution was sufficient to discern two distinct binding sites corresponding to each of the 14 SHL (where the DNA minor groove faces the octamer). Importantly, the study demonstrated marked inhomogeneity across the different binding sites. Moreover, the dwell-time peaks recorded during unzipping were found to exhibit significant asymmetry vis-à-vis opening from the two respective ends. As the nucleosome exhibits twofold symmetry, i.e., the two termini are energetically indistinguishable, this observation can only be rationalized with the nonpalindromic nature of the 601 sequence used in the experiment, which induces unbalanced wrapping energies.

In Appendix G, we show how the data from Ref. [89], see Fig. 11(a), can be used to extract the binding strengths of the various sites. Remarkably, by accounting for the highly asymmetric elastic energy of 601 positioning sequence [30,65], our model recovers near symmetric binding strength distributions; see Fig. 11(c). This confirms the assertion that the asymmetric distribution of the dwell times, depicted in Fig. 11(b), is mainly caused by the asymmetric DNA mechanics and shape of the nonpalindromic 601 sequence. By combining pairs of binding sites into single binding sites and then symmetrizing the distribution, by averaging over forward and backward directions, we arrive at the distribution of binding strengths shown in Fig. 6. Our analysis determines the adsorption strengths of the 14 binding sites up to a constant offset. This constant is determined further down via comparison to nucleosome breathing data. Importantly, as this distribution is

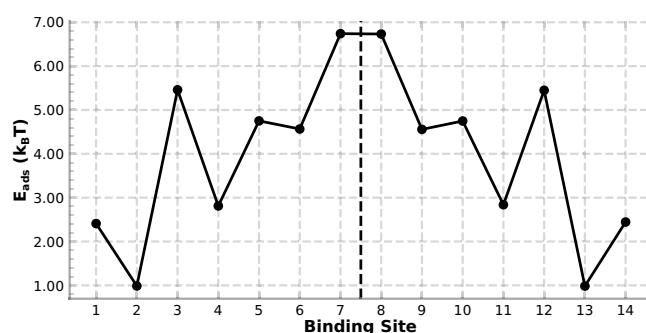


FIG. 6. Adsorption energy as a function of the binding site, 14 in total. Here, the constant offset is set to zero.

independent of the underlying DNA sequence, it can be used for any nucleosome in any partially unwrapped state.

Finally, we would like to emphasize that the analysis of the unzipping experiment presented here should be interpreted with caution. Details of the unzipping process into the nucleosome, such as the order of the opening of sites and the possible nucleosome rotations induced by the process, are not yet fully understood. A theoretical study of this experiment is urgently needed, but cannot be carried out with our current model, as it does not allow for bp separation.

D. Applications

In the following, we compare our analytical free-energy calculations to various experiments. These include binding free-energy differences for multiple pairs of DNA molecules, probability distributions of partial unwrapping states of nucleosomes, and the response of nucleosomes to external forces for various DNA sequences. We will demonstrate our model to be capable of explaining even rather detailed and, in some cases, puzzling experimental observations, both qualitatively and quantitatively.

1. Competitive reconstitution

As the first application, we combine data from two groups [82–84] that have determined free-energy differences between various pairs of DNA molecules, 22 pairs in total. In these *in vitro* competitive reconstitution experiments, nucleosome binding was achieved via lowering the salt concentration. It is known that the first step in the assembly of the nucleosome is the binding of the histone tetramer [82,109]. We therefore expect binding free energies derived from these *in vitro* competition experiments to primarily reflect the contributions from the tetramer. This is accounted for by applying constraints corresponding to $\text{SHL} \pm 2.5$, ± 1.5 , and ± 0.5 , which span the DNA region contacting the $(\text{H3-H4})_2$ tetramer, while considering the remaining DNA to be free. Binding free energies are then calculated by accounting for all possible positions of the tetramer on a given DNA molecule, weighted by the corresponding Boltzmann weight (due to elastic free energy).

In the following, we compare the predicted and experimental binding free-energy differences $\Delta\Delta G$ using the hybrid parametrization [100], as used in our earlier Monte Carlo simulation [64]. Using the hard-constraint model, we find

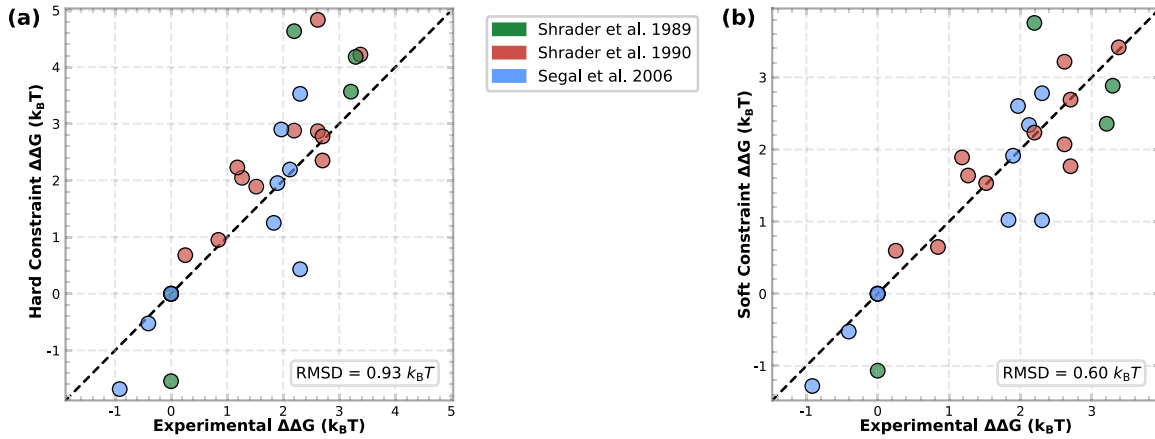


FIG. 7. Cross-correlation plot of experimental and theoretical free-energy differences ($\Delta\Delta G$), between pairs of DNA molecules. Plotted pairs are TG/GT, TG/TG-T, TG/TR-5, TG/TRGC [83], TG/ANISO, TG/TTT, TG/NOTA, TG/EXAT, TG/EXGC, TG/IAT, TG/IGC, TG/END, TG/ANNA, TG/34, TG/20 [84], c1/c2, c1/c3, d1/d2, d1/d3, d1/d4, d1/d5, e1/e2, and e1/e3 [82]. (a) Experimental vs hard-constraint model $\Delta\Delta G$, (b) experimental vs soft-constraint model $\Delta\Delta G$.

good agreement between theory and experiment, see Fig. 7, with a root-mean-square deviation of $0.93 k_B T$. Application of soft constraints instead reduces the RMSD to $0.6 k_B T$. Based on this observation, we will use the soft-constraint model throughout the rest of the paper.

Consideration of the enthalpic contribution only (as done in the simulation study [64]) results in markedly larger deviations of $1.62 k_B T$; see also Fig. 12(a) in Appendix H. This suggests that entropy contributes substantially to the binding free energy. Moreover, using the whole nucleosome instead of the tetramer significantly reduces the agreement between data and experiment, see Fig. 12(b) in Appendix H, where binding free-energy differences are overestimated in the theory. The corresponding root-mean-square deviation is $2.35 k_B T$.

2. Nucleosome breathing

As a second example, we compare our model to experimental data of thermally induced partial nucleosome unwrapping. This process, called site exposure or nucleosome breathing, has been first shown by measuring the accessibility of restriction sites inside nucleosomal DNA to the corresponding enzymes [28,85] and later by performing Förster resonance energy transfer (FRET) experiments [87,110–115] and atomic force microscopy [116,117]. Such experiments demonstrate nucleosomes to exhibit pronounced diversity resulting from the sequence-dependent mechanical properties of the wrapped DNA. However, in previous work, we argued that data from the various experimental approaches are challenging to interpret, which limits the testability of theoretical models [67,118].

However, an experimental approach based on small-angle x-ray scattering on a solution of nucleosomes [29,119] makes detailed observations that are more straightforward to interpret. This method facilitates the measurement of an ensemble of unwrapping states of nucleosomes containing the Widom 601 positioning sequence and even allows discrimination between wrapping from left and right. Moreover, the authors of Ref. [29] report how the ensemble of unwrapping states shifts with varying binding strength. By increasing the salt

concentration from 0.2 mM to 2.0 mM NaCl, the nucleosomes change from predominantly fully wrapped to unwrapped. Although this finding is not unexpected, one detail in the unwrapping scenario of the 601 nucleosome is striking: at intermediate salt concentrations, a highly asymmetric partially unwrapped state emerges with ≈ 65 bp unwrapped always from the same side. Termed by the authors of Ref. [29] as the “spring-loaded latch” mechanism, they argue that the observation is caused by a stiffer stretch of DNA along one arm that causes the wrapped nucleosome to jump discontinuously into this asymmetric state. Remarkably, our computational nucleosome model predicted the same effect [71]. However, our study considered only bending energies and did not account for entropy. Moreover, in that study, we assumed all binding sites to have identical adsorption energies.

We revisit this effect using our analytical model, which takes all of these contributions into account. In Fig. 8 we plot the relative occupancies of all possible unwrapping states which are characterized by the number of sites opened from the left, l , and from the right, r . Displayed are three cases differing by their average binding energy per site, $\langle E_{\text{ads}} \rangle$. For $\langle E_{\text{ads}} \rangle = 16.5 k_B T$ we find that the nucleosome is predominantly wrapped, see Fig. 8(a), whereas for $\langle E_{\text{ads}} \rangle = 13.5 k_B T$ it is predominantly unwrapped, Fig. 8(c). At intermediate binding strength, our analytical model also predicts the highly asymmetric state with $l = 0$ and $r = 5$, see Fig. 8(b), in close agreement with the experimental observations. For $\langle E_{\text{ads}} \rangle = 15.5 k_B T$, the value chosen in Fig. 8(b), the fully wrapped and this intermediate state have the same occupancies.

At first glance, this finding appears similar to the results obtained in Ref. [71]. However, closer inspection of individual contributions reveals additional insights. In Fig. 13(a) in Appendix H, we show the same plot as in Fig. 8, but considering only the enthalpic contribution and assuming uniform binding strength across all sites. This plot is almost identical to the plot produced by our earlier model; see Fig. 2 in Ref. [71]. The only difference between the two models lies in the nature of the constraints—rigid versus soft—which has here a negligible effect. Plotting next the occupancies accounting for

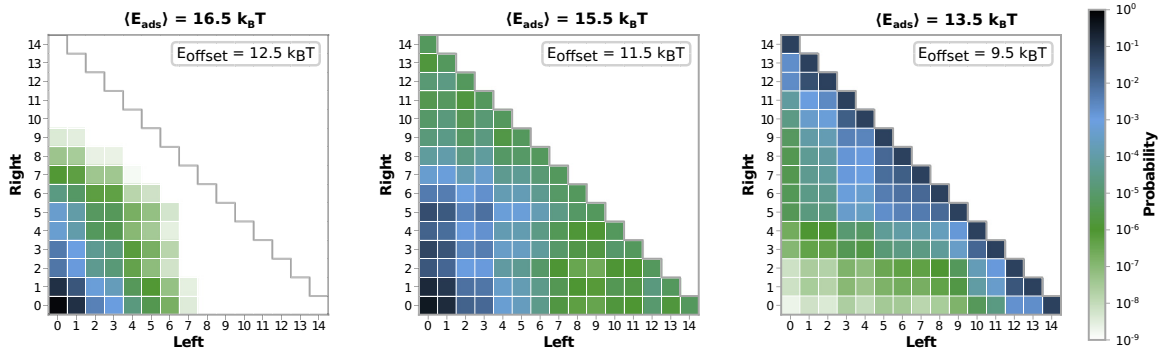


FIG. 8. Relative occupancies of unwrapping states for different average binding free energies plotted for different values of constant offsets E_{offset} to the binding energies from Fig. 6. Used values for E_{offset} are indicated in the respective plots, and resulting adsorption energies are provided in Fig. 14.

the full free energy of the DNA and still keeping the strengths of the binding sites homogeneous, we find surprisingly that the asymmetric state is never a total minimum but only occurs as a metastable state. Entropy washes out the “spring-loaded latch” mechanism; see Fig. 13(b) in Appendix H. Remarkably, including nonhomogeneous binding strengths of the binding sites into our model (obtained from the experimental data without fitting as explained in Sec. III C) recovers the effect, Fig. 8. This is caused by strong binding sites close to the dyad, which stop the unwrapping process; see Fig. 14 in Appendix H. In conclusion, our analytical model suggests that the “spring-loaded latch” is the result of the competition between several effects, with DNA elasticity and site-specific binding heterogeneity winning over entropy.

3. Nucleosomes under force

The third and final example concerns the response of nucleosomes to external forces. Early experiments suggested that nucleosomes are surprisingly stable in a micromanipulation setup where an array of 17 nucleosomes was pulled [108]. This led to the prediction of particularly strong binding sites that can withstand high forces. Subsequent theoretical modeling showed that the nucleosome’s remarkable stability arises not from exceptionally strong binding but from the interplay of DNA elasticity and its spool-like geometry; see Ref. [24] and subsequent work [25–27, 45, 65, 120–124]. As a nucleosome unwraps its last DNA turn under an external force f , the octamer needs to rotate by half a turn, leading to an energetically costly transition state. In this state, the octamer has flipped halfway while half a superhelical turn of wrapped DNA remains, requiring the DNA tails to describe sharp 90° bends. These bends amount to an energy barrier proportional to $\sqrt{f}$. As a result, under force, a metastable state arises, a nucleosome with one fully wrapped DNA turn and thus with the unwrapped DNA arms being nearly straight. In this state, five binding sites are open, allowing for five possible wrapping states depending on how many sites are opened from the left and the right; see Fig. 2 in Ref. [65]. Compared to the breathing experiment discussed above, the application of a force filters out all the unwrapping states where more or fewer than one DNA turn is wrapped, as this would lead to energetically costly bends in the DNA arms. Thus, we need only to compare

the free energies and corresponding occupancies of all the states with one full DNA turn. This corresponds in Fig. 8 to the diagonal connecting state $l = 5$ and $r = 0$ to state $l = 0$ and $r = 5$; the average strength $\langle E_{\text{ads}} \rangle$ of the binding sites is not important here, as it equally shifts the free energy of all these states.

An experiment combining FRET and micromanipulation on single nucleosomes [30] allows to distinguish between these various nucleosome states with one wrapped DNA turn. This experiment is particularly attractive for testing our model, as three different bp sequences were considered. Even though these sequences are closely related, they show entirely different occupancy profiles of the singly wrapped states. The traditional 601 sequence has the metastable state strongly preferring its DNA to be unwrapped from one end, the same end that shows enhanced breathing. Remarkably, for sequence 601MF, obtained by flipping the inner two quarters, the DNA is unwrapped preferentially from the other end. And finally, “symmetrizing” the 601 sequence by adding three TA steps to the inner quarter on the less stable half of the sequence (called sequence 601RTA), there seems to be no strong preference for unwrapping from a specific end.

We can test our model against these predictions by calculating the occupancies of all states with a total of five open sites (left and right). We enumerate these by the number of left-unbound sites, l , with $0 \leq l \leq 5$, which defines six states in total. These are plotted for all three sequences in Fig. 9(a). As can be seen, for 601, the free energy at the left end is lower than anywhere else. The right end has the highest free energy, about $5 k_B T$ higher than at the left end. Therefore, as long as the nucleosome is stuck in the metastable state, i.e., before the complete unspooling of the DNA at large forces, the FRET signal of a donor-acceptor pair at the left end is expected to be hardly affected by the pulling force. Conversely, FRET signals of donor-acceptor pairs at the right end should be much more sensitive to external tension and are expected to drop to small values once the applied force is large enough. This is indeed what is observed in the experiment [30]: whereas dyes probing the left end show a constant signal between 0 and 15 pN, the FRET efficiency for dyes at the right end drops already around 5 pN, reflecting the unwrapping of the outer turn. We expect that the states with $l = 0$ and $l = 1$ contribute equally to the FRET signal on the left end and neglect states $l > 1$,

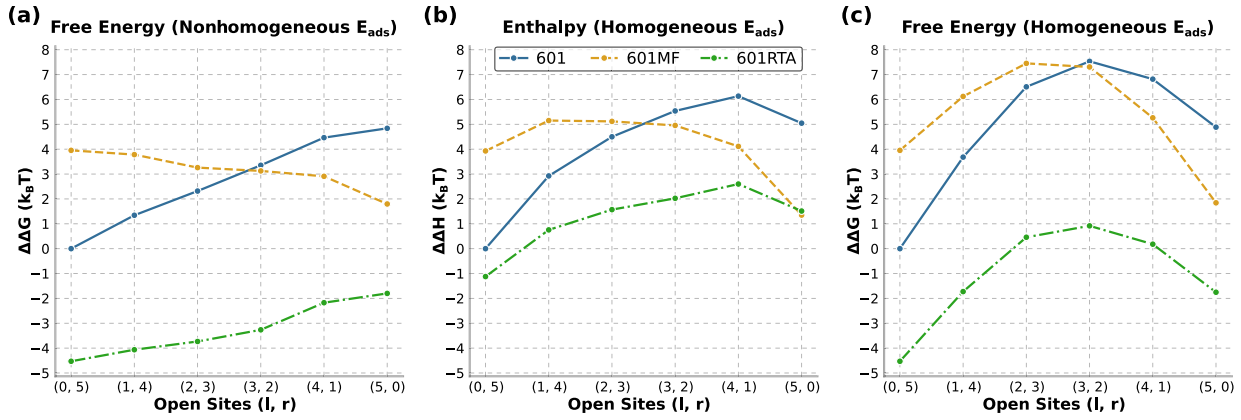


FIG. 9. Set of metastable states for nucleosomes 601, 601MF, and 601RTA under external tension. These states have a total of five open binding sites, l from the left end and r from the right end. (a) Free energy for nonhomogeneous binding energies, (b) enthalpy for homogeneous binding strength, and (c) free energy for homogeneous binding strength. $\Delta\Delta G$ and $\Delta\Delta H$ are the free energy and enthalpy relative to the 601 sequence in configuration $l = 0$ and $r = 5$.

see Ref. [118], and correspondingly, we account for two states when considering the right end. This suggests a nucleosome with a donor and acceptor fluorophore on the left end shows a 89% FRET efficiency and on the right of about 1%, close to the experimental values. For 601MF, the minimum lies now on the right side, Fig. 9(a), with a free energy that is about $2 k_B T$ lower than on the left side. In the experiment, FRET efficiencies are roughly constant between 5 and 15 pN with the right side being slightly reduced relative to the force-free case, whereas the left side shows only 20% of signal. Again, these results are reasonably close to what our model predicts, namely 12% FRET efficiency on the left and 64% on the right. Finally, for 601RTA, the insertion of the three TA steps indeed reduces the left-right asymmetry from $5 k_B T$ down to about $2.5 k_B T$. Whether this reduction of the free-energy difference is large enough to explain the observed FRET efficiencies for the 601RTA nucleosome remains an open question. The average of the data for both ends shows a continuous, roughly linear reduction of the FRET efficiencies with force. The authors of Ref. [30] explain this to result from averaging over two types of traces, a low and a high force response, similar to the response of the strong and the weak half of the 601 nucleosomes. However, comparing numbers between the left and right ends does not match up, and, especially, the number of traces for the pair of dyes on the left end, which is contained in the unmodified half of 601, is very low, indicating some problem with the 601RTA nucleosome. We therefore cannot give a quantitative comparison in this case.

Finally, Figs. 9(b) and 9(c) give the landscapes of the metastable states when one assumes a homogeneous binding strength of all sites, in Fig. 9(b) only for enthalpy and in Fig. 9(c) accounting also for entropy (including the entropy of the free DNA, neglecting the effect of tension). Going from Figs. 9(b) to 9(c), let us conclude that for all three sequences the effect of entropy is to increase the energetic barrier between the leftmost and rightmost states by an amount of about 1 to $2 k_B T$, without much affecting the asymmetry between these outer states. Then going from Figs. 9(c) to 9(a) teaches us that the nonhomogeneous strengths of the binding sites overcompensate for this entropic effect, leading to an almost linear dependence of the free energy on l .

Overall, the three applications provided in this section suggest that the analytical nucleosome model put forward here is capable of quantitatively predicting a wide range of experiments, including competitive nucleosome reconstitution, nucleosome breathing, and forced nucleosome unwrapping. It is worth emphasizing that we did not invoke any dependence of the strength of individual histone-DNA contacts on the local bp sequence. This suggests that the phenomena observed in these experiments can be explained by the sequence-dependent elastic and geometric properties of the involved DNA alone.

IV. BIOLOGICAL SIGNIFICANCE

The analytical framework developed here enables the calculation of the sequence-dependent free energy of nucleosomes in both fully wrapped and partially unwrapped states. Beyond the methodological advance, this capability has direct implications for understanding chromatin organization and gene regulation.

First, the free-energy landscape along genomic DNA encodes the intrinsic positioning preferences of nucleosomes. Regions of low nucleosome-free energy correspond to sequences that favor stable nucleosome formation, whereas regions with higher free energy are expected to disfavor nucleosome binding and thus correspond to nucleosome-depleted regions (NDRs). Such NDRs are characteristic features of promoters and regulatory elements in eukaryotic genomes and play a central role in determining which DNA regions remain accessible to the transcriptional machinery [125].

Second, the framework provides access to the free energies of partially unwrapped nucleosome states. These states correspond to transient nucleosome breathing, in which segments of nucleosomal DNA become temporarily exposed. The relative free energies of these states determine the probability and lifetime of DNA exposure and therefore provide a sequence-dependent measure of chromatin accessibility. This information is directly relevant for processes in which regulatory proteins, transcription factors, or repair enzymes interact with DNA sites that are otherwise buried within the nucleosome [28,111].

The model also provides a natural framework for understanding the interaction of transcription factors with nucleosomal DNA. Pioneer transcription factors such as SOX2 and OCT4 can recognize target motifs within nucleosomes and can exploit transient DNA exposure generated by nucleosome breathing [126]. In addition, their binding can locally deform the nucleosomal DNA and alter its rotational positioning, which may allosterically modulate the accessibility of nearby binding sites for other factors [127,128]. The sequence-dependent nucleosome free-energy landscape obtained from our model, therefore, defines the energetic background on which such cooperative and allosteric binding events occur.

Similar considerations apply to ATP-dependent chromatin remodelers, which exploit DNA flexibility and transient unwrapping to reposition or evict nucleosomes. Recent computational work, see Ref. [74], based on cryo-EM work [129], shows that chromatin remodelers reposition nucleosomes by generating twist defects in the wrapped DNA. These defects propagate through the nucleosome and shift the histone octamer along the DNA. Because the energetic cost of forming twist defects depends on DNA sequence, nucleosomes are shifted to new positions under the influence of the remodelers. This allows for the possibility that the bp-sequences contain “instructions” for remodelers to which positions to move nucleosomes away from their equilibrium position. A one-dimensional model of this mechanism is provided in Ref. [130]; our model would pave the way for a more accurate three-dimensional treatment.

Beyond transcriptional regulation in somatic cells, the model may also help to elucidate chromatin reorganization during spermiogenesis. During this process, most histones are replaced by protamines, yet a small fraction of nucleosomes (roughly 1–10%) is selectively retained at specific genomic loci associated with epigenetic information transfer and early embryonic development [131,132]. Whether the underlying DNA sequence contributes to this selective retention by providing intrinsically stable nucleosome configurations remains an open question. The sequence-dependent free-energy landscape provided by our framework offers a quantitative route to investigate this possibility.

Finally, the approach can be extended to study the effects of epigenetic and structural variations of chromatin. In particular, DNA modifications such as cytosine methylation can alter the mechanical properties of DNA and thereby modify nucleosome stability and breathing dynamics [53]. Likewise, histone variants or post-translational modifications can, in principle, be incorporated through modified binding geometries and interaction stiffness parameters. Such extensions will enable future applications of the framework to investigate how epigenetic modifications and histone variants reshape the nucleosome free-energy landscape and influence chromatin accessibility.

V. NOTE ON PARAMETERS

A key ingredient of our model is the choice of elastic parameters describing DNA mechanics. At present, no universally accepted parameter set exists for the RBP framework. The two most commonly used classes are derived either from protein-DNA cocrystal structures [57] or from MD simula-

tion of DNA oligomers in solution [91]. A commonly used compromise is the hybrid parametrization [100], in which the equilibrium geometry is taken from the crystallographic nucleosome structure and the stiffness from MD simulations. Because these parametrizations originate from different ensembles, they emphasize different aspects of DNA mechanics, and the appropriate choice depends on the observable under consideration.

For instance, both the competitive reconstitution experiments and the 601 “spring-loaded latch” experiment, discussed earlier, were performed under high-salt conditions. As a result, the configurational entropy gained through counter-ion release upon DNA wrapping is reduced, so the free energy is largely governed by the elastic cost of deforming a given DNA sequence into the nucleosomal superhelix. In this regime, the sequence-dependent shape of DNA encoded in the crystallographic nucleosome geometry plays a central role, and the hybrid parametrization provides a consistent description.

A similar situation arises in the force-induced unwrapping experiment of the 601 nucleosome and its variants. It was observed through independent cyclization measurements that asymmetric unwrapping is governed by the relative bendability of the two nucleosomal halves, identifying this as a shape-dominated observable [30]. Moreover, the Widom-601 sequence itself was identified in a SELEX experiment carried out under high-salt conditions [103], optimizing primarily for enthalpic binding affinity rather than entropy.

A qualitatively different situation arises for translational nucleosome positioning. As shown in Ref. [95], the GC-content dependence of nucleosome occupancy is largely entropy-driven, which is better captured by parametrizations that emphasize sequence-dependent entropy differences. An example of such a parametrization is the multiharmonic model [95] where the free DNA is modeled by MD parameters and the nucleosomal DNA by crystallographic parameters.

These considerations illustrate that the choice of parametrization must be guided by the physical quantity being studied. Moreover, when comparing to experiments, one needs to account for the experimental conditions.

VI. CONCLUSIONS AND PERSPECTIVES

In this work, we have introduced an analytical framework for calculating sequence-dependent nucleosome free energies. Our model extends approaches based on the RBP nucleosome modeling [64] and resolves the long-standing difficulty of imposing binding-site-mediated constraints while retaining analytical tractability [95]. We showed that a change of basis can be performed that separates the system into restrained and unrestrained parts. The degrees of freedom of the former are chosen to readily facilitate the imposition of histone-DNA contact constraints. The change of basis is based on prior work [88] that leverages linear approximations of composite junction coordinates. Known deficiencies of this approximation, i.e., systematic deviations in the translational components, are overcome by a two-step iteration that anchors the linear expansion on the rotationally relaxed state. Moreover, we show how the bound geometry can be enforced without imposing holonomic constraints (soft-constraint model), which,

by construction, present challenges for the calculation of the relative entropy across different wrapped states. Beyond the commonly employed elastic DNA parameters [57,91], the model is parametrized based on atomistic MD simulations and nucleosome unzipping data [89], yielding the stiffness matrix characterizing histone-DNA contacts K , and binding site-specific absorption energies, respectively. The framework is validated at multiple levels. Comparison to Monte Carlo simulations demonstrates excellent agreement for ground states and free-energy components and thereby validates the analytical framework. Furthermore, we show that the soft-constraint extension of the nucleosome model quantitatively captures experimental observations of (i) competitive nucleosome reconstitution, (ii) nucleosome breathing propensities, and (iii) force-induced unwrapping.

While the model parameters used in this work were tuned to account for canonical nucleosomes, extensions to noncanonical histones and epigenetic modifications can be achieved within the same framework and simply require reparametrization of the binding constraint parameters. In particular, this will require the characterization of histone-type-dependent constraint geometry and stiffness, which can be deduced from crystallographic data and atomistic simulations.

Observations of anti-cooperative nucleosome unwrapping [116,133], where the dissociation of either arm is found to strongly inhibit the unspooling of the other arm, suggest communication pathways through the nucleosome core. Such allosteric interactions are not captured by most nucleosome models. However, by construction, the stiffness matrix K introduced to limit DNA fluctuations at histone contact points allows for communication between different binding sites, i.e., it naturally incorporates potential allosteric interactions. Even though the data available for the present study was insufficient to obtain statistically meaningful insights on such communication, we envision these effects to be further explored in future studies.

We contemplate at least two avenues of extending the model beyond the RBP and nucleosomes. For one, the mathematical framework for analytically imposing binding constraints can be applied to any rigid-body-based model [58], not just the RBP. Especially, cgNA+ [76,77], a coarse-grained representation of DNA that extends the RBP by incorporating rigid bases and phosphates, lends itself as a promising candidate for our approach. In this context, additional components such as the influence of histone tails could be systematically explored. Moreover, we note that our scheme is not limited to histone-DNA contacts, but can be transferred to other DNA-protein complexes to explore the influence of DNA shape and elasticity on associated binding propensities. Therefore, we anticipate our approach may be instrumental in the broad exploration of the mechanical genome [134]—the encoding of biological function in the sequence-dependent mechanical properties of DNA.

ACKNOWLEDGMENTS

E.S., M.Y., and H.S. were supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—Grant No. EXC-2068-390729961. E.S. acknowledges funding

from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101202019. This research was supported in part by Grant No. NSF PHY-2309135 to the Kavli Institute for Theoretical Physics (KITP) and by the Gordon and Betty Moore Foundation Grant No. 2919.02. T.v.H. and J.V. were supported by the Dutch Research Council (NWO) via Grant No. OCENW.KLEIN.200.

E.S., M.Y., and H.S. conceived the study. E.S. developed the analytical framework. E.S. and M.Y. implemented the accompanying Python package. E.S. and M.Y. designed the MC protocols and performed the simulations. T.v.H. conducted the MD simulations. M.Y. and H.S. performed the comparative data analysis. E.S., M.Y., T.v.H., J.V., and H.S. wrote and edited the manuscript.

DATA AVAILABILITY

The free-energy calculations using the hard- and soft-constraint assumptions are implemented in a Python module available at Ref. [135]. The module includes three different libraries for the generation of elastic rigid-base-pair parameters of DNA [57,76,91].

APPENDIX A: ROTATIONS

A particularly convenient parametrization for elements of $SO(3)$ is provided by the exponential map

$$\mathcal{R} = \exp \hat{\Omega}, \quad (A1)$$

where $\hat{\Omega}$ are the antisymmetric tensors [rotation generators, $\mathfrak{so}(3)$]

$$\hat{\Omega} \equiv \begin{bmatrix} 0 & -\Omega_t & \Omega_v \\ \Omega_t & 0 & -\Omega_u \\ -\Omega_v & \Omega_u & 0 \end{bmatrix}, \quad (A2)$$

and $\exp$ is the matrix exponential

$$\exp \hat{\Omega} = \sum_{n=0}^{\infty} \frac{1}{n!} \hat{\Omega}^n. \quad (A3)$$

The inverse transformation of Eq. (A1), i.e., from a rotation matrix to an Euler vector, is given by the matrix logarithm

$$\hat{\Omega} = \log \mathcal{R}. \quad (A4)$$

As Ω_u , Ω_v , and Ω_t are simply the three components of the rotation vector Ω , there is a natural one-to-one relationship between Euler vectors and rotation generators that is encapsulated in the hat map and the vector map (also sometimes called the vee map)

$$\text{hat}(\Omega) \equiv \hat{\Omega}, \quad \text{and} \quad \text{vec}(\hat{\Omega}) \equiv \Omega. \quad (A5)$$

Under the action of the vector map, the Lie Bracket of $\mathfrak{so}(3)$ turns into the cross-product of $\mathbb{R}^3$, i.e., for any $\hat{\Omega}_1, \hat{\Omega}_2 \in \mathfrak{so}(3)$ one has

$$\text{vec}([\hat{\Omega}_1, \hat{\Omega}_2]) = \Omega_1 \times \Omega_2 = \hat{\Omega}_1 \Omega_2. \quad (A6)$$

This directly implies the relation

$$\hat{\Omega}_1 \Omega_2 = -\hat{\Omega}_2 \Omega_1, \quad (\text{A7})$$

which is frequently employed in this work. Another important relation is that for any $\mathcal{R} \in \text{SO}(3)$ and any $\hat{\Omega} \in \mathfrak{so}(3)$,

$$\text{vec}(\mathcal{R}\hat{\Omega}\mathcal{R}^\top) = \mathcal{R}\hat{\Omega}. \quad (\text{A8})$$

Central to this work is the notion of midstep triads as they constitute the anchor points for histone-DNA interactions in our model. Orientations of midstep triads are defined by reference to half rotations, i.e., rotations over the same axis, but over half the angle, for which we use the common shorthand notation

$$\sqrt{\mathcal{R}} \equiv \exp\left(\frac{1}{2} \log \mathcal{R}\right). \quad (\text{A9})$$

While using exponentiation and logarithms to relate the rotation vector and rotation matrix is convenient for the theoretical treatment in this work, it is computationally more efficient in practice to rely on expressions derived from Rodrigues' rotation formula

$$\mathcal{R} = \mathbb{1} + \sin(\Omega)\hat{\Omega} + (1 - \cos \Omega)\hat{\Omega}^2, \quad (\text{A10})$$

where $\Omega = \|\Omega\|$. The inverse relation, i.e., the analog to Eq. (A4) is given by

$$\Omega = \frac{\Omega \text{vec}(\mathcal{R} - \mathcal{R}^\top)}{2 \sin \Omega}. \quad (\text{A11})$$

Here, the rotation angle Ω can be computed via the expression

$$\text{tr} \mathcal{R} = 1 + 2 \cos \Omega. \quad (\text{A12})$$

Another important concept for this work is that of the right and left Jacobians, which account for the Baker-Campbell-Hausdorff relations between small perturbations in $\text{SO}(3)$ and $\mathfrak{so}(3)$:

$$\begin{aligned} \exp(\hat{\Omega} + \delta\hat{\Omega}) &\approx \exp(\hat{\Omega}) \exp(\text{hat}[J_R(\Omega)\delta\Omega]) \\ &\approx \exp(\text{hat}[J_L(\Omega)\delta\Omega]) \exp(\hat{\Omega}). \end{aligned} \quad (\text{A13})$$

The right Jacobian is explicitly given by

$$J_R(\Omega) = \mathbb{1} - \left(\frac{1 - \cos \Omega}{\Omega^2}\right)\hat{\Omega} + \left(\frac{\Omega - \sin \Omega}{\Omega^3}\right)\hat{\Omega}^2, \quad (\text{A14})$$

and it is related to the left Jacobian via the identities

$$J_L(\Omega) = J_R(-\Omega) = J_R^\top(\Omega) = \mathcal{R}J_R(\Omega), \quad (\text{A15})$$

where $\mathcal{R}$ is defined according to Eq. (A1). We further note that $\mathcal{R}$, $J_L(\Omega)$, $J_R(\Omega)$ are power series of $\hat{\Omega}$ which implies that they commute. Equation (A13) is exact only in the limit of infinitesimal perturbations. However, we will employ it throughout this work to construct first-order expansions for small fluctuating components.

APPENDIX B: COMPOSITE TRANSFORMATIONS

In this Appendix, we derive the linear transformation, Eq. (21), that transforms the system representation expressed fully in terms of dynamic bp step coordinates ($\tilde{\mathbf{Y}}_\Delta$) to system coordinates that include certain composite junctions spanning between midstep triads ($\tilde{\mathbf{E}}_\Delta$). This calculation is split into two steps: First, we will show how static transformations

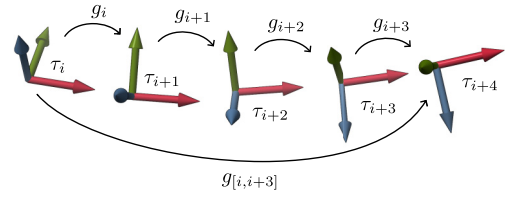


FIG. 10. Junction between triads τ_i and τ_{i+4} , is the product of the junctions g_i , g_{i+1} , g_{i+2} , and g_{i+3} .

and dynamic coordinates of such composite junctions can be written in terms of the coordinates of the individual bp step junctions. Then we will introduce the formalism that allows for midstep triads to be treated as ordinary triads, such that the results for general composites are directly applicable, which immediately leads to the linear transformation Eq. (21).

1. General composites

Central to the construction of composite coordinates is the principle that the relative transformation between any pair of triads say τ_i and τ_{j+1} (with $i < j + 1$) may be decomposed into the product of single junctions connecting adjacent triads within the range spanned by i and $j + 1$, see Fig. 10,

$$\tau_i^{-1} \tau_{j+1} = \prod_{l=i}^j g_l. \quad (\text{B1})$$

Following previous work [88], we introduce shorthand notation for such sequential products of transformations:

$$g_{[i,j]} \equiv \prod_{l=i}^j g_l. \quad (\text{B2})$$

Note that Eq. (B1) implies

$$g_{[i,i-1]} = \tau_i^{-1} \tau_i = \mathbb{1}. \quad (\text{B3})$$

The aim is to express $g_{[i,j]}$ again in terms of a static and a dynamic component

$$g_{[i,j]} = s_{[i,j]} d_{(i,j)}, \quad (\text{B4})$$

where the subscript bracket (i, j) indicates nontrivial component compositions, i.e., $d_{(i,j)}$ is not simply the product of the individual underlying dynamic junction components d_l .

As evidenced by considering the situations of zero deformations, $\tilde{\mathbf{Y}}_\Delta = \mathbf{0}$, implying $d_l = \mathbb{1}$ and therefore $g_l = s_l$ for all l , the static component of $g_{[i,j]}$ must be the simple composite

$$s_{[i,j]} \equiv \prod_{l=i}^j s_l = \begin{bmatrix} \mathcal{S}_{[i,j]} & \mathbf{s}_{(i,j)} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (\text{B5})$$

with

$$\mathbf{s}_{(i,j)} = \sum_{l=i}^j \mathcal{S}_{[i,l-1]} \mathbf{s}_l. \quad (\text{B6})$$

The dynamic composite is then defined as

$$d_{(i,j)} = s_{[i,j]}^{-1} g_{[i,j]} = \begin{bmatrix} \mathcal{D}_{(i,j)} & \mathbf{d}_{(i,j)} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (\text{B7})$$

and can usually not be written as the product of single-step deformations (unless $s_i = \mathbb{1}$ for all i)—which justifies the distinction in the composite bracket notations. Equation (B7) defines the dynamic composite parameters,

$$\mathcal{P}(d_{(i,j)}) \equiv \mathbf{Y}_{\Delta,(i,j)} = \begin{bmatrix} \Phi_{\Delta,(i,j)} \\ \mathbf{d}_{(i,j)} \end{bmatrix}, \quad (\text{B8})$$

which we seek to write in terms of the single junction deformations $\mathbf{Y}_{\Delta,l}$ [see Eq. (8)] with $i \leq l \leq j$. While this cannot be done exactly (in general, the relation will not be linear), one can construct a linear expansion of the transformation, which will yield a good approximation of $\mathbf{Y}_{\Delta,(i,j)}$ as long as the $\mathbf{Y}_{\Delta,l}$ are sufficiently small. This linear expansion was derived in Ref. [88]; since several steps are essential for the subsequent

development, we recapitulate the derivation below in a slightly more concise form.

A straightforward calculation shows that Eq. (B7) can be rewritten as

$$d_{(i,j)} = \prod_{l=i}^j (s_{[l+1,j]}^{-1} d_l s_{[l+1,j]}). \quad (\text{B9})$$

The explicit form of the static composites $s_{[l+1,j]}$ is given in Eqs. (B5) and (B6). Its inverse, $s_{[l+1,j]}^{-1}$, reads

$$s_{[i,j]}^{-1} = \begin{bmatrix} \mathcal{S}_{[i,j]}^\top & -\mathcal{S}_{[i,j]}^\top \mathbf{s}_{(i,j)} \\ \mathbf{0}^\top & 1 \end{bmatrix}. \quad (\text{B10})$$

Using this, each term in the product of Eq. (B9) is expressed explicitly as

$$s_{[l+1,j]}^{-1} d_l s_{[l+1,j]} = \begin{bmatrix} [\mathcal{S}_{[l+1,j]}^\top \mathcal{D}_l \mathcal{S}_{[l+1,j]}] & [\mathcal{S}_{[l+1,j]}^\top (\mathcal{D}_l \mathbf{s}_{(l+1,j)} + \mathbf{d}_l) - \mathcal{S}_{[l+1,j]}^\top \mathbf{s}_{(l+1,j)}] \\ \mathbf{0}^\top & 1 \end{bmatrix}. \quad (\text{B11})$$

Considering only the rotational part of Eq. (B9), one identifies the rotational composite

$$\mathcal{D}_{(i,j)} = \prod_{l=i}^j (\mathcal{S}_{[l+1,j]}^\top \mathcal{D}_l \mathcal{S}_{[l+1,j]}). \quad (\text{B12})$$

Expanding $\mathcal{D}_l$ to linear order in $\hat{\Phi}_{\Delta,l}$,

$$\mathcal{D}_l = \exp(\hat{\Phi}_{\Delta,l}) \approx \mathbb{1} + \hat{\Phi}_{\Delta,l}, \quad (\text{B13})$$

one further has

$$\begin{aligned} \mathcal{D}_{(i,j)} &\approx \prod_{l=i}^j (\mathbb{1} + \mathcal{S}_{[l+1,j]}^\top \hat{\Phi}_{\Delta,l} \mathcal{S}_{[l+1,j]}) \\ &\approx \mathbb{1} + \sum_{l=i}^j \mathcal{S}_{[l+1,j]}^\top \hat{\Phi}_{\Delta,l} \mathcal{S}_{[l+1,j]}, \\ &\approx \exp \left(\sum_{l=i}^j \mathcal{S}_{[l+1,j]}^\top \hat{\Phi}_{\Delta,l} \mathcal{S}_{[l+1,j]} \right), \end{aligned} \quad (\text{B14})$$

where in the second line, all terms of higher than linear order in the fluctuating components $\hat{\Phi}_{\Delta,l}$ were dropped.

Using

$$\mathcal{D}_{(i,j)} \equiv \exp(\hat{\Phi}_{\Delta,(i,j)}), \quad (\text{B15})$$

in conjunction with the relation (A8), one arrives at the final expression for the rotational composite coordinates

$$\Phi_{\Delta,(i,j)} \approx \sum_{l=i}^j \mathcal{S}_{[l+1,j]}^\top \Phi_{\Delta,l}. \quad (\text{B16})$$

Moving on to the translational component, we first combine Eqs. (B7), (B9), and (B11) to explicitly write out $\mathbf{d}_{(i,j)}$:

$$\mathbf{d}_{(i,j)} = \sum_{l=i}^j \left(\prod_{k=i}^{l-1} [\mathcal{S}_{[k+1,j]}^\top \mathcal{D}_k \mathcal{S}_{[k+1,j]}] [\mathcal{S}_{[l+1,j]}^\top (\mathcal{D}_l \mathbf{s}_{(l+1,j)} + \mathbf{d}_l) - \mathcal{S}_{[l+1,j]}^\top \mathbf{s}_{(l+1,j)}] \right). \quad (\text{B17})$$

Expanding $\mathcal{D}_l$ again to linear order and omitting all terms of higher than linear order in the fluctuating components, this expression simplifies to

$$\mathbf{d}_{(i,j)} \approx \sum_{l=i}^j \mathcal{S}_{[l+1,j]}^\top (\hat{\Phi}_{\Delta,l} \mathbf{s}_{(l+1,j)} + \mathbf{d}_l). \quad (\text{B18})$$

Finally, rearranging $\hat{\Phi}_{\Delta,l}$ to the right-hand side of the equation via [see also Eq. (A7)]

$$\hat{\Phi}_{\Delta,l} \mathbf{s}_{(l+1,j)} = -\hat{s}_{(l+1,j)} \Phi_{\Delta,l}, \quad (\text{B19})$$

one arrives at

$$\mathbf{d}_{(i,j)} \approx \sum_{l=i}^j \mathcal{S}_{[l+1,j]}^\top (\mathbf{d}_l - \hat{s}_{(l+1,j)} \Phi_{\Delta,l}). \quad (\text{B20})$$

As one may have anticipated, composite translations depend not only on the translational but also the rotational components of the single-step junctions. We remark that $\mathbf{s}_{(j+1,j)} = \mathbf{0}$ [see Eq. (B3)] such that the rotation of the last junction $\Phi_{\Delta,j}$ does not contribute to the composite translation.

The results of Eqs. (B16) and (B20) can be unified into the form (see also Ref. [88])

$$\mathbf{Y}_{\Delta,(i,j)} \approx \sum_{l=i}^j A_{l,(i,j)} \mathbf{Y}_{\Delta,l}, \quad (\text{B21})$$

with

$$A_{l,(i,j)} = \begin{bmatrix} \mathcal{S}_{[l+1,j]}^T & 0 \\ -\mathcal{S}_{[l+1,j]}^T \hat{s}_{(l+1,j)} & \mathcal{S}_{[l+1,j]}^T \end{bmatrix}. \quad (\text{B22})$$

We note that (B22) implies that the coordinates of the last junction in a composite require no transformation

$$A_{j,(i,j)} = \mathbb{1}_{6 \times 6}. \quad (\text{B23})$$

2. Midstep junctions

Composites constrained in our nucleosome model do not span between two bp triads τ_i , but between certain midstep triads μ_i and μ_j . Analogous to the treatment in the last section [Eq. (B2)], we would like to decompose the junctions spanning between these triads, $g^{(k)} = \mu_i^{-1} \mu_j$ [Eq. (15)], into a product of individual bp step junctions. However, in this case, the first and the last such junction only span over half of their respective bp steps [see also Fig. 2(a)]. To accommodate this notion, we introduce the decomposition of the junctions g_i into two half-step junctions:

$$g_i = g_i^{(\text{I})} g_i^{(\text{II})}. \quad (\text{B24})$$

The left-hand junction $g_i^{(\text{I})}$ and right-hand junction $g_i^{(\text{II})}$ are defined such that they precisely constitute the respective junctions between triads and midstep triads, i.e., $\mu_i = \tau_i g_i^{(\text{I})}$, and $\tau_{i+1} = \mu_i g_i^{(\text{II})}$, such that

$$g_i^{(\text{I})} = \tau_i^{-1} \mu_i, \quad (\text{B25})$$

$$g_i^{(\text{II})} = \mu_i^{-1} \tau_{i+1}. \quad (\text{B26})$$

Inserting the expressions for triads [Eq. (1)] and midstep triads [Eq. (11)] into Eqs. (B25) and (B26) one finds that

$$g^{(\text{I})} = \begin{bmatrix} \sqrt{\mathcal{R}} & \frac{1}{2} \mathbf{w} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (\text{B27})$$

$$g^{(\text{II})} = \begin{bmatrix} \sqrt{\mathcal{R}} & \frac{1}{2} \sqrt{\mathcal{R}}^T \mathbf{w} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (\text{B28})$$

i.e., the two junctions are not identical. Equipped with this definition, the full composite $g^{(k)}$ can be decomposed as

$$g^{(k)} = \mu_i^{-1} \mu_j = g_i^{(\text{II})} g_{[i+1,j-1]}^{(\text{I})}. \quad (\text{B29})$$

Throughout the remainder of this section, we will work out how the coordinates of the two half-step junctions can be written in terms of the respective full-junction coordinates, which in turn, enables the straightforward application of the linearized composite expression Eq. (B21). Note that for broader generality and consistency with the previous section, we have used the indices i and j instead of i_k and i_{k+1} . Moreover, since only single junction decompositions are considered here, we will drop the positional indices altogether for the remainder of this section.

To be able to apply the formalism of constructing composite transformations from the previous section, it is required to split the half-step junctions $g_i^{(\text{I})}$ and $g_i^{(\text{II})}$ into static and dynamic components

$$g^{(\text{I})} = s^{(\text{I})} d^{(\text{I})}, \quad (\text{B30})$$

$$g^{(\text{II})} = s^{(\text{II})} d^{(\text{II})}. \quad (\text{B31})$$

Specifically, we require expressions for the static components $s^{(\text{I})}$ and $s^{(\text{II})}$ in terms of the static component of the full junction s , and a linear expression for the parameters of the dynamic coordinates

$$\mathbf{Y}_{\Delta}^{(\text{I})} = \mathcal{P}(d^{(\text{I})}) = \begin{bmatrix} \Phi_{\Delta,i}^{(\text{I})} \\ \mathbf{d}_i^{(\text{I})} \end{bmatrix}, \quad (\text{B32})$$

$$\mathbf{Y}_{\Delta}^{(\text{II})} = \mathcal{P}(d^{(\text{II})}) = \begin{bmatrix} \Phi_{\Delta,i}^{(\text{II})} \\ \mathbf{d}_i^{(\text{II})} \end{bmatrix}, \quad (\text{B33})$$

in terms of the dynamic coordinates of the full junction

$$\mathbf{Y}_{\Delta} = \begin{bmatrix} \Phi_{\Delta} \\ \mathbf{d} \end{bmatrix}. \quad (\text{B34})$$

One observes that the half-step junctions are naturally parametrized in terms of the full junction coordinates [see Eqs. (3) and (A4)]

$$\mathbf{X} = \begin{bmatrix} \Omega \\ \mathbf{w} \end{bmatrix} = \mathcal{P}(g), \quad (\text{B35})$$

for which static and dynamic components are split at the coordinate level rather than the transformation level

$$\Omega = \Omega_0 + \Omega_{\Delta}, \quad (\text{B36})$$

$$\mathbf{w} = \mathbf{w}_0 + \mathbf{w}_{\Delta}. \quad (\text{B37})$$

To transform these coordinates to the dynamic half-step coordinates (B32) and (B33), we will rely on the transformations between the $\mathbf{X}$ and the $\mathbf{Y}$ coordinates developed in Ref. [88], which we will summarize here. The exact relation between the $\mathbf{X}$ and $\mathbf{Y}$ coordinates is given by

$$\mathcal{P}^{-1}(\mathbf{X}) = g = sd = \mathcal{P}^{-1}(\mathbf{Y}_0) \mathcal{P}^{-1}(\mathbf{Y}_{\Delta}). \quad (\text{B38})$$

By construction, the static components are identical in both representations, which can be seen by setting fluctuations in both coordinates to zero:

$$\Phi_0 = \Omega_0, \quad (\text{B39})$$

$$\mathbf{s} = \mathbf{w}_0, \quad (\text{B40})$$

such that

$$\mathbf{Y}_0 = \begin{bmatrix} \Omega_0 \\ \mathbf{w}_0 \end{bmatrix}. \quad (\text{B41})$$

The transformation from $\mathbf{X}$ to $\mathbf{Y}_{\Delta}$ is, in general, linear only for the translational component

$$\mathbf{d} = \mathcal{S}^T \mathbf{w}_{\Delta}, \quad (\text{B42})$$

while the exact relation for the rotational component is provided by the Baker-Campbell-Hausdorff Formula. However,

for sufficiently small Ω_Δ , it is well approximated by the right Jacobian

$$\Phi_\Delta \approx J_R(\Omega_0) \Omega_\Delta, \quad (\text{B43})$$

as defined in Eqs. (A13) and (A14).

a. Left-hand junction

First, we work out the left-hand junction. Setting the dynamic components in Eq. (B27) to zero, one directly identifies

$$s^{(I)} = \begin{bmatrix} \mathcal{S}^{(I)} & \mathbf{s}^{(I)} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{\mathcal{S}} & \frac{1}{2}\mathbf{s} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (\text{B44})$$

which implies

$$\mathcal{S}^{(I)} = \sqrt{\mathcal{S}}, \quad (\text{B45})$$

$$\Phi_0^{(I)} = \frac{1}{2}\Phi_0, \quad (\text{B46})$$

$$\mathbf{s}^{(I)} = \frac{1}{2}\mathbf{s}. \quad (\text{B47})$$

From here, the dynamic junction may be written as

$$d^{(I)} = s^{(I)-1} g^{(I)} = \begin{bmatrix} \sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}} & \frac{1}{2} \sqrt{\mathcal{S}^\top} (\mathbf{w} - \mathbf{s}) \\ \mathbf{0}^\top & 1 \end{bmatrix}. \quad (\text{B48})$$

Combining Eqs. (B40), (B42), and (B48) the translation reads

$$\mathbf{d}^{(I)} = \frac{1}{2} \sqrt{\mathcal{S}^\top} (\mathbf{w}_0 + \mathbf{w}_\Delta - \mathbf{s}) = \frac{1}{2} \sqrt{\mathcal{S}} \mathbf{d} = \frac{1}{2} \mathcal{S}^{(I)} \mathbf{d}. \quad (\text{B49})$$

Similarly, combining Eqs. (B39) and (B48) one identifies

$$\begin{aligned} \exp \hat{\Phi}_\Delta^{(I)} &= \sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}} \\ &= \exp \left(-\frac{1}{2} \hat{\Omega}_0 \right) \exp \left(\frac{1}{2} [\hat{\Omega}_0 + \hat{\Omega}_\Delta] \right) \\ &\approx \exp \left(\frac{1}{2} \text{hat}[J_R(\Omega_0/2) \Omega_\Delta] \right), \end{aligned} \quad (\text{B50})$$

where in the last step we used Eq. (A13) to separate the second exponential. After equating the exponents in Eq. (B50) and substituting Ω_0 and Ω_Δ for their expressions in terms of Φ_0 and Φ_Δ [via Eqs. (B39) and (B43)] the dynamic half-step rotation writes

$$\Phi_\Delta^{(I)} \approx \frac{1}{2} J_R \left(\frac{\Omega_0}{2} \right) \Omega_\Delta = \frac{1}{2} J_R \left(\frac{\Phi_0}{2} \right) J_R^{-1}(\Phi_0) \Phi_\Delta. \quad (\text{B51})$$

Once again, this expression should be read as a first-order expansion.

Jointly, Eqs. (B49) and (B51) can be expressed as the linear transformation that maps $\mathbf{Y}_\Delta$ into $\mathbf{Y}_\Delta^{(I)}$

$$\mathbf{Y}_\Delta^{(I)} = A^{(I)} \mathbf{Y}_\Delta, \quad (\text{B52})$$

with

$$A^{(I)} = \frac{1}{2} \begin{bmatrix} J_R \left(\frac{\Phi_0}{2} \right) J_R^{-1}(\Phi_0) & 0 \\ 0 & \sqrt{\mathcal{S}} \end{bmatrix}. \quad (\text{B53})$$

b. Right-hand junction

Moving on to the right-hand junction, we again start by setting the dynamic components in Eq. (B28) to zero to find the static pose

$$s^{(II)} = \begin{bmatrix} \mathcal{S}^{(II)} & \mathbf{s}^{(II)} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{\mathcal{S}} & \frac{1}{2} \sqrt{\mathcal{S}^\top} \mathbf{s} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad (\text{B54})$$

which yields

$$\mathcal{S}^{(II)} = \sqrt{\mathcal{S}}, \quad (\text{B55})$$

$$\Phi_0^{(II)} = \frac{1}{2}\Phi_0, \quad (\text{B56})$$

$$\mathbf{s}^{(II)} = \frac{1}{2} \sqrt{\mathcal{S}^\top} \mathbf{s}. \quad (\text{B57})$$

The dynamic transformation then becomes

$$\begin{aligned} d^{(II)} &= s^{(II)-1} g^{(II)} \\ &= \begin{bmatrix} \sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}} & \frac{1}{2} (\sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}^\top} \mathbf{w} - \mathcal{S}^\top \mathbf{s}) \\ \mathbf{0}^\top & 1 \end{bmatrix}. \end{aligned} \quad (\text{B58})$$

One recognizes the rotational component of the right-hand junction to be identical to the left-hand component [see rotation in Eq. (B48)], such that its coordinates obey the same expression [see Eq. (B51)]

$$\Phi_\Delta^{(II)} = \frac{1}{2} J_R \left(\frac{\Phi_0}{2} \right) J_R^{-1}(\Phi_0) \Phi_\Delta. \quad (\text{B59})$$

However, in this case, the translational component is slightly more involved

$$\begin{aligned} \mathbf{d}^{(II)} &= \frac{1}{2} \sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}^\top} \mathbf{w} - \frac{1}{2} \mathcal{S}^\top \mathbf{s} \\ &= \frac{1}{2} \sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}^\top} [\mathcal{S} \mathbf{d} + \mathbf{s}] - \frac{1}{2} \mathcal{S}^\top \mathbf{s} \\ &= \frac{1}{2} \sqrt{\mathcal{S}^\top} e^{-\hat{\Phi}_\Delta^{(II)}} \sqrt{\mathcal{S}} [\mathbf{d} + \mathcal{S}^\top \mathbf{s}] - \frac{1}{2} \mathcal{S}^\top \mathbf{s}, \end{aligned} \quad (\text{B60})$$

where we used the first line of Eq. (B50) in conjunction with the prior realization that $\hat{\Phi}_\Delta^{(I)} = \hat{\Phi}_\Delta^{(II)}$ to substitute

$$\sqrt{\mathcal{R}^\top} \sqrt{\mathcal{S}} = (\sqrt{\mathcal{S}^\top} \sqrt{\mathcal{R}})^\top = e^{-\hat{\Phi}_\Delta^{(II)}}. \quad (\text{B61})$$

Expanding the exponential on the right-hand side to first order and discarding quadratic terms in the fluctuating components, one sees that the constant terms cancel out

$$\begin{aligned} \mathbf{d}^{(II)} &\approx \frac{1}{2} (\mathbb{1} - \sqrt{\mathcal{S}^\top} \hat{\Phi}_\Delta^{(II)} \sqrt{\mathcal{S}}) (\mathbf{d} + \mathcal{S}^\top \mathbf{s}) - \frac{1}{2} \mathcal{S}^\top \mathbf{s} \\ &\approx \frac{1}{2} (\mathbf{d} - \sqrt{\mathcal{S}^\top} \hat{\Phi}_\Delta^{(II)} \sqrt{\mathcal{S}^\top} \mathbf{s}). \end{aligned} \quad (\text{B62})$$

To round up the derivation, we rearrange the term containing $\hat{\Phi}_\Delta^{(II)}$ such that the linear transformation acting on Φ_Δ can be readily identified. Combining Eqs. (A8) and (A7) with Eq. (B62) and substituting $\hat{\Phi}_\Delta^{(II)}$ according to Eq. (B59) one finally finds for the second dynamic half-step translation

$$\begin{aligned} \mathbf{d}^{(II)} &= \frac{1}{2} \left(\mathbf{d} + \frac{1}{2} \mathcal{S}^\top \hat{\mathcal{S}} \sqrt{\mathcal{S}} J_R \left[\frac{\Phi_0}{2} \right] J_R^{-1}[\Phi_0] \Phi_\Delta \right) \\ &= \frac{1}{2} \left(\mathbf{d} + \mathcal{S}^{(II) \top} \hat{\mathcal{S}}^{(II)} J_R \left[\frac{\Phi_0}{2} \right] J_R^{-1}[\Phi_0] \Phi_\Delta \right). \end{aligned} \quad (\text{B63})$$

Again, casting Eqs. (B59) and (B63) in the form of a single linear transformation acting on $\mathbf{Y}_\Delta$ one finds

$$\mathbf{Y}_\Delta^{(II)} = A^{(II)} \mathbf{Y}_\Delta, \quad (\text{B64})$$

with

$$A^{(II)} = \frac{1}{2} \begin{bmatrix} J_R \left(\frac{\Phi_0}{2} \right) J_R^{-1}(\Phi_0) & 0 \\ \frac{1}{2} \mathcal{S}^\top \hat{\mathcal{S}} \sqrt{\mathcal{S}} J_R \left[\frac{\Phi_0}{2} \right] J_R^{-1}[\Phi_0] & \mathbb{1} \end{bmatrix}. \quad (\text{B65})$$

c. Linear transformation

With the above framework in place, we can now proceed to construct the composite according to Eq. (B21). First, we note that while composites between triads τ_i to τ_j only involved by step junctions g_l with indices $i \leq l < j$, composites between midstep triads μ_i and μ_j also involve the junction g_j . That is why the composite defined by Eq. (B21) was constructed to span from τ_i to τ_{j+1} ; see Eq. (B1). For composites spanning from μ_i and μ_j , we can thus adopt the same expression as Eq. (B21), except that the half-step transformations have to be applied on the first and last components:

$$\mathbf{Y}_{\Delta}^{(k)} = \sum_{l=i}^j A'_{l,(i,j)} \mathbf{Y}_{\Delta,l}, \quad (\text{B66})$$

with

$$A'_{l,(i,j)} = \begin{cases} A'_{l,(i,j)} = A_{l,(i,j)} A_l^{(\text{II})} & \text{if } l = i, \\ A'_{l,(i,j)} = A_{l,(i,j)} A_l^{(\text{I})} & \text{if } l = j, \\ A'_{l,(i,j)} = A_{l,(i,j)} & \text{otherwise.} \end{cases} \quad (\text{B67})$$

To compute the individual linear transformations $A_{l,(i,j)}$ [Eq. (B22)], one has to make sure to use the right-hand and left-hand static poses for the first and last component, respectively:

$$s_i = s_i^{(\text{II})}, \quad (\text{B68})$$

$$s_j = s_j^{(\text{I})}. \quad (\text{B69})$$

These enter in every $A_{l,(i,j)}$ via the the static rotational and translational composites $\mathcal{S}_{[l+1,j]}$ and $\mathbf{s}_{(l+1,j)}$.

From Eq. (B66), it is straightforward to construct the full transformation matrix A_{Ξ} that maps the full state vector $\tilde{\mathbf{Y}}_{\Delta} \in \mathbb{R}^{6N}$ to the state vector containing the composites $\tilde{\mathbf{E}}_{\Delta} \in \mathbb{R}^{6N}$. Finally, we note that static composites $s^{(k)}$ are defined analogously to Eq. (B29) as the product of underlying static junctions

$$s^{(k)} = s_i^{(\text{II})} s_{[i+1,j-1]} s_j^{(\text{I})}. \quad (\text{B70})$$

APPENDIX C: TRANSLATION CORRECTION

The linear approximation of composites in Eq. (B21) achieves near-perfect agreement for rotational degrees of freedom, yet it struggles to capture translation accurately. In this Appendix, we demonstrate how the high fidelity achieved for the former can be leveraged to improve the latter. We will first provide the updated linear expansion for general composites that augments the formalism derived in Appendix B 1. Subsequently, we show how these changes translate to the hard-constraint and soft-constraint theory.

1. General composites

The excellent results of the expansion for the rotational components suggest that the approximation for the translations may be improved by employing a two-step iteration. First, we use the procedure outlined in Sec. II to calculate the nucleosome-bound compromise state around which fluctuations are centered. In the case of the hard constraint, this is $\bar{\mathbf{C}}_{\Delta}$ as given by Eq. (38); for the soft constraint, we

omitted the equation, but the procedure is analogous to the hard constraint. Next, we use the thus calculated compromise dynamic rotation values to improve the approximation for the translations by expanding rotations around these nonzero compromise states instead of zero, as was done previously.

We denote the single-step dynamic compromise rotations (the rotations in $\bar{\mathbf{C}}_{\Delta}$) as $\Phi_{\Delta,l}^{(\text{C})}$. Fluctuations around the compromise state are then written as

$$\Delta \Phi_{\Delta,l} = \Phi_{\Delta,l} - \Phi_{\Delta,l}^{(\text{C})}. \quad (\text{C1})$$

In the derivation, we will neglect all beyond linear contributions in $\Delta \Phi_{\Delta,l}$, but fully retain contributions of $\Phi_{\Delta,l}^{(\text{C})}$. The dynamic rotation matrix $\mathcal{D}_l$ then reads

$$\begin{aligned} \mathcal{D}_l &= \exp(\hat{\Phi}_{\Delta,l}^{(\text{C})} + \Delta \hat{\Phi}_{\Delta,l}) \\ &\approx \exp(\hat{\Phi}_{\Delta,l}^{(\text{C})}) \exp(\text{hat}[J_R(\Phi_{\Delta,l}^{(\text{C})}) \Delta \Phi_{\Delta,l}]), \end{aligned} \quad (\text{C2})$$

where the right Jacobian J_R [see Eqs. (A13) and (A14)] enables the splitting of the exponential. Introducing the shorthand notations

$$\mathcal{D}_l^{(\text{C})} \equiv \exp(\hat{\Phi}_{\Delta,l}^{(\text{C})}), \quad (\text{C3})$$

$$\delta_l \equiv J_R(\Phi_{\Delta,l}^{(\text{C})}) \Delta \Phi_{\Delta,l}, \quad (\text{C4})$$

we further expand Eq. (C2) as

$$\mathcal{D}_l \approx \mathcal{D}_l^{(\text{C})} (\mathbb{1} + \hat{\delta}_l). \quad (\text{C5})$$

Next, we rewrite Eq. (B17) as

$$\begin{aligned} \mathbf{d}_{(i,j)} &= \mathcal{S}_{[i,j]}^{\text{T}} \sum_{l=i}^j \left[\prod_{k=i}^{l-1} \mathcal{S}_k \mathcal{D}_k \right] \mathcal{S}_l [(\mathcal{D}_l \mathbf{s}_{(l+1,j)} + \mathbf{d}_l) - \mathbf{s}_{(l+1,j)}] \\ &\approx \mathcal{S}_{[i,j]}^{\text{T}} \sum_{l=i}^j \left[\prod_{k=i}^{l-1} \mathcal{S}_k \mathcal{D}_k^{(\text{C})} (\mathbb{1} + \hat{\delta}_k) \right] \\ &\quad \times \mathcal{S}_l [(\mathcal{D}_l^{(\text{C})} - \mathbb{1}) \mathbf{s}_{(l+1,j)} + \mathcal{D}_l^{(\text{C})} \hat{\delta}_l \mathbf{s}_{(l+1,j)} + \mathbf{d}_l]. \end{aligned} \quad (\text{C6})$$

Expanding to linear order in the $\hat{\delta}$, introducing the shorthand notation

$$\mathcal{R}_l^{(\text{C})} \equiv \mathcal{S}_l \mathcal{D}_l^{(\text{C})}, \quad (\text{C7})$$

and using the rearrangement of summations

$$\sum_{l=i}^j \sum_{k=i}^{l-1} x_{l,k} = \sum_{k=i}^{j-1} \sum_{l=k+1}^j x_{l,k}, \quad (\text{C8})$$

the leading product of rotations can be rewritten as

$$\begin{aligned} \mathcal{S}_{[i,j]}^{\text{T}} \sum_{l=i}^j \left[\prod_{k=i}^{l-1} \mathcal{S}_k \mathcal{D}_k^{(\text{C})} (\mathbb{1} + \hat{\delta}_k) \right] \mathcal{S}_l \\ = \mathcal{S}_{[i,j]}^{\text{T}} \sum_{k=i}^{j-1} \mathcal{R}_{[i,k]}^{(\text{C})} \hat{\delta}_k \sum_{l=k+1}^j \mathcal{R}_{[k+1,l-1]}^{(\text{C})} \mathcal{S}_l \\ + \mathcal{S}_{[i,j]}^{\text{T}} \sum_{l=i}^j \mathcal{R}_{[i,l-1]}^{(\text{C})} \mathcal{S}_l. \end{aligned} \quad (\text{C9})$$

Multiplying the preceding expression by the trailing term in Eq. (C6) produces a constant term; a term linear in $\mathbf{d}_l$; two

terms linear in $\widehat{\delta}_l$; and higher-order terms in $\widehat{\delta}_l$ and $\mathbf{d}_l$, which are omitted.

The constant term and the term linear in $\mathbf{d}_l$ read, respectively,

$$\mathcal{S}_{[i,j]}^\top \sum_{l=i}^j \mathcal{R}_{[i,l-1]}^{(C)} \mathcal{S}_l (\mathcal{D}_l^{(C)} - \mathbb{1}) \mathbf{s}_{(l+1,j)}, \quad (\text{C10})$$

$$\mathcal{S}_{[i,j]}^\top \sum_{l=i}^j \mathcal{R}_{[i,l-1]}^{(C)} \mathcal{S}_l \mathbf{d}_l. \quad (\text{C11})$$

Combining the two terms linear in $\widehat{\delta}_l$, one finds

$$-\mathcal{S}_{[i,j]}^\top \sum_{l=i}^{j-1} \mathcal{R}_{[i,l]}^{(C)} (\widehat{\lambda}_l + \widehat{s}_{(l+1,j)}) \delta_l, \quad (\text{C12})$$

$$A_{l,(i,j)}^{(C)} = \begin{bmatrix} \mathcal{S}_{[l+1,j]}^\top & 0 \\ -\mathcal{S}_{[i,j]}^\top \mathcal{R}_{[i,l]}^{(C)} (\widehat{\lambda}_l + \widehat{s}_{(l+1,j)}) J_R(\Phi_{\Delta,l}^{(C)}) & \mathcal{S}_{[i,j]}^\top \mathcal{R}_{[i,l-1]}^{(C)} \mathcal{S}_l \end{bmatrix}, \quad (\text{C15})$$

and the constant correction is given by

$$\mathbf{a}_{(i,j)} = \begin{bmatrix} 0 \\ \mathcal{S}_{[i,j]}^\top \sum_{l=i}^{j-1} \mathcal{R}_{[i,l]}^{(C)} [(\mathbb{1} - \mathcal{D}_l^{(C)\top}) \mathbf{s}_{(l+1,j)} + (\widehat{\lambda}_l + \widehat{s}_{(l+1,j)}) J_R(\Phi_{\Delta,l}^{(C)}) \Phi_{\Delta,l}^{(C)}] \end{bmatrix}. \quad (\text{C16})$$

APPENDIX D: SOFT-CONSTRAINT COORDINATE TRANSFORMATION

Here we derive the form of the linear transformation Eq. (47) used in the soft-constraint formalism to relate the coordinates of midstep triad fluctuations $\bar{\mathbf{Z}}_\Delta$ to the coordinates of junction fluctuations $\bar{\mathbf{Y}}_\Delta$. In this context, we assume the histone core to be fixed in space and assign innate positions, $\mu_0^{(k)}$, for the midstep triads. Their instantaneous pose $\mu^{(k)}$ is allowed to deviate within a small range around $\mu_0^{(k)}$, which is expressed by the fluctuations $\mu_\Delta^{(k)}$, such that the instantaneous midstep triad pose is given by

$$\mu^{(k)} = \mu_0^{(k)} \mu_\Delta^{(k)}, \quad (\text{D1})$$

with fluctuations parametrized by

$$\mathcal{P}(\mu_\Delta^{(k)}) = \mathbf{Z}_\Delta^{(k)} = \begin{bmatrix} \Theta_\Delta^{(k)} \\ \mathbf{z}_\Delta^{(k)} \end{bmatrix}. \quad (\text{D2})$$

Junctions spanning between these bound midstep triads are split into static and dynamic components

$$\mathbf{g}^{(k)} = \mathbf{s}^{(k)} \mathbf{d}^{(k)}, \quad (\text{D3})$$

with dynamic variables

$$\mathcal{P}(\mathbf{d}^{(k)}) = \mathbf{Y}_\Delta^{(k)} = \begin{bmatrix} \Phi_\Delta^{(k)} \\ \mathbf{d}^{(k)} \end{bmatrix}. \quad (\text{D4})$$

The connection between the two is established by

$$\mathbf{g}^{(k)} = \mu^{(k)-1} \mu^{(k+1)}. \quad (\text{D5})$$

This expression implies that the set of junction variables $\mathbf{Y}_\Delta^{(k)}$ is fully determined by the set of midstep triad parameters $\mathbf{Z}_\Delta^{(k)}$.

where

$$\lambda_l = \sum_{k=l+1}^j \mathcal{R}_{[l+1,k-1]}^{(C)} \mathcal{S}_k (\mathcal{D}_k^{(C)} - \mathbb{1}) \mathbf{s}_{(k+1,j)}. \quad (\text{C13})$$

Since $\delta_l = J_R(\Phi_{\Delta,l}^{(C)})(\Phi_{\Delta,l} - \Phi_{\Delta,l}^{(C)})$, Eq. (C12) gives rise to a constant term and a term linear in the rotation components $\Phi_{\Delta,l}$. Collecting all the terms, one finds that the linear combination Eq. (B21) is replaced by

$$\mathbf{Y}_{\Delta,(i,j)} \approx \sum_{l=i}^j A_{l,(i,j)}^{(C)} \mathbf{Y}_{\Delta,l} + \mathbf{a}_{(i,j)}, \quad (\text{C14})$$

where the corrected linear transformation reads

In this section, we will derive an approximate expression for the $\mathbf{Y}_\Delta^{(k)}$ as a linear combination of the $\mathbf{Z}_\Delta^{(k)}$.

Solving Eq. (D5) for $\mathbf{d}^{(k)}$ one has

$$\mathbf{d}^{(k)} = \mathbf{s}^{(k)-1} \mu_\Delta^{(k)-1} \mu_0^{(k)-1} \mu_0^{(k+1)} \mu_\Delta^{(k+1)}, \quad (\text{D6})$$

where the left-hand side corresponds to $\mathbf{Y}_\Delta^{(k)}$ and the right-hand side features $\mathbf{Z}_\Delta^{(k)}$ and $\mathbf{Z}_\Delta^{(k+1)}$ via $\mu^{(k)}$ and $\mu^{(k+1)}$, respectively. To abbreviate the notation, we define

$$\mu_0^{(k)-1} \mu_0^{(k+1)} \equiv \mathbf{v}_0^{(k)} = \begin{bmatrix} \Upsilon_0^{(k)} & \mathbf{v}_0^{(k)} \\ \mathbf{0}^\top & 1 \end{bmatrix}. \quad (\text{D7})$$

We will first work out the rotational component of Eq. (D6), which reads

$$\exp(\widehat{\Phi}_\Delta^{(k)}) = \mathcal{S}^{(k)\top} \exp(-\widehat{\Theta}_\Delta^{(k)}) \Upsilon_0^{(k)} \exp(\widehat{\Theta}_\Delta^{(k+1)}). \quad (\text{D8})$$

Assuming $\widehat{\Theta}_\Delta^{(k)}$ to be sufficiently small to justify linear expansion and applying the same reasoning as for the construction of composite junctions (see Sec. B 1), one has

$$\exp(\widehat{\Phi}_\Delta^{(k)}) \approx \mathcal{S}^{(k)\top} \Upsilon_0^{(k)} \exp(\widehat{\Theta}_\Delta^{(k+1)} - \Upsilon_0^{(k)\top} \widehat{\Theta}_\Delta^{(k)} \Upsilon_0^{(k)}). \quad (\text{D9})$$

To arrive at the final expression, this product of three consecutive rotations needs to be unified into one single rotation. Since the two leading rotations are constant, we are allowed to represent them collectively by a single rotation vector $\Psi^{(k)}$ as

$$\mathcal{S}^{(k)\top} \Upsilon_0^{(k)} \equiv \exp \widehat{\Psi}^{(k)}. \quad (\text{D10})$$

Applying the procedure that allows us to transform $\mathbf{X}$ coordinates into $\mathbf{Y}$ coordinates in Sec. B 2 (developed in

Ref. [88]), we can use Eq. (A13) to write

$$\Phi_{\Delta}^{(k)} = \Psi^{(k)} + J_R^{-1}(\Psi^{(k)})[\Theta_{\Delta}^{(k+1)} - \Upsilon_0^{(k)\top} \Theta_{\Delta}^{(k)}], \quad (\text{D11})$$

which is the final expression for the rotational component.

We proceed by writing out the translational component of Eq. (D6) using

$$\mu^{(k)-1} = \begin{bmatrix} e^{-\hat{\Theta}_{\Delta}^{(k)}} & -e^{-\hat{\Theta}_{\Delta}^{(k)}} \mathbf{z}_{\Delta}^{(k)} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \approx \begin{bmatrix} (\mathbb{1} - \hat{\Theta}_{\Delta}^{(k)}) & -\mathbf{z}_{\Delta}^{(k)} \\ \mathbf{0}^{\top} & 1 \end{bmatrix}, \quad (\text{D12})$$

and ignoring all beyond linear order terms in fluctuating components, to arrive at

$$\begin{aligned} \mathbf{d}^{(k)} &\approx \mathcal{S}^{(k)\top} (\Upsilon_0^{(k)} \mathbf{z}_{\Delta}^{(k+1)} + \mathbf{v}_0^{(k)} - \hat{\Theta}_{\Delta}^{(k)} \mathbf{v}_0^{(k)} - \mathbf{z}_{\Delta}^{(k)} - \mathbf{s}^{(k)}) \\ &= \mathcal{S}^{(k)\top} (\mathbf{v}_0^{(k)} - \mathbf{s}^{(k)}) \\ &\quad + \mathcal{S}^{(k)\top} (\Upsilon_0^{(k)} \mathbf{z}_{\Delta}^{(k+1)} - \mathbf{z}_{\Delta}^{(k)} + \hat{\mathbf{v}}_0^{(k)} \Theta_{\Delta}^{(k)}). \end{aligned} \quad (\text{D13})$$

In the second line, we separated constant and fluctuating terms and identified $-\hat{\Theta}_{\Delta}^{(k)} \mathbf{v}_0^{(k)} = \hat{\mathbf{v}}_0^{(k)} \Theta_{\Delta}^{(k)}$ [via Eq. (A7)].

The results of Eqs. (D11) and (D13) can be summarized as

$$\mathbf{Y}_{\Delta}^{(k)} \approx B_{-}^{(k)} \mathbf{Z}_{\Delta}^{(k)} + B_{+}^{(k)} \mathbf{Z}_{\Delta}^{(k+1)} + \mathbf{b}^{(k)}, \quad (\text{D14})$$

with

$$B_{-}^{(k)} = \begin{bmatrix} -J_R^{-1}(\Psi^{(k)}) \Upsilon_0^{(k)\top} & 0 \\ \mathcal{S}^{(k)\top} \hat{\mathbf{v}}_0^{(k)} & -\mathcal{S}^{(k)\top} \end{bmatrix}, \quad (\text{D15})$$

$$B_{+}^{(k)} = \begin{bmatrix} J_R^{-1}(\Psi^{(k)}) & 0 \\ 0 & \mathcal{S}^{(k)\top} \Upsilon_0^{(k)} \end{bmatrix}, \quad (\text{D16})$$

$$\mathbf{b}^{(k)} = \begin{bmatrix} \Psi^{(k)} \\ \mathcal{S}^{(k)\top} (\mathbf{v}_0^{(k)} - \mathbf{s}^{(k)}) \end{bmatrix}. \quad (\text{D17})$$

This single-step transformation can then be consolidated into the transformation of the full state vectors

$$\bar{\mathbf{Y}}_{\Delta}^{\text{M}} = B \bar{\mathbf{Z}}_{\Delta} + \bar{\mathbf{b}}, \quad (\text{D18})$$

with

$$B = \begin{bmatrix} B_{-}^{(0)} & B_{+}^{(0)} & 0 & \dots & 0 & 0 \\ 0 & B_{-}^{(1)} & B_{+}^{(1)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & B_{+}^{(N_{\text{M}}-2)} & 0 \\ 0 & 0 & 0 & \dots & B_{-}^{(N_{\text{M}}-1)} & B_{+}^{(N_{\text{M}}-1)} \end{bmatrix}, \quad (\text{D19})$$

and $\bar{\mathbf{b}} \in \mathbb{R}^{6N_{\text{M}}}$ defined as the state vector collecting all $\mathbf{b}^{(k)}$ into a single vector

$$\bar{\mathbf{b}}^{\top} = [\mathbf{b}^{(0)\top} \mathbf{b}^{(1)\top} \dots \mathbf{b}^{(N_{\text{M}}-1)\top}]. \quad (\text{D20})$$

APPENDIX E: SOFT-CONSTRAINT PARAMETER COMPUTATION

In this Appendix, we show how the parameters of the soft-constraint model, i.e., the intrinsically preferred poses $\mu_0^{(k)}$ introduced in Eq. (42), and the stiffness matrix K [see

Eq. (45)], can be computed from a thermal ensemble of nucleosome configurations. We will assume that the innate structure and elasticity of the associated DNA, i.e., the ground state s_i and the innate stiffness matrix M , are known—implying likewise knowledge of $s^{(k)}$ and $\tilde{M}_{\text{M}}$. Further, we assume that fluctuations of bound midstep triads $\mu^{(k)}$ are close to Gaussian, such that the model is applicable. The main challenge in this computation is that almost all components of the model, including the transformation from $\bar{\mathbf{Z}}_{\Delta}$ to $\bar{\mathbf{Y}}_{\Delta}$ and the effective stiffness matrix K_{comb} , depend on the a priori unknown $\mu_0^{(k)}$. We will first show how to compute K if the $\mu_0^{(k)}$ are known and subsequently outline the fitting procedure to determine $\mu_0^{(k)}$.

1. Computing stiffness with known innate poses

Given $\mu_0^{(k)}$, the calculation of K is straightforward. Applying Eq. (42) one finds the dynamic coordinates $\bar{\mathbf{Z}}_{\Delta}$ for each sample as

$$\bar{\mathbf{Z}}_{\Delta} = \text{stack}\{\mathbf{Z}_{\Delta}^{(0)}, \dots, \mathbf{Z}_{\Delta}^{(N_{\text{M}})}\}, \quad (\text{E1})$$

with

$$\mathbf{Z}_{\Delta}^{(k)} = \mathcal{P}(\mu_0^{(k)-1} \mu^{(k)}). \quad (\text{E2})$$

Observing Eq. (48) one finds that K_{comb} can be deduced via covariance inversion

$$K_{\text{comb}} = \langle (\bar{\mathbf{Z}}_{\Delta} - \langle \bar{\mathbf{Z}}_{\Delta} \rangle) (\bar{\mathbf{Z}}_{\Delta} - \langle \bar{\mathbf{Z}}_{\Delta} \rangle)^{\top} \rangle^{-1}, \quad (\text{E3})$$

where the average extends over all samples. In practice, the covariance matrix in Eq. (E3) may be ill-conditioned, especially when based on limited or highly correlated sampling. Such ill-conditioning can give rise to small or even negative eigenvalues due to numerical noise, in which case regularization may be required.

Since B and $\bar{\mathbf{b}}$ depend only on the static components $s^{(k)}$ and $\mu_0^{(k)}$ (see Appendix D for details), they can be readily evaluated. From Eq. (49) it then follows that

$$K = K_{\text{comb}} - B^{\top} \tilde{M}_{\text{M}} B. \quad (\text{E4})$$

Moreover, Eq. (51) can be applied to determine the compromise pose $\bar{\alpha}$ for the bound midstep triads. We note that according to Eq. (48) one also has

$$\bar{\alpha} = \langle \bar{\mathbf{Z}}_{\Delta} \rangle, \quad (\text{E5})$$

which defines the self-consistency relation based on which the unknown $\mu_0^{(k)}$ can be computed.

2. Fitting scheme

For arbitrary $\mu_0^{(k)}$, the self-consistency relation Eq. (E5) will generally not be satisfied. However, starting from an initial guess, one can conduct a gradient-descent procedure aimed at minimizing $\|\Delta\|^2$ with

$$\Delta = \bar{\alpha} - \langle \bar{\mathbf{Z}}_{\Delta} \rangle, \quad (\text{E6})$$

that yields consistent $\mu_0^{(k)}$ in just a few iteration steps. Since all involved operations are smooth and differentiable, the gradient can be calculated explicitly. However, in practice, it is sufficient to apply a block-diagonal Gauss-Newton scheme

TABLE I. Sequences of overlapping oligomers used for free DNA simulations. The overlaps are over- and underlined.

Index	Sequence
1	ATCAATATCCACCTGCAGATACTACCAAAAGT
2	ACCAAAAGTGTATTTGGAAACTGCTCCATCAA
3	CTCCATCAAAAGGCATGTTTCAGCTGGAATCCA
4	TGGAATCCAGCTGAACATGCCTTTTGATGGAG
5	TTGATGGAGCAGTTTCCAAATACACTTTTGGT
6	ACTTTTGGTAGTATCTGCAGGTGGATATTGAT

[99] that updates $\mu_{0,\{t\}}^{(k)}$ from iteration step t to $t + 1$ as

$$\mu_{0,\{t+1\}}^{(k)} = \mu_{0,\{t\}}^{(k)} \mathcal{P}^{-1}(-\gamma \Delta^{[t]}), \quad (\text{E7})$$

where $\gamma \in (0, 1]$ is a damping factor chosen small enough to avoid potential instabilities.

For the initial guess, average poses from the provided ensemble are usually sufficient. Here, the rotational average may, for example, be computed via the method provided in Ref. [99].

3. Symmetrization of K

To impose symmetry under wrapping-sense inversion, we average the forward- and backward-sense covariance matrices. The forward-sense covariance is given by the inverse of K . Its sense-inverted counterpart is obtained by applying the two defining operations of sense inversion. First, each 6×6 interaction block is moved from position (i, j) to $(145 - i, 145 - j)$. Second, sense inversion flips the sign of all coordinates except the first and the fourth; these sign changes are applied to the covariance by sign-flipping the corresponding rows and columns within each 6×6 block. Finally, the average of the forward- and backward-sense covariances is inverted to yield the wrapping-sense-symmetrized stiffness matrix K_{sym} . Note that this symmetrization does not guarantee positive semi-definiteness, such that small positive eigenvalues may become negative, and a further regularization step may be required.

APPENDIX F: MOLECULAR DYNAMICS SIMULATIONS

The matrix encoding the fluctuation restrictions imposed by the histone core on the nucleosomal DNA, K , is estimated from atomistic MD simulations, which we describe here.

1. Simulation protocol

MD simulations for the nucleosome core particle (NCP) with 147 bp DNA (PDB ID: 1KX5 [2]) were performed for a duration of 250 ns. In addition, the sequence of the nucleosome was divided into overlapping segments of 32 bp with 9 bp overlaps each, resulting in the oligomers listed in Table I.

The preparation of the systems for the MD simulation involved placing the structures in a periodic dodecahedron box, with the boundaries at least 1.0 nm away from the system, followed by the addition of water molecules. To mimic experimental conditions [136] and neutralize the system, 50 mM NaCl was added to the solvent by replacing water molecules

with ions. The interactions between atoms are described using the force field AMBER14sb-parmbnc1 [137,138] in combination with the TIP3P water model [139]. This specific force field combination was chosen because it covers the topologies for both amino acids and nucleotides and provides good representations of the static and dynamic properties of DNA under a wide range of conditions [138]. For nonbonded interactions, including both van der Waals and electrostatic interactions, we used a cutoff of 1.1 nm. Long-range electrostatic interactions were handled using the particle mesh Ewald method [140,141] with a grid spacing of 0.12 nm. To eliminate unfavorable interactions before starting the production MD runs, we performed energy minimization using steepest descent. By applying position restraints on the heavy atoms of the DNA (and nucleosome in case of the NCP) with a force constant in each direction of 1000 kJ/mol nm² and performing 0.1 ns of MD at a temperature of 298 K and a pressure of 1 bar, we relaxed the water and ions surrounding the initial structures.

After preparation, we performed MD simulations with different initial starting conditions, assigning new random starting velocities drawn from the Maxwell-Boltzmann distribution at 298 K. All simulations were carried out with GROMACS, version 2020.4 [142,143] in a locally maintained cluster at the University of Amsterdam, using the leap-frog integration scheme and a time step of 2 fs, with LINCS [144] applied to constrain the proteins and SETTLE [145] to constrain water bonds. All simulations were conducted in the isothermal-isobaric ensemble at a pressure of 1 bar, using the v-rescale thermostat [146] and the isotropic Parrinello-Rahman barostat [147,148].

2. MD trajectory analysis and calculation of bound midstep triads

MD-generated trajectories were preprocessed, e.g., translation of molecule in a single unit cell, with Python packages MDTraj [104]. Further, we employed MDNA [105] to map atomistic topologies onto the rigid-base-pair description, where each base pair is captured by a single triad τ_i [Eq. (1)]. Triad definitions follow the Curves+ convention [90] and we used the Kabsch algorithm [149] to fit canonical nucleobases onto the trajectory observed structures.

The six free DNA simulations were used to calculate the innate properties of the nucleosome-associated DNA, i.e., the ground state s_i and the stiffness matrix M [57,88,91,150].

For the evaluation of the nucleosome trajectory, we started by aligning (translation and rotation) the histone core structures over all snapshots relative to the first snapshot using the Kabash algorithm [149]. This is necessary since the intrinsically preferred poses $\mu_0^{(k)}$ are defined in the material frame of the nucleosome core. To measure midstep triad fluctuations relative to the nucleosome core, one has either to define a nucleosome reference frame [151] or superimpose all nucleosome structures. Here, we elected to employ the latter path to minimize the influence of fluctuations of individual atom groups. Employing Eq. (11), we then calculated the midstep triads $\mu^{(k)}$ corresponding to the 28 binding sites specified in Eq. (12). Given the ensemble of midstep triads $\mu^{(k)}$, we applied the procedure outlined in Appendix E to compute $\mu_0^{(k)}$ and K . Further required steps, i.e., regularization and wrapping-sense symmetrization, are described in Sec. III B and Appendix E 3.

APPENDIX G: SITE-SPECIFIC ADSORPTION

To compare free energies of nucleosomes in different wrapping states, one needs access to the adsorption energies of the binding sites of the DNA to the histone octamer. The individual strengths of these sites can be expected to vary significantly, as it is known from crystal structures that the contacts between the histone octamer and the DNA are very diverse [1]. Importantly, however, the contacts are made mostly via the DNA backbone phosphates rather than bp-specific contacts. This suggests that individual adsorption energies are independent of the bp sequence and sites that are symmetrically related by a rotation around the nucleosome dyad have the same strength. Our model allows this assumption to be tested.

To achieve this, we used the experimental data reported in Ref. [89] where DNA was mechanically unzipped with a constant force of 28 pN into a well-positioned 601 nucleosome, either from the left (forward) or right (reverse) direction, mapping the dwell times of the unzipping fork at bp resolution. Whenever the unzipping fork encountered binding sites, it was halted for extended times at the corresponding location [see Fig. 1(b) in Ref. [89] and Fig. 11(a)]. Because 28 pN is well above the 13–16 pN required to separate naked 601 DNA, most of the dwell time originates from breaking histone-DNA contacts rather than from bp melting. They further showed that halting the fork before it reached the dyad allowed the octamer to re-wrap spontaneously when the load was released, demonstrating that each intermediate can exchange reversibly with its neighbours and is therefore close to thermodynamic equilibrium under the clamp. Under these conditions, each paused state can be viewed as in quasiequilibrium. The thermally activated escape rate over a single barrier, whose height is set by the local histone-DNA binding strength, can then be determined from the Kramers' rate:

$$E_{\text{ads},i} = k_B T \ln \left(\frac{\tau_i}{\tau_0} \right) - \theta_1 E_{\text{el},i} + \theta_2 F \Delta x, \quad (\text{G1})$$

where $E_{\text{ads},i}$ is the site-specific adsorption energy, $E_{\text{el},i}$ is the local elastic bending cost, $F \Delta x$ is the work done by the force clamp, τ_i is the dwell time derived from Ref. [89], Fig. 1(b), for each DNA-histone contact site, and τ_0 is a site-independent microscopic time constant. Finally, θ_1 and θ_2 are “load distribution factors” [152] that describe how much the barrier against unwrapping of a site is biased by an internal load, the bending of the involved DNA stretch, and an external load, the force applied to the two DNA strands. We assume $0 \leq \theta_i \leq 1$, but without knowledge of how the unwrapping process proceeds in microscopic detail, their values cannot be restricted further. Of importance to us is only θ_1 , which we will estimate below via a symmetry argument.

Remarkably, the dwell time peaks are vastly different when comparing unzipping from the left and right; see Fig. 11(a). This raises the question whether the asymmetry can be explained by an asymmetry in the free energy of the two DNA halves of 601 or whether one needs to further assume adsorption energies to be sequence-dependent as well.

To extract dwell times per phosphate, τ_i , we digitized the data in Fig. 1(b) of Ref. [89] for both unzipping directions separately up to the dyad, as post-dyad dwell times are unre-

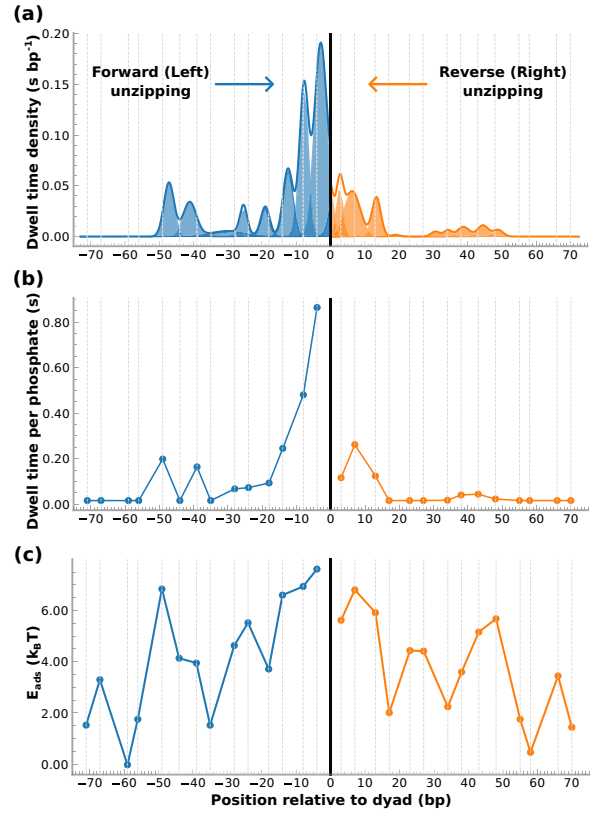


FIG. 11. Analysis of data from nucleosome unzipping experiments [89] used to map histone-DNA interactions and to determine adsorption energies. (a) Probability density of dwell times recorded during force-induced unzipping. Forward unzipping is indicated in blue, and reverse unzipping is indicated in orange. The signal is decomposed into a sum of individual Gaussians, each being the contribution of a single binding site. (b) Dwell times obtained by integrating the area under each Gaussian. (c) Effective adsorption strengths are estimated at the 28 binding sites (indicated by vertical dashed lines). Note that the signal is highly asymmetric relative to the dyad, reflecting the asymmetric nature of the 601 sequence (see text).

liable due to nucleosome instability and histone dissociation [89]. We identified the peaks, and Gaussian functions were fitted to each peak to decompose the signal; see Fig. 11(a). The integrated area under each Gaussian provides the dwell time per phosphate binding site and is shown in Fig. 11(b). For sites lacking peaks, we assigned 1/60 s, based on the 60 Hz low-pass filter used in the experiment.

Next, we estimated $E_{\text{ads},i}$ from Eq. (G1) by obtaining $E_{\text{el},i}$ from our nucleosome free-energy calculation. We absorb the work done by the force, $\theta_2 F \Delta x$, along with the term $-k_B T \ln(\tau_0)$ into a single offset factor E_{offset} . Since the force clamp introduces only a linear term to the underlying potential, it lowers every barrier by the same mechanical work $F \Delta x$ while leaving the local width and shape essentially unchanged. This new factor $E_{\text{offset}} = -k_B T \ln(\tau_0) + \theta_2 F \Delta x$ will be calibrated based on the observed 601 nucleosome breathing behavior after estimating $E_{\text{ads},i}$. For now, we set it to zero without affecting the relative energy landscape of E_{ads} across the sites.

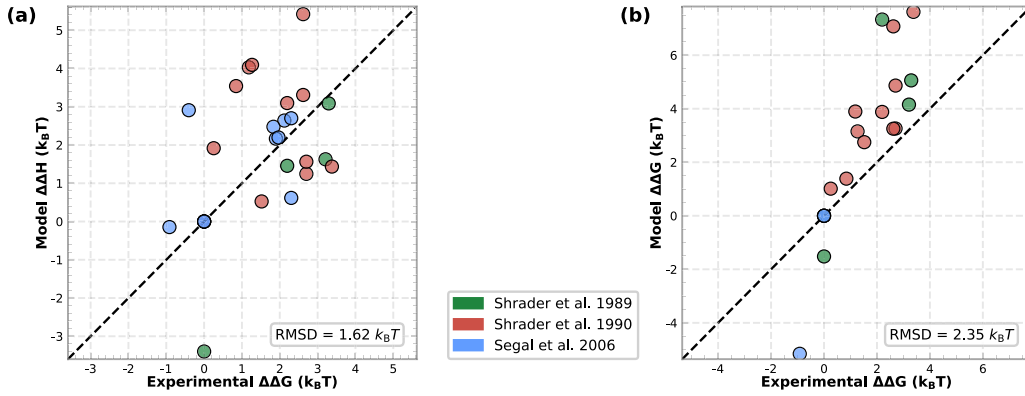


FIG. 12. (a) Comparison between experimental free energy, $\Delta\Delta G$, and model enthalpy of the tetramer, $\Delta\Delta H$. (b) Free-energy comparison between experiment and model, assuming for the model the fully wrapped nucleosome. Sequences shorter than 147 bp are excluded (8 in total).

We calculated the $E_{\text{ads},i}$ values of the 28 sites for various values of the load distribution factor θ_1 in the range $0 \leq \theta_1 \leq 1$. For each θ_1 value, we determined the total mean-squared difference between $E_{\text{ads},i}$ values of sites that are related by the nucleosome symmetry, i.e., sites with equal bp displacement from the dyad. The lowest value is obtained for $\theta_1 = 1$, which leads to a plot with a striking left-right symmetry; see Fig. 11(c). To understand why the full DNA free energy leads to the best value is outside the scope of the present work, but it deserves attention in a future study.

Recognizing the symmetry of the histone octamer and assuming that the DNA-histone association energies are sequence-independent (dominated by phosphate-histone interactions rather than bp-specific contacts), we symmetrize the adsorption energies by averaging corresponding values in Fig. 11(c) and its mirror image (relative to the dyad). Furthermore, the nucleosome features 28 distinct phosphate binding locations along the DNA backbone, but adjacent phosphate pairs are sufficiently close (spaced by ~ 5 bp) that their interactions can be consolidated into 14 effective binding sites, corresponding to the major histone-DNA contact points observed in structural studies [1,2]. The final adsorption energies for these 14 effective binding sites are shown in Fig. 6.

APPENDIX H: DISENTANGLING CONTRIBUTIONS

In Sec. III D, we compare our model to experimental data on competitive nucleosome reconstitution and on breathing, using the full free energy in our calculation. Heterogeneous site-specific absorption energies of binding sites between DNA and histone octamer were found to be necessary to account for experimental breathing data and are thus accounted for in Fig. 8. To disentangle which contributions lead to the excellent agreement between data and theory, we show here the corresponding plots where certain terms are missing.

Figure 12(a) provides a comparison between reconstitution data and our model when fluctuational contributions to the free energy are neglected, and Fig. 12(b) depicts the free energy when the full octamer, instead of just the tetramer, is considered. Evidently, the agreement between model and data is much worse than in Fig. 7 (see main text for the details).

Figures 13(a) and 13(b) provide the relative occupancies for nucleosome breathing, analogous to Fig. 8. In Fig. 8(a), only the enthalpic contribution is accounted for in our model and leads to the “spring-loaded latch” mechanism, first discovered in the experiment [29]. Adding the entropy in Fig. 8(b), this effect disappears. These plots, however, assume a homogeneous strength for the binding sites. Using the non-homogeneous values shown in Fig. 14, one obtains the three distributions displayed in Fig. 8. This full model recovers the “spring-loaded latch” effect, as there appears a stable, partially unwrapped and highly asymmetric state for intermediate offset values. Specifically, we chose $E_{\text{offset}} = 11.5 k_B T$ leading to $\langle E_{\text{ads}} \rangle = 15.5 k_B T$; see Fig. 14.

APPENDIX I: MONTE CARLO SIMULATIONS

We employed Monte Carlo simulations to generate configurations of a 147 bp DNA nucleosome using the RBP model. Every dinucleotide is captured by a single triad [see Eq. (1)]. Relative orientations between adjacent bp are parametrized by six helical degrees of freedom—tilt, roll, twist, shift, slide, and rise, which are split into intrinsic shape and deviations thereof. The latter are associated with an elastic energetic cost, as given by Eq. (10). We use hybrid parameters [100] for which intrinsic shape is derived from crystallographic data [57] and stiffness from atomistic molecular dynamics simulations [91]. Move acceptance is determined via the Metropolis algorithm.

The 28 DNA phosphate-histone contacts detailed in Sec. II A and shown in Fig. 2 were imposed by freezing the associated midstep frames. Four constraint-preserving moves were used to sample the configurational space: (i) random rotations and translations of single triads, (ii) midstep triad moves that rotate and translate the two neighboring triads such that the corresponding midstep triads remain unchanged, (iii) crankshaft moves that rotate segments involving several consecutive bp about the axis spanned between the two outermost bp in the sequence, and (iv) cluster translations that move a DNA segment (again spanning several bp) along one of the randomly chosen translational degrees of freedom. The latter two cluster moves are introduced to accelerate the decorrelation between successive configurations.

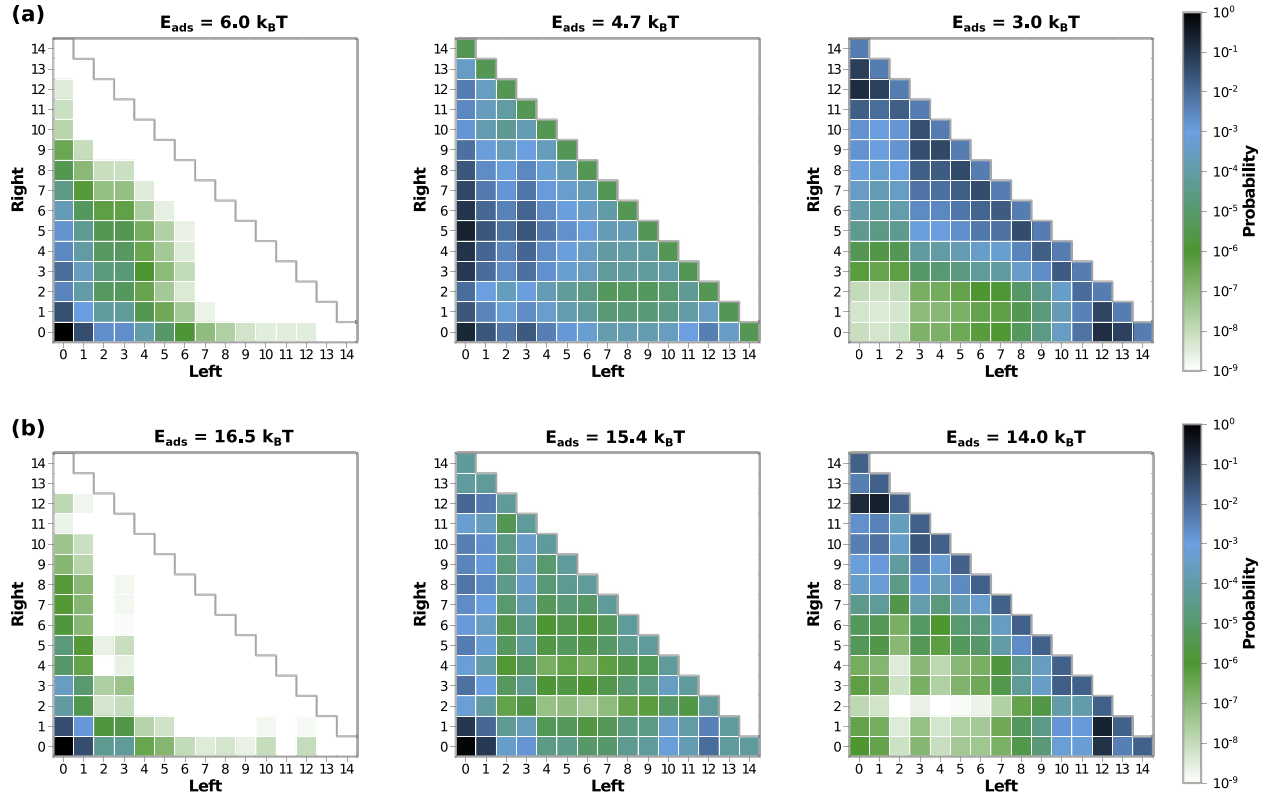


FIG. 13. Relative occupancies of all unwrapping states for a simplified version of the model, which assumes constant binding strength. (a) Enthalpy only and (b) full free energy. Analogous plots including site-specific absorption energies and fluctuational contributions to the free energy are shown in Fig. 8.

1. Thermodynamic integration

To compute the free-energy landscape for a nucleosome across the promoter region (-500 bp to $+1000$ bp around the transcription start site) of the YAL002W gene in *S. cerevisiae*, we employed TI [101,102]. TI exploits the fact that the free energy is a state function, making the free energy change between two sequences path-independent in the limit of equilibrium sampling. The free-energy difference (ΔF) between any two sequences, A and B , can be calculated by constructing a continuous, reversible path, parametrized by λ , between them. Here, we define the λ -dependent Hamiltonian as

$$H(\lambda, \tilde{\mathbf{Y}}_\Delta) = (1 - \lambda)E_A(\tilde{\mathbf{Y}}_\Delta) + \lambda E_B(\tilde{\mathbf{Y}}_\Delta), \quad (11)$$

where $E_A(\tilde{\mathbf{Y}}_\Delta)$ represents the elastic deformation energy [as defined in Eq. (10)] of sequence A with intrinsic deformation $s_{i,A}$ and stiffness matrix M_A . Similarly, $E_B(\tilde{\mathbf{Y}}_\Delta)$, $s_{i,B}$, and M_B are the corresponding quantities for sequence B . This path ensures continuity between the endpoints A and B .

Using the relation

$$\Delta F = (F_B - F_A) = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda, \quad (12)$$

where $\langle \cdot \rangle_\lambda$ denotes the ensemble average over configurations at fixed λ [101], we evaluate the free-energy difference between any two sequences. In our implementation, state A corresponds to the first sequence of the selected promoter

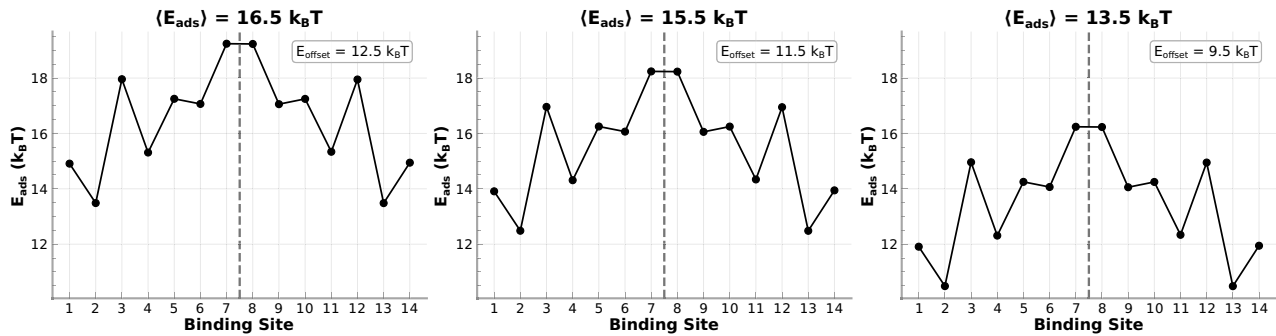


FIG. 14. Site-specific adsorption energies, for different values of the offset defined as $E_{\text{offset}} = -k_B T \ln \tau_0 + \theta_2 F \Delta x$. These data were extracted from Ref. [89].

region, a 147 bp sequence starting at -500 bp, which serves as the reference for all sequences generated by sliding a 147 bp window across this promoter region, yielding $1501 - 147 = 1354$ distinct B sequences. While sampling nucleosome configurations for a particular λ the derivative in Eq. (11) simply reads

$$\frac{\partial H(\lambda)}{\partial \lambda} = E_B(\bar{\mathbf{Y}}_\Delta) - E_A(\bar{\mathbf{Y}}_\Delta), \quad (13)$$

such that the corresponding thermal average can be readily evaluated.

To perform the integration in Eq. (11), we discretized the path into 20 equally spaced λ values. At each λ , we performed independent Monte Carlo simulations starting from an equilibrated configuration of the reference sequence A and sequence B , applying the constraint-preserving moves described earlier to sample a Boltzmann distributed ensemble. Averages $\langle E_B(\bar{\mathbf{Y}}_\Delta) \rangle_\lambda$ and $\langle E_A(\bar{\mathbf{Y}}_\Delta) \rangle_\lambda$ were computed over the production trajectories. Numerical integration was performed using the trapezoidal rule, yielding the free-energy difference for each sliding window relative to the reference sequence B . This approach was repeated for all 1354 windows in the 1501 bp promoter region to generate the free-energy landscape shown in Figs. 4(e) and 4(f).

-
- [1] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, *Nature* **389**, 251 (1997).
 - [2] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution, *J. Mol. Biol.* **319**, 1097 (2002).
 - [3] B. Fierz and M. G. Poirier, Biophysics of chromatin dynamics, *Annu. Rev. Biophys.* **48**, 321 (2019).
 - [4] G. Meersseman, S. Pennings, and E. M. Bradbury, Mobile nucleosomes—A general behavior, *EMBO J.* **11**, 2951 (1992).
 - [5] S. Rudnizky, H. Khamis, O. Malika, P. Melamed, and A. Kaplan, The base pair-scale diffusion of nucleosomes modulates binding of transcription factors, *Proc. Natl. Acad. Sci. USA* **116**, 12161 (2019).
 - [6] I. M. Kulić and H. Schiessel, Chromatin dynamics: Nucleosomes go mobile through twist defects, *Phys. Rev. Lett.* **91**, 148103 (2003).
 - [7] F. Mohammad-Rafiee, I. M. Kulić, and H. Schiessel, Theory of nucleosome corkscrew sliding in the presence of synthetic DNA ligands, *J. Mol. Biol.* **344**, 47 (2004).
 - [8] A. Fathizadeh, A. B. Besya, M. R. Ejtehadi, and H. Schiessel, Rigid-body molecular dynamics of DNA inside a nucleosome, *Eur. Phys. J. E* **36**, 21 (2013).
 - [9] J. Lequeieu, D. C. Schwartz, and J. J. de Pablo, In silico evidence for sequence-dependent nucleosome sliding, *Proc. Natl. Acad. Sci. USA* **114**, E9197 (2017).
 - [10] T. Niina, G. B. Brandani, C. Tan, and S. Takada, Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations, *PLoS Comput. Biol.* **13**, e1005880 (2017).
 - [11] G. B. Brandani, T. Niina, C. Tan, and S. Takada, DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations, *Nucleic Acids Res.* **46**, 2788 (2018).
 - [12] A. Z. Guo, J. Lequeieu, and J. J. de Pablo, Extracting collective motions underlying nucleosome dynamics via nonlinear manifold learning, *J. Chem. Phys.* **150**, 054902 (2019).
 - [13] H. Schiessel, J. Widom, R. F. Bruinsma, and W. M. Gelbart, Polymer reptation and nucleosome repositioning, *Phys. Rev. Lett.* **86**, 4414 (2001).
 - [14] I. M. Kulić and H. Schiessel, Nucleosome repositioning via loop formation, *Biophys. J.* **84**, 3197 (2003).
 - [15] J. Mozziconacci and J. M. Victor, Nucleosome gaping supports a functional structure for the 30 nm chromatin fiber, *J. Struct. Biol.* **143**, 72 (2003).
 - [16] T. T. M. Ngo and T. Ha, Nucleosomes undergo slow spontaneous gaping, *Nucleic Acids Res.* **43**, 3964 (2015).
 - [17] E. Gurovitch and P. Sens, Adsorption of polyelectrolyte onto a colloid of opposite charge, *Phys. Rev. Lett.* **82**, 339 (1999).
 - [18] E. M. Mateescu, C. Jeppesen, and P. Pincus, Overcharging of a spherical macroion by an oppositely charged polyelectrolyte, *Europhys. Lett.* **46**, 493 (1999).
 - [19] S. Y. Park, R. F. Bruinsma, and W. M. Gelbart, Spontaneous overcharging of macro-ion complexes, *Europhys. Lett.* **46**, 454 (1999).
 - [20] R. R. Netz and J.-F. Joanny, Complexation between a semiflexible polyelectrolyte and an oppositely charged sphere, *Macromolecules* **32**, 9026 (1999).
 - [21] T. T. Nguyen and B. I. Shklovskii, Overcharging of a macroion by an oppositely charged polyelectrolyte, *Physica A* **293**, 324 (2001).
 - [22] H. Schiessel, R. F. Bruinsma, and W. M. Gelbart, Electrostatic complexation of spheres and chains under elastic stress, *J. Chem. Phys.* **115**, 7245 (2001).
 - [23] K. K. Kunze and R. R. Netz, Complexes of semiflexible polyelectrolytes and charged spheres as models for salt-modulated nucleosomal structures, *Phys. Rev. E* **66**, 011918 (2002).
 - [24] I. M. Kulić and H. Schiessel, DNA spools under tension, *Phys. Rev. Lett.* **92**, 228101 (2004).
 - [25] T. Wocjan, K. Klenin, and J. Langowski, Brownian dynamics simulation of DNA unrolling from the nucleosome, *J. Phys. Chem. B* **113**, 2639 (2009).
 - [26] B. Sudhanshu, S. Mihađja, E. F. Koslover, S. Mehraeen, C. Bustamante, and A. J. Spakowitz, Tension-dependent structural deformation alters single-molecule transition kinetics, *Proc. Natl. Acad. Sci. USA* **108**, 1885 (2011).
 - [27] S. G. J. Mochrie, A. H. Mack, D. J. Schlingwein, R. Collins, M. Kamenetska, and L. Regan, Unwinding and rewinding the nucleosome inner turn: Force dependence of the kinetic rate constants, *Phys. Rev. E* **87**, 012710 (2013).
 - [28] J. D. Anderson and J. Widom, Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites, *J. Mol. Biol.* **296**, 979 (2000).
 - [29] A. W. Mauney, J. M. Tokuda, L. M. Gloss, O. Gonzales, and L. Pollack, Local DNA sequence controls asymmetry of DNA

- unwrapping from nucleosome core particles, *Biophys. J.* **115**, 773 (2018).
- [30] T. T. M. Ngo, Q. Zhang, R. Zhou, J. G. Yodh, and T. Ha, Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility, *Cell* **160**, 1135 (2015).
- [31] V. A. Bondarenko, L. M. Steele, A. Újvári, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse, and V. M. Studitsky, Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II, *Mol. Cell* **24**, 469 (2006).
- [32] F. Burgos-Bravo, A. B. Tong, C. Li, C. Diaz-Célis, C. D. Kaplan, G. LeRoy, D. Reinberg, and C. Bustamante, FACT weakens the nucleosomal barrier to transcription and preserves its integrity by forming a hexasome-like intermediate, *Mol. Cell* **85**, 2097 (2025).
- [33] B. Eslami-Mossallam, H. Schiessel, and J. van Noort, Nucleosome dynamics: Sequence matters, *Adv. Colloid Interface Sci.* **232**, 101 (2016).
- [34] S. Y. Ponomarev, V. Putkaradze, and T. C. Bishop, Relaxation dynamics of nucleosomal DNA, *Phys. Chem. Chem. Phys.* **11**, 10633 (2009).
- [35] C. K. Materese, A. Savelyev, and G. A. Papoian, Counterion atmosphere and hydration patterns near a nucleosome core particle, *J. Am. Chem. Soc.* **131**, 15005 (2009).
- [36] D. Winogradoff, H. Zhao, Y. Dalal, and G. A. Papoian, Shearing of the CENP-A dimerization interface mediates plasticity in the octameric centromeric nucleosome, *Sci. Rep.* **5**, 17038 (2015).
- [37] M. Pasi and R. Lavery, Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics, *Nucleic Acids Res.* **44**, 5450 (2016).
- [38] H. Kono, S. Sakuraba, and H. Ishida, Free energy profiles for unwrapping the outer superhelical turn of nucleosomal DNA, *PLoS Comput. Biol.* **14**, e1006024 (2018).
- [39] F. Cleri, F. Landuzzi, and R. Blossey, Mechanical evolution of DNA double-strand breaks in the nucleosome, *PLoS Comput. Biol.* **14**, e1006224 (2018).
- [40] D. Winogradoff and A. Aksimentiev, Molecular mechanism of spontaneous nucleosome unraveling, *J. Mol. Biol.* **431**, 323 (2019).
- [41] M. Pitman, Y. Dalal, and G. A. Papoian, Minimal cylinder analysis reveals the mechanical properties of oncogenic nucleosomes, *Biophys. J.* **118**, 2309 (2020).
- [42] D. Doğan, M. Arslan, T. Uluçay, S. Kalyoncu, S. Dimitrov, and K. Kale, CENP-A nucleosome is a sensitive allosteric scaffold for DNA and chromatin factors, *J. Mol. Biol.* **433**, 166789 (2021).
- [43] H. Kohestani and J. Wereszczynski, Effects of H2A.B incorporation on nucleosome structures and dynamics, *Biophys. J.* **120**, 1498 (2021).
- [44] H. Ishida and H. Kono, Torsional stress can regulate the unwrapping of two outer half superhelical turns of nucleosomal DNA, *Proc. Natl. Acad. Sci. USA* **118**, e2020452118 (2021).
- [45] R. Ettig, N. Kepper, R. Stehr, G. Wedemann, and K. Rippe, Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping, *Biophys. J.* **101**, 1999 (2011).
- [46] M. Biswas, K. Voltz, J. C. Smith, and J. Langowski, Role of histone tails in structural stability of the nucleosome, *PLoS Comput. Biol.* **7**, e1002279 (2011).
- [47] A. K. Shaytan, G. A. Armeev, A. Goncarencu, V. B. Zhurkin, D. Landsman, and A. R. Panchenko, Coupling between histone conformations and DNA geometry in nucleosomes on a microsecond timescale: Atomistic insights into nucleosome functions, *J. Mol. Biol.* **428**, 221 (2016).
- [48] Z. Li and H. Kono, Investigating the influence of arginine dimethylation on nucleosome dynamics using all-atom simulations and kinetic analysis, *J. Phys. Chem. B* **122**, 9625 (2018).
- [49] Y. Peng, S. Li, A. Onufriev, D. Landsman, and A. R. Panchenko, Binding of regulatory proteins to nucleosomes is modulated by dynamic histone tails, *Nat. Commun.* **12**, 5280 (2021).
- [50] K. Chakraborty, M. Kang, and S. M. Loverde, Molecular mechanism for the role of the H2A and H2B histone tails in nucleosome repositioning, *J. Phys. Chem. B* **122**, 11827 (2018).
- [51] J. Huertas, H. R. Schöler, and V. Cojocaru, Histone tails cooperate to control the breathing of genomic nucleosomes, *PLoS Comput. Biol.* **17**, e1009013 (2021).
- [52] G. A. Armeev, A. S. Kniazeva, G. A. Komarova, M. P. Kirpichnikov, and A. K. Shaytan, Histone dynamics mediate DNA unwrapping and sliding in nucleosomes, *Nat. Commun.* **12**, 2387 (2021).
- [53] S. Li, Y. Peng, D. Landsman, and A. R. Panchenko, DNA methylation cues in nucleosome geometry, stability and unwrapping, *Nucl. Acids Res.* **50**, 1864 (2022).
- [54] P. Khatua, P. K. Tang, A. G. Moulick, R. Patel, A. Manandhar, and S. M. Loverde, Sequence dependence in nucleosome dynamics, *J. Phys. Chem. B* **128**, 3090 (2024).
- [55] J. Huertas and V. Cojocaru, Breaths, twists, and turns of atomistic nucleosomes, *J. Mol. Biol.* **433**, 166744 (2021).
- [56] A. S. Fedulova, G. A. Armeev, T. A. Romanova, L. Singh-Palchevskaia, N. A. Kosarim, N. A. Motorin, G. A. Komarova, and A. K. Shaytan, Molecular dynamics simulations of nucleosomes are coming of age, *WIREs Comput. Mol. Sci.* **14**, e1728 (2024).
- [57] W. K. Olson, A. A. Gorin, X. Lu, L. M. Hock, and V. B. Zhurkin, DNA sequence-dependent deformability deduced from protein-DNA crystal complexes, *Proc. Natl. Acad. Sci. USA* **95**, 11163 (1998).
- [58] F. Lankaš, O. Gonzalez, L. M. Heffler, G. Stoll, M. Moakher, and J. H. Maddocks, On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations, *Phys. Chem. Chem. Phys.* **11**, 10565 (2009).
- [59] M. Y. Tolstorukov, A. V. Colasant, D. M. McCandlish, W. K. Olson, and V. B. Zhurkin, A novel roll-and-slide mechanism of DNA folding in chromatin: Implications for nucleosome positioning, *J. Mol. Biol.* **371**, 725 (2007).
- [60] S. Balasubramanian, F. Xu, and W. K. Olson, DNA sequence-directed organization of chromatin: Structure-based computational analysis of nucleosome-binding sequences, *Biophys. J.* **96**, 2245 (2009).
- [61] N. B. Becker and R. Everaers, DNA nanomechanics in the nucleosome, *Structure* **17**, 579 (2009).
- [62] A. V. Morozov, K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, and E. D. Siggia, Using DNA mechanics to predict *in vitro* nucleosome positions and formation energies, *Nucleic Acids Res.* **37**, 4707 (2009).

- [63] D. Norouzi and F. Mohammad-Rafiee, DNA conformation and energy in nucleosome core: A theoretical approach, *J. Biomol. Struct. Dyn.* **32**, 104 (2014).
- [64] B. Eslami-Mossallam, R. D. Schram, M. Tompitak, J. van Noort, and H. Schiessel, Multiplexing genetic and nucleosome positioning codes: A computational approach, *PLoS One* **11**, e0156905 (2016).
- [65] L. de Bruin, M. Tompitak, B. Eslami-Mossallam, and H. Schiessel, Why do nucleosomes unwrap asymmetrically? *J. Phys. Chem. B* **120**, 5855 (2016).
- [66] E. Skoruppa, S. K. Nomidis, J. F. Marko, and E. Carlon, Bend-induced twist waves and the structure of nucleosomal DNA, *Phys. Rev. Lett.* **121**, 088101 (2018).
- [67] J. Culkun, L. de Bruin, M. Tompitak, R. Phillips, and H. Schiessel, The role of DNA sequence in nucleosome breathing, *Eur. Phys. J. E* **40**, 106 (2017).
- [68] M. Zuiddam, R. Everaers, and H. Schiessel, Physics behind the mechanical nucleosome positioning code, *Phys. Rev. E* **96**, 052412 (2017).
- [69] G. Liu, Y. Xing, L. Zhao, L. Cai, and J. Wang, The implication of DNA bending energy for nucleosome positioning and sliding, *Sci. Rep.* **8**, 8853 (2018).
- [70] T. Parsons and B. Zhang, Critical role of histone tail entropy in nucleosome unwinding, *J. Chem. Phys.* **150**, 185103 (2019).
- [71] K. van Deelen, H. Schiessel, and L. de Bruin, Ensembles of breathing nucleosomes: A computational study, *Biophys. J.* **118**, 2297 (2020).
- [72] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo, An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization, *J. Chem. Phys.* **139**, 144903 (2013).
- [73] G. S. Freeman, J. P. Lequeieu, D. M. Hinckley, J. K. Whitmer, and J. J. de Pablo, DNA shape dominates sequence affinity in nucleosome formation, *Phys. Rev. Lett.* **113**, 168101 (2014).
- [74] G. B. Brandani and S. Takada, Chromatin remodelers couple inchworm motion with twist-defect formation to slide nucleosomal DNA, *PLoS Comput. Biol.* **14**, e1006512 (2018).
- [75] D. Petkevičiūtė, M. Pasi, O. Gonzalez, and J. H. Maddocks, cgDNA: A software package for the prediction of sequence-dependent coarse-grain free energies of B-form DNA, *Nucleic Acids Res.* **42**, e153 (2014).
- [76] R. Sharma, A. S. Patelli, L. de Bruin, and J. H. Maddocks, cgNA+web: A visual interface to the cgNA+ sequence-dependent statistical mechanics model of double-stranded nucleic acids, *J. Mol. Biol.* **435**, 167978 (2023).
- [77] R. Giniūnaitė and D. Petkevičiūtė-Gerlach, Predicting the configuration and energy of DNA in a nucleosome by coarse-grain modelling, *Phys. Chem. Chem. Phys.* **24**, 26124 (2022).
- [78] R. Giniūnaitė, R. Sharma, J. H. Maddocks, S. Kriaucionis, and D. Petkevičiūtė-Gerlach, Nucleosome wrapping energy in CpG islands and the role of epigenetic base modifications, *eLife* **13**, RP98468 (2025).
- [79] V. Miele, C. Vaillant, Y. d'Aubenton Carafa, C. Thermes, and T. Grange, DNA physical properties determine nucleosome occupancy from yeast to fly, *Nucleic Acids Res.* **36**, 3746 (2008).
- [80] G. Liu, Y. Xing, H. Zhao, Y. Wang, and L. Cai, A deformation energy-based model for predicting nucleosome dyads and occupancy, *Sci. Rep.* **6**, 24133 (2016).
- [81] S. C. Satchwell, H. R. Drew, and A. A. Travers, Sequence periodicities in chicken nucleosome core DNA, *J. Mol. Biol.* **191**, 659 (1986).
- [82] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom, A genomic code for nucleosome positioning, *Nature* **442**, 772 (2006).
- [83] T. E. Shrader and D. M. Crother, Artificial nucleosome positioning sequences, *Proc. Natl. Acad. Sci. USA* **86**, 7418 (1989).
- [84] T. E. Shrader and D. M. Crother, Effects of DNA sequence and histone-histone interactions on nucleosome placement, *J. Mol. Biol.* **216**, 69 (1990).
- [85] K. J. Polach and J. Widom, Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation, *J. Mol. Biol.* **254**, 130 (1995).
- [86] M. Tompitak, G. T. Barkema, and H. Schiessel, Benchmarking and refining probability-based models for nucleosome-DNA interaction, *BMC Bioinf.* **18**, 157 (2017).
- [87] H. S. Tims, K. Gurunathan, M. Levitus, and J. Widom, Dynamics of nucleosome invasion by DNA binding proteins, *J. Mol. Biol.* **411**, 430 (2011).
- [88] E. Skoruppa and H. Schiessel, Systematic coarse-graining of sequence-dependent structure and elasticity of double-stranded DNA, *Phys. Rev. Res.* **7**, 013044 (2025).
- [89] M. A. Hall, A. Shundrovsky, L. Bai, R. M. Fulbright, J. T. Lis, and M. D. Wang, High-resolution dynamic mapping of histone-DNA interactions in a nucleosome, *Nat. Struct. Mol. Biol.* **16**, 124 (2009).
- [90] R. Lavery, M. Moakher, J. Maddocks, D. Petkeviciute, and D. Zakrzewska, Conformational analysis of nucleic acids revisited: Curves+, *Nucleic Acids Res.* **37**, 5917 (2009).
- [91] F. Lankaš, J. Šponer, J. Langowski, and T. E. Cheatham III, DNA basepair step deformability inferred from molecular dynamics simulations, *Biophys. J.* **85**, 2872 (2003).
- [92] E. Skoruppa, A. Voorspoels, J. Vreede, and E. Carlon, Length-scale-dependent elasticity in DNA from coarse-grained and all-atom models, *Phys. Rev. E* **103**, 042408 (2021).
- [93] W. Laeremans, M. Segers, A. Voorspoels, E. Carlon, and J. Hooyberghs, Insights into elastic properties of coarse-grained DNA models: Q-stiffness of cgDNA vs cgDNA+, *J. Chem. Phys.* **160**, 144105 (2024).
- [94] P. Slavíková, M. Cuker, E. Matoušková, I. Čmelo, M. Zgarbová, P. Jurečka, and F. Lankaš, Sequence-dependent shape and stiffness of DNA and RNA double helices: Hexanucleotide scale and beyond, *J. Chem. Inf. Model.* **65**, 9208 (2025).
- [95] J. Neipel, G. Brandani, and H. Schiessel, Translational nucleosome positioning: A computational study, *Phys. Rev. E* **101**, 022405 (2020).
- [96] Edited by F. Zhang, *The Schur Complement and Its Applications*, Numerical Methods and Algorithms (Springer-Verlag, Berlin, 2005).
- [97] Y. A. Gutiérrez Fosado, F. Landuzzi, and T. Sakaue, Coarse graining DNA: Symmetry, nonlocal elasticity, and persistence length, *Phys. Rev. Lett.* **130**, 058402 (2023).
- [98] O. Gonzalez, D. Petkevičiūtė, and J. H. Maddocks, A sequence-dependent rigid-base model of DNA, *J. Chem. Phys.* **138**, 055102 (2013).

- [99] R. Hartley, J. Trumpf, Y. Dai, and H. Li, Rotation averaging, *Int. J. Comput. Vis.* **103**, 267 (2013).
- [100] N. B. Becker, L. Wolff, and R. Everaers, Indirect readout: Detection of optimized subsequences and calculation of relative binding affinities using different DNA elastic potentials, *Nucleic Acids Res.* **34**, 5638 (2006).
- [101] J. G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**, 300 (1935).
- [102] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic Press, San Diego, CA, 2002).
- [103] P. Lowary and J. Widom, New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning, *J. Mol. Biol.* **276**, 19 (1998).
- [104] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, MDTraj: A modern open library for the analysis of molecular dynamics trajectories, *Biophys. J.* **109**, 1528 (2015).
- [105] T. van Heesch, E. Skoruppa, P. G. Bolhuis, H. Schiessel, and J. Vreede, MDNA: A software module for DNA structure generation and analysis, *bioRxiv*: 10.1101/2025.07.26.666940 [Nucleic Acids Res. (to be published)].
- [106] B. van Rossum, W. Fischle, and P. Selenko, Asymmetrically modified nucleosomes expand the histone code, *Nat. Struct. Mol. Biol.* **19**, 1064 (2012).
- [107] P. Voigt, G. LeRoy, W. J. Drury III, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia, and D. Reinberg, Asymmetrically modified nucleosomes, *Cell* **151**, 181 (2012).
- [108] B. Brower-Toland, C. L. Smith, R. C. Yeh, J. Lis, C. L. Peterson, and M. D. Wang, Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA, *Proc. Natl. Acad. Sci. USA* **99**, 1960 (2002).
- [109] F. Dong and K. E. van Holde, Nucleosome positioning is determined by the (H3-H4)₂ tetramer, *Proc. Natl. Acad. Sci. USA* **88**, 10596 (1991).
- [110] G. Li and J. Widom, Nucleosomes facilitate their own invasion, *Nat. Struct. Mol. Biol.* **11**, 763 (2004).
- [111] G. Li, M. Levitus, C. Bustamante, and J. Widom, Rapid spontaneous accessibility of nucleosomal DNA, *Nat. Struct. Mol. Biol.* **12**, 46 (2005).
- [112] L. Kelbauskas, N. Chan, R. Bash, J. Yodh, N. Woodbury, and D. Lohr, Sequence-dependent nucleosome structure and stability variations detected by Förster resonance energy transfer, *Biochemistry* **46**, 2239 (2007).
- [113] W. J. A. Koopmans, A. Brehm, C. Logie, T. Schmidt, and J. van Noort, Single-pair FRET microscopy reveals mononucleosome dynamics, *J. Fluoresc.* **17**, 785 (2007).
- [114] G. Moyle-Heyrman, H. S. Tims, and J. Widom, Structural constraints in collaborative competition of transcription factors against the nucleosome, *J. Mol. Biol.* **412**, 634 (2011).
- [115] A. Gansen, S. Felekyan, R. Kühnemuth, K. Lehmann, K. Tóth, C. A. M. Seidel, and J. Langowski, High precision FRET studies reveal reversible transitions in nucleosomes between microseconds and minutes, *Nat. Commun.* **9**, 4628 (2018).
- [116] S. F. Konrad, W. Vanderlinden, W. Frederickx, T. Brouns, B. H. Menze, S. De Feyter, and J. Lipfert, High-throughput AFM analysis reveals unwrapping pathways of H3 and CENP-A nucleosomes, *Nanoscale* **13**, 5435 (2021).
- [117] S. F. Konrad, W. Vanderlinden, and J. Lipfert, Quantifying epigenetic modulation of nucleosome breathing by high-throughput AFM imaging, *Biophys. J.* **121**, 841 (2022).
- [118] L. Lenz, M. Hoenderdos, P. Prinsen, and H. Schiessel, The influence of DNA shape fluctuations on fluorescence resonance energy transfer efficiency measurements in nucleosomes, *J. Phys.: Condens. Matter* **27**, 064104 (2015).
- [119] H. Schiessel, Telling left from right in breathing nucleosomes, *Biophys. J.* **115**, 749 (2018).
- [120] A. H. Mack, D. J. Schlingman, R. D. Salinas, L. Regan, and S. G. J. Mochrie, Condensation transition and forced unravelling of DNA-histone H1 toroids: A multi-state free energy landscape, *J. Phys.: Condens. Matter* **27**, 064106 (2015).
- [121] J. Lequieu, A. Córdoba, D. C. Schwartz, and J. J. de Pablo, Tension-dependent free energies of nucleosome unwrapping, *ACS Cent. Sci.* **2**, 660 (2016).
- [122] M. Tompitak, L. de Bruin, B. Eslami-Mossallam, and H. Schiessel, Designing nucleosomal force sensors, *Phys. Rev. E* **95**, 052402 (2017).
- [123] F. Khodabandeh, H. Fatemi, and F. Mohammad-Rafiee, Insight into the unwrapping of the dinucleosome, *Soft Matter* **16**, 4806 (2020).
- [124] G. Reddy and D. Thirumalai, Asymmetry in histone rotation in forced unwrapping and force quench rewinding in a nucleosome, *Nucleic Acids Res.* **49**, 4907 (2021).
- [125] C. Jiang and B. F. Pugh, Nucleosome positioning and gene regulation: Advances through genomics, *Nat. Rev. Genet.* **10**, 161 (2009).
- [126] A. Mondal, S. K. Mishra, and A. Bhattacharjee, Nucleosome breathing facilitates cooperative binding of pluripotency factors Sox2 and Oct4 to DNA, *Biophys. J.* **121**, 4526 (2022).
- [127] A. Soufi, M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, and K. S. Zaret, Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming, *Cell* **161**, 555 (2015).
- [128] C. Tan and S. Takada, Nucleosome allostery in pioneer transcription factor binding, *Proc. Natl. Acad. Sci. USA* **117**, 20586 (2020).
- [129] X. Liu, X. Xia, X. Li, and Z. Chen, Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure, *Nature* **544**, 440 (2017).
- [130] S. Klempahn, R. Blossey, and H. Schiessel, Molecular motor in a box: Attractors, repellers, and ratchets in chromatin, *Phys. Rev. Res.* **6**, 023236 (2024).
- [131] S. S. Hammoud, D. A. Nix, H. Zhang, J. Purwar, D. T. Carrell, and B. R. Cairns, Distinctive chromatin in human sperm packages genes for embryo development, *Nature* **460**, 473 (2009).
- [132] L. Moritz and S. S. Hammoud, The art of packaging the sperm genome: Molecular and structural basis of the histone-to-protamine exchange, *Front. Endocrinol.* **13**, 895502 (2022).
- [133] S. Bilokapic, M. Strauss, and M. Halic, Histone octamer rearranges to adapt to DNA unwrapping, *Nat. Struct. Mol. Biol.* **25**, 101 (2018).
- [134] A. Basu, D. G. Bobrovnikov, B. Cieza, J. P. Arcon, Z. Qureshi, M. Orozco, and T. Ha, Deciphering the mechanical code of the genome and epigenome, *Nat. Struct. Mol. Biol.* **29**, 1178 (2022).

- [135] E. Skoruppa, RBPNE: Python module for rigid base-pair based nucleosome free energy calculations, <https://github.com/eskoruppa/RBPNE> (2026).
- [136] B. R. Gordon, Y. Li, A. Cote, M. T. Weirauch, P. Ding, T. R. Hughes, W. W. Navarre, B. Xia, and J. Liu, Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins, *Proc. Natl. Acad. Sci. USA* **108**, 10690 (2011).
- [137] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theory Comput.* **11**, 3696 (2015).
- [138] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, *et al.*, Parmbsc1: A refined force field for DNA simulations, *Nat. Methods* **13**, 55 (2016).
- [139] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **79**, 926 (1983).
- [140] T. E. Cheatham III, J. L. Miller, T. Fox, T. Darden, and P. A. Kollman, Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins, *J. Am. Chem. Soc.* **117**, 4193 (1995).
- [141] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* **103**, 8577 (1995).
- [142] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, GROMACS: Fast, flexible, and free, *J. Comput. Chem.* **26**, 1701 (2005).
- [143] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* **1-2**, 19 (2015).
- [144] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, LINCS: A linear constraint solver for molecular simulations, *J. Comput. Chem.* **18**, 1463 (1997).
- [145] S. Miyamoto and P. A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *J. Comput. Chem.* **13**, 952 (1992).
- [146] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* **126**, 014101 (2007).
- [147] M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.* **52**, 7182 (1981).
- [148] S. Nosé and M. Klein, Constant pressure molecular dynamics for molecular systems, *Mol. Phys.* **50**, 1055 (1983).
- [149] W. Kabsch, A solution for the best rotation to relate two sets of vectors, *Acta Crystallogr. A* **32**, 922 (1976).
- [150] E. Skoruppa, M. Laleman, S. K. Nomidis, and E. Carlon, DNA elasticity from coarse-grained simulations: The effect of groove asymmetry, *J. Chem. Phys.* **146**, 214902 (2017).
- [151] R. Sun and T. Bishop, The nucleosome reference frame and standard geometries for octasomes, *Biophys. Rev.* **16**, 315 (2024).
- [152] A. B. Kolomeisky and M. E. Fisher, Periodic sequential kinetic models with jumping, branching and deaths, *Physica A* **279**, 1 (2000).