Systematic coarse-graining of sequence-dependent structure and elasticity of double-stranded DNA

Enrico Skoruppa [1,*] and Helmut Schiessel [1,2]

[1]*Cluster of Excellence Physics of Life, TU Dresden, 01062 Dresden, Germany*
[2]*Institut für Theoretische Physik, TU Dresden, 01062 Dresden, Germany*

Coarse-grained models have played an important role in the study of the behavior of DNA at length scales beyond a few hundred base pairs. Traditionally, these models have relied on structurally featureless and sequence-independent approaches, such as the twistable wormlike chain. However, research over the past decade has highlighted the substantial impact of DNA sequence even at the kilobase pair scale. Several robust sequence-dependent models have emerged, capturing intricacies at the base pair-step level. Here we introduce an analytical framework for coarse-graining such models to lower resolution representations while preserving essential structural and dynamic features, enabling the efficient sampling of large molecules. When considering both rotational and translational degrees of freedom, coarse-graining is shown to yield excellent results up to about one helical repeat. For scenarios where the local stretch modulus is inconsequential, this range can be significantly increased by only considering the rotational degrees of freedom, which permits faithful coarse-grained representation up to several helical repeats. Rather than providing a fully parametrized model, we present the methodology and software necessary for mapping any base pair-step model to the desired level of coarse-graining. Finally, we provide application examples of our method, including estimates of the persistence length and effective torsional stiffness of DNA in a setup mimicking a freely orbiting tweezer, as well as simulations of helically curved DNA.

I. INTRODUCTION

It is well established that the base sequence of DNA carries significance beyond its encoding into amino acids and the biochemical recognition of sequence motifs via DNA-binding proteins. Both structure and elasticity of individual stretches of DNA have long since been shown to exhibit modulations vis-à-vis the underlying base pair sequence [1–3]. Such innate mechanical features constitute a rugged landscape of mechanical resistance that favors particular deformations at specific locations. For example, certain sequences display pronounced intrinsic curvature, i.e., the curvilinear center line is bent even in the absence of thermal fluctuations or externally induced deformations [4–10]. Achieving a collectively curved contour over a stretch of DNA is easiest at places where the molecule is already naturally bent. At these places adhering to the innate direction of curvature is strongly favored. A prominent example is nucleosome positioning, i.e., the sequence preference for the wrapping of 147 base pairs around histone octamers in 1.7 superhelical turns. Given the persistence length of roughly 150 bp the wrapping constitutes a significant elastic deformation. Consequently, the consideration of structure and elasticity has been successful in the theoretical evaluation of preferred nucleosome-wrapped sequences [11–14].

At large length scales the elastic response of DNA is well described by homogeneous semiflexible models and in particular the wormlike chain (WLC) [15–19] or the twistable wormlike chain (TWLC) when observables involving torsional and topological properties are considered [20–28]. However, even phenomena involving thousands of base pairs may exhibit appreciable sequence-specific behavior. For example, Kim *et al.* [29] demonstrated that DNA plectonemes tend to concentrate at specific sequence-encoded locations, which may carry profound implications for the spatial distribution of topological strain throughout the chromosome. Understanding the sequence-specific mechanical behavior of DNA may be crucial for comprehending the structural dynamics and functional fidelity of the chromosome.

Despite rapid advancements, experimental methods providing resolutions that reveal sequence-specific dynamics remain limited to date. Accordingly, the utility of molecular simulations as effective computational microscopes remains uncontested. Atomistic simulations have been successful in replicating local mechanical properties of DNA [30–32], and their application has been instrumental in unraveling a multitude of phenomena involving both bare DNA [33–40] and DNA-protein complexes [41–45]. However, the computational cost associated with this level of detail is prohibitive for

---

*Contact author: esk.phys@gmail.com

the study of systems involving more than a few hundred base pairs.

Implicit solvent simulations have significantly extended the attainable length scale [46]; however, reaching the length scales relevant to typical single-molecule experiments has thus far only been possible through the development of coarse-grained models. These models reduce the complexity of DNA to the necessary components for the particular phenomenon of interest, thereby reducing computational expense which in turn enhances scalability. A large variety of such models have emerged in recent years [47–53] some of which feature structural and elastic sequence-dependence [54–57].

A particularly prevalent way of coarse-graining DNA is the rigid base pair (RBP) model [58], which represents each base pair as a single rigid body (see Fig. 1). Relative rotations and translations between consecutive base pairs are parametrized in terms of six degrees of freedom: three rotational (tilt, roll, and twist) and three translational (shift, slide, and rise). Incidentally, these degrees of freedom coincide with the most common way of classifying local elastic properties of DNA [59–63]. In corresponding models, the sequence-dependent intrinsic structure is included via the ground states of the six respective degrees of freedom. Deformations away from this ground state are commonly penalized with a quadratic elastic energy [1,3,64–66]. This should be viewed as the lowest-order expansion of a generic underlying elastic energy. Truncation to quadratic order is warranted due to the stiff nature of double-stranded DNA. A given parametrization of an RBP model is fully characterized by a set of ground-state coordinates and a stiffness matrix.

To obtain sufficient statistics on systems involving thousands of base pairs the molecules under consideration are frequently coarse-grained to significantly larger sub-units than the basepair-step resolution of the RBP [25,27,28,67–69]. Not only do these coarser descriptions reduce the number of degrees of freedom, but coarse-graining also reduces the stiffness of the effective potentials, such that molecular dynamics simulations may be executed with larger time-steps [25,50,69] and Markov chain Monte Carlo simulations may feature larger cluster moves [28,70,71].

Many studies rely on top-down parametrizations for coarse-grained descriptions, that utilize the homogenous TWLC model which has only two free parameters: the bending stiffness A and the twist stiffness C. These parameters are usually adopted from experimental measurements or higher resolution simulations, yielding a bending stiffness of about 40–55 nm [18,27,72–74] and a torsional stiffness in the range of 60 to 110 nm [21,22,26,27,75–79] under physiological ionic conditions. Sequence-dependent simulations of models with subunits spanning several base pairs have been employed in various studies [80,81]. However, the used bottom-up parametrization required extensive all-atom molecular dynamics sampling for all considered sequences. Moreover, there have been several works providing schemes for the analytical coarse-graining RBP parameters to TWLC parameters [82–84]. While these approaches provide parametrizations that yield good agreement with the collective behavior of DNA molecules at large length scales, they are devoid of the structural and dynamic sequence features outlined before.
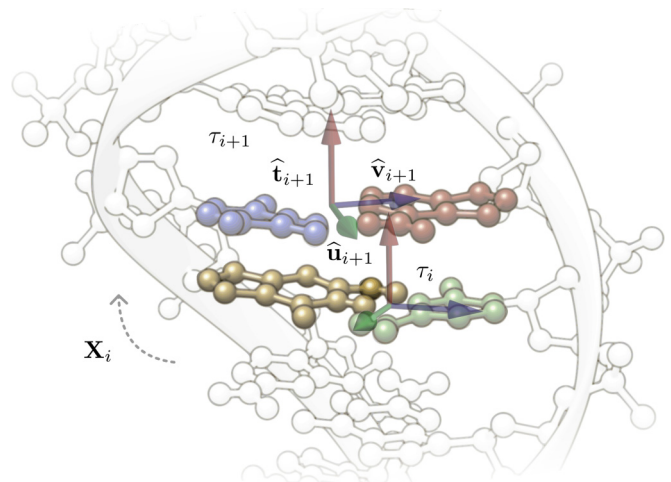

FIG. 1. Illustration of the rigid base pair description of DNA. Base pairs—depicted in blue-red and yellow-green—are captured by a single right-handed reference frame that captures the position and orientation of the base pair. Relative orientations and positions of adjacent base pairs are parametrized in terms of six-vectors ($\mathbf{X}_i$) consisting of three rotational and three translational parameters.

In this work, we present a systematic procedure for coarse-graining elastic RBP models to any resolution. Coarse-graining of rotational and translational degrees of freedom is shown to be effective for resolutions up to a single helical repeat. In contrast, focusing exclusively on rotational degrees of freedom extends the accuracy of coarse-grained representations which permits monomeric subunits spanning over multiple helical repeats. The procedure entails a parameter transformation into a self-similar system, where the functional form of the model remains unchanged. It can be applied to any set of RBP parameters such as, for example, those provided by Olson *et al.* [1], Lankaš *et al.* [3] or Sharma *et al.* [65] (cgNA+, which is technically a superset Model of the RBP as it resolves rigid bases and rigid phosphate groups).

The article is structured as follows: The first half is dedicated to the description of the model and the development of the coarse-graining procedure. Section II A introduces the RBP model framework and casts the model in the form that is most convenient for the coarse-graining procedure. The coarse-graining scheme is detailed in Sec. II B, and the specifics of the parameter transformation are provided in Sec. II C. A Python implementation for the parameter transformation is available at Ref. [85].

The second half of the paper presents various benchmark results assessing the effectiveness of the coarse-graining procedure. Section III A examines how well the coarse-grained system replicates individual distributions within unrestrained ensembles. Section III B discusses the accuracy of reproducing length-scale-dependent persistence lengths. Additionally, we simulate a freely orbiting magnetic tweezer setup for an experimentally studied sequence in Sec. III C, and analyze the behavior and response of helically curved DNA sequences in Sec. III D. Finally, the paper concludes by summarizing the observations, reflecting on the potential impact of the work, and discussing its potential applications.

II. THEORY

A. Rigid base pair model

In the rigid base pair description of DNA, each base pair is treated as a rigid body associated with a couple $(\mathcal{T}, \mathbf{r})$, capturing its orientation and position, respectively. Triads $\mathcal{T}$ are right-handed reference frames that reflect the local geometry of the base pair in question, i.e., the orientation of the quasiplanar Watson-Crick base pairs, and the location of groves and backbones relative to the center of mass of the respective base pair. While exact definitions may vary across different conventions, most implementations—based on atomistic [62,63,86] and coarse-grained descriptions [56,87,88]—seek to generally encapsulate these geometric features. Throughout this work, we will define the three orthonormal basis vectors of each right-handed frame to populate the columns of the triad

$$\mathcal{T} = [\ \widehat{\mathbf{u}}\ \widehat{\mathbf{v}}\ \widehat{\mathbf{t}}\ ], \tag{1}$$

such that global transformations act on triads in the same way as on any ordinary vector. Following standard convention [60], $\widehat{\mathbf{t}}$ is associated with the normal of the base pair plane (which for Watson-Crick base pairs is closely aligned with the curvilinear tangent of the molecule), $\widehat{\mathbf{u}}$ roughly points from the base pair center of mass to the major groove, and $\widehat{\mathbf{v}}$ is closely aligned with the vector connecting the two backbones. Note that this definition classifies triads as elements of the rotation group, SO(3).

Relative rotations and translations between adjacent base pairs may be expressed in terms of six parameters—three rotational and three translational. When considering neighboring base pairs these components are commonly referred to as tilt, roll, and twist—for the rotational components—and shift, slide, and rise—for the translational components. To express these components independently of the global orientation of the molecule as a whole, relative rotations and translations are expressed in terms of the local material frame. The orientation of triad $\mathcal{T}_{i+1}$ with respect to the frame of triads $\mathcal{T}_i$ is given by

$$\mathcal{R}_i = \mathcal{T}_i^{\mathsf{T}} \mathcal{T}_{i+1}. \tag{2}$$

These transformations between adjacent frames will occasionally be referred to as junctions or junction transformations. For this work, we will employ a definition of tilt, roll, and twist as the components of the rotation vector $\mathbf{\Omega}$ also known as Euler vector—associated with the rotation matrix R. The relationship between the rotation matrices and rotation vectors is provided by the Euler map, which is explicitly given by Rodrigues' rotation formula (see Appendix A) or equivalently by the exponential map

$$\mathcal{R}_i = \exp \widehat{\Omega}_i, \tag{3}$$

where exp indicates the matrix exponential and the antisymmetric generators of rotation

$$\widehat{\Omega} = \begin{pmatrix} 0 & -\Omega_t & \Omega_v \\ \Omega_t & 0 & -\Omega_u \\ -\Omega_v & \Omega_u & 0 \end{pmatrix}, \tag{4}$$

are elements of the Lie algebra $\mathfrak{so}(3)$. The components Ω_u, Ω_v, Ω_t—tilt, roll, and twist, respectively—are simply the entries of the rotation vector $\mathbf{\Omega}^{\mathsf{T}} = (\Omega_u \quad \Omega_v \quad \Omega_t)$. Hence,

there is a one-to-one correspondence between rotation generators [$\mathfrak{so}(3)$] and rotation vectors ($\mathbb{R}^3$),

$$\mathrm{hat}(\mathbf{\Omega}) = \widehat{\Omega}, \ \mathrm{and} \ \mathrm{vec}(\widehat{\Omega}) = \mathbf{\Omega}, \tag{5}$$

which we will refer to as hat-map and vector-map, respectively. An important relationship is that under the action of the vector map the lie bracket of $\mathfrak{so}(3)$ turns into the cross-product,

$$\mathrm{vec}([\widehat{\Omega}_1, \widehat{\Omega}_2]) = \mathbf{\Omega}_1 \times \mathbf{\Omega}_2 = \widehat{\Omega}_1 \mathbf{\Omega}_2, \tag{6}$$

where $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are arbitrary rotation vectors. A commonly employed alternative [58,64,65,89–91] to the Euler-map is the Cayley map—also called the Euler-Rodrigues formula—which is closely related to the definition of unit quaternions. In Appendix B we show how ground state and Gaussian elasticity, i.e., the stiffness matrix, can be transformed from the Cayley definition to the Euler definition and vice versa.

Frequently, translations are expressed in the coordinate system of midstep triads, which enables a definition for translations independent of the choice of reference strand [60,62,63]. However, for the coarse-graining procedure outlined further below, it is advantageous to depart from this invariance and instead express relative translations within the frame of the first triad in each respective pair,

$$\mathbf{w}_i = \mathcal{T}_i^{\mathsf{T}} (\mathbf{r}_{i+1} - \mathbf{r}_i). \tag{7}$$

Despite the deviation from common convention, we will continue to refer to the translational components as shift, slide, and rise. Transformation of structure and elasticity from midstep- to triad-definition is discussed in Appendix D.

Orientations and translations of triads may be cast in a unified representation within elements of the special Euclidean group [82], SE(3),

$$\tau_i = \begin{pmatrix} \mathcal{T}_i & \mathbf{r}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix}, \tag{8}$$

where $\mathbf{0}$ is the three-dimensional null vector. Junction transformations then naturally contain the previously introduced rotation matrix and translation vector

$$g_i \equiv \tau_i^{-1} \tau_{i+1} = \begin{pmatrix} \mathcal{T}_i^{\mathsf{T}} \mathcal{T}_{i+1} & \mathcal{T}_i^{\mathsf{T}} (\mathbf{r}_{i+1} - \mathbf{r}_i) \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \exp \widehat{\Omega}_i & \mathbf{w}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix}, \tag{9}$$

and are therefore parametrized by six-component vectors,

$$\mathbf{X}_i^{\mathsf{T}} = \begin{pmatrix} \mathbf{\Omega}_i^{\mathsf{T}} & \mathbf{w}_i^{\mathsf{T}} \end{pmatrix}. \tag{10}$$

For convenience, we introduce a map between transformations g_i and the corresponding parametrization vector,

$$\mathbf{X}_i \equiv \mathcal{P}(g_i). \tag{11}$$

The ground state of a particular molecule, which we also refer to as its structure, is characterized by certain sequence-specific junctions,

$$s_i \equiv \begin{pmatrix} \mathcal{S}_i & \mathbf{s}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix} = \begin{pmatrix} \exp \widehat{\Omega}_{0,i} & \mathbf{s}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix}, \tag{12}$$

where the sequence-dependent $\widehat{\Omega}_{0,i}$ and $\mathbf{s}_i$ are provided by the chosen RBP parameter set.

Intrinsic bending and intrinsic twist are encoded in the static rotational parameters $\mathbf{\Omega}_0$ while intrinsic translations, among which intrinsic rise is the most prominent, are contained in $\mathbf{s}_i$. Together these components are summarized in the ground-state vector

$$\mathbf{X}_{0,i}^{\mathsf{T}} \equiv \mathcal{P}^{\mathsf{T}}(s_i) = \begin{pmatrix} \mathbf{\Omega}_{0,i}^{\mathsf{T}} & \mathbf{s}_i^{\mathsf{T}} \end{pmatrix}. \tag{13}$$

Static and dynamic components—fluctuations away from the ground state—are usually distinguished by splitting the junction six-vectors, Eq. (11), into the respective components

$$\mathbf{X}_i = \mathbf{X}_{0,i} + \mathbf{X}_{\Delta,i}, \tag{14}$$

where $\mathbf{X}_{\Delta,i}$ is the dynamic component.

However, for this work, it turns out to be a more prudent choice to split static and dynamic components at the transformation level rather than the vector level. This implies splitting the junctions g_i into static components s_i and a dynamic components d_i as

$$g_i = s_i d_i, \tag{15}$$

with

$$d_i \equiv \begin{pmatrix} \mathcal{D}_i & \mathbf{d}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix} = \begin{pmatrix} \exp\widehat{\Phi}_{\Delta,i} & \mathbf{d}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{pmatrix}. \tag{16}$$

To differentiate from the commonly used definition of fluctuations as the excess part of a single vector, we introduce a distinct notation for the components within our definition:

$$\mathbf{Y}_{0,i} \equiv \mathcal{P}^{\mathsf{T}}(s_i) = \begin{pmatrix} \mathbf{\Phi}_{0,i}^{\mathsf{T}} & \mathbf{s}_i^{\mathsf{T}} \end{pmatrix}, \tag{17}$$

$$\mathbf{Y}_{\Delta,i} \equiv \mathcal{P}^{\mathsf{T}}(d_i) = \begin{pmatrix} \mathbf{\Phi}_{\Delta,i}^{\mathsf{T}} & \mathbf{d}_i^{\mathsf{T}} \end{pmatrix}. \tag{18}$$

The static component is, of course, independent of the definition of fluctuations such that

$$\mathbf{Y}_{0,i} = \mathbf{X}_{0,i}, \tag{19}$$

i.e., $\mathbf{\Phi}_{0,i} = \mathbf{\Omega}_{0,i}$.

To fully capture the state of a given molecule containing say $N + 1$ base pairs and N junctions, we introduce the system-wide static and dynamic state-vectors,

$$\bar{\mathbf{Y}}_0^{\mathsf{T}} \equiv \begin{pmatrix} \mathbf{Y}_{0,0}^{\mathsf{T}} & \mathbf{Y}_{0,1}^{\mathsf{T}} \cdots \mathbf{Y}_{0,N-1}^{\mathsf{T}} \end{pmatrix} \in \mathbb{R}^{6N}, \tag{20}$$

$$\bar{\mathbf{Y}}_\Delta^{\mathsf{T}} \equiv \begin{pmatrix} \mathbf{Y}_{\Delta,0}^{\mathsf{T}} & \mathbf{Y}_{\Delta,1}^{\mathsf{T}} \cdots \mathbf{Y}_{\Delta,N-1}^{\mathsf{T}} \end{pmatrix} \in \mathbb{R}^{6N}. \tag{21}$$

Fluctuations of the molecule away from its ground state, i.e., nonzero values of $\bar{\mathbf{Y}}_\Delta$, are characterized by an elastic Hamiltonian $\mathcal{H}(\bar{\mathbf{Y}}_\Delta)$. Since double-stranded DNA is a rather stiff molecule and fluctuations at the base pair-step level are generally small, it is customary to consider the lowest nontrivial order expansion of $\mathcal{H}(\bar{\mathbf{Y}}_\Delta)$, which takes the shape of a quadratic form [1,3,58],

$$\beta\mathcal{H}(\bar{\mathbf{Y}}_\Delta) = \tfrac{1}{2}\bar{\mathbf{Y}}_\Delta^{\mathsf{T}} M_{\bar{\mathbf{Y}}} \bar{\mathbf{Y}}_\Delta, \tag{22}$$

where $M_{\bar{\mathbf{Y}}}$ is a $6N \times 6N$ stiffness matrix and the inverse temperature $\beta = 1/k_{\mathrm{B}}T$ is assumed to be absorbed in the entries of the stiffness matrix. The exact entries of this stiffness matrix again depend on the chosen RBP parameter set. Unless the elasticity is assumed to be homogenous, this stiffness matrix
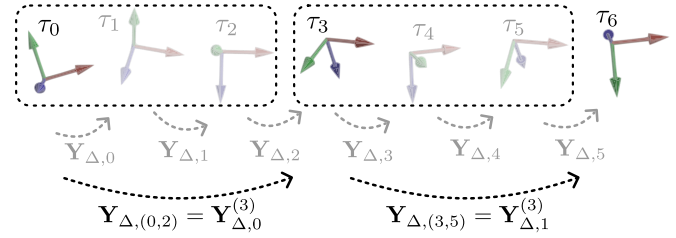

FIG. 2. Illustration of the coarse-graining scheme for $k = 3$. The original chain of base pair reference frames is decimated such that only one in three triads is retained. Deformations of the coarse-grained system are expressed in terms of the junctions $\mathbf{Y}_{\Delta,q}^{(k)}$ connecting the remaining reference frames. These composite junctions each account for the accumulative fluctuation of three original junctions.

will depend on the exact underlying sequence. Most studies define these stiffness matrices for deformations in the $\mathbf{X}$ definition [1,3,65]. The transformation of the stiffness matrix between the $\mathbf{X}$ and the $\mathbf{Y}$ definition of deformations can be carried out analytically by linearizing the transformation. A detailed discussion of this approach is given in Appendix C.

Finally, we note that in many studies the elastic energy is assumed to be local, i.e., fluctuations between distinct junctions are fully decoupled [1,3]. In this case, the elastic energy simplifies to

$$\beta\mathcal{H}(\bar{\mathbf{Y}}_\Delta) = \frac{1}{2}\sum_{i=0}^{N-1} \mathbf{Y}_{\Delta,i}^{\mathsf{T}} M_{\mathbf{Y}_i} \mathbf{Y}_{\Delta,i}, \tag{23}$$

with $M_{\mathbf{Y}_i}$ the individual 6×6 base pair-step stiffness matrices, which can usually be constructed from the 16 canonical dimers (of which only 10 are truly unique [1,3]). Previous work, however, has shown, that the assumption of elastic locality is not satisfied within the framework of the rigid base pair description [58,83,84,86,87,92,93] and that neighbor-couplings have to be considered to appropriately capture the long-range elastic properties [83,94]. Whether these couplings imply physical interactions spanning across such lengths or whether they are mere artifacts introduced by the marginalization process from underlying higher resolution descriptions, is still debated [58,95]. Regardless of the origin of the couplings, they incontrovertibly have to be considered if the correct large-scale elastic behavior is sought to be captured. The coarse-graining procedure outlined in the next section naturally incorporates any such couplings (if present), since it is based on the most general (Gaussian) model, Eq. (22), which may in principle feature couplings at all length scales.

B. Coarse-graining scheme

We propose a k-step coarse-graining scheme that splits the chain into groups of k triads and then eliminates all but the first triad in each group as is illustrated in Fig. 2 for the case of $k = 3$. The new—coarse-grained—junctions connect the remaining triads and therefore span over k of the original junctions. This effectively corresponds to a k-fold increase in the discretization length. Assuming the original molecule to contain $N = N^{(k)}k + 1$ base pairs, the coarse-graining scheme reduces the system to the set of $N^{(k)} + 1$

triads $\{\tau_0, \tau_k, \cdots, \tau_{(N^{(k)}-1)k}, \tau_{N^{(k)}k}\}$ connected by the k-step junctions

$$g_q^{(k)} = \tau_{qk}^{-1}\tau_{(q+1)k}, \tag{24}$$

with $q \in \{0 \ldots (N^{(k)} - 1)\}$. Following the same definition as in the base pair-step case, these junctions factor into static and dynamic components

$$g_q^{(k)} = s_q^{(k)} d_q^{(k)}, \tag{25}$$

each of which is associated with coarse-grained static and dynamic parameters

$$\mathbf{Y}_{0,q}^{(k)} \equiv \mathcal{P}(s_q^{(k)}), \tag{26}$$

$$\mathbf{Y}_{\Delta,q}^{(k)} \equiv \mathcal{P}(d_q^{(k)}). \tag{27}$$

Analogously, to the original system, the entire state of the reduced system is captured by the system-wide static and dynamics state-vectors $\bar{\mathbf{Y}}_0^{(k)}$ and $\bar{\mathbf{Y}}_\Delta^{(k)}$ [see Eqs. (20) and (21)]. This coarse-graining scheme is reminiscent of the decimation mapping from real space renormalization group theory [96].

The coarse-graining procedure itself consists of identifying the coarse-grained ground state $\bar{\mathbf{Y}}_0^{(k)}$ and the coarse-grained Hamiltonian $\mathcal{H}(\bar{\mathbf{Y}}_\Delta^{(k)})$. Calculation of the former is straightforward and will be shown further below. The latter should be identified via the condition that the free energy of the system has to remain unchanged under the transformation, which is equivalent to requiring equal canonical partition functions

$$Z = \int d\bar{\mathbf{Y}}_\Delta e^{-\beta\mathcal{H}(\bar{\mathbf{Y}}_\Delta)} = \int d\bar{\mathbf{Y}}_\Delta^{(k)} e^{-\beta\mathcal{H}(\bar{\mathbf{Y}}_\Delta^{(k)})}. \tag{28}$$

The main difficulty in the latter step stems from the requirement of finding a functional form for $\bar{\mathbf{Y}}_\Delta^{(k)}$ in terms of the original degrees of freedom $\bar{\mathbf{Y}}_\Delta$.

1. Composites

Junctions between two arbitrary frames, as considered in Eq. (24), may be written as a product of the intermediate junctions,

$$g_{[i,j]} \equiv \prod_{l=i}^{j} g_l = \tau_i^{-1}\tau_{j+1}. \tag{29}$$

For such composites, we make the explicit distinction between the notation of composites which may be written as a product (for transformations) or sum (for vectors) of all corresponding intermediate junction elements, indicated by square bracket subscripts $[i, j]$, and those for which we do not a priori assume a trivial junction element decomposition to be possible, indicated by ordinary bracket subscripts (i, j). For example, analogous to Eq. (9), there will be a generator of rotation $\widehat{\Omega}_{(i,j)}$ and a translation vector $\mathbf{w}_{(i,j)}$, such that

$$g_{[i,j]} = \begin{pmatrix} \exp\widehat{\Omega}_{(i,j)} & \mathbf{w}_{(i,j)} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}. \tag{30}$$

While the matrix $g_{[i,j]}$ is by definition a product of single junction matrices, the composite vectors may generally not be written as the sum of the single-junction vectors ($\Omega_{(i,j)} \neq \sum_{l=i}^{j}\Omega_l$ and $\mathbf{w}_{(i,j)} \neq \sum_{l=i}^{j}\mathbf{w}_l$), i.e., they are nonadditive, except for certain elect cases (e.g., when the tangents, $\widehat{\mathbf{t}}_i$, of all
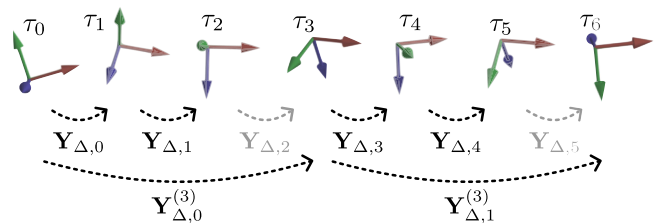

FIG. 3. This example configuration consisting of seven triads illustrates that the full state can be captured by a mixed set of original and coarse-grained dynamic junction parameters. In this example, a particular state of the system is fully parametrized by the set $\{\mathbf{Y}_{\Delta,0}, \mathbf{Y}_{\Delta,1}, \mathbf{Y}_{\Delta,0}^{(k)}, \mathbf{Y}_{\Delta,3}, \mathbf{Y}_{\Delta,4}, \mathbf{Y}_{\Delta,1}^{(k)}\}$ (assuming that the ground-state components are also given).

frames are aligned). For the construction of composites, it is most convenient to use the index range of the original junctions (from i to j) rather than the index of the coarse-grained junction q. Bear in mind, however, that the relationship

$$g_q^{(k)} = g_{[qk,(q+1)k-1]} \tag{31}$$

is implied.

2. Coarse-grained ground state

The coarse-grained ground-state junctions may be written as a product of the contained original junctions [see Eq. (12)] which by employing the notation for accumulative transformations, Eq. (29), can be written as

$$s_q^{(k)} = s_{[qk,(q+1)k-1]} = \prod_{l=qk}^{(q+1)k-1} s_l. \tag{32}$$

The corresponding ground-state vector is then simply

$$\mathbf{Y}_{0,q}^{(k)} = \mathcal{P}(s_q^{(k)}). \tag{33}$$

3. Coarse-grained Hamiltonian

In general, the dynamic coarse-grained state vectors $\bar{\mathbf{Y}}_\Delta^{(k)}$ are not simply a linear combination of the components of the original dynamic state vectors $\bar{\mathbf{Y}}_\Delta$. However, since fluctuations are generally small, we argue that expansion to linear order is warranted. The validity of this approximation will be further justified in the result section below.

As a first step, we note that the state of the original system can be equivalently captured if we substitute for each group of original junctions, a single junction by the coarse-grained junction $\mathbf{Y}_{\Delta,q}^{(k)}$ representing the respective group. For example, a particular configuration of the system depicted in Fig. 3 can either be captured by the set of original deformations $\{\mathbf{Y}_{\Delta,0}, \mathbf{Y}_{\Delta,1}, \mathbf{Y}_{\Delta,2}, \mathbf{Y}_{\Delta,3}, \mathbf{Y}_{\Delta,4}, \mathbf{Y}_{\Delta,5}\}$ or equivalently by the set $\{\mathbf{Y}_{\Delta,0}, \mathbf{Y}_{\Delta,1}, \mathbf{Y}_{\Delta,0}^{(k)}, \mathbf{Y}_{\Delta,3}, \mathbf{Y}_{\Delta,4}, \mathbf{Y}_{\Delta,1}^{(k)}\}$, where the elements $\mathbf{Y}_{\Delta,2}$ and $\mathbf{Y}_{\Delta,5}$ are replaced by the coarse-grained junctions $\mathbf{Y}_{\Delta,0}^{(k)}$ and $\mathbf{Y}_{\Delta,1}^{(k)}$, respectively.

Arranging all the coarse-grained components to the right, this reformulated state vector takes the form

$$\bar{\Xi}_\Delta^{(k)\mathsf{T}} = (\bar{\mathbf{Y}}_{\Delta,\mathrm{rem}}^\mathsf{T} \quad \bar{\mathbf{Y}}_\Delta^{(k)\mathsf{T}}), \tag{34}$$

where $\bar{\mathbf{Y}}_{\Delta,\mathrm{rem}}$ contains all the remaining components, i.e., the ones that have not been substituted. In the example depicted in Fig. 3, $\bar{\mathbf{Y}}_{\Delta,\mathrm{rem}}$ and $\bar{\mathbf{Y}}_{\Delta}^{(k)}$ are given by

$$\bar{\mathbf{Y}}_{\Delta,\mathrm{rem}}^{\intercal} = \begin{pmatrix} \mathbf{Y}_{\Delta,0}^{\intercal} & \mathbf{Y}_{\Delta,1}^{\intercal} & \mathbf{Y}_{\Delta,3}^{\intercal} & \mathbf{Y}_{\Delta,4}^{\intercal} \end{pmatrix}, \qquad (35)$$

$$\bar{\mathbf{Y}}_{\Delta}^{(k)\,\intercal} = \begin{pmatrix} \mathbf{Y}_{\Delta,0}^{(k)\,\intercal} & \mathbf{Y}_{\Delta,1}^{(k)\,\intercal} \end{pmatrix}. \qquad (36)$$

Note that according to the aforementioned approximation the transformation from $\bar{\mathbf{Y}}_{\Delta}$ to $\bar{\mathbf{\Xi}}_{\Delta}^{(k)}$ is a change of basis and there will, thus, be a linear transformation $\bar{A}^{(k)}$ such that

$$\bar{\mathbf{\Xi}}_{\Delta}^{(k)} = \bar{A}^{(k)}\bar{\mathbf{Y}}_{\Delta}. \qquad (37)$$

With this transformation in hand, we can transform the elastic energy Eq. (22) in the partition function as

$$\begin{aligned} Z &= \int d\bar{\mathbf{Y}}_{\Delta} e^{-\frac{1}{2}\bar{\mathbf{Y}}_{\Delta}^{\intercal} M_{\bar{\mathbf{Y}}}\bar{\mathbf{Y}}_{\Delta}} \\ &= \int \frac{d\bar{\mathbf{\Xi}}_{\Delta}^{(k)}}{\det \bar{A}^{(k)}} e^{-\frac{1}{2}\left(\bar{A}^{(k)\,-1}\bar{\mathbf{\Xi}}_{\Delta}^{(k)}\right)^{\intercal} M_{\bar{\mathbf{Y}}}\left(\bar{A}^{(k)\,-1}\bar{\mathbf{\Xi}}_{\Delta}^{(k)}\right)} \\ &= \int d\bar{\mathbf{\Xi}}_{\Delta}^{(k)} e^{-\frac{1}{2}\bar{\mathbf{\Xi}}_{\Delta}^{(k)\,\intercal} M_{\bar{\mathbf{\Xi}}^{(k)}}\bar{\mathbf{\Xi}}_{\Delta}^{(k)} - \log\det\bar{A}^{(k)}}, \qquad (38) \end{aligned}$$

where the the transformed stiffness matrix $M_{\bar{\mathbf{\Xi}}^{(k)}}$ is given by

$$M_{\bar{\mathbf{\Xi}}^{(k)}} = \left(\bar{A}^{(k)\,-1}\right)^{\intercal} M_{\bar{\mathbf{Y}}}\bar{A}^{(k)\,-1}, \qquad (39)$$

and $(\det \bar{A}^{(k)})^{-1}$ is the Jacobian of the transformation.

All that is left to do to obtain the stiffness matrix $M^{(k)}$ of the coarse-grained system is to integrate out the remaining original degrees of freedom, $\bar{\mathbf{Y}}_{\Delta,\mathrm{rem}}$, which is equivalent to marginalizing the stiffness matrix. This can, for example, be achieved by taking the Schur complement [58,84,97] of the matrix $M_{\bar{\mathbf{\Xi}}^{(k)}}$ with respect to $\bar{\mathbf{Y}}_{\Delta}^{(k)}$. By construction, the resulting elastic energy of the coarse-grained system will again be a quadratic form, i.e., it will have the same functional form as the original system.

C. Composite transformation

In this section, we derive the composite transformation $\bar{A}^{(k)}$ from Eq. (37). The first step will be to express dynamic composite vectors $\mathbf{Y}_{\Delta,(i,j)}$ in terms of the original dynamic vectors $\{\mathbf{Y}_{\Delta,i},\ldots,\mathbf{Y}_{\Delta,j}\}$, and after proper expansion we identify the linear transformation for single composites

$$\mathbf{Y}_{\Delta,(i,j)} = \sum_{l=i}^{j} A_{l}^{(i,j)}\mathbf{Y}_{\Delta,l}. \qquad (40)$$

Finally, these transformations are combined to construct the full system transformation $\bar{A}^{(k)}$.

1. Dynamic composite

We seek to express the dynamic composites

$$d_{(i,j)} = \begin{pmatrix} \mathcal{D}_{(i,j)} & \mathbf{d}_{(i,j)} \\ \mathbf{0}^{\intercal} & 1 \end{pmatrix}, \qquad (41)$$

with $\mathcal{D}_{(i,j)} = \exp\widehat{\Phi}_{\Delta,(i,j)}$, in terms of the single junction static and dynamic components

$$s_i = \begin{pmatrix} \mathcal{S}_i & \mathbf{s}_i \\ \mathbf{0}^{\intercal} & 1 \end{pmatrix}, \quad d_i = \begin{pmatrix} \mathcal{D}_i & \mathbf{d}_i \\ \mathbf{0}^{\intercal} & 1 \end{pmatrix}, \qquad (42)$$

where $\mathcal{S}_i = \exp\widehat{\Phi}_{0,i}$, and $\mathcal{D}_i = \exp\widehat{\Phi}_{\Delta,i}$ [see Eqs. (12), (19), and (16)]. Using Eq. (25) we can write $d_{(i,j)} = s_{[i,j]}^{-1}g_{[i,j]}$ and after working out the right-hand side explicitly in terms of the components of Eqs. (42) one finds

$$\mathcal{D}_{(i,j)} = \mathcal{S}_{[i,j]}^{\intercal}\mathcal{R}_{[i,j]}, \qquad (43)$$

and

$$\begin{aligned} \mathbf{d}_{(i,j)} &= \mathcal{S}_{[i,j]}^{\intercal}\mathbf{w}_{(i,j)} - \mathcal{S}_{[i,j]}^{\intercal}\mathbf{s}_{(i,j)} \\ &= \mathcal{S}_{[i,j]}^{\intercal}\sum_{l=i}^{j}\left(\prod_{m=i}^{l-1}(\mathcal{S}_m\mathcal{D}_m)(\mathcal{S}_l\mathbf{d}_l + \mathbf{s}_l)\right) - \sum_{l=i}^{j}\mathcal{S}_{[l,j]}^{\intercal}\mathbf{s}_l. \end{aligned}$$
$$(44)$$

We will calculate the rotational and translational components of $\mathbf{Y}_{\Delta,(i,j)}$ separately, starting with the former.

2. Rotation

A straightforward calculation shows that Eq. (43) can be rewritten as

$$\begin{aligned} \mathcal{D}_{(i,j)} &= \prod_{l=i}^{j}\left(\mathcal{S}_{[l+1,j]}^{\intercal}\mathcal{D}_l\,\mathcal{S}_{[l+1,j]}\right) \\ &= \prod_{l=i}^{j}\exp(\mathcal{S}_{[l+1,j]}^{\intercal}\widehat{\Phi}_{\Delta,l}\mathcal{S}_{[l+1,j]}), \qquad (45) \end{aligned}$$

where the last equality follows directly from the properties of matrix exponentials. To arrive at an expression for a single rotation matrix comprising all the fluctuating components, we make the approximation

$$\mathcal{D}_{(i,j)} \approx \exp\left(\sum_{l=i}^{j}\mathcal{S}_{[l+1,j]}^{\intercal}\widehat{\Phi}_{\Delta,l}\mathcal{S}_{[l+1,j]}\right), \qquad (46)$$

which is equivalent to discarding all terms of higher than linear order in the respective series expansions (or equivalently, ignoring all commutators in the Baker-Campbell-Hausdorff formula). Such an expansion is justified, because we separated the fluctuating components, which are assumed to be small, from the static components that assume appreciable values—especially intrinsic twist. This explains the unconventional choice of expressing the rigid base pair model in $\bar{\mathbf{Y}}$ coordinates rather than the usual $\bar{\mathbf{X}}$ coordinates. The fluctuational component of the composite step rotation matrix can therefore be written as a sum of transformed—approximately additive—generators

$$\widehat{\Phi}_{\Delta,l}' = \mathcal{S}_{[l+1,j]}^{\intercal}\widehat{\Phi}_{\Delta,l}\mathcal{S}_{[l+1,j]} = \mathrm{hat}\left(\mathcal{S}_{[l+1,j]}^{\intercal}\Phi_{\Delta,l}\right), \qquad (47)$$

where the second equality follows from general properties of rotation generators. Invoking the vector map, Eq. (5), shows that these transformed and approximately additive rotation vectors are obtained by rotating the original rotation vectors

$$\Phi_{\Delta,l}' = \mathcal{S}_{[l+1,j]}^{\intercal}\Phi_{\Delta,l}. \qquad (48)$$

Finally, summing over all the terms is Eq. (46) yields the sought transformation for the rotational components

$$\boldsymbol{\Phi}_{\Delta,(i,j)} \approx \sum_{l=i}^{j} \mathcal{S}_{[l+1,j]}^{\mathsf{T}} \boldsymbol{\Phi}_{\Delta,l}. \qquad (49)$$

3. Translation

To arrive at the expression for composite translations we expand all occurrences of the rotation matrices $\mathcal{D}$ in Eq. (44) to linear order

$$\mathcal{D} = \exp \widehat{\boldsymbol{\Phi}}_{\Delta} \approx \mathbb{1} + \widehat{\boldsymbol{\Phi}}_{\Delta}, \qquad (50)$$

which is again warranted as we assumed the components of $\boldsymbol{\Phi}_{\Delta,k}$ to be small. Discarding all terms of higher than linear order in any of the fluctuating components (both rotation and translation), one eventually arrives at (see Appendix E for details),

$$\mathbf{d}_{(i,j)} \approx \sum_{l=i}^{j-1} \left[\sum_{m=l+1}^{j} \mathrm{hat}\big(\mathcal{S}_{[m,j]}^{\mathsf{T}} \mathbf{s}_m^{\mathsf{T}}\big) \mathcal{S}_{[l+1,j]}^{\mathsf{T}} \right] \boldsymbol{\Phi}_{\Delta,l}$$
$$+ \sum_{l=i}^{j} \mathcal{S}_{[l+1,j]}^{\mathsf{T}} \mathbf{d}_l. \qquad (51)$$

4. Constructing the linear transformation

Jointly, the results of Eqs. (49) and (51) may be summarized in the form of the linear transformation Eq. (40), with

$$A_l^{(i,j)} = \begin{pmatrix} A_{l,\mathrm{rr}}^{(i,j)} & 0 \\ A_{l,\mathrm{rt}}^{(i,j)} & A_{l,\mathrm{tt}}^{(i,j)} \end{pmatrix} \qquad (52)$$

and entries

$$A_{l,\mathrm{rr}}^{(i,j)} = \mathcal{S}_{[l+1,j]}^{\mathsf{T}}, \qquad (53)$$

$$A_{l,\mathrm{tt}}^{(i,j)} = \mathcal{S}_{[l+1,j]}^{\mathsf{T}}, \qquad (54)$$

$$A_{l,\mathrm{rt}}^{(i,j)} = \sum_{m=l+1}^{j} \mathrm{hat}\big(\mathcal{S}_{[m,j]}^{\mathsf{T}} \mathbf{s}_m^{\mathsf{T}}\big) \mathcal{S}_{[l+1,j]}^{\mathsf{T}}. \qquad (55)$$

To determine the transformation $\bar{A}^{(k)}$ from Eq. (37) we will first construct the matrix of basis change that transforms dynamic state vectors spanning over a single compound, i.e., from junction i to junction j, from the original basis

$$\tilde{\mathbf{Y}}_{\Delta,(i,j)}^{\mathsf{T}} = \big(\mathbf{Y}_{\Delta,i}^{\mathsf{T}} \quad \cdots \quad \mathbf{Y}_{\Delta,j-1}^{\mathsf{T}} \quad \mathbf{Y}_{\Delta,j}^{\mathsf{T}}\big), \qquad (56)$$

to the basis in which the last entry is substituted by the respective dynamic compound vector

$$\tilde{\boldsymbol{\Xi}}_{\Delta,(i,j)}^{\mathsf{T}} = \big(\mathbf{Y}_{\Delta,i}^{\mathsf{T}} \quad \cdots \quad \mathbf{Y}_{\Delta,j-1}^{\mathsf{T}} \quad \mathbf{Y}_{\Delta,(i,j)}^{\mathsf{T}}\big). \qquad (57)$$

This transformation takes the form

$$\tilde{\boldsymbol{\Xi}}_{\Delta,(i,j)} = A^{(i,j)} \tilde{\mathbf{Y}}_{\Delta,(i,j)}, \qquad (58)$$

with

$$A^{(i,j)} = \begin{pmatrix} \mathbb{1} & 0 & \dots & 0 & 0 \\ 0 & \mathbb{1} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbb{1} & 0 \\ A_i^{(i,j)} & A_{i+1}^{(i,j)} & \dots & A_{j-1}^{(i,j)} & \mathbb{1} \end{pmatrix}. \qquad (59)$$

The right-most entry in the bottom row is identity since $A_j^{(i,j)} = \mathbb{1}$. The transformation of single compounds can be extended to the entire system via the block-diagonal matrix

$$\bar{A}'^{(k)} = \begin{pmatrix} A^{(0,k-1)} & 0 & \dots & 0 \\ 0 & A^{(k,2k-1)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A^{([N^{(k)}-1]k,N^{(k)}k-1)} \end{pmatrix}. \qquad (60)$$

Finally the $\bar{A}^{(k)}$ from Eq. (37) is obtained by rearranging the elements in $\bar{A}'^{(k)}$ via a permutation that arranges all the coarse-grained components to the back of the vector. This can be achieved by invoking a permutation matrix P,

$$\bar{A}^{(k)} = P\bar{A}'^{(k)}. \qquad (61)$$

Since $\bar{A}'^{(k)}$ is a block diagonal matrix with component blocks $A^{(i,j)}$, each of which satisfying $\det A^{(i,j)} = 1$, the transformation matrix $\bar{A}^{(k)}$ will also have unit determinant as per the orthogonality of permutation matrices.

III. RESULTS AND DISCUSSION

In this section, we assess the fidelity of the coarse-graining procedure by comparing various parameters sampled at different levels of coarse-graining. For this purpose a particular RBP model has to be chosen that provides sequence-dependent ground state and stiffness parameters. Throughout this work, we relied on parameters provided by the cgNA+ model [65,98]. We then compare results sampled at the original and coarse-grained resolutions. We emphasize that the purpose is not to assess the quality of the used parameter set but the agreement of the results obtained at different resolutions.

Note, that cgNA+ extends the RBP model by incorporating additional degrees of freedom beyond the six base pair-step parameters (tilt, roll, twist, shift, slide, and rise). Specifically, it treats bases and phosphate groups as rigid bodies, resulting in an additional six intra-base pair parameters (buckle, propeller, opening, shear, stretch, and stagger), along with six degrees of freedom encoding the orientation and position of the phosphate group relative to the corresponding base in each nucleotide. Stiffness matrices for the RBP model are obtained by marginalizing these excess degrees of freedom, employing appropriate Schur complements as outlined in previous work [58].

In the cgNA+ model, rotations are represented using the Cayley map. However, our coarse-graining procedure necessitates parameters expressed in terms of the Euler map. We therefore transform the ground state and the stiffness matrices to this representation (further details provided in Appendix B). Our approach assumes translations between pairs of triads

to be defined in the frame of the first triad of the respective pair, while cgNA+ expresses translations in the corresponding midstep frame. Additionally, as discussed in Sec. III A, we segregate static and dynamic components of rotations and translations at the transformation matrix level, rather than at the corresponding vector level, as commonly done in RBP models. To use the language and symbols introduced in this work we require the system to be parametrized in terms of $\bar{\mathbf{Y}}$ coordinates instead of the more common $\bar{\mathbf{X}}$ coordinates. This transformation of the ground state and stiffness from midstep triad to triad definition of translations, along with the redefinition of fluctuating components, is achieved via the transformations derived in Appendixes C and D. The outcome is a set comprising the ground state and a stiffness matrix $(\bar{\mathbf{Y}}_0, M_{\bar{\mathbf{Y}}})$, which are then transformed into coarse-grained sets $(\bar{\mathbf{Y}}_0^{(k)}, M_{\bar{\mathbf{Y}}}^{(k)})$ using the scheme outlined in Sec. II C 4.

A. Comparison with unrestrained Monte Carlo sampling

We first consider ensembles of unrestrained configurations ranging over single composite steps for composite sizes k ranging from 2 to 40. The analysis is based on a randomly generated 201 bp sequence [99] (containing $N = 200$ base pair steps) for which we consider partial oligomers selected between indices 80 and $80 + k$. The larger sequence context was introduced to avoid boundary effects. We generated a sample of 2.5×10^7 unrestrained configurations drawn according to the canonical measure. Specifically, we drew $\bar{\mathbf{Y}}_\Delta \in \mathbb{R}^{6N}$ from the multivariate Gaussian distribution

$$\rho(\bar{\mathbf{Y}}_\Delta) = \left(\frac{\det M_{\bar{\mathbf{Y}}}}{(2\pi)^{6k}} \right)^{\frac{1}{2}} e^{-\frac{1}{2}\bar{\mathbf{Y}}_\Delta^\mathsf{T} M_{\bar{\mathbf{Y}}} \bar{\mathbf{Y}}_\Delta}, \quad (62)$$

where the inverse temperature $\beta = (k_\mathrm{B}T)^{-1}$ is absorbed in $M_{\bar{\mathbf{Y}}} \in \mathbb{R}^{6N \times 6N}$. Configurations of triads are then generated by starting from a given first triad and iteratively applying static and dynamic base pair step junction transformations [see Eqs. (9), (12), (15), and (16)]. For each individual value of k, we then considered the relative position and orientation of the triads τ_{80} and τ_{80+k} encapsulated in the relative transformation $g^{(k)} = \tau_{80}^{-1}\tau_{80+k}$. Using the known composite ground-state transformation $s^{(k)}$ [see Eq. (32)] the dynamic composite components $\mathbf{Y}_\Delta^{(k)} = \mathcal{P}(d^{(k)})$ are then found as the parameters associated with the transformation

$$d^{(k)} = s^{(k)\,-1} g^{(k)} = s^{(k)\,-1} \tau_0^{-1} \tau_k. \quad (63)$$

According to the assumption of approximate linearity of the composite transformation, the six individual components of $\mathbf{Y}_\Delta^{(k)}$—for the sake of simplicity we will also call them tilt, roll, twist, shift, slide, and rise—should be multivariate Gaussian distributed. Histograms of the sampled values for composites composed of 10 junctions ($k = 10$) are shown in Fig. 4(a). For comparison, we coplot the distributions resulting from our analytically coarse-grained stiffness matrices $M_{\bar{\mathbf{Y}}}^{(k)} \in \mathbb{R}^{6 \times 6}$. The variances of individual degrees of freedom are obtained via marginalization. Following the construction of the $\mathbf{Y}_\Delta^{(k)}$ as deviations away from the ground state, their mean is assumed to be zero. Both the assumption of Gaussianity and the agreement between sampled and

predicted distribution are excellent in most components, except for the rise component, which exhibits pronounced left-skewness. We attribute this broken symmetry to the influence of bending fluctuation on the accumulative rise. Analogous to the entropic spring behavior of flexible and semiflexible polymers [16,19], local bending fluctuations reduce the end-to-end distance of a given oligomer; see Fig. 4(c) and Ref. [39]. This effect diminishes the composite rise between its two end triads thereby imposing left-skewness on the corresponding distribution. Accounting for this effect in the coarse-grained representation would require the introduction of cubic terms or higher-order in the elastic energy, which could be achieved by a perturbative calculation. Such higher-order treatment would significantly increase the complexity of the parameter transformation and is beyond the scope of this work.

For quantitative assessment we evaluated the variance, mean, and skewness of the six components of $\mathbf{Y}_\Delta^{(k)}$ for all considered values of k; shown in Figs. 4(d), 4(e), and 4(f), respectively. To put differences in variance into perspective we show relative differences between sampled and analytically coarse-grained variances as given by

$$\Delta \mathrm{var}_\mathrm{rel} = \frac{\mathrm{var}_\mathrm{cg} - \mathrm{var}_\mathrm{sampled}}{\mathrm{var}_\mathrm{sampled}}. \quad (64)$$

Since $\langle \mathbf{Y}_\Delta^{(k)} \rangle = 0$ by construction of the coarse-graining procedure, only the sampled mean values are displayed. Likewise, the analytical coarse-graining is constructed as a linear transformation which transforms the original Gaussian system into another Gaussian system. Therefore, the third central moment of the transformed system is assumed to be zero. Omitted higher-order contributions may lead to nonvanishing third central moments. Figure 4(f) shows the normalized sampled third central moments

$$\tilde{\mu}_3 = \frac{\langle (x - \langle x \rangle)^3 \rangle}{\langle (x - \langle x \rangle)^2 \rangle^{3/2}}, \quad (65)$$

which is sometimes referred to as Fisher's moment coefficient of skewness.

Rotational degrees of freedom display excellent agreement between sampled and predicted values, with relative differences in the variance of less than 2.5% for the largest considered coarse-graining size ($k = 40$). Mean and skewness of these values is generally small, with only the mean of the twist exhibiting increasing deviations, reaching about 0.58 deg for the largest considered coarse-graining size. This value constitutes a mere 3% of the standard deviation of the twist fluctuations at this level of coarse-graining.

Deviations between transformed and sampled distributions are significantly larger for the translational degrees of freedom. While the relative difference in variance reaches about 20% for the two lateral components shift and slide, deviations in rise compound to almost 100% for the largest considered composite size. This large deviation stems from the aforementioned symmetry breaking in rise fluctuations that manifests in appreciable left-skewness of the distribution. As a direct consequence, one observes a lower mean and a significantly larger variance.
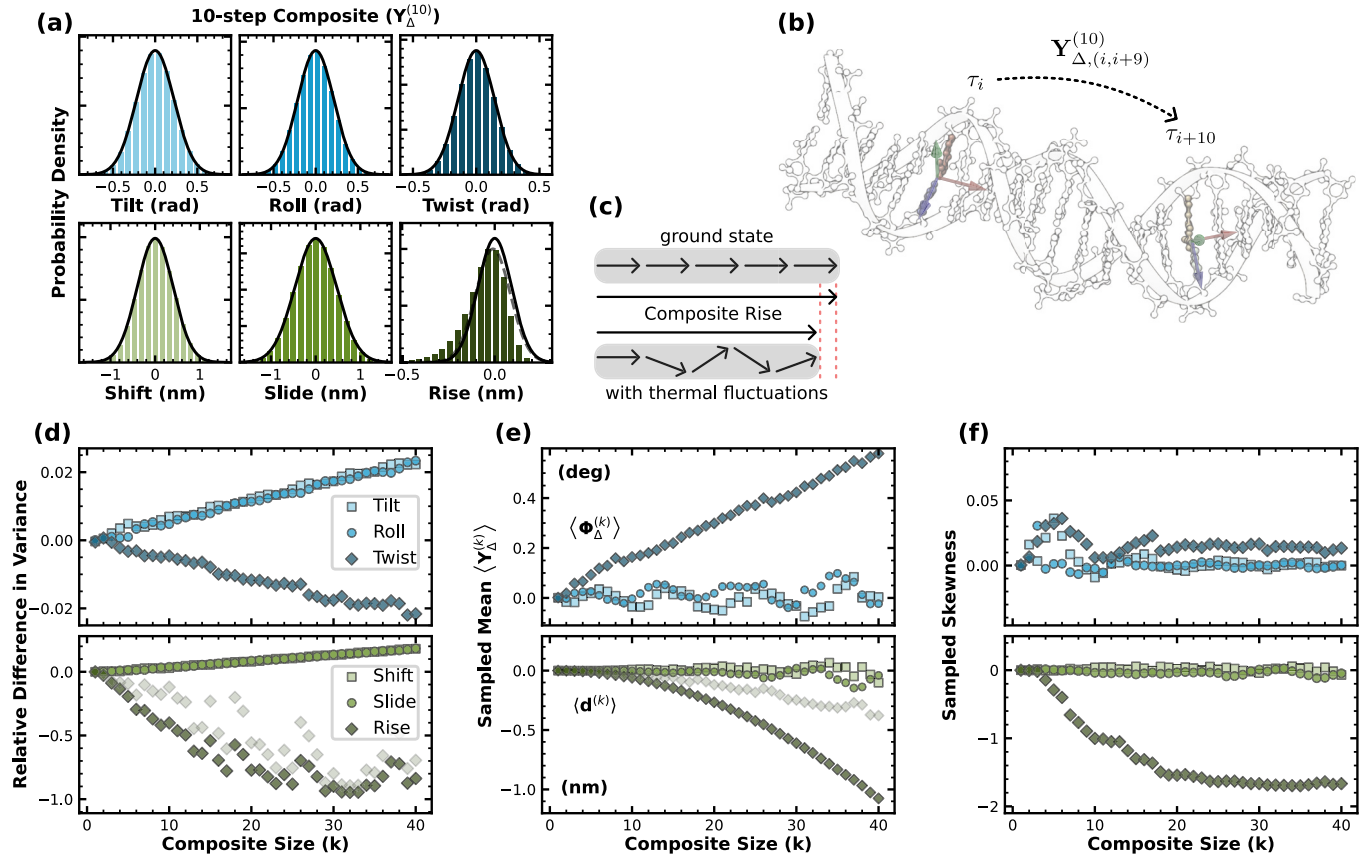
FIG. 4. Comparison between sampled distributions and those resulting from the analytical coarse-graining procedure. (a) Histograms for the six degrees of freedom of a single 10-step composite. The black lines are normal distributions with variances taken from the diagonal entries of the covariance matrix associated with the stiffness matrix of the coarse-grained system (the two matrices are related via matrix-inversion). (b) Depiction of the coarse-grained system represented in panel (a). (c) An illustration of the reduction in composite rise due to thermal bending fluctuations, which lead to the observed symmetry breaking in its distribution. Arrows indicate tangents; bending between consecutive segments leads to a contraction of the end-to-end distance. (d) Relative difference between the variance resulting from the coarse-graining procedure and the sampled variance according to Eq. (64) for values of k ranging from 2 to 40. Means and skewnesses [see Eq. (65)] of the same distributions are displayed in panels (e) and (f), respectively. Transparent scatters in panels (d) and (e) correspond to Gaussian fits centered at the value of maximum likelihood.

For closer comparison to the quadratic model, we fit the sampled rise distribution with a Gaussian centered around the most likely rise value (results shown as transparent scatters). This way of comparison reveals discrepancies to be less severe, especially for composite-step sizes up to $k = 10$.

B. Persistence lengths

The bending and torsional persistence lengths are the canonical measures of DNA elasticity quantifying the mechanical response of the molecule at the mesoscale. There are a variety of different definitions for the bending persistence length [74]. Here we employ a definition based on the exponential decay of the tangent-tangent correlation function [100]

$$\mathcal{C}_b(m) = \langle \widehat{\mathbf{t}}_n \cdot \widehat{\mathbf{t}}_{n+m} \rangle = e^{-\frac{am}{l_b}}, \qquad (66)$$

where subscripts are the positional indices of the considered tangents with m being the curvilinear distance expressed in

base pair-steps, and a is the discretization length, which we set to $a = 0.34$ nm.

Regular exponential decay of the tangent-tangent correlation function, as indicated in Eq. (66), is a behavior exhibited only by semiflexible polymers characterized by purely local elasticity [83]—which is, for example, the case for an elastic Hamiltonian of the form given by Eq. (23) and no structural features [92]. Previous work has shown that nonlocality in the elastic energy, i.e., couplings between neighboring junctions and beyond, give rise to a distinct length-scale dependence of the elastic properties [83,92,93,101,102]. Moreover, intrinsic bending components are known to induce additional deviations from the exponential behavior [92]. We, therefore, employ a length-scale-dependent definition of the persistence length based on Eq. (66), but evaluated for every m individually [87],

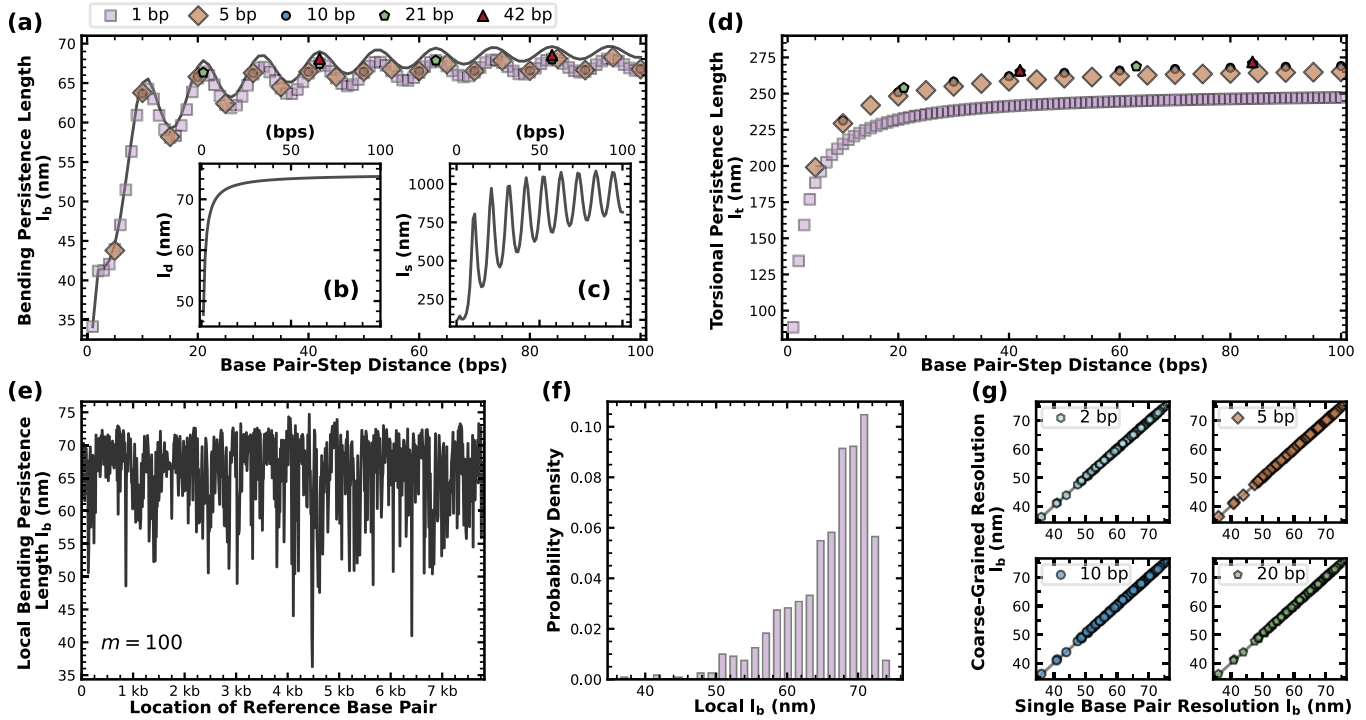$$l_b(m) = -\frac{am}{\log \mathcal{C}_b(m)}. \qquad (67)$$

FIG. 5. Persistence lengths of a 7922 bp sequence. (a) Sequence-averaged bending persistence length simulated at varying resolutions, indicated by different scatter symbols. The solid line represents the persistence length calculated using the Trifanof–Tan–Harvey expression [103], Eq. (70), based on separately calculated dynamic (b) and static (c) persistence lengths. (d) Sequence-averaged torsional persistence length calculated from the same simulations as in panel (a). (e) Sequence-specific persistence lengths obtained by sliding the reference tangent [n in Eq. (66)] through the 7922 bp molecule, calculated at a fixed base pair-step distance $m = 100$. (f) Histogram of the values shown in panel (e). (g) Cross-correlation plots between persistence lengths calculated at single base pair and coarse-grained resolutions (2, 5, 10, and 20 bp, respectively).

Analogous to Eq. (66) the twist-elasticity is characterized by the twist-correlation function [83]

$$\mathcal{C}_t(m) = \left\langle \cos\left(\sum_{i=n}^{n+m-1} [\boldsymbol{\Phi}_{\Delta,i}]_3 \right) \right\rangle = e^{-\frac{am}{l_t}}, \qquad (68)$$

where $[\boldsymbol{\Phi}_{\Delta,i}]_3$ is the third component of the excess rotation $\boldsymbol{\Phi}_{\Delta,i}$. The associated decay length, the torsional persistence length l_t, is related to the torsional stiffness by a factor of two: $l_t = 2C$ [50]. Equation (68) again only holds for twist-storing polymers with purely local elastic couplings. Nonlocality of the elastic energy breaks the monotonous decay of the twist-correlation function and introduces length-scale dependence, warranting a local definition analogous to the expression for the bending persistence length

$$l_t(m) = -\frac{am}{\log \mathcal{C}_t(m)}. \qquad (69)$$

To highlight the utility of the coarse-graining procedure, we compare the original base pair resolution model to coarse-grained parametrizations of various resolutions. In particular, we considered composites spanning 5, 10, 21, and 42 base pairs, corresponding to roughly half, one, two, and four helical repeats, respectively.

We first consider the sequence-averaged length-scale-dependent persistence lengths of a 7922 base pair sequence that has been employed in various experimental single-molecule studies [22,27,104]. In this context, the expectation

brackets in Eq. (66) indicate simultaneous thermal and sequence averages, i.e., the expression is averaged over all possible reference indices n, provided that the molecule contains at least $n + m$ base pairs. Following the procedure introduced in the previous section, we generated 10^8 independent configurations for each resolution (i.e., for each value of k).

Length-scale-dependent bending persistence lengths for base pair-step distances (m) ranging from 1 to 100 are displayed in Fig. 5(a). Consistent with previously reported findings, the simulation at base pair resolution ($k = 1$), exhibits low persistence, i.e., enhanced flexibility, at short distances (small m) while asymptotically converging towards larger stiffness for large m [83,84,87,92,101,102,105]. Moreover, the behavior of $l_b(m)$ is further modulated by a sinusoidal oscillation of wavelength corresponding to the helical repeat length (approximately 10.5 bp). This is a consequence of the base pair planes being slightly tilted on average relative to the helical axis (see, for example, Ref. [92]). Asymptotically, $l_b(m)$ converges to about 69 nm. While this value is significantly larger than the literature values of 40–55 nm, it is in line with previous studies reporting on the persistence length of cgNA+ [98,102] and similar to values found with atomistic simulations [83] conducted with the parmbsc1 forcefield [30] based on which cgNA+ is parametrized.

We emphasize again that this section is not intended to assess the accuracy of the underlying RBP parametrization, but rather to validate the fidelity of the methodology in

reproducing configurational fluctuations at the coarse-grained level. As depicted in Fig. 5(a), the coarse-grained parametrizations yield persistence length values that closely align with those of the original single base pair resolution model.

According to a heuristic argument due to Trifanof–Tan–Harvey [103] the bending persistence length may be expressed in terms of static and dynamic components, l_s and l_d, respectively, as

$$\frac{1}{l_{\rm b}} = \frac{1}{l_s} + \frac{1}{l_d}. \tag{70}$$

We tested the validity of this expression by calculating l_s and l_d independently. The dynamic persistence length l_d is obtained by running the Monte Carlo simulations with static components set to zero and then applying Eq. (67) to the generated ensemble of configurations [see Fig. 5(b)]. The static persistence length l_s, shown in Fig. 5(c), is obtained by applying the definition Eq. (67) to the static ground state of the 7922 base pair molecule. Constructing the bending persistence length according to Eq. (70) yields values very close to the ones found by the full simulations [see solid gray line in Fig. 5(a)].

The values of the twist-persistence length [see Fig. 5(d)] exhibit less agreement between simulations at different resolutions: at base pair resolution $l_t(m)$ converges to approximately 250 nm ($C \approx 125$ nm), whereas at lower resolutions $l_t(m)$ converges to somewhat larger values ($C \approx 135$–140 nm). We propose that these discrepancies do not arise from intrinsic deficiencies in the coarse-graining procedure, but rather from an inaccurate definition of twist as the third component of the junction rotation vector. For twist to be additive, it should quantify rotation around the helical axis of the molecule. However, due to fluctuations in the tangents of the reference frame away from the molecular center line and helical variations in base planes, twist fluctuations at larger length scales are influenced by geometric and topological features [106]. This point will be further illustrated in the following section.

Finally, we investigated the extent to which the coarse-graining methodology captures the sequence-specificity of the bending persistence length. To achieve this, we calculated the persistence length over individual segments of the same 7922 base pair sequence. Specifically, we evaluated Eq. (67) for fixed reference points $n = 0, 10, 20, \ldots, 7820$, at distance $m = 100$, which we assume to be close to the asymptotic value. As depicted in Fig. 5(e) the persistence length is found to exhibit appreciable positional dependence. This variability is further illustrated in the distribution of individual values; see histogram, Fig. 5(f). To directly compare the results at different resolutions with those obtained at single base pair resolution, we generated cross-correlation plots for coarse-grained resolutions of 2, 5, 10, and 20 base pairs, as depicted in Fig. 5(g). These plots demonstrate the near-perfect agreement between the results at various resolutions and highlight the efficacy of the coarse-graining methodology in capturing detailed sequence-dependent features.

C. Force extension and effective torsional stiffness

As an example for simulations more closely related to experiments, we consider the setup of freely orbiting magnetic tweezers (FOMT) [73] [see Fig. 6(a)]. In such experiments, a single double-stranded DNA molecule is tethered between a flow cell surface and a superparamagnetic magnetic bead. Linear stretching forces **f** can be applied to the DNA tether by proxy of exposing the bead to magnetic field gradients induced by a magnet located above the bead.

First, we will summarize the theory essential for rationalizing the experimental readout and extracting bending and torsional properties. Following this, we present the results obtained from extensive Monte Carlo simulations sampled at various resolutions.

1. Force extension

Measurement of the hovering height of the bead relative to the surface gives access to the instantaneous tether extension z. Elongation of the molecule is opposed by an accompanying loss in entropy, which results in a characteristic force-dependent equilibrium. For sufficiently large forces ($f \gtrsim 0.3$ pN) small fluctuation theory provides a simple expression for the mean extension $\langle z \rangle$ in terms of force f, tether length L, and the bending persistence length $l_{\rm b}$ [19,20],

$$\frac{\langle z \rangle}{L} = 1 - \frac{1}{2}\sqrt{\frac{k_{\rm B}T}{l_{\rm b}f}} + {\rm h.o.}, \tag{71}$$

where h.o. indicates higher order corrections.

To cover the full range of forces, several interpolation formulas that continuously connect the response at large forces to the behavior of ideal flexible polymers at low forces have been brought forward [19,107]. The first and possibly simplest of these was suggested by Marko and Siggia [19]

$$\frac{l_{\rm b}f}{k_{\rm B}T} = \frac{\langle z \rangle}{L} + \frac{1}{4(1 - \langle z \rangle/L)^2} - \frac{1}{4}. \tag{72}$$

Equations (71) and (72) allow for the fitting of the bending persistence length based on measured force-extension data.

2. Effective torsional stiffness

Suitable positioning of the magnet allows for the monitoring of rotations of the bead around the force director field (for details see Ref. [73]). Effectively, this gives access to the torsional fluctuations of the molecule as a whole, which can be applied to assess the molecule's innate torsional properties. However, rotation of the bead, or equivalently the molecular terminus, does not translate into an equal amount of accumulative twist strain distributed over the molecular contour. Instead, a part of the torsional strain will be absorbed in the form of writhe [108], which, simply put, is a form of chiral bending. As long as the total torsional strain is small and the extending force sufficiently large [21,109,110], writhe manifests in small helical fluctuations around the fully extended state. Beyond a certain force-dependent threshold torsional strain, these fluctuations compound to initiate a buckling transition, which leads part of the molecule to wrap around itself in superhelical conformations called
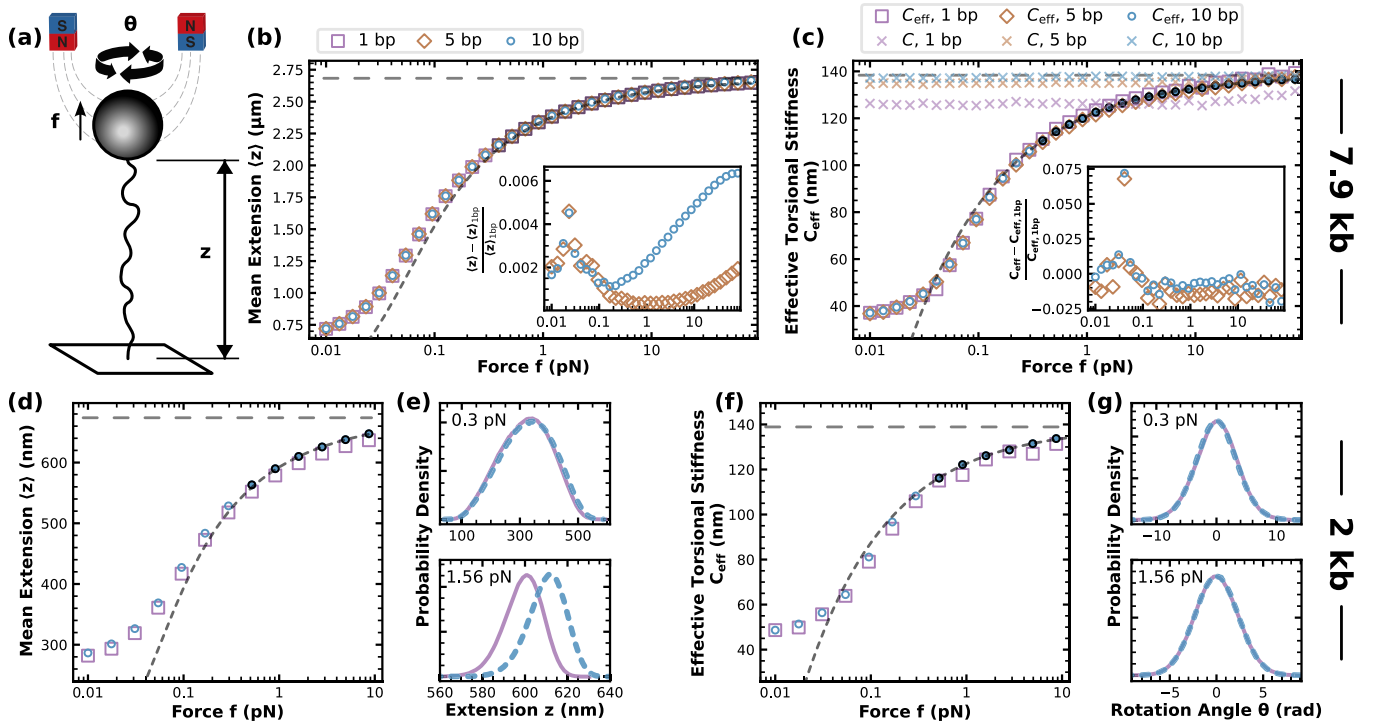
FIG. 6. Monte Carlo simulations of freely orbiting magnetic tweezer setup. (a) An illustration of a freely orbiting magnetic tweezer. (b), (c) Comparison of force-extension (b) and the effective torsional stiffness (c) for simulations of 7.9 kb molecules for three different resolutions. The underlying energy for these simulations only contains the rotational degrees of freedom tilt, roll, and twist and keeps the discretization length constant. Relative differences between the coarse-grained results from the single base pair-step resolution simulations are highlighted in the insets. The curved dashed line in panel (b) represents the interpolation formula, Eq. (72), using a value of $l_b = 68.4 \pm 0.2$ nm obtained by fitting Eq. (71) to the 1 bp resolution data for forces exceeding 0.3 pN. The horizontal dashed line indicates the length of the molecule, marking the maximum tether extension. In panel (c), the curved dashed line represents a fit of Eq. (74) to the 10 bp resolution data again for forces larger than 0.3 pN. Here, the horizontal dashed line indicates the fitted value of C, which is asymptotically approached by C_{eff} at large forces. The x-scatter points represent direct torsional stiffness C measurements, calculated from the total twist fluctuations according to Eq. (76). (d)–(g) 2 kb simulations including the translational degrees of freedom shift, slide, and rise. Data corresponding to the 1 bp resolution are shown in purple and those corresponding to 10 bp resolution are shown in blue. Panels (e) and (g) show probability distributions of extension and rotation angle θ, respectively, for two forces. The solid purple and dashed blue lines show the distributions observed for single-base pair and 10-base pair resolutions simulations, respectively.

plectonemes [28,109,111,112]. In this section, we are only interested in the former regime.

Absorption of torsional strain into writhe results in an effectively reduced torsional stiffness—usually referred to as effective torsional stiffness C_{eff}—relative to the angular fluctuations of the molecular termini [69],

$$\langle \Delta \theta^2 \rangle = \frac{L}{C_{\text{eff}}}, \tag{73}$$

where θ is the total angle traced out by the magnetic bead which is equal to the torsional fluctuation of the molecular endpoint. At low forces, writhe fluctuations are appreciable, leading to values of C_{eff} that are significantly lower than DNA's innate torsional stiffness C. Conversely, at large forces, bending fluctuations are largely suppressed, such that rotations of the magnetic bead are predominantly dictated by the torsional stiffness of the molecule. Moroz and Nelson showed that in the large force regime (once again $f \gtrsim 0.3$ pN for literature values of DNA elastic parameters) the effective

torsional stiffness can be expanded as [21,76]

$$C_{\text{eff}} = C - \frac{C^2}{4 l_b} \sqrt{\frac{k_B T}{l_b f}} + \text{h.o.} \tag{74}$$

Theoretical treatment in the low-force regime is also possible [113,114], albeit not in closed form. However, such analysis is beyond the scope of this work.

3. Limitations arising due to external forces

The coarse-graining procedure eliminates junctions by integrating out the corresponding degrees of freedom assuming them to be freely fluctuating. This assumption is violated by the application of external forces, which constitute effective constraints on the junction degrees of freedom. In principle, one would have to consider these constraints when carrying out the integration. However, linear stretching forces couple strongest to large length scale fluctuations while perturbing local junction fluctuations to much lesser degree [110,115]. This observation is conveniently expressed in the high force correlation length of lateral fluctuations [19], i.e.,

perpendicular to the force-director field,

$$\xi = \sqrt{\frac{k_B T l_b}{f}}, \tag{75}$$

which provides the length scale beyond which bending fluctuations are appreciably attenuated by the force. If the size of the composite segments exceeds this correlation length, then fluctuations within each such segment may no longer be assumed to be freely fluctuating, rendering the coarse-graining approach inadequate for accurately capturing the system's fluctuations. Conversely, for a given composite segment size, Eq. (75) provides an estimate for the upper limit of the applicable force range. For a persistence length of $l_b = 50$ nm and composite segments spanning 10 bp, this upper force limit is about 18 pN.

4. Monte Carlo simulations

We simulated the FOMT setup for the same 7.9 kb sequence [22,27,104] used for the calculation of persistence lengths using the Markov chain Monte Carlo method previously employed in the study of DNA plectonemes [27,28] (for more information see also Supplemental Material of Ref. [27]; code available at Ref. [116]). The underlying energetic model is identical to the RBP description discussed here, with the limitation that it only considers rotational fluctuations, assumes a constant discretization length, and no lateral translational components at the single junction level. We supplement the crankshaft and modified-pivot moves used in these works with a pivot move that allows for rotations of the terminal triad around its tangent. The orientation of this tangent remains constant and aligned with the force director field throughout the simulation, such that tracing the lateral triad vectors [$\widehat{\mathbf{u}}$ and $\widehat{\mathbf{v}}$ from Eq. (1)] gives direct access to the global torsional angle θ. Note that this procedure requires strand-crossing moves to be rejected since they would allow for the relaxation of torsional strain by other means than the rotation of the terminal triad. Moreover, the magnetic bead and the flow cell surface are explicitly considered in the form of two impenetrable surfaces anchored at the termini to avoid linking number changes via crossings beyond these termini.

We considered three separate resolutions, the original base pair-step parametrization, and coarse-grained representations at the 5 and 10 bp level ($k = 5$ and $k = 10$). Coarse-graining to 5 and 10 bp resolutions resulted in approximately fourfold and sixfold speedups of the MC simulation, respectively. Additionally, the sampling efficiency of the lower-resolution simulations is further enhanced by enabling larger MC moves, which enables the exploration of the configurational phase space with fewer such moves.

5. Force extension curves for 7.9 kb MC simulations

Force extension curves for these three respective resolutions and forces ranging from 0.01 pN to 83 pN are shown in Fig. 6(b). We note that double-stranded B-DNA undergoes an overstretching transition at about 65 pN [117]. However, the considered harmonic model does not feature such a transition, allowing for the application of larger forces. We include forces

beyond this point to illustrate that at large forces fluctuations of the coarse-grained system increasingly diverge from those of the original system as discussed above [see the discussion around Eq. (75)].

Base pair-step resolution extensions are faithfully reproduced by the coarse-grained simulations to the point that extension curves are visually indistinguishable. Closer inspection reveals relative differences to remain below 1% [see inset of Fig. 6(b)]. Relative differences exhibit an increasing tendency for large forces past a certain resolution-dependent threshold. This observation is in line with the discussion about the force-mediated suppression of lateral fluctuations as wavelengths below the respective discretization lengths commence to be attenuated by the stretching force. As expected, the threshold force is the lowest for the 10 bp resolution simulations. Calculation of the apparent persistence length via fitting of Eq. (71) for forces larger than 0.3 pN yields $l_b = 68.4 \pm 0.2$ nm for the original resolution simulation. This value is in excellent agreement with the asymptotic persistence length calculated via the tangent-tangent correlation function. Fitting of the coarse-grained data yields slightly lower values: $l_b = 67.7 \pm 0.3$ nm for the 5 bp resolution and $l_b = 65.3 \pm 0.4$ nm for the 10 bp resolution. Remarkably, imposing an upper limit on the force range used for the fit recovers the base pair resolution persistence length. We find $l_b = 68.3 \pm 0.1$ pN at 10 bp resolution when only considering forces lower than 1 pN and $l_b = 68.4 \pm 0.2$ pN at 5 bp resolution when only considering forces lower than 30 pN. Again these findings are in line with the permeation of force-induced suppression of lateral fluctuations below the length scale of the coarse-graining resolution. The dashed line in Fig. 6(b) is a plot of the interpolation formula Eq. (72) using the value of the persistence length deduced from the high force fit (for the 1 bp resolution). Clearly, Eq. (72) is not a good representation of the observed extension behavior for low forces as it predicts convergence to zero extension, while the simulations converge to a finite value. The low force behavior of Eq. (72) is valid for ideal flexible polymers, i.e., for non-self-avoiding polymers, and in the absence of the repulsion planes. In particular, the presence of the latter breaks the symmetry of the distribution at zero force thereby necessitating a positive and distinctly nonzero mean extension.

6. Effective torsional stiffness for 7.9 kb MC simulations

Values for C_{eff} calculated from the same simulations are displayed in Fig. 6(c). Once again, the agreement between base pair-resolution and coarse-grained simulations is excellent with relative differences not exceeding 2.5% [except for a single point; see inset of Fig. 6(c)]. Fitting of the 1 bp data for forces larger than 0.3 pN to Eq. (74) yields $C = 139.8 \pm 0.3$ pN and $l_b = 67.2 \pm 0.8$ pN. This value of C, which may be viewed as the high-force asymptote of C_{eff}, is indicated by the horizontal dashed line in Fig. 6(c). Quite remarkably, it agrees well with the direct measurements via the asymptotic long length-scale torsional persistence length for the coarse-grained representations calculated in the previous section (especially the 42 and 21 bp resolutions) but much less so with the results of the base pair-resolution model. The FOMT simulations give direct access to the twist fluctuations,
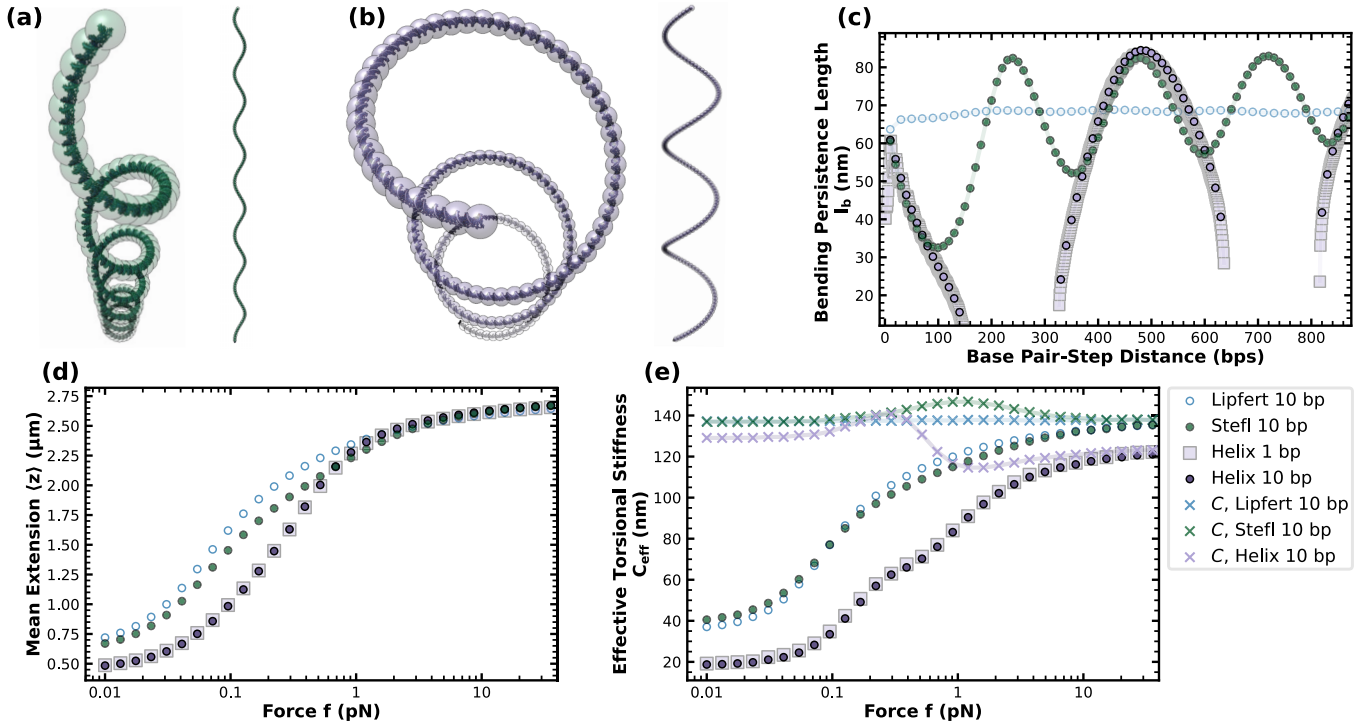
FIG. 7. Simulations with superhelically curved sequences. Panels (a) and (b) depict the top and side view of the ground state of the phased A-tract sequences CAAAATTTTG [118] and the phased sequence CGGGGGGCTTTTAGGGGGGCTTTTAGGGGGCTTT, respectively. (c) Comparison of the length-scale-dependent bending persistence lengths for the two superhelical sequences and the previously considered experimental sequence [22]. Force-extension and effective torsional stiffness data sampled with the FOMT setup for the same sequences are shown in panels (d) and (e), respectively.

which can then be related to the torsional stiffness via

$$\langle \Delta \mathrm{Tw}^2 \rangle = \frac{L}{4\pi^2 C}. \tag{76}$$

The resulting values of C [x-scatters in Fig. 6(c)] are independent of the force, as expected since there is no direct coupling between force and twist. Moreover, direct calculations of C stemming from the coarse-grained simulations closely align with the previously indicated asymptote (horizontal dashed line), while the base pair-resolution data underestimates the torsional stiffness. We conclude that the appropriate value for the torsional stiffness (for the cgNA+ provided RBP parametrization) is about 140 nm.

Fitting of the coarse-grained simulations to Eq. (74) yields almost the same values as the 1 bp resolution simulations. At 10 bp resolution we find $C = 138.4 \pm 0.4$ pN, and $l_b = 66.5 \pm 0.3$ pN, while the 5 bp resolution simulations yield $C = 138.2 \pm 0.3$ pN, and $l_b = 66.0 \pm 0.7$ pN. Reducing the force range used for the fit again slightly increases the obtained estimates for l_b.

7. Inclusion of translational degrees of freedom— 2 kb MC simulations

For the inclusion of the translational degrees of freedom in the simulations, we used a Python implementation of the MC package (available at Ref. [120]). However, its current limitations in computational efficiency made the generation of appreciable statistics for the 7.9 bp sequence unattainable. Therefore we limited the simulations to a 2 kb sequence taken

as the first 2 kb fragment of the 7.9 kb sequence. Force-extension curves [Fig. 6(d)] and values of C_{eff} [Fig. 6(f)] remain in reasonable agreement across 1 bp and 10 bp resolution simulations. Closer inspection reveals systematic discrepancies of the extension at large forces, where the distribution of z values is markedly shifted to the upside for the coarse-grained representation [see Fig. 6(e)]. This observation aligns with the prior observed pathology arising from the coarse-graining of the rise. The coarse-graining procedure systematically overestimates composite rise components by not accounting for the emergent asymmetry of composite-step rise distributions [see Figs 4(a) and 4(d)]. Note that the extension shift is approximately the same for all forces, but the narrowing of the distribution for large forces—see Refs. [27,28] for a more detailed discussion on the force-dependence of the extension variance—enhances the relative significance in the distribution. Distributions of the bead angle θ, however, are almost perfectly reproduced by the coarse-grained representations [Fig. 6(g)].

D. Superhelically curved DNA

As an example of a system where intrinsic curvature is manifestly relevant, we consider DNA sequences with an intrinsic helically wound ground state [see Figs. 7(a) and 7(b)]. Such superhelical sequences can be constructed by repeatedly concatenating small segments of strong intrinsic curvature. We consider two such sequences. The first consists of the phased A-tracts CAAAATTTTG and was considered

by Stefl *et al.* [118]. Its ground state is shown in Fig. 7(a). The second sequence, consisting of repeating segments of CGGGGGGCTTTTAGGGGGCTTTTAGGGGGCTTT was constructed in the present work by searching for sequences with a prominent helical diameter; see Fig. 7(b). For both superhelical sequences, we constructed 7.9 kb long molecules and conducted MC simulations in the unrestrained ensemble and the FOMT setup.

Figure 7(c) shows length-scale-dependent persistence lengths as given by Eq. (67) for the two superhelical sequences and the previously considered 7.9 kb sequence (taken from Lipfert *et al.* [22]). Unsurprisingly, the large-scale helical structure of the two superhelical sequences manifests in strong sinusoidal oscillations of the apparent persistence length. Missing points in the curve correspond to base pair-step distances m for which the tangent-tangent correlation function assumes negative values such that the logarithm in Eq. (67) is undefined.

In Fig. 7(d) the force-extension curves of the two superhelical sequences are compared with the largely featureless 7.9 kb sequence [22]. Both sequences exhibit markedly different elastic responses. Specifically, the artificially constructed sequence deviates from the response of the reference sequence for forces lower than approximately 1 pN. In this force range, the elastic behavior is influenced not only by entropy but also by the Hookean response of the superhelix, see also Ref. [119]. At higher forces, the superhelix is predominantly stretched out, leading the response to revert to that of the WLC model. The effective torsional stiffness C_{eff} is even more distinct for the strongly superhelical sequence, exhibiting a lower stiffness for all forces. This behavior persists to the largest considered forces.

IV. CONCLUSIONS

In this work, we presented a systematic coarse-graining scheme that enables the calculation of sequence-specific coarse-grained parameters. These parameters faithfully reproduce structural and elastic properties relevant to the chosen resolution. The original system is assumed to be described as a chain of rigid bodies, each captured by an intrinsic reference frame and characterized by a ground state and a stiffness matrix. Importantly, our procedure does not predict elastic parameters but transforms a given set of RBP parameters [1,3,58,64,65] to a lower resolution.

To achieve a k-fold reduction in resolution, the method retains every kth base pair associated reference frame while eliminating all other frames. The ground-state structure is then described in terms of the average relative orientation and position of these remaining reference frames. Fluctuations are captured through an effective stiffness matrix constructed to closely mimic the relative fluctuations of corresponding frames in the original system. Coarse-graining, therefore, entails mapping parameters from one Gaussian system into another. An extension of the coarse-graining procedure to incorporate higher-order effects, such as, for example, elastic multimodalities [121–123], kinks [33,35,124], and linear sub-elastic behavior [86,125], is possible in principle but requires more sophisticated approaches that go beyond the scope of the current work. Likewise, higher-order components of the

coarse-grained representation have been omitted, but could be included via a perturbative calculation. In particular, such an extension is expected to refine the description of composite rise fluctuations, where accurately capturing the emergent asymmetry requires introducing odd terms in the corresponding elastic Hamiltonian.

A Python module featuring implementations for all parameter transformations discussed in this work is available at Ref. [85]. These transformations include conversions between Cayley- and Euler-map definitions of rotations (see Appendix B), as well as between midstep-triad definitions of translations and the definition relative to the base pair triad employed in this work (see Appendix D).

Benchmark simulations of unrestrained molecules based on RBP parameters derived from the cgNA+ model [65] demonstrated that the coarse-grained parameters excellently capture the distributions of rotational degrees of freedom. The variances of coarse-grained rotational degrees of freedom were found to deviate from numerically obtained references by less than 2% for up to 40-fold reductions in resolution. However, translational fluctuations, particularly in rise, are less accurately reproduced. We attribute these discrepancies to the emergent asymmetry of composite-step translational fluctuations. Specifically, distributions of the total rise within a segment comprising multiple base pair steps exhibit marked left skewness, indicating that the end-to-end distance is more frequently contracted than extended. We argue this asymmetry to be a consequence of bending fluctuations, which, due to the entropic elasticity of a polymer [16], results in end-point contractions. Moreover, since DNA is found to be nearly inextensible for forces below about 10 pN [18], it is reasonable to assume rotational fluctuations to constitute the most relevant component of sequence-dependent elasticity under physiological conditions. We note that coarse-graining of both rotational and translations yields satisfactory results up to the 10 bp scale and argue that the 5-10 bp scale is a particularly relevant and natural choice as is aligns with the diameter of the double helix and has been employed by various studies in the past (see, e.g., Refs. [69,126,127]).

Sequence-specific and length-scale-dependent signatures in the bending persistence length were shown to be faithfully reproduced by the coarse-grained systems. Furthermore, we demonstrated that the coarse-graining approach makes it possible to simulate setups typically studied with single-molecule techniques, where large molecules containing thousands of base pairs are probed. We showed simulations of freely orbiting magnetic tweezers at 10 bp resolution to be virtually identical to corresponding simulations performed at full single-base pair resolution. This agreement is retained even if sequences with strong intrinsic curvature—in this case, a superhelically wound ground state—are considered. The coarse-graining approach, therefore, gives access to length scales far beyond what was previously attainable.

The proposed methodology is not limited to a particular model or parameter set and can, in principle, be applied to achieve coarse-grained representations for any base pair resolution (or higher) model. Switching to coarser representations may enable multi-resolution simulations, capturing the region of interest in full detail while simulating the rest of the system

at a lower resolution. Moreover, this approach could be used to expedite the equilibration of large systems.

We envision the access to faithful coarse-grained representations of DNA to enable the *in silico* study of many sequence-specific mesoscale phenomena. In particular, coarse-grained simulations could aid in elucidating the role of DNA sequences on the statistics and dynamics of DNA plectonemes. An implementation of the coarse-grained RBP model in a Molecular Dynamics framework might further enable the study of chromosomal organization guided by DNA protein interactions such as restriction factors [27], nucleosomes [14,128], and the localization of DNA loops [129], for example, guided by the action of structural maintenance of chromosomes (SMC) complexes [130,131].

APPENDIX A: RODRIGUES' ROTATION FORMULA

Matrix exponentials as in Eq. (3) are computationally expensive and in practice it is more convenient to use alternative formulations, such as Rodrigues' rotation formula

$$\mathcal{R} = \text{eul}(\boldsymbol{\Omega}) := \mathbb{1} + (\sin\Omega)\widehat{\Omega} + (1 - \cos\Omega)\widehat{\Omega}^2. \quad \text{(A1)}$$

The inverse transformation, which maps a rotation matrix to the corresponding rotation vector, is

$$\boldsymbol{\Omega} = \text{eul}^{-1}(\mathcal{R}) := \frac{\Omega \text{ vec}(\mathcal{R} - \mathcal{R}^{\mathsf{T}})}{2\sin\Omega}, \quad \text{(A2)}$$

where Ω is determined by the relation $\text{tr}\,\mathcal{R} = 1 + 2\cos\Omega$.

APPENDIX B: CONVERSION BETWEEN CAYLEY MAP AND EULER MAP DEFINITION OF ROTATIONS

Various studies [58,64,65,89,91] favor a definition of rotational components based on the Cayley map (also known as the Euler–Rodrigues or Gibbs formula),

$$\mathcal{R} = \text{cay}(\boldsymbol{\Theta}) := \mathbb{1} + \frac{4}{4 + \boldsymbol{\Theta}^2}\left(\widehat{\Theta} + \frac{1}{2}\widehat{\Theta}^2\right), \quad \text{(B1)}$$

where we denote the Cayley vectors by $\boldsymbol{\Theta}$ to distinguish them from the Euler-vectors $\boldsymbol{\Omega}$. The inverse relation takes the simple form

$$\boldsymbol{\Theta} = \text{cay}^{-1}(\mathcal{R}) = \frac{2\text{ vec}(\mathcal{R} - \mathcal{R}^{\mathsf{T}})}{1 + \text{tr}\,\mathcal{R}}. \quad \text{(B2)}$$

Computationally these definitions have an edge over the rotation vector definition chosen in the main text (in that its definition does not involve any infinite series expansions—or transcendental functions), but in most practical situations the difference is rather inconsequential.

The vector $\boldsymbol{\Theta}$ still indicates the axis of rotation, however, the rotation angle is no longer simply given by the vector's magnitude, but instead $|\boldsymbol{\Omega}| = 2\arctan(\Theta/2)$. This implies that the transformation of Cayley vectors into Euler vectors is given by

$$\boldsymbol{\Omega} = \mathbf{f}_{\text{eul}}(\boldsymbol{\Theta}) := 2\arctan\left(\frac{\Theta}{2}\right)\frac{\boldsymbol{\Theta}}{|\boldsymbol{\Theta}|}, \quad \text{(B3)}$$

and the corresponding inverse by

$$\boldsymbol{\Theta} = \mathbf{f}_{\text{eul}}^{-1}(\boldsymbol{\Omega}) := 2\tan\left(\frac{\Omega}{2}\right)\frac{\boldsymbol{\Omega}}{|\boldsymbol{\Omega}|}. \quad \text{(B4)}$$

Suppose that the geometry and Gaussian elasticity of a given molecule containing $N + 1$ base pairs is expressed in terms of a Cayley parametrization. The ground state will be specified by the system (Cayley) vector

$$\bar{\boldsymbol{\Theta}}_0^{\mathsf{T}} = \begin{pmatrix} \boldsymbol{\Theta}_{0,1}^{\mathsf{T}} & \cdots & \boldsymbol{\Theta}_{0,N}^{\mathsf{T}} \end{pmatrix}, \quad \text{(B5)}$$

where $\boldsymbol{\Theta}_{0,i}$ is the intrinsic rotation between base pairs i and $i + 1$ expressed in Cayley coordinates. Analogous to Eq. (22), the (lowest order expansion of the) elastic energy takes a (quadratic) form

$$\beta E = \frac{1}{2}\bar{\boldsymbol{\Theta}}_{\Delta}^{\mathsf{T}} M_{\Theta} \bar{\boldsymbol{\Theta}}_{\Delta}. \quad \text{(B6)}$$

Transformation of the static component from Cayley to Euler representation is straightforward,

$$\boldsymbol{\Omega}_0 = \mathbf{f}_{\text{eul}}(\boldsymbol{\Theta}_0). \quad \text{(B7)}$$

To identify how the stiffness matrix M_{Θ} transforms under the coordinate transformation, we expand $\mathbf{f}_{\text{eul}}$ to linear order around the ground state

$$\Omega_{\mu} = \mathbf{f}_{\text{eul},\mu}(\boldsymbol{\Theta}_0) + F_{\mu\nu}\Theta_{\Delta,\nu} + O(\Theta^2), \quad \text{(B8)}$$

where μ and ν indicate the dimensional subscripts of the single junction element, with summation over repeating indices implied. The positional subscript was omitted for ease of readability. Equations (B8) implies that up to linear order

$$\Omega_{\Delta,\mu} = F_{\mu\nu}\Theta_{\Delta,\nu}, \quad \text{(B9)}$$

with

$$F_{\mu\nu} = \frac{\partial \mathbf{f}_{\text{eul},\mu}(\boldsymbol{\Theta})}{\partial \Theta_{\nu}}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}_0}$$

$$= \left(\frac{4}{|\boldsymbol{\Theta}_0|^2 + 4} - \frac{|\boldsymbol{\Omega}_0|}{|\boldsymbol{\Theta}_0|}\right)\frac{\Theta_{0,\mu}\Theta_{0,\nu}}{|\boldsymbol{\Theta}_0|^2} + \frac{|\boldsymbol{\Omega}_0|}{|\boldsymbol{\Theta}_0|}\delta_{\mu\nu}, \quad \text{(B10)}$$

where $\delta_{\mu\nu}$ is the Kronecker δ and $|\boldsymbol{\Omega}_0|$ is given by Eq. (B3). Defining the block-diagonal matrix

$$\bar{F} = \begin{pmatrix} F^{(1)} & 0 & \cdots & 0 \\ 0 & F^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F^{(N)} \end{pmatrix}, \quad \text{(B11)}$$

which captures the transformation of the entire system, the transformation from system-wide Cayley to system-wide Euler vector may be written as

$$\bar{\boldsymbol{\Omega}}_{\Delta} = \bar{F}\,\bar{\boldsymbol{\Theta}}_{\Delta}. \quad \text{(B12)}$$

The elastic energy, Eq. (B6), may then be rewritten as

$$\beta E = \frac{1}{2}\bar{\boldsymbol{\Theta}}_{\Delta}^{\mathsf{T}} M_{\Theta}\bar{\boldsymbol{\Theta}}_{\Delta} = \frac{1}{2}\bar{\boldsymbol{\Omega}}_{\Delta}^{\mathsf{T}}[(\bar{F}^{-1})^{\mathsf{T}} M_{\Theta}\bar{F}^{-1}]\bar{\boldsymbol{\Omega}}_{\Delta}, \quad \text{(B13)}$$
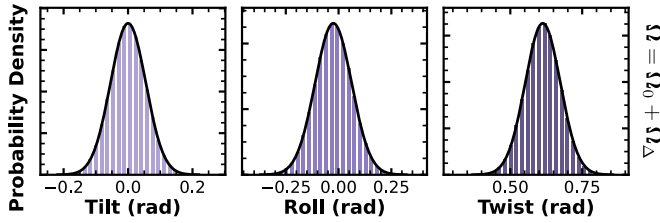
FIG. 8. Comparison between sampled and analytical distributions of Euler-vector component for a single base pair-step. The stiffness matrix in Cayley-representation was generated with cgNA+ [65] and corresponds to a single GpC step, which was generated in the context of a larger sequence (GACTGC**G**CGCCTCA) to avoid boundary effects. Tilt, roll, and twist were sampled in the original (Cayley) coordinates and then transformed to Euler coordinates via Eq. (B3).

where the inverse of $\bar{F}$ is also block-diagonal with entries explicitly given by

$$F_{\mu\nu}^{-1} = \left(\sec^2\left(\frac{|\boldsymbol{\Omega}_0|}{2} \right) - \frac{|\boldsymbol{\Theta}_0|}{|\boldsymbol{\Omega}_0|} \right) \frac{\Omega_{0,\mu}\Omega_{0,\nu}}{|\boldsymbol{\Omega}_0|^2} + \frac{|\boldsymbol{\Theta}_0|}{|\boldsymbol{\Omega}_0|}\delta_{\mu\nu}. \tag{B14}$$

Finally, we can identify the sought transformation of the stiffness matrix M_Θ,

$$M_\Omega = (\bar{F}^{-1})^\mathsf{T} M_\Theta \bar{F}^{-1}. \tag{B15}$$

In the present discussion, we assumed the energy to be composed of solely rotational components, but since the transformation from Cayley to Euler parameters leaves translations unaltered, the extension of the transformation of the stiffness matrix to include translations is straightforward (note however, that the entries of the stiffness matrix quantifying translational stiffness do not generally remain unaltered).

Last, we note that the Jacobian associated with the transformation, $\mathbf{f}_{\text{eul}}$,

$$\det G = \frac{4|\boldsymbol{\Omega}_0|^2}{(|\boldsymbol{\Theta}_0|^2 + 4)|\boldsymbol{\Theta}_0|^2}, \tag{B16}$$

takes values markedly different from 1 for typical values of $\boldsymbol{\Theta}_0$ (for double-stranded DNA) but varies by less than 5% over typical ranges of $\boldsymbol{\Theta}$, justifying it to be ignored in most practical cases [90].

We validate the quality of the analytical transformation of the stiffness matrix by comparing numerically sampled distributions (generated according to the canonical measure), with distributions obtained by marginalizing the analytically transformed stiffness matrix. Values are sampled in the Cayley representation and then transformed to Euler-vectors via Eq. (B3). Comparative histograms for a single base pair-step revealing a relative difference in the variance of less than 1% are shown in Fig. 8.

APPENDIX C: SPLITTING STATIC AND DYNAMIC COMPONENTS AT THE TRANSFORMATION LEVEL

As introduced in the main text, it is customary to express the transformation between consecutive base pair-related frames as a single six-vector, containing three rotational and

three translational components. In the compact SE(3) notation such a transformation takes the form

$$g = \begin{pmatrix} \mathcal{R} & \mathbf{w} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} = \begin{pmatrix} e^{\widehat{\Omega}_0 + \widehat{\Omega}_\Delta} & \mathbf{s} + \mathbf{w}_\Delta \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}. \tag{C1}$$

However, the coarse-graining procedure discussed in the main text requires static and dynamic components to be split at the transformation level, i.e.,

$$g = s\, d, \tag{C2}$$

with

$$s = \begin{pmatrix} \mathcal{S} & \mathbf{s} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} = \begin{pmatrix} e^{\widehat{\Phi}_0} & \mathbf{s} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \tag{C3}$$

and

$$d = \begin{pmatrix} \mathcal{D} & \mathbf{d} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} = \begin{pmatrix} e^{\widehat{\Phi}_\Delta} & \mathbf{d} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}. \tag{C4}$$

In this section, we will show how to achieve this decomposition and how the stiffness matrix transforms under such a redefinition of fluctuating variables.

For consistency, the ground state should be identical across both pictures, which leads to $\boldsymbol{\Phi}_0 = \boldsymbol{\Omega}_0$. Therefore, we can treat s as a known quantity, such that d may be expressed as

$$d = s^{-1}g = \begin{pmatrix} \mathcal{S}^\mathsf{T}\mathcal{R} & \mathcal{S}^\mathsf{T}(\mathbf{w} - \mathbf{s}) \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} = \begin{pmatrix} \mathcal{S}^\mathsf{T}\mathcal{R} & \mathcal{S}^\mathsf{T}\mathbf{w}_\Delta \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}. \tag{C5}$$

From Eq. (C5) we can deduce that

$$\mathbf{d} = \mathcal{S}^\mathsf{T}\mathbf{w}_\Delta, \tag{C6}$$

and

$$e^{\widehat{\Phi}_\Delta} = e^{-\widehat{\Omega}_0}e^{\widehat{\Omega}_0 + \widehat{\Omega}_\Delta}. \tag{C7}$$

The exact correspondence between $\widehat{\Omega}_0$ and $\widehat{\Phi}_\Delta$ is encapsulated in the Baker-Campbell-Hausdorff formula, which for three elements a, b, and c of a Lie algebra provides the solution of the equation $e^{a+b} = e^c$ in terms of the infinite series of commutators

$$c = a + b + \frac{1}{2}[a, b] + \frac{1}{12}([a, [a, b]] + [b, [b, a]]) + \text{h.o.} \tag{C8}$$

where h.o. is a placeholder for the infinite series of higher-order commutators. We seek the transformation from $\boldsymbol{\Omega}_\Delta$ to $\boldsymbol{\Phi}_\Delta$ up to linear order. Therefore, it will suffice to consider only terms linear in the fluctuating quantities. The first few terms in the Baker-Campbell-Haussdorf formula satisfying this requirement are

$$\begin{aligned} \widehat{\Omega}_\Delta &= \log\left(e^{\widehat{\Omega}_0}e^{\widehat{\Phi}_\Delta} \right) - \widehat{\Omega}_0 \\ &= \widehat{\Phi}_\Delta + \frac{1}{2}[\widehat{\Omega}_0, \widehat{\Phi}_\Delta] + \frac{1}{12}[\widehat{\Omega}_0, [\widehat{\Omega}_0, \widehat{\Phi}_\Delta]] \\ &\quad - \frac{1}{720}[\widehat{\Omega}_0, [\widehat{\Omega}_0, [\widehat{\Omega}_0, [\widehat{\Omega}_0, \widehat{\Phi}_\Delta]]]] \\ &\quad + \frac{1}{30240}[\widehat{\Omega}_0, [\widehat{\Omega}_0, [\widehat{\Omega}_0, [\widehat{\Omega}_0, [\widehat{\Omega}_0, [\widehat{\Omega}_0, \widehat{\Phi}_\Delta]]]]]] + \text{h.o.} \end{aligned} \tag{C9}$$

Note that

$$\mathrm{vec}\big([\widehat{\Omega}_0, \widehat{\Phi}_\Delta]\big) = \boldsymbol{\Omega}_0 \times \boldsymbol{\Phi}_\Delta = \widehat{\Omega}_0 \boldsymbol{\Phi}_\Delta, \qquad \text{(C10)}$$

nested commutators containing k static elements $\widehat{\Omega}_0$ and only a single $\widehat{\Phi}_\Delta$, which is arranged to be the right-hand element in the innermost commutator, take a simple form when expressed as vectors instead of antisymmetric tensors

$$\mathrm{vec}([\widehat{\Omega}_0, [\widehat{\Omega}_0, [\dots, [\widehat{\Omega}_0, \widehat{\Phi}_\Delta]\dots]]]) = \widehat{\Omega}_0^k \boldsymbol{\Phi}_\Delta. \qquad \text{(C11)}$$

After substitution of Eq. (C11) into Eq. (C9), one has up to linear order

$$\boldsymbol{\Omega}_\Delta = H^{-1}(\boldsymbol{\Omega}_0)\, \boldsymbol{\Phi}_\Delta, \qquad \text{(C12)}$$

with

$$\begin{aligned} H^{-1}(\boldsymbol{\Omega}_0) &= \mathbb{1} + \tfrac{1}{2}\widehat{\Omega}_0 + \tfrac{1}{12}\widehat{\Omega}_0^2 - \tfrac{1}{720}\widehat{\Omega}_0^4 \\ &\quad + \tfrac{1}{30240}\widehat{\Omega}_0^6 + \text{h.o.} \end{aligned} \qquad \text{(C13)}$$

We note that the Jacobian of this transformation is 1 since $\det \widehat{\Omega}_0 = 0$.

Defining the six-component basepair-step deformation vectors that include both the rotational and the translational components

$$\mathbf{X}_\Delta^\mathsf{T} \equiv \big(\boldsymbol{\Omega}_\Delta^\mathsf{T} \quad \mathbf{w}_\Delta^\mathsf{T}\big) \quad \mathbf{Y}_\Delta^\mathsf{T} \equiv \big(\boldsymbol{\Phi}_\Delta^\mathsf{T} \quad \mathbf{d}^\mathsf{T}\big), \qquad \text{(C14)}$$

one sees that the transformation from $\mathbf{X}_\Delta$ to $\mathbf{Y}_\Delta$ may be linearized as

$$\mathbf{Y}_\Delta = H(\mathbf{X}_0)\mathbf{X}_\Delta, \qquad \text{(C15)}$$

where the just derived linearized transformations of rotation [Eq. (C13)] and translation [Eq. (C6)] are the respective components of the matrix

$$H(\mathbf{X}_0) = \begin{pmatrix} H(\boldsymbol{\Omega}_0) & 0 \\ 0 & \mathcal{S}^\mathsf{T} \end{pmatrix}. \qquad \text{(C16)}$$

Defining

$$\bar{H}(\bar{\mathbf{X}}_0) = \begin{pmatrix} H(\mathbf{X}_{0,0}) & 0 & \dots & 0 \\ 0 & H(\mathbf{X}_{0,1}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & H(\mathbf{X}_{0,N-1}) \end{pmatrix},$$
$$\text{(C17)}$$

and following the same logic that led to Eq. (B15), one finds

$$M_Y = [\bar{H}^{-1}(\bar{\mathbf{X}}_0)]^\mathsf{T} M_X [\bar{H}^{-1}(\bar{\mathbf{X}}_0)]. \qquad \text{(C18)}$$

APPENDIX D: CONVERSION BETWEEN MIDSTEP-TRIAD AND TRIAD DEFINITION OF TRANSLATIONS

To achieve a parametrization of DNA configurations that are independent of the strand direction—up to a sign-flip—(i.e., assignment of Watson and Crick strand), translational base pair-step components are frequently expressed in terms of midstep coordinate frames that are found via half-way rotation between the two respective reference frames (rotational components are identical when defined relative to either of the two respective triads or the midstep frame). However, for this work, it proved convenient to express translations in the frame of the first triad of each respective pair of frames. In this section, we show how these definitions can be transformed

into one another and how the associated transformation of the stiffness matrix can be accommodated.

The midstep triad is usually defined as [62]

$$\mathcal{T}_{\mathrm{mid},i} = \mathcal{T}_i \sqrt{\mathcal{R}_i}, \qquad \text{(D1)}$$

where $\sqrt{\mathcal{R}_i}$ is the matrix of half rotation between the frames $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$. Specifically, if

$$\mathcal{R}_i = \mathcal{T}_i^\mathsf{T} \mathcal{T}_{i+1} = \exp \widehat{\Omega}_i, \qquad \text{(D2)}$$

then

$$\sqrt{\mathcal{R}_i} \equiv \exp \frac{1}{2}\widehat{\Omega}_i. \qquad \text{(D3)}$$

With respect to the midstep triad $\mathcal{T}_{\mathrm{mid},i}$ the translation between base pairs i and $i+1$, is given by [90]

$$\boldsymbol{\zeta}_i = \mathcal{T}_{\mathrm{mid},i}^\mathsf{T}(\mathbf{r}_{i+1} - \mathbf{r}_i). \qquad \text{(D4)}$$

Meanwhile, the triad frame definition as introduced in the main text is

$$\mathbf{w}_i = \mathcal{T}_i^\mathsf{T}(\mathbf{r}_{i+1} - \mathbf{r}_i). \qquad \text{(D5)}$$

Inspection of Eqs. (D4) and (D5) reveals the relation between the two definitions

$$\mathbf{w}_i = \sqrt{\mathcal{R}_i}\boldsymbol{\zeta}_i. \qquad \text{(D6)}$$

To ascertain the transformation of the stiffness matrix, we will once again follow the procedure of first splitting the transformation into static and dynamic components, where the dynamic component is assumed to be small, followed by the linearization of the dynamic part of the transformation.

Following previous treatment the rotation $\sqrt{\mathcal{R}_i}$ may be expanded to first order in the fluctuating rotational component $\boldsymbol{\Omega}_{\Delta,i}$,

$$\begin{aligned} \sqrt{\mathcal{R}_i} &= \exp\left(\frac{1}{2}\widehat{\Omega}_{0,i} + \frac{1}{2}\widehat{\Omega}_{\Delta,i}\right) \\ &\approx \sqrt{\mathcal{S}_i} \exp\left(\frac{1}{2}\mathrm{hat}\left[H\left(\frac{\boldsymbol{\Omega}_{0,i}}{2}\right)\boldsymbol{\Omega}_{\Delta,i}\right]\right) \\ &\approx \sqrt{\mathcal{S}_i}\left(\mathbb{1} + \frac{1}{2}\mathrm{hat}\left[H\left(\frac{\boldsymbol{\Omega}_{0,i}}{2}\right)\boldsymbol{\Omega}_{\Delta,i}\right]\right), \qquad \text{(D7)} \end{aligned}$$

where we used Eq. (C12) to split the exponential. Separating $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0 + \boldsymbol{\zeta}_\Delta$ into static and dynamic components Eq. (D6) takes the form

$$\mathbf{w}_i \approx \sqrt{\mathcal{S}_i}\left(\boldsymbol{\zeta}_{0,i} + \boldsymbol{\zeta}_{\Delta,i} + \frac{1}{2}\mathrm{hat}\left[H\left(\frac{\boldsymbol{\Omega}_{0,i}}{2}\right)\boldsymbol{\Omega}_{\Delta,i}\right]\boldsymbol{\zeta}_{0,i}\right). \quad \text{(D8)}$$

Following previous procedure the quadratic term coupling $\boldsymbol{\zeta}_{\Delta,i}$ and $\boldsymbol{\Omega}_{\Delta,i}$ was ignored. Finally, using

$$\widehat{a}\mathbf{b} = \mathrm{vec}([\widehat{a}, \widehat{b}]) = \mathrm{vec}([\widehat{b^\mathsf{T}}, \widehat{a}]) = \widehat{b}^\mathsf{T}\mathbf{a}, \qquad \text{(D9)}$$

one arrives at

$$\mathbf{s}_i = \sqrt{\mathcal{S}_i}\boldsymbol{\zeta}_{0,i}, \qquad \text{(D10)}$$

$$\mathbf{w}_{\Delta,i} = \sqrt{\mathcal{S}_i}\boldsymbol{\zeta}_{\Delta,i} + \frac{1}{2}\sqrt{\mathcal{S}_i}\widehat{\zeta}_{0,i}^\mathsf{T} H\left(\frac{\boldsymbol{\Omega}_{0,i}}{2}\right)\boldsymbol{\Omega}_{\Delta,i}. \qquad \text{(D11)}$$

The fluctuating component $\mathbf{w}_{\Delta,i}$ does not solely depend on the $\boldsymbol{\zeta}_{\Delta,i}$, but rotational fluctuations $\boldsymbol{\Omega}_{\Delta,i}$ lead to additional
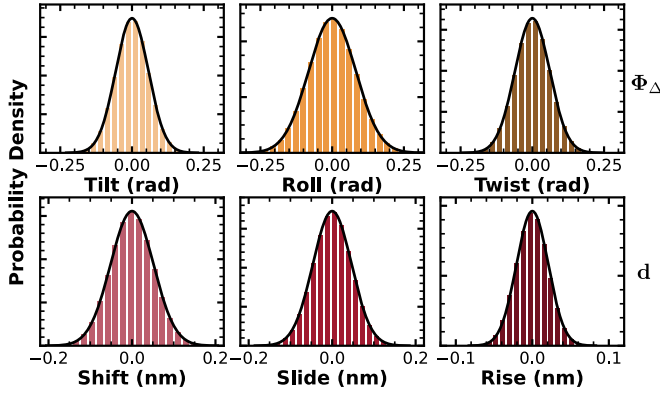
FIG. 9. Comparison between sampled and analytical distributions of the dynamic components of tilt, roll, and twist (in the $\boldsymbol{\Phi}_\Delta$ parametrization; expressed in radians) and shift, slide, and rise (in the $\mathbf{d}$ parametrization; expressed in nm) for a single base pair-step. The same sequence and procedure were considered as in Fig. 8.

translational fluctuation. In matrix form, the transformation takes the form

$$
\begin{pmatrix} \boldsymbol{\Omega}_{\Delta,i} \\ \mathbf{w}_{\Delta,i} \end{pmatrix} = \begin{pmatrix} \mathbb{1} & 0 \\ \frac{1}{2}\sqrt{\mathcal{S}_i}\widehat{\zeta_{0,i}^\mathsf{T}}H\left(\frac{\boldsymbol{\Omega}_{0,i}}{2}\right) & \sqrt{\mathcal{S}_i} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}_{\Delta,i} \\ \zeta_{\Delta,i} \end{pmatrix}. \quad \text{(D12)}
$$

The transformation of the stiffness matrix is then completely analogous to Eq. (C18).

We jointly assess the quality of the transformations of Appendixes C and D in the same way as was done for the transformation of the stiffness matrix from the Cayley map definition of rotations to the Euler map definition (see Appendix B). Rotations and translations were sampled according to the $\boldsymbol{\Omega}_\Delta$ and $\boldsymbol{\zeta}_\Delta$ parametrizations, respectively, and then transformed to $\boldsymbol{\Phi}_\Delta$ and $\mathbf{d}$ via Eqs. (D6), (C6), and (C7). Comparisons of the resulting histograms with the analytical predictions obtained via Eqs. (D12) and (C18) followed by the marginalization of the resulting stiffness matrix to the effective stiffness of the respective degrees of freedom are shown in Fig. 9. Deviations between sampled and analytical variances were not found to exceed 2% for any considered sequence out of a total sample of 500 sequences.

APPENDIX E: COMPOSITE TRANSLATIONS

This section serves to provide the details of the calculation of the composite translation [Eq. (51)]. Working out the translational component of

$$
d_{(i,j)} = s_{[i,j]}^{-1} g_{[i,j]}, \quad \text{(E1)}
$$

in terms of single junctions rotations and translations one finds

$$
\mathbf{d}_{(i,j)} = \mathcal{S}_{[i,j]}^\mathsf{T}\mathbf{w}_{[i,j]} - \mathcal{S}_{[i,j]}^\mathsf{T}\sum_{l=i}^{j}\mathcal{S}_{[i,l-1]}\mathbf{s}_l
$$

$$
= \mathcal{S}_{[i,j]}^\mathsf{T}\sum_{l=i}^{j}[\mathcal{R}_{[i,l-1]}(\mathcal{S}_l\mathbf{d}_l + \mathbf{s}_l)] - \sum_{l=i}^{j}\mathcal{S}_{[l,j]}^\mathsf{T}\mathbf{s}_l. \quad \text{(E2)}
$$

Applying the approximation used for the calculation of the rotational component (see Sec. II C 2) then yields

$$
\mathcal{S}_{[i,j]}^\mathsf{T}\mathcal{R}_{[i,l-1]} = \mathcal{S}_{[l,j]}^\mathsf{T}(\mathcal{S}_{[i,l-1]}^\mathsf{T}\mathcal{R}_{[i,l-1]}) = \mathcal{S}_{[l,j]}^\mathsf{T}\mathcal{D}_{(i,l-1)}
$$

$$
\approx \mathcal{S}_{[l,j]}^\mathsf{T}\exp\left(\text{hat}\left(\sum_{k=i}^{l-1}\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k}\right)\right)
$$

$$
\approx \mathcal{S}_{[l,j]}^\mathsf{T}\left(\mathbb{1} + \text{hat}\left(\sum_{k=i}^{l-1}\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k}\right)\right), \quad \text{(E3)}
$$

where the matrix exponential was expanded to first order in the last step. Substituting Eq. (E3) into Eq. (E2) one further finds

$$
\mathbf{d}_{(i,j)} \approx \sum_{l=i}^{j}\left[\mathcal{S}_{[l,j]}^\mathsf{T}\left(\mathbb{1} + \text{hat}\left(\sum_{k=i}^{l-1}\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k}\right)\right)\right.
$$

$$
\left.\times(\mathcal{S}_l\mathbf{d}_l + \mathbf{s}_l)\right] - \sum_{l=i}^{j}\mathcal{S}_{[l,j]}^\mathsf{T}\mathbf{s}_l
$$

$$
\approx \sum_{l=i}^{j}\mathcal{S}_{[l+1,j]}^\mathsf{T}\mathbf{d}_l + \sum_{l=i}^{j}\mathcal{S}_{[l,j]}^\mathsf{T}\text{hat}\left(\sum_{k=i}^{l-1}\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k}\right)\mathbf{s}_l
$$

$$
= \sum_{l=i}^{j}\mathcal{S}_{[l+1,j]}^\mathsf{T}\mathbf{d}_l + \sum_{l=i}^{j}\sum_{k=i}^{l-1}\mathcal{S}_{[l,j]}^\mathsf{T}\text{hat}(\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k})\mathbf{s}_l. \quad \text{(E4)}
$$

From the first to the second line, the terms coupling $\boldsymbol{\Phi}_{\Delta k}$ to $\mathbf{d}_l$ were discarded since they are of quadratic order in the fluctuations. Moreover, the constant term, $\sum_{l=i}^{j}\mathcal{S}_{[l,j]}^\mathsf{T}\mathbf{s}_l$, exactly cancels out.

Equation (E4) shows the dynamic composite translation, $\mathbf{d}_{(i,j)}$, to depend on both the dynamics junction translation, $\mathbf{d}_l$, and the dynamic junction rotation, $\boldsymbol{\Phi}_{\Delta,l}$. However, to identify the proper junction rotation transformation one has to rewrite the second term in Eq. (E4) as

$$
\mathcal{S}_{[l,j]}^\mathsf{T}\text{hat}(\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k})\mathbf{s}_l
$$

$$
= \mathcal{S}_{[l,j]}^\mathsf{T}\hat{s}_l^\mathsf{T}\mathcal{S}_{[k+1,l-1]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k} = \mathcal{S}_{[l,j]}^\mathsf{T}\hat{s}_l^\mathsf{T}\mathcal{S}_{[l,j]}\mathcal{S}_{[k+1,j]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k}
$$

$$
= \text{hat}(\mathcal{S}_{[l,j]}^\mathsf{T}\mathbf{s}_l)\mathcal{S}_{[k+1,j]}^\mathsf{T}\boldsymbol{\Phi}_{\Delta,k}, \quad \text{(E5)}
$$

where Eq. (D9) was used in the first step to shift $\boldsymbol{\Phi}_{\Delta,k}$ to the left-hand side of the expression. After combining Eqs. (E4) and (E5) and rearranging the order of the summation one finally arrives at

$$
\mathbf{d}_{(i,j)} = \sum_{l=i}^{j}\mathcal{S}_{[l+1,j]}^\mathsf{T}\mathbf{d}_l
$$

$$
+ \sum_{k=i}^{j-1}\left[\sum_{l=k+1}^{j}\text{hat}(\mathcal{S}_{[l,j]}^\mathsf{T}\mathbf{s}_l)\mathcal{S}_{[k+1,j]}^\mathsf{T}\right]\boldsymbol{\Phi}_{\Delta,k}. \quad \text{(E6)}
$$

[1] 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).

[2] F. Lankaš, J. Šponer, P. Hobza, and J. Langowski, Sequence-dependent elastic properties of DNA, J. Mol. Biol. **299**, 695 (2000).

[3] 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).

[4] P. J. Hagerman, Sequence-directed curvature of DNA, Nature (London) **321**, 449 (1986).

[5] H. C. M. Nelson, J. T. Finch, B. F. Luisi, and A. Klug, The structure of an oligo(dA)·oligo(dT) tract and its biological implications, Nature (London) **330**, 221 (1987).

[6] E. N. Trifonov, R. K. Tan, and S. C. Harvey, DNA bending and curvature, in *Structure and Expression*, edited by W. K. Olson, R. H. Sarma, M. H. Sarma, and M. Sundaralingam (Adenine Press, New York 1988), Vol. 3, pp. 243–253.

[7] H.-S. Koo, H.-M. Wu, and D. M. Crothers, DNA bending at adenine·thymine tracts, Nature (London) **320**, 501 (1986).

[8] C. R. Calladine, H. R. Drew, and M. J. McCall, The intrinsic curvature of DNA in solution, J. Mol. Biol. **201**, 127 (1988).

[9] J. A. Schellman and S. C. Harvey, Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature, Biophys. Chem. **55**, 95 (1995).

[10] J. P. Peters and L. J. Maher, DNA curvature and flexibility *in vitro* and *in vivo*, Q. Rev. Biophys. **43**, 23 (2010).

[11] 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).

[12] T. Dršata, N. Špačková, P. Jurečka, M. Zgarbová, J. Šponer, and F. Lankaš, Mechanical properties of symmetric and asymmetric DNA A-tracts: Implications for looping and nucleosome positioning, Nucleic Acids Res. **42**, 7383 (2014).

[13] 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).

[14] J. Neipel, G. Brandani, and H. Schiessel, Translational nucleosome positioning: A computational study, Phys. Rev. E **101**, 022405 (2020).

[15] O. Kratky and G. Porod, Röntgenuntersuchung gelöster Fadenmoleküle, Recl. Trav. Chim. Pays-Bas **68**, 1106 (1949).

[16] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics*, International series of monographs on physics (Clarendon Press, Oxford, UK, 1986).

[17] P. J. Hagerman, Flexibility of DNA, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).

[18] C. Bustamante, J. F. Marko, E. D. Siggia, and B. Smith, Entropic elasticity of λ-phage DNA, Science **265**, 1599 (1994).

[19] J. F. Marko and E. D. Siggia, Stretching DNA, Macromolecules **28**, 8759 (1995).

[20] J. F. Marko and E. D. Siggia, Bending and twisting elasticity of DNA, Macromolecules **27**, 981 (1994).

[21] J. D. Moroz and P. Nelson, Entropic elasticity of twist-storing polymers, Macromolecules **31**, 6333 (1998).

[22] J. Lipfert, J. W. J. Kerssemakers, T. Jager, and N. H. Dekker, Magnetic torque tweezers: Measuring torsional stiffness in DNA and RecA-DNA filaments, Nat. Methods **7**, 977 (2010).

[23] T. Brouns, H. De Keersmaecker, S. F. Konrad, N. Kodera, T. Ando, J. Lipfert, S. De Feyter, and W. Vanderlinden, Free energy landscape and dynamics of supercoiled DNA by high-speed atomic force microscopy, ACS Nano **12**, 11907 (2018).

[24] S. K. Nomidis, E. Skoruppa, E. Carlon, and J. F. Marko, Twist-bend coupling and the statistical mechanics of the twistable wormlike-chain model of DNA: perturbation theory and beyond, Phys. Rev. E **99**, 032414 (2019).

[25] K. Ott, L. Martini, J. Lipfert, and U. Gerland, Dynamics of the buckling transition in double-stranded DNA and RNA, Biophys. J. **118**, 1690 (2020).

[26] X. Gao, Y. Hong, F. Ye, J. T. Inman, and M. D. Wang, Torsional stiffness of extended and plectonemic DNA, Phys. Rev. Lett. **127**, 028101 (2021).

[27] W. Vanderlinden, E. Skoruppa, P. J. Kolbeck, E. Carlon, and J. Lipfert, DNA fluctuations reveal the size and dynamics of topological domains, PNAS Nexus **1**, pgac268 (2022).

[28] E. Skoruppa and E. Carlon, Equilibrium fluctuations of DNA plectonemes, Phys. Rev. E **106**, 024412 (2022).

[29] S. H. Kim, M. Ganji, E. Kim, J. van der Torre, E. Abbondanzieri, and C. Dekker, DNA sequence encodes the position of DNA supercoils, eLife **7**, e36557 (2018).

[30] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpí, C. González, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case, and M. Orozco, Parmbsc1: A refined force field for DNA simulations, Nat. Methods **13**, 55 (2016).

[31] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, and P. Jurečka, Refinement of the sugar–phosphate backbone torsion beta for AMBER Force fields improves the description of Z- and B-DNA, J. Chem. Theory Comput. **11**, 5723 (2015).

[32] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, and T. E. Cheatham, Assessing the current state of amber force field modifications for DNA, J. Chem. Theory Comput. **12**, 4114 (2016).

[33] F. Lankaš, R. Lavery, and J. H. Maddocks, Kinking occurs during molecular dynamics simulations of small DNA minicircles, Structure **14**, 1527 (2006).

[34] C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann, and R. Seidel, DNA-DNA interactions in tight supercoils are described by a small effective charge density, Phys. Rev. Lett. **105**, 158101 (2010).

[35] J. S. Mitchell, C. A. Laughton, and S. A. Harris, Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA, Nucleic Acids Res. **39**, 3928 (2011).

[36] K. Liebl, T. Dršata, F. Lankaš, J. Lipfert, and M. Zacharias, Explaining the striking difference in twist-stretch coupling between DNA and RNA: A comparative molecular dynamics analysis, Nucleic Acids Res. **43**, 10143 (2015).

[37] J. Vreede, A. Pérez de Alba Ortíz, P. G. Bolhuis, and D. W. H. Swenson, Atomistic insight into the kinetic pathways for Watson–Crick to Hoogsteen transitions in DNA, Nucleic Acids Res. **47**, 11069 (2019).

[38] M. Kim, S. Bae, I. Oh, J. Yoo, and J. S. Kim, Sequence-dependent twist-bend coupling in DNA minicircles, Nanoscale **13**, 20186 (2021).

[39] H. Dohnalová and F. Lankaš, Deciphering the mechanical properties of B–DNA duplex, Wiley Interdiscip. Rev. Comput. Mol. Sci. **12**, e1575 (2022).

[40] J. Yoo, S. Park, C. Maffeo, T. Ha, and A. Aksimentiev, DNA sequence and methylation prescribe the inside-out conformational dynamics and bending energetics of DNA minicircles, Nucleic Acids Res. **49**, 11459 (2021).

[41] S. Saurabh, M. A. Glaser, Y. Lansac, and P. K. Maiti, Atomistic simulation of stacked nucleosome core Particles: Tail bridging, the H4 Tail, and effect of hydrophobic forces, J. Phys. Chem. B **120**, 3048 (2016).

[42] Q. Liao, M. Lüking, D. M. Krüger, S. Deindl, J. Elf, P. M. Kasson, and S. C. Lynn Kamerlin, Long time-scale atomistic simulations of the structure and dynamics of transcription factor-DNA recognition, J. Phys. Chem. B **123**, 3576 (2019).

[43] J. Yoo, D. Winogradoff, and A. Aksimentiev, Molecular dynamics simulations of DNA–DNA and DNA–protein interactions, Curr. Opin. Struct. Biol. **64**, 88 (2020).

[44] J. Huertas and V. Cojocaru, Breaths, twists, and turns of atomistic nucleosomes, J. Mol. Biol. **433**, 166744 (2021).

[45] K. Coshic, C. Maffeo, D. Winogradoff, and A. Aksimentiev, The structure and physical properties of a packaged bacteriophage particle, Nature (London) **627**, 905 (2024).

[46] A. L. B. Pyne, A. Noy, K. H. S. Main, V. Velasco-Berrelleza, M. M. Piperakis, L. A. Mitchenall, F. M. Cugliandolo, J. G. Beton, C. E. M. Stevenson, B. W. Hoogenboom, A. D. Bates, A. Maxwell, and S. A. Harris, Base-pair resolution analysis of the effect of supercoiling on DNA flexibility and major groove recognition by triplex-forming oligonucleotides, Nat. Commun. **12**, 1053 (2021).

[47] T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo, A coarse grain model for DNA, J. Chem. Phys. **126**, 084901 (2007).

[48] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, DNA nanotweezers studied with a coarse-grained model of DNA, Phys. Rev. Lett. **104**, 178101 (2010).

[49] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, Sequence-dependent thermodynamics of a coarse-grained DNA model, J. Chem. Phys. **137**, 135101 (2012).

[50] C. A. Brackley, A. N. Morozov, and D Marenduzzo, Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS, J. Chem. Phys. **140**, 135103 (2014).

[51] N. Korolev, D. Luo, A. Lyubartsev, and L. Nordenskiöld, A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo, Polymers **6**, 1655 (2014).

[52] B. E. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, and A. A. Louis, Introducing improved structural properties and salt dependence into a coarse-grained model of DNA, J. Chem. Phys. **142**, 234901 (2015).

[53] Y. A. G. Fosado, D. Michieletto, J. Allan, C. A. Brackley, O. Henrich, and D. Marenduzzo, A single nucleotide resolution model for large-scale simulations of double stranded DNA, Soft Matter **12**, 9458 (2016).

[54] P. D. Dans, A. Zeida, M. R. Machado, and S. Pantano, A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics, J. Chem. Theory Comput. **6**, 1711 (2010).

[55] 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).

[56] G. S. Freeman, D. M. Hinckley, J. P. Lequieu, J. K. Whitmer, and J. J. de Pablo, Coarse-grained modeling of DNA curvature, J. Chem. Phys. **141**, 165103 (2014).

[57] S. Assenza and Rubén Pérez, Accurate sequence-dependent coarse-grained model for conformational and elastic properties of double-stranded DNA, J. Chem. Theory Comput. **18**, 3239 (2022).

[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 A El Hassan and C R Calladine, The assessment of the geometry of dinucleotide steps in double-helical DNA: A new local calculation scheme, J. Mol. Biol. **251**, 648 (1995).

[60] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, and H. M. Berman, A standard reference frame for the description of nucleic acid base-pair geometry, J. Mol. Biol. **313**, 229 (2001).

[61] X.-J. Lu and W. K. Olson, 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, Nucleic Acids Res. **31**, 5108 (2003).

[62] R. Lavery, M. Moakher, J.H. Maddocks, D. Petkeviciute, and D. Zakrzewska, Conformational analysis of nucleic acids revisited: Curves+, Nucleic Acids Res. **37**, 5917 (2009).

[63] S. Li, W. K. Olson, and X.-J. Lu, Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures, Nucleic Acids Res. **47**, W26–W34 (2019).

[64] 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).

[65] 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).

[66] Some of these models are technically super-sets of RBP models, that include additional degrees of freedom. The corresponding RBP model is in these cases obtained by straight-forward marginalization of the excess degrees of freedom.

[67] A. V. Vologodskii and J. F. Marko, Extension of torsionally stressed DNA by external force, Biophys. J. **73**, 123 (1997).

[68] T. Lepage, François Képès, and I. Junier, Thermodynamics of long supercoiled molecules: Insights from highly efficient Monte Carlo simulations, Biophys. J. **109**, 135 (2015).

[69] S. K. Nomidis, F. Kriegel, W. Vanderlinden, J. Lipfert, and E. Carlon, Twist-bend coupling and the torsional response of double-stranded DNA, Phys. Rev. Lett. **118**, 217801 (2017).

[70] K. V. Klenin, A. V. Vologodskii, V. V. Anshelevich, A. M. Dykhne, and M. D. Frank-Kamenetskii, Computer simulation of DNA supercoiling, J. Mol. Biol. **217**, 413 (1991).

[71] Z. Liu and H. S. Chan, Efficient chain moves for Monte Carlo simulations of a wormlike DNA model: Excluded volume, supercoils, site juxtapositions, knots, and comparisons with random-flight and lattice models, J. Chem. Phys. **128**, 145104 (2008).

[72] T. Strick, Jean-François Allemand, V. Croquette, and D. Bensimon, Twisting and stretching single DNA molecules, Prog. Biophys. Mol. Biol. **74**, 115 (2000).

[73] J. Lipfert, M. Wiggin, J. W. J. Kerssemakers, F. Pedaci, and N. H. Dekker, Freely orbiting magnetic tweezers to directly monitor changes in the twist of nucleic acids, Nat. Commun. **2**, 439 (2011).

[74] J. S. Mitchell, J. Glowacki, A. E. Grandchamp, R. S. Manning, and J. H. Maddocks, Sequence-dependent persistence lengths of DNA, J. Chem. Theory Comput. **13**, 1539 (2017).

[75] B. S. Fujimoto and J. M. Schurr, Dependence of the torsional rigidity of DNA on base composition, Nature (London) **344**, 175 (1990).

[76] J. D. Moroz and P. Nelson, Torsional directed walks, entropic elasticity, and DNA twist stiffness, Proc. Natl. Acad. Sci. USA **94**, 14418 (1997).

[77] F. Mosconi, J. F. Allemand, D. Bensimon, and V. Croquette, Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers, Phys. Rev. Lett. **102**, 078301 (2009).

[78] Z. Bryant, F. C. Oberstrass, and A. Basu, Recent developments in single-molecule DNA mechanics, Curr. Opin. Struct. Biol. **22**, 304 (2012).

[79] F. Kriegel, N. Ermann, R. Forbes, D. Dulin, N. H. Dekker, and J. Lipfert, Probing the salt dependence of the torsional stiffness of DNA by multiplexed magnetic torque tweezers, Nucleic Acids Res. **45**, 5920 (2017).

[80] F. Lankaš, N. Špačková, M. Moakher, P. Enkhbayar, and J. Šponer, A measure of bending in nucleic acids structures applied to A-tract DNA, Nucleic Acids Res. **38**, 3414 (2010).

[81] T. Dršata, K. Réblová, I. Beššeová, J. Šponer, and F. Lankaš, rRNA C-Loops: Mechanical properties of a recurrent structural motif, J. Chem. Theory Comput. **13**, 3359 (2017).

[82] N. B. Becker and R. Everaers, From rigid base pairs to semi-flexible polymers: Coarse-graining DNA, Phys. Rev. E **76**, 021923 (2007).

[83] 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).

[84] Y. A. Gutiérrez Fosado, F. Landuzzi, and T. Sakaue, Coarse graining DNA: Symmetry, nonlocal elasticity, and persistence length, Phys. Rev. Lett. **130**, 058402 (2023).

[85] https://github.com/eskoruppa/PolyCG

[86] A. Voorspoels, J. Vreede, and E. Carlon, Rigid base biasing in molecular dynamics enables enhanced sampling of DNA conformations, J. Chem. Theory Comput. **19**, 902 (2023).

[87] E. Skoruppa, M. Laleman, S. Nomidis, and E. Carlon, DNA elasticity from coarse-grained simulations: The

[88] H. Chhabra, G. Mishra, Y. Cao, D. Prešern, E. Skoruppa, M. M. C. Tortora, and J. P. K. Doye, Computing the elastic mechanical properties of rodlike DNA nanostructures, J. Chem. Theory Comput. **16**, 7748 (2020).

[89] O. Gonzalez, D. Petkevičiūtė, and J. H. Maddocks, A sequence-dependent rigid-base model of DNA, J. Chem. Phys. **138**, 055102 (2013).

[90] D. Petkevičiūtė, A DNA coarse-grain rigid base model and parameter estimation from molecular dynamics simulations, Ph.D. thesis, Lausanne, EPFL (2012).

[91] L. de Bruin and J. H. Maddocks, cgDNAweb: A web interface to the cgDNA sequence-dependent coarse-grain model of double-stranded DNA, Nucleic Acids Res. **46**, W5 (2018).

[92] A. Noy and R. Golestanian, Length scale dependence of DNA mechanical properties, Phys. Rev. Lett. **109**, 228101 (2012).

[93] M. Segers, A. Voorspoels, T. Sakaue, and E. Carlon, Mechanisms of DNA-mediated allostery, Phys. Rev. Lett. **131**, 238402 (2023).

[94] Y. A. G. Fosado, D. Michieletto, C. A. Brackley, and D. Marenduzzo, Nonequilibrium dynamics and action at a distance in transcriptionally driven DNA supercoiling, Proc. Natl. Acad. Sci. USA **118**, e1905215118 (2021).

[95] M. Segers, A. Voorspoels, T. Sakaue, and E. Carlon, Mechanical properties of nucleic acids and the non-local twistable wormlike chain model, J. Chem. Phys. **156**, 234105 (2022).

[96] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, Cambridge, UK, 2007).

[97] *The Schur Complement and Its Applications*, edited by F. Zhang, Numerical Methods and Algorithms (Springer-Verlag, Berlin, 2005).

[98] R. Sharma, cgNA+: A sequence-dependent coarse-grain model of double-stranded nucleic acids, Phd thesis, Lausanne, EPFL (2023).

[99] The detailed comparison of the distributions of composite steps in small oligomers is based on the sequence CGCGGACTGCGCGCCTCAAGTCTGTCCGTATT CTACAGCAGCACTGCTGGAATCACCTTGGGCT ATATTGTTGCTACGCTACTGCTCTCCGGGCCA ACAAAGACGCGAAATAACTCTCAATATATAAG ACGCCTCCTACCCCACTTAGATCTGTACGGCA GTTGATGGGGGGGTGCAGACTACCAGAAAACT GAGAAAGTC.

[100] A. K. Mazur, Wormlike chain theory and bending of short DNA, Phys. Rev. Lett. **98**, 218102 (2007).

[101] Y.-Y. Wu, L. Bao, X. i. Zhang, and Z.-J. Tan, Flexibility of short DNA helices with finite-length effect: From base pairs to tens of base pairs, J. Chem. Phys. **142**, 125103 (2015).

[102] 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).

[103] E. N. Trifonov, R. K. Tan, and S. C. Harvey, *In Structure and Expression (Vol. 3): DNA Bending & Curvature* (Adenine Press, New York, NY, 1988).

[104] F. Kriegel *et al.* The temperature dependence of the helical twist of DNA, Nucleic Acids Res. **46**, 7998 (2018).

[105] M. Segers, E. Skoruppa, J. A. Stevens, M. Vangilbergen, A. Voorspoels, and E. Carlon, Comment on "Flexibility of short

DNA helices with finite-length effect: From base pairs to tens of base pairs" [J. Chem. Phys. **142**, 125103 (2015)], J. Chem. Phys. **155**, 027101 (2021).

[106] F. B. Fuller, Decomposition of the linking number of a closed ribbon: A problem from molecular biology. Proc. Natl. Acad. Sci. USA **75**, 3557 (1978).

[107] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette, Estimating the persistence length of a worm-like chain molecule from force-extension measurements, Biophys. J. **76**, 409 (1999).

[108] F. Brock Fuller, The writhing number of a space curve, Proc. Natl. Acad. Sci. USA **68**, 815 (1971).

[109] J. F. Marko and E. D. Siggia, Statistical mechanics of supercoiled DNA, Phys. Rev. E **52**, 2912 (1995).

[110] J. F. Marko, Biophysics of protein-DNA interactions and chromosome organization, Physica A **418**, 126 (2015).

[111] J. F. Marko and Sébastien Neukirch, Competition between curls and plectonemes near the buckling transition of stretched supercoiled DNA, Phys. Rev. E **85**, 011908 (2012).

[112] M. Emanuel, G. Lanzani, and H. Schiessel, Multiplectoneme phase of double-stranded DNA under torsion, Phys. Rev. E **88**, 022706 (2013).

[113] C. Bouchiat and M. Mézard, Elasticity model of a supercoiled DNA molecule, Phys. Rev. Lett. **80**, 1556 (1998).

[114] C. Bouchiat and M. Mézard, Elastic rod model of a supercoiled DNA molecule, Eur. Phys. J. E **2**, 377 (2000).

[115] H. Schiessel, *Biophysics for Beginners: A Journey through the Cell Nucleus* (Jenny Stanford Publishing, Singapore, 2022).

[116] https://github.com/eskoruppa/PolyMC

[117] S. B. Smith, Y. Cui, and C. Bustamante, Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules, Science **271**, 795 (1996).

[118] R. Stefl, H. Wu, S. Ravindranathan, V. Sklenář, and J. Feigon, DNA A-tract bending in three dimensions: Solving the dA$_4$T$_4$ vs. dT$_4$A$_4$ conundrum, Proc. Natl. Acad. Sci. USA **101**, 1177 (2004).

[119] M. Tompitak, H. Schiessel, and G. T. Barkema, Force responses of strongly intrinsically curved DNA helices deviate from worm-like chain predictions, Europhys. Lett. **116**, 68005 (2016).

[120] https://github.com/eskoruppa/PMCpy

[121] P. D. Dans, A. Pérez, I. Faustino, R. Lavery, and M. Orozco, Exploring polymorphisms in B-DNA helical conformations, Nucleic Acids Res. **40**, 10668 (2012).

[122] K. Liebl and M. Zacharias, Accurate modeling of DNA conformational flexibility by a multivariate Ising model, Proc. Natl. Acad. Sci. USA **118**, e2021263118 (2021).

[123] K. López-Güell, F. Battistini, and M. Orozco, Correlated motions in DNA: Beyond base-pair step models of DNA flexibility, Nucleic Acids Res. **51**, 2633 (2023).

[124] A. Vologodskii and M. D. Frank-Kamenetskii, Strong bending of the DNA double helix, Nucleic Acids Res. **41**, 6785 (2013).

[125] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, High flexibility of DNA on short length scales probed by atomic force microscopy, Nat. Nanotechnol. **1**, 137 (2006).

[126] S. E. Farr, E. J. Woods, J. A. Joseph, A. Garaizar, and R. Collepardo-Guevara, Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions, Nat. Commun. **12**, 2883 (2021).

[127] D. Michieletto, M. Lusic, D. Marenduzzo, and E. Orlandini, Physical principles of retroviral integration in the human genome, Nat. Commun. **10**, 575 (2019).

[128] C. A. Brackley, J. Allan, D. Keszenman-Pereyra, and D. Marenduzzo, Topological constraints strongly affect chromatin reconstitution *in silico*, Nucleic Acids Res. **43**, 63 (2015).

[129] C. Battaglia and D. Michieletto, Loops are geometric catalysts for DNA integration, Nucl Acids Res. **52**, 8184 (2024).

[130] E. Orlandini, D. Marenduzzo, and D. Michieletto, Synergy of topoisomerase and structural-maintenance-of-chromosomes proteins creates a universal pathway to simplify genome topology, Proc. Natl. Acad. Sci. USA **116**, 8149 (2019).

[131] S. K. Nomidis, E. Carlon, S. Gruber, and J. F. Marko, DNA tension-modulated translocation and loop extrusion by SMC complexes revealed by molecular dynamics simulations, Nucl. Acids Res. **50**, 4974 (2022).