From rigid base pairs to semiflexible polymers: Coarse-graining DNA

Nils B. Becker
Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden, Germany

Ralf Everaers
Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden, Germany and Laboratoire de Physique, ENS Lyon, 46, Allée d'Italie, 69364 Lyon Cedex 07, France

(Received 16 November 2006; revised manuscript received 22 May 2007; published 22 August 2007)

The elasticity of double-helical DNA on a nm length scale is captured in detail by the rigid base-pair model, whose conformation variables are the relative positions and orientations of adjacent base pairs. Corresponding sequence-dependent elastic potentials have been obtained from all-atom MD simulation and from high-resolution structural data. On the scale of 100 nm, DNA is successfully described by a continuous wormlike chain model with homogeneous elastic properties, characterized by a set of four elastic constants which have been measured in single-molecule experiments. We present here a theory that links these experiments on different scales, by systematically coarse-graining the rigid base-pair model to an effective wormlike chain description. The average helical geometry of the molecule is accounted for exactly, and repetitive as well as random sequences are considered. Structural disorder is shown to produce a small, additive and short-range correction to thermal conformation fluctuations as well as to entropic elasticity. We also discuss the limits of applicability of the homogeneous wormlike chain on short scales, quantifying the anisotropy of bending stiffness, the non-Gaussian bend angle distribution and the variability of stiffness, all of which are noticeable below a helical turn. The coarse-grained elastic parameters show remarkable overall agreement with experimental wormlike chain stiffness. For the best-matching potential, bending persistence lengths of dinucleotide repeats span a range of 37–53 nm, with a random DNA value of 43 nm. While twist stiffness is somewhat underestimated and stretch stiffness is overestimated, the counterintuitive negative sign and the magnitude of the twist-stretch coupling agree with recent experimental findings.

DOI: 10.1103/PhysRevE.76.021923 PACS number(s): 87.15.La, 87.14.Gg, 87.17.Aa

I. INTRODUCTION

Local elastic properties of DNA on a nm length scale play a vital role in basic biological processes such as chromatin organization [1,2] and gene regulation, via indirect readout [3–6] or via DNA looping [7–9]. The structure and elasticity of double helical DNA on the nm scale is often described using rigid base-pair chain (RBC) models, in which the relative orientation and translation of adjacent base pairs (BPs) specify the conformation of the molecule [10,11]. Parameter sets for rigid base-pair step elastic potentials were obtained from molecular dynamics simulation [12,13] and from an analysis of high resolution crystal structure data [14]. We have found qualitative but not quantitative agreement between these different potentials in a recent study on indirect readout in protein-DNA binding [15].

On a mesoscopic length scale, it is possible to directly measure force-extension relations for DNA in single-molecule experiments [16]. For small external forces, DNA behaves as a wormlike chain (WLC) [17], i.e., an inextensible semiflexible polymer with a single parameter, the bending persistence length, and no explicit sequence dependence. An extension of the classical WLC model, reflecting the chiral symmetry of the DNA double helix, includes coupled twisting and stretching degrees of freedom [18–21]. These become important in a force regime where the DNA molecule is already pulled straight but not yet overstretched [22]. Interestingly, recent measurements indicate that DNA overtwists when stretched in the linear response regime [23,24].

The issue of relating atomistic and mesoscopic descriptions of DNA elasticity has been addressed mainly by simulation of oligonucleotides. Normal mode analysis using atomistic [25] or knowledge-based RBC potentials [26] can give an impression of global bending and twisting modes but does not offer a systematic coarse-graining prescription. In an MD simulation study in explicit solvent, a full set of elastic constants of a fluctuating global helical axis were determined [27]. A recent study [28] extended this approach, exposing technical difficulties concerning the very definition of the global helical axis, and deriving convergence criteria.

In this paper we present a theoretical approach to establish a relation between these different levels of detail. In a first step, we establish a method to systematically coarse-grain a homogeneous or repetitive sequence RBC to the WLC scale. In contrast to recent work [29], we take the full average helical geometry of the chain into account. As a result, we obtain exact expressions for the average helical parameters and the full set of stiffnesses for bend, twist, stretch, as well as twist-stretch coupling. We list their values for all six dinucleotide repeats.

It has been pointed out [30] that the total apparent persistence length of a WLC is composed of a static part which originates from the sequence-dependent equilibrium bends of the molecule, and a dynamic part induced by thermal fluctuations, and their relative contributions have been measured [31,32]. In a second step, we adapt this idea to the case of a random sequence RBC: Extending our coarse-graining procedure to also include structural variability, we calculate the conformational statistics of rigid base-pair chain ensembles with random, uncorrelated base sequence. We arrive again at an effective homogeneous WLC description. Finally, we also

quantify the deviations from the effective WLC due to stiffness variability on short scales.

The article is organized as follows: After a description of the model (Sec. II), our coarse-graining procedure is presented for the case of a homogeneous sequence in Sec. III. The extension to random DNA is made in Sec. IV. In Sec. V, the theory is related to observables measured in different experimental situations. Section VI contains detailed comparisons of the predictions of the different available parametrizations of the RBC model with experiments on the mesoscopic scale, as well as a discussion on the limitations of the WLC model on short scales. Conclusions are presented in Sec. VIII.

II. RIGID BASE PAIR MODEL OF DNA

In canonical double-stranded DNA, Watson-Crick base pairs are stacked into a helical column. We can fix a Cartesian coordinate frame to the center of each base pair in a standard way [33,34], effectively averaging out internal distortions within the base pair. By convention, the z axis of this right handed orthonormal frame is normal to the base pair plane and points towards the $3'$ direction of the preferred strand, while the x axis points towards the major groove.

The configuration in space of the chain is specified by the sequence of these frames, i.e., by a 3×3 rotation matrix R together with three Cartesian coordinates of the origin p, for each base pair step. Only for homogeneous, nonfluctuating DNA in an idealized B-form do all frames lie on a straight line, with their body z axes pointing in a single direction. Generically, the frames are displaced and rotated away from this idealized arrangement, due to both thermal fluctuations and sequence-dependent variations in the equilibrium conformations.

A. Homogeneous representation

We represent the rotation and translation of the $(k+1)$-th base pair frame relative to the kth frame by a 4×4 matrix, written in block form as

$$g_{kk+1} = \begin{bmatrix} R_{kk+1} & p_{kk+1} \\ 0\,0\,0 & 1 \end{bmatrix}. \tag{1}$$

Throughout the article, matrices in square brackets will have exactly this block structure. In idealized B-DNA along the z axis, $p_{kk+1} \propto d_3 = (0,0,1)$, and R_{kk+1} is a rotation about d_3.

This so-called homogeneous representation (see, e.g., Ref. [35]) has the advantage that the translation and rotation relating frames k and $l>k$ can be obtained by matrix multiplication along the chain

$$g_{kl} = g_{kk+1} g_{k+1k+2} \cdots g_{l-1l}. \tag{2}$$

For convenience we fix the lab frame on the first base pair, so g_{1k} represents the frame k relative to the lab. Observe that as a general rule $g_{kk+1} = g_{1k}^{-1} g_{1k+1}$. Throughout the article, the 4×4 identity matrix is denoted by e. For example, $g_{kk} = e$.

B. Exponential coordinates

At finite temperature, a base pair step $g = g_{kk+1}$ in a RBC fluctuates around a mean or equilibrium value g_0. To parametrize these fluctuations, we first introduce coordinates suitable to describe small deviations from g_0. We will then characterize thermal fluctuations in terms of their second moments. In our model, we neglect possible couplings between neighboring base-pair steps [36,37].

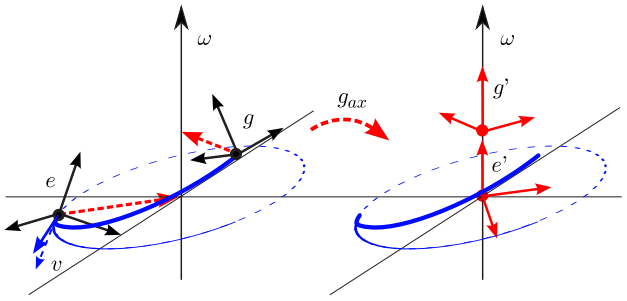

FIG. 1. (Color online) Frame geometry. A base pair step, connecting the base-pair fixed material frames e and g (left-hand side). The frame origin trace of the corresponding screw motion is shown in blue. It has initial tangent v. By right multiplication with g_{ax}, the same step can be described using the frames e' and g' (red, right-hand side). They lie on the helical axis and point into its direction ω.

Any continuous group can be locally parametrized by its infinitesimal generators via the exponential map. In the g-matrix representation, this is the ordinary matrix exponential and the group generators $\{X_i\}$ are 4×4 matrices. Explicitly, in block form,

$$X_i = \begin{bmatrix} \epsilon_i & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{with } (\epsilon_i)_{jk} = \epsilon_{jik} \tag{3a}$$

and

$$X_{i+3} = \begin{bmatrix} 0 & d_i \\ 0 & 0 \end{bmatrix}, \quad \text{with } (d_i)_j = \delta_{ij}. \tag{3b}$$

Here, ϵ_{ijk} and δ_{ij} are the antisymmetric and symmetric tensors, respectively, and $1 \leqslant i,j,k \leqslant 3$. A rotation around the d_i axis is generated by X_i while a translation along d_i is generated by X_{i+3}. The generators satisfy the usual commutation relations of angular and linear momentum. Any group element g can be written as

$$g = \begin{bmatrix} R(\xi) & p(\xi) \\ 0 & 1 \end{bmatrix} = \exp[\xi^i X_i] \tag{4}$$

which defines ξ^i as the exponential coordinates of g [38]. The coordinate vector can be split up into two three-dimensional parts $\xi = (\omega, v)$. Both have a geometrical meaning: ω points along the rotation axis of R with $\|\omega\|$ equal to the total rotation angle, and v is the initial tangent $\frac{d}{ds}|_0 p(s\xi)$, see Fig. 1. All of $SE(3)$ except for a measure zero set is covered one-to-one by the coordinate range $\{\xi \in \mathbb{R}^6 | \|\omega\| < \pi\}$. We denote the exponential coordinates of the equilibrium step by $\xi_0 = (\omega_0, v_0)$, so that $g_0 = \exp[\xi_0^i X_i]$.

C. Mean and covariance

Consider a base pair step subject to random fluctuations. The corresponding deformation probability distribution is

$p(g)dg$. To describe deviations from equilibrium, we use exponential coordinates also for $g_0^{-1}g$. Thus, a fluctuating step is written as

$$g = g_0 \exp[\xi^i X_i] = g_0(e + \xi^i X_i) + O(\|\xi\|^2). \tag{5}$$

The mean or equilibrium conformation is now *defined* by the requirement that the random deformations ξ have a distribution with zero mean. To determine g_0, from $p(g)$ we compute the distribution of $\xi=\log(g_0^{-1}g)$ and optimize g_0 until $\langle\xi\rangle=0$. This is always possible for not too wide step distributions [39], and can be implemented by a gradient search with no numerical problems. The corresponding deformation probability distribution is

$$p(\xi)dV_\xi = p(\xi)A(\xi)d\xi^1 \cdots d\xi^6. \tag{6}$$

Here, p is the probability density function (PDF) and $dV_\xi = A(\xi)d^6\xi$ is the invariant volume element on the group, which is the Jacobian factor corresponding to our choice of curvilinear coordinates [40]. We can approximate A as a constant, see Appendix B.

In a second step, the covariance matrix around g_0, is obtained as $C^{ij}=\langle\xi^i\xi^j\rangle$. Note that the very definition of ξ depends on the equilibrium conformation g_0, and so does C.

This formulation has the advantage that deformations with respect to different equilibrium positions are directly comparable and no distortions due to curvilinear coordinates occur. It is essential for our formalism which relates fluctuations given with respect to different frames (see below). Unfortunately, this definition of base-pair step deformations differs from those used in available software such as Refs. [41,42]. We explain in Appendix A how to convert between our exponential coordinates and the coordinate set used in Refs. [42,43]. A somewhat related approach makes use of exponential coordinates for the rotation part of the frame transformation only [44].

We emphasize that we have not specified the source of fluctuations yet. Below, we will consider either steps fluctuating thermally, or steps that in addition have fluctuations due to random sequence.

D. Combining steps

Using the matrix formalism described above, we can combine a chain of m consecutive steps into one compound step, which in turn is described in terms of its mean and covariance matrix. The latter can be computed in a straightforward way as long as the combined fluctuations of the compound step stay small. In other words, the short chain must be well approximated by a (helical) rigid rod.

In this section, we consider pure thermal deformation fluctuations. The thermal mean conformations $g_0(\sigma)$ and the thermal covariance matrices $C(\sigma)$ thus depend on step sequence $\sigma=bb'$ (e.g., $\sigma=5'\mathrm{AG}3'$).

Starting with $m=2$, let the double step sequence $\sigma_{13}=b_1b_2b_3$, and denote the means and covariances of the two steps by $g_0(\sigma_{kk+1})$ and $C(\sigma_{kk+1})$, respectively. The compound step can be written to first order in the fluctuations, as

$$\begin{aligned} g_{13} &= g_0(\sigma_{12})(e + \xi_{12}^i X_i)g_0(\sigma_{23})(e + \xi_{23}^i X_i) \\ &= g_0(\sigma_{12})g_0(\sigma_{23})[e + \xi_{12}^i g_0^{-1}(\sigma_{23})X_i g_0(\sigma_{23}) + \xi_{23}^i X_i]. \\ &= g_0(\sigma_{12})g_0(\sigma_{23})(e + \xi_{13}^i X_i). \end{aligned} \tag{7}$$

We introduce some standard notation. The 6×6 adjoint matrix $\mathrm{Ad}\,g$ is defined for any $g\in\mathrm{SE}(3)$ by $gX_ig^{-1} = (\mathrm{Ad}\,g)^j{}_i X_j$. Explicitly, if $g=(R,p)$, one finds

$$\mathrm{Ad}\,g = \begin{pmatrix} R & 0 \\ p^i\epsilon_i R & R \end{pmatrix}, \tag{8}$$

written in 3×3 blocks. The Ad matrices form a representation of the group, i.e., we have the general relation $\mathrm{Ad}(g^{-1}h)=\mathrm{Ad}^{-1}\,g\,\mathrm{Ad}\,h$.

We can now write $\xi_{13}=\mathrm{Ad}\,g_0^{-1}(\sigma_{23})\xi_{12}+\xi_{23}$. Finally, the mean of the compound step $g_0(\sigma_{13})=g_0(\sigma_{12})g_0(\sigma_{23})$ and the covariance matrix

$$C(\sigma_{13}) = \mathrm{Ad}\,g_0^{-1}(\sigma_{23})C(\sigma_{12})\mathrm{Ad}^{\mathsf{T}}\,g_0^{-1}(\sigma_{23}) + C(\sigma_{23}), \tag{9}$$

where we have used the model assumption that thermal fluctuations of different steps are uncorrelated.

This formula has a straightforward extension to $m>2$. Consider

$$g_{1m+1} = g_0(\sigma_{12})(e + \xi_{12}^i X_i) \cdots g_0(\sigma_{mm+1})(e + \xi_{mm+1}^i X_i). \tag{10}$$

Commuting all deformations on the right-hand side to the right and using the representation property of Ad, one arrives at the first order compound step

$$\xi_{1m+1} = \sum_{k=1}^{m} \mathrm{Ad}\,g_0^{-1}(\sigma_{k+1m+1})\xi_{kk+1}. \tag{11}$$

Here $g_0(\sigma_{kk'+1})=\Pi_{l=k}^{k'}g_0(\sigma_{ll+1})$, where the product is understood in increasing order. The compound mean conformation is $g_0(\sigma_{1m+1})$, and since all single step deformations are independent random variables, the compound covariance $C(\sigma_{1m+1})$ equals

$$\sum_{k=1}^{m} \mathrm{Ad}\,g_0^{-1}(\sigma_{k+1m+1})C(\sigma_{kk+1})\mathrm{Ad}^{\mathsf{T}}\,g_0^{-1}(\sigma_{k+1m+1}). \tag{12}$$

We have now characterized compound steps in terms of their mean and covariance. This will allow us to treat repetitive, poly-(σ_{1m}) DNA on the same footing as homogeneous DNA. The validity of this combination of steps is limited by the first order approximation for the deformations. For combining, it is necessary that the compound step angles stay small, $\|\omega_{1m}\|\ll1$.

III. MAPPING A HOMOGENEOUS RBC TO A HOMOGENEOUS WLC

What is the effective WLC model that corresponds to a given rigid body chain? We will address this question first for the case of homogeneous (or repetitive, see above) sequence.

Up to this point, step deformations and therefore also the covariance matrices were given with respect to the equilibrium step conformation g_0, i.e., as small changes of the end base-pair frame with respect to the start base-pair frame, see Eq. (5). Note that, in general, the end base-pair frame is both offset and tilted relative to the local helical axis defined by g_0. However to relate the RBC deformations to a coarse-grained WLC model, we are much more interested in the elastic properties of the centerline of the chain, which in general need not even intersect a base.

Once a covariance matrix for deformations of centerline segments is known, the large-scale elastic properties of the WLC are easily determined. For example, the bending persistence length of the WLC is defined as the decay length of bend angle correlations and thus depends only on the second moment of the centerline bend angle distribution.

A. Helical centerline

In the case of a nonfluctuating chain with identical steps, the centerline can be conveniently described using the matrix formalism introduced above. The screw motion $s \mapsto \exp[s\xi^i X_i]$ joins the identity frame e with g as s increases from 0 to 1, see Fig. 1. Its screw axis is determined by a vector from the origin of e to a point on the axis, given by $p_{\mathrm{ax}}=\|\omega\|^{-2}\omega\times v$, and by its direction, ω. It is the "local helical axis" [41] associated with the base pair step g. When concatenating many identical steps g one generates a RBC with frame origins lying on a regular helix with this axis.

In addition to p_{ax} we can define a matrix R_{ax} which rotates e such that ω becomes its third direction vector. One choice is to take p_{ax} as the second new direction. In combination, we then get

$$g_{\mathrm{ax}} = \begin{bmatrix} R_{\mathrm{ax}} & p_{\mathrm{ax}} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \dfrac{(\omega\times v)\times\omega}{\|(\omega\times v)\times\omega\|} & \dfrac{\omega\times v}{\|\omega\times v\|} & \dfrac{\omega}{\|\omega\|} & \dfrac{\omega\times v}{\|\omega\|^2} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (13)$$

which takes e to a frame $e'=eg_{\mathrm{ax}}=g_{\mathrm{ax}}$ sitting on the helix axis with its third direction pointing along it. One can check that $g'=gg_{\mathrm{ax}}$ shares these properties. The primed, on-axis frames are "local helical axis systems" in the terminology of Ref. [41], see Fig. 1

Under the influence of thermal fluctuations, the helical structure of the chain becomes irregular. It turns out that in this case the definition of a centerline is problematic in itself. One could try to define it as the local helical axis for each individual base pair step. This has the disadvantage that for a fluctuating chain, the local centerline pieces of consecutive steps do not form a continuous curve, since they are laterally offset. An alternative approach is to fit a continuous centerline globally to a stretch of a RBC, using the "Curves" algorithm [41], as carried out in Ref. [27]. The fitting procedure involves a free parameter, namely, the relative weight of translational and rotational deviations from an ideal helix shape. By a reasonable choice of this relative weight *a posteriori*, periodic artifacts in the analysis can be reduced but not eliminated [28]. Also, the fact that the resulting centerline depends nonlocally on the base pair step conformations introduces artificial correlations on the length scale over which the fitting procedure extends.

We circumvent these problems in three steps. First we transform all rigid base pairs of the chain to new frames of reference. These are chosen such that without fluctuations, all new BP frames lie exactly on, and point in the direction of, a single straight helical axis. We can then identify and average over the unwanted shear degrees of freedom. In a last step, this reduced model is averaged over the helical phase angle and mapped to the WLC models.

B. On-axis RBC

We would like to transform small deviations from an equilibrium conformation g_0 into small deviations from a version of g_0 which is on axis. Consider first a regular helix composed of identical g_0 steps. As explained in Sec. III B, the on-axis step between the kth and $(k+1)$-th on-axis frames is

$$g_{0\|} = (g_0^{k-1} g_{\mathrm{ax}})^{-1} g_0^k g_{\mathrm{ax}} = g_{\mathrm{ax}}^{-1} g_0 g_{\mathrm{ax}}, \quad (14)$$

where g_{ax} is the on-axis transformation (13) corresponding to g_0. Since $g_{0\|}$ is a transformation between on-axis frames, its rotation and displacement vectors point along the d_3 axis, $\omega_{0\|}=\|\omega_{0\|}\|d_3$ and $p_{0\|}=\|p_{0\|}\|d_3$.

For a step $g_{kk+1}=g_0\exp[\xi^i X_i]$ of a fluctuating RBC we calculate an on-axis version as

$$(g_{1k}g_{\mathrm{ax}})^{-1} g_{1k+1} g_{\mathrm{ax}} = g_{\mathrm{ax}}^{-1} g_{kk+1} g_{\mathrm{ax}} = g_{0\|} g_{\mathrm{ax}}^{-1} \exp[\xi^i X_i] g_{\mathrm{ax}}. \quad (15)$$

The three right-most factors in Eq. (15) clearly represent the deviation from the on-axis equilibrium step $g_{0\|}$. Pulling a similarity transformation inside the exponential series we can then rewrite Eq. (15) as

$$g_{0\|} g_{\mathrm{ax}}^{-1} \exp[\xi^i X_i] g_{\mathrm{ax}} = g_{0\|} \exp[\xi_\|^i X_i], \quad (16)$$

where the deviation from the on-axis equilibrium step $\xi_\| = \mathrm{Ad}\, g_{\mathrm{ax}}^{-1}\xi$. $\xi_\|$ has zero mean and covariance matrix $C_\|^{ij} = \langle \xi_\|^i \xi_\|^j \rangle$,

$$C_\| = \mathrm{Ad}\, g_{\mathrm{ax}}^{-1} C \mathrm{Ad}^{\mathsf{T}}\, g_{\mathrm{ax}}^{-1}. \quad (17)$$

The RBC composed of steps (16) is an equivalent description of the original chain, which we may call its on-axis version. Intuitively, to each fluctuating frame g_{1k} of the original chain, we rigidly connected a frame g'_{1k} in such a way that the primed, on-axis chain fluctuates about a straight, but still twisted, equilibrium conformation. This is illustrated in Fig. 2. The equilibrium conformations generate a tilted helix that is offset from the helical centerline. Thermal fluctuations distort it, producing an irregular helix. However, on average, the on-axis configuration is exactly lined up on a straight helical axis. Note that we had no need to compute a fluctuating axis explicitly, nor choose a weighting factor [28].

C. Averaging over shear variables

The on-axis RBC has the nice property that the translational fluctuations $(\xi_\|^4, \xi_\|^5) = (v_\|^1, v_\|^2)$ are now exactly transver-

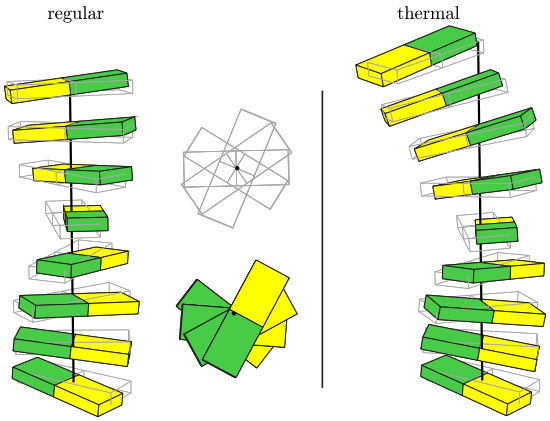

FIG. 2. (Color online) Equivalent descriptions of a poly-G RBC. Left: Colored blocks represent base pairs in their equilibrium conformations. Wireframe blocks represent their on-axis counterparts. Right: Thermal fluctuations distort the helix. (MP parameter set, base pair size scaled down by 1/2 for clarity.)

sal to the equilibrium helix axis. They are pure shear modes and do not contribute to compression fluctuations along the chain. Let $\eta=(\omega_\parallel, v_\parallel^3)$ be the vector of the four remaining variables. Noting that the volume element $A=A(\omega)$ depends only on the angular part (see Appendix B), we write/

$$\langle \eta^i \eta^j \rangle = \int \underbrace{d^3\omega_\parallel dv_\parallel^3 A(\omega_\parallel)}_{dV_\eta} \underbrace{\int dv_\parallel^1 dv_\parallel^2 p(\xi_\parallel)}_{p(\eta)} \eta^i \eta^j, \tag{18}$$

from which one can see that the 4×4 covariance matrix $\tilde{C}^{ij}=\langle \eta^i \eta^j \rangle$ is the same as $C_\parallel$ with its $v_\parallel^1, v_\parallel^2$ rows and columns deleted. Thus, η has a zero mean distribution with covariance matrix $\tilde{C}$. Here and in the following, $\tilde{\cdot}$ indicates deletion of the rows and columns 4 and 5 in a 6×6 matrix. E.g., $\tilde{\mathrm{Ad}}$ is the 4×4 adjoint matrix. Its on-axis version $\tilde{\mathrm{Ad}}_{g_{0\parallel}}$ has a particularly simple form. Using Eq. (8) and noting that $p_{0\parallel} \propto \omega_{0\parallel} \propto d_3$ we obtain

$$\tilde{\mathrm{Ad}}_{g_{0\parallel}} = \begin{pmatrix} \cos\|\omega_0\| & \sin\|\omega_0\| & 0 & 0 \\ -\sin\|\omega_0\| & \cos\|\omega_0\| & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{19}$$

D. Averaging over the helical phase

A shear-averaged, on-axis RBC still has a finite equilibrium twist and anisotropic bending stiffness. To relate it to a WLC with isotropic bending rigidity, we perform an average over a continuous helical phase angle rotation of the reference frame [45]. An on-axis covariance matrix which is rotated by a helical phase angle ϕ around the average local helical axis [see Eq. (17)], is

$$\tilde{C}_\phi = \tilde{\mathrm{Ad}}_{g_\phi} \tilde{C} \tilde{\mathrm{Ad}}_{g_\phi}^{\mathsf{T}}, \tag{20}$$

where $g_\phi=\exp[\phi X_3]$ is a pure rotation by an angle ϕ around d_3. Since $\tilde{\mathrm{Ad}}_{g_\phi}$ has the form (19), the helical phase average comes out as

$$\bar{C} = \frac{1}{2\pi}\int_0^{2\pi} \tilde{C}_\phi d\phi = \begin{pmatrix} \frac{\tilde{C}^{11}+\tilde{C}^{22}}{2} & 0 & 0 & 0 \\ 0 & \frac{\tilde{C}^{11}+\tilde{C}^{22}}{2} & 0 & 0 \\ 0 & 0 & \tilde{C}^{33} & \tilde{C}^{34} \\ 0 & 0 & \tilde{C}^{34} & \tilde{C}^{44} \end{pmatrix}. \tag{21}$$

From $\bar{C}$ one can read off the bend persistence length as $l_b=h_\parallel/\bar{C}^{11}$. For example the mean square end-to-end distance of a homogeneous chain $\langle R^2\rangle \propto 2l_b l$ for contour lengths $l \gg l_b$. The torsional modulus, normalized to units of length is called the twist persistence length $l_t=h_\parallel/\bar{C}^{33}$ [46] (see, e.g. [45]). Here, the on-axis helical rise $h_\parallel=\|p_{0\|}\|$. The WLC stiffness matrix $\beta\bar{S}=\bar{C}^{-1}$ can be found by inversion and has the same block structure as $\bar{C}$, see also Appendix B.

When the considered covariance matrix actually belongs to a compound step, $\bar{C}=\bar{C}_{1m+1}$, all of the elastic parameters can be extracted in the same way, the only difference being that $h_\parallel$ has to be taken as the total helical rise on the compound step. Also, $\bar{S}$ will be the compound step stiffness, which can be renormalized to one BP step by multiplying with m.

IV. MAPPING A RANDOM SEQUENCE RBC TO A HOMOGENEOUS WLC

Instead of homogeneous or repetitive sequences, we now turn our attention to random sequences, as a generic approximation to the properties of natural DNA. The crucial difference is that the relaxed conformation of any realization of random DNA is no longer a regular helix, and that the relaxed conformations of consecutive steps are correlated due to sequence continuity. To get around these complications, we introduce an ensemble average over sequence randomness in addition to the thermal average at fixed sequence.

A. Random sequence RBC

By a random sequence rigid base-pair chain we mean a sequence of rigid base-pair frames generated iteratively in the following way: Start with some choice of base at position $i=1$. Then, for each new base pair $i+1$,

(1) Choose a base identity b_{i+1} at random, following a fixed base distribution $p(b)$ [47].

(2) Generate the BP step conformation g_{ii+1}. Due to thermal fluctuations, this conformation is also random. It follows a PDF $p(g|\sigma)$ whose center and width depend parametrically on the step sequence $\sigma_{ii+1}=b_i b_{i+1}$.

After $m-1$ iterations, one has a realization $\sigma_{1m}=b_1 \ldots b_m$ of the random sequence and a corresponding realization $g_{1m}=g_{12} \cdots g_{m-1m}$ of conformations.

Generally, we denote by $\langle f(g_{1m})\rangle$ an expectation value of some function f over conformations of a thermal, random sequence RBC ensemble. It can be carried out sequentially:

$$\langle f(g_{1m})\rangle = \langle\langle f(g_{1m})|\sigma_{1m}\rangle\rangle = \sum_{b_1 \ldots b_m} p(\sigma_{1m})\langle f(g_{1m})|\sigma_{1m}\rangle. \tag{22}$$

Here the conditional expectation $\langle \cdots|\sigma\rangle$ [48] is the thermal average and $\langle \cdots\rangle$ denotes the global average over both thermal and sequence randomness. The second equality in Eq. (22) follows because $\langle f(g_{1m})|\sigma_{1m}\rangle$ is already averaged over thermal fluctuations.

A random sequence RBC captures the effects of sequence dependent structure and stiffness. It is a good model for DNA if we assume that (a) sequences of bases are independent, that (b) thermal conformations of base-pair steps are independent, and that (c) the step conformation distributions are independent of flanking base sequence. All of these assumptions are wrong in general, but may be considered reasonable first approximations. In particular, relaxing (a) requires extra knowledge about sequence statistics. Also, no parametrizations of conformational correlations are yet available that would allow to relax (b). In MD simulation studies [36,37], (c) was investigated, and a dependence of stiffness and equilibrium conformations on flanking base sequence was found. This dependence is, however, much weaker than the spread among the existing single step parametrizations (see Sec. VI B 1) and thus can be reasonably neglected in a first approximation.

B. Adapting the coarse-graining procedure

The method presented in Sec. III, consists of expressing the fluctuating conformations as deformations with respect to a helical reference structure, before transforming to an idealized, on-axis helix. Finally irrelevant degrees of freedom are identified and averaged over.

Two new difficulties arise in a random sequence RBC: The first is the choice of reference structure when structural disorder is present, since the chain no longer forms a regular helix in the absence of thermal fluctuations. This leads to the same problems of defining a centerline as discussed in Sec. III A. We will therefore *not* choose the approach of expressing thermal deformations of each sequence realization with respect to irregular on-axis frames. Rather, our strategy will be to describe random sequence RBC conformations, just like those of homogeneous sequences, as deformations from some sequence-averaged, regular helix. This issue is addressed in Sec. IV C.

The second difficulty comes from the fact that the sequence distribution features independent bases, while the conformation distributions depend on the base-pair steps. Loosely speaking, the sequence distribution lives on the nodes of the model while the conformation distribution lives on its links. We sketch in Sec. IV D and explain in Appendix C how to treat the short-range correlations introduced by the requirement of sequence continuity.

C. Combining thermal and sequence randomness for a single BP step

A base-pair step with sequence σ_{ii+1} in a chain fluctuates in a thermal environment. Its sequence-dependent thermal mean conformation as well as the covariance matrix are moments of the conditional PDF $p(g_{ii+1}|\sigma_{ii+1})$. What changes when σ_{ii+1} itself is a random variable?

To start with, we observe that the sequence-dependent variability in equilibrium conformations of B-DNA BP steps is in fact smaller than the average thermal fluctuation size. Since only the limit of small thermal deformations is considered throughout, it is consistent to use the same limit for the sequence induced conformational variability.

The basic idea then is to treat sequence variability on the same footing as thermally induced fluctuations; we add the sequence induced deviations from a global equilibrium conformation as another independent source of randomness. That is, we consider a random sequence step $g=g_0 \exp[\xi^i X_i]$ which now fluctuates around a sequence-independent global center g_0, characterized by a covariance matrix $C^{ij}=\langle \xi^i \xi^j\rangle$ resulting from both sequence and thermal fluctuations.

We now need to calculate the global center g_0 and the total covariance C from the thermal and sequence statistics. Recalling that $\langle \cdots\rangle$ denotes a thermal and sequence ensemble average, we can determine g_0 by the condition that $\langle \xi\rangle=0$.

We then split the deformation from g_0 into sequence plus thermal parts: $\xi=\langle \xi|\sigma\rangle+(\xi-\langle \xi|\sigma\rangle)$. Note that the thermal equilibrium deformation $\langle \xi|\sigma\rangle$ is a random variable, depending on σ, while $(\xi-\langle \xi|\sigma\rangle)$ is the random thermal deformation. We now discuss their relation.

Within a regime of linear response, the deformation energy of a step with fixed sequence σ is a quadratic function of the deviation from the thermal equilibrium value $\langle \xi|\sigma\rangle$. The associated thermal covariance matrix is sequence dependent:

$$C_{\mathrm{th}}^{ij}(\sigma) = \langle (\xi - \langle \xi|\sigma\rangle)^i (\xi - \langle \xi|\sigma\rangle)^j |\sigma\rangle. \tag{23}$$

Comparing this with the thermal fluctuations introduced in Sec. II D, one sees that $g_0(\sigma) \simeq g_0(e+\langle \xi^i|\sigma\rangle X_i)$. Also, Eq. (23) agrees with the $C(\sigma)$ used there to quadratic order in the deformations.

Similarly, the covariance of the thermal mean values can be written down. It is sequence independent:

$$C_0^{ij} = \langle \langle \xi|\sigma\rangle^i \langle \xi|\sigma\rangle^j\rangle, \tag{24}$$

where the outermost expectation is effectively taken with respect to $p(\sigma)$ only, see Eq. (22).

What is the total covariance C? The two sources of randomness are of independent physical origin, but are not independent random variables: The thermal conformation distribution depends on σ. Therefore, $p(\xi|\sigma)p(\sigma)=p(\xi,\sigma) \neq p(\xi)p(\sigma)$. Splitting up the deformation into thermal and sequence parts, one computes

$$\langle \xi^i \xi^j \rangle = \langle \langle \xi | \sigma \rangle^i \langle \xi | \sigma \rangle^j \rangle + \langle (\xi - \langle \xi | \sigma \rangle)^i (\xi - \langle \xi | \sigma \rangle)^j \rangle + \langle \langle \xi | \sigma \rangle^i (\xi - \langle \xi | \sigma \rangle)^j \rangle + \langle (\xi - \langle \xi | \sigma \rangle)^i \langle \xi | \sigma \rangle^j \rangle. \quad (25)$$

Now note that $\langle \langle \xi | \sigma \rangle^i (\xi - \langle \xi | \sigma \rangle)^j | \sigma \rangle = 0$ trivially. Using this with the identity $\langle \cdots \rangle = \langle \langle \cdots | \sigma \rangle \rangle$ in Eq. (25), one sees that the cross-terms actually vanish. The simple result is that the sequence induced static covariance and the sequence-averaged thermal covariance add up to C:

$$C = C_0 + \langle C_{\text{th}}(\sigma) \rangle. \quad (26)$$

In summary, given the covariance (or stiffness) matrices and equilibrium values of all sixteen dinucleotide steps, and a distribution of relative step frequencies $p(\sigma)$, by computing g_0 and C we have characterized a single, thermally fluctuating random sequence step in terms of its center and second moment. The global equilibrium step g_0 defines a regular helix which we take as a reference structure. Deformations from this reference are described by the total covariance C which includes a contribution from sequence-induced conformational variability. The relation (26) is the generalization of the well known additivity of inverse static and dynamic persistence length of the irregular WLC, to the rigid base-pair chain model.

D. Combining thermal and sequence randomness of a compound step

The basic idea of splitting the deformations into static disorder parts and thermal fluctuations can be carried out for a random sequence compound step of length $m+1$ bases. While by assumption thermal fluctuations of neighboring steps are independent random variables, the sequences of different bases, not steps, are independent. Any realization of a random sequence of dinucleotide steps must be "continuous," e.g., σ_{12}=AG implies that σ_{23} can only start with a G. This clearly correlates neighboring step sequences, and thus also their static offsets $\langle \xi | \sigma \rangle$.

The correlations introduced by sequence continuity can be captured by an effective uncorrelated RBC, with equilibrium steps g_0 and effective covariance $\hat{C}$. The matrix $\hat{C}$ is again a sum of static and thermal fluctuations, analogous to Eq. (26), and is derived in detail in Appendix C.

E. On-axis transformation and averaging

Having identified the regular reference structure to use, we now just follow the coarse-graining procedure from Sec. III. As a first step, we transform the total deformation fluctuations onto the average helical axis $\xi_\parallel = \mathrm{Ad}\, g_{\text{ax}}^{-1} \xi$, where g_{ax} is defined by the global equilibrium g_0 via Eq. (13). The on-axis deformation then still has zero mean $\langle \xi_\parallel \rangle = 0$ and covariance matrix

$$C_\parallel = \mathrm{Ad}\, g_{\text{ax}}^{-1} \hat{C}\, \mathrm{Ad}^{\mathsf{T}}\, g_{\text{ax}}^{-1}. \quad (27)$$

One realization of a random sequence RBC, together with its on-axis version, is shown in Fig. 3.

The next step in the coarse-graining procedure is to average over unwanted degrees of freedom. The first average is

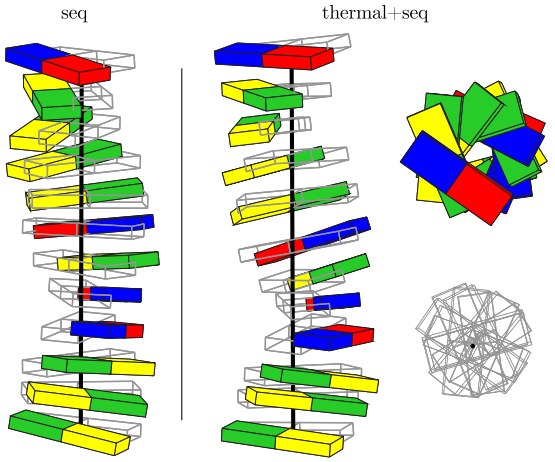

FIG. 3. (Color online) Equivalent descriptions of a realization of a random-sequence RBC. "seq": Colored blocks represent base pairs in their thermal equilibrium conformations. Wireframe blocks represent their on-axis counterparts, which do *not* lie on a straight line without sequence averaging. "thermal+seq": The same, but with added thermal fluctuations. The top views show the reduced helix axis offsets of the on-axis frames. (MD parameter set, base-pair size scaled down by 40% for clarity, sequence *GCGTTGTGGGCT*.)

that over the shear degrees of freedom $(v_\parallel^1, v_\parallel^2)$. As explained in Sec. III C, the result is that the remaining four variables $\eta = (\omega_\parallel, v_\parallel^3)$ have a 4×4 covariance matrix $\tilde{C}$ which equals $C_\parallel$ with its $(v_\parallel^1, v_\parallel^2)$ rows and columns deleted.

Finally, we perform an average over the helical phase, producing a version of the covariance that has isotropic bending as well as twist, stretch, and twist-stretch coupling covariances $\bar{C} = \frac{1}{2\pi} \int_0^{2\pi} \tilde{C}_\phi d\phi$, see Sec. III D.

V. CONFORMATIONAL FLUCTUATIONS AND EXTERNAL FORCES

A. Persistence lengths and numerical verification

The global offset g_0 and the combined covariance matrix $\hat{C}$ are constructed such that they capture the conformational statistics of an ensemble of thermally fluctuating, random sequence rigid base-pair chains, see Eq. (22). From the corresponding averaged covariance $\bar{C}$ one can read off the bend persistence length as $l_b = h_\parallel / \bar{C}^{11}$. The torsional modulus [49] normalized to units of length, we call the twist persistence length $l_t = h_\parallel / \bar{C}^{33}$ (see, e.g., Ref. [45]). Here, the on-axis helical rise $h_\parallel = \|p_{0\parallel}\|$. Since these persistence lengths include static variability, they are apparent persistence lengths in the terminology of Refs. [31,32]. For example, the square end-to-end distance, averaged over a random sequence ensemble $\langle \|p_{1m+1}\|^2 \rangle \propto 2 l_b l$ for long contour lengths $l = h_\parallel m \gg l_b$.

We tested the coarse-graining from RBC to WLC by performing a simple-sampling Monte Carlo (MC) simulation according to the algorithm in Sec. IV A. The raw, off-axis base-pair center end-to-end distances $\|p_{1m+1}\|$ were sampled. Their mean squares are plotted against the contour length in

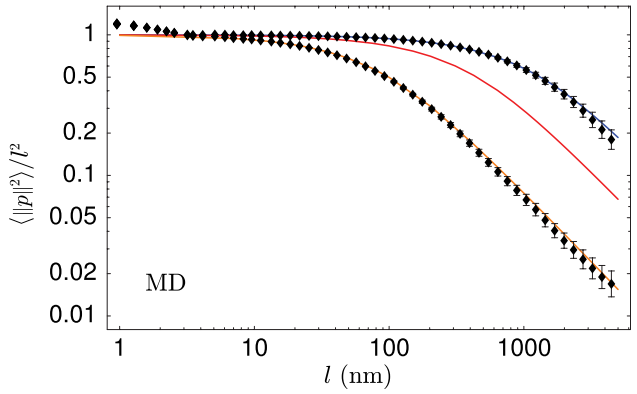

FIG. 4. (Color online) Comparison of an MC simulation of a random-sequence RBC to the coarse-grained effective WLC. Symbols designate the measured mean squared end-to-end distances for static disorder only (upper row) and for static plus thermal fluctuations (lower row). The theoretical curves assuming a WLC model are shown from top to bottom for static disorder (C_0 and C_1, blue), uncorrelated static disorder (C_0 only, red), and static plus thermal fluctuations ($\hat{C}$, orange), respectively. MD microscopic parameter set, as explained below.

Fig. 4. The theoretical curves $\langle \|p\|^2 \rangle = 2ll_b - 2l_b^2(1-e^{-l/l_b})$ for an inextensible WLC using the contour and bending persistence lengths l and l_b computed by coarse-graining, fit the simulation data to within numerical error. The only deviations occur below 3 nm, where the inextensible WLC model fails to reproduce the compressible and shearable helical RBC. In addition to the full covariance $\hat{C}$, simulations were also carried out for structural disorder only, setting all of the $C_{\mathrm{th}}(\sigma)=0$. The corresponding WLC using C_0 and the next-neighbor term C_1 [see Eq. (C1)] again fits the data.

Experiments that include a sequence ensemble average and thus measure apparent persistence lenghts include cryo-electron microscopy of frozen conformations of oligonucleotides [31], AFM tracing of adsorbed random-sequence DNA [50], and cyclization of random fragments [32]. Whenever such experiments are interpreted in terms of a intrinsically straight, homogeneous DNA, the apparent stiffness matrix extracted from experiment corresponds to the inverse of the total covariance $(\beta C)^{-1}$.

B. Stiffness

A slightly different situation arises in force-extension experiments carried out on single molecules (e.g., Ref. [23,24]). An external force tilts the elastic energy landscape of each step along the chain, introducing a bias towards those thermal fluctuations that lengthen the molecule. No such bias can be introduced on the sequence. Therefore the sequence randomness part of the total conformational covariance does not directly result in additional compliance to an external force.

What is the remaining effect of irregular sequence in micromanipulation experiments? We discuss this question in the weak static disorder limit, adapting the procedure of Ref. [51]. The basic idea is to expand the elastic Boltzmann factor $B \sim e^{-(\beta/2)(\xi-\langle\xi|\sigma\rangle)^{\mathsf{T}}[\langle S\rangle+\delta S(\sigma)](\xi-\langle\xi|\sigma\rangle)}$ for weak static disorder, and to interpret the result in terms of a homogeneous chain with renormalized stiffness. The calculation is carried out in Appendix D. The result is that the renormalized stiffness is the inverse total covariance $(\beta\hat{C})^{-1}$. That is, although the sequence disorder is quenched, the elastic response of the random chain to external forces is the same as if the sequence randomness were an additional elastic compliance. This result is valid for small sequence disorder in offsets and stiffness, which is a good approximation for DNA, and should hold for the entropic and enthalpic regimes of extension.

VI. DISCUSSION

A. Coarse-graining relations

We have derived all WLC elastic parameters starting from an arbitrarily oriented and offset homogeneous RBC. We now discuss in some detail how these coarse-grained parameters are related to the microscopic RBC parameters.

1. Equilibrium step

The transformation of the equilibrium step onto the helical axis Eq. (14) leaves the total rotation angle invariant. Therefore the equilibrium twist of $g_{0\|}$ is $\theta_\| = \|\omega_{0\|}\| = \|\omega_0\| \geqslant |\omega_0^3|$. That is, the twist per base pair of the WLC equals the total angle of rotation, not the Tw angle of the off-axis step. The equilibrium rise on axis is $h_\| = \|p_{0\|}\| = \omega_0^{\mathsf{T}} p_0 / \|\omega_0\|$ which is different from both off-axis quantities $\|p_0\|$ and p_0^3. These differences are of order $O(\omega_0^1+\omega_0^2)^2$ so they become important only when the equilibrium rotation axis ω_0 has significant roll and tilt with respect to the material frame, i.e., when the local helical parameters inclination and tip [33] are not negligible.

2. Fluctuations

Unlike the equilibrium step, the covariance matrix is changed not only by the rotation R_{ax} but also by the shift p_{ax} onto the average local helix axis. Intuitively, the on-axis frame g' is rigidly connected to g, see Figs. 1 and 2. Therefore, a rotational fluctuation of g with rotation vector ω' will result in an additional translational fluctuations of g' equal to $\omega' \times p_{\mathrm{ax}}$.

A familiar example of this geometrical effect is the stretching of an ordinary coil spring along its helix axis. In the wire material, this deformation corresponds mainly to torsion, i.e., a rotational deformation of consecutive wire segments. On a larger scale, the local torsion is levered into a translation of one coil end along the helix axis. The transformation (17) captures exactly this lever arm effect, which is proportional to the total axial displacement $\|p_{\mathrm{ax}}\|$ and so becomes relevant if the chain deviates from an idealized B-DNA form.

We calculate explicitly the 3×3 blocks $C_\|^{(ab)}$ of $C_\|$, (16), in terms of the corresponding blocks $C^{(ab)}$ of C, using (17) and (18). Here $a,b \in \{\omega,v\}$ stand for the set of rotational or translational components, respectively. Further, we let $C^{(ab)\prime} = R_{\mathrm{ax}}^{\mathsf{T}} C^{(ab)} R_{\mathrm{ax}}$ and $P'_{\mathrm{ax}} = R^i_{\mathrm{ax}j} p^j_{\mathrm{ax}} \epsilon_i$, an antisymmetric matrix. Using this notation,

$$C_{\|} = \begin{pmatrix} C^{(\omega\omega)\prime} & C^{(\omega\upsilon)\prime} + C^{(\omega\omega)\prime} P'_{\mathrm{ax}} \\ C^{(\upsilon\omega)\prime} - P'_{\mathrm{ax}} C^{(\omega\omega)\prime} & C^{(\upsilon\upsilon)\prime} - P'_{\mathrm{ax}} C^{(\omega\omega)\prime} P'_{\mathrm{ax}} + C^{(\upsilon\omega)\prime} P'_{\mathrm{ax}} - P'_{\mathrm{ax}} C^{(\omega\upsilon)\prime} \end{pmatrix}. \tag{28}$$

In this expression, the rotational block $C_{\|}^{(\omega\omega)}$ is merely a rotated version of the off-axis rotational block $C^{(\omega\omega)}$. In contrast, the translational block $C_{\|}^{(\upsilon\upsilon)}$ and the coupling block $C_{\|}^{(\omega\upsilon)}$ have "leverage terms," since rotational fluctuations about directions perpendicular to the offset vector contribute through a cross product with p_{ax}. For $C_{\|}^{(\upsilon\upsilon)}$, these involve the off-axis coupling $C^{(\upsilon\omega)}$ in first order and rotational fluctuations $C^{(\omega\omega)}$ in second order in $\|p_{\mathrm{ax}}\|$. The coupling block $C_{\|}^{(\omega\upsilon)}$ has contributions from $C^{(\omega\omega)}$ in first order. These leverage terms persist in the reduced WLC covariance matrix $\hat{C}$. They are the remainder of the microscopic description of fluctuations with respect to a material frame that is offset from the average helical axis.

Consider for example a base-pair step that exhibits x displacement but no inclination or tip, i.e., $p_{\mathrm{ax}} \propto d_1, \omega \propto d_3, R_{\mathrm{ax}} = I_3$. Then (28) implies that any coupled roll-rise (C^{26}) and roll (C^{22}) fluctuations will add to the stretching fluctuations $C_{\|}^{66}$ of the chain. In addition, the off-axis roll-twist fluctuation (C^{23}) contributes to twist-stretch coupling fluctuation on axis $C_{\|}^{36}$.

When inclination or tip are nonzero, then due to the additional rotation R_{ax} also shift and slide fluctuations contribute to the resulting WLC parameters. This mixing of local deformations makes it essential to transform to an on-axis frame before averaging over the shear degrees of freedom.

B. Comparison to experiment

1. Microscopic input

There are several different parameter sets available in the literature, extracted from analysis of x-ray crystal structures of DNA [14] and from molecular dynamics simulation [12,13]. For the stiffnesses obtained from structural data, the missing thermal energy scale is substituted by an "effective temperature." We here use the effective temperatures determined in a previous study [15] by equating the total, microscopic fluctuation strengths of the crystal and MD covariance matrices. The absolute magnitudes of all parameters derived from structural data (B for B-DNA crystal and P for protein•DNA cocrystals) therefore depend on our choice of effective temperature. Still, their relative magnitudes are properties of the microscopic structural data set independent of this choice. No such restrictions apply to the MD parameters, since here the temperature is set by the simulation. We also include a hybrid parametrization (MP) which combines the equilibrium values from the P•DNA dataset with the stiffness matrices from MD. This combined potential compared favorably to the others in binding affinity prediction [15]. It can be seen as a version of the MD potential which is corrected for the well known undertwist occurring in MD simulations. For MD and MP, our coarse-graining involves no free parameter. In the following, we give the full dinucleotide mesoscopic parameters only for the MP potential, while for random DNA, a comparison of the different potentials is made.

2. Conformational statistics of random DNA

In Table I we show the resulting effective WLC covariance parameters and geometry for random DNA. The values are relevant for experiments in which an ensemble average over sequence is implicitly performed, see Sec. V A.

For the crystal parameter sets, the equilibrium rise and twist are close to the commonly accepted values of 0.34 nm/ step and 10.5 BP/turn. The MD rise and twist are both low, a known effect for the force field used in that study [52]. The MD bending persistence length is smaller than the commonly accepted values at physiological conditions, around 48 nm [32]. The low equilibrium rise of the MD conformations accounts for half of this deviation. We remark that the elastic constants of the B and P parameter sets differ from the MD ones since the choice of effective temperature only fixes overall fluctuation strength, not relative stiffness of different modes.

For all parameters sets, the twist persistence length is similar to the bend persistence length, and smaller than the result of 58 nm extracted from cyclization data [32]. We remark that no rescaling by a different effective temperature can bring all crystal stiffness parameters into reasonable agreement with MD since the various deviations occur in opposite directions.

3. Thermal vs sequence randomness

Instead of the conformations of random DNA, we can consider thermal and sequence fluctuations separately. Table II shows the corresponding static and thermal persistence lengths [30]. In disagreement with the cryo-EM study [31]

TABLE I. Random sequence WLC geometry, persistence lengths, and conformational covariances for the considered RBC potentials.

	$\frac{2\pi}{\theta_{\|}}$	$h_{\|}$	l_b	l_t	$\bar{C}^{11}$	$\bar{C}^{33}$	$\bar{C}^{44}$	$\bar{C}^{34}$
B	10.1	0.334	27.1	15.2	12	22	0.79	0.67
P	10.5	0.334	43.4	35.7	7.7	9.4	0.86	0.85
MD	11.9	0.318	38.9	45.1	8.2	7	1.9	1.2
MP	10.5	0.334	42.8	47.8	7.8	7	1	0.55
units	1	nm	nm	nm	$\frac{\mathrm{rad}^2}{10^3}$	$\frac{\mathrm{rad}^2}{10^3}$	$\frac{\mathrm{nm}^2}{10^3}$	$\frac{\mathrm{nm\,rad}}{10^3}$

TABLE II. Thermal and static contributions to the apparent persistence length for different potentials. For comparison, the l' column shows the static persistence lengths when sequence continuity is disregarded.

	l_b	$l_{b,\mathrm{th}}$	$l_{b,0}$	$l'_{b,0}$	l_t	$l_{t,\mathrm{th}}$	$l_{t,0}$	$l'_{t,0}$
B	27.1	29.5	327	211	15.2	15.4	1260	88.3
P	43.4	45.3	1040	575	35.7	36.3	2430	172
MD	38.9	42	519	175	45.1	47.7	818	256
MP	42.8	44.6	1040	575	47.8	48.8	2340	172
units				nm				

we find that the static persistence lengths are much higher than the thermal ones, leading to a correction of only a few nm to the random DNA persistence lengths. This is in accordance with an analysis based on cyclization experiments [32]. Also, the static $l_{b,0}$ for the *P* parameter sets correctly reproduces the value found numerically in that study, using the same parameter set.

4. Stiffness of dinucleotide repeats

We can also extract the WLC stiffness parameters of any RBC with repetitive sequence. Here, stiffness and conformational covariance are just the inverse of each other. A detailed view of WLC geometry and stiffness for all dinucleotide repeats is given in Table III, for the MP parameter set. The number of BP steps per full turn $\frac{\pi}{\|\omega_{13}\|}$ and the contour length per BP step $\frac{1}{2}h_{\|13}$ vary by roughly 2%. Their respective values for the average step, obtained by averaging the equilibrium conformation and covariance, closely match commonly accepted values for *B*-DNA.

TABLE III. Comparison of WLC geometry and stiffness parameters of all six unique repetitive sequences of period two, for the MP hybrid parametrization. In the "av." row, the values for the average step is shown. A conversion to WLC units is appended, see text. MP parameter set.

	$\frac{\pi}{\|\omega_{13}\|}$	$\frac{1}{2}h_{\|\ 13}$	$\beta\bar{S}^{11}$	$\beta\bar{S}^{33}$	$\beta\bar{S}^{44}$	$\beta\bar{S}^{34}$	r_{resp}
AA	10.2	0.327	144	141	976	−38.3	0.59
AC	10.4	0.333	132	142	1140	−105	0.22
AG	10.5	0.334	139	159	1120	−103	0.25
AT	10.7	0.334	111	195	975	−80.1	0.39
GG	10.9	0.338	159	186	1090	−89.9	0.33
CG	10.3	0.338	124	126	831	−78.5	0.26
av.	10.5	0.334	134	153	1050	−87.9	0.28
units	BP	nm	rad^{-2}	rad^{-2}	nm^{-2}	$(\mathrm{rad\ nm})^{-1}$	nm
av.	10.5	0.334	45	51	1440	−120	
units	BP	nm	nm	nm	pN	$\frac{\mathrm{pN\ nm}}{\mathrm{rad}}$	

TABLE IV. Experimental stiffness parameters as given in the literature and average thermal stiffness (using the MP parameter set). The parameters B, C, g, S from Gore *et al.* [24] were multiplied by $\beta/h_{\|}$. The parameters B, C, D in Lionnet *et al.* [23] were multiplied by $\theta_{\|}^2/h_{\|}^3, 1/h_{\|}, \theta_{\|}/h_{\|}^2$, respectively. Beware of a missing 1/2 factor in the first formula of Ref. [23].

	$\beta\bar{S}^{11}$	$\beta\bar{S}^{33}$	$\beta\bar{S}^{44}$	$\beta\bar{S}^{34}$	r_{resp}
[24]	163±15	327±15	781±150	−64±15	0.5
[23]		294	710	−47±20	0.28
MP	128	149	1045	−82	0.29
units	rad^{-2}	rad^{-2}	nm^{-2}	$(\mathrm{nm\ rad})^{-1}$	$\frac{\mathrm{nm}}{\mathrm{turn}}$

The stiffness parameters are given in base-pair step units. Conversion to more commonly used WLC units is possible as follows: Multiplying βS^{11}, βS^{22}, S^{44}, and S^{34} by $\frac{1}{2}h_{\|13}$ gives, respectively, the bending persistence length in nm, the twist stiffness [53–56] in nm, the stretch modulus in pN and the twist-stretch coupling in pN nm. This is carried out for the average values in the last row.

Looking at the magnitudes, the poly-*AT* repeat stands out as the most bendable sequence which is at the same time torsionally rather stiff. Another common trend in our results is that poly-*G* DNA is comparatively stiff with respect to bending. The values are comparable to MD studies in which elastic constants of oligonucleotides were measured, with repeats *AA*, *AT*, *GC*, and *GG* [27] and with *AT* and *GC* [28]. There too, poly-*AT* is torsionally stiff but bendable. However, bending persistence lengths from Refs. [27,28] are up to two times bigger than either our or experimental values, which may be due to bending relaxation being too slow to be seen in that simulation [27]. The twisting persistence lengths in Refs. [27,28] are generally larger than our results by about a factor of 2, and show stronger sequence dependence, but with similar trends. The stretch modulus and the twist-stretch coupling depend on the sequence in a correlated way. Again comparing with Ref. [27], their stretch moduli agree qualitatively but show a different sequence dependence. Also, their twist-stretch coupling constants are *positive*, unlike our and recent single-molecule experimental results [24,23]. The rightmost column of Table III shows the ratio of elongation over overtwist in response to an external stretching force, $r_{\mathrm{resp}}=\bar{C}^{44}/(2\pi\bar{C}^{34})$ in nm/turn, as considered in Refs. [23,24].

5. Stiffness of random DNA

Recent single-molecule experiments at moderate applied tension have given new data on DNA stiffness [23,24]. We show the full elastic parameters collected in these articles in Table IV, together with the average stiffness of a random DNA computed from the MP parameter set, see Sec. V B. The bending modulus of 128 k_BT/rad^2 (i.e., $l_b=128$ BP) is lower than the result from Ref. [24] and still on the low end of the range of 132–138 k_BT/rad^2 found in previous [48–50]

TABLE V. Relative error in stiffness parameters made when using "naive" matrix elements instead of the coarse-grained parameters described above. Values are given in %. Average step BP parameters.

	e^{11}	e^{33}	e^{44}	e^{34}
MD	3	−13	59	50
MP	2	−7	−5	48

single-molecule experiments. (However, in Ref. [57] a lower experimental value is reported.)

The deviation in torsional rigidity is much more dramatic. Recent experimental values are about twice as high as our coarse-grained results, see also Ref. [16] for a review. This low twist rigidity is a feature of all parameter sets. For the crystal parameter sets one might argue this indicates that torsional deformations carry more elastic energy than bending deformations, thus "violating" an assumed equipartition of energy. However, for the MD parameter set, this is clearly not the case; the simulated DNA base-pair steps were indeed more twistable than experimental values for DNA suggest. A speculative explanation is that there may exist negative correlations between thermal twist deformations of neighboring steps which are neglected in our model, leading to an underestimation of twist stiffness. Negative twist-stretch coupling has been demonstrated in Refs. [23,24], a feature that is reproduced with good agreement by the microscopic data, and is also visible in the local Tw-Ri coupling of the microscopic parameter sets [29].

6. Effect of simplifications

Does our rather involved computation of macroscopic parameters actually make a noticeable difference? The calculations could be simplified in several ways.

One possibility is to leave away the correction for sequence continuity. This amounts to replacing the corrected structural covariance $C_0+C_\times$ by just C_0. The numerical error made in this approximation is listed in the l' columns of Table II. As can be seen, static variability is strongly overestimated (for twist, more than tenfold). The same error can also be seen in the middle line in Fig. 4, where the mean square displacement for pure uncorrelated static disorder is shown. Since the static fluctuations are only a small corrections to the dominating thermal fluctuations, the overall error made in assuming uncorrelated steps is not severe.

Another possible simplification is by disregarding the details of average helical geometry of the chain. Treating all base-pair steps as ideal *B*-DNA from the beginning as in Ref. [29], one would perform an average of the off-axis covariance matrix, over shift, slide, and helical phase angle. Inverting this, one obtains a "naive" stiffness matrix S_{na}. The relative error made in such a computation, $e^{ij}=(S^{ij}_{\mathrm{na}}-\bar{S}^{ij})/\bar{S}^{ij}$ is shown in Table V. While the bending and twisting stiffnesses are well approximated by the naive guess, the error in stretch modulus (for MD) and twist-stretch coupling (for MD, MP) is considerable. For these terms, leverage due to the axis offset becomes important (Sec. VI A). Especially the naive twist-stretch coupling is not negative enough. The effect is more pronounced with the pure MD parameter set [12,13], since it has unusual equilibrium conformations with bigger axis offset.

The procedure we describe involves no approximations regarding the geometry. This makes it directly applicable to alternative DNA structures, and indeed any polymer with average helical geometry, whenever microscopic covariance matrices are available. In fact, the more the average geometry deviates from idealized *B*-DNA, the greater is the need to treat the helical geometry correctly.

C. Limits of applicability of the WLC model

As a continuous model, the WLC is defined down to arbitrarily small length scales. However, the microscopic structure of DNA suggests that there must be a lower limit to its applicability. Indeed, recent experimental studies [50,58] have highlighted examples of strong bending on short scales, in disagreement with standard WLC elasticity. Reversing gears from the previous sections, we investigate at which length scale an isotropic, homogeneous WLC fails to reproduce the behavior of a random RBC. We start with the isotropy of bending in all directions, which is not a local symmetry of the RBC and comes about only after averaging over several turns. Another feature of the continuous WLC is the Gaussian bend angle distribution of short segments, which again is not representative of a random sequence RBC. Finally, the effective WLC of a random sequence RBC has homogeneous stiffness constants. If this were the whole truth, there would be no indirect readout. We quantify the deviations from this average behavior due to sequence fluctuations on short scales.

1. Anisotropic bending

A feature of short compound steps not captured by the coarse-grained WLC limit is their anisotropic bending stiffness. On scales much longer than a full turn, bending will be isotropic. Using the compound covariance $\tilde{C}$ [see Eq. (11)] we can quantify the decay of anisotropy.

The ratio of the two principal bending stiffnesses as a function of chain length is shown in Fig. 5. The oscillatory decay results from orientational averaging over fractional turns of the helix. Since linear response is always symmetric, the bending anisotropy has minima every *half* turn of the double helix. For exactly two full turns (21 BP), anisotropy is suppressed completely, but already a 5 BP compound step at almost a half turn is essentially isotropic. This behavior agrees nicely with that of the two principal bending stiffnesses measured in Ref. [27] for oligonucleotides of increasing length. Their stiffnesses are equal at around 6 BP, in line with the fact that MD potential produces a 12 BP/turn helix structure.

2. Bend angle distributions for short chains

The combined covariance matrix $\tilde{C}_{1m+1}$ gives the second moment of the distribution $p(\eta_{1m+1})$ of deformations, observed in a thermal ensemble of random sequence, length m

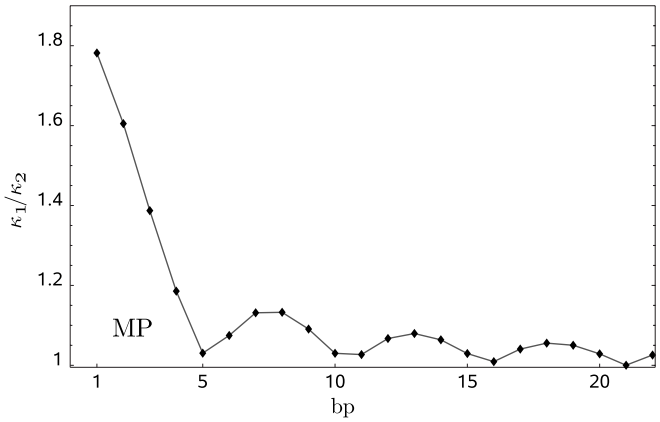

FIG. 5. Bending anisotropy. The ratio of larger over smaller bending stiffness decays in an oscillating fashion with compound step length. MP parameter set, average step geometry.

compound steps. Here it is not necessary that the single step deformation distributions have a Gaussian shape. Indeed such an assumption depends on the choice of coordinates.

Nevertheless, let us for the moment additionally assume that for each sequence, the single step thermal deformation distributions were in fact Gaussians. The deformation of a specific compound step with sequence σ_{1m+1} then again follows a Gaussian distribution $p(\eta_{1m+1}|\sigma_{1m+1})$, since for the small deformation angles we consider, it is the result of a convolution of the single step covariances.

Sequence randomness changes this picture. The deformation distribution of an ensemble of random compound steps $p(\eta_{1m+1})=\langle p(\eta_{1m+1}|\sigma_{1m+1})\rangle$ is a sequence average of several Gaussians with different offsets and widths and thus in general will deviate from a Gaussian shape.

In a recent AFM study of DNA adsorbed to a coverslip [50], bend angle distributions of DNA over short lengths have been found to favor large bend angles much more than expected from the WLC model. It is interesting to ask whether this can be explained as an effect purely of sequence randomness as outlined above. We show in Fig. 6 the effective potential U_{eff} for the total bend angle $\vartheta=[(\eta^1_{1,m+1})^2+(\eta^2_{1,m+1})^2]^{1/2}$ of random sequence compound steps of different lengths m. It was extracted from histograms of a simulation as described in Sec. IV A. For compound steps shorter that 5 BP, the effective potentials stay well below the respective harmonic potentials that correspond to an isotropic WLC model with the coarse-grained random DNA persistence length of $l_b=42.8$ nm. This is the combined result of the spread in bending stiffness resulting from sequence randomness and from anisotropic bending. However, above 5 BP the observed deviations are negligible and thus insufficient to explain the broader-than-Gaussian bend angle distributions observed in [50] for DNA as long as 15 BP.

3. Short-scale stiffness variability

We quantify the breakdown of the assumption of sequence-independent WLC stiffness for short random chains. The thermal covariance matrix $\tilde{C}_{\text{th}}(\sigma_{1m+1})$ (11) of a compound step with fixed sequence σ_{1m+1} was calculated in

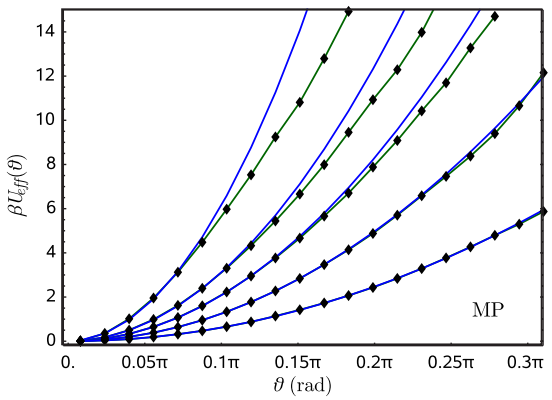

FIG. 6. (Color online) Effective potential for the total bend angle ϑ (curve with symbols, green). The curves without symbols (blue) show the harmonic approximation to the effective potential that results of a fine-graining of an isotropic WLC with the corresponding coarse-grained persistence length. Compound step length, from left to right: 1,2,3,5,10 BP. MP parameter set.

Sec. II D. While the mean thermal covariance matrix $M=\langle\tilde{C}_{\text{th}}(\sigma_{1m+1})\rangle$ is just the sequence average, the covariances of the 4×4 matrix entries are given by

$$V^{ijkl}_{1m+1}=\langle[\tilde{C}^{ij}(\sigma_{1m+1})-M^{ij}][\tilde{C}^{kl}(\sigma_{1m+1})-M^{kl}]\rangle. \quad (29)$$

In a lengthy but straightforward calculation (not shown), this expectation can be evaluated in terms of single-step and nearest-neighbor static covariances of the matrix entries. Again, including the nearest neighor cross covariances takes sequence continuity into account. For example, the fact that it is impossible to combine two of the comparatively soft pyrimidine-purine [14] steps in a row, reduces the variability of the average stiffness across random sequence compound steps.

From the result, we can characterize stiffness variability; the relative spread of angular stiffness coefficients of compound steps over all sequences is shown in Fig. 7. Explicitly,

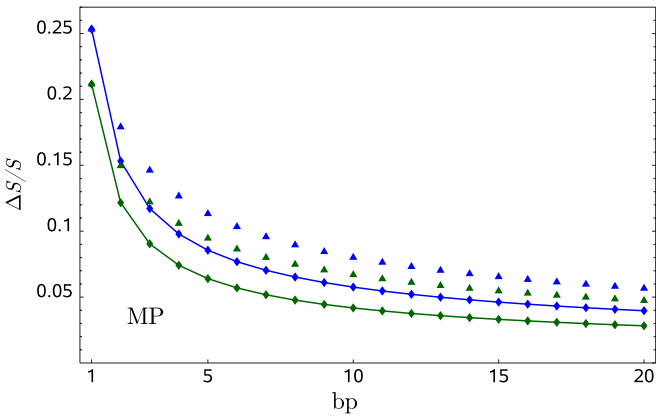

FIG. 7. (Color online) Relative spread $\Delta S/S$ of the bend (lower curve with diamonds, green) and twist (upper curve with diamonds, blue) stiffness coefficients vs compound step length. Ignoring sequence continuity leads to overestimation of the stiffness variability (bend, lower green triangles; twist, upper blue triangles).

$\Delta S/S=(V^{iiii}_{1m+1})^{1/2}/M^{ii}$, where $S=S^{ii}$ and $i=1,3$. After one full turn, variability in stiffness is down to 5%. The effect of sequence continuity is to reduce the variability compared to a model with independent step sequences, analogous to Table II.

D. Limitations of the RBC model

The main physical assumptions we have made come about by adopting the RBC model as starting point of our considerations. As a consequence, we treat thermal deformation fluctuations of neighboring steps as independent and also disregard internal deformations of a base pair such as propeller twist or buckle. These assumptions are not rigorously justified, but reasonable in view of the difficulties in obtaining reliable parameter sets even for the simpler RBC model.

Nevertheless our framework can be extended to improve on both of these points. Nearest-neighbor correlations in base-pair parameters may be included by extending the model to a full Markov chain. Internal deformations could be added by extending the configuration space, leading to a birod [59] in the continuum limit. However, for either of these interesting generalizations, a microscopic parametrization is an open challenge in itself. The fact that dinucleotide step stiffness depends overall rather weakly on the flanking sequence [36] and the encouraging agreement with mesoscopic data we found, suggest that the main features of coarse-grained DNA elasticity are captured already by our more basic model. Still, the low twist rigidity we found might be a result of missing negative twist correlations.

VII. SUMMARY

This work deals with a simple question: What is the long-wavelength limit of the rigid base-pair chain model for the local, sequence dependent structure and elasticity of DNA? In the following, we summarize the technical aspects of our work and the results of the comparison of microscopic potentials with direct mesoscopic experimental observations.

For coarse-graining the RBC model, the most intuitive starting point is to describe the irregular helical conformation of DNA in solution by a curved centerline. The fitting algorithm [41] defining its shape however inevitably introduces an arbitrary weighting factor and complicates a statistical analysis of thermally fluctuating conformations, see Sec. III A. We have tried to circumvent these problems by choosing a different approach in which irregular helix axes need not be considered. In essence, we reduce the general, irregular case to an ideal *B*-DNA geometry, by defining "on-axis versions" of the RBC model. For homogeneous ideal *B*-form DNA, it becomes trivial to integrate out the two lateral shear degrees of freedom of the RBC, to obtain all four mesoscopic elastic constants.

The first problem one encounters in this approach is the general helical geometry of DNA: Neither are the material frames oriented along the helical axis, nor are the base pairs centered on the axis. The consequences of this arrangement for the mesoscopic elastic properties are familiar from ordinary coil springs; e.g., a spring is much more extensible than the wire which it is made of (see also Sec. VI A).

A second complication arises in the case of heterogeneous DNA sequences which introduce small, intrinsic deformations as well as variable stiffness. In Sec. IV we have shown how to include the intrinsic variability of random DNA along the same lines as thermal fluctuations of an intrinsically straight helix. We have also incorporated a correlation effect inherent in the requirement of sequence continuity. Our theory reproduces the numerical and experimental observation that structural disorder contributes less than 10% of the total conformational fluctuations. Finally, we have extended our formalism (following Ref. [51]) to the case of an applied external force. We find that the covariance (or inverse stiffness) of the effective WLC is a sum of structural and thermal covariances, Eq. (26). This result generalizes the well-known inverse additivity of static and thermal bending persistence lengths.

Our results allow a direct comparison of the various existing microscopic potentials to AFM imaging, cyclization and single-molecule stretching and twisting experiments, with no free parameter, Sec. VI B. Given the unclear conceptual basis for extracting microscopic parameters from crystal structures and the long-term stability problems of the force fields employed in MD simulations, the overall agreement is striking. The microscopic bending persistence lengths match to within 5%; the predicted twist persistence is about 50% lower, and the magnitudes of compressional modulus and twist-stretch coupling are roughly 50% higher than the mesoscopic experimental values. Notably, our results also reproduce the order of magnitude and the negative sign of twist-stretch coupling.

We have also quantified the variability of WLC parameters across all possible dinucleotide repeats, Sec. VI B 4, finding pronounced variability especially in the twist-stretch coupling stiffness. Quite interestingly, the homogeneous WLC description seems to be applicable down to length scales corresponding to one or two turns of the double helix, Sec. VI C. On shorter scales, there occur a noticeable anisotropy in the bending rigidity and sequence dependent variability in the elastic constants. We have also shown that the bend angle distributions of a random ensemble has considerably bigger tails than the linear elastic assumption of a Gaussian.

VIII. CONCLUSIONS

In this article we have shown how to coarse-grain sequence dependent rigid base-pair chain models of DNA structure and elasticity to the wormlike chain level. In particular, we accounted for the helical geometry and the sequence disorder in random DNA, and we have discussed the range of validity of the coarse-grained model. Our results make it possible to quantitatively connect experiments (and simulations) on microscopic and mesoscopic length scales. While making this connection is conceptually important, we do not claim advantages in extracting the mesoscopic elastic constants from microscopic experiments. However, in a biological context there is certainly more to DNA than the

wormlike chain model can describe, and the need for a reliable model of short-scale, sequence dependent elasticity is evident. Here, the comparison of the predicted to the observed mesoscopic behavior constitutes an essential test of any microscopic parameter set, especially considering an experimental precision approaching one percent for the mesoscopic bending rigidity [32]. While more work is certainly needed, the present results add to the credibility of the available RBC parametrizations.

ACKNOWLEDGMENTS

N.B.B. thanks B. Lindner for helpful discussions. R.E. acknowledges support from the chair of excellence program of the Agence Nationale de la Recherche (ANR).

APPENDIX A: COORDINATE CONVERSION

How does one obtain the covariance matrix C and equilibrium conformations g_0 for a given collection $\{g_k\}_{1\leqslant k\leqslant N}$ of BP frame conformations? We can first determine g_0 by requiring that $\{g_0^{-1}g_k\}_k$ has mean 0 in exponential coordinates. For not too wide distributions, such a center always exists and is unique [39]. Then, $C^{ij}=\langle \xi^i \xi^j\rangle$ is the standard covariance matrix of $\{g_0^{-1}g_k\}_k$ in exponential coordinates.

However, for the potential parametrizations considered here, only the equilibrium values ζ_0 and covariance matrices $C_\zeta^{ij}=\langle(\zeta-\zeta_0)^i(\zeta-\zeta_0)^j\rangle$ with respect to the global coordinates $\zeta=(\Omega,\tau,\rho,q^1,q^2,q^3)$ as defined in Ref. [43] and used in Ref. [42], are given. Here, $\theta=(\Omega,\tau,\rho)$ are twist, tilt, and roll angles but differ from our choice of angles and the $q=(q^1,q^2,q^3)$ gives the translation vector with respect to the midframe R_m. The conversion formulas are

$$R(\zeta)=\exp\{[\Omega/2-\arctan(\tau/\rho)]\epsilon_3\}\exp(\sqrt{\rho^2+\tau^2}\epsilon_2)\times\exp\{[\Omega/2+\arctan(\tau/\rho)]\epsilon_3\},$$

$$R_m(\zeta)=\exp\{[\Omega/2-\arctan(\tau/\rho)]\epsilon_3\}\exp(\sqrt{\rho^2+\tau^2}/2\epsilon_2)\times\exp\{[\arctan(\tau/\rho)]\epsilon_3\},\quad \text{and } p(\zeta)=R_m(\zeta)q, \tag{A1}$$

together determining the frame conformation $g(\zeta)$. We checked that the variation of the volume element in the region of noticeable probability around g_0 is small compared to the variations in the probability density. Therefore neglecting the former, we get $g_0=g(\zeta_0)$. In linear order around the equilibrium position, we can then transform the covariance matrix C_ζ given in ζ coordinates to exponential coordinates using just the Jacobian matrix J_0 of the coordinate transition map $\zeta\mapsto\xi(\zeta)=\log[g(\zeta)]$. This gives $C=J_0C_\zeta J_0^\mathsf{T}$. We have calculated $J_0=\frac{\partial\xi}{\partial\zeta}|_{\zeta_0}$ analytically. Its 3×3 blocks are

$$\frac{\partial\omega^i}{\partial\theta^j}=1/2\ \mathrm{tr}(\epsilon_i R^\mathsf{T}\partial_{\theta^j}R),$$

$$\frac{\partial\omega^i}{\partial q^j}=0,$$

$$\frac{\partial v^i}{\partial\theta^j}=(R^\mathsf{T}\partial_{\theta^j}R_{\mathrm{mid}}q)^i,$$

$$\frac{\partial(v)}{\partial(q)}=R^\mathsf{T}R_{\mathrm{mid}}. \tag{A2}$$

All coarse graining calculations presented in this article use the matrices C converted in this way as a starting point.

The exponential coordinates of the equilibrium conformations have the usual symmetries under strand change and reading direction reversal: Denote by $\bar\sigma$ the sequence complementary to σ, e.g., $\overline{AG}=CT$, and let $E=\mathrm{diag}(-1,1,1,-1,1,1)$. Then as $\sigma\to\bar\sigma$, $\xi_0=(\mathrm{Ti}_0,\mathrm{Ro}_0,\mathrm{Tw}_0,\mathrm{Sh}_0,\mathrm{Sl}_0,\mathrm{Ri}_0)\to E\xi_0$. Due to the ξ_0 dependent coordinate conversion above, the body-frame covariance matrix does *not* obey the corresponding symmetries, $C\nmapsto ECE$. While this may seem a serious drawback of the coordinate system we use here, it turns out that in the on-axis, shear and helical phase averaged covariance matrices, the strand-exchange symmetry is reestablished. Therefore, our coarse-grained results are indeed independent of the reading sense.

APPENDIX B: VOLUME ELEMENT

In our coordinates, $\ln A(\xi)=-\frac{1}{6}\|\omega\|^2+O(\|\omega\|^4)$, so that in a Gaussian approximation

$$p(\xi)dV_\xi\propto e^{-(1/2)\xi^i(\beta S_{\sigma ij}+\bar A_{ij})\xi^j}d^6\xi,\quad \bar A=\begin{pmatrix}\frac{1}{3}I_3 & 0_3\\ 0_3 & 0_3\end{pmatrix}. \tag{B1}$$

Here, I_3 and 0_3 are the 3×3 identity and zero matrices, respectively. In DNA, the distributions $p(\xi)$ of single steps are very narrow. Therefore when computing moments, in particular, the covariance matrix $C^{ij}=\langle\xi^i\xi^j\rangle$, we can extend the integration boundaries to infinity with negligible error. Performing the integral we then get the relation $\beta S+\bar A=C^{-1}$. Since $\beta S\gg\bar A$, in making the approximation $\beta S=C^{-1}$, we introduce an error of less than 1% for typical B-DNA steps. That is, the stiffness matrix βS is indeed given by the inverse of the covariance.

APPENDIX C: CORRELATIONS INDUCED BY SEQUENCE

Consider the combined fluctuations of a short RBC consisting of m BP steps. By the assumption of independent bases, the joint pdf of sequence steps along the chain is the product of base PDFs, $p(\sigma_{12},\ldots,\sigma_{mm+1})=\Pi_{k=1}^{m+1}p(b_k)$. This implies that correlations between static offsets extend up to nearest neighbor steps. We write

$$\langle\langle \xi^i_{kk+1}|b_k b_{k+1}\rangle\langle \xi^j_{ll+1}|b_l b_{l+1}\rangle\rangle = \begin{cases} C_0^{ij}, & l=k, \\ C_1^{ij}, & l=k+1, \\ C_1^{ji}, & l=k-1, \\ 0, & \text{otherwise}. \end{cases} \tag{C1}$$

Here, the covariance of static offsets C_0 and the new nearest-neighbor term C_1 are defined by the left-hand side. They can be computed when $p(\sigma)$ is known.

There are no correlations in the model between thermal fluctuations of nearest neighbors; also analogous to the discussion after Eq. (25)

$$\langle\langle \xi^i_{l-1l}|b_{l-1}b_l\rangle(\xi^j_{ll+1} - \langle \xi^j_{ll+1}|b_l b_{l+1}\rangle)\rangle \tag{C2}$$

can be seen to vanish by taking the conditional expectation $\langle\cdots|b_l b_{l+1}\rangle$ first. As a result, static offsets are not correlated with thermal fluctuations even though the thermal PDF does depend on sequence.

We can now apply the procedure given in Sec. II D for combining the m base pairs of the chain steps into a compound step. The compound deformation

$$\xi_{1m+1} = \sum_{k=1}^{m} \mathrm{Ad}\, g_0^{k-m} \xi_{kk+1}. \tag{C3}$$

Considering the static part first, using Eq. (C1), we are left with a sum of appropriately transformed single-step covariances C_0 and in addition a sum of nearest neighbor cross-terms involving C_1. The covariance of static offsets $\langle\langle \xi^i_{1m+1}|\sigma_{1m+1}\rangle\langle \xi^j_{1m+1}|\sigma_{1m+1}\rangle\rangle$ is

$$C_{0;1m+1} = \sum_{l=0}^{m-1} \mathrm{Ad}\, g_0^{-l} C_0 \mathrm{Ad}^{\mathsf{T}}\, g_0^{-l} + \sum_{l=0}^{m-2} \mathrm{Ad}\, g_0^{-l} C_\times \mathrm{Ad}^{\mathsf{T}}\, g_0^{-l},$$

$$\text{where } C_\times = C_1 \mathrm{Ad}^{\mathsf{T}}\, g_0^{-1} + \mathrm{Ad}\, g_0^{-1} C_1^{\mathsf{T}}. \tag{C4}$$

Note that two neighboring compound steps are still correlated by sequence continuity at their interface. From Eq. (C4) we have the recursion relation

$$C_{0;1l+1} = \mathrm{Ad}\, g_0^{-1} C_{0;1l} \mathrm{Ad}^{\mathsf{T}}\, g_0^{-1} + C_0 + C_\times. \tag{C5}$$

The same relation is obeyed by a sequence of independent steps with static covariance matrix $C_0+C_\times$. We conclude that except for a boundary term $C_\times$ at the beginning of the chain, a RBC with independent static offsets and with covariance $C_0+C_\times$ exhibits the same effective static covariance as the original short range correlated chain. The relative error in effective compound covariance is of order $1/m$. We will neglect this error in the following.

Finally, we can combine the static and thermal randomness by summing their covariances, since they are uncorrelated. The total covariance matrix $C^{ij}_{1m+1} = \langle \xi^i_{1m+1} \xi^j_{1m+1}\rangle$ of the compound deformation is

$$C_{1m+1} = C_{0;1m+1} + C_{\mathrm{th};1m+1} = \sum_{l=0}^{m-1} \mathrm{Ad}\, g_0^{-l} \hat{C} \mathrm{Ad}^{\mathsf{T}}\, g_0^{-l}, \tag{C6}$$

where $\hat{C} = (C_0 + C_\times) + \langle C_{\mathrm{th}}(\sigma)\rangle$. In summary, we have described the conformational statistics of a compound step including sequence randomness by an effective, stepwise independent RBC whose covariance $\hat{C}$ is the sum of static and sequence parts and incorporates the additional requirement of sequence continuity.

APPENDIX D: RANDOM RBC RESPONSE

The expectation value of an observable $f(g_{1m})$, e.g., the z extension p^3_{1m}, for a fixed sequence σ_{1m}, is given by the multiple integral

$$\langle f|\sigma_{1m}\rangle_\epsilon = \frac{1}{\mathcal{Z}} \int \left(\prod_{k=1}^{m-1} dV_{\xi_{kk+1}} \right) f(g_{1m}) B_\epsilon e^{-\beta U(g_{1m})},$$

$$B_\epsilon = e^{-(\beta/2)\sum_{k=1}^{n-1} (\xi_{kk+1} - \epsilon\langle\xi|\sigma_{kk+1}\rangle)^{\mathsf{T}} S (\xi_{kk+1} - \epsilon\langle\xi|\sigma_{kk+1}\rangle)}. \tag{D1}$$

In this expression, $\mathcal{Z}$ is the partition sum and $U(g_{1m})$ is an external potential, e.g., $U=Fp^3_{1m}$ for linear stretching with a force F. For a start, the elastic Boltzmann weight B_ϵ, has sequence dependent offsets but constant stiffness matrix S. The auxiliary parameter ϵ is introduced here to keep track of orders in the following weak static disorder expansion:

$$\frac{B_\epsilon}{B_0} = 1 + \epsilon \sum_{k=1}^{m-1} \xi^{\mathsf{T}}_{kk+1} \beta S \langle\xi|\sigma_{kk+1}\rangle + \frac{\epsilon^2}{2} \sum_{k=1}^{m-1} -\langle\xi|\sigma_{kk+1}\rangle^{\mathsf{T}} \times \beta S \langle\xi|\sigma_{kk+1}\rangle + \frac{\epsilon^2}{2} \left(\sum_{k=1}^{m-1} \xi^{\mathsf{T}}_{kk+1} \beta S \langle\xi|\sigma_{kk+1}\rangle \right)^2 + O(\epsilon^3). \tag{D2}$$

We proceed to calculate the global expectation value $\langle f\rangle_\epsilon = \Sigma_{b_1 \ldots b_m} p(\sigma_{1m}) \langle f|\sigma_{1m}\rangle_\epsilon$, in a random sequence ensemble. Using Eqs. (D2) and (D1), and interchanging summation and integration, the result is

$$\langle f\rangle_\epsilon = \left\langle f \left[1 + \epsilon^2 \beta^2 \left(\frac{1}{2} \sum_{k=1}^{m-1} \xi^{\mathsf{T}}_{kk+1} S C_0 S \xi_{kk+1} + \sum_{k=2}^{m-1} \xi^{\mathsf{T}}_{k-1k} S C_1 S \xi_{kk+1} \right) \right] \right\rangle_{\epsilon=0}. \tag{D3}$$

As can be seen, the first order term from Eq. (D2) vanishes in the sequence average. The first quadratic term produces a global constant relevant only for normalization which was discarded. Finally, the second quadratic term in Eq. (D2) results in expressions involving the static covariance C_0 and nearest-neighbor covariance C_1 (see Appendix C), in Eq. (D3). The square bracket term can now be interpreted as the truncated expansion of an exponential. Then Eq. (D3) is, to second order, identical to an expectation value taken without static disorder but with renormalized elastic energy [46]

$$\langle f\rangle_\epsilon = \frac{1}{\tilde{\mathcal{Z}}}\int\left(\prod_{k=1}^{m-1} dV_{\xi_{kk+1}}\right) f e^{-\beta U} \times e^{-\beta/2[\Sigma_{k=1}^{m-1}\xi_{kk+1}^{\mathsf{T}}(S-\epsilon^2\beta S C_0 S)\xi_{kk+1}-2\Sigma_{k=2}^{m-1}\xi_{k-1k}^{\mathsf{T}}\epsilon^2\beta S C_1 S\xi_{kk+1}]}. \tag{D4}$$

We conclude that under arbitrary external forces, the random sequence RBC ensemble responds in the same way as a homogeneous RBC with renormalized stiffness. It is an exercise in Gaussian integrals to verify that in the special case $U=0$, the renormalized elastic energy in Eq. (D4) produces the covariances $\langle \xi_{kk+1}^i \xi_{kk+1}^j\rangle_{\epsilon=1} = (\beta S)^{-1ij} + C_0^{ij}$ and $\langle \xi_{k-1k}^i \xi_{kk+1}^j\rangle_{\epsilon=1} = C_1^{ij}$, to second order in ϵ. This matches nicely with the sum of static covariance [Eq. (C1)] and thermal covariance $\langle C_{\mathrm{th}}\rangle$ of the free chain, in agreement with the result in Sec. IV D. Following the discussion in Appendix C, the chain is equivalent to a homogeneous, independent-step RBC with stiffness matrix $(\beta\hat{C})^{-1}$ and global equilibrium step g_0.

What changes if the stiffness also depends on sequence? We split up the thermal covariance matrix (23) into its average and sequence-dependent parts $C_{\mathrm{th}}(\sigma) = \langle C_{\mathrm{th}}(\sigma)\rangle + \delta C_{\mathrm{th}}(\sigma)$. Since C scales as $(\xi)^2$, it is natural to assign an order of ϵ^2 to the term $\delta C_{\mathrm{th}}(\sigma)$, so that the variations in width of the distribution are of order ϵ. We then replace

$$\beta S \to [\langle C_{\mathrm{th}}\rangle + \epsilon^2 \delta C_{\mathrm{th}}(\sigma_{kk+1})]^{-1} = S_{\mathrm{th}} - \epsilon^2 \beta^2 S_{\mathrm{th}} \delta C_{\mathrm{th}} S_{\mathrm{th}}, \tag{D5}$$

in Eq. (D1), where $\beta S_{\mathrm{th}} = \langle C_{\mathrm{th}}\rangle^{-1}$. Repeating the expansion of B_ϵ as before, all occurrences of S in Eq. (D2) are replaced by S_{th}. The only extra term $-\epsilon^2\beta\Sigma\,\xi_{kk+1}^{\mathsf{T}} S_{\mathrm{th}}\delta C_{\mathrm{th}}(\sigma_{kk+1}) S_{\mathrm{th}}\xi_{kk+1}$ drops out in the sequence average (D3). Thus, sequence dependent stiffness drops out to second order [60].

In summary, to second order in ϵ, a random RBC ensemble with sequence disorder in offsets and stiffness, produces the same response to external forces or torques as a homogeneous chain with a renormalized elastic energy. In particular, at zero applied force, the effective covariance $\hat{C}$ is recovered.

[1] J. Widom, Q. Rev. Biophys. **34**, 269 (2001).
[2] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. Moore, J. Wang, and J. Widom, Nature (London) **442**, 772 (2006).
[3] G. Koudelka, S. Harrison, and M. Ptashne, Nature (London) **326**, 886 (1987).
[4] C. Hines, C. Meghoo, S. Shetty, M. Biburger, M. Brenowitz, and R. Hegde, J. Mol. Biol. **276**, 809 (1998).
[5] R. Hegde, Annu. Rev. Biophys. Biomol. Struct. **31**, 343 (2002).
[6] C. Prevost, S. Louisemay, G. Ravishanker, R. Lavery, and D. L. Beveridge, Biopolymers **33**, 335 (1993).
[7] R. Schleif, Proc. Natl. Acad. Sci. U.S.A. **69**, 3479 (1972).
[8] R. Schleif, Annu. Rev. Biochem. **61**, 199 (1992).
[9] K. Rippe, Trends Biochem. Sci. **26**, 733 (2001).
[10] C. Calladine and H. Drew, J. Mol. Biol. **178**, 773 (1984).
[11] B. D. Coleman, W. K. Olson, and D. Swigon, J. Chem. Phys. **118**, 7127 (2003).
[12] F. Lankaš, J. Šponer, J. Langowski, and T. Cheatham, 3rd, Biophys. J. **85**, 2872 (2003).
[13] F. Lankaš, in *Computational Studies of DNA and RNA*, edited by J. Šponer and F. Lankaš, Vol. 2 of *Challenges and Advances in Computational Chemistry and Physics* (Springer, Berlin, 2006).
[14] W. Olson, A. Gorin, X. Lu, L. Hock, and V. Zhurkin, Proc. Natl. Acad. Sci. U.S.A. **95**, 11163 (1998).
[15] N. Becker, L. Wolff, and R. Everaers, Nucleic Acids Res. **34**, 5638 (2006).
[16] G. Charvin, J. Allemand, T. Strick, D. Bensimon, and V. Croquette, Contemp. Phys. **45**, 383 (2004).
[17] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, Science **265**, 1599 (1994).
[18] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, Science **271**, 1835 (1996).
[19] J. F. Marko and E. D. Siggia, Biophys. J. **73**, 2173 (1997).
[20] R. D. Kamien, T. C. Lubensky, P. Nelson, and C. S. OHern, Europhys. Lett. **38**, 237 (1997).
[21] J. Moroz and P. Nelson, Proc. Natl. Acad. Sci. U.S.A. **94**, 14418 (1997).
[22] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. Viovy, D. Chatenay, and F. Caron, Science **271**, 792 (1996).
[23] T. Lionnet, S. Joubaud, R. Lavery, D. Bensimon, and V. Croquette, Phys. Rev. Lett. **96**, 178102 (2006).
[24] J. Gore, Z. Bryant, M. Nollmann, M. U. Le, N. R. Cozzarelli, and C. Bustamante, Nature (London) **442**, 836 (2006).
[25] A. Matsumoto and N. Go, J. Chem. Phys. **110**, 11070 (1999).
[26] A. Matsumoto and W. Olson, Biophys. J. **83**, 22 (2002).
[27] F. Lankaš, J. Šponer, P. Hobza, and J. Langowski, J. Mol. Biol. **299**, 695 (2000).
[28] A. Mazur, Biophys. J. **91**, 4507 (2006).
[29] T. Lionnet and F. Lankas, Biophys. J. **92**, L30 (2007).
[30] E. N. Trifonov, R. K. Z. Tan, and S. C. Harvey, in *Structure & Expression*, edited by W. K. Olson, M. H. Sarma, R. H. Sarma, and M. S. Sundaralinjam (Adenine, Schenectaky 1988), pp. 243–254.
[31] J. Bednar, P. Furrer, V. Katritch, A. Stasiak, J. Dubochet, and A. Stasiak, J. Mol. Biol. **254**, 579 (1995).
[32] M. Vologodskaia and A. Vologodskii, J. Mol. Biol. **317**, 205 (2002).
[33] R. Dickerson, Nucleic Acids Res. **17**, 1797 (1989).
[34] W. K. Olson *et al.*, J. Mol. Biol. **313**, 229 (2001).
[35] G. S.Chirikjian and Y. F. Wang, Phys. Rev. E **62**, 880 (2000).
[36] M. Arauzo-Bravo, S. Fujii, H. Kono, S. Ahmad, and A. Sarai, J. Am. Chem. Soc. **127**, 16074 (2005).
[37] S. Dixit *et al.*, Biophys. J. **89**, 3721 (2005).
[38] We conventionally always sum over all upper-lower index pairs.
[39] W. Kendall, Proc. London Math. Soc. **61**, 371 (1990).

[40] O. Gonzalez and J. Maddocks, Theor. Chem. Acc. **106**, 76 (2001).
[41] R. Lavery and H. Sklenar, J. Biomol. Struct. Dyn. **6**, 655 (1989).
[42] X. J. Lu and W. K. Olson, Nucleic Acids Res. **31**, 5108 (2003).
[43] X. Lu, M. El Hassan, and C. Hunter, J. Mol. Biol. **273**, 681 (1997).
[44] M. S. Babcock and W. K. Olson, J. Mol. Biol. **237**, 98 (1994).
[45] J. F. Marko and E. D. Siggia, Macromolecules **27**, 981 (1994).
[46] This is the modulus for unconstrained stretching degree of freedom.
[47] In our examples, we choose a flat distribution, although a sequence bias can be included.
[48] The conditional expectation of some function f is defined with respect to the conditional distribution $\langle f|\sigma\rangle = \int f(g)p(g|\sigma)dg$.
[49] For unconstrained stretching
[50] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, Nat. Nanotechnol. **1**, 137 (2006).
[51] P. Nelson, Phys. Rev. Lett. **80**, 5810 (1998).
[52] D. Beveridge *et al.*, Biophys. J. **87**, 3799 (2004).
[53] Note that due to twist-stretch coupling, only $(\bar{C}^{22})^{-1}$ has a physical interpretation as a correlation length.
[54] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, Biophys. J. **72**, 1335 (1997).
[55] C. Baumann, S. Smith, V. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).
[56] J. Wenner, M. Williams, I. Rouzina, and V. Bloomfield, Biophys. J. **82**, 3160 (2002).
[57] M. Salomo, K. Kegler, C. Gutsche, M. Struhalla, J. Reinmuth, W. Skokow, U. Hahn, and F. Kremer, Colloid Polym. Sci. **284**, 1325 (2006).
[58] F. Lankas, R. Lavery, and J. Maddocks, Structure (London) **14**, 1527 (2006).
[59] M. Moakher and J. H. Maddocks, Arch. Ration. Mech. Anal. **177**, 53 (2005).
[60] When $\delta C_{\mathrm{th}}(\sigma) = O(\epsilon)$, additional corrections occur. These involve correlations between stiffness and offsets which are outside the scope of this work.