Modeling DNA loops using the theory of elasticity

Alexander Balaeff*
Beckman Institute and Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

L. Mahadevan
Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Klaus Schulten†
Beckman Institute and Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

(Received 15 August 2005; published 20 March 2006)

An elastic rod model of a protein-bound DNA loop is adapted for application in multi-scale simulations of protein-DNA complexes. The classical Kirchhoff system of equations which describes the equilibrium structure of the elastic loop is modified to account for the intrinsic twist and curvature, anisotropic bending properties, and electrostatic charge of DNA. The effects of bending anisotropy and electrostatics are studied for the DNA loop clamped by the *lac* repressor protein. For two possible lengths of the loop, several topologically different conformations are predicted and extensively analyzed over the broad range of model parameters describing DNA bending and electrostatic properties. The scope and applications of the model in already accomplished and in future multi-scale studies of protein-DNA complexes are discussed.

DOI: 10.1103/PhysRevE.73.031919 PACS number(s): 87.14.Gg, 87.15.Aa, 87.15.La, 02.60.Lj

I. INTRODUCTION

Protein-DNA interactions are of primary importance for living organisms. Proteins are involved in organizing and packing genomic DNA, synthesizing new DNA, reading the information stored in the genes, and controlling the level of expression of each gene [1,2]. The amount of data on protein-DNA interactions both *in vivo* and *in vitro* keeps growing at a remarkable pace. This, paralleled by a similar growth in the available computation power, gives the biomolecular modeling community a superb opportunity to revise and advance the existing models of protein-DNA interactions.

A protein binding to DNA often results in formation of a DNA loop [3–5]. A segment of DNA folds into a loop when either its ends get bound by the same protein molecule or when the DNA gets wound around a large multi-protein aggregate, such as the nucleosome [6]. DNA loop formation is ubiquitous in both prokaryotic and eukaryotic cells; it plays a central role in controlling the gene expression, as well as in DNA recombination, replication, and packing inside the cells [1–7]. Understanding the structure and dynamics of DNA loops is therefore a prerequisite for studying the organization and function of the genomes of living cells.

A complete model of an interaction between a protein and a long DNA loop necessarily involves several spatial and temporal scales [8–12]. On the one hand, the protein complexes formed on the DNA typically do not exceed 100 Å in size; the interactions on the interface between the protein and the DNA, such as formation and breakage of hydrogen bonds or rearrangements in the local protein and DNA structure, occur on pico- to nanosecond time scales. These interactions are typically captured in molecular dynamics (MD) simulations of the all-atom models of the proteins bound to short segments of DNA [8–10]. On the other hand, the DNA loops induced by the bound proteins may measure hundreds of nanometers in length; the characteristic motions of such loops occur on micro- to millisecond time scales. The models of the DNA loops typically involve a certain degree of coarse-graining compared to the all-atom models [9–11,13]. Many such models are based on the Kirchhoff theory of elasticity [14]: they approximate the DNA helix by an elastic rod/ribbon, sometimes carrying an electric charge [9,10,13,15–24].

It has been demonstrated [25,26] that the all-atom and elastic rod models of DNA can be combined in a consistent multi-scale description of a protein-DNA complex. The structure of the DNA loop and the forces it exerts on the protein clamp are obtained from the elastic rod model of the loop that uses the boundary conditions resulting from the all-atom structure of the DNA segments directly bound by the protein. The subsequent all-atom MD simulations of the protein complex with the bound segments use the thus-computed forces, monitoring the resulting changes in the protein structure and dynamics and constantly updating the boundary conditions for the coarse-grained DNA model. The elastic rod model can take into account the electric field of the protein-DNA complex, and the all-atom simulation may similarly include the forces resulting from the electric field of the DNA loop. The multi-scale description yields a picture

*Present address: Department of Chemistry, Duke University, Durham, NC 27708

†Corresponding Author. Also Department of Physics, University of Illinois at Urbana-Champaign. Electronic address: kschulte@ks.uiuc.edu

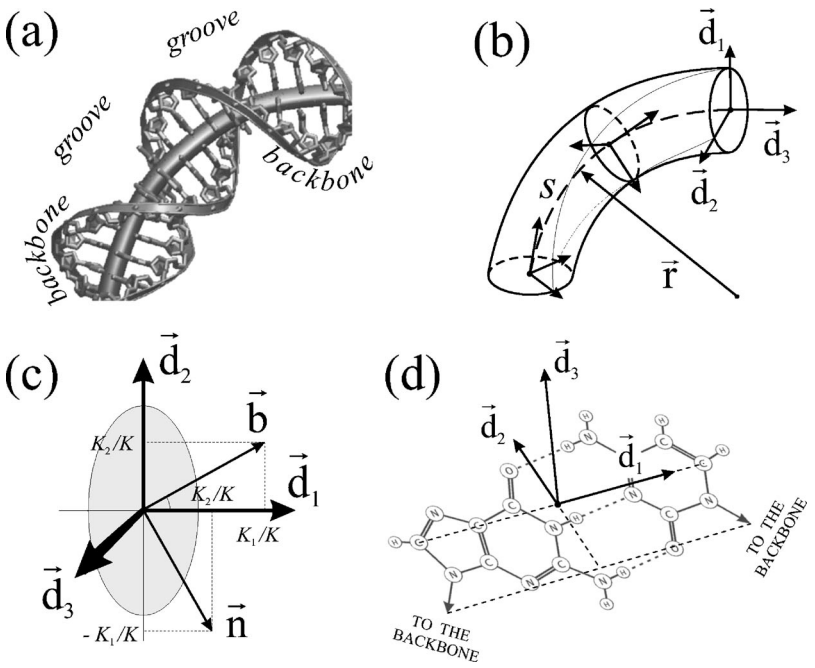

FIG. 1. Elastic rod model of DNA. (a) The elastic rod fitted into an all-atom structure of DNA. (b) Parametrization of the elastic rod: shown are the centerline $\mathbf{r}(s)$ and the intrinsic local frame $(\mathbf{d}_1,\mathbf{d}_2,\mathbf{d}_3)$. (c) The principal normal $\mathbf{n}(s)=\ddot{\mathbf{r}}/|\ddot{\mathbf{r}}|$ and the binormal $\mathbf{b}(s)=\mathbf{d}_3\times\mathbf{n}$ vectors, together with $\mathbf{d}_3$, form the natural (Frenet-Serret) local frame for the 3D curve $\mathbf{r}(s)$. (d) A coordinate frame associated with a DNA base pair. The vectors are defined according to the general convention stated in [44] except that $\mathbf{d}_1$ and $\mathbf{d}_2$ are swapped.

of the structure and dynamics of the protein-DNA complexes which is presumably closer to reality than what is separately predicted by either of the two models.

In the present paper, we describe in detail the Kirchhoff rod model of DNA as used in the multi-scale modeling studies [25,26]. The classical Kirchhoff equations are extended in order to describe such physical properties of DNA as the electric charge [27–30], intrinsic twist and bend [22,23,31], and the anisotropy of DNA bending [22,32–34]. To the best of our knowledge, this work is the first attempt to tie all these properties together in a single system of Kirchhoff equations [35]. All parameters are considered to vary along the DNA loop in order to account for variations in the DNA sequence. A fast computational procedure for numerically solving the extended Kirchhoff equations is developed on the basis of a continuation algorithm [22,36].

For a demonstration and analysis of the model, the elastic rod solutions are obtained for a DNA loop induced by the *lac* repressor, a celebrated *E. coli* protein that became a paradigm of genetic regulation [1,4,37,38]. The boundary conditions obtained from the protein structure [37] are used to solve Kirchhoff equations for the loop lengths of 76 and 385 base pairs (bp) which the *lac* repressor induces in genomic DNA [4,36,39]. The continuation algorithm yields two solutions for the shorter loop and four solutions for the longer loop; the solutions are used to extensively analyze such parameters of the model as the DNA bending anisotropy and electric charge, and to derive recommendations for future studies.

The manuscript is divided into five further sections. In Sec. II, the extended Kirchhoff equations are derived. In Sec. III, the *lac* repressor-DNA system is reviewed, the continuation algorithm is discussed, and the elastic rod solutions for the DNA loops folded by the *lac* repressor are obtained. In Sec. IV, the effect of bending anisotropy on the structure and energy of the DNA loops is analyzed. In Sec. V, adding the electrostatic interaction terms into Kirchhoff equations is discussed, including the required changes to the numerical algorithm, and the effect of electrostatics on the *lac* repressor loops is analyzed. In Sec. VI, the elastic rod DNA model and its applications in the multi-scale simulations are discussed.

II. THE ELASTIC ROD MODEL FOR DNA

Elastic rod theory [14,40,41] is a natural choice for modeling a long linear polymer, such as DNA [9,10,20]. The classical theory of elasticity describes the geometry of an elastic rod (ribbon) in terms of its centerline $\mathbf{r}(s)=(x(s),y(s),z(s))$, a three-dimensional curve parametrized by its arclength s, and a frame of three unit vectors $\mathbf{d}_1(s)$, $\mathbf{d}_2(s)$, $\mathbf{d}_3(s)$ associated with each cross section of the rod [Fig. 1(b)]. In the case of DNA, the centerline of the rod follows the axis of the DNA helix and Watson-Crick base pairs form cross sections of the DNA "rod" [Figs. 1(a) and 1(d)]. Below, we derive the system of equations that describes the mechanical equilibrium conformations of such a rod. The equations of the classical theory [14,17,35,42,43] are modified in order to account for the specific physical properties of DNA.

The centerline and the vectors $\mathbf{d}_{1-3}$ [45] describe the elastic rod conformation in terms of six variables, as the components of all the three vectors $\mathbf{d}_{1-3}$ can be expressed through three Euler angles $\phi(s)$, $\psi(s)$, $\theta(s)$, which define the rotation of the local coordinate frame $(\mathbf{d}_1,\mathbf{d}_2,\mathbf{d}_3)$ relative to the lab coordinate frame. Alternatively, one can use four Euler parameters $q_1(s)$, $q_2(s)$, $q_3(s)$, $q_4(s)$ [22,42], subject to the constraint

$$q_1^2+q_2^2+q_3^2+q_4^2=1. \tag{1}$$

The Euler parameters allow one to avoid the polar singularities inherent in the Euler angles [22,42] and are therefore employed in this paper.

If the elastic rod is inextensible, which is the approximation considered in this paper, then the tangent to the center

line $\dot{\mathbf{r}}$ coincides with the normal $\mathbf{d}_3$ [40,41], resulting in another geometric constraint,

$$\dot{\mathbf{r}}(s) = \mathbf{d}_3(s) \tag{2}$$

(the dot denotes the derivative with respect to s). The changes to the equations of elasticity due to using a model of an extensible rod are described in Appendix A.

Other important parameters that describe the geometry of the elastic rod are the curvature $K(s)$ of the centerline and twist $\Omega(s)$ of the rod around the centerline. The curvatures and the twist describe the *spatial* rate of rotation of the elastic rod cross section at each point s, as noted by the famous Kirchhoff's analogy between the sequence of the cross sections along the elastic rod and a motion of a rigid body [14,40]. The following equations ensue:

$$\dot{\mathbf{d}}_{1-3} = \mathbf{k} \times \mathbf{d}_{1-3}, \quad \mathbf{k} = \{K_1, K_2, \Omega\}, \tag{3}$$

where K_1 and K_2 are the principal components of the curvature so that $K = \sqrt{K_1^2 + K_2^2}$; the vector $\mathbf{k}$ is called the vector of strains [40,41]. The principal components of the curvature define the normal and binormal vectors:

$$\mathbf{n}(s) = (K_2/K)\mathbf{d}_1 - (K_1/K)\mathbf{d}_2, \tag{4}$$

$$\mathbf{b}(s) = (K_1/K)\mathbf{d}_1 + (K_2/K)\mathbf{d}_2, \tag{5}$$

which together with $\mathbf{d}_3$ form the natural (Frenet-Serret) local frame for the elastic rod [Fig. 1(c)].

Using (3), one can express the curvatures and the twist via the Euler parameters q_{1-4} (see [46], p. 16):

$$K_1 = 2(q_4\dot{q}_1 + q_3\dot{q}_2 - q_2\dot{q}_3 - q_1\dot{q}_4), \tag{6}$$

$$K_2 = 2(-q_3\dot{q}_1 + q_4\dot{q}_2 + q_1\dot{q}_3 - q_2\dot{q}_4), \tag{7}$$

$$\Omega = 2(q_2\dot{q}_1 - q_1\dot{q}_2 + q_4\dot{q}_3 - q_3\dot{q}_4). \tag{8}$$

Classically, a relaxed elastic rod (ribbon) is straight and untwisted so that $K=0$, $\Omega=0$. However, the relaxed shape of DNA is a helix with a pitch $H=36$ Å, or 10.4 bp [1]. The pitch is much smaller than the persistence length of DNA bending (500 Å) or twisting (750 Å) [47], so even a relatively straight segment of DNA is tightly twisted. Moreover, certain DNA sequences are known to be intrinsically curved [20,23,32,33]. Therefore, we separate the curvature and the twist into intrinsic and imposed components:

$$K_{1,2} = \kappa_{1,2} + \overset{\circ}{\kappa}_{1,2}, \quad \Omega = \omega + \omega^{\circ}. \tag{9}$$

The intrinsic twist and curvature of DNA is known to vary between different sequences [23,32,33,48], therefore the equations below will be derived in the general form, using arclength-dependent parameter-functions $\overset{\circ}{\kappa}_{1,2}(s)$, $\omega^{\circ}(s)$.

If the elastic rod is forced into a shape different from that of its relaxed state, then elastic forces $\mathbf{N}(s)$ and torques $\mathbf{M}(s)$ develop inside the rod:

$$\mathbf{N}(s) = \sum_{i=1}^{3} N_i \mathbf{d}_i, \quad \mathbf{M}(s) = \sum_{i=1}^{3} M_i \mathbf{d}_i. \tag{10}$$

The components N_1 and N_2 constitute the shear force $N_{sh} = \sqrt{N_1^2 + N_2^2}$; the component N_3 is the force of tension (if $N_3 > 0$) or compression (if $N_3 < 0$) at the cross section at the point s. M_1 and M_2 are the bending moments, and M_3 is the twisting moment. In our model, we adopt the widely used Bernoulli-Euler approximation [17,36,42] that linearly relates the elastic torque to the imposed curvatures and twist:

$$\mathbf{M}(s) = A_1\kappa_1\mathbf{d}_1 + A_2\kappa_2\mathbf{d}_2 + C\omega\mathbf{d}_3. \tag{11}$$

The linear coefficients A_1 and A_2 are called the bending rigidities of the elastic rod, and C is called the twisting rigidity [47]. If $A_1 = A_2$, we call the elastic rod isotropically bendable; the cross section of such a rod must have a rotational symmetry of the fourth order [40,41]. The experimental data on DNA elasticity are usually interpreted in terms of isotropically bendable DNA [49–52]. However, the atomic level structure of the DNA cross section does not exhibit the required symmetry, implying anisotropic bending [cf. Figs. 1(a) and 1(d)]. Hence, we derive our equations for the general case of $A_1 \neq A_2$ and vary the bending moduli in order to study the effect of DNA bending anisotropy (Sec. IV). Furthermore, the DNA elastic moduli are known to depend on its sequence [9,23,31,33,53]. Accordingly, Eq. (11) and all the subsequent equations employ parameter-functions $A_1(s)$, $A_2(s)$, and $C(s)$ rather than constant parameters A_1, A_2, C. The way to construct these parameter-functions for a DNA loop of a given sequence is discussed in Appendix C.

There are, conceivably, external body forces $\mathbf{f}(s)$ and torques $\mathbf{g}(s)$ acting upon the rod. In mechanical equilibrium, the elastic forces and torques balance the external forces and torques at every point s [17,22,24,36,43]:

$$\dot{\mathbf{N}} + \mathbf{f} = 0, \tag{12}$$

$$\dot{\mathbf{M}} + \dot{\mathbf{g}} + \dot{\mathbf{r}} \times \mathbf{N} = 0. \tag{13}$$

The body forces and torques of the classical theory usually result from gravity or from the weight of external bodies, e.g., in the case of construction beams. In the case of DNA, such forces are mainly of electrostatic nature and arise from either the self-repulsion of the negatively charged DNA or from interactions with biomolecular aggregates, such as protein molecules or lipid membranes. The treatment of electrostatic forces in our equations is described in Sec. V.

Equations (1), (2), and (11)–(13) form the basis of the Kirchhoff theory of elastic rods. We simplify the equations by first making all the variables dimensionless, i.e., defining

$$\bar{s} = s/l, \quad \bar{x} = x/l, \quad \bar{y} = y/l, \quad \bar{z} = z/l, \tag{14}$$

$$\bar{K}_1 = lK_1, \quad \bar{K}_2 = lK_2, \quad \bar{\Omega} = l\Omega, \tag{15}$$

$$\alpha = A_1/C_{\circ}, \quad \beta = A_2/C_{\circ}, \quad \gamma = C/C_{\circ}, \tag{16}$$

$$\bar{N}_{1-3} = N_{1-3} l^2 / C_\circ, \quad \bar{M}_{1-3} = M_{1-3} l / C_\circ, \tag{17}$$

where l is the length of the rod and $C_\circ$ is some reference value of the twisting modulus, for example, the average DNA value $C_\circ = 3 \times 10^{-19}$ erg cm. Second, we express the derivatives $\dot{q}_{1-4}$ through K_1, K_2, Ω, and q_{1-4}, using Eqs. (6)–(8) and constraint (1) differentiated with respect to s. Third, we eliminate the variables N_1 and N_2 using (13) and arrive at the following system of differential equations of 13th order [54]:

$$(\alpha\ddot{\kappa}_1) = (2\dot{\beta}\kappa_2\Omega) - (\gamma\dot{\omega}K_2) - \beta\kappa_2\dot{\Omega} + \alpha\kappa_1\Omega^2 - \gamma K_1\omega\Omega + K_1 N_3 + \Omega\dot{g}_2 - \dot{f}_2 - \ddot{g}_1, \tag{18}$$

$$(\beta\ddot{\kappa}_2) = -(2\alpha\dot{\kappa}_1\Omega) + (\gamma\dot{\omega}K_1) + \alpha\kappa_1\dot{\Omega} + \beta\kappa_2\Omega^2 - \gamma K_2\omega\Omega + K_2 N_3 - \Omega\dot{g}_1 + \dot{f}_1 - \ddot{g}_2, \tag{19}$$

$$(\gamma\dot{\omega}) = \alpha\kappa_1 K_2 - \beta K_1\kappa_2 - \dot{g}_3, \tag{20}$$

$$\dot{N}_3 = -(\alpha\dot{\kappa}_1)K_1 - (\beta\dot{\kappa}_2)K_2 - (\gamma\dot{\omega})\Omega - g_1 K_1 - g_2 K_2 - g_3\Omega - \dot{f}_3, \tag{21}$$

$$\dot{q}_1 = \tfrac{1}{2}(K_1 q_4 - K_2 q_3 + \Omega q_2), \tag{22}$$

$$\dot{q}_2 = \tfrac{1}{2}(K_1 q_3 + K_2 q_4 - \Omega q_1), \tag{23}$$

$$\dot{q}_3 = \tfrac{1}{2}(-K_1 q_2 + K_2 q_1 + \Omega q_4), \tag{24}$$

$$\dot{q}_4 = \tfrac{1}{2}(-K_1 q_1 - K_2 q_2 - \Omega q_3), \tag{25}$$

$$\dot{x} = 2(q_1 q_3 + q_2 q_4), \tag{26}$$

$$\dot{y} = 2(q_2 q_3 - q_1 q_4), \tag{27}$$

$$\dot{z} = -q_1^2 - q_2^2 + q_3^2 + q_4^2. \tag{28}$$

These equations, similar to those used by others [17,22,36,42,55] but derived here in a more general form, can be solved numerically or, in certain cases, analytically (e.g., for $\alpha=\beta$, $\mathbf{f}=0$, $\mathbf{g}=0$ [17,43]). A solution to the system consists of 13 functions: $\mathbf{r}(s)$, $q_{1-4}(s)$, $\mathbf{N}(s)$, and $\mathbf{M}(s)$ [the latter directly obtainable from $\mathbf{k}(s)$ by virtue of (11)]. These functions describe the geometry of the elastic rod and the distribution of stress and torques along it. In the case of a DNA loop with the ends bound to a single protein or a multiprotein aggregate, the positions and orientations of the ends are presumably known; therefore, one can deduce $\mathbf{r}(0)$, $\mathbf{r}(1)$, $q_{1-4}(0)$, $q_{1-4}(1)$, e.g., as explained in the next section, and solve the system (18)–(28) as a boundary value problem (BVP). The obtained—generally, multiple—solutions will represent the set of equilibrium conformations of the loop achievable under the given boundary conditions. The forces $\mathbf{N}(0)$, $-\mathbf{N}(1)$ and torques $\mathbf{M}(0)$, $-\mathbf{M}(1)$ correspond to the forces and torques which the DNA loop exerts upon the protein clamp at each end. These are the forces and torques that are sought in the multi-scale simulation method [25] and included into the MD simulation of a protein-DNA complex.

Which of the multiple solutions to the system (18)–(28) is to be used in the multi-scale simulation can be determined by an energy criterion, assuming that the ensemble of the DNA loop conformations is at thermodynamic equilibrium. The dimensionless elastic energy of each solution is computed, according to the Bernoulli-Euler approximation (11), as the quadratic functional of the curvatures and twist

$$U = \int_0^1 \left(\frac{\alpha\kappa_1^2}{2} + \frac{\beta\kappa_2^2}{2} + \frac{\gamma\omega^2}{2} \right) ds. \tag{29}$$

This functional may further include the electrostatic interaction terms (as described in Sec. V) and the terms due to the interactions of the DNA loop with other biomolecules (as expressed through the forces $\mathbf{f}$ and torques $\mathbf{g}$), in case those terms are not negligible. The solution with the lowest energy shall have the highest weight in the thermodynamic ensemble and the forces/torques derived from that are to be used in the multi-scale simulation. However, if multiple conformations of comparable energy (within 1–2 kT from the lowest one) are found, then their effect on the protein structure shall be studied separately, because the exchange time between the different loop conformations is much larger than the typical times of protein structural dynamics studied by MD.

The system (18)–(28) describes the elastic rod model of DNA in the most general terms. Not all of the options provided by such a model will be explored in the demonstration study presented below; most times the equations will be simplified in one way or another. The unexplored possibilities and situations when the various options might become relevant will be discussed in Sec. VI.

III. ELASTIC ROD SOLUTIONS FOR THE DNA LOOP CLAMPED BY THE *lac* REPRESSOR

In this section, we first describe our trial system, the complex of the *lac* repressor protein with a DNA loop. Then the DNA loop involved in the system is used to illustrate the numeric algorithm for solving the equations of elasticity (18)–(28). Finally, the different solutions obtained for the DNA loop are discussed.

A. The DNA complex with the *lac* repressor

In order to explore the extended Kirchhoff equations, we build an elastic rod model of the DNA loop induced in the *E. coli* chromosome by the *lac* repressor protein. The *lac* repressor functions as a switch that shuts down the lactose (*lac*) operon, a set of *E. coli* genes the study of which laid one of the cornerstones of modern molecular biology [1,4,37,38]. The genes code for proteins that are responsible for lactose digestion by the bacterium; they are shut down by the *lac* repressor when lactose is not present in the environment. When lactose is present, the molecules of it bind inside the *lac* repressor and deactivate the protein, thereby inducing the expression of the *lac* operon [Figs. 2(a) and 2(b)].

The *lac* repressor consists of two DNA-binding "hands," seen in the crystal structure of the protein [37] [Fig. 2(c)].

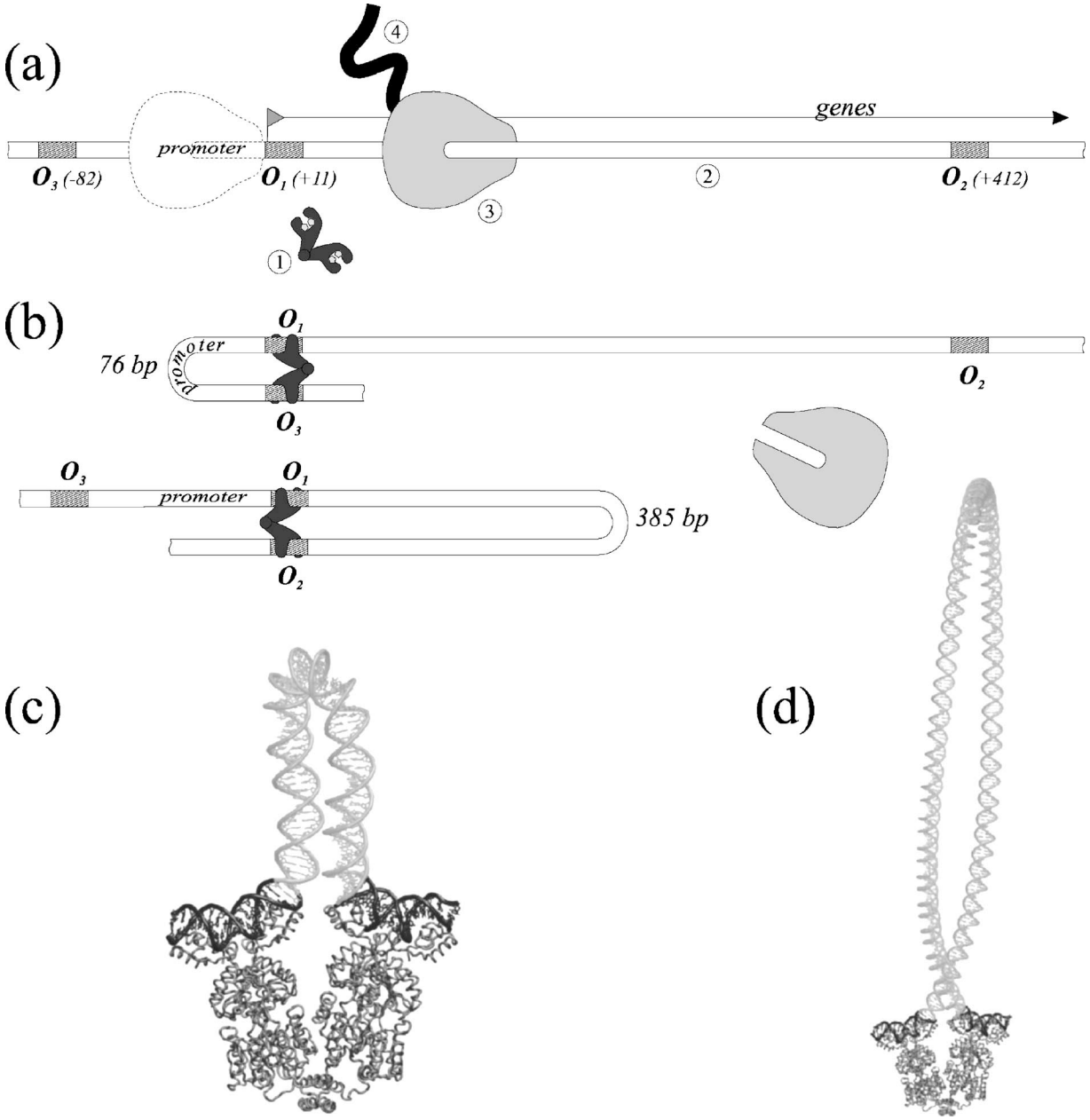

FIG. 2. (a) The expressed *lac* operon. The biomolecules involved are (1) *lac* repressor (shown deactivated by four bound lactose molecules), (2) the DNA of *E. coli*, (3) RNA polymerase (shown first bound to the promoter and then transcribing the *lac* operon genes), and (4) mRNA (shown as being transcribed by the RNA polymerase). The flag shows the "+1" base pair of the DNA where the genes of the *lac* operon begin. The operator sites, marked as O_{1-3}, are shown as shaded rectangles; the position of each operator's central base pair is shown in brackets. (b) The repressed *lac* operon. The *lac* repressor is shown binding two operator sites, either O_1 and O_3 or O_1 and O_2, and the RNA polymerase is shown released from the operon. The end-to-end length of the DNA loop formed in each case is indicated. (c) The crystal structure of the *lac* repressor [37]. The DNA loop, missing in the crystal structure and shown here in light color, corresponds to an all-atom model fitted to one of the two elastic rod structures predicted for the 76 bp loop (see [18] and Sec. III B). Small pieces of the crystal structure are omitted from the figure for clarity. (d) Same as (c) for the 385 bp loop; the all-atom DNA model is fitted to one of the four possible structures of the loop (see Sec. III C and Fig. 6).

Each "hand" recognizes a specific 21 bp long sequence of DNA, called the operator site. The *lac* repressor binds to two operator sites and causes the DNA connecting those sites to fold into a loop. There are three operator sites for the *lac* repressor in the *E. coli* genome: O_1, O_2, and O_3, all three are necessary for the maximum repression of the *lac* operon [4,5,39]. One hand of the *lac* repressor must bind to O_1 and the other to either O_2 or O_3; therefore, the resulting DNA loop has a length of either 385 bp (O_1-O_2) or 76 bp (O_1-O_3) [Fig. 2(b)]. While the long loop is the easier to form, the short loop contains the *lac* operon promoter, and the formation of this loop is very disruptive for the expression of the *lac* operon.

While the structure of the *lac* repressor itself is known, it

would hardly be possible to crystallize the induced DNA loops, merely because of their size, and thus to study both their structure and their effect on the *lac* repressor. The crystal structure [37] of the *lac* repressor complex with DNA includes only two disjoint operator segments bound to the protein. Bending a continuous piece of DNA into a loop puts the *lac* repressor under a stress that was shown to change the protein structure [26,56]. The elastic rod model provides a perfect tool for predicting the structure of the missing loops and the forces of protein-DNA interaction, and for the subsequent studying of the effect of the DNA loops on the protein structure by means of a multi-scale simulation [25,26].

B. Solving the equations of elasticity for the 76 bp long promoter loop O_1-O_3

The structure of the missing loops is built by numerically solving Eqs. (18)–(28) with the boundary conditions obtained from the crystal structure [37] of the *lac* repressor-DNA complex. The terminal [57] base pairs of the DNA segments bound to the *lac* repressor in the structure [37] are interpreted as the cross sections of the loop at the beginning and at the end, and orthogonal frames are fitted to those base pairs, as illustrated in Fig. 1(d). The positions of the centers of those frames and their orientations relative to the lab coordinate system (LCS) provide 14 boundary conditions: $\mathbf{r}(0)$, $\mathbf{r}(1)$, $q_{1-4}(0)$, $q_{1-4}(1)$. In order to match the 13th order of the system (18)–(28), a boundary condition for one of the q_i's is dropped; it will be automatically satisfied because the identity (1) is included into the equations.

The iterative continuation algorithm used for solving the BVP is the same as that used in our previous work [18] (with some modification when the electrostatic self-repulsion is included into the equations, as described in Sec. V). The solution to the problem is constructed in a series of iteration cycles. The cycles start with a certain set of boundary conditions and model parameters for which an exact solution is known. Then the boundary conditions and model parameters are changed towards the desired values during the iteration cycles; only a certain subset of parameters is normally changed during each cycle, e.g., only $\mathbf{r}(1)$ or only α/β. During the cycle, the chosen parameters evolve towards the desired values through a number of iteration steps; the number of steps is chosen depending on the sensitivity of the problem to the parameters being modified. At each step, the solution found on the previous step is used as an initial guess; with a proper choice of the iteration step, the two consecutive solutions are close to each other, which guarantees the convergence of the numerical BVP solver. For the latter, a classical software, COLNEW [58], is employed. COLNEW uses a damped quasi-Newton method to construct the solution to the problem as a set of collocating splines.

The known exact solution, from which the iteration cycles started, was chosen to be a circular closed $[\mathbf{r}(0)=\mathbf{r}(1)]$ elastic loop with zero intrinsic curvature $\overset{\circ}{\kappa}_{1,2}$, constant intrinsic twist $\omega^\circ=34.6$ deg/bp (the average value for classical B-form DNA [1]), constant elastic moduli $\alpha=\beta=\frac{1}{2}$, and zero electrostatic charge ($Q_{\mathrm{DNA}}=0$) [59]. This solution is shown in Fig. 3(a); the explicit form of the solution can be found in [60]. The loop started (and ended) at the center of the terminal base pair of one of the protein-bound DNA segments. The coordinate frame associated with that base pair, i.e., with the loop cross section at $s=0$, was chosen as the LCS. The orientation of the plane of the loop was determined by a single parameter $\psi_\circ$, the angle between the plane of the loop and the x axis of the LCS.

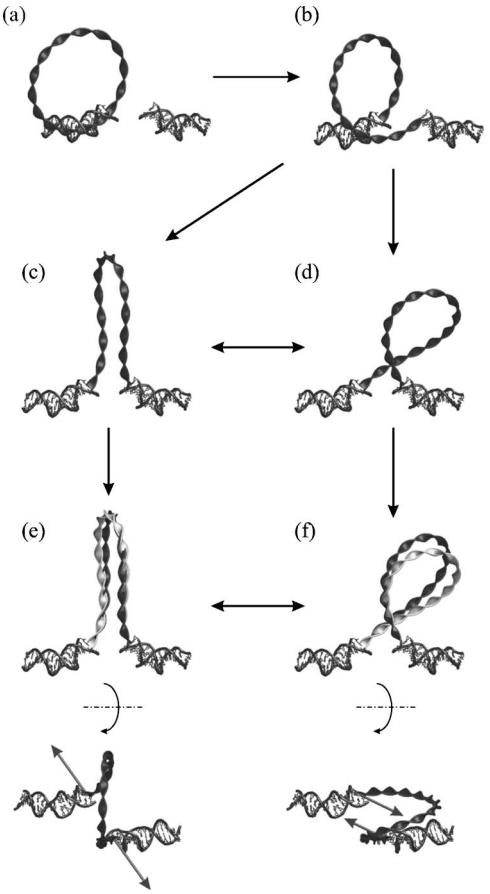

FIG. 3. Evolution of the elastic rod structure during the solution of the BVP for the short loop. (a) The initial solution: a closed circular loop. (b) The solution after the first iteration cycle. (c), (d) The solutions after the second iteration cycle, for the clockwise (c) and counter-clockwise (d) rotation of the $s=1$ end. (e), (f) The solutions after the third iteration cycle; the previous solutions are shown in light color; the views from the top include the forces that the DNA loop exerts on the DNA segments bound to the *lac* repressor. The protein-bound DNA segments from the *lac* repressor crystal structure are shown for reference only, as they played no role during the iteration cycles except for providing the boundary conditions.

In the first iteration cycle, the value of $\mathbf{r}(1)$ was changed, moving the $s=1$ end of the loop by 45 Å to its presumed location at the beginning of the second DNA segment [Fig. 3(b)]. In the second iteration cycle, the cross section of the elastic rod at the $s=1$ end was rotated to satisfy the boundary conditions for $q_{1-4}(1)$ [Figs. 3(c) and 3(d)]. The rotation consisted in simultaneously turning the normal $\mathbf{d}_3$ of the cross section to coincide with the normal to the terminal base pair and rotating the cross section around the normal in order to align the vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ with the axes of the base pair.

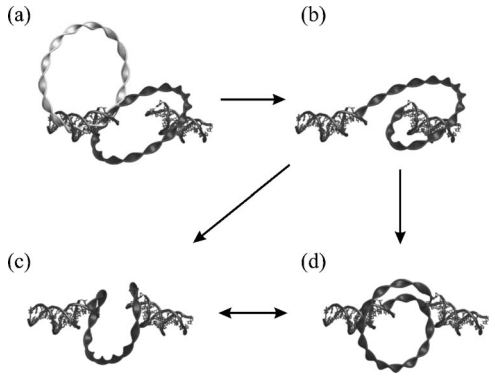

FIG. 4. Extraneous solutions to the BVP obtained for a different orientation $\psi_\circ$ of the initial circular loop. The initial loop from Fig. 3 (a) is shown in panel (a) in light color.

Depending on the direction of the rotation, two different solutions to the problem arise. Rotating the $s=1$ end clockwise results in the solution shown in Fig. 3(c). Rotating the end counter-clockwise results in the solution shown in Fig. 3(d). The former solution is underwound by $\omega=-1.4$ deg/bp on average and the latter solution is overwound by $\omega=1.6$ deg/bp on average; accordingly, the solutions will hereafter be referred to as "U" and "O." The two solutions may be transformed into each other through an additional iteration cycle, namely, turning the $s=1$ end around its normal by 2π. Continuous rotation of the $s=1$ end results in switching between the two solutions. Switching from the U to the O solution is accompanied by a self-crossing of the DNA loop, not prevented in the model at this point. Topologically, rotating the end by a whole turn clockwise increases the linking number [15,16,20,21] of the loop by 2π and a self-crossing reduces the linking by 4π; therefore, two full turns are canceled by one self-crossing, and the original solution gets restored after two turns.

In the third iteration cycle, the bending moduli α and β (so far, kept equal to each other) were changed from $\frac{1}{2}$ to $\frac{2}{3}$, the experimentally measured ratio between DNA bending and twisting moduli [47]. The resulting increase in the bending rigidity somewhat changed the geometry of the U and O solutions [Figs. 3(e) and 3(f)] and increased the unwinding/overwinding to -1.6 and 2.0 deg/bp, respectively. The change in ω has a clear topological implication: the deformation of the looped DNA is distributed between the writhe (bending) of the centerline and the unwinding/overwinding of the DNA helix. When the bending becomes energetically more costly, the centerline of the loop straightens up, on average, and the deformation shifts towards a bigger twist.

Notably, two more solutions may result from the described iteration procedure, depending on the orientation $\psi_\circ$ of the initial simplified circular loop (Fig. 4). However, for the 76 bp loop these solutions are not acceptable, because the centerlines of the corresponding DNA loops would have to run right through the structure of the *lac* repressor [cf. Figs. 4(c), 4(d), and 2(c)].

Therefore, only the two former solutions to the problem are acceptable in the case of the 76 bp loop. The solutions obtained after the third iteration cycle become our first-approximation answer to what the structure of the DNA loop created by the *lac* repressor must be. The solutions are portrayed in Figs. 3(e) and 3(f), and the profiles of their curvature, twist, and elastic energy density are shown in Fig. 5 (left column).

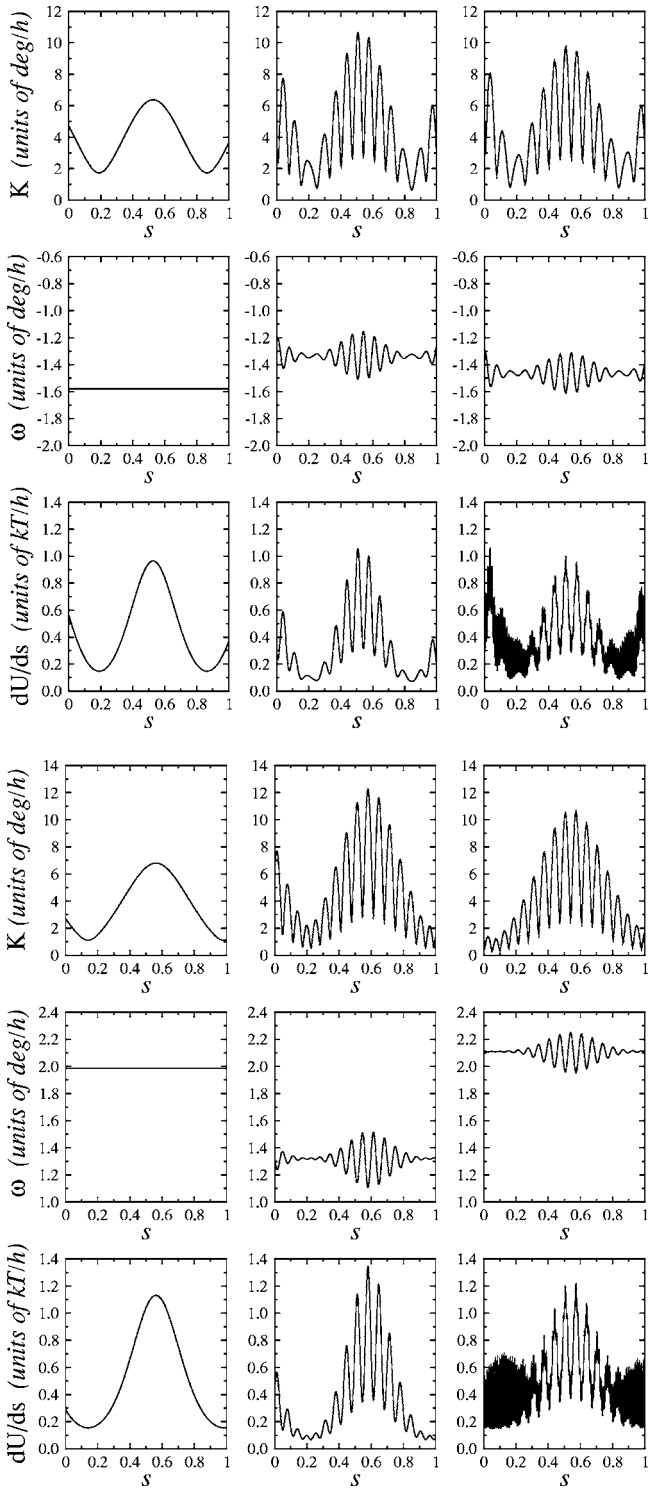

FIG. 5. Distribution of curvature, twist, and energy in the elastic rod BVP solutions (U—top panel, O—bottom panel) for the 76 bp DNA loop. In each panel, left column: the isotropic solution with $\alpha=\beta=2/3$; center column: the anisotropic solution with $\alpha=4/15$, $\beta=16/15$; right column: the anisotropic solution with electrostatics, for an ionic strength of 10 mM.

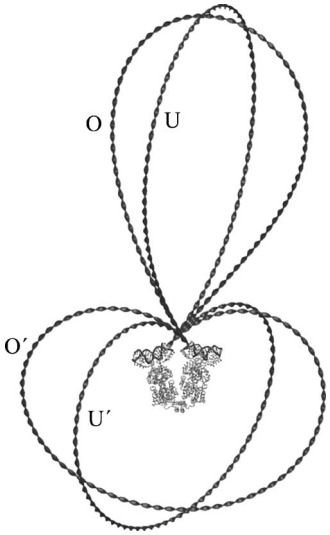

FIG. 6. Four solutions of the BVP problem for the 385 bp long DNA loop. Underwound solutions are marked by the letter "U," overwound ones by "O."

The U solution forms an almost planar loop, its plane being roughly perpendicular to the protein-bound DNA segments [Fig. 3(e)]. The shape of the loop resembles a semicircle sitting on two relatively straight segments connected by short curved sections to the *lac* repressor-bound DNA. Accordingly, the curvature of the loop is highest in the middle and at the ends and drops in between (Fig. 5). The average curvature of this loop is 3.7 deg/bp; the highest curvature, achieved in the middle of the loop, is around 6 deg/bp. Since $\alpha=\beta$, the unwinding is constant and the energy density profile simply follows the curvature plot. The total energy of the loop is 33.0 kT, of which 26.8 kT are accounted for by bending and 6.2 kT by twisting. The stress of the loop pushes the ends of the protein-bound DNA segments (and, consequently, the *lac* repressor headgroups) away from each other with a force of about 10.5 pN [Fig. 3(e)].

The O solution leaves and enters the protein-bound DNA segments in almost straight lines, connected by a semicircular coil of about the same curvature as that of the U solution, not, however, confined to any plane [Fig. 3(f)]. The average curvature equals 3.6 deg/bp. The energy of this loop is higher than that of the U loop: 38.2 kT, distributed between bending and twisting as 28.5 and 9.7 kT, respectively. The forces of the loop interaction with the protein-bound DNA segments equal 9.2 pN and are pulling the ends of each segment towards the other segment [Fig. 3(f)].

Since the energy of the U loop is 5 kT lower than that of the O loop, one could conclude that this form of the loop should dominate under conditions of thermodynamic equilibrium. Yet, both energies are too high: the estimate of the energy of the 76 bp loop from the experimental data [61] is approximately 20 kT at high salt concentration (see Appendix B). Therefore, one cannot at this point draw any conclusion as to which loop structure prevails, and further improvements to the model are needed, such as those described in Secs. IV and V.

TABLE I. Geometric and energetic properties of the four solutions of the BVP problem for the 385 bp long DNA loop, in the case $\alpha=\beta=2/3$.

Solution	$\langle\omega\rangle$ (deg/bp)	$\langle\kappa\rangle$ (deg/bp)	κ_{max} (deg/bp)	U (kT)	U_κ/U	U_ω/U
U	−0.24	0.73	1.31	6.2	0.88	0.12
O	0.40	0.80	1.49	8.7	0.76	0.24
U′	−0.33	1.21	1.59	14.4	0.90	0.10
O′	0.42	1.23	1.58	15.5	0.86	0.14

C. Solutions for the 385 bp long O_1-O_2 loop

Using the same algorithm, the BVP was solved for the 385 bp loop. Similarly, four solutions were obtained [cf. Figs. 3(e), 3(f), 4(c), and 4(d)]. With the longer loop, the previously unacceptable solutions are running around the *lac* repressor rather than through it and, therefore, are acceptable. All four solutions (denoted as U, U′, O, O′) are shown in Fig. 6. The solutions U and U′ are underwound, O and O′ are overwound. The geometric and energetic parameters of the four solutions are shown in Table I.

The elastic energy and the average curvature and twist of the longer loops are smaller than those of the shorter loops of the corresponding topology. That is not surprising. The curvature and the twist decrease because the same amount of topological change (linking number) has to be distributed over the larger length of the loop. As the energy density is proportional to the square of the local curvature/twist [cf. (29)], the integral energy is roughly inversely proportional to the length of the loop.

More interestingly, it is the U loop again that has the lowest energy and, *prima facie*, should be predominant under thermodynamic equilibrium conditions. The formerly extraneous U′ and O′ solutions clearly have such high energies that they should practically not be represented in the thermodynamical ensemble of the loop structures and could be safely discounted for the 385 bp loop as well.

Yet again, the conclusions, based on the obtained energy values, shall be postponed until the present elastic rod model is further refined.

IV. EFFECTS OF ANISOTROPIC BENDABILITY

As the first step towards refining our model, a closer look is taken into the DNA bending moduli A_1 and A_2 [or α and β (16)]. Having started with the model of the isotropically bendable DNA, i.e., the one with $\alpha=\beta$, we consider in this section the effect of anisotropic bendability $\alpha\neq\beta$ on the conformation and energy of the elastic rod.

A. Anisotropic moduli of DNA bending

Until recently, the experimental data on DNA bending have been interpreted in terms of a single effective bending modulus A [49–52], and many theoretical studies used

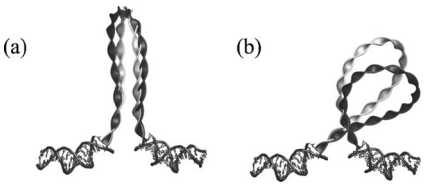

FIG. 7. Changes in the predicted structure of the elastic loop due to the bending anisotropy. (a) The underwound (U) solution. (b) The overwound (O) solution. Structures obtained for $\alpha=\beta=2/3$ are shown in light color; structures obtained for $\alpha=4/15$, $\beta=16/15$ are shown in dark color.

the approximation of isotropically bendable DNA [13,15,17,19,21,22,29,42,55,62]. Such an approximation simplifies the equations of elasticity and the resulting DNA geometry, e.g., by causing the unwinding/overwinding ω to be constant along the DNA loop whenever the external torque g_3 is constant [cf. Eq. (20) and Fig. 5, left column].

However, the atomic structure of DNA helix exhibits two sugar-phosphate backbone strands separated by two grooves [Fig. 1(a)]; clearly, bending towards the grooves (roughly, around $\mathbf{d}_1$) should cost less energy than bending over the backbone (roughly, around $\mathbf{d}_2$). Moreover, the bending properties of DNA are known to strongly depend on its sequence [23,31,33,53]. Therefore, we vary the two bending moduli $A_1 \neq A_2$ independently and study the effect of bending anisotropy on the structure and energy of the DNA loops [63].

The effective bending modulus A is related to A_1 and A_2 through a classical formula [16,23,34] that implies the independence of thermal fluctuations in the two principal bending directions and, consequently, the equidistribution of energy between those fluctuations:

$$\frac{1}{A}=\frac{1}{2}\left(\frac{1}{A_1}+\frac{1}{A_2}\right). \tag{30}$$

This relationship might be different in the case of a nonnegligible coupling between the bending fluctuations, e.g., due to the high intrinsic twist of DNA [60], and is further complicated when the effective modulus is measured for DNA with nontrivial intrinsic geometry [23]; such cases will not be considered in the present work.

In order to derive α and β from (30) one needs more information than the single experimental value of $A/C=2/3$ [47]. For example, one may demand to know the ratio $\mu=A_1/A_2=\alpha/\beta$. However, this ratio is less certain than the A/C ratio. In [34], the value of $\mu=1/4$ was suggested as both being close to experimental data and reproducing well the DNA persistence length in Monte Carlo simulations. But other estimates of μ result from comparison of the oscillations of roll and tilt (the angles of DNA bending in the two principal directions) [32], which are directly related to α and β. The dependence of both μ and A/C on the DNA sequence adds even more uncertainty as to what their effective values for a specific loop should be.

For these reasons, we choose to study the effect of bending anisotropy on the structure and energy of the *lac* repressor loops in a broad range of parameters $\mu=\alpha/\beta$ and $A/C=2\alpha\beta/(\alpha+\beta)$. The loops generated for a specific pair of values $\alpha=4/15$ and $\beta=16/15$ ($\mu=1/4$) are selected for a detailed structural and energetic analysis.

B. Structure of the 76 bp long U and O loops in the case of $\mu=1/4$

Let us first consider the short U and O loops for the selected moduli with $\mu=1/4$. The structures of the new loops were built in the course of another iteration cycle, during which the previously generated isotropic structures with $\alpha=\beta=2/3$ were transformed by simultaneously changing the bending moduli α and β towards the desired values of 4/15 and 16/15, respectively.

The structures of the thus constructed loops are shown in Fig. 7. As one can see, the U loop is not much different from the one in the isotropic case, the root-mean square deviation (rmsd) [64] between the two being only 1.1 h. However, the O loop differs from the isotropic case more significantly, by rmsd of 4.6 h; the loop bends over itself forming a point of near self-contact similarly to the isotropic O loop with $\alpha=\beta=1/2$ [cf. Fig. 3(f)]. The energy of the new U loop equals 23.3 kT, distributed between bending and twisting as 18.9 and 4.4 kT, respectively. The energy of the new O loop equals 26.5 kT, distributed between the bending and twisting as 22.2 and 4.3 kT, respectively. The elastic forces acting at the loop ends equal 7.9 pN for the U loop and 7.2 pN for the O loop.

The distribution of curvature and twist in the anisotropic loops is shown in Fig. 5 (central column). The previously smooth profiles acquire a seesaw pattern, observed in other studies, too [23,33]. This happens because the elastic rod—now better called elastic ribbon—twists around the centerline with a high frequency due to its high intrinsic twist. Accordingly, the vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ get in turn aligned with the principal normal $\mathbf{n}$ that points towards the main bending direction [cf. Fig. 1(c)]. In DNA terms, the double helix successively faces the main bending direction with the grooves and the backbone [cf. Figs. 1(a) and 1(b)].

When the vector $\mathbf{d}_2$ is aligned with $\mathbf{n}$, all bending occurs towards the grooves, resulting in the principal curvatures $|K_1|=|K|$, $K_2=0$, and the local energy density $dU_\kappa=\alpha K_1^2/2$. After a half-period of intrinsic twist, $\mathbf{d}_1$ gets aligned with $\mathbf{n}$ and all bending occurs towards the backbone, resulting in $K_1=0$, $|K_2|=|K|$, and $dU_\kappa=\beta K_2^2/2$. Since $\alpha<\beta$, bending towards $\mathbf{d}_1$ results in higher energetic penalty and elastic torques than bending towards $\mathbf{d}_2$. Therefore, the sections of the rod facing the main bending direction with the backbone ($\mathbf{d}_1$) become less bent, and those facing it with the grooves ($\mathbf{d}_2$) become more bent, resulting in the observed oscillations of the curvature and twist (Fig. 5). The structure of the rod becomes an intermediate between that of a smoothly bent loop and that of a chain of straight links, corresponding to the limits $\mu=1$ and $\mu=0$. The sections where the rod is facing the main bending direction with $\mathbf{d}_2$ play the role of "soft joints" where most of the bending is localized.

The local twist ω of the anisotropic loops also displays oscillations (Fig. 5). When all the bending occurs towards $\mathbf{d}_1$, the twist slightly increases, winding the "rigid face" away from the main bending direction. When the bending occurs

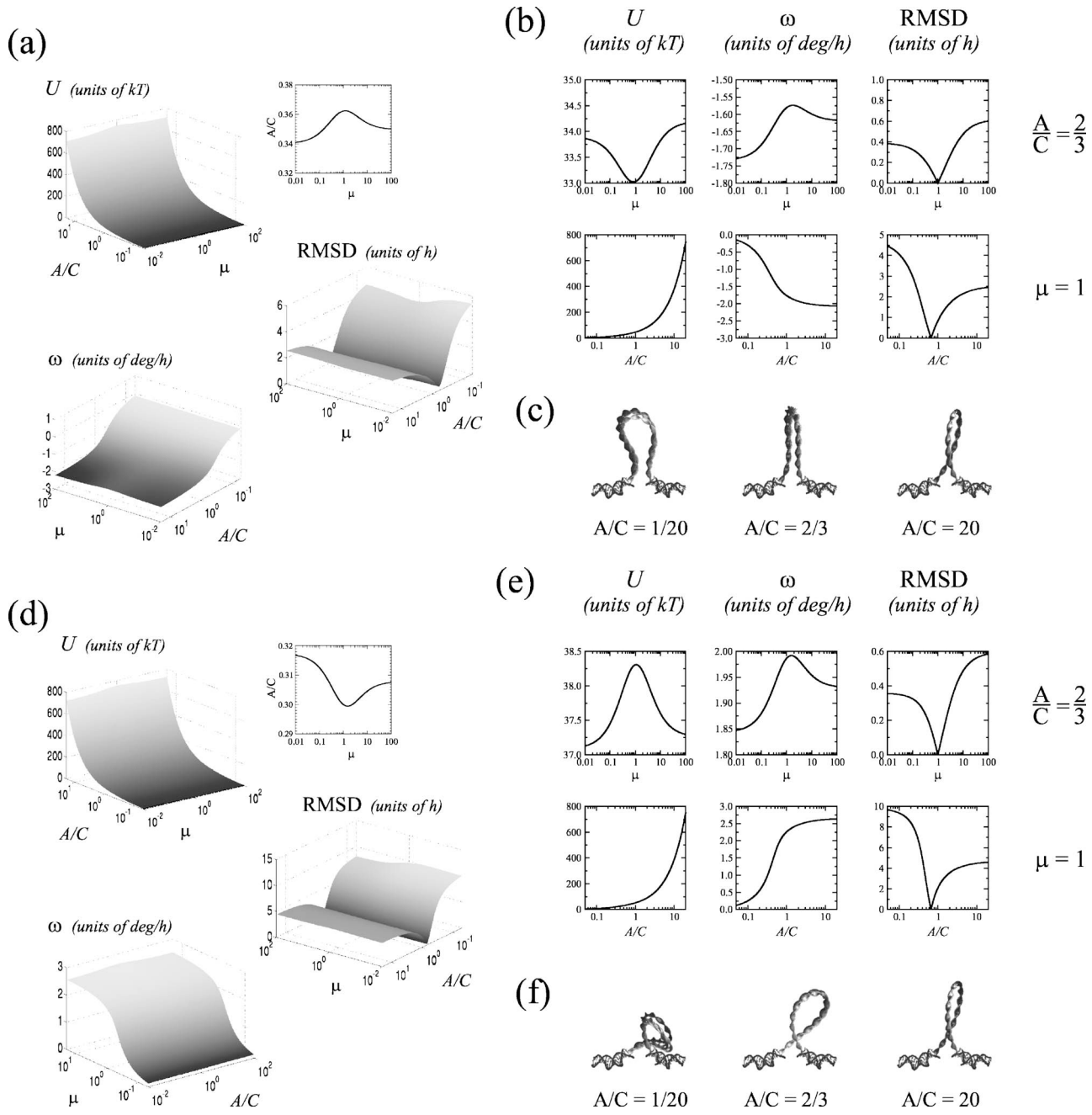

FIG. 8. Energy and geometry of the 76 bp loops in the broad range of elastic moduli α and β. The plots show the dependencies using coordinates $\mu=\alpha/\beta$ and $A/C=2\alpha\beta/(\alpha+\beta)$ (30). (a) 3D plots for the elastic energy U of the U loop, the average unwinding ω of the loop, and the rmsd of the loop centerline from that in the isotropic case $\mu=1$. The rmsd values are measured in DNA helical steps $h=3.4$ Å. To present the best view, the plots for rmsd and ω are shown from a different viewpoint than that for U. The inset shows the contour line for $U=20$ kT, which is the DNA looping energy estimated from experiment [61]. (b) Cross sections of the maps in (a) for the fixed values of $A/C=2/3$ (top row) and $\mu=1$ (bottom row). The plots illustrate the behavior of the loop structure in response to changing the rigidity (A/C) and bending anisotropy (μ) of the loop. (c) Snapshots of the loop structures for the stated values of A/C. The darkly colored loops correspond to $\mu=0.01$, and the lightly colored loops to $\mu=100$. For the U solution, the two loop structures practically overlap for all A/C values. (d)–(f) Same as (a)–(c) for the O solution. Note the more pronounced dependence of the O loop structure and energy on the bending anisotropy μ.

towards $\mathbf{d}_2$, the twist slightly decreases, extending the exposure of the "soft face" to the main bending. The oscillations of the twist cannot, however, be too large because they inflict a certain energetic penalty as well.

C. Changes to the 76 bp loops in the broad range of parameters μ and A/C

From the described specific case of $\mu=1/4$, we proceeded to studying the elastic rod conformations over a broad range

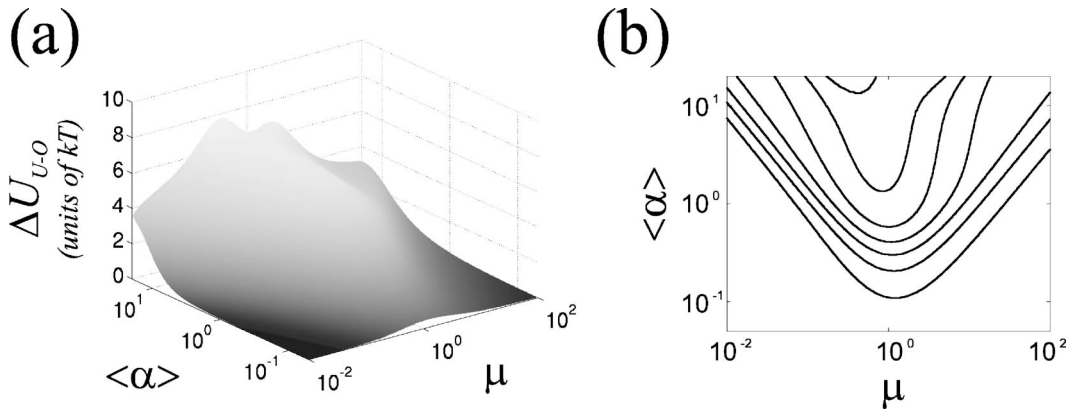

FIG. 9. (a) Dependence of the elastic energy difference $\Delta U = U_O - U_U$ between O and U solutions, plotted in coordinates $\mu = \alpha/\beta$ and $\langle\alpha\rangle = (\alpha+\beta)/2$. (b) Contours of the plot in (a) for ΔU values of $1, 2, \ldots, 7$ kT.

of parameters A/C and μ (or α and β). A/C was varied between 1/20 and 20, and μ between 10^{-2} and 10^2. In principle, such range is too broad as neither the DNA rigidity A/C significantly deviates from 1, according to all existing estimates [23,31,33,49–51,65], nor is μ likely to deviate from 1 by two orders of magnitude as the oscillations of DNA roll and tilt are normally of the same order [32]. The values of $\mu > 1$ are especially unlikely as the DNA bending towards the grooves should clearly take less energy than bending towards the backbone [cf. Fig. 1(a)]. Yet, the broad range of parameters was studied in regard to the behavior of twisted elastic rods in general.

The parameters were changed in two nested iteration cycles starting from the isotropic solutions with $\alpha = \beta = 2/3$. A/C was changed in the first cycle and μ in the second, nested cycle. The moduli α and β were then computed on each step according to the formulas above, and the new solutions were generated. The results of the computations are presented in Fig. 8.

The energy U of the elastic loops grows roughly linearly with the bending rigidity A/C. The changed A/C also means that the relative energetic cost of bending and twisting changes; as a result, the elastic loop changes its shape so as to optimally distribute the deformation between bending and twisting. When A/C is increasing, the average curvature of the loop decreases and the average unwinding/overwinding ω increases (Fig. 8). Conversely, when A/C is decreasing, the costly twisting falls to zero and the rod centerline becomes more significantly bent at every point. Yet, the rod cannot straighten itself to a line at high A/C, nor can $|\omega|$ fall lower than zero at low A/C; therefore, at some point the structure of the rod approaches an asymptotic shape and the average unwinding/overwinding and the rmsd from the initial structure become nearly constant [cf. the plots for $\mu = 1$ in Figs. 8(b) and 8(e)]. A similar effect has been observed in the studies of the bending anisotropy of a Möbius band [36].

Interestingly, the asymptotic shapes of the U and O loops are almost identical for the high bending rigidity [cf. Figs. 8(c) and 8(f) for $A/C = 20$]; the corresponding shape of the elastic loop achieves the least possible bending for the given boundary conditions. However, the twist of the two asymptotic loops is different: the total twist of the O solution is 2π less than that of the U solution. The energy difference between the asymptotic U and O loops is about 7 kT, arising mainly from the difference in the twisting energy.

Introducing bending anisotropy of the elastic rod causes the local curvature and twist to develop seesaw profiles similar to those described above, but otherwise does not have a significant effect on the loop structure and energy. Going from $\mu = 1$ to $\mu = 10^2/10^{-2}$ changes the total energy of the loop by only a few percent of its value [Figs. 8(b) and 8(e)]. The rmsd from the isotropic structures of the same rigidity never exceeded 1 h. The absolute value of the average twist increases (for the U loop) or decreases (for the O loop) by at most 10%. A more pronounced effect could be obtained if the coupling between the bending fluctuations in the two principal directions were taken into account and a relation between the effective rigidity A/C and the elastic moduli α and β different from (30) were used [60].

The experimentally measured energy of DNA looping by the *lac* repressor equals 20 kT [61]. This value can be reproduced in our computations for different sets of parameters A/C and μ, as can be seen from the contours on the μ -A/C plane that correspond to the cross sections of the 3D maps of the elastic energy for $U = 20$ kT [Figs. 8(a) and 8(d)]. Due to the observed small effect of the bending anisotropy on the energy of the elastic loops, the contours fit into a narrow A/C range near the values of 0.31–0.35, which is approximately one-half of the experimentally measured bending rigidity. Yet one has to be cautious when directly comparing the computed and observed energies since the structure of the *lac* repressor (and, therefore, the boundary conditions of the problem) might change under the stress of the bent DNA loop [18,25,26,56]. A multi-scale simulation in which the *lac* repressor structure would be allowed to change and to relax the DNA loop [25,26] is better suited for parametrizing the DNA elasticity moduli than the present simplistic calculations.

Yet, certain conclusions about the *lac* repressor loops can be made despite uncertainty about the values of α and β (or μ and A/C). The difference in the energy between the U and O loops, shown in Fig. 9, exceeds 1 kT over a wide range of α and β values, and in a significant part of the range even exceeds 3 kT. Therefore, it can be safely concluded that,

TABLE II. Geometric and energetic properties of the four solutions for the 385 bp long DNA loop, in the case of $\mu=1/4$ ($\alpha=4/15, \beta=16/15$). The rmsd values are computed with respect to the corresponding loops obtained for $\alpha=\beta=2/3$.

Solution	$\langle\omega\rangle$ (deg/bp)	$\langle\kappa\rangle$ (deg/bp)	κ_{max} (deg/bp)	U (kT)	rmsd (h)
U	−0.20	0.82	2.16	4.2	3.3
O	0.26	0.96	2.69	6.0	9.1
U′	−0.29	1.34	2.61	9.7	4.7
O′	0.34	1.38	2.66	10.5	7.6

under thermal equilibrium, the DNA loop formed by the *lac* repressor should preferably have the shape of the U solution. Incidentally, the shape of the U loop varies less for different α and β [Figs. 8(c) and 8(f)] and therefore can with more certainty be used to determine such global geometric parameters of the loop as, for example, the radius of gyration or an average protein-DNA distance.

A notable feature of the O solution is a point of near self-contact [Figs. 3, 7, and 8(f)], which becomes closer as the bending rigidity of the loop decreases [Fig. 8(f)]. If electrostatics were taken into account at this moment, this near self-contact would inflict a strong energetic penalty due to the self-repulsion of the negatively charged DNA. This happens indeed, as demonstrated in the next section. Thus, the open shape of the U loop lacking any self-contact becomes an additional argument in favor of predominance of the U loop in the ensemble of loop structures folded by the real *lac* repressor.

D. Structure of the 385 bp loops for the anisotropic values of μ and A/C

The effect of bending anisotropy on the structure and energy of the 385 bp loops is similar to that seen above for the 76 bp loops. The distributions of curvature, twist, and energy density along the four solutions developed oscillatory patterns, similar to those shown in Fig. 5; the bending concentrated in the "soft joints." For $\mu=1/4$, the solutions became more bent and less twisted on the average, as the data in Table II show (cf. Table I). The U solution was the one to undergo the least change from its isotropic shape, while the solutions O and O′ were those that changed the most.

Over the broad range of parameters α and β, the four long loops showed the same tendencies as the two short loops. The bending anisotropy introduced using (30) did not have a significant effect. Changes in the rod rigidity eventually brought the loops to asymptotic states. The asymptotic states for the loops with a small A/C ratio were strongly bent conformations with practically zero unwinding/overwinding, where the twisting energy was of the same order as the small bending energy. The asymptotic states for the loops with the large A/C ratio were the conformations with the least possible bending, where the twist achieved the highest possible value to compensate for the energetically costly bending, yet the bending still accounted for most of the elastic energy.

The underwound U solution was again the one to have the lowest elastic energy among the four solutions throughout the whole studied range of α and β values. The maps of the energy difference between U and the other three solutions are shown in Fig. 10. The energy of the O solution does not normally differ from that of the U solution by more than several kT; therefore, the O solution should contribute to a small extent to the thermodynamic ensemble of the loop structures when the *lac* repressor binds a 385 bp DNA loop. In contrast, the energies of the U′ and O′ loops are consistently 2–2.5 times larger than the energy of the U loop; the energy difference is small only for unlikely values of α and β. Therefore, one can safely conclude that these two loops, even though uninhibited by steric overlap with the *lac* repressor, are still extraneous solutions, as in the case of the 76 bp loop, and may be excluded from any multi-scale computation relying on the properties of the thermodynamic ensemble of the 385 bp loop conformations.

V. ELECTROSTATIC EFFECTS

The last but, perhaps, the most important extension of the classic theory employed here consists in "charging" the modeled DNA molecule. The phosphate groups of a real DNA carry a substantial electric charge, $-20.8e$ per helical turn, that significantly influences the conformational properties of DNA [20,27–29,66]. The DNA experiences strong self-repulsion that stiffens the double helix and increases the distance of separation at the points of near self-contact [67]. Also, all electrostatically charged objects in the vicinity of a DNA, such as amino acids of a bound protein or lipids of a nearby nuclear membrane, interact with the DNA charges and influence the DNA conformations. Below, we describe our model of the electrostatic properties of DNA and the effects of electrostatics on the conformation of the *lac* repressor DNA loops.

A. Changes to the equations of elasticity due to electrostatics

The electrostatic interactions of DNA with itself and any surrounding charges are introduced through body forces $\mathbf{f}$, such that

$$\dot{\mathbf{f}}(s) = \sigma(s)\mathbf{E}(\mathbf{r}(s)), \tag{31}$$

where $\mathbf{E}$ is the electric field at the point $\mathbf{r}(s)$ and $\sigma(s)$ is the density of DNA electric charge at the point s. The present simplified treatment places the DNA charges on the centerline, as was also done in other studies [27,42].

The charge density $\sigma(s)$ is modeled as a smooth differentiable function with relatively sharp maxima between the DNA base pairs, where the phosphate charges are located. The chosen (dimensionless) function

$$\sigma(s) = \tfrac{8}{3} Q_{\mathrm{DNA}} \sin^4(\pi N_{\mathrm{DNA}} s) \tag{32}$$

is somewhat arbitrary, but specifics are unlikely to significantly influence the results of our computations, as discussed

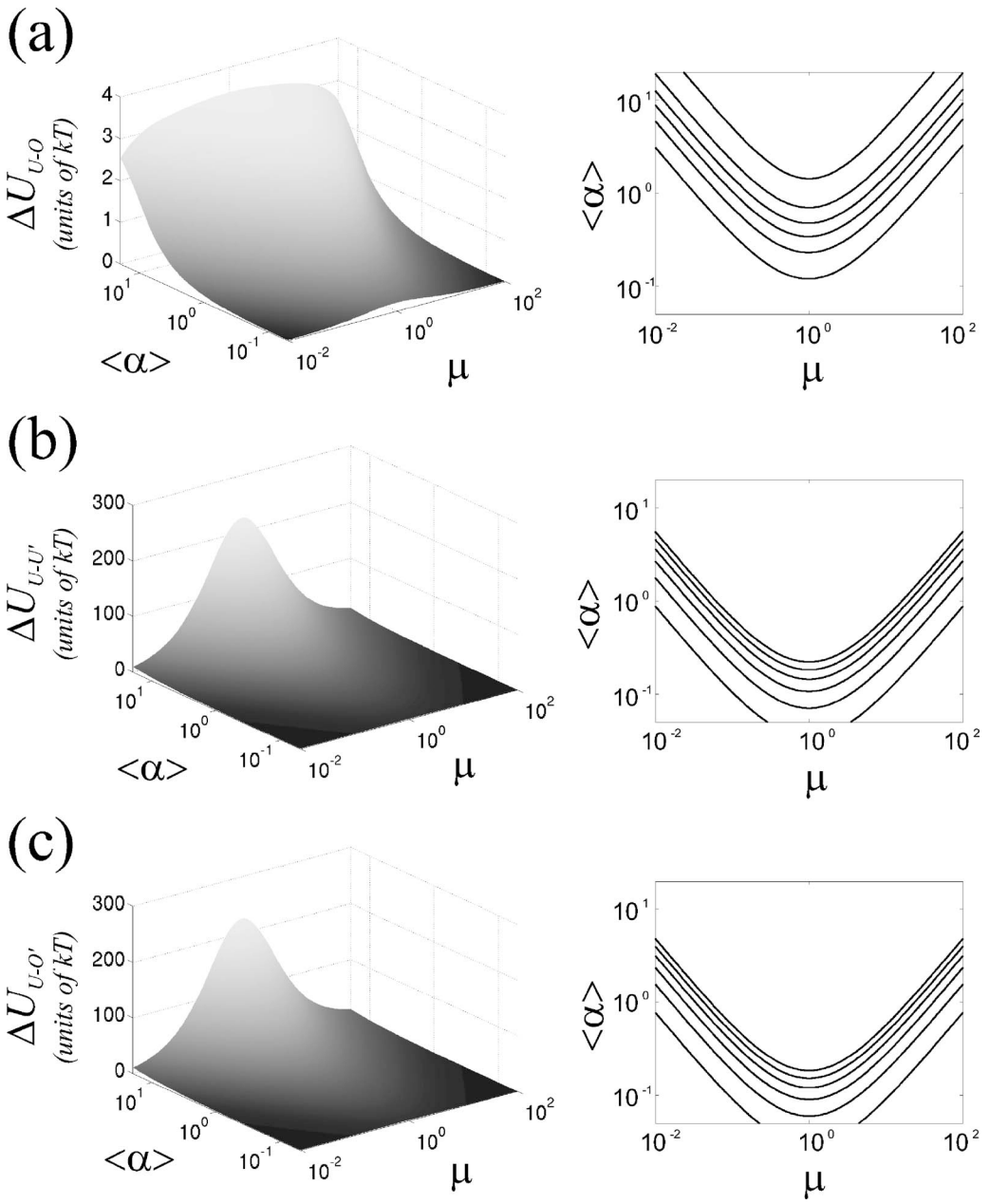

FIG. 10. Elastic energy difference between U and (a) O, (b) U′, and (c) O′ solutions for the 385 bp long DNA loop. 3D plots of the energy difference in the coordinates $\mu=\alpha/\beta$ and $\langle\alpha\rangle=(\alpha+\beta)/2$ are shown on the left, and contour maps of the 3D plots for the ΔU values of 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 kT are shown on the right.

below. N_{DNA} denotes the number of base pairs in the modeled DNA loop (which is assumed to begin and end with a base pair) and Q_{DNA} denotes the total charge of the DNA loop. That charge is reduced from its regular value of $2e$ per base pair due to Manning counterion condensation around the phosphates [66]: $Q_{\mathrm{DNA}}=2e\chi N_{\mathrm{DNA}}$. Here, we use the value of $\chi=0.25$, valid for a broad range of salt concentrations [66].

The electric field $\mathbf{E}$ is composed of the field of external electric charges, not associated with the modeled DNA loop, and from the field of the loop itself (Fig. 11). $\mathbf{E}$ is computed using the Debye screening formula

$$\mathbf{E}(\mathbf{r}(s)) = \frac{1}{4\pi\epsilon\epsilon_\circ}\left(\sum_i q_i \nabla \frac{\exp(-|\Delta\mathbf{r}_i(s)|/\lambda)}{|\Delta\mathbf{r}_i(s)|} + 2e\chi\sum_j{}' \nabla \frac{\exp(-|\Delta\mathbf{r}_j(s)|/\lambda)}{|\Delta\mathbf{r}_j(s)|}\right), \tag{33}$$

where $\Delta\mathbf{r}_i(s)=\mathbf{r}(s)-\mathbf{R}_i$, $\Delta\mathbf{r}_j(s)=\mathbf{r}(s)-\mathbf{r}(s_j)$, $\lambda=3\ \text{Å}/\sqrt{c_s}$ is the radius of Debye screening in an aqueous solution of monovalent ionic strength c_s at 25 °C [66], and $\epsilon=80$ is the dielectric permittivity of water.

The first term in Eq. (33) represents the DNA interaction with external charges q_i located at the points $\mathbf{R}_i$; the sum

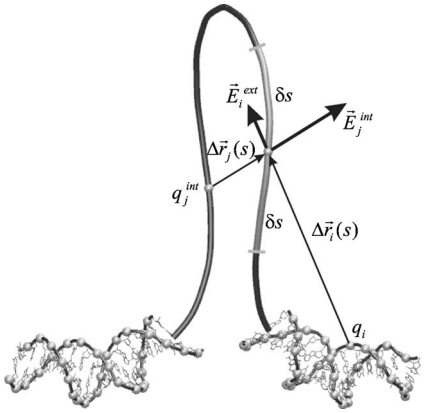

FIG. 11. Electrostatic interactions in the elastic rod problem. The electric field $\mathbf{E}(s)$ is computed at each point s of the rod as the sum of the "external" field $\mathbf{E}^{ext}$, produced by the charges q_i not associated with the elastic rod, and the "internal" field $\mathbf{E}^{int}$, produced by the charges $q_j^{int}=2e\chi$ placed in the maxima of the charge density $\sigma(s)$ of the elastic rod (32). $\Delta\mathbf{r}_i(s)=\mathbf{r}(s)-\mathbf{R}_i$ and $\Delta\mathbf{r}_j(s)=\mathbf{r}(s)-\mathbf{r}(s_j)$ [cf. Eq. (33)]. The part of the rod that lies within the cutoff δs does not contribute to $\mathbf{E}(s)$.

runs over all those charges. The second term represents the self-repulsion of the DNA loop and involves a sum over all the maxima of the charge density $\sigma(s)$, where the DNA phosphates, i.e., charges of $2e\chi$, are located. This sum approximates an integral over the charged elastic rod. Computing such an integral would be more consistent with the continuous $\sigma(s)$; however, the suggested discretization is rather accurate, as shown below, and reduces significantly the cost of the computation.

More importantly, certain phosphate charges are excluded from the summation in the second term (hence the prime sign next to the sum), namely, the charges located closer to the point s than a certain cutoff distance δs (Fig. 11). Similar electrostatic cutoffs were used in other studies [27,29,42]. The reason for introducing such a cutoff is that the DNA elasticity has partially electrostatic origin, so that the energetic penalty for DNA bending and twisting, approximated by the elastic functional (29), already includes the contribution from the electrostatic repulsion between neighboring DNA charges. It is debatable what "neighboring" implies here, i.e., how close should two DNA phosphates be in order to significantly contribute to DNA elasticity. In this work, the cutoff distance δs is chosen to be equal to the pitch H of the DNA helix ($H=36$ Å). This, on the one hand, is the size of the smallest structural unit of DNA, beyond which it does not make sense at all to use a continuum model of the double helix, so the phosphate pairs within such unit are naturally excluded from the explicit electrostatic term. On the other hand, the forces of interaction between the phosphates, separated by more than that distance from each other, are already much smaller than the elastic force, as shown below. Thus, even though the chosen cutoff δs might be too small, the resulting concomitant stiffening of the DNA is negligible. For comparison, calculations with cutoff values $\delta s=1.5H$ and $\delta s=2H$ were also performed.

Thus, the electric field $\mathbf{E}$, computed using (33), is substituted into (31), and the resulting body forces $\mathbf{f}$ appear in Eqs. (18), (19), and (21) in place of the previously zeroed terms [68]. The solutions of the modified equations minimize the energy functional

$$U = U_{elastic} + U_Q - U_{Q,straight}, \tag{34}$$

where $U_{elastic}$ is the elastic energy computed as in (29), U_Q is the electrostatic energy computed, in accordance with (33), as

$$U_Q = \frac{1}{4\pi\epsilon\epsilon_\circ}\int_0^1 \sigma(s)\left(\sum_i q_i \frac{\exp(-|\Delta\mathbf{r}_i(s)|/\lambda)}{|\Delta\mathbf{r}_i(s)|} + 2e\chi{\sum_j}' \frac{\exp(-|\Delta\mathbf{r}_j(s)|/\lambda)}{|\Delta\mathbf{r}_j(s)|}\right)ds, \tag{35}$$

and $U_{Q,straight}$ is the electrostatic "ground state" energy computed using (35) for a straight DNA segment of the same length as the studied loop. The continuous charge density $\sigma(s)$ used here is computed using (32).

B. Changes to the computational algorithm

The electrostatic forces introduced above depend on the conformation of the entire elastic loop due to the self-repulsion term in (33). This turns the previously ordinary differential equations of elasticity into integro-differential equations, and a new algorithm is required for solving them, implemented as follows.

We start with a solution built for an uncharged elastic loop and turn on the electrostatic interactions during yet another iteration cycle. On each step of that cycle, Eqs. (18)–(28) are solved with the electrostatic term (31) computed for the electric field $\mathbf{E}_i=w_{E,i}\mathbf{E}$, where the "electrostatic weight" w_E grows linearly from 0 to 1 during the cycle. Additionally, each step of the new iteration cycle becomes its own iterative subcycle. The electric field $\mathbf{E}_i$ is computed at the beginning of the subcycle and the equations are solved with this constant field. Then the field is recomputed for the new conformation of the elastic rod, the equations are solved again for the new field, and so on until convergence of the rod to a permanent conformation (and, consequently, of the field to a permanent value) is achieved. The weight w_E is kept constant throughout the subcycle. To enforce convergence, the field used in each round of the subcycle is weight averaged with that used in the previous round:

$$\mathbf{E}_{i,j} = w_a\mathbf{E}_{i,j(actual)} + (1-w_a)\mathbf{E}_{i,j-1}. \tag{36}$$

The averaging weight w_a is selected by trial and error so as to speed up convergence. For the *lac* repressor system, the trivial choice of $w_a=0.5$ turned out to be satisfactory.

Our approach to solving the integro-differential equations assumes that the elastic rod conformation changes smoothly with the growth of the electric field. For intricate rod conformations, which might change in a complicated manner with the addition of even small electrostatic forces, this approach may conceivably fail. Yet, it worked well for the studied case of the DNA loop clamped by the *lac* repressor and should presumably work well for simple elastic rod conformations corresponding to the lowest energy minima of short DNA loops.

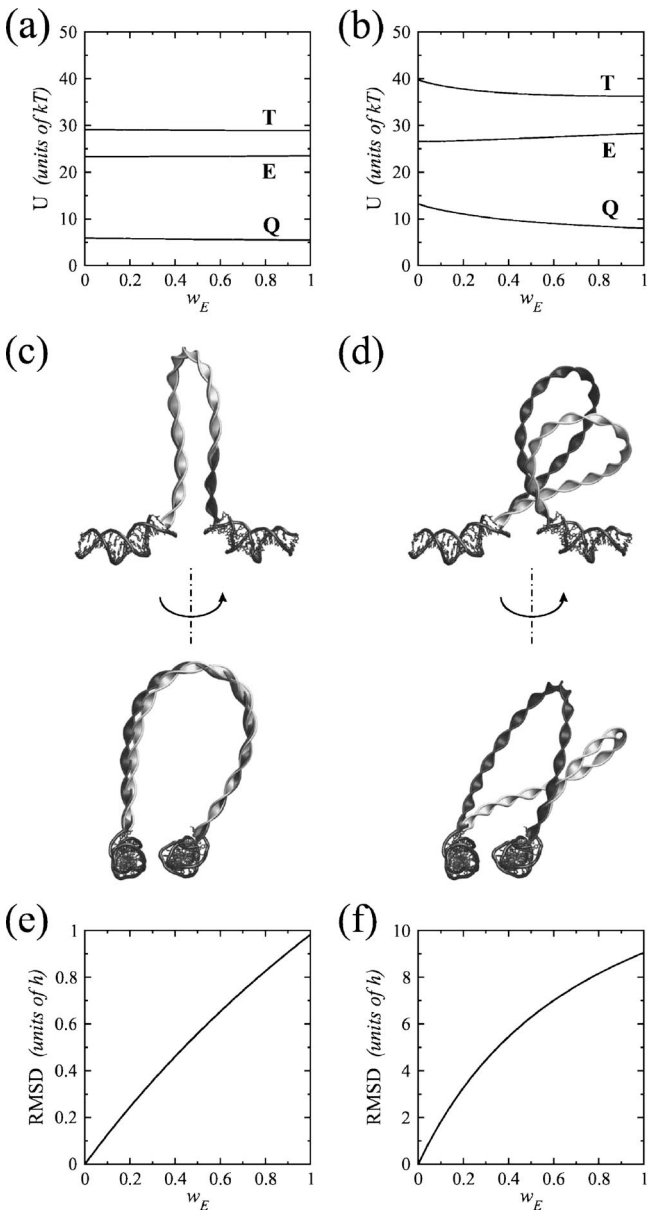

FIG. 12. Changes in structure and energy of the 76 bp DNA loops after electrostatic interactions are taken into account. Left column: U solution; right column: O solution. (a), (b) Elastic (E), electrostatic (Q), and total (T) energy of the modeled loop versus the electrostatic weight w_E. (c), (d) Uncharged ($w_E=0$, in light color) and completely charged ($w_E=1$, in dark color) structures of the loops. The bottom views are rotated by 70° around the vertical axis relative to the top views. (e), (f) The rmsd between the charged and uncharged loop structures. The data correspond to $\alpha=4/15$, $\beta=16/15$, ionic strength $c_s=10$ mM, and the electrostatics exclusion radius $\delta s=H$.

C. Electrostatic effects for the 76 bp loops

Equations (18)–(28) with the electrostatic term were solved for the 76 bp DNA loops for the ionic strength c_s in the range of 0–100 mM and three different cutoff values δs (H, $1.5H$, $2H$). The computations were performed with the previously used elastic moduli $\alpha=4/15$ and $\beta=16/15$. The external charges included in the model were those associated with the phosphates of the DNA segments from the crystal structure [37] (see Fig. 11). The iteration cycle was divided into 100 subcycles, which showed a remarkable convergence: the length of a subcycle never exceeded three iteration rounds.

The changes in structure and energy of the elastic loops due to electrostatic interactions are presented in Fig. 12 for an ionic strength of 10 mM and the exclusion radius of H. The structure of the U solution exhibits little change: the rmsd between the original ($w_E=0$) and the final ($w_E=1$) structures is less than 1 h. Neither the curvature nor the twist profiles of this loop change significantly (Fig. 5, third column). The energy of the loop changes by the electrostatic contribution of 6.1 kT, mainly accounted for by the interaction of the loop termini with each other and with the external DNA segments. The self-repulsion and the repulsion from the external DNA charges contribute about equally to the electrostatic energy.

In contrast, the near self-contact in the O structure of the loop [Figs. 7 and 12(d)] causes a significant change in the O loop structure and energy when electrostatics is turned on [Figs. 12(b), 12(d), and 12(f)]. The structure opens up, the separation at the point of the near self-crossing increases, the rmsd between the final and the original structures reaches 9 h [Fig. 12(f)], and the DNA overwinding almost doubles (Fig. 5, right column of the bottom panel). This allows the electrostatic energy to drop from 13.2 to 8.0 kT, yet the elastic energy grows by 1.7 kT [Fig. 12(b)]; the total energy of the charged O loop reaches 36.3 kT so that the energy difference from the U loop increases from 3.3 to 7.4 kT. As in the case of the U loop, the main contribution to the electrostatic interactions comes from the loop ends; the energy distribution between the self-repulsion and the repulsion from the external DNA charges is roughly equal.

Naturally, the calculated effect diminishes when the ionic strength of the solution increases and the electrostatics becomes better screened. Figure 13 shows structure and energy of the U and O loops when the ionic strength is increased from 10 to 100 mM (this range covers physiological ionic strengths). Structure and elastic energy of the U loop show almost no change again; the total energy of the loop decreases from 29.7 to 23.5 kT due to the drop in the electrostatic energy. The structure of the O loop returns to almost what it was before the electrostatics was computed (within the rmsd of 2.2 h); the elastic energy of the loop drops back to 26.6 kT and the electrostatic energy to a mere 0.5 kT. These results show that theoretical estimates of the energy of a DNA loop formation *in vivo* need to employ as good an estimate of the ionic strength conditions as possible.

The *lac* repressor loops were extensively used to analyze all the assumptions and approximations of our model and showed that those were satisfactory indeed. The calculations were repeated for the self-repulsion cutoffs of $\delta s=1.5H$ and $\delta s=2H$. The resulting changes in the loop energy at $c_s=10$ mM equal to only $\Delta U_{h-1.5h}=0.35$ kT and $\Delta U_{h-2h}=0.75$ kT for either loop; these values drop below 0.1 kT when the ionic strength rises to 100 mM. The difference lies mainly in the electrostatic energy, and the elastic energy is always within 0.1 kT of that of the structures obtained with $\delta s=H$. Accordingly, the rmsd from the uncharged structure

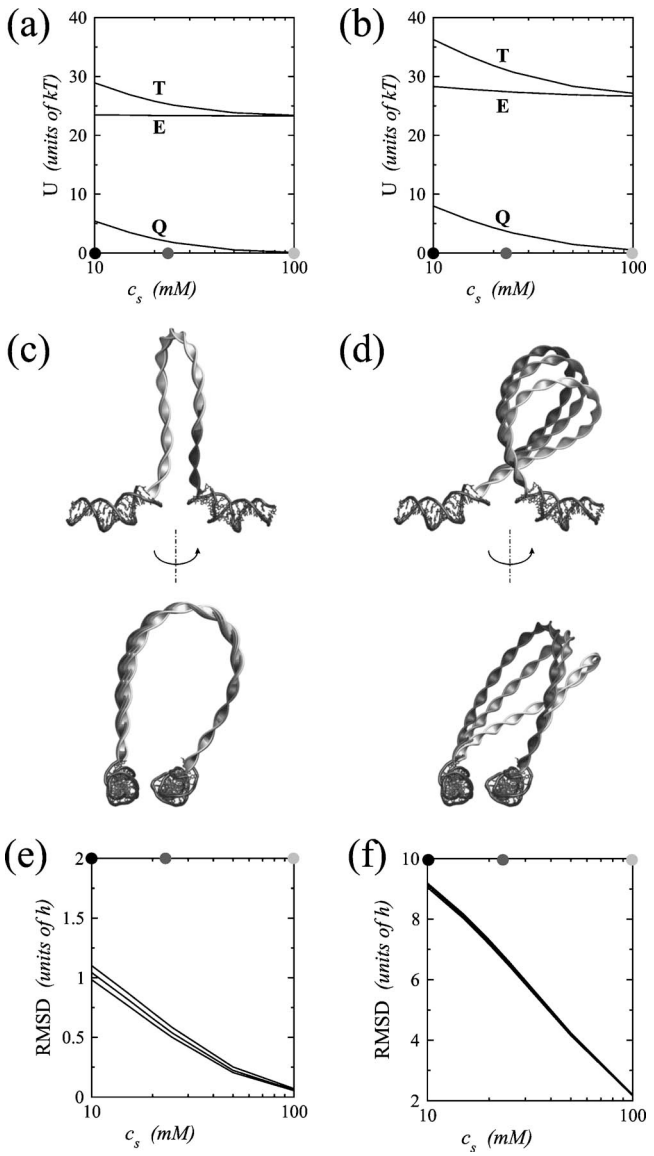

FIG. 13. The effect of salt on the structure and energy of the 76 bp DNA loops. Left column: U solution; right column: O solution. (a), (b) Elastic (E), electrostatic (Q), and total (T) energy of the elastic loop versus the ionic strength c_s. Shown are only the plots for the exclusion radius of H; the energy plots for the other exclusion radii are almost indistinguishable. (c), (d) Snapshots of the elastic loop structures for 10 mM (dark color), 25 mM (medium color), and 100 mM (light color). Points where the snapshots were taken are shown as dots of the corresponding colors on the axes of panels (a), (b), (e), (f). The bottom views are rotated by 70° around the vertical axis relative to the top views. (e), (f) The rmsd of the loop structures from those computed without electrostatics (equivalent to infinitely high salt concentration). The lines, from top to bottom, correspond to the exclusion radii of H, $1.5H$, and $2H$.

varies by at most 0.1 h for the different cutoffs [Figs. 13(e) and 13(f)]. Therefore, even the largest cutoff $\delta s=2H$ works satisfactorily for the electrostatic calculations, while at the same time increasing the speed of computations.

The force of electrostatic interactions did not exceed 1–2 pN per base pair, compared to the calculated elastic force in the range of 10–20 pN, and depended very little on the cutoff. Therefore, the conceivable stiffening of the rod due to possibly overcounting phosphate pairs in the electrostatic calculations is negligible. Evaluating the electric field and energy using the sums (33) and (35) instead of a more consistent integral over the loop centerline results in no significant error either. Test calculations showed that in all the studied cases the electrostatic field and energy evaluated by the integral and by the discrete sum differ by at most 2×10^{-4} of their values.

Finally, it was tested in how far the particular choice (32) for the charge density of DNA, $\sigma(s)$, influences the computational results. The calculations were repeated for the constant charge density $\sigma'(s)=Q_{\mathrm{DNA}}$ (in dimensionless representation). The energies of the loop conformation never changed by more than 10^{-4} of their values over the whole range of c_s and δs; the rmsd between the loop conformations obtained with different $\sigma(s)$ never exceeded 0.01 h. Therefore, the electrostatic properties of the elastic rod in the current model can safely be computed with constant electrostatic density, further saving computational costs.

D. Electrostatic effects for the 385 bp loops

The electrostatic computations were similarly performed for the 385 bp loops, in the same range of ionic strength and for the exclusion radii of $2H$ and $3H$. For the U and O loops, the results were qualitatively the same as in the case of the short loops. The loops became more open and straightened up; their energy increased by 0–6 kT, depending on the ionic strength (Fig. 14, cf. Table II). The U loop was again the one to change its structure and energy to the least extent upon turning on the electrostatics. The results of the computations were practically the same for both cutoff radii, were not influenced by the electric field discretization (33), and were not considerably changed by replacing the charge density function $\sigma(s)$ from Eq. (32) with the constant function $\sigma'(s)$.

Structurally, the long loops changed more significantly upon introducing the electrostatic repulsion than the short loops did. The rmsd values reached 10 h for the U loop and 25 h for the O loop [Fig. 14, cf. Figs. 13(e) and 13(f)]. As before, the major part of electrostatic repulsion came from the ends of the loop, including the protein-bound DNA segments, brought closely together by the protein. This repulsion tended to change the direction of the ends of the loop, bending them away from each other. In the case of the short loops, it was impossible to noticeably change the direction of the ends without significantly stressing the rest of the loop. Yet the long loops could more easily accommodate some opening up at the ends and, therefore, changed their structures more significantly.

The larger structural change necessitated longer calculations. For the long loops, the iteration steps typically consisted of five to six iteration subcycles, and even of a few dozen subcycles at especially stiff steps.

The U′ and O′ loops showed a similar response to the electrostatics at high ionic strength (above 25 mM). Their electrostatic energy was in the range 0–5 kT, and the rmsd from their uncharged structures amounted to up to 10 h for

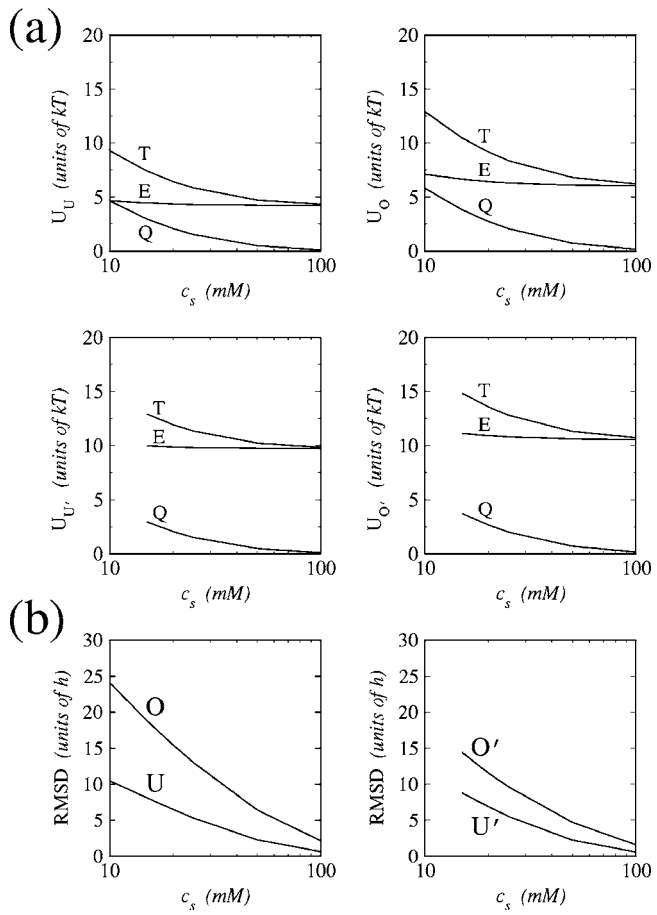

FIG. 14. The effect of ionic strength on structure and energy of the 385 bp DNA loops. (a) Elastic (E), electrostatic (Q), and total (T) energy of the U, O, U′, and O′ loops plotted versus the ionic strength c_s. (b) The rmsd of the loops from their uncharged structures. The missing points in the U′/O′ plot indicate solutions, missing due to nonconvergence of the iterative procedure (see text). The data correspond to $\delta s=3H$.

the U′ solution and up to 15 h for the O′ solution (Fig. 14). At low ionic strength, the stronger electrostatics rendered the solutions unstable: the U′ solution transformed into the U solution, and the O′ solution into the O solution, when computed with no salt screening (0 mM). What triggered the switch was apparently the ever increasing bending of the ends of the loops away from each other that, in combination with the bending anisotropy, also caused high twist oscillations near the loop ends (as described in Sec. IV B). The combination of twisting and bending caused the loops to flip up—as one can make a piece of wire flip up and down by twisting its ends between one's fingers.

For the intermediate values of ionic strength (10–20 mM), the iterative procedure for the U′ and O′ loops did not converge, caught in the oscillations between the up and down states of each loop. In this case, using the larger electrostatic exclusion radius improved convergence. With $\delta s=2H$, the iterations did not converge for the ionic strength of 15 and 20 mM; convergence for 10 mM was achieved, but resulted in flipping to the stable solutions. With $\delta s=3H$, the iterations successfully converged to U′ and O′ solutions (albeit somewhat changed due to the electrostatics) for the 15 and 20 mM ionic strength and did not converge for 10 mM.

Such instability of the U′ and O′ loops serves, of course, as another argument for disregarding them in favor of the stable U and O solutions.

Upon introducing the electrostatic self-repulsion, an interesting experiment could be performed. Self-crossing by the solutions during the iteration cycles, described in Sec. III B, was no longer possible. Therefore, one could explore whether superhelical structures of the loop could be built by further twisting the ends of the loop. One extra turn of the cross section at the $s=1$ end did generate new structures of the U and O loops. Yet, those structures were so stressed and had such a high energy (on the order of 50 kT higher than their predecessors) that it was obvious that those structures do not play any part at all in the realistic thermodynamic ensemble of the *lac* repressor loops. Any further twisting of the ends resulted in nonconvergence of the iterative procedure. Clearly, the 385 bp length of a DNA loop is insufficient to produce a rich spectrum of superhelical structures.

VI. DISCUSSION

In this section, we summarize and analyze the presented elastic rod model of DNA, describe its present and possible future applications, and, finally, summarize what has been learned about the *lac* repressor-DNA complex from the presented case study.

A. Summary of the model

The presented modeling method allows one to construct a numerical model of a DNA loop with fixed ends. The model approximates the DNA as an electrically charged elastic rod. The equilibrium conformations of the rod are computed by solving the modified Kirchhoff equations of elasticity with the boundary conditions obtained from the known positions and orientations of the ends of the loop. The numerical solutions to the equations yield the detailed geometric properties of the loop (the centerline coordinates, the curvature, and twist at each point along the loop) and the elastic stress and torque profiles along the loop. From these data, one can calculate further properties of the loop, such as the linking number and the radius of gyration, the electric field around the loop, distances between various sites of special interest on the loop, etc. The energy of the loop is estimated from the functional (29) or (34). The solutions to the Kirchhoff equations represent the equilibrium conformations of the rod that minimize the energy functionals [69], i.e., the zero-temperature structures of the DNA loop, the equilibrium points around which the loop fluctuates at a finite temperature.

The solutions to Kirchhoff equations are generated from previous or otherwise known solutions, e.g., those built for different parameters or boundary conditions, using the continuation procedure described above. A family of topologically different loop structures can be generated by varying the initial loop conformations or by subjecting the obtained

solutions to simple geometrical transformations, such as twisting and rotating the ends of the loop. The energy analysis of the obtained conformations allows one to focus on the lowest-energy structures of the loop, which presumably dominate in the Boltzmann ensemble of conformations of the real DNA loop.

The presented model advances the existing analytical and computational DNA models based on the theory of elasticity [17,22–24,42,55,62] by adding more versatility and flexibility to the extended Kirchhoff equations. The equations derived here describe DNA as an anisotropically bendable elastic rod with intrinsic twist and curvature, carrying electric charge and possibly interacting with external forces and torques [35]. The model parameters are considered to be the functions of the rod arclength, thus providing the means for modeling sequence-specific properties of DNA. Most of the earlier models excluded at least one of these DNA properties. Finally, the proposed numerical algorithm allows one to solve the extended Kirchhoff equations fast for an arbitrary set of boundary conditions, which is important during the multi-scale simulations when the assumed boundaries of the loop fluctuate randomly.

B. Analysis of the model

The case study of the DNA loops clamped by the *lac* repressor allowed us to analyze the importance of including the various physical properties of DNA in the model. It has been shown that properly accounting for anisotropic bendability of the elastic rod in combination with the intrinsic twist is important for the correct estimate of twist and bend distribution along the DNA loop. In particular, the seesaw pattern of the curvature distribution and the corresponding fluctuations of the twist are not observed for the isotropically bendable loop (cf. Fig. 5). The estimate of the overall energy and shape of the loop proved to be less sensitive to the bending anisotropy, at least when the model employs the classical relation (30) between the principal bending moduli A_1 and A_2 and the effective bending modulus A used traditionally to fit the experimental data.

It has been further shown that electrostatic interactions are screened out by physiological salt concentrations for the open conformations of the modeled loop and do not affect the shape of the loop, although they do contribute to some extent to the loop energy. In contrast, the loop conformations with near self-contacts are strongly affected by electrostatics, which therefore must be accounted for when dealing with such conformations. The effect of sequence-dependent elastic properties of DNA has not been studied here; other studies suggest that such an effect may be significant [23,31,33,53]. The sequence-dependent properties can be introduced into Eqs. (18)–(28) via parameter-functions of arclength $\alpha(s)$, $\beta(s)$, $\gamma(s)$, $\overset{\circ}{\kappa}_{1,2}(s)$, $\overset{\circ}{\omega}(s)$ constructed for each specific loop sequence using available experimental data [23,32,33,48] (cf. Appendix C).

Thus, the complexity of the more universal model can be adjusted for each specific problem regarding a DNA loop. For example, the simplification of an isotropically bendable elastic rod with constant elastic moduli should suffice for determining the global properties of the loop, such as its centerline geometry, its linking number, its radius of gyration, etc., with sufficient accuracy (cf. Figs. 3, 7, 12, and 13). On the other hand, if the local structure in a certain section of the loop needs to be predicted, e.g., with a goal to study how the formation of the loop changes the binding properties of a certain protein in that area [70], then the anisotropic loop model becomes essential, as has been demonstrated in Sec. IV. Finally, electrostatic interactions need to be used in the model for correctly predicting the structure and energy of DNA loops that exhibit either a close self-contact or a close contact with external electric charges involved in the modeled complex. Such conformations are exemplified by the O loop described here or by the tightly wound superhelical structures [10,13,15,20,24,27,42]. Electrostatics is also essential for studying the effect of salt concentration on the structure and energetics of DNA loops [20,27].

The parameter-functions included in the model, such as the elastic moduli and the intrinsic curvature/twist, in some cases can significantly influence the solution, as has been illustrated and discussed above. The simplified model, relying on the commonly accepted isotropic elastic moduli, is good for loop energy estimates within several kT and for selecting the best of the alternative DNA loop topologies. But quantitative predictions of more subtle structural and energetic properties of DNA loops, such as the local loop structure or the sequence-specific loop variations, require further refinement and adjustment of the model parameters on the basis of experimental data. Since some experimental results, such as the loop bending energies, can be reproduced by different combinations of the model parameters (cf. Figs. 9 and 10), the data for the model refinement should come from a variety of sources: from DNA micromanipulation experiments [50,52,65,71], from the analysis of DNA x-ray structures [32,33], from experiments on DNA interaction with DNA-binding proteins, including topoisomerases [72], and from the energetics of DNA minicircles [23,73].

The modeling method is best applied to DNA loops in the range of 50 to several hundred bp. The length of shorter DNA segments becomes comparable to their diameter, thus making the Kirchhoff rod approximation invalid. The longer DNA segments possess large families of structures [15,20,24] all of which may not be discovered by our simplified search procedure. Good representative structures of long loops can certainly be generated, but other structures of comparable low energy are likely to be missed. For the loops of intermediate lengths, like the ones folded by the *lac* repressor, only a few topologically different structures of comparable energy can exist and our simplified search of the conformational space should be sufficient for discovering all the members of the topological ensemble.

Since our method produces only zero-temperature structures of the loop, limiting the analysis to only these structures results in neglecting the thermal vibrations of the loop and the related entropic effects. For DNA loops not exceeding one to two persistence lengths, the thermal vibrations are likely to be small, and the entropic effects insignificant. For longer loops, the entropic effects may be larger and conformational sampling studies may become necessary, for which the generated equilibrium loop structures render excellent

starting points. A cheaper *ad hoc* way of accounting for the thermal oscillations consists in interpreting the loop properties, e.g., its energy or the forces of protein-DNA interaction, as a series of Gaussian random values distributed around the average values calculated for the equilibrium loop structure [25].

Building the family of the lowest-energy conformations of a DNA loop using the suggested numerical algorithm is a fast process that takes only several hours of computation on a single workstation if done from scratch. If done on the basis of a good initial guess for each loop conformation, e.g., the same conformation previously obtained for a slightly different set of parameters or boundary conditions, then generating the loop structures takes mere minutes. Of course, such approach works well for short loops only—for longer loops with numerous possible structures, an extensive conformational sampling is indispensable.

It should be noted that, while simple and fast, our continuation procedure may sometimes fail due to the nonlinear nature of Kirchhoff equations. The solutions to the equations are known to respond nontrivially to changes in the problem parameters or boundary conditions [16,20,24,62]. If a point of instability is reached, the solution will precipitously change in response to even a small change in the parameter, resulting in a failure of the numerical BVP solver. The solver will also fail when the evolving solution reaches a bifurcation point [22,24], as did the U′ and O′ solutions at low ionic strength (Sec. V D).

If such problems are encountered, it is recommended to thoroughly analyze the nature of the nonconvergence. A point of abrupt transition can sometimes be circumvented by choosing an alternative pathway through the conformational space. For instance, if rotating one end of the loop clockwise by a certain angle ϕ meets a transition point half-way through, then that end might instead be rotated by 2π-ϕ counter-clockwise, reaching the desired endpoint without going through the abrupt transition. In the case of a bifurcation, an approximate solution on each of the bifurcating branches can be guessed from the intermediate elastic rod structures produced by the nonconverging numerical solver. Then the continuation procedure can be reinitiated for each branch several steps away from the bifurcation point, starting from the guessed solutions.

C. Applications of the model

The speed and high adaptability of our modeling approach makes it a good candidate for multi-scale modeling simulations of protein-DNA complexes. The most straightforward application consists in studying the structure and dynamics of a multi-protein aggregate that clamps a long DNA loop [3–5,13,36]. Such a study has been conducted recently for the *lac* repressor-DNA complex [25,26]. The elastic rod model of the 76 bp DNA loop, built as described here, yielded the forces and torques that the loop exerted on the *lac* repressor. These forces and torques were included in the MD simulation of the all-atom structure of the *lac* repressor complex with the bound DNA segments, similarly to steered molecular dynamics simulations [74]. The forces were used either directly or served as a basis for generating a time series of random forces [26]. As the structure of the protein-DNA complex changed during the simulation, the boundary conditions were updated using the most recent coordinates of the protein-bound DNA segments and the Kirchhoff equations were re-solved yielding new values of the forces and torques to be used in the simulation. The multi-scale simulation provided a dynamic picture of the *lac* repressor complex with the DNA loop, revealing the main degrees of freedom in the *lac* repressor structural dynamics and suggesting a "protein lock" mechanism of the *lac* repressor cleft opening under the stress of the DNA loop (see [26,25] for further detail).

The elastic rod modeling method provides also a fast and convenient tool for mimicking the effect of protein binding within DNA loops. Another recent study [70] described the binding of the catabolite gene activator protein (CAP) [1,2,75] to a DNA site located within the loop folded by the *lac* repressor. The specific geometry that CAP enforces in DNA upon binding (two $\sim 45°$ kinks separated by 10 bp of straight DNA) was mimicked in the corresponding segment of the DNA loop by accordingly designed intrinsic curvature and twist parameter-functions $\overset{\circ}{\kappa}_{1,2}(s)$, $\omega^{\circ}(s)$ [70]. It was discovered that CAP binding switches the preferred loop topology from the U to the O structure; the resulting drop in the free energy of the loop explains the experimentally observed cooperativity in DNA binding between CAP and *lac* repressor [76]. Fitting an all-atom structure of CAP inside the constructed DNA loop and replacing the elastic rod DNA scaffold with an all-atom DNA structure (cf. Appendix D) resulted in an all-atom model of the whole ternary complex between CAP, *lac* repressor, and DNA, ready for further all-atom or multi-scale simulations.

All-atom structures of whole DNA loops can be built on the basis of the calculated elastic rod conformations of the loops as described in Appendix D and illustrated in Figs. 2(c) and 2(d). Knowing the all-atom structure at any point of the loop helps one to address many interesting issues. For example, if a binding site of a regulatory protein, such as CAP, is located within the loop, then the change in the structure of that site upon loop formation is readily assessed and the influence of the DNA looping on the binding of that protein can be deduced from the change. Likewise, any segment of the loop with a nontrivial deformation (e.g., an especially strong bend/twist) can be studied in a separate multi-scale simulation, using the elastic forces and torques at the ends of that segment predicted by the elastic rod model. The resulting corrections to the structure and elastic properties of that segment may in turn amend the predicted structure of the whole loop. Finally, the advent of much increased computational power may render even all-atom simulations of the protein complexes with whole DNA loops feasible. For such simulations, the all-atom loop structures, predicted on the basis of the elastic rod model of the loops, can serve as good starting points.

The described multi-scale modeling techniques and their combinations pave the way for extensive studies of multi-protein-DNA complexes that play a crucial role in the function of the genome of every living organism [1–7,75]. For example, multiple RNA polymerase components and tran-

scription factor proteins bind at or near the promoter region when transcription of a gene is being initiated; complex looping and buckling of the DNA between the binding sites ensues. Helicase and gyrase proteins concurrently alter DNA topology during replication causing diverse looping and coiling. A complex mesh of nucleosomes, loose DNA loops, and DNA-binding proteins forms the essence of eucaryotic chromatin. The growing number of known protein structures makes multi-scale simulations of such biomolecular systems possible. Such simulations, dealing with whole systems rather than focusing on the individual protein and DNA components, would yield a detailed picture of protein-DNA interaction and dynamics during key events in the life of a cell, leading to much improved understanding of these events.

D. Further development of the model

Several extensions of the model may prove necessary in order to achieve a realistic DNA description in the future studies suggested above. One such extension consists in adding terms accounting for DNA deformability to Eqs. (18)–(28), as discussed in Appendix A. DNA is known to be a shearable and stretchable molecule [17,32,33,42,50,71,77]; adding the deformability terms to Eqs. (18)–(28) may influence both the local and global structure of the modeled DNA loop.

Adding a steric repulsion force term to Eqs. (18)–(28) is critical for modeling DNA interactions with positively charged biomolecular entities, such as DNA-binding transcription factor proteins [75,78] or the nucleosome histone core [6]. Without the steric repulsion, a solution to Eqs. (18)–(28) may collapse onto a positive external charge included in the model, resulting in nonconvergence of the iterative BVP solver. Technically, the steric repulsion force is introduced on the basis of a 6–12 van der Waals potential in the same way as the electrostatic repulsion force (31)–(33) is introduced on the basis of the screened Coulomb potential. The resulting integro-differential Kirchhoff equations are solved using the iterative algorithm developed here.

Finally, the electrostatic interactions in our model can be rendered more realistic by placing the DNA charges on the periphery of the modeled DNA helix, where they belong in reality, rather than on the centerline as presently done [cf. Eq. (33)]. This modification would correct the DNA structure at the points of close contact with external charges or with itself, or at the points of a strong bend. The realistic charge placement can be achieved by using a three-dimensional charge density function instead of the simplified function (32). Moving the charges away from the centerline would give rise to electrostatic torques in each cross section of the rod, changing the local structure and the elastic force/torque distribution at the points of the rod where the electric field is strong.

E. *Lac* repressor loops

To conclude the paper, we summarize what has been learned with our method about the specific system, the *lac* repressor and its DNA loops. For both possible lengths of the loop, it has been shown that the underwound loop structure pointing away from the *lac* repressor should be predominant under thermodynamic equilibrium conditions, unless other biomolecules interfere. This conclusion holds true despite some uncertainty in the model parameters. The predicted structure of the U loop depends only slightly on the salt concentration, although the loop energy exhibits a stronger dependence. The experimentally observed energy of the loop can be obtained with the right combination of parameters. For further applications, the parameters have to be extensively tested with other protein-DNA systems.

The predicted loop structures yield the forces of *lac* repressor-DNA interaction, paving the way to multi-scale studies of the *lac* repressor-DNA complex, as described above. The already accomplished multi-scale studies gave important insights into the dynamics of the protein-DNA complex, elucidated the structural changes of the *lac* repressor in response to the stress of the bent DNA loop [25,26], and offered an explanation for DNA-binding synergy between the *lac* repressor and the major transcription activator CAP [70].

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Health (PHS 5 P41 RR05969). The figures in this manuscript were prepared using the molecular visualization program VMD [79]. The code for the numerical BVP solver COLNEW [58] was downloaded from the online software repository netlib.org

APPENDIX A

DNA deformability can be described in our model by three additional variables, combined into the shift vector $\boldsymbol{\epsilon}(s)$. Its components $\epsilon_{1,2}$ are the amount of shear in the two principal directions, and the component ϵ_3 is the amount of extension along the normal $\mathbf{d}_3$ [17,42]. The vector of shift $\boldsymbol{\epsilon}$ and the elastic force $\mathbf{N}$ are linearly related to each other, similarly to the vector of strains $\mathbf{k}$ and the torque $\mathbf{M}$ [cf. Eq. (11)]:

$$\mathbf{N}(s) = B_1\epsilon_1\mathbf{d}_1 + B_2\epsilon_2\mathbf{d}_2 + D\epsilon_3\mathbf{d}_3, \tag{A1}$$

where $B_{1,2}$ are the shear moduli in the two principal directions, and D is the extension modulus of DNA.

Thus introduced, the deformability terms change Eq. (2) into

$$\dot{\mathbf{r}} = \boldsymbol{\epsilon} + \mathbf{d}_3, \tag{A2}$$

propagating into Eq. (13) and Eqs. (18)–(28). The system (18)–(28) becomes of 16th rather than 13th order and could be similarly solved by the continuation method.

APPENDIX B

According to [61,80], the equilibrium constant of binding of the *lac* repressor to a single-operator DNA equals about 10^{-11} M for O_1 and 10^{-9} M for O_3 at high salt concentration (0.2 M). This results in the free energies of binding $\Delta G_{O_1} = \mathrm{kT}\log K_{O_1} = -25$ kT and $\Delta G_{O_3} = \mathrm{kT}\log K_{O_3} = -21$ kT. The

equilibrium binding constant of the *lac* repressor to the DNA promoter, containing both O_1 and O_3 sites, equals $3.4-6.2 \times 10^{-12}$ M [61], resulting in the free energy $\Delta G_{O_1-O_3} \approx -26$ kT. This results in the free energy of formation of the 76 bp DNA loop $\Delta G_{loop}=\Delta G_{O_1-O_3}-\Delta G_{O_1}-\Delta G_{O_3}=20$ kT.

APPENDIX C

The sequence-dependent DNA elasticity and geometry can be introduced into Eqs. (18)–(28) by constructing the appropriate parameter-functions $\alpha(s), \beta(s), \gamma(s), \overset{\circ}{\kappa}_{1,2}(s)$, and $\overset{\circ}{\omega}(s)$ for the specific sequence of each studied DNA loop. Such functions can be based on the elasticity parameters derived from the experimental data on DNA deformations and dynamics, e.g., as outlined in [23,32,33,48]. One can first assign the values $\alpha(s_{j+1/2})$, etc., according to the DNA sequence at the points $s_{j+1/2}=(j-1/2)h/l$, located midway between each successive [jth and $(j+1)$-th] base pairs. Then, the parameter-functions can be constructed as either smooth (e.g., spline-based) functions connecting the points $(s_{j+1/2}, \alpha(s_{j+1/2}))$, or piecewise functions, adopting the constant values $\alpha(s_{j+1/2})$ in the intervals $jh/l < s < (j+1)h/l$ between the successive base pairs and smoothly connected in narrow zones between the intervals.

APPENDIX D

For a DNA loop of a known sequence, the elastic rod model of that loop can be replaced by an all-atom model using the following algorithm. First, an idealized all-atom structure is built using standard biomodeling software (e.g., VMD [79] or Quanta [81]) for each base pair of the loop, including the phosphodeoxyribose backbone groups. Second, the elastic rod solution of the extended Kirchhoff equations, namely, $q_{1-4}(s)$, is used to obtain the local coordinate frames $(\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s))$ [46] at the points $s_i=ih/l$ along the loop centerline that correspond to the location of each (ith) base pair. Third, the built all-atom structures of the base pairs are centered at the points $\mathbf{r}(s_i)$ and aligned with the coordinate frames $(\mathbf{d}_1(s_i), \mathbf{d}_2(s_i), \mathbf{d}_3(s_i))$ as illustrated in Fig. 1(d). Fourth, several rounds of energy minimization are conducted for the thus-built all-atom DNA loop using standard molecular modeling packages (e.g., NAMD [82]) and forcefields (e.g., CHARM22 [83]). The minimization relieves bad interatomic contacts and chemical group conformations resulting from the idealized initial placement, especially in the DNA backbone. The resulting all-atom structure, while still stressed at certain points and overidealized at others, presents a good starting point for all-atom or multi-scale simulations, as described in Sec. VI C.

[1] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, 5th ed. (W. H. Freeman and Co., New York, 2002).
[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York, 2002).
[3] S. E. Halford, A. J. Welsh, and M. D. Szczelkun, Annu. Rev. Biophys. Biomol. Struct. **33**, 1 (2004).
[4] J. Vilar and L. Saiz, Curr. Opin. Genet. Dev. **15**, 136 (2005).
[5] S. Semsey, K. Virnik, and S. Adhya, Trends Biochem. Sci. **30**, 334 (2005).
[6] K. Luger, Curr. Opin. Genet. Dev. **13**, 127 (2003).
[7] M. Thanbichler, P. H. Viollier, and L. Shapiro, Curr. Opin. Genet. Dev. **15**, 153 (2005).
[8] T. E. Cheatham III, Curr. Opin. Struct. Biol. **14**, 360 (2004); J. Norberg and L. Nilsson, Q. Rev. Biophys. **36**, 257 (2003).
[9] W. K. Olson and V. B. Zhurkin, Curr. Opin. Struct. Biol. **10**, 286 (2000).
[10] T. Schlick, D. A. Beard, and J. Huang, Comput. Sci. Eng. **2**, 38 (2000).
[11] R. Phillips, M. Dittrich, and K. Schulten, Annu. Rev. Mater. Res. **32**, 219 (2002).
[12] A. A. Travers and J. M. T. Thompson, Philos. Trans. R. Soc. London, Ser. A **362**, 1265 (2004).
[13] J. Huang and T. Schlick, J. Chem. Phys. **117**, 8573 (2002).
[14] G. R. Kirchhoff, *Vorlesungen über Mathematische Physik, Mechanik, Vorlesung 28* (B.G. Teubner, Leipzig, 1977).
[15] A. V. Vologodskii and N. R. Cozzarelli, Annu. Rev. Biophys. Biomol. Struct. **23**, 609 (1994).
[16] T. Schlick, Curr. Opin. Struct. Biol. **5**, 245 (1995).
[17] Y. Shi, A. E. Borovik, and J. E. Hearst, J. Chem. Phys. **103**, 3166 (1995).
[18] A. Balaeff, L. Mahadevan, and K. Schulten, Phys. Rev. Lett. **83**, 4900 (1999).
[19] P. K. Purohit, J. Kondev, and R. Phillips, Proc. Natl. Acad. Sci. U.S.A. **100**, 3173 (2003).
[20] W. K. Olson, in *Oxford Handbook of Nucleic Acid Structure*, edited by S. Neidle (Oxford University Press, Oxford, UK, 1999), Chap. 16, pp. 499–531.
[21] J. F. Marko and E. D. Siggia, Phys. Rev. E **52**, 2912 (1995).
[22] P. B. Furrer, R. S. Manning, and J. H. Maddocks, Biophys. J. **79**, 116 (2000).
[23] W. K. Olson, D. Swigon, and B. D. Coleman, Philos. Trans. R. Soc. London, Ser. A **362**, 1403 (2004).
[24] B. D. Coleman and D. Swigon, Philos. Trans. R. Soc. London, Ser. A **362**, 1281 (2004).
[25] E. Villa, A. Balaeff, L. Mahadevan, and K. Schulten, Multiscale Model. Simul. **2**, 527 (2004).
[26] E. Villa, A. Balaeff, and K. Schulten, Proc. Natl. Acad. Sci. U.S.A. **102**, 6783 (2005).
[27] T. Schlick, B. Li, and W. Olson, Biophys. J. **67**, 2146 (1994).
[28] M. O. Fenley, W. K. Olson, I. Tobias, and G. S. Manning, Biophys. Chem. **50**, 255 (1994).
[29] A. Vologodskii and N. R. Cozzarelli, Biopolymers **35**, 289 (1995).
[30] J. Ubbink and T. Odijk, Biophys. J. **76**, 2502 (1999).
[31] A. A. Travers, Philos. Trans. R. Soc. London, Ser. A **362**, 1423 (2004).
[32] W. K. Olson, A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin, Proc. Natl. Acad. Sci. U.S.A. **95**, 11163 (1998).
[33] A. Matsumoto and W. K. Olson, Biophys. J. **83**, 22 (2002).

[34] W. K. Olson, N. L. Marky, R. L. Jernigan, and V. B. Zhurkin, J. Mol. Biol. **232**, 530 (1993).
[35] One important property, the extensibility/shearability of DNA [17,32,42,77], is not described by the equations presented here, yet it can be straightforwardly included in the problem, as discussed in Appendix A.
[36] L. Mahadevan and J. B. Keller, Philos. Trans. R. Soc. London, Ser. A **440**, 149 (1993).
[37] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, Science **271**, 1247 (1996).
[38] B. Müller-Hill, *The Lac Operon: A Short History of a Genetic Paradigm* (Walter de Gruyter, Berlin, 1996).
[39] S. Oehler, E. R. Eismann, H. Krämer, and B. Müller-Hill, EMBO J. **9**, 973 (1990).
[40] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover Publications, Inc., New York, 1927).
[41] L. D. Landau and E. M. Lifschitz, *Theory of Elasticity*, Vol. 7 of *Course of Theoretical Physics* (Pergamon Press, Oxford, New York, 1986), 3rd ed.
[42] T. P. Westcott, I. Tobias, and W. K. Olson, J. Chem. Phys. **107**, 3967 (1997).
[43] B. D. Coleman, I. Tobias, and D. Swigon, J. Chem. Phys. **103**, 9101 (1995).
[44] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked *et al.*, J. Mol. Biol. **313**, 229 (2001).
[45] All the variables in this paper, apart from a few clear exceptions, are considered to be functions of arclength s. Therefore, we frequently drop the explicit notation "(s)" from the equations throughout the paper.
[46] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, with an Introduction to the Problem of Three Bodies*, 4th ed. (Cambridge University Press, Cambridge, UK, 1960).
[47] The persistence lengths L_A and L_C of DNA bending and twisting are related to the elastic moduli A and C as $L_A=A/kT$ and $L_C=C/kT$ [84]. While most experiments agree on $A=2\times10^{-19}$ erg cm ($L_A=500$ Å) [31,49,52], there is less agreement on what the value of C should be [51,53,65,71], especially that it can be influenced by a strong DNA bend [71] or intrinsic curvature [23]. In this work, we use the value of $C=3\times10^{-19}$ erg cm ($L_C=750$ Å) resulting, e.g., from DNA micromanipulation experiments [60]. This gives the ratio of $A/C=2/3$.
[48] M. E. Hogan and R. H. Austin, Nature (London) **329**, 263 (1987).
[49] P. J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[50] T. R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, Annu. Rev. Biophys. Biomol. Struct. **29**, 523 (2000).
[51] P. J. Heath, J. B. Clendenning, B. S. Fujimoto, and J. M. Schurr, J. Mol. Biol. **260**, 718 (1996).
[52] S. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).
[53] Y. L. Zhang and D. M. Crothers, Biophys. J. **84**, 136 (2003).
[54] Here and further on, the bars over the dimensionless variables are dropped for simplicity.
[55] I. Tobias, D. Swigon, and B. D. Coleman, Phys. Rev. E **61**, 747 (2000).
[56] L. M. Edelman, R. Cheong, and J. D. Kahn, Biophys. J. **84**, 1131 (2003).
[57] More precisely, the boundaries of the loop were placed on the third base pair from the end of each DNA segment; the two terminal base pairs, seriously distorted in the crystal structure [37], were disregarded. The disregarded base pairs do not contact the *lac* repressor, therefore, the structure [37] is unlikely to be indicative of their structure and orientation within the continuous DNA loop.
[58] G. Bader and U. Ascher, SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **8**, 483 (1987).
[59] With symmetric boundary conditions and isotropic bendability ($\alpha=\beta$) of the elastic rod, an analytical solution to Eqs. (18)–(28) [24,43,73] can be chosen as the initial exact solution, rendering the first three iteration cycles unnecessary. In the studied case, the boundary conditions were asymmetric: e.g., the angles between $\mathbf{d}_3$ and the end-to-end vector $\mathbf{r}(1)-\mathbf{r}(0)$ were equal to 65° at $s=0$ and 99° at $s=1$.
[60] A. Balaeff, L. Mahadevan, and K. Schulten, http://arxiv.org/abs/physics/0301006 (2003).
[61] W. Hsieh, P. A. Whitson, K. S. Mathews, and R. D. Wells, J. Biol. Chem. **262**, 14583 (1987).
[62] B. D. Coleman, D. Swigon, and I. Tobias, Phys. Rev. E **61**, 759 (2000).
[63] A more sophisticated approach to modeling the bending anisotropy of DNA employs skewed elastic potentials [9,34].
[64] To characterize the difference between two elastic loop solutions $\mathcal{L}$ and $\mathcal{L}'$, this paper employs the root-mean square deviation (rmsd) between the centerlines $\mathbf{r}(s)$ and $\mathbf{r}'(s)$ of those loops, defined as $\sqrt{\int_0^1|\mathbf{r}(s)-\mathbf{r}'(s)|^2ds}$ and measured in DNA helical steps $h=3.4$ Å. h constitutes 1/76 of the length of the O_1-O_3 loop and 1/385 of the length of the O_1-O_2 loop.
[65] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature (London) **424**, 338 (2003).
[66] G. S. Manning, Q. Rev. Biophys. **2**, 179 (1978).
[67] An elegant alternative method of treating the forces of self-repulsion is presented in [24,55,62].
[68] The electrostatic torques $\mathbf{g}(s)$ are still effectively zeroed here due to placing the DNA charges on the centerline of the loop.
[69] Strictly speaking, the solutions to the Kirchhoff equations may correspond to maxima or saddle points of the energy functionals (29)/(34); only the solutions representing the energy minima, i.e., the mechanically stable solutions, are physically meaningful for any given DNA loop [24,55,62]. In the present work, the numerical solutions of Eqs. (18)–(28) are obtained by means of significant iterative variations of the boundary conditions and the problem parameters. For a mechanically unstable loop, such variations are likely to either flip the loop to a stable conformation (cf. Sec. V D) or to result in an outright nonconvergence, rendering any additional stability checks unnecessary. Yet, if the solutions are more numerous and their behavior is less certain than in the present case, then the analysis of their mechanical stability becomes important.
[70] A. Balaeff, L. Mahadevan, and K. Schulten, Structure (London) **12**, 123 (2004).
[71] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, Science **271**, 1835 (1996); S. B. Smith, Y. Cui, and C. Bustamante, *ibid.* **271**, 795 (1996).
[72] V. V. Rybenkov, C. Ullsperger, A. Vologodskii, and N. R. Cozzarelli, Science **277**, 690 (1997).

[73] D. Swigon, B. D. Coleman, and I. Tobias, Biophys. J. **74**, 2515 (1998).
[74] B. Israelewitz, M. Gao, and K. Schulten, Curr. Opin. Struct. Biol. **11**, 224 (2001).
[75] C. L. Lawson, D. Swigon, K. S. Murakami, S. A. Darst, H. M. Berman, and R. H. Ebright, Curr. Opin. Struct. Biol. **14**, 10 (2004).
[76] J. M. Hudson and M. G. Fried, J. Mol. Biol. **214**, 381 (1990); M. Perros, T. A. Steitz, M. G. Fried, and J. M. Hudson, Science **274**, 1929 (1996).
[77] J. F. Marko, Europhys. Lett. **38**, 183 (1997).
[78] J. C. Reese, Curr. Opin. Genet. Dev. **13**, 105 (2003).
[79] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).
[80] M. D. Barkley, and S. Bourgeois, in *The Operon*, edited by J. H. Miller and W. S. Reznikoff (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1980), pp. 177–220.
[81] Polygen, *Quanta*, Polygen Corporation, 200 Fifth Ave., Waltham, MA 02254 (1988).
[82] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, J. Comput. Phys. **151**, 283 (1999).
[83] A. D. MacKerell, Jr., J. Wiorkiewicz-Kuczera, and M. Karplus, J. Am. Chem. Soc. **117**, 11946 (1995).
[84] J. F. Marko and E. D. Siggia, Science **265**, 506 (1994).