Elastic Rod Model of a DNA Loop in the *Lac* Operon

Alexander Balaeff,[1,2] L. Mahadevan,[3] and Klaus Schulten[1,2,4]
[1]*Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*
[2]*Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*
[3]*Department of Mechanical Engineering, MIT, Cambridge, Massachusetts 02139*
[4]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*
(Received 22 June 1998; revised manuscript received 11 November 1999)

We use the theory of elasticity to compute the shape of the DNA loop bridging the gap in the crystal structure of the *lac* repressor-DNA complex. The Kirchhoff system of equations with boundary conditions derived from the crystal structure is solved using a continuation method. This approach can be applied effectively to find coarse-grained conformational minima of DNA loops.

PACS numbers: 87.14.Gg, 02.60.Lj, 87.15.Aa, 87.15.La

Introduction.—The *lac* repressor is one of the key enzymes in the lactose digestion chain of *E. coli* bacteria [1–4]. The protein turns off the genes responsible for lactose digestion when lactose is absent from the bacterial environment. The pioneering studies of this genetic switch, conducted in the 1950s, led to the fundamental concept of regulation of genetic activity [1–4].

The *lac* repressor functions through clamping two out of three target DNA sites (called the *operator* sites). The DNA between the sites is forced to form a loop [5], which interferes with reading the genes by another protein, the RNA polymerase [1,2], as shown in Fig. 1. The crystal structure of the *lac* repressor-DNA complex has been reported recently [6]. However, the crystallized protein holds two disjoint DNA segments, not connected with a DNA loop as they are in a living cell.

To reproduce the structure of the loop is essential for two reasons. (i) The repressor-DNA complex must absorb the stress of the looped DNA. Knowing the structure of the loop, one can estimate the stress and the resulting difference between the *in vivo* and the crystal structure of the protein-DNA complex. (ii) The DNA loop provides the scaffold for protein-DNA interactions of the complex genetic switch, of which the *lac* repressor is but a part. The structure of the loop provides a stepping-stone towards the study of these interactions at the molecular level.

In the present Letter, we predict the shape of the DNA loop induced by the *lac* repressor between the operator sites O_1 and O_3. We obtain the profiles of the DNA curvature and twist along the loop and estimate the energy of the loop and the force exerted by the loop on the protein-DNA complex. The operator sites are assumed to be positioned in the same way as the protein-bound DNA segments in the crystal structure [6].

The DNA in the loop is modeled as a naturally straight, inextensible elastic rod. Similar models have been used in Monte Carlo DNA simulations [7–9] and analytical studies of the conformations of DNA with an isotropic flexibility [10,11] or zero intrinsic twist [12]. Here we solve the equations of the theory of elasticity for anisotropically flexible DNA with the natural intrinsic twist [13], using a fast, computationally inexpensive approach that can be used to study DNA conformations in many protein-DNA systems.

Model.—The structure of the loop is derived from its elastic properties only, which effectively result from both internal DNA rigidity and the short-range [14] electrostatic self-repulsion of DNA. The long-range electrostatic interactions are screened out under physiological salt conditions and result in negligible forces in the case studied, as shown below. Entropic effects are also neglected since the studied loop has a length of only 26 nm, which is about half of the ~50 nm persistence length of DNA [15].

The elastic rod, approximating the DNA loop, is parametrized by its arclength s, according to the classical Kirchhoff approach [16]. The centers of the cross sections of the rod at each point s constitute the centerline $\vec{r}(s)$. The local coordinate frame $[\vec{d}_1(s), \vec{d}_2(s), \vec{d}_3(s)]$ determines the orientation of each cross section. The unit vectors $\vec{d}_1$ and $\vec{d}_2$ lie along the principal axes of the DNA cross section [17]; the vector $\vec{d}_3$ is normal to $\vec{d}_1$ and $\vec{d}_2$. The components of these vectors can be expressed in terms of the Euler angles or the Euler parameters [18], which determine the rotation of the local frame relative to the lab coordinate system.

The shape of the rod is described in terms of the curvature κ of its centerline and the twist Ω around the centerline [16]. The curvature is decomposed into its components κ_1 and κ_2 which account for the bending of the rod about $\vec{d}_1$ and $\vec{d}_2$, respectively. The twist is

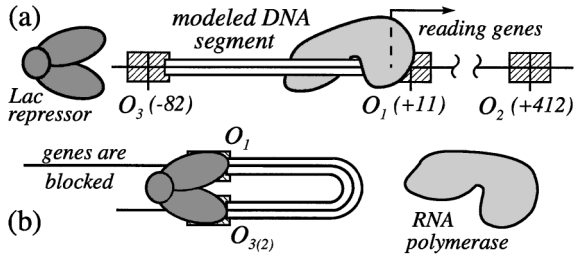

FIG. 1. The *lac* genetic switch [3,4] in its (a) active and (b) inactive states. The *lac* repressor operator sites are shaded.

 0031-9007/99/83(23)/4900(4)$15.00

decomposed into the equilibrium ("intrinsic") twist Ω_0 [13] of the relaxed rod and the unwinding (or overwinding) $\delta\Omega$, so that $\Omega = \Omega_0 + \delta\Omega$. The curvatures and the twist form the vector of strains $\vec{k}$ [16], which determines the spatial rate of rotation of the local coordinate frame along s:

$$\dot{\vec{d}}_i = \vec{k} \times \vec{d}_i, \qquad \vec{k} = (\kappa_1, \kappa_2, \Omega) \tag{1}$$

(the dot denotes the derivative with respect to s).

To complete the geometrical description of the rod, we use the constraint of inextensibility, which requires that the normal $\vec{d}_3$ coincide with the tangent to the centerline:

$$\dot{\vec{r}} = \vec{d}_3 . \tag{2}$$

The dynamical variables of the model are the elastic forces $\vec{n}(s)$ and torques $\vec{m}(s)$, which are required to be in equilibrium with all nonelastic forces $\vec{f}(s)$ and torques $\vec{g}(s)$ at every cross section in the rod [16]:

$$\dot{\vec{n}} + \vec{f}(s) = 0, \qquad \dot{\vec{m}} + \dot{\vec{g}} + \dot{\vec{r}} \times \vec{n} = 0 . \tag{3}$$

We assume that no body forces or torques act upon the DNA and disregard the long-range interactions, as stated above; therefore, $\vec{f} = 0, \vec{g} = 0$.

To complete the dynamical description of the rod, we assume the linear dependence of the elastic torques $\vec{m}$ on the curvatures κ_1, κ_2, and the unwinding $\delta\Omega$ [16]:

$$\vec{m} = A_1\kappa_1\vec{d}_1 + A_2\kappa_2\vec{d}_2 + C\delta\Omega\vec{d}_3 . \tag{4}$$

Here C is the DNA twisting modulus, and $A_1 = \alpha C$, $A_2 = \beta C$ are the moduli of DNA bending about $\vec{d}_1$ and $\vec{d}_2$, i.e., towards the DNA backbone and grooves, respectively [17]. We set $\alpha = 16/15$ and $\beta = 4/15$ [19] to account for the anisotropic flexibility of DNA [9].

Finally, the geometrical equations (1),(2) and the dynamical equations (3),(4) are made dimensionless [23] and combined into a 13th order Kirchhoff system of differential equations, as in [24]. The system poses a boundary value problem when supplemented with the location of the ends of the modeled loop and the orientation of the DNA cross section at either end, derived from the crystal structure [6].

The solutions to the system yield the geometry of the elastic rod $[\vec{r}(s), \vec{k}(s)]$ which equilibrates the elastic forces $\vec{n}(s)$ and torques $\vec{m}(s)$ and minimizes the energy

$$U = \int_0^1 \left(\frac{\alpha\kappa_1^2}{2} + \frac{\beta\kappa_2^2}{2} + \frac{\delta\Omega^2}{2} \right) ds . \tag{5}$$

Iterative procedure.—An iterative continuation method is used to solve the boundary value problem. The iterations start from a known exact solution, namely, from that for a self-equilibrated planar circular loop with coinciding ends and a circular cross section ($\alpha = \beta = 0.5$). Then the boundary conditions and the elastic moduli α and β are changed to achieve the desired values. Each change is gradually accomplished during a separate iteration cycle, consisting of a number of iteration steps. A solution obtained in each step is used as an initial guess for the next step. The next solution is computed by a numerical boundary value problem solver COLNEW [25], which uses a damped quasi-Newton method to construct the solution as a set of collocating splines.

Three iteration cycles are carried out. First, the far ($s = 1$) end of the loop is moved to the correct location. Second, the far end is rotated to adopt the correct orientation. Third, the elastic moduli α and β are changed to the chosen values $16/15$ and $4/15$. The resulting numerical solution smoothly connects the DNA segments in the crystal structure [6] and presents a minimum of the elastic energy functional (5).

However, the obtained solution is not unique, as the boundary conditions do not uniquely specify the DNA linking number (Lk), a topological property combining the writhe (coiling) of the centerline and the net twist of the rod [7–12]. In order to find the value of Lk minimizing the elastic energy (5), we split the continuation procedure after the second iteration cycle and rotate the far end of the loop by multiples of 2π around its axis $\vec{d}_3$ in a series of supplementary iteration cycles.

A new solution does emerge after one turn of the far end. The next turn, however, restores the original solution due to a self-crossing by the rod during the iterations. This event, not prevented in our model, reduces the Lk by 4π and turns out to be more energetically favorable than going to higher Lk solutions. Further turns of the far end flip the rod between the two solutions, hence termed "odd" and "even" according to the number of the supplementary iteration cycles.

The odd and the even solutions (Fig. 2a) remain different through the 3rd and 4th iteration cycles. At the end of

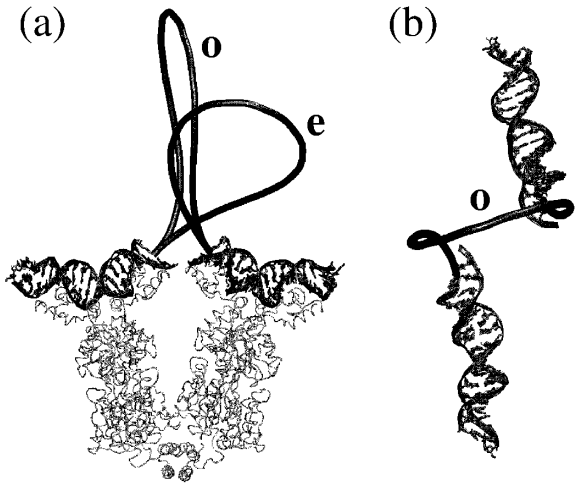

FIG. 2. (a) Elastic DNA loops, resulting from the odd (o) and the even (e) solutions of Eqs. (1)–(4) bridge the disjoint DNA segments in the crystal structure [6]. (b) Top view shows that the odd loop is roughly confined to the plane, perpendicular to the protein-bound DNA segments.

the iteration procedure, the odd solution has 14% (3.2 kT) lower energy than the even solution. Moreover, the even solution reveals a near self-crossing, which would impose a high energetic cost had electrostatics been taken into account. Therefore, from now on we focus on the odd solution, representing the energy minimum among all possible DNA conformations.

Structure and energetics of the loop.—The DNA loop described by the odd solution is nearly confined to the plane which is perpendicular to the direction of the protein-bound operator segments (Fig. 2b). This conformation differs from the smooth curvature model asserted in [6], where the best-fit plane to the DNA loop is apparently aligned with the operator segments. Otherwise, both models share the symmetry with respect to the symmetry axis of the protein-DNA complex. The loop in [6] is almost circular, and the elastic loop obtained here consists of a semicircle connected by two relatively straight sections to the operator DNA. Accordingly, the curvature rises in the middle and near the ends of the loop (Fig. 3).

The elastic energy of the loop amounts to $8.2C/l_0 =$ 23 kT, of which 81% is due to the bending of the loop and 19% is due to the unwinding. This energy is comparable to the free energy of the DNA looping by the *lac* repressor, estimated as 21 kT from [26]. The difference more likely results from the neglected entropic contribution and the approximations of our model rather than from significant discrepancies in the structure of the DNA loop and the protein clamp. Thus, the crystal structure [6], on which our model is based, should indeed closely resemble the *in vivo* structure of the protein-DNA complex.

The loop is stretched outwards with an average shear force $(n_1^2 + n_2^2)^{1/2}$ of $16.4C/l_o^2 = 7.4$ pN. The average tension (compression) n_3 of the straight (curved) segments is 0.9 (3.0) pN. This stress, partially absorbed by the protein clamp *in vivo*, should not lead to substantial changes of the internal DNA structure. For comparison, a force of 10–15 pN is required to separate the strands of the double helix [27], and a tension of 50–100 pN, to cause a DNA structural transition [28]. Hence, the values of the elastic moduli, measured for less severely bent DNA [7,21], are adequate here.

To clamp the DNA in the bent state, each protein subunit should hold its end of the loop with a force of 15.4 pN, assuming that the forces are similarly directed in the ends' coordinate frames and their vector sum yields the computed stress $\vec{n}$. This estimate is also reasonable. On one hand, proteins are able to produce such a force, e.g., the RNA polymerase can pull DNA with a force of up to 25 pN [29]. On the other hand, the protein should sustain such a stress without getting unfolded, which usually requires forces above 100 pN [29].

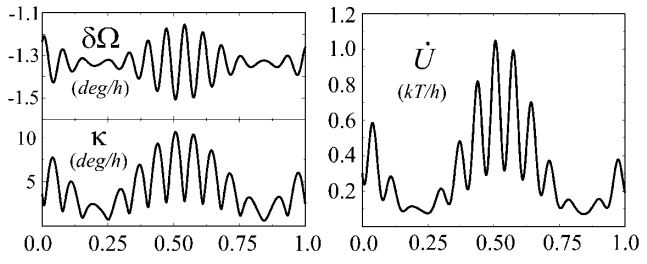

FIG. 3. Profiles of unwinding $\delta\Omega$, curvature $\kappa = \sqrt{\kappa_1^2 + \kappa_2^2}$, and elastic energy density $\dot{U}$ (5) along the odd solution.

The combination of the intrinsic DNA twist with the anisotropic DNA flexibility adopted in our model results in oscillations of the local structure of the loop (Fig. 3). The DNA segments bent towards the DNA backbone exhibit reduced curvature and increased twist, which relieve the difficult bending. When the backbone turns 90° away from the main bending direction, the curvature rises, the twist decreases, and the density of the elastic energy peaks.

Relevance of the method.—The obtained conformation of the DNA loop is dictated by the local elastic interactions, implicitly including a short-range electrostatic component. The neglected long-range electrostatics is unlikely to change significantly the loop conformation because no remote parts of the loop come close to each other so that the long-range interactions will be screened out by physiological salt concentrations. For example, a typical salt concentration of $0.1M$ [15] results in the Debye screening radius $\lambda = 10$ Å [30], and an all-atom model of the odd loop reveals no remote [32] phosphate groups (carrying the DNA negative charge) within 21 Å of each other. The operator DNA segments are even more distant and should cause an even lesser electrostatic effect.

However, the long-range electrostatics can have a significant influence on DNA conformation under different salt conditions or when near self-crossings occur. The latter become more likely for longer loops [7], but one is observed in our case as well. The closest phosphates in the discarded even solution come within 6 Å of each other, resulting in a repulsion of 40 pN. To address such cases correctly, one should include the screened Coulomb force in the equations of elasticity, e.g., in the form

$$\dot{\vec{f}}(s) = \frac{Q\rho(s)}{4\pi\epsilon\epsilon_0}\,\vec{\nabla}\,\frac{\exp[-|\vec{r}(s) - \vec{R}|/\lambda]}{|\vec{r}(s) - \vec{R}|}, \qquad (6)$$

where $\rho(s)$ is a smooth function with sharp maxima on the phosphates, modeling the charge distribution of DNA, and Q, $\vec{R}$ are, respectively, the magnitude and position of the force-inducing charge. The terms (6) should be summed over all the external electric charges, as well as those DNA phosphates which are separated by more than a certain number N_c of DNA steps from the point s. The closer phosphates are assumed to be adequately described by the elastic forces and therefore excluded from the sum.

Our preliminary calculations using this approach [33] result in the average shift of the loop centerline by 4.9 Å, or just 2% of the loop length. The average elastic force $n(s)$ acting on the phosphates changes by 0.25 pN only. These results justify our initial omission of the long-range electrostatic forces from the studied problem.

After including the long-range electrostatics, the method presented here can be used to study the DNA conformation and energetics in various biological systems, e.g., protein-induced DNA loops [5], or nucleosomal DNA [34]. If necessary, the elasticity equations can be modified to account for such DNA properties as the intrinsic curvature or the sequence-dependent elastic moduli [9]. The solution of the modified equations may however require additional iteration cycles during which the introduced DNA properties would be "turned on."

The data obtained from thus computed coarse-grained conformations can be used to study the structurally and functionally important parts of the biological systems by higher resolution modeling methods [35]. Such a multiresolution approach would make the computational studies of biomolecular aggregates much more feasible. For example, an all-atom simulation of the whole *lac* repressor-DNA complex would be of enormous computational cost, as the size of the fully solvated complex amounts to half a million atoms. Yet the repressor headpiece complexed with the operator DNA can readily be simulated separately [36], and the force computed here can be included in such a simulation to account correctly for the stress on the protein-DNA interface.

In summary, we have applied the theory of elasticity to find conformational minima of DNA loops using the continuation method. This approach allowed us to bridge the gap in the crystal structure of the *lac* repressor-DNA complex, revealing a realistic structure of the missing DNA loop. The results suggest a close similarity between the repressor-DNA complex in crystal and *in vivo*. The presented method opens a path to a multiresolution approach to the study of protein-DNA systems.

We thank Y. Lyanda-Geller for helpful discussions of the Letter. This work was supported by grants from the Roy J. Carver Charitable Trust, the NIH (PHS 5 P41 RR05969), and the NSF (BIR 94-23827EQ).

[1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland Publishing, New York, 1994), 3rd ed., Chap. 9.
[2] L. Stryer, *Biochemistry* (Freeman and Co., New York, 1995), 4th ed., Chap. 36.
[3] M. Ptashne, *A Genetic Switch* (Cell Press & Blackwell Scientific Publications, Cambridge, MA, 1992), 2nd ed.
[4] K. Matthews, Science **271**, 1245 (1996); *The Operon*, edited by J. Miller and W. Reznikoff (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1980).
[5] K. Matthews, Microbiol. Rev. **56**, 123 (1992).
[6] M. Lewis *et al.*, Science **271**, 1247 (1996).
[7] A. Vologodskii and N. Cozzarelli, Annu. Rev. Biophys. Biomol. Struct. **23**, 609 (1994).
[8] T. Schlick, Curr. Opin. Struct. Biol. **5**, 245 (1995).
[9] W. Olson, Curr. Opin. Struct. Biol. **6**, 242 (1996).
[10] J. Marko and E. Siggia, Science **265**, 506 (1994).
[11] Y. Shi and J. Hearst, J. Chem. Phys. **101**, 5186 (1994).
[12] B. Coleman, I. Tobias, and D. Swigon, J. Chem. Phys. **103**, 9101 (1995); Biophys. J. **74**, 2515 (1998).
[13] For DNA, $\Omega_0 = 34.6°$ per helical step $h = 3.4$ Å [1,2].
[14] The "range" of the DNA self-interactions refers to the length along the DNA between the interacting segments.
[15] P. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[16] A. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover Publications, New York, 1927), 4th ed., Chaps. 18–21; L. Landau and E. Lifshitz, *Course of Theoretical Physics* (Pergamon Press, New York, 1986), 3rd ed., Vol. 7, Secs. 16–21.
[17] The principal axes are associated with the DNA base pairs according to S. Diekman, EMBO J. **8**, 1 (1989). The axis $\vec{d}_1$ points towards the DNA major groove and $\vec{d}_2$ points towards the DNA backbone.
[18] E. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge University Press, Cambridge, England, 1960), 4th ed., Chap. 1.
[19] Conventional models of isotropically flexible DNA use a single bending modulus A. The ratio A/C varies between 0.5 and 1.5 in the literature [7,15,20,21], mainly due to the variance in C. We use more or less conventional values of $A = 2 \times 10^{-19}$ erg·cm and $C = 3 \times 10^{-19}$ erg·cm [7,20,21]. From the expression $A \approx (A_1 + A_2)/2$, which holds for a tightly twisted polymer [A. Balaeff (unpublished)], and the ratio of $\alpha/\beta = 4$ [22], we obtain $\alpha = 16/15$ and $\beta = 4/15$.
[20] P. Heath *et al.*, J. Mol. Biol. **260**, 718 (1996).
[21] T. Strick *et al.*, Science **271**, 1835 (1996).
[22] W. Olson *et al.*, J. Mol. Biol. **232**, 530 (1993).
[23] We scale all lengths by the length of the modeled DNA segment $l_0 = 26$ nm, the elastic moduli by $C = 3 \times 10^{-19}$ erg·cm [19], the energy by $C/l_0 = 1.16 \times 10^{-13}$ erg, and the forces by $C/l_0^2 = 0.45$ pN.
[24] L. Mahadevan and J. Keller, Proc. R. Soc. London A **452**, 1679 (1996).
[25] G. Bader and U. Ascher, SIAM J. Sci. Stat. Comput. **8**, 483 (1987).
[26] W-T. Hsieh *et al.*, J. Biol. Chem. **262**, 14 583 (1987).
[27] U. Bockelmann *et al.*, Phys. Rev. Lett. **79**, 4489 (1997).
[28] J. F. Léger *et al.*, Phys. Rev. Lett. **83**, 1066 (1999).
[29] A. Mehta *et al.*, Science **283**, 1689 (1999).
[30] $\lambda = 3$ Å$/\sqrt{c_s}$ for an aqueous solution of monovalent electrolyte of molar concentration c_s at 25° C [31].
[31] G. Manning, Q. Rev. Biophys. **II**, 179 (1978).
[32] The "remote" phosphates were defined as those separated by at least half a period of the DNA helix.
[33] The equations of elasticity with the terms (6) were solved for $\lambda = 10$ Å ($c_s = 0.1M$), $\epsilon = 80$ (water), $Q = -0.5e$ at each DNA step, as reduced by Manning counterion condensation [31], $\rho(s) = 8Q/3h \sin^4(\pi s l_0/h)$, and $N_c = 10$ DNA steps (one helical turn). The continuation algorithm was modified so that each iteration step was repeated until convergence of the DNA shape was achieved.
[34] K. Luger *et al.*, Nature (London) **389**, 251 (1997).
[35] A. Leach, *Molecular Modeling* (Longman, Harlow, England, 1996).
[36] V. Chuprina *et al.*, J. Mol. Biol. **234**, 446 (1993).