10.1098/rsta.2004.1384

THE ROYAL SOCIETY

Modelling DNA loops using continuum and statistical mechanics

By A. Balaeff[1]†, C. R. Koudella[2]‡, L. Mahadevan[2]‡ and K. Schulten[1]

[1]*Beckman Institute and Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

[2]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK* (lm@deas.harvard.edu)

Published online 13 May 2004

The classical Kirchhoff elastic-rod model applied to DNA is extended to account for sequence-dependent intrinsic twist and curvature, anisotropic bending rigidity, electrostatic force interactions, and overdamped Brownian motion in a solvent. The zero-temperature equilibrium rod model is then applied to study the structural basis of the function of the *lac* repressor protein in the *lac* operon of *Escherichia coli*. The structure of a DNA loop induced by the clamping of two distant DNA operator sites by *lac* repressor is investigated and the optimal geometries for the loop of length 76 bp are predicted. Further, the mimicked binding of catabolite gene activator protein (CAP) inside the loop provides solutions that might explain the experimentally observed synergy in DNA binding between the two proteins. Finally, a combined Monte Carlo and Brownian dynamics solver for a worm-like chain model is described and a preliminary analysis of DNA loop-formation kinetics is presented.

Keywords: DNA loops; elastic-rod model; *lac* repressor; catabolite gene activator protein (CAP); Brownian dynamics; loop-formation kinetics

1. Introduction

Although DNA is iconized as a straight double helix, it hardly ever exists in this pristine form in nature. Instead, it continuously interacts with a host of proteins involved in replication, transcription, repair and regulation and further must constantly be packed and unpacked into chromatin and other higher-order structures. These interactions make it bend, twist, supercoil, fold and loop dynamically even as it undergoes a variety of structural changes during these processes. All this occurs in a viscous environment as the highly charged molecule is constantly buffeted about by thermal fluctuations. To understand how the molecule works, we must build a dynamical

† Present address: IBM T. J. Watson Research Center, 1101 Kitchawan Rd, Route 134, Yorktown Heights, NY 10598, USA.

‡ Present address: Harvard University Division of Engineering and Applied Sciences, 29 Oxford Street, Cambridge, MA 02138, USA.

One contribution of 16 to a Theme 'The mechanics of DNA'.

© 2004 The Royal Society

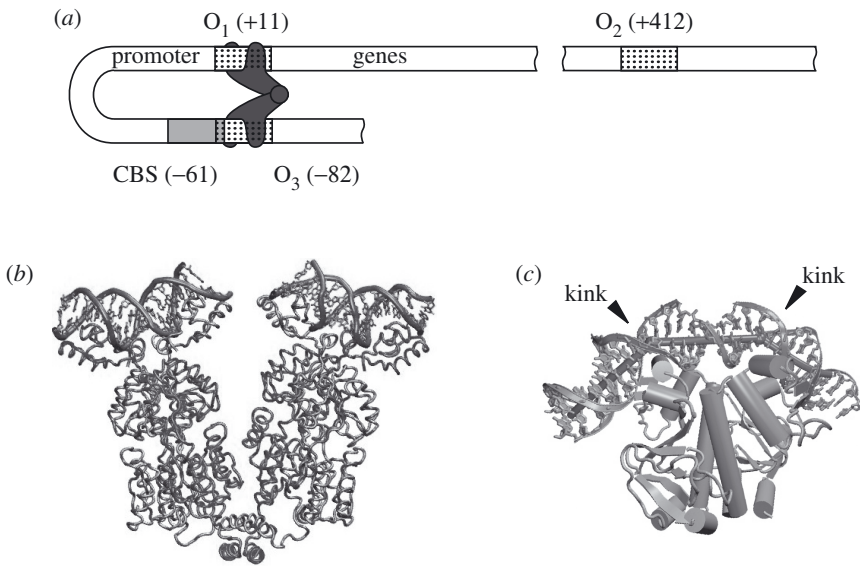

Figure 1. The *lac* operon genetic switch. (*a*) The *lac* operon showing the promoter bent by the *lac* repressor into a 76 bp-long loop. Hashed bars indicate the three *lac* repressor operators. The shaded bar represents the CAP binding site (CBS). The coordinates of the protein binding sites are shown in brackets. (*b*) The crystal structure of the *lac* repressor complex binding two operators (Lewis *et al*. 1996). (*c*) The crystal structure of the CAP–DNA complex (Parkinson *et al*. 1996) showing the location of the two kinks induced by CAP inside the binding site. The elastic-rod model of the CAP-bound DNA is shown as a tube fitted inside the DNA segment.

bridge using the scaffold provided by its structure. Since much of the action occurs on a scale that is neither truly molecular nor truly macroscopic, an important step in this venture is to have a hierarchy of models that weave into each other to span the large range of length-scales from the width of DNA (a few angstroms) to the size of tertiary structures such as DNA loops (a few micrometres).

A variety of mesoscopic coarse-grained models have been important in modelling DNA at these length-scales, and this paper looks at some of these in the context of a specific problem, the formation of loops. DNA loop formation is known to be an important mechanism for the regulation of gene expression (Matthews 1992; Schleif 1992). In many biological mechanisms involving loop formation, the characteristic length-scales are in the range of 1–5 times the persistence length of DNA, i.e. *ca.* 50–250 nm. This fact leads to some problems that lie at the boundary between statistical and continuum mechanics. Indeed, loop formation in a long piece of DNA is dominated by entropic effects, while loop formation in a short piece is dominated by the enthalpic cost of bending and twisting; in the intermediate range, on which our study is focused, both entropic and enthalpic effects have to be taken into account. The problem at hand leads to questions regarding the structure of loops, their equilibrium probability distribution, and the kinetics of their formation.

We address some of these issues in the context of one of the best-understood bacterial gene systems, the *lac* operon of *Escherichia coli* (see Ptashne 1992; Alberts *et al*. 2002), illustrated in figure 1, which is responsible for the regulation of the bacterial metabolism of lactose. We focus in particular on the action of *lac* repressor, a

protein that functions as a switch that shuts down the *lac* operon genes when the bacterial environment contains no lactose. *Lac* repressor achieves this by simultaneously binding two out of three operator sites and folding the DNA between those sites into a 76 bp- or 385 bp-long loop (figure 1*a*). The formation of both loops has been shown to be critical in order to achieve full repression (Oehler *et al*. 1990). In this paper, we focus on the shorter loop because it contains the *lac* operon promoter and its formation is thus certain to disrupt genetic expression. Besides, the 76 bp loop also contains a binding site for catabolite activator protein (CAP) (Busby & Ebright 1999), which has been shown to bind to the loop synergistically with *lac* repressor (Hudson & Fried 1990). The crystal structure of *lac* repressor is known (Lewis *et al*. 1996), but contains only disjoint DNA segments bound to the *lac* repressor, for the complex with the whole loop cannot be crystallized. We therefore calculate the unknown structure of the 76 bp-long DNA loop and investigate the properties of the loop using a Kirchhoff elastic-rod model. Similar models have been widely applied in DNA modelling (most recently reviewed in Olson & Zhurkin (2000) and Olson (1996)).

In § 2 we review the Kirchhoff model for the case of an anisotropic electrostatically charged elastic rod with intrinsic twist and/or curvature and subject to overdamped Brownian motion in a viscous solvent. In § 3 we use the equations of equilibrium to solve for the shape of the *lac* repressor-induced DNA loops using the *lac* repressor structure from Lewis *et al*. (1996). We also examine the changes on the loop structure induced by the binding of CAP inside the loop and reveal the structural basis of the synergy between the two proteins (Balaeff *et al*. 2003*b*). In § 4 we address the kinetics of the formation of a general loop in the context of the worm-like chain model using a combination of Monte Carlo and Brownian dynamics simulations. We conclude with a discussion of the rod model in the context of further simulations of the *lac* repressor–CAP–DNA complex.

2. The elastic-rod model

(*a*) *Rod kinematics*

We start with a consideration of the classical elastic-rod model (Kirchhoff 1883) modified to account for a possibly anisotropic cross-section, immersion in a viscous solvent, finite-temperature effects, electrostatic interactions and undergoing (overdamped) Brownian motion. We describe an elastic rod of length l in terms of its centreline and its cross-section, which, for expository simplicity, we assume to be circular and of constant radius a (figure 2*a*). The generalization to arbitrary cross-sections is straightforward. The centreline forms a space curve $\boldsymbol{r}(s,t)$ parametrized by arclength s and time t. The cross-sections are 'stacked' along the centreline. The director basis defined by three orthonormal unit vectors $\boldsymbol{d}_i(s,t)$, for $i = 1,\ 2$ or 3, uniquely defines the orientation of the cross-section given s and t. The vectors $\boldsymbol{d}_1(s,t)$ and $\boldsymbol{d}_2(s,t)$ lie in the plane of the cross-section and may be defined with respect to the reference configuration, while the vector $\boldsymbol{d}_3(s,t) = \boldsymbol{d}_1(s,t) \times \boldsymbol{d}_2(s,t)$ is the normal to that plane. Any point $\boldsymbol{x}(s,t)$ in the rod is consequently parametrized by

$$\boldsymbol{x}(s,t) = \boldsymbol{r}(s,t) + \boldsymbol{y}(s,t) = \boldsymbol{r}(s,t) + y_1\boldsymbol{d}_1(s,t) + y_2\boldsymbol{d}_2(s,t),$$

where the coordinates y_1 and y_2 define the location of the point relative to the centreline in a cross-section (figure 2*a*). We denote partial derivatives with respect

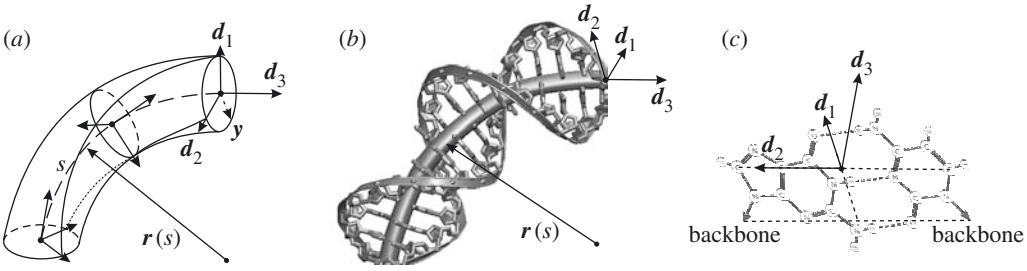

Figure 2. Elastic-rod model of DNA. (*a*) Parametrization of the elastic rod. Using the centreline position vector $\boldsymbol{r}(s,t)$ and the director basis defined by the unit vectors $\boldsymbol{d}_1(s,t)$, $\boldsymbol{d}_2(s,t)$ and $\boldsymbol{d}_3(s,t)$, where $\boldsymbol{d}_3 = \boldsymbol{d}_1 \times \boldsymbol{d}_2$, any given point in the rod may be written as $\boldsymbol{x}(s,t) = \boldsymbol{r}(s,t) + \boldsymbol{y}(s,t) = \boldsymbol{r}(s,t) + y_1\boldsymbol{d}_1(s,t) + y_2\boldsymbol{d}_2(s,t)$. In an application the director basis must be parametrized by a set of coordinates such as the singular Euler angles $\phi(s,t)$, $\psi(s,t)$, $\theta(s,t)$ or the four non-singular Euler parameters $q_1(s,t)$, $q_2(s,t)$, $q_3(s,t)$, $q_4(s,t)$ used in the equilibrium problems considered here. (*b*) The elastic rod fitted to an all-atom structure of DNA. (*c*) The coordinate frame associated with a DNA base pair (Olson *et al*. 2001).

to s and t by a prime ($'$) and a superposed dot ($\dot{}$), respectively. The space and time dependence of the director basis vectors $\boldsymbol{d}_i$ is implicitly assumed in what follows. In this work we consider the DNA to be unshearable and inextensible, so that $\boldsymbol{r}'(s,t)$, the tangent to the centreline, is always parallel to $\boldsymbol{d}_3$ and normalized:

$$\boldsymbol{r}' = \boldsymbol{d}_3. \tag{2.1}$$

The rod kinematics may then be characterized entirely by two bending modes and one twisting mode, encapsulated in the relationship

$$\boldsymbol{d}_i' = \boldsymbol{k} \times \boldsymbol{d}_i, \tag{2.2}$$

where the vector of strains $\boldsymbol{k}(s,t) = K_1(s,t)\boldsymbol{d}_1 + K_2(s,t)\boldsymbol{d}_2 + \Omega(s,t)\boldsymbol{d}_3$ has as components the projection of the centreline curvature onto the director basis vectors in the cross-section $K_1(s,t)$ and $K_2(s,t)$ and the local twist $\Omega(s,t)$. A similar relationship holds for the time dependence of the director basis:

$$\dot{\boldsymbol{d}}_i = \boldsymbol{\sigma} \times \boldsymbol{d}_i, \tag{2.3}$$

where $\boldsymbol{\sigma}(s,t) = \sigma_1(s,t)\boldsymbol{d}_1 + \sigma_2(s,t)\boldsymbol{d}_1 + \sigma_3(s,t)\boldsymbol{d}_3$ is the angular velocity vector.

(*b*) *Rod dynamics*

(i) *Elastic internal force and torque*

In order to proceed to a dynamical description of the rod we introduce the resultant internal elastic force $\boldsymbol{N}(s,t) = N_1(s,t)\boldsymbol{d}_1 + N_2(s,t)\boldsymbol{d}_2 + N_3(s,t)\boldsymbol{d}_3$ and the resultant internal elastic torque $\boldsymbol{M}(s,t) = M_1(s,t)\boldsymbol{d}_1 + M_2(s,t)\boldsymbol{d}_2 + M_3(s,t)\boldsymbol{d}_3$ acting on a cross-section. The force $\boldsymbol{N}(s,t)$ enters the theory as the Lagrange multiplier enforcing (2.1) and forbids pure extensional or shear deformations of the rod. The elastic torque $\boldsymbol{M}(s,t)$ is specified by a constitutive relation, generalizing Euler–Bernoulli beam theory (Love 1927) to

$$\boldsymbol{M}(s,t) = A_1(s)\kappa_1\boldsymbol{d}_2 + A_2(s)\kappa_2\boldsymbol{d}_1 + C(s)\omega\boldsymbol{d}_3, \tag{2.4}$$

where we define the curvature and twist deviations of the rod from its intrinsic shape defined by $\kappa^{\circ}_{1,2}$ and ω° as

$$\kappa_{1,2}(s,t) = K_{1,2}(s,t) - \kappa^{\circ}_{1,2}(s), \qquad \omega(s,t) = \Omega(s,t) - \omega^{\circ}(s). \tag{2.5}$$

$A_{1,2}$ are the bending rigidities along the 1, 2-directions and C is the twisting rigidity. The torque $\boldsymbol{M}(s,t)$ is seen to be proportional to the geometric deviation from the reference configuration and reflects the simplest possible dependence consistent with a quadratic energy

$$\mathcal{U} = \frac{1}{2}\int_0^l (A_1\kappa_1^2 + A_2\kappa_2^2 + C\omega^2)\,\mathrm{d}s, \tag{2.6}$$

with respect to body-fixed coordinates. Note that the rod model allows for s-dependent material properties $A_1(s)$, $A_2(s)$ and $C(s)$, as is manifest in (2.4). For a homogeneous circular cross-section the bending rigidities of the rod are $A_1 = EI_2$ and $A_2 = EI_1$, where E is Young's modulus and the moments of inertia are $I_1 = I_2 = \pi a^4/4$. In polymer science and biological applications it is common to define the bending rigidities as $A_1 = A_2 = k_{\mathrm{B}}Tl_{\mathrm{p}}$, where k_{B} is Boltzmann's constant, T is the absolute temperature, and l_{p} is the persistence length of the rod—the length-scale characterizing the orientational correlation of the rod. We use the commonly accepted value for DNA of $l_{\mathrm{p}} = 50\,\mathrm{nm}$ (Strick *et al.* 2000; Olson 1996) in all our computations. Note that the persistence length is sensibly defined only for isotropic rods, but when $A_1 \neq A_2$, it characterizes the bending only in some average sense. We show below that accounting for such anisotropic bending is important for a correct elastic-rod model of DNA. The twisting rigidity is similarly defined as $C = GJ$, where G is the shear modulus and $J = \pi a^4/2$ is the polar moment of inertia of the rod. We may similarly define C in terms of a twisting persistence length $C = k_{\mathrm{B}}Tl_{\mathrm{twist}}$ and we use $l_{\mathrm{twist}} = 75\,\mathrm{nm}$.

The physical reason underlying the constraint that the rod is unshearable and inextensible (2.1) may be summarized as follows. As can be seen from (2.6) and the moment of inertia $I = \pi a^4/4$, the bending energy per unit length-scales as $Ea^4\kappa^2$, where κ is the curvature of the rod. The energy associated with shearing/stretching on the other hand scales as $Ea^2\gamma^2$, where γ stands for the shearing/streching strain. It follows that the ratio scales as $\gamma^2/a^2\kappa^2$. For an elastic rod corresponding to the strongly bent DNA loops considered in our work, the typical curvature is $\kappa \sim 1/l$, and the cross-sectional stress scales as Ea^2/l^2, so that the shear strain is $\gamma \sim a^2/l^2$. Therefore, $\gamma^2/a^2\kappa^2 \sim a^2/l^2 \ll 1$, so that, compared with bending, shearing and stretching deformations may be neglected to a first approximation.

(ii) *External load per unit length*

Since the rod is immersed in solvent, we assume that any motion of the rod is resisted by a (Stokesian) viscous drag force and torque proportional to the linear and angular velocity, respectively. Accordingly, the viscous drag force per unit length acting on a flexible rod is (Cox 1970)

$$\boldsymbol{f}^{\mathrm{drag}}(s,t) = -\zeta(\boldsymbol{I} - \tfrac{1}{2}\boldsymbol{d}_3\boldsymbol{d}_3)\dot{\boldsymbol{r}}, \tag{2.7}$$

where $\zeta = 2\pi\eta/\ln(l/a)$ is a friction coefficient, η is the fluid viscosity, $\boldsymbol{I}$ is the identity tensor, and $\boldsymbol{d}_i\boldsymbol{d}_i$ is the tensor associated with direction i. The friction tensor

is diagonal but anisotropic, so that the drag force opposing motion in the plane of the cross-section is twice as large as for motion along the rod axis $\boldsymbol{d}_3$. Similarly, the viscous drag torque reads (Cox 1971)

$$\begin{aligned}\boldsymbol{g}^{\mathrm{drag}}(s,t) &= -\lambda(\boldsymbol{d}_1\boldsymbol{d}_1 + \boldsymbol{d}_2\boldsymbol{d}_2)\boldsymbol{d}_3 \times \dot{\boldsymbol{d}}_3 - \lambda \boldsymbol{d}_3\boldsymbol{d}_3(\boldsymbol{d}_1 \times \dot{\boldsymbol{d}}_1 + \boldsymbol{d}_2 \times \dot{\boldsymbol{d}}_2) \\ &= -\lambda\sigma_1\boldsymbol{d}_1 - \lambda\sigma_2\boldsymbol{d}_2 - 2\lambda\sigma_3\boldsymbol{d}_3, \end{aligned} \tag{2.8}$$

where $\lambda = 2\pi\eta a^2$ is a rotational friction coefficient. Here the first two terms describe the resistance to local tilting/bending motions of the rod associated with a viscous torque lying in the rod cross-sectional plane; the last term describes the resistance to local twisting motion associated with an axial viscous torque.

The Brownian motion of the rod is induced by a stochastic force $\boldsymbol{f}^{\mathrm{stoc}}(s,t)$ and a stochastic torque $\boldsymbol{g}^{\mathrm{stoc}}(s,t)$ modelling the effects of solvent molecules colliding with the rod. In order to specify the stochastic quantities we select expressions for the first moments and the cross-correlations consistent with analogous expressions used in Brownian dynamics of discrete bead-spring/rod polymer models (Öttinger 1996). The rod is in thermodynamic equilibrium under the action of the stochastic force and torque provided that both have zero mean and obey the fluctuation-dissipation relation

$$\langle f_i^{\mathrm{stoc}}(s,t)\rangle = 0, \qquad \langle f_i^{\mathrm{stoc}}(s',t')f_j^{\mathrm{stoc}}(s,t)\rangle = 2k_{\mathrm{B}}T\zeta_i\delta(s'-s)\delta(t'-t)\delta_{ij}, \tag{2.9}$$

$$\langle g_i^{\mathrm{stoc}}(s,t)\rangle = 0, \qquad \langle g_i^{\mathrm{stoc}}(s',t')g_j^{\mathrm{stoc}}(s,t)\rangle = 2k_{\mathrm{B}}T\lambda_i\delta(s'-s)\delta(t'-t)\delta_{ij}, \tag{2.10}$$

where k_{B} is Boltzmann's constant, T is the temperature, $\delta(\cdot)$ is the Dirac distribution, δ_{ij} is the Kronecker delta, $\zeta_{1,2} = 2\zeta$, $\zeta_3 = \zeta$, $\lambda_{1,2} = \lambda$, $\lambda_3 = 2\lambda$ and $\langle\cdot\rangle$ refers to ensemble averages.

Finally, the electrostatic interaction of the charged rod with its surroundings is modelled as

$$\boldsymbol{f}^{\mathrm{elec}}(s,t) = q(s)(\boldsymbol{E}^{\mathrm{self}}(s,t) + \boldsymbol{E}^{\mathrm{other}}(s,t)),$$

where $q(s)$ is the local electrostatic charge density on the rod. The electrostatic field has two origins, a rod self-interaction term, $\boldsymbol{E}^{\mathrm{self}}$, and an interaction between the rod and other charges present in the system, $\boldsymbol{E}^{\mathrm{other}}$. The implemented electrostatics model involves a Debye screening factor due to the effect of counterions surrounding DNA (see Balaeff *et al*. (2003) for full details on the electrostatics implementation in our model). Electrostatic torques, which may in principle result from electrostatic interactions, are neglected here because the electric charge is assumed to be located on the centreline of the rod.

(iii) *Equations of motion*

Collecting all the terms and assuming (as is common in polymer dynamics) that the rod motion is overdamped, i.e. the inertia terms are negligible compared with the forces involved, we arrive at the following force and torque balances for an *electrostatically charged elastic rod undergoing (overdamped) Brownian motion*:

$$-\boldsymbol{f}^{\mathrm{drag}} = \boldsymbol{N}' + \boldsymbol{f}^{\mathrm{elec}} + \boldsymbol{f}^{\mathrm{stoc}}, \tag{2.11}$$

$$-\boldsymbol{g}^{\mathrm{drag}} = \boldsymbol{M}' + \boldsymbol{r}' \times \boldsymbol{N} + \boldsymbol{g}^{\mathrm{stoc}}, \tag{2.12}$$

where the drag and stochastic interactions are given in (2.7), (2.8) and (2.9), (2.10), respectively.

In most cases relevant to DNA dynamics we may normally neglect the viscous and stochastic torques in the cross-sectional components of the angular momentum balance (2.12), as the following reasoning shows. Assuming that the elastic force is balanced by the viscous drag in (2.11), the typical elastic force scales as $N \sim U\eta l$, where U is a characteristic rod speed. From (2.12) it then follows that the ratio of the viscous torque to the torque arising from the stress-resultant is $g^{\mathrm{drag}}/(\boldsymbol{r}' \times \boldsymbol{N}) \sim \lambda_i \sigma_i / \eta l U$, where $\lambda_i \sim \eta a^2$ according to (2.8). Qualitatively, there are two types of rotational rod motion: spinning about the rod axis and bending/tilting perpendicular to the rod axis. For spinning motion, $U \sim \sigma_3 a$, resulting in the viscous torque being smaller than the stress-resultant by a factor $\varepsilon = a/l$. For bending/tilting, $\sigma_{1,2} \sim U/l$ and the resulting cross-sectional components of the viscous torque are a factor ε^2 smaller than the elastic stress-resultants. For problems in which we may neglect second-order effects, two of the equations in (2.12) reduce to time-independent constraints linking some of the components of the elastic force $\boldsymbol{N}(s,t)$ and moment $\boldsymbol{M}(s,t)$, respectively. Then two constraints determine the cross-sectional shear-resultant and consequently two of the kinematic constraints embodied in (2.1) become redundant. Relation (2.1) then reduces to the scalar inextensibility constraint $|\boldsymbol{r}'(s,t)| = 1$ commonly used in bead-rod polymer models such as the worm-like chain model discussed in § 4. We note that a further consequence of this approximation is that one cannot specify boundary conditions on the torques.

The complete set of equations are the kinematic relations (2.1), (2.2) and (2.3), the dynamical balances (2.11), (2.12), the constitutive relation (2.4) for the elastic torque $\boldsymbol{M}(s,t)$, and a set of boundary and initial conditions.

3. Equilibrium rod model applied to the study of *lac* operon promoter

We now review the results of our earlier work (Balaeff *et al.* 1999, 2003, 2004) on applying the Kirchhoff rod model to study the 76 bp loop folded by *lac* repressor. The 76 bp loop connects the ends of the operator DNA segments bound to a *lac* repressor molecule (figure 1), which is assumed to have the structure crystallized by Lewis *et al.* (1996). Because the loop is only of the order of $0.5 l_{\mathrm{p}}$ long, we neglect entropic effects to a first approximation and study its structure at equilibrium in the zero-temperature limit. The rod equations for equilibrium are obtained by discarding the time dependence of all variables and neglecting Brownian force and torque in (2.11), (2.12). Other authors have considered similar equilibrium models as a departure point for DNA studies (Manning *et al.* 1996; Olson 1996; Coleman *et al.* 1995; Schlick 1995). We rewrite the equilibrium equations as a 13th-order system using a quaternion parametrization of the director basis as in Mahadevan & Keller (1996), Manning *et al.* (1996), Westcott *et al.* (1997) and Balaeff *et al.* (2003). The coordinate frames fitted to the termini of the DNA segments (cf. figure 2*c*) provide us with the boundary conditions.

The loop solutions are constructed with an iterative continuation algorithm (Mahadevan & Keller 1996), as illustrated in figure 3. The iterations start with a simplified set of parameters and boundary conditions for which an exact equilibrium solution is known, namely, a closed circular loop and zero electrostatic charge (figure 3*a*). Then the boundary conditions and the elastic moduli are gradually changed in several iteration cycles (figure 3*b*, *c*) until the desired solution is obtained. Each

iteration cycle consists of using a standard boundary-value problem solver COLNEW (Bader & Ascher 1987) that uses the solution obtained at the previous step as an initial guess in a Newton–Raphson-like iterative algorithm.

Since the boundary conditions do not specify the linking number of the loop—a topological invariant combining the writhe (coiling) of the centreline and the net twist—we search for loop solutions with the smallest elastic energy by rotating one end of the loop an integral number of turns. Two solutions result, one that is slightly underwound, (figure 3*c*), and another one that is slightly overwound (figure 3*c*′). We denote these two solutions by 'U' and 'O', respectively, with the 'U' loop being energetically favourable. However, it should be noted that these estimates are based on ensemble averages obtained from bulk studies, although recent work on single-molecule experiments (Finzi & Gelles 1995; Lia *et al*. 2003) should eventually allow for better comparisons.

Both U and O loops point away from the *lac* repressor, as proposed by Lewis *et al*. (cf. figures 1*b* and 3), but in principle the loop may also wrap around the *lac* repressor tetramer (Tsodikov *et al*. 1999). Indeed, our algorithm can produce solutions with the loop pointing 'down', towards the *lac* repressor (see Balaeff *et al*. 2003). Yet, these loops are not only found to have significantly higher energies than the U and O loops above, but also to be sterically prohibitive, passing through the areas of space supposedly occupied by the *lac* repressor. Stretching these loops to avoid the steric clash and wrap them around the *lac* repressor would further increase their energies. Changes in the *lac* repressor geometry (Edelman *et al*. 2003) may presumably alter the energy balance and make the wrapped-around loops energetically more favourable. The present discussion, however, focuses solely on the structure by Lewis *et al*. and geometries other than the U and O loops are beyond the scope of this review.

(iv) *Anisotropy*

As an example of the various refinements to the basic equilibrium rod model that are possible, we first consider the effect of anisotropic bending rigidities based on the ratio $\mu = A_1/A_2 = 4$ (Olson *et al*. 1993; Balaeff *et al*. 2003). The new loops obtained after another iteration cycle, in which the bending moduli are adjusted to

Figure 3. Equilibrium elastic-rod solutions for the *lac* repressor-induced DNA loop generated using the iterative continuation algorithm (Balaeff *et al*. 1999, 2003). (*a*) The simplified circular loop initializes the iterative procedure. (*b*) One end of the loop is moved to the other receiving end of the *lac* repressor. (*c*), (*c*′) The moved end of the loop is rotated into the correct orientation. Depending on the direction of the rotation (or the number of subsequent turns of the end of the loop (see Balaeff *et al*. 1999, 2003 for details)) the underwound 'U' (*c*) and overwound 'O' (*c*′) solutions emerge. (*d*), (*d*′) The correct bending rigidity $A/C = 2/3$ is now introduced into the model. Loops obtained with isotropic bending moduli $A_1 = A_2$ are coloured light, loops obtained with anisotropic bending moduli $A_1/A_2 = 4$ are coloured dark. (*e*), (*e*′) Changes in the predicted loop structure due to electrostatic interactions at different ionic strengths. The loops obtained for the ionic strength of 100 mM are coloured light, those for 25 mM are coloured medium, and those for 10 mM are coloured dark. There is virtually no difference between the 100 mM structures and those obtained without the electrostatic term (dark solutions in (*d*), (*d*′)). (*f*), (*f*′) Total (T), elastic (E), and electrostatic (Q) energy of the rod at different values of ionic strength. The points corresponding to the snapshots in (*e*), (*e*′) are indicated with corresponding colours on the abscissa. In all panels, 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.

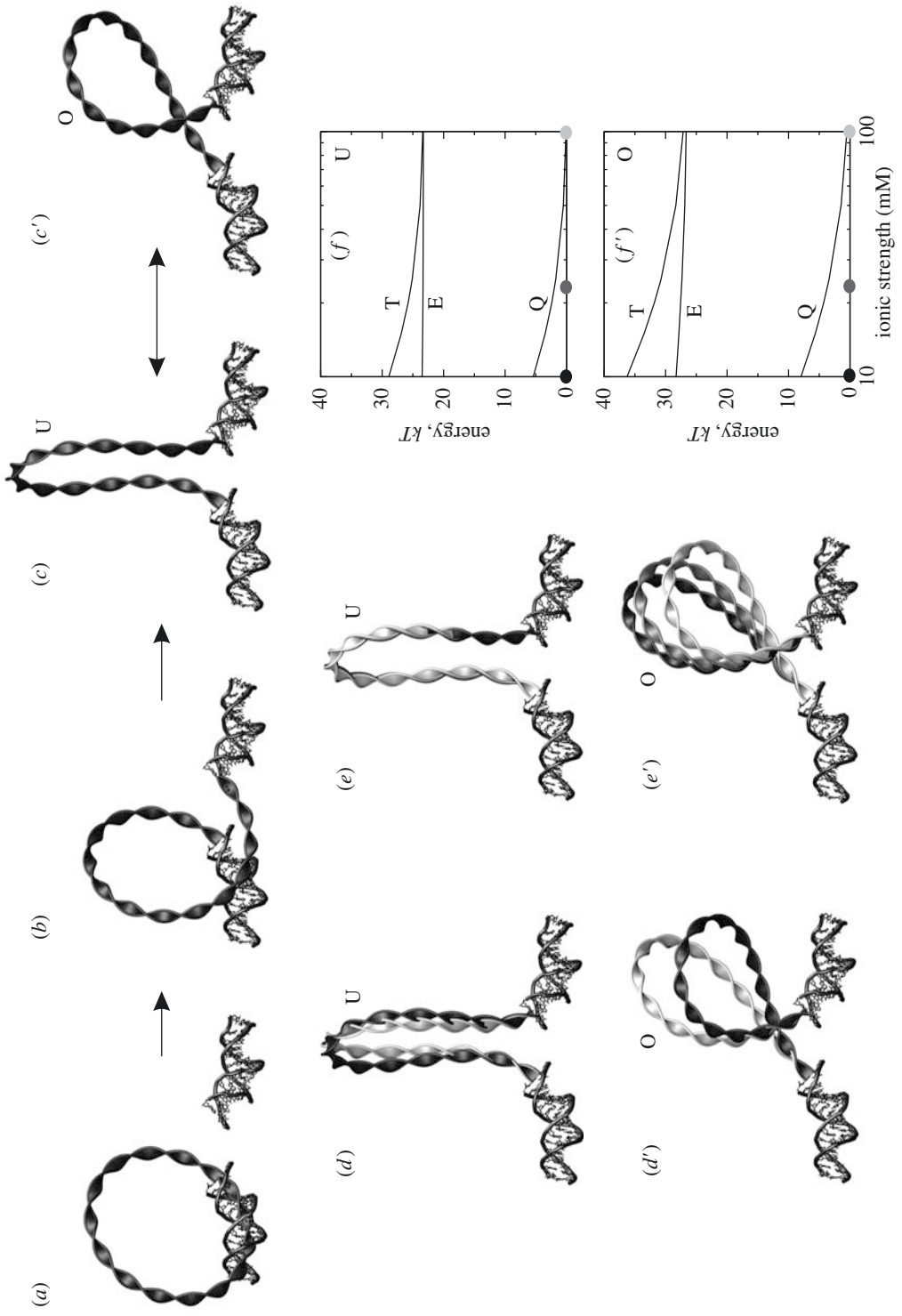

Figure 3. See opposite for description.

the anisotropic values, are shown in figure 3*d*, *d*′. Only the O loop is structurally altered to a significant degree, although the energies of both loops are reduced by a third. Microscopically, this leads to the formation of a kink-like structure since bending becomes effectively cheaper, and twist is exchanged for bend as the average twist ω decreases for both U and O solutions.

(v) *Electrostatics*

Next, we consider the effects of electrostatic interactions on the DNA loops. Including the force terms corresponding to the rod interaction with itself ($\boldsymbol{E}^{\mathrm{self}}$ above) and with the phosphates of the crystallized DNA segments ($\boldsymbol{E}^{\mathrm{other}}$) results in a set of integrodifferential equations (Westcott *et al*. 1997; Balaeff *et al*. 1999, 2003). Again an iterative algorithm is used to gradually increase the Debye screening radius from zero (infinite ionic strength) to infinity (zero ionic strength) and some sample results are presented in figure 3*e*, *f*′. The increased electrostatic repulsion gradually changes the shape of the O loop (but not the U loop) (figure 3*f*, *f*′). The feature that makes the O loop more sensitive to electrostatic interactions is its point of near self-crossing, which is absent in the U loop; thus, at low salt concentration the U loop becomes even more favourable. We find that electrostatic terms can be ignored except in the neighbourhood of near self-intersections, in accord with our intuition.

(vi) *Inhomogeneity*

Finally, we give an example of how to model the effect of a binding protein, CAP, which binds inside the 76 bp loop (see Balaeff *et al*. (2004) for details). This type of situation, where a second protein serves to assist another DNA binding protein, is very common in DNA–protein interactions. In order to mimic the structure of the CAP binding site (CBS) inside the U and O loops, we modify the intrinsic curvature and twist parameters $\kappa_1^\circ(s)$ and $\omega^\circ(s)$ locally. The resulting kinked and unwound rod section models the DNA steps kinked and unwound by the bound CAP (figure 1*c*). The desired 'intrinsic' structure is frozen inside the CBS segment of the loop by means of artificially increased bending rigidities in that section. The resulting structure of the U loop, shown in figure 4*a*, is able to accommodate the CAP protein while its energy remains almost unchanged. Yet the likely disruption of the protein–DNA interactions due to the close proximity of CAP and the upstream DNA-binding 'hand' of the *lac* repressor makes the viability of this ternary complex uncertain. The O loop, in contrast, can accommodate the CAP only if its length is increased to 82–84 bp (figure 4*b*), otherwise, the mimicked CBS forces the loop to adopt a sterically prohibitive structure. Physically, the increase in the loop length corresponds to the dislocation of the upstream DNA-binding hand of the *lac* repressor from its preferred site, accompanied by an increase in the interaction energy. However, the resulting drop in the elastic energy of the loop (figure 4*c*) is more than sufficient to compensate for the relocation cost. (In contrast, the U loop does not benefit from a similar increase in l, cf. figure 4*c*.) Our results conform with experimental observations that the formation of the ternary complex indeed results in the relocation of the binding hand of *lac* repressor by 6 bp (Perros *et al*. 1996; Hudson & Fried 1990), while our energetic estimates for ternary complex formation (Balaeff *et al*. 2004) agree with those of experiments (Hudson & Fried 1990).

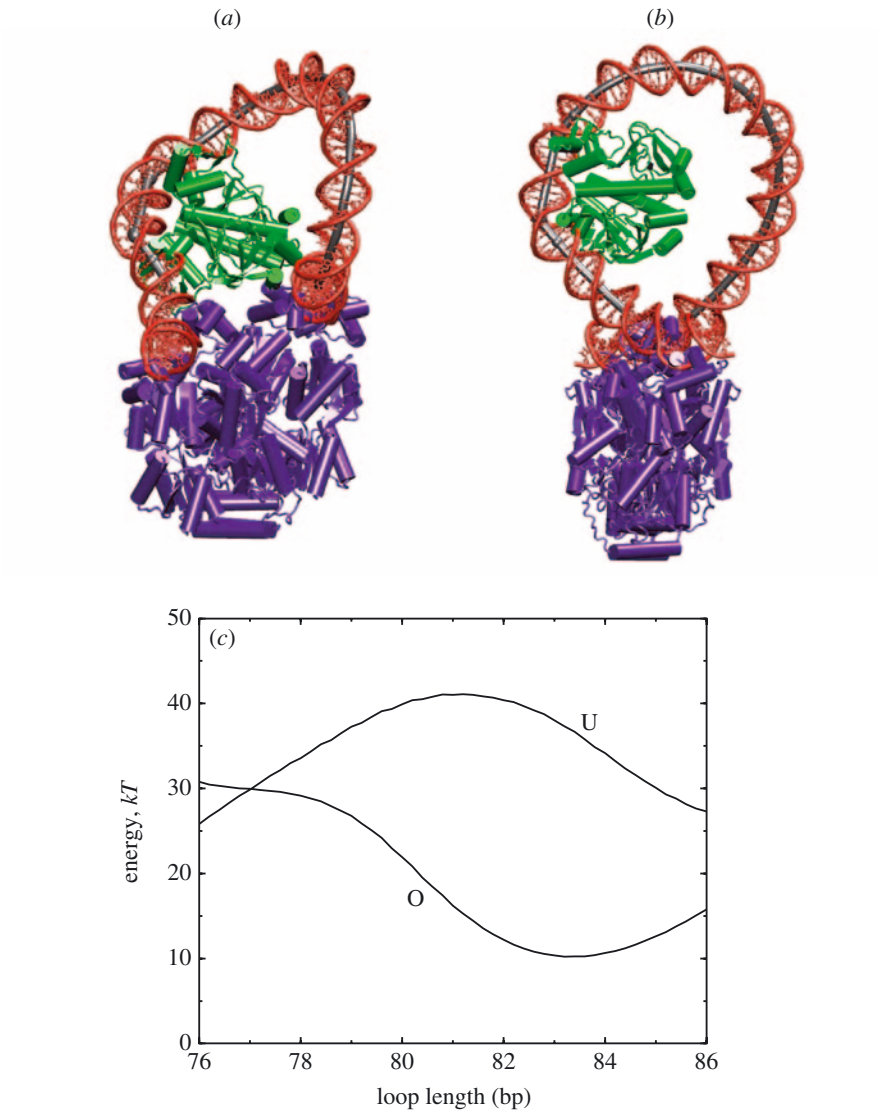

Figure 4. Structures of the ternary complex of CAP, *lac* repressor and the DNA loop. (*a*) A complex with the 76 bp-long underwound loop. (*b*) A complex with the 83 bp-long U loop. The views are rotated by *ca.* 90° around the vertical axis with respect to the *lac* repressor view in figure 1*b*. The *lac* repressor is coloured purple and CAP is coloured green. The grey tube shows the elastic-rod model of the 76 bp loop; the all-atom DNA structure (red) is built on top of the rod as described in Balaeff *et al.* (2004). (*c*) The elastic energy of the loops plotted versus l shows the optimal length of each loop.

4. Kinetics of DNA loop formation and finite-temperature effects

(*a*) *Worm-like chain model and loop-formation process*

In § 2 we presented the governing equations for the overdamped Brownian dynamics of a Kirchhoff rod as a simplified model for DNA with a length of the order of

one persistence length l_{p} or more. Here we begin the investigation of loop-formation kinetics at finite temperature with a yet more simplified rod model. Our motivation is guided by the high computational cost of a kinetic study of looping, which requires a detailed exploration of the configuration space accessible to the rod. Hence we consider the worm-like chain (WLC) model defined by the isotropic bending energy

$$\mathcal{U} = \tfrac{1}{2} k_{\mathrm{B}} T l_{\mathrm{p}} \int_0^l |\boldsymbol{r}''(s,t)|^2 \, \mathrm{d}s, \tag{4.1}$$

as opposed to (2.6). The WLC is further subject to the scalar inextensibility condition $|\boldsymbol{r}'(s,t)| = 1$, an approximation of (2.1). The WLC is described by three centreline coordinates only, compared with the 13 degrees of freedom needed to specify the full Kirchhoff model. Looping kinetics, also-called cyclization kinetics in polymer science, has received ample attention in the case of flexible polymers (e.g. Pastor *et al*. 1996 and references therein) and some attention in the case of (semi-flexible) worm-like chains (e.g. Merlitz *et al*. 1998; Dua & Cherayil 2002). The loop-formation process may be modelled as a two-state kinetic process composed of a reversible phase of approach of two active sites on the DNA within the reaction radius a_{reac} at rate constants k_{on} and k_{off}, and a phase where the actual chemical process 'closing' the loop takes place irreversibly at rate q. The process is said to be diffusion controlled when $q \gg k_{\mathrm{off}}$ and reaction controlled when $q \ll k_{\mathrm{off}}$. We assume from the outset that loop formation is diffusion controlled. Because DNA is relatively large and slowly diffusing this appears to be biologically relevant, although examples of reaction-controlled loop formations in DNA are also known. Prime quantities of interest are therefore the loop-formation rate k_{on} or the diffusional loop-formation time $\tau_{\mathrm{D}} \approx k_{\mathrm{on}}^{-1}$. Merlitz *et al*. (1998) used an extensible helical (i.e. twisted) WLC Brownian dynamics solver with hydrodynamic interactions in order to perform loop-formation kinetics experiments for a chain of fixed length. We focus instead on the dependence of τ_{D} on the WLC length l. The loop-formation time τ_{D} can be computed numerically by sampling a WLC configuration of a given length from its thermal equilibrium distribution and then by integrating the WLC equations of motion forward in time to measure τ_{D}.

(*b*) *Monte Carlo and Brownian dynamics algorithms*

In order to proceed we implemented a combined Monte Carlo (MC) and Brownian dynamics (BD) solver for the WLC. Both solvers are based on the discretized WLC energy for a chain of length l composed of N discrete segments with constant length b and $N+1$ vertices, each attached to a bead to model hydrodynamic behaviour:

$$\mathcal{U}_{\mathrm{disc}} = \frac{\tilde{l}_{\mathrm{p}} k_{\mathrm{B}} T}{2b} \sum_{i=1}^{N-1} \theta_i^2, \tag{4.2}$$

where θ_i is the angle between segments $i-1$ and i and where $\tilde{l}_{\mathrm{p}}$ is a modified persistence length corrected for discretization effects (Klenin *et al*. 1998). In order to generate the WLC thermal equilibrium distribution we use a pivot Metropolis MC algorithm (Madras & Sokal 1988), in which a new chain configuration is generated from an existing one by selecting one end of the chain of random length and pivoting it about a random axis by a random angle α within the interval $[-\delta_\alpha, \delta_\alpha]$. A

new configuration is accepted subject to the Metropolis criterion, which defines the acceptance probability as $p = \min[1, \exp[-(\mathcal{U}_{\rm new} - \mathcal{U}_{\rm old})/k_{\rm B}T]]$. The rotation angle interval for α is adjusted so that 30% of newly generated configurations are accepted. Note that chain inextensibility is automatically enforced by the rotational nature of the pivot moves.

In order to generate configuration space trajectories beginning with given equilibrium configurations we have adapted the flexible, freely draining BD bead-rod algorithm of Hinch (1994) to the discrete WLC. Hinch's algorithm enforces inextensibility by introducing constraint forces $\boldsymbol{T}_i$ collinear with segments $(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)$ and by solving for the overdamped equations of motion

$$\xi \dot{\boldsymbol{r}}_i = \boldsymbol{f}_i^{\rm bend} + \boldsymbol{f}_i^{\rm stoc} + \boldsymbol{T}_i - \boldsymbol{T}_{i-1} \tag{4.3}$$

for vertex $0 \leqslant i \leqslant N$, where ξ is the viscous drag coefficient, $\boldsymbol{f}_i^{\rm bend}$ is the net bending force derived from (4.2), and $\boldsymbol{f}_i^{\rm stoc}$ is the stochastic force verifying the discretized version of (2.9). The constraint forces $\boldsymbol{T}_i$ are resolved by applying the inextensibility condition $\dot{\boldsymbol{r}}_i \cdot \boldsymbol{r}_i = 0$, leading to the solution of a linear tridiagonal system for $\boldsymbol{T}_i$ at each time-step. Note that the viscous drag coefficient of the discretized WLC model ξ is consistent with the viscous drag coefficient per unit length ζ of the continuous WLC model, appearing in equation (2.7). The need for different drag coefficients naturally appears when equation (2.7) is discretized to yield (4.3). Further, observe that the locally anisotropic viscous drag force on the right-hand side of (2.7) is not modelled as such in the discrete WLC model. Here the effects of drag anisotropy are accounted for only in an average sense. Indeed, in the freely draining approximation ξ is modelled by the isotropic Stokes law of friction, $\xi = 6\pi\eta r_{\rm fric}$. However, the bead radius $r_{\rm fric}$ is computed so that the total viscous drag of the bead-rod chain is identical to that of a cylindrical rod with hydrodynamic radius corresponding to DNA (Klenin *et al.* 1998). Because of the great computational cost, we use a relatively coarse spatial discretization with $b = 10$ nm in this preliminary study. A midpoint rule algorithm is used to integrate the system in time (Hinch 1994).

(*c*) *Numerical results*

Figure 5*a* shows the probability density for the normalized end-to-end distance R/l for various WLC lengths l and a fixed persistence length $l_{\rm p} = 50$ nm, as computed from ensemble averages (MC) and time averages (BD). The results from both methods collapse well, although the convergence of the BD method is very slow compared with MC computations. The shorter the WLC length, the more restricted are its excursions away from its most probable configuration. This provides some justification for the zero-temperature approach used in § 3. In order to assess the likelihood of forming a loop at equilibrium the inset plot shows the cumulative probability for R to lie within a sphere of radius $a_{\rm reac} = 5$ nm and 10 nm, respectively. At equilibrium the maximum cumulative probability is realized for chains of length roughly $l = 3l_{\rm p}$ and there is a sharp drop in probability away from the maximum, consistent with our intuition. Loop formation is dominated by the enthalpic cost of bending for short chains $l < 3l_{\rm p}$, while it is dominated by entropic effects—reflected in the large number of open configurations—for long chains $l > 3l_{\rm p}$. The size of the reaction radius critically affects the loop-formation probability—in the example shown the maximum rises roughly by a factor of 5 in going from $a_{\rm reac} = 5$ nm to $a_{\rm reac} = 10$ nm,

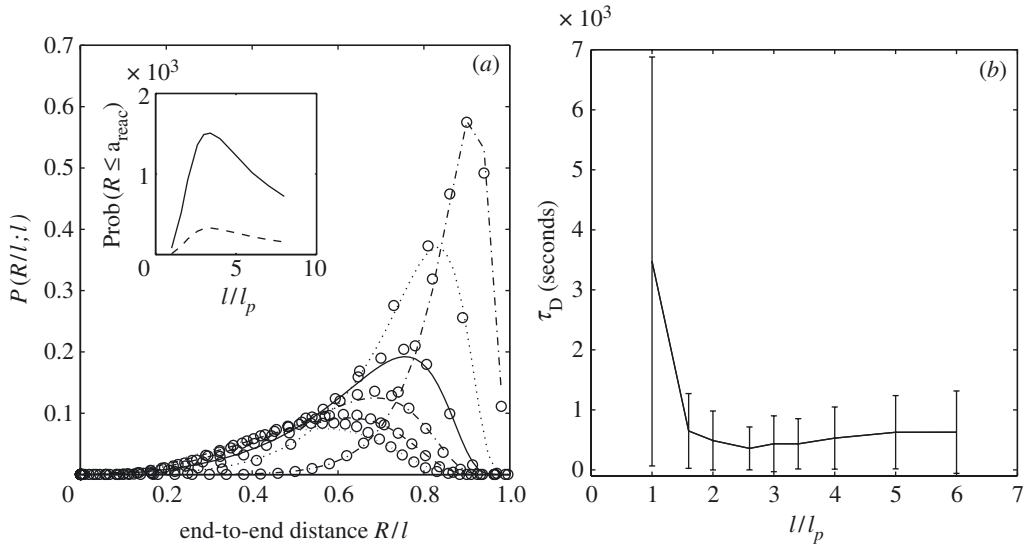

Figure 5. (*a*) Equilibrium probability density function $P(R/l;l)$ for the end-to-end distance R/l with $l_p = 50$ nm and various chain lengths $l/l_p = 6, 5, 4, 3, 2, 1$, from left to right. Lines and symbols represent MC (ensemble-averaged) and BD (time-averaged) results, respectively. The results for $l/l_p = 1$ have been multiplied by 0.5 for clarity of the figure. The inset shows the cumulative probability that R lies within a reaction radius of $a_{reac} = 5$ nm (– – –) and 10 nm (——), respectively. (*b*) The diffusional loop closure time τ_D as a function of the normalized WLC length l/l_p for a reaction radius $a_{reac} = 5$ nm. Vertical bars represent the dispersion of the times around their mean values.

with a corresponding increase in probability for longer chains. Finally, in figure 5*b* we show the diffusive loop-formation time τ_D as a function of the non-dimensional WLC length l/l_p, for a reaction radius of $a_{reac} = 5$ nm, corresponding roughly to the distance existing between the two binding hands of the *lac* repressor. The vertical bars indicate the dispersion of τ_D around the plotted mean values. As for the equilibrium data (figure 5*a*) there exists an optimal minimal loop closure time for chains with length between 2 and 4 persistence lengths, beyond which the enthalpic and entropic regimes dominate, respectively. Based on the WLC model we conclude that τ_D is of the order of a few milliseconds for a WLC with $l \sim l_p$. (Note that due to the coarse resolution of the WLC model, the smallest chain length considered is $l/l_p = 1$, which corresponds approximately to twice the length of the *lac* repressor loop discussed in § 3.)

5. Conclusion and outlook

In this paper, we have described a hierarchy of models starting with a generalized Kirchhoff elastic rod as it applies to the conformation and overdamped motion of DNA with length-scales of the order of the persistence length or more. Our results have focused on the equilibrium configurations of DNA loops constructed with an iterative continuation solver in the context of loops induced by *lac* repressor and CAP, two proteins active in the genetic regulation of the *lac* operon of *Escherichia coli*. We have also included a preliminary study of the equilibrium properties and

kinetics of loops in a worm-like chain using Monte Carlo and Brownian dynamics simulations. We close with a brief discussion of avenues for further work.

First, the estimated force of the protein–DNA interactions (see Balaeff *et al*. 2003) may be used to model the changes in the structure of the *lac* repressor that are likely to occur due to the stress of the bent DNA loop. This can be achieved via a molecular dynamics (MD) simulation for the all-atom structure of the *lac* repressor, in which the estimated elastic forces and torques are applied to the ends of the protein-bound DNA segments. The forces must be recomputed iteratively as the *lac* repressor structure and, consequently, the boundary conditions for the elastic rod change during the MD simulation. One can either employ the elastic forces obtained from the static calculations, or extract thermally averaged/instantaneous forces from an MC/BD simulation of the loop. Such a coupling between the all-atom and coarse-grained simulations of the different parts of the protein–DNA complex gives rise to a multiscale simulation technique that has a significant potential for further applications in biomolecular modelling (Villa *et al*. 2004).

Second, the elastic-rod solutions may serve as a scaffold on which to build all-atom structures of the whole DNA loop (Balaeff *et al*. 2004). The all-atom structures can be used as reasonable starting points for MD simulations of the whole ternary complex of CAP, *lac* repressor and DNA, which should become possible with the advent of massively parallel computers. Alternatively, one could subject the all-atom structure of a particularly interesting section of the loop (a binding site for a protein, or a site with a non-trivial intrinsic geometry) to multiscale simulations, similar to those described above.

Finally, the combination of Monte Carlo and Brownian dynamic simulations can be generalized to account for a more complex helical (i.e. twisted) WLC model as well as hydrodynamic interactions with the ultimate goal of comparison with single-molecule experimental data on DNA–protein interaction in the context of the *lac* repressor (e.g. Finzi & Gelles 1995; Lia *et al*. 2003).

A.B. and K.S. acknowledge support from the grants from the Roy J. Carver Charitable Trust, the NIH (PHS 5 P41 RR05969) and the NSF (BIR 94-23827EQ). C.R.K. thanks J. Langowski for making available CORCHY++ software (Klenin *et al*. 1998), which was used as a basis for developing the WLC MC and BD solvers. L.M. acknowledges the support of the US Office of Naval Research (YIP) and the US National Institutes of Health.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. 2002 *Molecular biology of the cell*, 4th edn. Garland Science.

Bader, G. & Ascher, U. 1987 A new basis implementation for a mixed order boundary value ODE solver. *SIAM J. Sci. Statist. Comput.* **8**, 483.

Balaeff, A., Mahadevan, L. & Schulten, K. 1999 Elastic rod model of a DNA loop in the lac operon. *Phys. Rev. Lett.* **83**, 1900.

Balaeff, A., Mahadevan, L. & Schulten, K. 2003 Modeling DNA loops using the theory of elasticity. E-print archive arXiv.org, pub. no. 0301006. (Available at http://arxiv.org/abs/physics/0301006.)

Balaeff, A., Mahadevan, L. & Schulten, K. 2004 Structural basis for cooperative DNA binding by CAP and lac repressor. *Structure* **12**, 123–132.

Busby, S. & Ebright, R. 1999 Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293**, 199.

Coleman, B. D., Tobias, I. & Swigon, D. 1995 Theory of the influence of end conditions on self-contact in DNA loops. *J. Chem. Phys.* **103**, 9101.

Cox, R. G. 1970 The motion of long slender bodies in a viscous fluid. Part 1. General theory. *J. Fluid Mech.* **44**, 791.

Cox, R. G. 1971 The motion of long slender bodies in a viscous fluid. Part 2. Shear flow. *J. Fluid Mech.* **45**, 625.

Dua, A. & Cherayil, B. J. 2002 The dynamics of chain closure in semiflexible polymers. *J. Chem. Phys.* **116**, 399.

Edelman, L. M., Cheong, R. & Kahn, J. D. 2003 Fluorescence resonance energy transfer over ≈130 basepairs in hyperstable lac repressor-DNA loops. *Biophys. J.* **84**, 1131.

Finzi, L. & Gelles, J. F. 1995 Measurement of lactose repressor-mediated loop formation and breakdown in sinngle DNA molecules. *Science* **267**, 378.

Hinch, E. J. 1994 Brownian motion with stiff bonds and rigid contraints. *J. Fluid Mech.* **271**, 219.

Hudson, J. M. & Fried, M. G. 1990 Co-operative interactions between the catabolite gene activator protein and the *lac* repressor at the lactose promoter. *J. Mol. Biol.* **214**, 381.

Kirchhoff, G. R. 1883 *Vorlesungen über mathematische Physik, Mechanik, Vorlesung 28*, Dritte Auflage. Leipzig: B. G. Teubner.

Klenin, K., Merlitz, H. & Langowski, J. 1998 A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.* **74**, 780.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, A. G. & Lu, P. 1996 Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247.

Lia, G., Bensimon, D., Croquette, V., Allemand, J.-F., Dunlap, D., Lewis, D. E. A., Adhya, S. & Finzi, L. 2003 Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc. Natl Acad. Sci. USA* **100**, 11 373.

Love, A. E. H. 1927 *A treatise on the mathematical theory of elasticity*. Dover.

Madras, N. & Sokal, A. D. 1988 The pivot algorithm: a highly efficient Monte Carlo algorithm for the self-avoiding random walk. *J. Stat. Phys.* **50**, 109.

Mahadevan, L. & Keller, J. B. 1996 Coiling of flexible ropes. *Proc. R. Soc. Lond.* A **452**, 1679.

Manning, R. S., Maddocks, J. H. & Kahn, J. D. 1996 A continuum rod model of sequence-dependent DNA structure. *J. Chem. Phys.* **105**, 5626.

Matthews, K. 1992 DNA looping. *Microbiol. Mol. Biol. Rev.* **56**, 123.

Merlitz, H., Rippe, K., Klenin, K. V. & Langowski, J. 1998 Looping dynamics of linear DNA molecules and the effect of DNA curvature: a study by Brownian dynamics simulation. *Biophys. J.* **74**, 773.

Oehler, S., Eismann, E. R., Krämer, H. & Müller-Hill, B. 1990 The three operators of the lac operon cooperate in repression. *EMBO J.* **9**, 973.

Olson, W. K. 1996 Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* **6**, 242.

Olson, W. K. & Zhurkin, V. B. 2000 Modeling DNA deformations. *Curr. Opin. Struct. Biol.* **10**, 286.

Olson, W. K. (and 14 others) 2001 A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229.

Olson, W. K., Marky, N. L., Jernigan, R. L. & Zhurkin, V. B. 1993 Influence of fluctuations on DNA curvature. A comparison of flexible and static wedge models of intrinsically bent DNA. *J. Mol. Biol.* **232**, 530.

Öttinger, H. C. 1996 *Stochastic processes in polymeric fluids*. Springer.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. E. & Berman, H. M. 1996 Structure of the CAP-DNA complex at 2.5 Å resolution: a complete picture of the protein–DNA interface. *J. Mol. Biol.* **260**, 395.

Pastor, R. W., Zwanzig, R. & Szabo, A. 1996 Diffusion limited first contact of the ends of a polymer: comparison of theory with simulation. *J. Chem. Phys.* **105**, 3878.

Perros, M., Steitz, T. A., Fried, M. G. & Hudson, J. M. 1996 DNA looping and lac repressor–lac CAP interaction. *Science* **274**, 1929.

Ptashne, M. 1992 *A genetic switch*, 2nd edn. Blackwell Science.

Schleif, R. 1992 DNA looping. *A. Rev. Biochem.* **61**, 199.

Schlick, T. 1995 Modeling superhelical DNA: recent analytical and dynamic approaches. *Curr. Opin. Struct. Biol.* **5**, 245.

Strick, T. R., Allemand, J.-F., Bensimon, D. & Croquette, V. 2000 Stress-induced structural transitions in DNA and proteins. *A. Rev. Biophys. Biomol. Struct.* **29**, 523.

Tsodikov, O. V., Saecker, R. M., Melcher, S. E., Levandoski, M. M., Frank, D. E., Capp, M. W. & Record Jr, M. T. 1999 Wrapping of flanking non-operator DNA in *lac* repressor–operator complexes: implications for DNA looping. *J. Mol. Biol.* **294**, 639.

Villa, E., Balaeff, A., Mahadevan, L. & Schulten, K. 2004 Multi-scale method for simulating protein-DNA complexes. *Multiscale Model. Simul.* (In the press.)

Westcott, T. P., Tobias, I. & Olson, W. K. 1997 Modeling self-contact forces in the elastic theory of DNA supercoiling. *J. Chem. Phys.* **107**, 3967.