Modeling the Lac repressor-operator assembly: The influence of DNA looping on Lac repressor conformation

David Swigon[†‡§¶], Bernard D. Coleman[‡¶], and Wilma K. Olson[†¶]

Departments of [†]Chemistry and Chemical Biology and [‡]Mechanics and Materials Science, Rutgers, The State University of New Jersey, Piscataway, NJ 08854; and [§]Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260

Communicated by Donald M. Crothers, Yale University, New Haven, CT, May 9, 2006 (received for review June 24, 2005)

Repression of transcription of the *Escherichia coli* Lac operon by the Lac repressor (LacR) is accompanied by the simultaneous binding of LacR to two operators and the formation of a DNA loop. A recently developed theory of sequence-dependent DNA elasticity enables one to relate the fine structure of the LacR–DNA complex to a wide range of heretofore-unconnected experimental observations. Here, that theory is used to calculate the configuration and free energy of the DNA loop as a function of its length and base-pair sequence, its linking number, and the end conditions imposed by the LacR tetramer. The tetramer can assume two types of conformations. Whereas a rigid V-shaped structure is observed in the crystal, EM images show extended forms in which two dimer subunits are flexibly joined. Upon comparing our computed loop configurations with published experimental observations of permanganate sensitivities, DNase I cutting patterns, and loop stabilities, we conclude that linear DNA segments of short-to-medium chain length (50–180 bp) give rise to loops with the extended form of LacR and that loops formed within negatively supercoiled plasmids induce the V-shaped structure.

lac operon | sequence-dependent DNA elasticity | DNase I footprinting

Many genetic processes are controlled by proteins that bind at separate, often widely spaced, sites on DNA and hold the intervening double helix in a loop (1–3). The classical example is the *lac* operon of *Escherichia coli* (4). The Lac repressor (LacR) is a tetrameric protein assembly that represses the expression of the *lac* operon by simultaneously binding to two DNA sites, i.e., operators, in the vicinity of the nucleotides at which transcription starts. The structure and elastic properties of DNA determine which spacings of the operators are optimal for functionality. For the *lac* operon, a change in spacing by five to six nucleotides can induce a 50-fold alteration in the efficiency of repression (5, 6).

Although there is a large amount of literature on genetic and biochemical aspects of expression in the *lac* system, less is known about the actual configuration of the LacR–DNA loop assembly. In the crystalline state the two dimer subunits of LacR are joined to form a V (7, 8), and contact with DNA is made at the tips of each arm of the V (Fig. 1). On the other hand, electron microscopy and solution studies (9–12) indicate that the angle between the dimer subunits, i.e., the angle of aperture α, can vary. A change in α affects the configuration of the DNA loop through its influence on the distance and orientation of the operators. We are here concerned with loops formed between the primary operator site O1 and the weaker auxiliary site O3. As each operator binds to the protein in one of two possible orientations, there are four distinct loop types that are analogous to those considered by Geanacopoulos *et al.* (13) in their treatment of DNA loops in the *E. coli gal* operon. We write A1, A2, P1, and P2 for these loop types, where the A and P refer to antiparallel and parallel orientations of operators (see Fig. 1*C*).

Published models for the LacR-mediated DNA loop (7, 14, 15) take into account selected aspects of the experimental literature and two of the four loop types. Balaeff and coworkers (14) used their model to calculate the effect of linking number on loop energies and configurations. Kahn and coworkers (10, 11) studied sequences other than the wild type and proposed models that account for the possibility of an extended LacR conformation with large angle of aperture. The models published to date do not account for the influence of DNA sequence and protein flexibility on the configuration of the O3–O1 loop as we do here. We consider all four orientations of the bound operators and the susceptibility of LacR to transitions between two states, V and E. In state V the tetramer adopts an essentially rigid V-shaped conformation with α fixed at a value near to 34°. In E the conformation fluctuates with α varying in a range that includes 180°, the value at which the tetramer is fully extended. We employ a recently developed theory of sequence-dependent DNA elasticity (16, 17) to calculate the

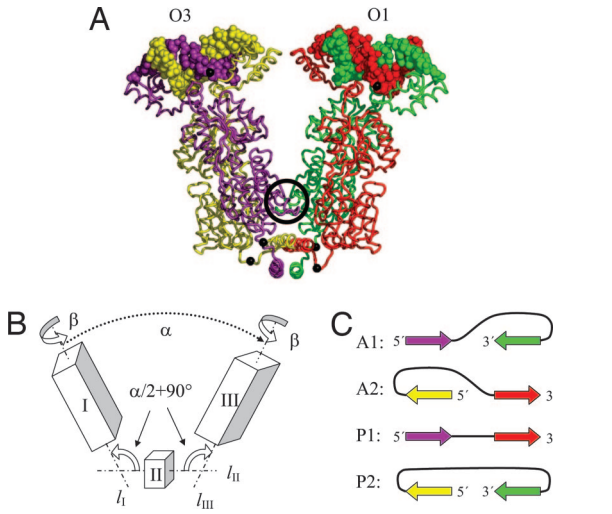

Fig. 1. Structure of the tetrameric Lac repressor protein in complex with O1 and O3 operator segments. (*A*) Atomic-level model obtained by composition of available x-ray data (see text). Chains A (yellow) and B (violet) form one dimer unit, and chains C (red) and D (green) form the other. The black spheres on protein represent the C^{α} atoms of Gln-335 and those on DNA the P atoms of the central base pairs. The black circle marks the dimer contact interface found in the crystal structure. (*B*) A schematic representation of LacR opening. The rigid domains I (residues 1–332 of chains A and B and the DNA bound to these chains) and III (residues 1–332 of chains C and D and the bound DNA) are connected to domain II (residues 340–354 of chains A, B, C, and D) by two hinges. The axes of rotational symmetry of the three domains are l_I, l_{II}, and l_{III}. (*C*) DNA loop types. The color-coded arrows depict the 5′–3′ direction of the sequence strand on LacR in the four possible orientations of DNA on the tetramer. The colors correspond to those of associated DNA and protein chains in part *A*.

Conflict of interest statement: No conflicts declared.

¶To whom correspondence may be addressed. E-mail: swigon@pitt.edu, bcoleman@jove.rutgers.edu, or olson@rutchem.rutgers.edu.

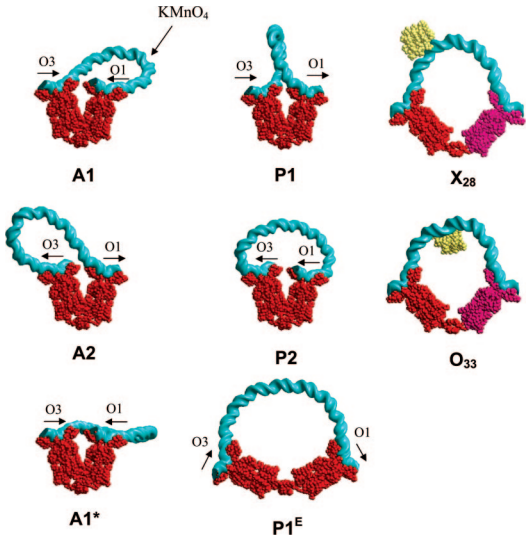

Fig. 2. Minimum energy configurations of DNA fragments complexed with the LacR tetramer. A1, P1, A2, P2, A1*, and $P1^E$: preferred arrangements of the wild-type Lac promoter with the specified loop types; X_{28} and O_{33}: optimum structures of a shortened, 74-bp construct (pHK74) bound concomitantly to state E of LacR and to a DNase I molecule located in a favorable (n = 28) position and an unfavorable (n = 33) position. The 5′–3′ orientations of operators in the loops of wild-type DNA are indicated by arrows. The properties of the wild-type loops are given in Table 1. The site of observed $KMnO_4$ hypersensitivity (22) is shown for A1.

configurations and free energies of DNA loops as functions of the linking number, the end conditions, and the loop length. Our calculations of the free energy account for thermal fluctuations and contributions from elastic and electrostatic energies, and imply that if the free energy penalty for a transition of LacR from V to E is sufficiently small, linear DNA segments of 50–180 bp induce state E of LacR, and loops formed in negatively supercoiled plasmids induce state V.

Supporting evidence for the existence of loops with the extended form of LacR comes from the analysis of DNase I footprinting experiments. Hochschild and Ptashne (18) were the first to note that the cleavage of looped DNA by DNase I restriction endonuclease differs from that found for the same DNA in an open state. The cleavage pattern shows sites of enhanced phosphodiester cutting that are spaced approximately 10.5 bp apart and sites of suppressed cutting that occur midway between the sites of enhancement. The crystal structure of DNase I complexed to DNA (19) shows an enzyme-induced bend in the targeted DNA in the direction of the major groove and away from the protein, indicating that DNase I has a preference for binding to the convex side of a DNA loop. As we explain below, knowledge of the location of cleavage enhancement or suppression sites can be used to obtain information about the structure of looped DNA in solution. Our analysis of the DNase I footprinting profiles of short, LacR-mediated loops (20) indicates that, under the conditions of these experiments, the tetramer adopts the E state.

We draw other new structural inferences about the conformation of LacR from the analysis of available measurements of (*i*) permanganate sensitivity and (*ii*) gel mobility, finding that the tetramer adopts the V state in *i* and the E state in *ii*. Because there are loop lengths for which our calculations yield several preferred configurations with distinct loop types but comparable free energies, we propose an experimental approach to the problem of determining the relative concentrations of the loop types.

Results

Configurations of the Wild-Type O3–O1 Loop. Calculated minimum energy configurations of the wild-type O3–O1 loop are shown in Fig. 2 for various combinations of orientation, linking number, LacR conformation, and nucleotide sequence. Because the O3–O1 sequence is not palindromic, the antiparallel loops (A1 and A2) are not congruent, and the theory we employ predicts that local DNA configurations can be very sensitive to orientation. For 92-bp wild-type DNA the configurations labeled A1, A2, P1, and P2 in Fig. 2 minimize G_{DNA}, the free energy of the DNA loop, over all values of Lk for the indicated loop types. (Whenever we refer to A1, A2, P1, or P2, the LacR complex is assumed to be in state V.) The loop labeled A1* in Fig. 2 has the same orientation and DNA sequence as that labeled A1 but differs in linking number [i.e., $Lk(A1^*) = Lk(A1) + 1$] and therefore has a different minimum energy configuration and free energy [$G_{DNA}(A1^*) > G_{DNA}(A1)$]. The loop labeled $P1^E$ in Fig. 2 minimizes G_{DNA} over all values of Lk and loop types for tetramers in state E bound to wild-type DNA.

For wild-type loops, calculated values of G_{DNA} are given in Table 1. If G_{LacR}, the free energy penalty for the transition of LacR from V to E, lies in the apparently feasible range, 1.8–9.4 kT, the $P1^E$-type loop has the lowest total free energy and is thus most likely to occur in solution. (The probability of occurrence of a loop depends on the energies, G_{O1} and G_{O3}, of LacR binding to operons O1 and O3, and the difference, ΔG, between G_{DNA} and the free energy of an unbound DNA segment of identical length.) If G_{LacR} is close to the upper limit of estimated values, the free energy of the $P1^E$ type is close to that of the two antiparallel types, A1 and A2, for LacR in state V. The P1 type configuration (with LacR in state V), which has been offered as a model for LacR-induced DNA looping in a

Table 1. Calculated energy values, in *kT*, and configurational parameters for different wild-type (O1–O3) LacR-mediated DNA loops

					Φ			G_{DNA}	
Loop	α, deg	β, deg	Lk	Ψ	100 mM	10 mM	G_{LacR}	100 mM	10 mM
A1	34	33	8	31.8	9.7	44.9	—	55.3	90.5
A2	34	33	8	32.9	9.7	44.8	—	55.8	90.9
P1	34	33	9	38.6	9.8	46.4	—	60.0	96.6
P2	34	33	9	45.5	9.8	44.8	—	69.1	104.1
A1*	34	33	9	38.9	9.7	44.3	—	62.1	96.7
$P1^E$	112	−4	9	22.9	9.6	42.0	1.8–9.4	47.5–55.1	79.9–87.5
Free	—	—	—	0	9.4	41.0	—	22.4	54.0

Ψ, elastic energy; Φ, electrostatic energy at high and low salt; G_{LacR}, free energy of LacR opening; G_{DNA}, free energy of the loop at room temperature under the given ionic conditions. ''Free'' refers to the unbound, linear DNA chain of the same wild-type (O1–O3) sequence: **GGCAGTGAGC G CAACGAATT** AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC AC*TTTATG*CT TCCGGCTCGT ATGTTGTGTG G **AATTGTGAGC G GATAACAATT.** Here, the O1 and O3 sequences are shown in boldface, and the KMnO4 hypersensitive element (22) is shown in italics.

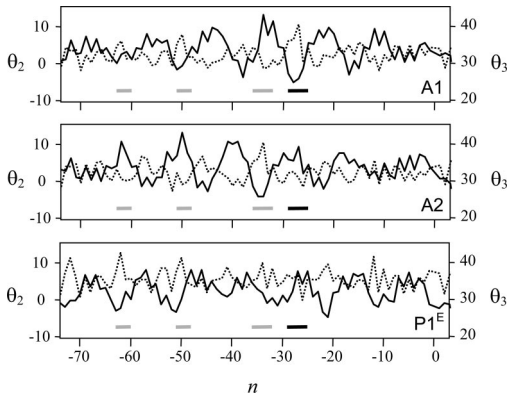

Fig. 3. Sequence-dependent deformations of roll θ_2 (solid curves) and twist θ_3 (dotted curves), in degrees, of DNA in the three wild-type LacR-mediated loops of lowest free energy (A1, A2, and $P1^E$). The location of the TTTAT fragment hypersensitive to attack by $KMnO_4$ in supercoiled plasmids (22) is indicated by a black bar. The locations of other TTA sites within the looped DNA are denoted by gray bars. Base-pair position *n* is defined with respect to the transcription start site as 0. Only A1 shows the high twist and negative roll expected at the hypersensitive site.

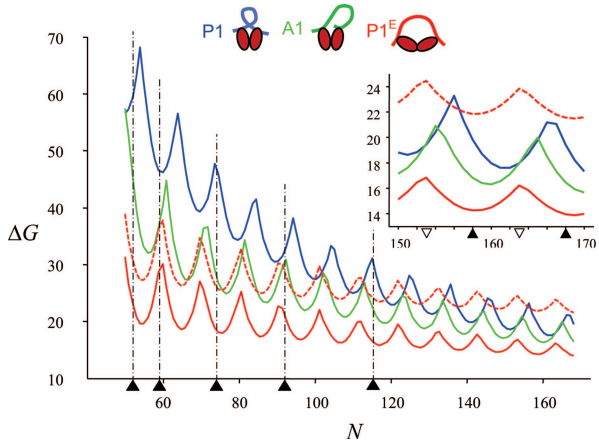

Fig. 4. Graphs of ΔG, the difference between G_{DNA}, and the free energy G_{free} of a free (unbound) DNA segment of the same length, versus loop length *N* (distance between operator sites) for three loop types corresponding to selected configurations in Fig. 2: A1 (green line), P1 (blue line), and $P1^E$ with $G_{LacR} = 1.8\ kT$ (solid red line) and $9.4\ kT$ (dashed red line). The solid triangles and vertical dash-dotted lines denote the chain lengths of experimentally characterized 52-, 59-, 74-, 92- (wild-type), and 114-bp loops. (*Inset*) Detail of the graph of ΔG versus *N*. Observed chain lengths of the most stable (158 and 168 bp) and least stable (153 and 163 bp) loops (20) are indicated by solid and hollow triangles. Data are obtained at 10 mM monovalent salt for a DNA homopolymer with properties based on averages over all base-pair steps. All ΔG values are expressed in units of *kT*.

number of papers (e.g., refs. 7 and 14), is appreciably higher in free energy than the antiparallel loops and the $P1^E$ loop and is therefore unlikely to occur in solution.

Permanganate Sensitivity. Addition of potassium permanganate, a probe for locally distorted DNA (21), to the wild-type LacR-mediated O3–O1 loop reveals a hypersensitive base-pair region, H = TTTAT, located ≈37 bp upstream of the O1 site (22). For DNA, TA steps are known to be sites of permanganate attack (23), and the configuration of such steps in the *H* region of the O3–O1 loop is apt to resemble the $KMnO_4$-sensitive TA step (with high twist and negative roll) found in the crystal complex of DNA with the Epstein–Barr virus origin-binding protein (24). Significantly, the *H* region of the optimized antiparallel A1 loop shows the requisite pattern of localized overtwisting and negative roll, i.e., exposure of the major groove atoms with $\theta_3 > 34°$, $\theta_2 < 0$ (Fig. 3). The AT-rich site of permanganate attack lies near the apex of the A1 structure. The corresponding thymines are less accessible to similar chemical attack in the two competing low-energy configurations (A2 and $P1^E$), with the potentially reactive major-groove atoms facing toward the interior of the closed loops. Other thymines, such as those on the TTTAC segment located 7 bp upstream of the reactive TTTAT site, are unexposed in the A1 configuration but are accessible to permanganate attack in the A2 and $P1^E$ structures.

Length Dependence of Loop Free Energy. Fig. 4 shows the results for LacR-induced DNA loops made up of DNA that is materially uniform, with elastic and intrinsic properties at each base-pair step corresponding to the average properties over all base-pair step combinations. The quantity plotted on the graph is ΔG, the difference between G_{DNA} and the free energy G_{free} of a free (unbound) DNA segment of the same length at 10 mM salt. The graphs shown are for the antiparallel (blue) A1-, parallel (green) P1-, and open (red) $P1^E$-type loops of Fig. 2, i.e., the topoisomers of lowest free energy for each loop type. The values of ΔG are somewhat lower than the values calculated when the sequence-dependent properties of the 92-bp wild-type loop are taken into account (e.g., for A1 circa 3 *kT* lower) owing to the smaller elastic constants of a generic base-pair step. Two sets of data are shown for $P1^E$, at the upper and the lower limit of the range of assumed values for G_{LacR}. The observed oscillations with a period of approximately one helical repeat (circa 10.5 bp) are characteristic of ring closure experiments (25, 26) and result from the twisting (i.e., torsional adjustment of DNA ends) required by the end conditions. Throughout the range of *N*, the antiparallel A1 loop has lower energy than the parallel P1 loop. The state that minimizes *G* depends, however, on the assumed value for G_{LacR}: at the lower limit of G_{LacR}, a $P1^E$-type configuration minimizes the free energy throughout the range of *N* shown on the graph.

Fig. 4 *Inset* shows details of the dependence of ΔG on *N* for loops of lengths 150–170 bp. Experimental measurements of the relative abundance of such loops in gel-mobility assays (20) indicate that the lengths of the most stable loops are 158 and 168 bp and that the lengths of the least stable loops are 153 and 163 bp. The close match of these lengths with the calculated locations of minima and maxima of ΔG for the $P1^E$ loop strongly suggests that loops of type $P1^E$ are observed in the cited experiment and hence that LacR adopts a flexible, open conformation in the protein–DNA assembly. As a consequence, the true value of G_{LacR} is likely to be close to the lower limit of the range 1.8–9.4 *kT*.

The dependence of the loop-closure probability on ΔG, the binding affinity of the operator sites, and the concentration of LacR can be determined by analyzing all binding states of the DNA and LacR (27, 28). The values of ΔG extracted from Fig. 4 can thus be checked against experimental data. For example, in the case of a 114-bp $P1^E$ loop, we find $\Delta G = 16.7\ kT + G_{LacR}$. Substitution of the estimated range of G_{LacR} implies that $18.5\ kT < \Delta G < 26.1\ kT$, a prediction in remarkably good agreement with the range of values (18.9–23.6 *kT*) determined from footprinting and gel mobility-shift measurements (27).

DNase I Footprinting. DNase I footprinting studies of LacR-induced DNA looping by Krämer *et al.* (20) examined the binding of LacR to constructs, called pHK74, pHK59, and pHK52, in which the distance between the centers of the operator sites was reduced from the wild-type spacing of 92 bp to 74, 59, and 52 bp, respectively. LacR forms loops with all three constructs, but only pHK74 and pHK52 give rise to apparent DNase I protection patterns like those described by Hochschild and Ptashne (18).

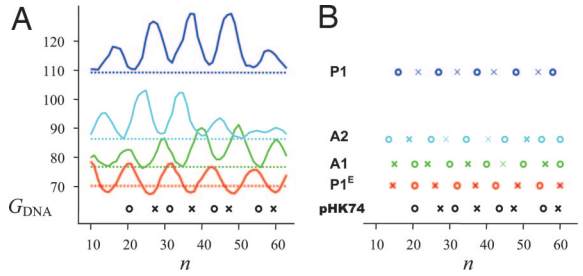

Fig. 5. Comparison of the calculated DNase I cutting patterns of different 74-bp LacR-mediated loops formed from the pHK74 construct with the observed protection pattern (20): **AATTGTGAGC GCTCACAATT** CCACACACTC TAGCAACTAG TGAGCTTGGC TGCAGGTCGA CGGATCCCCTAGA **AATTGTGAGC GCTCACAATT**. The high-affinity symmetrized LacR binding sites are shown in boldface, and the sites of enhanced and diminished DNase I activity are underlined by double and single lines, respectively. (*A*) Plots of free energy of DNA loops, G_{DNA}, versus enzyme binding location *n* for A1 (green line), A2 (cyan line), P1 (blue line), and $P1^E$ (red line) loop types. (The number *n* denotes the distance of bound enzyme from the center of the first operator.) In each case the dotted line corresponds to the free energy of the pHK74 loop in the absence of DNase I: A1, 77.5 *kT*; A2, 87.4 *kT*; P1, 109.4 *kT*; $P1^E$, 70.5 *kT*. (*B*) The predicted sites of enhanced (''x'') and diminished (''o'') DNase I sensitivity correspond, respectively, to the valleys and peaks in the plots of G_{DNA} versus *n* in *A*. (The valleys for which the minimum value of G_{DNA} is larger than the free energy of the pHK74 loop in the absence of DNase I, and hence where no enhancement is expected, are marked with a lightly shaded ''x.'') The experimentally observed sites (20) are noted by the label pHK74.

Calculated free energies $G_{DNA}(n)$ and predicted DNase I cutting patterns of various LacR-mediated loops formed by the 74-bp pHK74 construct are shown in Fig. 5. The observed protection pattern (20) is noted above the binding-site axis *n*, with the locations of enhanced cutting marked by "x" and the sites of reduced cutting by "o." The predicted cutting patterns are obtained by mapping the valleys and peaks of the graphs of G_{DNA} versus *n*. The configurations formed by the binding of LacR to the 74-bp sequence resemble those in Fig. 2 for the 92-bp O3–O1 loop and hence are not shown, but their calculated values of G_{DNA} are listed in the legend of Fig. 5. The $P1^E$-type configuration minimizes the free energy for all but the highest assumed cost of LacR opening. The $P1^E$ configuration also best matches the experimentally observed cutting pattern as well as that for the even shorter pHK52 construct (results not shown). Examples of minimum energy configurations of the loop with DNase I bound at favorable (X_{28}) and unfavorable (O_{33}) sites are shown in Fig. 2.

In the case of the wild-type (92 bp) O1–O3 loop and the intermediate 59-bp loop pHK59, $\Delta G_{DNA}(n)$, the difference in free energy between LacR-mediated loops in the presence and absence of DNase I is positive for all binding locations *n* in all loop types. Thus, the cutting activity of DNase I should not be affected by loop formation, and, indeed, no enhancement of DNase I cutting is observed for the wild-type O1–O3 loop (29) and the pHK59 loop (20).

Discussion

In this article, we have drawn attention to the fact that a DNA loop in the LacR–DNA complex can be one of five types shown in Fig. 2. For four of these, A1, A2, P1, and P2, the LacR is in its V-shaped conformation; for $P1^E$ the LacR has its extended form. Employing a base-pair-level theory of sequence-dependent DNA elasticity, we have calculated the deformational free energy of the DNA in the complex for each type of loop and used the results to relate protein and DNA fine structure to a wide range of heretofore-unconnected experimental observations, including (*i*) loop stabilities (20, 30), (*ii*) permanganate sensitivity (22), and (*iii*) DNase I cutting patterns (20). This work provides an understanding of the effects of chain length, base-pair sequence, protein binding, and supercoiling on DNA looping preferences and, in principle, allows one to design new ways to test specific looped structures in the laboratory.

Loops in Linear DNA Segments. Our analysis of DNase I footprinting and gel-mobility patterns indicates that linear segments of DNA of short-to-medium chain length (50–180 bp) form loops of type $P1^E$ with extended arrangements of the LacR tetramer. The calculations show that such loops have both lower elastic energy and higher entropy than configurations of DNA bound to the V-shaped structure seen in the crystal. The low free energy of the open loop stems from the smaller number of constraints imposed on DNA by the flexible complex.

Loops in Supercoiled Molecules. The situation changes if the loop is incorporated in a plasmid that is subject to the supercoiling found *in vivo* (31). Negative supercoiling is known to enhance and stabilize loop formation of DNA with LacR (30). Moreover, atomic force microscopy images of LacR-mediated loops in negatively supercoiled plasmids reveal the presence of crossings indicative of antiparallel configurations of DNA (12). One expects that antiparallel structures are favored in a negatively supercoiled plasmid. First of all, the linking numbers of the antiparallel A1 and A2 loops are lower than those of all other loop types (Table 1). Secondly, the DNA segments that enter and exit an antiparallel loop show a crossing of negative sign if incorporated in a longer chain segment (Fig. 2). The LacR-mediated antiparallel loops can thus occur as loops at the ends of the plectonemically wound arms of a negatively supercoiled DNA. The parallel P1 and P2 loops show positive crossing(s) and fit more easily into positively supercoiled DNA, whereas the open, parallel $P1^E$ loop does not show any intrachain crossings.

Our analysis of the distribution of roll and twist angles in optimized loop structures shows that published permanganate footprinting profiles (22) are compatible with the antiparallel A1 configuration. Preliminary calculations for 452-bp negatively supercoiled DNA plasmids with a bound LacR tetramer further indicate that the optimal spacing for stable A1 loop formation is 162 bp when $\Delta Lk = -1$, a result in accord with the gel observations of Krämer *et al.* (30). Although the free energies of the two antiparallel loop types are equal and both A1 and A2 orientations give rise to a negative crossing, comparison of the localized distortion of DNA with the permanganate footprinting data suggests a preference for the A1 over the A2 loop. It is possible that the two loop types occur in equal concentrations, but the A2 loop does not have a hypersensitive site for attack by permanganate and hence is not chemically visible. It is also possible that the A2 loop occurs in lower amounts because the associated strong bending of DNA near the O3 binding site (see Fig. 2) decreases the already low affinity of that site for LacR. A more complete analysis of the configurations and free energies of loops formed in supercoiled plasmids can be obtained by an extension of the methods used in the research reported here and will be the subject of future investigation.

Loops in Bacterial Repression. While this paper was under review, an interesting paper (32) was published on the dependence of the repression efficiency of LacR in *E. coli* on the distance between operator sites in the presence and absence of HU. Because reporter activity must be monotonically related to the free energy G_{DNA}, the minima and maxima seen in figures 4a and 5a of ref. 32 can be compared with the minima and maxima in ΔG versus *N* presented in our Fig. 4. Upon doing this, we find that in the presence of HU (figure 4a of ref. 32) the DNA loop is of type P1 (and the LacR in its V-shaped form), whereas in the absence of HU (figure 5a of ref. 32) the loop is of type $P1^E$ (and the LacR extended). One (and we believe the most likely) explanation of this behavior is that the presence of HU induces the sharp bending of DNA required for formation of a loop of type P1, whereas in the absence of HU, as

our results make clear, G_{DNA} is minimized when the loop and the LacR are fully extended.

Future Directions. The application, by Edelman *et al.* (11), of fluorescence resonance energy transfer techniques to distinguish open and closed configurations of LacR-mediated loops in specially designed DNA constructs points out the way to differentiate the dominant states of the wild-type O3–O1 loop in solution. Using a judicious positioning of fluorophores at different sites on DNA, in combination with selective placement of curved inserts, one should be able to discriminate among and promote the occurrence of different looped configurations. For example, curvature in the vicinity of the hypersensitive permanganate site can create a bias for the antiparallel A1 loop type. The optical signal associated with a probe at this location is expected to change in the presence or absence of curvature in the DNA (A. Kapanides, personal communication).

The ability to calculate the loop free energy for any DNA sequence makes it possible to address the role of motions of DNA headpieces in looping, as suggested by recent molecular dynamics simulations (33), and the issue of the experimentally observed dependence of repression efficiency on the number, quality, and distance of operators (e.g., refs. 5, 6, 32, and 34).

Methods

Lac Repressor. Each of the arms of the LacR tetramer (boxes I and III in Fig. 1*B*) is a dimer composed of two polypeptide chains. The two arms of the complex are held together by a four-helix bundle that is located at the base of the V (cube II in Fig. 1*B*) and is further stabilized by a small contact interface between the dimer units (located in the black circle in Fig. 1*A*). Disruption of that interface is required for the V → E transition. Ruben and Roos (9) obtained electron microscopic images of freeze-etched samples of LacR that suggest an opening of the V in such a way that α, instead of being fixed at circa 34°, fluctuates about a value of 180° and the DNA binding sites lie at opposite ends of the complex. Measurements by Kahn and associates (10, 11) of distances between fluorescent dyes attached to the ends of DNA constructs bound to LacR give strong evidence for the existence of these forms in solution. Atomic force microscopy images of 197-bp LacR-mediated loops recently observed by Virnik *et al.* (12) on a 599-bp negatively supercoiled DNA minicircle are more appropriate, however, to state V as described above.

We treat the V → E transition as a motion of three rigid domains (I, II, and III) about two hinges. For simplicity we assume that the angle between l_I and l_{II} equals the angle between l_{II} and l_{III} (see Fig. 1*B*) and is one-half the total angle of aperture α. In state E, α varies over a range that includes $\alpha = 180°$, and additional flexibility is provided by allowing rotation of domains I and III about axes l_I and l_{III} by equal amounts β, with $-90° \le \beta \le 90°$. For the fully extended conformation of state E, $\alpha = 180°$ and the axes l_I, l_{II}, and l_{III} are coincident. For state V, angles α and β are fixed with $\alpha = 34°$ and $\beta = 33°$. Thus, we assume that the tetramer can attain either a rigid state V (7, 8) or a flexible state E with two degrees of freedom, α and β.

Bound DNA. For our calculations we assign to the DNA segments that are bound to the protein a three-dimensional model structure that is in accord with available structural data. Because the crystal structure of the tetramer with one dimer bound to the O3 sequence and the other to O1 has not been determined, our model is obtained by superposing the 4-Å-resolution structure (35) of the LacR dimer complexed with the O1 operator (PDB entry 1JWL) and the 2.7-Å structure (7) of the LacR tetramer without DNA-binding headpieces (PDB entry 1LBI). The bound O3 operator is assumed to be congruent to the bound O1 operator.

Determinants of Loop Configuration. The configuration of the LacR–DNA loop complex depends on several factors: (*i*) the flexibility of the protein tetramer, (*ii*) the end conditions imposed on the loop, and (*iii*) the DNA linking number, a topological parameter that cannot be changed without release of at least one of the operators from the protein.

The DNA operators can be oriented in one of two ways with respect to each protein dimer, with the 5′–3′ direction of the coding strand pointing inside or outside the V-shaped reference state (Fig. 1*C*). Because the core regions of the dimer are congruent and each of them is axially symmetric, there appears to be no *a priori* preference for any orientation of the bound operators. The combination of possible DNA orientations for each dimer gives rise to four possible DNA types of looping for the repressor assembly: A1, A2, P1, and P2 (13).

The linking number, *Lk*, of the LacR–DNA loop is, by definition, the Gauss linking number (36, 37) of two closed curves formed by virtual closure of the two DNA strands through the tetramer assembly. We choose a virtual closure that originates at the phosphorus atom on one of the DNA strands attached to the central base pair of the O3 operator, passes through the Gln 335 C^α atom of the LacR chain that makes direct contact with the 5′ end of the strand (38), continues through a second Gln 335 C^α atom in the other half of the protein assembly, and terminates at the corresponding phosphorus atom on the O1 operator in such a way that the linked phosphorus atoms lie on the same DNA strand (see the legend of Fig. 1).

DNA Elasticity. The elastic energy of the protein-mediated loop is calculated by using a recently developed naturally discrete model for DNA (16, 17) that accounts for the dependence of DNA elastic properties on nucleotide sequence. In the model the basic structural units are base pairs and the configuration of a segment with N base pairs is specified by giving, for each of the $N - 1$ base-pair steps, six kinematical variables: tilt θ_1^n, roll θ_2^n, and twist θ_3^n, which measure the relative orientation of the nth and $(n + 1)$th base pairs, and shift ρ_1^n, slide ρ_2^n, and rise ρ_3^n, which describe the relative displacement of those base pairs. The elastic energy Ψ of the looped configuration is taken to be the sum of base-pair step energies Ψ^n. In the present research Ψ^n is a quadratic function of θ_i^n, ρ_j^n, i.e.,

$$\Psi = \sum_{n=1}^{N-1} \psi^n,$$

$$\psi^n = \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} [F_{ij}^{XY} \Delta\theta_i^n \Delta\theta_j^n + 2G_{ij}^{XY} \Delta\theta_i^n \Delta\rho_j^n + H_{ij}^{XY} \Delta\rho_i^n \Delta\rho_j^n]. \quad [1]$$

Here, XY is the nucleotide sequence (in the 5′–3′ direction of the coding strand) of the nth base-pair step; $\Delta\theta_i^n = \theta_i^n - \bar{\theta}_i^{XY}$ and $\Delta\rho_i^n = \rho_i^n - \bar{\rho}_i^{XY}$ are deviations from the intrinsic values $\bar{\theta}_i^{XY}$ and $\bar{\rho}_i^{XY}$, i.e., values appropriate to the stress-free state; and F_{ij}^{XY}, G_{ij}^{XY}, H_{ij}^{XY} are elastic moduli. Empirical estimates of the intrinsic parameters and moduli are based on the average values and fluctuations of structural parameters in crystals of B–DNA and DNA–protein complexes (39).

Electrostatic Interactions. Inspection of the LacR crystal structure (7, 35) reveals approximately equal numbers of positively and negatively charged amino acid residues distributed uniformly throughout the tetramer complex. Thus, for simplicity, the protein is assumed to be electrostatically neutral. The possibility of other interactions of DNA with LacR, e.g., wrapping (15), is not considered. In addition, because none of the DNA phosphates are neutralized by nearby cationic amino acids in the best-resolved

LacR–DNA structures (38, 40), each of the phosphate groups, including those on the operator sites, is assigned a negative charge.

The electrostatic energy of DNA is taken to be the sum of all pairwise screened interactions between phosphate groups, i.e.,

$$\Phi = \sum_{m=1}^{N-3} \sum_{n=m+2}^{N-1} \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\delta_i^m \, \delta_j^n \exp(-\kappa r_{ij}^{mn})}{4\pi\varepsilon r_{ij}^{mn}}, \qquad [2]$$

where r_{ij}^{mn} is the distance between the ith phosphate group of the mth base-pair step and the jth phosphate group of the nth base-pair step, ε is the permittivity of water at 300 K, and κ is the Debye screening parameter that, for monovalent salt such as NaCl, obeys the relation $\kappa = 0.329\sqrt{c}$ Å^{-1}, in which c is the molar salt concentration. Because we assume 76% charge neutralization by condensed cations (41), for the net charge δ_i^m associated with the ith phosphate of the mth base-pair step we have $0.24e^-$ or 3.85×10^{-20} C.

Calculation of Configurations. Configurations of DNA loops with specified linking number and anchoring conditions were calculated in two steps. First, a configuration that minimizes the elastic energy of the loop was obtained by a recursive solution of the variational equations expressing the laws of balance of forces and moments acting on the nth base pair (16); then, a configuration that minimizes the total energy $\Psi + \Phi$ was found by using a standard conjugate-gradient iteration procedure with the elastic equilibrium configuration taken as the initial guess.

Free Energy. The free energy G of the looped complex is taken to be the sum of the following quantities: (*i*) the free energies G_{O1} and G_{O3} of binding of the O1 and O3 operators to LacR, (*ii*) the free energy G_{DNA} of the deformed DNA loop, and (*iii*) the change G_{LacR} in free energy associated with the transition of the tetramer from its rigid state V to its flexible state E with the attendant creation of a surface area A. Using reported values of A (7, 8) and a formula of Chothia (42), we find that $1.8\,kT < G_{LacR} < 9.4\,kT$. Because G_{O1} and G_{O3} are constants independent of loop type, they need not be considered when two loops are compared.

For a given loop type, linking number, and choice of α and β, $G_{DNA} = -kT \ln Z$, where Z is the partition function

$$Z = \int \ldots \int \exp(-(\Psi + \Phi)/kT) d\theta_1 d\rho_1 \ldots d\theta_{N-1} d\rho_{N-1}, \qquad [3]$$

in which the integrations are over DNA configurations compatible with imposed constraints. Approximate values of Z are obtained by replacement of the expression for $\Psi + \Phi$ by the terms of order two in the expansion in (θ_i, ρ_i) about the minimum energy configuration and explicit evaluation of the resulting integral over the linear subspace of fluctuations compatible with the requirements of fixed linking number and end conditions, including loop type. A method of this type was used by Zhang and Crothers (43) in recent calculations of DNA ring-closure probabilities (but without electrostatic interactions taken into account).

Calculation of DNase I Cutting Patterns. In the modeling of the experiments of Krämer *et al.* (20), we assume that the binding of DNase I imposes the same restrictions on DNA structure as those observed in the crystal complex with the $d(GGTATACC)_2$ octamer (19), regardless of the sequence of the binding site. We further assume that the change in cleavage efficiency at a given base-pair step n decreases with the difference $\Delta G_{DNA}(n)$ between the free energy of a loop with DNase I centered at step n and the free energy G of a loop without DNase I. The presence of a negative minimum of $\Delta G_{DNA}(m)$ at $m = n$ indicates that the binding of DNase I and hence the cutting of the DNA is enhanced at step n, while a minimum at which $\Delta G_{DNA} > 0$ does not lead to enhancement. A maximum of $\Delta G_{DNA}(m)$ at n indicates that the cutting activity is suppressed there.

We thank Drs. Victor Zhurkin and Michael Tolstorukov for helpful discussions, Dr. Yun Li for sharing his analyses of protein–DNA contacts, and the anonymous referees for helpful suggestions. This work was supported by U.S. Public Heath Service Grants GM34809 and GM64375 and National Science Foundation Grants DMS-02-02668 and DMS-05-16646.

1. Adhya, S. (1989) *Annu. Rev. Genet.* **23,** 227–250.
2. Schleif, R. (1992) *Annu. Rev. Biochem.* **61,** 199–223.
3. Halford, S. E., Gowers, D. M. & Sessions, R. B. (2000) *Nat. Struct. Biol.* **7,** 705–707.
4. Müller-Hill, B. (1996) *The lac Operon* (de Gruyter, Berlin).
5. Bellomy, G. R., Mossing, M. C. & Record, M. T., Jr. (1988) *Biochemistry* **27,** 3900–3906.
6. Müller, J., Oehler, S. & Müller-Hill, B. (1996) *J. Mol. Biol.* **257,** 21–29.
7. Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996) *Science* **271,** 1247–1254.
8. Friedman, A. M., Fischmann, T. O. & Steitz, T. A. (1995) *Science* **268,** 1721–1727.
9. Ruben, G. C. & Roos, T. B. (1997) *Microsc. Res. Tech.* **36,** 400–416.
10. Mehta, R. A. & Kahn, J. D. (1999) *J. Mol. Biol.* **294,** 67–77.
11. Edelman, L. M., Cheong, R. & Kahn, J. D. (2003) *Biophys. J.* **84,** 1131–1145.
12. Virnik, K., Lyubchenko, Y. L., Karymov, M. A., Dahlgren, P., Tolstorukov, M. Y., Semsey, S., Zhurkin, V. B. & Adhya, S. (2003) *J. Mol. Biol.* **334,** 53–63.
13. Geanacopoulos, M., Vasmatzis, G., Zhurkin, V. B. & Adhya, S. (2001) *Nat. Struct. Biol.* **8,** 432–436.
14. Balaeff, A., Mahadevan, L. & Schulten, K. (1999) *J. Chem. Phys.* **83,** 4900–4903.
15. Tsodikov, O. V., Saecker, R. M., Melcher, S. E., Levandoski, M. M., Frank, D. E., Capp, M. W. & Record, M. T., Jr. (1999) *J. Mol. Biol.* **294,** 639–655.
16. Coleman, B. D., Olson, W. K. & Swigon, D. (2003) *J. Chem. Phys.* **118,** 7127–7140.
17. Olson, W. K., Swigon, D. & Coleman, B. D. (2004) *Philos. Trans. R. Soc. London* **362,** 1403–1422.
18. Hochschild, A. & Ptashne, M. (1986) *Cell* **44,** 681–687.
19. Weston, S. A., Lahm, A. & Suck, D. (1992) *J. Mol. Biol.* **226,** 1237–1256.
20. Krämer, H., Niemöller, M., Amouyal, M., Revêt, B., von Wilcken-Bergmann, B. & Müller-Hill, B. (1987) *EMBO J.* **6,** 1481–1491.
21. Iida, S. & Hayatsu, H. (1970) *Biochim. Biophys. Acta* **213,** 1–13.
22. Borowiec, J. A., Zhang, L., Sasse-Dwight, S. & Gralla, J. D. (1987) *J. Mol. Biol.* **196,** 101–111.
23. Nejedly, K., Sykorova, E., Diekmann, S. & Palecek, E. (1998) *Biophys. Chem.* **73,** 205–216.
24. Bochkarev, A., Bochkareva, E., Edwards, A. M. & Frappier, L. (1998) *J. Mol. Biol.* **284,** 1273–1278.
25. Shore, D. & Baldwin, R. L. (1983) *J. Mol. Biol.* **170,** 983–1007.
26. Crothers, D. M., Drak, J., Kahn, J. D. & Levene, S. D. (1992) *Methods Enzymol.* **212,** 3–29.
27. Brenowitz, M., Pickar, A. & Jamison, E. (1991) *Biochemistry* **30,** 5986–5998.
28. Vilar, J. M. & Leibler, S. (1993) *J. Mol. Biol.* **331,** 981–989.
29. Hudson, J. M. & Fried, M. G. (1990) *J. Mol. Biol.* **214,** 381–396.
30. Krämer, H., Amouyal, M., Nordheim, A. & Müller-Hill, B. (1988) *EMBO J.* **7,** 547–556.
31. Pettijohn, D. E. & Pfenninger, O. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1331–1335.
32. Becker, N. A., Kahn, J. D. & Maher, L. J., III (2005) *J. Mol. Biol.* **349,** 716–730.
33. Villa, E., Balaeff, A. & Schulten, K. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 6783–6788.
34. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Müller-Hill, B. (1994) *EMBO J.* **13,** 3348–3355.
35. Bell, C. E. & Lewis, M. (2001) *J. Mol. Biol.* **312,** 921–926.
36. Courant, R. (1936) *Differential and Integral Calculus* (Blackie, London), Vol. 2.
37. White, J. H. (1989) in *Mathematical Methods for DNA Sequences*, ed. Waterman, M. S. (CRC, Boca Raton, FL), pp. 225–253.
38. Kalodimos, C. G., Bonvin, A. M. J. J., Salinas, R. K., Wechselberger, R., Boelens, R. & Kaptein, R. (2002) *EMBO J.* **21,** 2866–2876.
39. Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M. & Zhurkin, V. B. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11163–11168.
40. Bell, C. E. & Lewis, M. (2000) *Nat. Struct. Biol.* **7,** 209–214.
41. Manning, G. S. (1978) *Q. Rev. Biophys.* **11,** 179–246.
42. Chothia, C. (1976) *J. Mol. Biol.* **105,** 1–12.
43. Zhang, Y. L. & Crothers, D. M. (2003) *Biophys. J.* **84,** 136–153.