
Relating Structure to Function Through the Dominant Slow Modes of Motion of DNA Topoisomerase II

R. L. JERNIGAN,¹ M. C. DEMIREL,^{1,2,3} I. BAHAR^{1,2,3}

¹*Molecular Structure Section, Laboratory of Experimental and Computational Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-5677*

²*Chemical Engineering Department & Polymer Research Center, Bogazici University, Istanbul, Turkey*

³*TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey*

Received 16 November 1998; revised 4 January 1999; accepted 5 January 1999

ABSTRACT: Type II DNA topoisomerases are enzymes capable of transporting one DNA duplex through another by performing a cycle of DNA cleavage, transport, and religation, coupled to ATP binding and hydrolysis. Here, we considered a coarse-grained model of the structure and investigated the motions within two structures, DNA topoisomerase II and DNA gyrase A. The coarse graining with only one point per residue means that motions in such large proteins can be thoroughly investigated. The overall motions are reflected in the crystallographic temperature factors, which are reproduced by the model. Also, with this approach, we can view the slowest, most cooperative, modes of motion, corresponding to the largest-scale correlated motions in the protein. These motions are nearly identical in the two proteins and are likely related to individual steps in the enzyme's complex mechanism of activity. © 1999 John Wiley & Sons, Inc.* Int J Quant Chem 75: 301–312, 1999

Key words: topoisomerase; Gaussian network model; DNA binding proteins; slow modes of motion

We dedicate this article to the memory of Bernard Pullman—one of the pioneers in computational biology—and to keeping alive his pioneering spirit.

Correspondence to: R. L. Jernigan.

Contract grant sponsor: NATO.

Contract grant number: 951240.

Contract grant sponsor: NCI.

Introduction

After the very first DNA and protein structures had been determined, there was an extreme contrast between the amount of direct information the two provided regarding function. The Watson–Crick double–helix structure immediately told us how the DNA functioned and how it carries the organism’s information, whereas the globular protein structures were highly disappointing in not providing any direct insights into the function of the individual protein.

Subsequently, molecular dynamics (MD) simulations have made some limited contributions in informing us about function. However, as demonstrated by Phillips and coworkers [1], state-of-the-art MD simulations are still quite limited because they do not provide an adequate sampling of conformations to permit consideration of all correlations in the largest-scale motions of a protein. Several years ago, the groups of Levitt, Karplus, Go, and others [2–5] performed normal-mode analyses on protein molecular dynamics runs. The present study is not closely related to those studies, but, instead, pursues a rather different model and an analysis with different intentions. One of our intentions here is to reach a level of understanding or to accumulate sufficient intuition about these slow modes so that we can suggest, immediately upon inspection of a structure, what are likely to be its most important modes of motion related to functional behavior.

As an alternative to MD, we have been investigating the equilibrium fluctuation dynamics of proteins with a recently proposed [6] simple model that agrees remarkably well with the X-ray temperature factors [6, 7] and hydrogen-exchange data [8]. We have made applications to several systems, including HIV-1 reverse transcriptase [9], the tRNA–cognate synthetase complex [10], and tryptophan synthase [11]. The approach is particularly informative about the largest-scale motions as well as the other extreme of the most rigid elements found to correspond to the enzyme active sites and the structural cores, as suggested by recent application to small globular proteins such as chymotrypsin inhibitor 2, cytochrome *c*, and Che Y [12]. Recently, Hinsen [13] presented a careful comparison of the residue-level Gaussian network model normal modes with atomic (Amber-94 po-

tential) normal modes for several small proteins. Similar behavior for the low-frequency modes was reported.

This simplest protein dynamics considers the fluctuations of a highly coupled, highly cooperative protein system. The structures are coarse-grained—for proteins, this usually means one point per residue, as taken here. Furthermore, all residues are taken to be identical. Close residues are simply linked to one another by identical springs subject to harmonic oscillations; hence, the name *Gaussian network model* (GNM). Because those close residues are linked, the model reflects, appropriately, the geometry of the protein structure. The distinction between nonbonded and bonded pairs occurs only through the detailed specific differences in residue packing density and geometry. The present type of mechanical model with identical residues is a rather crude approximation to the actual molecule. However, the largest-scale motions are expected to be less sensitive to the details of specific interactions, such as hydrophobic or electric type. As one passes to the smaller-scale motions, however, the effects of variable interaction types would certainly increase.

Individual modes of overall motion can be separately considered. The large- or small-scale modes, even though there is no explicit time considered here, must correspond physically to the slow or fast modes, respectively. The former modes have been proposed to give information on the mechanisms in motions relevant to function, and the latter, about regions critical for stability [14]. Generally, the GNM has proven to be a useful and appropriate model of the cohesive nature of a wide range of protein structures.

One example in which we have successfully related these large-scale motions to function is our recent investigations on the motions of HIV-1 reverse transcriptase [9], where we interpreted its largest-scale motion as corresponding directly to the stepwise, base-by-base processing of the RNA strand. These calculations were particularly exciting because the slow motion reflects the two halves of the molecule pulling the RNA in opposite directions in the essential processing step. In this motion, the RNA strand is released from the polymerase site and simultaneously pulled toward the RNase H site. This corresponds exactly to the stepwise, one base at a time, motion required for the processing of the RNA strand by this enzyme. Our success in investigating this NA–protein complex has encouraged us to look here at another case of

NA-protein complexes. In the present study, we present results from GNM analysis for the type II DNA topoisomerases, topoisomerase II (topo II), and gyrase A (GyrA). DNA topoisomerases are enzymes of vital importance to all cells, being capable of cleaving DNA single strands (type I) or double strands (type II) and allowing the passage of a second DNA through the gap opened between the broken pieces of the cleaved DNA duplex [15, 16]. The three steps of cleavage, gap opening, and transport of a second DNA segment through the gap are succeeded by the religation of the broken DNA. To achieve this process, the enzyme is proposed to serve as a "bridge" that spans the DNA break with a transient covalent 5'-phosphotyrosine link(s) and noncovalent binding of the 3'-hydroxyl ends.

Crystallographic and biochemical studies suggest that class II DNA topoisomerases act as ATP-modulated clamps with multiple joints and two sets of jaws at opposite ends [17–20]. The crystal structure of a large fragment (92 kDa) of yeast DNA topo II reveals a heart-shaped dimeric protein, with a large central hole, approximately 50 Å in diameter, proposed to hold the transported DNA segment (T-segment), before its release [21]. Each monomer (residues 410–1202) consists of two domains, A' and B', with the A' domain consisting, in turn, of three domains. The principal two of these are a catabolite-activator protein (CAP)-like part at the N-proximal part and a largely helical one at the C-proximal end (see these domains colored separately in Fig. 1). The first set of jaws is located at the interface between the two B' domains, at the upper part of the heart lobes, and the second at the base of the hole between the C-proximal domains of the two A' domains (Fig. 1). The latter is termed the *primary dimer contact* because of its extensive buried surface. The former set of jaws (N-gate) is proposed to be involved in admitting and cleaving the DNA duplex—which, subsequently, serves as a gate (G-segment) for the passage of the T-segment-, and the second (C-gate) in expelling the T-segment (see Fig. 2). The active-site residues, Tyr783 on each monomer, are located in the CAP-like domain, at the junction of the B' and A' domains. The full activity depends on ATP binding and hydrolysis. ATP binding induces a conformational change in the ATPase domains (residues 1–409) near the first set of jaws, from an open state to a closed state so as to capture the T-segment. Hydrolysis and ATP release, on the other hand, are accompanied by the opposite con-

formational change [19, 21]. In Figure 1(c, d), the structure of a 59 kDa fragment of GyrA is displayed [22], which is sequentially homologous to the C-terminal of two-thirds of topo II.

We wanted to apply GNM analysis to examine the available crystal structure of the large fragment of yeast DNA topo II [21] for assessing the mechanisms of cooperative motions directly implied by the overall molecular architecture of the enzyme, through its topology of contacts. The above mechanistic implications, summarized in Figure 2, are indeed deduced from crystallographic or biochemical studies. Also, a direct analysis of the dynamic characteristics of the enzyme inherently implied by its overall 3-dimensional structure is lacking. In particular, the fundamental (lowest-frequency/largest-amplitude) modes of motion at one end of the available spectrum of modes was explored here with the intention of extracting additional information regarding topoisomerase's mechanism of activity.

In the following section, a brief summary of the GNM is presented for explanatory purposes. This is followed by results for topo II in the third section and comparison with those for GyrA and a discussion and conclusions in the fourth section.

GNM Basic Assumptions and Methodology

MODEL AND THEORETICAL FOUNDATIONS

In the GNM, the protein structure is conceived as a network of interacting residues [6, 7], with each interacting pair constrained with a Gaussian function. By analogy to the classical model of a random network of polymer chains [23], on which the GNM of proteins is based, residues form the junctions of the network and the covalent or noncovalent forces that maintain the stability of the overall folded structure are viewed as the (hypothetical) chains linking these junctions. Again, in conformity with the Gaussian-distributed fluctuations of the end-to-end distances in random network chains [23–25], interresidue distances are assumed to obey Gaussian fluctuations, all with the same covariance. This assumption implicitly means that all interresidue interactions are accounted for by the same, single-parameter (force constant), harmonic potential. Inasmuch as no distinction among different types of side chains is made, a

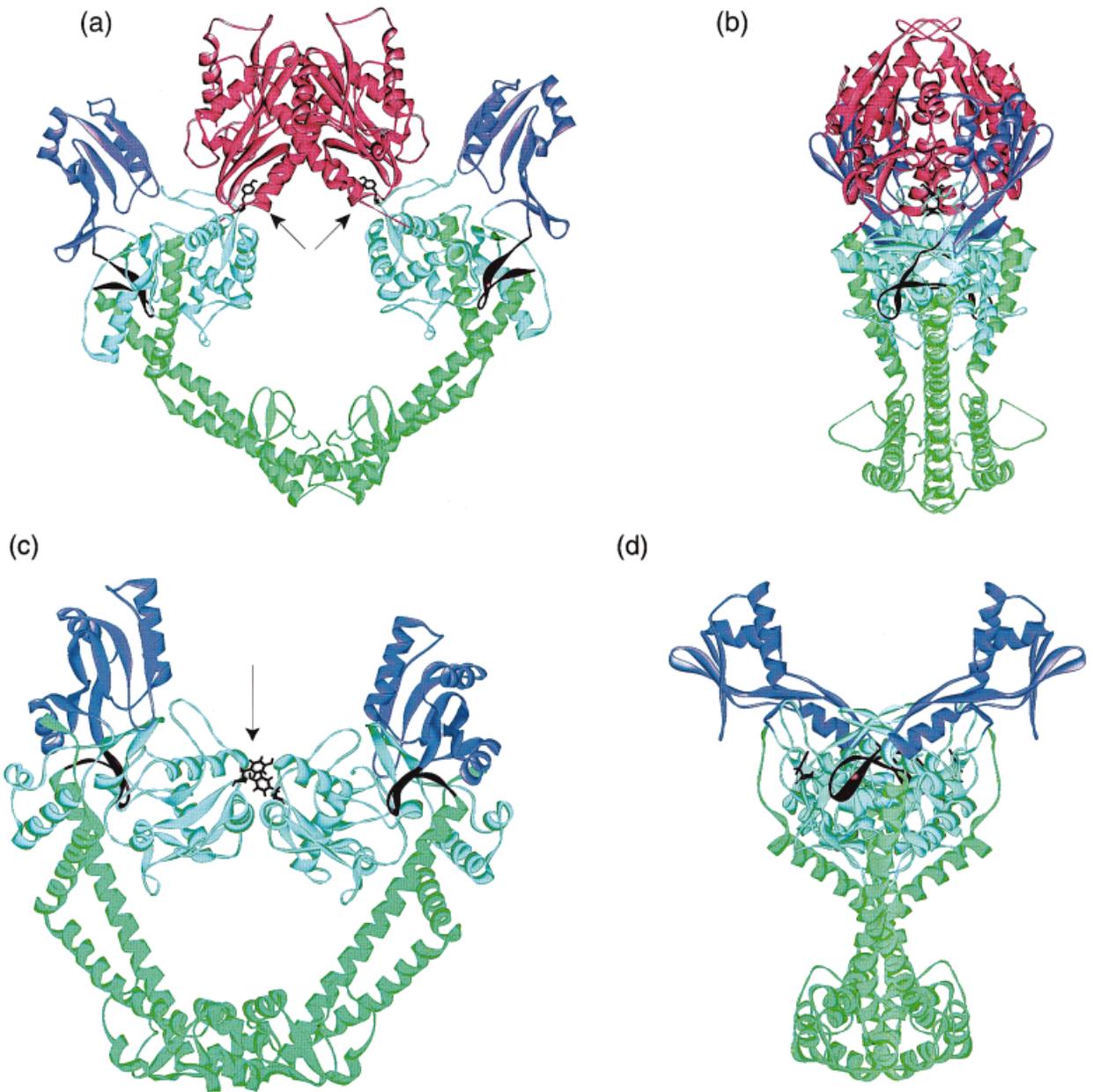


FIGURE 1. DNA topo II and GyrA crystallographically determined fragments shown in two different views each. Panels (a) and (b): topo II. The B' domain is given in red (group 1, residues 420–633). The A' domain is divided into four parts: group 2, residues 683–873 in cyan; group 3, residues 874–972 in purple; group 4, residues 973–990 in brown, and group 5, residues 991–1178 in green. Groups 2 and 4 form the CAP-like domain. In addition, the active-site residue Tyr783 is shown in black and indicated by arrows. In the left view, the plane of the figure is along the two largest dimensions of the protein and the two subunits are nearly separable as the left and right halves of the molecule. The side view obtained by a 90° rotation about a central vertical axis is displayed on the right. Panels (c) and (d): two perpendicular views of GyrA. The four regions are defined as group 1, residues 30–219 in cyan; group 2, residues 220–334 in purple; group 3, residues 335–346 in brown; group 4, residues 347–522 in green; and the active site Tyr122 in black. GyrA, overall, is smaller and differs most substantially in missing the B' domain of topo II and in having part of the analog of the A' domain coming into close contact with its counterpart in the second subunit. The figures utilize the respective Protein Data Bank (PDB) files 1bgw for topo II [21] and 1ab4 for GyrA [22].

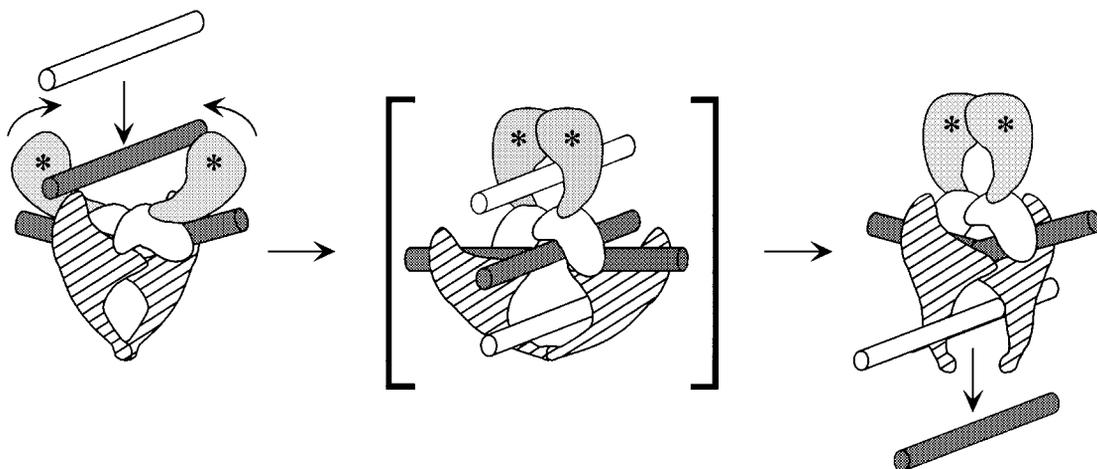


FIGURE 2. Proposed molecular mechanism for the reaction of DNA topo II. (Copied with permission from Fig. 5 of [20]). The ATPase domain is in light gray; the B' domain, in white; and the A' domain, cross-hatched. ATP binding is indicated by asterisks. The G-segment DNA is broken and reformed while the T-segment DNA is transported entirely through the protein, from top to bottom.

mechanistic description, purely entropic in origin, is followed [14].

In the GNM, each residue is represented by one interaction site, conveniently chosen at its α -carbon; the results are little affected by this choice. Two residues are assumed to interact if their α -carbons are separated by a distance $r \leq 10 \text{ \AA}$. This distance includes all the neighbors within a first coordination shell in the vicinity of a central residue, as indicated by statistical analyses of interresidue contacts in Protein Data Bank structures [26–29]. The pairs of residues located within this interaction range will be referred to as “contacting” residues. The first and second neighbors along the sequence always lie within 7 \AA , which ensures that chain connectivity is incorporated into the model. However, more important contributions to the observed dynamics are derived from nonbonded contacts, that is, those between residues sequentially distant, but spatially close. Our experience with various values of this cutoff distance between 6.5 and 10 \AA indicated only a weak dependence of the results on its exact value. A value of 10 \AA , at the high end of the usual range, was chosen simply because it corresponds to greater cooperativity and gives slightly smoother results.

The topology of contacts is accounted for by a Kirchhoff matrix of contacts (Γ). Γ is an $n \times n$ symmetric matrix for a structure of n sites (residues, nucleotides, and/or ligands). Its ij th off-diagonal element takes a value of -1 if sites i and

j are contacting and zero otherwise. In this, Γ has precisely the same information as do customary contact maps. The i th diagonal element of Γ is equal to the coordination number of the i th site, so that the rows (or columns) of Γ sum to zero. This matrix structure is characteristic of rate matrices in stationary stochastic processes. Thus, the Kirchhoff matrix contains two basic structural data: the local packing density in the neighborhood of each site and the sequence order of contacts. The former is expressed by the coordination numbers of residues, directly given by the diagonal elements. The latter refers to the separation, along the backbone sequence, between two contacting residues. The nonzero off-diagonal elements Γ_{ij} of Γ automatically describe the order of contacts as $|j - i|$.

The Kirchhoff matrix is identical in form to the mechanical stiffness matrices (of elastic materials) used in normal-mode analyses, provided that the force constants for all interacting elements are taken to have a fixed value, γ , for all interacting pairs. The internal Hamiltonian of the system may then be written as [14]

$$\mathcal{H} = \frac{1}{2} \gamma [\Delta \mathbf{R}^T (\Gamma \otimes \mathbf{E}) \Delta \mathbf{R}]. \quad (1)$$

Here, $\Delta \mathbf{R}$ is the $3n$ -dimensional vector of the x -, y -, and z -components of the fluctuations $\Delta \mathbf{R}_1, \Delta \mathbf{R}_2, \dots, \Delta \mathbf{R}_n$ in the positions of the n sites; the superscript T denotes the transpose; \otimes is the di-

rect product; and \mathbf{E} is the identity matrix of order 3. In parallel with normal mode analyses, the inverse Γ^{-1} is related to the auto- or cross-correlations between the motions of individual residues, as

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = (3kT/\gamma)[\Gamma^{-1}]_{ij}. \quad (2)$$

Here, k is the Boltzmann constant, T is the absolute temperature, and $[\Gamma^{-1}]_{ij}$ designates the ij th element of Γ^{-1} . Equation (2) follows from the equality

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{1}{Z} \int \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \exp\{-\mathcal{H}/kT\} d\{\Delta \mathbf{R}\}, \quad (3)$$

where Z is the configurational integral part of the vibrational partition function given by $Z = \int \exp\{-\mathcal{H}/kT\} d\{\Delta \mathbf{R}\}$.

Previously, we observed [6] excellent agreement between the mean-square (m.s.) fluctuations $\langle (\Delta R_i)^2 \rangle$ of individual residues evaluated for several proteins from the diagonal elements $[\Gamma^{-1}]_{ii}$ of Γ^{-1} [using Eq. (2) for $i = j$] and those deduced from the X-ray crystallographic temperature factors (or B factors) $B_i = 8\pi^2 \langle (\Delta R_i)^2 \rangle / 3$. These results indicate that the method is able to capture the essential motions of the protein and lend support to the use of the GNM as a simple and physically meaningful approach for generally investigating protein motions. Also, because the dimension of the matrix to be inverted is only the same as the number of residues in the protein, substantially larger proteins can be considered than with conventional atomic molecular dynamics.

MODAL DECOMPOSITION

Information on global dynamics is acquired by decomposing the motions into a series of modes and concentrating on the modes at the slowest/largest amplitude end of the spectrum. The latter are the most cooperative modes in the folded state, in that they dominate the coupled collective motions of large-size structural blocks. To elucidate the mechanism of these motions, as a first step, Γ is written as the product of the matrix \mathbf{U} of its eigenvectors \mathbf{u}_i ($1 \leq i \leq n$), and the diagonal matrix Λ of its eigenvalues λ_i as

$$\Gamma = \mathbf{U}\Lambda\mathbf{U}^T = [\mathbf{u}_1\mathbf{u}_2 \dots \mathbf{u}_n] \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \times [\mathbf{u}_1\mathbf{u}_2 \dots \mathbf{u}_n]^T. \quad (4)$$

We note that $\lambda_1 = 0$, and $\mathbf{U}^T = \mathbf{U}^{-1}$. The inverse of Γ is therefore easily found from $\Gamma^{-1} = \mathbf{U}\Lambda^{-1}\mathbf{U}^T$, and the correlation $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle_m$ contributed by the m th mode is found from

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle_m = (3kT/\gamma) [\lambda_m^{-1} \mathbf{u}_m \mathbf{u}_m^T]_{ij}. \quad (5)$$

The correlations conveyed by a subset $m_1 \leq m \leq m_2$ of modes of interest are evaluated by weighting the contribution of each mode by $1/\lambda_m$.

The first nonzero mode, also referred to as the fundamental mode, gives information on the most cooperative, global mechanism of motion. The dependence of the corresponding eigenvector on the residue index describes the motions of the individual residues in this mode. Likewise, an examination of a few eigenvectors u_i associated with the dominant, slow modes of motion provides information on the identity of structural elements acting as hinges, levers, etc., in the cooperative structural changes near the folded state, which are most probably relevant to biological activity and function. The regions acting as hinges in the global motion are distinguished by their having severely hindered fluctuations in the fundamental mode, that is, their displacements are small, and $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle_2 \approx 0$. The slow motions usually correspond to motions of rigid elements about a hinge or several hinges in the molecule. Therein, the hinge center is fixed in space, while the adjoining beams, of rigid structural elements undergoing coherent, coupled movements, exhibit motions increasing in amplitude at increased distances from the hinge center. A hinge center possesses a local rotational mobility in determining the motion of the surrounding structural elements.

Cross-correlations $\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle$ between the fluctuations of residues i and j ($1 \leq i < j \leq n$) are calculated from Eq. (2) and normalized following the equality

$$C_{ij} \equiv \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle / [\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle \times \langle \Delta \mathbf{R}_j \cdot \Delta \mathbf{R}_j \rangle]^{1/2}. \quad (6)$$

The normalized C_{ij} values vary in the range $-1 \leq C_{ij} \leq 1$. The upper and lower limits correspond to pairs of residues exhibiting fully correlated (same direction, same sense) and fully anticorrelated (same direction, opposite sense) motions, respectively. The particular case of $C_{ij} = 0$ refers to uncorrelated, or orthogonal, motions. Cross-correlations involved in the individual modes can be

examined separately, which provide information on the types and strengths of couplings between different structural elements in the specific modes being examined.

Results

TEMPERATURE FACTORS

Figure 3 compares the theoretical m.s. fluctuation amplitudes, $\langle(\Delta R_i)^2\rangle$, evaluated from Eq. (2) (for $i = j$) with the experimental ones reported [21] with the X-ray structure. With the Gaussian network model, we reconstruct fluctuations of DNA topo II using the slowest-mode shapes obtained from the eigenvalue decomposition of the contact matrix. The agreement is excellent, and it is interesting that the GNM is highly effective in representing the motions for a broad range of structures, from the compact globular to the open-ring form of the topo II structure.

To show quantitatively the motions, we utilize a color distribution for the motions of each residue of topo II. Six different colors are used to represent different levels of displacement: white, cyan, green, yellow, magenta, and red, where the smallest displacement level corresponds to white and the

highest to red. The results for the slowest four modes of motion, calculated from Eq. (5) for $2 \leq m \leq 5$, are presented in Figure 4. The first mode indicates the symmetric motions of the two monomers about a central vertical axis, shown by a dashed line, and could lead to strains to open the gates at the top and bottom of the structure. The second mode reveals the mobility of domains B' around a horizontal axis of rotation; domain B' appears therein to be rather decoupled from domain A'. Simultaneously, a rotation about a second axis placed at thin points in the structure permits motion of the lowest section of the molecule. Yet, for precisely identifying the rotation axes and/or hinge sites, as well as the type of correlations between the movements of different structural elements in the individual modes, we need to examine the auto- and cross-correlations between residue fluctuations, as will next be elaborated.

LOCATING THE HINGE AXIS FOR EACH MODE

For each of the slow modes, it is straightforward to locate the sites of the hinges in the structures, by combining the results of the colored

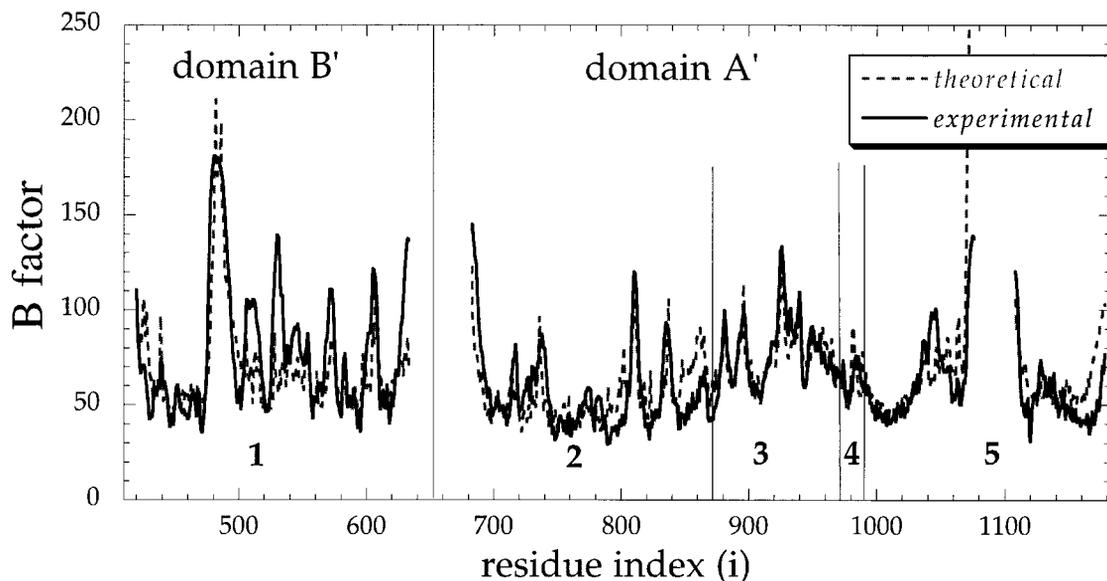


FIGURE 3. The B factors of the 92K fragment of topo II dimer. Experimental data from [21] (PDB structure 1bgw) shown as a solid line and our calculated fluctuations shown as a dotted line. The agreement is excellent overall. The domains are indicated by the vertical line completely dividing the figure and the labels at the top; the subdomains defined in the caption of Figure 1 are given by the shorter vertical lines and the numbers 1–5 inside the bottom of the figure.

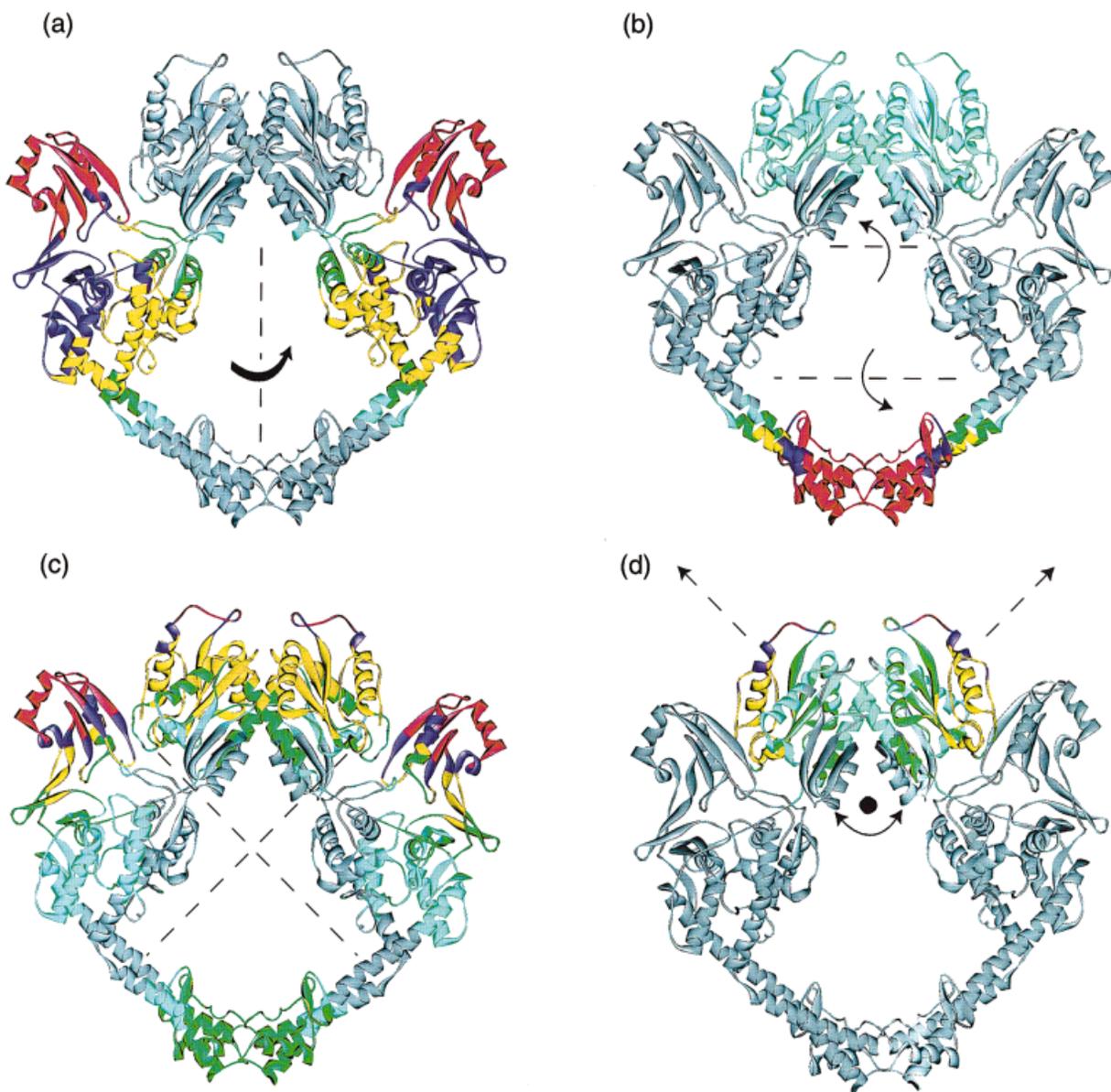


FIGURE 4. Rotation axes for the four slow modes (a)–(d) of topo II displayed on the structure—shown as blue lines. The structure itself is colored in a qualitative way according to the extent of rms deviations for the respective modes, colored gray, cyan, green, yellow, purple, and red, from the smallest to largest displacement. The location of the rotation axes are inferred at the loci of lowest displacement between the coherently moving rigid blocks identified in the correlation maps of Figure 5.

displacement figures (Fig. 4) with the cross-correlations plots given in Figure 5, together with the calculated fluctuations of each residue. In Figure 5, the regions that move en bloc are demarcated by lines, and the regions exhibiting positive or negative correlations [see Eq. (6)] are indicated. See the

caption of Figure 1 for the residue ranges of the groups 1–5 indicated on the axes. By comparing the corresponding parts in Figures 5 and 6 for an individual mode, it is straightforward to locate the site of the hinge, if also the most rigid (hingelike) residues are selected from the corresponding

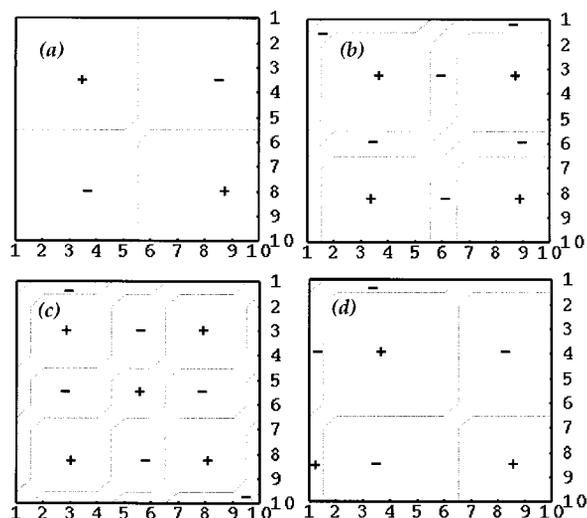


FIGURE 5. Cross-correlations for the four slow modes (a)–(d) of topo II dimer, with indices for the residue groupings in the first monomer given in the caption of Figure 1 and numbering in the same order, from 6 to 10, in the second monomer. Regions marked + have cross-correlation values of +1 and those marked – have values of –1.

curves of fluctuations. These rigid residues must lie close to the hinge axis. Hinge axes are indicated in Figure 4 by the pink lines.

It is interesting to see that the four slowest modes correspond to distortions out of a plane. Mode 1 [Fig. 4(a)] corresponds to a simple hinge rotation about an axis running between the top and bottom of the structure. Mode 2 [Fig. 4(b)] has two rotation axes approximately perpendicular to the first one, but displaced upward and downward from the center of the protein because of the protein's asymmetry along this vertical direction. Mode 3 [Fig. 4(c)] is the most complex of the four modes being shown. It is particularly interesting because it involves two composite perpendicular hinge axes in which the plane shown in Figure 4 is divided into four parts, with the external corners exhibiting the largest displacements to form a "saddle." These move in such a way that the top and bottom are correlated and the left and right are correlated, but these two pairs are anticorrelated with respect to one another. This distortion of the plane can be described by the two intersecting rotation axes shown. Mode 4 [Fig. 4(d)] localizes most of the motion in the upper B' domains and corresponds to an upper-gate opening. The motion between the two B' segments corresponds to a hinge perpendicular to the plane shown in

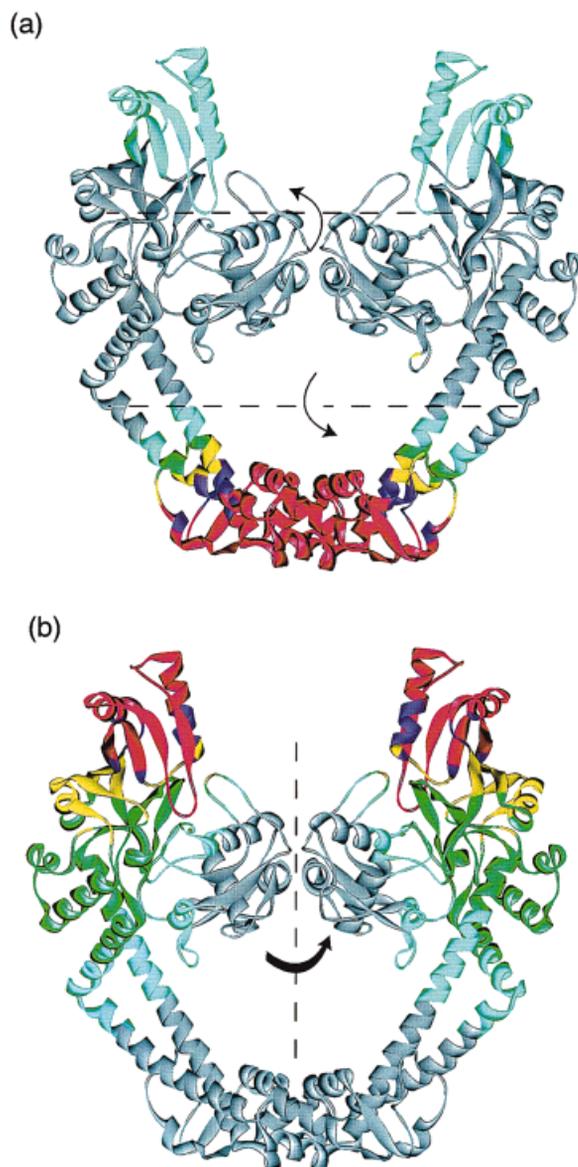


FIGURE 6. Rotation axes for the two slowest modes of GyrA. These are derived in the same way as those in Figure 4 and closely resemble those of the two slowest modes of topo II shown in Figure 4, but are in opposite order: The slowest mode observed here corresponds to mode 2 displayed in Figure 4(b) while the second slowest mode corresponds to mode 1 shown in Figure 4(a). The coloring is the same as in Figure 4.

Figure 4 in which the two colored parts in Figure 4 move in an anticorrelated fashion with respect to each other. It is quite interesting that all of these motions can be described with respect to the plane shown in Figure 4. It is likely so because the dimensions of the dimer structure are approximately $120 \times 120 \times 55$ Å; consequently, this slab-

like nature is reflected directly in the dynamic behavior of the structure.

A more direct way to obtain the detailed directions of motion in each mode would be to expand the matrix to include Cartesian coordinates for each point residue, thereby expanding the matrix to $3n \times 3n$ in dimension. The present way utilized here permits shorter calculations since smaller matrices of dimension $n \times n$ are sufficient.

The slowest mode could be involved in the assembly process with the missing ATPase domains and also with the DNA, since it could enhance the formation of their encounter complexes. It can also be imagined that this motion is related to the opening of both the upper and lower gates, and that following the binding of DNA in the upper-gate region, this motion would subsequently enhance the opening of the lower gate. The second mode of motion could be involved in refining interactions in both the upper and lower DNA interaction sites. Both modes 2 and 3 induce a large-scale distortion of the opening between domains B' and A', which may be essentially important for accommodating the G-segment. The fourth mode is more directly related to the opening and closing of the upper gate as needed for taking in the DNA at the N-gate of the topo II structure near the B' domain. Mode 3 involves overall stress to the structure, but is not so obviously directly related to the individual steps shown in the hypothetical mechanism of Figure 2.

The comparisons between the GyrA and the topo II structures given by Cabral et al. [22] considered the two structures to be closely related to one another. The domain B' is not present in the GyrA structure, and so they produced a model to form the remainder of the dimer interface in the topo II structure, based on the GyrA structure. This gave, in turn, a model of the topo II, with the tails open at the primary dimer interface at the bottom of our structures because of the large distortions in helices $\alpha 14$, $\alpha 18$, and $\alpha 19$.

GYRASE A

When we repeated the GNM calculations on DNA GyrA, we found slow modes that are quite similar to those of DNA topo II. However, the specific similar modes are not quite in the same order. We can understand the order of modes from the eigenvalues: When the adjacent eigenvalues in

the ranked list are close together, then they can easily be interchanged in order between two similar structures. For example, the first two modes appear in the reversed order between the two molecules topo II and gyrA.

Another observed difference is the slightly higher mobility of the primary contact region in GyrA, compared to that of topo II at the same region. The enhancement to the amplitude of motions at this region may be attributed to an indirect effect caused by the closer positioning of the CAP-like domains of the two monomers in GyrA. As seen in panels (c) and (d) of Figure 1, these domains are now in close contact near the active-site residue (Tyr122), and the tighter packing at the upper part of the molecule, near the N-gate, may place a relatively higher strain on this region. As a result, the mechanism of motion inherently driven by the dominant mode 2 (in topo II) tends to activate now (in GyrA) the bottom part of the molecule, near the C-gate. We looked for modes that corresponded to anticorrelated motions between the two subunits in the primary dimer interface. For topo II and GyrA, we found that the respective modes 7 and 5 have such localized fluctuations which might directly lead to the opening of the C-gate. Therein, the long helices connected to the primary contact region were distinguished by large-amplitude, opposite-direction motions. However, the actual interface residues 1042, 1043, and 1121–29 were rigid in this mode. The packing at the interface is quite dense and close. In particular, residues N1043 and T1126 between the two monomers have C α atoms that are only 4.3 Å apart. It is conceivable that these modes would be directly responsible for the opening of the C-gate, provided that these close interactions were somewhat weaker, an effect that might appear if atoms were considered.

In the modes elucidated in the present calculations, we observed only rotational motions. There are two possible reasons for this: Either these proteins are so cohesive that translational motions involving expansion are not permitted or the Gaussian function used here strongly enforces energetic penalties for translation modes, leading to lost interactions. If the latter is the reason for the missing translational modes in our calculations, modifications to the present calculations might uncover these.

Discussion and Conclusions

The 92K fragment of DNA topo II presently analyzed is an incomplete structure in that the ATPase domains in the N-terminal parts and the last parts of the C-terminal domains are missing. The results that we obtained refer to the intrinsic dynamic characteristics of this fragment whose crystal structure has been utilized. It is conceivable that the dynamic characteristics of the overall topo II (164 kDa) and that of the DNA-bound forms might differ. They could differ possibly to two different extents: Either the mode(s) involved in the binding step to the DNA might be frozen out and removed from the accessible motions, the remainder being substantially unaffected, or all modes could be changed. However, our previous examinations of DNA (or tRNA)-bound and free forms of enzymes [9, 10] revealed that the dominant modes of motion were preserved by having the same qualitative shape, in both substrate-bound and free forms. Essentially, the hinge sites were maintained while the relative amplitudes of motion of the different regions were affected by binding or complex formation. Thus, the qualitative conclusions reached in the present study might be expected to be relevant to the motions of the molecular machinery underlying the reaction cycle of the enzyme.

One particularly appealing aspect of the present modes is the prospect of relating them to the individual mechanistic steps in the action of these proteins. However, we do not know with certainty whether the individual modes of the unbound protein can be directly related to the mechanistic steps or whether these modes of motion need to be recalculated following each step. The present study is an initial study of these relationships and issues.

The comparison between the two proteins provides important confirmation that the motions that we are observing are, indeed, a direct consequence of the overall structure and do not depend on computational artifacts. A strong similarity is observed between the motions in the two proteins, despite the numerous small differences between them. Other than an interchange of neighboring modes in the rank order, and the relative mobilities of the two interfacial regions, the slowest modes for the two molecules are found to be nearly identical.

The present motions studied are coarse-grained, since only one point per residue is considered. There is the possibility that some of these large-scale motions might be inconsistent with some details of the atomic structure; however, this is highly unlikely, inasmuch as these motions involve domain motions and any atomic conflicts could presumably be relaxed without affecting the coarse graining through local rearrangements of atoms. It is a sufficiently simple approach that future extensions may be made to much more complex systems and these further studies ought to foster a more comprehensive understanding of protein function.

ACKNOWLEDGMENTS

This work was supported by NATO Grant #951240 and completed during visits supported by that grant and by the NCI.

References

1. Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettitt, B. M.; Phillips, J. G. N. *Proc Natl Acad Sci USA* 1995, 92, 3288.
2. Levitt, M.; Sander, C.; Stern, P. S. *J Mol Biol* 1985, 181, 423.
3. Brooks, B.; Karplus, M. *Proc Natl Acad Sci USA* 1985, 82, 4995.
4. Horiuchi, T.; Go, N. *Prot Struct Funct Genet* 1991, 10, 106.
5. Ben-Avraham, D. *Phys Rev B* 1993, 47, 14559.
6. Bahar, I.; Atilgan, A. R.; Erman, B. *Fold Des* 1997, 2, 173.
7. Haliloglu, T.; Bahar, I.; Erman, B. *Phys Rev Lett* 1997, 79, 3090.
8. Bahar, I.; Wallquist, A.; Covell, D. G.; Jernigan, R. L. *Biochemistry* 1998, 37, 1067.
9. Bahar, I.; Erman, B.; Jernigan, R. L.; Atilgan, A. R.; Covell, D. G. *J Mol Biol*, 1998, 285, 1023.
10. Bahar, I.; Jernigan, R. L. *J Mol Biol* 1998, 281, 871.
11. Bahar, I.; Jernigan, R. L., *Biochemistry* 1999, in press.
12. Demirel, M. C.; Atilgan, A. R.; Jernigan, R. L.; Erman, B.; Bahar, I. *Prot Sci* 1998, 7, 2522.
13. Hinsen, K. *Prot Struct Funct Genet* 1998, 33, 417.
14. Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. *Phys Rev Lett* 1998, 80, 2733.
15. Sharma, A.; Mondragon, A. *Curr Opin Struct Biol* 1995, 5, 39.
16. Berger, J. M.; Fass, D.; Wang, J. C.; Harrison, S. C. *Proc Natl Acad Sci USA* 1998, 95, 7876.
17. Roca, J.; Berger, J. M.; Harrison, S. C.; Wang, J. C. *Proc Natl Acad Sci USA* 1996, 93, 4057.
18. Berger, J. M.; Wang, J. C. *Curr Opin Struct Biol* 1996, 6, 84.

19. Maxwell, A. *Nature Struct Biol* 1996, 3, 109.
20. Berger, J. M. *Curr Opin Struct Biol* 1998, 8, 26.
21. Berger, J. M.; Gamblin, S. J.; Harrison, S. C.; Wang, J. C. *Nature* 1996, 379, 225.
22. Cabral, J. H. M.; Jackson, A. P.; Smith, C. V.; Shikotra, N.; Maxwell, A.; Liddington, R. C. *Nature* 1997, 388, 903.
23. Flory, P. J. *Proc R Soc Lond A* 1976, 351, 351.
24. Pearson, D. S. *Macromolecules* 1977, 10, 696.
25. Kloczkowski, A.; Mark, J. E.; Erman, B. *Macromolecules* 1989, 22, 1423.
26. Miyazawa, S.; Jernigan, R. L. *Macromolecules* 1985, 18, 534.
27. Jernigan, R. L.; Bahar, I. *Curr Opin Struct Biol* 1996, 6, 195.
28. Miyazawa, S.; Jernigan, R. L. *J Mol Biol* 1996, 256, 623.
29. Bahar, I.; Jernigan, R. L. *J Mol Biol* 1997, 266, 195.