Spatial density distributions for illustrating the base sequence dependent features of double helical DNA: Computer graphic visualization of Monte Carlo chain simulations

A. R. Srinivasan and Wilma K. Olson*

Department of Chemistry, Rutgers, the State University, New Brunswick, NJ 08903, USA

The sequence-directed flexibility of double helical DNA is examined with color-coded representations of the spatial probability density distributions of the chain ends. The distributions are derived from Monte Carlo simulations that incorporate local sequence-dependent bending of neighboring Watson-Crick base pairs. The density functions are compared with typical rigid representations of the double helix and with sample Monte Carlo trajectories. Applications are presented for three short fragments of kinetoplast DNA from Crithidia fasciculata, *which exhibit dramatically different behavior on nondenaturing polyacrylamide gels. The distributions (based on 10^6 configurations per chain) are useful descriptors of overall chain flexibility, illustrating the effects of chain length and base sequence on macromolecular configuration and revealing characteristic differences between curved and rodlike DNA.*

Keywords: DNA; density distributions; macromolecular flexibility; Monte Carlo simulations; sequence effects

Received 28 March 1988
Accepted 23 May 1988

INTRODUCTION

Computer graphics is proving to be an essential tool in understanding the conformations and properties of biopolymers. The bulk of molecular applications to date have been static representations of crystallographically observed or energy minimized three-dimensional structures. A growing number of procedures, however, are being developed to study chain dynamics.[1,2] Particularly popular is the animation of Newtonian molecular dynamics simulations,[3] providing a visual description of the small-scale fluctuations of specific low energy structures. The coordinates and velocities of individual atoms are changed in response to the forces exerted on them by other atoms in the chain, generating a complex vibrational pattern with occasional large-scale changes of local molecular conformation (e.g., the crossing of torsional barriers about the single bonds of the system). Detailed study of large-scale molecular motions by such methods, however, is not yet computationally practical. Large-scale flexibility is generally described in terms of simplified artificial models[4,5] or through computational short cuts[6] that force the system over preselected energy barriers. Little, if any, attention has been given to visualization of the large-scale changes of macromolecular structure (involving the crossing of multiple torsional barriers) expected on the basis of local conformational flexibility.

Large-scale macromolecular motions are best deduced from direct Monte Carlo simulations.[4,7-14] Such methods are particularly appropriate for studying polymers with properties determined by the chemical constitution of individual side groups. Specific chain conformations are generated from randomly chosen combinations of backbone structural parameters, usually the torsions about the single bonds of the chain. Sequence-dependent features of structure are included in terms of the statistical weights describing various combinations of residue torsions. The variation of structure through the stretching of bonds and bending of valence angles

*To whom correspondence should be addressed

is usually ignored. Individual torsional parameters can be varied in small fixed angular increments, generating smooth variations of structure that can be monitored graphically.[15] The residues to be moved and the direction of angular change are chosen by random number techniques. The approach, unfortunately, is feasible only for small molecules in view of the time it takes to sample the range of low energy states available to individual chain units. The generation of structure is slowed, although not to the same extent, by the same barriers that impede interconversion between different conformations in molecular dynamics simulations. Other approaches that sample broader regions of torsion angle space must be employed in the study of long chains.

To examine the broad spectrum of configurations accessible to a macromolecule, it is necessary to sample local conformation angle space more widely. Local conformations can be chosen on the basis of individual statistical weights, avoiding the crossing of high energy barriers. The smooth folding of the macromolecule is, unfortunately, lost in such an approach. The Monte Carlo sample is merely a collection of arbitrary, unrelated three-dimensional structures. Sequential configurations of the Monte Carlo sample are random snapshots of overall chain movement, offering no clues to the transitional pathways that link them. Information concerning chain flexibility can, nevertheless, be extracted from catalogs of conformational data accumulated during the sampling process. Structures can be organized on the basis of some criterion (conformational energy, end-to-end extension, terminal bond orientation, etc.) and distributions of the relevant parameters accumulated. The distributions can then be compared against ideal models or with those of related polymers.

One such probe of macromolecular flexibility is the spatial density distribution $W_0(\mathbf{r})$. This quantity is a three-dimensional function describing the probability of finding the terminus of a chain molecule at vectorial location $\mathbf{r}$ relative to a reference frame embedded at its origin ($\mathbf{0}$).[16] The location of the chain terminus in this matter is analogous to the use of distribution functions in describing the electron density of a molecular orbital. The characteristic shape of $W_0(\mathbf{r})$ is tied to local chain properties just as the shape of a molecular orbital is linked to the quantum states of its electrons. In principle, if the polymer is flexible enough, the chain end can be found with equal likelihood at any point within a sphere of radius equal to the maximum chain extension ($r_{max} = xv$, where x is the number of chain residues and v the length of the chemical or virtual bonds spanning each repeating unit). Because of the constraints of chemical bonding and the restrictions on local bending and twisting of chain residues, this ideal Gaussian limit is normally attained only at very long chain lengths. The distributions are skewed in shorter chains to shapes determined by the polymer architecture.[17-22] The shapes can be correlated with observed measures of chain extension and flexibility and the densities can be used to estimate the likelihood of polymer cyclization and looping as a function of chain length and sequence.[20-22]

The conformation and properties of double helical DNA are intimately tied to the linear sequence of its heterocyclic base side groups. A number of models[23-32] have been offered to account for the subtle irregularities of local conformation in crystalline oligomers[33] and the observed twisting and bending of the chain in solution.[34-41] Computer programs[40-42] have also been developed to translate the primary sequence of bases into three-dimensional models on the basis of these rules. The flexible nature of the double helix is generally ignored in these representations with individual repeating residues of the chain described by fixed local geometries. Adjacent base pairs are found in experimental and theoretical studies,[33,43-48] however, to adopt a broad range of accessible conformations rather than a single narrowly defined minimum energy state. The conformation of the DNA as a whole is more aptly described by a Monte Carlo computer sample of the accessible low energy domains than a static macromolecular picture and is readily monitored by the distribution of $W_0(\mathbf{r})$.

We describe below how we have taken advantage of color graphics techniques to study the conformation and mobility of selected DNA sequences. We employ a series of recent potential energy estimates[48] of the local flexibility of adjacent nucleic acid base pairs to generate rotational isomeric state representations and Monte Carlo samples of the double helix. We monitor the chain flexibility with $W_0(\mathbf{r})$, distinguishing regions of high and low probability density on the basis of color. We color code the DNA to examine effects of chain sequence on overall structure and flexibility.[48] We study three short fragments of kinetoplast DNA from *Crithidia fasciculata* that exhibit dramatically different behavior on nondenaturing polyacrylamide gels.[49] We find characteristic differences in the distributions of conformations between curved and rodlike sequences and in the overall flexibility of AT and GC rich regions. We additionally superimpose selected trajectories and various static representations of the double helix on the density distributions in an effort to understand the flexibility of the DNA as a whole.

METHODS

Potential energies

The potential energies of interaction of adjacent base pairs are computed, as detailed elsewhere,[48] as a function of relative orientation and translation. The energies are a sum of all pairwise van der Waals and electrostatic interactions between atoms in the two residues. The orientation is described in terms of the so-called twist, roll, and tilt angles[49,50] of the base pair plane. The twist τ is the rotation about the base pair normal, while the roll ρ and the tilt λ are rotations about in-plane axes. The roll axis is chosen to run approximately parallel to the vector connecting purine C8 and pyrimidine C6 of the base pair. It is defined by the cross product $\mathbf{y} \times \mathbf{z}$ where $\mathbf{y}$ is the pseudodyad or tilt axis and $\mathbf{z}$ the twist (normal) axis of the base pair. The translation between residues is assumed, for simplicity, to extend

3.4 Å along the base pair normal. The base pairs are also treated as planar rigid bodies, ignoring the possible effects of propeller twisting, buckling, or other small perturbations of the hydrogen bonded complex.

Residues are found to roll more easily about their long axes than to tilt about their short (dyad) axes, in accordance with the anisotropic temperature factors[52] characterizing the local mobility of the crystalline B-DNA dodecamer. The predicted range of rolling, however, is considerably greater for certain residues, AT sequences bending with almost equal likelihood into the major and minor grooves of the duplex but GC sequences exhibiting a pronounced tendency to flex into the minor groove.[48] The range of rolling motions within 5 kcal/mole of the AT energy minimum is approximately ±20° and within 1 kcal/mole ±10°. The energetically preferred range of GC rolling, in contrast, is confined to positive values of ρ, diminishing the local flexibility by a factor of 2–3. Tilting, however, is roughly comparable in all sequences, conformations up to ±7.5° being within 5 kcal/mole. The bending partition functions (Equation 1) of the AT and GC sequences at 298 K are 5.1–7.1 and 2.6–5.0, respectively.[48] The greater flexibility of AT dimers compared to GC base pairs is also consistent with the observed triplet anisotropy decay[53] of fluorescent probes intercalated in poly dA · poly dT and poly dG · poly dC.

Monte Carlo array selection

The probability $w(\rho_0, \lambda_0)$ of bending a particular base pair in the ρ_0, λ_0 angular state is computed from the ratio of its Boltzmann factor, $\exp[-V(\rho_0, \lambda_1)/RT]$, to the partition function $z_{\rho,\lambda}$ in the (ρ, λ) conformational plane. The latter quantity is obtained by summing over ρ and λ at 5° increments yielding:

$$z_{\rho,\lambda} = \sum_{\rho} \sum_{\lambda} \exp[-V(\rho,\lambda)/RT] \tag{1}$$

Here $V(\rho_0, \lambda_0)$ is the energy of the ρ_0, λ_0 state, R the universal gas constant, and T the absolute temperature (298 K). Since contributions to $z_{\rho,\lambda}$ are negligible from points beyond the low energy minimum (i.e., where $V(\rho,\lambda) \geqslant 5$ kcal/mole), the partition function is further approximated by the summation over angular values between ±30°. The probabilities are converted into frequencies $F(\rho,\lambda)$ by multiplying the $w(\rho,\lambda)$ by 1000. Only nonzero frequency values are saved, generating N_F rotational isomeric state combinations for each of the sixteen different dimeric chain sequences (e.g., AA, AT, AG, AC, etc.). The angular values are stored in the 16 000 × 2 Monte Carlo selection array **A**, the sum of the $F(\rho,\lambda)$ over all dimers equaling 16 000. Elements of the array are assigned specific ρ and λ values in proportion to their frequency of occurrence. Each of the N_F combinations of ρ_0 and λ_0 is assigned to $F(\rho_0, \lambda_0)$ rows of **A**, roll angles being saved in one column and tilt angles in the other.[54]

Model building

Coordinates of successive base pairs of a DNA sequence are obtained from the transformation matrices $\mathbf{T}_{i,i+1}$ (τ, ρ, λ) relating neighboring coordinate frames and the translation vectors $\mathbf{v}_i$ = (0, 0, 3.4 Å) linking successive units. A standard twist of 36° is assumed for all residues. This value is close to that found to characterize poly dA · poly dT in solution[34] but is somewhat greater than that reported for poly dG · poly dC under the same conditions.[34] As noted elsewhere,[48] the bending tendencies of GC sequences are only slightly altered from those used here when τ is decreased by 2–3°. Local coordinate axes are chosen along the rotation axes specified above. The transformation matrix is arbitrarily defined by the product of the matrices of right-handed rotation about **z**, **x** and **y**, respectively, and is given by

$$\mathbf{T}_{i,i+1}(\tau, \rho, \lambda) = \begin{bmatrix} \cos\tau\cos\lambda - \sin\tau\sin\rho\sin\lambda & -\sin\tau\cos\rho & \cos\tau\sin\lambda + \sin\tau\sin\rho\cos\lambda \\ \sin\tau\cos\lambda + \cos\tau\sin\rho\sin\lambda & \cos\tau\cos\rho & \sin\tau\sin\lambda - \cos\tau\sin\rho\cos\lambda \\ -\cos\rho\sin\lambda & \sin\rho & \cos\rho\cos\lambda \end{bmatrix} \tag{2}$$

Coordinates $\mathbf{X}_m$ of the mth base pair are then transformed into the frame of the first base pair by simple matrix manipulations of the form:

$$\mathbf{X}'_m = [\mathbf{E}_3\mathbf{0}]\left\{\prod_{i=1}^{m-1}\begin{bmatrix} \mathbf{T}_{i,i+1} & \mathbf{v}_i \\ \mathbf{0} & 1 \end{bmatrix}\right\}\begin{bmatrix} \mathbf{X}_m \\ 1 \end{bmatrix} \tag{3}$$

where the $\mathbf{X}'_m$ are transformed coordinates, the **0**'s null matrices of orders required to conform, and $\mathbf{E}_3$ the identity matrix of order 3. The $\mathbf{T}_{i,i+1}$'s are defined in terms of the specific (ρ, λ) rotational pairs detailed above.

Monte Carlo chain generation technique

The Monte Carlo chain generation technique is outlined in Figure 1. Given the primary sequence of the DNA fragment to be studied, the elements of **A** corresponding to the sequence of the first dimer in the chain are identified. Random numbers (N_R) between 1 and N_F are then generated. The roll and tilt values of the N_Rth element of the respective sequential element of the array are used in Equation 3 to position the second hydrogen bonded base pair with respect to the first. The succeeding dimer sequence (involving base pairs 2 and 3) is next identified and the relevant segment of **A** located. The random number generation and selection process are repeated, and the third base pair unit is added to the second unit. This procedure is continued until the terminal base pair is added. The spatial displacement of the chain terminus **r** is recorded and used in the determination of density distribution. Unperturbed end-to-end chain dimensions $\langle r^2 \rangle_0$ and terminal residue orientation correlations[54] ($\langle \varphi_1, \rangle$, $\langle \varphi_2 \rangle$, and $\langle \gamma \rangle$) are used in the compilation of average values. The process is then repeated to generate additional configurations.

The average chain extension, or persistence vector, is determined from the end-to-end vectors of the Monte Carlo sample,

$$\langle \mathbf{r} \rangle = \frac{1}{N}\sum_{k=1}^{N} \mathbf{r}_k \tag{4}$$

the unperturbed mean-square end-to-end distance from the average scalar product of **r**,

$$\langle r^2 \rangle_0 = \frac{1}{N} \sum_{k=1}^{N} \mathbf{r}_k \cdot \mathbf{r}_k \tag{5}$$

and the orientational correlations from the scalar products of unit vectors in the first and last residues of the chain.

$$\langle \varphi_1 \rangle = \frac{1}{N} \sum_{k=1}^{N} \left\{ [1 \quad 0 \quad 0] \prod_{i=1}^{x-1} \mathbf{T}_{i,i+1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}_k \tag{6}$$

$$\langle \varphi_2 \rangle = \frac{1}{N} \sum_{k=1}^{N} \left\{ [0 \quad 1 \quad 0] \prod_{i=1}^{x-1} \mathbf{T}_{i,i+1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}_k \tag{7}$$

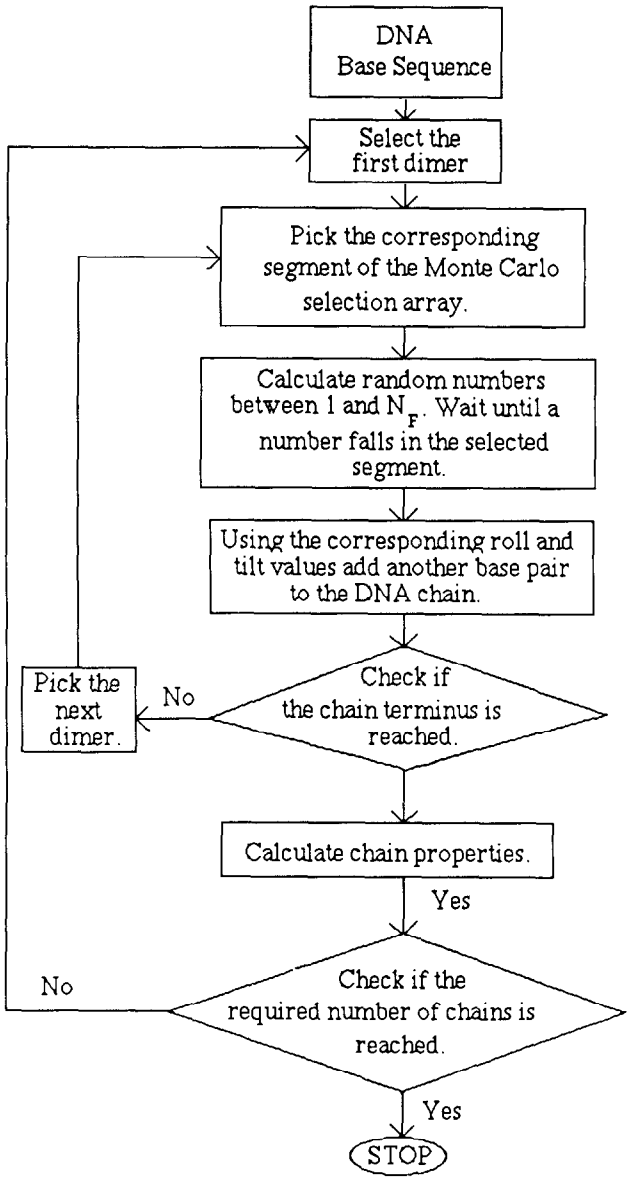

Figure 1. Flow chart of base sequence dependent Monte Carlo DNA chain generation program

$$\langle \gamma \rangle = \frac{1}{N} \sum_{i=1}^{N} \left\{ [0 \quad 0 \quad 1] \prod_{i=1}^{x-1} \mathbf{T}_{i,i+1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}_k \tag{8}$$

Here N is the size of Monte Carlo sample, and x is the number of residues in the chain fragment.

Configurations are generated until the average extension and orientation of the DNA match those determined by exact matrix generator techniques.[42,55,56] The random number generators are saved, along with the coordinates of the end-to-end vector **r** and the terminal **xyz** axes, every 1000 chains.

Spatial density distributions

The spatial probability density function $\mathbf{W_0(r)}\ \Delta\mathbf{r}$ is determined by counting the number of chains that terminate within specific volume elements inside a sphere of radius r_{max}, the maximum chain extension (3.4x Å). The sphere is divided into 3432 (11 × 13 × 24) wedges of volume $\Delta\mathbf{r} = r^2 \sin\theta\Delta\theta\Delta\psi\Delta r$, where $\Delta r = 0.1 r_{max}$ and $\Delta\psi = \Delta\theta = \pi/12$ rad. The chain coordinates are transformed to the spherical reference frame located at **0** and the appropriate volume element is identified. The probability density of each spherical wedge is then determined by the quotient of the number of chains that terminate in it to the total number of chains in the Monte Carlo sample. The radial density distribution function $R(r)\ \Delta r$ is obtained by summing over all θ and ψ for a fixed value of r. The spatial densities are displayed together with various representative chain trajectories (see below) on an Evans and Sutherland PS330 color graphics system using original interactive programs developed in this laboratory.

RESULTS

Color-coded representations of the spatial density distributions of three fragments of kinetoplast DNA from *Crithidia fasciculata* are illustrated in stereo in Color Plate 1. Fragment I of 150 base pairs (Color Plate 1) is known to be rodlike in solution, while fragment II of 211 base pairs (Color Plate 1) is among the most highly curved DNA sequences reported to date.[49] Fragment III of 274 base pairs (Color Plate 1) is intermediate in observed behavior.[49] The specific sequences of the chains are detailed in Table 1. The 211 residue chain is noticeably richer in AT base pairs and the 274 residue fragment richer in GC base pairs compared to the other chains, the proportion of A + T in fragments I–III being 0.52, 0.63 and 0.45, respectively. Fragment II is also characterized by several stretches of five to six recurring A's or T's. The distributions in Color Plate 1 are based on Monte Carlo samples of 10^6 spatial configurations of each chain at 298 K. These samples are found to reproduce within 0.1–0.5% the persistence lengths of the chain fragments computed by exact matrix generator techniques.[42] Indeed, chain dimensions are well reproduced (within 0.7–6.2%) in samples of as few as 10^2 chains. The average orientations of terminal

Table 1. DNA base sequence of kinetoplast fragments

Fragment	Chain length	Sequence				
I	150	CTACCACCCG	GGTCGGTTAA	ATATGTCGGC	CGTATTAAAG	CTCACATGAC
		AAAATCATGC	GATGAGGTTA	GGAAAGGGAA	AGACAATGTA	CCTTTCCGCG
		ATCACCATCC	TATATCTGTT	GCTTGACCCC	CTCTGTCTCC	AGGCAAGCTT
II	211	CCTAAAATTC	CAACCGAAAA	TCGCGAGGTT	ACTTTTTTGG	AGCCCGAAAA
		CCACCCAAAA	TCAAGGAAAA	ATGGCCAAAA	AATGCCAAAA	AATAGCGAAA
		ATACCCCGAA	AATTGGCAAA	AATTAACAAA	AAATAGCGAA	TTTCCCTGAA
		TTTTAGGCGA	AAAAACCCCC	GAAAATGGCC	AAAAACGCAC	TGAAAATCAA
		AATCTGAACG	T			
III	274	GCGGAAAATG	TCAGAAGTCC	ATTTCTGTCA	AACCCCCCAA	AAATCCAGAA
		TACGACCGAA	CGGAATACAC	ATAAACACAC	CCAGAAGCGA	AACAGCAACC
		AGGAAAGCGG	TGAAAACACC	CCCACCAAAC	CCAAGGCAGG	CCCAAAGTAC
		CAAACCAGCG	CAAATCACCT	CTGTCCAGCA	CAAACCCCGT	CCAAACCAGC
		ACTCAGACCC	AGGAAAACCC	CTCCCGGAGG	CCCCGAAATC	GGGCTAGAAC
		CCCGCCAAAC	CCCCTGCCAG	GAGG		

bonds, however, are less satisfactorily accounted for, the computed values of $<\gamma>$, $<\varphi_1>$, and $<\varphi_2>$ in the chain samples falling respectively within 3.0–7.5%, 7.2–21.3%, and 2.8–55.0% of the exact values.[42] The disagreement may stem, in part, from choosing configurations entirely at random, rather than on the basis of the relative energies of successive states of the sample.[7] The number of high energy conformations may therefore be overestimated in the present calculations. The agreement between Monte Carlo and exact orientational averages found here is best for the rodlike chains and worst with curved DNA.

If the distributions are Gaussian, the DNA is equally likely to terminate in any of the spatial wedges, with an expected population of roughly 300 states per wedge (10^6 configurations/3432 wedges = 291 configurations/wedge). The observed distributions in Color Plate 1 are defined with respect to this reference, states of lower probability being colored blue and states of comparable and higher probability magenta and red, respectively. Wedges with 100 or fewer configurations are not colored. The states included in the figures, nevertheless, account for 97.8, 96.5, and 96.4% of the respective Monte Carlo samples. Yellow rotational isomeric state models of each fragment are superimposed on the density distributions. The latter rigid representations are typically used in DNA modeling.[23-32,39-41] Successive base pairs are represented by characteristic roll and tilt angles rather than by the entire range of low energy conformations available to the Monte Carlo sample. The angles used in the current representation are chosen to reproduce, as closely as possible, the average transformation matrices[48] describing the orientation of adjacent base pairs. As evident from Color Plate 1, the Monte Carlo chains are clearly more flexible than the static models, even when confined to their more probable ranges of conformations. Furthermore, in none of the three fragments is the rotational isomeric state representation found in one of the most probable chain termination domains. The static models are also more extended than the most highly probable Monte Carlo (red) configurations. The average chain configurations described by the persistence vector $<\mathbf{r}>$ are significantly less extended than the rotational isomeric state models, the computed end-to-end distances of the rigid models being 493.3, 532.3, and 819.4 Å *versus* persistence vector lengths (e.g., $|<\mathbf{r}>|$) of 291.7, 273.1, and 363.5 Å, respectively.

The empty portions of the spheres are indicators of the least probable chain configurations. The unfilled volumes are also measures of the flexibility of the DNA fragments, a large empty space typifying a highly rigid chain. Fragment I (x = 150) is more rigid than fragments II (x = 211) and III (x = 274) on this basis. Only 731 of the 3432 spatial wedges of fragment I are colored in Color Plate 1 compared to 1105 and 1143 wedges of fragments II and III in Color Plate 1, respectively. Fragment I is also characterized by a larger number (103) of red dots, corresponding to the most highly probable spatial configurations, compared to fragments II (63) and III (21). Red is used to distinguish spatial volumes with 3000 or more configurations in the samples. The few red dots in Color Plate 1 (bottom) are also indicative of the greater flexibility of the 274 *versus* the 150 and 211 residue fragments. The most probable configurations of the longer chains are spread over a larger proportion of its configuration space compared to those of the shorter chains. The 21 most populous states in Color Plate 1 account for 13.8% of the Monte Carlo configurations of fragment I but only 10.5% of the configurations of fragment II and 6.6% of those of fragment III.

The locations of the spatial density distributions within their spheres of configuration space are indicative of the different morphologies of the three fragments. The cluster of dots are restricted almost exclusively to the positive ($\theta \leqslant \pi/2$ rad) hemispheres of Color Plate 1 (I and III), whereas the distribution of the 211 residue DNA is skewed and nearer the equator of the sphere in Color Plate 1 (II). The 150 residue chain is typically rodlike, tending to follow straight trajectories with termination points toward the "north pole" of the sphere preferred. The 274 residue fragment is, however, more flexible than the 150 residue chain, the chain ends

spanning a broader range of northerly spatial configurations. A chain of this length is more aptly described as a wormlike coil.[57] Shorter subfragments of the chain, however, are clearly rodlike. The rod- and wormlike behavior of the two fragments is also apparent from the rotational isomeric state models of the chains in Color Plate 1 (I and III). The bending of the two static models is fairly limited with radii of curvature over successive 10 residue segments spanning ranges of 231–8192 Å and 131–839 Å, respectively. The 211 residue fragment, in contrast, is intrinsically curved with the ends of the chain in much closer proximity, the relative extension of the sequence (r/r_{max}, where r is the end-to-end distance) being 0.74 *versus* 0.97 and 0.88 in fragments I and III. The rotational isomeric state model of fragment II is a superhelical trajectory with local radii of curvature ranging between 118 and 1450 Å. As noted above, the more compact chain II is less flexible than the extended fragment III, countering the common notion that polymer extension is determined by chain flexibility.

The tendency of the DNA to adopt closed circular forms is determined, in part, by the spatial probability density near $\mathbf{r} = \mathbf{0}$. Chain cyclization additionally requires the chain ends to adopt angular orientations that preserve standard valence bond angles and preferred single bond torsions of the double helix.[18–22] The computed probabilities of chain cyclization based upon the number of chain termini falling within a sphere of radius $0.1r_{max}$ are $10^{-5.7}$, $10^{-3.7}$, and $10^{-3.8}$ for fragments I, II, and III, respectively. The 150 residue sequence is not long enough to bend back on itself and form a cyclic species. Only two of the 10^6 configurations of the 150 residue chain are within the central volume elements of $W_0(\mathbf{r})$ compared to 185 of the 211 residue chain configurations and 163 of the 274 residue chain configurations. The predicted probabilities for Gaussian chains of the same length falling within the ring closure volume are $10^{-3.5}$, $10^{-3.1}$, $10^{-3.3}$. As evident from Figure 2, where the radial density distributions of the three chains are compared with the corresponding Gaussian distributions, the Monte Carlo systems are consistently more extended than the random models. The deviation from ideal behavior, however, is smaller for the fragment II than I and III. The Gaussian curves, $R(r)\,dr = (3/2'\pi < r^2 >_0)^{3/2}\, 4\pi r^2 \exp(-3r^2/2 < r^2 >_0)\, dr$, are derived from the characteristic ratio of the mean-square end-to-end dimensions of the chains ($< r^2 >_0/xv^2 = 98.8$, 75.1, and 127.3 Å, respectively, where x is the number of chain residues and $v = 3.4$ Å), and the probabilities of chain extension are based on subintegrals of the curves between $r = 0$ and r_{max}.

The density distributions are further compared in Color Plate 2 with a series of typical Monte Carlo configurations of the three sequences. The curved or rodlike character of three fragments are preserved in the various snapshots, even though the fragments deviate widely from the rotational isomeric state configurations. A number of the more highly probable configurations of fragment II, for example, are superhelical or circular. The trajectories are colored to match the wedges in

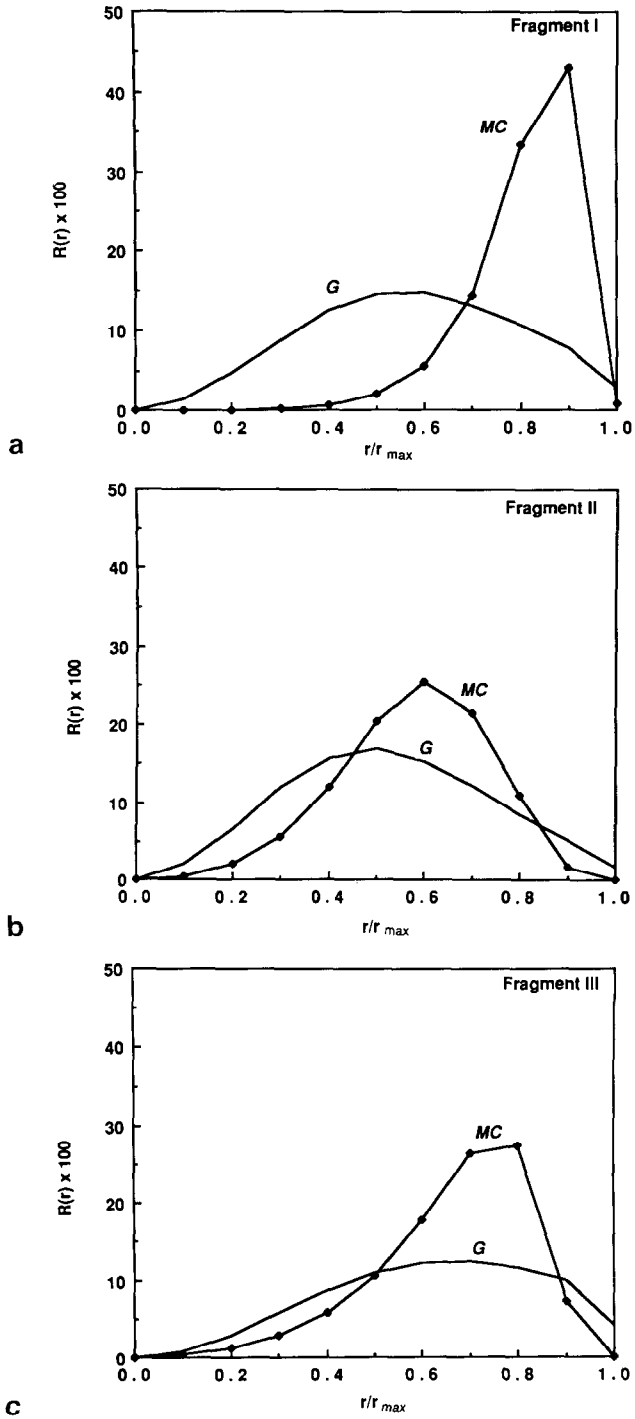

Figure 2. Radial density distributions $R(r)\,\Delta r$ of the end-to-end distances of kinetoplast fragments I–III as a function of relative chain extension based on the Monte Carlo (MC) sample and an ideal Gaussian (G) model

which they terminate and are then superimposed on the density distributions. Only points of higher than random probability (e.g., the red and magenta points of Color Plate 1, where $W_0(\mathbf{r}) > 300$) are included in Color Plate 2. The three-dimensional distributions are, nevertheless, representative of the Monte Carlo samples, the probabilities of the states of the Color Plate 2 I–III accounting respectively for 93.2, 89.0, and 89.3% of the computed

populations. The distributions in Color Plate 2 are also slightly tilted relative to those in Color Plate 1 to emphasize the mushroom capped shapes of the density distributions. The most probable configurations of the DNA are nearer the surface than the centers of the spheres. The population of states near the center is clearly greater for the curved 211 residue chain compared to the rodlike fragments and for the more flexible 274 residue chain than the 150 residue sequence. Finally, the Monte Carlo representations included in the figure are not necessarily typical of all configurations that terminate at the specified spatial locations. The chains falling in a particular spatial wedge are likely to bend in a variety of ways owing to the considerable flexibility of individual chain units. Differences are expected to be more pronounced in AT rich chains owing to the greater local flexibility (e.g., both positive and negative rolling motions) of the constituent repeating sequences.

CONCLUSIONS

The color-coded spatial density distributions described above appear to be useful descriptors of macromolecular flexibility. The distributions incorporate a vast quantity of data which cannot be comprehended at the molecular level. The Monte Carlo sampling technique also provides a more complete picture of the sequence dependent properties of a chain molecule than typical static models. Chains limited to narrow ranges of local residue mobility span significantly larger volumes of configuration space than indicated by rigid representations. Adjacent base pairs of the DNAs considered in this study, for example, bend, on the average, 5–10°. The end of the chains constructed from these residues however, adopt 20–30% of the theoretically possible spatial locations, depending upon length and sequence. Furthermore, very long DNAs of the same sequence obey Gaussian statistics and terminate with equal probability in any part of the normalized density sphere.

The graphical representations additionally offer a simple means to study the effects of chain length and residue sequence on overall chain properties. The symmetry and size of the distribution provide immediate indication of the macroscopic character of the polymer. Rodlike chains, for example, terminate within small symmetric mushroom shaped density volumes, whereas curved sequences define skewed densities with respect to the north-south axis of the probability sphere. Longer chains span broader regions of space than shorter chains. The longer chains not only spread over larger areas of the density surface but also terminate within larger portions of the spherical interior. The density of points within the centers of the probability spheres provides a useful estimate of the cyclization tendencies of different chains. Quantitative comparison of experimental data with theoretical models, however, requires incorporation of conditional orientational probabilities determined on the basis of the sample of chains with appropriate chain extension.[21,22] A study of this sort requires a greater number of chains than available in the current Monte Carlo samples.

Finally, the density distributions detailed here are qualitatively consistent with the observed gel mobilities of three kinetoplast DNA fragments. The more rodlike 150 residue fragment is expected to meander more easily through a gel matrix than the curved 211 residue chain. The relative stiffness of the 211 residue fragment compared to the 274 residue chain compounded with the difference in curved and wormlike character may account for the relative ease of the longer chain to pass through the gel. Flexibility has heretofore been ignored in models of curved versus rodlike DNA. Both chain size and mobility must be considered in the development of theories of the movement of macromolecules through dense gels.

ACKNOWLEDGEMENT

Sponsorship of this research by the U.S. Public Health Service under grant GM–20861 is gratefully acknowledged. Calculations were performed at the Rutgers Center for Computational Chemistry and the John von Neumann Supercomputer Center.

REFERENCES

1 McCammon, J. A. and Karplus, M. Simulation of protein dynamics. *Ann. Rev. Phys. Chem.*, 1980, **31**, 29–45
2 Levitt, M. Protein conformation, dynamics, and folding by computer simulation. *Ann. Rev. Biophys. Bioeng.*, 1982, **11**, 251–271
3 Weiner, P. K. *et al.* Visualization of energetics and conformations from molecular computer simulations. *J. Mol. Graphics*, 1986, **4**, 203–207
4 Vologodskii, A. V. *et al.* Statistical mechanics of supercoils and the torsional stiffness of the DNA double helix. *Nature*, 1979, **280**, 294–298
5 Ermak, D. L. and McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 1982, **69**, 1352–1360
6 Northrup, S. H. *et al.* Dynamical theory of activated processes in globular proteins. *Proc. Natl. Acad. Sci. USA*, 1982, **79**, 4035–4039
7 Metropolis, N. A. *et al.* Equation of state calculations by fast computing machines. *J. Chem Phys.*, 1953, **21**, 1087–1092
8 Fixman, M. and Alben, R. Polymer conformational statistics. I. Probability distribution. *J. Chem. Phys.*, 1973, **58**, 1553–1558
9 Premilat, S. and Hermans, J., Jr. Conformational statistics of short chains of poly(L-alanine) and poly-(gycine) generated by Monte Carlo method and the partition function of chains with constrained ends. *J. Chem. Phys.*, 1973, **59**, 2602–2612
10 Premilat, S. and Maigret, B. Effects of long range interactions on the conformational statistics of short polypeptide chains generated by a Monte Carlo method. *J. Chem. Phys.*, 1977, **66**, 3418–3425
11 Leclerc, M. *et al.* Interpretation of energy-transfer experiments by theoretical studies of model compounds using semiempirical potential functions.

II. Monte Carlo calculation on oligopeptides. *Biopolymers*, 1977, **16**, 531–544
12 Rubin, H. and Kallenbach, N. R. Conformational statistics of short RNA chains. *J. Chem. Phys.*, 1975, **62**, 2766–2776
13 Hagerman, P. J. Analysis of the ring closure probabilities of isotropic wormlike chains: Application to duplex DNA. *Biopolymers*, 1985, **24**, 1881–1897
14 Levene, S. D. and Crothers, D. M. Ring closure probabilities for DNA fragments by Monte Carlo simulation. *J. Mol. Biol.*, 1986, **189**, 61–72
15 Bassolino, D. A. *et al.* Determination of protein structure in solution using nmr data and IMPACT. *Int. J. Supercomputer Applicns.*, 1988, **2**, 41–46
16 Flory, P. J. Moments of the end-to-end vector of a chain molecule, its persistence and distribution. *Proc. Natl. Acad. Sci. USA*, 1973, **70**, 1819–1823
17 Flory, P. J. and Yoon, D. Y. Moments and distribution functions for polymer chains of finite length: Theory. *J. Chem. Phys.*, 1974, **61**, 5358–5365
18 Yevich, R. and Olson, W. K. The spatial distributions of randomly coiling polynucleotides. *Biopolymers*, 1979, **18**, 113–145
19 Olson, W. K. The flexible DNA double helix. I. Average dimensions and distribution functions. *Biopolymers*, 1979, **18**, 1213–1233
20 Olson, W. K. The flexible DNA double helix, in *Stereodynamics of Molecular Systems*, Sarma, R.H. (Ed.), Pergammon Press, New York, 1979, pp. 297–314
21 Marky, N. L. and Olson, W. K. Loop formation in polynucleotide chains. I. Theory of hairpin loop closure. *Biopolymers*, 1982, **21**, 2329–2344
22 Marky, N. L. and Olson, W. K. Loop formation in polynucleotide chains. II. Flexibility of the anticodon loop of $tRNA^{Phe}$. *Biopolymers*, 1987, **26**, 415–438
23 Calladine, C. R. Mechanistics of sequence dependent stacking of bases in B–DNA. *J. Mol. Biol.*, 1982, **161**, 343–352
24 Calladine, C. R. and Drew, H. R. A base centered explanation of the B to A transition in DNA. *J. Mol. Biol.*, 1984, **178**, 773–782
25 Calladine, C. R. and Drew, H. R. Principles of sequence dependent flexure of DNA. *J. Mol. Biol.*, 1986, **192**, 907–918
26 Dickerson, R. E. and Drew, H. R. Kinematic model for B–DNA. *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 7318–7322
27 Dickerson, R. E., Kopka, M. L. and Pjura, P. A random walk model for helix bending in B–DNA. *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 7099–7103
28 Trifonov, E. N. and Sussman, J. L. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 3816–3820
29 Arnott, S. *et al.* Heteronomous DNA. *Nucleic Acids Res.*, 1983, **11**, 4141–4155
30 Prunell, A. *et al.* The smaller helical repeat of poly(dA) · poly(dT) relative to DNA may reflect the wedge property of the dA·dT base pair. *Eur. J. Biochem.*, 1984, **138**, 253–257
31 Jernigan, R. L. *et al.* Hydrophobic interactions in the major groove can influence DNA local structure. *J. Biomol. Str. and Dynamics*, 1986, **4**, 41–48
32 Ulanovsky, L. E. and Trifonov, E. N. Estimation of wedge components in curved DNA. *Nature*, 1987, **326**, 720–722
33 Shakked, Z. and Rabinovich, D. The effect of the base sequence on the fine structure of the DNA double helix and references cited therein. *Prog. Biophys. and Molec. Biol.*, 1986, **41**, 159–195
34 Peck. L. J. and Wang, J. C. Sequence dependence of the helical repeat of DNA in solution. *Nature*, 1981, **292**, 375–378
35 Rhodes, D. and Klug, A. Sequence dependent helical periodicity of DNA. *Nature*, 1981, **292**, 378–380
36 Kabsch, W., Sander, C. and Trifonov, E. N. The ten helical twist angles of B–DNA. *Nucleic Acids Res.*, 1982, **10**, 1097–1104
37 Wu, H.-M. and Crothers, D. M. The locus of sequence directed and protein induced DNA bending. *Nature*, 1984, **308**, 509–513
38 Koo, H.-S., Wu, H.-M. and Crothers, D. M. DNA bending at adenine · thymine tracts. *Nature*, 1986, **320**, 501–505
39 Zahn, K. and Blattner, F. R. Direct evidence for DNA bending at the lambda replication origin. *Science*, 1987, **236**, 416–422
40 Levene, S. D. and Crothers, D. M. A computer graphics study of sequence-directed bending in DNA. *J. Biomol. Str. and Dynamics*, 1983, **1**, 429–435
41 Tung, C.-S. and Harvey, S. C. Computer graphics programs to reveal the dependence of the gross three-dimensional structure of the B–DNA double helix on primary structure. *Nucleic Acids Res.*, 1986, **14**, 381–387
42 Olson, W. K. and Srinivasan, A. R. The translation of DNA primary base sequence into three-dimensional structure. *Computer Applications in the Biosciences*, 1988, **4**, 133–142
43 Ulyanov, N. B. and Zhurkin, V. B. Anisotropic flexibility of DNA depends upon base sequence. Conformational calculations of double stranded tetranucleotides AAAA:TTTT, $(AATT)_2$, $(TTAA)_2$, GGGG:CCCC, $(GGCC)_2$, $(CCGG)_2$, *Mol. Biol. USSR* (Eng. Ed.), 1984, **18**, 1366–1384
44 Ulyanov, N. B. and Zhurkin, V. B. Sequence dependent anisotropic flexibility of B–DNA: A conformational study. *J. Biomol. Str. and Dynamics*, 1984, **2**, 361–385
45 Olson, W. K. *et al.* The effects of base sequence and morphology upon the conformation and properties of double helical DNA, in *Biomolecular Stereodynamics IV*, Sarma R. H. and Sarma, M. H. (Eds.) Adenine Press, Guilderland, NY, 1985, pp. 75–100
46 Tung, C.-S. and Harvey, S. C. Base sequence, local helix structure, and macroscopic curvature of A–DNA and B–DNA. *J. Biol. Chem.*, 1986, **261**, 3700–3709
47 von Kitzing, E. and Diekmann, S. Molecular

mechanics calculations of dA_{12}. dT_{12} and of the curved molecule $d(GCTCGAAAAA)_4 \cdot d(TTTTTCGAGC)_4$. *Eur. Biophys. J.*, 1987, **15**, 13–26

48 Srinivasan, A. R. *et al.* Base sequence effects in double helical DNA. I. Potential energy estimates of local base morphology. *J. Biomol. Str. and Dynamics*, 1987, **5**, 459–496

49 Kitchin, P. A. *et al.* A highly bent fragment of *Crithidia fasciculata* kinetoplast DNA. *J. Biol. Chem.*, 1986, **261**, 11302–11309

50 Arnott, S., Dover, S. D. and Wonacott, A. J. Least-squares refinement of the crystal and molecular structures of DNA and RNA from X-ray data and standard bond lengths and angles. *Acta Cryst.*, 1969, **B25**, 2192–2206

51 Fratini, A. V. *et al.* Reversible bending and helix geometry in a B–DNA dodecamer: $CGCGAATT^{Br}CGCG$. *J. Biol. Chem.*, 1982, **257**, 14686–14707

52 Holbrook, S. R., Dickerson, R. E. and Kim, S.–H. Anisotropic thermal-parameter refinement of the DNA dodecamer CGCGAATTCGCG by the segmented rigid-body method. *Acta Cryst.*, 1985, **B41**, 255–262

53 Hogan, M., LeGrange, J. and Austin B. Dependence of DNA helix flexibility on base composition. *Nature*, 1983, **304**, 752–754

54 Srinivasan, A. R. and Ponnuswamy, P. K. On the existence of *cis/trans* peptide mixtures in poly (N–methyl glycine). *Polymer*, 1977, **18**, 107–110

55 Maroun, R.C. and Olson, W.K. Base sequence effects in double helical DNA. II. Configurational statistics of rodlike chains. *Biopolymers*, 1988, **27**, 561–584

56 Maroun, R. C. and Olson, W. K. Base sequence effects in double helical DNA. III. Average properties of curved DNA. *Biopolymers*, 1988, **27**, 585–603

57 Kratky, O. and Porod, G. Röntgenuntersuchung gelöster fadenmoleküle. *Rec. Trav. Chim.*, 1949, **68**, 1106–1122