Biophysical Chemistry 11 (1980) 329–337

THE FLEXIBILITY OF DNA. II. SPONTANEOUS AND LIGAND INDUCED DISTORTIONS ☆

John A. SCHELLMAN
Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403

DNA can bend not only as a result of smooth worm-like deviations from linearity, but also in an abrupt fashion because of natural spontaneous fluctuations in structure (breathing modes, kinks, etc.) and because its interaction with other molecules (dyes, proteins, other ligands). It is shown how all these structural deformations can be treated with a single geometrical model and the statistical theory of such deformations is developed into specific formulas for persistence and mean directions. The statistical theory does not apply to bending which is cooperative or site specific.

Glossary

$\langle h^2 \rangle_0$	Mean square end to end distance of undeformed DNA.
$\langle h_d^2 \rangle_0$	Mean square end to end distance of DNA deformed by ligands or fluctuations.
k	The number of DNA base pairs which are a structural part of a deformation.
l_0	The link length of DNA, 3.38 Å.
l_i	The link vector of link i.
l_β	The link vector of a β link, fig. 3.
N_d	The number of deformations in the DNA chain.
N_f	The number of free or normal base pairs in a DNA.
N_ℓ	The number of links, $N_d + N_f$, in the DNA molecule.
N_t	The total number of base pairs in the DNA.
p	The probability of a free link.
P_∞	The normal persistence length of DNA.
P_α	The persistence length starting on a normal (α) link in the presence of deformation.
P_i^+, P_i^-	The persistence length calculated in the forward and backward direction, starting on link i.
P_β^+, P_β^-	The forward and backward persistence length starting on a β link (fig. 3).
P_γ^+, P_γ^-	The forward and backward persistence length starting on a γ link (fig. 3).
q	The probability of encountering a deformation in the chain.
$r = N_d/N_t$	The number of ligands or deformations per base pair.
α	$\langle \cos\theta_\alpha \rangle$.
β	$l_0^{-1}\langle l_\beta \cos\theta_\beta \rangle$, see fig. 3.
$\bar{\beta}$	$l_0^{-1}\langle\langle l_\beta \cos\theta_\beta \rangle\rangle$, the second average is over the orientation of the ligand on the DNA chain.
$\bar{\gamma}$	$\langle \cos\theta_\gamma \rangle$, see fig. 3.
θ_α	angle between a normal link and a normal link which directly precedes it.
θ_β	angle between the vector l_β which spans the deformation and the normal link which precedes it (fig. 3).
θ_γ	angle between the helix axis of the DNA which enters the deformation and the helix axis of the DNA which leaves it.

☆ Research supported in part by grants from the National Institutes of Health (GM 20195, GM 15423, GM 00715) and the National Science Foundation (PCM 76-80741).

1. Introduction

The preceding paper [1] has dealt with the flexibility acquired by DNA by virtue of small local structural variations. These small deviations from rigidity result from thermal excursions on the conformational free energy surface which remain close to the minimum free energy which determines the static equilibrium structure.

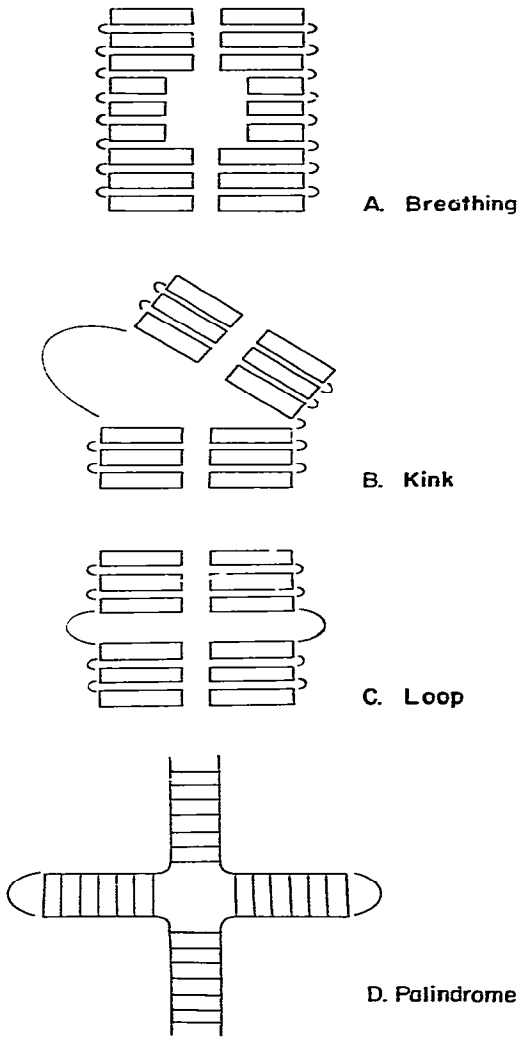

Fig. 1. Schematic illustrations of spontaneous DNA deformations. A. A breathing mode [4] Bases are still stacked but are flipped around to allow hydrogen exchange. The change in structure will cause a change in local flexibility properties. B. Kinked DNA. There are a vast number of possibilities but our discussion will use as examples two specific structures which have been proposed in connection with chromatic folding [5,6]. C. Internal loops. These are probably not of great importance except near the helix–coil transition. D. Palindromic structures. These clearly produce regions of changed flexibility and chain propagation.

This is the equivalent of the smooth bending models of Kratky and Porod [2] or Landau and Lifschitz [3]. The present paper will investigate the effects of larger deviations from the stable equilibrium structure on the chain properties of DNA and its flexibility. Two sources of the large local deformation will be considered: structural defects which arise from spontaneous fluctuations in the DNA itself and structural changes induced by ligands. Before beginning it will be useful to survey the types of deformations which may be included in the model presented in the next section.

Spontaneous structural defects which have been proposed for DNA include kinks, breathing modes, local unstacking, loops, etc. Schematic illustrations and descriptions are provided in fig. 1. The important thing which must be remembered is that DNA molecules are very long so that even relatively improbable events will occur at some or many positions along the chain. It can be calculated that a structural defect which requires 7 kcal/mole of free energy for its formation will occur with a frequency of unity for a DNA of 10^8 Daltons. Lower free energies lead to much higher frequencies. For example, Mandal et al. [7] have recently estimated that the structural defect which permits the exchange of a certain class of protons in double stranded nucleic acids occurs with a probability of about 5%.

Deviations from the wormlike propagation of DNA can also arise from the binding of a wide variety of ligands. Some examples are depicted in fig. 2 including intercalation which changes the length of the chain as well as introducing bends and changes in flexibility. Any ligand is likely to cause changes in the overall flexibility of DNA by bending or straightening, loosening or stiffening the local structure.

2. A geometric model

For the purpose of chain statistics, all of the physical situations of fig. 1 and fig. 2 as well as many others can be represented by a single geometric construction which is shown in fig. 3. The important features of the model are the direction and position where the normal DNA chain terminates as it enters the liganded or deformed region (Base pair b) and the position and orientation where it resumes normal propagation (Base pair m). The rectangles indicate base pairs which are part of the normal DNA chain, i.e., segments displaying the normal flexibility properties relative to

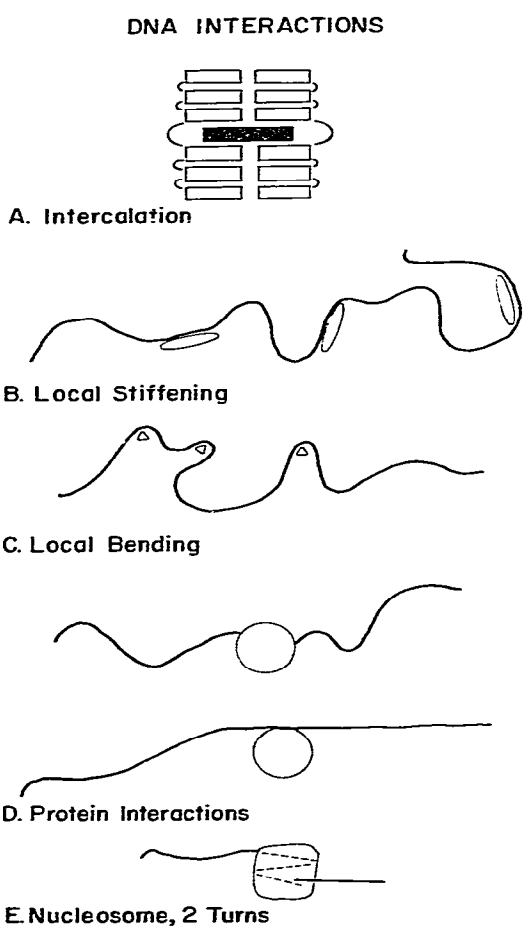

Fig. 2. Schematic illustrations of DNA deformations caused by ligands. A. Intercalation. The chain is lengthened accompanied by a change in direction and presumably a change in local flexibility. B. Local stiffening. Interaction with ligands conceivably stiffen and/or straighten the DNA helix. In the present context these changes from normal flexing behavior are deformations. C. Local bending. The DNA is partially wrapped around a ligand causing a bend and a change in flexibility. D. Protein interactions. These can be extensive and involve many base pairs. As will be seen the important factors are the geometrical relationships of the DNA helices which enter and leave the complex. E. The nucleosome. This is an extreme case of a protein interaction. For the model in which there are two turns of DNA wrapped about the nucleosome, there is not much change in direction but a large dislocation and about 160 base pairs are tied up in the complex [8]. For the model in which there are $1\frac{3}{4}$ turns about the nucleosome (not shown in figure) there is again a large dislocation but also a 90° change in helix direction [8], which, as is shown in the text, means a complete loss of memory of chain direction.

contiguous base pairs. The small arrows along the chain are the link vectors for the base pairs [9,1]. The vector $\boldsymbol{l}_\beta$ spans the liganded or distorted region. The two parameters of interest are the projection of $\boldsymbol{l}_\beta$ on the last normal link ratio (base pair b) and the projection of the link vector for base pair m on the link vector of base pair b. Since there will normally be flexibility in the deformation, averages are required. The two dimensional parameters of importance will be $\beta \equiv \langle l_\beta/l_0 \cos\theta_\beta \rangle$ and $\gamma \equiv \langle \cos\theta_\gamma \rangle$. The figure is not planar. In general there will be dislocations of the helices before and after the deformation but these do not enter into persistence properties.

3. Stoichiometry

In performing chain averages it will be required not only to average over the conformations of each link but to average as well over the kind of links which arise. Links will be divided into three kinds: α-links for normal propagation, e.g. base pairs b or n in fig. 3; β-links which span the ligand or deformation and γ-links which follow β links. (Note that a γ-link going in one direction becomes on α link going in the other and that parameter β depends on direction for unsymmetrical ligands or deformations. These features of the model will be taken into consideration). The assumption will be made that the link probabilities are independent of neighboring links with two exceptions. The first is that a β link is always followed by a γ link except at the terminus of the chain which will be ignored. The second is that two contiguous β links will not be allowed in the analysis. Clearly two contiguous β links indicate an enlargement of the deformed region and introduce propagation properties characteristic of ligand–ligand interaction. This interesting case requires a broader analysis than the following.

The probability of a deformation ($\beta+\gamma$-link) will be designated by q and of an α-link by p with $p+q=1$. q is related to the number of ligands per base pair or deformations per base pair either of which will be symbolized by r. By definition, $r \equiv N_d/N_t$ where N_d is the number of deformations in the chain and N_t is the total number of base pairs in the DNA. Now $q \neq r$ in general because there are an average of k base pairs of bound or deformed DNA in the β link. We shall define a deformation link, $\boldsymbol{l}_\beta + \boldsymbol{l}_\gamma$, as a pair of adjacent β and γ links.

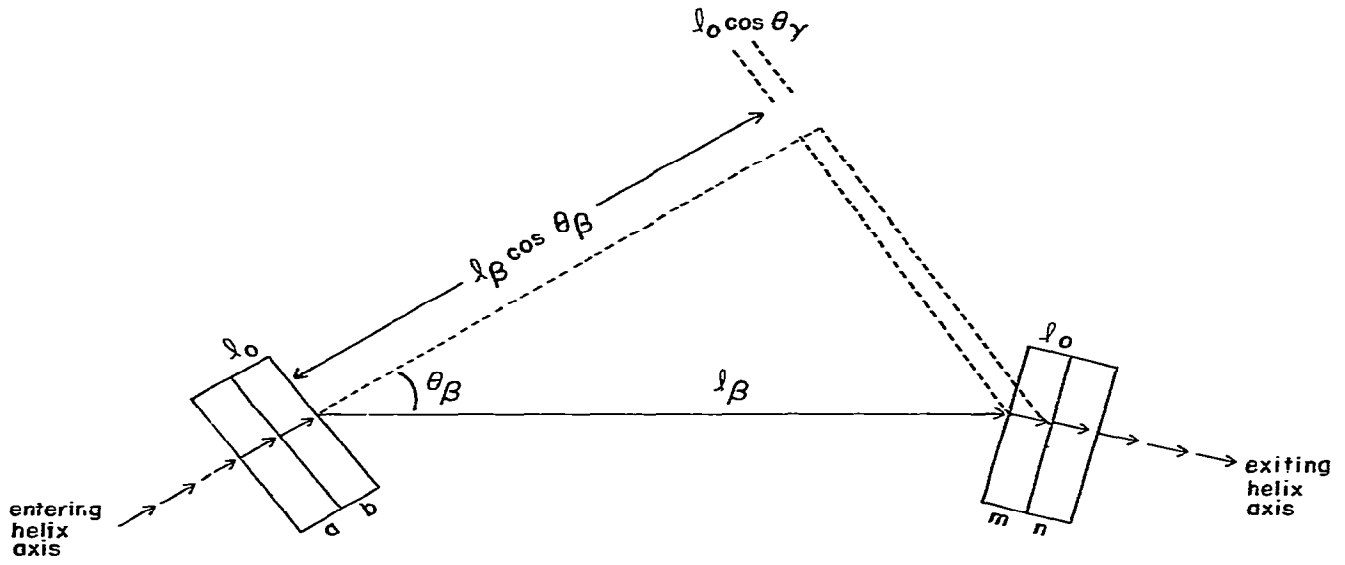

Fig. 3. A chain displacement caused by a ligand or spontaneous deformation. Normal chain propagation ends at base pair b and resumes again after base pair m. Intervening is a ligand (or deformation) and a number of DNA segments, k, which do not propagate like normal DNA segments either because they are interacting with ligand (fig. 2) or because they are part of a deformation (fig. 1). k may be zero for a kink or intercalation.

This gives a conditional probability of unity that a β link is followed by a γ link in the discussion which follows, which is otherwise based on independent statistics. The number of normal links, N_f, is $N_f = N_t - (k+1)N_d$ where the unity in $(k+1)$ takes the γ link into account. The total number of links, N_ϱ, will be the sum of N_f and N_d. Thus

$$p = \frac{N_f}{N_\varrho} = \frac{N_f}{N_d + N_f} = \frac{1-(k+1)r}{1-kr}, \quad q = \frac{N_d}{N_\varrho} = \frac{r}{1-kr}. \tag{1}$$

These are the formulas connecting the stoichiometric quantities k and r with the link probabilities. In some instances, especially breathing modes, the number of base pairs in the deformation may be variable so that k is an average.

4. The persistence length of a deformed DNA chain

The persistence length of a homogeneous chain is defined by the relation

$$P_\infty = \left\langle \boldsymbol{l}_1 \cdot \sum_{i=1} \boldsymbol{l}_i \right\rangle / l_0 , \tag{2}$$

where $\boldsymbol{l}_1$ may be any link in the chain except those too close to ends where the sum is cut off before it attains its limiting value. The average is over chain conformation. For the present problem the starting link might be an α-link or a deformation consisting of β and γ links. Consequently there will be three kinds of persistence calculations. We first consider the persistence length starting with a normal link. Excluded volume contributions to the statistics will be ignored so that the calculations are valid only for a theta solvent. Many workers in the field believe that DNA is sufficiently stiff that 0.2 M salt or higher is effectively a theta solvent for it, but this opinion is not universally accepted. The theories of Yamakawa and coworkers [10,11], which appear to give an accurate account of the sedimentation and viscosity of DNA solutions, neglect excluded volume. Writing eq. (2) out explicitly,

$$P_\alpha = l_0(1 + \boldsymbol{l}_2 \cdot \boldsymbol{l}_1 / l_0^2 + (\boldsymbol{l}_3 \cdot \boldsymbol{l}_1)/l_0^2 + \ldots) . \tag{3}$$

To provide an initial simplification, contribution from β links will be ignored. The persistence is then composed of α links with probability p and γ links with probability q. (This case would cover, for example,

kinks where l_β is the gap in the broken DNA helix). A term like $(\boldsymbol{l}_{n+1}\cdot\boldsymbol{l}_1)/l^2$ is of the form $g(i,j)\alpha^i\gamma^j$ where $i+j=n$ and where $g(i,j)$ is the probability that there are i α-links and j γ-links in the chain of n projections of $\boldsymbol{l}_{n+1}$ on $\boldsymbol{l}_1$. $\alpha=\langle\cos\theta_\alpha\rangle$ is the mean projection of an ordinary DNA link on its predecessor and γ is defined in fig. 3. Order does not matter. Because of the assumption of independence $g(i,j)=\binom{n}{j}p^iq^i$ where $\binom{n}{j}$ is the binomial coefficient. Thus the sum over all possibilities for $(\boldsymbol{l}_{n+1}\cdot\boldsymbol{l}_1)/l_0^2$ is $(p\alpha+q\gamma)^n$, and the persistence length is given as the sum

$$P_\alpha/l_0=1+\sum_{n=1}^{\infty}(p\alpha+q\gamma)^n=\frac{1}{1-p\alpha-q\gamma}. \qquad (4)$$

The contribution of β-links to the sum can be taken into account in the following way. A main point is that though a β-link is projected downward onto $\boldsymbol{l}_1$, it always terminates a sequence since nothing is projected onto it. As can be seen from fig. 3, the segments on the right side of the deformation are projected onto base pair m and thence directly onto b, bypassing the β-link. A typical β contribution at link $\boldsymbol{l}_{n+2}$ is $\alpha^i\gamma^j\beta$ with $i+j=n$, where the β factor is always in the $n+2$ position. The γ-contributions arise from j deformations intervening between $\boldsymbol{l}_1$ and $\boldsymbol{l}_{n+2}$. This gain yields a binomial contribution $(p\alpha+q\gamma)^n q\beta$. Summing over all values of n

$$q\beta\sum_{n=0}^{\infty}(p\alpha+q\gamma)^n=q\beta\left(1+\sum_{n=1}^{\infty}(p\alpha+q\gamma)\right)$$

$$=q\beta/(1-p\alpha-q\gamma)\,.$$

Adding this to the contribution of the α and γ terms gives

$$P_\alpha/l_0=(1+q\bar{\beta})/(1-p\alpha-q\gamma)\,. \qquad (5)$$

This formula can be verified by expanding the denominator and showing that it systematically generates all possible sequences of links. Because of the assumed independence it also generates sequences like $\gamma\gamma(\beta+\gamma)$. This describes a situation in which there are three consecutive deformations separated only by single base pairs, i.e., base pair m of one deformation is also base pair b of the next. This type of crowding will not be allowable in a number of situations, but is permissable in others. In the interest of the simple formulas which result from the present line of analysis, this complication will be avoided until a later paper. In any case, the formulas will be applicable at low $\bar{r}$ where the terms associated with near contact of deformation on the DNA make negligible contributions to the sums.

We also require a persistence length that begins on the l_β and l_γ links of fig. 3. If the ligand is symmetric, then the geometric relationship of l_β to base pair b is the same as its relation to base pair m. Under these circumstances, proceeding either way along the chain will give the same persistence contribution. If the ligand is not symmetric then a different persistence will be obtained for going forward or backward along the chain (actually for going forward or backward along the ligand, since the DNA chain is assumed to have no directionality). This situation can be handled by assigning a nominal positive and negative direction and using the symbols $P_\beta^+, P_\beta^-, \beta^+, \beta^-$ etc. The same problem arises in eq. (5) where the β contribution will depend on the directionality of the deformation. Since the two directions occur with equal probability, $\bar{\beta}$ in the equation is to be interpreted as $(\beta^+ + \beta^-)/2$.

The β persistences are easily seen to be

$$P_\beta^+ = l_\beta + \beta^- P_\alpha\,, \quad P_\beta^- = l_\beta + \beta^+ P_\alpha\,. \qquad (6)$$

The β^- results from the fact that the relation between a γ and a β link in the forward direction is the same as an α and a β link going in the opposite direction. The P_α terms arise because the links adjacent to a β link are DNA base pairs by the restrictions of the model, and the projection of the rest of the chain on this unit is P_α.

For the γ links one needs a persistence going both toward and away from the adjacent β link. The latter is just P_α

$$P_\gamma^- = l_0\bar{\beta} + \gamma P_\alpha\,, \quad P_\gamma^+ = P_\alpha\,. \qquad (7)$$

Here $\bar{\beta}=(\beta^+ + \beta^-)/2$ since we are averaging over both orientations of the ligand. The γ link is considered to be on the + side of the β link by convention. Later $l_\beta(P_\beta^+ + P_\beta^-)$ will be encountered which must be averaged over the conformation of l_β. From eq. (6) the average is

$$\langle l_\beta(P_\beta^+ + P_\beta^-)\rangle = 2\{\overline{l_\beta^2} + l_0\bar{\beta}P_\alpha\}\,. \qquad (8)$$

5. Chain dimensions

The mean square end to end distance of a chain of links is given by

$$\langle h^2\rangle_0 = \left\langle \left(\sum l_i\right)\cdot\left(\sum l_j\right)\right\rangle = \left\langle \sum_i \left(\sum_j l_j\right)\right\rangle$$

$$= \sum_i (l_i P_i^+ + l_i P_i^- - l_i^2)\,, \qquad (9)$$

where P_i^+ is the forward persistence measured from link i and P_i^- is the backward persistence length. The P_i may be finite persistences (i.e., sums which have not converged because of insufficient chain length) if link i is near one of the ends of the chain. If the chain is very long, end effects can be ignored and P_i in eq. (8) can be equated to the persistence length for infinite chains. In the deformed DNA chain there will be pN_ϱ links of standard length l_0 and qN_ϱ deformations consisting of adjacent β and γ links. Thus

$$\langle h_d^2\rangle_0 = N_\varrho\{2pl_0P_\alpha + ql_\beta(P_\beta^+ + P_\beta^-) + ql_0(P_\gamma^+ + P_\gamma^-) - pl_0^2 - q(\overline{l_\beta^2} + l_0^2)\}\,. \qquad (10)$$

Compare with eq. (5) of ref. [1]. Substituting from eqs. (7) and (8)

$$\langle h_d^2\rangle_0 = N_\varrho\{2l_0P_\alpha[p + q(\bar\beta + (1+\gamma)/2)] + q(\overline{l_\beta^2} + l_0\overline{l_\beta}) - l_0^2\}\,, \qquad (11)$$

and eliminating p, q and N_ϱ via eqs. (1),

$$\langle h_d^2\rangle_0 = N_t\{2l_0P_\alpha[1 + r(\bar\beta - k + (\gamma-1)/2)] + r(\overline{l_\beta^2} + l_0l_\beta + kl_0^2) - l_0^2\}\,, \qquad (12)$$

where P_α is given by eq. (5). The quantities p and q can also be removed from P_α but it is probably easier to calculate them from r and k and substitute into eq. (5).

Eq. (12) is the final result. It relates the m' dimension of a DNA molecule to the persistence of native DNA via the parameter α, to the structure of its deformations or binding complexes via l_β, β and γ, and to stoichiometry via r and k. Since P_α is large, it is frequently possible to drop the last two terms within the braces of eq. (12). This is standard for ordinary DNA where $P_\infty l_0$ is clearly much greater than l_0^2. It is less clear, however, that $P_\alpha l_0$ will always be much larger than l_β^2, especially for proteins.

6. Applications

6.1. Spontaneous deformations

We first consider the flexibility behavior of DNA in the absence of ligands. This is usually expressed in terms of the persistence length, P_∞. The persistence length is usually interpreted in terms of a worm like coil though in fact it is not actually known to what extent DNA flexibility arises from spontaneous fluctuations of the kind shown in fig. 1.

Consider first a kink in the DNA chain. We define a kink as a sudden localized change in DNA direction which arises from an alternative local structure of higher free energy (fig. 1A–C). Structural models for kinks have been proposed [5,6] in connection with the chromatin problem, though we prefer to consider kinks as a class of deformations. Fig. 3 can be construed as a kink with l_β, the vector which spans the gap between the two segments of DNA. For simplicity we assume that the kink only affects the structure and flexibility of bases b and m and not base pairs which are further removed. If this is not a good assumption the kink must be defined to include the perturbed base pairs. To simplify the discussion, we also assume that the effect of lengthening the chain by the gap vector l_β is small compared with the change in chain direction given by γ. We can then ignore β and consider only the persistence

$$P_\alpha/l_0 \cong \frac{1}{1-p\alpha-q\gamma} = \frac{1}{1-\alpha-q(\gamma-\alpha)}, \qquad (13)$$

or inverting

$$l_0/P_\alpha = (1-\alpha) - q(\gamma-\alpha)\,. \qquad (14)$$

$1-\alpha$ is the contribution produced by the normal smooth bending of the chain. The q term is a measure of the contribution of the kink. For DNA l_0/P_∞ is of the order of 0.005. If the free energy of a kink is of the order of 5 kcal, then q is about 0.0002 at room temperature and kinks will make a small contribution to P_∞.

If by contrast we make the drastic supposition that there is virtually no smooth bending of DNA, i.e., $\alpha = 1$, then the flexibility only comes from kinks (or other spontaneous deformations) and l_0/P_∞ is given by $q(1-\gamma)$. Thus if the kinks are at 90°, they are required in one of 200 base pairs, i.e., one per persistence length.

In this far fetched model the chain would consist of perfectly straight segments of average length P_∞, with a complete loss of persistence (memory of chain direction) at each kink.

Another model of a kink [6] has been proposed in which $\gamma = 0.77$ ($\theta_\gamma = 40°$). If this type of kink was the only source of flexibility then $q \cong 0.02$, i.e., 2% of all links would be kinked implying a free energy of about 2500 kcal.

Another type of chain deformation is associated with the mechanism of hydrogen exchange, which occurs rapidly in DNA. This is not possible without a structural change and the kinetic data indicates that there is an equilibrium between ordinary helical DNA and an exchangeable state (fig. 1A). This type of deformation evidently does not unstack bases [4] through a recent investigation [7] does not rule out the possibility that uridine (thymine in DNA) is partially unstacked. The surprising thing about this deformation is its high frequency. It has been proposed that it occurs with a probability of about 0.05 in RNA's and the same frequency has been suggested for DNA. If a deformation of this frequency is responsible for flexibility then it would require a γ of 0.9 or an angle in the neighborhood of 25°. If these deformations occur as breathing modes exposing k base pairs, then $q = 0.05/k$ and a very large angle of bend would be required per deformation.

There are other natural deformations. Short sections of A-DNA could form as a fluctuating structure or short runs of AT or of GC base pairs could form helical forms which differ from the B-form [12,13]. leading to a change in helix direction.

The possibilities are great and the fact is that we do not know to what extend deformations contribute to the flexibility. There is no reason to suspect that DNA is extremely rigid so smooth bending is probably the principle mechanism. This is supported by generations of EM photographs of DNA which do not show straight line segments. More intensive studies will be required to clarify the situation. For example, it would be interesting to see if there is a correlation between the presence of the open state for hydrogen exchange of DNA and its persistence length when temperature or solvent conditions are changed.

As far as chain statistics are concerned, the origins of the flexibility are not of great importance. If the ql_β factor can be ignored, then eq. (13) is valid. The factor $p\alpha + q\gamma$ is, after all, equal to $\langle\cos\theta\rangle$ where the two terms represent the (mathematically) arbitrary separation into small and large angles for physical reasons. On the other hand, hydrodynamic methods which we use to measure the persistence length may depend on the relative contributions of $p\alpha$ and $q\gamma$ since a system consisting of straight links with large angles between will behave differently from a worm-like coil.

6.2. *Intercalation*

This is a special and interesting case. Since intercalating agents are usually planar π systems, they have the thickness of a base pair so that l_β is approximately l_0 and $\beta = \langle\cos\theta_\beta\rangle$. Also $k = 0$, so that $p = 1 - r, q = r$. In most cases the structures are presumed to have a 2-fold symmetry axis (actinomycin is an exception [6]. For actinomycin, l_β will span the entire complex so that $k \neq 0$). If the intercalation has 2-fold symmetry, then the average projection of the α link on l_β equals the average projection of γ link on l_β with the result that $\gamma = \beta^2$. Thus, assuming a very long chain,

$$\langle h_d^2\rangle_0 = 2N_t l_0 P_\alpha [1 + r(\beta - (\bar{\beta}^2 - 1)/2], \qquad (15)$$

with

$$P_\alpha/l_0 = (1 + r\beta)/(1 - \alpha - r(\beta^2 - \alpha)) . \qquad (16)$$

Thus there is only one parameter, β, for the isotherm. The bracket in eq. (15) essentially follows the increase in length of the duplex caused by the intercalation while eq. (16) accounts for the changes in persistence which result because intercalating groups are encountered in the sums. In fact, if the intercalating group acted exactly like a base pair as far as chain propagation is concerned, then $\beta = \alpha$ and eq. (16) reduces to $1/(1 - \alpha)$ which is the result for unliganded DNA.

The effect of intercalating agents has been discussed previously by a number of workers in terms of chain lenthening and the increased stiffness (or limpness) caused by the intercalating agent, but never in a quantitative way.

Sobell and coworkers [6] have obtained crystal structures of several intercalating systems and from these it is possible to obtain estimates for the parameters $\bar{\beta}$ and γ by model building procedures. Sobell estimated, for example, that $\gamma = 8°$ for the intercalation of ethidium$^+$. This indicates that this compound would produce very

small changes in P_α. Using $\alpha = 0.99437$ (appropriate for $P_\infty = 600$ Å), $\gamma = \cos 8^\circ$, estimating $\bar{\beta} = \sqrt{\cos 8^\circ}$ and using eq. (16), yields the result that P_α is relatively insensitive to r in the region of $r = 0$–0.2, increasing only by a few percent. This is because $\bar{\beta}$ has a value that is quite close to the normal propagation parameter l. It would be of considerable interest to obtain values of the parameters by using eqs. (15) and (16) on sets of experimental data, and compare them with the parameters from the crystal structure determinations. They need not be the same. Recall that DNA flexibility arises in general from deviations from the equilibrium structure.

One problem might arise experimentally. If excluded volume exists in DNA in 0.2M salt solution, it is presumably largely coulombic in origin. The intercalation of positive ions would reduce the excluded volume and cause an additional perturbation on $\langle h^2 \rangle$.

6.3. External ligands

Substances, such as proteins, which bind to DNA are often quite large. Many of the known ligands will tend to fix, or at least affect, the conformation of stretches of DNA though this is not necessarily so. The number of base pairs, k, that are affected can vary from a few to more than a hundred in the case of nucleosomes. All base pairs which do not behave with normal flexibility statistics are part of the complex spanned by I_β. The structural parameters are β, γ and k which are normally specified in model structures. Studies of $\langle h_d^2 \rangle_0$ versus r may serve to verify or eliminate model structures though it is doubtful that data will be good enough to determine all three parameters. There is, in fact, little data at present on this type of system.

If the ligand is large, the excluded volume will increase with increasing r, eventually invalidating the formalism. Therefore, for large ligands, the theory will only be good at low coverage with the magnitude of "low" bearing an inverse relation to the size of the ligand. For example, nucleosomes fit the model of fig. 3 so that eq. (12) is suitable for describing the behavior of nucleosome-DNA systems with randomly distributed, low coverage. On the other hand, the equation is probably inaccurate for chromatin because of the large excluded volume.

7. Conclusion

Formulas have been derived for the chain statistics of DNA which contain normal smooth flexing regions as well as deformations. These deformations may be spontaneous fluctuations from the normal DNA structure or may be induced by ligands. There are restrictive assumptions which have been made in the derivation of these formulas and we restate them here to avoid misinterpretation or misapplication.

1) The statistical calculations are all for "unperturbed" chains with no excluded volume. If excluded volume is a contributing factor in experiments, it will have to be superposed on the unperturbed statistics in terms of expansion parameters such as Flory's α or Peterlin's ϵ.

2) There is no sequence specificity. All binding sites on the DNA are considered to be equivalent.

3) Independent statistics have been employed so that all cooperative effects are excluded. To a limited extent certain cooperative effects can be artificially introduced. For example, the excluded site model for intercalation (anti-cooperativity) can be introduced by defining the complex to contain both a base pair ($k = 1$) and the intercalating base. On the other hand, very interesting effects result when cooperative interactions tend to form clusters of ligands on the DNA. The formulas have no applications to these situations. In fact, the model was set up in such a way (fig. 3) that contiguous ligands, cooperative or not, are excluded from consideration. Inclusion of cooperativity and the new structural features which arise when ligands are in contact is under investigation at the present time.

Acknowledgement

The author is grateful for the generous hospitality of the Laboratory for Structural Studies, MRC, Cambridge, England and the Department of Chemical Physics, Weizmann Institute, Rehovot, Israel during the course of this investigation.

References

[1] J. Schellman, J. Biophys. Chem. 11 (1980) 321.
[2] O. Kratky and G. Porod, Rec. Trav. Chim. 68 (1949) 1106.

[3] L. Landau and E. Lifschitz, in: Statistical physics, (Addison Wesley, Reading, Mass., 1958) sec. 148, 149.
[4] B. McConnell and P. von Hippel, J. Mol. Biol. 50 (1970) 297.
[5] F. Crick and A. Klug, Nature 255 (1975) 530.
[6] H. Sobell, in: Biological regulation and development, ed. R. Goldberger (Plenum Publishing Corp., New York, New York, 1978).
[7] C. Mandal, N. Kallenback and S.W. Englander, in press, 1979.
[8] J. Finch, D. Rhodes, R. Brown, B. Rushton, M. Levitt and A. Klug, Nature 269 (1977) 29.
[9] J. Schellman, Biopolymers 13 (1974) 217.
[10] H. Yamakawa and M. Fujii, Macromolecules 6 (1973) 407.
[11] H. Yamakawa and M. Fujii, Macromolecules 7 (1974) 128.
[12] S. Arnott and E. Selsing, J. Mol. Biol. 88 (1974) 509.
[13] S. Arnott, R. Chandrasekharan, D. Hukins, P. Smith and L. Watts, J. Mol. Biol. 88 (1974) 523.