A Monte-Carlo Estimate of DNA Loop Formation

CAROL BETH POST,* *Department of Chemistry, University of California, San Diego, La Jolla, California 92093*

Synopsis

The ends of rather short double-helical DNA segments (approximately two persistence lengths) can be enzymatically joined to form closed circles. Such covalent closure into circles is a measure of the likelihood of the two ends of the DNA coming into close contact. There is a length of DNA for which loop formation is most likely to occur. We have determined the chain-length dependence of loop formation for stiff chains using computer-generated chains of cylinders. The distribution from which the values for the angles between cylinders were chosen relates the chain parameters to a given chain persistence. Our results are compared with those of other theories, including a statistical wormlike chain model, and with the experimental measurements for ring closure of DNA restriction fragments.

The likelihood of the two ends of a segment of double-helical DNA meeting to form a loop depends on the length of the segment; segments less than a minimum size are unable to bend sufficiently due to the chain stiffness, while loops of much longer lengths rarely occur, since the large solution volume available to the segment ends greatly reduces the chances of the ends coming in contact. The balance of these two dependences results in a maximum in the probability of loop formation as a function of DNA length.

We have studied loop formation of DNA using a Monte Carlo procedure to generate, in free space, chains of cylinders with restricted bending. Finding a cylinder that contacted a second cylinder in the chain determined the occurrence of a loop, and the number of intervening cylinders was used to calculate the loop size. This work followed directly from our earlier study,[1] which examined excluded-volume effects on the chain dimensions of intermediate-sized DNA; the same methods for chain generation and for detection of cylinder overlap were used in this work. As discussed in Ref. 1, the excluded-volume effects predicted by the Monte Carlo analysis are very similar to those determined by a procedure[2] based on the polymer theory of Yamakawa and Stockmayer[3] for wormlike chains. Loop formation in computer-generated DNA chains is also compared in this communication with theoretical predictions[3,4] and with the experimental measurements of Shore et al.[5] for the covalent ring closure of short DNA restriction fragments.

The DNA model for the Monte Carlo procedure was a chain of cylinders linked with free rotation and with bend angles, θ, which follow a Gaussian

* Present address: Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138.

Biopolymers, Vol. 23, 601–605 (1984)
 CCC 0006-3525/84/030601-05$04.00

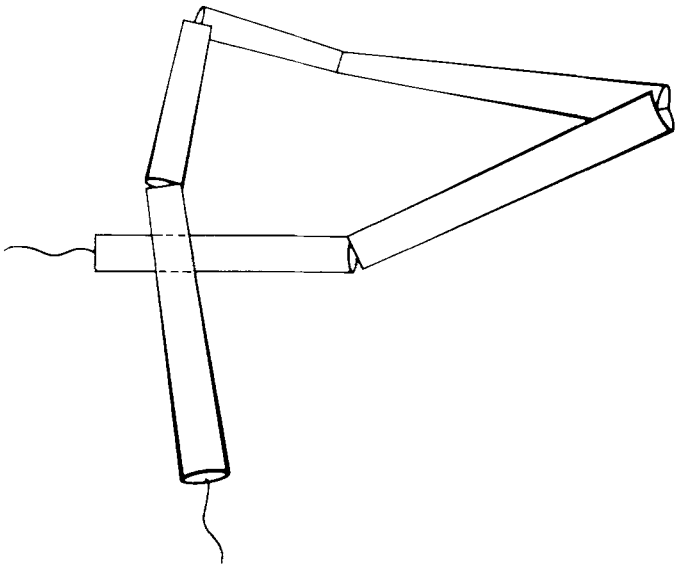

Fig. 1. Loop formed in a Monte Carlo-generated chain. Only part of the chain of 75 cylinders is shown. The loop size is $n = 5$ cylinders, or contour length $L = 148.5$ nm, and corresponds to the size found most frequently (see Fig. 2). The chain was generated using $l/p = 1 - \langle \cos\theta \rangle = 0.5$. The cylinder dimensions are $l = 29.7$ nm and $d = 4.4$ nm. Using these values for L, $\langle \cos\theta \rangle$ and l in Eq. (2) give a wormlike persistence length $a = 29.7$ nm.

distribution. Values from 0 to 2π for the rotation angle between two cylinders were randomly generated. The values for θ from 0 to π were generated according to the distribution $\sin(\theta) \exp(-q\theta^2)$, where q is related to the ratio of the cylinder length, l, to the persistence of the chain, p. The chain flexibility depended on the degree of bending, which was set by l/p and q, with the value chosen for l/p being 0.5. A chain consisted of 75 cylinders 29.7 nm long with an effective exclusion diameter,[10] d, of 4.4 nm. These chain parameters were chosen so that the average radius of gyration from the subset of chains for which no contacts occurred was approximately equal to the radius of gyration measured by light scattering[6] for *Eco*R1 DNA in 0.2*M* Na^+. Each cylinder of a given chain was tested for overlap with previously generated cylinders. Upon overlap, the length of the loop formed was set equal to the number of cylinders separating the overlapping cylinders plus one, assuming that the average position of contact was midway along a cylinder. Figure 1 is an illustration of a loop of size n equal to five cylinders. The details of the chain generation and the method for testing for overlapping cylinders have been reported earlier.[1]

The frequency of loop formation as a function of length was obtained from a set of 1200 chains. The probability of a loop of n cylinders was calculated from the frequency of occurrence divided by 1200 $(75 - n + 1)$, the factor $75 - n + 1$ being needed to allow for the number of possible ways to form loops of size n.

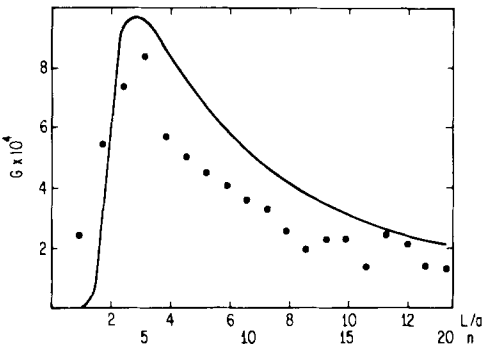

Fig. 2. Probability of loop formation determined by (●) Monte Carlo-generated chains and by (solid line) Eq. (1) from the theory of Yamakawa and Stockmayer (Ref. 3). The number of cylinders forming a loop is n. $L = n \times 29.7$ nm and a is given in Table I (the determination of a is described in the text).

Examination of the occurrence of loop sizes plotted in Fig. 2 (●) finds that loops of 4 to 6 cylinders, corresponding to 119- and 178-nm lengths, form most frequently. There is a decline in cyclization with fewer cylinder lengths, while cyclization of longer lengths is a more slowly decreasing function.

In order to compare the Monte Carlo loop-formation frequencies with the distribution derived from the statistical wormlike chain model,[3] one must consider the size of the cylinders and recognize that two cylinders come in contact over a large volume. The segments of a finite volume chain do not meet at a point in space, as is the case for the statistical chain, but rather exclude a certain volume from each other. In view of this difference in the two methods, the volume element multiplying the ring-closure probability density is assigned the volume in which the overlap of two cylinders occurs, that is, a rigid-rod excluded volume.[7] We use an expression obtained by Isihara,[7] Eq. (1b), for the excluded volume of short rods (for the Monte Carlo cylinder, $l/d \approx 7$). Thus, the following expression for the probability of ring closure occurring within a volume equal to the rigid-rod excluded volume is obtained from Eq. (62) of Ref. 3:

$$G(t) = \frac{3}{8\pi t a^2}\sqrt[3]{2}\left(1 - \frac{5}{8t} - \frac{79}{640t^2}\right)dV \tag{1a}$$

$$dV = \frac{\pi d l^2}{2}\left[\frac{d}{l} + \left(1 + \frac{d}{2l}\right)\left(1 + \frac{\pi d}{2l}\right)\right] \tag{1b}$$

where a is the persistence length of a wormlike chain and t equals $L/2a$, one-half the number of persistence lengths in contour length, L. Equation (1) is given in terms of real lengths, not reduced length; it includes the factor $(2a)^{-2}$, which does not appear in Eq. (62) of Ref. 3.

The Monte Carlo probability for the formation of a loop of size n cylinders was compared to the probability given by Eq. (1) for the ring closure

TABLE 1
Corresponding Values of Persistence Length, a, for a Loop Formed from n Cylinders with Contour Length L

n	L (nm)	a (nm)	n	L (nm)	a (nm)
2	59	63.8	11	327	45.5
3	89	52.5	12	356	45.3
4	119	49.1	13	386	45.3
5	149	47.7	14	416	45.2
6	178	46.9	15	445	45.2
7	208	46.3	16	475	45.1
8	238	46.0	17	505	45.1
9	267	45.8	18	535	45.0
10	297	45.6	19	564	45.0
			20	594	45.0

of size $L = n \times l$. We relate the parameter t of the theoretical ring-closure probability to the computer model by equating the expression for the end-to-end length of a wormlike chain[8] to that for a chain with restricted bend angle[9]:

$$2aL - 2a^2\left[1 - \exp\left(\frac{L}{a}\right)\right] = l^2\left[\frac{n - \langle\cos\theta\rangle(2 + n\langle\cos\theta\rangle - 2\langle\cos\theta\rangle^n)}{\langle 1 - \cos\theta\rangle^2}\right] \tag{2}$$

With $\langle\cos\theta\rangle = (1 - l/p) = 0.5$, as determined by the distribution of θ used in generating the chains, Eq. (2) can be solved iteratively for a to find the persistence length corresponding to the Monte-Carlo chains. The resulting values for a are shown in Table I.

Like the good agreement reported earlier[1] in estimating excluded-volume effects, the Monte Carlo-determined probability of loop formation is consistent with the theoretical probability for ring closure of Yamakawa and Stockmayer.[3] The Monte Carlo frequencies are compared in Fig. 2 with the distribution function given by Eq. (1) (solid line). The two methods agree well, particularly considering the contrast between the discrete bends separating adjacent cylinders in the Monte Carlo chain and the continuously curving wormlike chain (perhaps a better model for DNA) used in the statistical theory. Both the Monte Carlo results and Eq. (1) have a maximum at L/a approximately equal to 3.0 and show a similar dependence on loop size. Overall, the Monte Carlo method appears to give a somewhat lower frequency for loop formation than is predicted by Eq. (1). The direction in which the Monte Carlo-predicted ring-closure probabilities differ from the wormlike-chain probability function is opposite to that found for the small differences in estimates of the effects of excluded volume on molecular dimensions; our Monte Carlo chains showed slightly larger excluded-volume effects than were found using the Yamakawa and Stockmayer theory. This inconsistency suggests that Eq. (1b) may not be an

accurate estimate for the differential volume element, which, in effect, "scales" the analytical probability to the Monte Carlo frequencies.

Olson has also examined the probability of cyclization as a function of DNA chain length.[4] Her study of flexibility, in which the DNA was modeled as a chain of virtual bonds linking the backbone phosphates, involved chains generated both with and without angular correlations. Since angular correlations were not taken into account in the work presented here, we compare Olson's results for which angular correlations were not considered. We find a peak in the distribution of circle sizes occurs at 256 base pairs. This gives L/a approximately equal to 1.6 (a = 55 nm, rise per base pair = 0.34 nm). Thus, this computer model, with nearly the same persistence length but a shorter step length than the cylinder-chain model, predicts a maximum at a DNA length of about half that found by the cylinder model or by Eq. (1) for the wormlike chain.

Shore et al.[5] have measured the formation of covalently closed rings of DNA by the enzymatic joining of the ends of restriction fragments. They found the above analyses to be consistent with their experimental results (see Fig. 5 of Ref. 5), in that the frequency of ring closure from fragments between 1.6 and 30 persistence lengths did not vary by more than a factor of 10. The apparent maximum in the experimental measurements at 4.5 persistence lengths, however, is somewhat longer than seen for the Monte Carlo or wormlike model. This discrepancy may result because neither the computer analysis reported here nor the statistical theory[3] consider the dependences on the angular and rotational orientation of the segment ends, which must be a factor in the experimental measurements.

The author is particularly grateful to Prof. Bruno H. Zimm for many helpful discussions and comments on this work and for bringing to the author's attention the excluded-volume expression for short chains. This work was supported by Public Health Service Grant GM-11916.

References

1. Post, C. B. (1983) *Biopolymers* **22,** 1087–1096.
2. Manning, G. S. (1981) *Biopolymers*, **20,** 1751–1755.
3. Yamakawa, H. & Stockmayer, W. H. (1972) *J. Chem. Phys.* **57,** 2843–2854.
4. Olson, W. K. (1979) *Stereodynamics of Molecular Systems*, Sarma, R. H., Ed., Pergamon Press, New York, pp. 297–314.
5. Shore, D., Langowski, J. & Baldwin, R. L. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 4833–4837.
6. Borochov, N., Eisenberg, H. & Kam, Z. (1981) *Biopolymers* **20,** 231–235.
7. Ishihara, A. (1950) *J. Chem. Phys.* **18,** 1446–1449.
8. Kratky, O. & Porod, G. (1949) *Rec. Trav. Chim.* **68,** 1106–1111.
9. Yamakawa, H. (1971) *Modern Theory of Polymer Solutions*, Harper & Row, New York.
10. Stigter, D. (1977) *Biopolymers* **16,** 1435–1448.

Received April 8, 1983
Accepted July 29, 1983