The randomly broken chain as a semiflexible macromolecular model. Computer simulation of statistical properties

Jesús J. García Molina  
Escuela Universitaria de Informática, Universidad de Murcia, Espinardo, Murcia, Spain  
José García de la Torre  
Departamento de Química Física, Facultad de Ciencias Químicas y Matemáticas, Universidad de Murcia, Murcia, Spain

(Received 4 November 1985; accepted 19 December 1985)

At any repeating unit of the randomly broken chain model, the direction of the chain can either follow that of the preceding bond, or change at random. The contour-length dependence of the mean-squared end-to-end distance and radius of gyration, when expressed in terms of a persistence length, is found to be the same as that for worm-like chains. However, other moments of the end-to-end distance have a clearly distinct length dependence, as found by computer simulation. Due to the finite probability of a fully extended, straight conformation, the distribution function of the end-to-end distance has a peculiar sharp peak. The similarities and differences between this model and the worm-like chains are discussed.

INTRODUCTION

The solution properties of stiff macromolecules, whose behavior is intermediate between a rigid rod and a Gaussian coil, are usually interpreted in terms of the worm-like chain (WC) model. In this model, the flexibility of the macromolecule is assumed to be like that of an elastic wire, so that in an instantaneous conformation the macromolecule bends smoothly along all its contour length.

One can envision a quite different mechanism of flexibility. Due to various reasons (like thermal fluctuations in internal degrees of freedom, or binding of ligands), there may be spontaneous, transient discontinuities in a chain that otherwise would be essentially rigid. These discontinuities, that should be very localized (affecting only to one or a few repeating units), could act as flexible joints, or the chain could just change its direction at them.

This idea is applicable to a variety of situations, like the stiff chain limit of rotational-isomeric chains, correlated walks on lattices, and liquid crystalline ordering in solutions of semiflexible chains. There are also interesting potential applications in the field of biological macromolecules. Thus, although the flexibility of DNA is usually and successfully interpreted in terms of the WC model, in a series of papers Mannings has put forward a different mechanism of DNA flexibility based on the possibility of transient opening of base pairs. There are also interesting potential applications in the field of biological macromolecules. Thus, although the flexibility of DNA is usually and successfully interpreted in terms of the WC model, in a series of papers Mannings has put forward a different mechanism of DNA flexibility based on the possibility of transient opening of base pairs. Such openings would give rise to bending fluctuations, or kinks, and the DNA chain would present transient discontinuities that could act as flexible joints. In an instantaneous conformation, the axis of the double helix would follow a straight line, changing sharply its direction at the discontinuities. Then the conformation would look like a line broken at random sites (see Fig. 1 in Ref. 8). Even if this picture of DNA flexibility were not correct, local sharp bendings could take place in DNA, for instance, upon binding of certain ligands.

In this paper, we present a semiflexible macromolecular model that we call the randomly broken chain, in which every repeating unit can either be in a rigid conformation, or act as a universal joint. This model is reminiscent of the broken chain model of Stockmayer and co-workers, in which the joint elements are uniformly distributed and their number is predetermined.

An important question is to what extent the randomly broken chain differs in practice from the classical WC. In the present paper we examine this question from the point of view of the statistics and molecular weight (or length) dependence of chain dimensions. The analysis of dynamic properties is left for further work.

THEORY

The randomly broken chain (RBC) is a string of \( N + 1 \) elements connected by \( N \) rigid bonds. Any element can present two states, which are called "rigid" and "flexible." In the rigid state, the chain advances in the same direction as the previous bond. In other words, if element \( k \) is rigid, then bond \( k \) is collinear with bond \( k - 1 \). If the element is flexible, it acts as a universal joint and the chain takes a new, random direction. The probability that any element is in the rigid state is \( p < 1 \), and that of the flexible state is \( 1 - p \).

Obviously, the state of the terminal elements (1 and \( N + 1 \)) has not to be specified since they do not affect the conformation of the chain. In an instantaneous conformation, the chain appears as a broken line having \( \nu \) straight segments, \( \nu - 1 \) being the number of flexible elements other than the terminal ones. An example is shown in Fig. 1.

It is evident that the RBC reproduces the two limiting cases of chain flexibility. If \( p = 1 \), all the elements will be rigid and the chain reduces to a rigid rod (\( \nu = 1 \)). If \( p = 0 \), all the elements are flexible and the RBC behaves as a freely jointed chain with \( \nu = N \) bonds.

The mean squared end-to-end distance of the RBC can be obtained analytically. We start from

\[
\langle r^2 \rangle = \sum_{j=1}^{N+1} \langle r_j^2 \rangle = \sum_{j=1}^{N} \left( \frac{R_g^2}{j} \right) + R_g^2
\]

\[
\langle r_j^2 \rangle = \begin{cases} \frac{R_g^2}{j} & \text{if } j \leq \nu \\ R_g^2 & \text{if } j > \nu \end{cases}
\]

\[
R_g = \frac{1}{2\sqrt{3}} \langle r \rangle
\]

To whom correspondence should be addressed.

\[\text{Received } 4 \text{ November } 1985; \text{ accepted } 19 \text{ December } 1985\]
Substitution of Eq. (3) into Eq. (1) leads to

\[
\langle r^2 \rangle = N b^2 \left[ \frac{1 + p}{1 - p} - \frac{2p}{N} \frac{1 - p^N}{(1 - p)^2} \right].
\]  

This result is the same form as that for freely rotating chains, with \( p \) playing the role of the cosine of the fixed bond angle of those chains.

As for any other flexible model, one can formulate the persistence length, \( a \). The expression is also equivalent to that for freely rotating chains, i.e.,

\[
a = b / (1 - p)
\]  

and

\[
L / a = N(1 - p),
\]  

where \( L = N b \) is the contour length of the chain.

Although the dependence of \( \langle r^2 \rangle \) on \( N \) has the same form as for freely rotating chains, it must be noted that the RBC is a completely distinct model. A specific quantity of the RBC is the number of straight segments, \( v \). It can be easily shown that the probability that, for a given value of \( p \), the number of segments is \( v \) in a chain of \( N \) elements, denoted \( f(v; N, p) \), is a binomial distribution:

\[
f(v; N, p) = \binom{N}{v} p^v (1 - p)^{N - v}
\]  

and the average value of \( v \) is given by

\[
\langle v \rangle = (N - 1) (1 - p) + 1.
\]  

In the applications of the RBC to macromolecular chains, \( N \) should be high, and one can therefore use limiting expressions for \( N \to \infty \). To obtain such expressions, the total length and the flexibility parameter of the chain must be kept fixed. Thus, in the limit \( N \to \infty \) with the restriction \( Nb = \) constant and \( N(1 - p) = \) constant, we obtain

\[
f(v; L / a) = e^{-L/a} (L / a)^{-1} \binom{v}{v - 1}
\]  

which is a Poisson distribution with

\[
\langle v \rangle = 1 + L / a
\]  

and

\[
\langle r^2 \rangle = 2aL \left[ 1 - \frac{a}{L} (1 - e^{-L/a}) \right].
\]  

Equation (11) is the same as that for the WC. Thus, the variation of \( \langle r^2 \rangle \) with contour length (or with molecular weight), expressed in terms of the persistence length as the flexibility parameter, is the same for the RBC and the WC models. As a consequence, the variations of the mean squared radius of gyration, \( \langle s^2 \rangle \),

\[
\langle s^2 \rangle = \frac{La}{3} \left[ 1 - \frac{3a}{L} - \frac{6a^2}{L^2} - \frac{6a^3}{L^3} (1 - e^{-L/a}) \right]
\]  

are also identical.

Other moments of the end-to-end distance, \( \langle r^n \rangle \) with \( n = 1, 4, 6, \) and 8 have been obtained by computer simulation. The RBC are generated step by step. Element 1 is placed at the origin of coordinates and element 1 is on the z axis. The state of the \( i \)th element is decided picking a random \( \theta_k \), \( \phi_k \), and \( \psi_k \) are chosen at random in the (0, \( \pi \)) interval. If \( \phi_k \leq \theta_k \) the element is rigid and bond \( b_i \) is in the same direction as bond \( b_{i-1} \). Otherwise, the element is flexible and the orientation of \( b_i \) is determined by the polar and azimuthal angles \( \beta \) and \( \psi \). \( \theta \) is chosen at random in the (0, \( \pi \)) interval, and \( \beta \) and \( \phi \) is generated from a cos \( \beta \) distribution in the (0, \( \pi \)) interval. Thus the distribution of the orientation of chain segments is spherically uniform.

For each conformation generated, the end-to-end distance was obtained as the distance from element \( N + 1 \) to the origin, and the squared radius of gyration was calculated from a list of element coordinates. The statistical sample was divided into four subsamples, for which the various averages were obtained. As a measure of the uncertainty of the sample averages, we have taken the standard deviations of the averages for each subsample.

**RESULTS AND DISCUSSION**

As commented above, the situation of interest is that of high \( N \) with the restriction \( N(1 - p) = \) constant. Then, we made preliminary calculations to find a value of \( N \) that, be-
ing high enough, is also reasonable for practical simulations. Table I shows results for simulations with varying \( N \), in which \( p \) has been adjusted to keep fixed \( L/a = 10 \) and \( \langle v \rangle = 11 \). We see that when \( N \) increases from 200 to 1000, the averages increase very slightly. Therefore we have taken \( N = 200 \) for subsequent calculations whose results, according to Table I, should underestimate the limiting values for \( N \to \infty \) by only a few percent.

Next, we performed simulations of RBCs with varying flexibility, for \( N = 200 \) and various values of \( p \). The results are presented in Table II. In addition to \( \langle r^2 \rangle \) and \( \langle s^2 \rangle \), the distribution functions of \( v \) and \( r \) were also computed by the simulation algorithm. The observed \( f(v;N,p) \) were in excellent agreement with the values calculated from Eq. (7) for \( N = 200 \), and these in turn were very close to the limiting results of Eq. (9). Furthermore, the observed values of \( \langle v \rangle \) and those calculated for the same \( N \) and \( p \) are in perfect agreement, as shown in Table II.

The distribution function of the end-to-end distance, \( P(r) \), plotted in Fig. 2 in the form of a frequency histogram has the expected aspect, with a maximum closer to \( r = L \) for higher \( p \)'s. However, in the region of \( L/a < 4 \) the histograms present a peculiarity: there is a spike at \( r = L \). The reason is that there is a finite probability for the \( r = L \) case; particularly, that probability is 2% for \( L/a = 4 \) from Eq. (9). Then, \( P(r) \) is discontinuous at \( r = L \), where it has a Dirac's delta contribution. For moderate or high values of \( L/a \) the discontinuity is irrelevant because its contribution is very small.

The results for the RBC presented in Table II are to be compared with those of the WC. The comparison is presented in Fig. 3. For \( \langle r^2 \rangle \) and \( \langle s^2 \rangle \) of the WC we have used the expressions of Hermans and Ullman and Heine et al., respectively, and \( \langle r^{-1} \rangle \) was calculated from the series expansions derived by Yamakawa and Fujii.

As pointed out in the preceding section, the analytical results for \( \langle r^2 \rangle \) and \( \langle s^2 \rangle \) [Eqs. (11) and (12)] are the same for the RBC as for the WC. The simulation results of \( \langle r^2 \rangle \) and \( \langle s^2 \rangle \) of the RBC coincide with the analytical ones, as they should. Thus, in regard to \( \langle r^2 \rangle \) and \( \langle s^2 \rangle \), the RBC and the WC are equivalent models. In other words, and from the point of view of the flexibility of DNA, our oversimplified version of Manning's model predicts the molecular-weight dependence of the radius of gyration of DNA as well as the continuous, uniformly bent worm-like model.

However, Fig. 3 shows that for moments of the end-to-end distance other than the second one, the results for the two models are clearly distinct. This is a consequence of the difference between the \( P(r) \) functions. Although these functions are unknown for the two models, they must be different, as evidenced by the peculiarity of \( P(r) \) for the RBC described above. Therefore, one can expect that the models may not be equivalent in regard to other solution properties. We remark particularly the discrepancy found for \( \langle r^{-1} \rangle \), which is closely related to hydrodynamic properties.

In Fig. 4 we have plotted the dimensionless quantities \( \langle r^{-1} \rangle \) \( \langle r^2 \rangle \) and \( \langle s^2 \rangle /\langle r^2 \rangle \), which are bound by values corresponding to the limits of straight rod and Gaussian coil. As the contour length increases, these quantities approach the

### Table I

Results for RBCs for several combinations of \( N \) and \( p \) with constant \( \langle v \rangle = 11 \) [Eq. (8)] and \( L/a = 10 \) [Eq. (10)], showing the convergence to the continuous-chain limit.

<table>
<thead>
<tr>
<th>( p )</th>
<th>( N )</th>
<th>( \langle 1/r \rangle L )</th>
<th>( \langle r^2 \rangle /L^2 )</th>
<th>( \langle s^2 \rangle /L^2 )</th>
<th>( \langle r^4 \rangle /L^4 )</th>
<th>( \langle r^6 \rangle /L^6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.977</td>
<td>50</td>
<td>3.66 ± 0.04</td>
<td>0.163 ± 0.001</td>
<td>0.0237 ± 0.0002</td>
<td>0.049 ± 0.001</td>
<td>0.020 ± 0.001</td>
</tr>
<tr>
<td>0.899</td>
<td>100</td>
<td>3.54 ± 0.05</td>
<td>0.172 ± 0.004</td>
<td>0.0244 ± 0.0002</td>
<td>0.053 ± 0.002</td>
<td>0.023 ± 0.001</td>
</tr>
<tr>
<td>0.950</td>
<td>200</td>
<td>3.47 ± 0.02</td>
<td>0.174 ± 0.001</td>
<td>0.0247 ± 0.0002</td>
<td>0.055 ± 0.003</td>
<td>0.024 ± 0.001</td>
</tr>
<tr>
<td>0.980</td>
<td>500</td>
<td>3.51 ± 0.02</td>
<td>0.175 ± 0.001</td>
<td>0.0246 ± 0.0001</td>
<td>0.055 ± 0.003</td>
<td>0.024 ± 0.002</td>
</tr>
<tr>
<td>0.990</td>
<td>1000</td>
<td>3.47 ± 0.04</td>
<td>0.177 ± 0.003</td>
<td>0.0250 ± 0.0002</td>
<td>0.056 ± 0.002</td>
<td>0.025 ± 0.001</td>
</tr>
</tbody>
</table>

### Table II

Results for RBCs with different flexibility (\( N = 200 \)).

<table>
<thead>
<tr>
<th>( p )</th>
<th>( L/a )</th>
<th>( \langle v \rangle )</th>
<th>( \langle 1/r \rangle L )</th>
<th>( \langle r^2 \rangle /L^2 )</th>
<th>( \langle s^2 \rangle /L^2 )</th>
<th>( \langle r^4 \rangle /L^4 )</th>
<th>( \langle r^6 \rangle /L^6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>10</td>
<td>10.95</td>
<td>3.24 ± 0.03</td>
<td>0.177 ± 0.002</td>
<td>0.0250 ± 0.0001</td>
<td>0.050 ± 0.001</td>
<td>0.019 ± 0.001</td>
</tr>
<tr>
<td>0.960</td>
<td>8</td>
<td>8.96</td>
<td>3.29 ± 0.03</td>
<td>0.218 ± 0.004</td>
<td>0.0294 ± 0.0003</td>
<td>0.074 ± 0.003</td>
<td>0.031 ± 0.002</td>
</tr>
<tr>
<td>0.970</td>
<td>6</td>
<td>6.97</td>
<td>3.19 ± 0.03</td>
<td>0.272 ± 0.003</td>
<td>0.0352 ± 0.0002</td>
<td>0.113 ± 0.003</td>
<td>0.060 ± 0.003</td>
</tr>
<tr>
<td>0.975</td>
<td>5</td>
<td>5.97</td>
<td>2.93 ± 0.02</td>
<td>0.317 ± 0.005</td>
<td>0.0394 ± 0.0004</td>
<td>0.150 ± 0.005</td>
<td>0.089 ± 0.005</td>
</tr>
<tr>
<td>0.980</td>
<td>4</td>
<td>4.98</td>
<td>2.13 ± 0.04</td>
<td>0.376 ± 0.007</td>
<td>0.0465 ± 0.0005</td>
<td>0.203 ± 0.008</td>
<td>0.133 ± 0.007</td>
</tr>
<tr>
<td>0.985</td>
<td>3</td>
<td>3.98</td>
<td>1.90 ± 0.02</td>
<td>0.457 ± 0.002</td>
<td>0.0512 ± 0.0003</td>
<td>0.287 ± 0.004</td>
<td>0.212 ± 0.004</td>
</tr>
<tr>
<td>0.990</td>
<td>2</td>
<td>2.99</td>
<td>1.65 ± 0.04</td>
<td>0.563 ± 0.013</td>
<td>0.0588 ± 0.0009</td>
<td>0.408 ± 0.014</td>
<td>0.333 ± 0.014</td>
</tr>
<tr>
<td>0.995</td>
<td>1</td>
<td>1.99</td>
<td>1.36 ± 0.03</td>
<td>0.733 ± 0.010</td>
<td>0.0695 ± 0.0006</td>
<td>0.627 ± 0.012</td>
<td>0.370 ± 0.012</td>
</tr>
<tr>
<td>0.999</td>
<td>0.2</td>
<td>1.20</td>
<td>1.07 ± 0.01</td>
<td>0.937 ± 0.003</td>
<td>0.0809 ± 0.0002</td>
<td>0.908 ± 0.003</td>
<td>0.891 ± 0.003</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.0842 ± 0.0000</td>
<td>1</td>
</tr>
</tbody>
</table>

J. Chem. Phys., Vol. 84, No. 7, 1 April 1986
Gaussian limit. We see, however, that the approach is faster for the RBC than for the WC. Figure 4 displays better than Fig. 3 the deviation between the RBC and WC values. The difference is found to be relatively more important in the region of small \( L / a \). Thus, the possible discrepancy between other solution properties of the two models can be expected to be more pronounced for chains whose contour length is only a few times the persistence length. Precisely, this is the case for the short, monodisperse DNA fragments which have been recently studied.\(^7\)

In the present paper we have compared the WC and RBC models from the point of view of the length dependence of chain dimensions. The possibility of calculating hydrodynamic properties by Monte Carlo simulation\(^{17-19}\) will allow to reach other results that can be helpful in the interpretation of experimental data.

**ACKNOWLEDGMENTS**

This work was supported by Grant 561/84 from the Comisión Asesora de Investigación Científica y Técnica to J. G. T. We are grateful to Dr. B. Hammouda for encouraging comments on the applicability of the RBC to scattering problems.

\(^6\)G. S. Mannings, in *Nucleic Acids and Proteins*, edited by E. Clementi and
13See Ref. 1, p. 35.