An iterative convolution approach to the statistics of polymer conformation

1984 J. Phys. A: Math. Gen. 17 2129

(http://iopscience.iop.org/0305-4470/17/10/025)

J. Phys. A: Math. Gen. 17 (1984) 2129–2135. Printed in Great Britain

An iterative convolution approach to the statistics of polymer conformation

Clive A Croxton

Department of Mathematics, Statistics and Computer Science, The University of Newcastle, NSW 2308, Australia

Received 26 October 1983, in final form 21 February 1984

Abstract. An iterative convolution description of polymer conformations for short to intermediate length self-avoiding sequences is reported. The intrachain spatial distribution functions and the mean square end-to-end separation for such systems are in the closest reported agreement with Monte Carlo estimates, with which they are compared. The simplicity, versatility and quantitative accuracy of the model is emphasised, and applications to problems of physical and biophysical interest are indicated.

The radius of gyration and the expansion of internal subsequences with respect to their end-to-end counterparts are also determined.

1. Introduction

Descriptions of self-interacting polymer sequences have been given on the basis of a variety of approaches ranging from exact enumerations on regular lattices, through diagrammatic and renormalisation techniques to direct machine simulations, both on and off lattices. Effort has been primarily directed at the determination of asymptotic properties of such sequences in the form of exponent representations as the number of steps $N \to \infty$, whilst relatively little attention has been given to the no less important description of short to intermediate length chains, particularly those characterised by a variety of intrachain stereochemical interactions, fixed bond angles, etc. External constraints such as boundaries, solvent effects, etc—precisely those features which characterise many systems of technological and biophysical interest, including tertiary structures in protein conformation and ligand–substrate interactions—are rarely investigated and, moreover, are not readily modelled other than by direct simulation.

There is, then, considerable motivation in seeking a description of such systems which is substantially simpler than direct machine simulation, is versatile, yet is both quantitatively and qualitatively reliable and whose approximations are physically readily accessible.

The principal quantities which characterise the conformations of the sequence are the inter-segmental spatial distributions $Z(r_{ij}|N)$ between monomers i, j within the N-mer, and their associated moments $\langle R_{ij}^2 \rangle$, $\langle S_N^2 \rangle$—the mean square intersegmental separation and radius of gyration, respectively. Since the general preoccupation in the literature is with asymptotically long sequences, little attention is paid to the distributions $Z(ij|N)$ themselves. Here, however, we are concerned more with the detailed structure of the sequence, and indeed, the estimate of limiting exponents is beyond the scope of the present investigation.

0305-4470/84/102129+07$02.25

We have previously reported some of these features based on a primitive version of the convolution approximation (Croxton 1979a, b, c). We now report an iterative approach which represents a substantial improvement upon the previous treatment, and offers a number of specific quantitative and qualitative advantages, particularly if the sequence is composed of different species. The present iterative approach brings the results into the closest reported coincidence with Monte Carlo estimates, also determined here, and this we regard as an important endorsement of the model.

2. Theory

The intrachain spatial correlation $Z(il|N)$ of segments (i, l) within an isolated sequence may be expressed in terms of the propagation of such correlations through all possible indirect routes between the interacting pair of segments. Thus the diagrams represent the propagation of correlation between segments (i, l) through one (k), two (j, k), . . . , intrachain segments. The straight bonds represent the Z-functions whilst the wiggly bonds represent the direct segment–segment interaction, $H(il) = \exp(-\Phi(il)/kT)$ where $\Phi(il)$ is the intersegmental potential.

Unlike a bulk, homogeneous fluid, the mediating particles k, j etc are *not* equivalent, but instead depend sensitively upon their location within the sequence. Accordingly, we propose that the (i, l) correlation develops in the mean field of these indirect correlations. Thus, denoting by M the formation of this mean over the allowable convolution products, as specified by the indices, we may write

$$Z(il|N) = H(il) \underset{\substack{k \neq i,l \\ =1}}{\overset{N}{\mathrm{M}}} \int Z(ik|N)Z(kl|N)\,\mathrm{d}\boldsymbol{k} + H(il)H(ik)H(jl) \underset{\substack{k \neq l,j \\ =1}}{\overset{N}{\mathrm{M}}} \underset{\substack{j \neq k,l \\ =1}}{\overset{N}{\mathrm{M}}} \int\int Z(ij|N)Z(jk|N)Z(kl|N)\,\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{j} + \ldots . \tag{1}$$

Equation (1) essentially represents the defining relation for the intrachain correlation functions. It is an approximation in that a mean field representation of the remainder of the chain upon the interacting pair (i, l) is introduced. For reasons we shall discuss later, we choose to adopt the *geometric* mean.

The correlation between segments (j, l) may be formed by analogy with (1) as

$$Z(jl|N) = H(jl) \prod_{\substack{k \neq j,l \\ =1}}^{N} \int Z(kl|N)Z(kj|N)\,\mathrm{d}\boldsymbol{k} + H(jl)H(jk)H(ml) \prod_{\substack{k \neq m,l \\ =1}}^{N} \prod_{\substack{m \neq k,l \\ =1}}^{N} \int\int Z(jm|N)Z(mk|N)Z(kl|N)\,\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{m} + \ldots \tag{2}$$

where Π_k denotes the geometric mean of the k convolution integrals. Multiplying (2) by $H(il)Z(ij|N)$, integrating $\int \mathrm{d}\boldsymbol{j}$ and forming the geometric mean Π_j gives

$$H(il)\prod_j \int Z(jl|N)Z(ij|N)\,\mathrm{d}\boldsymbol{j} = H(il)H(jl)\prod_j\prod_k \int\int Z(ij|N)Z(jk|N)Z(kl|N)\,\mathrm{d}\boldsymbol{k}\,\mathrm{d}\boldsymbol{j}. \tag{3}$$

The RHS of (3) may be closely identified with the second and subsequent terms of (1), whereupon we may write

$$Z(il|N) \sim 2H(il)\prod_k \int Z(ik|N)Z(kl|N)\,\mathrm{d}\boldsymbol{k}. \tag{4}$$

The physical nature of the approximation in (4) is readily understood in terms of the neglect of H-bonds. We may designate the *order* of the route of propagation as the number of field points n or centres of integration involved in the connection of (i, l). The number of H-bonds involved in a route of order n may easily be shown to be $\Sigma_{i=0}^{n}(n-i)$; approximation is incurred in neglecting $(n-1)$ H-bonds at the nth order of propagation. We therefore anticipate that self-avoiding distributions will be collapsed with respect to their simulated counterpart. However, we also note that the range of the kth convolution product arising in the geometric mean, e.g.

$$\int Z(ik|N)Z(kj|N)\,\mathrm{d}\boldsymbol{k} \tag{5}$$

will be $|i-k|+|k-j|$, which may substantially exceed the range of the internal section $|i-j|$. Nevertheless, the correct range of the geometric mean is ensured by the presence of terms of the form

$$\int \delta_{i,i+1}Z(i+1, j|N)\,\mathrm{d}(i\pm 1), \qquad \int Z(i, j-1|N)\delta_{j-1,j}\,\mathrm{d}(j\pm 1) \tag{6}$$

where δ-bonds represent sequential connection of adjacent segments.

The long-range distributions (5) result in internal distributions shifted outwards towards larger separations. Whilst inspection of Monte Carlo simulations of internal distributions also show such an outward shift, it appears that the present approximation overestimates this aspect of the distribution, particularly at larger N when very long range convolutions (5) will arise. In consequence, we anticipate an overestimate of the second moment $\langle R_N^2\rangle$ with increasing N.

The formation of the geometric rather than the arithmetic mean of the convolution products has a simple physical interpretation as follows. If we express the component convolutions of Π_k in terms of a potential of mean force:

$$\exp\left(\frac{-\Psi_k(il)}{kT}\right) = \int Z(ik|N)Z(kl|N)\,\mathrm{d}\boldsymbol{k} \tag{7}$$

then formation of the geometric mean implies

$$Z(il|N) \sim \exp\frac{-\Phi(il)}{kT}\exp\left(\sum_k \frac{-\Psi_k(il)}{kT}\right)$$

where Σ_k denotes the *arithmetic* mean of the potentials of mean force developed within the sequence: an intuitively appealing result. No such simple interpretation appears possible on the basis of an arithmetic mean of the convolution products. Moreover,

the geometric mean ensures that $Z(il|N)$ goes smoothly to zero at large r_{il}, as it should, whilst for the arithmetic mean this will not generally be the case: indeed, the range of $Z(il|N)$ will substantially exceed $|i-l|$.

Now, we may further write, to the present degree of approximation,

$$\begin{aligned} Z(ik|N) &= 2H(ik)\prod_l \int Z(il|N)Z(lk|N)\,\mathrm{d}\boldsymbol{l} \\ Z(kl|N) &= 2H(kl)\prod_i \int Z(ki|N)Z(il|N)\,\mathrm{d}\boldsymbol{i}\ . \end{aligned} \tag{8}$$

Inserting (8) in (4) we obtain

$$Z(il|N) = aH(il)\prod_k \int H(ik)\prod_l \int Z(il|N)Z(lk|N)\,\mathrm{d}\boldsymbol{l}\,H(kl)\prod_i \int Z(ki|N)Z(il|N)\,\mathrm{d}\boldsymbol{i}\,\mathrm{d}\boldsymbol{k} \tag{9}$$

where the numerical factor on the RHS of (9) has been replaced by the constant a: the distribution $Z(il|N)$ has to be subsequently normalised, and the value of a is arbitrary for present purposes.

Equation (9) is, of course, nothing other than a once-iterated form of equation (4); the equation could be further refined by repeated insertions of the form (8) for the convolution integrals. However, our objective is simplicity, and our primary intention of casting equation (1) into iterative form has been achieved. For, subject to initial guesses for the various internal distributions, equation (9) may be solved iteratively. Either rectangular functions of the appropriate range, or previously determined distributions from the $(N-1)$th sequence may be used. Convergence is somewhat more rapid in the latter case, taking only four or five iterations whilst the choice of rectangular functions tends to yield convergence after about eight or ten iterations. In all cases the converged distribution is essentially independent of the initial choice, and fast Fourier transform techniques may be used in their evaluation.

Finally, we should point out that more realistic systems embodying steric effects, fixed bond angles, variable bond lengths, etc are readily incorporated in this model, simply by specifying the various $\Phi(i, l)$. For example, to fix the bond angle between segments $i-1, i, i+1$, one needs only to specify the distance $R_{i-1,i+1}$ and introduce the interaction

$$\begin{aligned} \Phi(i-1, i+1) &= +\infty \qquad r_{i-1,i+1} \neq R_{i-1,i+1} \\ &= 0 \qquad\qquad = R_{i-1,i+1} \end{aligned} \tag{10a}$$

whilst to introduce a harmonic interaction between sequential segments, the $\delta_{i,i+1}$ bond introduced earlier needs only to be replaced by

$$\Phi(i, i+1) = k_{i,i+1}(r_{i,i+1} - r^0_{i,i+1})^2 \tag{10b}$$

where $r^0_{i,i+1}$ is the equilibrium separation of sequential segments $i, i+1$, and $k_{i,i+1}$ is the appropriate bond constant.

Here, however, we restrict ourselves to the investigation of the model for the simplest of systems, a perfectly flexible hard sphere sequence. Unless the analysis can adequately describe such a system, discussion of more exotic systems is obviously unwarranted.

3. Results

Throughout, comparison of the calculated quantities determined on the basis of the distributions (9) is made with Monte Carlo simulations of the corresponding hard sphere self-avoiding sequences.

Briefly, the Monte Carlo technique adopted in these investigations consisted in the sequential addition of hard sphere segments of unit diameter, uniformly distributed about the centre of the preceding sphere. If the generated sequence embodied no violations of the excluded volume condition (no geometric overlap of segments), the chain conformation was accepted. If the excluded volume condition was violated, the entire sequence was rejected, and the sequential development of a self-avoiding system started anew. In the specification of an N-mer it is *not* correct to salvage an initial trial sequence embodying no violations and to repetitively attempt trial extensions until successful, since this implies independence of the initial sequence upon subsequent extensions.

In figure 1 we compare the second moment of the distribution

$$\langle R_{1N}^2 \rangle = 4\pi \int_0^\infty Z(1N|N) r_{1N}^4 \, \mathrm{d}r_{1N} \tag{11}$$

(random walk $\langle R_{1N}^2 \rangle = N - 1$)

determined on the basis of the iterative convolution approach with the Monte Carlo result (10^6 successful configurations per chain). Also shown are estimates based on the arithmetic (rather than geometric) mean, Curro's estimate based on the Percus–Yevick approximation, and the previously reported (Croxton 1979a) (non-iterative) convolution approximation. The present approximation clearly provides the best

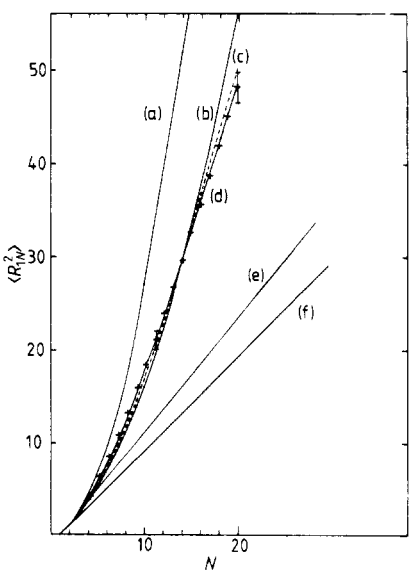

Figure 1. The mean square end-to-end distance $\langle R_{1N}^2 \rangle$ as a function of chain length on the basis of various approximations and compared with the Monte Carlo estimate. (a) Percus–Yevick (Curro), (b) geometric convolution, (c) arithmetic convolution, (d) Monte Carlo, (e) non-iterative convolution, (f) random walk.

representation yet of $\langle R^2_{1N}\rangle$ for short to intermediate length sequences ($N \leqslant 20$), crossing the Monte Carlo curve at $N = 15$. Below $N = 15$, $\langle R^2_N\rangle$ is slightly under-estimated, undoubtedly due to the partial neglect of exclusion processes, whilst above this value the details of the convolution product probably imply more extended configurations than is truly the case, as discussed earlier, resulting in an over-estimate of $\langle R^2_{1N}\rangle$. This suggestion is supported by a direct comparison of $Z(1N|N)$ for $N = 10$ (figure 2) with its Monte Carlo counterpart. We see quite clearly that for small $r_{1,10}$ the calculated distribution is collapsed into the origin, undoubtedly attributable to the under-estimate of exclusion processes operating within the sequence. At larger $r_{1,10}$, however, we note that the tail of the calculated distribution is sustained beyond its Monte Carlo counterpart with an associated over-estimate of $\langle R^2_{1N}\rangle$ as N increases. It is apparent from figure 1 that the *arithmetic* estimate of $\langle R^2_{1N}\rangle$ is in closer agreement with the Monte Carlo result than the geometric mean. The internal spatial distribution $Z(ij|N)$ determined on the arithmetic basis is quantitatively inferior to the geometric estimate, despite yielding a better $\langle R^2_{1N}\rangle$. Since we are not simply concerned with the moments of the distributions, but also with the distributions themselves, we choose to adopt the *geometric* mean.

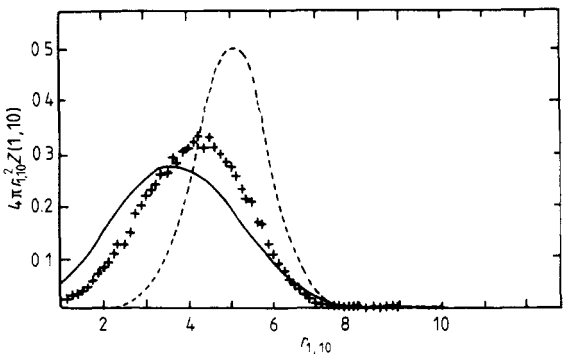

Figure 2. The spatial end-to-end probability distribution $Z(1, 10)$ for a sequence of ten self-avoiding hard sphere segments (Percus–Yevick (Curro), broken curve; geometric convolution, full curve). The Monte Carlo estimate is also shown (crosses).

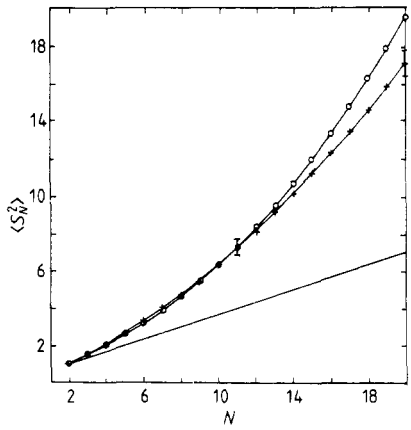

Figure 3. The radius of gyration $\langle S^2_N\rangle$ as a function of chain length on the basis of various approximations. (Crosses, Monte Carlo; open circles, geometric convolution; full curve, random walk.)

Incidentally, it is not appropriate to compare our numerical estimates of $\langle R^2_{1N}\rangle$ with the earlier MC data of Bruns (1977), since we believe the 'dimerisation' technique adopted therein to be incorrect. In that treatment randomly selected sequences of eight segments are 'dimerised' (provided there are no geometric violations) to yield sequences of integral multiples of 8-mer. Clearly, $\langle R^2_{1,8}\rangle$ for example, remains identical to its isolated end-to-end value, even in its multiply dimerised form.

All internal distributions $Z(ij|N)(i \neq j \neq 1, N)$ are determined in the course of numerical evaluation of (9) from which the internal second moments $\langle R^2_{ij}\rangle_N$ may be calculated. We find the subsequence of segments $|i-j|$ within the N-mer is generally *expanded* with respect to an isolated sequence of $|i-j|$ segments, in agreement with

previous investigations (Croxton 1979b, Redner 1980). Moreover, we find this conclusion is confirmed by Monte Carlo simulation. A particularly sensitive test of the accuracy of the internal distributions and their associated moments is provided by calculating the mean square radius of gyration $\langle S_N^2 \rangle$ of the chain, where

$$\langle S_N^2 \rangle = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \langle R_{ij}^2 \rangle \tag{12}$$

which may be compared directly with the random walk result

$$\langle S_N^2 \rangle = \tfrac{1}{3}(N+1). \tag{13}$$

We see from figure 3 that the geometric convolution estimate of $\langle S_N^2 \rangle$ on the basis of equation (12) is in very good agreement with the Monte Carlo data—a particularly important result since the radius of gyration is experimentally accessible through light scattering studies.

Whilst this iterative version yields a substantial quantitative improvement upon the previous convolution approximation, the model retains its simplicity and versatility. Boundary effects (on the basis of the previous approximation) have been investigated by allowing the diameter of the first segment $\to \infty$ (Croxton 1983), whilst the ith bond angle may be fixed by applying δ-function bonds between segments $i-1$, $i+1$. Any central, pairwise interaction may be specified between any given pair of segments, whose diameters may also be regarded as parameters of the sequence.

References

Bruns W 1977 *J. Phys. A: Math. Gen.* **10** 1963
Croxton C A 1979a *J. Phys. A: Math. Gen.* **12** 2475
—— 1979b *J. Phys. A: Math. Gen.* **12** 2487
—— 1979c *J. Phys. A: Math. Gen.* **12** 2497
—— 1983 *J. Phys. A: Math. Gen.* **16** 4343
Curro, J G, Blatz P J and Pings C J 1969 *J. Chem. Phys.* **50** 2199
Redner S 1980 *J. Phys. A: Math. Gen.* **13** 3525