The Distribution of End-to-End Separations of Self-Avoiding Walks on a Lattice[1]

Frederick T. Wall* and R. Allen White

William Marsh Rice University, Houston, Texas 77001.
Received October 16, 1974

ABSTRACT: To obtain a further understanding of macromolecular configurations, additional calculations have been carried out for self-avoiding walks on a tetrahedral lattice. Exact formulas are given for numbers of self-avoiding walks with one component of length approaching its maximum value. Two different Monte Carlo methods for obtaining distributions of end-to-end separations are described. The statistical studies, in conjunction with the exact calculations, give accurate length distributions for walks of 21 through 24 steps and somewhat less accurate information for walks of 30 steps. The results gained so far will improve extrapolations for long-chain behavior.

The problem of predicting theoretically the distribution of end-to-end separations of flexible polymer chains has proved to be exceedingly difficult.[2] Even assuming a very simple model, such as a self-avoiding random walk on a lattice, the problem remains quite complex indeed. Nevertheless, considerable progress has been made through a variety of methods including limited exact enumerations,[3–7] Monte Carlo techniques,[8–12] and asymptotic approaches.[13–15]

The results of analytic studies have been encouraging, but not fully satisfactory. Although pure counting techniques[5–7] appear to be most direct, they rapidly become overly cumbersome as the contour lengths of the chains increase. Generating function techniques have given rise to an improved insight of the problem by simplifying old results but without yielding much new information.[14,15] In this paper we shall describe some further efforts designed to extend our range of knowledge of the distribution of end points for self-avoiding random walks on a tetrahedral lattice.

It is not difficult in principle to set up procedures for the exact enumeration of lattice chain configurations, but the tremendous number of configurations encountered as contour lengths are increased overwhelms the capacity of generally available computers. On the other hand, sampling techniques can be used to good practical advantage, provided necessary precautions are observed in such a statistical approach.

In an effort to extend the range of exact counts of distinguishable configurations on a lattice, we proceeded by what we shall call a "stride" method. A stride is defined as an acceptable sequence of a fixed number of steps; the strides in the work here reported were six steps long. On a tetrahedral lattice, there are a total of 948 non-self-intersecting six-step strides of which only 79 are rotationally nonsuperimposable. The computational procedure consists first of determining which strides can follow other strides making sure that no interstride double occupancy of lattice sites occurs when one stride is attached to another. By combining two strides in this way, one can identify all self-avoiding walks of 12 steps. At the same time, one establishes in essence the transition matrix for the coupling of strides, subject to the constraint that no double site occupancies occur between adjacent strides.

The next step in the procedure is to determine what third strides can be attached to the allowable combinations of two strides. Nothing new is involved in asserting which third strides might be attached to the second strides, since that is established in the two-stride problem. However, it is necessary to find out if a third stride interferes with the first one when a particular second stride intervenes. Upon excluding configurations involving interference between first and third strides, one can then establish all permissible configurations of 18 steps.

In principle, this procedure can be extended indefinitely. Let s equal the number of strides and suppose the s-stride problem has been all worked out. To solve the $s + 1$ stride problem, one must first add strides to the acceptable s-stride configurations subject to the condition that the new s-stride combinations obtained by omitting the first stride are likewise acceptable. Tests are then carried out to determine whether the last, or $(s + 1)$th, strides are compatible with the first ones. Actually the scheme ultimately breaks down because of the sheer volume of data that must be stored and the prodigious amount of computation required if one seeks to count all possible configurations. In fact, we were unable to carry it through for four strides, so it was necessary to take recourse to Monte Carlo sampling of strides, utilizing information gained in solving the two-stride problem.

Relatively few chains have lengths approaching their maximum values and the great majority of them have smaller end-to-end separations. Accordingly, a sampling technique will yield, a priori, few long chains but many relatively short ones. Happily, it is possible to analyze limited exact enumeration studies to deduce formulas that give exact counts for chains with long components of length. Such formulas used in conjunction with samples obtained statistically for shorter chains can give a fairly accurate overall picture. Suppose $N_n(x)$ is the number of configurations of n steps with x equal to the corresponding component of end-to-end separation. Utilizing data reported by Wall and Hioe[7] on exact counts for walks on a tetrahedral lattice up to 20 steps, we have been able to set forth equations for $N_n(x)$ for $x = n, n - 2, n - 4, \ldots n - 12$. It is assumed that each step has an x component equal to ± 1; the maximum component is thus equal to n, with all others differing from n by an even number. The appropriate formulas for $N_n(x)$ are given in Table I.

The first statistical attack we pursued is one we shall call the total distribution approach. To obtain total distribution samples for $n = 24$, we proceeded as follows. All of the independent six-step walks were each followed by equal numbers of randomly selected trial strides. In each instance, the trial stride was checked to see if it was compatible with the first one. Each of the allowable two-stride configurations of 12 steps so obtained was similarly augmented by randomly chosen trial strides to give walks of 18 steps. The process was then repeated once more to yield 3,483,818 configurations of 24 steps.

In another set of calculations, using a somewhat different method of pseudo-random processing, we generated 4,662,440 additional walks of 24 steps. All configurations in

Table I
Numbers of Self-Avoiding Walks of n Steps with Final Coordinate x, $N_n(x)$, on a Tetrahedral Lattice

$N_n(n) = 2^n$
$N_n(n-2) = 2^{n-2}(n+2)$
$N_n(n-4) = (2^{n-4}/2!)(n^2 + 9n + 4)$
$N_n(n-6) = (2^{n-6}/3!)(n^3 + 21n^2 + 14n + 84);\ n > 6$
$N_n(n-8) = (2^{n-8}/4!)(n^4 + 38n^3 + 107n^2 + 412n + 606);\ n > 10$
$N_n(n-10) = (2^{n-10}/5!)(n^5 + 60n^4 + 435n^3 + 570n^2 + 7949n + 19{,}410);\ n > 13$
$N_n(n-12) = (2^{n-12}/6!)(n^6 + 87n^5 + 1225n^4 + 795n^3 + 47{,}644n^2 + 112{,}203n + 209{,}745);\ n > 16$

Table II
Numbers of Self-Avoiding Walks on a Tetrahedral Lattice Computed Statistically Using the Weighted Total Distribution Method[a]

	$N_n(x)$ with n =	
x	24	30
0	20,207,000,000	10,655,000,000,000
2	20,096,000,000	10,584,000,000,000
4	19,068,000,000	10,100,000,000,000
6	16,932,000,000	9,281,000,000,000
8	14,105,000,000	8,097,000,000,000
10	10,825,000,000	6,675,000,000,000
12	7,569,400,000 (7,575,395,584)	5,131,000,000,000
14	4,697,000,000 (4,699,631,616)	3,687,000,000,000
16	2,543,100,000 (2,537,373,696)	2,416,000,000,000
18	1,154,900,000 (1,150,812,160)	1,428,200,000,000 (1,421,137,395,714)
20	417,400,000 (417,333,248)	742,700,000,000 (746,375,348,225)
22	105,200,000 (109,051,904)	339,200,000,000 (339,959,873,536)
24	17,600,000 (16,777,216)	128,000,000,000 (129,754,988,544)
26		38,900,000,000 (39,392,903,168)
28		8,300,000,000 (8,589,934,592)
30		1,000,000,000 (1,073,741,824)
Total	2.1527×10^{11}	1.28×10^{14}

[a] Exact numbers calculated using formulas of Table I are given in parentheses underneath statistical numbers.

this latter group were in turn augmented by six trial strides, thereby generating 24,244,714 walks of 30 steps.

The results obtained for $n = 24$ and $n = 30$ were then analyzed in the following way. The total number of walks obtained with x components in the range $n \geq |x| \geq n - 12$ were grouped together and divided into the exact total to establish a weight factor. The exact total was calculated using the equations of Table I. The number of statistical samples obtained for each value of x in the long-component range was small, but the consolidated group was of sufficient size to be more meaningful. The actual numbers of configurations obtained for $n - 14 \geq |x| \geq 0$ were then multiplied by the weight factor to bring them into scale with the long component configurations for which we had exact information. The results of these computations are given in Table II.

The total distribution approach described above gives information for the whole range of end-to-end separations. However, if one seeks the probabilities of obtaining particular end-to-end separations, another method appears to be more attractive. This involves a procedure which we shall call the fixed coordinate technique.

Suppose, for example, we want to obtain information about self-avoiding walks of 24 steps with x components of length equal to 10 units. This can be accomplished by checking pairs of 12-step walks so chosen that the sum of their x components equals 10. For this purpose we need only consider the pairs (12,−2), (10,0), (8,2), and (6,4); because of symmetry, it is unnecessary to work with (4,6), (2,8), (0,10), and (−2,12). Since there are a tremendous number of such combinations, we found it necessary, once again, to take recourse to statistical methods. By pseudorandom selection of pairs of 12-step walks, subject to the constraint of a fixed component sum and to the elimination of those configurations involving intersections, we were able to obtain further results as shown in Table III. The fixed coordinate technique was carried out for walks with n = 21, 22, 23, and 24 steps.

A comparison of the results obtained by the two methods for $n = 24$ provides some idea as to the reliability of the results. For the range $0 \leq |x| \leq 10$ (for which we do not have exact formulas) the largest relative difference between the fixed coordinate and the total distribution methods was 0.81% for $x = 6$ and the smallest was 0.17% for $x = 0$. Another check on the reliability of the two methods can be made by comparing the weighted numbers obtained for the range $n - 12 \leq |x| \leq n$ with the results of exact calculations. For the fixed coordinate method, the maximum error is −0.59% for $N_{24}(12)$ and the root-mean-square error for N_{24} is 0.2%. Because of the paucity of samples, equally good agreement would not be expected for the fringe areas using the total distribution technique. Nevertheless, except for $N_{24}(22)$, $N_{24}(24)$, and for $N_{30}(x)$ with x = 24, 26, 28, and 30, the results are most gratifying. This suggests that, as x becomes smaller, the relative calculated populations should be even better.

Although both methods give substantially the same results, it is apparent that, for values of $|x|$ approaching n, the fixed coordinate method is more effective; on the other hand, the total distribution method is best suited for $|x|$ near zero. Both methods effect major computational time savings by utilizing the transition matrix of stride on stride acceptability. This is true because single yes–no calls in an array can be done much faster than point-by-point checks for possible looping. However, the time gain is made at the cost of storage, so this looms as a limiting factor in the computations. Since the transition matrix consists only of ones and zeros (to denote that two strides are or are not compatible), it is useful to store the array in bit string form. Because the programming language known as PL/I is highly suitable for manipulating such data, that language was used on an IBM 370-155 for the computations.

In addition to the distribution data, it is of interest to know something of the total number of configurations. In Table IV we see how the total number of self-avoiding walks, N_n, on a tetrahedral lattice changes with the number of steps for the range studied. An analysis of the data reported earlier by Wall and Hioe[7] for $n \leq 20$ showed that for extrapolation purposes the best equation found for a three-point fit was given by

$$\ln (N_n/4^n) = \alpha \ln n + \beta n + \gamma \quad (1)$$

where the three points must correspond to all even or all odd numbers. The fact that there is an even–odd zig zag in

Table III
Numbers of Self-Avoiding Walks on a Tetrahedral Lattice Computed Statistically Using the Fixed Coordinate Method[a]

	$N_n(x)$ with n =			
x	21	22	23	24
0		2,560,500,000		20,241,000,000
1	883,050,000		7,172,600,000	
2		2,533,400,000		20,017,000,000
3	855,310,000		6,922,000,000	
4		2,381,200,000		18,950,000,000
5	774,200,000		6,332,300,000	
6		2,088,500,000		17,070,000,000
7	653,000,000		5,423,000,000	
8		1,701,700,000		14,182,000,000
9	504,800,000 (505,074,944)		4,274,400,000	
10		1,259,000,000 (1,255,697,856)		10,858,000,000
11	(345,074,432)		3,085,000,000 (3,096,091,904)	
12		(829,816,832)		7,530,000,000 (7,575,395,584)
13	(205,770,752)		(1,981,318,144)	
14		(478,097,408)		4,678,000,000 (4,699,631,616)
15	(103,219,200)		(1,104,379,904)	
16		(231,604,224)		2,537,000,000 (2,537,373,696)
17	(41,549,824)		(517,391,184)	
18		(89,915,392)		1,152,000,000 (1,150,812,160)
19	(12,058,624)		(193,986,560)	
20		(25,165,824)		419,700,000 (417,333,248)
21	(2,097,152)		(52,428,800)	
22		(4,194,304)		109,100,000 (109,051,904)
23			(8,388,608)	
24				16,777,216 (16,777,216)

[a] Exact numbers calculated using formulas of Table I are given in parentheses underneath statistical numbers.

Table IV
Total Number of Configurations, N_n, of Self-Avoiding Walks of n Steps on a Tetrahedral Lattice

Method used	n	N_n
Exact	18	362,267,652
	19	1,052,271,732
	20	3,051,900,516
Fixed coordinate statistics	21	8,760,900,000
	22	25,799,000,000
	23	74,157,000,000
	24	215,410,000,000
Total distribution statistics	24	215,270,000,000
	30	128,000,000,000,000

certain properties, such as the moments, has been noticed before[4] and is not surprising. Utilizing exact data[7] for n = 18, together with the total distribution data for n = 24 and n = 30, we find that eq 1 can be used to fit N_{18}, N_{24}, and N_{30} with $\alpha = -5.569972 \times 10^{-3}$, $\beta = -0.32148126$, and $\gamma = 0.5573574$.

Finally, even though the regular moments of the distribution are functions of n, it appears[4,7] that certain reduced moments are not; accordingly, we include data on those reduced moments. Let the $2p$th reduced moment be defined by $M_{2p} = \langle x^{2p}\rangle / \langle x^2\rangle^p$. In Table V we show values of M_4, M_6, and M_8 for various n, including exact values for n = 18, 19, and 20 for comparison.

Discussion

The Monte Carlo results for the distribution of self-avoiding walks on a tetrahedral lattice are believed to be

Table V
Reduced Moments of Distributions for Self-Avoiding Walks of *n* Steps on a Tetrahedral Lattice

	n	M_4	M_6	M_8
Total distribution statistics	18	2.4802	8.8920	39.541
Exact	18	2.4785	8.8814	39.497
	19	2.4897	8.9926	40.446
	20	2.4972	9.0752	41.193
Fixed coordinate statistics	21	2.495	9.068	41.25
	22	2.519	9.292	43.04
	23	2.522	9.326	43.36
	24	2.523	9.359	43.78
Total distribution statistics	24	2.531	9.429	44.31
	30	2.562	9.751	47.27
	∞^a	2.65	10.8	58.0

[a] Taken from ref 7.

quite accurate for values of n up to 24. (We are less sure of the results for $n = 30$.) A comparison of the exact and statistical values, displayed in Tables II and III, supports this conclusion.

Another check on the general validity is obtained by considering the total number of walks calculated in two different ways. In preparing Tables II and III, we used the numbers of fringe walks, for which we have exact formulas, to establish weight factors. Alternatively, we could examine the overall attrition attending interstride exclusions. Suppose after 18 steps we try t different six-step strides, chosen randomly, to bring each walk up to 24 steps. If there are s successes, then we could calculate the total number of walks for $n = 24$ as follows:

$$N_{24} = N_{18}948(s/t) \qquad (2)$$

The factor 948 appears because there are precisely that many six-step strides of which t are tried. The computation can, of course, be generalized to cover the addition of more than one stride. Under those circumstances, we would calculate for p additional strides

$$N_{n+6p} = N_n \prod_i (948 s_i/t_i) \qquad (3)$$

where i denotes an added stride.

Since we know N_{18} exactly, we used eq 2 and found that $N_{24} = 2.1521 \times 10^{11}$; this is about 0.06% less than the number calculated using the weight factor obtained from exact formulas for fringe configurations. The results for N_{30} and the distribution associated with $n = 30$ are, of course, less reliable than those for $n = 24$. Nevertheless, we feel that the method used is not only valid but highly effective, and only more computation time is necessary to reach high precision. Once accurate values are obtained for $n = 30$, extrapolation can be carried out with greater confidence than heretofore.

References and Notes

(1) This work was supported by Grant No. C-521 of the Robert A. Welch Foundation.
(2) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969; C. Domb, *Adv. Chem. Phys.*, **15**, 229 (1969).
(3) M. E. Fisher and M. F. Sykes, *Phys. Rev.*, **114**, 45 (1959).
(4) C. Domb, J. Gillis, and G. Wilmers, *Proc. Phys. Soc., London*, **85**, 625 (1965).
(5) M. F. Sykes, J. W. Essam, B. R. Heap, and B. J. Hiley, *J. Math. Phys. (N.Y.)*, **7**, 1557 (1966).
(6) F. T. Hioe, *J. Phys. Chem.*, **74**, 4401 (1970).
(7) F. T. Wall and F. T. Hioe, *J. Phys. Chem.*, **74**, 4410 (1970).
(8) F. T. Wall and J. Erpenbeck, *J. Chem. Phys.*, **30**, 634 (1959).
(9) F. T. Wall, S. Windwer, and P. J. Gans, *J. Chem. Phys.*, **38**, 2220 (1963).
(10) Z. Alexandrowicz, *J. Chem. Phys.*, **51**, 561 (1969).
(11) S. D. Stellman and P. J. Gans, *Macromolecules*, **5**, 516 (1972).
(12) R. Grishman, *J. Chem. Phys.*, **57**, 220 (1973); 5309 (1973).
(13) F. T. Wall and S. G. Whittington, *J. Phys. Chem.*, **73**, 3953 (1969).
(14) M. E. Fisher, *J. Chem. Phys.*, **44**, 616 (1966).
(15) T. R. Chay, *J. Chem. Phys.*, **52**, 1025 (1970).

Viscoelastic Properties of Straight Cylindrical Macromolecules in Dilute Solution

Hiromi Yamakawa

Department of Polymer Chemistry, Kyoto University, Kyoto, Japan. *Received December 3, 1974*

ABSTRACT: The rotatory diffusion coefficient, intrinsic viscosity, and rigidity of straight cylindrical macromolecules are evaluated by an application of the Oseen–Burgers procedure of hydrodynamics. Plots of the reduced intrinsic viscosity and rigidity against the reduced frequency are rather insensitive to the change in the ratio of length to diameter. In particular, the high frequency limit of the latter is exactly independent of the molecular weight and takes the Kirkwood–Auer limiting value of ⅗ for an infinitely long rod.

In previous papers,[1] the steady state transport coefficients of stiff chains have been evaluated by an application of the Oseen–Burgers procedure of classical hydrodynamics to wormlike cylinder models with some comments[2] on Ullman's treatment[3] of the same problem. For example, the phenomenological friction constant per unit chain contour length remains in his final results, though his equations in the nondraining limit are equivalent to ours. In the present paper, it is shown that the kernels of his[4] and our integral equations are also different in the case of the rotatory diffusion, dynamic viscosity, and rigidity of rodlike or straight cylindrical macromolecules.

The present work has also been motivated by the very recent experimental results. The data obtained by Nemoto[5] for tobacco mosaic virus show that the high frequency limit of the reduced intrinsic rigidity is independent of the molecular weight and takes the Kirkwood–Auer limiting value[6] of ⅗ for an infinitely long rod. However, the Ullman