OLE J. HEILMANN

MATHEMATICAL POLYMERS I

Synopsis

The model introduced by VERDIER for the change of configuration of a polymer-molecule in solution has been examined and a new possibility for the motion of the polymer-segments has been tried. It is concluded that the long-ranged effect of the excluded-volume has almost no influence on the motion, while the rules chosen for changing bond-angles and bond-directions seems to be the determining factors. It is suggested that the relaxation-time for the change of configuration of a linear polymer with n segments should be proportional either to n^2 or to n^3, but the theory gives no possibility for preferring one possibility instead of the other.

The article includes a proposal for an extended least-squaremethod of estimation, which may be of general value, especially in Monte Carlo calculations.

This work has been inspired by an article by VERDIER and STOCKMAYER (1), which was later followed by two additional articles by VERDIER (2–3).

The polymers considered are assumed to consist of simple chains of n units (atoms). A cubic-lattice model is used for the configuration of the polymer. In a Cartesian reference-system this means that the atoms of the polymer are only allowed to be on points with integral coordinates, the distance between neighbouring atoms in the polymer being one. In models where the so called "excluded volume effect" is taken into consideration the further constraint that no two atoms are permitted to have identical coordinates is added. Within these restrictions all configurations are equally probable.

Several different models have been tried for the polymer changing its configuration, the general feature of all the models being the following: At equal time-intervals (t = 1, 2, 3, . . .) one of the n atoms of the polymer is chosen at random and moved according to the specific rules of the model, which may imply that no motion takes place. The philosophy behind these rules is that independent of the actual configuration of the polymer, any part of it has, within each unit of time, an equal probability of being affected by the surroundings (the solvent-molecules or other parts of the polymer) to such an extent that it changes its configuration significantly. If the time-unit of the model corresponds to this probability being 1/n, the model should be a reasonable discrete analogue of the actual physical process. The time-unit of the model will then be α/n real time-units (seconds), where α is an unknown constant.

The detailed rules of the seven models, which have been tried are as follows:

Model I (which is identical with the model introduced by VERDIER and STOCKMAYER (1)) is a model with excluded volume. Let the chosen atom be numbered i. If it is not an end-atom ($i \neq 1 \wedge i \neq n$) then the local configuration is either as shown in Fig. 1 or as shown in Fig. 2. For the case shown in Fig. 1 no movement is possible. For the case shown in Fig. 2 the configuration

1*

is changed to the configuration shown in Fig. 3 (meaning that the directions from atom no. $i-1$ to atom no. i and from no. i to no. $i+1$ are interchanged) if the new configuration of the polymer does not conflict with the excluded-volume constraint. If on the other hand the chosen atom is an end-atom, the direction from the neighbouring atom to the end-atom is changed to one of the directions perpendicular to the original one, the choise being made at random between the four possibilities, and again the change is only carried out if it does not result in a conflict with the excluded volume constraint.

In Model II an additional type of motion is included, while the excluded-volume-effect is maintained. If the local configuration is as shown in Fig. 4a or Fig. 4b then no change occurs in Model I, while in Model II a 90°-rotation to either the configuration of Fig. 5a (respective 5b) or the configuration of Fig. 6a (respective 6b) is attempted, the choice between Fig. 5 and Fig. 6 being random and the actual decision, whether to move or not to move is made so that the structure remains consistent with the excluded-volume constraint. The physical significance of the difference between the two models seems quite small, from a mathematical viewpoint however the difference is very important. In Model I the number of "bonds" in any of the six possible directions is conserved except for the movement of the end-atoms, while this is not the case in Model II.

In the Models III–VI only the excluded-volume-effect among next-neighbouring atoms in the polymer is maintained, such that atoms being separated by more than one atom are allowed to occupy the same lattice point. Except for this difference, the rules in Model III are the same as those of Model I and the rules in Model IV are analogous to those of Model II. In Model IV the special question of how to treat a configuration like the one shown in Fig. 7 arises. It was arbitrarily decided to treat it as the configuration shown in Fig. 4a and not as the configuration shown in Fig. 4b.

In the Models V and VI the probability of the movement, which leads from the configurations shown in Fig. 4 to those shown in Fig. 5 and 6, is diminished as compared to Model IV. In Model V it is decreased by a factor 5/9 which is obtained by leaving out the movement if the configuration is the one shown in Fig. 5b.

In Model VI a further reduction by a factor 1/2 as compared to Model V is obtained by introducing a random choice of whether to move or not to move.

Finally in Model VII the effect of the excluded volume is totally neglected, the rules otherwise being the same as the rules of Model I and III except for a small change in the rules for moving the end-atoms. The choice for this movement is now made between all six possible directions of the bond to
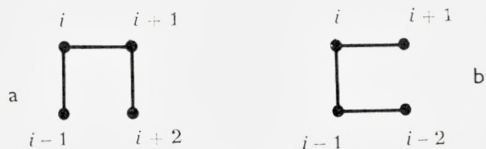
$i-1$ i $i+1$

Fig. 1

i $i+1$

a

$i-1$ $i+2$

i $i+1$

b

$i-1$ $i-2$

Fig. 4

i $i+1$

$i-1$

Fig. 2

$i-1$ $i+2$

a

i $i+1$

$i+1$

i

$i-2$ b

$i-1$

Fig. 5

$i+1$

$i-1$ i

Fig. 3

i $i+1$

a

$i-1$ $i+2$

i

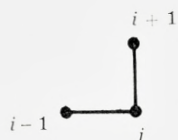$i+1$ b

$i+1$
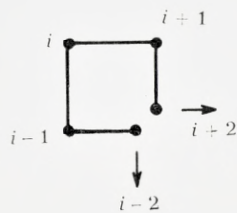
$i-2$

Fig. 6

i $i+1$

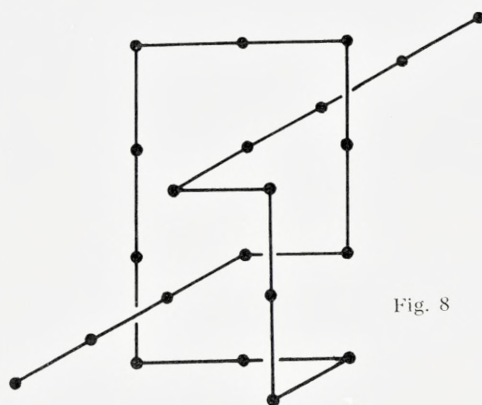$i-1$ $i+2$

$i-2$

Fig. 7

Fig. 8

the end-atoms. This change was introduced in the hope that it would then be possible to solve the model theoretically.

All the models for the movement of the polymer are Markovian, the process being discrete with the allowed configurations of the polymer as the states of the process. Taking configurations to be identical if they can be mapped onto each other by a simple translation (taking the numbering of the atoms into account), it is readily seen that the stochastic matrix for the Markov-process is symmetrical, and since not all atoms can be moved in all configurations not all the diagonal-elements of the matrix can be zero. The implication of this can be summarized as follows: The stochastic matrix is diagonalizable, all the eigenvalues are real, there are no transient states, there is no eigenvalue equal to -1 and the process is not cyclic. (See e. g. HOUSEHOLDER (4)). The configurations are numbered from 1 to N, where for a specific polymer N depends on the extend to which the excluded volume is taken into account. If p_{ji} is the conditional probability that a polymer which at time τ has configuration no. i will have configuration no. j at time $\tau + 1$, then the stochastical matrix is:

$$P = \{p_{ji}\}$$

Note that the matrix is transposed as compared to normal statistical nomenclature. When we introduce the probability-vector $p^{(\tau)}$ the i'th component of which is equal to the probability of finding the polymer in the configuration no. i to time τ and e as the vector having all unit components, the theory of Markov-processes gives as usual (the suffix T stands for transposing and all non-transposed vectors are taken to be column-vectors):

$$p^{(\tau)} = Pp^{(\tau-1)} = P^\tau p^{(0)} \qquad (1)$$

$$e^T p^{(\tau)} = 1. \qquad (2)$$

Numbering the eigenvalues by their numerical value:

$$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_N|; \quad \lambda_1 = 1$$

and numbering the eigenvectors, s_i, accordingly a suitable normalization gives:

$$s_i^T s_j = \delta_{ij} \qquad (3)$$

$$s_1^T = \left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}}, \ldots, \frac{1}{\sqrt{N}} \right) \qquad (4)$$

$$p^{(t)} = \sum_{i=1}^{N} (s_i^T p^{(0)}) \lambda_i^t s_i. \qquad (5)$$

In the case where the polymer is able to change from any configuration to any other configuration in a finite number of steps, the eigenvalue 1 is non-degenerate, with the result that:

$$\lim_{t \to \infty} p^{(t)} = \frac{1}{N} e. \tag{6}$$

This means that the equilibrium-distribution of the Markov-process always would be identical to the static equilibrium distribution, where all con-figurations are equally probable.

This is not the case however, for the models I and II. Consider e. g. the configuration shown in Fig. 8 in which the polymer has tied itself in a tight knot, which it is not possible to untie by the movements allowed in the two models mentioned. (It should be noticed that it is essential that the knot is tight, and that the number of knots in general is not a constant of the motion.) On the other hand, since the knot consists of a total of 18 atoms, having very limited possibilities of moving, the number of configurations with tight knots is a very small fraction, only, of the total number of configurations. Therefore disregarding these configurations totally and changing N and the stochastic matrix accordingly, probably introduces only a negligible difference. In the following it will consequently be assumed that $|\lambda_2| < 1$ and that eqn. (6) is valid for the models.

We consider now a measurable property, f, for the polymer, which means that f is a stochastic variable for the Markov-process. Let f be the vector, the i'th component of which has the value assumed by f when the polymer is in configuration no. i, then the expectation-value of f at time t is given by:

$$E\{f(t)\} = f^T p^{(t)} = \sum_{i=1}^{N} (s_i^T p^{(0)}) \lambda_i^t (f^T s_i) \tag{7}$$

$$\lim_{t \to \infty} E\{f(t)\} = E\{f(\infty)\} = (f^T e)/N \tag{8}$$

$$E\{f(t)\} = E\{f(\infty)\} + \sum_{i=2}^{N} \alpha_i \lambda_i^t \tag{7a}$$

$$\alpha_i = (s_i^T p^{(0)})(f^T s_i).$$

As it stands eqn. (7) is an expression which is not particularly useful since N is of the order 5^n. If, however, one or two of the eigenvalues are much closer to one than any of the other eigenvalues ($-1/\ln |\lambda_2| \gg -1/\ln |\lambda_j|$ if $|\lambda_j| \neq |\lambda_2|$)then it would be possible to describe the process by one or two relaxation-times, meaning that the summation in (7) is cut off at 3 (if $|\lambda_3| \neq |\lambda_2|$).

It is of course a very queer assumption and at the present no attempt shall be made to defend it except that it is a very usual approximation (see ref. 5).

At a first glance eqn. (7) would give the impression that the relaxation-time does not depend on the chosen property. This need not to be the case, however, if $f^T s_2$ is zero for some properties and not for others. Considering the form of the stochastic matrix, it is found that the matrix must be invariant under the transformation-group obtained by taking the direct product of the full cubic-symmetry group (O_h) and a group with two elements corresponding to the possibility of renumbering the polymer-atoms from the other end. Normal group-theoretical arguments then show that only properties trans-forming in the same manner under this group can be expected to have equal relaxation-times. On the other hand it should not be unreasonable, among properties transforming in the same manner to choose that property, which it is most convenient to work with. Since $p^{(\infty)}$ is invariant under the group, all properties having no invariant component should converge to zero as t goes to infinity.

Another way of getting rid of some of the relaxation-times by making some of the α's zero, would be to make some of the scalar products ($s_i^T p^{(0)}$) zero. This could be obtained by choosing a starting configuration (or distribution of configurations) with a special symmetry. It will never be possible to exclude the eigenvalues having invariant eigenvectors in this manner, however.

Instead of attempting to press the theoretical considerations further it has been tried to get additional insight in the problems by a Monte-Carlo calcula-tion, using a direct simulation of the Markov-process. The calculations have been carried out on a GIER-computer and the programs have so far been written in GIER-Algol 3 with parts of the program in machine-code to keep the computing-time within days. The randomness has been introduced by using a pseudo-random-number-generator:

$$x_{n+1} = 23 x_n \mod (2^{39} + 1) \tag{9}$$

where x_n is the n'th random number (ZELEN and SEVERO (6)).

The detailed accomplishment of the simulation was as follows. For each model and each chosen n, a polymer was started several times in the same, fixed starting-configuration and allowed to move according to the rules of the model. At equal time-intervals (in model-time-units) the actual configura-tion was registered and the value of the selected property was calculated. Up till now only a single property has been tried, namely the square of the end-to-end distance of the polymer, which is invariant under the symmetry-group of the stochastic matrix. This property has the advantage of being

the most thoroughly examined property of the equilibrium distribution. A property depending more explicitly on the positions of all the atoms of the polymer, such as the radius of gyration would be expected to have a smaller variance on the average-values and thus give better estimates.

The starting-configuration has been rather close to a three-fold-axis in the cubic-lattice, consisting of alternating bonds in the directions $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. This configuration has the advantage of having a very high value of the end-to-end distance.

The decision of spacing the "observations" equally in time was made to simplify the administration in the programs, the smaller efficiency being compensated by a rather close spacing. The disadvantages of this strategy should be compensated for by the sophisticated method of estimation which was used (see Appendix).

Applying the hypothesis of two eigenvalues being sufficient, the squared end-to-end distance of the i'th registration, d_i, should have the expectation-value:

$$E\{d_i\} = d_{eq} + \alpha_2 e^{-\gamma_2 i\Delta} + \alpha_3 e^{-\gamma_3 i\Delta} \tag{10}$$

(where d_{eq} is the equilibrium-value of the squared end-to-end distance, and Δ is the time spacing between two consecutive registrations of d).

The estimation of the five parameters d_{eq}, α_2, α_3, γ_2 and γ_3 is exactly the kind of problem treated in the Appendix. To ensure the reliability of the result a graphical test was made.

In order to test whether the method of estimation outlined in the Appendix was applicable in the present case a presumably typical example was selected, Model II with n = 32, and this was thoroughly examined, using a material of all together 640 starts. Three problems were of special interest: Which of the three formulas (I. 12), (I. 15) and (I. 17) should be used to estimate the variance? Would the number of observations during a single start be important, if the spacing between the observations was adjusted to keep the total running-time of the starts constant? And would varying the number of starts used for estimation give the expected results?

To settle the first question both formula (I. 12) and formula (I. 17) were used in all the cases which were also used to answer the two other problems. Since the difference between the resulting standard deviations was of the order 10 % and since q (formula (I. 3a)) also showed to be of that order, it was decided to use (I. 17), since it was the simplest formula. Consequently all standard deviations quoted will refer to this formula.

To see the influence of the number of observations, estimations were carried out with 8, 16 and 32 observations during the 8192 model-time-units

TABLE 1.

Number of observations	Number of starts	q	d_{eq}	α_2
8	160	$2.28_{10} - 2$	63 ± 4.1	$4.65_{10} - 4 \pm 6.1_{10} - 5$
16	160	$7.90_{10} - 2$	60 ± 4.2	$4.17_{10} - 4 \pm 4.6_{10} - 5$
32	160	$1.81_{10} - 1$	62 ± 3.2	$4.46_{10} - 4 \pm 3.4_{10} - 5$
8	640	$9.48_{10} - 3$	57 ± 1.3	$2.92_{10} - 4 \pm 1.6_{10} - 4$
16	640	$2.74_{10} - 2$	61 ± 2.9	$3.62_{10} - 4 \pm 2.7_{10} - 5$
32	640	$5.21_{10} - 2$	65 ± 2.0	$4.49_{10} - 4 \pm 2.0_{10} - 5$

Table 1 shows that the squared end-to-end distance at equilibrium, d_{eq}, and the reciprocal relaxation time, γ_2, are essentially independent of the number of observations.

in which each start was followed. To ensure that the conclusions would not depend strongly on the number of starts used, both the total of all 640 starts and a smaller group of 160 starts were used for the estimation. The results are shown in Table I. The differences do not seem to be significant. To use 8 observations for estimating 5 parameters is however rather unfavourable, and this leaves us with the choice between 16 and 32 observations. In general 16 has been used since this gives a smaller computation-time and a better numerical stability in finding the minimum for q. Finally, to see the effect of varying the number of starts, estimations were made on four groups of 40 starts each, on four groups of 160 starts each and on all 640 starts together. The results are shown in Table 2. It should be remarked that the estimations on the groups of 40 starts were made with 16 observations per start, while the others were made with 32 observations per start, meaning that the q's are not directly comparable. The two rows denoted mean give the mean of the four rows just above. The uncertainties quoted in these two rows are calculated from square-sums of the deviation from the mean, not using the uncertainties on the single estimates. The behaviour of the results seems satisfactory and especially it seems that the method of estimating the standard-deviation is reasonable.

For the general choice of the number of starts to use for estimation, the following considerations were essential. It was desirable to have approximately the same relative error on the estimates. If the number of starts was too low, it was extremely difficult to find minimum for q. The computation per start was however growing rapidly with the number of atoms in the polymer. It was found that when the number of atoms was equal to 8, 16, 32, 64 and 128 respectively, a reasonable compromise was something around 1200, 600, 300, 150 and 60 starts respectively.

TABLE 2.

	Numbers of starts	q	d_{eq}	α_2
	40	$2.68_{10} - 1$	48 ± 6.6	$3.28_{10} - 4 \pm 6.0_{10} - 5$
	40	$9.19_{10} - 2$	42 ± 95	$2.37_{10} - 4 \pm 7.0_{10} - 4$
	40	$1.28_{10} - 1$	72 ± 16	$7.70_{10} - 4 \pm 1.9_{10} - 4$
	40	$3.77_{10} - 2$	67 ± 21	$4.86_{10} - 4 \pm 1.7_{10} - 4$
mean	160	$1.32_{10} - 1$	57 ± 7.3	$4.55_{10} - 4 \pm 1.2_{10} - 4$
	160	$1.81_{10} - 1$	62 ± 3.2	$4.46_{10} - 4 \pm 3.4_{10} - 5$
	160	$2.36_{10} - 1$	60 ± 4.0	$3.80_{10} - 4 \pm 3.4_{10} - 5$
	160	$2.62_{10} - 1$	61 ± 3.0	$5.14_{10} - 4 \pm 3.4_{10} - 5$
	160	$1.94_{10} - 1$	67 ± 3.7	$5.08_{10} - 4 \pm 3.9_{10} - 5$
mean	640	$2.18_{10} - 1$	63 ± 1.6	$4.62_{10} - 4 \pm 3.1_{10} - 5$
	640	$5.21_{10} - 2$	65 ± 2.0	$4.49_{10} - 4 \pm 2.0_{10} - 5$

Table 2 shows that the estimated standard deviation is essentially proportional to the square root of the number of starts.

As a further check on the method the estimated values of d_{eq} were compared with the values known from equilibrium data:

$$d_{eq} = 1.067(n - 1)^{6/5} - 0.0915 \qquad (11)$$

for the models I and II (DOMB (7)),

$$d_{eq} = \frac{3}{2}(n - 1) - \frac{5}{8} + \frac{1}{8 \cdot 5^{n-2}} \qquad (12)$$

for the models III–VI, and

$$d_{eq} = n - 1 \qquad (13)$$

for model VII. As can be seen from table 3 the agreement is satisfactory. The method of estimation thus being confirmed, the results of the simulations will now be considered. Of the parameters estimated, γ_2 is the most interesting since γ_2^{-1} supposedly has some connection with the relaxation time in certain experiments on polymers. All the estimated values of γ_2 with the estimated standard deviations are shown in Table 4. (The values of γ_3 were generally larger by an order of magnitude.)

It is of course not really the absolute values of γ_2, but only the way in which it depends on n (the number of atoms in the polymer), which can be predicted by the simulations. Naturally the functional form to be chosen is open for discussion. It is easily seen that:

TABLE 3.

n	8	16	32	64	128
Formula (11)	10.8	27.2	65.5	154	358
Model I	$10 \pm .3$	$27 \pm .6$	67 ± 4	111 ± 20	489 ± 80
Model II	$12 \pm .3$	$27 \pm .9$	61 ± 3	166 ± 9	319 ± 24
Formula (12)	9.88	21.9	44.9	93.9	190
Model III	$10 \pm .3$	26 ± 1.6	62 ± 9	268 ± 20	990 ± 150
Model IV	$11 \pm .2$	22 ± 7.7	45 ± 1.3	98 ± 4	201 ± 10
Model V	$11 \pm .3$	$23 \pm .8$	45 ± 1.7	95 ± 4.3	190 ± 15
Model VI	$20 \pm .2$	19 ± 2	48 ± 2	101 ± 6	171 ± 25
Formula (13)	7	15	31	63	127
Model VII	$7 \pm .1$	$15 \pm .4$	30 ± 1.7	59 ± 5	107 ± 14

Table 3 shows that the values of the squared end-to-end distance at equilibrium, d_{eq}, estimated from the relaxation curves are essentially identical with values known from equilibrium data (formula (11), (12) and (13)).

$$\gamma_2 \approx an^{-\nu} \tag{14}$$

is a possibility, and that $\nu = 4$ for model I and III and $\nu = 3$ for the other five models would not be unreasonable. This is a result which for the models I and VII agrees with VERDIER (3). For the models II, IV, V and VI however the addition of a further term in eqn. (14) seems necessary to get agreement with the values for γ_2 for $n = 8$, so the forms shown in Table 5 are tentatively proposed as a representation of the values of γ_2. (The factors are least-square estimates and the uncertainties are standard deviations). It should be noticed that the time-unit for γ_2 is the model-time unit. To convert the results to real time-units, the values of γ_2 should be multiplied by n.

TABLE 4.

Model \ n	8	16	32	64	128
I	$(3.8 \pm .6)_{10} - 2$	$(2.2 \pm .2)_{10} - 3$	$(1.3 \pm .2)_{10} - 4$	$(5.5 \pm .9)_{10} - 6$	$(4.9 \pm .7)_{10} - 7$
II	$(4.5 \pm .7)_{10} - 2$	$(3.6 \pm .4)_{10} - 3$	$(3.6 \pm .3)_{10} - 4$	$(5.5 \pm .5)_{10} - 5$	$(5.0 \pm .4)_{10} - 6$
III	$(2.3 \pm .3)_{10} - 2$	$(2.1 \pm .3)_{10} - 3$	$(1.3 \pm .2)_{10} - 4$	$(1.2 \pm .1)_{10} - 5$	$(4.7 \pm .8)_{10} - 7$
IV	$(5.0 \pm .6)_{10} - 2$	$(3.9 \pm .3)_{10} - 3$	$(4.5 \pm .4)_{10} - 4$	$(5.6 \pm .3)_{10} - 5$	$(7.8 \pm .4)_{10} - 6$
V	$(4.7 \pm .6)_{10} - 2$	$(3.7 \pm .3)_{10} - 3$	$(3.5 \pm .2)_{10} - 4$	$(5.1 \pm .3)_{10} - 5$	$(5.3 \pm .3)_{10} - 6$
VI	$(2.3 \pm .3)_{10} - 2$	$(1.9 \pm .3)_{10} - 3$	$(3.1 \pm .2)_{10} - 4$	$(3.7 \pm .2)_{10} - 5$	$(4.2 \pm .2)_{10} - 6$
VII	$(4.4 \pm .4)_{10} - 2$	$(5.2 \pm .3)_{10} - 3$	$(5.9 \pm .3)_{10} - 4$	$(7.9 \pm .3)_{10} - 5$	$(8.9 \pm .3)_{10} - 6$

Table 4 shows the estimated valves of the reciprocal relaxation time, γ_2, in model-time-units.

TABLE 5.

Model	α_2
I	$(129 \pm 7)\ n^{-4}$
II	$(10.2 \pm .7)\ n^{-3} + (81 \pm 20)\ n^{-4}$
III	$(129 \pm 8)\ n^{-4}$
IV	$(14.8 \pm .6)\ n^{-3} + (35 \pm 18)\ n^{-4}$
V	$(10.8 \pm .5)\ n^{-3} + (74 \pm 18)\ n^{-4}$
VI	$(9.0 \pm .4)\ n^{-3} + (17 \pm 12)\ n^{-4}$
VII	$(19.8 \pm .4)\ n^{-3}$

Table 5 shows the estimated dependence of the reciprocal relaxation time, γ_2, on the number of atoms in the polymer. Remark that the time-unit is the model-time-unit.

It seems possible to draw the following three conclusions from the material shown in Table 4 and 5.

The real excluded volume has only an imperceptible influence on the relaxation (compare Model I and II with Model III and IV).

The possibility of a valence-angle equal to $0°$, however, has an appreciable influence (compare Model III with Model VII)*).

The additional possibility of movement introduce in Models II and IV as compared with models I and III causes also a radical change in the relaxation behaviour.

The general result is then that it is the local structure of the polymer rather than the long-range effect of the excluded volume, which is of importance for understanding the movements of polymers.

Acknowledgements

* These conclusions should be compared with the analogous conclusions by VERDIER (3).

Appendix

A General Least Square Method

Consider n uncorrelated observations $y^{(i)}$ (i = 1, 2, . . . , n) of a p-dimensional stochastic variabel having the same p-variate distribution, F, with mean value μ given as a function of m parameters, β_i(i = 1, 2, . . . , m), where m < p:

$$\mu_j = g_j(\beta) \qquad\qquad j = 1, 2, \ldots, p \qquad\qquad (I.\,1)$$

The problem is that of finding reasonable estimates for the parameters, β_i, and for the variance of the estimators used.

The sample mean:

$$y = \frac{1}{n} \sum_{i=1}^{n} y^{(i)} \qquad\qquad (I.\,2)$$

will in general, following the central-limit theorem, be asymptotically (n → ∞) normally distributed with mean μ and a dispersion matrix $\frac{1}{n}\Sigma$, Σ being the dispersion matrix of F. Argueing from this or from the general theory of least squares a good estimate of β should be the value of β that minimizes:

$$q' = (y_r - g_r(\beta))\sigma^{rs}(y_s - g_s(\beta)) \qquad\qquad (I.\,3)$$

(y_r without superscript stands for component of y defined in (I. 2), σ^{rs} are written for elements of Σ^{-1}, and here as well as in the following we use the convention that repeated indices in a product means summation over these indices, the limits of the summations being self-evident).

In order to be able to solve the problem of minimizing q it is indispensable to know Σ. Since Σ cannot be assumed known it is necessary to use an estimate of Σ.

An unbiased estimate for the elements of Σ is (conf. e. g. Rao (8))

$$s_{rs} = \frac{1}{n-1}\left[y_r^{(i)}y_s^{(i)} - n\,y_r y_s \right] \qquad\qquad (I.\,4)$$

Using this in (I. 3), the quadratic form to be minimized becomes:

$$q = (y_r - g_r(\beta))s^{rs}(y_s - g_s(\beta)) \qquad\qquad (I.\,3\,a)$$

(superscripts on a matrix-element are again used to designate the elements of the inverse matrix).

We shall not here consider the intricate numerical problem of finding minimum for q if $g(\beta)$ is not linear in β. The solution, β^*, will in any case satisfy the equations:

$$-2 \frac{\partial g_r(\beta)}{\partial \beta_l} s^{rs}(y_s - g_s(\beta)) = 0 \qquad \left.\begin{array}{c} \\ l = 1, 2, \ldots, m \end{array}\right\} \quad (\text{I. 5})$$

β^* being an implicit function of y by (I. 5). A simple extension of the proof by CRAMER for variance of functions of moments (CRAMER (9) p. 353 ff.) gives:

$$C\{\beta_l^*, \beta_k^*\} = \frac{\partial \beta_l^*}{\partial y_r}(\sigma_{rs}/n)\frac{\partial \beta_k^*}{\partial y_s} + O(n^{-3/2}) = d'_{lk} \quad (\text{I. 9})$$

where σ_{rs} is an element of Σ.

By (I. 5):

$$\frac{\partial \beta_k^*}{\partial y_r} = 2a^{kl}(\beta^*)\frac{\partial g_s(\beta)}{\partial \beta_l}\bigg|_{\beta = \beta^*} s^{rs} \quad (\text{I. 10})$$

where

$$a_{kl}(\beta) = \frac{\partial^2 q}{\partial \beta_l \partial \beta_k} = 2 \frac{\partial g_r(\beta)}{\partial \beta_l} s^{rs}\frac{\partial g_s(\beta)}{\partial \beta_k} - 2 \frac{\partial^2 g_r(\beta)}{\partial \beta_l \partial \beta_k} s^{rs}(y_s - g_s(\beta)) \quad (\text{I. 11})$$

Hence using s_{sr} for σ_{sr} and neglecting terms of the order $n^{-3/2}$:

$$d'_{lk} = \frac{4}{n}a^{kj}(\beta^*)\frac{\partial g_s(\beta)}{\partial \beta_j}\bigg|_{\beta = \beta^*} s^{rs}\frac{\partial g_r(\beta)}{\partial \beta_i}\bigg|_{\beta = \beta^*} a^{il}(\beta^*). \quad (\text{I. 12})$$

To be correct it is not only y also s^{rs} that is estimated.

Making allowance for this is somewhat complicated. However, using the normal approximation, y and $\{s^{rs}\}$ becomes uncorrelated and the covariance of s_{rs} and s_{tu} becomes:

$$C\{s_{rs}, s_{tu}\} = (s_{ru}s_{st} + s_{rt}s_{su})/(n-1). \quad (\text{I. 13})$$

The total-effect is then the addition of

$$d''_{lk} = \frac{n}{n-1}qd'_{lk} \quad (\text{I. 14})$$

to d'_{1k} (For a proof see the end of the Appendix). Hence the total covariance-matrix is

$$d_{1k} = d'_{1k}\left(1 + \frac{n}{n-1}q\right). \tag{I. 15}$$

This should be compared with the analogous result by Rao (10, 11) for the case where $g(\beta)$ is linear in β.

Since q is of the order n^{-1} and the relative error of omitting terms of order $n^{-3/2}$ in (I. 9) is of the order $n^{-1/2}$, (I. 12) will actually do just as well as (I. 15).

However, reasonning this way (I. 12) can be simplified even more. As the second terms in (I. 11) is of the order $n^{-1/2}$ as compared to the first, $a_{1k}(\beta)$ can be approximated accurately enough by:

$$2b_{1k}(\beta) = 2\frac{\partial g_r(\beta)}{\partial \beta_1} s^{rs} \frac{\partial g_s(\beta)}{\partial \beta_k} \tag{I. 16}$$

and using this, (I. 12) becomes

$$d^+_{1k} = b^{kl}(\beta^*)/n. \tag{I. 17}$$

Finally a proof of eqn. (I. 14) under the assumption of a normal distribution shall be given.

Introducing $S^{(rs)}$ for the complement of s_{rs} i $S(S = \{s_{rs}\})$ and a double complement $S^{(rs)(tu)}$ being equal to the complement of s_{tu} in $S^{(rs)}$, if $r \neq t$ and $s \neq u$, and zero otherwise, and using $|S|$ for the determinant of S, straightforward calculations give:

$$S^{(rs)(tu)} = S^{(ts)(ru)} = S^{(tu)(rs)} = S^{(ru)(ts)} \tag{I. 18}$$

$$s^{rs} = S^{(rs)}/|S| \tag{I. 19}$$

$$\frac{\partial s^{rs}}{\partial s_{tu}} = \frac{S^{(rs)(tu)}}{|S|} - \frac{S^{(rs)}S^{(tu)}}{|S|} \tag{I. 20}$$

$$S^{(rs)(tu)}s_{tu} = S^{(ru)}\delta_{sv}(1 - \delta_{su}) + S^{(rs)}\delta_{uv}(1 - \delta_{su}) \tag{I. 21}$$

(factors $(1 - \delta_{su})$ and $(0 - \delta_{su})$ will not cause summation irrespective of the indices s and u being repeated).

$$\frac{\partial s^{rs}}{\partial s_{tu}}s_{tv} = s^{ru}\delta_{sv}(1 - \delta_{su}) + s^{rs}\delta_{uv}(0 - \delta_{su}) \tag{I. 22}$$

$$\frac{\partial s^{rs}}{\partial s_{vw}}s_{vx}s_{wy}\frac{\partial s^{tu}}{\partial s_{xy}} = s^{ru}s^{ts}. \tag{I. 23}$$

Then on evaluating d''_{1k} in the same way as d'_{1k} we get:

$$
\left.
\begin{aligned}
d''_{ek} &= \frac{4}{n-1} a^{kj}(\beta^*) \frac{\partial g_r(\beta)}{\partial \beta_j}\bigg|_{\beta=\beta^*} (y_s - g_s(\beta^*)) \\[2mm]
&\quad \frac{\partial s^{rs}}{\partial s_{vw}} (s_{vx}s_{wy} + s_{vy}s_{wx}) \frac{\partial s^{tu}}{\partial s_{xy}} \\[2mm]
&\quad (y_t - g_t(\beta^*)) \frac{\partial g_u(\beta)}{\partial \beta_i}\bigg|_{\beta=\beta^*} a^{il}(\beta^*).
\end{aligned}
\right\}
\quad\text{(I. 24)}
$$

Using (I. 23), (I. 5), (I. 3 a) and (I. 12), (I. 24) can be transformed into (I. 14). The essential content of (I. 14), namely that d'_{1k} is smaller than d''_{1k} by a factor n, can of course be seen very easily by a direct comparison of (I. 24) with (I. 12), which shows that (I. 24) contains two extra factors of the form $(y_j - g_j(\beta^*))$, each being of the order $n^{-1/2}$.

Kemisk Laboratorium III
H. C. Ørsted Institutet
Københavns Universitet

References

1. P. H. VERDIER and W. H. STOCKMAYER, J. Chem. Phys. **36** 227 (1962).
2. P. H. VERDIER, J. Chem. Phys. **45** 2118 (1966).
3. P. H. VERDIER, J. Chem. Phys. **45** 2122 (1966).
4. A. S. HOUSEHOLDER, "The Theory of Matrices and Numerical Analysis", New York (1964).
5. See e. g. S. K. KIM, J. Chem. Phys. **28** 1057 (1958), K. E. SHULER, G. H. WEISS and K. ANDERSEN, J. Mat. Phys. **3** 550 (1962) and ref. 1.
6. M. ZELEN and N. C. SEVERO, "Probability Functions" in M. ABRAMOWITZ and I. A. STEGUN (ed.) "Handbook of Mathematical Functions", New York (1965).
7. C. DOMB, J. Chem. Phys. **38** 2957 (1963).
8. C. R. RAO, "Linear Statistical Inference and its Applications", New York (1965).
9. H. CRAMER, "Mathematical Methods of Statistics", Princeton (1946).
10. C. R. RAO, Biometrika **46**, 49 (1959).
11. C. R. RAO in L. Lecam & J. Neymann (Ed.) "Proceedings of the Fifth Berkeley Symposion on Mathematical Statistics and Probability". Vol I California (1967) p. 355.