

MONTE CARLO GENERATION OF POLYPEPTIDE RANDOM COIL CONFORMATIONS; THE PERSISTENCE VECTOR AND THE CHAIN VECTOR DISTRIBUTION

F.Th. HESSELINK*

Van 't Hoff Laboratorium der Rijksuniversiteit, Trans. III, Utrecht, The Netherlands

Received 2 January 1974

Polypeptide random coil conformations of various chain lengths ($N = 5, 10, 20, 40, 80$ peptide units) are generated by a Monte Carlo procedure. The characteristic ratio obtained for the sets of generated conformations is identical with the exact value calculated with the average transformation matrix procedure, indicating the equivalence of the two treatments. On the basis of the generated sets of conformations the length and direction of the persistence vector (the averaged chain vector expressed in the reference frame of the first two skeletal bonds) are investigated for various chain lengths. The radial distribution function for the chain vector shows the length of the chain vector for small polypeptides ($N = 5, 10$) not to deviate far from its most probable value. Also for larger chains up to chains of 80 peptide units very significant deviations from a gaussian distribution are observed.

The distribution of the length of the vector connecting the remote end of the chain with the end of the persistence vector exhibited behavior much closer to the gaussian approximation, an improvement especially significant for the short chains.

1. Introduction

Recently, Flory [1] has called attention to the mean displacement vector α , the conformational average of the chain vector r expressed in a reference frame fixed with respect to the first two skeletal bonds of the polymer chain. The interest in this vector α resides in Flory's assumption that, especially for short chains, the density distribution of the vector ρ , with $\rho = r - \alpha$, should behave in better approximation according to gaussian statistics than the distribution of the chain vector itself. In this paper we shall investigate this assumption for polypeptide random coil conformations, using a previously described [2] Monte Carlo procedure.

Monte Carlo generation of polymer chains has become an increasingly popular method to obtain information of the conformational behavior of polymer molecules. This technique has been applied to various conformational problems, such as the influence of volume exclusion on chain dimensions [3–7], the conformational

behavior of macromolecules near interfaces [8, 9], the behavior of short chains [10] and ring closure problems [11], etc. Recently, a set of Monte Carlo generated poly L-lysine random coil conformations served as a basis for the calculation of the electrical free energy of a poly-lysine random coil [2]. However, this set contained only a rather small number (200) of conformations, and so a second purpose of this paper is to investigate further the validity of the proposed generation procedure by a calculation of the chain vector length and distribution for larger sets of conformations and for various chain lengths. Our calculations are based on a set of preferred conformations for the L-lysine residue, but presumably they apply to any (co)poly α -L-amino acid containing C^β atoms.

2. Method of calculation

The Monte Carlo generation procedure has been described elsewhere [2]; it is based on the assumption that the conformation of a residue in a randomly coiled polypeptide chain is independent of neighboring residues [12],

* Present address: Shell Research, P.O. Box 60, Rijswijk, The Netherlands.

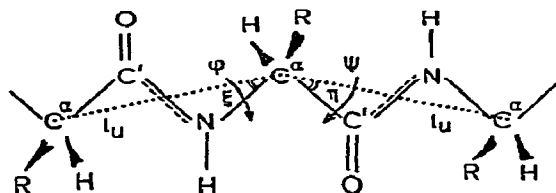


Fig. 1. Polypeptide chain indicating the meaning of the fixed angles η and ξ and of the variable angles φ and ψ .

whereas the conformation of each residue (i.e. the rotational angles φ and ψ , see fig. 1) is determined by the short range interatomic interaction within the peptide residue. The probabilities for various (φ, ψ) combinations are tabulated [2], and the Monte Carlo procedure consists essentially in the random picking of (φ, ψ) combinations for the consecutive residues in the polypeptide chain. The coordinates of the atoms of the macromolecule are then calculated using the peptide geometry and transformation matrix procedure described by Ooi et al. [13]. This procedure should generate the unperturbed dimensions of the polypeptide random coil.

The characteristic ratio, C_N , given by

$$C_N = \langle r^2 \rangle / Nl_u^2, \quad (1)$$

$\langle r^2 \rangle$ being the mean square end-to-end distance of a chain of N peptide units each of length l_u , has been calculated from the generated sets of conformations for various chain lengths, with $l_u = 3.80 \text{ \AA}$. Exact values for C_N for various N are also calculated on the basis of the preferred set of (φ, ψ) combinations (table 4 in ref. [2]), using the average transformation matrix procedure described by Brant and Flory [14, 12]. In these calculations the value used for the angle between the virtual bond and the C-N bond ξ was 13.5° and the value of the angle between the virtual bond and the C-C' bond η was 21.9° (see fig. 1). The peptide bond between C'-N has been taken as flat and rigid.

3. Results and discussions

3.1. Characteristic ratio

Sets of random coil conformations have been generated for polypeptides with chain lengths, ranging from

$N = 5$ up to $N = 80$ peptide units. In table 1 the number of generated conformations and the obtained values for the characteristic ratio are given, together with the exact values for C_N , obtained using Brant and Flory's average transformation matrix procedures ([12] p. 25; [14]) and our set of preferred (φ, ψ) combinations. The agreement between the values of the characteristic ratio obtained with these two different methods is such that we can state confidently that the generated sets of conformations are indeed representative for the polypeptide random coil. The present Monte Carlo procedure to generate polypeptide conformations proves to be a significant improvement over previous procedures [4, 5], which produced too low values for the characteristic ratio.

Due to differences in the chosen peptide geometry and in the set of preferred (φ, ψ) angles, our values for the characteristic ratio are somewhat higher than those obtained by Brant and Flory [14], possibly indicating that the dimensions of our sets of random coil conformations are somewhat larger than would correspond with "unperturbed dimensions". However, since polypeptide molecules have the tendency to form regular structures (α -helix, β -structure) in poor solvents, we usually only find random coil conformations in good solvents, i.e., under conditions when polypeptides have much larger dimensions than would correspond with unperturbed dimensions. Therefore, the differences between the generated conformations and the unperturbed dimensions, if any, are thought to be of little importance.

3.2. Persistence vector

When the chain vector \mathbf{r} , which connects to two ends of the chain, is expressed in terms of an external reference frame, the averages of the components of \mathbf{r} become zero. However, expressed in a reference frame defined by the first two skeletal bonds of the chain the averages of the components of vector \mathbf{r} represent a finite displacement \mathbf{a} from the origin. Flory [1] calls this vector \mathbf{a} the persistence vector of the chain, since it is obviously related with the persistence length a of the Kratky-Porod [15] chain model. For a discussion of this relationship we refer to Flory's paper [1]. From our sets of random coil conformations values for this persistence vector are obtained for polypeptide chains of various chain lengths, and these values are given in table 1. As Flory expected, the length of \mathbf{a} increases with chain length for short chains,

Table 1
Results of the Monte Carlo generation of sets of polypeptide random coil conformations

Number of peptide units	Number of generated conformations	$\langle C_N \rangle_{\text{M.C.}}$	$\langle C_N \rangle_{\text{exact}}$	Persistence vector \mathbf{a} (Å)	$ \mathbf{a} $ (Å)	$\langle r^2 \rangle^{1/2}$ (Å)	$\langle \rho^2 \rangle^{1/2}$ (Å)
5	20000	2.981	2.982	(10.9 -3.5 - 5.6)	12.7	14.7	7.31
10	10000	4.900	4.899	(16.3 -4.5 - 9.6)	19.4	26.6	18.2
20	5000	7.166	7.174	(20.8 -6.3 -12.6)	25.1	45.5	37.9
40	5000	9.113	9.053	(23.2 -5.4 -13.4)	27.4	72.6	67.2
80	2000	10.112	10.151	(20.9 -7.8 -15.0)	26.9	108.1	104.7

and it levels off for longer chains to a finite length we find to be of about 27 Å.

Figure 2 shows the orientation of these persistence vectors for various chain lengths in the (x, y) plane, the plane defined by the also shown orientation of the first peptide unit. The origin is chosen at the first C^α atom.

As can be seen, the length of the persistence vector increases up to a certain length, but its orientation is predominantly determined by the orientation of the first residue. Because of the asymmetric character of chains composed of L-amino acids, the z -coordinate of the persistence vector (see table 1) deviates from the orientation of the first residue.

Because of the non-zero values of the average coordinates of the chain vector when expressed in the frame of reference of the first peptide unit ($\langle x \rangle$, $\langle y \rangle$, $\langle z \rangle$), the distribution of the cartesian components of the chain vector is also far from symmetric around the origin. As an example, fig. 3 shows the distribution of the x , y , and z component of the chain vector for a chain of 20 peptide units. Note that these distributions are not symmetric either around $\langle x \rangle$, $\langle y \rangle$ or $\langle z \rangle$ respectively.

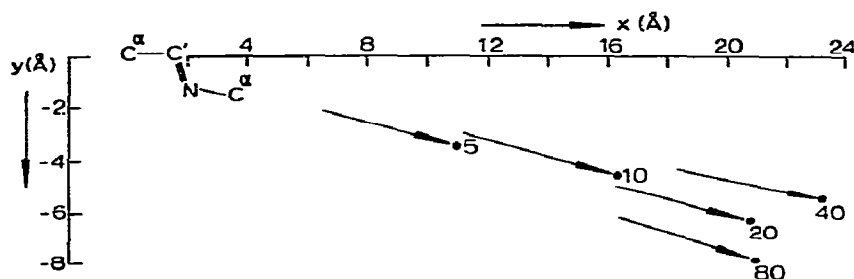


Fig. 2. Orientation of the first peptide unit in the x - y plane and the persistence vectors for polypeptide chains of 5, 10, 20, 40, and 80 peptide units.

3.3. The chain vector \mathbf{r} and its radial distribution function

The radial distribution function of the chain vector for chains of 5, 10, 20, 40 and 80 peptide units is shown in figs. 4–8, whereas the rms end-to-end distance $\langle r^2 \rangle^{1/2}$ corresponding with these distributions is given in table 1. Using these values for $\langle r^2 \rangle^{1/2}$, radial distribution functions $R(r)$ are calculated according to the gaussian approximation with

$$R(r) = 4\pi(3/2\pi\langle r^2 \rangle)^{3/2} r^2 \exp(-3r^2/2\langle r^2 \rangle) \quad (2)$$

and these distributions are also shown in figs. 4–8.

Comparison of the actual and the gaussian distributions indicates the extreme deviation from gaussian behavior for short chains. For chains up to 10 peptide units the actual distribution is extremely peaked around a most probable value for the end-to-end distance, a value shown to be not very much shorter than the length of a completely extended polypeptide chain (with $\phi = \psi = 180^\circ$). For longer chains the actual and gaussian distri-

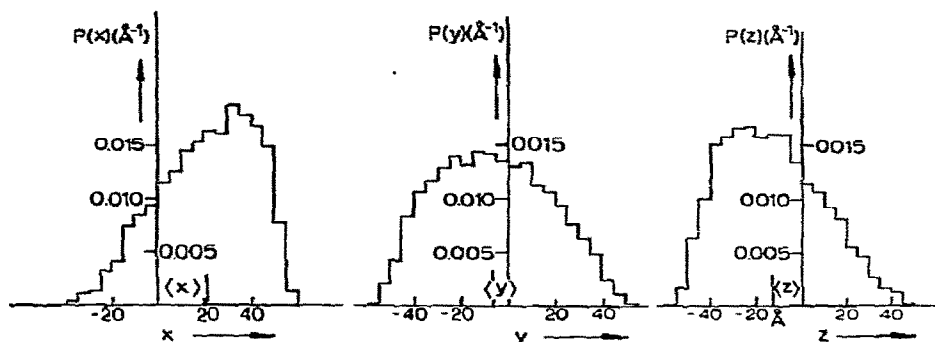


Fig. 3. Distribution $P(x)$, $P(y)$ and $P(z)$ of respectively the x , y and z component of the chain vector for a chain of 20 peptide units, expressed in the frame of reference of the first two skeletal bonds of the chain.

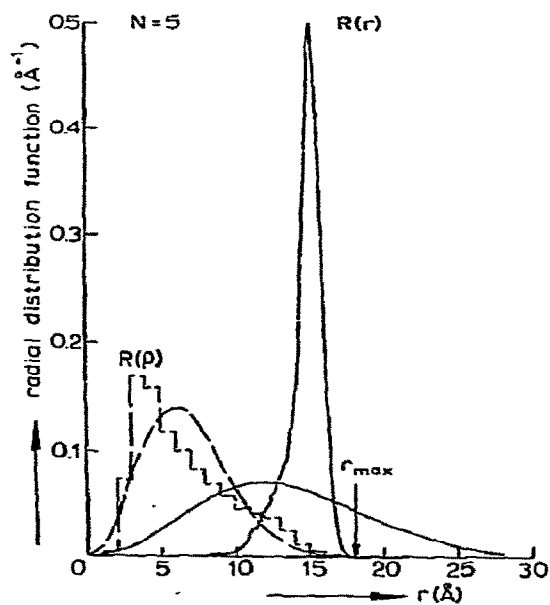


Fig. 4. The radial distribution function $R(r)$ for the chain vector and its gaussian approximation (full curves); further, the distribution $R(\rho)$ of Flory's ρ vector and its gaussian approximation (broken curves) both for a chain of 5 peptide units. r_{\max} indicates the length of the fully extended chain.

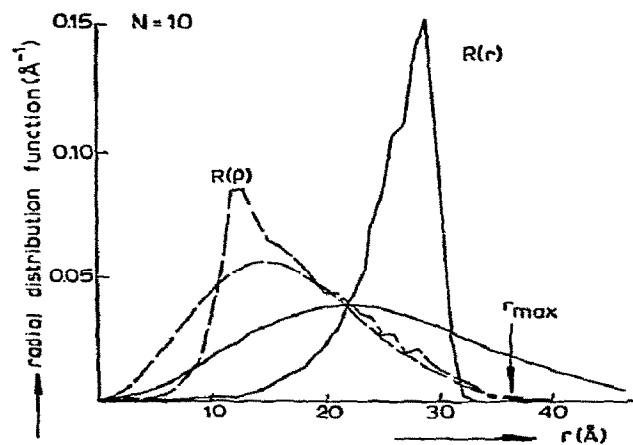


Fig. 5. Same as fig. 4, for a chain of 10 peptide units.

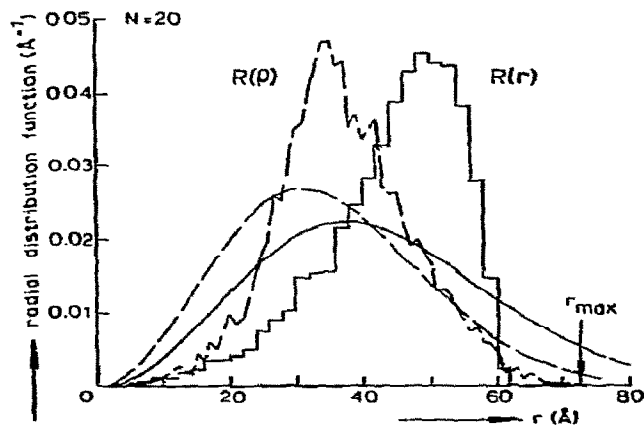


Fig. 6. Same as fig. 4, for a chain of 20 peptide units.

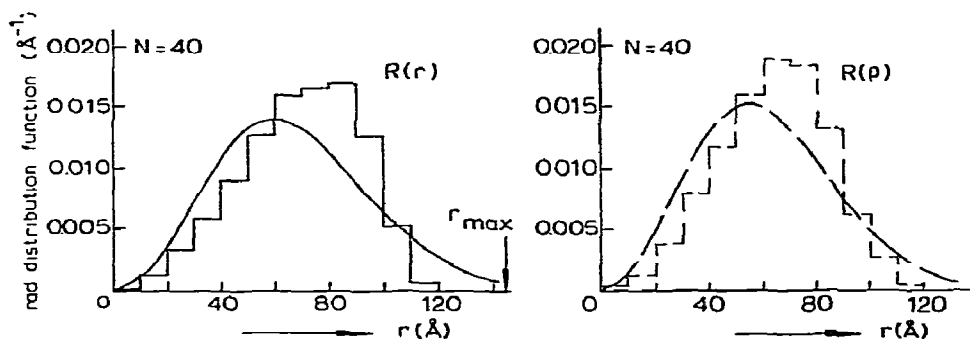


Fig. 7. Same as fig. 4, for a chain of 40 peptide units.

butions are closer together, but we still see systematic deviations in three regions of the distribution curves.

Deviation I: The actual probability of chains having a chain vector longer than about $2\langle r^2 \rangle^{1/2}$ is overestimated by the gaussian distribution, a well known deficiency of this distribution which is only valid for chains not extended too much.

Deviation II: The actual distribution exceeds the gaussian distribution in the region where r is somewhat larger than $\langle r^2 \rangle^{1/2}$. This hump found in the actual distribution over the gaussian distribution (see figs. 4–8) becomes smaller for longer chains, and it has almost vanished for $N = 80$.

Deviation III: The probability of a short chain vector ($r < \langle r^2 \rangle^{1/2}$) is overestimated by the gaussian distribu-

tion. This effect is very important for short chains and it has again almost vanished for chains of 80 peptide units. It has also been found for other types of polymers, such as polynucleotides [11], polymethylene and polysiloxanes [16] and it plays an important role in the formation of polymer rings. Actually, Delisi and Crothers [11] obtained a radial distribution function for a polynucleotide of 7 units, which looks very similar to our results for a chain of 20 peptide units.

Comparison of the actual distributions with more sophisticated treatments of the freely jointed chain ([11] ch. 8; [17]) is not expected to remove the observed deviations II and III, since Jernigan and Flory [18] showed the gaussian approximation to function very well as long as $r \ll Nl_u$.

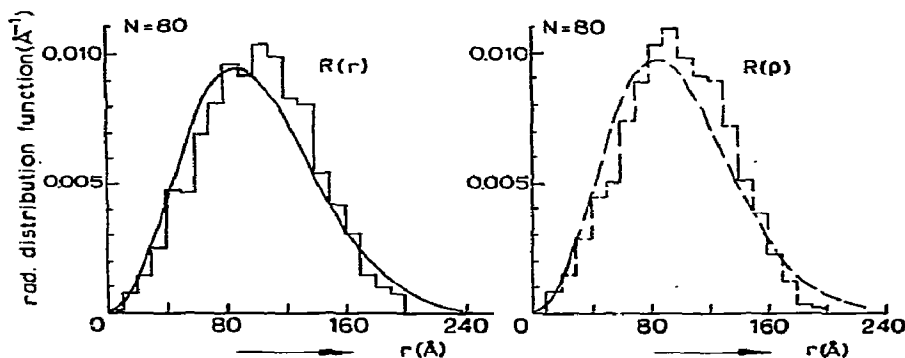


Fig. 8. Same as fig. 4, for a chain of 80 peptide units.

3.4. The vector ρ and its distribution

Now, we consider the distribution of the remote end of the chain relative to the terminus of the persistence vector, defining [1] a vector ρ according to

$$\rho = r - a \quad (3)$$

or

$$\rho = \begin{bmatrix} x - \langle x \rangle \\ - \langle y \rangle \\ |z - \langle z \rangle| \end{bmatrix}, \quad (4)$$

where x , y and z are the coordinates of the remote end of the chain in the reference frame of the first residue and $\langle x \rangle$, $\langle y \rangle$ and $\langle z \rangle$ the elements of the persistence vector.

The mean square value of ρ is then given by [1]

$$\langle \rho^2 \rangle = \langle r^2 \rangle - a^2. \quad (5)$$

After generation of the sets of conformations for each chain length the elements of the persistence vector are calculated. Then the elements of the vector ρ are calculated for each conformation according to eq. (4), and the resulting value for $\langle \rho^2 \rangle^{1/2}$ is given in table 1. These values for $\langle \rho^2 \rangle$ obey eq. (5).

The radial distribution function for the vector ρ , $R(\rho)$, is shown for various chain lengths in the figs. 4–8, together with the gaussian distribution for the value of $\langle \rho^2 \rangle$ (broken lines). For short chains ($N = 5$ and $N = 10$, chains with respectively 16 and 31 backbone atoms) the distribution $R(\rho)$ exhibits behavior very much closer to the gaussian approximation than the distribution for the chain vector $R(r)$. For longer chains, however, the distribution $R(\rho)$ behaves only slightly more like the gaussian approximation than the distribution of the chain vector $R(r)$. Therefore, the vector ρ could be useful for ring closures of small rings, whereas for larger rings the use of ρ instead of r would not result in great improvements.

After completion of this manuscript a paper appeared by Premilat and Hermans, Jr. [19], in which similar results are reported for the radial distribution function $W(r)$ for polypeptide chains of 5, 10 and 20 peptide units. These results, however, are not discussed in terms of Flory's persistence vector but in terms of the Kratky–Porod wormlike chain model.

Acknowledgement

The author wishes to thank Professor H.A. Scheraga for the inspiration which led to the present work.

References

- [1] P.J. Flory, Proc. Natl. Acad. Sci. 70 (1973) 1819.
- [2] F.Th. Hesselink, T. Ooi and H.A. Scheraga, Macromolecules 6 (1973) 541.
- [3] J. Mazur and F.L. McCrackin, J. Chem. Phys. 49 (1968) 648.
- [4] H.E. Warvari, K.K. Knaell and R.A. Scott, J. Chem. Phys. 57 (1972) 1161.
- [5] S. Tanaka and A. Nakajima, Macromolecules 5 (1972) 708.
- [6] S.D. Stellman and P.J. Gans, Macromolecules 5 (1972) 720.
- [7] Z. Alexandrowicz and Y. Accad, Macromolecules 6 (1973) 251.
- [8] E.A. DiMarzio and F.L. McCrackin, J. Chem. Phys. 43 (1965) 539.
- [9] E.J. Clayfield and E.C. Lumb, Macromolecules 1 (1968) 133.
- [10] M. Fixman and R. Alben, J. Chem. Phys. 58 (1973) 1553.
- [11] C. Delisi and D.M. Crothers, Biopolymers 10 (1971) 1809.
- [12] P.J. Flory, Statistical mechanics of chain molecules (Interscience, New York, 1969) Ch. VII.
- [13] T. Ooi, R.A. Scott, G. Vanderkooi and H.A. Scheraga, J. Chem. Phys. 46 (1967) 4410.
- [14] D.A. Brant and P.J. Flory, J. Am. Chem. Soc. 87 (1965) 2791.
- [15] O. Kratky and G. Porod, Rec. Trav. Chim. 68 (1949) 1106.
- [16] M.S. Beevers and J.A. Semlyen, Polymer 13 (1972) 385.
- [17] H. Yamakawa, Modern theory of polymer solutions (Harper and Row, New York, 1971) p. 12.
- [18] R.L. Jernigan and P.J. Flory, J. Chem. Phys. 50 (1969) 4185.
- [19] S. Premilat and J. Hermans, Jr., J. Chem. Phys. 59 (1973) 2602.