A NOVEL VIRTUAL BOND SCHEME TO PROBE ORDERED AND RANDOM COIL CONFORMATIONS OF NUCLEIC ACIDS: CONFIGURATIONAL STATISTICS OF POLYNUCLEOTIDE CHAINS

R. MALATHI AND N. YATHINDRA[+]

Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Madras 600 025, India

ABSTRACT

A new two virtual bond scheme spanning the three chemical bonds, P–O5–C5′–C4′ and C4′–C3′–O3′–P on either side of the C4′ atom of the repeating nucleotide unit has been developed to study the ordered as well as random coil conformational properties of polynucleotide chains. An important feature of this scheme is its ability to effectively account for the main sources of flexibility in a polynucleotide chain and to incorporate the near neighbour (long range) conformational dependence that exists among them. Using this scheme, unperturbed end-to-end dimensions and persistence lengths of randomly coiling polynucleotides have been computed and are in excellent agreement with the experimentally determined values. The results most importantly, demonstrate that the random coil consists of a large proportion of compactly stacked helices (A or distorted A-form) characterised by $g^- g^-$ and C3′ *endo* conformations for the phosphodiester and the nucleotide respectively interspersed by small proportions of loosely wound helical segments characterised by tg^- and C2′ *endo* geometries. It is argued that the occurrence of other conformational blends dictated by near-neighbour (long range) interactions in small percentage decreases the chain dimensions. These results are in sharp contrast to all the earlier interpretations which invoke a high percentage of extended tg^- conformations.

INTRODUCTION

THEORETICAL as well as experimental studies on nucleic acid constituents have established that rotations around the various chemical bonds of the repeating nucleotide unit are greatly restricted[1]. Detailed insights concerning their dependence on the rotations about adjacent and near-neighbour bonds have also been obtained more recently[2-6]. These together with the results of x-ray structure of yeast tRNA[phe] have demonstrated[7-9] that the main source of variety in nucleic acid chains arises predominantly from rotations around the P–O bonds of the internucleotide phosphodiester, the exocyclic C4′–C5′ bonds and the sugar ring pucker. In other words, rotations around the two adjacent C5′–C4′ and C4′–C3′ (ψ and ψ') and O5′–P and P–O3′ (ω and ω') bonds (Fig. 1) mainly dictate the conformations of nucleic acid chains both in the ordered helical as well as loop and bend structures. The rotations around the two remaining C5′–O5′ and C3′–O3′ bonds (ϕ and ϕ') have been shown to exist predominantly in the preferred *trans* conformational domain. The correlations between ψ and ψ' (P) and ω and ω' and also the dependence of the latter two on the former, viz., ψ' and ω' and ω and ψ are also recognised. These stereochemical considerations are strongly suggestive of treating the nucleotide unit in terms of two blocks, comprising three bonds each, viz., P–O5–C5′–C4′ and C4′–C3′–O3′–P spanning the four atoms P to C4′ on the 5′-end and C4′ to P on the 3′-end of the repeat-

ing nucleotide unit (Fig. 1). The virtual bond lengths P–C4′ and C4′–P calculated from crystal structures of 5′-nucleotides, 3′-nucleotides, di- and oligo-nucleotides and yeast tRNA[phe] show that they are around 3·9 A and 3·8 A respectively[10]. This "blocked nucleotide" has some similarities to the situation in peptides[11].

Recognition of these stereochemical features inherent to the nucleotide unit readily suggests that the entire polynucleotide chain can be regarded as a sequence of two virtual bonds P–C4′ and C4′–P (Fig. 1) per repeating nucleotide and its conformation being governed by two pairs of rotation angles (ψ, ψ') *within* the repeating unit and (ω', ω) *between* the repeating units. This situation greatly simplifies the analysis of ordered helical as well as random coil conformations of nucleic acid chains. The present scheme of virtual bonds is unique in that, it permits as a natural follow up (i) implicit consideration of variation of the sugar pucker, in contrast to earlier schemes[12-15] where independent treatment was necessary, (ii) correlation between the nature of the sugar pucker and C4′–C5′ bond orientation (indirect estimate of base sequence effect) and (iii) near-neighbour (long range) correlation between ψ' and ω' and ψ and ω.

Using this virtual bond scheme, we report here unperturbed end-to-end dimensions and also persistence lengths of polynucleotide chains in excellent agreement with experimentally determined values[16-21]. Results of our studies show that the random coil conformation of polynucleotides exists predominantly in the compactly stacked A-form (or distorted A-form) possessing the preferred $g^- g^-$ conformation for the phosphodiester concomitant with the C3′–

+ For correspondence.

endo geometry for the nucleotide backbone in sharp contrast to earlier models[12-17] which invoked very high percentage of extended and unstacked backbone conformation.
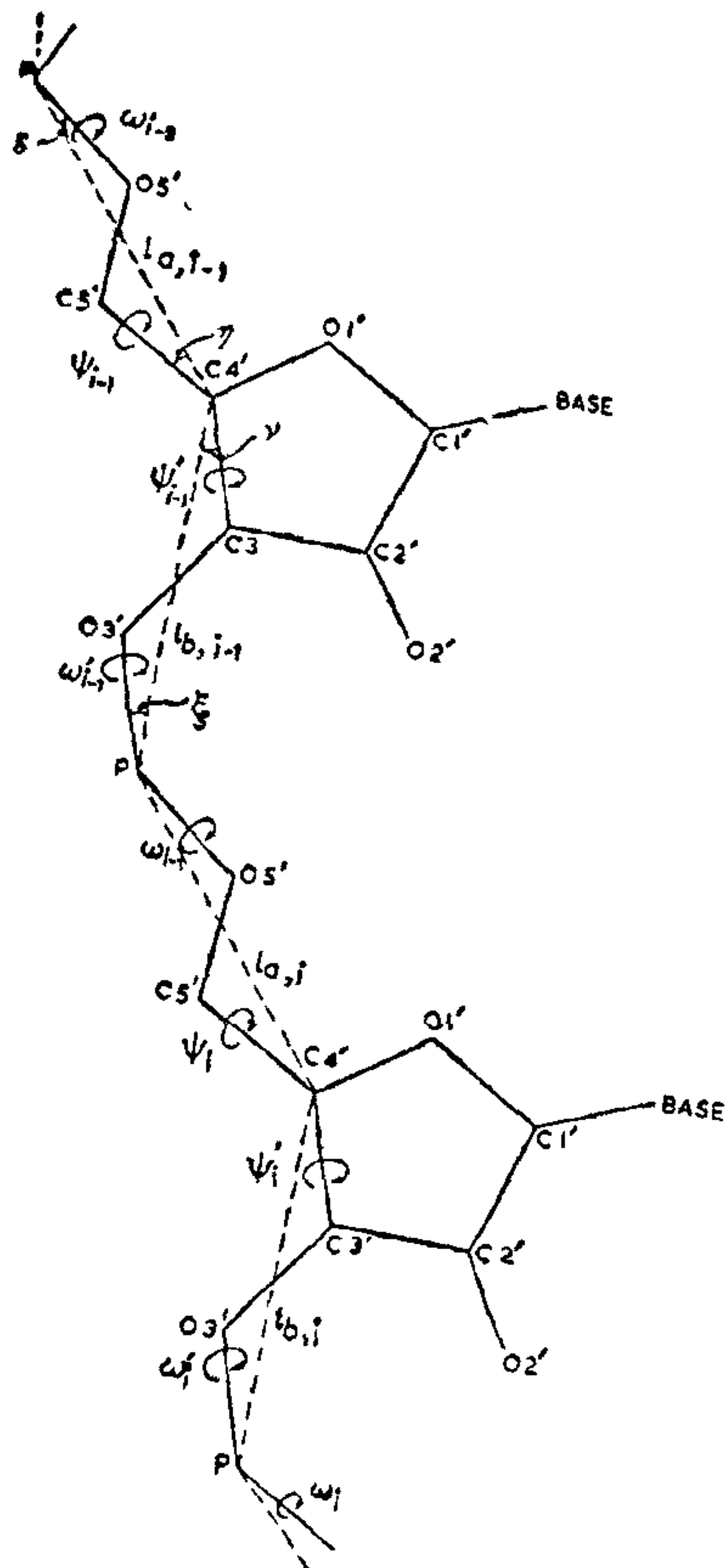

FIG. 1. Schematic representation of the new two virtual bond scheme in a section of a polynucleotide chain. Virtual bonds spanning the atoms P to C4' on the 5'-side and C4' to P on the 3'-side are indicated by dotted lines l_a and l_b respectively. The rotations (ϕ, ϕ') around the two C–O bonds are assigned the most preferred *trans* conformation. The pairs of rotations (ψ, ψ') and (ω', ω) completely describe the spatial configuration of a polynucleotide chain. The long range conformational relations between (ω, ψ) and (ψ', ω') can easily be seen. The virtual bond scheme is reminiscent of that in peptides.

THEORETICAL TREATMENT

Virtual Bond Parameters

According to the above scheme of virtual bonds shown in Fig. 1 by dotted lines, the magnitude of

their lengths l_a and l_b are independent of the rotation angles (ψ, ψ') and (ω', ω), but depend on the values of ϕ and ϕ'. They have been assigned their preferred values of 180° and 210° consistent with theoretical, n.m.r. and x-ray studies[1]. The bond lengths and bond angles used here are the same as reported earlier[2]. The bond lengths and other parameters associated with the virtual bonds are given in Table I.

TABLE I

Virtual bond parameters for the blocked nucleotide

Virtual bond lengths		Geometrical Parameters			
l_a (Å)	l_b (Å)	$\eta°$	$\nu°$	$\xi°$	$\delta°$
3·97	3·85	23·73	28·22	20·46	14·73

Mean Square Dimensions

The scheme of virtual bonds described above permits one to describe the spatial configuration of a poly-nucleotide chain in terms of the orientation of the successive alternating virtual bonds l_a and l_b. The relative orientation of the virtual bonds are determined by successive (ψ, ψ') and (ω', ω) values. The mean square end-to-end distance of such a chain is given by

$$\langle r^2 \rangle_0 = \sum_{i=1}^{x} \sum_{j=1}^{x} \langle (l_{a,i} + l_{b,i}) \cdot (l_{a,j} + l_{b,j}) \rangle$$

$$\langle r^2 \rangle_0 = x l_a^2 + x l_b^2 + 2 x l_a^T \langle T_a \rangle \, l_b + 2 \sum_{i<j} (l_a^T + l_b^T)$$

$$\cdot (l_a + l_b) \tag{1}$$

where l_a and l_b are the two virtual bond lengths constituting the repeating unit and x is the degree of polymerisation. The value of $\langle r^2 \rangle_0$ can easily be obtained by resorting to matrix techniques. When the rotations in the successive residues are independent of each other or in the absence of near-neighbour correlation between (ω', ψ') and (ω, ψ), the above can be expressed following the methodology of Flory[22] as

$$\langle r^2 \rangle_0 = 2 \, [1\ 0\ 0\ 0\ 0] \, [G_a \, G_b]^x \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{2}$$

where G_a and G_b are the generator matrices of the order 5 × 5 defined by

$$G_a = \begin{bmatrix} 1 & l_a^T \langle T_a \rangle & l_a^2/2 \\ 0 & \langle T_a \rangle & l_a \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$G_b = \begin{bmatrix} 1 & l_b^T & \langle T_b \rangle & l_b^2/2 \\ 0 & \langle T_b \rangle & l_b \\ 0 & 0 & 1 \end{bmatrix}$$

$\langle T_a \rangle$ and $\langle T_b \rangle$ are the statistical mechanical averaged matrices averaged over the rotational states of (ψ, ψ') and (ω', ω) respectively. T_a and T_b are the transformation matrices which transform any vector in the coordinate system of the virtual bonds l_a and l_b into their preceeding one.

Using the above expression the characteristic ratio of an infinitely long chain is expressed as

$$\langle r^2 \rangle_0 / N\overline{L}^2$$

$$= \underset{N \to \infty}{Lt} \frac{2}{N\overline{L}^2} [1\ 0\ 0\ 0\ 0]\ [G_a\ G_b]^x \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

where $N = 2x$, represents the total number of virtual bonds in the chain and $\overline{L}^2$ is the average squared virtual bond length.

Consideration of interdependency between the rotations of the neighbouring residues, namely, (ψ', ω') and (ω, ψ), the above expression reduces to

$$\frac{\langle r^2 \rangle_0}{N\overline{L}^2} = \underset{N \to \infty}{Lt} \frac{2}{N\overline{L}^2} [1\ 0\ 0\ 0\ 0]\ G_a\ [G_{ba}]^{x-1}\ G_b$$

where G_{ba} is a 5×5 matrix which represents (ω, ψ) interdependence and is given by

$$G_{ba} = \begin{pmatrix} 1 & l_a^T \langle T_a \rangle + l_b^T \langle T_b T_a \rangle \frac{l_a^2}{2} + l_b^T \langle T_b \rangle l_a + \frac{l_b^2}{2} \\ 0 & \langle T_b T_a \rangle & \langle T_b \rangle l_a + l_b \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

Similarly, expression for G_{ab} correlating (ψ', ω') dependence can be obtained. However, in order to completely consider the inter-relation between the rotations of the successive nucleotide residues, transformation matrices have to be averaged over all the combined rotational states of $(\psi\psi' \omega'\omega\ \psi\psi')$. For a polynucleotide chain, described by the above scheme of double virtual bonds, the persistence vector P is given by

$$P = \underset{x \to \infty}{Lt} [E\ O]\ [A_a\ A_b]^x \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (5)$$

where E is the identity matrix of order 3×3 and O

is a null matrix of required order. A_a and A_b are 4×4 matrices expressed as

$$A_a = \begin{bmatrix} \langle T_a \rangle & l_a \\ O & 1 \end{bmatrix} \text{ and } A_b = \begin{bmatrix} \langle T_b \rangle & l_b \\ O & 1 \end{bmatrix} \quad (6)$$

The components of the persistence vector, namely, $\langle x \rangle, \langle y \rangle$ and $\langle z \rangle$ are given by (1, 4), (2, 4) and (3, 4) elements of the product matrix obtained from (5). The persistence length is given by the component $\langle x \rangle$ of the persistence vector P.

The transformation matrices T_a and T_b are evaluated using standard procedures.

RESULTS AND DISCUSSION

Assignment of Statistical Weights

Two rotational states have been assigned for ψ' with values of $\psi' = 86°$ (g^+) and $\psi' = 156°$ (t) corresponding to the C3'-endo (3_E) and C2'-endo (2_E) sugar puckers in accordance with overwhelming experimental[1],[23-25] and theoretical evidence[1]. However, it is possible to assign other states along the pseudo-rotation path such as O (1') endo, etc., in this scheme. Similarly, three states corresponding to the well known g^+, t and g^- conformations have been considered for rotation around C4'-C5' bond (ψ). While there is a strong preference for the g^+ rotamer, possibilities of a small proportion of t and g^- conformers have also been suggested[1],[23-25]. Further n.m.r. studies[23-25] find that there is a tendency for increased t populations when there is bias towards 2_E sugar pucker. An analysis[26] of x-ray structures has earlier led to similar dependence of ψ and ψ'. This has been taken into account in assigning the statistical weights for ψ and ψ'.

The favoured phosphodiester $(\omega'\omega)$ rotations analysed earlier as a function of sugar pucker (ψ') as well as exo-cyclic C4'-C5' (ψ) torsions showed marked restriction for gauche (g^+ and g^-) rotations of ψ compared to trans for both 3_u and 2_s sugar puckers. Further, P-O3' rotations (ω') are greatly influenced by changes in the sugar pucker (ψ') preceding it while P-O5' torsions (ω) are influenced by C4'-C5' rotations (ψ) following it[a]. Similar correlations between ω' and ψ' and ω and ψ have also been established with regard to stacked conformations from helical parameter calculations[27]. This near-neighbour bond correlations arise because of the proximity of the P-O5' (ω) and C5'-C4' (ψ) bonds and P-O3' (ω') and C4'-C3' (ψ') bonds concomitant with their parallel orientations induced by the preferred trans character of the bonds which link them. These correlations have to be considered while assigning the statistical

weights to P–O bond rotations when the interdependence of rotations in the neighbouring residues are incorporated.

Characteristic ratio (C_f) evaluated under the approximation of free rotation, by assigning equal statistical weights to all the rotational states has a value of 1·4. C_f values of 1·5 and 1·7 have been obtained for C3'-endo and C2'-endo polynucleotide chains The higher values of C_f compared to 1·1 obtained by single virtual bond treatment are indicative of the restrictions imposed by the sugar pucker.

Values of characteristic ratios (C_∞) evaluated in the absence of free rotation and long range dependency of ψ', ω' and ω, ψ are shown in Table II, along with the various statistical weight combinations for ψ, ψ' and ω', ω. The most favoured g^+ state has been assigned the maximum statistical weight followed by t. The g^- rotational state for ψ which represents a high energy conformation due to unfavourable electrostatic interactions between the adjacent phosphate groups, especially for the 3_g sugars is not considered. For similar reasons, the phosphodiester conformations g^+g^+, g^+g^-, g^-g^+ and tt are not included. Phosphodiester conformations g^-g^- and tg^- are assigned relatively higher statistical weights compared to g^-t and tg^+ because of their higher probability of occurrence in accordance with theoretical calculations[28]. It should be mentioned here that the upfield shifts in the P[11] resonances observed at high temperatures in certain oligonucleotides have been interpreted[29] as due to the tg^- phosphodiester.

It is clear from Table II that values of characteristic ratios (C_∞) depend on the nature of the sugar pucker and phosphodiester conformation. A small fraction of chain-reversing g^-t or the tg^+ phosphodiester sharply lower the chain dimensions. Value of C_∞ varies from 10–20 depending upon the combination of phosphodiester, sugar and C4'–C5' bond conformations. The experimental value of $C_\infty \simeq 18$ is

realised when the contribution from g^-g^- is over 70% concomitant with similar contribution of 3_g sugar puckers. Increase in 2_E and tg^- phosphodiester population gradually increases the chain dimensions beyond the experimentally determined value, due to the extended nature of 2_g sugar pucker and tg^- phosphodiester. This clearly demonstrates that the random coil conformation of polynucleotides exists predominantly in the stacked helical states similar to A-form or distorted A form along with a small proportion of loosely wound helical segments corresponding to tg^- phosphodiester and a C2'-endo sugar pucker. This is in sharp contrast to earlier results[12-17] obtained from a different two virtual bond and single virtual bond treatment wherein a very large percentage of extended tg^- conformation was invoked to obtain experimental agreement of C_∞.

Characteristic ratio (C_x) computed as a function of the degree of polymerisation (X) is shown in Fig. 2. It is clear that saturation value is reached for X values around 2^7–2^8.
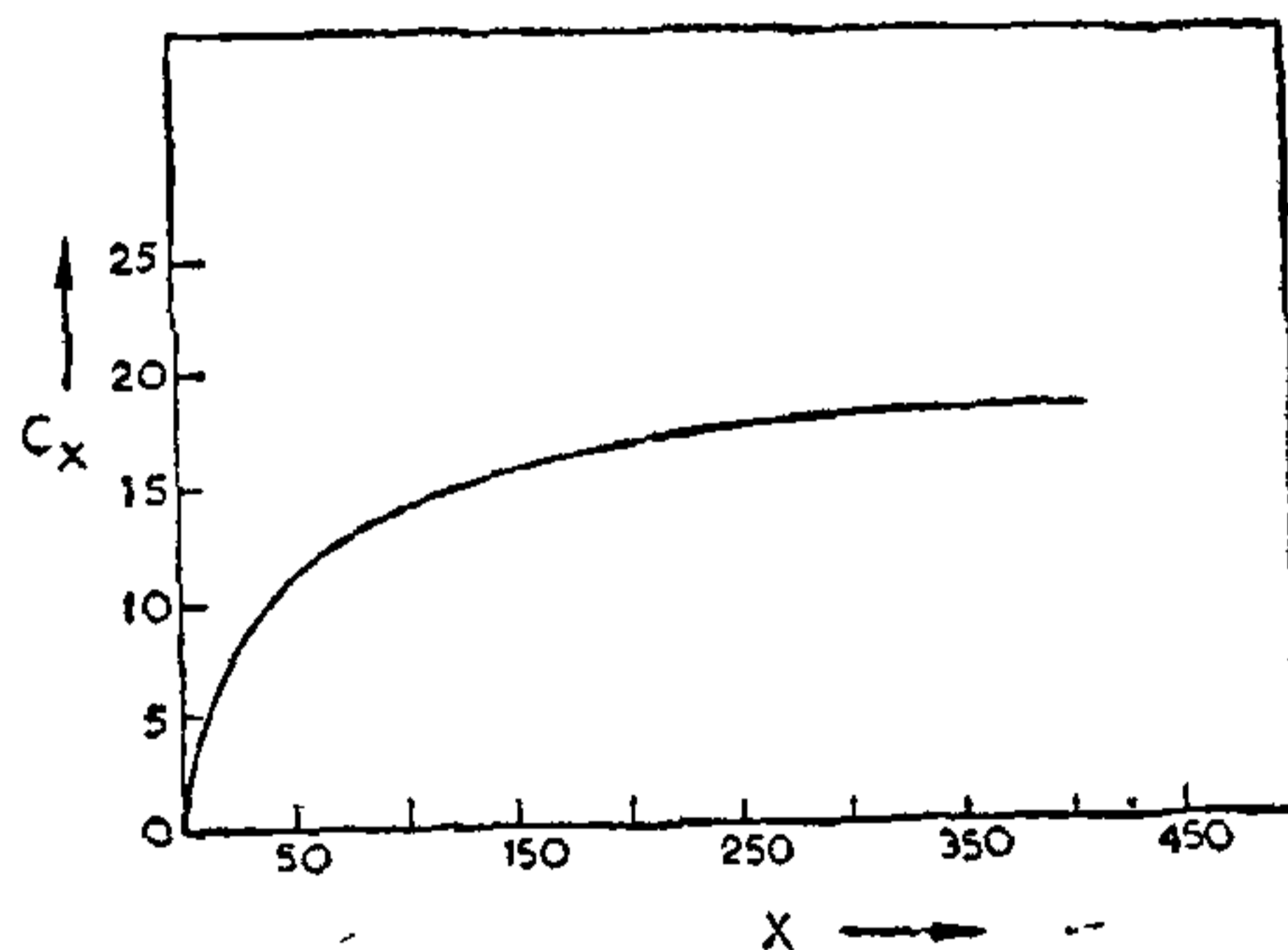

FIG. 2. Plot of characteristic ratio C_x computed as a function of degree of polymerisation (X) corresponding to III set of statistical weight in Table II.

TABLE II

Characteristic ratios C_∞ obtained as a function (ψ, ψ') and (ω', ω) rotational states

Statistical weights for the rotational states of (ψ, ψ')				Statistical weights for the rotational states of (ω', ω)				C_∞
$g^+/3E$	$g^+/2E$	$t/3E$	$t/2E$	tg^+	g^-t	tg^-	g^-g^-	
0·8	0·2	..	..	..	..	0·25	0·75	16·38
0·75	0·25	..	..	..	..	0·25	0·75	18·1
0·75	0·25	..	..	..	..	0·2	0·8	17·63
1	..	..	..	..	..	0·3	0·7	14·65
0·7	0·3	..	..	0·001	0·001	0·3	0·698	14·64
0·6	0·4	..	..	0·001	0·001	0·3	0·698	16·74
0·73	0·24	0·01	0·02	0·02	0·02	0·2	0·76	10·5

Persistence lengths of about 55–65 A have been obtained by the use of same statistical weight combinations which reproduce experimentally observed C_∞. This is again in agreement with the experimentally observed[30] value of 75 A, which is not corrected for solvent effects. It may be mentioned that previous calculations[16,17] using different virtual bond schemes led to low values of persistence lengths. Incorporation of interpendence of rotations in the neighbouring residues (long range correlation) invokes tedious calculations because of statistical mechanical averaging of a very large number of combined rotational states. The results of such calculations are to be reported elsewhere[31].

CONCLUSION

Representation of the nucleotide repeat in terms of the above two virtual bond scheme greatly simplifies the analysis and interpretation of random coil conformations of polynucleotides consistent with the preferred rotational states. The distinct advantage of the new scheme lies in its ability to include simultaneously various combinations of sugar puckers and long range conformational correlations. The scheme can also be effectively utilised for the analysis of ordered helical, loop and bend structures.

ACKNOWLEDGEMENT

1. For a detailed discussion and references see Sundaralingam, M., In *Structure and Conformation of Nucleic Acids and Nucleic Acid Protein Interactions*, Univ. Park Press, Eds. Sundaralingam, M. and Rao, S. T., 1974, p. 487.

2. Yathindra, N. and Sundaralingam, M., *Ibid.*, 1975, p. 649.

3. — and —, *Biopolymers*, 1974, 14, 2387.

4. Perahia, D., Pullman, B., Vasilescu, D., Cormillon R. and Broch, H., *Biochim. Biophys. Acta*, 1977, 473, 244.

5. Broyde, S. and Hingerty, B., *Nucl. Acids. Res.*, 1979, 6, 2165.

6. Yathindra, N., In *Structure, Conformation, Function and Evolution*, Pergamon Press, Eds. Srinivasan, R., Subramanian, E. and Yathindra, N. 1980, p. 379.

7. Stout, C. D., Mizuno, H., Rao, S. T., Swaminathan, S., Rubin, J., Brennan, T. and Sundaralingam, M., *Acta Cryst.*, 1978, B34, 1529.

8. Hingerty, B. E., Brown, R. S. and Jack, A., *J. Mol. Biol.*, 1978, 124, 58.

9. Holdbrook, S. R., Sussman, J. L., Warrant, R. W. and Kim, S. H., *Ibid.*, 1978, 123, 631.

10. Arangarajan, M., *M.Sc. Thesis*, Madras University, Madras, June 1979.

11. Sundaralingam, M., In *Structure, Conformation, Function and Evolution*, Pergamon Press, Eds. Srinivasan, R., Subramanian, E. and Yathindra, N., 1980, p. 259.

12. Olson, W. K. and Flory, P. J., *Biopolymers*, 1972, 11, 1.

13. —, *Ibid.*, 1975. 14, 1797.

14. —, *Macromolecules*, 1975.

15. Yevich, R. and Olson, W. K., *Biopolymers*, 1979, 17, 113.

16. Hingerty, B. and Broyde, S., *Nucl. Acids. Res.*, 1978, 5, 3249.

17. Broyde, S. and Hingerty, B., In *Stereodynamics of Molecular Systems*, Pergamon Press, Ed. Sarma, R. H., 1979 (Private Communication).

18. Tewari, R., Nanda, R. K. and Govil, G., *Biopolymers*, 1974, 13, 2015.

19. Eisenberg, H. and Felsenfeld, G., *J. Mol. Biol.*, 1967, 30, 17.

20. Inners, L. D. and Felsenfeld, G., *Ibid.*, 1970, 50, 373.

21. Achter, E. K. and Felsenfeld, G., *Biopolymers*, 1971, 10, 1625.

22. Flory, P. J., *Statistical Mechanics of Chain Molecules*, Interscience, N.Y., 1969.

23. Sarma, R. H., Mynott, R. J., Wood, D. J. and Hruska, F. E., *J. Am. Chem. Soc.*, 1973, 95, 6457.

24. Davies, D. B. and Danyluk, S. S., *Biochemistry*, 1974, 13, 4417.

25. Evans, F. E. and Sarma, R. H., *Nature*, 1976, 263, 567.

26. Yathindra, N. and Sundaralingam, M., *Biopolymers*, 1974, 13, 2061.

27. Yathindra, N. and Sundaralingam, M., *Nucl. Acids. Res.*, 1976, 3, 729.

28. Srinivasan, A. R., Yathindra, N., Rao, V. S. R. and Prakash, S., *Biopolymers*, 1980, 19, 165.

29. Patel, D. J., *Ibid.*, 1976, 15, 533.

30. Mngot, F., Jorcano, J., Acuna, M. and Davila, C., *Biochim. Biophys. Acta*, 1976, 413, 315.

31. Malathi, R. and Yathindra (submitted).

32. Olson, W. K., *Macromolecules*, 1980, 13, 721.