Configurational Statistics of Polynucleotide Chains. A Single Virtual Bond Treatment

Wilma King Olson

Department of Chemistry, Douglass College, Rutgers, the State University, New Brunswick, New Jersey 08903. *Received December 9, 1974*

ABSTRACT: A simplified single virtual bond scheme has been developed for the calculation of mean-square unperturbed dimensions in polynucleotide chains. As a consequence of the structural rigidity of the sugar residues in the chain, it is possible to represent the six chemical bonds comprising the chain backbone repeating unit by a single virtual bond (connecting successive phosphorus atoms). The mutual orientation of a pair of adjoining virtual bonds is determined by the angles of rotation about the phosphodiester bonds adjoining intervening phosphorus atoms and is independent of the orientation of all other virtual bonds in the chain. Computed values of chain dimensions based on the single virtual bond scheme are comparable to those calculated previously using a two virtual bond model which permits rotational flexibility in the sugar moieties of the chain.

Recent experimental studies of conformation in low molecular weight nucleic acid analogs as well as numerous theoretical estimates of conformational energy suggest that the polynucleotide chain backbone is composed of repeating units of relative structural rigidity.[1-3] As a consequence of geometric constraints imposed by the skeletal structure the rotations of at least four of the six chemical bonds comprising the sugar–phosphate repeat unit are restricted to narrow regions of conformation space. The rotations ϕ' and ϕ about the C–O bonds of the polynucleotide (see Figure 1) are each confined to a single conformation range centered approximately about the planar trans orientation (here defined as 0° rotation). The ψ' rotation about the $C_{3'}$–$C_{4'}$ bond is found in two conformation ranges only, the trans and the gauche$^-$ (where $\psi' \sim 240$ or $-120°$), owing to the presence of the pentose ring in the chain backbone. The former range of ψ' is observed when the ring puckering is $C_{2'}$-endo and the latter when the puckering is $C_{3'}$-endo. The ψ rotation about bond $C_{4'}$–$C_{5'}$ is threefold with minima occurring in the normal trans, gauche$^+$, and gauche$^-$ domains. The gauche$^-$ minimum, however, is almost exclusively favored over the trans and gauche$^+$ ranges, which thus will be omitted in the following treatment.

The restrictions to the internal rotations ω' and ω about the two P–O bonds of each polynucleotide repeating unit are much less severe than those about the four bonds discussed above. The phosphodiester rotations are threefold with minima occurring in the usual trans, gauche$^+$, and gauche$^-$ locations and, together with the sugar pucker, they thus provide the sources of flexibility in the polynucleotide backbone. The polynucleotide may then be regarded, as Sundaralingam has suggested,[1,5] as one of two kinds of chains, $C_{3'}$-endo or $C_{2'}$-endo, in which "rigid nucleotide moieties" are connected by flexible phosphodiester links. This circumstance, as we shall show, leads to considerable simplification in the theoretical treatment of configuration-dependent properties of the polynucleotide chain.

In this paper we develop a virtual bond scheme based on the rigidity of the sugar–phosphate repeating unit for the calculation of the unperturbed dimensions of the polynucleotide chain. We represent the six skeletal bond repeating unit of the polymer backbone by a single virtual bond of fixed length. This model presents a further simplification from the two virtual bonds per six chemical bond repeat unit scheme that has been employed previously in the treatment of configuration-dependent properties of polynucleotides.[13] In addition, the representation of the nucleotide repeating unit by a single virtual bond reduces the analysis of polynucleotide chains to a treatment similar to those developed previously for polypeptides[6] and polysaccharides.[7-9] In the latter instances a single virtual bond is rigidly fixed within the structure of the peptide or saccharide unit, respectively.

The Virtual Bond Scheme

Rotational States. Description of the conformation ranges accessible to the C–O and C–C rotations in the polynucleotide in terms of discrete rotational isomeric states is both convenient and well founded for bonds such as these having distinct conformational minima separated by energy barriers substantially exceeding RT.[10] The conformational ranges accessible to the ϕ, ϕ', and ψ rotations may then be replaced by single values centrally located within their respective confines. The rotational isomeric states chosen here, $\phi = 0°$, $\phi' = 35°$, and $\psi = 240°$, are consistent with the ranges of these angles observed in X-ray, NMR, and theoretical analyses.[2] The pentose ring pucker as described by the ψ' rotation angle is also represented here by a single rotational isomeric state. In the $C_{2'}$-endo form when this atom is displaced toward atom $C_{5'}$ and above a plane described by the four remaining ring atoms, the ψ' angle is fixed here at 317° and in the $C_{3'}$-endo pucker where atom $C_{3'}$ is similarly displaced, the ψ' rotation is taken to be 266°. The $C_{3'}$-exo and $C_{2'}$-exo conformers in which the atom of note is displaced away from atom $C_{5'}$ and below a plane described by the four remaining ring atoms are very similar to the $C_{2'}$-endo and $C_{3'}$-endo forms, respectively, as evidenced from the value of ψ'. In the calculations that follow the ψ' rotation is fixed at 328° for the $C_{3'}$-exo pucker and at 278° for the $C_{2'}$-exo pucker. These four values of ψ' are based on an ideal ribose geometry described previously[11] in which four atoms of the pentose ring are precisely coplanar and the fifth puckered atom is displaced from the plane exactly 0.5 Å.

Assignment of the ϕ, ϕ', and ψ angles to their rotational isomeric state values and the ψ' angle to one of the four possibilities open to it fixes the distance that separates successive phosphorus atoms in the chain. It is appropriate then to introduce a virtual bond vector **v** joining these atoms as shown by the heavy solid lines in Figure 2. The length v of the virtual bond vector depends upon the value assigned to the angle ψ' as detailed below. The polynucleotide is then regarded as a sequence of identical virtual bonds joining the phosphorus atoms in the chain.

The mutual orientation of any two successive virtual bonds is determined by the pair of rotation angles ω' and ω about the phosphodiester bonds adjoining the intervening phosphorus atom. A given configuration of the chain is completely determined by specifying the values for every such pair of angles in the chain. Moreover, as a conse-

Table I
Geometrical Parameters of the Virtual Bond Model

Ring pucker	ψ', deg	v, Å	κ, deg	Φ, deg	ζ, deg	τ, deg	Ψ, deg
$C_{3'}$-endo	266	5.63	43	83	48	31	114
$C_{2'}$-exo	278	5.89	41	79	44	29	121
$C_{2'}$-endo	317	6.60	32	78	27	28	150
$C_{3'}$-exo	328	6.69	30	81	21	31	160

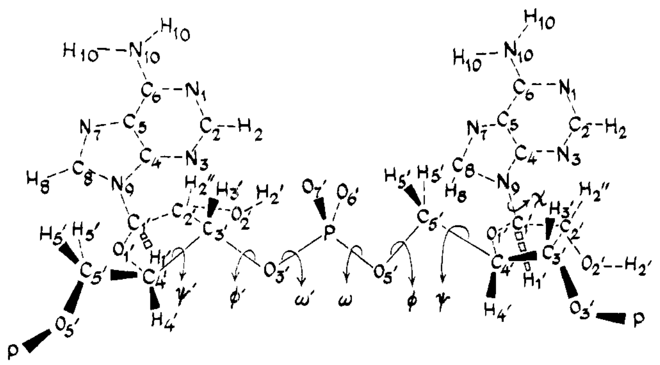

Figure 1. Section of an extended polyribonucleotide chain showing chain atoms and rotation angles (nomenclature introduced in ref 4). The rotation angle convention adopted here (trans = 0°) differs from that described in ref 4.

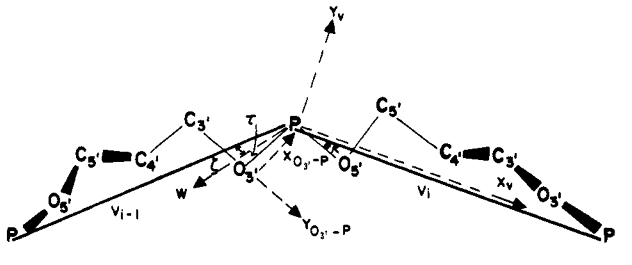

Figure 2. Section of a polynucleotide chain showing virtual bonds (heavy lines) and coordinate axes (dashed lines) of the chain backbone. The parameters **w**, ζ, τ, and κ are defined in the text.

quence of the large distances separating the successive phosphorus atoms in the chain, the rotation angle pair governing the mutual orientation of adjacent virtual bonds is independent of the orientation of all other virtual bonds in the chain. Statistical mechanical treatment of the random coil configuration of the chain is greatly simplified by this circumstance (cf. the following).

Virtual Bond Length. Calculation of the phosphorus-phosphorus distances (i.e., the virtual bond lengths) in the polynucleotide is readily accomplished using matrix methods. For this purpose a right-handed Cartesian coordinate system is defined for each chemical bond of the chain repeating unit as follows. The x axis is parallel to the direction of the chemical bond and the y axis is located in the plane defined by the chemical bond and its predecessor in the chain as shown in Figure 2 for chemical bond $O_{3'}$–P. The direction of the y axis is chosen to make an acute angle with the preceding bond vector. Premultiplication with the orthogonal matrix $\mathbf{T}_{AB,BC}$ (Θ^B,ψ) transforms the representation of a chemical bond vector in the B–C coordinate system into the coordinate system of its predecessor (A–B) in the chain. This matrix is a function of Θ^B, the angle between the A–B and B–C bond vectors, and of ψ, the angle of rotation about A–B. The matrix $\mathbf{T}_{AB,BC}$ is expressed as the matrix product

$$\mathbf{T}_{AB,BC}(\Theta^B, \psi) = \mathbf{X}(\psi - 180°)\mathbf{Z}(-\Theta^B) \tag{1}$$

where

$$\mathbf{Z}(\Theta) = \begin{bmatrix} \cos\Theta & -\sin\Theta & 0 \\ \sin\Theta & \cos\Theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

and

$$\mathbf{X}(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{bmatrix} \tag{3}$$

This transformation involves a rotation of the reference frame through an angle $-\Theta^B$ about the z axis affixed to bond BC followed by a rotation of $\psi - 180°$ about the x axis of BC, which then coincides with the x axis of AB.

The matrix product $\mathbf{T}_{A-E}$ of the form

$$\mathbf{T}_{A-E} = \mathbf{T}_{AB,BC}\mathbf{T}_{BC,CD}\mathbf{T}_{CD,DE} \tag{4}$$

then serves to transform a vector in the coordinate system of chemical bond DE into that of bond AB. The vector **v** connecting successive phosphorus atoms is thus given in the coordinate system of chemical bond P–$O_{5'}$ by the sum of skeletal bond vectors as

$$\mathbf{v} = \mathbf{l}_{P-O_{5'}} + \mathbf{T}_{P-C_{5'}}\mathbf{l}_{O_{5'}-C_{5'}} + \mathbf{T}_{P-C_{4'}}\mathbf{l}_{C_{5'}-C_{4'}} + \mathbf{T}_{P-C_{3'}}\mathbf{l}_{C_{4'}-C_{3'}} + \mathbf{T}_{P-O_{3'}}\mathbf{l}_{C_{3'}-O_{3'}} + \mathbf{T}_{P-P}\mathbf{l}_{O_{3'}-P} \tag{5}$$

Each bond vector **l** in this expression is represented in its own coordinate system and the serial products of transformation matrices, as defined in eq 4, effect the representation of each chemical bond vector into the coordinate system of bond P–$O_{5'}$. The virtual bond **v** is then the sum of the six chemical bond vectors and the interatomic distance between the phosphorus atoms (or virtual bond length) is given by $(\mathbf{v}\cdot\mathbf{v})^{1/2}$. Values of the virtual bond lengths thus calculated for the four kinds of polynucleotide chains ($C_{3'}$-endo, $C_{2'}$-exo, $C_{2'}$-endo, $C_{3'}$-exo) are listed in Table I along with the values of the parameters κ, Φ, ζ, τ, and Ψ (cf. the following) that relate the directions of the virtual bonds and the phosphodiester bonds of the polymer. In these computations the bond lengths and valence bond angles of the chain backbone were taken from the compilation of Arnott.[12]

Mean Square Dimensions. For the polynucleotide chain comprising x virtual bonds of length v the mean-square unperturbed end-to-end distance $\langle r^2\rangle_0$ is given by the average scalar product[10]

$$\begin{aligned} \langle r^2\rangle_0 &= \sum_{i=1}^{x}\sum_{j=1}^{x}\langle \mathbf{v}_i^T\cdot\mathbf{v}_j\rangle \\ &= xv^2 + 2\sum_{i<j}\langle \mathbf{v}_i^T\cdot\mathbf{v}_j\rangle \end{aligned} \tag{6}$$

Evaluation of the scalar products between the transpose of virtual bond vector $\mathbf{v}_i$ and virtual bond vector $\mathbf{v}_j$ appearing in eq 6 requires that all virtual bonds in the chain be expressed in the same coordinate system. For this purpose let a right-handed Cartesian coordinate system be affixed to each virtual bond in the chain (see Figure 2) with its x axis in the direction of the given bond and its y axis in the plane defined by the virtual bond and chemical bond $P-O_{5'}$. The positive direction of the y axis makes an acute angle with the x axis of the $P-O_{5'}$ coordinate system.

The matrix

$$\mathbf{T}_{\mathrm{P-O}_{5'},v}(\kappa, \omega - \Phi) = \mathbf{X}(\omega - \Phi)\mathbf{Z}(\kappa) \tag{7}$$

effects the transformation of a vector from the coordinate system of the virtual bond into the coordinate system of the $P-O_{5'}$ chemical bond. The parameter κ is the angle formed by the x axes of the virtual bond and the $P-O_{5'}$ bond and Φ is the supplement to the angle formed by the positive z axes of the virtual bond and chemical bond $P-O_{5'}$ when the rotation angle ω is equal to 0°.

The matrix product

$$\mathbf{T}_{v,\mathrm{O}_{3'}\text{-P}}(\zeta, \Psi, \tau) = \mathbf{Z}(\zeta)\mathbf{X}(\Psi)\mathbf{Z}(-\tau) \tag{8}$$

transforms a vector given in the coordinate system of chemical bond $O_{3'}-P$ into that of the virtual bond. The angles ζ and τ are formed by the x axis of the virtual bond or that of the $P-O_{3'}$ bond, respectively, with a vector $\mathbf{w}$ (see Figure 2) that is normal to the plane defined by the positive z axes of these two bonds (i.e., $\mathbf{w}$ is equal to the vector product ($\mathbf{z}_{\mathrm{O}_{3'}\text{-P}} \times \mathbf{z}_v$) where the $\mathbf{z}$'s represent bond vectors along the positive z axes of the given bonds). The angle Ψ is the supplement of the angle formed by the positive z axes of the virtual bond and of chemical bond $P-O_{5'}$. The matrix $\hat{\mathbf{T}}$ representing the transformation of a vector in the coordinate system of one virtual bond into that of the preceding virtual bond in the chain involves three successive transformations of vectors and is given by

$$\hat{\mathbf{T}} = \mathbf{T}_{v,\mathrm{O}_{3'}\text{-P}}(\zeta, \Psi, \tau)\mathbf{T}_{\mathrm{O}_{3'}\text{-P},\mathrm{P}\text{-}\mathrm{O}_{5'}}(\Theta^{\mathrm{P}}, \omega')\mathbf{T}_{\mathrm{P}\text{-}\mathrm{O}_{3'},v}(\kappa, \omega - \Phi) \tag{9}$$

This matrix is a function of the $\omega'\omega$ rotation pair of angles only; all other parameters in the above expression are constants dependent upon the value of ψ' (i.e., ring pucker). The values of these parameters for all four sugar conformations are listed in Table I.

The scalar product in eq 6 can now be formulated so that the $\langle r^2 \rangle_0$ is expressed as[10]

$$\langle r^2 \rangle_0 = xv^2 + 2\sum_{i<j} \mathbf{v}_i^{\mathrm{T}} \langle \hat{\mathbf{T}}_i \hat{\mathbf{T}}_{i+1} \ldots \hat{\mathbf{T}}_{j-1} \rangle \mathbf{v}_j \tag{10}$$

Independence of each pair of $\omega'\omega$ rotation angles in the chain renders each $\hat{\mathbf{T}}$ in the serial products of eq 10 independent. The averaged product of these matrices may thus be replaced by products of the averaged matrix $\langle \hat{\mathbf{T}} \rangle$. The characteristic ratio $\langle r^2 \rangle_0/n\bar{l}^2$, where $n = 6x$ is the number of bonds along the chain backbone and $\bar{l}^2$ is the mean-square bond length, may then be evaluated as

$$C_n = \langle r^2 \rangle_0/nl^2 = v^2/6l^2[(\mathbf{E} + \langle \hat{\mathbf{T}} \rangle)(\mathbf{E} - \langle \hat{\mathbf{T}} \rangle)^{-1} - (2/x)\langle \hat{\mathbf{T}} \rangle(\mathbf{E} - \langle \hat{\mathbf{T}} \rangle^x)(\mathbf{E} - \langle \hat{\mathbf{T}} \rangle)^{-2}]_{11} \tag{11}$$

The parameter $\mathbf{E}$ in this expression is the identity matrix of order 3 and the 11 denotes the 1,1 element of the averaged matrix product. In the limit of an infinite chain length the expression reduces to

$$C_\infty = v^2/6l^2[(\mathbf{E} + \langle \hat{\mathbf{T}} \rangle)(\mathbf{E} - \langle \hat{\mathbf{T}} \rangle)^{-1}]_{11} \tag{12}$$

Results of Numerical Calculations

Chain Type and Chain Length Dependencies. Values of the characteristic ratio C_n calculated using eq 11 for randomly coiling polynucleotides composed of identical pentose repeating units are plotted as a function of the degree of polymerization $x = n/6$ in Figure 3. The averaged matrix $\langle \hat{\mathbf{T}} \rangle$ required for the calculation was evaluated by summation at 30° intervals over classical $\omega'\omega$ conformational potential energy surfaces as described previously.[13] The potential surfaces employed in the computation were the $\omega'\omega$ contour maps detailed in the previous two virtual bond treatment of polynucleotides (there described as the $\psi'\psi''$ surfaces).[13] These energy estimates lead to calculated dimensions which conform to the experimental behavior of polynucleotides in a Θ solvent.[14-16] Calculations based upon quantum mechanical estimates of the $\omega'\omega$ conformational energy or upon the frequency of X-ray crystallographic structures, however, yield chain dimensions significantly smaller than the experimental values.[2,3] A discussion of these observations is presented elsewhere.[17]

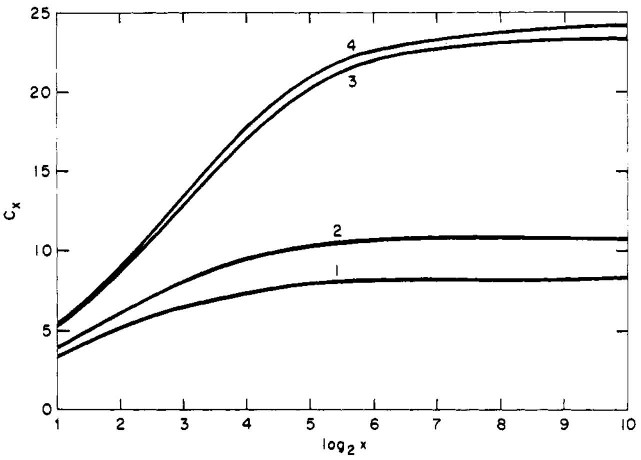

Figure 3. The calculated dependence of the characteristic ratio on the number of repeat units in $C_{3'}$-exo (4), $C_{2'}$-endo (3), $C_{3'}$-endo (2), and $C_{2'}$-exo (1) chains.

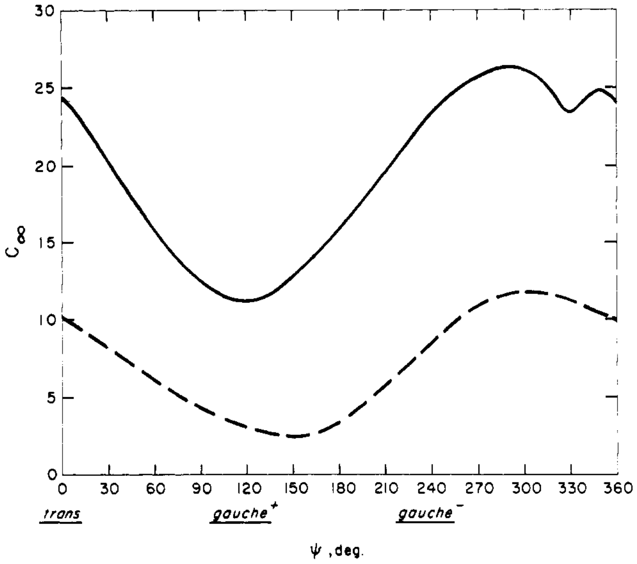

Figure 4. Characteristic ratios for the $C_{2'}$-endo (—) and the $C_{3'}$-endo (- - -) chains as a function of the rotation angle ψ about chemical bond $C_{5'}-C_{4'}$.

The characteristic ratios of both the $C_{3'}$-exo and the $C_{2'}$-endo chains, represented by curves 4 and 3, respectively, approach limiting values of 24.1 and 23.3. The 11° variation in ψ' (see Table I) associated with the different ring puckerings thus has only a minor influence on chain dimensions. Furthermore, the dependence of C_n upon x is similar in the two chains. In both cases a polymer of at least 2^7 or 128 repeating units is required before the asymptotic values of C_∞ are substantially (99%) attained. In view of the similarities of these chains it is thus not necessary to differentiate between the $C_{3'}$-exo and $C_{2'}$-endo puckerings in treating properties of the polynucleotide as a whole. These two conformations may then be described by a single rotational isomeric state. In the calculations that follow only the $C_{2'}$-endo chain will be treated.

The characteristic ratios for the $C_{3'}$-endo and $C_{2'}$-exo chains, represented in Figure 3 by curves 2 and 1, respectively, are considerably smaller than those of the $C_{3'}$-exo or $C_{2'}$-endo chain and also approach their asymptotic limits more slowly. The limiting values of C_∞ are 11.1 and 8.3 for the $C_{3'}$-endo and $C_{2'}$-exo chains, respectively, and these curves reach 99% of their asymptotic values in chains of 2^8 or 256 repeating units. Although the dependence of C_∞ upon ψ' is somewhat greater when this angle is in the gauche$^-$ range of $C_{3'}$-endo–$C_{2'}$-exo puckering rather than in the trans range of $C_{3'}$-exo–$C_{2'}$-endo puckering, it is still reasonable to represent the $C_{3'}$-endo–$C_{2'}$-exo puckered forms by a single rotational isomeric state. Thus, only the $C_{3'}$-endo chain will be considered below.

Dependence of C_∞ Upon ψ. As pointed out above, experimental studies of conformation in low molecular weight nucleic acid analogs indicate that the rotation angles ψ about the $C_{5'}$–$C_{4'}$ bonds of the polynucleotide backbone may assume conformation ranges other than the preferred gauche$^-$ region.[2,3] The extent to which variation of this angle affects the mean-square dimensions of the polynucleotide is evident from Figure 4 where C_∞ of $C_{2'}$-endo and $C_{3'}$-endo chains is plotted as a function of ψ. These quantities were determined using eq 12 in conjunction with the geometric parameters υ, κ, Φ, ζ, τ, and Ψ required for the evaluation of the averaged $\langle \hat{\mathbf{T}} \rangle$ matrices associated with each value of ψ. The upper solid line in Figure 4 represents the variation in C_∞ for a $C_{2'}$-endo chain. Variation of ψ within the gauche$^-$ rotation range (centered about the ψ = 240° rotational isomeric state chosen in the model above) as well as angular rotation in the trans range (centered about ψ = 0°) has little effect on the calculated dimensions. The value of C_∞, however, is significantly depressed if ψ is restricted to the gauche$^+$ rotational range (centered about ψ = 120°). The unperturbed dimensions of the $C_{3'}$-endo chain, represented by the lower dashed curve in Figure 4, show a similar dependence on ψ. The calculated values of C_∞ are similar when ψ is found in the trans and gauche$^-$ ranges but are significantly lower when the rotation is gauche$^+$.

Of the three staggered conformations of ψ, the gauche$^+$ is least favored according to both theoretical energy estimates[2,3] and experimental studies[1-3] of model nucleic acid systems. The very small extent (less than 10%) to which the gauche$^+$ state is populated would then have little effect on the unperturbed dimensions of the chain. It is thus not surprising that the unperturbed dimensions of polynucleotides calculated on the basis of a two virtual bond scheme (which allows for angular variation of ψ as well as rotations about the phosphodiester bonds and $C_{5'}$–$O_{5'}$ bond) are similar to those reported above. The values determined in the two virtual bond treatment are 24.8 and 9.0 for $C_{2'}$-endo and $C_{3'}$-endo chains, respectively.[13]

Discussion

The treatment of polynucleotides in terms of a single virtual bond repeating unit as set forth above rests upon the premise that the sugar residues of the chain are structurally rigid moieties. The fact that the unperturbed dimensions calculated on the basis of a two virtual bond model which allows for flexibility in the sugar are similar to those determined with the one virtual bond scheme lends strong support to this assumption. In fact, the type of sugar (pucker) comprising the chain appears to be a much more important factor than the flexibility of a given sugar in determining configuration-dependent properties of the polynucleotide. Treatment of spatial properties of the chain only requires that the polynucleotide be regarded as a sequence of two kinds of rigid sugar repeating units, $C_{2'}$-endo or $C_{3'}$-endo, connected by flexible phosphodiester links.

Acknowledgment. The support of a U.S. Public Health Service Grant (GM 20861) is gratefully acknowledged. The author thanks the Rutgers University Center for Computer and Information Services for the necessary computational support and Elizabeth Matthews for technical assistance. The helpful comments of the referees are also greatly appreciated.

References and Notes

(1) M. Sundaralingam, *Jerusalem Symp. Quantum Chem. Biochem.*, **5**, 417 (1973).
(2) W. K. Olson, *Biopolymers*, in press, and references cited therein.
(3) R. Tewari, R. K. Nanda, and G. Govil, *Biopolymers*, **13**, 2015 (1974).
(4) B. Pullman, W. Saenger, V. Sasisekharan, M. Sundaralingam, and H. R. Wilson, *Jerusalem Symp. Quantum Chem. Biochem.*, **5**, 815 (1973).
(5) N. Yathindra and M. Sundaralingam, *Proc. Natl. Acad. Sci. U.S.*, **71**, 3325 (1974).
(6) D. A. Brant and P. J. Flory, *J. Am. Chem. Soc.*, **87**, 2791 (1965).
(7) V. S. R. Rao, N. Yathindra, and P. R. Sundararajan, *Biopolymers*, **8**, 325 (1969).
(8) C. V. Goebel, W. L. Dimpfl, and D. A. Brant, *Macromolecules*, **3**, 644 (1970).
(9) S. G. Whittington, *Biopolymers*, **10**, 1481, 1617 (1971).
(10) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969.
(11) W. K. Olson, *Biopolymers*, **12**, 1787 (1973).
(12) S. Arnott, *Prog. Biophys. Mol. Biol.*, **21** (1970).
(13) W. K. Olson and P. J. Flory, *Biopolymers*, **11**, 25 (1972).
(14) H. Eisenberg and G. Felsenfeld, *J. Mol. Biol.*, **30**, 17 (1967).
(15) L. D. Inners and G. Felsenfeld, *J. Mol. Biol.*, **50**, 373 (1970).
(16) E. K. Achter and G. Felsenfeld, *Biopolymers*, **10**, 1625 (1971).
(17) W. K. Olson, *Biopolymers*, in press.