On the Role of Uniform and Mixed Sugar Puckers in DNA Double-Helical Structures

Shashidhar N. Rao and Peter Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. *Received May 16, 1984*

Abstract: Molecular mechanical studies of the base-paired deoxyhexanucleoside phosphates $d(ATATAT)_2$, $d(TATATA)_2$, $d(GCGCGC)_2$, $d(CGCGCG)_2$, and $dA_6 \cdot dT_6$ have been carried out to assess the energetic consequences of C2′ endo and C3′ endo sugars in a right-handed deoxyribopolynucleotide chain. Consistent with experiment, the alternating AT polymers, $d(ATATAT)_2$ and $d(TATATA)_2$, have a tendency to adopt mixed sugar puckers, with the sugar attached to adenine C3′ endo and thymine C2′ endo, whereas $d(CGCGCG)_2$ and $d(GCGCGC)_2$ are calculated to prefer structures with uniform C2′ endo puckering. In contrast, for the polymer $dA_6 \cdot dT_6$, the lowest energy structure is predicted to have the sugars attached to adenine in a C2′ endo conformation and those attached to thymine C3′ endo. All these results can be nicely rationalized by considering the following energy contributions: (a) the intrinsic sugar puckering preference, which favors a C2′ endo conformation by ~0.5–1.0 kcal/mol; (b) phosphate–phosphate repulsions, which are less destablizing for uniform C2′ endo conformations than for mixed C2′ endo–C3′ endo conformations, and (c) thymine–phosphate interactions, which are more favorable if the deoxyribose sugar on the 5′ end of the thymine has a C3′ endo conformation. The competition between effects (a) and (b), which favor uniform C2′ endo B-DNA conformations, and (c), which favors C3′ endo conformations on the 5′ side of thymine bases, allows one to rationalize the available experimental data and to *predict* that poly(dA)·poly(dT) will prefer a conformation in which the poly(dT) strand has C3′ endo sugars to the one in which the poly(dA) strand has C3′ endo sugar geometries.

DNA is known to exhibit several polymorphic forms both in solution and in the solid state. The proposed B-DNA structures[1,2] for polynucleotides with random sequences of the bases have an approximate mononucleotide repeating unit. However, crystal structure analysis[3] of the tetranucleotide fragment $d(pApTpApT)_2$ has indicated the existence of an "alternating B" form of DNA. Based on that structure,[3] it was hypothesized[4] that poly(dA-dT)·poly(dA-dT) adopted the right-handed double-helical structure, but sugars attached to the thymine bases had C2′ endo puckers and those attached to the adenine bases had C3′ endo puckers. The phosphodiester conformation in the ApT fragment was (gauche$^-$, gauche$^-$), while that in the TpA fragment was intermediate between (gauche$^-$, gauche$^-$) and (trans, gauche$^-$). These conformational variations were suggested to modify the environment around the 5′ end of the thymine base so as to enhance the binding of poly(dA-dT)·poly(dA-dT) to the lac repressor of *E. coli* by several orders of magnitude compared to the corresponding binding by calf thymus DNA.[4] On the basis of these conformational features, a model for the spatial relationship between the parent and daughter stems at the fork at a particular stage of the replication process[5] has been proposed.

The occurrence of alternate sugar puckers in a dinucleoside monophosphate fragment is believed to be important in drug–nucleic acid interactions, particularly those involving intercalating drugs, although model building studies suggest that drug–polynucleotide complexes could have either alternating sugar pucker or uniform C3′ endo sugar pucker.[6-8]

Several NMR studies carried out in recent years have also supported the dinucleotide repeat of B-type structures for poly-(d(AT))·poly(d(AT)). Shindo et al.[9] and Patel et al.[10] have demonstrated that at low molar concentration of salt, poly(dA-dT)·poly(dA-dT) exhibits a dinucleotide repeat, which does not correspond to the left-handed Z conformation but to an alternating right-handed structure of the B type. Recently, Vorlickova et al.[11] have concluded that poly(dA-dT)·poly(dA-dT) could exist in a dinucleoside repeat conformation, altogether different from the B, A, and Z forms, both at high and low salt concentrations.

On the other hand, low salt conditions are believed to favor the B-DNA structure (with mononucleotide repeat) for poly(dG-dC)·poly(dG-dC).[12] This synthetic polynucleotide is known to take up the now well-known Z conformation at high salt concentrations, which is characterized by a dinucleotide repeat in which the purine and the pyrimidine bases are respectively attached to C3′ endo and C2′ endo sugars. Model building studies on the synthetic polynucleotides and nucleic acids with random sequences, based on the X-ray fiber diffraction data, have led to the proposal of right- and left-handed double-helical structures with both mono- and dinucleotide repeat units.[1,13,14]

The importance of sequence-dependent differences in sugars puckers has been found in molecular mechanics studies on various polymers of DNA.[15-17] In models of B-DNA poly(dA)·poly(dT) (e.g., $dA_6 \cdot dT_6$, $dA_{12} \cdot dT_{12}$), the sugars in the homopyrimidine strand tend to repucker to C3′ endo geometries from the conventional C2′ endo geometries. Support for this hypothesis has been found in recent Raman spectroscopic analyses[18] of poly(dA)·poly(dT), which have indicated the occurrence of both C3′ endo and C2′ endo sugars at low temperatures (5 °C or lower). It has been pointed out that in poly(dA)·poly(dT), both of these sugar conformations were found at both low and high salt concentrations, while in poly(dA-dT)·poly(dA-dT), high salt concentrations and low temperatures were essential for the occurrence of both C2′ endo and C3′ endo sugar puckers. In calf thymus DNA, only C2′

(1) Gupta, G.; Bansal, M.; Sasisekharan, V. *Int. J. Biol. Macromol.* **1980**, *2*, 368.
(2) Arnot, S.; Campbell-Smith, P.; Chandrasekaran R. In "CRC Handbook of Biochemistry"; Fasman, G. D., Ed.; CRC: Cleveland, OH, 1984; Vol. II, p 411.
(3) Viswamitra, M. A.; Kennard, O.; Jones, P. G.; Sheldrick, G. M.; Salisbury, Ls. A.; Falvello, L.; Shakked, Z. *Nature (London)* **1978**, *273*, 687.
(4) Klug, A.; Jack, A.; Viswamitra, M. A.; Kennard, O.; Shakked, Z.; Steitz, T. A. *J. Mol. Biol.* **1979**, *131*, 669.
(5) Viswamitra, M. A.; Shakked, Z.; Jones, P. G.; Sheldrick, G. M.; Salisbury, S. A.; Kennard, O. *Biopolymers* **1982**, *21*, 513.
(6) Sobell, H. M. In "Nucleic Acid Geometry and Dynamics"; Sarma, R. H., Ed.; Pergamon Press: New York, 1980; p 289.
(7) Shieh, H.-S.; Berman, H. M.; Dabrow, M.; Neidle, S. *Nucleic Acids Res.* **1980**, *8*, 85.
(8) Neidle, S. In "Topics in Nucleic Acid Structure"; Neidle, Ed.; Wiley: New York, 1981; p 177.
(9) Shindo, H.; Simpson, R.; *Cohen*, J. *J. Biol. Chem.* **1979**, *254*, 8125.
(10) Patel, D. J.; Kozlowski, S. A.; Suggs, J. W.; Cox, S. D. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4063.
(11) Vorlickova, M.; Kypr, J.; Sklenar, V. *J. Mol. Biol.* **1983**, *166*, 85.
(12) Cohen, J. S.; Wooten, J. S.; Chatterjee, C. L. *Biochemistry* **1981**, *20*, 3049.
(13) Leslie, A. G. W.; Arnott, S.; Chandrasekaran, R.; Ratliff, R. L. *J. Mol. Biol.* **1980**, *143*, 49.
(14) Sasisekharan, V.; Gupta, G. *Curr. Sci.* **1980**, *49*, 43.
(15) Kollman, P. A.; Weiner, P.; Dearing, A. *Biopolymers* **1981**, *20*, 2583.
(16) Kollman, P. A.; Keepers, J. W.; Weiner, P. *Biopolymers* **1982**, *21*, 2345.
(17) Tilton, R. F., Jr.; Weiner, P.; Kollman, P. A. *Biopolymers* **1983**, *22*, 969.
(18) Thomas, G. A.; Peticolas, W. L. *J. Am. Chem. Soc.* **1983**, *105*, 993.

0002-7863/85/1507-1611$01.50/0

endo sugar puckers were found. In all these polynucleotides, only C2′ endo sugars have been concluded to exist at high temperatures, independent of salt concentration. However, recently, Arnott et al.[19] have pointed out from their fiber diffraction studies that poly(dA)·poly(dT) could take up a B-like conformation, with all the purine nucleotides having C3′ endo sugars and the pyrimidine nucleotides having C2′ endo sugars, which is the opposite puckering preference suggested by earlier calculations.[15-17]

In view of the experimental evidence for the importance of the alternate sugar puckers in polynucleotides with purine–pyrimidine sequences as well as synthetic homopolynucleotides, a systematic molecular mechanics investigation on the energetics of a number of model compounds has been undertaken. Conformations of hexanucleotide duplexes with various uniform and alternating sugar puckers have been analyzed. The two major goals of this study are firstly to understand the tendency of poly(dA-dT)·poly(dA-dT) to adopt the alternating repeat B form of DNA in contrast to the preference of mononucleotide repeat B-DNA in poly(dG-dC)·poly(dG-dC) and secondly to predict the sugar puckers of the lowest energy structure of AT homopolymers poly(dA)·poly(dT) and to understand why such a sugar pucker is a lower energy structure than alternative sugar pucker models.

Methods

The calculations used the method of molecular mechanics and the program AMBER.[20] The energies were determined by using eq 1 given below, and the structures were refined until the root mean square gradient was less than 0.05 kcal/mol Å. The force field parameters pres-

$$E_{tot} = \sum_{bonds} K_r(R - R_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{\substack{i<j \\ nonbonded}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{H\ bonds} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] \quad (1)$$

ented by Weiner et al.[21] were employed in all the calculations, as was a distance-dependent dielectric constant, $\epsilon = R_{ij}$.

We have investigated the conformational features of the following hexanucleotides: $d(ATATAT)_2$, $d(TATATA)_2$, $d(GCGCGC)_2$, $d(CGCGCG)_2$, and $d(A_6)\cdot d(T_6)$. As before,[15] the hexamers containing A and T residues have been referred to as AT polymers and those containing G and C residues have been referred to as GC polymers. The residues corresponding to adenine, thymine, guanine, and cytosine bases have been referred to as ADE, THY, GUA, and CYT, respectively, together with their sequence numbers along a given strand (in the 5′–3′ direction). For example, in $d(ATATAT)_2$, the residues would be referred to as ADE1, THY1, ADE2, THY2, ADE3, and THY3, respectively. For the sake of convenience of description, the structures with the above five sequences have been referred to as AT, TA, GC, CG, and A6T6, respectively, in the rest of this paper. Also, the phosphate groups have been referred to as P_{n-m}, where n and m are the serial numbers of the bases at, respectively, 5′ and 3′ ends. For example, the phosphate group between ADE2 and THY2 in $d(ATATAT)_2$ is designated as $P_{(3-4)}$.

We have added additional constraint energy terms $E_{con} = k \cos(\psi - \psi_{target})$ (where ψ_{target} is the desired value of the ψ torsional angle and $k = 100$ kcal/mol; $\psi'_{target} = 80°$ to force C3′ endo geometries; $\psi'_{target} = 140°$ to force C3′ endo geometries) to eq 1 in order to force different sugar pucker geometries in the polynucleotide. In this way we can construct and analyze polynucleotide geometries with any desired combination of sugar pucker, while allowing all degrees of freedom to reach their optimum values.

The energy refined structures of the first four of these hexanucleotides were obtained in a number of different ways. In the first one, we started with the B-DNA geometry,[22] while in the second approach the starting conformation was the alternating B form.[4] In both of these cases, no constraints were imposed in obtaining the final energy minimized structures which are identified by the subscripts 1 and 2, respectively. In the third approach, the alternate sugars were constrained to have C3′ endo and C2′ endo puckers (for purine and pyrimidine nucleotides, respectively). The constrained structures were refined, until the earlier stated condition on the rms gradient was achieved (leading to structures AT_3, TA_3, GC_3, and CG_3), and then the constraints were relaxed and the refinement was continued (leading to structures AT_4, TA_4, GC_4, and CG_4). In the case of the two alternating AT polymers, similar calculations were also carried out, with the sugars in the purine and pyrimidine nucleotides being constrained to C2′ endo and C3′ endo puckers, respectively. These structures have been denoted by AT_5 and TA_5. These structures were further refined with the above stated constraints removed, to obtain a new set of structures denoted by AT_6 and TA_6, respectively. In addition, $d(TATATA)_2$ was also refined with the sugars constrained to have the C2′ endo geometries, and the resultant structure has been designated as TA_7.

The homopolymer $dA_6\cdot dT_6$ was also energy refined with three different sets of constraints. The first of them was C3′ endo and C2′ endo sugar puckers for the purine and pyrimidine nucleotides, respectively, while the other specified the sugar geometries in the reverse order. In the third approach, each strand had alternate sugars with C2′ endo and C3′ endo puckers. These three structures have been denoted by $(A6T6)_1$, $(A6T6)_2$, and $(A6T6)_3$, respectively. These three structures have been further refined after removing the constraints on the sugar puckers, giving rise to structures which have been designated $(A6T6)_4$, $(A6T6)_5$, and $(A6T6)_6$, respectively. Also, this homopolymer has been refined without any constraints, starting from the B-DNA geometry,[22] and the resultant structure has been designated as $(A6T6)_7$. Further, calculations were also carried out by constraining all the sugars to C2′ endo geometries, resulting in a structure designated as $(A6T6)_8$.

We note that the goal of these various calculations is to create structures with various sugar pucker combinations and, after removing the energy constraints, to assess whether such sugar pucker combinations correspond to local minima.

Table I. Sugar Puckering Properties and Total Energies (in kcal/mol) of the AT and GC Hexanucleotides Investigated in the Present Study[a]

	E^3	I	E^2	tot energies
		$d(ATATAT)_2$		
AT_1		THY1 (113)	all ADE, THY(2,3)	−645.4
AT_2		THY1 (112)	all ADE, THY(2,3)	−644.7
AT_3	all ADE		all THY	−645.1
AT_4	all ADE		all THY	−647.4
AT_5	all THY		all ADE	−634.5
AT_6	THY(2,3)	THY1 (93)	all ADE	−636.9
		$d(TATATA)_2$		
TA_1		ADE1 (107)	all THY, ADE(2,3)	−644.0
TA_2	ADE2	ADE1 (81)	all THY, ADE3	−644.7
TA_3	all ADE		all THY	−637.8
TA_4	ADE(2,3)	ADE1 (87)	all THY	−641.0
TA_5	all THY		all ADE	−632.9
TA_6	THY(1,3)	THY2 (120)	all ADE	−643.1
TA_7			all ADE, all THY	−640.0
		$d(GCGCGC)_2$		
GC_1		CYT1 (122)	all GUA, CYT(2,3)	−541.0
GC_2		CYT1 (120)	all GUA, CYT(2,3)	−540.9
GC_3	all GUA		all CYT	−531.2
GC_4	GUA3	GUA1 (115)	all CYT, GUA2	−538.7
		$d(CGCGCG)_2$		
CG_1			all CYT, all GUA	−539.3
CG_2			all CYT, all GUA	−538.9
CG_3	all GUA		all CYT	−526.0
CG_4	GUA2, GUA3	GUA1 (119)	all CYT	−530.8
CG_5			all CYT, all GUA	−540.7
CG_6	GUA2	CYT2 (94)	all except CYT2, GUA2 of first strand	−539.7

[a] The phases (in deg) of the sugars with intermediate puckers are indicated in parentheses. The nomenclature for the various structures has been detailed in the text.

Results

Conformational Characteristics of the Alternating Polymers. Because the twofold symmetry in these structures, the discussion on conformational parameters will be confined to one of the strands only. The sugar puckering properties of these four hexanucleotides have been listed in Table I. The W values in the ranges 0–54°

(19) Arnott, S.; Chandrasekaran, R.; Banerjee, A. K.; He, R.; Walker, J. K. *J. Biomol. Struct. Dyn.* **1983**, *1*, 437.
(20) P. Weiner, P.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287.
(21) Weiner, S. J.; Kollman, P. A.; Case, D.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(22) Arnott, S.; Chandrasekaran, R.; Hall, I. H.; Puigjaner, L. C. *Nucl. Acids Res.* **1983**, *11*, 4141.

and 126–180° correspond, respectively, to C3′ endo and C2′ endo sugar puckers (marked E^3 and E^2, respectively, in Table I). The intermediate sugar puckers designed by I have a range of 54–126° for *W*.

d(ATATAT)$_2$. For the lowest energy conformation AT_4, the three adenine residues in each strand have their corresponding sugar puckers as C3′ endo-C_4′ exo, C3′ endo-C_4′ exo, and C3′ endo, in the 5′–3′ direction of the polynucleotide. The thymine residues have all C2′ endo sugar puckers. AT_2 has none of its sugars with C3′ endo puckers. The sugars attached to adenine bases have either a C2′ endo geometry or a mixed pucker, C2′ endo-C1′ exo. The C2′ endo sugars attached to thymine bases repuckered to either O1′ endo-C1′ exo or C1′ exo-C2′ endo geometries. The sugar geometries in AT_1 are very similar to those in AT_2. In AT_3, the sugar geometries do not deviate significantly from the starting conformations of C3′ endo and C2′ endo puckers for purine and pyrimidine nucleotides, respectively.

The other backbone torsion angles in AT_1 and AT_2 are almost identical with one another and are in the range found in previous calculations on DNA double helixes.[14-16] Also, the backbone torsions in AT_3 and AT_4 have very similar values. When these conformational parameters in AT_1 and AT_4 are compared, only the phosphodiester conformations show variations of as large as 20–30° between the two structures.

d(TATATA)$_2$. As in d(ATATAT)$_2$, the backbone conformational parameters (other than the sugar puckers) do not very more than 25–30° in the various structures. However, TA_1, TA_2, TA_3, and TA_4 have a few characteristics different from those of their counterparts for d(ATATAT)$_2$. There are major differences in the phosphodoiester conformations and sugar puckers between TA_1 and TA_2. On the other hand, almost all the backbone torsions are similar in TA_2 and TA_3. Only the sugar puckers of the ADE3 residues in the two structures differ significantly, being C2′ endo and C3′ endo, respectively.

d(GCGCGC)$_2$. In GC_1 (as in AT_1 and TA_1) and in GC_2, all the sugar puckers are either C2′ endo or C1′ exo. As in d(ATATAT)$_2$, GC_1 and GC_2 have almost identical backbone conformations. Unlike the AT polymers, GC_3 and GC_4 have a few significant differences. In GC_3, as is to be expected, the alternating sugars have C3′ endo (attached to guanines) and C2′ endo (attached to cytosines) puckers, while in GC_4, the sugars corresponding to GUA1 and GUA2 residues have intermediate geometries C1′ exo–C2′ endo and C1′ exo, respectively. These changes in sugar pucker are accompanied by changes in the phosphodiester conformations. While ω' at the 3′ end of GUA1 decreases by about 20° in GC_4 (over GC_3), that at the 3′ end of GUA3 decreases by 40°. When GC_1 (or GC_2) is compared wtih GC_4, ω' at the 3′ end of GUA3 is the only other backbone torsion which differs significantly (by about 40°) in the two structures.

d(CGCGCG)$_2$. As in the case of d(GCGCGC)$_2$, the structures CG_1 and CG_2 are very similar. The sugar puckers are either C2′ endo or C1′ exo. Only the sugar corresponding to GUA1 changes pucker to a C1′ exo geometry from a C3′ endo geometry. The other two sugars attached to the guanine bases have their puckers retained in the C3′ endo region. The only significant difference in the rest of the backbone conformational parameters is the P–O3′ torsion at the 3′ end of GUA1, which differs in the two structures by about 25°. Comparing CG_1 (or CG_2) with CG_4 reveals that the sugar puckers of the GUA3 residue in the two structures differ significantly, being C2′ endo in the former and C3′ endo in the latter. These differences lead to differences in the P–O3′ torsion at the 3′ and 5′ ends of these sugar moieties, whereas the values of the other backbone torsions are almost identical with those in the structures of d(GCGCGC)$_2$. In all the structures, the χ values for C2′ endo sugars are around 60°, whereas those for C3′ endo sugars are around 20–30°.

Conformational-Dependent Energies of the Alternating Polymers. The total energies of the various alternating polymers are listed in Table I. It is found that only for the AT polymers, the structures with both C3′ endo and C2′ endo sugar puckers are comparable in energy to those with uniform C2′ endo sugar puckers throughout. In contrast, for the GC polymers, the structures with uniform C2′ endo (or closely related C1′ exo) puckers are significantly more stable than those with mixed sugar puckers. It is not surprising that the "unrestrained" calculations have neither completely uniform sugar geometries of a single kind or strictly an alternating arrangement of C2′ endo and C3′ endo puckers. In view of this fact, it is not feasible in a simple fashion to attribute the total energies of such structures of specific sugar geometry changes. However, a general trend of variation of energies in relation to sugar geometries can be noted.

In d(ATATAT)$_2$, the structure with the sugar constrained to C2′ endo and C3′ endo puckers (AT_3) is almost equal in energy to AT_1 with uniform C2′ endo sugar puckers. Further, AT_4, which is obtained from AT_3 after relaxing the constraints with subsequent energy refinement, is lower in energy than AT_1 by almost 3 kcal/mol. Thus, d(ATATAT)$_2$ has mixed sugar puckers with C3′ endo and C2′ endo geometries in ADE and THY residues, respectively, in its lowest energy conformation.

In contrast to d(ATATAT)$_2$, in d(TATATA)$_2$ the structure with constrained sugars (TA_3) is destablized by nearly 7 kcal/mol relative to TA_2. TA_4 and TA_7 are also significantly less stable than TA_2. Each of them is higher in energy than TA_2 by about 4 kcal/mol. Surprisingly, TA_1 and TA_2 differ only by 0.7 kcal/mol despite the fact that the latter has a central adenine sugar with C3′ endo geometry, with the sugar attached to the same base at the 5′ end with an intermediate O1′ endo–C4′ exo pucker. Thus, it appears that in an AT polymer, the adenines would prefer at least some sugars with C3′ endo (or closely related C4′ exo geometries), while the thymines prefer C2′ endo (or closely related C1′ exo) geometries.

In d(GCGCGC)$_2$, the lowest energy of the four structures has mostly C2′ endo puckers, with the cytosine bases at the 5′ ends in both the strands being attached to the C1′ exo sugars. Also, GC_2 has a backbone conformation and a total energy very similar to those of GC_1. These two facts suggest that GC polymers prefers a uniform sugar pucker in the framework of right-handed double-helical B-DNA. This is substantiated by the energies of the structures GC_3 and GC_4. In GC_3, where the sugars attached to guanine bases are constrained to C3′ endo geometries, the destabilization relative to that of GC_1 is about 9 kcal/mol. In GC_4, where some of the C3′ endo sugars repucker to C2′ endo geometries, the corresponding destabilization is only 1.3 kcal/mol.

In d(CGCGCG)$_2$, its lowest energy conformation has the same features as that of d(GCGCGC)$_2$. Here too, CG_1 and CG_2 (the latter differs from the former only by 0.4 kcal/mol) have mostly C2′ endo puckers. Interestingly, however, the difference in energies between CG_1 and CG_4 is considerably more than that between the "GC" counterparts, presumably because one of the sugars in the guanine residues of each strand repuckers to C2′ endo geometries, while the other two retain the predominantly C3′ endo character. In view of this, it is easy to rationalize a much larger difference in energy between CG_1 and CG_3.

Conformations and Conformational-Dependent Energies of dA$_6$-dT$_6$. Table II lists the total energies of structures (A6T6)$_1$ through (A6T6)$_7$, of which (A6T6)$_5$ is the most stable. In this structure, all the sugars on the ADE strand have C2′ endo geometries, while most of them on the THY strand have C3′ endo geometries. Only the sugars corresponding to THY1 and THY2 residues have intermediate geometries, C1′ exo–O1′ endo and O1′ endo, respectively. (A6T6)$_4$ is less stable than (A6T6)$_5$ by about 7 kcal/mol. In (A6T6)$_4$, the sugars attached to most of the adenine bases have C3′ endo–C2′ exo puckers, while those attached to ADE1 and ADE2 have, respectively, O1′ endo–C1′ exo and C1′ endo puckers. On the THY strand, the sugars corresponding to THY2, THY3, and THY4 have, respectively, O1′ endo–C1′ exo, and C1′ exo–C2′ endo puckers. The other three sugars have C2′ endo geometries.

Thus, for an AT homopolymer, the sugars attached to adenine bases tend to prefer C2′ endo pucker, while those corresponding to thymine residues tend to prefer C3′ endo pucker. This conclusion is substantiated by the observation that (A6T6)$_1$ is higher in energy than (A6T6)$_5$ by about 15 kcal/mol; in the former, sugars attached to adenine and thymine bases are constrained to

Table II. Sugar Puckering Properties and Total Energies (in kcal/mol) of $dA_6 \cdot dT_6$ Structures Obtained in the Present Investigations[a]

	E^3	I	E^2	tot energy
$(A6T6)_1$	all ADE		all THY	−638.23
$(A6T6)_2$	all THY		all ADE	−648.23
$(A6T6)_3$	ADE(2,4,6)		ADE(1,3,5)	
	THY(2,4,6)		THY(1,3,5)	
$(A6T6)_4$	ADE(3,4,5,6)	ADE1(119)	ADE2,	−645.20
			THY-(1,3,4,5,6)	
		THY2(112)		
$(A6T6)_5$	THY(3,4,5,6)	THY1(119)	all ADE	−652.50
		THY2(100)		
$(A6T6)_6$	ADE(4,6)	THY2(100)	ADE(1,2,3,5)	−647.60
	THY4		THY(1,3,5,6)	
$(A6T6)_7$		THY2(108)	all ADE	
		THY3(115)	THY(1,5,6)	−650.10
		THY4(100)		
$(A6T6)_8$			all ADE	−647.50
			all THY	

[a] The phase (W in deg) of the sugars corresponding to intermediate puckers have been indicated in parentheses.

C3′ endo and C2′ endo geometries, respectively. When $(A6T6)_1$ is refined with the constraints on sugar puckers removed, the resultant structure $(A6T6)_4$ is higher in energy than even $(A6T6)_2$ by about 2 kcal/mol. In the latter, the sugars corresponding to ADE and THY residues are constrained to C2′ endo and C3′ endo geometries, respectively.

The structure with alternate C2′ endo and C3′ endo sugars on each strand, $(A6T6)_3$, is about 10 kcal/mol higher in energy than $(A6T6)_5$. Further refinement with constraints removed led to a reduction of this difference to about 5 kcal/mol and was accompanied by repuckering of sugars corresponding to residues THY2, THY3, and THY6 to O1′ endo–C1′ exo, C1′ exo, and C1′ exo–C2′ endo geometries. However, on the adenine strand, only the sugar corresponding to ADE2 repuckers to C1′ exo geometry, whereas the other sugars have retained very nearly their starting geometries. The structure $(A6T6)_7$ is found to be only 2.4 kcal/mol higher in energy than $(A6T6)_5$. In this structure, all the sugars attached to adenine bases have C2′ endo or closely related C1′ exo puckers. On the other hand, the sugars corresponding to THY3, THY4, and THY5 residues have O1′ endo–C1′ exo puckers, while those corresponding to the other three thymine residues have C2′ endo and C1′ exo–C2′ endo puckers. Thus, during the course of refinement, a tendency on the part of some of the sugars attached to the thymine bases to change their pucker from C2′ endo toward C3′ endo is seen.

As in alternating purine–pyrimidine hexanucleotides, in the homopolymers too, the C4′–C5′, C5′–O5′, P–O5′, and C3′–O3′ torsions are confirmed to narrow ranges of values in the gauche$^+$, trans, gauche$^-$, and trans regions, respectively. the P–O3′ torsion is largely restricted to the gauche$^-$ region in all the structures of $dA_6 \cdot dT6$, with the range of variation being 245–300°. Thus, except for sugar pucker, the backbone torsions generally resemble those in the standard B-DNA geometry.[2] As in the other four hexanucleotides, here, too, the glycosydic torsions are in the anti region, with values of χ around 60 for C2′ endo sugars and around 20° for C3′ endo sugars.

Energy Component Analysis of the Hexanucleotides. We have carried out energy component analyses on these structures, concentrating our attention on the central two base pairs. Some of the common features of the energy minimized structures are as follows: (1) The phosphate–phosphate interactions around a C3′ endo sugar are higher in energy than those around a C2′ endo sugar by about 2–2.5 kcal/mol, and sugar–phosphate interactions are energetically more favorable in the case of C2′ endo than in the case of C3′ endo geometries. In general, interactions involving 5′-phosphates differ more significantly (around 2 kcal/mol) than those involving 3′-phosphates (0.5–1.0 kcal/mol). These two facts imply that in a hexanucleotide with larger amounts of C2′ endo sugars, the backbones are stabilized by the above interactions to a larger extent than when the sugars have more C3′ endo geometries. (2) As we have discussed earlier,[19] a deoxyribonucleotide is per se about 0.6 kcal/mol more stable in a C2′ endo than a C3′ endo geometry. (3) The interactions between thymine bases and their corresponding 5′-phosphate groups are sensitive to sugar geometry in the residue at the 5′ end. For example, in the case of a C3′ endo sugar, they are more favorable than in the case of C2′ endo sugar by about 1.5 kcal/mol. This difference is itself not very sensitive to the nature of the pucker of the sugar to which the thymine base is attached. The corresponding interactions involving adenine bases are not very different for different sugar puckers. The above factors help in understanding the relative stabilities of various hexanucleotides, as has been detailed below.

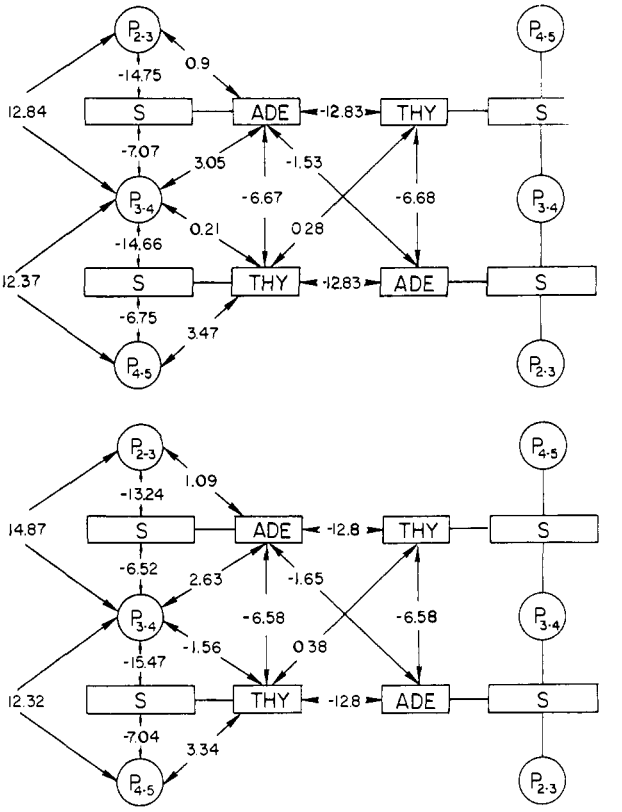

Figure 1. Component analysis of the interaction energies in the central two base pairs of two $d(ATATAT)_2$ structures (a, top) with predominantly C2′ endo sugar geometries in ADE and THY residues and (b, bottom) with C2′ endo and C3′ endo sugar puckers in THY and ADE residues, respectively. In these and the other component analysis diagrams, the sugars and bases are represented by rectangular boxes (marked S and the base name as in text, respectively). The phosphate groups (atoms O3′, O5′, P, OA, and OB) are represented by circles (marked P_{n-m} as defined in text). The interactions and their energies between various groups are represented by arrows connecting the appropriate boxes.

AT Polymers. In AT_3 and TA_3, the adenines and the thymines are attached to, respectively, C3′ endo and C2′ endo sugars. Yet the total energies of the two structures differ by about 7 kcal/mol. This difference is found to arise mainly due to differences in interactions between thymines and phosphate groups at the 3′ and 5′ ends, respectively, of the two hexanucleotides. For example, THY3–$P_{(5-6)}$ interactions at AT_3 are favored over THY1–$P_{(1-2)}$ interactions in TA_3 by about 5 kcal/mol. Further, this difference is contributed less significantly through nonbonded (1.2 kcal/mol) than the electrostatic interactions (3.6 kcal/mol). This is because the atoms N1, C6, C5, and C4 (of THY1) are farther from the atoms of $P_{(1-2)}$ in TA_3 than are the corresponding atoms of THY3 from $P_{(5-6)}$ at AT_3. However, the corresponding differences for adenine bases in the two structures are less significant. They contribute only around 1 kcal/mol in a way as to offset the above mentioned difference involving thymine bases.

Figure 1 shows the energy component analysis of the central two A···T base pairs of the two structures AT_4 and AT_1. Ster-

Figure 2. Stereopairs of two d(ATATAT)$_2$ structures (a, top) AT_1 and (b, bottom) AT_4. These two structures correspond to the component analysis diagrams of Figure 1, respectively.

eopairs corresponding to these structures are shown in Figure 2. We note that the former has alternating sugars and the latter has uniform sugar puckers in the C2′ endo region. AT_4 is more stable than AT_1, which can be attributed mainly to differences in interactions between thymine bases and their phosphates in the two structures. In AT_4, THY2–$P_{(3-4)}$ interaction is energetically more favorable by around 1.8 kcal/mol than in AT_1. Furthermore, the interactions between the adenine bases and their 5′-phosphates do not vary significantly (less than 0.2 kcal/mol) in the two structures.

In contrast, in the case of d(TATATA)$_2$, TA_1 is lower in energy than TA_4 by about 3 kcal/mol. While in d(ATATAT)$_2$, there are three thymine phosphate interactions which stabilize the conformations of each strand, here only two such interactions exist. This obviously leads to a reduced total interaction energy between the thymines and 5′-phosphates. The phosphate–phosphate interactions are obviously more favorable in TA_1 than in TA_4, for the reasons stated earlier. Also, the sugar–phosphate interactions stabilize TA_1 compared to TA_4 to a larger extent than they do in d(ATATAT)$_2$. Thus, the balance of differential energies involving the thymine–phosphate interactions on one hand and phosphate–phosphate and sugar–phosphate interactions on the other render TA_1 energetically more favorable than TA_4. In the case of AT_1 and AT_4, these energies were closer to equal.

Energy component analysis of AT_5, in which the sugar puckers have been constrained to C2′ endo (ADE) and C3′ endo (THY) geometries, reveals that its lower stability than AT_4 arises mainly from differences in interactions involving phosphates and the thymine bases. Base THY3–$P_{(5-6)}$ interaction is higher in energy in AT_5 than in AT_4 by about 1.8 kcal/mol. The corresponding difference for interactions base THY2--$P_{(3-4)}$ and base THY1–$P_{(1-2)}$ are, respectively, around 1.3 and 2.5 kcal/mol. In addition, the sugar–phosphate interactions are more favorable in AT_4 than in AT_5 because the former has three sugars with C2′ endo geometries involved in interactions with 5′-phosphates, while the latter has only two such sugars. In the case of d(TATATA)$_2$ also, the difference between the total energies of TA_3 and TA_5 is mainly due to a "competition" between the thymine–phosphate interactions and the phosphate–phosphate and sugar–phosphate interactions.

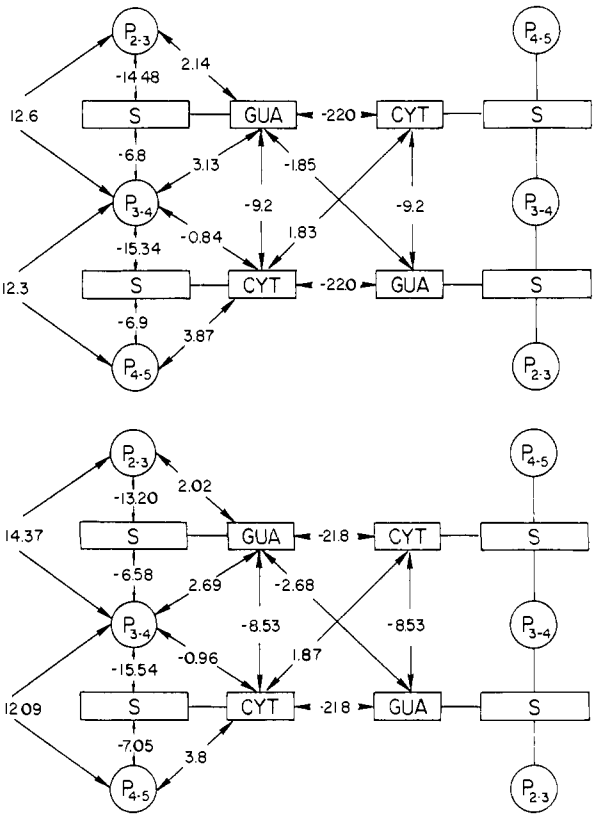

Figure 3. Component analysis of the interaction energies in the central two base pairs of two d(GCGCGC)$_2$ structures (a, top) with predominantly C2′ endo sugar geometries in GUA and CYT residues and (b, bottom) with C2′ endo and C3′ endo sugar puckers in CYT and GUA residues, respectively.

GC Polymers. We present the energy component analysis of GC_1 and GC_3 in Figure 3. GC_3 is higher in total energy than GC_1 by ~10 kcal/mol, and the major contributor to this difference is the phosphate–phosphate interactions. Also, the interaction energies between adjacent base pairs are slightly lower in GC_1 than in GC_3. In the case of AT polymers, we have noted that similar differences were compensated by the differences in pyrimidine–5′-phosphate interactions. For example, the interactions between the thymine base and $P_{(5-6)}$ differed in AT_1 and AT_3 by about 2.0 kcal/mol. However, in the case of the corresponding GC polymer, this difference is negligible. In order to obtain further insight into the reason for this difference, we have evaluated interaction energies between the atoms N1, C6, C5, C4, and C7 (in thymine only) and P, O5′, and OA (in phosphate group) in the four structures AT_1, AT_3, GC_1, and GC_3. We find that while it is not possible to single out a specific interaction between two specific atoms, one each on the phosphate and the base, it appears that the presence of a methyl group in thymine at C5 plays a role in this difference, as does the difference in charge distributions on these atoms in the two bases. Because of the absence of the stabilizing base–backbone interactions found in the AT polymers, in the GC polymers, the hexamer with alternate sugar puckers is considerably less stable than the hexamer with uniform C2′ endo sugar puckers.

dA_6·dT_6. Adenine and thymine bases have their corresponding sugars with C3′ endo and C2′ endo geometries in $(A6T6)_1$, while reverse is the case in $(A6T6)_2$, which is ~10 kcal/mol higher in total energy. In Figure 4 is shown stereopairs corresponding to these two structures. The energy component analyses diagrams corresponding to these two structures are shown in Figure 5a and b, respectively. Not surprisingly, the phosphate–phosphate and sugar–phosphate interactions in the two structures are similar, since they have an equal number of C2′ endo and C3′ endo sugars. The nature of the sugar pucker does not greatly influence the interaction of adenine with either 5′- or 3′-phosphate. However, in the case of thymine, the situation is quite different. Figure 5a

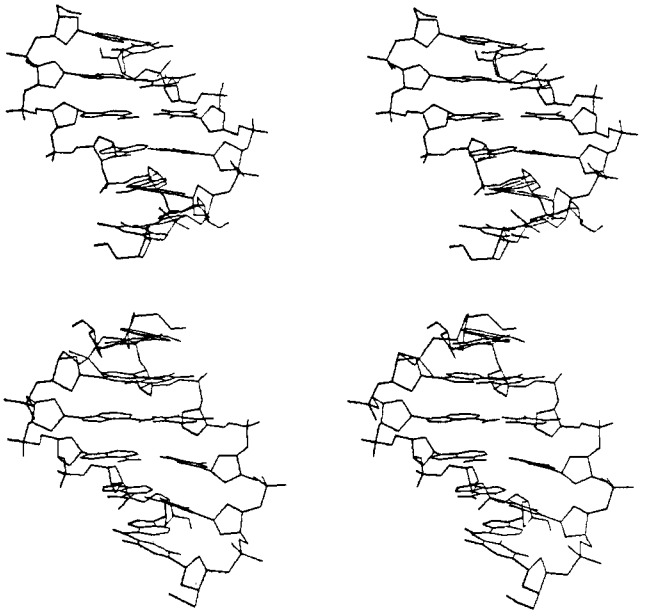

Figure 4. Stereopairs of two dA_6·dT6 structures (a, top) $(A6T6)_1$ and (b, bottom) $(A6T6)_2$. These two structures correspond to the component analysis diagrams of Figure 5 respectively.

and b reveals that thymine 5′-phosphate interactions differ in the range of roughly 1–2 kcal/mol, while the corresponding interactions involving 3′-phosphates differ by about 1 kcal/mol. Thus, thymine–phosphate interactgions are seen to be influenced to a much larger extent by the nature of the sugar geometries than are the adenine–phosphate interactions.

Further, we have carried out a component analysis on $(A6T6)_8$ (Figure 5c), which is less than a kilocalorie/mole higher in total energy than $(A6T6)_2$. Due to the intrinsic stability of C2′ endo sugars over C3′ endo sugars in a DNA structure,[19] the total internal energy of various groups is higher by about 2 kcal/mol in $(A6T6)_2$ than in $(A6T6)_8$. This difference is compensated by interactions between phosphates, sugars, and bases. Comparing Figure 5b and c, we find that the base-stacking interactions in the two structures are nearly the same. In $(A6T6)_8$, the favorable sugar–phosphate interactions (compasred to $(A6T6)_2$) are offset by thymine–phosphate interactions.

Discussion and Conclusions

Given the simplicity of our molecular mechanical calculations, it is encouraging that they give results that are generally consistent with experimental data. In the case of $d(ATATAT)_2$, a mixed sugar pucker model (AT_4) is somewhat more stable than the corresponding uniform sugar pucker (AT_1) model by ~2 kcal/mol, consistent with the X-ray structure (mixed sugar pucker) observed for $d(pApTpApT)_2$. In $d(TATATA)_2$, the uniform sugar pucker model (TA_1) is more stable than the mixed (TA_4) by ~3 kcal/mol. Taken together, these numbers would suggest that poly(d(AT))·poly(d(AT)) would be slightly more stable in a uniform than in a mixed sugar pucker geometry. However, this energy difference is small and, even if taken literally, would suggest a nonnegligible fraction of the mixed sugar pucker conformation. The results of the energy component analysis also suggest *why* high salt concentrations stabilize the mixed sugar pucker conformation, in that phosphate–phosphate repulsions are an energy term favoring a uniform C2′ endo geometry. These calculations suggest that in poly(d(AT))·poly(d(AT)), uniform and mixed sugar pucker conformations are nearly equal in energy and that high salt (implying lessened phosphate–phosphate repulsion) will tend to favor the mixed sugar conformation observed in $d(ATAT)_2$,[3,4] in which the adenine sugar is C3′ endo and the thymine sugar is C2′ endo. Part of the preference for mixed sugar geometries (thymine–phosphate interaction) also has an electrostatic component, but, because it is a mucher shorter range interaction without intervening solvent/counterions, it would be

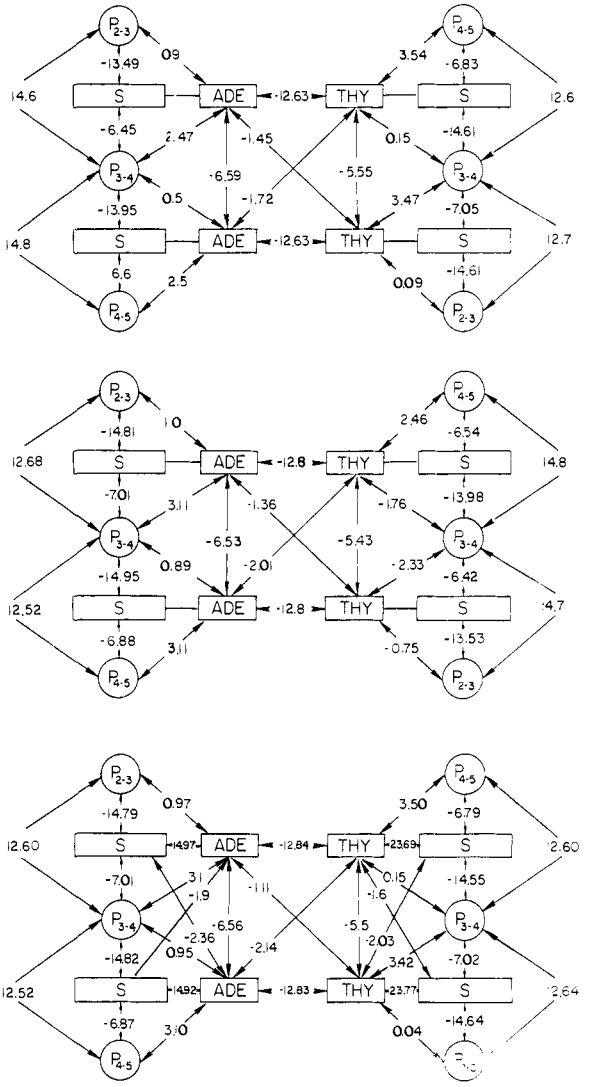

Figure 5. Component analysis of the interaction energies in the central two base pairs of three dA_6·dT_6 structures (a, top) with sugars attached to purine and pyrimidine bases having C3′ endo and C2′ endo geometries, respectively, (b, middle) with sugars attached to purine and pyrimidine bases having C2′ endo and C3′ endo geometries, respectively, and (c, bottom) with sugars on each strand being constrained to C2′ endo puckers.

expected to be much less effected by salt conditions.

The GC polymers $d(CGCGCG)_2$ and $d(GCGCGC)_2$ are found to be much more stable in uniform compared to mixed sugar geometries, apparently consistent with experiments which suggest that under salt conditions which do not favor a Z conformation, only a single ^{31}P resonance is found for poly(d(GC))·poly(d(GC)), in contrast to comparable studies of poly(d(AT))·poly(AT). We suggest that this is due to the much less favorable cytosine than thymine base–backbone interactions in the mixed sugar pucker conformation. A very recent NMR study of $d(CGCGCG)_2$ at low salt has suggested that the individual sugars in this structure are 70–80% C2′ endo for both guanine and cytosine sugars. This suggests that a structure with *one* of the sugars of $d(CGCGCG)_2$ repuckered to C3′ endo will be 0.5–0.8 kcal/mol less stable than that with all C2′ endo sugars. We thus carried out two additional sets of calculations on this molecule: first, we restrained all sugars to be C2′ endo, refined, removed the restraints, and rerefined (CG_5), and secondly we restrained the central cytosine and guanine sugars (CYT2 and GUA2) on the first strand to C3′ endo and the remaining sugars to C2′ endo, refined, removed the restraints, and rerefined (CG_6). In CG_6, GUA2 remains C3′ endo but CYT2 has an intermediate O1′ endo geometry. It is encouraging that the energy difference between the pure C2′ endo structure and the one with one C3′ endo and one O1′ endo sugar is 1.0 kcal/mol,

qualitatively consistent with the percent C2′ endo inferred by Cheng et al.[23] However, it is not clear how consideration of some percent of O1′ endo character would effect the NMR analysis.

Cohen et al.[12] found a ^{31}P doublet in poly(d(AT), poly(d(AU^{5Br})), and poly(d(IC)) and singlets in poly(d(AU)) and poly(d(GC)). One could imagine the observation of ^{31}P singlets even in structures with different sugar puckers or ^{31}P doublets despite uniform sugar puckers, which come from differences in other backbone angles. However, if we assume that the ^{31}P pattern is related to sugar pucker differences, our calculations allow one to rationalize the surprising difference between poly(d(AU)) and the two 5-substituted polymers poly(d(AT)) and poly(d(AU^{5Br})), since we have noted the importance of the 5 substituent in the attractive thymine–phosphate interaction which stabilizes the alternating sugar geometries. The reason for the difference between poly(d(GC))·poly(d(GC)) and poly(d(IC))·poly(d(IC)) cannot be surmised without detailed calculations on poly(d(IC)) structures.

Our calculations find that $dA_6 \cdot dT_6$ prefers a conformation in which the thymine sugars have C3′ endo puckers and adenine sugars have C2′ endo puckers over the uniform sugar model (by 2 kcal/mol), over a conformation in which the adenines have C3′ endo sugars and thymine C2′ endo sugars (by 10 kcal/mol), and over a conformation in which both strands have alternating C2′ endo–C3′ endo conformations (by 7 kcal/mol). These results are consistent with the experimental observation by Raman spectroscopy of an equal amount of C2′ endo and C3′ endo sugar puckers in poly(dA)·poly(dT). However, the calculations are able to go further and suggest that the model with thymine sugars C3′ endo and adenine sugars C2′ endo is the *only* one more favorable than the uniform C2′ endo models. This result is consistent with our earlier, less extensive calculations on $dA_6 \cdot dT_6$ and $dA_{12} \cdot dT_{12}$, where we noted a tendency for thymine C2′ endo → C3′ endo sugar repuckering, even during unconstrained energy minimization starting with uniform C2′ endo conformations. The results of our calculations are in contrast with those of Arnott et al.,[19] who suggest that the conformation of poly(dA)·poly(dT) had adenines C3′ endo and thymines C2′ endo, based on a model building fit to fiber diffraction data. They also rejected the model in which the sugars attached to adenine and thymine bases had C2′ endo and C3′ endo puckers, respectively, because of short interstrand contacts between neighboring unpaired A and T bases, even though such a model was shown to give an *R* factor similar to that of the proposed model.[21] Our investigations *have* obtained a model of the latter type which is energetically more favored than the one corresponding to fiber diffraction data. We stress that our model is "B-DNA"-like in its properties, unlike that of ref 19, and note that the existence of C3′ endo sugar within a B-DNA helix does not imply an "A-DNA"-like geometry of the rest of the helix.

In this context, a referee has suggested that we calculate the fiber diffraction patterns expected for the various $dA_6 \cdot dT_6$ structures presented here. We have not done so for two reasons. Firstly, the limits of fiber diffraction data would not allow a definitive establishment of geometry, whatever the result of the calculation. Secondly, the fiber diffraction study may not be relevant to the solution structure, and a definitive NMR study of a $dA_n \cdot dT_n$ polymer such as carried out by Cheng et al.[23] on $d(CGCGCG)_2$ is needed to determine unequivocably the conformation(s) of poly(dA)·poly(dT). It is intriguing that the calculations on $d(ATATAT)_2$ suggest sugar puckers (A-C3′ endo, T-C2′ endo) whereas those on $dA_6 \cdot dT_6$ suggest the opposite sugar puckers (A-C2′ endo, T-C3′ endo). We have been able to show that both results can be consistently explained by invoking a stabilizing thymine–phosphate interaction which occurs only when the sugar at the 5′ side of the thymine is C3′ endo rather than C2′ endo. This has structural consequences, also, in that in the lowest energy conformation of $dA_6 \cdot dT_6$, only the two thymines on the 5′ endo, the last of which has no 5′-phosphate at all to interact with, are in O1′ endo rather than C3′ endo conformations, with the remaining thymines C3′ endo.

In summary, we have noted three important factors which influence sugar puckering properties in B-DNA helices: (a) an intrinsic preference of ~0.5 kcal/mol for deoxyribose rings for C2′ endo geometries, (b) greater phosphate–phosphate repulsions when a sugar is C3′ endo, which we calculate to be ~2 kcal/mol but whose magnitude is likely to be exaggerated in our calculations, and (c) more attractive thymine–phosphate interactions, when the sugar on the 5′ end of the thymine is C3′ endo than when it is C2′ endo, of 1–2 kcal/mol. Again, this latter figure is likely to be an upper bound, but we suggest that the effect is qualitatively important in understanding sequence-dependent DNA sugar puckering. A final point to be emphasized is the fact that the presence of C3′ endo sugars does *not* imply an A-DNA-like helix, as illustrated in Figures 2 and 4.

Acknowledgment. We gratefully acknowledge the support of the National Cancer Institute (Grant CA-25644) in this research and Larry Waggoner, who carried out some of the earlier calculations in this study. The use of the facilities of the UCSF Computer Graphics Lab (R. Langridge, director, and T. Ferrin, faculty manager), supported by NIH RR-1081, is also gratefully acknowledged.

Registry No. d(ATATAT), 79695-11-7; d(TATATA), 79695-10-6; d(GCGCGC), 76186-50-0; d(CGCGCG), 58927-26-7; $d(A_6) \cdot d(T_6)$, 80601-48-5; deoxyribose, 533-67-5.

(23) Cheng, D. M.; Kan, L.-S.; Frechet, D.; Tso, P.; Uesugi, S.; Shida, T.; Ikehara, M. *Biopolymers* **1984**, *23*, 775.

Ab Initio Study of the Photodissociation of Nitrosoalkanes and Nitrosamines

Renzo Cimiraglia,* Maurizio Persico, and Jacopo Tomasi

Contribution from the Istituto di Chimica Fisica dell'Università di Pisa, Via Risorgimento 35, I-56100 Pisa, Italy. *Received July 5, 1984*

Abstract: The primary step of the photodissociation in CH_3NO and NH_2NO has been investigated by ab initio SCF-CI techniques. The first singlet excited state presents a barrier along the dissociation coordinate both for CH_3NO and for the planar form of NH_2NO. However, NH_2NO prefers a twisted and pyramidalized structure in S_1, with no barrier to dissociation. These features of the S_1 surface explain the observation of a vibrational structure in the $n \rightarrow \pi^*$ spectrum of $(CH_3)_2NNO$, dominate the dissociation mechanism (vibrational predissociation vs. intersystem crossing or internal conversion), and may influence the quantum yields in the condensed phase.

In this paper we present an ab initio SCF-CI study of the primary step in the photochemistry of nitroso compounds, i.e., the dissociation of the X–NO bond. We have concentrated on two of the most important classes of nitroso compounds, nitrosoalkanes

0002-7863/85/1507-1617$01.50/0