Volume 8 Number 1 1980 Nucleic Acids Research

Loopstructures in synthetic oligodeoxynucleotides[1]

Cornelis A.G.Haasnoot, Jeroen H.J.den Hartog, Jan F.M.de Rooij, Jacques H.van Boom and Cornelis Altona

Gorlaeus Laboratories, State University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

Received 17 October 1979

ABSTRACT.

A comparative nuclear magnetic resonance study of the hydrogen-bonded imino protons in a series of synthetic DNA fragments is presented. The fragments ATCCTA(T_n)TAGGAT are in principle capable of forming either a self-complementary hairpin loop structure (monomer form) or an interior loop structure (dimeric form). It has been shown[2], that for n=1 only the dimer structure is present in aqueous solution, whereas the exclusive existence of the hairpin loop structure is indicated for n=3, 4 & 5. Surprisingly, for n=2 two different structures appear to be present in solution. Concentration studies show that both monomers and dimers exist side by side in this case. Hairpins as well as interior loops form extra "melting sites" in addition to the well-known fraying phenomenon at the terminus of the double helix.

INTRODUCTION.

Double helical complexes formed from tailor-made synthetic oligodeoxynucleotides are convenient models for the investigation of the structural features and dynamics of base recognition in nucleic acids. In a previous paper[2] we investigated the helix-coil transition of the (partially) self-complementary oligodeoxynucleotides fragment 12 (fig. 1a, n=0) and fragment 13 (fig. 1a, n=1) by monitoring the Watson-Crick hydrogen-bonded ring imino protons by high-field nuclear magnetic resonance (NMR) techniques. It was found[2] that the insertion of a TxT wobble pair (*i.e.* an interior loop) in the middle of the strands introduces an extra melting site in the double helical structure (Form II, fig. 1b; see also fig. 2), which gives rise to an overall lowering of the NMR "melting" temperature of the double helix.

The relative thermodynamic stability of the various types of loops which can occur in DNA is a matter of debate. For RNA it has been found that interior loops are more stable than hairpin loops[3]. If this were true for DNA, elongation of the

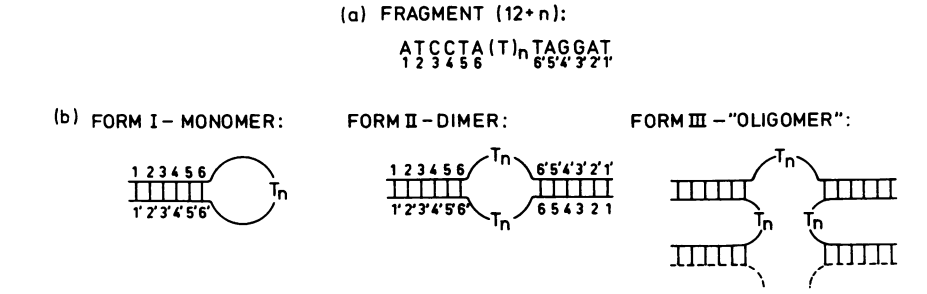

Figure 1. (a) General formula denoting the base-sequence of the synthetic oligonucleotides studied.
(b) Three possible secondary structures for these oligonucleotides. Base pair numbering as shown; note the pairwise equivalency of the base pairs in Form II and III because of the symmetry of the double helix

non-complementary site of thymidylyl residues in the middle of fragment 13 (T_n, n=1,2,...) would not give rise to major conformational changes, *i.e.* Form II (fig. 1b) being preferred for all values of n. On the other hand, if for a certain chain length n the hairpin loop (Form I, fig. 1b) becomes more stable than the interior loop one expects a discontinuity in conformational behaviour to occur.

In the present paper we wish to report the results of a comparative proton magnetic resonance study of fragments 13 (n=1), 14 (n=2), 15 (n=3), 16 (n=4) and 17 (n=5), undertaken in

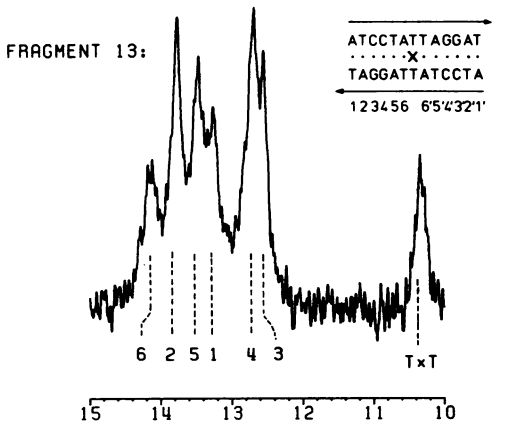

Figure 2.
Low field region of the 360 MHz NMR spectrum of fragment 13 (*ca.* 3 mM), taken at 10 °C.

order to evaluate the conformational behaviour of these fragments as function of chain length and temperature. The possible formation of "oligomers" (Form III, fig. 1b) is also investigated.

MATERIALS AND METHODS.

The oligodeoxynucleotides were synthesized *via* an improved phosphotriester approach[4]. Full details of preparation and purification are published elsewhere[5]. The DNA fragments were converted into their Na^+-salts by slow passage over a column (12 cm x 1 cm^2) of Dowex cation-exchange resin, Na^+-form; elution was with distilled water. Lyophilization gave the nucleotide material as white fluffy compounds.

The NMR samples were prepared by dissolving the appropriate amount of the oligodeoxynucleotide in 270 µl of a buffer solution containing 0.5 M NaCl, 0.1 M sodium cacodylate (pH = 7.0) and 5 % v/v D_2O. High-resolution ^1H NMR spectra were recorded on a Bruker HX-360 NMR spectrometer interfaced with a BNC-12 computer system and equipped with a standard temperature controller, using micro NMR tubes with a cylindrical cavity (Wilmad 508-CP). The spectrometer was operated in the correlation mode[6], using an internal D_2O field/frequency lock. Chemical shifts were measured relative to the solvent H_2O peak and converted to DSS-reference by correcting for the H_2O to DSS chemical shift, using the appropriate temperature calibration curve. All reported shifts are denoted according to the δ-convention, *i.e.* positive values represent shifts to *lower* field.

RESULTS.

Secondary structure of the synthetic DNA fragments.

The chemical shift of a given hydrogen-bonded imino proton in an intact double-helical DNA fragment is mainly determined by the effects of ring current and diamagnetic anisotropies[7,8] of the aromatic bases which it links together and those directly above and below in the vertical stacking array (nearest neighbour effects), with the next-nearest neighbours playing a relatively minor role. Various additivity rules have been proposed as an

aid in predicting imino-proton shifts in RNA and DNA helices[9,10] based upon effects exerted by nearest-neighbour bases. Barring gross changes in either geometry or stability of the double helix for different values of n, one should expect that the DNA fragments presently under investigation display a very similar NMR-pattern, as all fragments contain the same sequence of six hydrogen-bonded base pairs and differ only in the number of thymidylyl residues in the middle of the strands. Our expectation is indeed borne out by experiment. Let us leave aside for the moment the results for fragment 14 (for reasons to be discussed below) and concentrate on the six signals displayed by fragments 13, 15, 16 and 17 in the 12.6 - 14.3 ppm region (figs. 2 & 3). The chemical shifts observed at low temperature are collected in Table 1. It is readily seen that the signal patterns of the three larger fragments are quite similar to that of fragment 13 and we therefore conclude that our previous assignment of the

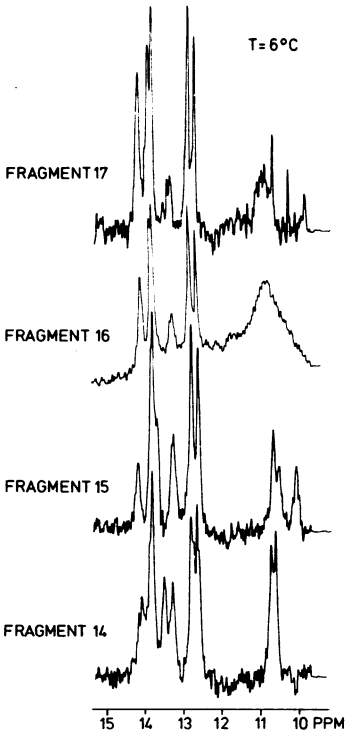

Figure 3.
The 360 MHz ^{1}H-NMR spectra (low field part) of fragments 14, 15, 16 and 17 (∿3 mM), recorded at 6 ^{o}C.

Table 1.

Chemical shifts of fragments 13-17 (∿3 mM) at 2 °C, estimated accuracy ∿0.05 ppm.

	Base pair 1		Base pair 2		Base pair 3		Base pair 4		Base pair 5		Base pair 6	
Set Fragment	*A* monomer	*B* dimer	*A* monomer	*B* dimer	*A* monomer	*B* dimer	*A* monomer	*B* dimer	*A* monomer	*B* dimer	*A* monomer	*B* dimer
13		13.28		13.78		12.53		12.67		13.45		14.19
14	13.34	13.30	∿13.79†	∿13.79†	12.60	12.55	12.76	12.70	∿13.79†	13.47	--*	14.24
15	13.35		13.84		12.64		12.82		13.66		14.30	
16	13.34		13.85		12.65		12.82		13.78		14.21	
17	13.34		13.85		12.64		12.82		13.75		14.25	

†Specific assignment not possible due to serious overlap.

*Considered to be sterically impossible.

proton signals in fragment 13 is also valid for fragments 15, 16 and 17. It should be stressed that the assignment of fragment 13 was solely based on experimental grounds using the principle of chemical modification[2]. The "melting" behaviour of fragments 15, 16 and 17 (*vide infra*) is fully consistent with this assignment.

A more detailed comparison of the spectra of the above mentioned fragments, taken at various temperatures, reveals the existence of small, but persistent, differences in chemical shift displayed by fragment 13 on the one hand and fragments 15, 16 and 17 on the other hand. This fact in itself would hardly deserve comment but for the most interesting finding that the spectrum of fragment 14 shows *more than six signals* in the 12.6-14.3 ppm region. Close inspection of the NMR-temperature profiles of fragment 14 (not shown) reveals that these spectra must be analyzed in terms of two subsets. All proton resonances (except those due to the A-T base pair 6) are split into two components, of which the ones at lower field (set A) correspond with the resonances found in fragments 15, 16 and 17; the high field components (set B) practically coincide with the resonances occurring in fragment 13. Moreover, the NMR-temperature profiles also show that the two subsets "melt out" independently of each other. The signals belonging to set A disappear at slightly

lower temperatures than the corresponding ones of set B. The exchange between the two subsets is thus found to be slow on the NMR time scale over the entire temperature range monitored (0 °C to 43 °C). From these observations we draw the rather surprising conclusion that *fragment 14 occurs in two different molecular species in aqueous solution.*

As mentioned in the introduction, three different secondary structures may be envisaged:

(1) A hairpin loop structure, the molecule folds back onto itself (monomeric Form I, fig. 1b).

(2) An interior loop structure in which two molecules participate, forming a head-to-tail complex (dimeric Form II, fig. 1b).

(3) Oligomeric structures (Form III, fig. 1b), several ($\geq$3) molecules are involved in complex formation.

From the accordance in chemical shifts with fragment 13, the high-field subset B of fragment 14 is assigned to the same secondary structure (interior loops; Form II, fig. 1b) as was derived[2] for fragment 13. This leaves the low-field subset A to be ascribed to either a hairpin loop or to an "oligomer". In order to decide between the two alternatives, NMR-temperature profiles of fragment 14 at several concentrations were recorded. Figure 4 shows a representative concentration profile of fragment 14 at *ca.* 18 °C. It is seen that at low concentrations (0.5-1 m$\underline{M}$) the signals belonging to subset A predominate; it is at higher concentrations that the intensity of the signals due to the high-field subset B gains importance. In other words, the equilibrium between the two secondary structures of fragment 14 shifts towards the dimeric form at higher concentrations. This implies that the signals of subset A represent a secondary structure with lower order reaction kinetics than the dimeric form, *viz.* the monomeric Form I (fig. 1b).

Of course, it should not be deduced that the hairpin loop structure of fragment 14 contains only the two central T-bases within the loop. Such an arrangement is thought to be sterically impossible[11]. Rather, a structure may be envisaged in which the A and T residues at position 6 are no longer hydrogen-bonded to each other, but form a part of the hairpin loop.

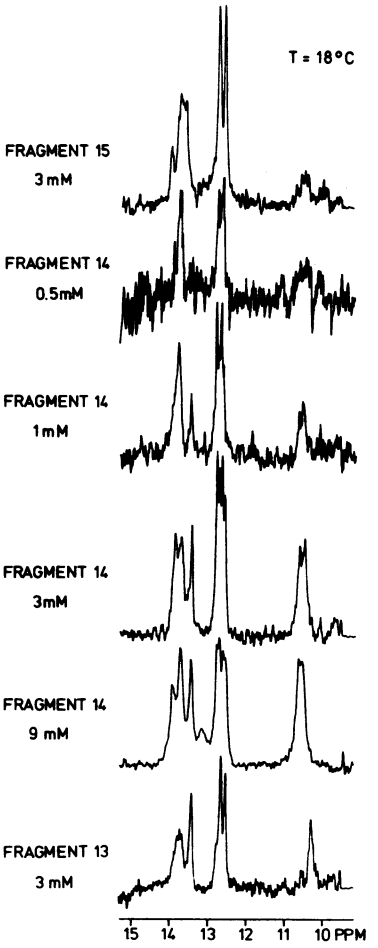

Figure 4.
The 360 MHz NMR concentration-profile of fragment 14 recorded at 18 °C. The corresponding spectra of fragment 15 (∿3 mM, top) and fragment 13 (∿3 mM, bottom) are included for purposes of comparison.

The NMR concentration profile of fragment 14 (fig. 4) may further be used to summarize the differences in chemical shift between a loop structure and a dimer structure with an interior loop. On lowering the concentration, the resonance of base pair 5 (at 13.5 ppm) decreases in relative intensity, accompanied by an increase in intensity of the resonance at *ca.* 13.8 ppm. Combining this finding with the forementioned shifts to lower field (*vide supra*), we may state that the difference in chemical shifts between a hairpin loop structure and a dimer structure with an interior loop is small ($\leq$ 0.1 ppm) with an exception for

base pair 5 for which the downfield shift amounts to approximately 0.3 ppm (Table 1).

Close examination of the spectra of fragments 15, 16 and 17 (see *e.g.* fig. 3) shows that these spectra exhibit all features ascribed above to the monomer structure. Therefore it is concluded that the fragments 15-17 adopt a hairpin-loop secondary structure in aqueous solution. Our preliminary molecular weight determinations by ultra-centrifuge experiments carried out at 10 °C support this conclusion as it is evinced that the molecular weight of fragment 12 (dimer[2]) is *higher* than that of fragment 16 (monomer).

The upfield resonances (10-11 ppm) of the "free" imino protons (T•) of the thymidylyl residues in the middle of the strand deserve some attention. Fragment 15 in particular displays three well-resolved signals in this range, indicating a (hairpin-) structure in which each of the three T• protons takes up a specific non-equivalent site, in some measure protected against rapid exchange with the solvent protons. We note that the relative intensity of the "free" imino protons *decrease* upon further insertions of thymidylyl residues in the middle of the strand (fragment 15 → fragment 17, fig. 3). Tentatively, we take this to indicate that upon enlarging the loop loses rigidity and becomes more penetrable to the water molecules (*vide infra*).

Still, several puzzling questions remain unsolved for the time being. For example, one expects to observe several signals from the thymidylyl imino protons (T•) in the hairpin loop of fragment 14, whereas only a single line that can be ascribed to this species is seen: the intensity of the second (downfield) T• signal increases with increasing concentration and is therefore assigned to the dimer structure.

Another unexpected phenomenon concerns the intensity of the signal due to base pair 6. Figure 3 clearly shows that, at a given temperature, this intensity increases in the series fragment 15 → fragment 17, contrary to the rationale that a smaller hairpin loop should be tighter and induce a greater stability and/or less exchange with the surrounding water than might be the case with a larger hairpin. Still, the "melting" behaviour of this particular signal accords with expectations (*vide infra*).

Further studies are currently being planned in order to shed more light upon these problems.

Melting behaviour of the loop structures.

Figure 5 presents the high-resolution NMR spectrum of fragment 15 as a function of temperature. On raising the temperature from *ca.* 1 to 15 °C the signal at *ca.* 13.3 ppm disappears (A-T base pair 1). In the range between about 10 to 20 °C the intensity of the signal at *ca.* 13.8 ppm collapses from two to one proton. As the relative intensity of the signal at 14.3 ppm (A-T base pair 6) has not changed yet, the faded signal must be

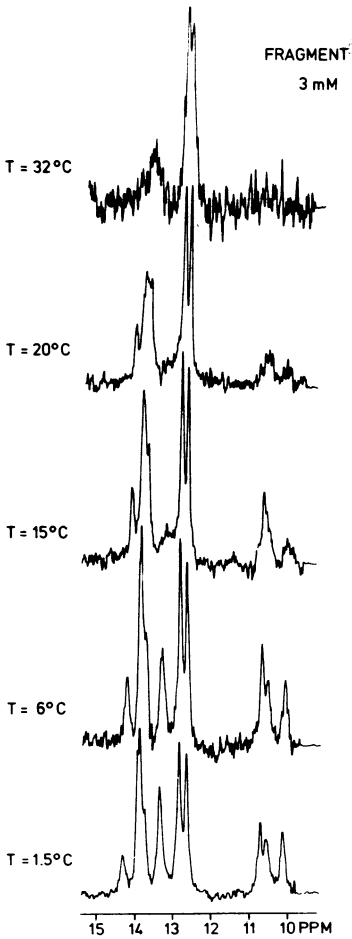

Figure 5.
The 360 MHz proton NMR spectra of fragment 15 (∿3 mM) as a function of temperature (only a representative set of spectra is shown).

due to A-T base pair 2. Further raising of the temperature (up to *ca.* 26 °C) causes A-T base pair 6 (at 14.3 ppm) to disappear, followed next by A-T base pair 5 (34 °C). Finally, the two central C-G pairs 3 and 4 (centered around 12.7 ppm) disappear simultaneously (*ca.* 40 °C).

The NMR-temperature profiles of fragment 16 and 17 (not shown) are very similar to that of fragment 15. Except for A-T base pair 6 (*vide infra*), the hydrogen-bonded imino protons of the duplex "melt out" within the same temperature range as was found in fragment 15, the two central C-G pairs are the last ones being disrupted (at about 40 °C). In other words: the breaking of the Watson-Crick hydrogen bonds occurs on both sides of these C-G pairs. This observation implies that the loop in itself forms *an extra melting site* in addition to the normal "fraying" phenomenon[12] starting from the terminus of the double helical structure.

Of interest in this context are the differences between fragments 15, 16 and 17 emerging from the melting behaviour of A-T base pair 6 and the "free" imino protons of the thymidylyl residues which constitute the loop. In fragments 16 and 17, the signal due to A-T base pair 6 already disappears below 20 °C, whereas in the tightest loop (fragment 15) this signal is observed up to a temperature of about 26 °C. Moreover, in going from fragment 15 to fragment 17 the "melting" temperatures of the T• proton signals display a regular decreasing trend (from 22 °C to 10 °C). These findings perhaps endorse the higher accessibility of larger *vs.* smaller loops with respect to the water molecules and point again to the forementioned decrease in rigidity of the loop upon enlarging.

DISCUSSION.

The results reported in this study reveal several interesting points. The NMR temperature-concentration profiles of fragment 14 (fig. 4) show that at *ca.* 3 mM the monomeric and dimeric form have an approximately equal free energy content. Yet in the monomeric form the double strand may involve only 5 base pairs, as a loop consisting of only two monomeric units is said to be

sterically impossible[11]. As the enthalpy term (ΔH^{o}) in the free energy content (ΔG^{o}) of the double helix, in principle, depends on the chain length (as well as on the base composition and the base sequence) it is evident that the entropy term (ΔS^{o}) plays a decisive role in this process. This may be visualized when the formation of a double helix is dissected into two steps[11]: the formation of the first base pair (initiation) and the formation of subsequent base pairs (propagation). The probability of initiation is much larger for the intramolecular process (leading to the hairpin loop) than for the intermolecular (dimerization) process.

This rationalization is further corroborated by the NMR-temperature profile of the duplex formed by the two hexanucleotides (6+6) which designate the ends of the synthetic oligonucleotides studied (*i.e.* ATCCTA and TAGGAT). It is found[13] that the two central G-C base pairs are the last ones to disappear from the spectrum, as is the case for fragment 15, 16 and 17 (*vide supra*). However, the temperature at which this occurs is about 15 ^{o}C lower for the hexanucleotide mixture than for the fragments 15, 16 and 17. As the enthalpy term (apart from possible small extra stacking interactions in the central T-region of the fragments 15-17) must be equal for the various molecules under consideration, the difference in stability is mainly due to the difference in entropy of formation of open-ended *vs.* hairpin-closed duplexes.

The relative stabilities of the hairpin loops in DNA fragments 15-17 merit some discussion. In all three fragments the signals due to the central C-G pairs collapse at about the same temperature (*ca.* 40 ^{o}C). Yet, the "melting" behaviour of the T• bases constituting the loops and of base pair 6 (next to the hairpin loops) appears to be affected by the size of these loops. The data suggest that the number of nonbonded bases in the loop (three to five in our case) has a strictly *local* influence on the stability of base pairs and does not affect the overall stability of duplex regions removed by two or more bases from the hairpin loop.

Summarizing, the apparently decreasing stability of the hairpin in going from fragment 15 to 17 is in accordance with

the rules originally deduced by Tinoco *et al.*[11] in which hairpin loops that contain three bases are assigned a "stability number" of -5 kcal/mole (ΔG^{o}) and hairpins of 4-7 bases are less stable by 1 kcal/mole. It is noted that these rules, intended for the estimation of secondary structures in RNA molecules, were derived from experimental data concerning *DNA and oligo-d(AT)* melting behaviour. Remarkably, in a later publication[14] it was asserted, based upon data from *RNA* sequences, that a hairpin loop of three bases closed by an A-U base pair is *less* stable than a hairpin consisting of 4-5 bases, thus contrary to the present conclusions. In our opinion these findings support the thesis that experimental data derived from DNA fragments should not be used to predict the behaviour of RNA molecules and *vice versa*. Perhaps the rationale that explains in part the intriguing differences between DNA and RNA on various levels of biological activity should be sought in the outspoken difference in conformational flexibility of the deoxyribose- *vs.* the ribose sugar ring[15,16].

ACKNOWLEDGEMENT.

This research was supported by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement for Pure Research (Z.W.O.). We wish to thank Mr. K. Dijkstra and Dr. R. Kaptein for their technical assistance.

REFERENCES.

1. This paper is number XIV in the series "Nucleic Acids Constituents" from this laboratory. For number XIII see: Leeuw,H.P.M. de, Haasnoot,C.A.G. & Altona,C. (1980) *Isr. J. Chem.* - accepted for publication.
2. Haasnoot,C.A.G., Hartog,J.H.J. den, Rooij,J.F.M. de, Boom,J.H. van & Altona,C. (1979) *Nature 281*,235-236.
3. Gralla,J. & Crothers,D.M. (1973) *J. Mol. Biol. 78*, 301-319.
4. Boom,J.H. van, Burgers,P.M.J. & Deursen,P.H. van (1976) *Tetrahedron Letters*, 869-872.
5. Rooij,J.F.M. de, Wille-Hazeleger,G., Deursen,P.H. van, Serdijn,J. & Boom, J.H. van (1979) *Recl. Trav. Chim. Pays-Bas 98*, 537-548.
6. Dadok,J. & Sprecher,R.F. (1974) *J. Magn. Res. 13*, 243-248.
7. Pople,J.A. (1964) *J. Chem. Phys. 41*, 2559-2560.
8. Giessner-Prettre,C. & Pullman,B. (1976) *Biochem. Biophys. Res. Commun. 70*, 578-581.
9. Borer,P.N., Kan,L.S. & Ts'O,P.O.P. (1975) *Biochemistry 14*, 4847-4863.
10. Arter,D.B. & Schmidt,P.G. (1976) *Nucl. Acids Res. 3*, 1437-1447.
11. Tinoco,I., Uhlenbeck,O.C. & Levine,M.D. (1971) *Nature 230*, 362-367.
12. Patel,D.J. & Hilbers,C.W. (1975) *Biochemistry 14*, 2651-2656.

13. Haasnoot,C.A.G., Hartog,J.H.J. den & Altona,C. - *unpublished results.*
14. Tinoco,I., Borer,P.N., Dengler,B., Levine,M.D., Uhlenbeck,O.C., Crothers, D.M. & Gralla,J. (1973) *Nature New Biology 246,* 40-41.
15. Altona,C., Boom,J.H. van & Haasnoot,C.A.G. (1976) *Eur. J. Biochem. 71,* 557-562.
16. Altona,C. & Sundaralingam,M. (1973) *J. Amer. Chem. Soc. 95,* 2333-2344.