Volume 2 number 9 September 1975 Nucleic Acids Research

The sequence dependence of circular dichroism spectra of DNA duplexes

Struther Arnott

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received 16 May 1975

ABSTRACT

The three satellite DNAs of *Drosophila virilis*, that approximate to poly d(CAAACTA)·poly d(TAGTTTG), poly d(TAAACTA)·poly d(TAGTTTA), poly d(CAAATTA)·poly d(TAATTTG), the satellite DNA of *Drosophila melanogaster* that approximates to poly d(AATAT)·poly d(ATATT), the synthetic DNA duplexes, poly dG·poly dC, poly d(AT)·poly d(AT), poly d(AAT)·poly d(ATT), poly d(AAC)·poly d(GTT), poly d(TAC)·poly d(GTA) and the block copolymer $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$ all have circular dichroism spectra consistent with the propositions that they have the same molecular geometry in solution and that it is the kind and frequency of nucleotide triplet sequences that determines their spectral characteristics. Poly dA·poly dT is apparently an exception.

INTRODUCTION

DNA duplexes with different distributions of nearest neighbouring bases have distinctive ultraviolet circular dichroism spectra.[1,2] The spectra[3] of the *Drosophila virilis* satellites *II* and *III*, which have the same composition but are sequence isomers, are good examples of this. The CD spectrum of a DNA of random nucleotide sequence is apparently[4] a weighted sum of the spectral contributions of the ten different hydrogen-bonded dimer pairs (e.g. d(AC)·d(GT)). However the analogous calculations that Gray and Gall[3] have made for four *Drosophila* satellites show that this cannot generally be the case. The failure of the theory could be because the satellite DNAs and some or all of the synthetic DNAs from which the dimer contributions are derived are not isogeometrical, or because triplet or higher nucleotide sequence components are needed. Although it is now known from fiber diffraction studies that poly d(AT)·poly d(AT)[5] and poly d(GC)·poly d(GC),[5] poly dA·poly dT[6] and poly d(AAT)·poly d(ATT)[7] can indeed assume forms with markedly

different conformations from other DNAs, it is doubtful if these results should be invoked to support the first explanation of the anomalous CD results. Firstly, poly d(AT)·poly d(AT), poly d(GC)·poly d(GC) and poly d(AAT)·poly d(ATT) in fibers containing small amounts of salt have the classical *B*-DNA conformation above 92% relative humidity, indicating that the conformation in aqueous buffers is likely to be *B*. Secondly, X-ray fiber diffraction studies[8] of a number of satellite DNAs, including *D. virilis I*, show only classical conformations. It therefore appears that repetitious polyoligonucleotides are likely to have orthodox molecular geometries especially in solution.
[Poly dA·poly dT may be an exception since its distinctive structure in fibers (a ten-fold helix with h=0.325nm) persists at the highest relative humidities[6] and its measured CD spectrum[1] is the only one that is inconsistent with the additivity rule established below which accommodates ten other different DNA duplexes.]

It therefore appears more likely that CD spectra of DNA duplexes depend (at least) on what triplets of consecutive nucleotide pairs are present and this is the proposition I have tested assuming that poly dA·poly dT is the only geometrically peculiar polymer in the set of eleven considered.

DATA

If we consider that each Watson-Crick paired triplet of nucleotides embedded in an infinite polymer has a characteristic molar ellipticity $[\theta]_n$ then the molar ellipticity $[\theta]$ of any polyoligonucleotide with N nucleotide pairs in the repeating segment should be given by eqn.(1).

$$N[\theta] = \Sigma\ [\theta]_n,\ n = 1 \text{ through } N \qquad (1)$$

a) Poly dA·poly dT and poly dG·poly dC.

In principle the molar ellipticities [θ][d(AAA)·d(TTT)] and [θ][d(GGG)·d(CCC)] should (applying eqn.(1)) be obtainable directly from the CD spectra of poly dA·poly dT and poly dG·poly dC.

$$[\theta][\text{poly dG·poly dC}] = [\theta][\text{d(GGG)·d(CCC)}] \qquad (2)$$

$$[\theta][\text{poly dA·poly dT}] = [\theta][\text{d(AAA)·d(TTT)}] \qquad (3)$$

However Gray and Bollum[9] have shown that the poly dG·poly dC spectrum of Wells *et al.*[1] probably contains contributions from self-complexes of oligo dG and of oligo dC. I have therefore

used the corrected poly dG·poly dC spectrum from Gray and Bollum to provide eqn.(2). I have also assumed that poly dA·poly dT may have the *B'*-DNA[6] conformation in solution and that therefore it would be inappropriate to use the CD spectrum of Wells *et al.*[1] (see Fig. 1). Instead I have assumed that the block copolymer[10] $d(C_{15}A_{15})\cdot d(T_{15}G_{15})$ contains sufficient CG pairs to ensure that it is geometrically orthodox and that therefore eqn.(4) can be used to supply [θ][d(AAA)·d(TTT)] (see Fig. 1).

$$2[\theta][d(C_{15}A_{15})\cdot d(T_{15}G_{15})] = [\theta][d(AAA)\cdot d(TTT)] + [\theta][d(GGG)\cdot d(CCC)] \quad (4)$$

Eqn.(4) assumes that "end effects" will be negligible despite the short length of the block copolymer and that the contributions of two different triplet sequences at the junction of the two blocks can be ignored.

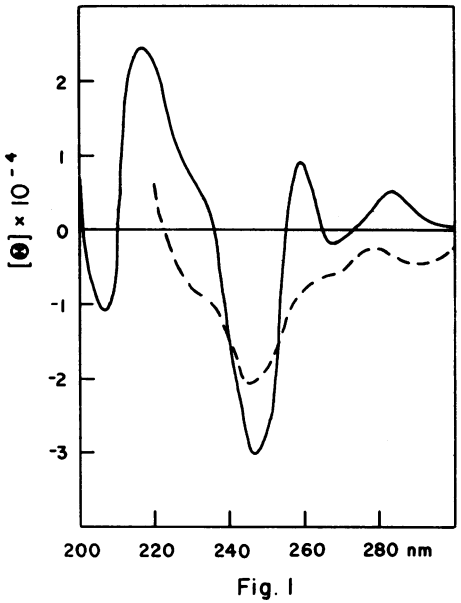

FIG. 1. CD spectra of poly dA·poly dT: (a) observed by Wells *et al.* (——); (b) calculated using eqn.(4) (- - - -).

b) Poly d(AT)·poly d(AT).

The CD spectrum of poly d(AT)·poly d(AT)[1] is the only poly-dinucleotide spectrum that is needed for our calculations. It provides the sum of two triplet components of molar ellipticity.

$$2[\theta][\text{poly d(AT)}\cdot\text{poly d(AT)}] = [\theta][d(ATA)\cdot d(TAT)] + [\theta][d(TAT)\cdot d(ATA)] \quad (5)$$

Because A and T are complementary in the Watson-Crick sense the two components turn out to be identical as they would in the case of poly d(GC)·poly d(GC) but not in the case of, say, poly d(AC)·poly d(GT).

$$2[\theta][\text{poly d(AC)}\cdot\text{poly d(GT)}] = [\theta][\text{d(ACA)}\cdot\text{d(TGT)}] + [\theta][\text{d(CAC)}\cdot\text{d(GTG)}] \quad (6)$$

c) Poly d(AAT)·poly d(ATT), poly d(TAC)·poly d(GTA), poly d(CAA)·poly d(TTG).

The CD spectra of polytrinucleotide duplexes each provide the sum of three triplet components. The ones that will be used in later calculations, poly d(AAT)·poly d(ATT),[11] poly d(TAC)·poly d(GTA)[1] and poly d(CAA)·poly d(TTG)[1] provide eqns.(7)-(9).

$$3[\theta][\text{poly d(AAT)}\cdot\text{poly d(ATT)}] = [\theta][\text{d(TAA)}\cdot\text{d(TTA)}] + [\theta][\text{d(ATT)}\cdot\text{d(AAT)}] + [\theta][\text{d(ATA)}\cdot\text{d(TAT)}] \quad (7)$$

$$3[\theta][\text{poly d(TAC)}\cdot\text{poly d(GTA)}] = [\theta][\text{d(CTA)}\cdot\text{d(TAG)}] + [\theta][\text{d(TAC)}\cdot\text{d(GTA)}] + [\theta][\text{d(ACT)}\cdot\text{d(AGT)}] \quad (8)$$

$$3[\theta][\text{poly d(CAA)}\cdot\text{poly d(TTG)}] = [\theta][\text{d(ACA)}\cdot\text{d(TGT)}] + [\theta][\text{d(CAA)}\cdot\text{d(TTG)}] + [\theta][\text{d(AAC)}\cdot\text{d(GTT)}] \quad (9)$$

d) Drosophila melanogaster *satellite DNA.*

Gall and Atherton[12] proposed that the sequence of this satellite approximates to poly d(AATAT)·poly d(ATATT). The molar ellipticity is therefore resolvable into the triplet components given in eqn.(10).

$$[\theta][\textit{D. melanogaster}] = [\theta][\text{d(TAA)}\cdot\text{d(TTA)}] + [\theta][\text{d(AAT)}\cdot\text{d(ATT)}] + [\theta][\text{d(ATA)}\cdot\text{d(TAT)}] + [\theta][\text{d(TAT)}\cdot\text{d(ATA)}] + [\theta][\text{d(ATA)}\cdot\text{d(TAT)}] \quad (10)$$

Combining eqns.(5),(7),(10) gives eqn.(11).

$$[\theta][\textit{D. melanogaster}] = (3/5)[\theta][\text{poly d(AAT)}\cdot\text{poly d(ATT)}] + (2/5)[\theta][\text{poly d(AT)}\cdot\text{poly d(AT)}] \quad (11)$$

In other words the molar ellipticity of *D. melanogaster* should be this particular linear combination of the molar ellipticities of poly d(AT)·poly d(AT) and poly d(AAT)·poly d(ATT).

e) Drosophila virilis *DNA satellite I.*

The sequence of this satellite approximates to poly d(CAAACTA)·poly d(TAGTTTG)[12] and therefore the components of the CD spectrum are given by eqn.(12).

$$7[\theta][\textit{D. virilis I}] = [\theta][d(ACA)\cdot d(TGT)] + [\theta][d(CAA)\cdot d(TTG)] + [\theta][d(AAA)\cdot d(TTT)] + [\theta][d(AAC)\cdot d(GTT)] + [\theta][d(ACT)\cdot d(AGT)] + [\theta][d(CTA)\cdot d(TAG)] + [\theta][d(TAC)\cdot d(GTA)] \quad (12)$$

By combining eqns.(2),(4),(8),(9),(12) we obtain eqn.(13) which shows that the CD spectrum of *D. virilis* DNA satellite *I* is also a particular linear combination of known CD spectra.

$$[\theta][\textit{D. virilis I}] = (3/7)[\theta][\text{poly } d(CAA)\cdot\text{poly } d(TTG)] + (3/7)[\theta][\text{poly } d(TAC)\cdot\text{poly } d(GTA)] + (1/7)\{2[\theta][d(C_{15}A_{15})\cdot d(T_{15}G_{15})] - [\theta][\text{poly dG}\cdot\text{poly dC}]\} \quad (13)$$

f) Drosophila virilis *DNA satellites* II *and* III.

Gall and Atherton[12] have assigned the sequences poly d(TAAACTA)·poly d(TAGTTTA) and poly d(CAAATTA)·poly d(TAATTTG) to satellite DNA *II* and *III* respectively. In these two cases the triplet components of the CD spectra (eqns.(14),(15)) *cannot* be recombined in terms of the known spectra of simpler polymers.

$$7[\theta][\textit{D. virilis II}] = [\theta][d(ATA)\cdot d(TAT)] + [\theta][d(TAA)\cdot d(TTA)] + [\theta][d(AAA)\cdot d(TTT)] + [\theta][d(AAC)\cdot d(GTT)] + [\theta][d(ACT)\cdot d(AGT)] + [\theta][d(CTA)\cdot d(TAG)] + [\theta][d(TAT)\cdot d(ATA)] \quad (14)$$

$$7[\theta][\textit{D. virilis III}] = [\theta][d(ACA)\cdot d(TGT)] + [\theta][d(CAA)\cdot d(TTG)] + [\theta][d(AAA)\cdot d(TTT)] + [\theta][d(AAT)\cdot d(ATT)] + [\theta][d(ATT)\cdot d(AAT)] + [\theta][d(TTA)\cdot d(TAA)] + [\theta][d(TAC)\cdot d(GTA)] \quad (15)$$

However, by combining eqns.(2),(4),(7),(8),(9),(14),(15) we obtain eqn.(16) that shows that the *average* molar ellipticity of satellites *II* and *III* should be expressible as a weighted sum of known spectra and that therefore a third test of triplet nucleotide sequence dependence is possible.

$$(1/2)[\theta][\textit{D. virilis II}] + (1/2)[\theta][\textit{D. virilis III}] = (3/7)[\theta][\text{poly } d(AAT)\cdot\text{poly } d(ATT)] + (3/14)[\theta][\text{poly } d(CAA)\cdot\text{poly } d(TTG)] + (3/14)[\theta][\text{poly } d(TAC)\cdot\text{poly } d(GTA)] + (1/7)\{2[\theta][d(C_{15}A_{15})\cdot d(T_{15}G_{15})] - [\theta][\text{poly dG}\cdot\text{poly dC}]\} \quad (16)$$

Moreover by combining eqns.(13),(16) we obtain eqn.(17) that allows a fourth test of the theory and eliminates the need to use the poly dA·poly dT spectrum derived from eqns.(2),(4).

$$[\theta][\textit{D. virilis I}]$$
$$= (1/2)[\theta][\textit{D. virilis II}] + (1/2)[\theta][\textit{D. virilis III}]$$
$$- (3/7)[\theta][\text{poly d(AAT)}\cdot\text{poly d(ATT)}]$$
$$+ (3/14)[\theta][\text{poly d(CAA)}\cdot\text{poly d(TTG)}]$$
$$+ (3/14)[\theta][\text{poly d(TAC)}\cdot\text{poly d(GTA)}] \qquad (17)$$

RESULTS AND DISCUSSION

Using eqns.(11),(13),(16),(17) we can make four tests of the proposition that the CD spectra of *Drosophila* satellite DNAs depend on the sequences of consecutive triplets of nucleotide pairs. The results of these tests are shown in Figs. 2,4,6,7

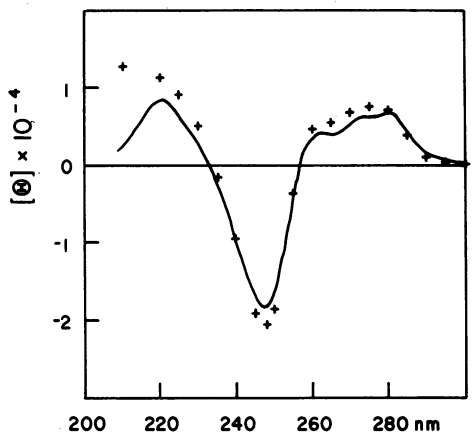

Fig. 2

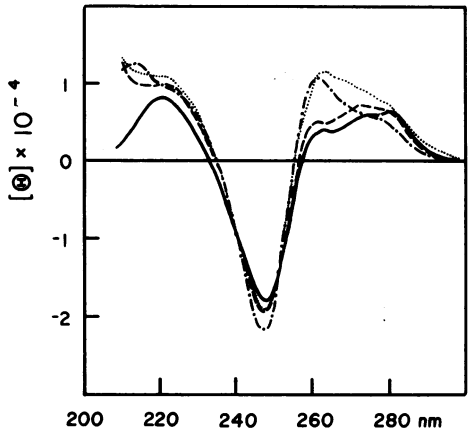

Fig. 3

FIG. 2. CD spectrum of *D. melanogaster* satellite DNA [3] (solid line) compared with calculated values (+) from eqn.(11).
FIG. 3. The CD spectrum of *D. melanogaster* satellite DNA (solid line) compared with first neighbour calculations made by Gray and Gall[3] using eqns.(18) (-·-·-), (18) and (19) (·····), and (18) and (20) (----).

For the *D. melanogaster* satellite DNA (Fig. 2) the fit is very good in the region 220-300nm although the calculated spectrum does not show the dip observed below 220nm.

It is interesting to compare the result obtained using the triplet nucleotide hypothesis with those obtained by Gray and Gall using the doublet hypothesis (Fig. 3). In this case eqn.(18) should apply.

[θ][*D. melanogaster*]
= (4/5)[θ][poly d(AT)·poly d(AT)]
+ (1/5)[θ][poly dA·poly dT] (18)

As can be seen (Fig. 3) a very poor fit is obtained in the region of 255-280nm as one would expect if the triplet hypothesis were true. From eqns.(11) and (18) the difference would correspond to (3/5)[θ][poly d(AAT)·poly d(ATT)] - (2/5)[θ][poly d(AT)·poly d(AT)] - (1/5)[θ][poly dA·poly dT].

In a second trial (Fig. 3) Gray and Gall replaced the [θ][poly dA·poly dT] contribution using eqn.(19)

[θ][poly dA·poly dT]
= 3[θ][poly d(AAC)·poly d(GTT)]
- 2[θ][poly d(AC)·poly d(GT)] (19)

The triplet hypothesis would predict that this should not improve the fit, as indeed it does not. In a third calculation Gray and Gall replaced the [θ][poly dA·poly dT] contribution in eqn.(18) using eqn.(20)

[θ][poly dA·poly dT]
= 3[θ][poly d(AAT)·poly d(ATT)]
- 2[θ][poly d(AT)·poly d(AT)] (20)

It is interesting that this is equivalent to using eqn.(11) which follows directly from the assumption that *D. melanogaster* satellite DNA, poly d(AAT)·poly d(ATT), and poly d(AT)·poly d(AT) all have the same conformation in solution and that the triplet hypothesis is valid.

Good fitting is achieved using eqn.(13) to calculate the spectrum of *D. virilis* satellite *I* (Fig. 4) according to the triplet dependence hypothesis. This fit may be contrasted with those obtained[3] using the doublet hypothesis. According to the latter the spectrum of *D. virilis* satellite *I* should be given by eqn.(21)

[θ][*D. virilis I*]
= (3/7)[θ][poly d(TAC)·poly d(GTA)]
+ (2/7)[θ][poly d(AC)·poly d(GT)]
+ (2/7)[θ][poly dA·poly dT] (21)

Finding that eqn.(21) did not provide a good fit (Fig. 5) for the observed spectrum Gray and Gall[3] replaced the poly dA·poly dT contribution firstly using eqn.(22) and secondly using eqn.(23).

$$[\theta][\text{poly dA}\cdot\text{poly dT}] = 3[\theta][\text{poly d(AAT)}\cdot\text{poly d(ATT)}] - 2[\theta][\text{poly d(AT)}\cdot\text{poly d(AT)}] \quad (22)$$

$$[\theta][\text{poly dA}\cdot\text{poly dT}] = 3[\theta][\text{poly d(AAC)}\cdot\text{poly d(GTT)} - 2[\theta][\text{poly d(AC)}\cdot\text{poly d(GT)}] \quad (23)$$

The hypothesis of triplet dependence would predict that no better fit would be obtained by these manipulations and that is indeed the case (see Fig. 5).

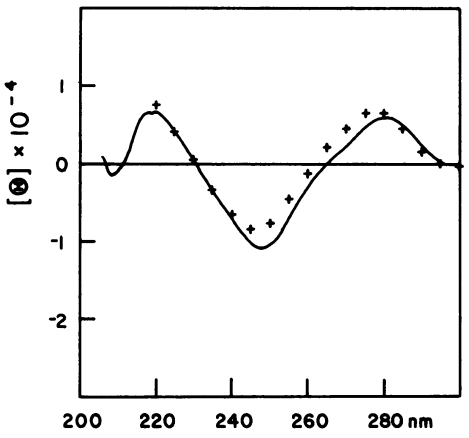

Fig. 4

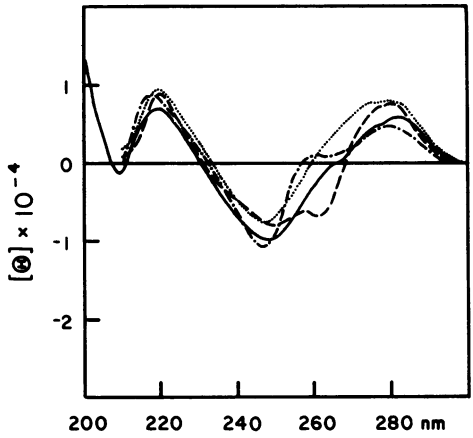

Fig. 5

FIG. 4. CD spectrum of *D. virilis* satellite DNA *I* [3] (solid line) compared with calculated values (+) from eqn.(13).
FIG. 5. CD spectrum of *D. virilis* satellite DNA *I* (solid line) compared with first neighbour calculations made by Gray and Gall using eqn.(21) (-·---), eqn.(21) and (22) (----), and (21) and (23) (·····).

The average of the spectra of *D. virilis* satellites *II* and *III* and prediction of the triplet hypothesis (from eqn.(16)) are compared in Fig. 6.

The excellent fitting (Fig. 7) obtained by using eqn.(17) to predict the spectrum of *D. virilis* satellite *I* is particularly satisfying since it implies that the three *D. virilis* satellite spectra are consistent with the triplet sequence hypothesis and in this case there is no dependence on the derived poly dA·poly dT spectrum.

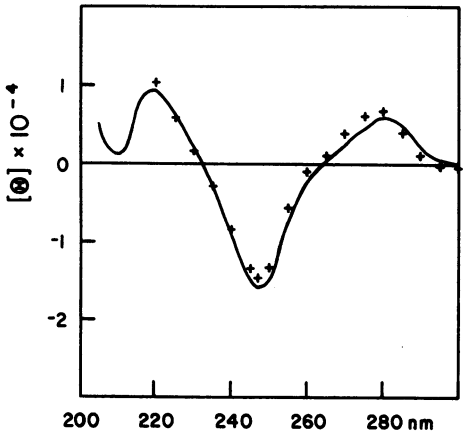

Fig. 6

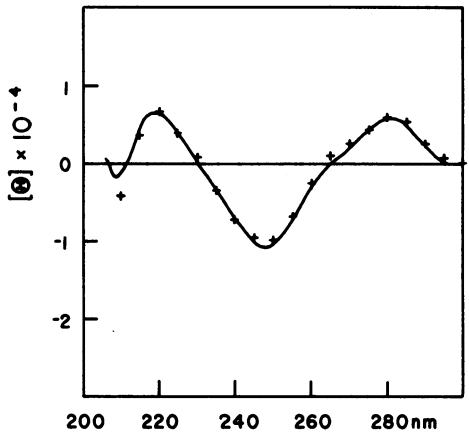

Fig. 7

FIG. 6. Average CD spectrum of *D. virilis* satellite DNAs *II* and *III* [3] (solid line) compared with calculated values (+) from eqn. (16).
FIG. 7. CD spectrum of *D. virilis* satellite DNA *I* [3] compared with calculated values from eqn.(17).

It is reasonable to conclude that the molar ellipticities are indeed a function of the kind and frequencies of paired nucleotide triplets. The small discrepancies that persist are not unexpected: there may be errors of this magnitude in the estimated spectra used to derive some of the components. There are also approximations in deriving the various equations: I have, for example, had to ignore the small [θ][d(CCA)·d(TGG)] and [θ][d(CAA)·d(TTG)] contributions in eqn.(4), and also have had to assume that the satellite DNA sequences are as ideal as stated (which is almost certainly not the case).

It also follows from the self-consistency demonstrated by the tests that all the polymers involved must indeed have essentially identical conformations as we assumed.

It might be argued that all we have done is to use a set of simple polymers with more appropriate geometries. In other words in poly d(ABCXYZ)·poly d(Z'Y'X'C'B'A') not only would the fragments like d(ABC)·d(C'B'A') and d(XYZ)·d(Z'Y'X') have the same

conformation as in poly d(ABC)·poly d(C'B'A') and poly d(XYZ)·poly d(Z'Y'X') respectively, but the overlapping fragments like d(BCX)·d(X'C'B') would have the same geometry as in poly d(BCX)·poly d(X'C'B'). This is highly improbable unless the conformations of all the fragments are the same. This is indeed what is observed by X-ray diffraction of fibers of (for example) poly d(AAT)·poly d(ATT) [7], poly d(AT)·poly d(AT) [5], and *D. virilis* satellite DNA *I* [8]. These DNAs have the same regular helical *B*-DNA conformation at the high relative humidities that correspond most closely to conditions in aqueous solution. Even when, under less polar conditions, these DNAs change to *A* or *D*-DNA then *regular* helical chains (in which every nucleotide has the same conformation) are retained.

There therefore seems little doubt that all the DNAs discussed (except poly dA·poly dT), with compositions ranging from 0-100% AT and with sequences as different as oligo-G and oligo-AT, will not usually have different molecular geometries in solution.

ACKNOWLEDGMENTS

This work was supported by a grant (GM 17371) from the National Institutes of Health. Miss Stephanie Tsivoglou assembled the data and prepared the Figures.

REFERENCES

1 Wells, R.D., Larson, J.E., Grant, R.C., Shortle, B.E. and Cantor, C.R. (1970) *J. Mol. Biol.* 54, 465-497
2 Gratzer, W.B., Hill, L.R. and Owen, R.J. (1970) *Eur. J. Biochem.* 15, 209-214
3 Gray, D.M. and Gall, J.G. (1974) *J. Mol. Biol.* 85, 665-679
4 Gray, D.M. and Tinoco, Jr., I. (1970) *Biopolymers* 9, 223-244
5 Arnott, S., Chandrasekaran, R., Hukins, D.W.L., Smith, P.J.C. and Watts, L. (1974) *J. Mol. Biol.* 88, 523-533
6 Arnott, S. and Selsing, E. (1974) *J. Mol. Biol.* 88, 509-521
7 Selsing, E., Arnott, S. and Ratliff, R.L. (1974) *J. Mol. Biol.* (in press)
8 Arnott, S., Selsing, E., Gall, J.G., Skinner, D.M. and Southern, E.M. (1974). (in preparation)
9 Gray, D.M. and Bollum, F.J. (1974) *Biopolymers* 13, 2087-2102
10 Burd, J.F. and Wells, R.D. (1974) *J. Biol. Chem.* 249, 7094-7101
11 Gray, D.M., Ratliff, R.L. and Williams, D.L. (1973) *Biopolymers* 12, 1233-1245
12 Gall, J.G. and Atherton, D.A. (1974) *J. Mol. Biol.* 85, 633-664