SHAPIRO, H. S. 1968. Distribution of purines and pyrimidines in DNAs. *In* Handbook of Biochemistry. H. A. Sober, ed. CRC-Press, Cleveland, Ohio. page H-39.

SOBELL, H. 1973. The stereochemistry of actinomycin binding to DNA and its implications in molecular biology. *Prog. Nucleic Acid Res. Mol. Biol.* **13**:153.

WYATT, G. R., and S. S. COHEN. 1952. A new pyrimidine base from bacteriophage nucleic acids. *Nature (Lond.).* **170**:1072.

CHRISTIAN MARCK
Service de Biophysique
WILHELM GUSCHLBAUER
Service de Biochimie
Départment de Biologie, B.P.N°2,
Centre d'Etudes Nucléaires de Saclay
F-91190 Gif-sur-Yvette, France

LETTER TO THE EDITOR

A Response to "Problems Involved in the Computation of the 10 Elementary First-Neighbor Interaction Circular Dichroism Signals of DNA"

Dear Sir:

The above letter calls into question the formalism of the first-neighbor approximation. Specifically, the authors question the use of the reentrant condition as a constraint upon the circular dichroism (CD) contributions of the first neighbor units which comprise a double-stranded DNA. The application of these conditions results in eight independent CD contributions. The authors suggest that there are 10 such components.

Within the first neighbor approximation we consider that the CD of a base pair is due entirely to its own intrinsic nature and its interactions with the base pairs above and below it in the duplex structure. This means that only first-neighbor units are involved in the determination of the CD spectrum. There are only eight independent first-neighbor frequencies. Because no nonfirst-neighbor interactions are allowed to influence the CD, there can be no more than eight independent contributions to the CD. Central to the first-neighbor approximation is the view that the CD contributions of the first-neighbor units are vector representations of the first neighbors themselves. This results in the disputed relationships given as Eqs. 6 and 7 above.

When we consider the Watson-Crick constraints for double-stranded DNA, it is evident that $f_{AA} = f_{TT}$, $f_{AC} = f_{GT}$, etc. Hence, a given nonself-complementary first-neighbor unit and its complement will each occur exactly the same number of times in any double-stranded polynucleotide. Thus, these two first-neighbor units become linearly dependent, and we are not able to separate the actual CD contribution of an ApA unit from that of a TpT. Consequently, we choose one of the two first-neighbor units as independent, say ApA, and define $T_{AA} = T_{TT}$. At this point we have redefined both T_{AA} and T_{TT} as the average of the actual contributions of ApA and TpT. This causes no difficulty whatsoever, because in any double-stranded DNA with which we will work ApA and TpT will occur in the same numbers, and we can express their actual CD contributions in terms of the average of both with no error.

A similar situation obtains in the case of the reentrant conditions. Inasmuch as the equation

$$f_{AC} + f_{AG} + f_{AT} = f_{TA} + f_{GA} + f_{CA}$$

is true for every double-stranded DNA, the first-neighbor combinations on the left and right sides of this equation (or any algebraic rearrangement thereof) are linearly dependent, and we will not be able to resolve the CD contributions of the two sides. Consequently, the measured first-neighbor CD contributions for the left and right sides must be identical. Hence,

$$T_{AC} + T_{AG} + T_{AT} = T_{TA} + T_{GA} + T_{CA};$$

this is Eq. 6 in the above letter and an identical agrument for the other reentrant condition produces Eq. 7 above.

This is not to say that the CD contributions of the six first-neighbor units appearing in Eqs. 6 or 7 above are inextricably mixed together. Because of the different sequences of the DNA molecules the average contribution of the six first neighbors is different in each case. For example, in poly dAT:dAT, T_{AT} and T_{TA} are mixed together but there are no others; in poly dATC:dGAT, T_{AT} is mixed with T_{GA} and T_{CA} but not T_{TA}. As a consequence, resolution of the CD contributions of the eight independent first-neighbor units is possible.

The effect of the application of the reentrant conditions as CD constraints is difficult to visualize. Some mixing of the absolute first-neighbor unit contributions may occur. The binding of a drug can be a useful probe as to the extent of this mixing. The perturbation caused in a relatively small number of first-neighbor units by the binding may be mixed uniformly, or to a very small extent into the other first-neighbor units involved in the reentrant conditions. Examination of the results for the binding of actinomycin-D to DNA (Allen et al., 1977, referenced above) indicate that this mixing is small. The binding perturbations on ApT and TpA are very different from ApC and GpA. Similarly, the binding of actinomycin D perturbs CpG and GpC to a very different extent from ApG and TpG. In fact, ApG and GpA behave quite differently as the binding progresses. As a result of the considerable differences in the behavior of first-neighbor units involved in the reentrant conditions, I conclude that mixing caused by these conditions is not an effect of primary importance.

If one makes the assumption that the 10 CD contributions of Marck and Guschlbauer obtained by application of the Watson-Crick constraints are all independent, and applies the reentrant conditions in the form

$$f_{TA} T^{\lambda}_{TA} = f_{AC} T^{\lambda}_{TA} + f_{AG} T^{\lambda}_{TA} + f_{AT} T^{\lambda}_{TA} - f_{GA} T^{\lambda}_{TA} - f_{CA} T^{\lambda}_{TA}$$

and

$$f_{GC} T^{\lambda}_{GC} = f_{CG} T^{\lambda}_{GC} + f_{TG} T^{\lambda}_{GC} + f_{AG} T^{\lambda}_{GC} - f_{GA} T^{\lambda}_{GC} - f_{GT} T^{\lambda}_{CG},$$

one obtains

$$\begin{aligned} S^{\lambda} = {} & 2T_{AA} f_{AA} + (2T^{\lambda}_{AC} + T^{\lambda}_{TA} - T^{\lambda}_{GC}) f_{AC} \\ & + (2T^{\lambda}_{AG} + T^{\lambda}_{TA} + T^{\lambda}_{GC}) f_{AG} + (T^{\lambda}_{AT} + T^{\lambda}_{TA}) f_{AT} \\ & + (2T^{\lambda}_{CA} - T^{\lambda}_{TA} + T^{\lambda}_{GC}) f_{CA} + 2T^{\lambda}_{CC} f_{CC} + (T^{\lambda}_{CG} + T^{\lambda}_{GC}) f_{CG} \\ & + (2T^{\lambda}_{GA} - T^{\lambda}_{TA} - T^{\lambda}_{GC}) f_{GA}. \end{aligned} \tag{1}$$

For a group of DNAs one again obtains a matrix equation. Carrying out the matrix algebra for known S and F matrices gives a numerical expression for T as a matrix of n_{λ} rows and eight columns, where n_{λ} is the number of measured wavelengths.

The eight columns give the linear combination of the 10 independent first-neighbor contributions shown in Eq. 1. There is a requirement for two more relationships in order to determine all 10 contributions. For the first-neighbor hypothesis, Eqs. 6 and 7 in the above letter provide those two relationships. In any first-neighbor-based formalism, two further constraints of some type must be determined.

In fact, it can be shown that the 10 CD contributions we have supposed here to be independent are not. This question can be resolved by examination of the rank of the matrix of the input CD spectra, S. In the original work of Allen et al. (1972), referenced above, the rank of this matrix was determined to be eight. Since $T = SF^t(FF^t)^{-1}$ and no matrix on the right-hand side has a rank >8, the rank of T can be no >8. Hence there are only eight independent components.

I do not feel that the above letter of Marck and Guschlbauer necessitates any changes in the first-neighbor formalism we have used nor any reinterpretation of the results of this formalism.

Received for publication 8 May 1978.

FRITZ ALLEN
The University of New Mexico
Albuquerque, New Mexico 87131 U.S.A.