LETTER TO THE EDITOR

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

Dear Sir:

In a recent paper, Allen et al. (1977) measured perturbations of the circular dichroism (CD) spectra of DNA due to actinomycin binding. From the selective perturbation of the computed elementary first-neighbor CD contributions to the DNA CD spectra, these authors have concluded that actinomycin (AM) is equally well bound to GpC/GpC and CpG/CpG sequences in DNA and, only to a slightly lesser extent, also to GpA/TpC sequences. From their results the authors question the validity of the AM-DNA binding model of Sobell (1973).

Without wanting to enter into an argument on the mode of AM binding to DNA, there are two main points with which we do not agree: (*a*) We contest the validity of the formalism used by Allen et al. in this and previous publications for the computation of the 10 elementary first-neighbor CD signals. This formalism introduces two additional constraint relations on the elementary CD contributions that are mathematically and physically unfounded and wrong. (*b*) We question that selective perturbation of the elementary contributions of the CD spectra of DNA can be obtained and interpreted in the way Allen et al. (1977) do, because this would upset the very relations that are the basis of the authors' formalism.

(*a*) Allen et al. (1977) use essentially the mathematical formalism of Allen et al. (1972) which states that the 10 elementary CD contributions of the 10 base-paired first-neighbor configurations can be computed from any 8 (or more) CD spectra of DNAs, the nearest-neighbor frequencies of which are known. Using these 10 elementary CD contributions, the nearest-neighbor frequencies of any DNA can be computed from its CD spectrum. We have demonstrated that it is impossible to compute these elementary contributions, but that it is still possible to obtain the nearest-neighbor frequencies without knowing all the 10 elementary CD contributions (Marck and Guschlbauer, 1978*a*).

The first-neighbor approximation as defined by Gray and Tinoco (1970) assumes that any set of CD spectra of double-stranded DNAs can be written as

$$S = T \cdot F \tag{1}$$

"where S is an $n \times m$ matrix whose columns are the measured CD spectra of m DNAs and F is a $10 \times m$ matrix of the DNA first-neighbour frequency information" (Allen et al., 1972). The columns of the T matrix are considered as vector representations of the elementary CD signals of base-paired first-neighbor configurations. For this reason, the frequencies appearing in the F matrix have to present base-paired first-neighbor frequencies.

Because of the complementarity of DNA, the 16 single-stranded first-neighbor configurations lead to only 10 possible base-paired first-neighbor configurations. From the reentrant con-

---

Dr. Marck is a Collaborateur Temporaire de Thèse du Commisariat à l'Energie Atomique. His present address is: Max-Planck-Institute für Experimentelle Medizin, Abteilung Chemie, D-3400 Göttingen, W. Germany.

dition on the nearest-neighbor frequencies (Josse et al., 1961), it can be shown that there are only eight independent single-stranded frequencies in the above system. These reentrant conditions are:

$$f_{AT} + f_{AC} + f_{AG} = f_{TA} + f_{CA} + f_{GA}. \tag{2}$$

$$f_{CG} + f_{AG} + f_{TG} = f_{GC} + f_{GA} + f_{GT}. \tag{3}$$

For the base-paired first-neighbor frequencies (which form the F matrix in Eq. 1), these equations become (Gray and Tinoco, 1970):

$$f_{\substack{AT\\TA}} + 1/2.f_{\substack{AC\\TG}} + 1/2.f_{\substack{AG\\TC}} = f_{\substack{TA\\AT}} + 1/2.f_{\substack{CA\\GT}} + 1/2.f_{\substack{GA\\CT}}. \tag{4}$$

$$f_{\substack{CG\\GC}} + 1/2.f_{\substack{AG\\TC}} + 1/2.f_{\substack{TG\\AC}} = f_{\substack{GC\\CG}} + 1/2.f_{\substack{GA\\CT}} + 1/2.f_{\substack{GT\\CA}}. \tag{5}$$

The correctness of these conditions can be verified on any simple sequence, like poly[d(TAC)]·poly[d(GTA)].

Eqs. 4 and 5 show that the F matrix of Eq. 1 is always singular (i.e. of order 8); therefore the T matrix cannot be computed and solution of Eq. 1 is impossible. This does not imply, however, that the T matrix is singular. Allen et al. (1972, 1977), however, have supposed this and have introduced two additional constraint relations on the T matrix that would allow them to compute the elementary CD contributions of each of the 10 configurations:

$$T_{\substack{AT\\TA}} + T_{\substack{AC\\TG}} + T_{\substack{AG\\TC}} = T_{\substack{TA\\AT}} + T_{\substack{CA\\GT}} + T_{\substack{GA\\CT}}. \tag{6}$$

$$T_{\substack{CG\\GC}} + T_{\substack{AG\\TC}} + T_{\substack{TG\\AC}} = T_{\substack{GC\\CG}} + T_{\substack{GA\\CT}} + T_{\substack{GT\\CA}}. \tag{7}$$

Eqs. 6 and 7 are constraints on the vectors T and have been presented (Allen et al., 1972, 1977) as a "consequence" of relations 2 and 3 which are (experimentally verified) constraints on the scalars f, due to the topological features of the DNA strands of opposite polarity. The application of a constraint relation valid for the scalars f to the vectors T is mathematically unjustified and physically meaningless.

It is known (Allen and Daub, 1974) that $T_{\substack{AT\\TA}} \neq T_{\substack{TA\\AT}}$, $T_{\substack{AG\\TC}} \neq T_{\substack{GA\\CT}}$ etc. Therefore, in order to accept Eqs. 6 and 7, one has to imagine some physical reason why certain sums should lead to equalities.

Many DNAs have part or all of a given base modified; the electronic transitions of these are different from those of the mother base and thus their CD contributions to the CD spectrum of DNA. In many plant DNAs (Shapiro, 1968) or in T-even phage DNAs (Wyatt and Cohen, 1952), part or all of the cytosines are modified. In such cases Eq. 6 would become

$$T_{\substack{AT\\TA}} + T_{\substack{AC\\TG}} + T_{\substack{AG\\TC}} + T_{\substack{AZ\\TG}} + T_{\substack{AG\\TZ}} = T_{\substack{TA\\AT}} + T_{\substack{CA\\GT}} + T_{\substack{GA\\CT}} + T_{\substack{ZA\\GT}} + T_{\substack{GA\\ZT}} \tag{6a}$$

or

$$T_{\substack{AT\\TA}} + T_{\substack{AZ\\TG}} + T_{\substack{AG\\TC}} = T_{\substack{TA\\AT}} + T_{\substack{ZA\\GT}} + T_{\substack{GA\\ZT}} \tag{6b}$$

These equations are only correct, if $T_{\substack{AZ\\TG}} + T_{\substack{AG\\TZ}} = T_{\substack{ZA\\GT}} + T_{\substack{GA\\ZT}}$ or nil. Thus $T_{\substack{AT\\TA}} = T_{\substack{TA\\AT}}$, which is evidently wrong.

It can also be shown that any sequence, like the one shown above, that lacks certain nearest-neighbor configurations would not fulfill Eqs. 6 and 7. Allen et al. (1977) claim that the

columns of the T matrix obtained as a solution of their Eq. 4 were the CD contributions of the individual base-paired first-neighbor configurations. This is only true if the 10 actual CD contributions would obey Eq. 6 and 7. On the contrary, if—as we think—the 10 actual CD signals are linearly independent, the solution of their Eq. 4 does not yield these CD signals, but gives only linear combinations of them, except for $T_{\substack{AA \\ TT}}$ and $T_{\substack{GG \\ CC}}$, which are correctly obtained. Thus the column vector of the T matrix corresponding presumably to the signal of the $\frac{AC}{TG}$ configuration will be a sum of

$$-1/8 \cdot T_{\substack{AT \\ TA}} + 1/8 \cdot T_{\substack{TA \\ AT}} + 3/4 \cdot T_{\substack{AC \\ TG}} + 1/4 \cdot T_{\substack{CA \\ GT}} - 1/8 \cdot T_{\substack{GC \\ CG}} + 1/8 \cdot T_{\substack{CG \\ GC}}$$

(Marck, 1978).

(b) The analysis of binding of AM to DNA of Allen et al. (1977) passes by the computation of a perturbed T matrix due to the binding of AM. Therefore the reentrant conditions 6 and 7 must also hold for the perturbed T matrix, as they are used to compute it. From this perturbed T matrix the nonperturbed T matrix is subtracted; the consequence of this is that the computed perturbation signals are also linked by relations 6 and 7.

Even if the constraint relations 6 and 7 could be proven correct for the nonperturbed T matrix, inspection of these relations shows that perturbation of a single column vector of the T matrix is impossible, because it would have to be counter-balanced by a perturbation of at least one other column vector of the T matrix. For instance, perturbation of GpA/TpC would have to be either equaled by an identical perturbation of ApG/CpT, or by a more complex distribution of perturbations between several configurations. It is rather illogical to presume in such a way on the binding specificity of AM or any other drug. The selective perturbation of any one of the configurations cannot be obtained. Therefore, the results on the preferential binding of AM to specific DNA sequences and the questioning of the Sobell model (Allen et al., 1977) are not warranted, and the data have to be reconsidered.

Although it is impossible to obtain the 10 separate individual CD contributions of the 10 configurations (Marck and Guschlbauer, 1978*a*), it is possible to obtain the 10 nearest-neighbor frequencies correctly, by using Eqs. 4 and 5 instead of Eqs. 6 and 7. This work is performed in our laboratory and published elsewhere (Marck and Guschlbauer, 1978*b*).

Received for publication 8 April 1978.

REFERENCES

ALLEN, F. S., and G. W. DAUB. 1974. Analysis of circular dichroism spectra of homopurine:homopyrimidine DNAs for sequence information and comparison of results with heterobase DNA. *Biopolymers.* **13**:241.

ALLEN, F. S., D. M. GRAY, G. P. ROBERTS, and I. TINOCO, JR. 1972. The ultraviolet circular dichroism of some natural DNAs and an analysis of the spectra for sequence information. *Biopolymers.* **11**:853.

ALLEN, F. S., M. B. JONES, and U. HOLLSTEIN. 1977. First-neighbor specificities of actinomycin-DNA bindings by circular dichroism. *Biophys. J.* **20**:69.

GRAY, D. M., and I. TINOCO, JR. 1970. A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers.* **9**:223.

JOSSE, J., A. D. KAISER, and A. KORNBERG. 1961. Enzymatic synthesis of DNA. VIII. Frequencies of nearest neighbour sequences in DNA. *J. Biol. Chem.* **236**:864.

MARCK, C. 1978. Le problème du calcul des signaux de dichroisme circulaire des dix configurations de plus proches voisins de l'ADN. *C. R. Acad. Sci. (Paris).* **286D**:1531.

MARCK, C., and W. GUSCHLBAUER. 1978*a*. Le problème du calcul fes frèquences de plus proches voisins des ADN à partir de leurs spectres de dichroime circulaire. *C. R. Acad. Sci. (Paris).* **286D**:713.

MARCK, C. and W. GUSCHLBAUER. 1978*b*. A simple procedure for the computation of first nearest neighbour frequencies of DNAs from CD spectra. *Nucleic Acids Res.* **5**:2013.

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.