APPENDIX

Calculation of the Circular Dichroism of Bihelical Ribonucleic Acids as a Function of Nucleotide Composition

By ROBERT A. COX and WILLIAM HIRST

National Institute for Medical Research, Mill Hill, London NW7 1AA, U.K.

(Received 10 October 1975)

The c.d. of bihelical RNA may, in principle, be calculated on the basis of the assumption that the c.d. of a base-pair is affected by its nearest neighbour only and not by more distant residues (see, e.g., Blum *et al.*, 1972). Neglecting minor bases, there are ten unique pairs of nearest neighbours that may be found in a bihelix stabilized by Watson-Crick base-pairs (see Table 1). The c.d. at any wavelength λ of a bihelical RNA of mole fraction f_G of guanine nucleotides is given by eqn. (1):

$$\Delta\epsilon_{f_G}(\lambda) = \sum_{i=1}^{10} (N_i/N) C_i(\lambda) \quad (1)$$

where N is the number of base-pairs, N_i is the number of times each of the ten nearest neighbours occurs ($\sum_i N_i = N-1$), and $C_i(\lambda)$ is the contribution of couple i per mole of nucleotide residues to the c.d.

We make two principal assumptions (see also the Theoretical section of the main paper): (1) the c.d. spectrum of bihelical RNA is a linear combination of the spectral contributions of the different component couples as stated in eqn. (1); (2) we assume that nearest neighbours occur with the same frequency as if the nucleotide sequence were random. Thus the probability P_i of the i th couple occurring may be calculated from the overall nucleotide composition. Thus:

$$P_i = N_i/N \quad (2)$$

The restrictions of Watson-Crick pairing allow P_i to be expressed solely in terms of f_G , the mole fraction of guanine nucleotides, since $f_G = f_C$ and $f_A = f_U = (0.5 - f_G)$ (see Table 1). Thus P_i may replace N_i/N in eqn. (1).

Values of C_i (see eqn. 1) or of combinations of C_i may be obtained from the c.d. of model polynucleotides. The models and their relation to C_i are also given in Table 1.

The data of Table 1 allow eqn. (1) to be formulated in terms of f_G and the c.d. of model polynucleotides (see eqn. 3):

$$\Delta\epsilon_{f_G}(\lambda) = af_G^2 + bf_G + c \quad (3)$$

where

$$a = 2(\Delta\epsilon_A + \Delta\epsilon_B + \Delta\epsilon_C + \Delta\epsilon_D - \Delta\epsilon_E - \Delta\epsilon_F);$$

$$b = \Delta\epsilon_E + \Delta\epsilon_F - 2(\Delta\epsilon_C + \Delta\epsilon_D);$$

$$c = 0.5(\Delta\epsilon_C + \Delta\epsilon_D)$$

Except for poly(A-G)·poly(U-C), the c.d. of the model RNA species are known. However, the c.d. for this polynucleotide ($\Delta\epsilon_E$) may be calculated by means of eqn. (3) from the c.d. of any bihelical RNA species

Fig. 1. C.d. spectrum calculated for poly(A-G)·poly(U-C) by means of eqn. (3) from the c.d. spectrum of bihelical RNA species isolated from virus-like particles of *Penicillium chrysogenum*

Eqn. (3) was rearranged to make $\Delta\epsilon_F(\lambda)$ the subject, i.e.:

$$\Delta\epsilon_E(\lambda) = (\Delta\epsilon_{f_G} - a'f_G^2 + b'f_G - C)/(f_G - 2f_G^2)$$

where $a' = 2(\Delta\epsilon_A + \Delta\epsilon_B + \Delta\epsilon_C + \Delta\epsilon_D + \Delta\epsilon_F)$ and $b' = \Delta\epsilon_F - 2(\Delta\epsilon_C + \Delta\epsilon_D)$. The c.d. spectra of RNA of $f_G = 0.19$ and $f_G = 0.255$ (Cox *et al.*, 1970) were both used to calculate $\Delta\epsilon(\lambda)$ and the mean values (○) are presented. Both estimates of $\Delta\epsilon_E$ agreed closely, the 'fit' being 0.1. For comparison, the c.d. spectrum of poly(A-C)·poly(U-G) (Tinoco, 1973) is also given (▽).

Table 1. The ten unique-first-neighbour couples arising from Watson-Crick base-pairs, their abundance calculated on the basis of a random distribution of nucleotides and their contribution to c.d. expressed in terms of bihelical model polynucleotides

Unique first-neighbour couple	1	2	3	4	5	6	7	8	9	10
	$\rightarrow$ G-G C-C $\leftarrow$	$\rightarrow$ G-C C-G $\leftarrow$	$\rightarrow$ C-G G-C $\leftarrow$	$\rightarrow$ A-A U-U $\leftarrow$	$\rightarrow$ A-U U-A $\leftarrow$	$\rightarrow$ U-A A-U $\leftarrow$	$\rightarrow$ G-A C-U $\leftarrow$	$\rightarrow$ A-G U-C $\leftarrow$	$\rightarrow$ C-A G-U $\leftarrow$	$\rightarrow$ A-C U-G $\leftarrow$
$P_1 = N_i/N$ (see eqn. 2)	$2f_G^2$	$f_G f_C$	$f_C f_G$	$2f_A^2$	$f_A f_U$	$f_U f_A$	$2f_G f_A$	$2f_A f_G$	$2f_C f_A$	$2f_A f_C$
P_1 expressed in terms of $f_G(f_G = f_C; f_A = f_U = 0.5 - f_G)$	$2f_G^2$	f_G^2	f_G^2	$2(0.5 - f_G)^2$	$(0.5 - f_G)^2$	$(0.5 - f_G)^2$	$2f_G(0.5 - f_G)$	$2f_G(0.5 - f_G)$	$2f_G(0.5 - f_G)$	$2f_G(0.5 - f_G)$
Contribution per nucleotide of couple to c.d. spectrum	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Complex giving rise to C_i	Poly(G)-poly(C)	Poly(G-C)-poly(G-C)	Poly(G-C)-poly(G-C)	Poly(A)-poly(U)	Poly(A-U)-poly(A-U)	Poly(A-U)-poly(A-U)	Poly(A-G)-poly(U-C)	Poly(A-G)-poly(U-C)	Poly(A-C)-poly(U-G)	Poly(A-C)-poly(U-G)
Abbreviation for complex	A	B	B	C	D	D	E	E	F	F
Value of C_i expressed in terms of c.d. per nucleotide of complex ($\Delta\epsilon_A$ etc.)	$C_1 = \Delta\epsilon_A$	$C_2 + C_3 = 2\Delta\epsilon_B$	$C_4 = \Delta\epsilon_C$	$C_5 + C_6 = 2\Delta\epsilon_D$	$C_7 + C_8 = 2\Delta\epsilon_E$	$C_9 + C_{10} = 2\Delta\epsilon_F$

and the c.d. of the other reference compounds. Two estimates of $\Delta\epsilon_B$ were obtained from the c.d. of bihelical RNA of 38% G+C and of 51% G+C (Cox *et al.*, 1970). Both estimates were in close agreement (see Fig. 1). Values of $\Delta\epsilon_B$ obtained in this way were then used to calculate the c.d. spectrum of bihelical RNA of chosen base composition. The spectrum calculated for bihelical RNA of 78% G+C was substantially different from the c.d. of RNA of 42% G+C, e.g. $\Delta\epsilon$ values at 260nm were respectively 5.7 and 7.8 and the values of c.d. at 280nm were respectively 1.3 and 2.7.

The data used in Table 1 may be used in another way, namely to provide the values of $\Delta\epsilon$ for RNA of 100% A+U and 100% G+C that may then be used to construct a semi-empirical relation between $\Delta\epsilon_{f_G}$ and % G+C. These additional data are needed because the bihelical RNA available ranges from approx. 38% to approx. 51% G+C only.

For the average bihelical RNA the frequency with which nearest neighbours occur is close to that calculated for a random nucleotide sequence (Cantor *et al.*, 1966). We define the 100% G+C analogue as poly(G,C)-poly(G,C) that is characterized by the property that each of the three unique couples (1, 2 and 3) occurs with the frequency P_i indicated in Table 1. Hence the c.d. ($\Delta\epsilon_{G,C}$) of poly(G,C)-poly(G,C) is given by eqn. (4):

$$\Delta\epsilon_{G,C}(\lambda) = 0.5(\Delta\epsilon_A + \Delta\epsilon_B) \quad (4)$$

Similarly $\Delta\epsilon_{A,U}$, the c.d. of poly(A,U)-poly(A,U)

that is homologous to poly(G,C)-poly(G,C), is given by eqn. (5):

$$\Delta\epsilon_{A,U}(\lambda) = 0.5(\Delta\epsilon_C + \Delta\epsilon_D) \quad (5)$$

The values of $\Delta\epsilon_{A,U}(\lambda)$, $\Delta\epsilon_{G,C}(\lambda)$ and $\Delta\epsilon_{f_G}$ of known bihelical RNA may be plotted against % G+C. The regression line provides a semi-empirical relation between $\Delta\epsilon_{f_G}(\lambda)$ and % G+C and does not require a knowledge of $\Delta\epsilon_E(\lambda)$ or $\Delta\epsilon_F(\lambda)$.

The non-linear relation between $\Delta\epsilon_{f_G}(\lambda)$ and % G+C [see Fig. 2, and also Fig. 2 of the main paper (Cox *et al.*, 1976)] shows that the c.d. of bihelical RNA is not obtained simply by mixing $\Delta\epsilon_{A,U}(\lambda)$ and $\Delta\epsilon_{G,C}(\lambda)$ in the proportions $f_A:f_G$. A third factor in the analysis arising from the combinations of neighbouring A·U and G·C base-pairs appears to be essential.

References

- Blum, A. D., Uhlenbeck, O. C. & Tinoco, I., Jr. (1972) *Biochemistry* **11**, 3248–3256
- Cantor, C. R., Jaskunas, S. R. & Tinoco, I., Jr. (1966) *J. Mol. Biol.* **20**, 39–62
- Cox, R. A., Kanagalingam, K. & Sutherland, E. A. (1970) *Biochem. J.* **120**, 549–558
- Cox, R. A., Hirst, W., Godwin, E. & Kaiser, I. (1976) *Biochem. J.* **155**, 279–291
- Samejima, T., Hashizume, H., Imahori, K., Fujii, I. & Miura, K. A. (1968) *J. Mol. Biol.* **34**, 39–48
- Tinoco, I., Jr. (1973) in *Fundamental Aspects and Recent Developments in ORD and CD* (Ciardelli, F. & Salvadori, P., eds.), pp. 66–86, Heyden, London
- Wells, B. D. & Yang, J. T. (1974) *Biochemistry* **13**, 1317–1321

Fig. 2. Dependence of c.d. of bihelical RNA on % G+C

The continuous line was calculated by means of eqn. (3); it is also the regression line relating $\Delta\epsilon_{G,C}(\lambda)$ (see eqn. 4), $\Delta\epsilon_{A,U}(\lambda)$ (see eqn. 5) and the c.d. of bihelical RNA species. Measurements were made at the following wavelengths: (a) 300nm; (b) 295nm; (c) 290nm; (d) 285nm; (e) 280nm; (f) 275nm; (g) 270nm; (h) 265nm; (i) 260nm; (j) 210nm. $\diamond$, Poly(A)-poly(U) and poly(G)-poly(C) (Tinoco, 1973); $\blacklozenge$, poly(A-U)-poly(A-U) and poly(G-C)-poly(G-C) (Tinoco, 1973) [the data ($\diamond$ and $\blacklozenge$) are used in eqns. (4) and (5)]; $\circ$, data for the bihelical RNA of 38% and 51% G+C (Cox *et al.*, 1970); $\square$, rice dwarf virus RNA (43% G+C) (Samejima *et al.*, 1968; Wells & Yang, 1974); $\blacksquare$, cytoplasmic polyhedrosis RNA (43% G+C) (Wells & Yang, 1974).