Comparative Studies on the Secondary Structure of Ovalbumin Messenger RNA and Its Complementary DNA Transcript†

Nguyen T. Van, John J. Monahan, Savio L. C. Woo, Anthony R. Means, and Bert W. O'Malley*

ABSTRACT: The secondary structure of ovalbumin messenger RNA and its complementary $cDNA_{ov}$ transcript was compared using three independent physical techniques: secondary structure probing with ethidium bromide, thermal denaturation, and circular dichroism. The technique of fluorescence probing of secondary structure was reinvestigated with the help of 22 model compounds. The information obtained permitted assessment of the relative contributions of base composition, base sequence, and base mismatching toward a quantitative estimate of base pairing in naturally occurring RNA. An estimate of 41 ± 2% base pairing was deduced from a new technique for measuring base pairing in ovalbumin mRNA. Circular dichroism data showed subtle differences in the helical structure between ovalbumin mRNA and globin mRNA. Thermal denaturation and fluorescence probing provided the first physical evidence for the presence of a terminal hairpin loop in $cDNA_{ov}$. Such evidence is in agreement with earlier findings based on biochemical evidence that cDNA molecules also contain a terminal hairpin loop (Efstratiadis, A., et al. (1976), *Cell 7*, 270–288). The $cDNA_{ov}$ transcript proper (i.e., excluding this hairpin loop) cannot form stable duplexes under physiological conditions, notwithstanding the fact that base complementarity in the $cDNA_{ov}$ transcript is identical with that in $mRNA_{ov}$. Quantitative estimation of secondary structures other than the hairpin loop in $cDNA_{ov}$ was unsuccessful. The base pairing in $cDNA_{ov}$ which is unstable under physiological conditions can be partially stabilized by excess ethidium bromide.

We reported previously that ovalbumin mRNA contained many double helical regions with an average length of at least 4–5 base pairs and that the base composition in these paired regions was 50–54% GC (Van et al., 1976). The present study extends the earlier investigation of the secondary structure of ovalbumin mRNA with results obtained from two additional physical techniques: secondary structure probing with ethidium bromide and circular dichroism. The former technique allows a quantitative estimate of the percent of base pairing in ovalbumin mRNA; the latter provides information about the helical structure of both the single-stranded and the double-stranded regions in this RNA.

Based on the secondary structure models and thermal denaturation data, the base pairing in yeast 5.8S rRNA can approach but not exceed the base pairing of mammalian 5.8S rRNA (Van et al., 1977). Fluorescence data, however, conflict with the above conclusion. This prompted us to investigate the base composition, base sequence, and base mismatching effects on fluorescence enhancement of ethidium bromide to understand how these factors might affect the quantitative estimation of secondary structure in naturally occurring RNAs. Twenty-two model compounds described in Table I were used to study each of the individual effects mentioned above as independently from the other two as possible. A new technique to quantitate base pairing in naturally occurring RNAs which takes into account the base composition effect is described.

Recently, we demonstrated the synthesis of a complete complementary DNA copy of ovalbumin mRNA using AMV reverse transcriptase (Monahan et al., 1976a,b). Since this cDNA was an exact complement of $mRNA_{ov}$, it should, in theory, form all the base pairing (i.e., internal hairpin loops or short duplexes) found in the latter. However, the presence of deoxynucleotides in place of ribonucleotides in $cDNA_{ov}$ might be expected to change the nature and extent of base pairing in the complementary $cDNA_{ov}$. Our data revealed that the $cDNA_{ov}$ molecule contains a terminal hairpin loop that is conspicuously absent in $mRNA_{ov}$. Physical measurements based on three independent techniques (thermal denaturation, fluorescence probing, and circular dichroism) indicate that large differences in the secondary structure of the two nucleic acids exist.

Materials and Methods

Nucleic Acids. The preparation and purity of ovalbumin mRNA and its complete cDNA transcript have been reported previously (Woo et al., 1975; Rosen et al., 1975; Monahan et al., 1976a,b). Data for ovalbumin mRNA were derived from two separate $mRNA_{ov}$ preparations. Physical data obtained from these two preparations were identical within experimental errors. The $cDNA_{ov}$ preparations, on the other hand, showed a large variation in the amount of the bases found in the terminal hairpin loop. The size of this hairpin loop appeared to depend on the conditions used to synthesize the $cDNA_{ov}$ molecule (e.g., temperature and concentration of actinomycin D, nucleotides, monovalent, and divalent cations).

In the present study, three types of $cDNA_{ov}$ molecules were utilized: (1) $cDNA_{ov}$ preparations containing a very short terminal hairpin, hereafter referred to as $cDNA_{ov}$ transcript proper (similar preparations of globin cDNA have been shown by Efstratiadis et al., 1976, to contain a hairpin loop less than 20 base pairs in length); (2) $cDNA_{ov}$ preparations containing 12–27% of the bases in the hairpin loop [or $cDNA_{ov}$ (H), the letter H referring to a terminal hairpin of 100–250 base pairs in length]; and (3) the full-length double-stranded $cDNA_{ov}$ transcript, hereafter referred to as double-stranded $cDNA_{ov}$. The percents of bases found in the hairpin loop were estimated from the area of the thermostable peak in the derivative profiles at 260 nm (see Results, section 2).

The model compound, yeast 5.8S rRNA, was a generous gift

† From the Department of Cell Biology and Center for Population Research, Baylor College of Medicine, Houston, Texas 77030. *Received January 24, 1977.* These studies were supported by the Population Center Research Grants HD 07495-04-8A and HD 07495-04-7A, National Institutes of Health Grant HD 08188-04, American Cancer Society Research Grant BC-101, and a grant from the Robert A. Welch Foundation (O-611).

TABLE I: Model Compounds Used.

Compound	Source	Information derived from this compound	Description of compound
Poly[d(A-T)]	P-L Biochemicals	Model for A-T pairs	Alternating copolymer, 100% base pairing
Poly[d(A)]·poly[d(T)]	P-L Biochemicals	Model for A-T pairs; nearest neighbor effect	Equimolecular mixture of poly[d(A)]·poly[d(T)]; 100% base pairing
Poly(A)·poly(U)	P-L Biochemicals	Model for A-U pairs	Equimolecular mixture of poly(A) and poly(U)
Poly(A-U)	Synthesized from poly[d(A-T)] using *E. coli* polymerase	Model for A-U pairs	Alternating copolymer, 100% base pairing
Poly(A,U) no. 1 & poly(A,U) no. 2	Miles Laboratories	Models for A-U pairs; effect of mismatching	Random copolymer with A/U = 1:1; contained substantial mismatchings. No. 1 contains more mismatchings than No. 2
Poly[d(G-C)]	P-L Biochemicals	Model for deoxy G-C pairs	Alternating copolymer, 100% base pairing
Poly[d(G)]·poly[d(C)]	P-L Biochemicals	Model for deoxy G-C pairs; nearest neighbor effect	Equimolecular mixture of poly[d(G)] & poly[d(C)] 100% base pairing
Poly[(G-C)]	Biogenic Research Corp.	Model for G-C pairs in RNA	Alternating copolymer 100% base pairing
f-Met-tRNA	Oak Ridge Laboratories	Model for G-C pairs in naturally occurring RNA	Contained 17 G-C pairs, 1 A-U and 1 G-U pair
5.8S rRNA	Nazar et al. (1975)	Model for a GC-rich naturally occurring RNA	Contained 33 GC, 12 AU pairs, 63% of the bases are paired (57% base pairing if G-U pairs are excluded)
Yeast 5.8S rRNA	Nazar et al. (1975)	Model for AU-rich naturally occurring RNA	Contained 20 AU and 21 GC pairs; 59% of the bases are paired (52% base pairing if G-U pairs are excluded); 54% of the bases are A,U.
M. lysodeikticus DNA	Prepared according to Marmur (1961)	Model for a GC-rich DNA	Contained 72% GC, 100% base pairing
Calf thymus DNA	Prepared according to Marmur (1961)	Model for rich DNA AT	Contained 58% AT, 100% base pairing
Poly(A), (U), (G), (C), d(A), d(T), d(G), and d(C)	Miles Laboratories	Models for single-stranded nucleic acids	Zero base pairing

from Dr. R. Nazar, National Research Council, Toronto, Canada. Chick DNA was prepared from fresh chick liver by phenol extraction, RNase treatment followed by two chloroform isoamyl alcohol extractions, and precipitation in cold ethanol (Marmur, 1961). The 22 nucleic acids shown in Table I were obtained from sources indicated in column 2 of the table. The purity of the synthetic homopolymers and copolymers is the highest available, according to the manufacturers. The purity of the two 5.8S rRNAs was published elsewhere (Nazar et al., 1975). The homogeneity of calf thymus and *Micrococcus lysodeikticus* DNA was assessed by T_m analysis; the T_m values were found to be identical with published values (Marmur and Doty, 1962). Ethidium bromide was a gift from Dr. T. I. Watkins of Boots Pure Drug Co., Nottingham, England, and is of the same degree of homogeneity as the samples used by Lepecq and Paoletti (1967). All chemicals used for buffer preparations are of reagent grade. Water was distilled with an all-glass Corning "Mega Pure" apparatus.

To minimize errors from solution preparation in drug binding experiments, volumetric flasks and pipets were calibrated gravimetrically by weighing the volume of water contained or delivered.

Instrumentation. Absorbance measurements were carried out with a Gilford spectrophotometer Model 240 modified for automatic recording (Van et al., 1976). Fluorescence measurements were performed in a self-constructed system consisting of the following compounds: a pair of 0.25-m grating monochromators (Jarrell-Ash Model 82-410) for excitation and emission; a 200-W xenon–mercury arc lamp (Hanovia 901B-1) powered by a constant current power supply (Electro Power Pack Model 346); a bialkali photomultiplier (EMI 9635 Q) powered by a highly regulated (±0.1%) HV power supply (SSR 1106); and a single photon counting system (Digital Synchronous Computer Model SSR1110). The excitation of ethidium bromide was achieved with the intense green mercury line near 546 nm with a slit set at 1 mm to attain a pass band of 1.6 nm. Fluorescence at 590 nm was measured at right angles to the emission monochromator. Source fluctuation was estimated to be less than 2% within the entire period of measurement (generally 3–6 h). A calibration curve was obtained prior to each set of measurements to take into account the long-range changes in the intensity of the source, the sensitivity of the photon detector and/or possible disturbance in the wavelength setting and geometry of the system. If a set of

measurements extended for a period longer than 6 h, a new calibration curve was obtained.

Calibration and Fluorescence Measurement. During a typical calibration, 1000 ± 0.005 g of buffer (to avoid pipetting error, the quantity of buffer or of nucleic acid solution was weighed with a Mettler balance before addition of ethidium bromide) was introduced into a quartz cuvette, and fluorescence was counted for 20 or 100 s after each addition of a 10-μL aliquot of an ethidium bromide solution (concentration 40 μM in the same buffer). During the titration, the 1000 ± 0.005 g buffer was replaced by a solution of the nucleic acid containing 15–20 μM of DNA or RNA as phosphate. It should be noted that the same calibrated automatic pipet and the same ethidium bromide solution were used for both calibration and titration steps to minimize errors. Each addition was followed by vigorous vortexing (Vortex Genie, Scientific Industries) to ensure a thorough mixing. The fluorescence readings were in duplicate with an interval of 30–60 s between the two readings to ensure that the counts did not increase with time. A change in the number of photons emitted with time was indicative of the lack of equilibrium. Such changes were corrected by longer waiting periods. Since fluorescence measurement is sensitive to temperature, the sample solution was maintained at 25 °C by a circulating bath and a jacketted cuvette holder. Turbidity errors were minimized by centrifuging the sample for 10 min at 6000 rpm prior to measurement. Data were corrected for background counts and for volume expansion but generally not for "inner filter effect" (Parker, 1968) since the absorbance of added ethidium bromide at 546 nm remained low throughout the titration.

Concentration Determination for Ethidium Bromide and Nucleic Acid Solutions. An accurate knowledge of concentration was essential in comparing various nucleic acids with respect to their fluorescence enhancement capacity and their number of binding sites for ethidium bromide. A molar absorbancy determined for an RNA with a high GC content could not be utilized to determine the concentration of an RNA with a low GC content. Therefore, the molar absorbancy of the two eukaryotic 5.8S rRNAs, $mRNA_{ov}$ and $cDNA_{ov}$, was determined with a Fiske–Subbarow reaction after mineralization of the nucleic acid by a mixture of H_2SO_4 and HNO_3 (Kirkpatrick and Bishop, 1971). The molar absorbancies at 260 nm (expressed in mol^{-1} L cm^{-1} units) are: yeast 5.8S rRNA (8000); mammalian 5.8S rRNA (7700); ovalbumin mRNA (8000); and $cDNA_{ov}$ transcript proper (10 000). The molar absorbancies for the various synthetic nucleic acids were taken from the literature or supplied by the manufacturers. The molar absorbancy of ethidium bromide at 480 nm was 5600 mol^{-1} L cm^{-1} (Waring, 1965).

Circular Dichroism Measurements. Circular dichroism spectra were obtained on a Jasco Model 5 CD spectrometer. Each nucleic acid solution was dissolved in 0.1 M KCl–1 mM phosphate buffer at pH 7 to achieve a final $A_{260nm} = 0.6 \pm 0.1$. The solutions were centrifuged for 15 min at 6000 rpm to remove particulate matter. Samples were contained in a jacketted cylindrical cuvette maintained at 25 °C with a constant temperature circulating bath. Spectra were recorded in duplicate from 200 to 320 nm and corrected for a blank. The reproducibility of duplicate recordings was better than 1% at 265 nm and 2% at 230 nm. Since the Jasco spectrometer was able to directly read in $\Delta\epsilon$, the latter was converted into ellipticity units using the relation:

$$[\theta]\ (\text{deg cm}^2\ \text{dmol}^{-1}) = 3300 \times \Delta\epsilon$$

Definitions and Abbreviations. Abbreviations and symbols for polynucleotides and their constituents were drawn from the IUPAC-IUB Commission recommendations (*J. Mol. Biol.* (1971), *55*, 299–310). $H\% = (A_T/A_{T_0}) - 1$, where A_T and A_{T_0} are the absorbances of the nucleic acid solution at temperature T and at the starting temperature T_0, respectively. $H\%$ at complete denaturation is equal in magnitude to the hypochromicity of the nucleic acid. For other thermal denaturation symbols and parameters (e.g., IDS, index of double strandedness; melting cooperativity and slope of T_m vs. pM), see Van et al. (1976). F = fluorescence; V = fluorescence enhancement ratio, which is the ratio of the fluorescence of a nucleic acid/ethidium bromide mixture to the fluorescence of a solution of free ethidium bromide of identical concentration; V_f = fluorescence enhancement factor, which is the value of V at large ethidium bromide excess. V_f is identical with the constant V defined by Lepecq and Paoletti (1967).

When fluorescence (F) is plotted against ethidium bromide concentration for a solution of nucleic acid, the resulting plot is a relative fluorescence (or F) profile. When the fluorescence profiles of various nucleic acids are normalized to identical nucleic acid concentrations (as phosphate), the fluorescence at the saturating level of drug is a function of the fluorescence enhancement capacity of the nucleic acid. The latter is a function of the concentration of binding sites and of the intrinsic fluorescence enhancement factor of the various binding sites in the nucleic acid.

$$\text{fluorescence enhancement capacity} = \sum_{i=1}^{i=N} n_i(V_f)_i$$

where n_i and $(V_f)_i$ are respectively the mole fraction and the fluorescence enhancement factor of the ith binding site species, and N is the number of binding species with different V_f (this relation states that each base pair contributes independently to the observed fluorescence, but holds only if base sequence effect is negligible). In general, the fluorescence enhancement capacity of a nucleic acid depends on instrument factors (such as intensity of the excitation source and the sensitivity of the photon detector) and the units used to express fluorescence change. To permit interlaboratory comparison, a common standard (such as a solution of free ethidium bromide of known concentration) should be used to establish a reference fluorescence level.

Quantitative Analysis of the Data and Scatchard Plots. The determination of the concentration of drug bound (C_b) and the Scatchard treatment of the binding data have been described in detail by Lepecq and Paoletti (1967). The same approach was followed in the present study.

$$C_b = \frac{I_i - I_0}{k(V_f - 1)}$$

where I_i is the fluorescence of the nucleic acid/ethidium mixture, I_0 is the fluorescence of a solution of free ethidium bromide of equal concentration as the above, k is an instrument factor determined from the calibration curve, and $I_0 = K \times C_0$, where C_0 is the total concentration of ethidium bromide. V_f is the fluorescence enhancement factor of the nucleic acid.

The Scatchard plot was utilized in the following form: $r/C_f = K_a(n - r)$, in which r is the fraction bound, and $r = C_b/[NA]$, where [NA] is the total nucleic acid concentration. The association constant, K_a, and the number of binding sites per nucleotide are given respectively by the slope and the intercept of the Scatchard line (plot of r/C_f vs. r). Only the nucleic acids which give a clear plateau in their fluorescence enhancement profile (Figure 2) yielded a straight line in a Scatchard plot. Figure 4A gives the plots of two binding parameters (C_f and r) which should be taken into consideration during the treat-

TABLE II: Fluorescence Enhancement Capacity and Fluorescence Enhancement Factor of Nucleic Acids.

Compounds	% base pairing	Fluorescence enhancement capacity[a]	Fluorescence enhancement factor[b]	No. of nucleotides required to bind 1 EB[c]
Poly(A-U)	100	100	80 (100%)	3.5-4
Poly(A)·poly(U)	100	57	80 (100%)	7-8
Poly(A,U) No. 1[d]	29	29	80 (100%)	11-13
Poly(A,U) No. 2[e]	60	60	80 (100%)	7
Poly[d(A-T)]	100	88	74 (92%)	3.5-4
Poly[d(A)]·poly[d(T)]	100	45	38 (48%)	3.5-4
Poly[d(G-C)]	100	52	58 (72%)	3.5-4
Poly[d(G)]·poly[d(C)]	100	30	58 (72%)	7-8
Poly(G-C), 0.01 M KCl	100	21	37 (46%)	?
Poly(G-C), 0.50 M KCl	100	10	10 (12%)	?
Poly(A), (U), (G), (C), d(A), d(T), d(G), and d(C)	0	5	5 (6%)	?
Calf thymus DNA	100	75	63 (79%), 65[g]	4.5-5.5
M. lyso. DNA	100	68	58 (73%), 63[g]	4.5-5.5
f-Met-tRNA	50%[f]	13	30 (37%)	?
Yeast 5.8 S	52%[f]	20	46 (58%)	?
Novikoff 5.8 S	57%[f]	18	42 (52%)	?

[a] Expressed as percent of fluorescence enhancement capacity for poly(A-U). [b] Figure in parentheses expresses the V_f factor as a percent of that for poly(A-U). [c] Calculated from Scatchard plots as in text. [d] This random copolymer contained substantial mismatchings as evidenced by 20 °C lower T_m compared with that of poly(A-U). [e] This random copolymer sample contained much less mismatchings as evidenced by 4 °C lower T_m compared with poly(A-U). [f] Percents of base pairing for naturally occurring RNA were estimated from secondary structure models (Van et al., 1977). [g] The fluorescence enhancement factors calculated from the (AT) and (GC) mole fractions according to the relation V_f (calcd) = f(AT) × V_f (A-T) + f(GC) × V_f(G-C). Calf thymus = 42% GC; *M. lysodeikticus* = 72% GC.

ment of the data according to the Scatchard method. In the initial portion of the fluorescence profile where virtually all ligands added are bound, C_f, which is equal to $C_0 - C_b$, is very sensitive to small errors in C_0 and C_b (e.g., a 1% error in C_b is magnified into a 100% error in C_f if 99% of the added ligands are bound). Certain investigators have circumvented this problem by using only the region of the titration curve where a significant amount of the ethidium bromide present remains free (Waring, 1965). Others have fit the experimental data with a least-square regression line and calculated the Scatchard plot from the fitted curve (Rhoads, 1975). Our study used an iterative approach to solve this problem. We first obtained n from the double-reciprocal plot of $1/r$ vs. $1/C_f$. A computer program tested various association constants, K_a, to select the one which yielded a fluorescence profile most similar to the experimental profile maintaining n, V_f, and all the concentration terms identical with the corresponding parameters for the experimental curve. Figure 4B shows the Scatchard plots for poly(A-U), poly(A,U) no. 1, and poly[d(G-C)]. Table II contains some of the estimates of binding site concentration for ethidium bromide in various nucleic acids.

Thermal Denaturation. Details on instrumentation and methodology were given in Van et al. (1976).

Results

(1) *Fluorescence Probing of Secondary Structure in Nucleic Acids.* Thermal denaturation cannot give an accurate estimate of base pairing in naturally occurring RNA since the physical parameter of interest in this technique, hypochromicity, is a function of both base pairing and base stacking. Hypochromicity, when observed at a single wavelength, is also a function of base composition (Van et al., 1976). It is generally agreed that fluorescence probing with ethidium bromide can give a better estimate of base pairing in RNA (Waring, 1965; Lepecq and Paoletti, 1967; Urbanke et al., 1975; Feunteun et al., 1975). However, we found a conflicting result in an earlier study (Van et al., 1977) in which yeast 5.8S rRNA with 52% base pairings (as determined from the secondary structure model for yeast 5.8S rRNA) produced 10% more fluorescence enhancement in ethidium bromide than an RNA which was 57% base paired (i.e., mammalian 5.8S rRNA). This prompted a reinvestigation of the effect of base composition, base sequence and base mismatching on the quantitation of base pairing in nucleic acids.

(a) *Base Pairing and Base Composition Effects.* The base pairing effect can be demonstrated easily by comparing the fluorescence enhancement capacity (see Materials and Methods) of single-stranded model compounds to that of double-stranded model compounds. The fluorescence enhancement capacity of single-stranded nucleic acids is significantly lower than that for all double-stranded model compounds and is less than 5% of that for poly(A-U) (Figure 1A-D and Table I). This result is in good agreement with earlier studies which showed that only base paired nucleic acids can provide strong binding sites to ethidium bromide (Waring, 1965; Lepecq and Paoletti, 1967).

To estimate the base composition effect on fluorescence enhancement of ethidium bromide, the fluorescence enhancement factors (V_f, see Materials and Methods) were calculated from the fluorescence profiles of various nucleic acids which contain only one type of base pairs. Perfect duplexes might differ in their concentration of binding sites for ethidium bromide (see Results, section 1b). Therefore only V_f, the fluorescence enhancement factor of a nucleic acid, can be used as an indicator of base composition effect. The fluorescence enhancement capacity is not suitable for this purpose since it is a function of both the number of binding sites and the intrinsic fluorescence enhancement factor of the latter.

Our results show that adenine-containing pairs are more efficient in increasing the fluorescence of ethidium bromide than GC pairs. The difference in V_f between deoxyribo-(AT) and (GC) pairs is a factor of two when measured under identical conditions. The corresponding difference between ribo-(AU) and (GC) pairs is even larger, especially at high ionic strength (Figure 1B-D; Table II). A tRNA very rich in GC pairs, such as f-Met-tRNA, shows a small fluorescence en-

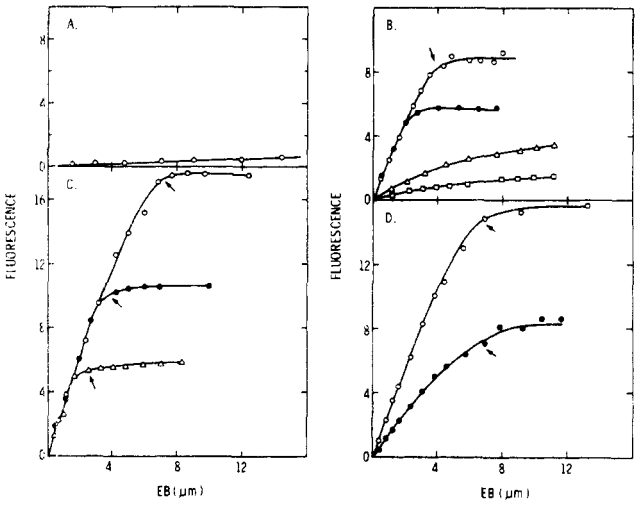

FIGURE 1: Fluorescence profiles of 17 nucleic acids. (A) Fluorescence profiles of eight single-stranded polyribo- and polydeoxyribonucleic acids: Poly(A), (U), (G), (C), d(A), d(T), d(G), and d(C). The plot is for poly(A). All other plots (not shown to simplify presentation) are below that for poly(A). (B) Fluorescence profiles for poly[d(G-C)] (open circle); poly[d(G)]·poly[d(C)] (closed circle); poly(G-C) at 0.01 M KCl (open triangle); and poly(G-C) at 0.5 M KCl (open square). Data for poly(G)·poly(C) (not shown) are similar to data for poly(G-C). (C) Fluorescence profiles for poly(A-U) (open circle); poly(A)·poly(U) (closed circle); and poly(A,U) no. 1 (open triangle). Data for poly(A,U) no. 2 (not shown) are very similar to that for poly(A)·poly(U). The initial parts of the three plots are coincidental within experimental errors. (D) Fluorescence profiles for poly [d(A-T)] (open circle) and poly[d(A)·poly[d(T)] (closed circle). All data are normalized to 1 μM of DNA to facilitate direct comparison. Arrows point to 95% saturation. The abscissa of the arrow is proportional to the concentration of binding sites for ethidium bromide. Data are averages of three to five determinations. Fluorescence units = photons × 10^{-4}/s.

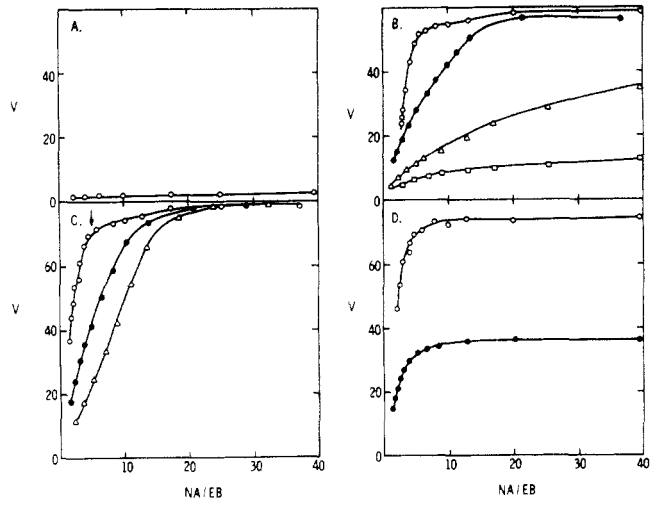

FIGURE 2: Fluorescence enhancement (or V) profiles for the 17 nucleic acids calculated from data in Figure 1. Refer to Figure 1 for symbol identifications. Ordinate: V, defined in text. Abscissa: nucleic acid to drug ratio.

hancement capacity (Table II). The fluorescence enhancement factors of naturally occurring DNA are intermediate between those for poly[d(A-T)] and poly[d(G-C)]. The twofold difference between the fluorescence enhancement factors of poly[d(A-T)] and poly[d(A)]·poly[d(T)] is indicative of the fact that the binding sites in these two duplexes are not identical. Yeast 5.8S rRNA shows a 10% higher fluorescence enhancement capacity than mammalian 5.8S RNA. This result indicates a higher concentration of AU pairs in yeast 5.8S rRNA (Van et al., 1977) rather than a higher number of base pairs in mammalian 5.8S rRNA.

(b) *Base Sequence Effect and Effect of the Coexistence of Two Types of Base Pairs in the Duplexes.* In its simplest form, a base sequence effect can be demonstrated with complementary homopolymers and alternating copolymers. Figure 1 and Table II indicate clearly that duplexes containing the same type of base pairs but different base sequence can show a twofold difference in their fluorescence enhancement capacity for ethidium bromide. Thus poly(A)·poly(U) and poly[d(G)]·poly[d(C)] show 50–60% of the fluorescence enhancement capacity of poly(A-U) and poly[d(G-C)], respectively. Poly[d(A)]·poly[d(T)] also shows half the fluorescence enhancement of poly[d(A-T)]. In all three examples cited above, the comparison is between perfect duplexes containing only one type of base pair. The only difference is that, in alternating copolymers, a purine base is sandwiched between two pyrimidine bases, whereas in complementary homopolymers it is sandwiched between two purine bases. In poly(A-U) and poly[d(G-C)], the number of base pairs required to bind an ethidium bromide molecule is 2, whereas in poly(A)·poly(U) or poly[d(G)]·poly[d(C)], the corresponding number increases to 4 base pairs (Table II). Since every space between the base pair planes is a potential binding site for ethidium bromide, it appears that, in the examples cited above, the binding of an ethidium bromide to one site would preclude the next two (in alternating copolymers) or four (in complementary homopolymers) adjacent sites from binding ethidium bromide. However, in the next example, the nature of the binding sites in poly[d(A)]·poly[d(T)] seems changed, as evidenced by a 50% lower V_f compared with that of poly[d(A-T)]; the number of binding sites of the former is unchanged with respect to that of the latter. The more stacked structure of poly[d(A)]·poly[d(T)] might be unfavorable to a complete immersion of the drug molecule in the hydrophobic environment offered by the base planes of this polymer. Such steric hindrance might reduce the energy exchange between the nucleic acid and the drug with subsequent decrease in the fluorescence quantum yield of the drug bound (Lepecq and Paoletti, 1967).

In duplexes containing two types of base pairs, such as in naturally occurring DNA, the fluorescence enhancement capacity and the fluorescence enhancement factor are 75–85% of the corresponding parameters for poly[d(A-T)]. A DNA which is GC-rich shows, as expected, lower fluorescence enhancement capacity and V_f than an AT-rich DNA (Table II). However, the fluorescence enhancement factors calculated from the mole fractions of AT and GC in these DNAs and the intrinsic fluorescence enhancement factors of AT (V_f = 74) and GC (V_f = 58) pairs do not agree perfectly with the observed V_f factors for chick liver DNA and *M. lysodeikticus* DNA (Table II). Furthermore, naturally occurring DNA also shows a 20–25% smaller concentration of binding sites for ethidium bromide compared with that observed for alternating copolymers. Taken together, these data are interpreted to mean that there is sequence effect in naturally occurring DNA.

(c) *Effect of Base Mismatchings.* The effect of base mismatchings can be studied independently of other factors by comparing the fluorescence parameters of alternating copolymers to that of random copolymers of similar base composition. Only a fraction of the bases in the random copolymer poly(A,U) with A/U = 1:1 are paired since base randomization creates a large percentage of base mismatchings. In Figures 1C and 2C and Table II, the fluorescence parameters are shown for two different samples of random copolymer poly(A,U) containing different amounts of mismatching along

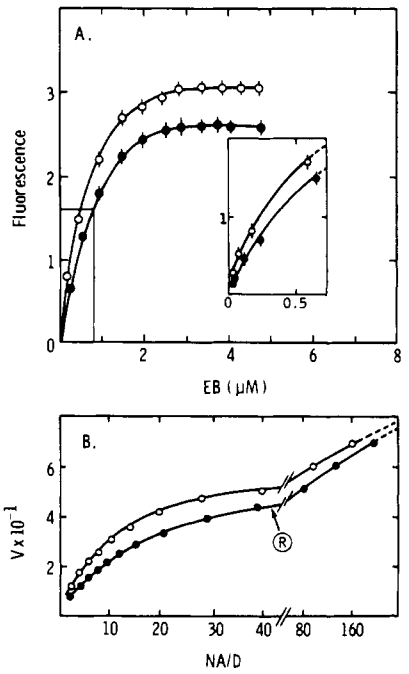

FIGURE 3: (A) Fluorescence profiles for 5.8S yeast rRNA and for $mRNA_{ov}$. Ordinate axis: photons × 10^{-4}/s. Counting time: 20 s. Nucleic acid concentrations: 20 μM ± 2 μM in 0.1 M KCl. The concentration of the two nucleic acids was determined from their molar absorbancy at 260 nm: 8000 and 7700 mol^{-1} L cm^{-1} for $mRNA_{ov}$ and yeast 5.8S rRNA, respectively. Data displayed are the average of four independent determinations. The spread of the data is indicated by error bars. All data are normalized to 1 μM nucleic acid to facilitate comparison. Insert: Detailed view of the first part of the titration. These points were included in the fluorescence enhancement profiles shown in B. (B) Fluorescence enhancement profiles for $mRNA_{ov}$ (closed circle) and yeast 5.8S rRNA (open circle). Arrow points to the V_f value chosen by Rhoads (1975) to calculate C_b.

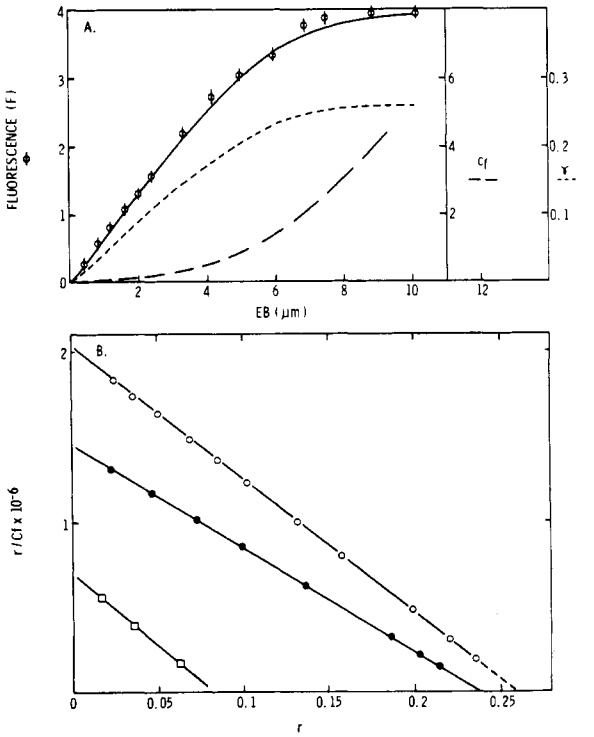

FIGURE 4: Plots of fluorescence and binding parameters for poly(A-U) in 0.01 M KCl (A). Scatchard analysis of binding data for three RNAs (B). (A) Profiles of fluorescence, fraction bound (r) and concentration of free ethidium bromide (C_f) vs. total ethidium bromide concentration. Fluorescence (open circles); fraction of ethidium bromide bound (r) (- - - -); and free ethidium bromide (C_f) (— — —). The solid line is a computer fitted profile (see text) to generate the Scatchard plot below (open circle). (B) Scatchard plots for poly(A-U) (open circle): poly[d(G-C)] (closed circle); and poly(A,U) no. 1 (open square). Fluorescence scale reduced by a factor of 0.24× compared with that in Figure 1.

with the values for the 100% duplex poly(A-U). The differing degree of mismatching in the two random copolymer samples is clearly revealed by a 20 °C lower T_m for poly(A,U) no. 1 and a 4 °C lower T_m for poly(A,U) no. 2 than that of poly(A-U). The fluorescence (F) profiles (Figure 1C) of these random copolymers resemble that observed for the alternating copolymer in the initial portion of the titration curves but plateau at a lower level. The ordinate of the saturation level is a measure of the fluorescence enhancement capacity of the nucleic acid. It is clear that the more mismatchings a random copolymer contains, the lower its fluorescence enhancement capacity. In Table II, the V_f factors for the two random copolymer samples are equal to that for the alternating copolymer ($V_f = 80$). This suggests that the binding sites are the same in these three samples. The percent of base pairing in a random copolymer is therefore equal to the ratio of its fluorescence enhancement capacity to that for the corresponding alternating copolymer. This technique of calculating % base pairing can be extended to duplex populations containing two types of base pair (AU and GC) providing that the V_f factors of these duplexes are similar (see Results, section 1d).

(d) *Quantitation of Secondary Structure in $mRNA_{ov}$ and $cDNA_{ov}$.* In naturally occurring RNA, AU and GC pairs coexist with single-stranded structures and base mismatchings. When these two effects are combined, such as in naturally occurring RNA, the resulting effect is a fluorescence enhancement profile which does not plateau even at a high nucleic acid to drug concentration (Figure 3B). An accurate V_f factor cannot be estimated from such a V profile. A Scatchard treatment of the fluorescence probing profile will yield binding parameters that are entirely dependent on the arbitrary choice of V_f. Attempts to analyze binding data for naturally occurring RNA with a Scatchard plot have not been successful since a single V_f factor cannot be used to calculate the concentration of the drug bound for various populations of binding sites which differ in their fluorescence enhancement factor.

A better estimate of the percent base pairing in an unknown RNA may be derived from the comparison of its fluorescence enhancement capacity to that for a standard RNA of known secondary structure and of similar base composition. Yeast 5.8S rRNA is an ideal reference RNA to use for estimating the percent base pairing in ovalbumin mRNA for two reasons. First, chemical analysis reveals a similar base composition for the two RNAs (42% and 46% G + C for ovalbumin mRNA and yeast rRNA, respectively). Second, both RNA samples display similarity between their derivative denaturation profiles recorded at 260 and 278 nm, indicating that the secondary structures of these RNAs are formed of short duplexes with AU/GC = 1:1. The duplex populations in both RNAs should therefore display similar fluorescence enhancement factors and the technique of estimating base pairing described in Results, section 1c, can be applied if the % base pairing of the reference RNA is known. The fluorescence profile of ovalbumin mRNA shows a 20% lower fluorescence enhancement capacity than that for 5.8S yeast rRNA (Figure 3A). Since the percent of base pairing (excluding G-U pairs) estimated from the secondary structure model (Van et al., 1977) was 52% for yeast 5.8S rRNA, there must be 41 ± 2% base pairing in ovalbumin mRNA.

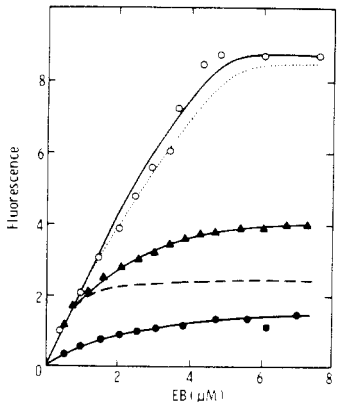

FIGURE 5: Fluorescence profiles for $cDNA_{ov}$, double-stranded $cDNA_{ov}$, and chick liver DNA. Units and buffer as in Figure 3. $cDNA_{ov}$ transcript proper (closed circle); $cDNA_{ov}$ with 22% of the bases as hairpin (closed triangle); chick DNA (open circle); and double-stranded $cDNA_{ov}$ (dotted line). The interrupted line was the difference between closed triangle and closed circle. It corresponded to the fluorescence enhancement capacity of the hairpin proper. The percent of the bases found in the terminal hairpin loop could be calculated as the ratio of the corrected F for cDNA(H) to that for double-stranded $cDNA_{ov}$.

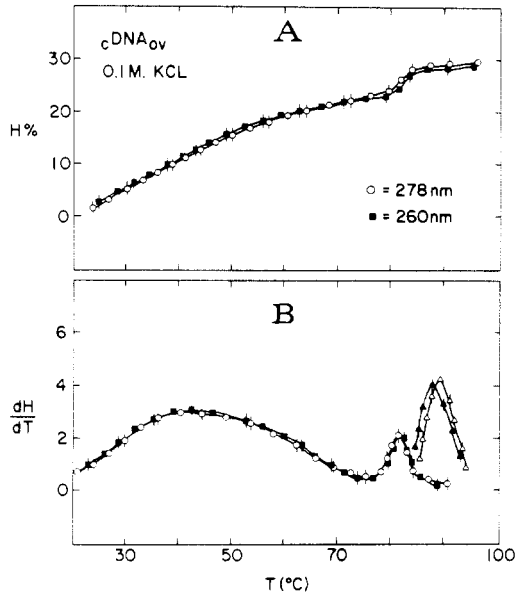

FIGURE 6: H% and derivative profiles dH/dT for $cDNA_{ov}$ at 260 and 278 nm. Both derivative profiles exhibited two distinct peaks, a thermostable peak with T_m near 83 °C and a broad peak with T_m near 40 °C. The thermostable peak accounted for 12% of the total area under the derivative profiles. The similarity in the T_m and the size of the peaks obtained at the two wavelengths 260 nm and 278 nm was within 10% (indicated by error bars). Only one in eight data points was represented by a symbol to simplify presentation. cDNA concentration: 60 μM. Melting buffer: 0.1 M KCl; 260 nm profiles (open circle) and 278 nm profiles (closed square). When the melting buffer also contained 3 μM of ethidium bromide, the T_m values of the two peaks were 60 and 92 °C, respectively. Heating and cooling experiments: A solution of 20 μM of $cDNA_{ov}$ (H) containing about 20% of its bases as a hairpin loop was heated in 0.5 M KCl and allowed to renature. This sample, when reheated, yielded a thermostable peak (closed triangle) similar to the freshly heated sample (open triangle). Data below 80 °C not shown to simplify presentation. Explanatory diagrams shown in Figure 8.

The $cDNA_{ov}$, on the other hand, shows a strikingly different fluorescence profile compared with that of ovalbumin mRNA (Figure 5). The cDNA transcript proper (see Materials and Methods) produces weak fluorescence. Unlike ovalbumin mRNA (or any other naturally occurring RNA tested by us) which exhibits a clear saturation plateau in its *relative* fluorescence profile (Figure 3A), the fluorescence profile of this $cDNA_{ov}$ transcript does not reach a saturation plateau at three times higher ethidium bromide concentration (Figure 5). This lack of saturation is not due to a high concentration of binding sites for ethidium bromide in $cDNA_{ov}$ but might be the result of the stabilization effect of excess ethidium bromide on the unstable, AT-rich duplexes in $cDNA_{ov}$ [the T_m of the thermolabile peak in the thermal denaturation profile of $cDNA_{ov}$ (Figure 6B) increased by 20 °C in the presence of 3 μM of ethidium bromide; see Results, section 2]. Therefore, we conclude that the *internal* hairpin loops (the counterparts of the short duplexes in ovalbumin mRNA) in $cDNA_{ov}$ are not stable under physiological ionic strength.

The fluorescence enhancement capacity of the $cDNA_{ov}$ preparations which contain a terminal hairpin loop [$cDNA_{ov}$ (H); see Materials and Methods] is directly proportional to the length of the hairpin loop (Figure 5). The length of the hairpin loop in a $cDNA_{ov}$ (H) preparation is equal to the ratio of the fluorescence enhancement of the latter [corrected for the contribution of the single-stranded region (Figure 5)] to that of the full length cDNA. The above technique of hairpin loop quantitation is valid since the V_f of the reference DNA (full length double-stranded $cDNA_{ov}$) is identical with that of the hairpin loop region in $cDNA_{ov}$ (H). Such length can also be estimated from the analysis of the derivative profiles (see Results, section 2). The completely double-stranded $cDNA_{ov}$, as expected, shows virtually an identical profile when compared with that of chick liver DNA (Figure 5).

(2) *Thermal Denaturation of $cDNA_{ov}$.* The basis for interpreting thermal denaturation data in this section has been described in detail previously (Van et al., 1976). When the various $cDNA_{ov}$ preparations were subjected to the same systematic thermal denaturation studies as performed on ovalbumin mRNA (Van et al., 1976), the results summarized in Table III were obtained. Data for ovalbumin mRNA and chick DNA were also included to facilitate a direct comparison.

(a) *$cDNA_{ov}$ Transcript Proper.* To detect duplex regions of different base composition in ovalbumin cDNA, we applied a technique described in a preceding paper (Van et al., 1976) which consists of comparing the hypochromicity of ovalbumin cDNA at 260 and 278 nm, respectively. The basis for this technique was derived from the work of Fresco and his collaborators (Fresco, 1963; Fresco et al., 1963).

Ovalbumin mRNA and its cDNA transcript display similar hypochromicity at both wavelengths examined, 260 nm and 278 nm (~30%, Table III), but the former is significantly more thermostable than the latter. It can be seen, for example, that the T_m of ovalbumin mRNA is 6 °C higher than that of $cDNA_{ov}$. The change in the T_m with ionic strength is twice as large with ovalbumin mRNA as that observed with the $cDNA_{ov}$. The melting cooperativity of the latter is also significantly lower than that of ovalbumin mRNA (Table III). The index of double-strandedness (IDS) [which combines hypochromicity with melting cooperativity and slope of the change in T_m with ionic strength (Van et al., 1976)] is almost four times smaller for the $cDNA_{ov}$ than for the messenger RNA (Table III). An IDS of 7 would place the $cDNA_{ov}$ molecule in the same group as the random copolymer poly(A,U,G,C) which forms a separate class between single-stranded structures and naturally occurring RNA (Van et al., 1976).

(b) *$cDNA_{ov}$ Preparations Containing a Hairpin Loop [$cDNA_{ov}$ (H)].* Recently, we have shown that by plotting together the derivative profiles recorded at 260 and 278 nm, duplex regions containing different base compositions can be detected (Van et al., 1976). Such data plots are presented in

TABLE III: Thermal Denaturation Data for $mRNA_{ov}$, $cDNA_{ov}$, and Chick DNA.

Thermal denaturation parameters	$mRNA_{ov}$	$cDNA_{ov}$ (H) (transcript region)	$cDNA_{ov}$ (H) (hairpin region)	Double-stranded $cDNA_{ov}$	Chick DNA
T_m in 0.1 M KCl	46 °C	40 °C[a]	82 °C	83 °C	80 °C
Sensitivity of slope of T_m with pM[b]	17 °C/pM	8 °C/pM	13 °C/pM	NA[e]	15 °C/pM
Melting cooperativity at 0.1 M KCl (transition half width)	20 °C	28 °C	4 °C	4 °C	5 °C
Hypochromicity at 260 nm	30%	26%[a]	4%[c]	36%	39%
Index of double strandedness (IDS)[d]	26	7		135	117

[a] Excluding the hairpin loop (Figure 6). The cDNA transcript proper showed 30% hypochromicity. [b] pM is the logarithm of the reciprocal ionic concentration. The slope of T with pM was determined at 5 ionic concentrations for $mRNA_{ov}$ and chick DNA (0.01, 0.02, 0.05, 0.1, 0.5 M). The T_m for $cDNA_{ov}$ was determined at only two ionic strengths (0.01 and 0.1 M KCl) for sample scarcity reason. [c] Determined from the denaturation profile at 260 nm for a sample containing about 13% of the bases as hairpin loop. [d] IDS = (hypochromicity at 260 nm × slope of T_m with pM)/melting cooperativity. [e] NA = Not available for reason of sample scarcity. For the definitions of hypochromicity, melting cooperativity, and index of double strandedness, refer to Van et al. (1976).

Figures 6A and 6B for a cDNA preparation with about 10% of the bases in the hairpin loop. Both derivative profiles show two distinct peaks: a broad peak with T_m at 40 °C and a thermostable peak with T_m near 83 °C. Furthermore, two profiles recorded at 260 and 278 nm are almost identical and are more alike than the corresponding profiles observed for ovalbumin mRNA (Van et al., 1976). The derivative profile for the ovalbumin mRNA displays a comparable thermolabile peak with a T_m near 46 °C (Van et al., 1976; Table III) but only a small thermostable peak which melts at 82 °C. The latter peak is present in the 278 nm recording but not in the derivative profile at 260 nm. This peak has been correlated with a GC-rich duplex population accounting for about 4% of the bases in ovalbumin mRNA (Van et al., 1976). The thermostable peak in the derivative profiles for $cDNA_{ov}$ is not considered to be a counterpart of the same GC-rich peak in ovalbumin mRNA for several reasons. The following summary presents physical evidence that the thermostable peak arises from the melting of a terminal hairpin loop in the $cDNA_{ov}$ molecule.

To prove that the thermostable peak in ovalbumin cDNA is not due to the melting of a GC-rich duplex population, the area under the derivative profiles of this molecule recorded at 260 and 278 nm was compared. In the derivative profiles shown in Figure 6B, the similarity in the size, the T_m and the form of the two thermostable peaks is remarkable. A GC-rich duplex population is revealed as a very large peak in the derivative profile recorded at 278 nm compared with a very small peak in the profile at 260 nm (Van et al., 1976). The possibility that the high T_m peak might arise from the melting of an AT-rich duplex [formed between the poly(A) track of the $mRNA_{ov}$ molecule (present as a potential contaminant) with oligo[d(T)] added as primer] can be ruled out since the peak recorded at 260 nm should be much larger than that recorded at 278 nm.

When various $cDNA_{ov}$ (H) preparations were denatured, the size of the thermostable peak varied from a few percentage points to 28% of that for the full-length double-stranded $cDNA_{ov}$ (Figures 6 and 7). The variability in the size of the thermostable peak with $cDNA_{ov}$ preparations might conceivably be due to a single nucleic acid sequence assuming conformations with different amounts of secondary structures. To test this hypothesis, heating and cooling experiments were performed. In such experiments, the $cDNA_{ov}$ was heated in a high ionic strength buffer (which favors renaturation) until complete denaturation of the thermostable duplex population was accomplished. This denatured sample was then allowed to renature and then was reheated to yield a second denaturation profile. Our assumption was that, if the $cDNA_{ov}$ sequence

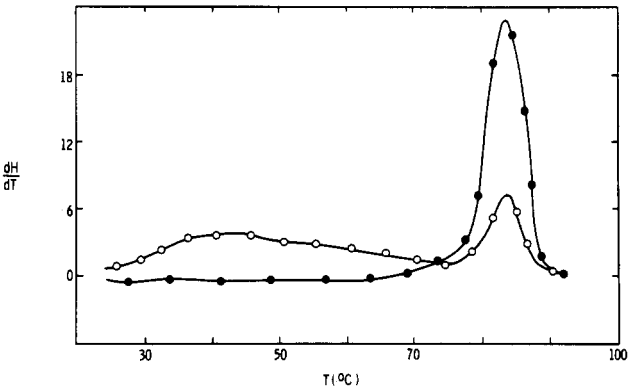

FIGURE 7: Derivative profiles at 260 nm for $cDNA_{ov}$ containing 28% hairpin loop and the double-stranded $cDNA_{ov}$. The high T_m and high cooperativity for the hairpin loop were shown in Table II. These parameters compared favorably with that for double-stranded chick DNA. Double-stranded $cDNA_{ov}$ (closed circle) and $cDNA_{ov}$ with 28% of the bases in the hairpin loop (open circle). Melting buffer = 0.1 M KCl.

can renature to different conformations with different amounts of secondary structure, the reheated sample should yield a different derivative profile when compared with the initial profile. The data in Figure 6B show that the two profiles are essentially the same. (An explanatory diagram for the heating and cooling experiments is given in Figure 8.)

It is also unlikely that the thermostable peak in $cDNA_{ov}$ resulted from the melting of an internal loop (or loops) in this molecule. Such a loop, if it were to exist, should also be present in the denaturation profile of the mRNA template and should be observed as an even higher T_m peak since double-stranded RNA is known to be more stable than double-stranded DNA of similar base composition (Figure 8C). The possibility that such a peak might be too stable to be observed in the range 0–100 °C can be excluded since the melting of $mRNA_{ov}$ at very low ionic strength (1 mM KCl) failed to reveal such a peak (Van et al., 1976). The evidence for a unique loop and not a population of short duplexes is given by the high T_m and the high melting cooperativity of this thermostable peak (Table III). These two parameters compare favorably with the corresponding parameters for double-stranded chick DNA.

(c) *Double-Stranded $cDNA_{ov}$.* Figure 7 shows the thermal denaturation profile for a full-length double-stranded cDNA. Table III contains information about the melting cooperativity and the thermal stability of this DNA. These two parameters compare favorably with those for chick DNA. The base composition of the double-stranded $cDNA_{ov}$ calculated from its T_m at 0.1 M KCl using Schildkraut and Lifson's equation (1965) is 44% GC. This estimate is close to the chemically

TABLE IV: Ellipticities (deg cm^2 dmol^{-1}) and Wavelengths of Maximum Optical Activity for Five Nucleic Acids.

	$mRNA_{ov}$	$cDNA_{ov}$	Hb mRNA*[a]	Chick liver DNA	Poly(U)
Positive Extrema					
1	20000 (267 nm)	12300 (275 nm)	23000 (267 nm)	9500 (275 nm)	17700 (275 nm)
2	8000 (219 nm)	3500 (225 nm)	5000 (220 nm)	4500 (220 nm)	NA
Negative Extrema					
1	−100 (300 nm)	0	0	0	0
2	−3500 (242 nm)	−10000 (245 nm)	−4000 (240 nm)	−12000 (245 nm)	−11000 (245 nm)
$[\theta]\lambda_1/[\theta]\lambda_2$	5.7	1.23	5.75	0.79	1.6

[a] From Bobst et al. (1974), in the absence of Mg^{2+}. [b] λ_1 = 275 nm for DNA, cDNA, and poly(U); 267 nm for $mRNA_{ov}$ and globin mRNA. λ_2 = 245 nm for chick DNA and $cDNA_{ov}$; 242 nm for mRNA.

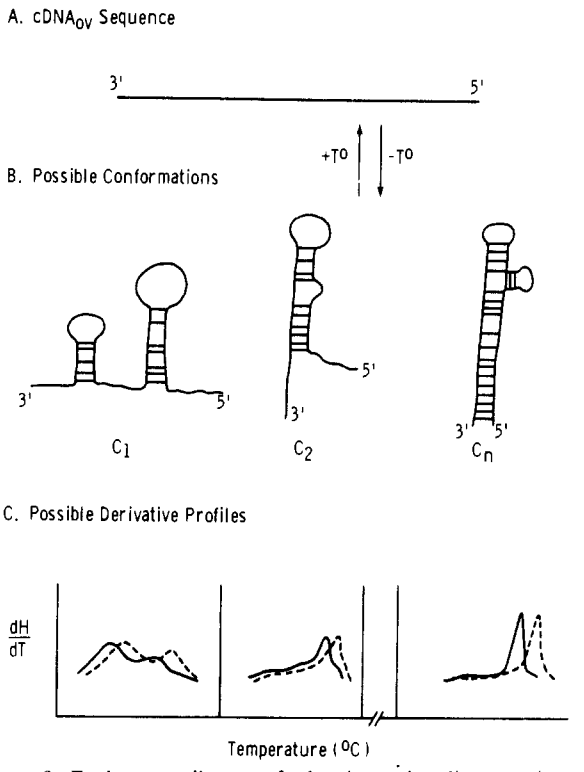

FIGURE 8: Explanatory diagrams for heating and cooling experiments. (1) A $cDNA_{ov}$ sequence (A) might possess such base complementarity that various conformations C_1, C_2 ... C_n, could be assumed by this $cDNA_{ov}$ molecule (B). Since the extent and nature of base pairing are different with various conformations, each of these conformations is expected to yield a distinct derivative profile, dH/dT (C, solid lines). If only one of the above conformations is thermodynamically probable (stable in solution), then an identical profile is expected for both "native" and reheated samples (e.g., the size, shape, and T_m of the thermostable peak should be invariant). If more than one conformation is stable, then the derivative profiles of the "native" sample should be different from that for a renatured sample since the histogram for conformation species in a renatured $cDNA_{ov}$ sample is expected to be different from that for the "native" sample. (2) Expected derivative denaturation profiles for the RNA counterparts of the $cDNA_{ov}$ conformations shown in B (interrupted lines) (C).

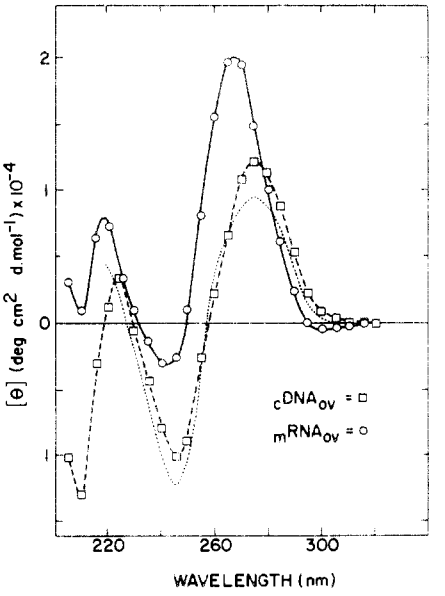

FIGURE 9: Circular dichroism spectra for $mRNA_{ov}$, $cDNA_{ov}$, and chick liver DNA. Spectra parameters were presented in Table II. $mRNA_{ov}$ (open circle); $cDNA_{ov}$ (open square); and chick liver DNA (dotted line). Buffer: 0.1 M KCl.

determined GC content of ovalbumin mRNA (42%; Woo et al., 1975).

(3) *Circular Dichroism of $mRNA_{ov}$ and the $cDNA_{ov}$ Transcript.* The helical structure of a nucleic acid [which reflects not only the conformation in the base pairing regions but also the stacking structures in the single-stranded region of nucleic acids (Brahms and Brahms, 1970)] has been correlated with its CD spectrum in a theory developed by Johnson and Tinoco (1969). It was thus interesting to compare the CD spectra of $mRNA_{ov}$ and the $cDNA_{ov}$ transcript to obtain additional insight into their helical structure. The CD spectra of these two molecules are shown in Figure 9. To facilitate a comparison, CD spectral parameters for $cDNA_{ov}$ are presented in Table IV along with the corresponding data for globin mRNA, calf thymus DNA, and poly(U).

The unequal amplitude of the two extrema at 267 nm and 242 nm observed with $mRNA_{ov}$ is characteristic of naturally occurring RNA (Brahms and Brahms, 1970). The CD spectrum for $mRNA_{ov}$ would, therefore, be described as nonconservative (Bush and Brahms, 1967). The 10% higher ellipticity observed near 265 nm with globin mRNA might be due to a higher content of poly(A) in this RNA (10–12% vs. 3–4%). When corrected for the difference in poly(A) content as described by Bobst et al. (1974), the molar ellipticity at 265 nm for these two eukaryotic mRNAs was very close (18 000 deg cm^2 dmol^{-1}). However, the poly(A) contribution correction did not nullify the difference noted at the second extremum at 242 nm. Furthermore, $mRNA_{ov}$ also displayed a small negative ellipticity at 300 nm which was not present in the CD spectrum of globin mRNA when the latter was in a buffer not containing Mg^{2+} (Bobst et al., 1974). The CD data, therefore, suggested that some subtle differences in the helical structures of these eukaryotic mRNAs do exist.

The CD spectrum for a $cDNA_{ov}$ transcript sample is shown in Figure 9 along with the spectrum for double-stranded chick DNA. The molar ellipticity at 275 nm was about 25% higher for the $cDNA_{ov}$ than for chick DNA. We were curious about the possibility that a higher molar ellipticity in $cDNA_{ov}$ might

be caused by an underestimate of the $cDNA_{ov}$ concentration. Such an error would be reflected in the absolute value of the molar ellipticity but not in the ratio of the molar ellipticities at different wavelengths. The $[\theta]_{275nm}/[\theta]_{245nm}$ ratio for chick DNA was less than unity (0.79), whereas that for $cDNA_{ov}$ was larger than unity (1.23). Such a difference reflects a real distribution between the helical structure of $cDNA_{ov}$ and double-stranded DNA. On the other hand, the $[\theta]_{267nm}/[\theta]_{242nm}$ ratio was much higher with all naturally occurring RNAs (i.e., 6) (Table IV). Taken together, these data indicate that the helical structure of $cDNA_{ov}$ is representative of a structure which is distinct from that observed for double-stranded DNA and naturally occurring RNA. Other nucleic acids in the synthetic series such as poly(A), poly(G), and poly(C) display drastically different CD spectra with respect to both molar ellipticities and $[\theta]_{275nm}/[\theta]_{245nm}$ ratios. However, the CD data for the single-stranded homopolymer poly(U) were close to those for $cDNA_{ov}$ with respect to two important criteria: the position and the amplitude of the extrema at 275 nm and 245 nm. The ellipticity ratio $[\theta]_{275nm}/[\theta]_{245nm}$ also agreed within 30%. We therefore speculate that the helical structure of $cDNA_{ov}$ might be similar to that assumed by poly(U) in solution (Richards et al., 1963; Simkins and Richards, 1967; Thrierr et al., 1971). It should be noted, though, that the absence of base pairing does not imply that the $cDNA_{ov}$ sequence was in a fully random state as is attributed to denatured nucleic acid at very high temperature.

Discussion

Two conclusions emerged from the present study. First, the level of base pairing in the $cDNA_{ov}$ transcript proper was quite low compared with the 40–42% estimate of base pairing in ovalbumin mRNA. Second, the existence of a terminal loop was demonstrated by physicochemical criteria in the $cDNA_{ov}$ but not in the ovalbumin mRNA molecule itself. Data from all three physical techniques were in general agreement that there was little secondary structure in $cDNA_{ov}$ except for the hairpin loop. However, the large hypochromicity of $cDNA_{ov}$ indicated that this molecule was not in a random conformation such as that prevailing in denatured DNA at high temperature. The CD results supported such an interpretation.

An earlier estimate for the extent of base pairing for ovalbumin mRNA (Rhoads, 1975) was about 18–20% higher than our estimate. We have ruled out the possibility that the difference between these two estimates might be due to instrumental factors or to the difference in the ethidium bromide samples used in the two studies. In support for this claim, a very similar estimate of binding sites for ethidium bromide in poly(A)·poly(U) was derived from both studies (7–8 nucleotides per ethidium bromide). We believe that the earlier higher estimate was the consequence of two arbitrary assumptions. First, the fluorescence enhancement factor of 47 which was chosen to calculate the concentration of the drug bound is still in the nonsaturating portion of the fluorescence enhancement profile for ovalbumin mRNA (Figure 3B). However, even using such an estimate for V_f, the concentration of bound drug would still be underestimated (see Results, section 1). Second, Rhoads (1975) assumed that the complementary homopolymer mixture poly(A)·poly(U) could be used as a reference compound to estimate the percent of base pairing in ovalbumin mRNA. However, the compound poly(A)·poly(U) happens to contain only about half the number of binding sites as the alternating copolymers and/or naturally occurring DNA (see Results, section 1). The net effect of these two assumptions resulted in an overestimate of the percent base pairing for ovalbumin mRNA.

An estimate of 55–62% base pairing was derived by Holder and Lingrel (1975) for rabbit globin messenger RNA from thermal denaturation data using 18S rRNA as a reference compound. The Favre et al. estimate for both duck and rabbit globin messenger RNA was 45–60% base pairing (Favre et al., 1975). According to these authors, the above wide margin of uncertainty is justified in view of the base composition effect which can introduce large errors in the estimate of base-pairing obtained with the fluorescence titration technique. Another source of error might be the presence of contaminants, especially as DNA, in the messenger RNA samples. Recent advances in the field of messenger RNA purification have reduced this second source of error to negligible proportion, at least in the case of ovalbumin messenger RNA. However, any estimate of base pairing (including our own) in messenger RNA still suffers from the inherent inaccuracy of the techniques utilized and also from the inadequacy of the RNA molecule used as the reference compound. A single physical technique is clearly not sufficient to establish an accurate estimate of base pairing. Thus the estimate of base pairing of the reference RNA molecule should be obtained by an independent technique other than that proposed for the study of the unknown RNA.

Efstratiadis et al. (1976) have presented the following biochemical evidence for the existence of a terminal hairpin loop in globin cRNA: (i) when globin RNA was subjected to S1 nuclease action, resistant fragments could be isolated; (ii) hybridization kinetics indicated that the duplex fragments were from intramolecular pairing; and (iii) electrophoretic analysis of cDNA preparations of varying lengths showed that the hairpin loop could not be an internal loop. Our studies provide additional physical evidence for the presence of a terminal hairpin loop in ovalbumin cRNA. The terminal nature of this hairpin loop was suggested by the variability in the length of this loop (as assessed from the derivative profiles) in different $cDNA_{ov}$ preparations if the two following results are also taken into consideration: first, freshly denatured and reheated $cDNA_{ov}$ (H) samples from the same preparation yielded identical derivative denaturation profiles, indicating that the variable size of the thermostable peak obtained with different $cDNA_{ov}$ (H) preparations was not a result of multiple conformations of the $cDNA_{ov}$ sequence; second, the absence of similar thermostable structures in the derivative profiles on the $mRNA_{ov}$ molecule (i.e., the exact complement of the $cDNA_{ov}$ molecule). The hypothesis that this thermostable peak corresponded to the melting of a GC-rich duplex population which was known to be present in ovalbumin mRNA has been excluded. Finally, our data support the presence of a unique loop rather than a population of loops since both a high T_m and a high melting cooperativity of the thermostable peak were noted.

We have no data at this time which will permit us to conclude that the AMV reverse transcriptase will always proceed to the extreme 5′ end before synthesizing a hairpin loop. Indeed, much of our evidence is consistent with the conclusion that premature loops can also be initiated by the AMV reverse transcriptase enzyme when it encounters some undefined turnback signal in its path. Such a signal could be in the form of a secondary structure and/or a specific sequence in the messenger RNA molecule. However, the capability of the AMV reverse transcriptase enzyme to copy bidirectionally was clearly demonstrated. This is also a property shared by other replicase enzymes (Masters and Broda, 1971). Even in the presence of actinomycin D, which stabilizes the mRNA/cDNA hybrid duplex, the AMV reverse transcriptase enzyme might have a built-in mechanism (possibly in the form of an

unwinding protein, or more likely, of an RNase H activity (Leis et al., 1973)) to displace a substantial fraction (up to 28%) of the $mRNA_{ov}$ template from the newly synthesized $cDNA_{ov}$ strand.

Last, the observation that short DNA *internal* hairpin loops are not stable under physiological conditions cannot be generalized to long palindromic structures (Wilson and Thomas, 1972) which involve the folding back of inverse repetitive sequences.

Acknowledgments

The authors wish to thank Drs. A. T. Ansevin, H. C. Towle, and J. M. Rosen for helpful discussions, Dr. Tom Su for the use of a Jasco Model 5 spectrometer, and Mr. Tim Yu for excellent technical assistance.

References

Bobst, A. M., Pan, Y. C., and Phillips, D. J. (1974), *Biochemistry 13,* 2129.

Brahms, J., and Brahms, S. (1970), in Fine Structure of Proteins and Nucleic Acids, Biological Macromolecules Series, Vol. 4, Fasman, G. D., and Timasheff, S. N., Ed., New York, N.Y., Marcel Dekker.

Bush, C. A., and Brahms, J. (1967), *J. Chem. Phys. 46,* 79.

Efstratiadis, A., Kafatos, F. C., Maxam, A. M., and Maniatis, T. (1976), *Cell 7,* 270.

Favre, A., Morel, C., and Scherrer, K. (1975), *Eur. J. Biochem. 57,* 147.

Feunteun, J., Monier, R., Garrett, R., Le Bret, M., and Lepecq, J. B. (1975), *J. Mol. Biol. 69,* 217.

Fresco, J. R. (1963), in Informational Macromolecules, Vogel, H. J., et al., Ed., New York, N.Y., Academic Press, p 121.

Fresco, J. R., Klotz, L. C., and Richards, E. G. (1963), *Cold Spring Harbor Symp. Quant. Biol. 28,* 83.

Holder, J. W., and Lingrel, J. B. (1975), *Biochemistry 14,* 4209.

Johnson, W. C., and Tinoco, Jr., I. (1969), *Biopolymers 7,* 727.

Kirpatrick, D. S., and Bishop, S. H. (1971), *Anal. Chem. 43,* 1707.

Leis, J., Hurwitz, J., Schincariol, A. L., Stone, M., and Joklik, W. K. (1973), in The Biology of Tumor Viruses, Proceedings of the 34th Annual Biology Colloquium, Corvallis, Ore., Oregon State University Press, pp 77–106.

Lepecq, J. B., and Paoletti, C. (1967), *J. Mol. Biol. 27,* 87.

Marmur, J. (1961), *J. Mol. Biol. 3,* 208.

Marmur, J., and Doty, P. (1962), *J. Mol. Biol. 5,* 109.

Masters, M., and Broda, P. (1971), *Nature (London), New Biol. 232,* 137.

Monahan, J. J., Harris, S. E., Woo, S. L. C., Robberson, D. L., and O'Malley, B. W. (1976a), *Biochemistry 15,* 223.

Monahan, J. J., McReynolds, L. A., and O'Malley, B. W. (1976b), *J. Biol. Chem. 251,* 7355.

Nazar, R. N., Sitz, T. O., and Busch, H. (1975), *J. Biol. Chem. 250,* 8591.

Parker, C. A. (1968), in Photoluminescence of Solutions, Amsterdam, Elsevier.

Rhoads, R. E. (1975), *J. Biol. Chem. 250,* 8088.

Richards, E. G., Flessel, C. P., and Fresco, J. R. (1963), *Biopolymers 1,* 431.

Rosen, J. M., Woo, S. L. C., Holder, J. W., Means, A. R., and O'Malley, B. W. (1975), *Biochemistry 14,* 69.

Schildkraut, C., and Lifson, S. (1965), *Biopolymers 3,* 1975.

Simpkins, H., and Richards, E. G. (1967), *Biopolymers 5,* 551.

Thrierr, J. C., Dourlent, M., and Leng, M. (1971), *J. Mol. Biol. 58,* 815.

Urbanke, C., Romer, R., and Maas, G. (1975), *Eur. J. Biochem. 55,* 439.

Van, N. T., Holder, J. W., Woo, S. L. C., Means, A. R., and O'Malley, B. W. (1976), *Biochemistry 15,* 2054.

Van, N. T., Nazar, R. N., and Sitz, T. O. (1977), *Biochemistry 16* (in press).

Waring, M. J. (1965), *J. Mol. Biol. 13,* 269.

Woo, S. L. C., Rosen, J. M., Liarakos, C. D., Robberson, D. L., Choi, Y. C., Busch, H., and O'Malley, B. W. (1975), *J. Biol. Chem. 250,* 7027.