DETERMINATION OF MELTING SEQUENCES IN DNA AND DNA-PROTEIN COMPLEXES BY DIFFERENCE SPECTRA

ANTHONY P. RUSSELL, ROBERT L. HERRMANN, *and* LENEAL E. DOWLING

From the Veterans Administration Hospital, Bedford, Massachusetts 01730 and the Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts 02118

ABSTRACT A graphical formula is presented for determining the base ratio of melted DNA. By use of this formula, the composition of sequences which melt in different portions of the melting curves of *Clostridium* DNA, *Escherichia coli* DNA, and mouse DNA were determined. As the DNA melts, the per cent of adenine and thymine (AT) in the melted sequences decreases linearly with temperature. The average composition of sequences which melt in a given part of the melting curve is proportional to the base ratio of the DNA. The concentration and average composition of sequences were determined for three parts of the melting curves of the DNA samples, and a frequency distribution curve was constructed. The curve is symmetrical and has a maximum at about 56% AT. The distribution of GC-rich sequences on the *E. coli* chromosome was estimated by shearing, partially melting, and fractionating the DNA on hydroxylapatite. GC-rich sequences appear to occur every thousand base pairs, and have a maximum length of about 180 base pairs. The graphical formula was applied to the determination of the composition of sequences which melt in different parts of the melting curve of chromatin. Throughout the melting curve, the composition of the melting sequences is about 60% AT, which appears to suggest that relatively long sequences are melting simultaneously. Their melting temperature may be a function of the composition of the protein on different parts of the DNA. The problem of light scattering in DNA-protein and DNA was also investigated. A formula is presented which corrects for light scattering by relating the intensity of the scattered light to the rate of change of absorbance of DNA with wavelength.

INTRODUCTION

DNA undergoes an increase in ultraviolet absorption, or hyperchromic shift, if it is heated to a temperature which causes the strands to separate. In 1962, Felsenfeld and Sandeen (1) showed that the spectrum which resulted from this increase in absorption could be used to measure the concentration of the base pairs adenine

plus thymine (AT) and guanine plus cytosine (GC) in the DNA. They found that AT pairs have a lower thermal stability than GC pairs, and studied the melting behavior of these bases in both double- and single-stranded DNA. The spectral techniques were further developed by Fresco, Klotz, and Richards (2) and Mahler, Klein, and Mehrotra (3). Felsenfeld and Hirschman (4) have studied the effect of interactions of neighboring bases on the DNA strands on spectra. A method for determining the concentration of the base pairs in DNA under various conditions has been published by Hirschman and Felsenfeld (5).

In this paper we develop a graphical formula for the determination of the composition of sequences which melt in DNA. This formula is applied to the melting curve of the DNA from two species of bacteria and the mouse. The characteristics of the melting curve are compared. As *Escherichia coli* DNA is rich in GC, we investigate the distribution and length of the GC-rich sequences by fractionation of the DNA on hydroxylapatite.

Olins, Olins, and von Hippel (6, 7) have shown that spectra of DNA-protein contain an anomalous distortion due to light scattering. In order to apply our spectral method to the study of chromatin, we develop a formula for the correction of light scattering, and apply it to the study of the composition of melting sequences.

METHODS

Calf thymus, *E. coli*, and *Clostridium perfringens* DNA were purchased from Worthington Biochemical Corp. (Freehold, N. J.). The DNA from the C57BL/6J mice were isolated from liver nuclei. The nuclei were isolated according to the method of Blobel and Potter (8). Chromatin was extracted from the nuclei according to the method of Commerford et al. (9), and DNA was isolated directly from nuclei by the method of Marmur (10).

The DNA concentration was determined by the Burton modification of the diphenylamine procedure (11). Protein concentration was measured by the Lowry procedure (12), using histone as a standard. The base ratio of the mouse DNA was 57.4% AT as determined by chromatography, 57% as determined by the method of Mindich and Hotchkiss (13), and is given by Marmur and Doty (14) as 59%.

The DNA was dissolved in either 0.03 M NaCl—0.003 M sodium citrate or 0.0025 M NaCl—0.00025 M sodium citrate, and dialyzed for 24 hr against two changes of the solvent. Solutions were degassed by vacuum.

Optical measurements were carried out in a Beckman DU spectrophotometer (Beckman Instruments, Inc., Fullerton, Calif.). The cell compartment was equipped with a Beckman thermal spacer set, and the temperature of the cell compartment was controlled by circulating water from a Tamson water bath (made by P. M. Tamson N.V., The Netherlands, supplied by Neslab Instruments, Durham, N.H.). Temperature of the sample was monitored by a YSI telethermometer (Yellow Springs Instrument Co., Yellow Springs, Ohio). 20 min were permitted for temperature equilibration. After equilibration, the temperature of the sample varied less than 0.1°C over a period of 15 min.

Sample and blank were placed in Beckman glass-stoppered silica cells with a 1 cm light path. The glass stopper of the sample cell was replaced with a machined nylon stopper. A glass bead thermistor passed through the stopper and into the solution. The YSI telethermometer was accurate to 0.5°C with a reproducibility of 0.1°C. Spectra of bases were taken from Beaven et al. (15).

Polylysine was complexed with DNA according to the method of Olins et al. (6), with the exception that their cacodylate buffer was replaced with phosphate buffer. Poly-L-lysine, mol wt 150,000, was purchased from Mann Research Labs Inc (New York). DNA in 0.03 M phosphate buffer, pH 6.8, was made 4 M in NaCl. Polylysine was dissolved in 0.03 M phosphate buffer, pH 6.8. Polylysine was added to DNA so that the mixture contained the following equivalents of lysine to DNA phosphorous: 0.17, 0.33, 0.50, 0.67, and 1.00. The mixtures were dialyzed at 4°C against 0.4 M NaCl and 0.3 M NaCl for 4 hr each, overnight against 0.15 M NaCl, and for 24 hr against 0.03 M phosphate buffer. Precipitates were removed by centrifugation for 15 min at 1,700 $\times$ *g*. The resultant ratios of lysine to DNA phosphorous were 0.14, 0.27, 0.71, 0.84, and 1.16.

DNA was chromatographed on hydroxylapatite by a technique similar to that of Bernardi (16). The column was made of SERVA hydroxylapatite on cellulose (Gallard-Schlesinger Chemical Mfg. Corp., L. I., N. Y.). As the column was used to fractionate partially melted DNA in 0.15 M phosphate buffer, pH 6.5, the temperature of the column was maintained at 88°C in one experiment, and 90°C in another. DNA was eluted from the column at 0.15, 0.3, and 0.9 M phosphate buffer. Thermal melting profiles and spectra of the fractions were recorded.

RESULTS

Method of Analysis of DNA Spectrum

The spectrum at a neutral pH of the deoxyribonucleosides published by Beaven et al. (15) has been used to determine the concentrations of the bases in melted DNA. The following graphical formula is used, which is similar to that of Fresco et al. (2):

$$\frac{A_\lambda}{E_{(AT)\lambda}} = M_{(AT)} + M_{(GC)} \frac{E_{(GC)\lambda}}{E_{(AT)\lambda}}, \qquad (1)$$

where A_λ = absorbance at wavelength λ, E = the extinction coefficient at the wavelength λ of either the adenine and thymine (AT) or guanine and cytosine (GC) pairs, $M_{(AT)}$ = moles of adenine plus thymine, $M_{(GC)}$ = moles of guanine plus cytosine. The extinction coefficients of the nucleoside pairs at a neutral pH, taken from the data of Beaven et al. (15) are given in Table I.

By use of the formula, the number of bases which have melted between two temperatures on the melting curve of DNA may be determined. The absorbance at the wavelengths indicated in Table I is measured at two temperatures. The difference between the spectrum at the higher temperature and lower temperature (difference spectrum) is found, and the resulting absorbancies are divided by the extinction of AT at corresponding wavelengths, as indicated in the formula. These values are plotted as the ordinate, and the extinction of the GC pairs divided by the AT pairs at different wavelengths is plotted on the abscissa. A straight line can generally be drawn between the points. The intercept on the ordinate is equal to the concentration of the AT pairs and the slope of the line is equal to the concentration of the GC pairs. The concentrations of the AT and GC pairs are summed, and the per-

TABLE I

MOLAR EXTINCTION COEFFICIENTS AT NEUTRALITY FOR THE BASE PAIRS ADENINE PLUS THYMINE AND GUANINE PLUS CYTOSINE

Wavelength	Extinction A + T	Extinction G + C
nm		
245	12.3 × 10^3	18.4 × 10^3
250	17.0 "	19.6 "
255	21.5 "	20.0 "
260	23.5 "	18.8 "
265	23.2 "	18.4 "
270	19.8 "	18.5 "
275	14.4 "	17.5 "
280	8.8 "	15.0 "

See text for details.

centage of AT determined. This percentage will be referred to as the "relative percentage" of AT to stress the fact that this number is not the per cent AT of the DNA sample. The relative per cent refers only to the AT pairs which melt between two temperatures in the melting range of the DNA.

Points determined by this formula are linear within these wavelengths, as found by others (1, 2, 4). Nonlinearity may occur outside these wavelengths (5). However, if the concentrations of the AT and GC pairs melting at several temperatures over the melting curve of DNA are summed, the total base ratio is correct within 2 to 3 mole per cent.

The method of Hirschman and Felsenfeld (5) for determining base concentration has been compared with the graphical method. The two methods agree within 2 to 4 mole per cent. However, the graphical method consistently underestimates the per cent AT calculated from the spectrum of double-stranded DNA.

An advantage of this graphical formula is that spectral abnormalities are indicated by nonlinearity of points. The nonlinearity can appear as scattering of points, a curved line, or two separate lines, one associated with wavelengths above 260 nm and one with wavelengths below 260 nm. A scattering of points is typically observed when plotting the data from hyperchromic spectra associated with the very beginning and end of melting curves.

Relationship Between Base Composition and the Melting Curve of DNA

The graphical analysis has been applied to the melting curve of double-stranded DNA. Spectra were recorded at different temperatures within the melting curve, and the relative number of moles of AT and GC pairs which melted between two

consecutive temperatures was calculated. In all cases the changes in relative per cent AT (or GC) melting appears to be a linear function of temperature.

In Fig. 1 are presented the melting curves, and immediately above each of them, the relative per cent AT melting at different temperatures for *Clostridium perfringens* DNA (73.5% AT [14]), mouse DNA (57% AT), and *E. coli* DNA (50% AT[14]) The DNA samples were dissolved in 0.03 M NaCl—0.003 M sodium citrate.

The melting curve for each DNA sample contains a portion which appears linear. From other published melting curves (14), this is a fairly consistent characteristic of DNA. The linear portion of the melting curves corresponds to an interval on the relative per cent AT curve. In *Clostridium* DNA, the relative per cent AT at the start of the linear part of the melting curve is about 90%, and at the end of the linear part, it is about 40% AT. In mouse DNA, the linear part covers a range from about 75 to 25% AT, and in *E. coli* the range is from 55 to 5% AT. It appears that the relative per cent AT melting at the start of the linear portion of the melting curve

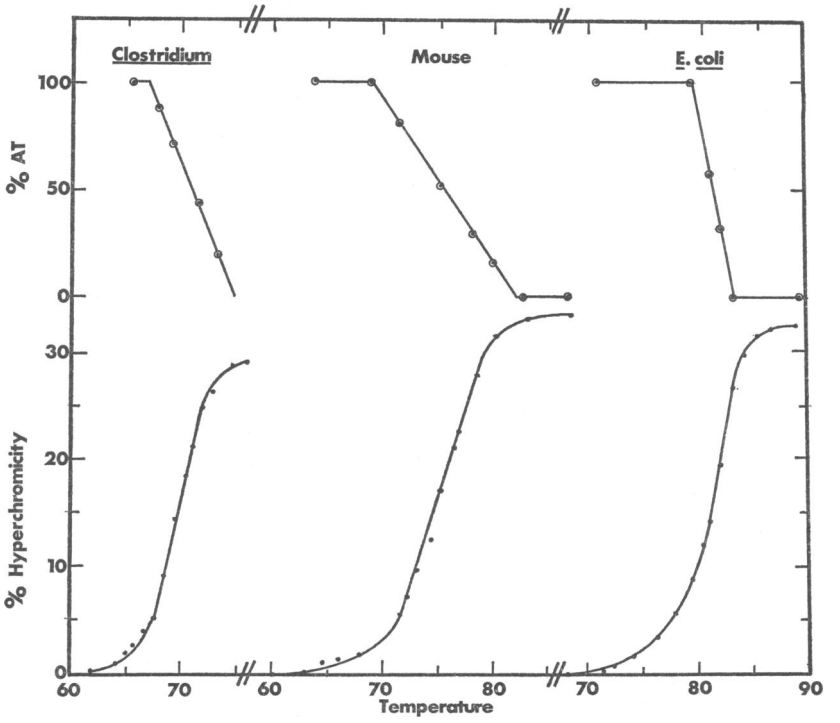

FIGURE 1 The relative per cent AT of sequences which melt during the thermal dissociation of DNA. Solvent: 0.03 M NaCl—0.003 M sodium citrate. The abscissa represents the overlapping temperature scales for dissociation of the three different DNA samples. Lower curves represent the per cent hyperchromic shift of the DNA at 260 nm. The upper curves correspond to the lower curves, but show the relative per cent AT of sequences which cause the hyperchromic shift. See the text for further details.

is proportional to the per cent AT of the DNA sample. The change in relative per cent AT melting over the linear portion of these curves is about 50%. This would mean that in the linear portion of the melting curve the amount of AT pairs melting equals the amount of GC pairs melting.

In the three DNA samples, the linear portion represents a different per cent of the total melting curve. As the linear portion also covers a range of per cent AT melting, the average of this range will also represent a different per cent of these melting curves. For instance, in *Clostridium* DNA, the linear portion of the melting curve represents 60% of the total melting curve. The average relative per cent AT melting in this region is 65%. Therefore, 60% of *Clostridium* DNA has melting sequences which average 65% AT. Similar calculations can be made for the nonlinear ends of the melting curve, although values for the average relative per cent AT are less accurate in these areas. These calculations estimate the relative concentration of sequences with a given base ratio. This represents a kind of frequency distribution. That is, in *Clostridium* DNA, sequences with an average composition of 65% AT would have a frequency of 0.6. We have used similar calculations to arrive at the frequency distribution of the different AT sequences in mouse and *E. coli* DNA. The numerical values are given in Table II. These values are shown in graphical form in Fig. 2. The average per cent AT melting in one of the portions of the melting curve is plotted on the abscissa, and its per cent of the total DNA is plotted on the

TABLE II

THE FREQUENCY AND AVERAGE COMPOSITION OF SEQUENCES WHICH MELT IN THE INITIAL CURVED SECTION, LINEAR MIDSECTION, AND TERMINAL CURVED SECTION OF THE MELTING CURVES OF THREE DNA SAMPLES

DNA sample and base ratio	Section of melting curve	Average per cent AT	Frequency
Clostridium (73.5% AT)	initial	94%	0.17
	mid	65%	0.69
	terminal	19%	0.14
Mouse (58% AT)	initial	87.5%	0.21
	mid	48%	0.70
	terminal	12.5%	0.09
E. coli (50% AT)	initial	79%	0.46
	mid	30%	0.35
	terminal	00	0.19

See text for explanation.

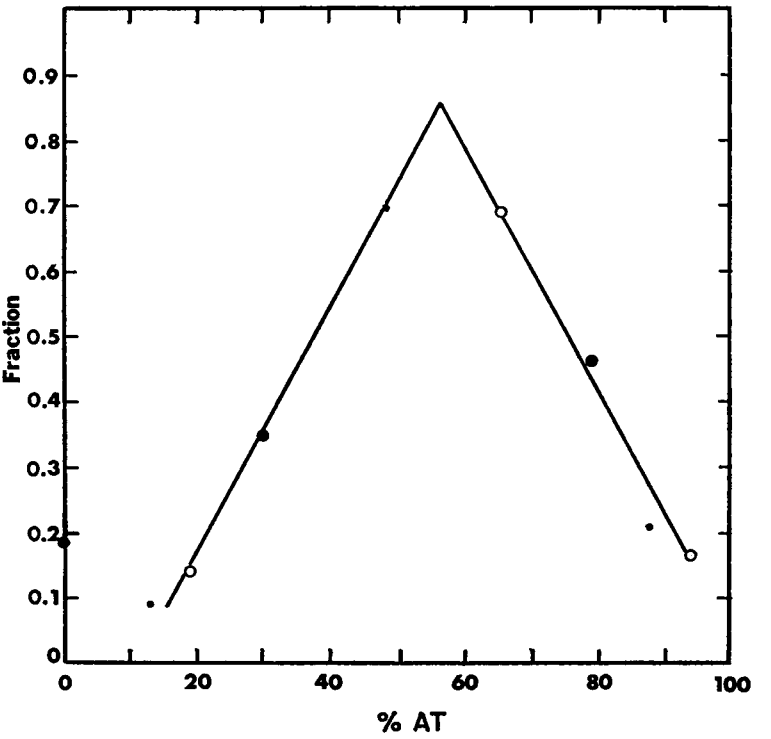

FIGURE 2 The frequencey distribution curve for sequences of different base ratio in DNA from three organisms. Abscissa: the average of the relative per cent AT which melts in the initial, mid, and terminal sections of the melting curve. Ordinate: the fractional concentration of the three sections of the melting curve. The points represent portions of the melting curves of *Clostridium* DNA (open circles), *E. coli* DNA (closed circles), and mouse DNA (small dots). See the text for further details.

ordinate. It can be seen that all but one of the values appear to lie on two lines which intersect at 56% AT. This value is of interest since we calculate that a DNA sequence containing all the codons for 21 amino acids could have a base ratio in this range.

Distribution of GC-Rich Sequences on the E. coli Chromosome

According to the data presented in Fig. 1, 14% of the melting curve of the DNA of *E. coli* has a difference spectrum which corresponds to pure GC. An experiment was performed to determine how the GC-rich sequences are distributed on the chromosome.

E. coli DNA in 0.15 M phosphate buffer, pH 6.8, was sheared to a molecular weight of approximately 5×10^5. The melting characteristics of the DNA were determined. At 88°C, 79% of the DNA was denatured. At 90°C, 94% of the DNA was denatured. According to the above calculations, the unmelted regions should be rich in GC sequences.

TABLE III
CHARACTERISTICS OF DNA ELUTED FROM HYDROXYLAPATITE BY 0.9 M BUFFER

Column temperature	Characteristics of total fraction		Characteristics of double-stranded component	
	Total DNA	Base ratio	Double stranded	base ratio
88°C	69%	52% AT	20%	unknown
90°C	35%	46% AT	18%	18% AT

See text for details.

In separate experiments, the DNA was brought to 88°C and 90°C for 10 min, and placed on hydroxylapatite columns held at these temperatures. Fractions were collected at 0.15, 0.3, and 0.9 M phosphate buffer, pH 6.8. The fractions were cooled and remelted in the spectrophotometer. Spectra were recorded at different temperatures, including the temperature of maximum denaturation (5). The base ratio was determined by the graphical method, and by the method of Hirschman and Felsenfeld (5). The two methods give similar results although the AT content as determined by the graphical formula was consistently slightly lower than that determined by the method of Hirschman and Felsenfeld.

The 0.15 and 0.3 M fractions were single-stranded. The 0.9 M fraction was partially double-stranded. The characteristics of this fraction are summarized in Table III.

In the 88°C column, 69% of the sheared DNA was held by regions 20% double-stranded. We were unable to measure the base ratio of the double-stranded regions, but the base ratio of the total fraction was 52% AT. As this DNA had been sheared to a length corresponding to about 1000 base pairs, it is concluded that double-stranded sequences occur nearly every thousand base pairs along the chromosome.

35% of the DNA applied to the 90°C column was eluted at 0.9 M. 18% of the DNA was double-stranded. The total per cent AT of this fraction was 46%, while the double-stranded component was 82% GC. However, we had calculated that the unmelted component should be 100% GC. There is evidence that there had been some reversible renaturation (17) of the partially melted strands upon cooling, since they melted over a temperature span of 11°C. When the original melting curve of the DNA was measured, the fraction of the DNA which had not melted at 90°C melted completely over a range of 3°C.

Composition of Melting Sequences in DNA-Protein

It would be of interest to apply the graphical formula to the study of chromatin. However, due to light scattering, it is not possible to measure the base ratio of DNA

containing bound protein from the spectrum of the complex. If the spectrum is plotted according to formula 1, a line results with little scatter of points, but the slope is steep, and the intercept is negative.

Light scattering is an exponential function of wavelength (18). The relationship derived by Rayleigh is $A = k/\lambda^a$, where A is the apparent absorption due to scatter, k is a constant, and a a number which lies between two and four in the case of DNA-protein (6). Leach and Scheraga (19) have successfully used this relationship to correct the spectra of protein for light scatter. The correction involves measuring the absorption due to light scatter at wavelengths above 300 nm where protein does not absorb. The absorbancies are plotted vs. wavelength on log-log paper. A straight line results which is extrapolated to wavelengths below 300 nm. However, Olins et al. (6) have shown that in the region of absorption of DNA, light scattering appears to have a maximum around 280 nm. Consequently, the Rayleigh relationship cannot be used directly to correct the spectrum of DNA-protein for scatter.

We obtained spectra of DNA-polylysine which had been complexed according to the method of Olins et al. (6). As found by these authors, the complex exhibits strong light scattering that goes through a maximum around 280 nm. We measured the light scattering of these samples from 340 to 400 nm, and extrapolated the scatter to lower wavelengths on a log-log plot (6, 19). We found that the actual scatter of the sample appears to be a function of the extrapolated scatter multiplied by the rate of change of absorption of DNA with wavelength. The formula is as follows:

$$A = cD_\lambda + kSe_\lambda(cD_\lambda - cD_{\lambda-5\text{nm}}) \qquad (2)$$

where λ = the wavelength in nanometers, A = the measured absorbance at wavelength λ, D = the absorbance of a standard concentration of DNA at wavelength λ, $D_{\lambda+5\text{nm}}$ = the absorbance of DNA at a wavelength 5 nm higher, Se = the extrapolated scatter, determined according to Leach and Scheraga (19), c = concentration of DNA relative to the standard concentration.

If both sides are divided by cD_λ, the formula assumes a form suitable for plotting:

$$A_\lambda/cD_\lambda = 1 + kSe_\lambda(1 - D_{\lambda+5\text{nm}})/D_\lambda \qquad (3)$$

A plot according to this formula of spectra of DNA-polylysine complexes containing 0.7 and 1.16 equivalents of polylysine per DNA nucleotide is given in Fig. 3. As the factor $(1 - D_{\lambda+5\text{nm}}/D_\lambda)$ assumes negative values below 260 nm, the graph has both a positive and negative abscissa. The formula fits the positive values above 260 quite well. The intercept is at 1 as required by the formula, and the graph is linear. There is a slope change at 280 nm. The plot is less satisfactory below 260 nm. The slope is near zero, and the intercept is greater than 1. Below 245 nm, the line becomes scattered. This may be due to the absorption spectrum of polylysine which starts near 240 nm.

The absorption values we have used for DNA, and the scatter correction terms

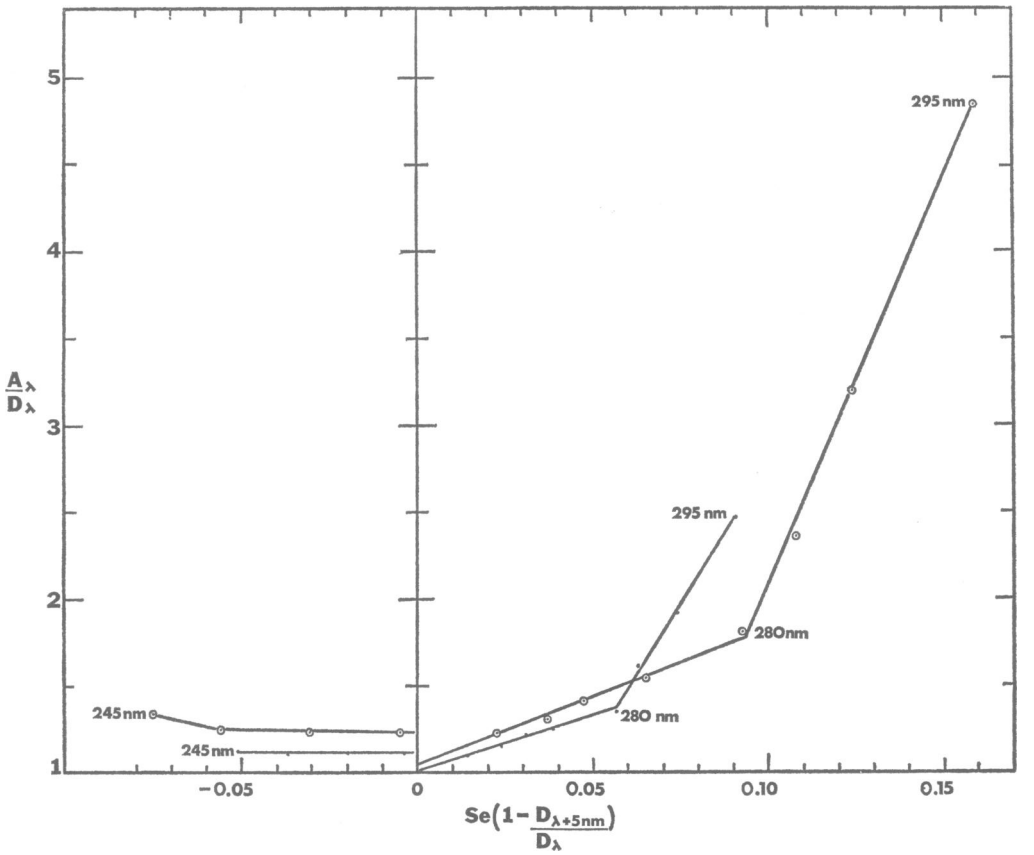

FIGURE 3 The graph illustrates the relationship between the light scattering of solutions of DNA polylysine and the absorption spectrum of DNA, according to formula 3 of the text. The open circles represent a complex of DNA-polylysine which contains 1.16 equivalents of lysine/DNA phosphorous, and the dots, a complex containing 0.7 equivalents of lysine/DNA phosphorous. The absorption spectrum of the complex divided by the calculated absorption of DNA is plotted on the ordinate. On the abscissa is plotted the theoretical scattering of the complex in the ultraviolet, *Se*, times the rate of change of absorption of DNA at intervals of 5 nm. The change in absorption is positive from 295 to 260 nm, and negative from 260 to 245 nm, which is reflected in positive and negative values on the abscissa. In order to identify these wavelengths, the wavelengths of three points are given. See the text for further discussion.

are given in Table IV. The scatter correction terms at wavelengths above 280 nm have been corrected so that a graphical plot of the data is linear between 260 and 295 nm.

The absorbancies for DNA given in Table IV have been normalized to 1.000 at 260 nm. According to our measurements of DNA concentration, this would represent 0.082 mg DNA/ml. This corresponds to 12.2 OD units/mg which is slightly low for DNA. If this value is used, the intercept of the plotted line is at 1. The inter-

TABLE IV

SCATTER CORRECTION TERMS FOR DNA-PROTEIN. VALUES FOR THE ABSORBANCE OF DNA AND RATE OF CHANGE OF ABSORBANCE OF DNA WITH WAVELENGTH IN THE ULTRAVIOLET

Wavelength	Absorbance of 0.082 mg of DNA	Correction factor $1 - (D_{\lambda+5nm})/D_{\lambda}$
nm		
295	0.120	1.547*
290	0.225	0.943*
285	0.369	0.555*
280	0.546	0.333*
275	0.690	0.209
270	0.819	0.157
265	0.931	0.121
260	1.000	0.069
255	0.979	−0.021
250	0.899	−0.089
245	0.754	−0.191
240	0.602	−0.253
235	0.513	−0.172
230	0.457	−0.123
225	0.514	0.109
220	0.706	0.273

See text for details.

* Corrected values to compensate for slope change.

cept is of interest because if the measured DNA concentration, c_{measured}, is different from the actual concentration, c_{actual}, the intercept (and the slope of the line) will deviate by the ratio $c_{\text{actual}}/c_{\text{measured}}$.

Above 260 nm, the slope of the line, which is equal to the constant k in the formula, appears to be a function of the ratio of DNA to protein in the sample. The evidence for this is as follows: the slope k was determined for three samples with different amounts of polylysine to DNA, and one sample of chromatin. These different values of k were multiplied by the ratio of DNA to protein in milligrams. The product was 15.3 ± 0.8, which is essentially a constant.

According to formula 3, the spectrum of DNA may be obtained from the spectrum of chromatin by dividing the measured absorbance of chromatin by

$$1 + kSe(1 - D_{\lambda+5nm}/D_{\lambda}).$$

The absorption of a sample of mouse liver chromatin was measured from 400 to 260 nm. The ratio of DNA to protein was 0.626. Therefore, the slope of the line, k, is 15.3 ÷ 0.626 = 24.4. The extrapolated scatter was determined and multi-

plied by the correction factor given on Table IV and the slope k. The spectrum of DNA was calculated. From the calculated spectrum of DNA, the base ratio was determined by the graphical formula to be 36% AT. This value is low, but the graphical formula underestimates the per cent AT of double-stranded DNA. For instance, the DNA spectrum given on Table IV gives a value of 32% AT by the graphical formula.

In a subsequent experiment this sample was melted, and the spectrum was recorded at the temperature of maximum denaturation. The same value of k was used to correct for the extrapolated scatter. The absorption due to the melted DNA was calculated, and plotted according to the graphical formula for determining the base ratio. The points representing the absorbancy at 260 and 280 nm were slightly high. The other points fell on a straight line which gave a base ratio of 56.5% AT, which is close to the actual base ratio of mouse DNA.

Two samples of chromatin were partially stripped of protein, and a melting curve and spectra were taken, Changes in light scattering occurred throughout the melt. Scatter corrections were made, and the spectrum of DNA reconstructed between the first half and second half of the melting curve.

Points calculated by the graphical formula from the spectrum of the first half of the hyperchromic shift were scattered. Consequently, it was only possible to estimate the per cent AT at 40% ± 10%. Similar calculations of the spectrum of the second half of the hyperchromic shift were more satisfactory, and gave a base ratio

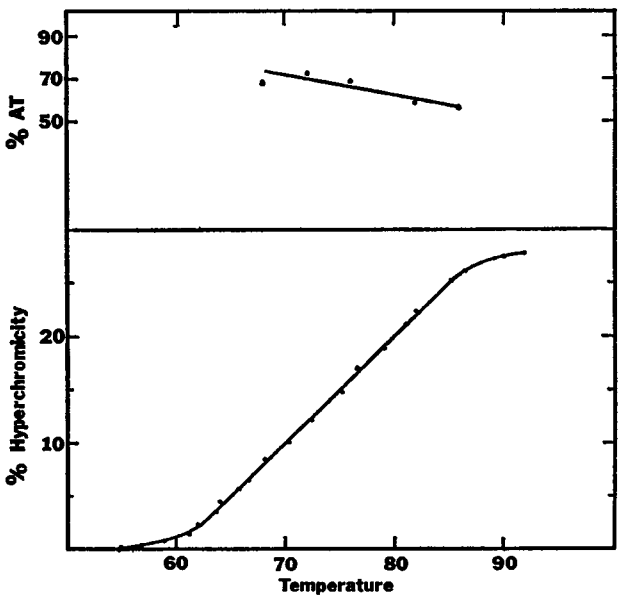

FIGURE 4 The graph, similar to Fig. 1, of the relative per cent AT of sequences which melt during the thermal dissociation of mouse-liver chromatin. The chromatin contained 1.57 mg protein/mg DNA. Solvent: 3.3×10^{-4} M NaCl—3.3×10^{-5} M sodium citrate.

of 20% AT. Consequently, it would appear that AT melts prior to GC. This interpretation was confirmed when a sample of unstripped chromatin was melted. The extent of light scattering did not change with heating, so it was possible to obtain base ratios directly from difference spectra. A plot of the per cent AT at different temperatures and the hyperchromic shift are shown in Fig. 4. AT melts prior to GC under these conditions. However, in chromatin, the temperature stability of AT and GC pairs are much closer together than in pure DNA, since the change in relative per cent AT varies only between 70 and 55%.

As light scattering of DNA-protein appears to be related to the rate of change of absorption of DNA with wavelength, it would appear reasonable that high molecular weight DNA alone might also show a similar light-scattering distortion. We found that if the molecular weight of DNA is decreased by sonication, relative to the absorption of the unsheared DNA, there is a decrease in absorption in range the of 270–300 nm which is maximal at 290 nm. There is also a decrease from 255 to 220 nm. To investigate this further, the molecular weight of DNA was reduced with pancreatic DNase. The hyperchromic shift associated with the hydrolysis was partially corrected by normalizing the absorption of the hydrolyzed DNA to that of the unhydrolyzed DNA at 260 nm. From 265 to 285 nm the absorption of the hydrolyzed DNA is greater than the unhydrolyzed. From 285 nm to 300 nm the absorption of the hydrolyzed DNA was considerably below that of the unhydrolyzed, with a maximum difference at 290 nm, as observed in the sonicated DNA. The hydrolyzed DNA also showed a decrease in absorption from 245 to 225 nm. These differences are as large as 10%. These are the wavelengths where DNA-protein shows a maximal distortion due to light scatter (6) which, according to formula 2, appears to be a function of the rate of change of absorbance of DNA with wavelength. It would therefore appear that the spectrum of DNA contains a similar light-scattering distortion which will vary with molecular weight of the DNA.

DISCUSSION

As DNA melts, the difference spectrum, which is the spectral change between two temperatures, closely approximates the spectrum of the bases which melt between the two temperatures. Consequently, the graphical formula is satisfactory for determining the concentration of AT and GC pairs which have melted. Felsenfeld and Hirschman (4) have shown that the spectral change accompanying the helix to coil transition of DNA is a linear function of the concentration of the bases. Our method, and that of Hirschman and Felsenfeld (5) give, within about 2 mole per cent, the same relative base ratios from difference spectra.

However, the total hyperchromic spectrum which results when DNA is entirely melted cannot be used as successfully to determine the base ratio by the graphical formula. This appears to be due to the contribution from light scattering to the spectrum of the melted DNA. It would also appear that there is a distortion in the spec-

trum of double-stranded DNA from light scattering. Such light scattering could account for slight differences which are occasionally observed in spectra of DNA prepared from the same organism.

The formula for the correction of light scattering, which is based on the spectrum of DNA-polylysine may be applicable to all large particles which absorb in the ultraviolet. However, the formula is only applicable when the particle contains a chromophore. Starch amylopectin scatters light strongly, but does not appear to aggregate with DNA. The spectrum of a mixture of the two is the sum of the spectrum of DNA and amylopectin alone. This is quite different from a mixture of polylysine and DNA (6).

It may only be fortuitous that the light scattering of DNA-protein complexes appears to be related to the rate of change of absorption of DNA. However, as shown in Fig. 3 of the Results, application of formula 3 to the spectrum of DNA-polylysine results in a straight line with three slope changes. The linearity extends from 245 to 295 nm in the case of DNA-polylysine containing 0.7 equivalents of lysine It seems improbable that such agreement between the formula and the spectra of DNA protein is due to chance. Furthermore, it appears that the slope k in the formula is a function of the ratio of DNA to protein. This is reasonable since we would not expect that a complex of protein with a small amount of DNA would show as strong a scatter dependence on the DNA spectrum as a complex with a greater amount of DNA. Finally, the formula results in a straight line plot for chromatin, which indicates that it is applicable to systems other than DNA-polylysine, for which it was derived.

A possible explanation for the apparent dependence of light scattering on the rate of change of absorbance of DNA with wavelength is suggested by Olins et al. (6). These authors point out that in the absorbing region there is an anomalous dispersion of the index of refraction, and therefore of scattering. It is possible that the rate of change of absorbance of DNA may approximate the change of refractive index with wavelength.

The purpose in developing this formula was to enable us to study the melting behavior of base pairs in chromatin and partially stripped chromatin. Unfortunately, the application of the formula to the melting of partially stripped chromatin is at present only approximate. However, by use of the formula, it was found that AT melts prior to GC in partially stripped chromatin, which agrees with our spectral analysis of native chromatin. In chromatin, the thermal stability of AT pairs is close to that of GC pairs. Olins et al. (6) have shown that some of the homopolymer-DNA complexes have a similar melting behavior; that is, AT pairs have nearly the same thermal stability as GC pairs. It is of considerable theoretical interest that native chromatin and the DNA-protein homopolymer should behave similarly. It is of further interest to note that the melting must involve fairly long sequences, since the per cent AT is nearly constant over the total melting curve. Also, the melting curve of chromatin extends over approximately 30°C, so it would appear that

the melting temperature of the sequences is determined by the composition of the protein rather than the base ratio of the sequences.

In DNA free of protein, AT pairs have a thermal stability considerably below that of the GC pairs. Felsenfeld and Sandeen (1) have shown that the difference in thermal stability is approximately 5°C. Our melting curves show that the melting temperature of sequences is determined by the relative per cent AT, which is analagous to the dependence of the average temperature of denaturation (Tm) of DNA on base ratio (13). This also agrees with the findings of Felsenfeld and Sandeen.

The shape of the melting curve is determined by the concentration of the sequences which melt at a given Tm . The linear section of the melting curves would represent the melting of an equal number of sequences from one temperature interval to the next. As mentioned in the Results, within the linear range of the melting curve, the concentration of AT pairs which melt equals the concentration of GC pairs which melt. If the melting sequence is taken to be 100 nucleotides, in *Clostridium* DNA melting groups of equal concentration range from 90 AT + 10 GC to 40 AT + 60 GC. In *E. coli*, the linear portion of the melting curve would represent groups ranging from 55 AT + 45 GC to 5 AT + 95 GC. The composition of these groups reflects the base ratio of the DNA, which is 73% AT in *Clostridium* DNA and 50% AT in *E. coli* DNA. Mouse DNA shows values intermediate to these.

It is not certain how many base pairs comprise a sequence which melts at one temperature. If the sequences are short and involve just a few base pairs, then all the genes of *E. coli*, mouse, and *Clostridium* could have a very different composition. For instance, in the linear part of the melting curve of *Clostridium* DNA the melting sequences might range from 9 AT + 1 GC to 4 AT + 6 GC while those for *E. coli* would range from about 6 AT + 4 GC to 1 AT + 9 GC.

The hydroxylapatite chromatography of *E. coli* DNA shows that GC-rich sequences occur frequently throughout the chromosome. For instance, 70% of the sheared DNA had helical regions at 88°C, a temperature at which the DNA is 79% denatured. The maximum length of a GC-rich sequence at 90°C is 180 base pairs. This number was arrived at by assuming the average length of the sheared DNA to be 1000 base pairs based on a molecular weight of 5×10^5. This DNA was found to be 18% double-stranded, which represents 180 base pairs. Goel and Maitra[1] have calculated the length of a melting sequence to be about 40 base pairs. The figures do not disagree since the 180 pairs melt over a temperature of several degrees, and are therefore not a homogeneous melting unit.

With melting sequences of this length, there is a greater possibility for different organisms to share genes in common, which is a hypothesis that we favor (20). The frequency distribution curve shown in Fig. 2 does not necessarily contradict this possibility. Such a curve could be generated if most melting sequences were

[1] Goel, N. S., and S. C. Maitra. 1969. *J. Theor. Biol.* 22.

composed of two parts. One part would average about 55% AT, and the other part would be high in AT or GC, depending on the base ratio of the DNA.

Our results give some indication of how DNA sequences melt. In order to correlate this data with the arrangement of genes on DNA, it is necessary to construct a model of DNA which takes into account the influence of the Tm of a sequence on its neighbors, as is being developed by Goel, Montroll, Rein, and their co-workers (footnote 1, references 21–24). The application of our data to such a model will be the subject of a later paper.

We would like to thank Dr. Ann R. O'Meara for the sample of mouse-liver chromatin which was used in the study of native chromatin. We would also like to thank Dr. David I. Bachrach for his interest in this work, and his critical discussion of the frequency distribution curve.

Dr. Herrmann is with the Boston University School of Medicine, Mr. Dowling with the Veterans Administration Hospital. Dr. Russell is with both institutions.

Received for publication 2 February 1968 and in revised form 29 November 1968.

REFERENCES

1. FELSENFELD, G., and G. SANDEEN. 1962. *J. Mol. Biol.* **5**:587.
2. FRESCO, J. R., L. C. KLOTZ, and E. G. RICHARDS. 1963. *Cold Spring Harbor Symp. Quant. Biol.* **28**:83.
3. MAHLER, H. R., B. KLINE, and B. D. MEHROTRA. 1964. *J. Mol. Biol.* **9**:801.
4. FELSENFELD, G., and S. Z. HIRSCHMAN. 1965. *J. Mol. Biol.* **13**:407.
5. HIRSCHMAN, S. Z., and G. FELSENFELD. 1966. *J. Mol. Biol.* **16**:347.
6. OLINS, D. E., A. L. OLINS, and P. VON HIPPEL. 1967. *J. Mol. Biol.* **24**:151.
7. OLINS, D. E., A. L. OLINS, and P. VON HIPPEL. 1968. *J. Mol. Biol.* **33**:265.
8. BLOBEL, G., and V. R. POTTER. 1966. *Science.* **154**:1662.
9. COMMERFORD, S. L., W. J. HUNTER, and J. L. ONCLEY. 1963. *J. Mol. Biol.* **238**:2123.
10. MARMUR, J. 1961. *J. Mol. Biol.* **3**:208.
11. BURTON, K. 1956. *Biochem. J.* **62**:315.
12. LOWRY, O. H., N. J. ROSEBROUGH, A. L. FARR, and R. J. RANDALL. 1951. *J. Biol. Chem.* **193**:265.
13. MINDICH, L., and R. D. HOTCHKISS. 1964. *Biochim. Biophys. Acta.* **80**:73.
14. MARMUR, J., and P. DOTY. 1962. *J. Mol. Biol.* **5**:109.
15. BEAVEN, G. H., E. R. HOLIDAY, and E. A. JOHNSON. 1955. *In* The Nucleic Acids. E. Chargaff and J. N. Davidson, editors. Academic Press Inc., New York. Vol. 1.
16. BERNARDI, G. 1965. *Nature.* **206**:779.
17. GEIDUSCHEK, E. P. 1961. *Proc. Nat. Acad. Sci. U.S.A.* **47**:950.
18. TANFORD, C. 1961. Physical Chemistry of Macromolecules. John Wiley and Sons, Inc., New York. 277.
19. LEACH, S. J., and H. A. SCHERAGA. 1960. *J. Amer. Chem. Soc.* **82**:4790.
20. RUSSELL, A. P. 1966. *Comp. Biochem. Physiol.* **18**:961.
21. MONTROLL, E. W., and N. S. GOEL. 1966. *Biopolymers.* **4**:855.
22. GOEL, N. S., and E. W. MONTROLL. 1968. *Biopolymers.* **6**:731.
23. GOEL, N. S., N. FUKUDA, and R. REIN. 1968. *J. Theor. Biol.* **18**:350.
24. REIN, R., N. S. GOEL, N. FUKUDA, M. POLLAK, and P. CLAVERIE. 1969. Conference on Perspectives of π-Interaction in Biological Sciences, March 28, 1967. *Ann. N.Y. Acad. Sci.*