peptides. Glycine, alanine, and arginine also occurred in significant amounts. Glycine is again the predominant amino acid in the fourth position which indicates that a significant portion of the collagen molecule should have the sequence -glycyl-X-Y-glycyl-Z... This agrees with X-ray diffraction findings.

Schrohenloher *et al.* (1959) isolated glycylprolylhydroxyproline containing 23% of the total hydroxyproline from collagenase digest of collagen. This amount is consistent with, but rather lower than, our findings. Low yields are to be expected from their work since it involved several purification steps with accompanying losses. Our findings from Edman degradation enable us to establish a maximum permissible frequency for a given peptide. However part of the 82% of the total hydroxyproline in the third position (reported here) could be in peptides other than glycylprolylhydroxyproline. The glycine at the third position (10% of the total) indicates the existence of -glycylglycyl- sequences in collagen. Kroner *et al.* (1955) reported finding glycylglycine in acid hydrolysates of collagen. The Edman degradation of a mixture of peptides is not in itself conclusive evidence for the existence of any particular peptide but reflects the distribution of amino acids in the peptide mixtures.

Presently accepted interpretations of the X-ray-diffraction patterns and of the "collagen-fold" requirements postulate sequences similar to those established by the present work, but we emphasize that such sequences are not absolutely fulfilled in reality. In particular proline and hydroxyproline do not occupy equivalent positions: most of the proline found is in position two, a position not occupied at all by hydroxyproline whereas most of the hydroxyproline we find in position 3 with little or none in any other position.

The findings have facilitated the examination of peptides separated from the digest of collagen and may serve as a basis for quantitative estimates of particular peptides. It is confirmed that collagen is a highly ordered protein.

References

Acher, R., Laurila, U. R., and Fromagoet, C. (1956), *Biochim. Biophys. Acta 19*, 97.

Dunn, M. S., and Loshakoff, A. (1936), *J. Biol. Chem. 113*, 359.

Edman, P., and Sjöquist, J. (1956), *Acta Chem. Scand. 10*, 1507.

Eriksson, S., and Sjöquist, J. (1960), *Biochim. Biophys. Acta 45*, 290.

Fraenkel-Conrat, H., and Harris, J. I. (1954), *J. Am. Chem. Soc. 76*, 6058.

Gallop, P. M. (1955), *Arch. Biochem. Biophys. 54*, 486.

Gallop, P. M., Seifter, S., and Meilman, E. (1957), *J. Biol. Chem. 227*, 891.

Grassmann, W., Nordwig, A., and Hörmann, N. (1961), *Z. Physiol. Chem. 323*, 48.

Greenstein, J. P., and Winitz, M. (1961), *in* Chemistry of the Amino Acids, Vol. I, Greenstein, J. P., and Winitz, M., eds., New York, Wiley, p. 489.

Konigsberg, W., and Hill, R. J. (1962), *J. Biol. Chem. 237*, 2547.

Kroner, T. D., Tabroff, W., and McGarr, J. J. (1955), *J. Am. Chem. Soc. 77*, 3356.

Levy, A. L. (1954), *Biochim. Biophys. Acta 15*, 589.

Light, A., and Smith, E. L. (1960), *J. Biol. Chem. 235*, 3151.

Mandl, I. (1961), *Advan. Enzymol. 23*, 163.

Margoliash, E. (1962), *J. Biol. Chem. 237*, 2161.

Michaels, S., Gallop, P. M., Seifter, S., and Meilman, E. (1958), *Biochim. Biophys. Acta 29*, 450.

Moore, S., and Stein, W. H. (1948), *J. Biol. Chem. 176*, 367.

Piez, K. A., and Morris, L. (1960), *Anal. Biochem. 1*, 187.

Schrohenloher, R. E., Ogle, J. D., and Logan, M. A. (1959), *J. Biol. Chem. 234*, 58.

Sjöquist, J. (1953), *Acta Chem. Scand. 7*, 447.

Sjöquist, J. (1957), *Arkiv Kemi 11*, 129.

Smyth, D. G., Stein, W. H., and Moore, S. (1963), *J. Biol. Chem. 238*, 227.

Spectral Characterization of Oligonucleotides by Computational Methods

Arnold W. Pratt, J. Nicolet Toal, George W. Rushizky, and Herbert A. Sober

From the Laboratory of Physiology and the Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Md.

Received July 21, 1964; revised October 2, 1964

Computational methods are presented which permit the quantitative determination of base composition of a hydrolyzed oligonucleotide using only spectrophotometric data. The analysis can be done rapidly with a minimal amount of material (absorbance at 260 mμ of 0.5–1.0 in 3 ml) using any selected catalog of pure component spectra, and the linear programming formulation is applicable in the presence of minimal impurities or blanks. The computational methods are suitable for further spectral characterization. These methods have been used on compounds of known base composition and sequence up to the tetranucleotide level; however it is expected that the spectral analysis and characterization of longer-chain compounds will be equally practicable. The computational methods were used to characterize a homologous series of adenine oligoribonucleotides. The entire absorbance function was found to increase linearly with chain length, indicating that the formation of the phosphodiester linkages completely accounted for the absorbance changes seen and that no secondary structure was present up to a chain length of four.

Improved enzymatic hydrolysis and chromatographic separation procedures make it possible to prepare a large number of oligonucleotide fragments of both ribonucleic acid and deoxyribonucleic acids. For example, the two-dimensional "mapping" procedure (Rushizky and Knight, 1960; Rushizky and Sober, 1962a) permits separation and semiquantitative identification of the mono-, di-, tri-, and tetranucleotides obtained from the enzymatic digestions of RNA and DNA. To characterize these oligonucleotides further usually requires separation and quantitative determination of the constituent base composition and the

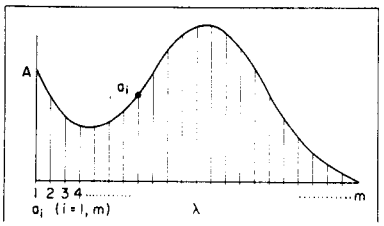

FIG. 1.—A schematic presentation of the ordered set of points (vector $A = a_1, a_2 \ldots a_m$) which comprise an absorbance spectrum.

terminal residues by comparison and correlation of physical and chemical properties of the samples against a library of known oligonucleotides.

It is known that the recognized purine and pyrimidine bases which occur in natural and synthetic polynucleotides absorb ultraviolet light. This physical property of the bases together with the available instrumentation for measuring ultraviolet-absorbance spectra provide a reliable system capable of yielding quantitative data for computational analysis. The current study was devised to explore the applicability of computational methods as adjunct, analytical procedures for the ultraviolet-spectral characterization of the oligonucleotides.

EXPERIMENTAL

Mono- and Oligonucleotides.—These were isolated and identified from alkaline and enzymatic hydrolysates of polyadenylic acid (Singer *et al.*, 1962) of yeast ribonucleic acid (Crestfield *et al.*, 1955) and of tobacco mosaic virus by the "mapping" procedure of Rushizky and Knight (1960; Rushizky and Sober, 1962a). Some of these hydrolysates were separated on DEAE-cellulose columns prior to "mapping" (Rushizky and Sober, 1962b). The compounds thus isolated on Whatman No. 3 paper were located under ultraviolet light, cut out together with appropriate blanks, and stored in this form.

Enzymatic Hydrolysis of Oligonucleotides.—Compounds on paper strips were eluted for 3 hours at 23° with 7–10 ml of deionized water, and centrifuged. The optical density of the supernatant solution at 260 mμ ranged from 0.3 to 1.2. Three ml of this solution was adjusted to *p*H 4.6 with 0.05 M ammonium acetate, and 100 μg of a concentrated ribonuclease T_2 solution (Naoi-Tada *et al.*, 1959) was added so as to obtain 0.02–0.04 unit of enzyme/ml (Rushizky and Sober, 1963). At this dilution, the contribution of enzyme to the absorbance at 280 mμ is less than 0.015. After complete digestion (30 minutes at 23°), the solution was adjusted to *p*H 7 with 0.1 M sodium phosphate, then to *p*H 1 with 6 N HCl, and finally to *p*H 12 with 10 N NaOH. Spectra of the hydrolyzed and intact compounds as well as appropriate blanks were obtained at each *p*H in silica cells of 1 cm light path. Thus a total of six spectra were obtained from each sample. Each spectrum was comprised of 91 observations obtained over a wavelength range of 220–310 mμ at 1-mμ intervals. At least two different samples of each oligonucleotide were examined.

Data Processing.—Spectral data were measured with an automatically scanning/recording Cary Model 14 spectrophotometer. In addition to the pen-drawn strip-chart record of sample absorbance, a numeric record of the same data was obtained simultaneously by linking a voltage source and a voltage-dividing circuit to the pen-drive balance mechanism of the recorder, thereby providing a constantly generated analog voltage which was proportional to sample absorbancy at all times. At appropriate intervals, this voltage was automatically sampled, measured, and properly scaled into numeric absorbance units in an analog-to-digital converter and finally recorded as binary-coded decimal information in an 8-channel paper tape which served as compatible input to both (IBM 1620 and Honeywell 800) data processing systems used.

RESULTS

Composition Analyses.—It is possible in principle to determine the composition of a mixture from its spectrum, providing the spectrum of each pure component is known and providing the spectra are linearly additive. The computation represents an analytical curve-fitting problem in which an unknown mixture of mononucleotides represented by the ultraviolet spectrum of the hydrolyzed oligonucleotide is resolved in terms of a standard catalog of spectra of the pure components; for the ribonucleotide analysis the catalog would contain the spectra of the common monoribonucleotides. The absorbance curves of the pure components together with selected concentration values are used to calculate an absorbance curve which is the best approximation to the absorbance curve of the unknown mixture. In these studies, two computational methods have been employed to solve this multicomponent system: a least-squares method and a linear-programing method (Sternberg *et al.*, 1960; Reid and Pratt, 1960; Pratt *et al.*, 1963; and White *et al.*, 1963).

The ultraviolet-absorption spectra of the common ribonucleotides are smooth, relatively unstructured curves which tend to be highly overlapping. For regression analysis, such data tend to be rather compactly grouped, creating a mathematically ill-conditioned system which can be very sensitive to experimental error and which makes precision of solution very difficult to obtain. For these reasons, any computation made on the multicomponent ribonucleotide system must be based on a large number of absorbance observations spaced at regular intervals over a wide range of wavelengths. Figure 1 shows that an absorbance spectrum obtained in this manner at a single *p*H may be considered as an ordered set of observations which for computational purposes may be represented as a vector, A, composed of m elements, a_i:

$$A = (a_1, a_2, a_3 \ldots\ldots a_m) = a_i \quad (i = 1, m)$$

If measured at the same wavelengths and the same *p*H values, spectra of either pure components or mixtures may be so represented and compared directly if first adjusted for concentration effects. It is our practice to construct extended absorbance vectors consisting of a number of absorbance measurements obtained on the same preparation over the same wavelength range but at three different *p*H values (i.e., a minimum of 84 points—28 points from each of the three 91-point spectra) to provide sufficient information on which to base the computational analysis.

It is difficult to establish concentration values for an unknown oligonucleotide preparation, therefore concentration effects have been eliminated by normalizing the data and expressing composition as percentages. In these studies, spectra were normalized to unit area; the normalized elements, $\hat{a}_i$, of the absorbance vector A, were developed by summing over the m elements of the vector and then dividing each observation by the sum:

$$\hat{a}_i = a_i \Big/ \sum_{i=1}^{m} a_i$$

TABLE I
CALCULATED COMPOSITION[a]

	Least Squares Solutions (%)				
	Ap	Cp	Gp	Up	
Ap/Cp	54.8	46.1	1.2	−2.0	
Ap/Cp	54.5	46.8	0.5	−1.9	
Ap/Gp	47.1	−1.5	54.2	0.1	
Ap/Ap/Gp	63.9	−0.2	36.1	0.2	
Cp/Ap/Gp	34.9	30.7	35.1	−0.9	
pAp/A	101.0	0	−1.9	0.9	
	Linear Programing Solutions (%)				
	Ap	Cp	Gp	Up	Slack
Ap/Cp	50.1	50.1			2.3
Ap/Cp	50.1	50.1			2.0
Ap/Gp	49.9		49.9		1.0
Ap/Ap/Gp	66.2		33.1		0.8
Cp/Ap/Gp	33.3	33.3	33.3		1.5
pAp/A	100.1				1.2

[a] Composition of di- and triribonucleotides. The upper portion was calculated by the least-squares method from their hydrolyzed forms. The lower part was calculated by solution 2 of the linear-programing method. Blank items indicate that these mononucleotides were excluded from the catalog for this solution.

All computations were done using normalized data. Given the absorbance spectrum of each pure compound in a mixture, each adjusted for concentration, it is theoretically possible to calculate the concentrations of each of the compounds in the mixture from the mixture spectrum. If the system satisfies the Beers-Bouguet-Lambert law, that is, at any wavelength, the absorbance of all the components in the mixture is directly proportional to their concentration in the mixture, then the solution (b_i) is a linear combination of the known absorbances and is given by:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (i = 1, m)$$

where the absorbance at unit concentration of the pure jth component at the ith wavelength is a_{ij}, x_j is the concentration of the jth component, and b_i is the absorbance of the mixture at the ith wavelength. The values of n and m specify the number of components in the system and the number of observations (wavelengths), respectively.

In the least-squares method (Bauman, 1959; Opfell and Sage, 1958), the accepted criterion for a "best fit" is that the sum of the squares of the residuals (r) between the observed values and the calculated values over the entire set of observations must be a minimum:

$$r_i = b_i - \sum_{j=1}^{n} a_{ij}x_j \qquad (i = 1, m)$$

where $m \geq n$. The results obtained with the least-squares method for some known hydrolyzed oligonucleotide preparations against a standard catalog of adenylic, cytidylic, guanylic, and uridylic acids are given in Table I.[1] An adequate estimate of mononucleotide composition of each of the oligonucleotides was obtained in every instance; however, the precision of solution, while adequate for the short-chain compounds, could be inadequate for the analysis of longer-chain polynucleotides. Limitations of the method, as applied to the analysis of this multicomponent monoribonucleotide system, should be noted. The method is not applicable if the catalog of pure spectra does not adequately describe the unknown mixture or if the experimental error tends to be large or nonrandom.

An alternative way to approach the calculation recognizes that there is always uncertainty as to the adequacy with which the catalog spectra of pure ribonucleotides can exactly describe the spectrum of an unknown hydrolysate because of experimental error, contamination, the presence of an unknown component, and the like. Any inadequacy of the catalog alters the theoretical set of linear equalities to a set of linear inequalities in which the additive sum of the known components may be less than, equal to, or greater than the absorbance of the hydrolyzed mixture. The existence of an inequality, and the desired restriction that the concentration values for each of the components be nonnegative values are recognized to be the standard constraints of a linear-programing formulation (Dantzig, 1963). Such a system can be solved if the linear inequalities are converted to equations by the addition of new variables known as slack variables (s). The slack variables are additional unknowns which must be determined and provide at every wavelength a measure of residual absorbance which must be added (s_i) or subtracted (s_{m+i}) to satisfy the equations (constraints) in the system. The objective of the linear-programing technique is to obtain an optimal solution of the system by maximizing or minimizing a linear combination of the variables. Thus, an additional constraint called an *objective function* must be included in the system (see Z); the objective function in this application requires that the sum of the absolute values of the slack variables over the entire spectrum be a minimum:

$$\sum_i \left| \sum_j a_{ij}x_j - b_i \right|$$

The linear-programing formulation which solves for the concentration of the individual components and minimizes the absolute sum of the total slack, while allowing either positive or negative slack at any wavelength, is given by:

$$\begin{aligned}
a_{11}x_1 + \ldots + a_{1n}x_n + s_1 \quad & - s_{m+1} & = b_1 \\
a_{21}x_1 + \ldots + a_{2n}x_n \quad + s_2 \quad & - s_{m+2} & = b_2 \\
\cdot \quad \cdot \quad \cdot \quad & \cdot \quad \cdot & \cdot \\
\cdot \quad \cdot \quad \cdot \quad & \cdot \quad \cdot & \cdot \\
\cdot \quad \cdot \quad \cdot \quad & \cdot \quad \cdot & \cdot \\
a_{m1}x_1 + \ldots + a_{mn}x_n \quad + s_m \quad & - s_{2m} & = b_m
\end{aligned}$$

where all the x_j's ≥ 0 and all s_j's ≥ 0. The objective

[1] Abbreviations used in this work; A, C, G, U denote the ribonucleosides adenosine, cytidine, guanosine, and uridine; (p) denotes the phosphoryl group. The 3′ nucleotide of adenosine is shown as Ap and the 5′ nucleotide of adenosine is shown as pA. The residues found after enzymatic hydrolysis are denoted by use of the (/), e.g. Ap/Cp or pAp/A.

TABLE II
LINEAR PROGRAMING SOLUTION OF CP/AP/GP %

Constraint	Slack	Ap	Cp	Gp	Up	Slack
(1) None	±	34.6	30.5	34.8	0	0.8
(2) $x_A = x_C = x_G$	±	33.3	33.3	33.3		1.5

[a] Linear-programing method applied to the solution of a hydrolyzed triribonucleotide (Cp/Ap/Gp). Slack = the absolute sum of the catalog mismatch at every wavelength.

function, Z, is given by:

$$Z = s_1 + s_2 + \ldots + s_m + s_{m+1} + \ldots + s_{2m} = \text{min.}$$

This method allows selection of the "best" k wavelengths ($k = n$) on which to base the computational solution of the system involving k known nucleotides where "best" is in the sense of minimizing the total slack.

Computational analysis of an unknown mixture spectrum by the linear programing method, subject to the single constraint of minimizing the absolute sum of the total slack, yields a percentage concentration for each of the catalog components present in the mixture (x_j's) plus the residual absorbance or slack present at each wavelength:

$$s_i = r_i = b_i - \sum_{j=1}^{n} a_{ij}x_j \qquad (i = i, m)$$

The percentage composition data provide the biochemist with a quantitative estimate of the base composition; the size and distribution of the residual absorbance values provide a quantitative estimate of the purity of the sample. In the ideal system, a single pure oligonucleotide preparation which has been completely hydrolyzed and which has zero "blank" absorbance would, theoretically, be precisely described by the catalog and the ratio of the percentage concentrations would be integers (small whole numbers) and the total slack would be zero. For example, the theoretical percentages for Ap and Gp in the mixture Ap/Ap/Gp are 66.6% and 33.3%, respectively with a relationship between the values of 2:1. From these data, the biochemist would know the base composition and the purity of the sample and could estimate the chain length as 3 or a multiple of 3.

In practice however, a variable amount of "blank" absorbance is inherent to the system. Therefore, an integer solution is rarely obtained if the system is solved subject only to constraint that the total slack be a minimum. By forcing a maximum fit of the catalog components, the percentage concentrations yielded will be distorted as a function of any unexplained absorbance. The equivalent of the "integer solution" in the system which has a modest "blank" absorbance can be obtained by introducing additional system-related constraints in the form of equations. In the constrained (integer) solution the "blank" or impurity absorbance must be seen in the slack (residual absorbance); therefore the total slack will be larger than that obtained for the same system solved without additional constraints (see Table II, solutions 1 and 2). The applicable constraints can be derived from experimental information available, i.e., the additional constraints would be derived from the semiquantitative base composition afforded by the chemical separation procedures employed. As an example, the computational sequence for the hydrolyzed trimer CpApGp is illustrated below.

The spectral analysis of the hydrolyzed sample for base composition, satisfying the condition of minimizing the absolute sum of the total slack did not quantitatively identify the preparation as being the pure trimer (Table II, solution 1). The results suggested, however, that the mixture contained equal percentages of Ap, Cp, and Gp, and that Up was absent from the mixture—a result compatible with the data obtained with the chromatographic procedures. To test whether the concentrations of Ap, Cp, and Gp were equal, the following constraints were added to the formulation:

$$x_A - x_G = 0.000$$
$$x_C - x_G = 0.000$$

and the mixture spectrum was resolved against a catalog containing the spectra of adenylic, cytidylic, and guanylic acids, allowing both positive and negative slack (Table II, solution 2). The result obtained confirmed that, within experimental limits, the mixture consisted of equal parts of Ap, Cp, and Gp; the concentration values gave an acceptable integer solution, and the total slack, while nearly doubled (0.8 versus 1.5), was acceptably small and was distributed as small positive and negative deviations over the entire set of observations. Adding the constraints limited the concentration values to be equal only, and did not otherwise restrain the values. Had the mixture contained an additional absorbing entity, and the like, the concentration values would have been equal but would not have given an integer solution as defined, and/or the total amount of slack would have been unacceptably large. Our criterion for rejection is a total slack value greater than 2.5–3.0%.

Table I gives other typical results computed with the linear-programing method, in each case a "solution 2," obtained in the spectral analysis of a set of known oligonucleotides. It is apparent that the results represent a quantitative identification in every case and, in every instance, are more acceptable than the results obtained with the least-squares method.

Spectral Comparisons.—Correlative techniques are used to examine the spectral similarities among all quantitatively identified oligonucleotide preparations. To examine the similarity of a pair of spectra, the variance of the difference from zero was computed by subtracting the corresponding elements of the two normalized vectors and summing the squares of the differences over the entire set of observations:

$$\text{variance} = \sum_{i=1}^{m} (\hat{a}_i - \hat{c}_i)^2$$

where $\hat{a}_i$ and $\hat{c}_i$ are the ith elements of the normalized elements of the absorbance functions A and C, respectively. This procedure was used to compare the spectra of oligonucleotides before and after hydrolysis, to compare among the spectra of different oligonucleotides, and to compare among replicate samples for estimates of purity. It was considered that two spectral waveforms were indistinguishable if the variance was equal to or less than 1×10^5. This numerical criterion was established empirically by comparing several samples of a known compound prepared according to the complete experimental procedure. Table III contains comparison data on all possible pairs of six sampes of cytidylic acid (*p*H 7). From these results one concludes that the preparations are identical within experimental

TABLE III
REPLICATE SAMPLES OF CYTIDINE-3′-PHOSPHATE

	pH 7					
	S_1	S_2	S_3	S_4	S_5	S_6
S_1	0					
S_2	0.1	0				
S_3	0.3	0.2	0			
S_4	0.6	0.6	0.3	0		
S_5	0.2	0.1	0.1	0.2	0	
S_6	0.6	0.6	0.1	0.2	0.2	0

[a] Replicate samples of cytidine-3′-phosphate mathematically examined by the comparison method. Calculations were made from pH 7 spectra using 91 points. Correlations of 1.0 or less are considered to establish similarity. Scale = variance × 10^5.

TABLE IV
COMPARISON OF OLIGORIBONUCLEOTIDES

	pH 7				
	ApGp	ApCp	pApA	ApCpGp	ApApGp
ApGp	0				
ApCp	74	0			
pApA	60	120	0		
ApCpGp	34	18	130	0	
ApApGp	4	62	43	34	0

[a] Di- and triribonucleotides examined by the comparison method. Calculations were made from pH 7 spectra using 91 points. Correlations of 1.0 or less are considered to establish similarity. Scale = variance × 10^5.

variability. Similar data obtained with some known unhydrolyzed compounds containing adenylic acid are shown in Table IV. According to our experimental criterion (variance $\geq 1 \times 10^5$) none of these compounds can be considered identical by direct spectral comparisons one with another; however the similarity of the spectra of the intact ApGp and ApApGp is apparent, indicating that the correlation procedures alone can have limited capability for distinguishing among spectra but may indicate similarities or relationships.

The computational procedures set forth above were employed in the spectral characterization and identification of a series of known adenine oligoribonucleotides. The series included samples of the mono-, di-, tri-, and tetranucleotides which were of the form pApA, pApApA, and so on, having a 5′-monophosphate-terminal group and 3′,5′-phosphodiester linkages exclusively. The absorbance data of the tetranucleotide before and after enzymatic hydrolysis at the three pH values are given in Figure 2. Certain features of these absorbance data were common to the di- and trinucleotides as well: the spectra at neutral and the spectra at alkaline pH were highly similar to each other but were significantly different from the spectra at acid pH; also, each of the adenine oligonucleotides showed a difference in molar absorptivity between the intact and hydrolyzed forms and in each case the effect was greatest at neutral and alkaline pH.

With reference to the spectrum of adenylic acid, the spectra of the intact dimer, trimer, and tetramer at neutral and alkaline pH showed minimal shifts in the wavelengths of maximum and minimum absorbance, 259–258 mμ and 227–226 mμ, respectively, and the values of the 280 mμ/260 mμ, 250 mμ/260 mμ absorbance ratios progressively increased with increasing chain length. No change in the wavelengths of maximum and minimum absorbance nor change in these ratios was observed at acid pH. Our data agree with that of Singer *et al.* (1962) who have reported spectral properties for a homologous series of intact adenine ribonucleotides ranging in chain length from 2 to 11.

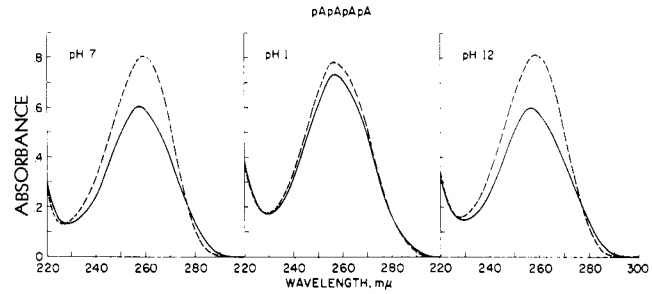

FIG. 2.—Ultraviolet spectra of the tetramer of adenylic acid before and after enzymatic hydrolysis. Solid line denotes spectra of intact tetramer. Broken line denotes spectra of hydrolysis products (pAp/Ap/Ap/A) resulting from incubation at pH 4.6 with ribonuclease T_2.

Enzymatic hydrolysis (ribonuclease T_2) of these adenine oligoribonucleotides yields one residue of adenosine plus one residue of adenosine-3′,5′-diphosphate and, if the chain length is sufficient, one or more residues of adenosine-3′-phosphate, viz:

$$\text{pApApA} \longrightarrow \text{pAp/Ap/A}$$

After enzymatic hydrolysis, all compounds solved (linear-programing method) as adenylic acid exclusively; in each instance the concentration values were effectively 100% adenylic acid and the total slack values were small (Table V). Spectral comparison

TABLE V
LINEAR-PROGRAMING SOLUTIONS

	Ap (%)	Slack (%)
pAp/A	100.2	1.0
pAp/Ap/A	99.2	1.3
pAp/Ap/Ap/A	99.9	1.0

[a] Composition of oligoribonucleotides from polyadenylic acid. Per cent composition was calculated by the linear programing method (solution 2) from their hydrolyzed form.

data (correlative method) supported this finding; the spectra of the hydrolyzed dimer, trimer, and tetramer were indistinguishable from the spectrum of the 3′-phosphate form of adenosine (variance, 1×10^5) (Table VI).

TABLE VI
COMPARISON OF ADENINE OLIGORIBONUCLEOTIDES

	pH 7			
	Ap	pAp/A	pAp/Ap/A	pAp/Ap/Ap/A
Ap	0	<1	1	<1
pApA		12		
pApApA			17	
pApApApA				23

[a] Di-, tri-, and tetraadenylic acids from polyadenylic acid examined by the comparison method. Calculations were made from pH 7 spectra using 91 points. Correlations of 1.0 are considered to establish similarity. Scale = variance × 10^5.

Analysis of adenosine spectra against the spectra of the Ap catalog gave a variance of 1×10^5 and solved as 100.2% Ap with a slack of 1.1%. These results indicate that the spectral differences among these adenosine-containing hydrolysis products are minimal when taken over the entire spectra and the presence or position of the phosphate group is not detectable by the spectral analysis.

TABLE VII
ABSORBANCE RATIOS

	*p*H 7 Wavelength (mμ)					*p*H 1 Wavelength (mμ)					*p*H 12 Wavelength (mμ)				
	240	250	260	270	280	240	250	260	270	280	240	250	260	270	280
pApA	1.11	1.11	1.16	1.13	0.81	1.03	1.02	1.02	1.02	1.00	1.09	1.10	1.15	1.12	0.84
pApApA	1.20	1.21	1.27	1.23	0.78	1.05	1.05	1.05	1.03	1.01	1.22	1.25	1.29	1.24	0.79
pApApApA	1.29	1.28	1.37	1.31	0.76	1.09	1.07	1.06	1.03	1.01	1.32	1.29	1.38	1.36	0.80

[a] Absorbance ratios of di-, tri-, and tetraadenylic acid. Absorbance Ratio = absorbance of hydrolyzed nucleotide/absorbance of intact nucleotide.

The differences in molar absorptivity which accompanied hydrolysis of the adenine oligonucleotides are given in Table VII. The results are expressed as ratios of the absorbance of the compound after hydrolysis to the absorbance of the same compound before hydrolysis (Michelson, 1963) and cover a wavelength range from 240 to 280 mμ at 10-mμ intervals. For each of the oligonucleotides in the series, the magnitude and character of the absorbance changes at neutral and alkaline *p*H were large and nearly identical; the changes at the corresponding wavelengths at acid *p*H were much smaller.

At neutral and alkaline *p*H, the absorbance ratios changed progressively as a function of chain length of the oligonucleotides (Singer *et al.*, 1962). The change in the ratios, however, was not the same at all wavelengths. At 280 mμ the ratios progressively decreased, indicating a decreased absorbance after hydrolysis; at all the other wavelengths recorded the absorbance ratios increase as the chain length increased. The maximum changes were seen with the tetramer. By difference, the largest change in ratio values as a function of chain length occurred with the formation of the dimer. Thereafter, adding another residue of Ap to the dimer and then to the trimer produced essentially a linear change in the absorbance ratios at the wavelengths studied; note, for example, the molar absorptivity ratios at 260 mμ at neutral *p*H and acid *p*H, although at this latter *p*H the ratio changes are small (Table VII). The change at alkaline *p*H is undoubtedly linear; the deviation from linearity is caused by a minimal contamination of the alkaline spectra of the trimer (Table VII, *p*H 12, wavelengths 240–260). The presence of the contaminant was indicated by the linear-programing solution. Note that the concentration value is slightly lower, and the total slack value is slightly higher compared to the other compounds in the analogous series (Table V).

The changes of the wavelength ratios and the absorbance ratios appear to be a linear function of chain length and suggest that the entire absorbance function changes linearly with chain length. Support for this hypothesis is found in the comparison data, which was obtained by comparing the normalized spectra of the same oligonucleotides before and after hydrolysis (Table VI). Clearly, except for the monomer, the absorbance function of a compound before and after hydrolysis was not identical for any of the compounds. Since the waveform of the spectra of each of the hydrolyzed oligonucleotides is that of adenylic acid, it follows that the spectra of the intact oligonucleotides were compared against a common reference. Thus, the increase in variance seen as a function of chainlength becomes significant and shows that the absorbance function does change progressively with increasing chain length, at least up to 4 residues. By difference alone, the largest increase in variance was seen with formation of the dimer and, if considered as a function of adenylic acid residues, the change in variance is a linear function of chain length.

DISCUSSION

From the analysis data, it can be concluded that the series of adenine oligoribonucleotides were of a high order of purity and were completely hydrolyzed by ribonuclease T_2. In each instance the ultraviolet-absorbance spectrum of the hydrolysis product was characteristic of the spectrum of adenylic acid. Furthermore, since the purity of the compounds has been established (i.e., by analysis of the hydrolysate) it follows that the linear change in absorbance ratios at 260 mμ, *p*H 7, can be used to identify compounds having from 2 to 4 adjacent adenylic acid residues.

The spectral-characterization findings show that the absorbance function of the adenine oligoribonucleotides changes linearly with chain length through a chain length of 4, presumably only as a result of the phosphodiester linkages known to be formed. There were no spectral changes which would indicate the presence of any secondary structure.

In this study, the utility of selected computational procedures in the spectral characterization including the quantitative compositional analysis of oligonucleotides has been examined using compounds of known base composition and sequence. The experimental results demonstrate that a quantitative base-composition analysis of the hydrolyzed oligonucleotide can be achieved with computational methods using only spectrophotometric data without recourse to prior physical separation. Once having provided a quantitative identification of a compound by spectral analysis, it is the choice of the investigator as to what spectral features and which of several correlative techniques are most suitable for further spectral characterization. The value of comparing portions or entire spectra of the intact and hydrolyzed compound seems apparent from the experimental results.

The computational methods afford certain advantages: the analyses can be done rapidly, a minimal amount of material is required (absorbance at 260 mμ of 0.5 – 1.0 in 3 ml), the analysis can be done with any selected catalog of pure component spectra and the linear-programing formulation modifies the stringent requirement that there be a virtual absence of "blank" values. So far, these methods have been used only with compounds up to the tetranucleotide level; however, it is expected that the spectral analysis and characterization of longer-chain compounds will be equally practicable.

REFERENCES

Bauman, R. P. (1959), *Appl. Spectry. 13*, 156.
Crestfield, A. M., Smith, K. C., and Allen, F. W. (1955), *J. Biol. Chem. 216*, 185.
Dantzig, G. B. (1963), Linear Programming and Extensions, Princeton, N. J., Princeton University Press.

Michelson, A. M. (1963), The Chemistry of Nucleosides and Nucleotides, London and New York, Academic.
Naoi-Tada, M., Sato-Asano, K., Egami, F. (1959), *J. Biochem. (Tokyo) 46*, 757.
Opfell, J. B., and Sage, B. H. (1958), *Ind. Eng. Chem. 50*, 803.
Pratt, A. W., White, W. C., and Toal, J. N. (1963), Proceedings of the 4th IBM Medical Symposium, October 22–26, 1962, Endicott, N. Y.
Reid, J. C., and Pratt, A. W. (1960), *Biochem. Biophys. Res. Commun. 3*, 337.
Rushizky, G. W., and Knight, C. A. (1960), *Virology 11*, 236.
Rushizky, G. W., and Sober, H. A. (1962a), *J. Biol. Chem. 237*, 834.
Rushizky, G. W., and Sober, H. A. (1962b), *J. Biol. Chem. 237*, 2883.
Rushizky, G. W., and Sober, H. A. (1963), *J. Biol. Chem. 238*, 371.
Singer, M. F., Heppel, L. A., Rushizky, G. W., and Sober, H. A. (1962), *Biochim. Biophys. Acta 61*, 474.
Sternberg, J. C., Stills, H. S., and Schwendeman, R. H. (1960), *Anal. Chem. 32*, 84.
White, W. C., Shapiro, M. B., and Pratt, A. W. (1963), *Commun. Assoc. Computing Machinery 6*, 66.

Studies on the Mechanism of Biological Methylation of Nucleic Acids*

BURTON E. TROPP†, JOHN H. LAW, AND JOHN M. HAYES

From the James Bryant Conant Laboratory, Harvard University, and the Department of Chemistry, Massachusetts Institute of Technology, Cambridge

Received August 10, 1964

Escherichia coli K-12 W-6 cells were grown in the presence of L-methionine-methyl-d_3. The methylated ribosides, thymine riboside and 6-methylaminopurine riboside, isolated from the soluble ribonucleic acid of these cells were analyzed by mass spectrometry and shown in each case to contain three atoms of deuterium. Thus transmethylation both to the 6-amino group of adenosine and to the 5 position of uridine involves the transfer of an intact methyl group.

Among the most interesting of the several new biological reactions discovered in recent years are those which give rise to the methylated bases of nucleic acids. Mandel and Borek (1963) and Fleissner and Borek (1962, 1963) have shown that this enzymatic process involves a transfer of the methyl carbon of *S*-adenosyl-L-methionine to various bases of intact macromolecular ribonucleic acid, especially the type known as soluble RNA (s-RNA). Similar reactions involving deoxyribonucleic acid were discovered by Gold *et al.* (1963). Since the methylation of s-RNA results in the formation of both carbon-carbon bonds and carbon-nitrogen bonds one might expect to find that two different reaction mechanisms were involved. For example, Keller *et al.* (1949) and DuVigneaud *et al.* (1956) found that methylations of nitrogen involved transfer of an intact methyl group with all three protons, while in the carbon-alkylation reactions leading to tuberculostearic acid and ergosterol, Jauréguiberry *et al.* (1964, 1965) have shown that only two of the three protons of the methionine methyl group were transferred to the products. Therefore it was of interest to study a system where alkylation reactions involving the transfer of methyl groups to both nitrogen and carbon were occurring in the same organism.

We have investigated the formation of methylated bases in s-RNA by the use of methionine-methyl-D_3. The results described in this paper show that the methylation of the 5 position of uridine, unlike the methylation of oleic acid or fungal sterols (Jauréguiberry *et al.*, 1964, 1965), involves the transfer of an intact methyl group with three protons, and thus is more closely related to the methylation of nitrogen and other atoms with nonbonding electrons.

* Supported by a research grant from the National Science Foundation (GB 952) and, in part, by a U. S. Public Health Research Career Program Award (8521).

† Predoctoral fellow of the U. S. Public Health Service.

Materials

L-Methionine-methyl-D_3 was prepared by the method of Melville *et al.* (1947), using methyl iodide-D_3 (Volk Radiochemical Co., Chicago, Ill.) and *S*-benzyl-L-homocysteine (Cyclo Chemical Corp., Los Angeles, Calif.). The twice-recrystallized product gave a single spot corresponding to methionine on paper chromatography (Schlenk and DePalma, 1957). Analysis[1] gave 27.30 atom % excess deuterium (100% of theory for $C_5H_8D_3SN$), and mass spectrometry (Biemann, 1962) confirmed that the sample consisted nearly exclusively of the trideuterated species.

L-[methyl-^{14}C]Methionine was purchased from New England Nuclear Corp., Boston, Mass. Adenosine and 6-methylaminopurine were purchased from California Corp. for Biochemical Research, Los Angeles, Calif.; adenine and thymine were obtained from Nutritional Biochemicals Corp., Cleveland, Ohio; and 6-methylaminopurine riboside was the product of Cyclo Chemical Corp., Los Angeles, Calif. Uridine and thymidine were purchased from Schwartz Bioresearch, Inc., Orangeburg, N.Y. Alkaline phosphatase (*E. coli*) and other enzymes were products of Worthington Biochemical Corp., Freehold, N.J.

Experimental Procedure

Escherichia coli, strain K 12 W-6, a methionineless auxotroph, was a gift of Dr. Ernest Borek. The cells were cultured in 40-liter batches in a Biogen (American Sterilizer Co., Erie, Pa.) at 37°, with good aeration. The synthetic medium (Law *et al.*, 1963) was supplemented with 20 mg of L-methionine/liter. Nonisotopic methionine, [methyl-^{14}C]methionine, and methionine-methyl-D_3 were employed for different batches. Near the end of the logarithmic phase of growth the cells

[1] Analysis for deuterium was performed by the falling-drop method by Mr. Josef Nemeth, Urbana, Ill.