Determination of Nucleotide Composition of Polyribonucleotides by Spectrophotometric Analysis*

Wilhelm Guschlbauer,† Edward G. Richards,‡ Karin Beurling, Alice Adams, and Jacques R. Fresco§

ABSTRACT: An analytical method for the determination of ribonucleotides in mixtures has been developed, based on fitting a linear combination of the spectra of the components to that of the sample. High accuracy is achieved by utilizing spectra of solutions before and after near-ultraviolet irradiation, which serves to better differentiate the otherwise similar spectra of some nucleotides. The method, which is readily applicable to 10–15 μg of sample, was tested on known nucleotide mixtures and simulated alkaline hydrolysates of ribonucleic acid (RNA), for which it gave computed mole fractions generally accurate to 0.01 or better. When employed for RNA analysis, in conjunction with an alkaline hydrolysis procedure, the method gave compositional data in reasonable agreement with literature values, but with indications of slightly lower uridine-5′-phosphate (UMP) and higher cytidine-5′-phosphate (CMP) values, and slightly higher guanosine-5′-phosphate (GMP) values. The former differences are attributed to deamination of CMP, which was found to occur under the hydrolytic conditions in general usage. In the present method, such an artifact can be accurately compensated, but not so in the conventional analytical methods. The latter difference is presumed to be due to losses of GMP that can occur in conventional procedures, but not in this method. The experimental procedure and mathematical approach are described in detail. Particular attention is paid to the influence of experimental errors on the computed results, and to criteria for evaluating the validity of results.

Accurate knowledge of the base composition of nucleic acids and polynucleotides and their partial hydrolysis products is important in many investigations of the structure and properties of these macromolecules. The methods hitherto employed for this purpose involve chromatographic or electrophoretic separation of the constituent bases or nucleotides, followed by their spectrophotometric determination (Cohn, 1955; Smith, 1955; Wyatt, 1955). These procedures are tedious, require relatively large amounts of material, and the recovery of the separated nucleotides is often incomplete.

Multicomponent analysis using least-squares curve-fitting methods is a widely employed statistical technique. This type of analysis has been applied to the resolution of spectra of complex mixtures. In particular, mixtures of steroids have been successfully analyzed using this approach by exploiting the highly differentiated character of their infrared spectra (Rogoff, 1957). A similar analysis was used, with moderate success, to resolve the ultraviolet spectra of steroid mixtures (Sternberg *et al.*, 1960). The intense and characteristic ultraviolet absorption of purines and pyrimidines suggests that multicomponent analysis might be applied to mixtures of nucleic acid components as well. If such an approach could be made experimentally feasible, a method would result having the distinct advantages of simplicity, rapidity, and very high sensitivity.

Kerr and Seraidarian (1945) and Steiner (1961) have exploited the differences in spectral characteristics of nucleotides at various *p*H values and selected wavelengths to determine an *approximate* base composition of nucleotide mixtures (see also Loring, 1955). Reid and Pratt (1960) suggested that the base composition of nucleic acids might be determined by resolution of the ultraviolet spectra of their hydrolysates. These workers utilized absorbances at many wavelengths and employed the least-squares fitting method, solved by an electronic computer, to analyze known nucleotide mixtures. A similar, but simpler, approach was reported by Vasilenko *et al.* (1962). However, because these workers restricted their analysis to only nine wavelengths, their method is extremely sensitive to small experimental errors.

Although the work of Reid and Pratt (1960) suggests that multicomponent spectral analysis of nucleotide mixtures is theoretically possible, we found, in attempting to repeat their results, that large errors arise from two sources. The major source of error is the great

* From the Department of Chemistry, Princeton University, Princeton, N.J. *Received December 28, 1964.* This investigation was supported by grants from the National Science Foundation, the U.S. Public Health Service, and the American Heart Association. The facilities of the Princeton University Computer Center were supported in part by the National Science Foundation (Gp 579).

† Present address: Institut de Biologie Physico-Chimique, Paris 5^{e}.

‡ Present address: M.R.C. Biophysics Research Unit, Kings College, London.

§ To whom inquiries should be sent. Established Investigator of the American Heart Association (1958–1963).

similarity between the AMP[1] and UMP spectra at neutral *p*H, so that small experimental errors lead to a confusion of these components. The other is the presence of ultraviolet absorbing impurities in the mixture.

Both these difficulties have been overcome in the present work. By selectively altering the spectra of the pyrimidines, using intense ultraviolet radiation, the confusion of the AMP and UMP spectra is avoided. Moreover, by employing computational tests to detect ultraviolet absorbing impurities, including unexpected nucleotides, errors due to such contaminants can be properly evaluated or rejected. With these modifications, it has been possible to obtain satisfactory analyses of known nucleotide mixtures, and of hydrolysates of polyribonucleotides and RNA's on as little as 10 μg of sample.

Principle of the Method

The present method is, in principle, similar to that described by Reid and Pratt (1960), except it makes use not only of the spectrum of the polynucleotide hydrolysate in neutral solution, but also of the spectrum obtained after irradiation with ultraviolet light.

After purification to remove ultraviolet absorbing impurities, the polynucleotide sample is hydrolyzed with alkali to yield quantitatively a mixture of the nucleotide components. The hydrolysate is brought to neutrality and absorption spectra are determined before and after irradiation with ultraviolet light, using a recording spectrophotometer. Standard solutions of the individual nucleotides are treated in the same way. The set of molar extinction coefficients so obtained for these standard solutions comprises the *library spectra*.

The best linear combination of the library spectra of the components which fits the spectrum of the hydrolysate is calculated employing a least-squares criterion. Spectra of both the unirradiated and irradiated hydrolysates are utilized. The calculations are performed with the aid of a digital computer.

Several critical assumptions are involved in such an analysis: (1) Each component nucleotide in the mixture obeys Beer-Lambert's law. (2) The additivity law holds; i.e., the absorbance of the mixture is the sum of the absorbances of the components. (3) The absorption spectrum of UMP is altered quantitatively and reproducibly by the irradiation procedure. It should be noted that this assumption is not required for CMP (*vide infra*). (4) The only ultraviolet absorbing components in the mixture are those included in the library.

[1] Abbreviations used: AMP, GMP, UMP, CMP = adenylic, guanylic, uridylic, and cytidylic acids; A, G, U, C symbolize the foregoing nucleotides in the tables; poly-AU and poly-GU are copolymers containing the indicated nucleotides, the synthesis of which was catalyzed by polynucleotide phosphorylase; TMV = tobacco mosaic virus; TYMV = turnip yellow mosaic virus.

Experimental

Analytical Procedure

Library Spectra. Paper chromatography on Whatman No. 3MM paper of commercial samples of the 2′(3′)-mononucleotides in 2-propanol–1% aqueous ammonium sulfate (2:1, v/v) showed that, in general, they contain significant amounts of ultraviolet absorbing impurities. Therefore, the 2′(3′)-mononucleotide preparations were chromatographed on paper, eluted with water, and were, together with appropriate blanks, subjected to the standard hydrolysis procedure (*vide infra*), and their spectra were recorded. Absorbance ratios obtained from these spectra are given in Table I.

TABLE I: Absorbance Ratios of 2′(3′)-Ribonucleotides at Selected Wavelengths.[a]

Nucleotide	A_{240}/A_{260}	A_{250}/A_{260}	A_{270}/A_{260}	A_{280}/A_{260}
AMP: obs	0.40	0.79	0.66	0.14
lit[b]		0.80		0.15
GMP: obs	0.82	1.16	0.83	0.68
lit[b]		1.15		0.68
UMP: obs	0.38	0.76	0.81	0.32
lit[b]		0.78[c]		0.30
		0.73		0.35
CMP: obs	0.96	0.87	1.18	0.88
lit[b]		0.90[c]		0.85
		0.86		0.93

[a] The solvent is 0.05 M phosphate (Na^+), *p*H 7. [b] Volkin and Cohn, 1954. [c] The first value is for the 2′-nucleotide, the second for the 3′-nucleotide.

They are in good agreement with those reported by Volkin and Cohn (1954). The molar extinction coefficients at 260 mμ given by these workers were assumed. The use of these generally employed values allows the results of this work to be compared with the analyses of previous workers.

Polymer Sample Preparation. In our experience, most polynucleotide materials, even when purified by conventional procedures, require the removal of contaminants that distort their hydrolysate spectra. Consequently, after deproteinization with phenol (Schramm and Gierer, 1957), polynucleotides are routinely dialyzed against the following solvents (in order): 8 M urea, 0.01 M Versene (to remove metal contaminants), 0.2 M NaCl, and finally, five changes of distilled water.

Hydrolysis of Polymers. The hydrolysis procedure is that generally used for the preparation of hydrolysates for subsequent chromatographic or electrophoretic separation (Davidson and Smellie, 1952). The reference library components (*vide supra*), known nucleotide

mixtures, and polynucleotide samples are incubated at 37° for 24 hours in 0.3 N KOH. These conditions are sufficient to hydrolyze quantitatively even those polynucleotides containing large amounts of guanylic acid residues (see Results and Table VI), which have been reported to be relatively resistant to digestion by alkali (Lane and Butler, 1959; Mii and Warner, 1960).

Hydrolyses are carried out in polyethylene or polypropylene tubes which have been incubated previously for 48 hours in 1.0 N KOH, in order to remove any ultraviolet absorbing components. (Alkali-resistant glass tubes are unsatisfactory because they contribute ultraviolet absorbing material.) After incubation, an equivalent amount of $HClO_4$ is added, and the mixtures are shaken, chilled, and centrifuged in the cold. The neutralized supernatant is decanted and allowed to attain room temperature. Since $KClO_4$ is virtually insoluble, the salt content is kept to a minimum. Davidson and Smellie (1952) have reported that no significant coprecipitation of nucleotides occurs in this step; we have confirmed this observation.

To aliquots of the hydrolysate is added a standard amount of phosphate buffer (0.2 M, *p*H 7.4) to attain a *p*H of 7.0–7.4 and a concentration of 0.05 M phosphate. The solutions are placed in stoppered quartz cuvets, and a spectrum of each solution is recorded on a Cary 14 spectrophotometer against a blank that has also been carried through the entire procedure. The amounts of polynucleotide used are such that the absorbance never exceeds 1.0 in the wavelength range recorded, 238–284 mμ.

For reasons discussed later, it is preferable to record the spectra at about 4°.[2] To this end the sample and reference chamber of the spectrophotometer are thermostated at this temperature; misting of the cuvets is prevented by continuously flushing the chambers with dry, filtered air.

Irradiation Procedure. After the spectrum is obtained, the hydrolysate, still in the same cuvet, is irradiated at room temperature for 150 minutes with a short-wave ultraviolet lamp (Mineralight, Model SL 2537) or for 75 minutes with one such lamp facing each cuvet window. The cuvets are placed on a magnetic stirrer exactly 1.5 cm from the lamp filter and the solutions are stirred with small Teflon-covered magnets. After irradiation the absorption spectrum is recorded, again at 4°.

Selection of Wavelengths. In selecting the most suitable wavelengths at which to obtain absorbance data, several factors were considered: (1) In general, contaminants absorb significantly below 235 mμ. Consequently, absorbance data are not obtained at shorter wavelengths. (2) Wavelengths at which *all* four nucleotides have low absorbance do not serve to differentiate them. Therefore data are not collected at wavelengths above 284 mμ. (3) Sternberg *et al.* (1960) have shown that it is preferable to use absorbance data at wavelengths distributed uniformly over the range not excluded by the first two considerations, instead of picking unequally spaced wavelengths that are differentially sensitive to each of the components in the mixture.

The absorbances of both the unirradiated and irradiated solutions are read off at 2-mμ intervals in the range 238–284 mμ, corrected for solvent blank, and punched on IBM cards. It would be possible to eliminate this step and the possibility of reading errors by employing a digital recorder with punched-card output, but this has not been done in the present work.

Thus data are collected from each spectrum at twenty-four wavelengths, giving rise to forty-eight absorbance measurements which are utilized in the computer analysis.

Precautions. To obtain reproducible results, several precautions are necessary: (1) The glassware used throughout must be extremely clean and free of ultraviolet absorbing contaminants. Consequently detergents cannot be used; instead, all glassware is cleaned in concentrated H_2SO_4 and rinsed with large quantities of distilled water. (2) Polynucleotide samples must be freed of ultraviolet absorbing contaminants. In particular, traces of protein and phenol and Versene must be removed. All reagents must be of spectroscopic grade or highest purity. (3) Because dust and tiny air bubbles significantly distort spectra below 250 mμ, owing to light scattering, it is advisable to filter hydrolysates through a Millipore filter (Type VM, pore size 50 mμ) after the neutralization step. No loss of nucleotides occurs in this step. (4) The library and sample spectra must be determined on the identical wavelength scale and with the same photometric accuracy. Therefore wavelength and photometric calibrations of the spectrophotometer must be standardized at regular intervals. (5) Even though the efficiency of different Mineralight lamps was found to be very similar, it is advisable to check periodically that the irradiation dose is sufficient to complete the UMP photoreaction.

Basis for Irradiation Procedure

When dilute solutions of UMP and CMP are irradiated with near-ultraviolet light having a maximum intensity at 2537 A, hydration across the 5,6 double bond of the pyrimidine ring is thought to be the primary reaction (Sinsheimer, 1954, 1957). The reversal of this photochemical reaction occurs readily at elevated temperatures, considerably faster for CMP than for UMP. Indeed, at room temperature and neutral *p*H, using the irradiation source employed in the analytical method, the reverse reaction proceeds for UMP at a negligible rate, whereas for CMP it proceeds at a rate comparable with the forward reaction. Thus for a given radiation intensity a sufficiently large dose converts UMP quantitatively to its stable photoproduct; but for CMP only an equilibrium mixture of the unchanged nucleotide and its photoproduct results. The position of this equilibrium can be shifted toward completion either by increasing the radiation intensity,

[2] Errors increased when spectra were recorded at room temperature.

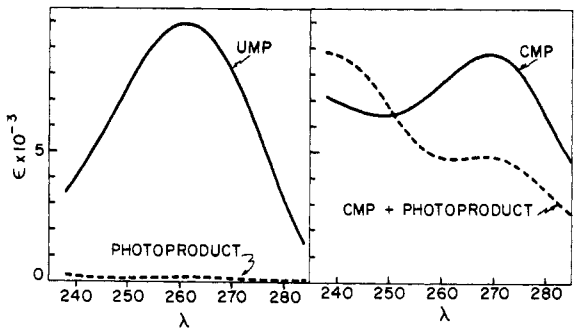

FIGURE 1: Ultraviolet absorption spectra of pyrimidine nucleotides in 0.05 M phosphate (Na^+) buffer, *p*H 7, before and after ultraviolet irradiation (see text) for about 150 minutes.

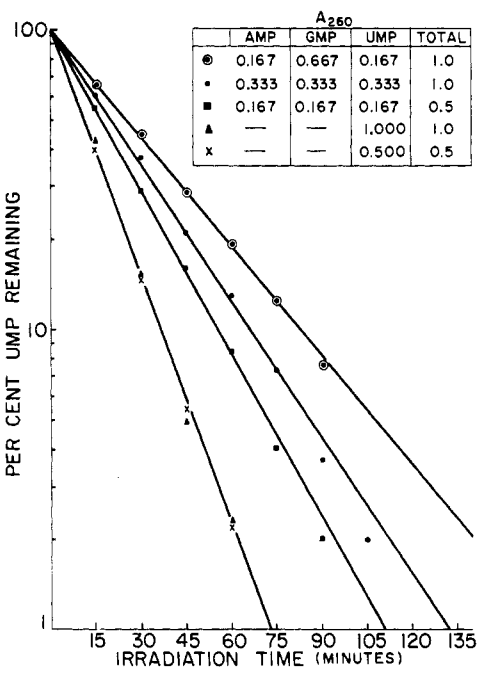

FIGURE 2: Kinetics of ultraviolet photolysis of 2′(3′)-UMP alone and in nucleotide mixtures at different total absorbances. CMP was excluded to simplify the calculations (see text). Per cent remaining UMP was determined at 260 mμ as $[A_{t_{1,2,\ldots(n-1)}} - A_{t_n}]/A_{t_0} - A_{t_n}$, where A_{t_0}, $A_{t_1} \ldots A_{t_n}$ indicate absorbances at times 0, 1, . . . *n*. A_{t_n} was defined as having been reached when an additional 10 minutes of irradiation caused no further absorbance change. When two lamps are used for irradiation (see text) the rate of photolysis was found to double.

or decreasing the rate of the back reaction. The latter can be accomplished by lowering the temperature.

Besides these primary photochemical processes, there are other slow, irreversible changes produced in both purine and pyrimidine nucleotides. These changes, which give rise to complications, are mainly caused by radiation of shorter wavelengths produced in small amounts by the mercury lamp. For this reason it is necessary to use the filter supplied with the lamp, which has a cutoff at about 240 mμ. Even so, small changes in the spectra of the purine nucleotides are observed. Hence spectra of irradiated solutions of the purine (as well as of the pyrimidine) nucleotides must be incorporated in the spectral library.

The ultraviolet spectra of solutions of the pyrimidine nucleotides before and after the irradiation step are shown in Figure 1. It is evident that irradiation causes a marked change in the CMP spectrum, and essentially destroys that of UMP. The latter effect results in the desired improved resolution of adenylic and uridylic acids when spectra of both irradiated and nonirradiated solutions are included in the library and for analysis.

Figure 2 shows typical first-order plots of the photoconversion of UMP alone and in mixtures, under the conditions employed in the irradiation step of the analysis. The rate of photoconversion of UMP in the absence of other nucleotides is the same when the absorbance is 1.0 or less at 260 mμ. But when the absorbance exceeds this value (not shown), there is a significantly slower rate initially, which, however, approaches that shown in Figure 2 as the solution is bleached. The rate of UMP photoconversion is also markedly diminished in the presence of other nucleotides, even when the total initial absorbance does not exceed 1.0. This may be ascribed to the filtering action of these absorbing substances, resulting in reduced intensity of the radiation available for the photochemical reaction. From many such experiments, it was found that irradiation for 150 minutes is sufficient to convert 99% or more of the UMP present in any mixture whose initial absorbance does not exceed 1.0 and whose UMP content exceeds 10% of the total absorbance at 260 mμ.

The rate of photoconversion of CMP also varies with the composition of the mixture. Indeed, for a total absorbance at 260 mμ of 1.0 or less, the reaction is never complete after 150 minutes of irradiation under the prescribed conditions, because of the reversibility of the reaction. Consequently, irradiated solutions contain an indeterminate mixture of CMP and its photoproduct.

It should be emphasized that this indeterminacy presents no problem in carrying out the analyses of mixtures. Thus, the spectral library does not need to contain the spectrum of the pure CMP photoproduct in addition to that of CMP, because the sum of the concentration of the unchanged CMP and of the photoproduct is equal to the original concentration of CMP before irradiation. This equivalence is made use of in the mathematical analysis (*vide infra*) in such a way that the total CMP concentration is obtained. In effect, this analysis of the spectra of irradiated solutions is based on a five- rather than a four-component mixture, but it does not assume any particular proportion of CMP to its photoproduct either in the hydrolysate or in the irradiated part of the CMP library.

It can be seen then, that by using this algebraic device

to calculate the indeterminate mixture of CMP and its photoproduct, the actual intensity and time of irradiation are not critical to the analysis so long as sufficient time is allowed for UMP photoconversion to go to completion. Thus, 150 minutes was chosen as the standard irradiation period under the prescribed conditions.

The only problem associated with the irradiation of CMP, then, relates to the rapid reversibility of its photoreaction at room temperature. Since the proportion of CMP to its photoproduct changes significantly during the time it takes to record a spectrum, a valid reproducible spectrum cannot be obtained at room temperature. This problem is circumvented by recording the spectra of irradiated solutions at 4°. The spectra of the nonirradiated solutions are likewise determined at that temperature simply for convenience.

Mathematical Considerations

If the absorbances of the components of a nucleotide mixture are additive and obey Beer-Lambert's law, then

$$D_\lambda = \Sigma_i c_i E_{i_\lambda} \tag{1}$$

where D_λ is the absorbance at wavelength λ, c_i is the concentration of the ith component, and E_{i_λ} is its extinction coefficient at wavelength λ. For a mixture of four nucleotides, four simultaneous equations, generated from (1) at four different wavelengths, are sufficient to determine the concentration of each base; but, small errors in D_λ would cause large errors in these values. These errors may be reduced by using equations from many more wavelengths and employing a statistical approach to estimate the "best" value for c_i. For reasons to be discussed, the absorbances for the nonirradiated and irradiated solutions at twenty-four wavelengths for each are employed.

The mathematical method used is described by Sternberg *et al.* (1960). The principle is to select values of c_i such that

$$s^2 = \Sigma_\lambda (D_\lambda^m - \Sigma_i c_i E_{i_\lambda})^2 \tag{2}$$

is a minimum. Here, D_λ^m is the observed absorbance, and thus, s^2 is the sum of the squares of the deviations between the observed and actual absorbances at all wavelengths. The solution of this problem is best formulated in matrix notation. It makes use of a matrix $\underline{M}$ defined as

$$\underline{M} = (\underline{E}\underline{E}^T)^{-1}\underline{E} \tag{3}$$

where $\underline{E}$ is the matrix of the elements E_{i_λ}, given by the library spectra. The values of c_i are then given by

$$c_i = \Sigma_\lambda \underline{M}_{i_\lambda} D_\lambda^m \tag{4}$$

The millimolar extinction coefficients of the four nonirradiated nucleotides are calculated from the absorbances at the twenty-four wavelengths used, assuming the extinction coefficients at 260 mμ (Volkin and Cohn, 1954). From the absorbances of the irradiated solutions of AMP, GMP, and UMP the corresponding extinction coefficients are calculated, assuming the concentrations to be unchanged.

The irradiated solution of CMP consists of a mixture of unreacted CMP and CMP photoproduct, the sum of whose concentrations is equal to the original CMP concentration. If x is the fraction of CMP which remains unreacted, the *absorbance of the irradiated library solution*, D_λ^l, is given by

$$D_\lambda^l = c_4 x E_{4_\lambda} + c_4(1 - x)E_{5_\lambda} \tag{5}$$

where E_{4_λ} and E_{5_λ} are the extinction coefficients of CMP and pure CMP photoproduct at wavelength λ, and c_4 is the initial concentration of CMP. Dividing by c_4, an apparent extinction coefficient, E'_{4_λ}, for the irradiated solution of CMP can be defined as

$$E'_{4_\lambda} = xE_{4_\lambda} + (1 - x)E_{5_\lambda} \tag{6}$$

In an irradiated unknown mixture of nucleotides, CMP also exists in both unreacted and photoreacted forms. If y is the fraction of the CMP that is unreacted, then the *total absorbance of a mixture*, D_λ^m, of the four nucleotides can be expressed by

$$D_\lambda^m = \sum_1^3{}_i c_i E_{i_\lambda} + c_4 y E_{4_\lambda} + c_4(1 - y)E_{5_\lambda} \tag{7}$$

Solving equation (6) for E_{5_λ} and substituting in equation (7) gives

$$D_\lambda^m = \sum_1^3{}_i c_i E_{i_\lambda} + c_4 y E_{4_\lambda} + c_4(1 - y) \times \left[\frac{E'_{4_\lambda}}{1 - x} - \frac{xE_{4_\lambda}}{1 - x}\right] \tag{8}$$

Rearranging

$$D_\lambda^m = \sum_1^3{}_i c_i E_{i_\lambda} + c_4 \left\{\frac{1 - y}{1 - x} E'_{4_\lambda} + \left[y - \frac{x(1 - y)}{1 - x}\right] E_{4_\lambda}\right\} \tag{9}$$

The sum of the coefficients of E'_{4_λ} and E_{4_λ} is a constant, c_4, which is independent of both x and y.

The irradiated solution is thus analyzed for five components, irradiated AMP, GMP, and UMP, irradiated CMP with apparent extinction coefficient E'_{4_λ}, and nonirradiated CMP with extinction coefficient E_{4_λ}. The total CMP concentration is then given by the sum of the concentrations of the last two com-

TABLE II: Construction of the Five-Component Library Matrix.

	Component[a]				
	1	2	3	4	5
Nonirradiated	E_{AMP}	E_{GMP}	E_{UMP}	E_{CMP}	E_{CMP}
Irradiated	E'_{AMP}	E'_{GMP}	E'_{UMP}	E'_{CMP}	E'_{CMP}

[a] E_{AMP} and following are the extinction coefficients for the nonirradiated solutions of AMP and following. E'_{AMP} and following are the extinction coefficients for the irradiated solutions of AMP and following.

ponents. For convenience in the computation, the matrix $\underline{E}$ is arranged so that effectively the nonirradiated solutions are also analyzed for five components. This is possible because the absorbance of the nonirradiated CMP *can be expressed in two terms* each with the same extinction coefficient, $E_{4\lambda}$, and with the concentration, c_4, expressed the same as in (9). The construction of the library matrix employed is set out in Table II.

The calculation proceeds in two steps: first, the elements of the matrix $\underline{M}$ are calculated from equation (3); once this has been performed for a given set of library spectra, it need not be done again. Then, the concentrations, c_i, of the components are calculated from equation (4). In the present work, both parts of the calculation were performed on an IBM 7090 digital computor using a FORTRAN program.[3]

The computor program was also written to give (a) the standard deviation of the absorbance,

$$\sigma = \sqrt{s^2/(N - C)} \tag{10}$$

where N is the number of wavelengths, 48, C is the number of components, (five), and s^2 is obtained from equation (2); (b) the standard deviations of the concentrations, σ_{ci}; (c) the mole fractions of the components, f_i, and (d) the standard deviations of these mole fractions σ_{fi}. These standard deviation terms are readily obtained from equation (4). Here σ is a measure of the goodness of fit between the measured spectrum of the mixture and the best combination of the library spectra. For a perfect fit it would be zero but it is generally greater on account of errors in the absorbances and the presence of ultraviolet absorbing contaminants, including substances not present in the library. Values of σ up to 0.006 can be accounted for solely in terms of instrumental and reading errors in the absorbance measurements; larger errors generally indicate the presence of contaminants.

The σ_{ci} parameters are particularly useful measures of the errors, and they are each proportional to σ by constants that depend only on the library spectra.[4] These constants range between 0.05 and 0.1.

[3] Full details of this computer program may be obtained from this laboratory upon request.

[4] These constants are the square roots of the diagonal elements of the matrix $(\underline{EE^T})^{-1}$.

Differentiation of Nucleotide Spectra

A correlation term P_{ij},[5] indicative of the similarity between the spectra of any two components i and j, has been employed in determining the conditions that afford maximal differentiation of the common nucleotides.

$$P_{ij} = \frac{\Sigma_\lambda E_{i\lambda}E_{j\lambda}}{\sqrt{\Sigma_\lambda E_{i\lambda}^2 \Sigma_\lambda E_{j\lambda}^2}} \tag{11}$$

Here P_{ij} attains a limit of unity in the case of the identity of i and j, for which the fit is naturally perfect. With increasing dissimilarity of i and j, P_{ij} approaches zero. (Mathematically, when $P_{ij} = 0$, the vectors are orthogonal and the spectra are maximally dissimilar.)

Values of P_{ij} were calculated for all possible nucleotide pairs using spectra obtained at (a) *p*H 7 (thirty-six wavelengths), (b) *p*H 11 (twenty-three wavelengths), (c) *p*H 7 and *p*H 12 (two times thirty-six wavelengths), and (d) *p*H 7 before and after ultraviolet irradiation (two times thirty-six wavelengths). These values, given in Table III, show that for any combination of two nucleotides the P_{ij} values for spectra at neutral *p*H are hardly affected when the corresponding spectra at alkaline *p*H are used in addition, i.e., the spectra at conditions (a), (b), and (c) are not greatly different. By contrast, marked reductions in P_{ij} result for all uridine and cytidine-containing combinations, when the spectra of nonirradiated and irradiated solutions are used.[6] Spectra determined in acid solution do not offer a similar advantage because the UMP spectrum is the same at neutral and low *p*H, and the AMP spectrum differs only very slightly.

These observations suggest that analyses employing the spectra of both normal and irradiated solutions will give more accurate compositional data than those ob-

[5] P_{ij} is the scalar product of the two n-dimensional vectors $E_{i\lambda}$ and $E_{j\lambda}$; it is equal to the cosine of the angle between them.

[6] Although values of P_{ij} less than unity (for $i \neq j$) are necessary for the success of this method of analysis, this condition is not sufficient since it is also necessary that each library spectrum be dissimilar from any linear combination of the other three. A measure of this dissimilarity is provided by the value of the determinant of the square matrix with elements P_{ij}. For mutually orthogonal spectra the determinant would be unity while for linearly dependent spectra the determinant would be zero.

TABLE III: Degree of Similarity of Nucleotide Library Spectra, P_{ij} Correlation Terms, under Different Conditions.[a]

Nucleotide Combinations	*p*H 7 Spectra	*p*H 11 Spectra	*p*H 7 and 12 Spectra	*p*H 7 Spectra before and after Irradiation
A + A	1.00	1.00	1.00	1.00
A + G	0.94	0.95	0.95	0.94
A + U	0.99	0.92	0.96	0.73
A + C	0.87	0.87	0.87	0.79
G + G	1.00	1.00	1.00	1.00
G + U	0.95	0.96	0.95	0.71
G + C	0.94	0.96	0.96	0.87
U + U	1.00	1.00	1.00	1.00
U + C	0.91	0.96	0.93	0.74
C + C	1.00	1.00	1.00	1.00

[a] P_{ij} is unity when spectrum of i is identical to spectrum of j, and approaches zero with increasing dissimilarity.

TABLE IV: Comparison of Analyses of Known Nucleotide Mixtures from Spectra Obtained under Different Conditions

Nucleotide Components		Nucleotide Composition (mole fractions) A	G	U	C
A + G	Actual	0.539	0.461		
	[a]	0.737	0.177	−0.135	0.219
	[b]	0.546	0.447	0.011	−0.004
A + U	Actual	0.683		0.317	
	[a]	0.747	−0.010	0.252	0.009
	[b]	0.738	−0.030	0.299	−0.007
G + U	Actual		0.649	0.351	
	[a]	−0.058	0.632	0.453	−0.031
	[b]	−0.008	0.653	0.348	0.008
A + G + C	Actual	0.210	0.180		0.610
	[a]	0.234	0.119	0.041	0.603
	[b]	0.214	0.164	−0.019	0.641
A + G + U + C	Actual	0.265	0.227	0.123	0.384
	[a]	0.335	0.255	0.072	0.335
	[b]	0.289	0.188	0.131	0.391

[a] Neutral and alkaline spectra used (*p*H 7 and 12). [b] Normal and irradiated spectra used (*p*H 7). These analyses were performed at an early stage in the investigation, when the analytical procedure was less precise.

tained with either a normal spectrum alone or a combination of the spectra in neutral and alkaline solution. Analyses of nucleotides in simulated hydrolysates indicate that this is the case. Some examples are presented in Table IV.

Results

Analysis of Known Nucleotide Mixtures and Simulated Hydrolysates. In order to test the overall analytical procedure, nucleotide mixtures of known composition, containing some or all of the nucleotides in the library, were either analyzed directly using an untreated library, or first taken through the hydrolysis step and then analyzed with a library derived from nucleotides taken through the same hydrolysis step. In this way, any chemical alterations introduced by the hydrolytic conditions, e.g., deamination of CMP to UMP (*vide infra*), would be compensated by their comparable effect on the library components and thus introduce no error in the analyses. Analytical results were computed using the appropriately obtained complete nucleotide library, and also, for mixtures containing fewer than four components, using only those library

TABLE V: Analysis of the Nucleotide Composition of Simulated Hydrolysates.

Nucleotide Components		A	G	U	C	A	G	U	C
		(mole fractions)				(mole fractions)			
		5-Component library used				2-Component library used			
A + G	Actual	0.502	0.498			0.502	0.498		
	Computed	0.483	0.503	0.008	0.006	0.488	0.512		
	Difference	0.019	0.005	0.008	0.006	0.014	0.014		
	SD[a]	0.003	0.003	0.002	0.005	0.002	0.002		
A + U	Actual	0.419		0.581		0.419		0.581	
	Computed	0.417	−0.001	0.567	0.016	0.425		0.575	
	Difference	0.002	0.001	0.014	0.016	0.006		0.006	
	SD[a]	0.006	0.008	0.008	0.011	0.002		0.002	
G + U	Actual		0.417	0.583			0.417	0.583	
	Computed	0.008	0.389	0.571	0.032		0.418	0.582	
	Difference	0.008	0.028	0.012	0.032		0.001	0.001	
	SD[a]	0.009	0.013	0.016	0.020		0.004	0.004	
						3-Component library used			
U + C	Actual			0.465	0.535			0.465	0.535
	Computed	−0.004	0.007	0.462	0.535			0.469	0.541
	Difference	0.004	0.007	0.003	0.000			0.004	0.006
	SD[a]	0.002	0.003	0.003	0.004			0.002	0.002
A + G + U	Actual	0.297	0.293	0.411		0.297	0.293	0.411	
	Computed	0.292	0.305	0.412	−0.008	0.290	0.298	0.412	
	Difference	0.005	0.012	0.001	0.008	0.007	0.001	0.001	
	SD[a]	0.002	0.003	0.003	0.005	0.002	0.002	0.002	
						4-Component library used			
A + U + C	Actual	0.251		0.348	0.400	0.251		0.348	0.400
	Computed	0.259	0.010	0.346	0.385	0.262		0.343	0.395
	Difference	0.008	0.010	0.002	0.015	0.011		0.005	0.005
	SD[a]	0.002	0.002	0.002	0.002	0.001		0.002	0.002
G + U + C	Actual		0.249	0.350	0.402		0.249	0.350	0.402
	Computed	0.000	0.256	0.349	0.395		0.256	0.349	0.395
	Difference	0.000	0.007	0.001	0.007		0.007	0.001	0.007
	SD[a]	0.001	0.002	0.001	0.002		0.001	0.001	0.002
A + G + C	Actual	0.279	0.276		0.445	0.279	0.276		0.445
	Computed	0.277	0.283	0.008	0.432	0.280	0.281		0.439
	Difference	0.002	0.007	0.008	0.013	0.001	0.005		0.006
	SD[a]	0.002	0.002	0.001	0.002	0.002	0.002		0.003
A + G + U + C	Actual	0.234	0.232	0.162	0.372				
	Computed	0.226	0.243	0.158	0.373				
	Difference	0.008	0.011	0.004	0.001				
	SD[a]	0.003	0.004	0.003	0.005				
A + G + U + C	Actual	0.240	0.237	0.332	0.191				
	Computed	0.236	0.244	0.327	0.194				
	Difference	0.004	0.007	0.005	0.003				
	SD[a]	0.001	0.002	0.002	0.003				
A + G + U + C	Actual	0.201	0.201	0.279	0.321				
	Computed	0.204	0.204	0.275	0.317				
	Difference	0.003	0.003	0.004	0.004				
	SD[a]	0.001	0.001	0.001	0.002				

[a] SD of computed = σ_{I_i} (see text).

components known to be present. In this way it was possible to assess the specificity as well as the accuracy of the multicomponent analysis scheme.

Table V contains a sampling of the results of such single analyses. It is evident that the determined mole fractions are in good agreement with the actual compositions. No significant differences in accuracy were observed between analyses performed with or without the hydrolysis step. The absolute differences between the actual and computed mole fractions, δ_i, vary between zero and 0.032; in most cases (34 of 44) $\delta_i \leq 0.008$. This indicates a level of accuracy not generally attainable by conventional procedures employed for nucleotide analysis in hydrolysates. As δ_i and σ_{f_i} both indicate errors in the analytical results, a rough correlation between them would be expected.[7] In other words, the worse the fit between the spectrum of the mixture and the best combination of the library spectra, the larger σ and the greater the error in the analyses. It is seen that in most of the analyses, $\delta_i \leq 3\sigma_{f_i}$, as would be expected from statistical theory. The absolute error in the mole fraction of a component in an unknown sample may therefore be expected not to exceed $3\sigma_{f_i}$ more frequently than in these control analyses.

For the 4-nucleotide and 3-nucleotide mixtures, the errors are consistently very low (<4%), even when the smallest component present (any one of the nucleotides) amounts to only 16 mole %. With replicate analyses such errors could be expected to be even smaller, especially since most errors in single analyses are 1–2.5%. The errors in the mole fractions of components actually present in the 2-nucleotide mixtures are no greater (<3%) than those obtained for the more complex mixtures; and these are further reduced when the computations are made with the appropriately restricted libraries. Moreover, in most cases, when the complete library is employed, only very small amounts (<1 mole % in 10 of 12 cases) of absent nucleotides are "found."

While there are too few analyses to exclude any small specific trends in errors, these results do attest to the great specificity of the multicomponent analysis computation and the high accuracy of the overall analytical method.

Analysis of Polyribonucleotides and RNA's. The analysis of polyribonucleotides, including RNA's, requires a hydrolytic step that the control analyses did not evaluate. The conversion of CMP to UMP during the hydrolytic step was assessed by appropriately incubating samples of 2′(3′)-CMP for varying periods of time, and analyzing them for both CMP and UMP, employing a library derived from nucleotides *not* taken through the hydrolysis step. The data in Table VI, top, indicate that such deamination does occur, since the amount of UMP increases with time of incubation at the expense of CMP. The deamination amounts to 5% for the incubation time employed in the analytical procedure, 24 hours. The errors which such deamination could introduce in polyribonucleotide analyses are avoided by employing a library derived from nucleotides taken through the standard hydrolysis step.

The adequacy of the hydrolytic conditions (Davidson and Smellie, 1952) was reevaluated in this work because of recent reports (Lane and Butler, 1959; Mii and Warner, 1960) that guanosine-containing dinucleotides are relatively resistant to alkaline hydrolysis. Aliquots of two random copolymers, poly-AU and poly-GU, were incubated under the hydrolytic conditions for varying times and then analyzed using the standard library spectra of the relevant nucleotides. Table VI, bottom, shows that the compositional data are essentially constant by 6 hours of hydrolysis for both polymers, and that there is little if any effect of further hydrolysis on the accuracy of the data. Since the guanosine content of the poly-GU is very high, this copolymer must contain a high proportion of repeating guanosine sequences. As the hydrolysis of guanine dinucleotides would lead to a large spectral change (Michelson, 1959) it is inferred that these observations do not indicate a marked resistance to alkaline hydrolysis of such guanosine sequences. Similar results have been obtained with RNA samples. It is concluded from all these observations that hydrolysis for 24 hours under the conditions employed is an adequate and safe procedure.

Table VII contains RNA composition data obtained in the present work, together with some comparable literature values. Again, the standard deviations are generally noticably small. In most cases, the compositional data obtained in this laboratory fall well within the range of literature values, although there appears to be a distinct tendency for the present analyses to be higher in cytidine and lower in uridine. This tendency is consistent with deamination of CMP, which in the conventional analytical methods is not taken into account. Slightly higher values for guanosine in this work will also be noticed. This may be owing to low solubility of GMP, which would introduce errors in the conventional procedures.

Discussion

Since the nucleotides of interest follow Beer's law and the additivity law in the concentration range of interest,[8] and the irradiation procedure is reproducible, the analytical method described should, in principle, be capable of high accuracy; and the control analyses indicate that this potentiality has been realized. Two main sources of error should be recognized as con-

[7] Here σ_{f_i} is proportional to σ, which is directly related to s^2, which in turn is a measure of the difference between the observed and computed absorbances of all wavelengths (see Mathematical Considerations).

[8] At neutral *p*H, deviations from Beer's law and the additivity law have been found at very low concentrations for the purine and pyrimidine bases and nucleosides, but not for the nucleotides (M. Bramwell, R. J. Williams, L. Klotz, and J. R. Fresco, unpublished observations).

TABLE VI: Effect of Incubation Time on Deamination of CMP and on Hydrolysis of Polyribonucleotides in 0.3 N KOH at 37°.

Sample	Time of Hydrolysis (hours)	Nucleotide Composition[a] (mole fractions) A	G	U	C
2′(3′)-CMP	0			− 0.004 ± 0.001	1.004 ± 0.001
	6			0.021 ± 0.002	0.979 ± 0.002
	12			0.026 ± 0.003	0.974 ± 0.003
	18			0.029 ± 0.002	0.971 ± 0.002
	24			0.057 ± 0.002	0.943 ± 0.002
	36			0.067 ± 0.002	0.933 ± 0.002
Poly-AU	6	0.692 ± 0.012		0.309 ± 0.012	
	12	0.700 ± 0.012		0.300 ± 0.012	
	24	0.700 ± 0.009		0.300 ± 0.009	
	36	0.703 ± 0.008		0.297 ± 0.008	
	48	0.711 ± 0.009		0.289 ± 0.009	
Poly-GU	6		0.841 ± 0.008	0.159 ± 0.008	
	12		0.826 ± 0.012	0.174 ± 0.012	
	24		0.816 ± 0.008	0.184 0.008	
	36		0.824 ± 0.010	0.176 ± 0.010	
	48		0.833 ± 0.013	0.167 ± 0.013	

[a] Calculated with the relevant unhydrolyzed library spectra.

tributing to inaccuracies in the method even when the library is accurately determined, the analytical procedure is carefully followed, and all precautions are taken to avoid impurities. Random errors in the measurement of sample spectra, which arise from small inaccuracies in spectrophotometric response and in chart reading (±0.002–0.003 absorbance unit), will amount to ±0.005 mole %. Ultraviolet absorbing impurities that have not been removed from the sample being analyzed (or have been introduced during the procedure) or any nucleotides of the sample not present in the library will give rise to nonrandom errors, which if large (>0.02 mole %), will be indicated by higher σ values. The cumulative effect of the random errors have been seen from the control analyses (Table V) to be generally ±0.01 mole % or less, which is better than that attainable by the conventional procedures; and errors of the nonrandom type were also seen to be small in the RNA analyses shown in Table VII.

The method presented here possesses additional advantages over the conventional methods for nucleotide analysis, when the qualitative content is already known. It is much faster and involves fewer operations, and a large number of analyses can be performed concurrently. Separation and then quantitative elutions (which often involve losses, especially of GMP) are obviated. An outstanding advantage is that the size of the sample required for a single analysis is very small; for cuvets requiring 1 ml of solution, only 10–25 μg of sample is needed. With microcuvets, the method ought to be adaptable to the 1– 5μg level without any sacrifice of accuracy. Finally, by utilizing a library

TABLE VII: Base Composition of Ribonucleic Acids.[a]

Sample	Nucleotide Composition (mole fractions) A	G	U	C	Reference
Yeast Ribosomal	0.268	0.276	0.240	0.216	Found
RNA	± 0.003	± 0.003	± 0.003	± 0.005	
	0.250	0.280	0.270	0.200	Lit.[b]
Escherichia coli	0.244	0.334	0.178	0.244	Found
Ribosomal	± 0.006	± 0.007	± 0.005	± 0.012	
RNA	0.270	0.350	0.170	0.210	Lit.[c]
	0.252	0.315	0.217	0.216	Lit.[d]
TMV-RNA	0.259	0.286	0.263	0.192	Found
	± 0.011	± 0.013	± 0.011	± 0.022	
	0.298	0.253	0.263	0.185	Lit.[e]
	0.287	0.254	0.282	0.178	Lit.[f]
TYMV-RNA	0.221	0.170	0.231	0.378	Found
	± 0.002	± 0.002	± 0.002	± 0.001	
	0.224	0.172	0.221	0.383	Lit.[g]
Neurospora crassa	0.258	0.294	0.226	0.222	Found
Ribosomal	± 0.005	± 0.006	± 0.005	± 0.010	
RNA	0.249	0.274	0.249	0.228	Lit.[h]
	± 0.005	± 0.005	± 0.008	± 0.009	
Aspergillus niger	0.248	0.312	0.200	0.242	Found
Ribosomal	± 0.004	± 0.005	± 0.003	± 0.008	
RNA	0.243	0.301	0.227	0.229	Lit.[i]
	± 0.004	± 0.006	± 0.005	± 0.003	

[a] The data in this work are the means of single analyses of two separately hydrolyzed samples. The standard deviation values given, σ_{fi}, are means for these two analyses. [b] Crestfield *et al.* (1955). [c] Pardee and Prestidge (1956). [d] Spahr and Tissières (1959). [e] Knight (1952). [f] Reddi (1957). [g] Symons *et al.* (1963). [h] Henney and Storck (1963). [i] Moyer and Storck (1964).

derived from nucleotides subjected to all the chemical steps in the analysis, errors that might otherwise arise from chemical alterations of components in the unknown sample are obviated.

Two disadvantages of the method should be noted. One is that the expected error (0.002–0.02 mole %) gives the analytical data on a recognized (included in the library) small component a large uncertainty. The other disadvantage stems from the absence from the library of many newly discovered (or as yet undiscovered) *minor components* present in certain nucleic acids. If they are truly *minor components*, their absence from the library will give rise to only very small distortions of the actual proportions of those nucleotides present in both the sample and the library; if present in larger amounts, such *minor components* should lead to unsatisfactory analyses (large σ values). Moreover, it should be possible to determine some of these minor components by including their appropriate spectra in the library and making small modifications in the procedure that enable accurate assessment of their contribution to the total spectrum. This would involve extending the number of absorbance data within and/or outside the present wavelength range used; for as the number of components in the library is increased, the opportunities for resolving their spectra must also be increased, if high accuracy is to be maintained. The possibility of exploiting the far-ultraviolet region for additional spectral differentiation exists.

The method of multicomponent analysis of ribonucleotide mixtures described here has been applied successfully in this laboratory to the analysis of ribosomal and viral RNA's and many different polyribonucleotides. It should also be readily applicable to the products of enzymic digestion of such substances and, after suitable modification, to the analysis of deoxyribonucleotide mixtures and hence DNA. Indeed, it should be possible to analyze for bases, or nucleosides as well, by similar procedures. However, for each system to which the method may be applied, Beer's law must be shown to hold rigorously.[8] The use of least-square fitting calculations for establishing the identity of certain di- and trinucleotides with known spectra has already been reported (Pratt *et al.*, 1962, 1964).

Acknowledgment

We are grateful to Drs. A. B. Pardee and H. Sober, and the staff of the Princeton University Computer Center for helpful discussions, and to Dr. David Straus for a critical reading of the manuscript. RNA samples were graciously provided by Drs. H. Boedtker, H. G. Wittmann, C. G. Kurland, R. Haselkorn, N. D. Zinder, R. Storck, and S. Mandeles.

References

Cohn, W. (1955), *in* The Nucleic Acids, Vol. I, Chargaff, E., and Davidson, J. N., eds., New York, Academic, p. 211.

Crestfield, A. M., Smith, U. C., and Allen, F. W. (1955), *J. Biol. Chem. 216*, 185.

Davidson, J. N., and Smellie, R. M. S. (1952), *Biochem. J. 52*, 599.

Henney, H., and Storck, R. (1963), *J. Bacteriol. 85*, 222.

Kerr, S. E., and Seraidarian, K. (1945), *J. Biol. Chem. 159*, 211.

Knight, C. A. (1952), *J. Biol. Chem. 197*, 241.

Lane, B. G., and Butler, G. C. (1959), *Can. J. Biochem. Physiol. 37*, 1329.

Loring, H. S. (1955), *in* The Nucleic Acids, Vol. I, Chargaff, E., and Davidson, J. N., eds., New York, Academic, p. 199.

Michelson, A. M. (1959), *J. Chem. Soc.*, 1371.

Mii, S., and Warner, R. C. (1960), *Federation Proc. 19*, 3 7.

Moyer, R. C., and Storck, R. Y. (1964), *Arch. Biochem. Biophys. 104*, 193.

Pardee, A. B., and Prestidge, L. S. (1956), *J. Bacteriol. 71*, 677.

Pratt, A. W., Toal, J. N., Rushizky, G. W., and Sober, H. A. (1964), *Biochemistry 3*, 1831.

Pratt, A. W., Toal, J. N., and Sober, H. A. (1962), *Federation Proc. 21*, 376c.

Reddi, K. K. (1957), *Biochim. Biophys. Acta 25*, 528.

Reid, J. C., and Pratt, A. W. (1960), *Biochem. Biophys. Res. Commun. 3*, 377.

Rogoff, M. (1957), *Ann. N.Y. Acad. Sci. 69*, 27.

Schramm, G., and Gierer, A. (1957), *Nature 177*, 702.

Sinsheimer, R. (1954), *Radiation Res. 1*, 505.

Sinsheimer, R. (1957), *Radiation Res. 6*, 121.

Smith, J. D. (1955), *in* The Nucleic Acids, Vol. I, Chargaff, E., and Davidson, J. N., eds., New York, Academic, p. 267.

Spahr, P. F., and Tissières, A. (1959), *J. Mol. Biol. 1*, 237.

Steiner, R. F. (1961), *J. Biol. Chem. 236*, 3037.

Sternberg, J. C., Stills, H. S., and Schwendemann, R. D. (1960), *Anal. Chem. 32*, 84.

Symons, R. H., Rees, M. W., Short, M. N., and Markham, R. (1963), *J. Mol. Biol. 6*, 1.

Vasilenko, S. K., Kamzolova, S. G., and Knorre, D. G. (1962), *Biokhimyia 27*, 142.

Volkin, E., and Cohn, W. (1954), *Methods Biochem. Analy. 1*, 287.

Wyatt, G. R. (1955), *in* The Nucleic Acids, Vol. I, Chargaff, E., and Davidson, J. N., eds., New York, Academic, p. 243.