VECTOR ANALYSIS OF ULTRAVIOLET MIXTURE SPECTRA: THE COMPOSITION OF RIBONUCLEIC ACID

James C. Reid and Arnold W. Pratt

National Cancer Institute, National Institutes of Health,
Bethesda, Maryland

Received September 16, 1960

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. If n components are present, measurements at n or more points permit the establishment of a set of m simultaneous equations which can be solved for the concentration of each component. Such treatment works well with highly differentiated spectra such as mass spectra or nuclear spectra where n and m can commonly be made equal and relatively simple analogue devices can be used to ease the equation-solving burden.

Application of such technique to ultraviolet spectra rarely yields useful results on mixtures of more than two components where $n \backsim m$ because the amount of information is not adequate for the precision required by the comparatively undifferentiated character of the spectra. If the entire spectrum is utilized, however, rather than a few points, and the component spectra are treated as vectors to be fitted to the mixture spectrum, a considerable increment in precision must accrue. Such technique is not feasible by manual computation but lends itself readily to machine methods. The best fit is conveniently determined by the method of least squares.

One application of least squares to infrared spectra of steroid mixtures has been reported by Rogoff (1957) and during the course of our own work an application to ultraviolet spectra has appeared by Sternberg, Stills and Schwendeman (1960), also working with steroids. Readers unacquainted with least squares analysis of this type will find some discussion of it in Sternberg et al. (1960), and we shall not dwell upon mathematical aspects in this communication.

The first point we wish to emphasize is the excellence of the precision attainable when the component spectra truly form a basis for the mixture spectrum, that is, when the mixture contains no absorbing components which are not in the library

TABLE I

Computational Analysis of Synthetic Mixtures

Mixture	Xanthine		Cytidine		Guanine	
	A_{260} Theor.	A_{260} Found as % of Theor.	A_{260} Theor.	A_{260} Found as % of Theor.	A_{260} Theor.	A_{260} Found as % of Theor.
$T_1 = S_1+S_2+S_3$	.322	101	.220	100	.153	97.5
1/3 T_1	.107	103	.074	100	.051	100
1/9 T_1	.036	105	.024	102	.017	94.3
$T_2 = T_1+3S_1$	.806	99.9	.055	98.5	.038	96.6
$T_3 = T_1+3S_2$	.081	104	.551	99.6	.038	96.2
$T_4 = T_1+3S_3$	.081	103	.055	103	.382	103
$T_5 = T_2+T_3$	.443	100	.301	99.5	.038	98.6

Three stock solutions were prepared: S_1 (xanthine), S_2 (cytidine), and S_3 (guanine). Mixture T_1 consisted of equal parts of the stocks; 1/3 T_1 and 1/9 T_1 are 1:3 and 1:9 dilutions of T_1. Other mixtures were made in the proportions indicated; for example, "$T_1 + 3S_1$" means one part of T_1 plus three of S_1. All measurements were made in 0.4 M HCl in a 1 cm. cell; A_{260} designates absorbance at 260 mμ.

of components against which the mixture is being analyzed. Table I demonstrates this for a series of test mixtures made up of cytidine, guanine and xanthine mixed in various proportions to cover the range from minor to major constituent. The excellence of the agreement is readily apparent but what is particularly striking is the fact that no more than 5% error is encountered at absorbances as low as 0.025, a region which one would not ordinarily regard as significant. We have obtained similar results in alkaline medium; these data are not included.

The computations of table I were made from readings obtained at 37 points on each spectrum, that is, at 2.5 mμ intervals through the UV range. This does not exhaust the information available however, for readouts can be spaced as close as 1.0 mμ. Moreover, alkali and acid data can be combined to form extended vectors and indeed a whole series of pH values could in principle be used to increase the power of the computation. We are engaged in a program of exploring the limits of practical applicability of such methodology in particular connection with a study of urinary excretion of nucleic acid congeners in leukemia. For this purpose the study of RNA analysis provides a convenient model, as well as offering the possibility of developing a useful application of the technique.

Table II shows results obtained when two samples of a single batch of commercial RNA were subjected to Schmidt-Thannhauser digestion and the spectra of the digests were run against a matrix formed from the spectra of the four component nucleotides (commercial specimens). Aliquots were also analyzed by Cohn's anion-exchange column method. Table II further illustrates the potential precision of the method; the observed replication error or ca. 3% is considerably better than that of paper chromatographic methods commonly employed. The absolute values are seen to be comparable to those obtained by the chromatographic method to within the precision limits of the latter with the exception of the somewhat high figure for uridylic acid. It is to be anticipated, however, that this can be substantially improved, for

TABLE II

Comparison of Computational With Chromatographic Analysis of RNA

Component	Chromatographic Analysis			Computational Analysis			
	Composition as % of Total A_{260} of Digest		Ratio Sample 2/Sample 3	Composition as % of Total A_{260} of Digest		Ratio Sample 2/Sample 1	Ratio Av. of Computed Values / Av. of Chrom. Values
	Sample 2	Sample 3		Sample 2	Sample 1		
Cytidylic Acid	12.2	13.3	.920	12.3	12.2	1.01	.963
Adenylic Acid	30.5	30.6	.996	30.4	30.5	.996	1.01
Uridylic Acid	24.8	24.1	1.03	29.1	29.4	.990	1.12
Guanylic Acid	24.6	26.5	.928	27.9	27.2	1.03	1.08

the commercial stocks used for the library curves may not have been altogether pure. Moreover, the column chromatography revealed the presence of at least two additional components, comprising about 6% of the total and the library therefore did not exactly describe the mixture.

TABLE III

Identification of Single Components

Library Component	% of Total Absorbance	
	Adenylic Acid b fraction	Guanylic Acid b fraction
Cytidylic Acid Cut A	- 5.3	- 3.0
" " " B	0.3	- 2.7
Adenylic Acid - a	99.5	- .1
Uridylic Acids	4.3	- 8.9
Guanylic Acid - a	- 3.8	101.9
Resin bkgd.	.7	4.2
HCl strip	4.7	13.5
HCl bkgd.	.3	- 2.0
$FeCl_3$	.5	- 2.8

A specimen of RNA was subjected to Schmidt-Thannhauser digestion, then resolved on Dowex-1 chloride by Cohn's dilute hydrochloric acid system. A matrix was composed of spectra of all fractions except the b isomers of adenylic and guanylic acids (whose spectra are very similar to the a isomers). Certain background entities were also included as indicated. The spectra of the b fractions were then run against the matrix as a non-redundant computation showing the retention of accuracy in a large system.

In table III is illustrated further the practical applicability of vector analysis to multicomponent UV spectra. As is indicated by the text below the table, the computation should indicate substantially 100% adenylic acid and guanylic acid respectively in the two fractions. Lack of a perfect fit with the library is evident in the negative numbers, although the numerical values obtained for the components actually present are quite good. An important cause of the lack of fit is hydrolysis of the nucleotides during chromatography, as was evidenced by paper chromatography of the column fractions. Probably for this reason too, computation of the composition of the digests against this library failed to give results significantly better than those in table II (data not shown).

The reproducibility and absolute accuracy demonstrated here relative to the chromatographic data indicates the feasibility of using computational technique as a new method for nucleic acid analysis, provided fairly pure specimens are used. There is no apparent reason why the same technique, which is of course applicable to very small samples, could not be used for base-ratio determinations on DNA by employing enzymatic hydrolysis and of course including spectra for the enzymes in the library. It is obvious that the determination of the absorption spectrum of a hydrolysate is very much simpler to carry out than the chromatographic methods of analysis presently employed.

References

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

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