BASE COMPOSITION OF INTACT NUCLEIC ACID OLIGOMERS*

by

Stanley Mandeles and Charles R. Cantor**, Space Sciences Laboratory and Department of Chemistry, University of California, Berkeley, California.

SYNOPSIS

The base composition of intact, purified, oligonucleotides has been determined by comparing the absorption spectrum of each oligomer with calculated curves. The spectral curve of each oligomer was measured in 7 M urea at three pH values and digitized at 1 mμ intervals. The calculated curves consisted of the sums of the absorption spectra of mononucleotides in 7 M urea at the same pHs. A computer was used to make the comparisons and establish the best fit. Results are presented for nine of the ten possible dimer composition isomers; trimers from pancreatic and T_1 Ribonuclease hydrolysates of TMV-RNA; poly-A, poly-C, and poly-U; and intact tobacco mosaic virus-RNA.

INTRODUCTION

A necessary procedure in the determination of base sequences in nucleic acids is the base compositional analysis of pure oligomers from nucleic acid hydrolysates. Methods commonly used involve hydrolysis of the purified oligomer to the constituent mononucleotides, separation of the nucleotides chromatographically, and estimation of each nucleotide by its ultraviolet absorbance[1]. Recently, Pratt, et al[2], Lee, et al[3], and Guschlbauer, et al[4], have described spectrophotometric techniques for base com-

---

*Supported by grants NsG 479 by the National Aeronautics and Space Administration and GM 12158-02 by the National Institutes of Health to the Space Sciences Laboratory, University of California, Berkeley, California.

**National Science Foundation, Cooperative Pre-doctoral Fellow.

positional analysis that can be applied directly to the hydrolyzed oligomer. These procedures eliminate the need for a chromatographic separation of the individual nucleotides. The requirement for hydrolysis of the oligomer, however, remains.

Because methods of base sequence determination require an intact oligomer, we wished to adapt a method of spectrophotometric analysis for use with unhydrolyzed single oligomers. In principle, this can be accomplished under conditions where the optical properties of the intact oligomer is the sum of the optical properties of its component mononucleotides. These conditions may be satisfied by using solvents such as 7 M urea or ethanol[5]. These solvents tend to prevent the formation of hydrogen-bonded base pairs, and minimize stacking between nighboring bases. Thus, one can expect that, in 7 M urea or in ethanol, spectral properties of oligomers will approach sums of monomer properties.

This report details some of our findings on the use of 7 M urea as a solvent for the spectrophotometric determination of the base composition of intact oligonucleotides. Seven molar urea is a particularly convenient solvent because many of the effective separation techniques yield oligonucleotides as solutions in weakly buffered 7 M urea[6,7]. We have analyzed spectra of chromatographically purified oligomers with the aid of a simple computer program which quantitatively compares an experimental absorption curve with a set of possible spectra. However, least squares[4] or linear programming[2] could easily be applied to the analysis of the spectra of unhydrolyzed oligomers.

EXPERIMENTAL

Materials: Most of the studies reported here were made on nucleic acid

dimers and trimers from pancreatic RNase[8] (EC 2.7.7.16) and T_1 RNase[9] (EC 3.1.4.8) digests of TMV-RNA. Some of the data were taken with dimers obtained commercially from Gallard-Schlesinger, New York. Poly-A, poly-C, and poly-U were obtained from Miles Laboratories, Elkhart, Indiana. TMV-RNA was prepared from the virus by a modification[10] of a method of Stanley and Bock[11].

Purification of Urea: Reagent grade urea, Baker and Adamson, was dissolved in water to a final concentration of 7 M. This solution was decolorized by the addition of 2 grams of decolorizing charcoal (Norit A) per liter of solution, followed by heating to 70° and filtering the cooled solution. This solution was then passed through a DEAE-cellulose column (2 x 20 cm), bicarbonate form, to remove interfering anions and the last traces of charcoal. A small amount (500 ml) of this urea solution was freeze-dried. Appropriate amounts of this dried material were added to reagent grade concentrated HCl and 2M KOH, respectively, to make stock 7 M urea-acid or base which were used to adjust the pH of the 7 M urea buffers.

7 M urea buffers: Portions of the 7 M urea solution were adjusted to pHs 1.60, 7.05, and 11.05 by the addition of the stock urea-acid or urea-base reagents. The pH adjustments were made on a Corning model 12 pH meter with expanded scale. After pH adjustment, the buffers were filtered through Whatman No. 40 filter paper. It was found that the pH 1.60 7 M urea solution remained at constant pH for several days, at 25°. The urea solution at pH 11.05 decreased 0.3 pH units in 24 hours, while the solution at pH 7.05 increased 0.1 pH units in four hours. The pHs of these latter solutions were therefore checked and adjusted at least twice a day during the spectral measurements.

Methods: A) Preparation of oligomers. Approximately 100 - 200 mg samples of TMV-RNA were hydrolyzed at pH 7.5 with pancreatic RNase or T_1 RNase at 40° in a Radiometer TTT-1c pH-stat. The extent of the hydrolysis was followed by recording the amount of standard alkali used to neutralize the H^+ released, and enough of the enzyme was added so that 95% of the theoretical H^+ was released within six hours. The hydrolysates were then acidified with HCl to pH 2 and maintained at this pH for three hours at 40°, to cleave any cyclic phosphates that remained. The hydrolysate was then freeze-dried, the solid residue was dissolved in 10 ml of 7 M urea and this solution was chromatographed on a Sephadex-A-25 column at pH 7.6 in the presence of 7 M urea and a NaCl gradient[9]. This procedure separates the nucleic acid oligomers according to charge. The fractions corresponding to each charge were combined and the 7 M urea was removed by rechromatography of a diluted solution on AG 1X2 or A-25. Elution of the oligomers from A-25 was accomplished with 0.5M NH_4HCO_3. Most of the NH_4HCO_3 was removed when this solution was evaporated under reduced pressure at 40°. The remainder volatilized during the subsequent freeze-drying process. The residue of each fraction consisted mainly of oligomers which had the same charge at pH 7.6. These oligomers were then separated according to base composition on a column of AG 1X2 using an HCl-NaCl gradient.

Spectral measurements: Dimers and trimers purified in the above fashion were freeze-dried, dissolved in a small quantity of unbuffered 7 M urea and added dropwise to quartz cuvettes containing 7 M urea at the three pHs described above. Reference samples were made by isolating blank portions of the AG 1X2 chromatogram adjacent to peak areas. These samples were treated in the same manner as the oligomers and added to quartz cuvettes that were matched

with the sample cuvettes. The UV spectrum was recorded in a Cary 15 spectrophotometer equipped with a Cary-Datex No. 56-220-300 analog to digital attachment. This permitted the recording of the absorbance and wavelength on punched paper tape. The recording were made at one millimicron intervals from 320 mμ to 225 mμ. In general, enough of the oligomer was added to provide a maximum absorbance of 1 in the 10 mm path length cells.

The punched tape was then converted to IBM cards and the data were analyzed by computer.

Analysis of spectral data: A computer program was devised[12] so that data from the spectral curve of an unknown oligomer in 7 M urea could be compared with the calculated spectral curves of a variety of oligomers. The program computes the spectra of the $(n+3)!/n!3!$ compositional isomers of chain length n (assuming only four bases). The spectra are calculated by summing the spectra of the monomers in 7 M urea, weighted according to the base composition. These calculated oligomer spectra are each normalized to an absorbance of 1.000 at 260 mμ and compared with the normalized absorption curves of the unknown oligomer. The use of normalized absorption spectra somewhat diminishes the sensitivity of the method. However, this greatly simplifies the analysis, since concentrations need not be determined. The difference between each normalized calculated spectra and unknown is computed at 1 mμ intervals, squared, and summed. Finally the sums of the squares of the deviations at three pH's are tabulated in ascending order, together with the corresponding base composition. Thus, the possible base compositions are ordered from best-to-worst overall fit.

The library of monomer spectra used for the calculations was constructed in the following manner: Adenosine-2'-(3')-phosphate, cytidine-2'-(3')-

phosphate, guanosine-2'-(3')-phosphate and uridine-2'-(3')-phosphate were dissolved, separately, in each of the three 7 M urea buffers, and UV spectra were taken as described above. Each spectrum used for the library is the average of three samples. The values of the absorbances at each wavelength were converted to molar extinction by assuming that the relative molar extinction coefficients of the four mononucleotides in 7 M urea are essentially the same as in dilute aqueous buffer[13].

RESULTS AND DISCUSSION

Spectra of Dimers: T_1 RNase dinucleotides were obtained from the chain length separation procedure by isolation of the three negative charge fractions and rechromatography of this material, after removal of excess urea and salt, on an AG 1X2 column. The pattern of this chromatographic separation, which resolves oligomers according to base composition, is shown in Figure 1. Dinucleotides from the chain length fractionation of a pancreatic RNase hydrolysate were similarly separated on AG 1X2. In addition, commercial dinucleoside phosphates were purified by chromatography on AG 1X2. Figure 2 is a pattern obtained when a mixture of the commercial dimers, combined to simulate dimers from a pancreatic RNase hydrolysate, is chromatographed. Each of the dimers was isolated, treated as described in the methods section and compared with calculated oligomers from the library mononucleotides. The results of these comparisons are shown in Table 1. It can be seen that a positive identification has been made in each case, since the square root of the sum of the squares of the deviations of the best fit is less than 1/3 of the sum for the next best fit. This is expressed in column 3 of Table 1 as the ratio of the square root of the second best fit to the square root of the best fit. In general, a higher ratio means a higher probability of a correct identification.

Trimer Spectra: Trinucleotides from pancreatic and T_1 RNase hydrolysates were isolated as described above and resolved according to base composition as shown in Figures 3 and 4. The individual trimers were isolated as before and the spectra taken in the 7 M urea buffers. These results are shown in Table II. It can be seen that here again, correct identifications have been made in every case. In general, the ratios of the second best fit to the best fit are lower for trimers than for dimers. This is particularly noticeable because most of the trimers studied contain A or U. In these cases, the certainty of the identification can be increased by noting the sum of the squares of the deviations at pH 11. It is only at this pH that the absorption curves of A and U have fairly different shapes.

Our results show that the base composition of pure dimers and trimers can be determined satisfactorily without hydrolysis. However, the resolving power of the method decreases with chain length. Correct identification of tetranucleotides and larger oligomers thus may be more difficult, unless the deviations between observed spectra and sums of monomer spectra can be made smaller. As shown in Tables I and II, the average error between a dimer spectrum and the calculated spectrum corresponds to an absorbance of 0.018 O.D. The corresponding error for trimers is 0.022 O.D.

A major source of error may be impurities in the oligomers used. Rechromatography of a dinucleoside phosphate fraction has often resulted in a spectrum which agrees much more closely with the sum of the monomers. One source of impurities in the above procedures for isolating oligomers may be solubilized AG 1X2 which has not been adequately blanked out. The AG 1X2 chromatography was used as an example of an isolation procedure. It could presumably be supplemented or replaced by methods such as two-dimensional electrophoresis-chromatography[14,15]. Another major source

of deviations would be any residual hypochromicity in 7 M urea at 25°C. Stanley and Bock[16] have recently reported on the effect of temperature on the hypochromicity of aqueous solutions of dimers and trimers from pancreatic ribonuclease digests. They found that raising the temperature decreases the hypochromicity, but not to the point where the optical properties of the oligomers were equal to the sum of monomer properties. Presumably a combination of 7 M urea and increased temperature would result in oligomer spectra which more closely fit the sums of monomer properties. Davis[5] has observed that the optical rotatory dispersion (ORD) of alcohol solutions of ApA and ApAp are very close to that of the monomer. Thus ethanol or methanol merits consideration as an alternative or adjunct to urea. However, the insolubility of long oligomers in alcohol may limit its usefulness.

Since the hypochromism of small oligonucleotides arises from base stacking, it should be possible to refine the above calculations of oligomer spectra by including any remaining nearest neighbor interactions[17]. Trimer and higher oligomer spectra would then be used to fit the experimentally determined dimer spectra. This should increase the accuracy of the determination, but would also involve much greater experimental and computational complexity. We do not feel that these refinements are called for at this time.

Spectra of Polymers: It is apparent that the methods we have used for the analysis of the base composition of oligomers would be unsatisfactory for the analysis of polynucleotides: it is inefficient to have to spell out all of the possible base compositions for chain lengths greater than 5 or 6. In addition, the accuracy of determining an arbitrary base composition is hindered by the strong similarities between the spectra of A and U.

It has been shown by Pratt, et al[2] that the methods of linear programming are capable of determining accurate base compositions of random mixtures of mononucleotides. It is therefore of interest to see whether experimental conditions can be found in which the spectrum of a polymer is essentially the sum of the spectra of its monomers. Towards this end, we have examined the spectra of several polynucleotides in the 7 M urea buffers described earlier.

The spectra of poly-A, poly-C, and poly-U, and TMV-RNA were compared with those of the monomers in 7 M urea at pH 7 and 11, and the results are shown in Table III. We were unable to use pH 1.6 because of the insolubility of the polymers in acid. In the case of poly-A and poly-C, the sum of the squared deviations of the monomer and polymer spectra at both pH 7 and 11 differ by about 0.1 to 0.2. This difference is five to ten times greater than our average results for dimers and trimers. Inspection of the spectra shows that the absorption curve of poly-A, in 7 M urea is shifted 2 mμ to the blue when compared with A-2'-(3')-phosphate. Similar effects have been observed with aqueous solutions[18]. Apparently poly-A still retains some secondary structure, even in the presence of such strongly denaturing conditions as 7 M urea. The long wavelength bands of poly-C and monomer C almost superimpose. When these spectra are normalized to 1 at 260 mμ, the magnitude of abosrbance at the shorter wavelength is slightly different in the two cases. This accounts for most of the spectral difference shown in Table III. The spectrum of poly-U at pH 7 in 7 M urea is very close to the spectrum of U-2'-(3')-phosphate. The deviation between the two curves is no larger than that observed with an average dimer or trimer. At pH 11 the spectrum of the polymer is very different from that of the monomer, and in fact, seems to be closer to the spectrum of U-2'-(3')-phosphate at pH 7. This effect may well

be due to a shift in the pK of the bases due to their presence in a polyelectrolyte. For comparison with the spectrum of TMV-RNA, the spectra of the four mononucleotides were weighted according to the base composition determined by several workers,[4,19,20,21] and summed. From Table III it can be seen that the agreement between the polymer spectrum in the 7 M urea and the sum of the monomers is excellent at pH 7. The spectra of two different samples of TMV-RNA were measured. In both cases, the best fit is obtained using the base composition data of Reddi[21]. From the four sets of results shown, it is evident that this method has reasonable sensitivity to small changes in base composition. The spectral agreement of TMV-RNA and the sum of its monomers at pH 11 is no better than that observed for poly-A and poly-C, but this may well be due to the same effects that contribute to the deviations with poly-U. This point could be examined more thoroughly by studying spectra as a function of salt concentration. While added salt might stabilize secondary structure slightly, it should diminish any shifts in the alkaline pH of the bases in poly-U or of the U's and G's in RNA.

In view of the fact that this method analyzes the base composition of intact oligomers, it should be adaptable to automated procedures where the determinations are made directly on column effluents.

The authors are pleased to acknowledge the advice and encouragement of Drs. T. H. Jukes and I. Tinoco, Jr., and the excellent technical assistance of Mr. Frank Fearney.

REFERENCES

1. Volkin, E., and Cohn, W.E., J. Biol. Chem., 205:767 (1953).

2. Pratt, A.W., Toal, J.N., Rushizky, G.W., and Sober, H.A., Biochem., 3:1831 (1964).

3. Lee, S., McMullen, D., and Brown, G.L., Biochem. J., 94:314 (1965).

4. Guschlbauer, W., Richards, E.G., Beurling, K., Adams, A., and Fresco, J.R., Biochem., 4: 964 (1965).

5. Davis, S., Ph.D. Thesis, University of California (1965).

6. Rushizky, G.W., and Sober, H.A., Biochem. Biophys. Res. Comm., 14:276, (1964).

7. Solymosy, F., Tener, G.M., and Reichmann, M.E., Virology, 27:409 (1965).

8. Tomlinson, R.V., and Tener, G.M., Biochem., 2:703 (1963).

9. Bartos, E.M., Rushizky, G.W., and Sober, H.A., Biochem., 2:1179 (1963).

10. Mandeles, S., and Bruening, G., Biochem. Preps., (in press).

11. Stanley, Jr., W.M., and Bock, R.M., Biochem., 4:1302 (1965).

12. Cantor, C.R., Ph.D. Thesis, University of California (1966).

13. "Ultraviolet Absorption Spectra of 5'-Ribonucleotides", Circular OR-10, Pabst Laboratories, Milwaukee, Wisconsin.

14. Rushizky, G.W., and Knight, C.A., Virology, 11:236 (1960).

15. Sanger, F., Brownlee, G.G., and Barrell, B.G., J. Mol. Biol., 13:373 (1965).

16. Stanley, Jr., W.M., and Bock, R.M., Anal. Biochem., 13:43 (1965).

17. Cantor, C.R., and Tinoco, Jr., I., J. Mol. Biol., 13:65 (1965).

18. Holcomb, D.M., and Tinoco, Jr., I., Biopolymers, 3:121 (1965).

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

20. Black, F.L., and Knight, C.A., J. Biol. Chem., 202:51 (1953).

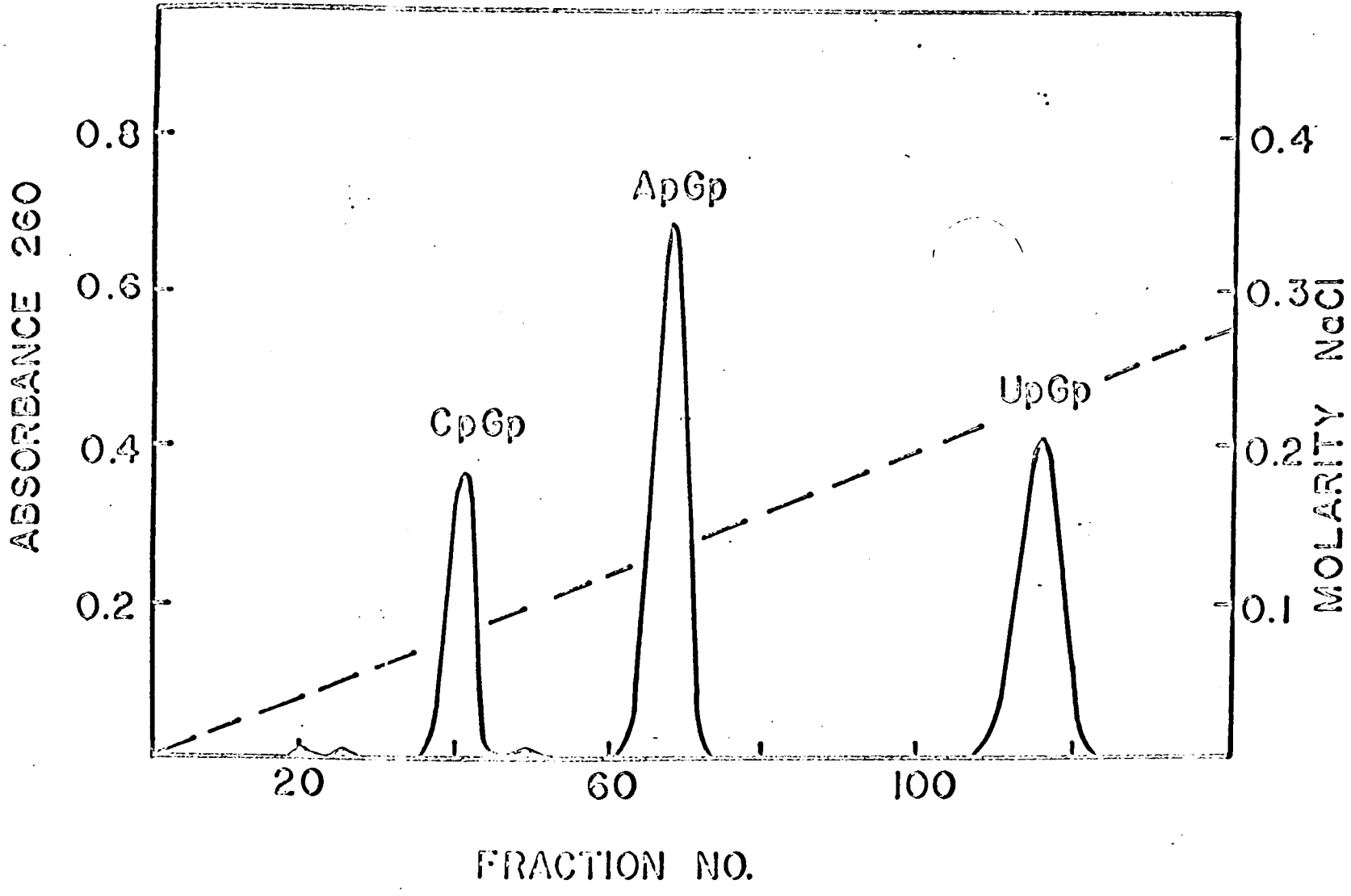
21. Reddi, K.K., Biochim. Biophys. Acta, 25:528 (1957).

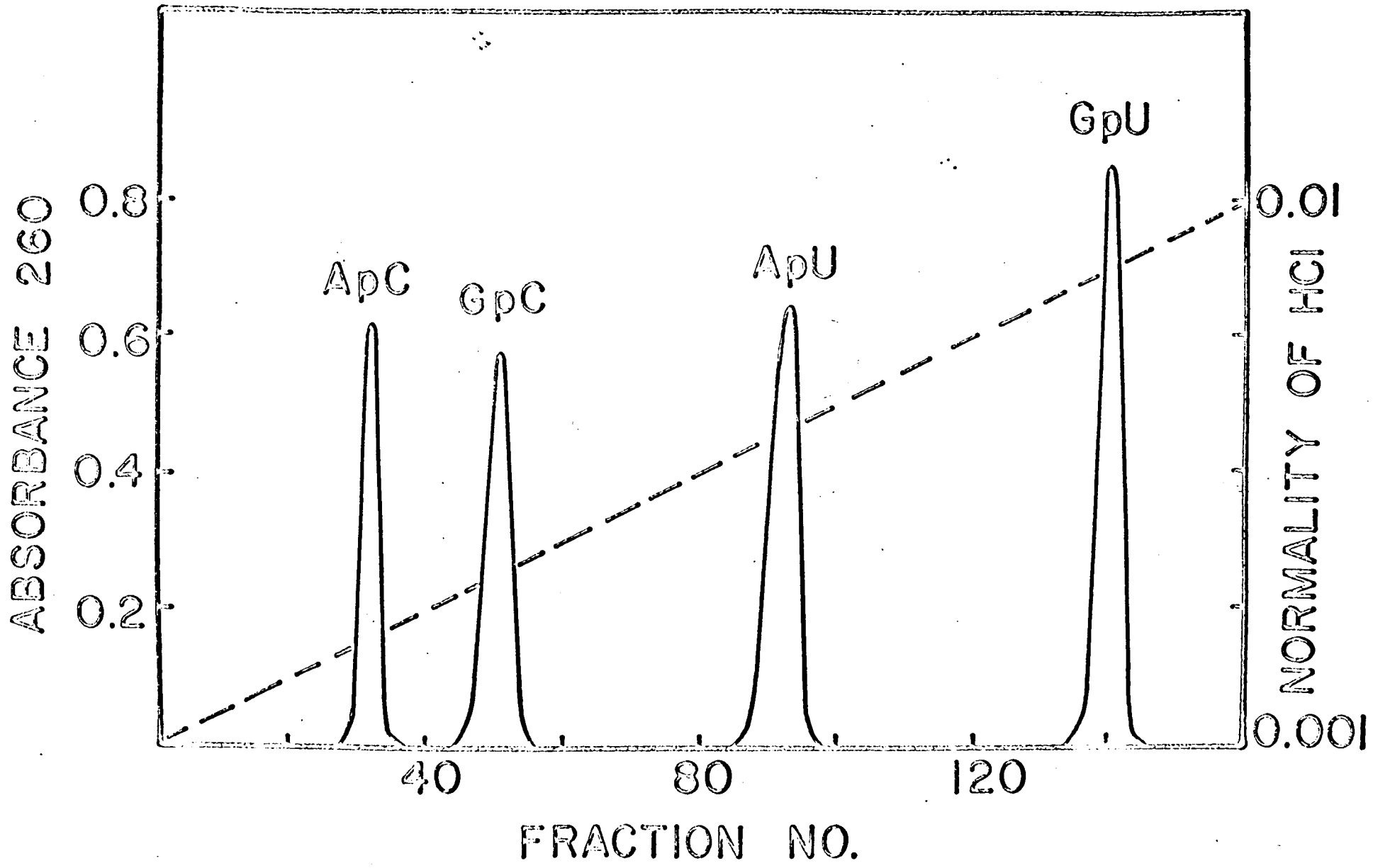
Figure 1. Separation of dimers from a T_1 RNase hydrolysate of TMV-RNA on an AG 1X 2, 200-400 mesh, column (1.1 X 50 cm). The fraction volume was 10 ml and the flow rate was 70 ml/hr. Linear gradient as indicated from 0.001 M HCl to 0.01 M HCl.

Figure 2. Separation of dimers from a commercial source mixed to simulate a pancreatic RNase hydrolysate. Column dimensions and chromatography conditions as shown in Figure 1.

Figure 3. Separation of trimers from a T_1 RNase hydrolysate of TMV-RNA. Column dimensions and flow rate are the same as in Figure 1. Linear gradient from 0.01 M HCl to 0.01 M CHI- 0.4 M NaCl.

Figure 4. Separation of trimers from a pancreatic RNase hydrolysate of TMV-RNA. Column dimensions, and chromatography conditions the same as in Figure 3.

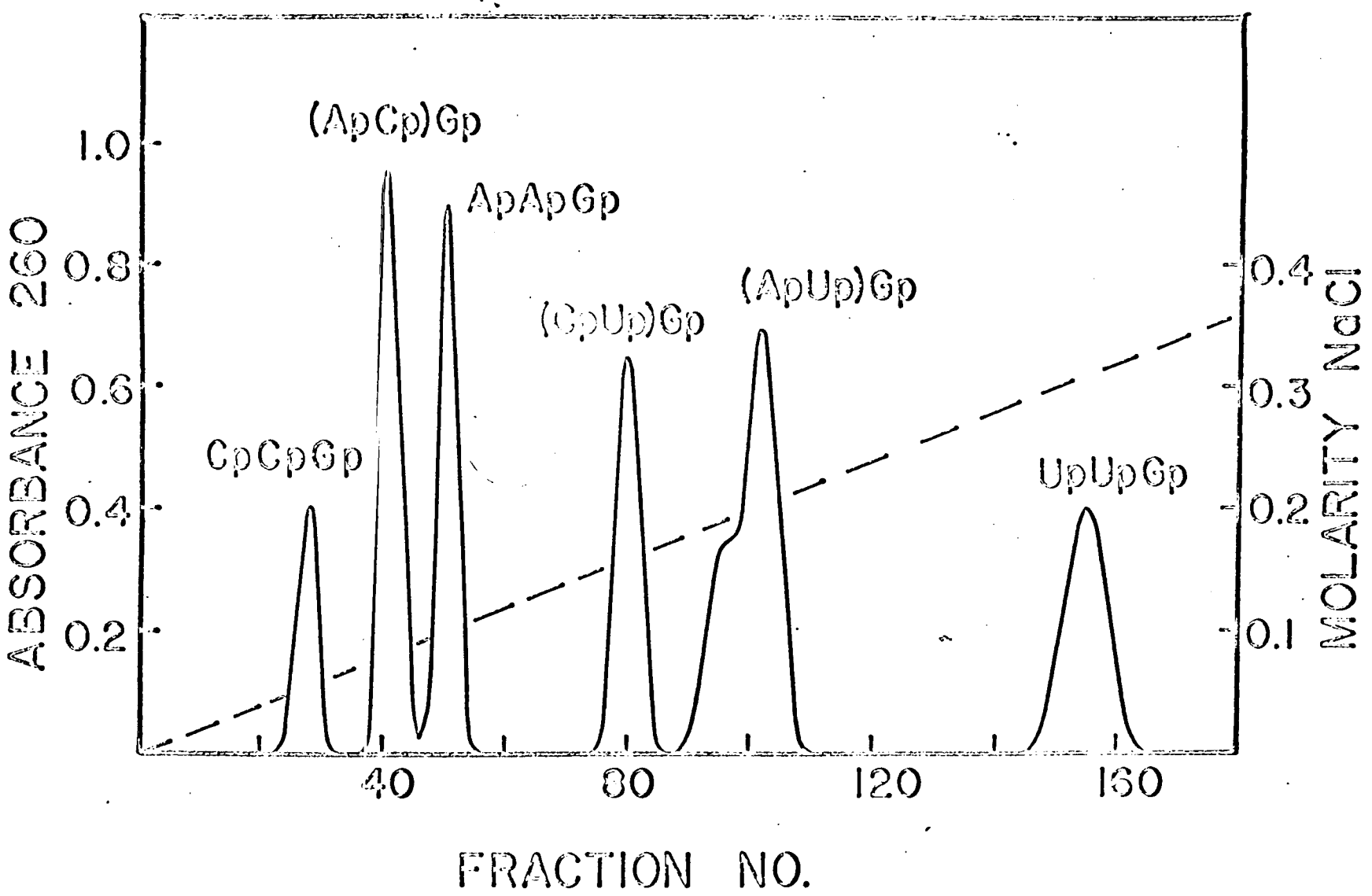

ABSORBANCE 260
0.8
0.6
0.4
0.2
CpGp
ApGp
UpGp
0.4
0.3
0.2
0.1
MOLARITY NaCl
20
60
100
FRACTION NO.

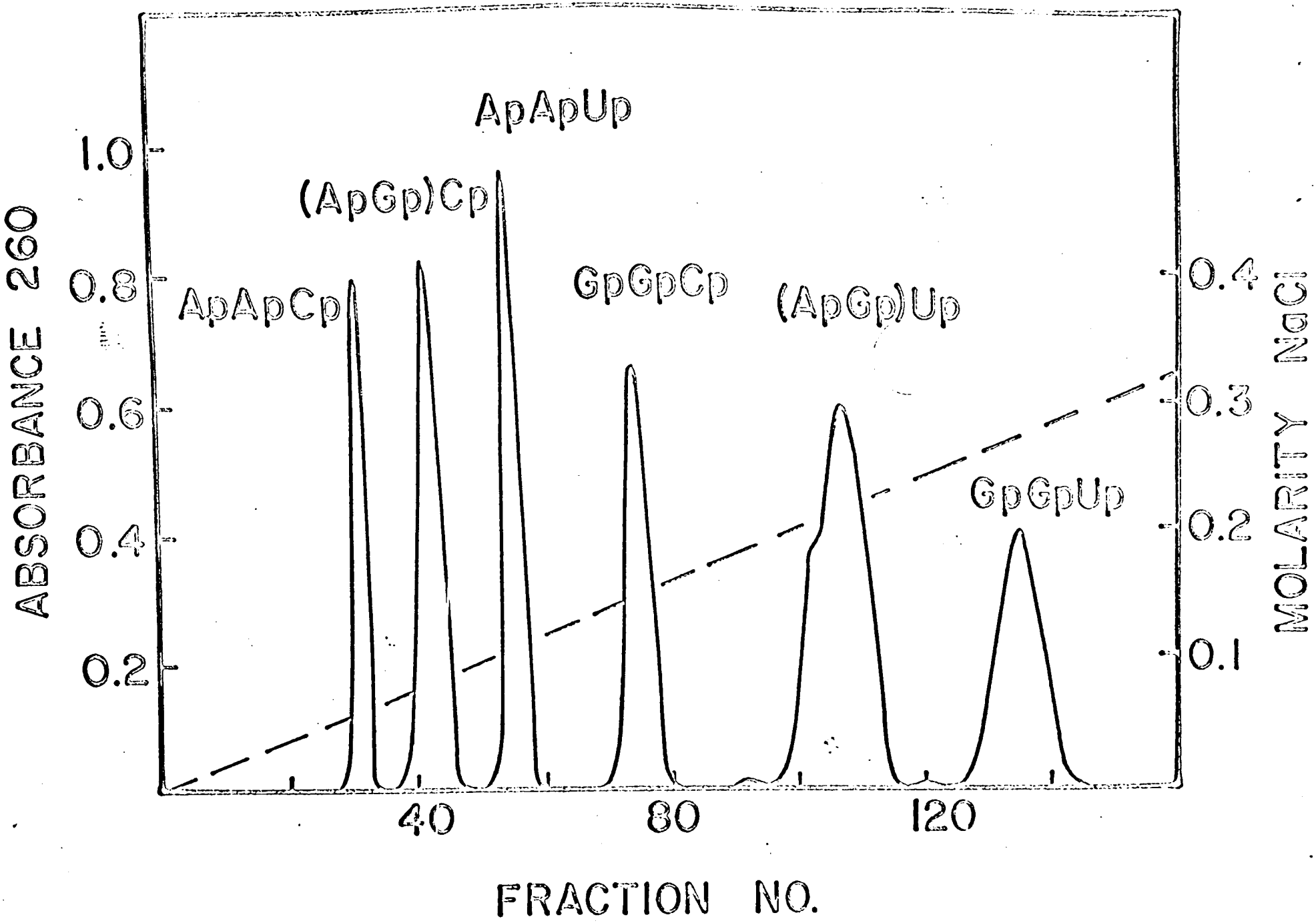

ABSORBANCE 260
0.8
0.6
0.4
0.2
ApC
GpC
ApU
GpU
0.01
0.001
NORMALITY OF HCl
40
80
120
FRACTION NO.

TRIMER DEVIATIONS

Trimer	pH 1-6 SSQ	pH 7 SSQ	pH 11 SSQ	$\sqrt{TSSQ_1}$	$\sqrt{TSSQ_2}$	$\frac{\sqrt{TSSQ_2}}{\sqrt{TSSQ_1}}$
AAC	0.026	0.015	0.035	0.277	0.817	2.95
AAG	0.013	0.009	0.021	0.206	0.760	3.69
AAU	0.018	0.017	0.044	0.281	0.635	2.26
(AG)C	0.026	0.033	0.007	0.355	1.129	3.18
CCG	0.096	0.069	0.069	0.485	1.422	2.93
(CU)G	0.129	0.077	0.072	0.528	1.283	2.43
GGC	0.048	0.051	0.073	0.415	1.155	2.78
GGU	0.043	0.054	0.082	0.423	0.786	1.86
(AU)G	0.059	0.058	0.054	0.415	0.827	1.99
UUG	0.046	0.038	0.070	0.392	1.181	3.01
(AG)U	0.014	0.019	0.073	0.326	0.622	1.91
Average	0.047	0.040	0.060	0.373		

Table II. Column headings are explained in Table I. Average error $= \frac{\sqrt{TSSQ_1}}{\sqrt{(3)\ (96)}} = 0.022$ O.D.


(ApCp)Gp
ApApGp
(CpUp)Gp
(ApUp)Gp
CpCpGp
UpUpGp
ABSORBANCE 260
1.0
0.8
0.6
0.4
0.2
0.4
0.3
0.2
0.1
MOLARITY NaCl
40
80
120
160
FRACTION NO.


ApApUp
(ApGp)Cp
ApApCp
GpGpCp
(ApGp)Up
GpGpUp
1.0
0.8
0.6
0.4
0.2
ABSORBANCE 260
0.4
0.3
0.2
0.1
MOLARITY NaCl
40
80
120
FRACTION NO.

DIMER DEVIATIONS

Dimer	pH 1.6 SSQ*	pH 7 SSQ	pH 11 SSQ	$\sqrt{TSSQ_1}$**	$\sqrt{TSSQ_2}$***	$\frac{\sqrt{TSSQ_2}}{\sqrt{TSSQ_1}}$
AA	0.067	0.028	0.023	0.343	1.123	3.27
CA	0.023	0.039	0.022	0.291	1.869	6.42
AG	0.036	0.022	0.041	0.315	1.452	4.61
AU	0.010	0.019	0.022	0.226	0.967	4.28
CC	0.038	0.008	0.007	0.230	5.403	23.5
CG	0.057	0.021	0.172	0.500	1.825	3.65
CU	0.042	0.025	0.071	0.372	1.954	5.25
UG	0.020	0.020	0.009	0.219	1.389	6.34
UU	0.001	0.015	0.046	0.250	1.602	6.41
Average	0.032	0.022	0.046	0.305		

Table 1. *Is the sum of the squares of the deviations at the pH indicated in each column.

**Is the square root of the square deviations totalled over the three pHs.

***Is the square root of the squared deviations for the dimer that gave the second best fit.

Average error $= \frac{\sqrt{TSSQ_1}}{\sqrt{(3)\ (96)}} = 0.018$ O.D.

TRIMER DEVIATIONS

Trimer	pH 1-6 SSQ	pH 7 SSQ	pH 11 SSQ	$\sqrt{TSSQ_1}$	$\sqrt{TSSQ_2}$	$\frac{\sqrt{TSSQ_2}}{\sqrt{TSSQ_1}}$
AAC	0.026	0.015	0.035	0.277	0.817	2.95
AAG	0.013	0.009	0.021	0.206	0.760	3.69
AAU	0.018	0.017	0.044	0.281	0.635	2.26
(AG)C	0.026	0.033	0.007	0.355	1.129	3.18
CCG	0.096	0.069	0.069	0.485	1.422	2.93
(CU)G	0.129	0.077	0.072	0.528	1.283	2.43
GGC	0.048	0.051	0.073	0.415	1.155	2.78
GGU	0.043	0.054	0.082	0.423	0.786	1.86
(AU)G	0.059	0.058	0.054	0.415	0.827	1.99
UUG	0.046	0.038	0.070	0.392	1.181	3.01
(AG)U	0.014	0.019	0.073	0.326	0.622	1.91
Average	0.047	0.040	0.060	0.373		

Table 11. Column headings are explained in Table 1. Average error $= \frac{\sqrt{TSSQ_1}}{\sqrt{(3)\ (96)}} = 0.022$ O.D.

POLYMER DEVIATIONS

Polymer	pH 7 SSQ	pH 11 SSQ	$\sqrt{TSSQ}$	pH 7 SSQ	pH 11 SSQ	$\sqrt{TSSQ}$
poly-A	0.113	0.102	0.464			
poly-C	0.286	0.232	0.719			
poly-U	0.036	1.811	1.359			
TMV-RNA[21]	0.036	0.128	0.406	0.044	0.174	0.467
TMV-RNA[20]	0.039	0.115	0.392	0.048	0.157	0.452
TMV-RNA[19]	0.048	0.142	0.436	0.057	0.191	0.499
TMV-RNA[4]	0.071	0.153	0.473	0.082	0.204	0.534

Table III. The homopolymer values represent comparisons between the polymer spectra and the monomer spectra. The TMV-RNA values were calculated according to the base compositions given in each of the references cited, and compared with experimental curves obtained from intact TMV-RNA.