ANALYTICAL BIOCHEMISTRY 201, 17–29 (1992)

A Semi-Micromethod for the Determination of the Extinction Coefficients of Duplex and Single-Stranded DNA[1]

Jeffrey T. Henderson,* Albert S. Benight,† and Sue Hanlon*,[2]
*Department of Biochemistry and †Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60612

Received May 29, 1991

We have developed a rapid and convenient procedure for the determination of concentrations and extinction coefficients of oligo- and polynucleotides. It offers significant advantages over other methods in terms of precision and the ability to detect artifactual or erroneous results. Samples are first completely digested with appropriate enzymes to mononucleotides and nucleosides. Using the multicomponent linear regression capabilities of commonly available spreadsheet programs, the absorbance spectrum of the digest can be analyzed as a linear combination of the contribution of the possible constituent monomers. If all the spectral components present have been included, the analysis yields the concentration of each of the monomer species whose sum is the concentration (in monomer units) of the original undigested sample. When combined with the predigest absorbance spectrum, the extinction coefficients of the intact sample can then be calculated. The analysis also yields the fractional base composition of the oligomer or polymer. The extensive spectral data provided by digital readouts of modern spectrophotometers permit the application of sensitive tests of the goodness of fit, thus facilitating the detection of artifacts and sample inhomogeneity. Both single-strand and duplex structures can be analyzed comfortably in sample sizes of 25 to 35 nmol (total) of mononucleotides with a precision of 1%. The concentrations obtained by this method agree, on the average, within 0.2% with those determined by phosphate analysis of the same sample. The method also yields the base composition with an accuracy of ca. 5% for high-molecular-weight polymers and 2% for short oligomers (15–20 bp) when compared to the predicted values.

The advent of rapid procedures for large-scale synthesis and purification of DNA oligomers has made possible a number of biophysical studies with these compounds. Such studies usually require a concentration of the species more accurate than the concentration that can be calculated from an absorbance and a "rule of thumb" extinction coefficient or that calculated for nearest neighbor contributions at some fixed wavelength. It is thus desirable to have a versatile semi-microexperimental procedure for the accurate estimation of concentrations of oligomeric and polymeric forms in terms of monomer units. Data gathering and processing capabilities of the modern laboratory have made it feasible to develop a variation of an approach that has been frequently employed in the nucleic acid field for the estimation of polymer concentration in terms of the concentration of its constituent monomers. In the procedures currently in use, a polymer or oligomer is degraded to its constituent monomers with appropriate reagents. The absorbance per unit path of the digest at a characteristic wavelength is used, together with the absorbance of the predigest spectrum and the number average extinction coefficient of the monomers at the same wavelength, to solve for the concentration and/or the extinction coefficient of the intact polymer/oligomer. (See, for example, Pohl and Jovin (1) and Nadeau and Gilham (2)). Because these procedures utilize spectral data from only one wavelength, the results are sensitive to a variety of sample and instrumental artifacts compared to those obtained from methods relying on direct chemical or weight analyses. In general, accuracy suffers. This approach has the additional dis-

[1] Taken in part from the Ph.D. thesis of J.T.H., University of Illinois at Chicago.

[2] To whom reprint requests should be addressed at Department Biochemistry MC 536, University of Illinois at Chicago, Chicago, IL 60612.

0003-2697/92 $3.00

advantage of relying on the assumed fractional composition of the sample in question.

A given spectrum of the digested sample, however, provides a wealth of data in terms of absorbance vs wavelength. This is particularly true for those spectra collected by modern spectrophotometers with digital readout. When analyzed appropriately, these data can be readily used to solve for both fractional base content and monomer concentration. Furthermore, the excess data readily permit the detection of artifacts and errors that an analysis based on a single wavelength value of the absorbance will miss. In this paper, we describe such an analysis for evaluating the extinction coefficients and base compositions of both duplex and single-stranded deoxyoligo- and deoxypolynucleotides that takes advantage of the full spectral display of nuclease digests and a reference set of nucleotides. The accuracy of the method is equal to that of the classical phosphate method, but is inherently more precise and convenient to use.

MATERIALS AND METHODS

Commercial Products

Calf thymus DNA, poly(dGdC), and poly(dIdC) were the highest purity preparations obtained from Sigma. The 5′ deoxynucleotide monophosphates (A, G, T, and C) were also the highest purity preparations available from Fisher. Pancreatic DNase was obtained from Sigma, snake venom phosphodiesterase from Cooper Biochemicals, and P1 nuclease (from *Penicillium citrinum*) from Boehringer-Mannheim. All these substances were prepared as concentrated stock solutions and stored in several small frozen aliquots at −20°C until used. The polynucleotides were dissolved at a concentration of ca. 1 mg/ml and dialyzed against several changes of 200 mM NaCl, 10 mM EDTA, pH 7, followed by exhaustive dialysis against 20 mM NaCl, 10 mM Tris, pH 7.5, in order to remove extraneous uv-absorbing material. All chemicals were reagent-grade quality.

M13 Isolation

Single-stranded, closed circular bacteriophage M13-mp19 DNA was isolated in a single preparation, by infection of *Escherichia coli* strain JM101 (New England Biolabs, Beverly, MA) using standard procedures (3,4) as described in Wilson *et al.* (5). Homogeneity and purity of the sample were verified by electrophoresis on native and alkaline agarose gels.

Oligonucleotides

Oligonucleotides whose sequences are shown in Table 1 were synthesized on an Applied Biosystems Model 380B DNA Synthesizer, using β-cyanoethylphosphoramidite derivatives and reagents obtained from Glen Research Corp. After detritylation and deblocking, the products were dialyzed extensively in 2K cutoff tubing against a buffer of 200 mM NaCl, 1 mM EDTA, and 10 mM Tris, pH 7.5. In the final change of solvent, the salt content was decreased to 100 mM NaCl. These stock solutions, at a concentration of ca. 2 mM (in monomer units), were stored at −20°C in several aliquots.

TABLE 1

Sequence of Synthetic Oligomers Analyzed in These Studies

Name	Sequence
MRE I	5′-CCTTTGCGCCCGGCC-3′
MRE I′	5′-GGCCGGGCGCAAAGG-3′
INV II	5′-CCGGCCCGCGTTTCC-3′
INV II′	5′-GGAAACGCGGGCCGG-3′
RAN III	5′-CCGGCGACGACTGCC-3′
RAN III′	5′-GGCAGTCGTCGCCGG-3′
$(dTGG)_5$	5′-TGGTGGTGGTGGTGG-3′
$(dCCA)_5$	5′-CCACCACCACCACCA-3′
$(dTTG)_5$	5′-TTGTTGTTGTTGTTG-3′
$(dCAA)_5$	5′-CAACAACAACAACAA-3′
$(dGA)_8$	5′-GAGAGAGAGAGAGAGA-3′
$(dCG)_3$	5′-CGCGCG-3′
$(dA)_{16}$	5′-AAAAAAAAAAAAAAAA-3′
$(dC)_{16}$	5′-CCCCCCCCCCCCCCCC-3′
$(dT)_{16}$	5′-TTTTTTTTTTTTTTTT-3′

Oligonucleotide Homogeneity

The oligomers were end labeled with [γ-^{32}P]dATP using T4 polynucleotide kinase (6) and electrophoresed on a 20% polyacrylamide gel in tetrabromoethane buffer (pH 8.0) at 45°C. With the exception of INV II′ (see Table 1 for sequence), all samples exhibited less than 2% lower molecular weight products representing prematurely terminated sequences. This was confirmed by HPLC chromatography on a molecular sieve column, Zorbax GF-250 (DuPont), using an isocratic gradient of 100 mM NaCl, 10 mM Tris, pH 8.05, at room temperature. INV II′ had a sizable fraction (ca. 15%) of lower molecular weight material that was removed by preparative gel electrophoresis followed by elution on a hydroxyapatite column.

In addition to the lower molecular weight material, the HPLC profiles of the repetitive trimer oligonucleotides all showed variable amounts of higher molecular weight products, suggesting self-association in the concentrated stock solutions. Precautions were taken to irreversibly dissociate these aggregates by heating the more dilute sample solutions (65°C for 5 min) before analysis.

Digestion Conditions

Different sets of enzymes and digestion conditions were employed for duplex and single-stranded or dena-

tured substrates. Both sets, however, yield 5′ mononucleotides and a terminal nucleoside, if the 5′ end is not phosphorylated. Pancreatic DNase I and snake venom phosphodiesterase were used for polynucleotides at temperatures at which the duplex form was stable. P1 nuclease was employed for single-strand oligomers and polymers under conditions of duplex instability. P1 nuclease has several advantages for these studies: (i) It is active under conditions (pH 7 in 100 μM $ZnCl_2$) that do not significantly perturb the absorbance spectra of the nucleotides. (ii) It shows little sequence bias. (iii) The exceptional heat stability of this enzyme permits its use with a variety of duplex or hairpin substrates near their melting temperatures. Although the digestions were usually allowed to proceed for a longer time, digestion conditions were selected for both sets of enzymes to yield fully digested sample within 5 h of initial addition, as evaluated by electrophoretic and/or chromatographic methods. The standard procedures that we followed are given below.

(a) Duplex substrates. After an initial spectrum of the DNA sample in 300 mM Tris, pH 7.9, was obtained, aliquots of stock solutions of $CaCl_2$, $MgCl_2$, pancreatic DNase, and snake venom phosphodiesterase (pretreated to remove contaminating phosphatases) were added to 3.00-ml volumes of duplex DNA and polynucleotides at ca. 70 μM in nucleotide concentration to bring the salt concentration to 2 mM $CaCl_2$ and 8 mM $MgCl_2$ and a final concentration of 10 μg/ml for the DNase and 15 units/ml of the phosphodiesterase. Equivalent additions were made to a spectrophotometric reference solution and the tightly stoppered cuvettes were incubated at 37°C. Spectra were obtained periodically until no further hyperchromic increase due to digestion was noted (generally, 5 h). At the end of 24 h, an additional aliquot of the two enzymes was added to both sample and reference solutions and the sample solution was checked for additional hyperchromic increase. If no further increase occurred, the spectra were analyzed as described under Analysis. Separation and careful quantitation of the digestion products of a few samples of the native calf thymus DNA digested in this manner revealed that the digestion procedure resulted in a 98 ± 3% yield of 5′ nucleotides.

(b) Single-strand and denatured substrates. As in the procedure above, an initial spectrum of a sample at ca. 70 μM in a 0.5-ml volume of 100 mM NaCl, 10 mM Tris, pH 7, was obtained, after which appropriate aliquots of concentrated stock solutions of $ZnCl_2$ and P1 nuclease were added, sequentially, to bring the concentrations to 100 μM for the former and 10 units/ml for the latter. (The volume change caused by the reagent additions amounted to no more than 2%). A matching reference solution containing all components except the oligo- or polynucleotide was also prepared and used for the spectral baseline. For purposes of monitoring the activity of different batches of enzyme, spectral changes at the wavelength maximum were sometimes followed for 5 h at 25°C. At the end of this time, a full spectrum of the digest was obtained and sample and reference solutions were then removed from the spectrophotometer, placed in tightly capped Eppendorf tubes, and incubated at 45°C overnight.

In those instances where kinetic data were not acquired, both sample and reference were routinely incubated at 45°C immediately after enzyme addition to the sample. (In the case of the calf thymus DNA digestions with P1 nuclease, however, the incubation temperature was at 65°C in order to destabilize the duplex structure). After 12 to 24 h, the sample solution(s) was removed from the incubator and a final spectrum was collected against the appropriate reference solution which had been subjected to a similar incubation procedure.

Completeness of digestion under these conditions was indicated by the fact that (i) additional aliquots of enzyme added to the sample generally resulted in no further hyperchromic increase and (ii) there was no evidence of any oligomeric products on 20% polyacrylamide gels. When kinetic data were obtained, the 5-h spectrum was usually identical to the overnight digest. The latter result substantiates the fact that no significant evaporation of solvent occurs over the course of the incubation period. Incubation of calf thymus DNA at 65°C gave similar results.

Phosphate Determinations

Concentrations of selected samples and nucleotides were also determined by phosphate analysis using two variations of the colorometric method of Fiske and Subbarow (7). The duplex polynucleotide samples were assayed by a procedure of Ames and Dubin (8). The intact oligomers, oligomer digests, and mononucleotide standards were assayed by the procedure of Chen *et al.* (9). The standard curves were generated using dilutions of a 1.000 mM KH_2HPO_4 solution prepared from the dried salt. A secondary standard of 5′-AMP that gave a standard curve indistinguishable from the inorganic phosphate standard was also employed. In some cases, the P1 nuclease digests were used for phosphate analysis. Because such solutions contained $ZnCl_2$, the standard phosphate solution contained an equivalent amount of $ZnCl_2$ found in the digest samples.

Spectrophotometry

Both a double-beam Cary 14 and a single-beam Beckman DU 40 spectrophotometer were employed in this work. Spectral data obtained from the Cary 14 were read manually at 2.5-nm intervals as it had no data-capturing capabilities. They were processed by a regression program for an IBM mainframe that we had previously

developed (10). The DU 40 instrument had a serial port and was interfaced with an IBM AT personal computer using data capture software developed by Beckman Instruments. Spectral data were acquired at 0.5-nm intervals and stored in Beckman standard format. For analysis, they were converted to Lotus 1-2-3 format by the data capture program and analyzed by the multicomponent linear regression subprogram of the Lotus 1-2-3 Release 2 software package.

Spectra were obtained at 25°C from 220 to 360 nm in 1.00 ± 0.001-cm path quartz cuvettes. The DU 40 experiments routinely employed microcuvettes that could be used with 0.5-ml volumes. The geometry of the Cary 14 instrument was better suited to the standard 3-ml volume cuvettes. Temperature was monitored by a Yellowsprings Instrument Co. thermistor and bridge assembly and did not fluctuate more than ±0.5°C for any given experimental set. In the region of the nucleotide absorption bands, data in the range of 230 to 290 nm were routinely utilized for the concentration analyses. The additional data acquired above 310 nm in a region in which the nucleotides did not absorb permitted the detection—and correction, if necessary—of extraneous light scattering by the solution under investigation.

ANALYSIS

Mathematical Procedures

The principle of our analysis is based on the assumption commonly used in other spectral techniques that the observed absorbance, $A_{\lambda j}$, of a given path, L, at a given wavelength, λ_j, is a linear sum of the contributions of the i constituents of the solution. In a mixture of mononucleotides and nucleosides resulting from complete digestion by nucleases or chemical reagents, these contributions are a function of the monomer concentrations, c_i, and their extinction coefficients, $E^i_{\lambda j}$, when the reagent blank contributions are subtracted. That is,

$$[A_{\lambda j}/L] = \sum_i c_i E^i_{\lambda j}. \qquad [1]$$

This basic equation can be used to calculate concentration of the intact sample in terms of monomer concentration, after correction for minor dilution factors, as long as the implicit assumptions are valid. That is, the sample solution exhibits no uncompensated spectral component and there are no interactions between the spectral components that have been accounted for in the spectral range of interest. When combined with the predigest spectrum, extinction coefficients of the intact sample, which are more convenient for repeated estimates of concentration, can be obtained.

The sequence of operations in this procedure is as follows: A spectrum of an intact sample is obtained prior to digestion in a standard solvent. The same (preferably) or an identical dilution of a stock solution of the oligo- or polynucleotide is then subjected to digestion by the appropriate combination of nucleases or reagents. After digestion is complete, the product spectrum is read against a blank containing the same concentration of enzymes and all other digest components except the sample. The spectra of dilutions of standard stock solutions of the constituent nucleotides and nucleosides are read in the same digest buffer with the same cuvette and spectrophotometer. These latter spectra are converted to extinction coefficients throughout the wavelength range of analysis, using concentrations of the mononucleotides evaluated from their absorbance at the wavelength maximum, λ_{max}, of the absorption spectrum and their known extinction coefficients at this position. (This extra step avoids the problems of a miscalibrated monochromator). If the digest buffer is acidic or contains appreciable concentrations of a perturbing ion such as Zn, it may be necessary to determine concentrations of the mononucleotide/nucleoside standard solutions by performing identical dilutions (in triplicate) into a standard buffer for which extinction coefficients are known.

One thus has an $(i + 1) \times j$ matrix of data consisting of observed absorbances of the digest and extinction coefficients of the i nucleotide/nucleoside components in the same digest buffer at every wavelength, λ_j. This matrix can be analyzed as a multicomponent linear regression matrix, with or without an intercept, with no biased input of the fractional content of nucleotides/nucleosides. The independent variables (x) for the analysis are the values of $E^i_{\lambda j}$ for each of the i components. The dependent variables (y) are the values of $A_{\lambda j}$. The analysis will yield the best fitting coefficients, which should be the concentrations, c_i, of each of the i nucleotides/nucleosides. If all of the absorbing components have been accounted for, the sum of the coefficients will be the total concentration, C_0. After correction for any dilutions caused by reagent addition, this concentration is used together with the absorbance spectrum of the predigested oligomer or polymer to yield the extinction coefficients of the polymer. The ratio of the coefficients of the x variable will yield the base ratios and the value of $[c_i/C_0]$ will be the fractional base content.

In the above discussion, we have described the most general case. In practice, however, the independent variables for most nuclease and alkaline digestions of RNA and DNA polymers need be represented only by the mononucleotide content. Within the error of the analysis, the spectral properties of the nucleotides are essentially identical to those of the mononucleotides in the spectral range of interest. Furthermore, the position of the phosphate group or the identity of the pentose does not significantly alter the absorption properties of the nucleotides (see Table 2). Thus, the number of i components employed in the analysis can be reduced to four

TABLE 2

Extinction Coefficients of 5′ Mononucleotides and Mononucleosides

Nucleotide	λ_{max}	Extinction coefficients (mM^{-1} cm^{-1}) at 25°C, pH 7.0	
		Literature[a]	This work (±0.1)
5′-dAMP	259	15.3	15.4
5′-AMP	259	15.4	
Adenosine	259	15.4	
5′-dGMP	252	13.5	13.7
5′-GMP	252	13.7	
Guanosine	252	13.7	
5′-dCMP	271	9.3	9.17
5′-CMP	271	9.0	
Cytidine	271	8.9	
5′-TMP	268	9.8	9.80
Thymidine	268	9.7	

[a] Averages of values taken from Refs. (12–14).

independent variables if the sample contains no modified bases. When the number of data points is not extensive, complementary duplex polymers can be restricted even further to two components. For those samples that are either known to or suspected of containing modified bases, obviously the reference basis set of extinction coefficients must be expanded to included the modified nucleotide/nucleosides.

When the samples's base content is known and it is homogeneous, the monomer concentration can, in principle, be obtained more simply by calculating the concentration from a number average extinction coefficient, $E^n_{\lambda j}$, where the latter is $\sum_i f_i E^i_{\lambda j}$. f_i is the mole fraction of each of the i nucleotides. A simple one-component linear regression analysis of the equation

$$[A_{\lambda j}/L] = C_0 E^n_{\lambda j} \qquad [2]$$

can then be run over a suitable wavelength range. The total concentration, C_0, will be the slope function. Reliance on this approach, however, is not advisable as the concentrations obtained will be very sensitive to shifts in baseline position, inhomogeneity in base composition, and other systematic errors. In general, we recommend the use of the regression analysis, appropriate for Eq. [1], for the calculation of total nucleotide concentration rather than the simpler procedure based on Eq. [2]. As is shown, however, the rearrangement of Eq. [2] to calculate C_0 at every wavelength in the analysis range can be used to advantage as a sensitive test of the absence of these problems.

Major Sources of Error

A variety of systematic errors that universally plague methods utilizing digest data can also create problems in this procedure. Some are obvious and easily detectable, such as incomplete digestion of the sample or the evaporation of water at the elevated temperatures and long incubation times in some of the P1 digestions. Others are more subtle yet equally serious. In contrast to other methods, however, the extensive data collected in these measurements almost always permit the detection of such errors. When present, significant systematic variations are usually seen in the concentration or base composition when different spectral windows are employed in the analysis. Performance tests described in another section also show deviant behavior and the patterns frequently permit a diagnosis of the source. Frequently, the data also permit their minimization or rectification as well. The most common of these systematic errors is discussed below.

(a) Components not included in the analysis that give rise to nonuniform absorption in the wavelength region of analysis. A common source of this problem is the presence in the sample solution of small uncompensated concentrations of EDTA, which has a strong absorption below 260 nm. This compound is a frequent contaminant of many polymeric and oligomeric preparations as it is difficult to remove by dialysis, especially at low ionic strengths. Since its concentration is unknown, its contribution is difficult to compensate in the solvent blank. Its contribution can be minimized by discarding the spectral data below 260 nm.

Light scattering in the sample solution can also distort the spectrum of the digestion mixture. The presence of scattering is usually evidenced by the presence of signal above 320 nm, where the nucleotides should exhibit no absorbances. As long as the scattering is not due to the formation of microaggregates of nucleotides in the digest, the data can be corrected prior to analysis by the method of Oster (11). In this treatment, the exponential wavelength dependence of the absorbance due to turbidity, $A^S_{\lambda j}$, under the instrumental and experimental conditions, is expressed empirically as

$$\log A^S_{\lambda j} = Q + m \log \lambda_j. \qquad [3]$$

The slope and intercept, m and Q, in Eq. [3] are evaluated by a linear regression of the values of $\log A^S_{\lambda j}$, between 330 and 400 nm. The corrected values of the absorbances, $A^C_{\lambda j}$, in the wavelength range of the nucleotide absorption band can then be calculated as the difference $(A_{\lambda j} - A^S_{\lambda j})$. If the scattering is due to column fines, dust, or extracted polymeric gel components, its wavelength dependence is small and it can be more simply corrected by the incorporation of a constant in the regression analysis.

A third common source of this type of error is the unexpected presence of modified nucleotides/nucleosides in the digest. Such modifications can arise either in the oligonucleotide synthetic procedure or by the action of enzymes and/or the digest conditions on the oligomer or polymer. This is the more serious of problems in this class because, although it can be readily detected, it is impossible to correct without extensive effort to separate and identify the unsuspected modifications.

(b) Errors of an instrumental origin. These errors afflict all spectrophotometric assays. They include those due to (i) monochromator and cell misalignment or path variations, (ii) baseline drift, and (iii) spectral noise. Each of these is discussed below.

Differential path lengths and/or wavelength shifts in the sample and reference spectra due to systematic errors in the wavelength calibration can be avoided completely by simply reading each member of the spectral set (i.e., that set that includes the spectra of the mononucleotide standards, the predigested sample, and the digested sample) in the same cuvette on the same instrument. The wavelength shift is generally the more significant of these errors and, occasionally, arises in cheaper spectrophotometers because of variability of mechanical positioning of the monochromator in repetitive scans. This type of mechanical artifact is not as crucial for the estimation of total concentration of nucleotides in the digest, since a small constant incremental shift in wavelength position from instrument to instrument usually cancels out when data are taken over a wide wavelength range centered about the spectral maximum of the digest. Such wavelengths shifts, however, will significantly affect the accuracy of the base composition derived from the analysis.

Occasionally electronic or mechanical effects (drift, misaligned cells, etc.) will give rise to absorbance offsets. These shifts are usually uniform in the wavelength of interest and can be compensated by including a constant in Equations [1] and [2]. If the goodness of fit does not improve with the incorporation of a constant, then it should not be included. The calculation of an additional parameter, if unnecessary, will usually lower the accuracy of the determinations of the concentration and base composition.

The spectral noise problem can best be handled by restricting the range of analysis to one in which the noise represents no more than 5 to 10% of the observed absorbance signal. For the nucleotide mixtures at the concentrations examined in this present analysis, this means a range of analysis of 230 to 280 nm.

(c) Inaccurate extinction coefficients of the reference nucleotide/nucleoside components. This is probably the most serious problem and represents the true limitations of the analysis. This is the inherent error in the values of the extinction coefficients of the reference nucleotides at their maximum in a standard solvent. As described, these values are used to generate the extinction coefficients of the reference nucleotides in the digest buffer. The accuracy of the concentration analysis depends crucially on the accuracy of these values. If the digest buffer has a significant effect on the nucleotide absorption spectra, then this error is compounded by the demand for dilution accuracy from the nucleotide stock solutions into the digest buffer. This, together with the various other dilution errors due to miscalibrated pipettes, is probably the major source of systematic error in most determinations. It should be noted, however, that since the initial and final spectra can be obtained on the same sample solution with little or no volume dilution, initial errors in the dilution of the concentrated stock solutions of oligo- or polynucleotides into the digest buffer are usually irrelevant with respect to the determination of the extinction coefficients of the sample.

Tests for the Presence of Errors Described Above

Because of the wealth of data available from the digests and the reference solutions, the usual criteria of "goodness of fit," the regression correlation coefficient and the F test, were essentially useless as tests of performance. We therefore developed more sensitive tests for the detection of aberrations arising from the systematic errors described above. Three of these are described below. Two are convenient graphical displays and one is a standard deviation function. The production of these graphs and tabular values can be readily automated using the programming script of commonly available spreadsheet software, thus enhancing the convenience of the method.

(a) Absorbance ratio plots. A sensitive indicator of the presence of all but two of the systematic errors discussed above is the shape and appearance of a plot of observed absorbance, A^{OBS}, divided by the calculated absorbance, A^{CALC}, over the spectral range of data acquisition. In this test, A^{CALC} is the best fitting absorbance at a given wavelength, calculated with the coefficients and intercept (if any) obtained in the regression analysis of Eq. [1] for a given spectral window. If only nucleotide components are present in the digest and the reference spectral set is inclusive, then this plot should have a slope of 0 (i.e., wavelength invariant) and an intercept of 1. Deviations of the plot characteristics from these values coupled with a systematic variation with wavelength and a dependence on the wavelength window of analysis indicate that one or more of the problems discussed in the previous section are present. Strategies for their remedy can be attempted as described.

(b) C_0 test. If the sample does not have the expected sequence or fractional base content, the absorbance ratio test will not detect this fact as long as the nucleotide

identities are included in the reference spectra. Nor will it detect errors arising from systematic errors in the extinction coefficients of one or more of the mononucleotide standards. When the base composition is known, however, heterogeneity or the use of a deviant set of mononucleotide extinction coefficients is suggested if the coefficients of the independent variables yield base compositions that do not agree, within the estimated experimental error, with the predicted values.

An informative graphical display of the presence of errors of this type is the application of a rearrangement of Eq. [2] to calculate C_0 at every wavelength over the same range as that of the first regression test described above. When analyzed with an appropriate set of reference spectra, the spectrum of a homogeneous sample that contains the nucleotide components in the predicted ratio should yield a wavelength-invariant value of C_0. (If a constant, however, was required for better fit of the regression analysis of Eq. [1], the values of A^{OBS} used to calculate C_0 should be corrected for this constant, which represents a consistent baseline shift). If the plot of C_0 vs λ_j does show systematic deviations that are not observed in the absorbance ratio test, then either the sample does not have the expected nucleotide content or one or more of the mononucleotide extinction coefficient sets is in error. Because of its sensitivity, this test is useful only if the sample passes the less sensitive absorbance ratio test. If the sample fails the latter test (i.e., shows systematic deviations), it will also fail the C_0 test.

(c) Standard deviation functions. In addition to the graphical displays, standard deviations and standard error functions are also useful for providing a single numerical value that permits an immediate assessment of poor performance in the analysis. One of the more useful of such functions, SD A, is defined as

$$\text{SD } A = \{[(f^P - f^O)^2]/n\}^{1/2}, \qquad [4]$$

where f^P and f^O represent the predicted and the observed fractional nucleotide content, respectively, and n is the number of reference components included in the analysis.

Application of Diagnostic Tests to Artificial Data

In order to assess the effect that common instrumental artifacts such as noise and small baseline shifts would introduce, artificial digest data were generated using the sequence appropriate for the RAN III oligomer, at a concentration of 61.5 μM. Two levels of random noise with boundaries of up to ± 0.003 and ± 0.01 in absorbance and a systematic baseline shift of $+0.005$ were introduced into the calculated perfect data, which were generated by combining the extinction coefficients of the reference set of nucleotides determined from the data in Table 2 in the proportions appropriate for RAN III. The spectral range 240 to 280 nm was used in order to restrict the noise level to 5% or less of the signal. Failure to include an intercept in the regression analyses in order to accomodate the baseline shift led to values of C_0 of 62.1 μM at either level of random noise. Inclusion of an intercept in the analysis gave a value of 61.3 μM. SD A values ranged from 0.001 to 0.01 for such data, with the higher level corresponding to the larger value of the random noise increment and those regressions run without an intercept included.

The effects of these artifacts on the absorbance ratio plot and the C_0 plots are shown in Figs. 1A and 1B. Except for the data above 280 nm, where the noise levels were a significant fraction of the signal, the absorbance ratio plot is not dramatically affected for this degree of baseline shift and noise. The C_0 plot has a slight systematic curvature that is corrected by the incorporation of an intercept in the regression of Eq. [1]. Even without the correction, however, the value of C_0 calculated from the regression of Eq. [1] is only 1% higher than the true value. It is clear that errors of this magnitude will not have a dramatic effect on the accuracy of the results.

EXPERIMENTAL RESULTS AND DISCUSSION

Nucleotide Standards

A survey of the reported literature values for extinction coefficients at the absorbance maxima of the common unmodified 5′ deoxy- and ribonucleotides as well as matching nucleosides at pH 7 in a nonperturbing solvent (100 mM NaCl) is presented in Table 2. Our determinations for the individual 5′ deoxynucleotides by the same phosphate analysis as that used for the DNA and oligonucleotide species are reported in the last column in Table 2. These experimentally determined values were used for the estimates of concentrations and the generation of the reference spectra of the deoxynucleotide standards in our spectral analyses.

The oligomers examined in these studies have no terminal 5′ phosphate residue and thus the products of P1 digestion will result in a nucleoside, in addition to the 5′ mononucleotides. Nucleosides can also arise by the spontaneous hydrolysis of the nucleotides. As discussed, they were not considered separate components in our analysis. The entries in Table 2 demonstrate that their spectral properties at pH 7 are essentially identical to those of the corresponding mononucleotides.

Oligonucleotide Digestions

The several single-strand oligomers examined in this study, shown in Table 1, represented various levels of sequence complexity. MRE I, MRE I′, INV II, INV II′, RAN III, and RAN III′ were 15-bp complex hetero-oli-

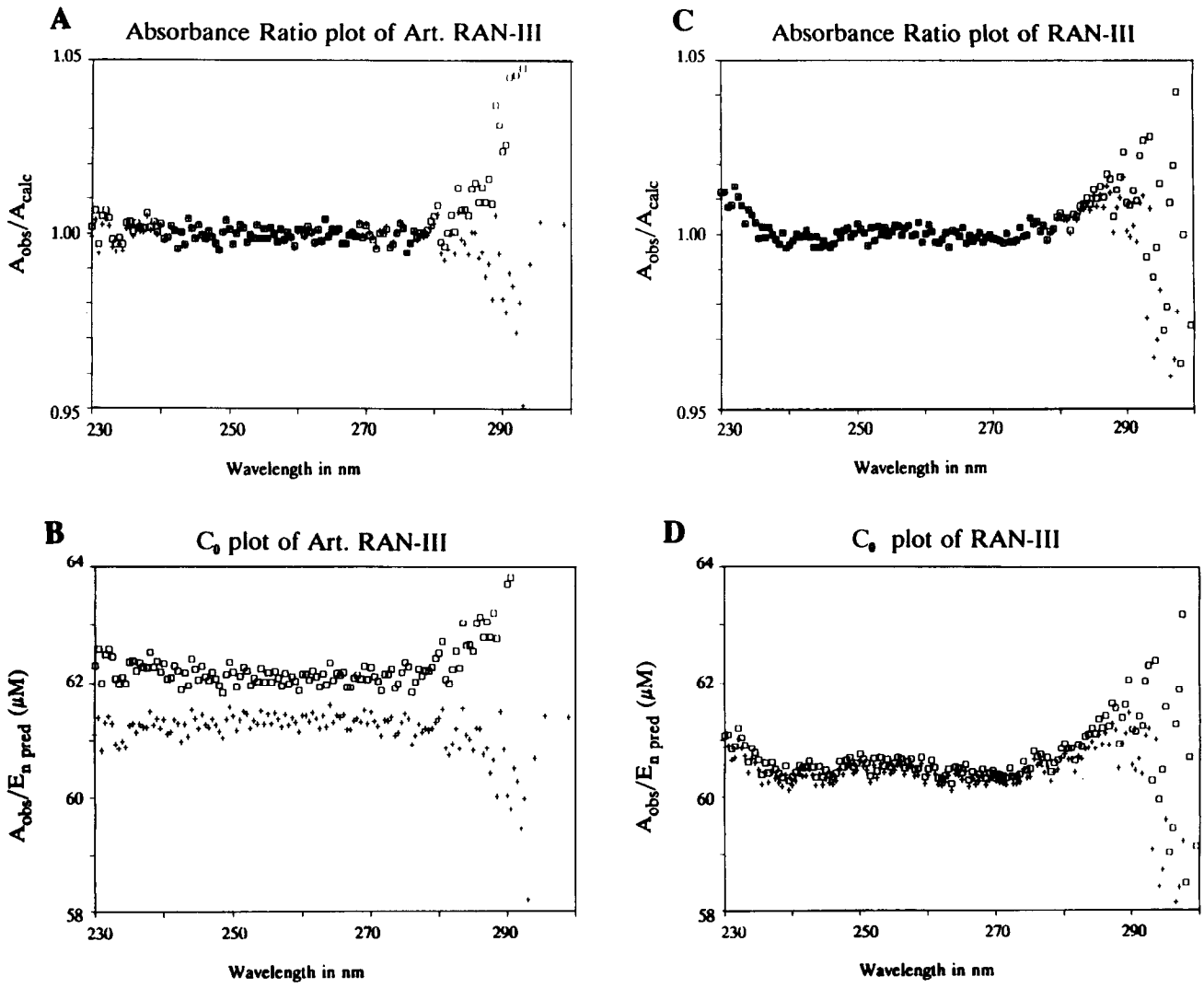

FIG. 1. Absorbance ratio and C_0 plots for RAN III digest. The simulated data for RAN III (A and B) are compared with the experimental results in (C and D). Results are shown for the regressions run with (+) and without (□) an intercept. The simulated data incorporated random noise levels of ±0.003 and a baseline shift of +0.005 over the wavelength range of 230 to 360 nm. The P1 nuclease digestions were conducted as described under Materials and Methods. The wavelength range of the regression analyses was 240 to 280 nm.

gomers. The primed numerals represent the complement of the unprimed numerals and have the higher purine content. The other oligomers examined were the homo-oligomers, $(dA)_{16}$, $(dT)_{16}$, $(dC)_{16}$, and simple repetitive sequences, $(dGdC)_3$, $(dGdA)_8$, $(dTdGdG)_5$, $(dTdTdG)_5$, $(dCdCdA)_5$, and $(dCdAdA)_5$.

Table 3 shows typical results of the analysis of base compositions and extinction coefficients on the basis of the P1 nuclease determination. In all cases, the regression analysis included all four nucleotides analyzed over the optimal range of 240 to 280 nm. Using smaller subsets of data within this range resulted in the same total concentration, C_0, and corresponding extinction coefficients, but less accurate base compositions. The observed base compositions given in this table are in excellent agreement with those predicted from the sequences. The values of SD A fall within the range that would be expected if only small baseline shifts and minimal spectral noise were corrupting the accuracy of the spectral signals.

Additional verification of the validity of the results is also based on the appearance of the absorbance ratio and the C_0 plots. As expected for the samples in Table 3, these plots are normal. Figures 1C and 1D shows typical plots for experiments with RAN III, for which the same plots of its artifical counterpart with random spectral noise and baseline shifts were shown in Figs. 1A and 1B. The scattering of the points seen in all four panels at the high-wavelength end of the wavelength window of analysis is attributable to the effects of spectral noise in regions of low sample signal. The scattering of the real data for RAN III in Figs. 1C and 1D at the low-wave-

TABLE 3

Extinction Coefficients and Base Compositions of Single-Strand Oligomers at 25°C, pH 7, Whose Analyses Passed All Tests

Oligomer	Base Compositions A	G	C	T	SD A ($\times 10^2$)	λ_{max} (nm)	E ($mM^{-1}\ cm^{-1}$)	E_{avg} (± 0.08) ($mM^{-1}\ cm^{-1}$)
Hetero-oligomers								
MRE I								
Predicted	0.000	0.267	0.533	0.200				
Observed	0.002	0.275	0.524	0.198	0.53	262	8.02	7.99
MRE I′								
Predicted	0.200	0.533	0.267	0.000				
Observed	0.200	0.531	0.266	0.002	0.13	254	9.30	9.22
INV II								
Predicted	0.000	0.267	0.533	0.200				
Observed	−.002	0.278	0.517	0.207	0.89	262	7.99	
INV II′								
Predicted	0.200	0.533	0.267	0.000				
Observed	0.197	0.529	0.269	0.005	0.36	254	9.40	
RAN III								
Predicted	0.133	0.333	0.467	0.067				
Observed	0.131	0.337	0.464	0.068	0.25	258	8.35	8.38
RAN III′								
Predicted	0.067	0.467	0.333	0.133				
Observed	0.064	0.471	0.326	0.138	0.49	255	8.49	8.48
Repetitive oligomers								
$(dTGG)_5$								
Observed	−0.007	0.679	0.008	0.320	1.04	254	9.80	9.80
$(dCCA)_5$								
Observed	0.334	−0.008	0.672	0.002	0.49	254	8.60	8.63
$(dTTG)_5$								
Observed	−0.016	0.339	0.011	0.666	1.01	259	8.43	
$(dCAA)_5$								
Observed	0.659	−0.007	0.336	0.012	0.81	258	9.40	9.44
$(dCG)_3$								
Observed	0.002	0.508	0.495	−0.006	0.70	255	8.73	8.76
$(dGA)_8$								
Observed	0.499	0.510	−0.009	0.000	0.66	254	9.99	9.99
$(dA)_{16}$								
Observed	0.998	0.007	−0.012	0.018	1.27	258	10.07	10.01
$(dC)_{16}$								
Observed	−0.001	0.011	0.993	−0.003	0.69	270	8.19	8.19
$(dT)_{16}$								
Observed	−0.008	−0.002	−0.066	1.007	5.07	267	8.50	

length end is probably due to reference solvent inbalances.

The ability of these tests to detect aberrations in the data is demonstrated by the results shown in Figs. 2 and 3. The INV II′ preparation whose characteristics are presented in Table 3 was the preparation purified by gel electrophoresis. The unpurified sample with 15% of its nucleotide content in lower molecular weight products gave a value of SD A of 0.049, indicating a significantly deviant base composition. Interestingly enough, the absorbance ratio plot for the unpurified sample, shown in Fig. 2A, appears normal although the C_0 plot, shown in Fig. 2B, is clearly aberrant. The data are not improved by the incorporation of an intercept. This suggests that the impurities in the intact sample are mono- or oligonucleotides of different fractional base content arising from prematurely terminated synthesis.

The absorbance ratio and the C_0 plots for the purified sample of INV II′ are shown in Figs. 2C and 2D. The analysis shows that the best fit of the digest data was obtained by the incorporation of an intercept in the regression. Examination of the absorption spectra revealed that the sample exhibited a slight elevation in baseline above 300 nm attributable to fines from the gel used in the purification. The incorporation of an intercept corrected for this essentially wavelength-independent scattering component. The calculated value of the extinction coefficient is now within 1% of that for the

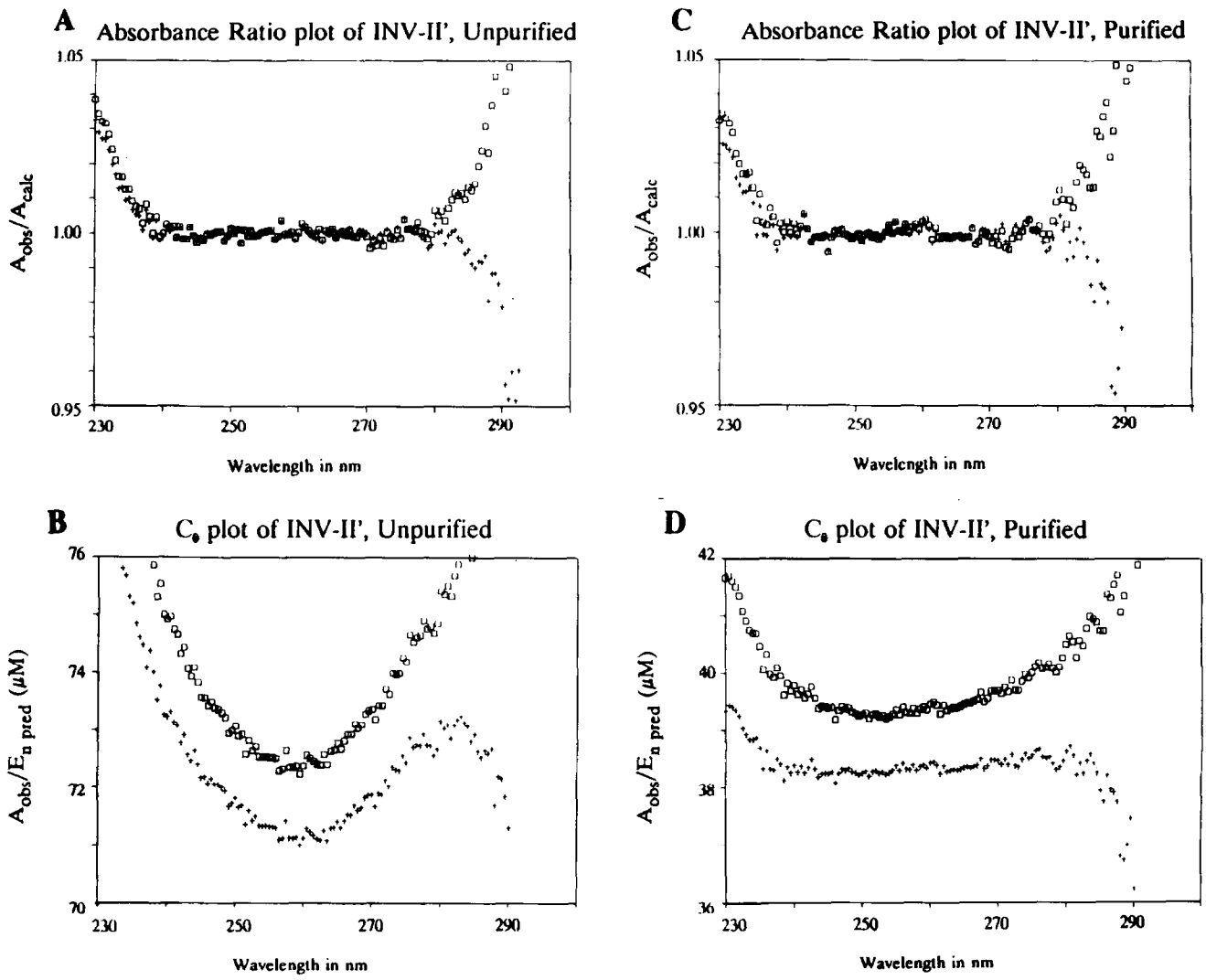

FIG. 2. Absorbance ratio and C_0 plots for the INV II' digest. (A and B) Results for the oligomer prior to removal of the contaminating lower molecular weight impurities. (C and D) Data obtained on the final purified product. Results are shown for the regressions run with (+) and without (□) an intercept. The wavelength range of the regression analyses was 240 to 280 nm.

inverse sequence, MRE I'. The same is true for the pair MRE I and INV II. It is interesting to note that the reversal of the sequences of these oligomers has essentially no effect on the character of their absorption spectra.

The effects of another artifact that is occasionally encountered are shown in Figs. 1D and 3. Small shifts in wavelength (ca. 0.2 nm) between the reference components and the digest solution generally lead to minor periodic variations of C_0 with wavelength, of the type seen in Fig. 1D, but have little effect on the results of the final analysis. A more significant shift, however, will result in a very dramatic appearance in even the less sensitive absorbance ratio plot. Figure 3A shows the effect of a 1-nm shift between the reference spectra and the digest spectra for the repetitive oligomer $(dGA)_8$. The value of the SD A for the oligomer analyzed with the shifted reference set was 0.104, indicative of a very skewed base composition. A shift of the digest spectrum 1 nm to the red results in the plots shown in Figs. 3C and 3D and excellent agreement between the predicted and the observed composition, as reported in Table 3.

The concentration and extinction coefficients of a number of samples that passed the tests of the P1 analyses were also determined by conventional phosphate analysis. These results are presented in Table 4. Columns 4 and 5 of the same table give the values based on the phosphate concentration assay, corrected for the lack of a terminal phosphate at the 5' end of the oligomer. The last column reports the ratio of the extinction coefficient determined by the P1 digestion procedure to that by the phosphate determination. The overall agreement is excellent. The average of these ratios is 1.002, with a population standard deviation of

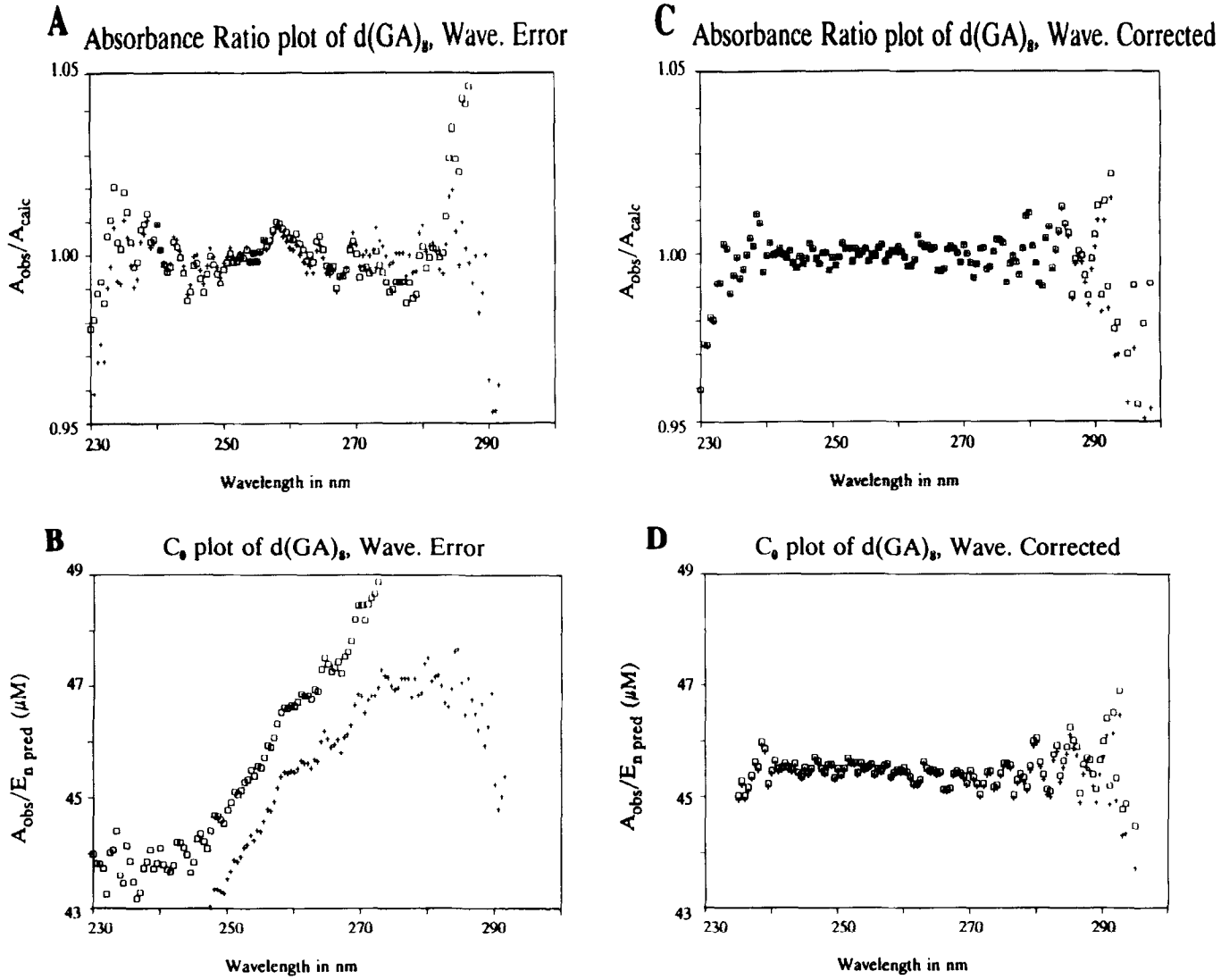

FIG. 3. Absorbance ratio and C_0 plots for the $(dGA)_8$ digest. (A and B) Data corrupted with a 1-nm wavelength shift of the reference spectral set. (C and D) The results of wavelength correction of the shift. Solvent and digestion conditions for the oligomer are those described under Materials and Methods. Results are shown for the regressions run with (+) and without (□) an intercept. The wavelength range of the regression analyses was 240 to 280 nm.

±0.028. This variability is due to the greater variability of the phosphate analysis, which is usually between 2 and 3%, rather than the P1 analysis, which is only ca. 1%.

Polynucleotide Digestions

Three duplex species, poly(dIdC), poly(dGdC), and calf thymus DNA, as well as heat-denatured DNA and single-strand M13mp19, were examined. The three duplex samples were digested with a mixture of pancreatic DNase I and snake venom phosphodiesterase, as described under Materials and Methods, under conditions where their duplex structure was intact. Because of the paucity of data collected by the Cary 14 spectrophotometer, these spectra were analyzed with a minimal number of independent variables. Except for the GC content of calf thymus DNA, base composition analyses were not attempted. These results are shown in the upper part of Table 5 together with values for the extinction coefficients taken from the literature. Duplex calf thymus DNA at denaturing temperatures, heat-denatured calf thymus DNA, and single-strand M13 DNA were digested with P1 nuclease and digital spectral data were collected with the DU 40 spectrophotometer. The results of these investigations are shown in the lower part of Table 5. It is interesting to note that despite the differences in spectrophotometers, modes of digestion, and quantities of data processed, the results for calf thymus, digested as either the native duplex species or the heat-denatured form, are very similar.

TABLE 4

Comparison of Individual Sample Concentrations and Extinction Coefficients Determined by the P1 Nuclease and the Phosphate Procedures

Oligomer	P1 nuclease		Phosphate		$E_{\lambda max}$ ratio[a] P1/phosphate
	Concentration (μM)	$E_{\lambda max}$ (mM^{-1} cm^{-1})	Concentration (μM)	$E_{\lambda max}$ (mM^{-1} cm^{-1})	
MRE I	47.4	8.06	47.1	8.10	0.995
MRE I	59.5	8.02	61.3	7.78	1.031
MRE I	66.0	7.96	64.4	8.15	0.977
MRE I	60.3	7.99	60.3	7.99	1.000
MRE I	58.9	7.99	60.2	7.81	1.023
		Average = 8.00 ± 0.04		Average = 7.97 ± 0.17	
MRE I′	75.0	9.15	72.0	9.53	0.960
MRE I′	48.0	9.14	47.9	9.17	0.997
MRE I′	46.3	9.36	45.1	9.60	0.975
		Average = 9.22 ± 0.12		Average = 9.43 ± 0.23	
INV II	60.6	7.99	62.9	7.70	1.038
$(dTGG)_5$[b]	65.1	8.71	63.8	8.89	0.980
$(dCCA)_5$	50.7	8.66	50.6	8.68	0.998
$(dCAA)_5$	48.6	9.40	51.3	8.90	1.056
					Average = 1.002 ± 0.028

[a] Ratio of $E_{\lambda max}$ by P1 nuclease method to $E_{\lambda max}$ by phosphate method.
[b] Not premelted.

Polynucleotide analysis, in general, presents special problems that result in less accurate base compositions. Most of these problems are attributed to a greater degree of light scattering, protein contamination, and other extraneous absorbing components carried along in the isolation procedure. The deviations in observed base composition are always greater than that normally observed for the oligomers. These artifactual contribu-

TABLE 5

Extinction Coefficients and Base Compositions of Polynucleotides

Polymer	Base composition				SD A ($\times 10^2$)	λ_{max} (nm)	$E_{\lambda max}$ (mM^{-1} cm^{-1})	
	A	G	C	T			This work	Literature
Duplex forms: DNase I + snake venom phosphodiesterase digestions								
Poly(dGdC), B form						256	7.22	7.10[a]
Poly(dIdC)						252	7.29	6.90[b]
Calf thymus DNA						258	6.63	6.67[c]
Predicted GC = 0.42								
Observed GC = 0.45								
Single-strand and denatured forms: P1 nuclease digestions								
Calf thymus DNA								
Predicted	0.29	0.21	0.21	0.29				
Observed								
Denatured	0.267	0.221	0.212	0.300	1.30	258	6.70	
Digested at 70°C	0.275	0.216	0.228	0.282	1.29	258	6.76	
M13mp19 DNA								
Predicted	0.245	0.206	0.203	0.346				
Observed	0.226	0.202	0.198	0.370	1.56	258.5	8.36	

[a] Taken from Pohl and Jovin (1).
[b] Taken from Grant *et al.* (15).
[c] Taken from Johnson *et al.* (16).

tions, however, do not appear to have a dramatic effect on the concentration assay and, consequently, the value of the extinction coefficient obtained by the analysis.

CONCLUSIONS

The method described in this publication offers significant advantages over the colorimetric methods or other digestion procedures that rely on the analysis of data from one or, at most, a very few wavelengths for the quantitation of oligo- and polynucleotide concentrations and extinction coefficients. Its accuracy and sensitivity are comparable to those of colorimetric phosphate analyses but its convenience and precision are considerably greater. Although its precision is comparable to that of the simpler digestion procedures, the capacity to detect significant artifacts and errors permits greater confidence in the results. Thus, less sample is ultimately required, compared to other procedures which usually demand multiple determinations for reliable estimates of concentration. For short oligomers of known base composition, the data are accurate enough to serve as indicators of purity and homogeneity. The ease of quantitation of these oligomers will be a great advantage in studies that demand accurate values of their concentration.

ACKNOWLEDGMENTS

We thank Dr. Douglas Maibenco and Dr. Catherine Chen for obtaining the analytical data on the duplex polynucleotides. We are indebted to the personnel of the Research Resources Laboratory for advice and data-processing facilities. This research was supported by Grants GM 30284 (to S.H.) and GM 39471 (to A.S.B.) from the National Institutes of Health.

REFERENCES

1. Pohl, F. M., and Jovin, T. M. (1972) *J. Mol. Biol.* **67,** 375–396.
2. Nadeau, J. G., and Gilham, P. T. (1985) *Nucleic Acid Res.* **13,** 8259–8275.
3. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **XLVII,** 405–407.
4. Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., and Treiber, G. (1970) *Virology* **40,** 734–744.
5. Wilson, D. H., Price, H. L., Henderson, J., Hanlon, S., and Benight, A. S. (1990) *Biopolymers* **29,** 357–376.
6. Berkner, K. L., and Folk, W. R. (1977) *J. Biol. Chem.* **252,** 3176–3184.
7. Fiske, C. H., and Subbarow, Y. (1925) *J. Biol. Chem.* **66,** 370–375.
8. Ames, B., and Dubin, D. (1960) *J. Biol. Chem.* **235,** 770–775.
9. Chen, P. S., Toribara, T. Y., and Warner, H. (1956) *Anal. Chem.* **28,** 1756–1758.
10. Hanlon, S., Brudno, S., Wu, T. T., and Wolf, B. (1975) *Biochemistry* **14,** 1648–1660.
11. Oster, G. (1948) *Chem. Rev.* **43,** 319–365.
12. Sober, H. A., (Ed.) (1968) Handbook of Biochemistry, pp. G21–G86, Chemical Rubber Co., Cleveland, OH.
13. Pabst Laboratories (1972) Circular OR-7, pp. 1–7, PL Biochemicals, Milwaukee, WI.
14. Voet, D., Gratzer, W. B., Cox, R. A., and Doty, P. (1963) *Biochemistry* **2,** 193–208.
15. Grant, R. C., Harwood, S. J., and Wells, R. D. (1968) *J. Am. Chem. Soc.* **90,** 4474–4476.
16. Johnson, R., Chan, A., and Hanlon, S. (1972) *Biochemistry* **11,** 4347–4358.