Precise Measurement of the G+C Content of Deoxyribonucleic Acid by High-Performance Liquid Chromatography

MOSTAFA MESBAH,[†] USHA PREMACHANDRAN, AND WILLIAM B. WHITMAN*

Department of Microbiology, University of Georgia, Athens, Georgia 30602

High-performance liquid chromatography is a promising alternative for determining the G+C content of bacterial deoxyribonucleic acid (DNA). The method which we evaluated involves enzymatic degradation of the DNA to nucleosides by P1 nuclease and bovine intestinal mucosa alkaline phosphatase, separation of the nucleosides by high-performance liquid chromatography, and calculation of the G+C content from the apparent ratios of deoxyguanosine and thymidine. Because the nucleosides are released from the DNA at different rates, incomplete degradation produces large errors in the apparent G+C content. For partially purified DNA, salts are a major source of interference in degradation. However, when the salts are carefully removed, the preparation and degradation of DNA contribute little error to the determination of G+C content. This method also requires careful selection of the chromatographic conditions to ensure separation of the major nucleosides from the nucleosides of modified bases and precise control of the flow rates. Both of these conditions are achievable with standard equipment and C18 reversed-phase columns. Then the method is precise, and the relative standard deviations of replicate measurements are close to 0.1%. It is also rapid, and a single measurement requires about 15 min. It requires small amounts of sample, and the G+C content can be determined from DNA isolated from a single bacterial colony. It is not affected by contamination with ribonucleic acid. Because this method yields a direct measurement, it may also be more accurate than indirect methods, such as the buoyant density and thermal denaturation methods. In addition, for highly purified DNA, the extent of modification can be determined.

The measurement of G+C content is one of the most common measurements performed on bacteria. Its greatest values are probably its usefulness as a taxonomic marker and its usefulness in distinguishing phenotypically similar microorganisms. G+C content is also a fundamental property of cellular deoxyribonucleic acid (DNA) and is correlated with the amino acid composition of proteins, codon usage in messenger ribonucleic acid, auxotrophy for specific bases, and other properties of general biological interest (1, 8, 9, 19, 25, 26, 29, 36). Therefore, interest in this parameter is not limited to microbial taxonomists.

Although the G+C content of DNA has a precise analytical meaning, the techniques for determining this parameter are often indirect and lack a high degree of precision. For instance, the standard deviations of the most widely used methods, the buoyant density centrifugation and thermal denaturation methods, are usually about ± 1.0 and ± 0.4 mol% G+C, respectively (5, 20, 21, 23, 28, 34). The accuracy of both of these methods is also affected by the purity of the DNA and the presence of modified bases (20, 28). The other methods which are occasionally used, the chemical modification and ultraviolet absorption methods, also present difficulties under certain experimental conditions (3, 4, 11, 16, 31-33).

Recently, high-performance liquid chromatography (HPLC) has been proposed as an alternative procedure (17, 24, 30). Chromatography has an immediate advantage in that it allows direct measurement of the base composition. Thus, sources of error are predictable and easily identifiable. In addition, modern instruments make measurements rapid and reproducible. However, little work has been performed to determine either the accuracy or the precision of this method. Notable exceptions have been the studies of Kuo

and collaborators on the application of liquid chromatography to the determination of modified bases in DNA (6, 15, 18). In this report we describe an HPLC method for determining G+C content, and we identify a number of potential sources of error.

MATERIALS AND METHODS

Apparatus. The equipment used for HPLC was obtained from Beckman Instruments, Inc., Berkeley, Calif. It included two model 110A high-pressure pumps, a model 420 controller, an Altex model 210 injector with a 20- μ l sample loop, and a model 160 fixed-wavelength absorbance detector with a 254-nm filter and a quartz-mercury ultraviolet lamp. The data were collected with a model 3390A integrator (Hewlett-Packard Co., Avondale, Pa.). The integrator used the default conditions except for attenuation and chart speed, which were set at 4 and 0.5 cm/min, respectively. For the standard chromatographic conditions, an Econosphere C18 reversed-phase column (Alltech Associates, Inc., Deerfield, Ill.) was used. The particle size was 5 μ m, and the column dimensions were 250 by 4.6 mm. The column temperature was controlled with a water jacket and a refrigerated circulating water bath.

Chromatography conditions. Unless specified differently below, the conditions which we used included a flow rate of 1.0 ml/min at a temperature of 37°C. Each sample contained between 0.5 and 1.5 μ g of nucleosides in 5 to 25 μ l. The solvent contained 12% methanol and 20 mM triethylamine phosphate (pH 5.1). The solvent was prepared by combining 40 ml of 0.5 M triethylamine phosphate (pH 5.1) with about 750 ml of glass-distilled water. HPLC-grade methanol (120 ml; J. T. Baker Chemical Co., Phillipsburg, N.J.) was added, and the volume was adjusted to 1 liter. The solvent was then filtered through a 0.45- μ m cellulose triacetate membrane (GA-6 Metrice; Gelman Sciences, Inc., Ann Arbor, Mich.). Triethylamine (minimum concentration, 99%) was obtained from Kodak Chemical Co., Rochester,

* Corresponding author.

[†] Present address: Department of Pharmacognosy, Faculty of Pharmacy, University of Assiut, Assiut, Egypt.

N.Y. If the reagent had a noticeable yellow color, it was purified by vacuum distillation. To prepare the 0.5 M triethylamine phosphate solution, triethylamine was diluted with water, the pH was adjusted to 5.1 with 85% phosphoric acid, and the solution was brought to its final volume. When the column was not in use, the flow rate was reduced to 0.1 ml/min. When the machine was not to be operated for more than 3 days, the column was washed with water and then with 70% (vol/vol) methanol. When the column pressure exceeded 2,000 lb/in², the filters and precolumn were changed. Piston seals were also changed frequently.

Preparation of DNA. "*Methanococcus aeolicus*" was grown as described previously (35). *Escherichia coli* was grown on LB broth or agar plates. Bacterial cells were washed once in fresh medium prior to centrifugation and determination of the wet weight of the cells. The cellular dry weight was taken to be 0.2 times the cellular wet weight. DNA was then extracted by one of three methods. In the first (phenol) method, 40 mg (dry weight) of cells was suspended in 2 ml of TEN buffer, which contained 0.01 M tris(hydroxymethyl)aminomethane (Tris) chloride, 0.20 M NaCl, and 0.01 M sodium ethylenediaminetetraacetic acid (EDTA) (pH 7.2). A 1-ml portion of TEN buffer containing 25% (wt/vol) sodium dodecyl sulfate and 3 μ l of a pancreatic ribonuclease solution (1 mg/ml in 10 mM Tris chloride [pH 7.5]–15 mM NaCl) was then added. The ribonuclease solution had been heat treated to remove residual deoxyribonuclease activity (22). The sample was then incubated at 50°C for 1 h. Proteinase K (15 μ l of a 20-mg/ml solution in water) was added, and the sample was incubated for an additional 30 min at 37°C. An equal volume of buffer-saturated phenol was then added (22). After gentle mixing for 15 min, the aqueous phase was separated by centrifugation at room temperature at 5,000 $\times$ g for 5 min. The aqueous phase was collected and extracted further with an equal volume of a 1:1 mixture of phenol and chloroform and finally with an equal volume of chloroform-isoamyl alcohol (24:1, vol/vol). After the addition of sodium acetate to a concentration of 0.1 M, the DNA was precipitated with 2 volumes of 95% ethanol at –20°C for 1 h. After centrifugation, the pellet was suspended in 0.8 ml of 0.1 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl plus 15 mM trisodium citrate). Then 0.1 volume of 1 M NaCl was added, and the DNA was precipitated by adding an equal volume of isopropyl alcohol at –20°C for 1 h. After centrifugation, the DNA was suspended in a minimum volume of 0.01 $\times$ SSC (usually 0.05 ml) and stored at –20°C. When the initial cell weight was different from 40 mg, the volumes of the solutions used in the procedure were adjusted proportionally.

The second (NaOH) method was a modification of the procedure of Beji et al. (2). Washed cells (40 mg, dry weight) were suspended in 1 ml of 0.03 M NaOH. After incubation for 3 min at room temperature, 3.3 ml of saline-EDTA buffer (0.15 M NaCl, 0.1 M disodium EDTA, pH 7.0) containing 2.5% sodium dodecyl sulfate was added. The sample was then treated with 6 μ l of ribonuclease for 30 min at 60°C and with 12 μ l of proteinase K for 15 min at 37°C. Then 0.25 volume of 5 M NaCl was added, and the sample was treated with an equal volume of chloroform-isoamyl alcohol for 5 min. After centrifugation, the DNA was collected from the aqueous phase by ethanol and isopropyl alcohol precipitations and suspended in a minimum volume of 0.01 $\times$ SSC as described above for the phenol method. When the initial cell weight was less than 40 mg, the volumes of the solutions were reduced proportionally.

In the third (hydroxylapatite [HA]) method, cells were treated with sodium dodecyl sulfate and enzymes as de-

scribed above for the phenol method. After two extractions with buffer-saturated phenol, 0.25 volume of 1 M sodium phosphate buffer (pH 7.1) and 0.3 g of HA (or about 0.1 g/0.1 mg of nucleic acid) were added. The mixture was gently mixed on a rotary shaker for 30 min at room temperature. The HA was collected by centrifugation for 3 min at 5,000 $\times$ g and washed four times in 3 ml of 0.1 M sodium phosphate buffer (pH 7.1). The DNA was then eluted with 3 ml of 0.5 sodium phosphate buffer (pH 7.1). After centrifugation to remove the HA, the supernatant was filtered through a 0.5- μ m Millex-HV filter (Millipore Corp., Bedford, Mass.), and the DNA was collected by precipitation with ethanol. After centrifugation, the DNA was suspended in a minimum volume of 0.1 $\times$ SSC (about 0.2 ml) and dialyzed overnight at 4°C against 100 volumes of 15 mM sodium acetate buffer (pH 5.2).

DNA was purified from a single colony by the NaOH method. A colony (diameter, 2 mm) of *E. coli* was scraped off a plate with a needle and suspended in 50 μ l of 0.03 M NaOH by vortexing. After 2 min, 100 μ l of saline-EDTA buffer containing sodium dodecyl sulfate was added. The sample was then treated with 6 μ l of ribonuclease for 20 min at 55°C and with 3 μ l of proteinase K for 15 min at 37°C. The sample was then extracted with chloroform, and the DNA was precipitated with ethanol and isopropanol as described above. The pellet from the isopropanol precipitation was then suspended in 10 μ l of 0.01 $\times$ SSC and dialyzed overnight against 15 mM sodium acetate (pH 5.2).

Degradation of DNA. In the standard method, 25 μ l of a solution containing 2 to 25 μ g of DNA in a 1.5-ml Microfuge tube was heated in a boiling water bath for 2 min and rapidly cooled in an ice water bath. Then 50 μ l of 30 mM sodium acetate buffer (pH 5.3), 5 μ l of 20 mM ZnSO₄, and 3 μ l of P1 nuclease (1 mg/ml in sodium acetate buffer; 340 U/ml) were added. The sample was incubated for 2 h at 37°C, and 5 μ l of 0.1 M glycine hydrochloride buffer (pH 10.4) and 5 μ l of bovine intestinal mucosa alkaline phosphatase (200 U/ml in glycine buffer) were added. With these additions, the pH of the sample was between 7.5 and 8.5. The sample was incubated for 6 h at 37°C, centrifuged at 10,000 $\times$ g for 4 min, and stored at –20°C until it was chromatographed. The P1 nuclease was purchased as the lyophilized protein. After suspension in buffer, it was stored in small portions at –20°C. After thawing, each portion was used immediately, and any remaining enzyme was discarded. The alkaline phosphatase was purchased in a 3.0 M NaCl solution. Immediately prior to use, a small volume was diluted in glycine buffer.

For some experiments, the standard procedure was modified. To study the P1 nuclease reaction, the reaction was terminated after 4 min by adding an equal volume of 0.2 M Tris chloride buffer (pH 9.0). The formation of deoxyadenosine monophosphate (dAMP) was measured by using the same chromatographic conditions used for the nucleosides. When the effect of pH was studied, the following buffers were used (pH values in parentheses); sodium acetate (3.0, 3.5, 4.0, 4.5, 5.0, and 5.0), sodium 2-(*N*-morpholino)ethanesulfonic acid (MES) (6.0 and 6.5), and sodium piperazine-*N,N'*-bis(2-ethanesulfonate) (PIPES) (6.5, 7.0, and 7.5). To determine the nucleoside ratios, the nuclease reaction was terminated by boiling, and the nucleotides were degraded by the phosphatase for 24 h at 37°C. To study the phosphatase reaction, DNA was first treated with nuclease P1. After the solution was boiled for 4 min, the phosphatase reaction was initiated with enzyme and glycine buffer.

To degrade the nucleic acids from single colonies, 10 μ l of

each dialyzed DNA sample was denatured in a boiling water bath and diluted with 12 μ l of sodium acetate buffer, 2 μ l of 20 mM ZnSO₄, and 2 μ l of P1 nuclease. After incubation at 37°C for 1 h, 2 μ l of alkaline phosphatase and 2 μ l of glycine buffer were added. After 7 h at 37°C, 20 μ l of the sample was injected into the HPLC apparatus.

Calculation of G+C content. G+C content can be calculated from the total composition of DNA or from the ratio of certain bases. G+C content is defined as $100 \times M$, where M is the mole fraction of deoxyguanosine (dGuo) plus deoxycytidine (dCyd). Thus,

$$M = (G + C)/(G + C + A + T)$$

where G , C , A , and T are the mole fractions of the nucleosides dGuo, dCyd, deoxyadenosine (dAdo), and thymidine (dThd), respectively. When modified bases are present, G , C , A , and T are the sums of the mole fractions of the modified and unmodified nucleosides. The HPLC method does not report the true mole fraction, but a value related to the true mole fraction. If the HPLC response is linear, the apparent value is a multiple of the true mole fraction and a constant which depends upon the extinction coefficient of the bases and the operating parameters, including the wavelength of the detector, cuvette shape, and elution profile of the nucleosides. When modified bases are absent, wG , xC , yA , and zT are the apparent mole fractions of the nucleosides, where w , x , y , and z are the associated constants. M_a , the apparent mole fraction of G+C, is then defined as follows:

$$M_a = (wG + xC)/(wG + xC + yA + zT)$$

Because $G = C$ and $A = T$,

$$M = G/(G + T)$$

$$M_a = G/(G + TX)$$

where $X = (y + z)/(w + x)$. If the composition of a standard DNA is known, this equation can be solved for X , as follows:

$$X = (G/M_a - G)/T$$

For bacteriophage lambda DNA, which has a G+C content of 49.858 mol%, X was found to be about 0.981 under the conditions described above.

Once X is known, M or the true mole fraction of G+C for an unknown DNA may be determined. Because $2G = M$ and $2T = 1 - M$,

$$M_a = M/[M + X(1 - M)]$$

$$M = [(M_a^{-1} - 1)/X + 1]^{-1}$$

The G+C content may also be calculated from the ratios T/G , A/G , T/C , and A/C . The calculations shown below are for the T/G ratio, but similar arguments can be made for other bases. Because $M = G/(G + T)$,

$$M = (1 + T/G)^{-1}$$

To determine T and G from the apparent mole fractions zT' and wG' , the constant Y must be determined by chromatography of standard DNA. Thus,

$$Y = (wG'/zT') (T'/G') = w/z$$

where zT' and wG' are the apparent mole fractions for the DNA standard. For bacteriophage lambda DNA under the conditions described above, Y was found to be about 2.04. The mole fraction of G+C for a DNA of unknown composition is now

$$M = [1 + Y(zT'/wG')]^{-1}$$

This expression may be generalized to other pairs of nucleosides. Let $n_1 = yA$ or zT , $n_2 = xC$ or wG , and Y_i be the appropriate correction factor for each pair of nucleosides; then,

$$M = [1 + Y_i(n_1/n_2)]^{-1}$$

Because bacterial DNA is frequently modified at dCyd or dAdo, additional correction factors must be calculated when these nucleosides are considered. The correction factors for modified dCyd, Z_c , and for modified dAdo, Z_a , can be shown to be

$$Z_c = (xC/wG)/(xC'/wG')$$

$$Z_a = (zT/yA)/(zT'/yA')$$

where wG' , xC' , yA' , and zT' are the apparent mole fractions for the standard unmodified DNA. Then,

$$M = [1 + Z_c Y_i (n_1/xC)]^{-1}$$

$$M = [1 + Z_a Y_i (yA/n_2)]^{-1}$$

$$M = [1 + Z_c Z_a Y_i (yA/xC)]^{-1}$$

The mole fractions of modified dCyd, M_c , and of modified dAdo, M_a , may also be calculated as follows:

$$M_c = M(1 - Z_c)/2$$

$$M_a = (1 - M)(1 - Z_a^{-1})/2$$

The ratio of two nucleosides, like zT/wG , may be determined experimentally in the following two ways: (i) it is simply the quotient of the machine responses for dThd and dGuo during the chromatography of a single sample; and (ii) zT/wG is also the slope of plots of the machine responses for dThd versus the responses for dGuo. The second method may be more accurate if the machine response for the nucleosides is not linear at low concentrations. In this case, the x and y intercepts of these plots are significantly different from zero.

Materials. Salmon sperm DNA was obtained from Sigma Chemical Co., St. Louis, Mo., and calf thymus DNA was obtained from Calbiochem-Behring, La Jolla, Calif. DNA was purified from *Methanococcus voltae* PS by centrifugation in CsCl gradients as described previously (35). Nucleic acid from *Myxococcus xanthus* DK1622 was a generous gift from L. J. Shimkets, University of Georgia. Nucleic acid from *Acinetobacter calcoaceticus* BD413 was a generous gift from M. Singer, University of Georgia. Nucleic acid from *E. coli* K-12 was purified by HA chromatography. Methyl-free DNA from bacteriophage lambda was generously provided by R. Makula, University of Georgia. Yeast DNA from *Candida parapsilosis* stored at -20°C in 1× SSC was generously provided by S. Meyer, Georgia State University.

RESULTS

The nucleosides produced upon enzymatic degradation of bacterial nucleic acids were readily separated by HPLC (Fig. 1). The G+C content of DNA could then be calculated from the ratios of certain bases (see below). Because dCyd and dAdo were frequently modified in bacterial DNA, the ratio of dGuo to dThd was most useful for this purpose. Therefore, the chromatographic conditions were optimized for the resolution of dGuo and dThd, and the reliability of the method was evaluated. The following four major parameters were investigated: chromatography conditions, equipment performance, degradation, and sample preparation.

FIG. 1. Isocratic separation of nucleosides. (A) Separation of ribonucleosides from deoxyribonucleosides at 24°C in 12% methanol. The nucleosides were obtained by enzymatic degradation of *Methanococcus voltae* nucleic acid. (B) Separation of nucleosides under the optimal conditions for resolution of dGuo and dThd. The chromatography conditions were 38°C and 12% methanol. The nucleoside mixture was prepared from standards. (C) Separation of nucleosides from nucleotide monophosphates. The chromatography conditions were 38°C and 12% methanol. The nucleoside mixture contained 2% nucleotide monophosphates. dC, Deoxycytidine; dG, deoxyguanosine; dA, deoxyadenosine; dT, thymidine; G, guanosine; C, cytidine; A, adenosine; U, uridine; mC, 5-methyldeoxycytidine; dGMP, deoxyguanosine monophosphate; X, unidentified compound.

Chromatography conditions. When we used a C18 reversed-phase column at 37°C and 12% methanol, dGuo and dThd were well resolved in a short time (Fig. 1B). Importantly, 5-methyldeoxycytidine, which was a common minor component of DNA, was well separated from dGuo. Without temperature control and careful selection of the chromatographic conditions, 5-methyldeoxycytidine frequently coeluted with other bases and produced large errors in the calculation of G+C content. In contrast to 5-methyldeoxycytidine, we observed no interference by *N*⁶-methyldeoxyadenosine, which eluted several minutes after dAdo (data not shown).

If the degradation was incomplete, small amounts of nucleotide monophosphates remained in the samples. Resolution of dAMP and adenosine monophosphate from dThd

FIG. 2. Resolution of critical components for determining the apparent dGuo/dThd ratio. Resolution was with 12% methanol on an Econosphere C18 column (particle size, 5 μ m). For abbreviations see the legend to Fig. 1. AMP, Adenosine monophosphate.

and dGuo, respectively, was also important for the calculation of the nucleoside ratio. Without careful selection of the chromatographic conditions, these nucleotides coeluted with the nucleosides and produced large errors (Fig. 2). Under the conditions chosen, dAMP eluted just after dThd (Fig. 1C). Therefore, it also served as a diagnostic tool for the efficiency of degradation.

When the chromatography system was optimized for the resolution of these critical components (dGuo, dThd, 5-methyldeoxycytidine, adenosine monophosphate, and dAMP), the presence of a 10-fold excess of ribonucleosides had no effect on the measurement of the dGuo/dThd ratios (data not shown). Moreover, the presence of 40% (wt/wt) ribonucleotide monophosphates or 10% (wt/wt) deoxyribonucleotide monophosphates also had no effect (data not shown). Because uridine coeluted with dCyd, the latter nucleoside could not be measured in the presence of ribonucleosides (Fig. 1).

Equipment performance. The reproducibility of the chromatographic system was tested over a period of 9 months. Periodically, the dGuo/dThd ratio for a mixture of nucleosides was determined. The ratios found ($\pm$ standard deviations) were as follows: 2.094 ± 0.005 , 2.093 ± 0.001 , 2.093 ± 0.001 , 2.094 ± 0.003 , and 2.089 ± 0.003 . Thus, these ratios were remarkably stable even though three different columns were employed during this time.

Measurements of the nucleosides and the apparent dGuo/dThd ratio were not affected by the flow rate under most conditions. For instance, for flow rates of 0.5, 1.0, 1.5, and 2.0 ml/min the apparent dGuo/dThd ratios ($\pm$ standard deviations) for a nucleoside mixture were 2.216 ± 0.002 , 2.211 ± 0.006 , 2.212 ± 0.003 , and 2.219 ± 0.002 , respectively. In contrast, the standard deviations of the measurements were very sensitive to fluctuations in the flow rate during the course of a single run. Upon wear, small leaks typically developed at the piston seals of the high-pressure pumps. In one experiment, before the seals were changed, the apparent dGuo/dThd ratio was 2.229 ± 0.011 (or $\pm 0.5\%$) for 45 samples. When the seals were replaced, the relative standard deviation dropped to 0.09% for 15 samples. Thus, changing the piston seals led to a threefold decrease in the relative standard error of the mean (from 0.067 to 0.023%), while only one-third as many samples were analyzed.

Degradation. To obtain the G+C content, DNA was

FIG. 3. Effect of pH and salt on the degradation of salmon sperm DNA by P1 nuclease. Symbols: ●, 30 mM buffer (sodium salt); ▲, 100 mM buffer (sodium salt); ■, 30 mM buffer plus 100 mM NaCl. The buffers used are described in the text.

degraded with P1 nuclease and alkaline phosphatase to nucleosides. Initially, two methods were used for degradation, the 40-h three-enzyme system and the 4-h two-enzyme system of Gehrke et al. (15, 18). For our purposes, the 4-h system was unsuitable, and replicate degradations produced highly variable results. With highly purified DNA, the 40-h system gave consistent results. However, it too produced variable results with crude preparations of DNA (data not shown). To determine the sources of these errors, the degradation procedure was examined further.

The P1 nuclease from *Penicillium citrinum* hydrolyzed nucleic acids to the 5'-nucleotide monophosphates (12, 13). However, the activity was sensitive to the nature of the substrate (ribonucleic acid [RNA] or DNA), the base composition of the substrate, the pH, and the concentration of salts (14). Under the conditions used to determine the G+C content, the pH optimum of the enzyme was about 5.0 (Fig. 3). The addition of 0.1 M NaCl lowered the overall activity and shifted the pH optimum to about 4.0. This effect was due primarily to the addition of Cl^- ions, because increases in the concentrations of the buffers, which were all Na^+ salts, had little effect (Fig. 3). Furthermore, the addition of 0.1 M KCl also lowered the overall activity and the pH optimum (data not shown). The nuclease also had a specific requirement for Na^+ . When the sodium acetate buffer was replaced by Tris acetate or ammonium acetate, the activity was reduced by more than 50% at pH 4 and 5 (data not shown). Likewise, activity was also reduced by 50% in sodium citrate buffer. Therefore, the presence of salts was an important factor in degradation.

Because the rate of degradation varied with the base composition, incomplete degradation of DNA might have a

FIG. 4. Time course of P1 nuclease degradation of salmon sperm DNA. The DNA (200 μg) was degraded under standard conditions with 3 U of P1 nuclease. After the reactions were terminated by boiling, the nucleotide monophosphates (NMPs) released were determined as the nucleosides after phosphatase treatment. Symbols: ●, deoxycytidine monophosphate; ○, deoxyguanosine monophosphate; ▲, dAMP; △, thymidine monophosphate. A value of 100% nucleotide monophosphate released was equal to the amount of nucleotide monophosphate released after 20 h. The bar in panel B represents a change in the apparent ratio of 0.1. For abbreviations see the legend to Fig. 1.

large effect on the ratios of the nucleotides released. To test this hypothesis, the rate of release of the nucleotide monophosphates from salmon sperm DNA was determined (Fig. 4). Under the standard conditions used for degradation, deoxyguanosine monophosphate was released much more slowly than the other nucleotides, and the nucleotide ratios fluctuated widely early in the degradation. However, after 14 min, degradation was complete, and the nucleotide ratios did not change for up to 1 h (data not shown). Similarly, when low amounts of the nuclease were used in the standard degradation procedure, the ratios of the nucleotides released

FIG. 5. Time course of alkaline phosphatase degradation of deoxynucleotide monophosphates. The nucleotide monophosphates were obtained by P1 nuclease degradation of salmon sperm DNA. The degradation mixture included 3 U of phosphatase under standard conditions. Symbols: ●, dCyd; ○, dGuo; ▲, dAdo; △, dThd. A value of 100% nucleosides released was equal to the nucleosides obtained after 24 h. For abbreviations see the legend to Fig. 1.

also fluctuated greatly (data not shown). However, with between 1 and 2 U of nuclease per 50 µg of DNA, degradation was complete, and the ratios were constant. Therefore, incomplete degradation by P1 nuclease can be a major source of error in determining the nucleoside ratios.

Initially, degradations were performed with bacterial alkaline phosphatase from *E. coli* (18). However, good results were also obtained with bovine intestinal mucosa alkaline phosphatase. Because of its lower price and greater availability, the bovine enzyme was substituted in subsequent experiments. Like the P1 nuclease activity, the activity of the phosphatase showed specificity among the different nucleotide monophosphates, and its activity was sensitive to salts and pH (10). Under the conditions used for determining G+C content, deoxyguanosine monophosphate was hydrolyzed more slowly than the other deoxynucleotides (Fig. 5).

Therefore, the ratio of the nucleosides released early in degradation fluctuated greatly. After 4 h, degradation was complete, and the nucleoside ratios did not change for up to 24 h (data not shown). Likewise, after degradations with less than 1 U of phosphatase per 50 µg of DNA, the nucleoside ratios were variable (data not shown). However, no change in the ratios was observed at concentrations between 1 and 4 U of phosphatase per 50 µg of DNA. Thus, the nucleoside ratios were very sensitive to incomplete degradation by the phosphatase. Overdegradation, or a change in the ratios due to contaminating enzymes, was not observed with purified DNA.

Sample preparation. Degradation of DNA samples which contained high concentrations of salts produced variable results. Thus, DNA suspended in 1× SSC or 0.14 M phosphate buffer (pH 6.8) required dialysis prior to degradation (data not shown). Likewise, 30 mM Tris chloride (pH 8.5) also interfered with degradation. Thus, salts or buffers remaining in the samples were a major source of error during determination of G+C content.

To determine whether other sources of error were present, nucleic acid was isolated from cell pastes of "*Methanococcus aeolicus*" by three methods, and G+C contents were calculated from the dGuo/dThd ratios (Table 1). For each method, four samples were prepared. The G+C contents of the replicate samples were not significantly different. However, the G+C contents of the samples prepared by the HA method were slightly higher than the G+C contents of the samples prepared by the phenol and NaOH methods. Because the G+C contents of fragments of chromosomal DNA vary over a wide range, the HA method may have resulted in a slight enrichment of high-G+C-content DNA (37). However, this difference was very small, about 0.03 to 0.04 mol%, and can probably be neglected for most purposes. In addition, the 95% confidence levels for the phenol, NaOH, and HA samples were ± 0.043 , ± 0.022 , and ± 0.026 mol%, respectively. Thus, the determinations from the samples prepared by the phenol method were more variable than the determinations from the samples prepared by the other methods. Presumably, this variability resulted from a higher proportion of RNA or other contaminants in these samples. Importantly, the standard deviations of the determinations of the replicate NaOH samples were very close to the error associated with the chromatography itself. Thus, sample preparation and degradation did not introduce additional sources of error.

DNA samples were stored frozen for many years without a change in apparent G+C content. For three methanococcal DNAs stored for 2 years at -10°C in 10 mM Tris chloride (pH 7.4)–1 mM sodium EDTA, the G+C contents of the old and new samples, respectively, were as follows: 29.67 ± 0.07 and 29.67 ± 0.06 , 29.52 ± 0.06 and 29.61 ± 0.04 , and 29.23 ± 0.07 and 29.40 ± 0.06 mol%. Likewise, for purified yeast DNA stored at -20°C in 1× SSC for up to 10 years, the G+C content did not appear to change (data not shown). In contrast, a great deal of variability was encountered in samples stored for several years after degradation of the DNA.

G+C contents of DNAs. The G+C contents of selected eucaryotic and procaryotic DNAs were determined from the nucleoside ratios by using bacteriophage lambda as a standard (Table 2). This standard was chosen because its composition is known precisely from the DNA sequence (27). For salmon sperm, calf thymus, and *Methanococcus voltae*, highly purified DNAs were used. In these cases, the G+C contents calculated from all four nucleoside ratios were in

TABLE 1. G+C contents of replicate preparations of "*Methanococcus aeolicus*" DNA

Method of DNA preparation ^a	Yield (mg/g) ^b	Purity		G+C content (mol%) for:				
		Absorbance at 260 nm/absorbance at 280 nm	Amt of RNA (%) ^c	Sample 1 ^d	Sample 2	Sample 3	Sample 4	All samples ^e
Phenol	9.0	1.91	60–75	29.83 ± 0.04 ^f	29.87 ± 0.03	29.80 ± 0.03	29.86 ± 0.05	29.84 ± 0.05
NaOH	9.5	1.85	12–30	29.82 ± 0.02	29.85 ± 0.02	29.84 ± 0.02	29.82 ± 0.02	29.83 ± 0.02
HA	5.0	1.86	4–6	29.86 ± 0.03	29.89 ± 0.02	29.87 ± 0.03	29.85 ± 0.02	29.87 ± 0.03

^a DNA was prepared by one of three methods, as described in the text.^b Total amount of nucleic acid (in milligrams) per gram of cell dry weight.^c Range of percentages of RNA encountered in the four samples. The percentage of RNA was calculated as follows: $100 \times [\text{Ado}/(\text{dAdo} + \text{Ado})]$.^d Replicate samples (30 mg, dry weight) of cells. G+C contents were calculated from dThd/dGuo ratios. One degradation of each sample was performed, and the product of each degradation was analyzed four times. As determined by analysis of variance, the F value for a difference among the phenol (I) samples had a probability of <0.076%. The F value for a difference among the NaOH-chloroform (II) samples had a probability of <0.241. The F value for a difference among the HA (III) samples had a probability of <0.217. The F value for a difference among I, II, and III had a probability of <0.0127.^e Means ± standard deviations for all measurements determined for samples 1 through 4.^f Mean ± standard deviation.

good agreement with each other. For salmon sperm and calf thymus DNAs, the values were close to the average values cited in the literature, 44.0 and 44.4 mol%, respectively (18). However, the mole fraction of modified dCyd was somewhat lower than reported previously. Thus, for salmon sperm and calf thymus DNAs, the reported values for modified dCyd were 1.57 and 1.42 mol%, respectively. By using the indirect method described in this paper, the values found were 1.16 ± 0.09 and 1.03 ± 0.11 mol%, respectively. The source of this discrepancy is not obvious. One possibility is that the previously reported values are in error because they were based upon the weight of standard samples. If the standard samples were contaminated with salts or other nucleosides, these values would be overestimates. Modified dAdo has not been detected in these DNAs (18). In agreement with these observations, the amounts of modified dAdo were found to be 0.06 ± 0.04 and 0.02 ± 0.13 mol% in salmon sperm and calf thymus DNAs, respectively, values which were not significant.

Methanococcus voltae DNA is not modified, and its G+C content calculated from the nucleoside ratios can be compared with the value calculated from the apparent G+C content (M_a). By using this latter method, the G+C content was found to be 29.09 ± 0.04 mol%, which was in good agreement with the values obtained from the nucleoside ratios and the measurements of other investigators (7) (Table 2). In addition, the G+C contents of three other bacterial DNAs, which contained between 5 and 70% RNA, were determined (Table 2). Although the amount of dCyd could

not be measured, the G+C contents determined from the dThd/dGuo and dAdo/dGuo ratios were in good agreement with each other and with values published previously. Moreover, *Methanococcus voltae* and *Myxococcus xanthus* DNAs represent extremes in base composition. Thus, the method appears to be reliable over a wide range.

G+C contents from single colonies. Because only small amounts of DNA were required to determine G+C contents, it was possible to obtain sufficient DNA from a single bacterial colony. Colonies of *E. coli* were grown on LB plates for 28 h. When the colony diameter was about 2 mm, DNA was prepared by a modification of the NaOH method. About 1.8 µg of DNA was obtained from a single colony. For DNAs from four colonies from a single plate, the G+C contents were 50.31, 50.28, 50.17, and 50.25 mol%. These values were not significantly different from the value obtained from a broth culture of the same strain (50.28 mol%).

DISCUSSION

The HPLC method is a promising alternative for determining G+C contents. The major advantages of this method are given below. It is rapid. Analysis of a single sample requires less than 15 min, and multiple analyses can be performed easily. With an automatic injector, little operator attention is required. The analysis can be performed with standard equipment, which may also be used for other purposes in a laboratory. The analysis can be performed on small samples. For instance, G+C content can be deter-

TABLE 2. Calculation of G+C contents from nucleoside ratios^a

Source of DNA	G+C content calculated from the following apparent ratios:			
	dGuo/dThd	dGuo/dAdo	dCyd/dThd	dCyd/dAdo
Salmon sperm ^b	43.32 ± 0.00	43.39 ± 0.04	43.32 ± 0.00	43.39 ± 0.04
Calf thymus ^c	44.03 ± 0.05	44.06 ± 0.12	44.03 ± 0.05	44.06 ± 0.12
<i>Methanococcus voltae</i>	29.13 ± 0.05	29.11 ± 0.06	29.06 ± 0.05	29.04 ± 0.05
<i>E. coli</i> ^d	50.64 ± 0.08	50.84 ± 0.19	ND ^e	ND
<i>Myxococcus xanthus</i> ^f	67.55 ± 0.02	67.52 ± 0.18	ND	ND
<i>A. calcoaceticus</i> ^g	40.07 ± 0.09	40.07 ± 0.09	ND	ND

^a G+C contents were calculated from the averages of five measurements.^b Corrected for the modified dCyd value, which was 1.16 ± 0.09 mol%.^c Corrected for the modified dCyd value, which was 1.03 ± 0.11 mol%.^d Samples contained 5% RNA.^e ND, Not determined.^f Corrected for the modified dAdo value, which was 0.84 ± 0.10 mol%. Samples contained 70% RNA.^g Corrected for the modified dAdo value, which was 0.37 ± 0.05 mol%. Samples contained 40% RNA.

mined from the nucleosides after degradation of 0.5 µg of DNA. The analysis is not sensitive to contamination by RNA. Because dGuo and dThd are well separated from the ribonucleosides, a 10-fold excess of RNA has no effect on the determination. Likewise, if the G+C content is calculated from the dGuo/dThd ratio, modifications of dCyd and dAdo have no effect. The method is precise. The relative standard deviations of replicate measurements are generally close to 0.1%. Importantly, the major sources of error can be readily identified.

Three major sources of error were encountered in the HPLC method. (i) Comigration of the deoxynucleosides from the minor bases frequently interfered with the determinations. This error was eliminated by careful selection of the chromatographic conditions. In addition, this error may be tested for by determining the G+C content from more than one base ratio. Modification may also be tested for by determining the dCyd/dGuo and dThd/dAdo ratios. (ii) Equipment failure was a common source of error. Erratic pumping rates caused by leaking piston seals and inaccurate calibration of the pumps were both encountered. These problems were first detected as increases in the standard deviations of replicate samples and changes in the nucleoside ratios of standards over the course of a day. Likewise, the reversed-phase columns have a limited lifetime, and after several months of continuous usage, resolution may be lost. This error was detected by a change in the apparent nucleoside ratios of standards, an increase in column pressure, and visual inspection of the chromatograms. (iii) Reproducibility of degradation proved to be an important source of error. Incomplete degradation was detected by the appearance of peaks for the nucleotides on the chromatograms. This problem was encountered frequently with impure preparations, and it was largely due to inhibition of the enzymes by salt. However, when care was taken to ensure complete degradation, determinations of G+C contents were very reproducible.

The G+C contents determined by the HPLC method agree well with the values determined by other methods over a wide range. Moreover, the accuracy of this method may be superior to the accuracy of the buoyant density and thermal denaturation methods. Both of the latter methods are indirect, and their validity was established by demonstrating a high correlation with older chemical methods (5). However, the variability of the older chemical determinations was such that the accuracy was only tested over a fairly broad range. For instance, the 95% confidence limits for the accuracy of both methods compared with the chemical determinations span about 8 mol% (5). Thus, it is not surprising that these methods frequently give somewhat different values. In contrast, the accuracy of the HPLC method depends greatly on the choice of standards. The availability of DNA of known sequence and high purity greatly increases the accuracy.

The HPLC method can be applied to samples heavily contaminated with RNA and small samples. From samples of about 40 mg (dry weight) of cells, three methods of DNA preparation yield very similar results. However, the phenol method contributes somewhat to the overall error in the determination of the G+C content, and the HA method has a very small systematic error. Because only microgram amounts of DNA are required, this method should be suitable for measuring the G+C content of DNA from a single bacterial colony. In this case, extraction of the DNA is likely to be a critical parameter. While procedures for DNA isolation were not investigated in detail, it was possible to obtain the G+C contents of DNAs from single colonies of

E. coli. Therefore, the HPLC method may have potential applications in bacterial identification, as well as in systematics.

ACKNOWLEDGMENTS

This work was supported by grant DMB83-51355 from the National Science Foundation, by the Georgia Power Company, and by the University of Georgia Research Foundation.

We thank Bob Ivarie for calculating the G+C content of bacteriophage lambda DNA.

LITERATURE CITED

1. Bak, A. L., J. F. Atkins, and S. A. Meyer. 1972. Evolution of DNA base compositions in microorganisms. *Science* **175**:1391-1393.
2. Beji, A., D. Izard, F. Gavini, H. Leclerc, M. Leseine-Delstanché, and J. Krembel. 1987. A rapid chemical procedure for isolation and purification of chromosomal DNA from Gram-negative bacilli. *Anal. Biochem.* **162**:18-23.
3. Cashion, P., M. A. Holder-Franklin, J. McCully, and M. Franklin. 1977. A rapid method for the base ratio determination of bacterial DNA. *Anal. Biochem.* **81**:461-466.
4. DeLey, J. 1967. The quick approximation of DNA base composition from absorbancy ratios. *Antonie van Leeuwenhoek J. Microbiol. Serol.* **33**:203-208.
5. DeLey, J. 1970. Reexamination of the association between melting point, buoyant density, and chemical base composition of deoxyribonucleic acid. *J. Bacteriol.* **101**:738-754.
6. Ehrlich, M. 1979. Separation of six DNA bases by ion pair-reversed phase high pressure liquid chromatography. *J. Chromatogr. Sci.* **17**:531-534.
7. Ehrlich, M., M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, and C. W. Gehrke. 1985. DNA methylation in thermophilic bacteria: N⁴-methylcytosine, 5-methylcytosine, and N⁶-methyladenine. *Nucleic Acids Res.* **13**:1399-1412.
8. Elton, R. A. 1973. The relationship of DNA base composition and individual protein composition in micro-organisms. *J. Mol. Evol.* **2**:263-276.
9. Elton, R. A. 1974. Theoretical models for heterogeneity of base composition in DNA. *J. Theor. Biol.* **45**:533-553.
10. Fernley, H. N. 1971. Mammalian alkaline phosphatases, p. 417-447. *In* P. D. Boyer (ed.), *The enzymes*, vol. 4, 3rd ed. Academic Press, Inc., New York.
11. Fredericq, E., A. Oth, and F. Fontaine. 1961. The ultraviolet spectrum of deoxyribonucleic acids and their constituents. *J. Mol. Biol.* **3**:11-17.
12. Fujimoto, M., A. Kuninaka, and H. Yoshino. 1974. Purification of a nuclease from *Penicillium citrinum*. *Agric. Biol. Chem.* **38**:777-783.
13. Fujimoto, M., A. Kuninaka, and H. Yoshino. 1974. Identity of phosphodiesterase and phosphomonoesterase activities with nuclease P₁ (a nuclease from *Penicillium citrinum*). *Agric. Biol. Chem.* **38**:785-790.
14. Fujimoto, M., A. Kuninaka, and H. Yoshino. 1974. Substrate specificity of nuclease P₁. *Agric. Biol. Chem.* **38**:1555-1561.
15. Gehrke, C. W., R. A. McCune, M. A. Gama-Sosa, M. Ehrlich, and K. C. Kuo. 1984. Quantitative reversed-phase high-performance liquid chromatography of major and modified nucleosides in DNA. *J. Chromatogr.* **301**:199-219.
16. Huang, P. C., and E. Rosenberg. 1966. Determination of DNA base composition via depurination. *Anal. Biochem.* **16**:107-113.
17. Ko, C. Y., J. L. Johnson, L. B. Barnett, H. M. McNair, and J. R. Vercellotti. 1977. A sensitive estimation of the percentage of guanine plus cytosine in deoxyribonucleic acid by high performance liquid chromatography. *Anal. Biochem.* **80**:183-192.
18. Kuo, K. C., R. A. McCune, C. W. Gehrke, R. Midgett, and M. Ehrlich. 1980. Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.* **8**:4763-4776.
19. Lesk, A. M. 1973. On hypothesized selective pressure by u.v. on

- DNA base compositions. *J. Theor. Biol.* **40**:201–202.
20. Mandel, M., and J. Marmur. 1968. Use of ultraviolet absorbance-temperature profile for determining the guanine plus cytosine content of DNA. *Methods Enzymol.* **12**:195–206.
 21. Mandel, M., C. L. Schildkraut, and J. Marmur. 1968. Use of CsCl density gradient analysis for determining the guanine plus cytosine content of DNA. *Methods Enzymol.* **12**:184–195.
 22. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 23. Marmur, J., and P. Doty. 1962. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* **5**:109–118.
 24. Mischke, C. F., and E. Wickstrom. 1980. Deoxynucleoside composition of DNAs and modified nucleoside composition of tRNAs determined at nanomole sensitivity by reversed-phase liquid chromatography. *Anal. Biochem.* **105**:181–187.
 25. Muto, A., and S. Osawa. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* **84**:166–169.
 26. Presber, W., C. Schroeder, and D. H. Krueger. 1976. The evolution of mutation rates via DNA base composition and mutator genes: a proposition. *J. Theor. Biol.* **59**:353–360.
 27. Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen. 1982. Nucleotide sequence of bacteriophage λ DNA. *J. Mol. Biol.* **162**:729–773.
 28. Schildkraut, C. L., J. Marmur, and P. Doty. 1962. Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. *J. Mol. Biol.* **4**:430–443.
 29. Singer, C. E., and B. N. Ames. 1970. Sunlight ultraviolet and bacterial DNA base ratios. *Science* **170**:822–826.
 30. Tamaoka, J., and K. Komagata. 1984. Determination of DNA base composition by reversed-phase high-performance liquid chromatography. *FEMS Microbiol. Lett.* **25**:125–128.
 31. Ulitzur, S. 1972. Rapid determination of DNA base composition by ultraviolet spectroscopy. *Biochim. Biophys. Acta* **272**:1–11.
 32. Wang, S. Y. 1968. The determination of nucleic acid base composition by chemical reactivity. *Methods Enzymol.* **12**:178–184.
 33. Wang, S. Y., and J. M. Hashagen. 1964. The determination of the base composition of deoxyribonucleic acids by bromination. *J. Mol. Biol.* **8**:333–340.
 34. Wetmur, J. G., and N. Davidson. 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* **31**:349–370.
 35. Whitman, W. B., J. Shieh, S. Sohn, D. S. Caras, and U. Premachandran. 1986. Isolation and characterization of 22 mesophilic methanococci. *Syst. Appl. Microbiol.* **7**:235–240.
 36. Woese, C. R., and M. A. Bleyman. 1972. Genetic code limit organisms—do they exist? *J. Mol. Evol.* **1**:223–229.
 37. Yamagishi, H. 1970. Nucleotide distribution in the DNA of *Escherichia coli*. *J. Mol. Biol.* **49**:603–608.