Determination of guanine-plus-cytosine content of bacterial DNA by dual-laser flow cytometry

CYNTHIA A. SANDERS,1,* DAVID M. YAJKO,1 WILLIAM HYUN,2 RICHARD G. LANGLOIS,3 PATRICIA S. NASSOS,1 MACK J. FULWYLER2 and W. KEITH HADLEY1

1,2Department of Laboratory Medicine, University of California, San Francisco General Hospital Medical Center1 and Laboratory for Cell Analysis,2 San Francisco, CA 94110, USA
3Biomedical Sciences Division, Lawrence Livermore National Laboratory, University of California, Livermore, CA 94550, USA

(Received 23 March 1989; revised 15 August 1989; accepted 27 September 1989)

A dual-laser flow cytometer was used to analyse different species of bacteria for the molar percentage of guanine-plus-cytosine (% G + C) without the need for DNA extraction or purification. Ethanol-fixed bacterial cells were stained with a combination of DNA-specific fluorochromes, Hoechst 33258 and chromomycin A3, which bind to AT- and GC-rich regions of DNA, respectively. A linear relationship (r = 0.99) was demonstrated between the log of the ratio of chromomycin A3 to Hoechst 33258 fluorescence and the log of the % G + C as determined by thermal denaturation (Tm) or buoyant density centrifugation (Bd) methods. Linearity was maintained for all bacterial species tested over the range of 28–67 % G + C. A standard curve was constructed using five strains whose % G + C had been determined by other methods. From the equation describing this line, the % G + C values of nine other strains with known DNA base composition, together with the five strains used to construct the curve, were calculated using the chromomycin A3 to Hoechst 33258 ratio and were in agreement with values obtained by Tm, Bd or HPLC. The reproducibility of flow cytometric analysis (mean error 0.7 % G + C) compared well with the reproducibility of other methods. Mixtures containing two species were also analysed. Two cell populations could be discerned in mixtures containing two species which differed in base composition by as little as 4 % G + C. Dual-laser flow cytometric analysis of stained bacteria is a rapid, simple and accurate method for determining the % G + C of bacterial DNA and can be used to distinguish populations of bacteria with differing % G + C content.

Introduction

Determination of the molar percentage of guanine-plus-cytosine (% G + C) of DNA is often useful in the characterization of bacteria. Several methods exist for the determination of % G + C (Johnson, 1985), the most common procedures being thermal denaturation (Tm) (Marmur & Doty, 1962) and buoyant density centrifugation (Bd) (Schildkraut et al., 1962). In recent years, HPLC has been employed to determine the % G + C of bacterial DNA (Ko et al., 1977; Mesbah et al., 1989). These and most other analytical methods require DNA to be extracted from the bacterial cell and purified, which is both laborious and time consuming. A new method to determine % G + C using flow cytometry is described here. It has the potential to allow base composition analyses to be done directly on individual cells of a bacterial sample without extraction of DNA.

Flow cytometry is a versatile technique of cellular analysis that permits a variety of parameters and constituents of cells to be analysed rapidly. Mammalian chromosomes have been studied by flow cytometry using the DNA-specific fluorochromes Hoechst 33258 (H33258) and chromomycin A3 (CA3). H33258 fluorochrome has been shown to bind preferentially to regions of DNA rich in AT base pairs (Comings, 1975; Muller & Gautier, 1975) while CA3 binds preferentially to GC-rich regions of DNA (Behr et al., 1969). Using the two fluorochromes in combination, chromosomes can be analysed and separated based upon overall DNA base composition (Gray et al., 1987; Lalande et al., 1985; Langlois et al., 1980; van den Engh et al., 1985).

Van Dilla et al. (1983) adapted this method of chromosome analysis to the study of bacteria. Using flow
cytometry, they analysed six species of bacteria with differing DNA base composition and were able to demonstrate a relationship between the CA3:H33258 fluorescence ratio and % G + C. The data showed that the CA3:H33258 fluorescence ratio increases with increasing % G + C. Although the authors did not mathematically define this relationship, their findings suggested that it may be possible to calculate the % G + C value for a variety of bacteria using the CA3:H33258 fluorescence ratio.

The goals of the present study were: (i) to characterize mathematically the relationship between the CA3:H33258 fluorescence ratio and the % G + C of bacterial DNA; (ii) using this relationship, to construct a standard curve which spans a wide range of % G + C of bacteria; (iii) to determine whether this standard curve could be used to obtain a reproducible % G + C value from the CA3:H33258 fluorescence ratio; (iv) to compare the % G + C values obtained using flow cytometry with published values obtained by $T_m$, Bd and HPLC.

Table 1. Comparison of % G + C content of 14 bacterial strains determined by dual-laser flow cytometry with % G + C values from the literature

<table>
<thead>
<tr>
<th>Strain analysed*</th>
<th>(Other strain designations)</th>
<th>% G + C calculated from CA3:H33258 ratio</th>
<th>% G + C value in the literature</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteroides fragilis VPI 2553</td>
<td>(ATCC 25285)</td>
<td>42.6 ± 0.8</td>
<td>42</td>
<td>$T_m$</td>
<td>(Johnson, 1978)</td>
</tr>
<tr>
<td>B. thetaiotaomicron VPI 5482</td>
<td>(ATCC 29148)</td>
<td>42.7 ± 0.9</td>
<td>42</td>
<td>$T_m$</td>
<td>(Johnson, 1978)</td>
</tr>
<tr>
<td>Clostridium butyricum VPI 3266</td>
<td>(ATCC 19398)</td>
<td>25.4 ± 0.7</td>
<td>28</td>
<td>$T_m$</td>
<td>(Cammmis &amp; Johnson, 1971)</td>
</tr>
<tr>
<td>C. curnsi VPI 1635B</td>
<td>(ATCC 25777)</td>
<td>26.0 ± 1.2</td>
<td>28</td>
<td>$T_m$</td>
<td>(Johnson &amp; Francis, 1975)</td>
</tr>
<tr>
<td>Escherichia coli ATCC 11775†</td>
<td>50.3 ± 0.1</td>
<td>51.7</td>
<td>$T_m$</td>
<td>(Starr &amp; Mandel, 1969)</td>
<td></td>
</tr>
<tr>
<td>E. coli CDC K-12</td>
<td>50.4 ± 0.4</td>
<td>50.8</td>
<td>$T_m$</td>
<td>(Starr &amp; Mandel, 1969)</td>
<td></td>
</tr>
<tr>
<td>Klebsiella pneumoniae CDC III†</td>
<td>55.5 ± 0.9</td>
<td>55.6</td>
<td>$T_m$</td>
<td>(Starr &amp; Mandel, 1969)</td>
<td></td>
</tr>
<tr>
<td>Lactobacillus acidophilus VPI 6033</td>
<td>(ATCC 19992)</td>
<td>34.3 ± 0.2</td>
<td>33</td>
<td>$T_m$</td>
<td>(Johnson et al., 1980)</td>
</tr>
<tr>
<td>Propionibacterium acnes VPI 0389</td>
<td>(ATCC 6919†)</td>
<td>58.7 ± 0.7</td>
<td>59</td>
<td>$T_m$</td>
<td>(Johnson &amp; Cummins, 1972)</td>
</tr>
<tr>
<td>Proteus vulgaris CDC 636-73†</td>
<td>38.3 ± 0.4</td>
<td>37%</td>
<td>$T_m$</td>
<td>(Slatowsky et al., 1977)</td>
<td></td>
</tr>
<tr>
<td>Pseudomonas aeruginosa ATCC 8707</td>
<td>67.2%</td>
<td>67.3</td>
<td>Bd</td>
<td>(Mandel, 1966)</td>
<td></td>
</tr>
<tr>
<td>Serratia marcescens CDC SM6-Red</td>
<td>54.3 ± 0.8</td>
<td>57.5%</td>
<td>$T_m$</td>
<td>(Slatowsky et al., 1977)</td>
<td></td>
</tr>
<tr>
<td>Staphylococcus aureus ATCC 12600</td>
<td>(NCTC 8532)†</td>
<td>30.5 ± 0.2</td>
<td>31.2</td>
<td>$T_m$</td>
<td>(Silvestri &amp; Hill, 1965)</td>
</tr>
<tr>
<td>Vibrio parahaemolyticus ATCC 17802</td>
<td>43.4 ± 0.7</td>
<td>45.6</td>
<td>$T_m$</td>
<td>(Colwell, 1970)</td>
<td></td>
</tr>
</tbody>
</table>

* VPI, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA; ATCC, American Type Culture Collection, Rockville, Maryland, USA; CDC, Centers for Disease Control, Atlanta, Georgia, USA; NCTC, National Collection of Type Cultures, London, UK.
† Strains used to construct standard curve.
‡ Value calculated using the formula % G + C = G/(A + C).
§ Value calculated using the formula % G + C = (G + C)/(2A + G + C).
∥ Mean value based on one determination.
* Personal communication, D. Brenner & A. Steigerwalt, Centers for Disease Control, Atlanta, Georgia, USA.

Methods

Bacterial strains. Bacterial strains used in this investigation and their sources are listed in Table 1. The % G + C of the strains was previously determined by one or more of the following methods: Bd, $T_m$ or HPLC (Colwell, 1970; Cummins & Johnson, 1971; Johnson, 1978; Johnson & Cummins, 1972; Johnson & Francis, 1975; Johnson et al., 1980; Ko et al., 1977; Mandel, 1966; Silvestri & Hill, 1965; Starr & Mandel, 1969; Voss, 1970).

Preparation of bacteria for flow cytometric analysis. Aerobic and facultatively anaerobic bacteria were grown at 35°C for 20-24 h in 5-0 ml Todd–Hewitt broth medium (Difco). Serial tenfold dilutions of the overnight cultures were prepared in sterile water and 0-1 ml samples were plated out to obtain viable bacterial counts. The Todd–Hewitt broth cell suspensions were centrifuged, then resuspended in sterile water and fixed by adding 95% (v/v) ethanol to obtain a final ethanol concentration of 70% (v/v). Ethanol-fixed bacterial suspensions were stored at 4°C for up to 7 weeks before staining.

Anaerobic bacteria were grown on Brucella agar plates (Difco) supplemented with 5% (v/v) sheep blood, vitamin K1 (1 mg per 100 g solution) and haemin (0.5 mg per 100 g solution) and incubated anaerobically for 48 h at 35°C. Bacterial growth was scraped from the surface of the plate and suspended in 5-Om1 Todd-Hewitt broth cell suspensions were centrifuged, then resuspended in sterile water and fixed by adding 95% (v/v) ethanol to obtain a final ethanol concentration of 70% (v/v). Ethanol-fixed bacterial suspensions were stored at 4°C for up to 7 weeks before staining.

Methods used to construct the standard curve: (i) to mathematically define the relationship between the CA3:H33258 fluorescence ratio and % G + C of bacteria; (ii) to determine the relationship between the CA3:H33258 fluorescence ratio and % G + C; (iii) to compare the % G + C values obtained using flow cytometry with published values obtained by $T_m$; Bd and HPLC.
and ethanol fixation of cells were done as described above for aerobic strains.

**Bacterial DNA staining procedure.** A 1.0 ml portion of ethanol-fixed cell suspension was centrifuged and the cell pellet was resuspended in an equal volume of Tris/magnesium/saline (TMS) buffer (10 mM-Tris pH 7.2, 1.5 mM-MgCl₂, 150 mM-NaCl). A sample of resuspended cells was diluted in TMS buffer to give a final concentration equivalent to approximately 10⁷ c.f.u. ml⁻¹ in a total volume of 1.0–1.1 ml. The volume of the cell suspension added to the buffer was based on previously determined viable counts. To the resulting suspension, 0.01 ml each of aqueous solutions of 0.3 mM-H33258 (Calbiochem–Behring) and 1 mM-CA3 (Calbiochem–Behring) were added. The bacterial cells were stored in the staining solution and shielded from light for 24 h at 4°C. Adequate staining of bacteria, however, could be obtained in as little as 45 min at 4°C.

**Flow cytometry.** The ethanol-fixed bacterial cells stained with the two fluorochromes were analysed using a FACS 440 flow cytometer (Becton Dickinson) equipped with dual argon-ion lasers (Spectra Physics). The primary and secondary lasers were used to excite the H33258 and CA3 fluorochromes at excitation wavelengths of 351 nm and 457 nm, respectively. Laser power levels were 200 mW at 351 nm and 150 mW at 457 nm wavelengths. Two 480 nm long pass interference filters (Omega Optical) were used to filter the emitted fluorescence of the H33258 and the CA3. The instrument was aligned using 4-8 pm fluorescein microspheres (Polysciences) for the 351 nm laser line and 1.96 µm fluorescein microspheres (Polysciences) for the 457 nm laser line. The instrument was standardized by adjusting instrument parameters so that the peak H33258 and CA3 fluorescence signals from stained E. coli ATCC 11775 cells were both centred in fluorescence channel number 50.

A total of 10000 cells were analysed in each sample at a rate of 700–800 cells s⁻¹ through a 60 µm nozzle. In experiments where dilutions of E. coli ATCC 11775 were tested, it was necessary to increase the relative sample pressure to obtain an acceptable rate of cell analysis.

**Flow cytometric data collection and analysis.** Flow cytometric data in list mode were collected and analysed using a DEC VAX 11/750 computer system and Electric Desk data analysis program (developed by Wayne Moore, Department of Genetics, Stanford University, Stanford, California, USA). The program presents the data as a bivariate contour plot showing the relative fluorescence of CA3 versus relative fluorescence of H33258 versus the number of cells. An example of the presentation of flow cytometric data as a contour plot is shown in Fig. 1(a). The location of a bacterial cell population on the bivariate contour plot depends on DNA base composition and cellular DNA content. The radial distance of a population of cells from the origin is a function of the amount of DNA per cell, while angular displacement from the H33258 axis is proportional to the % G + C (Van Dilla et al., 1983).

The data used to create Fig. 1(a) can also be represented as a frequency distribution curve. Frequency distribution curves which show the ratio of CA3 to H33258 relative fluorescence versus the number of cells were generated using Energy Transfer (data analysis program developed by Marty Bigos, Laboratory for Cell Analysis, University of California, San Francisco, California, USA). The log of the mean CA3:H33258 fluorescence ratio associated with a frequency distribution curve is the value that was used to calculate the % G + C. An example of bivariate contour plot data presented as a frequency distribution curve of the fluorescence ratio is shown in Fig. 1(b).

**Results**

**Relationship of CA3 : H33258 fluorescence ratio to % G + C**

Initially, the relationship between the CA3:H33258 fluorescence ratio and % G + C determined by conventional methods was examined. Fourteen bacterial strains whose % G + C was determined by Tₘ, Bd or HPLC (Table 1) were stained with CA3 and H33258 and analysed by dual-laser flow cytometry. Mean CA3:H33258 fluorescence ratios were plotted against the % G + C values reported in the literature for each strain. A linear relationship was demonstrated (r = 0.99, P < 0.001) between the log of the CA3:H33258 fluorescence ratio and the log of the % G + C of bacterial DNA over the range of 28–67% G + C (Fig. 2). The line in Fig. 2 is defined by the equation: log₁₀ fluorescence ratio = 3.696(log % G + C) - 6.331. A plot of the residuals was constructed which indicated that logarithmic transformation of the data was appropriate.
Flow cytometric determination of % G + C

After the relationship between the CA3:H33258 fluorescence ratio and % G + C was established, we tested the ability of the flow cytometric method to determine the % G + C of bacterial strains treated as unknowns. To accomplish this, a standard curve was constructed using five strains of bacteria whose % G + C had been previously determined by conventional methods. The strains chosen as standards, Staph. aureus ATCC 12600, Prot. vulgaris CDC 636-73, E. coli ATCC 11775, K. pneumoniae CDC II and Prop. acnes VPI 0389, represent a wide range of DNA base composition. The standard curve was constructed using the log of the mean of three CA3:H33258 fluorescence ratio values for each bacterial strain versus the log of the % G + C determined by Tm, Bd or HPLC. Using the equation of the regression line obtained from these data, the % G + C was calculated for the other nine strains listed in Table 1. The % G + C values were also calculated for the five strains used as standards in the construction of the line. The % G + C values obtained by flow cytometry for all fourteen strains are shown in Table 1 along with the % G + C value for each strain determined by other methods. Each flow cytometric % G + C value given in Table 1, with the exception of Ps. aeruginosa ATCC 8707, represents the mean of three determinations obtained on separate days. A determination consists of the mean value obtained by analysing a sample three times on the same day. Ps. aeruginosa ATCC 8707 appeared to deteriorate in the 70% ethanol during the 6 weeks which elapsed between analyses. For this reason, the % G + C value given in Table 1 for this strain is based upon one determination.

Reproducibility of flow cytometric % G + C values

The mean standard error associated with determinations in which the same preparation was analysed three times on the same day was 0-7% G + C. The mean error associated with three % G + C determinations on the same ethanol-fixed cell preparation analysed on three separate days was also 0-7% G + C for the 14 strains tested (Table 1). Seven different culture preparations of K. pneumoniae CDC II were also ethanol-fixed and stained and each preparation analysed once on the same day. The mean standard error associated with the % G + C values of the seven preparations was 0-3% G + C, a value which was less than that of the single preparation of K. pneumoniae CDC II analysed three times on a single day. These results demonstrate that (i) the magnitude of error associated with growing and staining different batches of the same strain is small, and (ii) the degree of imprecision associated with instrument variability can be reduced by running the same preparation several times.

Effect of bacterial cell concentration on % G + C determinations

Serial tenfold dilutions of an ethanol-fixed suspension of E. coli ATCC 11775 were prepared and 1-0 ml samples stained with CA3 and H33258. Portions of these samples (20–70 µl) were then analysed in the flow cytometer. Preparations which contained the equivalent of 2 × 10^7, 2 × 10^6 and 2 × 10^5 c.f.u. E. coli ATCC 11775 ml⁻¹ formed clearly discernible cell populations on the data contour plots. Background fluorescence began to appear on the contour plot at 2 × 10^6 c.f.u. ml⁻¹ and increased as the cell concentration decreased, such that at 2 × 10^4 c.f.u. ml⁻¹, a discrete cell population was barely discernible. At concentrations of ≤2 × 10^5 c.f.u. ml⁻¹, only background fluorescence was detected and contour plot patterns of these data were virtually the same as those of preparations containing no bacteria. Furthermore, as the cell concentration decreased, there was a decrease in the CA3:H33258 fluorescence ratio. This resulted in a corresponding decrease in calculated % G + C values from 51-6 to 47-2 to 46-9 at concentrations of 2 × 10^7, 2 × 10^6 and 2 × 10^5 c.f.u. ml⁻¹, respectively.

Analysis of mixtures of two bacterial strains

Mixtures containing two ethanol-fixed preparations of bacterial strains were stained and analysed. Pairs of strains were chosen such that decreasing differences in % G + C values existed between the two strains. The contour plot representations of the results are shown in Fig. 3. Two stained cell populations could be distinguished when the pairs V. parahaemolyticus ATCC 17802...
(45.6% G + C) and Staph. aureus ATCC 12600 (31.2% G + C) (Fig. 3a) and Prop. vulgaris CDC 636-73 (37% G + C) and Staph. aureus ATCC 12600 (Fig. 3b) were analysed. When the mixture of E. coli ATCC 11775 (51% G + C) and K. pneumoniae CDC II (55% G + C) was analysed, the contour plot (Fig. 3c) showed two barely discernible cell populations that had considerable overlap of their distributions. Two populations could not be discerned when the combination of Staph. aureus ATCC 12600 (31.2% G + C) and L. acidophilus VPI 6033 (33% G + C) was analysed (Fig. 3d). The populations seen at the top of Figs 3(a), 3(b) and 3(d) represent brightly-stained cells whose relative fluorescence intensities are off the scale used in these plots. These cells are most likely aggregates of bacteria, because when a logarithmic scale is used they have the same base composition as those populations shown in Fig. 3.

**Discussion**

This study demonstrates that the % G + C of bacterial DNA can be determined with considerable precision using dual-laser flow cytometry without extraction of DNA. The results show that the reproducibility of % G + C determinations by flow cytometry is comparable to the commonly used T<sub>m</sub> and Bd methods. The error (SEM) associated with % G + C determinations is reported to be about 1% G + C when T<sub>m</sub> or Bd is the analysis method (Mandel et al., 1968; Marmur & Doty, 1962), while the error associated with flow cytometric determinations of % G + C was found to be 0.7%. Ko et al. (1977), reported a relative standard deviation associated with replicate determinations of 1.3 to 1.7% G + C using HPLC. After our manuscript was submitted, Mesbah et al. (1989) reported obtaining improved precision in determining % G + C by HPLC such that the relative standard deviation associated with replicate measurements was reduced to approximately 0.1% G + C. Although the reproducibility of HPLC appears to be somewhat greater than that of flow cytometry, sample preparation required for HPLC analysis is considerably more laborious than that required for flow cytometric determination of % G + C.

The accuracy of % G + C values determined using flow cytometry is difficult to assess. Comparisons were made between % G + C values determined using T<sub>m</sub> or Bd and values determined by flow cytometry. Discrepancies exist in % G + C values when different methods or even the same method is used to analyse bacterial DNA from the same strain. For example, the % G + C of Prop. acnes VPI 0389 was reported to be 59% when analysed by T<sub>m</sub> (Johnson & Cummins, 1972), 57% (Ko et al., 1977) and 60.8% (Voss, 1970) when Bd was used, and 60% and 59.9% when analysed by HPLC (Ko et al., 1977). The value we obtained using flow cytometry was 58.7%. Among the 14 strains of bacteria that we analysed, there was generally good agreement between % G + C values determined by flow cytometry and values determined using T<sub>m</sub>, Bd and HPLC. In those cases where values differed, it was unclear whether the difference was due to the inaccuracy of the flow cytometric method, inaccuracies associated with other methods, or both.

Several factors influence the reliability of % G + C values determined using flow cytometry. Our results suggest that the reproducibility of determinations of % G + C using the FACS 440 appears to be affected primarily by instrument alignment and standardization. We found that reproducibility is improved by using commercially available microspheres for alignment. Reproducibility was further improved by standardizing...
the instrument prior to sets of analyses by adjusting instrument parameters so that the peak fluorescence of one of our strains (E. coli ATCC 11775) was positioned in specific channels.

To obtain reliable results, it is also important to prepare a standard curve just prior to performing sets of analyses. Thus, the equation given earlier in this paper for the data in Fig. 2 should not be used with data obtained in other experiments. We recommend that a standard curve be constructed using bacterial strains with known base composition and should include bacteria which span a wide range of \( \% \) G + C. The standard curve must be linear over the range of \( \% \) G + C being tested.

Our results also indicate that the concentration of the bacterial cell suspension that is used may affect the value of the \( \% \) G + C that is obtained using the FACS 440. As the cell concentration was decreased from \( 2 \times 10^7 \) to \( 2 \times 10^5 \) c.f.u. ml\(^{-1}\), the CA3:H33258 fluorescence ratio decreased. One explanation for this finding is related to the necessity to increase the relative sample pressure when analysing preparations which contain low cell concentrations. Increased sample pressure may decrease the CA3:H33258 fluorescence ratio either by increasing the background relative to the cellular fluorescence or by changing the ability to detect one of the two fluorescent signals. Consequently, when analysing samples with cell concentrations < \( 10^7 \) c.f.u. ml\(^{-1}\), it may be necessary to decrease the relative sample pressure or to use a bacterial cell concentration equivalent to that of the bacterial standards to obtain a reliable \( \% \) G + C value using this instrument.

It is not yet clear what effect the presence of more than one bacterial species has on the reliability of \( \% \) G + C values obtained using flow cytometry. We were able to show that preparations containing mixtures of two bacterial species formed discernible cell populations when \( \% \) G + C values differed by as little as 4 percentage points. These results suggest that \( \% \) G + C determinations may be feasible on samples which contain mixtures of two, and possibly more than two, bacterial species when \( \% \) G + C values differ by 4 percentage points or more; further work is needed to determine the reliability of \( \% \) G + C values obtained from analysis of mixtures of bacteria. For this reason, we recommend that flow cytometric determination of \( \% \) G + C be done on samples containing only one bacterial species.

The major advantage of flow cytometric determinations of \( \% \) G + C is that bacterial samples may be analysed without extraction of DNA. Preparation of samples for flow cytometric analysis typically requires about one hour and the actual analysis time using the FACS 440 is less than 1 min per sample. This allows for rapid analysis of batches of samples.

This study was supported in part by an award from the Program for Analytical Cytology, Lawrence Livermore National Laboratory, and the University of California, San Francisco, USA.

References


Determination of $\% G + C$ by flow cytometry


