Bull Yamaguchi Med Sch 31 : 13-20, 1984

A Look-up Table for Coefficient of Variation in DNA Flow Cytometry[1]

Manabu Takahashi and Tomoyuki Murakami

Department of Pathology, Yamaguchi University School of Medicine, Ube, Yamaguchi 755, Japan
(Received November 6, 1984)

Abstract A look-up table for estimation of the coefficient of variation (CV) during flow cytometry of cellular DNA content distribution was made by a computer simulation of the shape of the G1 fluorescence peak. The basic notion is that the accuracy of measurement is indicated by the sharpness of the G1 fluorescence peak, and in contrast to the well-known FWHM method (full width of the peak at half maximal ordinate value method), it utilizes the height of three channels at the G1 peak; channels which are least affected by the presence of early S cells or cell debris. The table is two-dimensional and requires entry of the channel number of the G1 peak and of a bluntness index which is defined as the ratio of mean height of two adjacent channels to the height of the peak.

Key Words: Look-up table, coefficient of variation, CV, DNA content distribution, flow cytometry

Introduction

A coefficient of variation (CV) indicates the accuracy of DNA content measurements made by flow cytometry and every effort should be made to reduce the CV. The guiding principle is to obtain a sharp and slender G1 fluorescence peak. However, a visual inspection of the histogram is liable to give a deceptive impression of the CV which is influenced by the location of the peak and by the vertical scale used to plot the data. To facilitate immediate estimation of the measurement CV, we present here a table that can be used at the experimental site. Use of this table reduces the need for an on-line computer for CV estimation during instrument adjustment.[1),2),3),4),5)]

The most favored method for manual CV estimation is the FWHM (a full width of the peak at half maximal ordinate value) method[6)]. This method is based on the assumption that the shape of the G1 fluorescence peak is unchanged by the presence of early S cells or by grouping continuous data into discrete channels. The three-

1 [1]This work was supported in part by the Grant-in-Aid for Cancer Research (59-12) from the Ministry of Health and Welfare.

Presented at US-Japan Cooperative Cancer Research Program on the Application of Cytology Automation in Cancer Cytology and Cell Biology, February 1983, held in Memorial Sloan-Kettering Cancer Center.

channel method that we describe here is an alternative which serves for the same purpose but utilizes the height of three channels at the G1 peak, instead of the width, to measure the sharpness (or bluntness) of the peak. It has the advantage that these channels are least affected by the presence of early S cells or cell debris. This paper describes the method how the look-up table was made and compares the results obtained with the FWHM method.

Methods

A variety of fluorescence histograms of known CV's were synthesized numerically for cell populations, assuming a rectangular DNA distribution for cells in the S and G2M regions (Fig. 1) using a multi-Gaussian model as described previously[3], in order to correlate the measurement CV with the bluntness of the G1 peak.

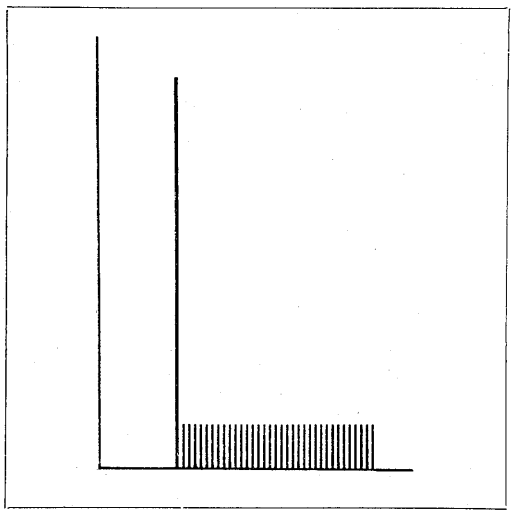

Fig. 1 A model of DNA distribution: The model assumes that the cells in the S and G2M regions are evenly distributed (rectangular distribution) whereas the G1 DNA content is not distributed (like the delta function of Dirac). In this model, the G1 DNAcontent, x_1, as well as the G1 cell fraction, F_{G1}, can be changed and cell density in other phases is determined accordingly. Based on this model and using a multiple Gaussian simulation technique, fluorescence histograms were synthesized to examine effects of early S cells on the shape of the G1 peak.

The bluntness index (b_n) of the G1 peak which is at the n-th channel was defined by

$$b_n = \frac{1}{2}(h_{n-1} + h_{n+1})/h_n \quad (1)$$

where h_{n-1}, h_n, and h_{n+1} are the height of histogram elements (cell numbers) in channels, n-1, n and n+1.

Fluorescence Histogram for G1 Cells: The probability density for a cell with DNA content x_i to be scored at fluorescence intensity y is given by

$$p_{cv}(x_i, y) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{1}{2}\left(\frac{x_i - y}{\sigma_i}\right)^2\right) \quad (2)$$

$$\text{where} \quad \sigma_i = CV.x_i \quad (3)$$

However, each fluorescence channel has a finite width so that all cells whose fluorescence intensities are within $n-1 < y \leqq n$ are grouped in channel n. Thus, in the absence of S-phase cells, the number of cells in the n-th channel is

$$h_n = f(x_1) \cdot \int_{n-1}^{n} p_{cv}(x_1, y)\,dy \quad (4)$$

where $f(x_1)$ is the number of cells with DNA content x_1 (Fig. 2).

Histogram of G1 Peak in the Presence of S Cells : Under the assumption of a rectangular DNA distribution in the S and G2M phases (Fig. 1), the shape of the histogram depends on the fraction of cells in G1 (F_{G1}), the G1 DNA content (x_1) and on the measurement CV. Cell density in the S phase is uniquely determined from F_{G1}. The histogram is simulated by summing up fluorescence of cells with discrete DNA contents, i.e. x_1, x_2, ..., x_{100}:

$$h_n = \sum_{i=1}^{100} f(x_i) \cdot \int_{n-1}^{n} p_{cv}(x_i, y)\,dy \quad (5)$$

assuming that $f(x_i)$ represents the number of cells which belong to the i-th DNA column and that

$$F_{G1} = f(x_1)/\sum_{i=1}^{100} f(x_i) \quad (6)$$

Fjg. 2-a Fjg. 2-b Fjg. 2-c

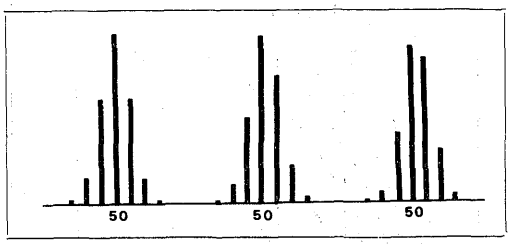

Fig. 2 Discretization effects: A fluorescence histogram is a set of columns that represent the number of cells with fluorescence intensities being in each discrete channel width. In making a histogram by grouping continuous data into columns, a distribution pattern is somehow distorted. Figs. 2-a, 2-b, and 2-c are the histograms which were derived from Gaussian distributions having a same CV (=2%) but slightly different modes, i.e. x_1=49.5, 49.7, and 49.9, respectively. All of them have a peak at channel 50. The bluntness indices are b_{50}=62.5%, 63.9%, and 67.3%, respectively. It increases to 69.9% when the histogram is maximally skewed (m=49.99) indicating that the bluntness index is variable depending on where the mode is located within the width of a channel. This discretization effect will be more pronounced when the peak is skewed by the presence of early S cells.

In order to see how the shape of the G1 peak is affected by factors other than the CV, several histograms were calculated by changing the G1 peak near x_1 by 1/10 of a channel width around 30, 40, 50, 60, 70, 80, 100, and, by varying F_{G1} from 0.3 to 0.9. When the CV was large and when there were many early S cells, the modal channel often shifted to the right. When this occurred, x_1 was changed (decreased) to keep the mode of the histogram at the desired position. The bluntness index was calculated for each simulated histogram and correlated with the measurement CV.

Construction of CV Table : The data were grouped according to the modal channel number, n, of the G1 fluorescence peak. Therefore, each group contains data for histograms with differing G1 means, x_1, and G1 phase fractions, F_{G1}. Bluntness indices, b_n's, in each group were plotted against CV's. The b_n depends to some extent on the variables x_1 and F_{G1} which are not known, even if n and CV are specified, so that only the upper and lower bounds of b_n's can be determined. In reverse, by specifying a b_n one can read a possible range of CV (Fig. 3).

Estimation of CV with the FWHM Method: For comparison with the look-up table method, CV was estimated using the FWHM method[6] which calculates CV by the following formula:

$$CV=\frac{FWHM}{2.35\times n}\times 100(\%) \qquad (7)$$

where FWHM is the width of the G1 peak at half maximal ordinate value expressed in unit of channel width and n is the modal channel number of the G1 peak.

Results

Bluntness Index and CV: When n and F_{G1} were held constant, the bluntness index b_n increased almost linearly with CV up to b_n=60%. Above this level, however, the rate of increase decreased and the curve leveled off towards 95-97%. As was expected, the bluntness index for a given n and CV involved a considerable range of uncertainty and the uncertainty was more marked with smaller CV's. That it is due partly to a discretization effect was suggested by the finding that the uncertainty was reduced by taking into account a symmetry of the G1 fluorescence peak. However, when the CV is large, the b_n depends to large extent on the F_{G1} which cannot be estimated accurately without computer analysis.

Specification of the symmetry does not appreciably decrease the uncertainty but increases complexity of the CV table (Table 1). Therefore, the symmetry was disregarded. Fig. 3 shows the graphs that correlate b_n with CV. Fig. 3-a gives a lower bound on the CV within a realistic range of variation in F_{G1} (=30%) and Fig. 3-b gives the upper bound on the CV that corresponds to F_{G1}=90%. The various curves show these bounds for histograms with G1 peak modes

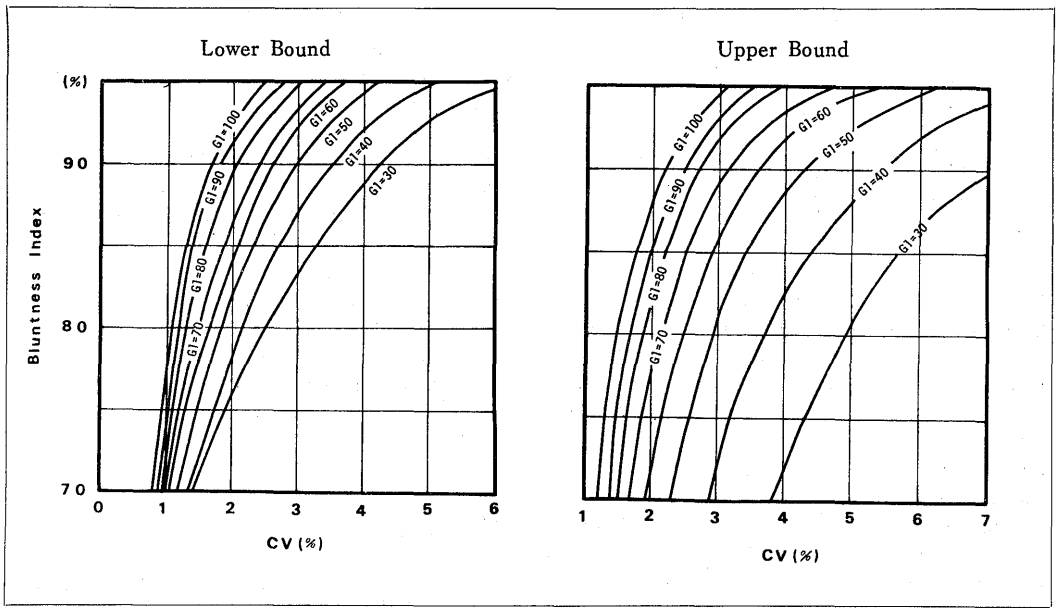

Fig. 3 CV vs. bluntness index: The calculated bluntness indices were plotted against CV's so that the latter can be read.
Fig. 3-a (left) is for F_{G1} =30% and gives a lower bound of realistic range of CV.
Fig. 3-b (right) is for F_{G1} =90% and gives an upper bound of CV.

at 100, 90, 80, 70, 60, 50, 40, or 30th.

Use of Table 1 : Table 1 was used to estimate the CV of L5178Y cells[3]. The G1 peak was at channel number 36. The cell numbers in three channels around the peak were 1134, 1452, and 1305. Hence, its bluntness index was

$$b_{36}=\frac{1}{2}\left(\frac{1134+1305}{1452}\right)=0.84$$

If we may assume that the G1 peak was at 40th or 30th channel, we get CV=2.6-4.2% or CV=3.1-5.5%, by entering the columns for n=40 and 30 (in Table 1) with b=84 %. Therefore, if the peak channel is not accurately located within these columns, uncertainty of CV ranges from 2.6% to 5.5 %. However, by interpolation for n=36, we can restrict the uncertainty to CV=2.8-4.7 %. The analysis with DHA8005[4] yielded CV=4.4% whereas the FWHM method[6] gave CV=5.7%. Five other examples were shown in Table 2, which indicate that the results with DHA8005 fall within the CV range estimated by this three-channel method.

Statistical Variation of the Bluntness Index: Since the b varies among experimentally measured histograms in a way that can hardly be simulated numerically, an experiment was performed to determine how b changes and how this variation affects the estimates of CV. For these experiments, L5178Y cells were used which were fixed with 70% ethanol and stained with propidium iodide solution (40 μg/ml). As shown in Table 3, DNA histograms of 5000 cells was measured 7 times and CV's were estimated for each histograms. Then, these histograms were added (Table 4) and CV's were calcu-

Table 1 Look-up Table for CV(%)

Bluntness index	Channel Number of G1 Peak							
	30	40	50	60	70	80	90	100
95%	6.4-	5.1-	4.2-	3.7-	3.4-	3.0-3.9	2.8-3.5	2.5-3.1
94%	5.6-	4.7-7.0	3.9-5.7	3.4-4.7	3.1-4.2	2.7-3.6	2.5-3.3	2.2-2.9
93%	5.1-	4.4-6.4	3.6-5.2	3.1-4.4	2.8-3.9	2.5-3.3	2.3-3.0	2.1-2.7
92%	4.8-	4.1-6.0	3.4-4.8	2.9-4.1	2.6-3.6	2.3-3.0	2.1-2.8	1.9-2.5
91%	4.5-	3.9-5.7	3.2-4.6	2.8-3.8	2.6-3.4	2.2-2.9	2.1-2.6	1.8-2.3
90%	4.2-7.1	3.6-5.5	3.0-4.3	2.6-3.6	2.4-3.2	2.1-2.7	1.9-2.5	1.7-2.2
89%	4.0-6.8	3.3-5.2	2.8-4.1	2.5-3.7	2.3-3.1	2.0-2.6	1.8-2.4	1.6-2.1
88%	3.8-6.5	3.2-5.0	2.7-4.0	2.4-3.3	2.2-2.9	1.9-2.5	1.8-2.3	1.6-2.0
87%	3.6-6.3	3.0-4.8	2.6-3.8	2.3-3.1	2.1-2.8	1.8-2.4	1.7-2.2	1.5-2.0
86%	3.5-6.0	2.9-4.6	2.4-3.6	2.2-3.0	2.0-2.6	1.7-2.3	1.6-2.1	1.5-1.9
85%	3.3-5.7	2.7-4.4	2.3-3.4	2.1-2.9	1.9-2.5	1.6-2.2	1.5-2.0	1.4-1.8
84%	3.1-5.5	2.6-4.2	2.2-3.3	2.0-2.8	1.8-2.4	1.5-2.1	1.4-1.9	1.3-1.7
83%	3.0-5.3	2.5-4.0	2.1-3.2	1.9-2.7	1.7-2.4	1.5-2.1	1.4-1.9	1.3-1.7
82%	2.8-5.2	2.4-3.9	2.1-3.1	1.9-2.6	1.7-2.3	1.4-2.0	1.3-1.8	1.2-1.6
81%	2.7-5.0	2.3-3.8	2.0-3.0	1.8-2.6	1.6-2.3	1.4-2.0	1.3-1.8	1.2-1.6
80%	2.5-4.8	2.2-3.7	1.9-2.9	1.7-2.5	1.5-2.2	1.3-1.9	1.2-1.7	1.1-1.5
79%	2.4-4.7	2.1-3.6	1.8-2.8	1.6-2.4	1.5-2.1	1.3-1.8	1.2-1.7	1.1-1.5
78%	2.3-4.6	2.0-3.5	1.8-2.8	1.5-2.4	1.4-2.1	1.2-1.8	1.1-1.6	1.0-1.4
77%	2.1-4.5	2.0-3.4	1.7-2.7	1.5-2.3	1.4-2.0	1.2-1.7	1.1-1.6	1.0-1.4
76%	2.0-4.4	1.9-3.3	1.7-2.7	1.4-2.3	1.3-2.0	1.1-1.7	1.0-1.5	0.9-1.3
75%	1.9-4.3	1.8-3.2	1.5-2.6	1.3-2.2	1.2-1.9	1.1-1.6	1.0-1.5	0.9-1.3
74%	1.8-4.2	1.7-3.1	1.5-2.5	1.3-2.1	1.2-1.9	1.1-1.6	1.0-1.5	0.9-1.3
73%	1.7-4.1	1.6-3.1	1.4-2.5	1.2-2.1	1.1-1.8	1.0-1.5	1.0-1.4	0.9-1.3
72%	1.6-4.0	1.6-3.0	1.4-2.4	1.2-2.0	1.1-1.8	1.0-1.5	0.9-1.4	0.8-1.2
71%	1.5-3.9	1.5-3.0	1.3-2.4	1.1-2.0	1.0-1.8	0.9-1.5	0.9-1.4	0.8-1.2
70%	1.4-3.8	1.4-2.9	1.2-2.3	1.1-1.9	1.0-1.7	0.9-1.5	0.9-1.4	0.8-1.2

* The bluntness index is defined as $0.5\ (h_{n-1}+h_{n+1})/h_n$ where n is the channel number of G1 peak. The data were arranged from top to bottom in a decreasing order of bluntness index, b, (see the leftmost column) because the lower the channel heights on both sides of the peak the lower row is to be looked at.

lated for histograms in which the total number of cells increased at intervals of 5000 cells. The result indicated that the estimate by this method is not subject to a marked statistical fluctuation provided some 1000 cells per channel are scored at the peak.

Comparison with the FWHM Method: Table 2 shows some examples which indicate that the CV values estimated by the FWHM method were greater than those obtained from the CV table. The former were even greater than the upper limit of probable range of the latter: in one example, a CV by the FWHM method exceeded the upper limit by a factor of 1.7. In contrast, the results of the computer analysis always fell within the ranges of uncertainty.

Discussion

A sharp G1 fluorescence peak in a DNA content distribution measured flow-cytometrically indicates high resolution. However, the reverse is not necessarily true since broadening may also be caused by the presence of a large population of early S cells. In addition, a visual inspection is often misleading since the peak characteristics are affected

Table 2 The CV's Estimated by Different Methods

Channel No.	Cell Number in Each Channel				
	CHO				
34	741				
35	1148				
36	1415				KHT
37	1619	1025			640
38	1796	1462			1110
39	2046	1688			1407
40	2193*	2006			1460*
41	1743	2186	HeLa S3		1309
42	1020	2576	650		1039
43		2604*	1080	775	809
44		2073	1607	1255	652
45		1241	1881	1622	
46			2108*	1908	
47			1881	2015*	
48			1454	1802	
49			964	1275	
50				814	
FWHM	7.4%	7.2%	5.5%	5.5%	6.7%
Computer CV Table	4.3%	3.8%	3.9%	5.0%	5.3%
bluntness	2.9–4.6% b_{40}=86%	3.2–4.9% b_{43}=89%	3.0–4.5% b_{46}=89%	3.6–5.2% b_{47}=92%	4.4–6.4% b_{40}=93%

Two sets of data are shown for each of the CHO and HeLa S3 and one data set for KHT cells. These are the examples of the G1 fluorescence peak which are sharp at the top and broad at the base. The asterisk indicates a peak channel and the underline locates the position of half maximal ordinate value of the peak.

not only by the CV but also by many other factors and quantification of the peak sharpness may facilitate estimation of the CV in these instances.

Visual inspection of DNA content histograms reveals that the sharpness of the peak can be estimated qualitatively by the number of data points on the left half of the G1 peak: the sharper the peak the fewer data points can exist in this region. However, quantification of the CV from the left half of this histogram is difficult because the position at which the histogram leaves the baseline can not be identified exactly. This is especially true when the CV is large or when there is a large number of cell debris as we often observe with clinical materials. The well-known FWHM method[6], a peak-width method, reduces the difficulty by elevating the estimator level. However, the FWHM method requires use of the data on the opposite side of the peak which is broadened by the presence of early S cells. Thus, it inevitably overestimates CV as

Table 3 Statistical Variation of Estimated CV's In Repeated Measurements of the Same Sample

Total No. of cells measured	No. of cells in the channels around the G1 peak(*)				Bluntness index	CV (from Table 1)
	Channel No.					
	48	49	50	51		
5000	368	452	473*	405	b_{50}=0.91	3.2-4.6%
5000	383	458*	439	422	b_{49}=0.90	3.1-4.4%
5000	407	449	499*	411	b_{50}=0.87	2.6-3.8%
5000	412	447*	439	387	b_{49}=0.95	4.3% or more
5000	385	422*	342	316	b_{49}=0.86	2.5-3.7%
5000	390	408*	377	320	b_{49}=0.94	4.0-5.8%
5000	398	444*	404	322	b_{49}=0.90	3.1-4.4%

Using L5178Y cells from the same lot, CV was measured for 7 different histograms. In each measurement, a total of 5000 cells were scored and CV was read from the CV Table. This table shows the number of cells in several channels around the G1 fluorescence peak (indicated by*).

Table 4 Fluctuation of CV in Cumulative Data

Total No. of cells measured	No. of cells in the channels around the G1 peak(*)				Bluntness index	CV (from Table 1)
	Channel No.					
	48	49	50	51		
5000	368	452	473*	405	b_{50}=0.91	3.2-4.6%
10000	751	910	912*	827	b_{50}=0.95	4.2 or over
15000	1158	1360	1402*	1238	b_{50}=0.93	3.6-5.2%
20000	1570	1807	1841*	1625	b_{50}=0.93	3.6-5.2%
25000	1955	2229*	2183	1941	b_{49}=0.93	3.6-5.2%
30000	2345	2637*	2560	2261	b_{49}=0.93	3.6-5.2%
35000	2743	3081*	2964	2583	b_{49}=0.93	3.6-5.2%

The data were taken from Table 3 and cumulative histograms were made at the interval of 5000 cells to see how the peak characteristic is stabilized with the progression of data accumulation.

described in this report.

The present method, called "channel height method", adopts the highest possible estimator level, i.e. the cell numbers in three channels of the G1 peak which are statistically most stable and least contaminated with both early S cells and cell debris. Nevertheless, it is important to note that the estimate of CV by this method is not completely free from the influence of cell debris and S-phase cells. The infleuence of the latter cannot be compensated for without

analyzing the entire histogram. Even with the use of the computer, it is not possible without making some artificial assumptions [1),4)] to separate cells in the lower portion of the so-called G1 fluorescence peak into the G1 and early S compartments. As far as the effects of the early S cells and of the CV are unresolvable, it is better to estimate the CV with the range of uncertainty.

The uncertainty range shown in Table 1 was calculated based on a realistic assumption that 30-90% of cells are in G1 and other cells are evenly distributed in the rest of the cycle. The upper bound of the CV corresponds to the case in which 90% of cells are in G1. This is an extreme case in which the bluntness can be attributed almost exclusively to the error of measurement. Generally, it applies to the data from growth-arrested cell populations like peripheral lymphocytes. The lower bound, on the other hand, applies to the cell populations in which 70% of cells are (evenly distributed) on the right side of G1. However, a true CV may be even less because of the possibility that more of the S cells lie very close to G1 and they increase the bluntness. This precaution is particularly important when the G1 peak is very sharp on the left side. In order to stand on a safer side, it is advisable to read the upper limit and take it as such. However, when the histogram is high in the S region as is often the case with a tumor cell population, the lower limit of the table gives a more accurate estimate.

The advantage of this method is its ease of application with no need of calculation. The only parameters necessary to enter the CV table are the channel number of the G1 peak and the bluntness index. Although the latter is not given numerically on the machine, its approximate value can be estimated by eye and it suffices for a practical purpose. In using the CV table, it is usually the case that the G1 peak mode is not at one of the channel numbers listed in the table. However, in order to have a rough estimate of the CV during experiment, interpolation in between the channels is not very important. It is partly because inherent uncertainty of the estimate, although minimized in this method, is still considerable. Furthermore, it is also possible to bring the peak to one of these channels by controlling the gain of the amplifier. For many reasons, the three-channel method described here seems to be more useful than the FWHM method when CV is small. However, for the data with a flat-topped G1 peak, the FWHM method remains more practical.

The authors wish to express thanks to Dr. K. Isiglo of Department of Pathology, Fukui Medical School and Dr. F. Dolbeare of Biomedical Sciences Division, Lawrence Livermore National Laboratory (LLNL) for their generous supply of culture cells, to Mr. M. Okabe, Pharmaceuticals Research Laboratory, Kyowa Hakko Kogyo Co. Ltd. for kindly permitting us to use the data of HeLa S3 cells, and also to Dr. Joe W. Gray (LLNL) for his valuable suggestions.

References

1) Dean, P.N. and Jett, J.H.: Mathematical analysis of distributions derived from flow microfluorometry. *J. Cell Biol.*, 60 : 523-527, 1974.
2) Fried, J.: Method for the quantitative evaluation of data from flow microfluorometry. *Comput. Biomed. Res.*, 9 : 263-276, 1976.
3) Takahashi, M.: Multichannel method for the analysis of DNA content distribution using flow cytometric data. *Comput. Biomed. Res.*, 14: 506-517, 1981.
4) Takahashi, M.: DNA histogram analysis of exponentially growing and perturbed cell populations. *In*: Biomathematics and Cell Kinetics, Rotenberg M (ed.) Elsevier/North-Holland ,Amsterdam, 1981, p. 211-221.
5) Zietz, S. and Nicolini, C.: Flow microfluorometry and cell kinetics: a review. *In*: Biomathematics and Cell Kinetics, Valleron A-J, Macdonald PDM (eds.) Elsevier/North-Holland, Amsterdam, 1978, p. 357-394.
6) Vindelφv, L.L.: Flow microfluorometric analysis of nuclear DNA in cells from solid tumors and cell suspensions. *Virchow Arch. B*, 24 : 227-242, 1977.