Chapter 50
Collection, Display, and Analysis of Flow Cytometry Data

DAVID R. PARKS AND MARTIN BIGOS

Objectives of data collection and analysis

The biological objectives of FACS experiments are now extremely varied, but there are common features that we emphasize here. When collecting flow cytometry data, the objectives are to ensure that annotations and documentation are adequate to make the data interpretable in the long term. It is also critical to ensure that the measurements are correct and that enough cells have been recorded to support the analyses to be done.

In the data analysis process, we are usually interested in identification and/or enumeration of cell populations on interest, characterization of these populations in terms of one or a few numerical values, and production of clear visual representations of data results too complex to be conveyed in a few numerical values. Another objective, which we do not deal with in this chapter, is extraction of parameters to fit mathematical models for biological processes.

Before collecting data

Three critically important steps affecting the success of flow cytometry data analysis should be done before any data are collected. These are inclusion of appropriate and informative controls, documentation of the experiment and cell samples, and standardization of the flow cytometer.

Control samples

Biological experiment design and sample preparation are discussed elsewhere in this volume (particularly in Chapter 49 [1]). Here we simply emphasize the importance of carefully selected control samples. There are two general types of controls to be considered: biological controls to verify reagent staining patterns or specificity and reagent controls that are needed to set fluorescence compensation values. Including processed but totally unstained cells is an often overlooked but sometimes critical biological control that can be valuable in understanding background staining of fluorescent reagents. Single-color stains for adjusting or confirming fluorescence compensation should use the brightest stain in each color to achieve the best accuracy.

Documentation and annotation

The designers of most flow data systems have viewed documentation as a poor stepchild to the glamor of graphical display. Often, all they provide is a pedestrian editor for keyword values to be stored with the data file. When later analyzed, these keywords, such as date run, reagents used, and so on, are usually displayed with the data. Unfortunately, these editors tend to be oriented toward single samples and not taking sample grouping and experiment structure into account. Once the annotations are entered, there may be no facility for browsing or searching the keywords to group or retrieve relevant files when doing later analysis. Moreover, most flow facilities rely on the underlying structure of the computer file system to manage the data itself. The organization of data files into experiments and projects is not ensured in this arrangement. These management issues are problems related not only to flow cytometry. Other disciplines in which large amounts of instrument data are generated are experiencing similar problems [2].

As an experimenter and experiment designer, it is worth considering these issues before any work is done. If a very limited number of flow cytometry experiments with a relatively small number of samples are performed, then the documentation provided by a good laboratory notebook will probably suffice. However, if the use of flow cytometry will extend through many experiments and samples, setting up a secondary database on a personal computer to track this information, and keeping that database current, is a worthwhile endeavor to avoid many future headaches.

Instrument standardization

Instrument standardization should guarantee that if "identical" samples were run at different times, the measurements would be the same. The first step in carrying this out is to establish a set of standard operating conditions and measurement expectations. The optical conditions of the instrument should be optimized using high uniformity test particles, and signal levels should be adjusted, running typical cell samples and using amplifier gains and PMT voltages, to set good levels for measurements on cells. Measurement values for the test particles can then be recorded under these conditions to establish signal level and uniformity (CV) targets to be reproduced in future instrument standardizations. (We have found that multi-dye polystyrene microspheres, such as those obtainable from Spherotech, Inc., Libertyville, IL, USA, work well for this purpose, and with the right dyes, a single lot can be used with UV, 488 nm, and dye laser excitations. Types of standard particles are discussed in Chapter 47 [3].)

At later times, the instrument conditions, such as laser powers, are reproduced as well as possible, and samples of the standard microspheres are used for alignment optimization and target signal level matching (by PMT voltage adjustment, etc.). Comparing the CVs obtained in this way with the previously measured values and comparing the PMT voltages required to match the target signal levels with those of the initial run gives us a diagnosis of how well the running conditions have been matched. We found it useful to run the standard microspheres and standardize the signals before each FACS experiment and to recheck the standardization when any problems occur during an experiment. It is also important to record the standardization values (PMT voltages, CVs, etc.) for each run to track trends and help in the investigation of unexpected results.

If standard conditions are re-established accurately enough, fluorescence compensation settings for particular dyes should

have the same value from day to day [4]. However, if possible, fluorescence compensations should be checked with a single stained sample for each dye.

The signal scales in the standard conditions can be calibrated in relation to numbers of dye molecules or numbers of dye coupled antibody molecules using microspheres that carry a known amount of dye or that bind a known number of antibody molecules (discussed in Chapter 47 and available, for example from Flow Cytometry Standards Corp., San Juan, Puerto Rico). Another form of signal calibration, providing a measure of the relation between photoelectrons generated in the PMT and output signal levels, is discussed below under "Absolute signal levels—photoelectron estimation".

Data collection

Format, resolution, and range of data

An emerging standard for data storage and interchange is the FCS format [5]. Files written to this standard should be readable by any program that supports FCS (almost all commercial analysis programs make this claim); however, at this time, full intercompatibility has not been realized. As an alternative, many programs will allow the export of binary list data as a text list, which can then be imported into almost any third-party spreadsheet or analysis program for computations not available in the program of origin.

The simplest and generally most useful way to store FACS data are as data lists. In this form, the measurement values for each of the cell-related data events are recorded in order so that the data collection may be processed repeatedly in a variety of different ways to optimize data gating and to obtain different graphical and numerical representations of the sample. Well-documented list data can often be reanalyzed long after the original experiment to test new ideas or interpretations. Most software also allows data storage in one- or two-dimensional histograms. When computer storage was expensive and data were usually limited to two or three measurements per cell, this was a necessary and adequate alternative, but currently there is little reason to record anything short of list data.

For most measurements in flow cytometry, analog voltage levels representing the measurements of interest are processed with analog-to-digital converters (ADCs) to obtain digital values that can be conveniently processed and stored. The maximum value a of a binary integer is determined by the number of bits allocated to it. For the digitization process, the maximum size integer (in bits) is referred to as the resolution of the data. The number of signal levels is 2^n where n is the bit resolution, so that 8-bit data have levels ranging from 0 to 255, whereas 10-bit data range from 0 to 1023. The ADCs used in flow cytometry currently produce data with a resolution of 8 to 16 bits.

It is common to use a 12- or 16-bit ADC but to record only 8- or 10-bit data from it. This is because the types of ADCs commonly used, although accurate in terms of signal level evaluation, have variation in the "widths" of the successive digitized levels (i.e., the range of analog signal levels recorded as a particular digital value is not constant). This results in ragged-looking histograms when the raw data are plotted at full resolution. Ignoring several of the finest bit levels smoothes out such displays.

Calculations and displays of data are sometimes not done at the full data storage resolution. For example, many two-dimension displays (see below) are computed at 6-bit (64×64 levels) or 7-bit (128×128 levels) resolution for computational efficiency even though the underlying data are stored at higher resolution. However, for statistical parameters calculated from the data, it is almost always appropriate to use the full resolution. When stored data are likely to be used in producing computed dimensions (see "Data transformations" below), it may be useful to record at higher bit resolution than is really needed for direct use of the primary data.

During data collection

We found that the two most important tasks during data collection are monitoring of the data to verify its integrity and ensuring that the documentation and annotation of the sample/experiment information is complete and correct. We do not recommend that extensive analysis of the data be carried out during data collection for the following reasons. First, many flow instruments are time-shared, and doing a full data analysis during collection prevents others from using the instrument. Second, with most instruments and experiments, there are enough aspects to monitor that doing serious analysis may shift the focus of the experimenter enough to miss key events. Of course, when physically sorting out cells, analysis is necessary to delineate the desired population(s). However, in most sorting experiments, there are a limited number of samples, and preliminary work requiring detailed analysis, such as reagent titration and evaluation of controls (see Chapter 49 [1]) should already have been done. One of the advantages of routinely storing full-list data is that the experimenter is free to focus on the data integrity during data collection, knowing that various analysis options are available later.

Monitoring of data during collection involves visualization of the data on displays that are updated in real time (as the sample is run). The most common and useful display is called a dot plot, in which two (user selected) parameters form the x and y axes of a two-dimensional display (as in Fig. 50.2D). Dot plots give useful information about the distribution pattern of the data and are quick to compute. Dot plots are not ideal for final data display but are usually quite adequate for monitoring data patterns to ensure that the right sample is being run, that the results are qualitatively as expected for the sample and reagents, and in, longer runs, that the pattern does not change over time. Forward scatter is a very useful parameter to detect instrument changes, partly because it is used as a linear signal (as opposed to log signal; see below). A small change in the instrument conditions will generally result in a more recognizable difference in the forward scatter distribution than in the distributions of log fluorescences.

Gating of data during collection—how many cells to collect

Although we generally recommend collecting all events from a sample in list format for later analysis, from the monitoring displays one can usually see events that will not be relevant to any analysis. These usually involve cellular debris, dead cells, or perhaps platelets if one is running diluted blood. When these are a small fraction of the total events, they may as well be recorded and excluded in later analysis. Then, if there is any question about the results, the whole set of measurements is available for re-examination. Sometimes, however, well-defined undesired events occur at a higher frequency than all possibly desirable

Collection, Display, and Analysis of Flow Cytometry Data

Table 50.1. Minimum cell sample sizes (in thousands) needed to reveal significant differences in population frequencies

Measured frequency	Frequency difference to be detected as significant					
	3%	1%	0.3%	0.1%	0.03%	0.01%
90%	1	7	78	700	7700	
70%	2	17	180	1600		
50%	2	20	216	1950		
30%	2	17	180	1600		
10%	1	7	78	700	7700	
3%		2.5	26	230	2500	
1%		1	9	78	860	7600
0.3%			3	25	260	2300
0.1%				9	90	790

Note that the required sample sizes for a frequency F greater than 50% are the same as for a frequency 100%-F because this change represents only a reversed designation of which events are in and out of the population of interest.

events. Most instruments (or the data collection systems associated with them) allow one to exclude certain data regions from the data collection file. This exclusion before storage is also useful for analyzing very-low-frequency subpopulations in samples where most of the events are uninteresting.

It is useful to estimate how many events are needed to provide a desired level of statistical accuracy in population estimates. Although in many cases the statistical variations will not be the dominate source on uncertainty in evaluating the experiment, this at least makes it possible to ensure that enough data are taken to support the intended analysis. Table 50.1 gives cell counts needed for measuring population frequencies with varying degrees of certainty and illustrates the usefulness of gating out debris. For example, measuring a 10% population with 1% certainty requires only 7000 events, but, if 90% of the sample is debris, then 70,000 ungated events are needed.

Data transformations

Many types of transformations have been used for FACS data, but the common ones we mention here are linear/logarithmic, fluorescence compensation, and ratios.

The choice of logarithmic or linear data

Logarithmic signal amplification is widely used for immunofluorescence measurements as a way to evaluate and visualize signals over a wide range, typically about four decades. Digitization of the logarithmic amplifier output results in an approximately logarithmic data distribution. Depending on the cell populations involved and the purpose of the analysis, logarithmic or linear presentation of the data may be more appropriate. Linear is preferred for data with limited dynamic range or narrow peaks (both frequently encountered, for example, in cellular DNA measurements). Logarithmic presentation retains the relative shape of distributions regardless of signal level, making it easy to compare distributions in which the shape is more important than their absolute signal levels. Thus, in a logarithmic display, a population whose signals vary over a sixfold range will be broader than one with only a threefold range, regardless of which is stained more brightly. In a linear display, the width of a population in the display is proportional to the brightness of its staining (as well as to its intrinsic variation). In a logarithmic presentation, each successive digital level corresponds to a constant fractional increase in original

signal level (about 4% per level in 8-bit data and 1% in 10-bit data). In linear presentation, of course, each successive level represents a constant added signal increment over the previous one.

In multicolor immunofluorescence, it is often easier to identify, delineate, and visually estimate frequencies of subpopulations of cells in one- or two-dimension displays using log rather than linear presentation (Fig. 50.1). Another problem with linear presentation is that when a cell population is centered near midscale, higher level signals within the population will often go offscale. (Offscale events are shown as a small peak at the right end of the upper right panel in Fig. 50.1.) The offscale events are counted, but their true values are unknown, and a mean or coefficient of variation calculated for the population will be biased. In logarithmic presentation, it is usually easier to keep all the relevant signals on scale. Log presentation may also give more easily interpreted orthogonal light scatter distributions even though the dynamic range of the signals does not require four decades. In this case, we found that a two-decade log display may be optimal.

Fluorescence compensation

Analog fluorescence compensation is discussed in Chapter 47. Off-line linear transformation equivalent to fluorescence compensation can be useful, usually in situations where stored fluorescence data are uncompensated or undercompensated. In multiple laser systems, a dye may be excited by more than one laser, leading to unwanted signal contributions that cannot be corrected by ordinary analog fluorescence compensation. For example, there are problems in using PE-CY5 with 488 nm excitation in the same system with allophycocyanin excited at 600 nm. The CY5 in PE-CY5 is excited directly by a dye laser at 600 nm, the allophycocyanin excites a little bit at 488 nm, and both dyes emit in the 660- to 670-nm range. The two separately timed signals from the two laser transits will each be the sum of contributions from both dyes. Analysis of data from singly stained cells of each type provides the subtraction coefficients that can be used to correct the mixed dye results.

Ratios

Ratios are synthetic measurements derived from two primary signals. Ratios are particularly useful in analyzing measurements with metachromatic dyes. These dyes generally measure physiological properties of cells by having emission spectra that change with varying physiological conditions. Thus, the fluorescence emission is measured in two different wavelength regions, and the desired parameter for data analysis is the ratio of these signals. A common application is the use of Indo-1 to measure intracellular free $[Ca^{2+}]$ [6].

The ratio can be computed either in the instrument hardware or in software. The advantage of the hardware approach is that the full functionality of the instrument is available for the ratio signal, that is, real time monitoring and sorting, as well as data collection. Depending on the system used, software ratioing may not offer all of these capabilities. Analog electronic ratios, however, generally cannot provide the ratio as an accurate on-scale output over the full range of primary signals. Software ratios also have problems maintaining the synthesized parameter on-scale. We have proposed and demonstrated a solution by using the difference of the logs of the primary signals as a measure of the log of the ratio [7]. Our solution was done in hardware, but it could also be done by
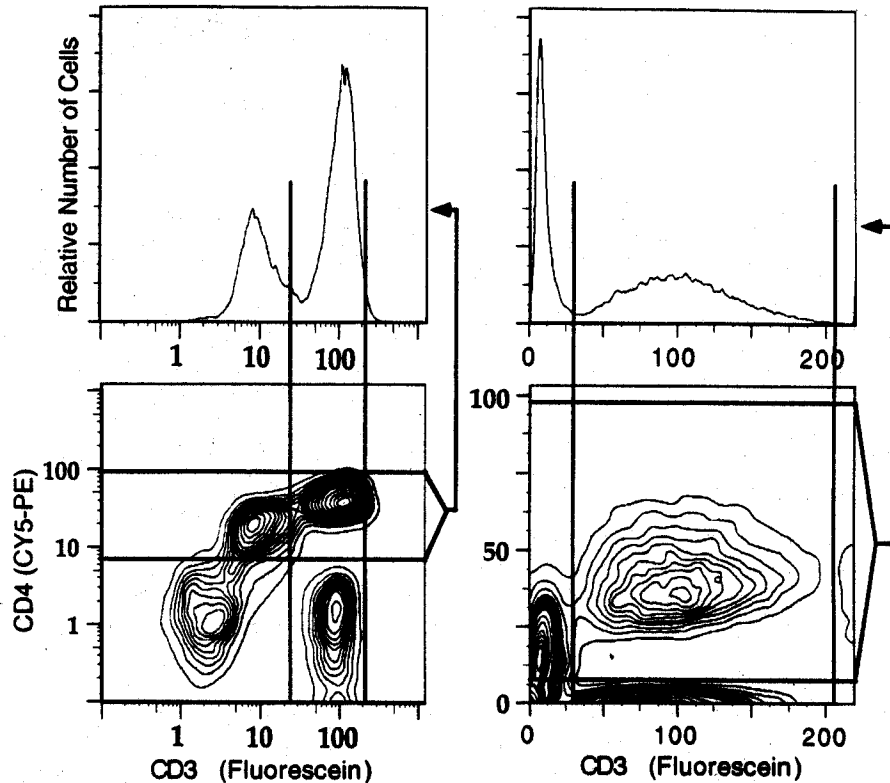
Fig. 50.1. *Comparison of logarithmic and linear data display.* A single sample of human blood mononuclear cells was stained for CD3 and CD4, fixed with paraformaldehyde, and analyzed using logarithmic (left) and linear amplification (right). Light scatter gating was used to exclude all events other than lymphocytes and monocytes. The cell population centered at about 9 units CD3 and 18 units CD4 consists of monocytes; the other populations are lymphocytes. The background fluorescence levels (CD3-,CD4- appearing at about 2 units and 1 unit, respectively, in the lower left log-log panel) are elevated due to the fixation process. The single dimension displays are gated CD3 distributions for cells in the CD4 range marked on the two-dimension plots. The marker lines in both CD3 and CD4 are drawn at equivalent signal levels in the logarithmic and linear displays.

analysis software. We found that this approach overcame the limitations described above.

Data display and analysis

Gating and multidimensional data analysis

The term "gating" is commonly used for the process of segmenting cell data events into subpopulations for subsequent display or numerical evaluation. This frequently involves exclusion of undesired debris, dead cells, and/or aggregates followed by delineation of populations of interest among the remaining events. Gating regions are typically defined as intervals on one-dimension displays and as rectangles, polygons, or other regions on two-dimension displays. Compared with single-dimension displays, two-dimension display methods often provide clearer indications of optimal demarcation between populations even when only a single fluorescent label is measured (such as using a forward light scatter versus fluorescence contour plot for selecting high and low fluorescence populations rather than just a fluorescence histogram). Analysis strategies for multicolor immunological data are presented in Chapter 49.

Density estimation

In flow cytometry, we use measurements on a limited number of cells to estimate the properties of the full population. Statistical fluctuation in the number of cells allocated to each digital bin

(channel) limits the ability of raw data histograms to represent the true population distributions. Appropriate density estimation functions can give better estimates of the true distributions and make it easier to see real features in the data that might otherwise be obscured by the statistical noise. In addition, the smoothing provided by a density estimation function is often useful when several one-dimension curves are overlaid, and it' is practically necessary for producing easily interpreted contour plots from moderate-sized data sets (Fig. 50.2, B versus F). For light scatter, immunofluorescence and most other measurements (except high-resolution DNA content and measurements on very uniform test particles), we found that a count-dependent kernel method works very well [8]. This method adjusts smoothing in relation to the number of data events in each bin so that sharp features are retained where there are many events per bin but purely statistical fluctuations are diminished. With any smoothing method, it is important ensure that real features in measurement are retained. In general, sample statistics should be computed on the raw data even if gating displays have been computed at lower resolution.

Two-dimension displays

Two-dimension displays of flow cytometry data have several functions, and each display method is better adapted to some functions than others. In this section, we describe and, in some cases, illustrate these methods. The functions include quickly
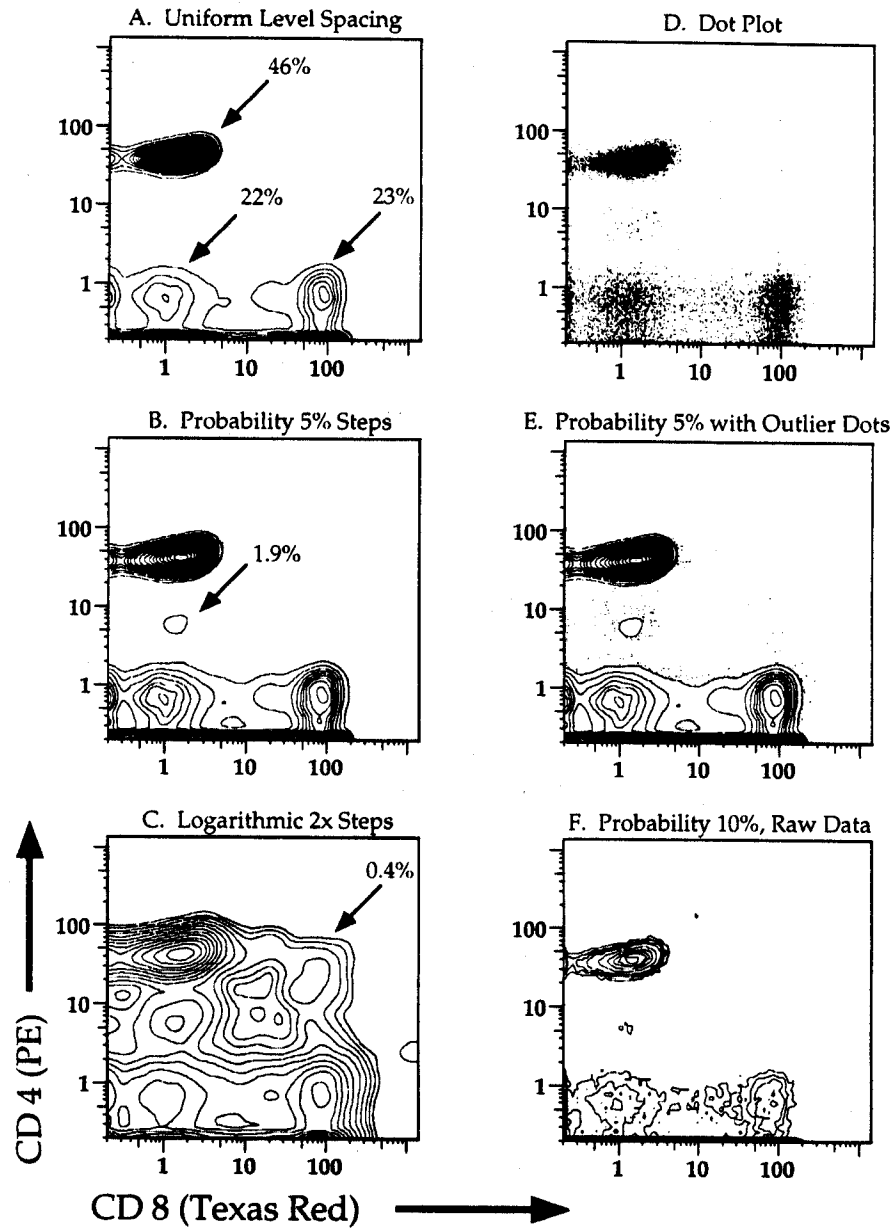
Fig. 50.2. *Comparison of different contour and dotplot methods.* A single data set is represented in six types of plots. The sample consists of human peripheral blood lymphocytes (light scatter gated) stained for CD8 (Leu-2a-Texas Red) and CD4 (Leu-3-Phycoerythrin). Five identifiable subpopulations with frequencies of 46%, 23%, 22%, 1.9%, and 0.4% are marked on the plots. Except for the raw data plot in F, the contour plots are computed on density estimated surfaces at 128 × 128 resolution. To obtain even marginally acceptable contours, it was necessary to compute the raw data plot (F) at 64 × 64 resolution. For probability contours with outlier dots (E), the 5% of data events falling outside the lowest contour are plotted as individual points.

computed displays for interactive analysis, displays used for "gating" data and defining subpopulations, and displays used for presentation/publication of results. Further discussion and illustrations can be found in Chapter 49 by Kantor and Roederer [1].

Dot displays (see Fig. 50.2D), in which each cell is represented by a dot at the appropriate coordinates for its measurements in the displayed dimensions, are quickly produced and easy to interpret, but they suffer from limited dynamic range. The number of dots in any region of the plot represents the frequency of cells with the corresponding characteristics. Low-frequency popula-

tions will not be visible if too few events are plotted; high-frequency populations lose visible structure if too many events are plotted, and plots of the same cell sample using different numbers of events will look different. However, because they can be rapidly computed at the full data resolution, dot displays are ideal for real-time monitoring (described above) and for some types of interactive analysis.

Color dot displays offer the possibility of overlaying two-dimension data distributions by representing each population in a different color. No other normal two-dimension displays lend
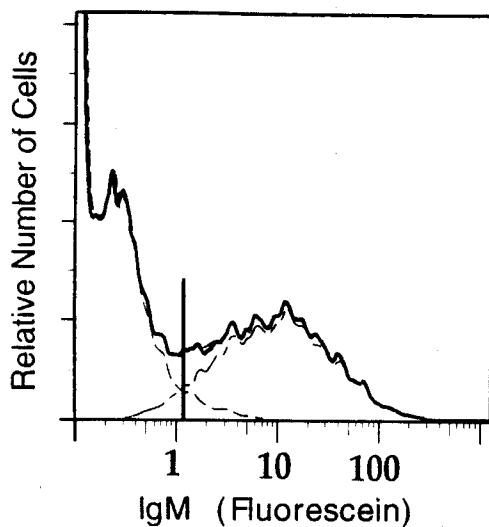
Fig. 50.3. *Estimating positive and negative cell frequencies in overlapping distributions.* Peritoneal cells from a CBA/Ca mouse were stained for IgM (fluorescein) and IgD (Texas Red). Total IgM staining distribution is shown for light scatter gated live cells and separate IgD+ and IgD− components overlaid onto the total. Because the cell types represented in this sample are either IgM+,IgD+ or IgM−,IgD−, the IgD+ and IgD− components can be used to identify the true IgM+ and IgM− cell populations even though their measurements in the fluorescein dimension are overlapping. If the total distribution is divided between "positive" and "negative" at the indicated line, the classification errors of the true IgM positive and negative cells balance and the measured frequencies of the populations equal the true frequencies.

themselves to overlays. Linked color dot plots displaying different pairs of data dimensions from a single sample (as implemented, for example, by Paint-A-Gate, Becton Dickinson, San Jose, CA) can be very useful in exploratory analysis of multiparameter data. When a gating region is defined on one of the displays, the data events in that region are marked with a particular color in all of the displays, making it easy to trace the population characteristics in different views. Moreover, by choosing the "painting" colors carefully, color combining rules can easily show populations that satisfy more than one gate.

Contour plots require more computation, but they can maintain details over a wide dynamic range and can be normalized to adjust for different numbers of cells in the set of data being plotted. This is important when comparable displays are needed for cell populations that have different numbers of events in different cell samples. The usefulness of contour plots depends a great deal on the methods used to derive the surface on which the contours are drawn and on the method used to specify the contour levels. The surface must be smooth enough to avoid having the contours strongly affected by local statistical fluctuations in the recorded events, but it should also preserve statistically significant sharp features. As mentioned above, variable kernel methods have been developed that are generally quite successful in meeting these conditions [8].

Methods for defining contour levels include uniform steps, logarithmic, and probability levels. These are illustrated in Figure 50.2. Uniform density steps are comparable with the uniformly spaced elevation lines of topographic maps (see Fig. 50.2A). They are the easiest to conceptualize, but, we found, the least informative. Logarithmic contouring starts at a level equal to the highest

channel value and then successively decreases by a fixed ratio until the level is at the height of one cell (see Fig. 50.2C). This results in easy visualization of very-low-frequency populations while maintaining a valid representation of large populations. Probability contour levels are chosen so that an equal number of cells falls in the zone between each pair of contour levels [8]. We found that a single default form with probability contours at 5% spacing usually provides good views of immunological data without any user adjustments (see Fig. 50.2B). A useful hybrid to aid in visualization of low-frequency subpopulations while using probability contour levels can be obtained by showing dots for all of the data events outside the lowest contour (see Fig. 50.2E).

Grayscale and color-shaded two-dimension histograms are intermediate between dot displays and contour plots in terms of computation requirements. Like dot plots, these displays rely on visual integration for density estimation. We have not used these methods sufficiently to compare their ability to visualize data with that of the methods described above, but we have seen choices of gradients and color shading that produce markedly inferior displays.

Another type of display closely related to two-dimensional displays is one where a third dimension is added that represents the "channel counts" in the primary two dimensions. These plots are visualized with colors or as "fishnet" models and are useful primarily as final displays to convey an overall impression of the cell populations. They are not readily used for defining gating regions.

Researchers have been working with other methodologies of analyzing and displaying more than two parameters at once, such as principal components [9], but none have come into widespread use yet.

One-dimensional displays

The simple one-dimensional histogram, plotting cell frequency as a function of signal level, has been a venerable analysis tool since the beginnings of flow cytometry. It is still used extensively in DNA analysis [9] and in situations where data can be reduced by either gating or synthesizing to one critical dimension to view a population or several populations of interest. Compared with any of the two-dimension display forms, one-dimension histograms provide easier visual estimation of population frequencies because the area under a region of the curve corresponds to the integrated cell frequency. One of the most useful features on one-dimension displays, as is illustrated in Figure 50.3, is that several may be overlaid on a single set of axes to compare different samples or different populations within a sample. This works well also for sample series analysis as in monitoring antibody titration experiments. In general, density estimation procedures are useful in reducing the effects of statistical fluctuations in one parameter displays, and the smoother density estimated curves produce more readable overlays than when raw data histograms are used.

Sample and subsample statistics

The purpose of deriving statistics from flow cytometry data are to reduce the multidimensional data distribution to one or a few salient features that can be used for evaluation and comparison of results between data sets. The commonly used statistics include population and subpopulation frequencies (including quadrant statistics), measures of central tendency or typical signal levels for

A Gating for Peak Position and CV Evaluation
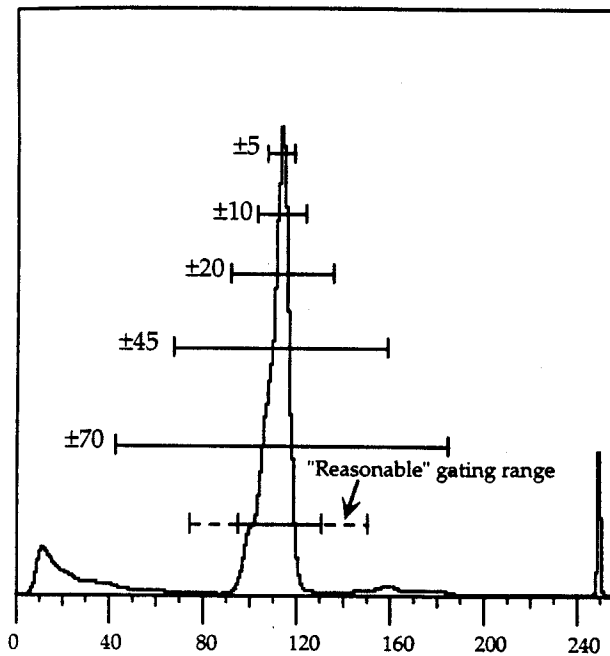
B Mean and Median as Peak Location Indicators
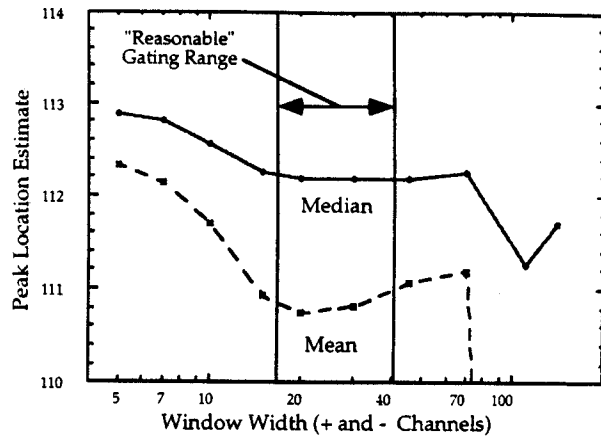
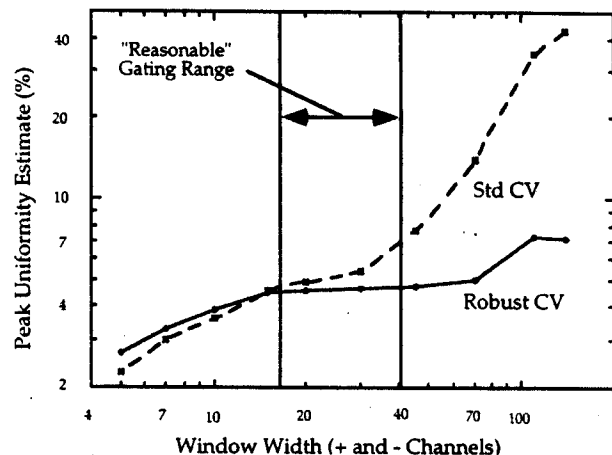**C** Comparison of "Robust CV" with Standard CV

Fig. 50.4. *Robustness of estimators: median and "robust CV" versus mean and CV.* Illustration of how order statistics (based on percentiles) are less sensitive to data gating selection than moment statistics. (**A**) Histogram of the forward light scatter distribution for a microsphere sample. The regions indicated are ± channels from the main population peak position. The "reasonable" gating range covers the range of selections that one would normally expect if different individuals selected gates. (**B**) Comparison of the mean with the median for the different gating ranges shown in A. (**C**) Comparison of the CV of events in the gating ranges with the "robust CV", as defined in the text. The median and robust CV are essentially constant over the "reasonable" gating range and gatings outside the "reasonable" range have less effect on the median and robust CV than on the mean and ordinary CV.

the population (median, mean, mode), and measures of population uniformity (coefficient of variation [CV] and robust CV; defined in *Uniformity/variation within a population*, below). Higher moments of distributions such as skewdness and kurtosis can be computed, but in flow cytometry, graphical displays of the data are usually preferred for conveying the more complex aspects of data distributions.

Robustness and order-based versus moment-based statistics

Important considerations in deciding what statistics to use are their appropriateness to the task at hand and their robustness. For example, an appropriate measure of staining intensity of a fluorescent antibody would be the mean of the linear data if one were comparing flow results to a bulk assay. However, in other situations, the median is often a better value to report for central

tendency (Fig. 50.4). Robustness, in this sense, means that the analysis results should not be too sensitive to limitations in the data or deviations from the assumptions of the analysis. This is relevant when statistics are to be extracted from gated populations, because, in most cases, gating is done visually by the experimenter, leading to variations in the selection of the data values included in the final analysis. A statistic that is very sensitive to changes in the gates has to be used with special care to produce consistent results.

When the population of interest is clearly separated from others by one or a combination of the flow cytometry measurements, the accuracy of the evaluation is limited only by random counting variation in the distribution of recorded events. This variation can be minimized, as mentioned above, by collecting data from a sufficient number of relevant cells (see Table 50.1).

However, even with well-separated populations, there may be ambiguous events due to debris, dead or dying cells, doublets, and so on. If you investigated the distribution of fly wing lengths by measuring the wings of 400 flies, you would know that each data value corresponds to a fly wing length. In this case, the common moment-based statistics (mean, variance, etc.) are valid for characterizing the distribution. In flow cytometry, however, the first problem is to decide which data events properly should be included in the population of interest. The questionable events will usually have measurement values on the margins of the main population. In this circumstance, it is preferable to rely on statistical measures that emphasize the central parts of the distribution and are not greatly affected by inclusion or exclusion of a few events on the periphery. In particular, as illustrated in Figure 50.4, order statistics (median, robust CV, etc.) tend to be less affected than moments (mean, CV, etc.).

Subpopulation frequencies

When positive staining and negative cell populations are not well resolved, there are several ways to estimate the actual fraction of positive cells. In a mixed positive/negative sample, any cutoff chosen to define positive cells will underestimate the actual positive population when the frequency of negative control cells above the cutoff has been subtracted. A more balanced approach is described below. In cases where test and control staining patterns overlap but the biology and/or staining pattern indicate signal levels for all test cells have been elevated from their control levels, it is not appropriate to evaluate a fraction of positive cells. Evaluating the mean or median signal levels for the test and control populations is more reasonable.

Quadrant statistics, frequencies in the four regions obtained by dividing a two-dimension display with a vertical and a horizontal line, are sometimes used as a quick way to evaluate two-color staining. This approach often does not provide optimal delineation of all the cell populations in the sample. The appropriate separation level in one dimension may be different for subpopulations that are high or low in the other dimension, especially if significant fluorescence compensation is required. In addition, the quadrants often contain peripheral events that are not part of any of the primary cell populations.

Population signal levels

Depending on the purpose, several different statistics, including arithmetic mean, geometric mean, median, and mode, may be computed to express the signal level for a cell population. The arithmetic mean (or simply "mean") is needed for comparison of FACS staining levels with corresponding bulk cell assays. In practice, the accuracy of evaluating means is often limited by their sensitivity to small numbers of events with signal levels far from the center of the population. Thus, reasonable differences in gating choices may have a significant effect on the mean, and a significant fraction of offscale events will bias the mean to an unknown extent.

Direct calculation of the mean on logarithmic data leads to what is effectively a geometric mean. For populaions where log presentation is appropriate, this may be a better representation of the signal level of typical cells than the arithmetic mean.

For many purposes, the median is the best choice to represent and compare typical signal levels of cell populations. It automatically falls at corresponding points in log or linear presentation. It

is also robust in that including a few false signals in the calculation or excluding a few outlying events that are really part of the population of interest has minimal effect. The difference between the median and the mean in this respect is illustrated in Figure 50.4. As long as offscale events in the population are fewer than 50%, the median will be completely unaffected.

The mode of a distribution is simply the most common signal value (the "peak" of the distribution). It is not particularly useful because it can be affected by statistical fluctuations in the data bins and it depends strongly one how the data are expressed in that the modes of linear and log presentations of the same data can be quite different.

Uniformity/variation within a population

In flow cytometry, signal uniformity statistics are most commonly used to evaluate measurements on test particles or to express the uniformity of DNA measurements. Because the SD of a population is proportional to its overall signal level, it is common to express relative variation in signal levels in a population by normalizing to the mean, yielding what is conventionally called the coefficient of variation or CV ($=$(SD)/(mean)). The CV is another moment-based statistic like the mean, and its value is even more sensitive to the inclusion or exclusion of a few outlier events and to the loss of information represented by offscale events. Therefore, care should be taken to optimize data gating so that all events that are really part of the population of interest are included and extraneous events are excluded.

For most purposes, and particularly for evaluations and comparisons of standard microspheres, we found the "robust CV" (RCV) [10] to be a useful and reliable substitute for the standard CV. For a normal distribution, the 25th and the 75th percentile are 0.68 SDs from the median. (The 25th and 75th percentiles are particularly easy to compute because they are just the medians of the subpopulations above and below the population median.) In that case, the interquartile range, the difference in signal level between the 25th and 75th percentiles, divided by the median will be 1.36 times the CV, leading to the definition

$$RCV = (\text{interquartile range})/(1.36 \times \text{median}).$$

Figure 50.4 shows a comparison of standard CV and robust CV over a range of data gating choices, illustrating how the robust CV is much more stable. We found the robust CV to be a good monitor for cell sorter alignment using standard microspheres and an automated instrument standardization program. The automated peak-finding algorithm does not have to be extremely accurate in delineating the main microsphere population because we rely on medians and robust CVs in the evaluation process.

Special analysis methods

Evaluation of overlapping populations in one dimension

When there are overlapping populations, any choice of demarcation line will result in misclassification of some events. However, if a separator between two populations can be chosen that gives equal numbers of misclassified events in each direction, the population frequency estimates will be correct. Figure 50.3 shows an overlapping distribution of IgM+ and IgM- cells in an example where we have other data to identify the true IgM+ and IgM− populations. The dividing line marked on the plot gives correct "positive" and "negative" frequencies. In this case, the medians of
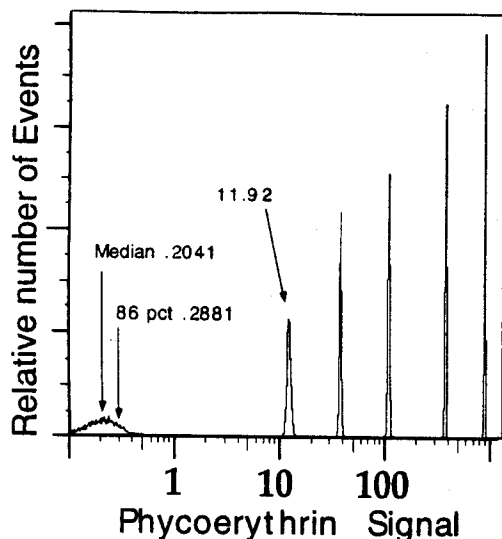
Fig. 50.5. *Estimating fluorescence sensitivity and signal-to-background.* Illustration of a mixture of microspheres with five levels of dye plus a blank population with no added dye (Rainbow Fluorescent Particles, RCP-30-5 from Spherotech, Inc., Libertyville, IL). Light scatter gating has been used to accept only single particles of the main microsphere population. Taking the blank microspheres to represent the background signal level for this type of measurement, we evaluate the median and 86th percentile of the blank event group, obtaining values of 0.2041 and 0.2920 units, respectively. Twice the difference between these values is 0.176 fluorescence units, which we take as the background variation measure (or sensitivity unit). The least fluorescent of the dye labeled groups has median fluorescence of 11.92 units, which gives a net signal about 67 times the background level.

the positive and negative parts of the total distributions are both within 1/2% of the true values, but calculations of mean signal levels will be biased to a greater extent. This is another case in which more robust estimators like the median give more stable results. Of course, in the normal situation for estimating frequencies from an overlapping distribution, we do not have another marker to separate the positive and negative groups, but any controls or extrapolations that allow us to approximate the distributions in the overlap zone can be used estimate the division point that will balance the classification errors and yield correct frequencies.

Sensitivity and signal-to-background evaluation

Continuing our emphasis on robust statistics, we propose the following procedure (illustrated in Fig. 50.5) for evaluating measurements in relation the their appropriate background signal distributions. The method should be useful for comparisons of measurement sensitivity on different instruments. It is also appropriate for signal-to-background evaluations in comparing different reagents, comparing different conjugations of the same reagent, or comparing different concentrations of a single reagent (i.e., in evaluating reagent titrations). The selection of a truly appropriate control (blank microspheres, unstained cells, antigen-negative cells, isotype stained control cells, etc.) is often crucial in obtaining a useful result.

1. Record measurements where test and control cells are both on scale and identify the test and background/control cell populations (or microspheres, etc.) by appropriate gating.

2. Find the medians of the test (Tm) and control (Cm) population signals for the measurement of interest.
3. Find the 86th percentile of the control population (C86). In a normal distribution, this percentile would be one SD above the mean. (Evaluate the medians and 86th percentile in linearized units if they were measured on a logarithmic scale.)
5. The "yardstick" (Y) for comparison is twice the signal difference between the 86th percentile and the median of the control population:

$$Y = 2*(C86 - Cm).$$

6. The ratio (R) of the difference between the test population median and the control population median to this yardstick, that is,

$$R = (Tm - Cm)/Y,$$

can be used as a measure of signal quality, something like a signal-to-noise ratio.

For example, the antibody concentration giving the highest ratio of positive staining to background by this measure would be expected to provide the most sensitive detection for low levels of antigen. The yardstick unit can be expressed as a dye sensitivity for the particular type of measurement if the signal scale has been calibrated with test particles of known equivalent dye content.

The "yardstick" corresponds to two SDs for a normal distribution. If we mixed particles with a background fluorescence distribution (median Cm) with equal numbers of particles with background plus Y added signal (median Cm + Y), we would expect about 95% of the particles measured above signal level Cm + Y to be from the "real" labeled population.

Compared with a direct evaluation of the two SD point, the 86th percentile is better defined and less vulnerable to gating choices that include or exclude a few events on the tail of the control distribution.

Absolute signal levels—photoelectron estimation

One of the fundamental characteristics defining the quality of fluorescence measurements is the number of photoelectrons that contributed to defining the measurement value. For example, even if perfectly uniform light pulses were generated, the measurement uniformity would be limited by the statistical variation in the numbers of photoelectrons produced at the photomultiplier tube photocathode. Thus, knowledge of the relation between number of photoelectrons and signal level allows us to estimate the best possible measurement uniformity at any particular signal level.

The following procedure provides an estimate of the absolute signal level in terms of the number of photoelectrons contributing to the measurement value for a standard particle, establishing a value that can be used to compare different instruments (using the same standard particle sample under similar excitation and emission wavelength conditions), different electronic circuitry (such as peak height versus peak area), or different signal channels on a single instrument. The method presented here is similar to that described by Steen [11]. The procedure works best for particles with good uniformity (CV less than 5%) because it depends on the increase in measurement variability as the amount of light reaching the detector is decreased.
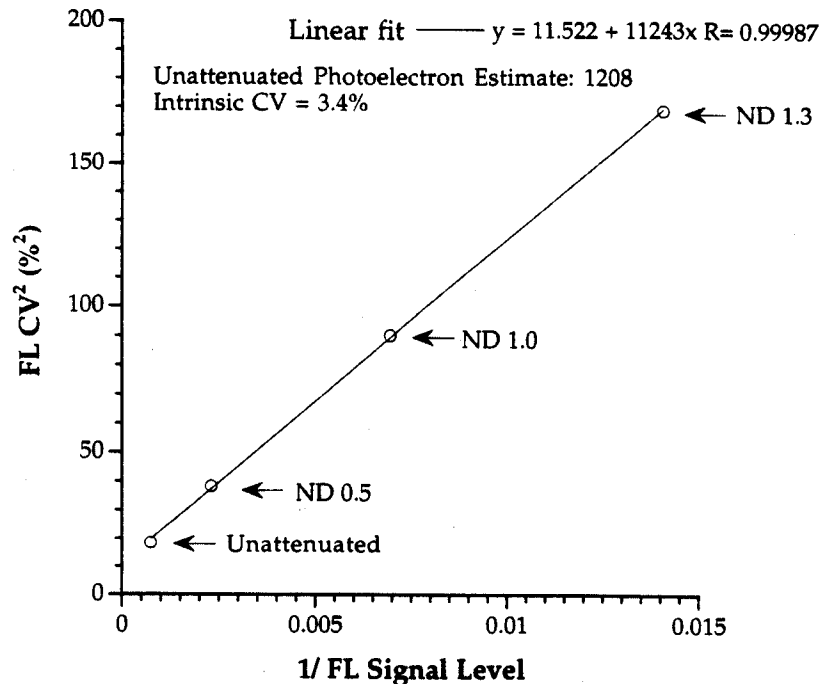
Fig. 50.6. *Photoelectron signal estimation using uniform microspheres.* The standard microspheres were the Rainbow Fluorescent Particles, 3.2 micron (Spherotech, Inc., Libertyville, IL). They were measured on a Becton Dickinson FACStarPlus using 488-nm excitation and a 530/30-nm bandpass filter. Measurements were taken with no attenuation and with ND 0.5, 1.0, and 1.3 for a signal range of about 20-fold. Light scatter gating was used to isolate the main single particle peak in each data set. The mean signal level and CV of that peak in green fluorescence were calculated with exclusion of clear outlier events. The plot shows the reciprocal of the mean signal level versus the square of the CV. A linear fit to the four points gives a constant equaling the square of the intrinsic CV and a slope that is the reciprocal of the photoelectron estimate per unit of signal level.

Materials. Flow cytometer, standard particles, and several neutral density filters in the optical density range from ND 0.3 to 2.0 (such as from Melles-Griot, Irvine CA).

Methods

1. Optimize the instrument alignment using the standard microspheres.
2. Set the selected fluorescence measurement channel to the lowest linear gain (or at least a gain at least 10-fold below the highest available). Adjust the PMT voltage to position the standard particles near but not at the top of the scale. Record data and evaluate the peak signal mean P_0 (=peak position/linear gain) and coefficient of variation (CV) for this configuration. (Light scatter gating will probably be helpful in limiting these evaluations to the main population of single particles.)
3. Insert the lowest o.d. value neutral density filter in the measurement channel. Adjust the gain on the linear amplifier for that channel to bring the microsphere peak to somewhere near midscale. Evaluate the CV and peak signal mean for this condition.
4. Repeat 3 for each neutral density filter. (Stop if the CV goes over about 20%.)
5. Plot $1/P_0$ versus CV^2 as is illustrated in Figure 50.6. Fit a line to these points. The estimate of the photoelectron events (PE) at the PMT on this channel is given by

$$PE = (P_0/m) * 10^4,$$

where P_0 is the unattenuated signal mean and m is the slope of

the fitted line. The 10^4 factor comes from using CV as a percent.

6. The constant in the linear fit is the square of the intrinsic CV of the measurement. The intrinsic CV (CVI) includes inherent variation in the particles plus flow and laser variations but excludes photoelectron statistics. (Because the *x* axis is 1/signal, the intercept corresponds to the projected measurement at infinite signal level.)

In the example shown in Figure 50.6, the fit is y = 11.522 + 11243x and the unattenuated signal mean is 1358 units. This leads to

$$PE = 1358/11,243 * 10,000 = 1208 \text{ photoelectrons}$$

$$CVI = (11.522)^{1/2} = 3.4\%.$$

Acknowledgments

References

1. KANTOR AB, ROEDERER M: FACS analysis of leukocytes. In: Weir DM, Herzenberg LA, Blackwell C, Herzenberg LA, eds. *Handbook of Experimental Immunology*, 5th Ed. Boston: Blackwell, 1996, chap. 49

2. COOK T: It's 10 O'Clock—do you know where your data are? *Technol Rev* 98(1):48–53, 1995

3. PARKS DR: Flow cytometry instrumentation and measurements. In: Weir DM, Herzenberg LA, Blackwell C, Herzenberg LA, eds. *Handbook of Experimental Immunology*, 5th Ed. Boston: Blackwell, 1996, chap. 47

4. PARKS DR, BIGOS M: Automatic standardization, compensation adjustment and instrument diagnosis for flow cytometers. *Cytometry* Suppl 4:70, 1990

5. DATA FILE STANDARDS COMMITTEE OF THE SOCIETY FOR ANALYTICAL CYTOLOGY: Data file standard for flow cytometry. *Cytometry* 11:323–332, 1990

6. TSIEN R: Fluorescent indicators of ion concentrations, in *Methods in Cell Biology*, edited by Taylor DL, Wang Y-L, Academic Press, 1989:127–156

7. ROEDERER M, BIGOS M, NOZAKI T, STOVEL RT, PARKS DR, HERZENBERG LA: Heterogeneous calcium flux in peripheral T cell subsets revealed by 5 color flow cytometry. *Cytometry* 21:187–196, 1995

8. MOORE WA, KAUTZ RA: Data analysis in flow cytometry, in *The Handbook of Experimental Immunology*, edited by Weir DM, Herzenberg, LA, Blackwell, CC, Herzenberg, LA, 4th Ed. Edinburgh: Blackwell Scientific, 1986, pp 30.1–30.11

9. DEAN PN: Data processing, in *Flow Cytometry and Sorting, 2nd Ed*, edited by Melamed M, Lindmo T, Mendelsohn M, New York, Wiley-Liss, 1990, pp 415–444

10. SHAPIRO HM: *Practical Flow Cytometry*, 3rd Ed. New York, Wiley-Liss, 1994

11. STEEN H: Noise, sensitivity, and resolution of flow cytometers. *Cytometry* 13:822–830, 1992