Evaluation of a Modeling System for S-Phase Estimation in Breast Cancer by Flow Cytometry

Lynn G. Dressler, Larry Seamer, Marilyn A. Owens, Gary M. Clark, and William L. McGuire

University of Texas Health Science Center, Department of Medicine, Division of Oncology, San Antonio, Texas 78284 [L. G. D., G. M. C., W. L. M.], and Nichols Institute, San Juan Capistrano, California 92675 [L. S., M. A. O.]

ABSTRACT

Using software programs provided by Coulter Electronics, we have developed an analysis system that would address problems encountered in DNA flow cytometric analysis of heterogeneous solid tumor populations, especially where the G2-M phase of the diploid population contaminates the S-phase of the aneuploid population, causing an overestimation of cells in S phase. We used the PARA 1 and PARA 2 programs in concert and developed three analysis models: (a) for euploid tumors; (b) for hyperdiploid tumors with overlapping populations; and (c) for near-diploid aneuploid tumors. Our purpose in this paper is to define the limits and reproducibility of this analysis system with an emphasis on tumors with overlapping populations. Aliquots of frozen, pulverized breast tumor tissue (50 to 100 mg), routinely used in the steroid receptor assay, were used for routine flow cytometric measurement of the DNA index and S-phase fraction. To determine the accuracy of the analysis when overlapping populations were present, we mixed an aneuploid breast cancer cell line with human blood lymphocytes in varying ratios. A 10% mixture of aneuploid cells, the lowest mixture tested, still allowed analysis results within 95% confidence limits. Reproducibility of the system was assessed on frozen breast tumor tissue by intra- and interassay variation studies measuring cell cycle parameters and coefficient of variation of the G0-G1 peak width. Within any sample the variation (±2 SD) for the G0-G1 value was ±2.40 for intraassay and ±2.60 for interassay, and the amount of variation for S phase was ±3.0 and ±3.2 for intraassay and interassay, respectively. There was no difference in the variation of estimates for G2-M (±2.6 for both intra- and interassay). In this study, the coefficient of variation of the G0-G1 peak greater than 5% was defined as unacceptable for accurate analysis, with the conclusion that S-phase fractions in aneuploid tumors can be routinely analyzed in human breast tumor biopsies despite tumor cell heterogeneity.

INTRODUCTION

Recent studies have shown that DNA measurements of S phase and ploidy are prognostically significant in breast cancer (1-11) and in other neoplasias (12, 13). Improved instrumentation and methodologies in flow cytometry in the last decade have made routine DNA analysis of clinical material feasible.

Single-parameter DNA histograms obtained by flow cytometry can be compartmentalized into three phases: G0-G1; S; and G2-M. Because these phases overlap, certain assumptions must be made in constructing an analysis algorithm to estimate the relative size of each phase. Many algorithms have been developed (14-18, 19, 20), using different sets of assumptions. Most work well with asynchronous cell culture or synthetic data. However, problems arise when these algorithms are used to analyze histograms from clinical material such as solid tumors. One of the most often encountered problems is that of heterogeneous populations. This is commonly seen when a tumor has an aneuploid population in addition to a normal diploid population. The G2-M phase of the diploid population contaminates the S phase of the aneuploid population, causing an overestimation of the aneuploid S phase. A second problem is encountered when ploidy level or DNA content of the aneuploid population is close to that of the diploid population, resulting in significant overlap of both populations in all phases.

It was our objective to develop and define the limits and reproducibility of an analysis system that would address these problems. We first began by analyzing DNA breast cancer histograms using the analysis program designated PARA 1, included in the EASY software package developed by Dr. C. Bruce Bagwell for Coulter Electronics (Hialeah, FL). It was found, however, that problems of overlapping populations in solid tumor histograms prevented the analysis of many tumors. We then explored the possibilities of using another EASY software package designated PARA 2 in concert with the PARA 1 program. The PARA 2 analysis, based on linear/nonlinear least-squares techniques, allows users to construct models representing the data and to change various parameters of the analysis system. Although this program was designed to analyze DNA cell cycle data, we were not aware of any successful attempts to apply this modeling system to clinical data. Consequently, with the help of Dr. Bagwell, we combined the PARA 1 program with the PARA 2 program and developed three modeling systems. The purpose of this paper is to define the limits of this modeling system for analyzing cell cycle compartments to include evaluation of the reproducibility of the system through intraassay and interassay studies, and sensitivity through mixing experiments.

MATERIALS AND METHODS

Tissue Preparation

Breast tumor tissue obtained at surgery from various hospitals was received on dry ice and stored in liquid nitrogen until analysis. While still frozen, the tissues were minced with a scalpel and ground to a coarse powder by mortar and pestle. Approximately 100 mg of powder were then gently hand homogenized with a loose-fitting Teflon pestle in 1.5 ml of homogenization buffer (0.04 M Tris:0.5 M sucrose: 1.5 mM MgCl2, pH 7.4) (21). The homogenate was filtered through successive layers of 210- and 53-μm nylon mesh (Small Parts, Inc., Miami, FL). Subcellular debris was removed from the filtrate by differential centrifugation through a double cushion of sucrose, 1.50 and 1.75 M, respectively. Following centrifugation at 1500 × g for 45 min, the supernatant was carefully removed, and the cellular pellet was resuspended and washed with MEM2 containing 10% FBS (Irvine Scientific, Irvine, CA). Cell counts were performed using a hemocytometer, averaging 2 × 105 (range, 2 × 104 to 2 × 106) per 100 mg of frozen, pulverized tumor. CRBC (Colorado Serum Co.) in phosphate-buffered saline (0.01 M NaH2PO4:0.15 M NaCl:0.1% sodium azide, pH 7.4) were added as an internal standard (final concentration of 5% in the total sample population). Cells were lysed and stained by incubation in a modified Krishan

Received 7/21/86; revised 7/8/87; accepted 7/13/87.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Supported by NIH Grant CA30195.
2 Present address: University of New Mexico Cancer Center, Solid Tumor Facility, Albuquerque, NM 87131.
3 Present address: Stanford University, Department of Biology, Stanford, CA 94305.
4 To whom requests for reprints should be addressed, at University of Texas Health Science Center, Department of Medicine/Oncology, 7703 Floyd Curl Drive, San Antonio, TX 78284-7884.
hypotonic sodium citrate buffer (3, 22) containing 0.1% sodium citrate (J. F. Baker Chemical Co.), 0.3% Nonidet P-40, nonionic detergent (Particle Data Laboratories, Elmhurst, IL), 0.05 mg/ml propidium iodide (Calbiochem), and 0.02 mg/ml RNase A (Sigma). Tubes were immediately vortexed for 10 s and allowed to incubate for at least 30 min on ice. Following centrifugation, the nuclei were resuspended in fresh staining buffer, syringed through a 27-g needle to break up any clumps, and filtered through 37-μm nylon mesh immediately prior to flow analysis.

Cell Preparation for Mixing Experiment

We determined the accuracy of analysis and assessed the minimal percentage of aneuploid cells that would give reproducible results by performing mixing experiments. hpBLs were mixed with a hyperdiploid breast cancer cell line, designated T47D. The following mixtures of T47D:hpBL were achieved: 100:0; 80:20; 70:30; 60:40; 50:50; 40:60; 30:70; 20:80; and 10:90. hpBLs were prepared fresh by ficoll separation from heparinized whole blood. T47D cells were harvested by EDTA, washed in MEM, and stored frozen in MEM plus 10% glycerol until assay. On the day of the assay both cell preparations were washed with MEM plus 10% FBS, mixed in their appropriate ratios, and stained in the modified Krishan buffer as described above. Cells were counted in a Model S Coulter Counter prior to mixing to ensure appropriate mixing ratios. DNA analysis was performed using Model 2 (PARA 2 analysis).

Analysis

Propidium iodide-stained nuclei were analyzed on a EPICS V flow cytometer (Coulter Electronics, Hialeah, FL) fitted with a single Inova 90 argon ion laser (Coherent Laser Products Division, Palo Alto, CA). Laser emission was 400 mW at 488 nm. Fifty thousand sample events were acquired on a single-parameter 256-channel integrated fluorescence histogram. CRBc were added as an internal control to help identify the G0-G1 diploid population. These were not included in the count of 50,000 sample events. Photomultiplier tube high voltage was adjusted to maintain the CRBC peak in Channel 20. The presence of a diploid population in every breast cancer specimen run provided an additional internal control. CRBC nuclei served as a threshold for acceptable fluorescence intensity; events having less fluorescence than CRBC nuclei were considered debris and were gated out of the histogram.

The DNA histograms obtained as described above were analyzed for cell cycle distribution on a Terak 8600 color computer (Terak, Inc., Scottsdale, AZ) using the PARA 1 and PARA 2 programs provided by Coulter Electronics in their EASY software package.

PARA 1 Analysis

Each histogram was first analyzed using the PARA 1 or Parametric Analysis 1 program (Easy 1.2 Software Program; Coulter Electronics). PARA 1 is a system of subroutines designed to analyze cell cycle distributions in asynchronous DNA histograms. The analysis is parametric in the sense that the 256 numbers that constitute a DNA histogram are reduced to 4 biologically meaningful parameters: percentages of G0-G1; S; G2-M; and CV. It uses a least-squares analysis system to find three gaussian peak positions with an efficiency of about 0.1%. For our purpose, peaks assigned for PARA 1 analysis were: Peak 1, G0-G1 peak of the population of interest; Peak 2, G2-M of Peak 1 population; and Peak 3, G0-G1 peak of the diploid population in aneuploid tumors. In diploid tumors, Peaks 1 and 3 were identical. PARA 1 calculates the peak height and SD of each gaussian. To avoid error due to skewness caused by debris or S-phase contamination, the program calculates the SD based on the steepest side of each gaussian. The percentages of G0-G1 and G2-M are assessed by determining the area under those two peaks. The S phase is calculated by subtracting the total number of events in the G0-G1 and G2-M phases from the observed data. The program stores these values in an analysis result register for later access by the PARA 2 program. Details of the mathematics associated with these analyses can be found in Dr. Bagwell's thesis dissertation (23).

PARA 2 Analysis

The output from PARA 1 analysis can be used as estimates for Parametric Analysis 2 (PARA 2 EASY 1.2 Software Program; Coulter Electronics). Each histogram was subsequently analyzed using the PARA 2 program. As described by Dr. Bagwell in the Coulter User's Manual, the PARA 2 is a hybrid analysis system. This program uses the Marquardt least-squares algorithm (24) to drive a linear least-squares procedure. The program takes advantage of the fact that complex models are generally a mixture of linear and nonlinear parameters. To calculate the optimal value of linear parameters is a relatively simple and fast operation. However, the determination of optimal nonlinear parameters is much more time consuming. The linear parameters turn out to be the heights of the various model components (e.g., gaussian peak heights, trapezoid heights, and exponential height). The nonlinear parameters are the channel positions of the model components and the linespread function (25).

The philosophy of the PARA 2 program is to allow the biologist or clinician to make assumptions concerning data (not the programmer). Model components (gaussians, exponentials, and one or more sets of trapezoids or rectangles) are put together by the user and can be displayed at any time during the construction process. The model components are chosen to distinguish and quantitate overlapping distributions. The PARA 2 analysis system is a trilogy composed of: (a) the program designated MODSET, that portion of the program which allows the user to edit parameters; (b) the program designated PARA 2, the nonlinear/linear least-squares analysis program which is actually doing the calculations of the compartments after fitting the raw data to the model; and (c) the program designated GRAFP2, which enables the user to "see the fit" of the modeled or processed data to the raw data as a graphical representation and shows the user the distribution of under- and overfitting by plotting the reduced chi-squared residuals. Reduced chi-squared residuals are computed using methods described by Bevington (26).

The total reduced chi-squared statistic that tests the goodness of fit of the model is decomposed into single degree of freedom chi-squared statistics for each of the 256 channels. Residuals from the expected value of 1.0 are graphically displayed for each channel.

It is within the MODSET program that the user's modifications are made. There are three areas of modifications: (a) editing model components; (b) editing least-squares analysis controls; and (c) editing algebraic equations to define the model compartments. It is not the purpose of this paper to discuss the definitions of each of these MODSET parameters, as this can be found in detail in the User's Manual from Coulter Electronics. There are, however, some salient points which require definitions here. First, model components (gaussians, exponentials, or trapezoids to model S phase) can be one of three types: (a) floating, designated by "0" in the program and a double-headed arrow on the graph (→), which indicates that the parameter "floats" or is free to be changed by the analysis program to find its optimal value; (b) fixed, designated by "1" in the program and by the downward arrow (↓) on the graph, indicating that the parameter is fixed and does not change at all during the analysis; and (c) dependent, designated by a number greater than zero, corresponding to a particular parameter number as given in the program, and designated by a horizontal bar (—) in the graph, indicating that the parameter is dependent on some other parameter, either multiplicatively or additively. S phase can be modeled by single or multicompartamental trapezoids depending on the data; the height of the sides of the trapezoid may vary, and therefore it has the potential to be a variety of shapes. In the three models described below, the S phase is pictured as being modeled to a rectangle, the most common shape observed to model the data. However, this rectangular shape has the potential to become more trapezoidal in nature if the data warrant.

When the linear/nonlinear least-squares program, PARA 2 (Part 2 of the trilogy), is executed, it requires exact information on how it is to operate. This information comes from the least-squares analysis controls, another section which the user can edit. The analysis controls are extracted from Marquardt's nonlinear least-squares algorithm, which has been modified by Bruce Bagwell within the Coulter Electronics program. These controls include 13 parameters which are listed in Table 1. As per Dr. Bagwell's suggestions, only 5 of these parameters

5295
were altered: Parameter 1, initial $\lambda$, was changed to a value of 0.25; Parameter 2, maximum iterations, was changed to 3; Parameter 5, $\delta$ parameter factor, was changed to 1.00; Parameter 9, Marquardt $\lambda$ factor, was changed to 2; and Parameter 13, $\delta$ stabilization factor, was changed to 2 (Table 1). The reader is referred to the Coulter User's Manual for a full description of these parameters.

In addition to editing model components and least-squares analysis controls, the user has the ability to write algebraic equations defining the model components. The equation-building variables are explained in detail in the User's Manual. Using the PARA 2 trilogy, we developed three models to allow maximum analysis of breast cancer histograms. All three models shared the same least-squares analysis controls (Table 1), but differed in their model components and component equations.

**Model 1.** Model 1 is the simplest of the three models (Fig. 1A). It consists of two gaussians and a single trapezoid. The lower and upper boundaries of interest are represented as vertical dotted lines. The peak position of the gaussian is represented as solid vertical lines; the height of the line is equal to the height of the broadened gaussian. Vertical lines (a and b) model the G$_0$-G$_1$ and G$_2$-M peak positions, respectively. The S phase is modeled by a single trapezoid (c), whose upper and lower boundaries are defined as the mean position of the two gaussians (a and b). Debris is modeled by a negative slope exponential (d; slope $= -0.04$) with its lower boundary (e) set between the CRBC peak (not pictured in model) and the diploid G$_0$-G$_1$ peak. Symbols indicate the type of dependency which is assigned to each model component or parameter: fixed, designated by the downward arrow (v); floating, designated by the double-headed arrow ($\leftrightarrow$); and dependent, designated by the horizontal bar (---). In Model 1, the boundaries of interest are fixed, the left boundary of the exponential is dependent upon the left boundary of interest, the G$_0$-G$_1$ peak position (a) is floating, and the G$_2$-M peak position (b) is fixed. The boundaries of the trapezoid (c) that models the S phase are dependent upon the position of the G$_0$-G$_1$ and G$_2$-M. The standard deviation of the peak position is represented by the angle of the upward arrows ($\uparrow$) at the top of the graph. The initial position and standard deviation of G$_0$-G$_1$ and G$_2$-M peaks are obtained from the PARA 1 analysis register. The G$_0$-G$_1$ peak position is allowed to deviate from these values in order to obtain minimal overfitting and underfitting with the raw data as measured by reduced $\chi^2$ residuals (GRAFP2). The program is allowed to go through 3 iterations to find this fit. This model was used to analyze histograms where no overlapping populations were apparent (e.g., diploid tumors) and those where it was mathematically impossible to distinguish elements of one population from another (e.g., tetraploid tumors). Table 2 indicates algebraic equations for Model 1. Fig. 1B shows this model superimposed on a diploid DNA histogram obtained from a frozen breast cancer tumor. The peak at Channel 20 represents CRBC as an internal standard.

**Model 2.** Model 2 is most useful when dealing with overlapping DNA populations (Fig. 2A). It consists of gaussians modeled for the diploid G$_0$-G$_1$, and G$_2$-M and gaussians for the aneuploid (hyperdiploid) population. Model 2. Model 2 is most useful when dealing with overlapping DNA populations (Fig. 2A). It consists of gaussians modeled for the diploid G$_0$-G$_1$, and G$_2$-M and gaussians for the aneuploid (hyperdiploid) population. Model 2. Model 2 is most useful when dealing with overlapping DNA populations (Fig. 2A). It consists of gaussians modeled for the diploid G$_0$-G$_1$, and G$_2$-M and gaussians for the aneuploid (hyperdiploid)

**MODELS OF BREAST CANCER PROLIFERATION**

**Table 1 Least-squares analysis control parameters used for all three models**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>Initial $\lambda$</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>Maximum iterations</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Gaussian matrix stabilizer ($0 = ff$)</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>X-axis transform ($0 = linear$, $1 = log$)</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>$\delta$ parameter factor</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>Terminating reduced $\chi^2$ change</td>
<td>0.10</td>
</tr>
<tr>
<td>7</td>
<td>Initial $\lambda$ parameter factor</td>
<td>0.01</td>
</tr>
<tr>
<td>8</td>
<td>Compartment size type ($0 = linear$, $1 = exponential$)</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>Marquardt $\lambda$ factor</td>
<td>2.00</td>
</tr>
<tr>
<td>10</td>
<td>Trapezoid 1-matrix stabilizer</td>
<td>0.001</td>
</tr>
<tr>
<td>11</td>
<td>Trapezoid 2-matrix stabilizer</td>
<td>0.00</td>
</tr>
<tr>
<td>12</td>
<td>Precision of calculations (no. of SD)</td>
<td>4.00</td>
</tr>
<tr>
<td>13</td>
<td>$\delta$ stabilization factor</td>
<td>2.00</td>
</tr>
</tbody>
</table>

* These parameters were altered in our models.
Compartment equations were constructed for each model developed as per the instructions in the Coulter Electronics User’s Manual for Easy 1, 2 software programs. G(1) = area of the “ith” gaussian such that: G(1) = gaussian of the G0-G1 of population of interest; G(2) = gaussian of the G2-M of population of interest; G(3) = gaussian of the G0-G1 of diploid population; G(4) = gaussian of the G2-M of diploid population; T1 = total area of trapezoid Set 1; M(1) is the value of the “ith” model parameter where: M(1) = peak position of gaussian 1; M(2) = peak position of gaussian 2; and M(7) = SD of peak position of gaussian 1. Note the percentage of S phase is calculated for the population of interest; i.e., in Model 2, the percentage of S phase in Rectangle 1 reflects the S phase of the aneuploid population. As in Model I debris was removed by modeling to a negative slope until a fixed region is reached. The shaded region in Fig. 3C illustrates this type of fit, superimposed on the raw data (unshaded region). Fixing the peak positions and SDs obtained from the PARA 1 analysis results in an accurate fitting of the aneuploid G0-G1 peak (Fig. 3D, shaded region). Algebraic equations for Model 3 are found in Table 2.

### Data Analysis

The results from the PARA 1 and PARA 2 programs were compared using paired *t* tests. The parameters analyzed included the percentage of nuclei found in the G0-G1, S, and G2-M phases of the cell cycle; DI; and CV of the G0-G1 peak.

Inter- and intraassay variabilities were assessed using one-way analysis of variance, where the individual tumor specimens represented the grouping factor and the replicate observations represented the within-specimen factor. The total variance was partitioned into components representing “between tumor” and “within tumor” variability. The “within tumor” standard deviation was used to compute approximate 95% confidence intervals for future measurements.

The effect of different sample weights was examined using one-way repeated-measures analysis of variance, where the repeated factor consisted of the various tumor weights. PAIRWISE comparisons were performed using the Newman-Keuls multiple comparisons procedure.

### RESULTS

Comparison of PARA 2 and PARA 1 Analyses. We compared cell cycle estimates (percentages of G0-G1, S, and G2-M), DI, and coefficient of variation of the width of the G0-G1 peak (CV G0-G1) obtained from the PARA 1 and PARA 2 programs. A total of 14 different tumors (7 diploid and 7 aneuploid) was prepared and analyzed with both programs. Each tumor had 3 to 10 replicate runs for a total of 101 pairs of estimates. For the PARA 2 analyses, Model 1 was used for the diploid tumors and 2 aneuploid tumors having a tetraploid DNA content, and Model 2 was used for the remaining aneuploid tumors, all of which were hyperdiploid.

No significant differences were observed between the PARA 1 and PARA 2 estimates of DI or CV G0-G1, indicating that the modeling programs were not significantly altering peak positions of the raw data. In addition, no significant differences were observed in the estimate of percentage of tumor nuclei having a G2-M DNA content.

However, significant differences in estimates of the percent-
MODELS OF BREAST CANCER PROLIFERATION

Fig. 3. PARA 2 Model 3. Model 3 is very similar to Model 1 but is not allowed to deviate from designated fixed conditions. This model allows analysis of near-diploid tumors. Symbols are defined as in Fig. 1. In A, as in Model 1, the S phase is modeled by a single trapezoid (c) with the upper and lower boundaries defined as the peak positions of the G0-G1 gaussian (a) and the G2-M gaussian (b). Unlike Model 1, the initial peak positions and standard deviations (obtained from PARA 1 analysis, see text) are not allowed to deviate. In B, Model 3 (shaded region) is superimposed on a near-diploid aneuploid tumor (unshaded region). The gaussian models the G0-G1 peak of the aneuploid population (G0/G1). In C, Model 3 is allowed to deviate from fixed gaussians and standard deviations as calculated by PARA 1 (see text). The shaded area illustrates that the fit of the model is incorrect, locating the G0-G1 peak at a point midway between the diploid (G0/G1) and aneuploid (G0/G1) populations. In D, accurate fitting of Model 3 is illustrated by the shaded region superimposed on the raw data (unshaded) after fixing the peak positions and standard deviations.

Fig. 4. Cell cycle values: PARA 1 versus PARA 2. Fourteen different tumors (Nos. 1 to 14) were prepared and analyzed by PARA 1 (∥) and PARA 2 (∥∥) Models 1 and 2) programs. Each column represents the mean result obtained from at least 3 replicates (range, 3 to 10 replicates/tumor) of 100-mg aliquots of frozen breast tumor; bars, SE. A. G0-G1 Values. PARA 1 analysis demonstrates consistently lower G0-G1 cell cycle values as compared to PARA 2 analysis. B. S-phase values. PARA 1 demonstrates consistently higher S-phase values as compared to PARA 2 analysis. The actual differences in S-phase results are dependent upon the individual tumor. G2-M results from PARA 1 and PARA 2 (not shown) were practically superimposable.

age of nuclei in G0-G1 and S phases were observed (Fig. 4). Fig. 4A graphically illustrates the mean G0-G1 results of replicate runs on 14 tumors obtained by PARA 1 analysis (∥) and PARA 2 analysis (∥∥). One observes that estimates from the PARA 1 program were consistently lower than those obtained from the PARA 2 program. The mean difference and standard error between the two programs for estimation of the percentage of G0-G1-phase nuclei were 7.0 ± 0.3 (P < 0.0001). This calculation was based on all replicate runs for the 14 tumors, a total of 101 pairs of estimates. Conversely, Fig. 4B shows that estimates of S-phase values were consistently higher using the PARA 1 analysis (∥) as compared to the PARA 2 analysis (∥∥). The mean difference and standard error of the 101 pairs of estimates were 7.3 ± 0.3 (P < 0.0001). Fig. 5 shows an example of a DNA histogram from one of these 14 tumors, where the G2-M of the diploid population (G2M) overlaps into the S phase of the aneuploid population. The inset shows a comparison of cell cycle values obtained by PARA 2 (P2; Model 2) and PARA 1 (P1) analysis. We observe a striking decrease in the estimation of the percentage of S-phase nuclei using PARA 2, with a concomitant increase in the percentage of G0-G1 of the hyperdiploid population. We decided to further test the accuracy of the PARA 2 analysis, in particular Model 2, by performing mixing experiments of hyperdiploid and diploid cell lines.

Mixing Study. In order to examine the accuracy of the PARA 2 (Model 2) analysis as well as the percentage of aneuploid cells evaluable by our model and methodology, we measured cell cycle parameters of the aneuploid breast cancer cell line, T47D, mixed with hPBLs at various ratios and compared these results to those obtained from the pure (100%) T47D population. Fig. 6 graphically illustrates G0-G1 values (∥) and S-phase values (∥∥) obtained at the following T47D:hPBL mixing ratios: 10:90; 20:80; 30:70; 40:60; 50:50; 60:40; 70:30; and 80:20. The shaded regions represent the 95% confidence limits defined as 2 SD from the values obtained for the pure (100%) aneuploid population. For the 100% aneuploid population, the percentage of DNA Fluorescence (Channel Number)

![Cell Cycle Analysis](http://example.com/dna-fluorescence)

![Histogram of a hyperdiploid tumor illustrating contamination of the aneuploid S phase with the diploid G2-M peak (G2M). The inset compares cell cycle analysis results obtained from PARA 1 (P1) and PARA 2 (P2) (Model 2).](http://example.com/histogram)

Downloaded from cancerres.aacrjournals.org on April 6, 2017. © 1987 American Association for Cancer Research.
MODELS OF BREAST CANCER PROLIFERATION

Fig. 6. Accuracy of cell cycle analysis by mixing experiment. Accuracy of cell cycle analysis was assessed by comparing the Go-G1 and S-phase cell cycle values obtained at each mixing ratio of the hyperdiploid T47D breast cancer cell line with normal peripheral blood lymphocytes (T47D:hPBL: 10%; 20%; 30%; 40%; obtained at each mixing ratio of the hyperdiploid T47D breast cancer cell line cycle analysis was assessed by comparing the G0-G1 and S-phase cell cycle values were within the 95% confidence interval at each mixing ratio.

Fig. 7. Accuracy of model fitting in mixing experiment. The hyperdiploid breast cancer cell line, T47D, was mixed with normal human peripheral blood lymphocytes (hPBLs) at various ratios and incubated in hypotonic propidium iodide buffer as described in the text. Three mixing ratios of T47D to hPBLs are pictured: 10% (A); 40% (C, D); and 70% (E, F). An EPICS V flow cytometer (Coulter Electronics) equipped with an argon laser (488 nm, 200 mW) was used to generate DNA histograms. The shaded regions represent the 95% confidence limits. The percentage of S phase (not picured) likewise showed little variability and remained within the 95% confidence interval at each mixing ratio.

nuclei in Go-G1, was 79.5% (——). One observes that the variability of the Go-G1 percentage at any one mixing ratio was either within or very close to the 95% confidence limits of the 100% aneuploid population. The fraction of nuclei in S phase likewise showed very little variability, even less than that observed for the percentage of Go-G1. The 100% aneuploid population had a percentage of S phase of 13.0%. At every mixing ratio, the percentage of S phase was always within the 95% confidence limits. The percentage of nuclei in G2-M (not pictured) likewise showed little variability and remained within the 95% confidence interval at each mixing ratio. Fig. 7 illustrates the histograms obtained from three of these mixing ratios (10%, 40%, and 70%) and the 100% aneuploid cells, as well as the corresponding “fit” of Model 2 to the raw data (graphs displaying the “fit” were generated using the GRAFP2 software program provided by Coulter Electronics). The shaded regions in the graphs represent the processed data (the actual peaks the modeling program found) superimposed on the raw data or nonshaded region. All graphs showing the “fit” (B, D, F, H) have an expanded ordinate scale to allow visualization of processed to raw data. Fig. 7A illustrates that 10% mixing generates a well-defined aneuploid peak (Go/G1A) in the DNA histogram and that there is a near superimposition of the processed data (shaded) to the raw data (Fig. 7B, unshaded). Mixing of 40% aneuploid cells (C and D) and 70% aneuploid cells (E and F) likewise shows a close fit of the processed data to the raw data, although slight underfitting of G2-M peaks is observed. Results from the mixing study showed that the PARA 2 analysis (Model 2) was able to provide accuracy of analysis with a 95% confidence interval for this particular mixed population and that we were able to detect at least 10% aneuploidy. We wanted to try to define the variability of the analysis system in a variety of “real,” human breast cancer biopsies and proceeded to do reproducibility studies as described below.

Reproducibility. Reproducibility of the PARA 2 modeling algorithms was evaluated by analyzing intra- and interassay variation. Intraassay variation was determined by comparing the analysis results obtained from at least 3 separate 100 mg aliquots (range, 3 to 5 replicates) of frozen breast tumor powder from each of 9 tumor biopsies processed on the same day. Interassay variation was evaluated by comparing results obtained from at least 7 replicates (range, 7 to 10 replicates) of frozen breast tumor powder from each of 8 tumor biopsies processed on nonconsecutive days. (Tumor biopsies for interassay were different specimens from those used for intraassay.) Preliminary statistical analyses (not shown) indicated that the SDs for the parameters of interest did not depend on the magnitude of the parameter being estimated. For example, an S-phase value of 7.95 from a particular tumor showed the same variability (SD) as an S-phase value of 23.77 from another tumor.

In the intraassay studies, 4 of the 9 tumor samples used were aneuploid. S-phase values for all 9 tumors ranged from 2.9 to 23.8% with a mean value of 9.2% (median, 8.4%). Table 3 illustrates results of the intraassay study. The mean between-tumor SD and within-tumor SD for Go-G1, S, and G2-M cell cycle analysis as well as the CV of the Go-G1 peak width were calculated for all 9 tumors. The between-tumor SD is simply an indication that values from the 9 tumors were different. From the table we observe the greatest between-tumor SD in Go-G1 and S-phase estimates. This reflects a heterogeneous distribution of cell cycle values in these 9 tumors. The relatively small between-tumor SD of the CV Go-G1 indicates little change in the width of the Go-G1 peak, reflecting consistency of sample preparation and instrumentation. The within-tumor SD gives a measure of the variability one observed in replicates of the same sample. We observed within-tumor SD of 2.2 for the Go-G1-phase estimates. Variability of S-phase estimates was slightly lower.

Fig. 7. Accuracy of model fitting in mixing experiment. The hyperdiploid breast cancer cell line, T47D, was mixed with normal human peripheral blood lymphocytes (hPBLs) at various ratios and incubated in hypotonic propidium iodide buffer as described in the text. Three mixing ratios of T47D to hPBLs are pictured: 10% (A); 40% (C, D); and 70% (E, F, H). An EPICS V flow cytometer (Coulter Electronics) equipped with an argon laser (488 nm, 200 mW) was used to generate DNA histograms. The shaded regions (shaded region) to unprocessed or raw data (unshaded region) (B, D, F, H). All mixing ratios showed a good fit of processed data to raw data. The pure or 100% aneuploid population of T47D nuclei is pictured in G and H.
Go-G. We would expect S-phase measurements to be accurate with 95% confidence and G2-M-phase measurements to be accurate to about ±3.0% of all preparations are performed on the same day or intraassay and interassay estimates of cell cycle phases and CV Go-G1. The similarity of the SDs for intra- and interassay suggests that the 95% confidence intervals obtained from the intraassay study describe the major variance inherent to this estimation and preparation system. Both intra- and interassay studies indicate a high degree of reproducibility.

Aneuploidy and CVs. In our initial pilot studies (n = 40), we observed an occasional “hidden” aneuploid peak in the Go-G1 region of a seemingly diploid tumor, or a second (biclonal) aneuploid peak in a seemingly uniclonal aneuploid during repeat analyses. We were able to detect a second peak on at least one replicate of 6 different tumors run at least 5 different times. We reviewed the CV of the Go-G1 peaks of these tumors and compared them with CV Go-G1 from single-peak tumors to see if a CV value could be obtained which might predict a possible hidden aneuploid peak. The mean CV of the Go-G1 peak of these 6 tumors was as follows: 5.35% for “diploid” tumors (n = 4, range, 4.25 to 6.13); 5.02% for a hyperdiploid tumor (n = 1); and 5.70% for a tetraploid tumor (n = 1). Overall, the mean CV of the Go-G1 peak of these hidden “double peak” tumors was 5.36%. Tumors that demonstrated only single peaks (n = 27) on repeated assay had the following CVs: 4.27% for diploid tumors (n = 10; range, 3.75 to 5.07); 4.17% for hyperdiploid tumors (n = 12; range, 3.52 to 5.03); and 4.91% for tetraploid tumors (n = 5; range, 3.60 to 5.40). These single peak tumors had a mean CV of 4.34%. A CV of 5.0% was selected as the practical cutoff point for acceptable reproducibility.

Furthermore, we determined the smallest amount of specimen necessary for accurate analysis. We compared cell cycle values of 5 different tumors at 100, 50, and 25 mg. Values obtained at each of these weights were not statistically different from each other (Table 6); however, we noted the highest degree of variability or standard error in the estimates of the fraction of nuclei in Go-G1, and S phase in the 25-mg sample. In addition, the 25-mg sample was difficult to homogenize, and cell yield was sometimes very low. We chose the 50-mg weight as the minimal acceptable weight for accurate analysis and consistency of sample preparation.

### DISCUSSION

If DNA cell cycle and ploidy measurements are to establish their full potential in clinical medicine, a dependable and reproducible analysis system must be developed. To make this technology available to the greatest number of patients, the analysis system must begin with a tissue sample that is easily shipped. Many institutions are already accustomed to shipping frozen
breast tumors to clinical laboratories for their cytosol steroid receptor determinations. This method of shipment has been in use for over a decade and has proven simple and reliable. In light of recent data (1, 2, 4, 5, 7, 8, 10, 11, 27, 28) showing the impact of both the percentage of S phase and ploidy status on risk of relapse and overall survival in the breast cancer patient, the need for assessing these 2 parameters in each case is well recognized. However, analysis of DNA histograms from breast cancer, as well as other solid tumors, has been difficult because of overlapping cell populations having different DNA contents. Most investigators have had to simply omit these difficult cases from their studies.

Using an Epics V flow cytometer (Coulter Electronics) and the EASY software programs supplied with the instrument, we attempted cell cycle analysis and DNA index measurements. Our initial histogram analysis began with the PARA 1 program. Using this program, the S phase is calculated as that percentage of cells remaining after subtraction of the G0-G1 and G2-M gaussians from the total number of events in the histogram. Skewing of the G0-G1 peak in this model can cause overestimation of the S phase, as can an overlapping G2-M peak from a diploid population in the S phase of an aneuploid population. Even by visual inspection, an overestimation of S-phase fractions was observed in many of our aneuploid populations.

We then looked at the PARA 2 analysis (Coulter Electronics), a linear/nonlinear least-squares analysis system first described by Marquardt in 1963 (25) and subsequently modified by Dr. Bruce Bagwell for DNA analysis (23). The PARA 2 system allows the user to modify or edit various parameters and to build models to fit the data. The models are made up of 3 components: gaussians to model G0-G1 and G2-M phases; trapezoids to model S phase; and an exponential to model debris particles. The model fits each of these compartments to the raw data. We initially attempted to model the S phase with a 4-compartment trapezoid, but found that the model had too much flexibility and allowed the trapezoid to bend itself to unwanted peaks. We proceeded to model S phase with three-, two-, and then a one-compartment trapezoid. We found that a single compartment trapezoid, which often appears graphically as a rectangle (Model 1), gave our model the least flexibility in fitting itself to unwanted data, but the best fit by reduced x^2 residuals. In tumors with a contaminating peak in aneuploid S phase, the model allowed the trapezoid to overfit most of the S phase and underfit the unwanted peak. This approach gave us a significantly better estimation of S phase than PARA 1, but still appeared to be an overestimation. To reduce this overestimation, we developed a slightly more complex model (Model 2) with two sets of gaussians that modeled the diploid G0-G1 and G2-M regions and the aneuploid G0-G1 and G2-M regions and two overlapping trapezoids to model the diploid and aneuploid S phases, respectively. This model correctly assigns the diploid G2-M peak to a gaussian and does not contaminate the aneuploid S-phase fitting routine. There are identification and simultaneous processing of the diploid G0-G1, S, and G2-M phases, and the aneuploid G0-G1, S, and G2-M phases. The program searches to find the best gaussian to fit the diploid G2-M, while finding the height of the trapezoid to model the S phase of the aneuploid population. If the G2-M of the diploid were not identified by the model and subsequently processed, the program would try to adjust the height of the rectangle fitting the aneuploid S phase to fit the top of the diploid G2-M peak.

With Model 2, we were able to obtain a more accurate S-phase estimate in most hyperdiploid tumors with overlapping populations. Tetraploid tumors were analyzed using Model 1; however, this model has no way of separating diploid G2-M nuclei from aneuploid G2-G1-phase nuclei. Using these two models, we were able to analyze 78% of our tumors. The small subset of aneuploid tumors unanalyzed by these models had a DI so close to diploid that the G0-G1 peak position would usually be assigned midway between the aneuploid and diploid peaks to minimize the reduced x^2 residuals. To correct this problem we modified Model 1 so the peak positions of the G0-G1 and G2-M gaussians could not deviate from the positions found by the PARA 1 analysis. The modification (Model 3) can identify near-diploid G0-G1 peaks but cannot separate diploid S-phase from aneuploid S phase. Therefore we anticipate underestimation of S phase in some of these tumors due to dilution from diploid S-phase nuclei. Using these three models, we could successfully analyze 88% of all tumors tested for cell cycle.

In order to study the accuracy of S-phase estimation and determine proportion of aneuploid cells that we could confidently assess, we mixed the hyperdiploid breast cancer cell line, T47D, with normal human peripheral blood lymphocytes in various ratios. We obtained a definite aneuploid peak from the 10% mixture of aneuploid cells, the lowest ratio tested, and results from cell cycle analysis using Model 2 were within the 95% confidence limits as assessed by comparison with analysis results from the pure (100%) aneuploid population. We still need to determine the lower limit of sensitivity for detection of aneuploidy and accurate analysis. This requires not only lower mixing ratios, but aneuploid cell populations which provide a range of DNA index measurements, as well as diploid cell populations with different (increased) proportions of G2-M cells contaminating the S phase of the aneuploid population, as well as varying amounts of debris.

Reproducibility of the analysis programs, as well as sample preparation and instrumentation, was assessed by intraassay and interassay studies. A similar amount of variance of cell cycle estimates was observed in both intraassay and interassay studies. Although this variability might be higher than what we would consider optimum, it allowed us to determine a range of normal values expected for any one tumor. We are presently working on improving this variability. Interassay and intraassay experiments, although performed on two different groups of tumors, also helped to provide a reference for quality control of sample preparation and instrumentation. We were able to establish certain criteria for definition of acceptable histogram analysis. This includes: (a) CV of the G0-G1 cutoff of 5.0 to minimize incorrect interpretation of a wide G0-G1 peak; (b) a percentage of aneuploidy cutoff of 10%, defined as a minimum of 10% of the total sample events that must be found in the G0-G1 peak of a population to define it as aneuploid; and (c) a minimum sample weight of 50 mg of frozen pulverized tumor tissue for accurate cell cycle analysis. Additional criteria are being established for interpretation of histograms, which should facilitate comparisons of histograms and data analysis from different laboratories, independently of the sample preparation or modeling algorithms, thereby reducing subjectivity of interpretation.

The frozen powder technique and the models described here are presently being applied to a larger cohort of samples. Improvements in technique, optics, and analysis programs should soon increase the total percentage of solid tumor cases that can be assessed for the percentage of S phase and ploidy. These results should help us to better understand the clinical implications of these measurements in the prognosis and treatment of the breast cancer patient.
MODELS OF BREAST CANCER PROLIFERATION

ACKNOWLEDGMENTS

The authors gratefully acknowledge the expert advice and helpful discussions from Dr. John Meyer, Dr. Robert McDivitt, and Dr. Kenneth Stone of St. Louis, MO, and Dr. Bartholomew Barlogie and Dr. Todd Johnson of Houston, TX.

REFERENCES

Evaluation of a Modeling System for S-Phase Estimation in Breast Cancer by Flow Cytometry


Updated version  Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/47/20/5294

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.