Title

FLOW CYTOMETRIC ANALYSIS OF MITOTIC CYCLE PERTURBATION BY CHEMICAL CARCINOGENS IN CULTURED EPITHELIAL CELLS

Permalink

Author

Pearlman, A.L.

Publication Date

1978-08-01

Replacement

LBL-8169 c.3

FLOW CYTOMETRIC ANALYSIS OF MITOTIC CYCLE PERTURBATION BY CHEMICAL CARCINOGENS IN CULTURED EPITHELIAL CELLS

Andrew Leonard Pearlman
(Ph. D. thesis)

August 1978

Prepared for the U. S. Department of Energy
under Contract W-7405-ENG-48

LBL-8169 c.3

Flow Cytometric Analysis of Mitotic Cycle Perturbation by Chemical Carcinogens in Cultured Epithelial Cells

Doctoral Dissertation of Andrew Leonard Pearlman

ABSTRACT

A system for kinetic analysis of mitotic cycle perturbation by various agents, described below, was developed and applied to the study of the mitotic cycle effects and dependency of the chemical carcinogen benzo[a]pyrene-diol-epoxide, "DE", upon a mouse liver epithelial cell line, NMu-Li.

The study suggests that the targets of DE action are not confined to DNA alone, but may include cytoplasmic structures as well. DE was found to affect cells located in virtually every phase of the mitotic cycle, with cells that were actively synthesizing DNA showing the strongest response. However, the resulting perturbations were not confined to S-phase alone. DE slowed traversal through S-phase by about 40% regardless of the cycle phase of the cells exposed to it, and slowed traversal through G_2M by about 50%. When added to G_1 cells, DE delayed recruitment of apparently quiescent (G_0) cells by 2 hours, and reduced the synchrony of the cohort of cells recruited into active proliferation.

The kinetic analysis system consists of four elements: tissue culture methods for propagating and harvesting cell populations; an elutriation centrifugation system for bulk synchronization of cells in various phases of the mitotic cycle; a flow cytometer (FCM), coupled with appropriate staining protocols, to enable rapid analysis of the DNA distribtution of any given cell population; data reduction and analysis methods for extracting information from the DNA histograms produced by the FCM. The elements of the system are discussed.

Detailed methods are presented for bulk cell cycle synchronization of NMuLi cells by elutriation centrifugation. When run under the conditions described, G_2M cells sediment at about 2.5 times the rate of G_1 cells, with most S-phase cells sedimenting at about 1.6 to 2.0 times the G_1 rate.

Optimal methods for sample preparation and fluorescent staining of the NMuLi cells are presented, along with an analysis of the FCM system used in this work.

A mathematical analysis of DNA histograms obtained by FCM is presented, in which the histograms are shown to be well modeled as a convolution of the true DNA distribution with a single overall spread function, provided that their abscissas have been transformed to represent the logarithm of fluorescent intensity. The analysis leads to the detailed implementation of a new modeling approach, in which a given histogram is fit using the spread function extracted

from the G_1 peak of a related reference histogram. The central algorithms in the approach are presented in detail, along with a discussion of model stability with respect to various sources of error. Results are shown for various DNA histograms.

The above modeling approach is applied to the estimation of cell cycle kinetic parameters from time series of DNA histograms, and methods for the reduction and interpretation of such series are suggested.

Signed, D A Glaser 9/5/78

Committee Chairperson Date

Flow Cytometric Analysis of Mitotic Cycle Perturbation by Chemical Carcinogens in Cultured Epithelial Cells

by

Andrew Leonard Pearlman

TABLE OF CONTENTS

FOREWORD

The structure of this dissertation reflects the central theme of my work reported in it: the assembly of several new methodologies and approaches in a kinetic analysis system, applied to the study of the interactions of a chemical carcinogen with the mitotic cycle of epithelial cells grown in culture. Hence, Chapters I through IV describe and analyse the various components of the system. Chapter V presents a brief review of pertinent recent work with the chemical carcinogen studied here, and serves as background to the principal experiments of this dissertation, which are reported and discussed in Chapter VI. The Appendices present detailed analyses of the optics of the flow cytometry system used in this study, and of my attempts at mathematical modeling of cell cycle kinetics using time series of DNA histograms.

ACKNOWLEDGEMENTS

Since the various sections of this dissertation were written so as to stand somewhat independently, each is equipped with its own set of acknowledgements for assistance generously accorded me in connection with the work described. However, there are still many otherwise "unsung heroes" without whom this thesis would have been ever so more difficult, if not altogether impossible.

A warm and hearty THANKS, JIM! goes to Dr. James Bartholomew, my de facto research advisor, colleague and dear friend, through whom I have learned many crucial lessons not only about the ways of a tissue culture laboratory, but about working with other people in a variety of settings. Dr. Bartholomew is a master of calmness and clarity, some of which has hopefully rubbed off, and it is in large measure due to his steady guidance that this dissertation has seen its way to completion. It is sad irony that, for bureacratic procedural reasons, Dr. Bartholomew's signature could not grace the title page of this document, for it is he more than any other who knew the intricacies of the work described. I nonetheless wish to thank him both formally and personally for his invaluable help and support, and I wish him the very best in his future work.

I should also like to express my deep gratitude to Dr. Melvin Klein, who brought me to the Laboratory of Chemical Biodynamics (LCB), pointed me in the direction of Dr. Bartholomew, kept in close touch with my work as it progressed, and lent me both support and useful criticism throughout. It's been a pleasure and an honor to work with such a MENSCH of a scientist and human being; thanks, Mel.

The completion of my graduate career is much indebted to Dr. Edwin Alpen, my academic advisor, who kept me on track and on schedule so that my efforts could be brought to a fruitful culmination in the form of a doctoral dissertation. In the often nebulous world of an inter-disciplinary graduate curriculum, the strong and clear guidance of the sort I received from Dr. Alpen is both sorely needed and deeply appreciated, and I thank him for having provided it to me.

Within LCB, the number of individuals to whom I owe a debt of thanks is far too large to list here, but I should like to single out the following people who have become my CHAVERIM and have made life in the laboratory a pleasure: Mr. Bruce Navsky, a research assistant at LCB, was my right-hand man (and often the other hand, too!) during most of the work of this dissertation, lending critical support in the computational aspects of the work, as well as invaluable assistance in most of the actual experiments. To Bruce, who has also been my close friend, goes a special note of thanks for all this and more, and I wish him every success in his new career in computer science. Sue (Dr. Susan) Hawkes, has been a dear friend and confidant, helping me to acclimate to the laboratory human environment, and brightening up the place with her natural cheerfulness. She has assisted me in numerous cell biology, biochemistry, and text editing facets of my work, and it has been a pleasure to help her with the flow cytometry and mathematical aspects of hers. Mr. Hisao Yokota deserves a special "thank you" for having taught me most of what I know about tissue culture technique, and for having provided a sharp eye and well-tuned ears for critical evaluation and counseling on questions relating to the cell biology aspects of this dissertation. Dr. Gordon Parry, a visiting post-doctoral fellow from Britain, has been a warm and considerate colleague, observer and friend, lending frequent help in the planning and interpretation of experiments, and in the critical review of papers written by me. To all of you very special people, I extend my warmest appreciation and best wishes.

Finally, there were people outside of my laboratory setting who contributed vitally to this achievement: Dr. Arthur Zingher, who also completed his dissertation during this period, afforded me enormous assistance in areas ranging from his insights in numerical analysis and statistics to the UNIX word processing system, on which this document has been composed, and to which Arthur introduced me. His help has often far exceeded the call of friendship, but it has always been deeply appreciated by me, and I wish to express my thanks to one of the most imagineative and creative scientists I've met; my good friend, Arthur Zingher. To Dr. Lawrence Scherr, fellow former biophysics graduate student working in flow cytometry applications, and YEDID NEFESH, I am grateful for his strong personal support and encouragement throughout my work on this dissertation. I wish him DERECH TSLECHA and much happiness in his future work.

There have been many moments during these past five years when it was anything but evident that my work would culminate in a thesis; times of great personal doubt and not inconsiderable fear. It is to my wonderfully supportive wife, Debbie, ESHET CHAYIL, that I owe the deepest gratitude for helping me through those times, and for being with me in so many ways at all times.

To all those whose names have gone unmentioned either here, or in the separate sectinons of acknowledgement, and to whom I owe a debt of gratitude, I extend my apologies for not having included you here specifically, but hope you will accept my warmest thanks for your help and support during my graduate career.

DEDICATION

In June of 1972, I dedicated my then Senior Thesis to the memory of the late Prof. Aharon Katzir-Katchalsky, beloved member of the faculty of the Department of Medical Physics and Biophysics at U.C. Berkeley, and one of the 27 victims who were brutally cut down by pro-Arab terrorists at Israel's Lod Airport on May 30, 1972. Then, as now, the dedication was not to the memory of the way Prof. Katzir died, but most decidedly to the contrary: to honor and remember the way he lived, and the splendid example he provided for all; scientists or not.

Aharon Katzir was a whole human being, a man of many dimensions, a modern day Renaissance Man. My life has been greatly enriched by attempting to conduct my affairs in accordance with the example of Prof. Katzir, and with a living philosophy best summarized by the great sage, Rabbi Hillel:

> "If I am not for myself, who will be?
>
> But if I am only for myself, what am I?
>
> And if not now, when?"

It is to the memory and cause of great men such as these that I dedicate not only this dissertation, but also the conduct of my life.

Berkeley, California, 4 September, 1978

CHAPTER I

The Flow Cytometry System at The Laboratory of Chemical Biodynamics, University of California at Berkeley

INTRODUCTION

During the past several years, numerous high speed cell analysis systems (1-7) for measurement of physical and biological properties of single cells have been developed, providing new and useful techniques for performing a wide variety of biological experiments. Such systems have been used, for example, in studies relating DNA content per cell to chromosome number (8,9), quantitation of cell surface binding of plant lectins such as concanavalin A (10), determination of the effects of chemo-therapeutic drugs and other agents upon cell cycle traverse (11), immunofluorescent detection of antigen-binding cells (12), screening of gynecologic cytology specimens (13), and myriad other applications.

The common feature of these systems is their use of monodisperse cellular suspensions, which are made to flow through a defined region of observation, where measurements are made on individual cells in succession. Hence, the generic name by which these devices have come to be known is "flow cytometer", or FCM, and the field of study relating to their development and use is known as "flow cytometry."

Typical parameters measured by FCM are cell size, DNA content, RNA content, specific surface antigen content, etc., and in general, FCM systems trade low resolution on individual cells for large sample size and hence high statistical precision in characterizing whole populations of cells. An FCM thus provides complementary information to that obtained by more conventional methods such as light microscopy, in which a far smaller number of cells may be studied in much greater detail.

A given sample population is FCM analysed by first preparing a monocellular preparation of the sample (often a difficult step), staining it with appropriate fluorogenic or absorbing molecular labels, and streaming the cells, one by one, through an excitation light source (generally a laser), where their absorption, emitted fluorescence, scattering and possibly other output are detected and measured at rates of 1-10,000 cells per second. The measured quantities usually represent average values for each whole cell, with little subcellular detail, due to the optical limitations of the excitation light source, whose dimensions generally cannot be reduced below 5 to 10 μm (4), or about one cell diameter.

So many variations on this theme have emerged in the past few years that even the rather general definition given above is inadequate to cover all the important kinds of measurements made or in development. However, the focus in this chapter will be on the FCM at the Laboratory of Chemical Biodynamics (LCB), and its use in obtaining DNA histograms of cell populations, in particular. The FCM device at LCB is illustrated graphically in Fig. 1, and schematically in Fig. 2.

The principal goal of the FCM at LCB, as regards DNA histograms, has been to enable the rapid estimation of cell cycle distributions of cell populations under study, and use them in extracting descriptive kinetic information about the behavior of the population under various conditions and treatments of interest. Prior to the advent of FCM, the only means for obtaining information of this nature was autoradiography, which still serves as the principal benchmark against which FCM results are tested.

Autoradiography is the method whereby growing populations are exposed to radioactively labeled substrate molecules, in this case tritiated thymidine, which are taken up by the cells and used as would non-labeled molecules of the same type. Thus, in theory, only cells that are actively synthesizing DNA take up the labeled thymidine and incorporate it into the new DNA copy being replicated. The cells, which are usually grown in monolayers on transparent plastic or glass, are usually exposed to the label for a short "pulse"(15-30 min.), then fixed and covered with a thin layer of special photographic emulsion, and left in the dark to expose the emulsion for several days (typically 7 to 21 days). Upon developing, those portions of the emulsion situated above labeled nuclei show a larger density of silver grains (due to tritium emissions) than elsewhere. The fraction of cells that were actively synthesizing DNA during the time that the cells were exposed to the labeled tritium can be estimated by determining the fraction of labeled nuclei (FLN). This is usually done by establishing a grain count threshold for classifying any given nucleus as "labeled," and examining a large number of cells to approximate the FLN.

There are several drawbacks to autoradiography which have motivated our choice of FCM as an alternative. First, the time lag between a given experiment and the FLN estimates is often two weeks or more, due to the long exposure time needed, which precludes the use of the FLN information as "feedback" during the course of most experiments. Second, the FLN estimate is completely dependent upon the grain count threshold and exposure time used, both of which are arbitrary in value, thus bringing the reliability of the estimate obtained into question (14). Third, the procedure of counting the labeled nuclei usually involves extensive visual examination of the material by light microscope, which is both tedious and given to human error (especially after the first few thousand cells counted). When counting by eye, it is very difficult to consistently apply the grain count threshold, often leading to mis-classification of cells, especially if there is an appreciable background of grains. Fourth, the method detects only those cells that are actively sythesizing DNA, so that G_1 and G_2 are not distinguished from one another, and cells blocked during synthesis (an effect of interest caused by many agents) are not recognized as being in S-phase. Finally, the presence of grains over the nucleus of a given cell is not a completely unambiguous indication that the cell was actually synthesizing DNA at the time of tritiated thymidine addition.

FCM analysis of the cell cycle distribution of a given cell population makes use of cytochemical methods for fluorescently staining the DNA in a stoichiometric manner, such that the number of fluorescent dye molecules bound per 100 DNA bases is approximately constant. Hence, measurement of the amount of bound dye in a given cell yields an estimate of the relative amount of DNA contained in the cell, which should double as a cell traverses the cycle from G_1 to G_2.

In practice, the measured quantity is relative fluorescent intensity per cell, which is digitized in a set of discrete values (typically 256). As each cell is analysed, the count in the appropriate discrete channel is incremented by one, and in this manner, a historam describing the whole population is built up. FCM can thus estimate the fraction of cells in G_1, S, and G_2M (FCM cannot distinguish G_2 from mitosis, so G_2M is the combination of the two), with high accuracy (typically about 3-6% in our system). Because of its small measurement error and large statistical sample size, FCM analysis can also determine the detailed distribution of cells within S-phase, including early-, mid-, and late-S. The whole FCM analysis procedure for a given cell population can be as short as 20 minutes (for the mithramycin staining protocol of Crissman and Tobey (18)), or as long as 14 hours (for the acriflavin-Feulgen protocol). In this thesis, a propidium iodide staining protocol (described below and in chapters II and VI) was employed, with a resultant analysis lag time of 2.5 to 3 hrs for a given cell population.

Cell cycle kinetic information about a given cell population is obtained by FCM analysis of the population at sequential timepoints. The resulting set of histograms is stored on magnetic tape and analysed by various modeling routines to extract the information sought. The overall flow cytometry system at LCB is thus seen to consist of sample preparation and staining methods, an FCM (which lacks the sorting capability present on other systems), hardware and software for long term storage of the histograms, and associated programs for modeling and other analyses of the data.

Resolution of the FCM is measured in terms of the sharpness of the histogram peak corresponding to G_1 cells, which, theoretically, all posses the same DNA content. The coefficient of variation (CV) of the G_1 peak, defined as the ratio of the standard deviation of the peak (if it were a Gaussian distribution) to its mean channel, is approximated by the ratio of full-width-at-half-maximum to modal channel number, divided by 2.35. Acceptable CV's are typically 2% - 6%.

GENERAL DESCRIPTION OF OUR FCM: OVERVIEW

Sample preparation and staining are described below. The flow cell, detection electro-optics and lenses for the excitation beam of our system were all obtained from a former employee of the Los Alamos Scientific Laboratory (LASL), who constructs these items to the specifications used at LASL as of 1973. Thus, the heart of our system is equivalent to a LASL Flow System II, described by Holm and Cram (4).

The operational principles of the sample feed and observation cell (henceforth termed the "flow cell") are illustrated in Fig. 3. Briefly, the flow cell operates by generating a concentric laminar sheath of water around the cylindrical sample flow stream, and accelerating both by forcing them through a 250 μmeter nozzle into the region of illumination. The stream flows vertically, forced against gravity, so as to maximally separate adjacent cells. The sample and sheath are drawn through the flow cell by two forces: gravity, and air pressure differential between inlet and outlet sides. The differential pressure between the disposal vessel (which is partially evacuated by means of a peristaltic pump), and the two respective reservoirs (which are exposed to the atmosphere), determines the absolute final velocity with which the sample traverses the region of illumination. The relative heights of the flow cell to the respective reservoirs also drive the two flows, and for fixed sheath reservoir-to-flow-cell height difference, the concentration of sample fluid in the final stream through the nozzle is determined by the sample reservoir-to-flow-cell height, seen in Fig. 3. Hence, counting rate, for a sample of a given density of cells per ml, is varied by adjusting the height of the sample reservoir relative to the flow cell.

For height differences of 25 cm for the sheath reservoir-to-flow-cell, and [10-15] cm for the sample, a differential pressure of 2 PSI results in a final flow velocity of 4-5 meters/second through the region of observation. At these values, sample is drawn into the flow cell at rates of about 1 ml/min, but is easily varied by adjusting the sample height relative to the flow cell.

The sample height also modulates the minimum width achieved by the flow stream as it traverses the excitation beam. When drawn through at minimum rate, stream diameters of 8 to 10 μmeters are obtained, while raising the sample to 40cm above the flow cell increases the diameter to as much as 60 μmeters. As discussed below, the resolution of the FCM declines markedly with increased stream diameter in the excitation beam, thus it is preferable to use highly concentrated samples (1 to 10 milllion cells/ml) fed into the flow cell at rates of about 0.1 ml/min, when possible. This sample feed combination yields a counting rate of 1.5 to 3 KHz, at which the 4 to 8 μsecond long pulses from succeeding cells rarely coincide, and has been found to yield better resolution than higher rates, and as good as that obtained at lower rates.

As a given cell enters the region of illumination, it traverses a relatively flat excitation beam, of eliptical cross section, formed by two crossed cylindrical lenses of focal lengths 3cm and 23cm, respectively. The 3cm focal length lens focuses in the vertical direction, in the plane containing the sample stream axis and the laser beam axis of propagation, and we term it the "vertical lens." As seen in Fig. 4, the sample stream is directed through the focal plane of the vertical lens, which focuses the 2mm diameter laser beam from a Spectra Physics Model 171 Argon Ion laser to a minimum width of about 18 μmeters. The focal point of the horizontal lens is displaced slightly from the intercept with the sample stream, so that the horizontal beam width is about 150 micrometers. The crossed lens combination yields an eliptical excitation region for the stream, of semi-axes 9 and 75 μmeters in the vertical (y) and horizontal (x) directions, respectively.

The argon ion laser is capable of 0.5 watt or greater output at several lines of interest, the choice of which is dictated by the choice of dye used in staining the DNA of the sample cells. We have used mithramycin, Chromomycin-A, and propidium iodide stains at various times, the optimal lines for which are the 458 nm, 458 nm, and the 488 nm wavelengths, respectively. For propidium iodide, for example, we routinely use a 2 watt excitation beam at 488 nm (though we have found a 3 watt 514 nm line does just as well), which causes the stained cells to emit a brilliant red fluorescence upon traversing the beam.

A portion of the emitted fluorescence is collected via a lens system, seen in Fig. 2, along with scattered excitation beam. The latter is eliminated as the light passes through a barrier filter chosen to absorb at the excitation wavelength, and the remainder encounters a dichroic filter/mirror that reflects wavelengths above 540 nm, while transmitting shorter wavelengths. The two-color components of the light are parallel processed, each being imaged onto the cathodes of matched photomultiplier tubes, the output from which is pulse-shaped and amplified by separate electonics. The two color capability of the LCB FCM system has been employed in correlating a second fluorescently tagged parameter with DNA content, and of course has a great many other possible applications, but will not be further discussed in this paper as only one channel at a time has been used in this thesis work.

The original photomultiplier assembly used a preamplifier with an integrating time constant of 50 μseconds which, compared with the 4 to 8 μsecond duration of the individual cell pulses, is far too long to allow any detailed information to be gained from the pulse shapes produced by it. Hence, an additional alternative preamplifier was designed and built by Dr. Bojan Turko* of LBL, with an integrating time constant in the 100 nsec range ("non-integrating amplifier"), to enable real-time measurement of the detailed pulse shapes generated as the cells traverse the excitation beam. The main amplification is done by a Research Amplifier (Northern Scientific), which allows for separate integrating and differentiating time constants from 0 to 10 μseconds, and we found that the best resolution was achieved when the non-integrating amplifier was used in conjunction with the 10 microsecond integrating and differentiating time constants.

The shaped, amplified pulses are digitized by an ADC converter (Northern Scientific, Chicago, USA), and assigned to one of 256 bins of a Northern pulse height analyser (PHA). The PHA has a 4096 channel memory, enabling the temporary storage of up to 16 256-channel histograms, which can be transferred to permanent storage on tape via an interface to a Sigma-2 computer. A program for the documenting and transfer of up to 16 histograms was developed by Sheldon Wong* of LCB, incorporating the values of vital parameters (laser wavelength, laser power, vacuum setting, etc.), as well as general experiment commentary and comments on each histogram. The total time necessary to encode this highly useful information for a typical set of 16 histograms is about 10 minutes, which can generally be done while samples are being run.

SAMPLE PREPARATION AND STAINING PROCEDURES

Details of these methods are presented in chapters II and VI, so they will be only briefly presented here, with some theoretical background.

All experiments done in connection with this thesis used cells grown in monolayer tissue culture. Cells are removed from the surface of either plastic culture dishes or roller bottles by brief trypsinization, followed by washing in a trypsin-neutralizing solution. For propidium iodide staining, which was the method of choice for this work, the resulting cell suspension was spun down, and resuspended in 25% ethanol with 15mM $MgCl_2$. Non-specific ribonuclease is added to the fixative to 50 μgrams/ml final concentration, and the suspension held at $37^{\circ}C$. for 60 min. The ribonuclease is nescessary to eliminate any double stranded RNA that might be present, as it competes with DNA in binding propidium iodide and thus can generate a false signal. The ethanol and divalent cations increase membrane permeability so that the enzyme can pass freely into the dead cells. No ribonuclease treatment is used with either mithramycin or chromomycin-A, since they show much stronger specificity for DNA over other double-stranded nucleic acids. However, we found that the propidium iodide (P.I.) protocol yields 50% better resolution for several cell types (WI-38, NMuLi, Balb 3T3, etc.) than the other stains, so we elected to endure the extra preparatory step necessary to obtain P.I. stained cells.

All three stains, like many others, show greatly enhanced quantum efficiencies when bound to the target molecules as compared to that in free solution. Since these fluorochromes also exhibit strong affinity and specificity for DNA (or double stranded nucleic acids in the case of P.I.), virtually all the emitted fluorescence comes from stain molecules bound to target. However, as Van Dilla, et al (15) have shown, cell geometry can play a dominant role in determining the emitted intensity in the direction of the detection optics, especially for highly non-spherical cells such as sperm.

Microscopic observation of the suspended cells obtained from tissue culture for FCM analysis at LCB shows them to be generally spherical in shape, displaying little elipticity or directionality. Hence, it would seem reasonable to assume that shape would not be as serious a matter for concern as in the case of sperm cells, though Kerker (16) has shown that cytoskeletal structure can seriously affect the emitted fluorescent intensity from even perfectly spherical model cells. It appears that such geometric considerations must somehow cancel each other out, as the histograms obtained from well characterized cell populations (quiescent, exponentially growing, chemically or otherwise synchronized populations, etc.) all conform rather well to what one would predict from the basic model of the cell cycle and previous knowledge about the populations studied. Therefore, the role of cellular and subcellular shape in determining the final detected fluorescent intensity from cells shall not be further considered in this thesis, but assumed to be a small, random perturbation on the true fluorescent signal. This, along with other sources of error, comprises the FCM spread process which gives rise to smoothed histogram shapes from presumably spike-like initial DNA distributions for G_1 and G_2M cells. (discussed at length in chapter III)

FCM EXCITATION OPTICAL SYSTEM

Cells with identical DNA content should yield identical detected fluorescent intensity. However, as mentioned above, the final detected intensities are scattered about the (true) mean intensity due to errors in sample preparation, staining and the FCM measurement itself. The detection electro-optics have been tested and shown to be superbly linear and stable, and are unlikely to produce significantly different outputs for identical fluorescent emissions detected. Therefore, it may be presumed that the only significant contibution to overall error from the FCM device arises from the flow stream and laser beam configurations.

A thorough analysis of the properties of the excitation optics of the LCB FCM is presented in Apendix A, from which several important points are summarized here. The principal assumption was that for cells containing identical amounts and disposition of fluorophore (i.e., identical fluorescent objects) the intensity of emitted fluorescence is proportional to the local excitation intensity. Hence, variations in position within the region of excitation, corresponding to variations in local beam intensity, could give rise to variations in fluorescence emission. The aim of the analysis was to identify the sources of possible signal variation due to the illumination and flow configurations, and to estimate their relative size compared to the main signal.

The illuminating laser beam is presumed to be in the TEM_{00} mode, and to have a Gaussian intensity cross sectional profile, of spot size 2mm (defined in Apendix A). The wavelength is taken to be 488 nm, and the intensity distribution in the region of excitation is assumed to be entirely determined by the laser and the two crossed cylindrical lenses, neglecting any perturbation by the quartz window and water sheath that must be traversed by the illuminating beam prior to reaching the cell stream. Finally, the analysis is performed for the case of perfect focusing, in which the focal planes of the two lenses coincide with the sample stream axis. In practice, only the vertical lens is thus focused, and the horizontal lens is deliberately set slightly off focus, in order to broaden the intensity distribution in the horizontal direction and thus reduce sensitivity to the positional fluctuations of the flow stream.

The intensity along the laser axis, z, is shown to vary as a Lorentzian function of position about the focal

$$I(z) = \frac{I_0}{1 + \left|\frac{z}{110}\right|^2}$$

where z is the deviation in micrometers from stream center. Since deviations in the z-component of cell position within the stream are typically on the order of 5 to 10 μmeters, the maximum intensity variation suffered due to this source is 1%. For a stream radius of 5 μmeters, this error would be 0.2%

With the above assumptions, the minimum vertical and horizontal semi-axes of the elipsoidal Gaussian intensity distribution in the excitation region are found to be about 4 and 32 μmeters, respectively. These dimensions are from the beam axis to the points where the intensity falls to e^{-2} times the maximum value, and should be about half the size of the detectable width. In practice, the total visible beam diameter (in the vertical direction) has been measured by telescoping microscope, equipped with calibrated reticule, as 17 to 18 μmeters, which compares well with the calculated value. Thus, the smallest beam dimension obtainable under these conditions is about 8 to 10 μmeters, or roughly the diameter of a large cell. Hence, the spatial resolution of the FCM system cannot achieve much better than average values for a whole cell, with subcellular detail being beyond the detection capabilities of the system.

Once in the beam, the only components of cell position variation that contribute to emitted fluorescence variations are those along the axis of laser propagation (z), and in the horizontal direction (x). Variations in the z-direction were treated above. The intensity distribution in the y-direction is Gaussian with a standard deviation of 16 μmeters for the perfectly focused case. The actual full beam width in the x-direction was measured on the LCB FCM, and shown to be about 150 μmeters, so we estimate the actual e^{-2} full width to be somewhat smaller than this, or 100 μmeters. Using this value, which corresponds to a standard deviation of 50 micrometers, we have superimposed various stream diameters upon the y-axis Gaussian intensity profile, as is shown in Fig. 5. From this plot, we see that the intensity discrepancy between the boundary and the axis of the flow stream varies with stream radius "r" as:

$$\text{Intensity Difference} = 1 - G(0,50;r)$$

where G(0,50;r) represents a Gaussian distribution with 0 mean and standard deviation 50 μmeters, evaluated r μmeters away from the mean. (G = 1 when r = 0) Thus, for some typical stream radii the variance over x is:

Radius (μmeters)	% variation
4 (best obtainable)	0.5
10 (typical)	2.0

15 (poor)	4.5

Hence, we conclude that under minimum stream radius conditions, the variation attributable to illumination optics, considering both the x- and z- components of flow stream positional uncertainty, is about 1.5%. Thus, for a histogram recorded with optimum stream and optics configurations and displaying a CV of 2.5%, probably only about 1% of that is due to sample preparation and staining variance.

RESULTS AND SUGGESTIONS FOR IMPROVEMENT

Typical histograms for the epithelial cell line used in this thesis are shown in Fig. 6. The performance of the FCM was generally adequate for the purposes of this thesis, but improvements are needed in several areas. There is always some uncertainty as to the composition of the apparent G_2M peak, since the FCM device, as described above, cannot distinguish a doublet of G_1 cells from a single G_2M cell. This situation can be further exacerbated by excessively high counting rates, since succeeding pulses from closely packed cells can overlap, yielding a summation of their individual signals. Because of overlap summation, high counting rates can also degrade the overall CV and increase any shoulder on the left side of the G_1 peak. These problems can be minimized by using a standard sample cell concentration, whenever possible, and a standard sample height above the flow cell, to obtain the minimum flow stream radius. However, to discriminate between doublets and singlets, additional information must be detected; fluorescent pulse shape and scattered signal are two possible parameters for the delineation.

A pulse shape discriminator has been developed by Dr. Bojan Turko, of Lawrence Berkeley Laboratory, and is currently being tested on sample cell populations. The method is illustrated in Fig. 7, and works as follows: A single cell should generate a unimodal pulse as it traverses the excitation beam, and its duration can be characterized by the time interval over which the pulse height exceeds a fixed percent of its maximum value. In this manner, the absolute amplitude of the pulse is not taken into consideration. The hydrodynamics of the flow cell are such that elongated objects tend to orient along the axis of flow, so that a cell doublet would likely be aligned with the flow axis. Thus, one would expect a bimodal pulse shape from a doublet, folded with the intensity distribution in the x-direction of the exciting beam, as shown in Fig. 7. The corresponding interval during which the bimodal pulse exceeds the same fixed percentage of its maximum value should be significantly longer than that for any conceivable single cell of the same line, and is thus a reasonable parameter for distinguishing singlets from doublets.

Another common problem is the occurence of gain shifts during the course of a given experiment, due to such causes as positional variation of the sample stream or laser power fluctuations, for example. Such machine-related shifts mask any actual change in the fluorescence intensity per cell among the different cell samples being analysed. Since altered fluoresecence emission from cells stained in a standard manner may reflect differences in binding, quenching or other phenomena related to the microenvironment of the fluorophore, potentially valuable biological information is lost if the true gain is not known. Furthermore, histograms lacking a prominent G_1 or G_2M peak are respective peak means of previous or subsequent histograms can be "adopted," and this is reasonable only if the gain has remained stable.

The use of a fluorescent standard together with the sample cells could greatly alleviate both problems. By selecting a standard whose signal is clearly distinguishable from that of the sample, while still appearing in the same histogram, one could detect any gain shift occuring between samples by the position of the histogram peak corresponding to the standard. After recording the histogram, it is a relatively simple matter to simulate a gain-shift by computer, so that the standard peaks are aligned. The same effect could be achieved in real time by use of a digital gain stabilizing cirurcuit, gated on the standard signal. Locking the laser power, flow cell alignment and electronic gain setting at fixed values is a poor substitute for such an internal standard, since the occurence of bubbles and other irregularities invariably requires some form of rescue operation that later involves re-alignment of the system. Fluorescent polystyrene spheres of 1.88 μmeter diameter (DuPont) have yielded a CV of 1.8% on the LCB FCM, and though they do not appear on the histogram with the cells (they flouresce in the green, while P.I. fluoresces red) they can be used to align the system prior to running the next sample. Some authors use stained chicken erithrocytes as a standard, which if stained in parallel with the cell samples, provide an excellent control against which to detect real changes in fluorescent output from the cells (17).

FIGURES

1. Graphic illustration of FCM flow and optics at LCB

2. Schematic diagram of FCM device at LCB Device can simultaneously measure forward scattering and emitted fluorescent intensity at two different wavelengths. A lens observes the cells at right angles to the direction of the excitation laser beam, and the collected light is first passed through a filter to

remove scattered laser light (barrier filter), then split into long and short wavelength components by a dichroic filter. The two components are parallel processed by photomultplier tube (PMT), preamplification and amplification electronics, leading into a multiparameter signal processing unit.

3. Operation of the flow cell. Sample fluid is fed into the central stream of the concentric flow configuration, with distilled water from the sheath water reservoir fed into the surrounding sheath about the sample stream. The sample and sheath streams merge at the indicated flow junction, and the two streams are gravity-driven by the heights of their respective reservoirs above the junction.

4. Excitation optical configuration. Argon-ion laser beam, propagating in the positive Z-direction, encounters cylindrical lens L_1, which focuses in the horizontal or X-direction, at point Z_1. The refracted beam continues to Z_2, where it enters cylindrical lens L_2, which focuses in the vertical or Y-direction. The resulting focal region has elipsoidal equi-intensity scontours, and is traversed by the flow stream at $\overline{Z}_c$.

5. Gaussian radiant intensity profile of the excitation laser beam, in the horizontal (x) direction, with flow stream diameter superimposed upon it. Gaussian

standard deviation based on assumption of perfect focusing: in practice, deliberate defocusing widens the Gaussian to a standard deviation 2 or 3 times this figure, thus reducing errors due to large stream diameters. Details discussed in text.

6. Typical histograms of the cell population used in this study.

7. Constant fraction single-cell versus doublet discriminator. As a single cell traverses beam, a unimodal fluorescence pulse is emitted, whose intensity exceeds a given constant fraction of maximum for a characteristic time $\underline{\triangle}t_1$. Doublets are aligned with the direction of flow, and hence generate a bimodal pulse, whose corresponding characterstic time, $\underline{\triangle}t_2$ is longer, enabling it to be discriminated from the single cell shape.

REFERENCES

1. Kamensky, L.A., Melamed, M.R., and Derman, H. Spectrophotometer: new instrument for ultrarapid cell analysis. Science 150:630, 1965.

2. Van Dilla, M.A., Trujillo, T..T., Mullaney, P.F., and Coulter, J.F. Cell microfluorometry: a method for rapid fluorescence measurement. Science 163:1213,

1969.

3. Bonner, W.A., Hulett, H.R., Sweet, R.G., and Herzenberg, L.A. Fluorescence activated cell sorting. Rev Sci Instrum 43:404, 1972.

4. Holm, D.M., and Cram, L.S. An Improved Flow Microfluorometer for Rapid Measurement of Cell Fluorescence. Exp Cell Res 80:105-110 (1973)

5. Mullaney, P.F., Steinkamp, J.A., Crissman, H.A., and Holm, D.M. Laser flow microfluorometers for rapid analysis and cell sorting of individual cells. In Laser Applications in Medicine and Biology. Vol.II, M.L. Wolbarsht, editor. Plenum, New York, 1974.

6. Epstein, M., Norman, A., Pinkel, D., and Udkoff, R. Flow System Fluorescence Polarization Measurements on Fluorescein Diacetate- Stained EL4 Cells. J Histochem Cytochem 25(7):821, 1977.

7. Shapiro, H.M., Shildkraut, E.R., Curbelo, R., Turner, R.B., Webb, R.H., Brown, D.C., and Block, M.J. Cytomat-R: A computer-controlled multiple laser source multiparameter flow cytophotometer system. J Histochem Cytochem 25(7):836, 1977.

8. Kraemer, P.M, Petersen, D.F., and Van Dilla, M.A. DNA constancy in heteroploidy and the stem-line theory of tumors. Science 174:714, 1971.

9. Kraemer, P.M., Deaven, L.L., Crissman, H.A., and Van Dilla, M.A. DNA constancy despite variability in chromosome number. In Advances in Cell and Molecular Biology, Vol.II. E.J. DuPraw, editor. Academic Press, Inc., New York, p. 47. 1972.

10. Kramer, P.M., Tobey, R.A., and Van Dilla, M.A. Flow microfluorometric studies of lectin binding to mammalian cells. J Cell Physiol 81:305, 1973.

11. Tobey, R.A., and Crissman, H.A. Use of flow microfluorometry in detailed analysis of effects of chemical agents on cell cycle progression. Cancer Res 32:2726, 1972.

12. Julius, M.A., Masuda, T., and Herzenberg, L.A. Demonstration that antigen-binding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter. Proc Natl Acad Sci USA 69:1934, 1972.

13. Fowlkes, B.J., Herman, C.J., and Cassidy, M. Flow microfluorometric system for screening gynecologic cytology specimens using propidium iodide-fluorescein isothiocyanate. J Histochem Cytochem 24(1):322, 1976.

14. Shackney, S.E., Ford, S.S., and Wittig, A.B. The effects of counting threshold and emulsion exposure duration on the percent-labeled mitosis curve and

their implications for cell cycle analysis. Cancer Res 33:2726, 1973.

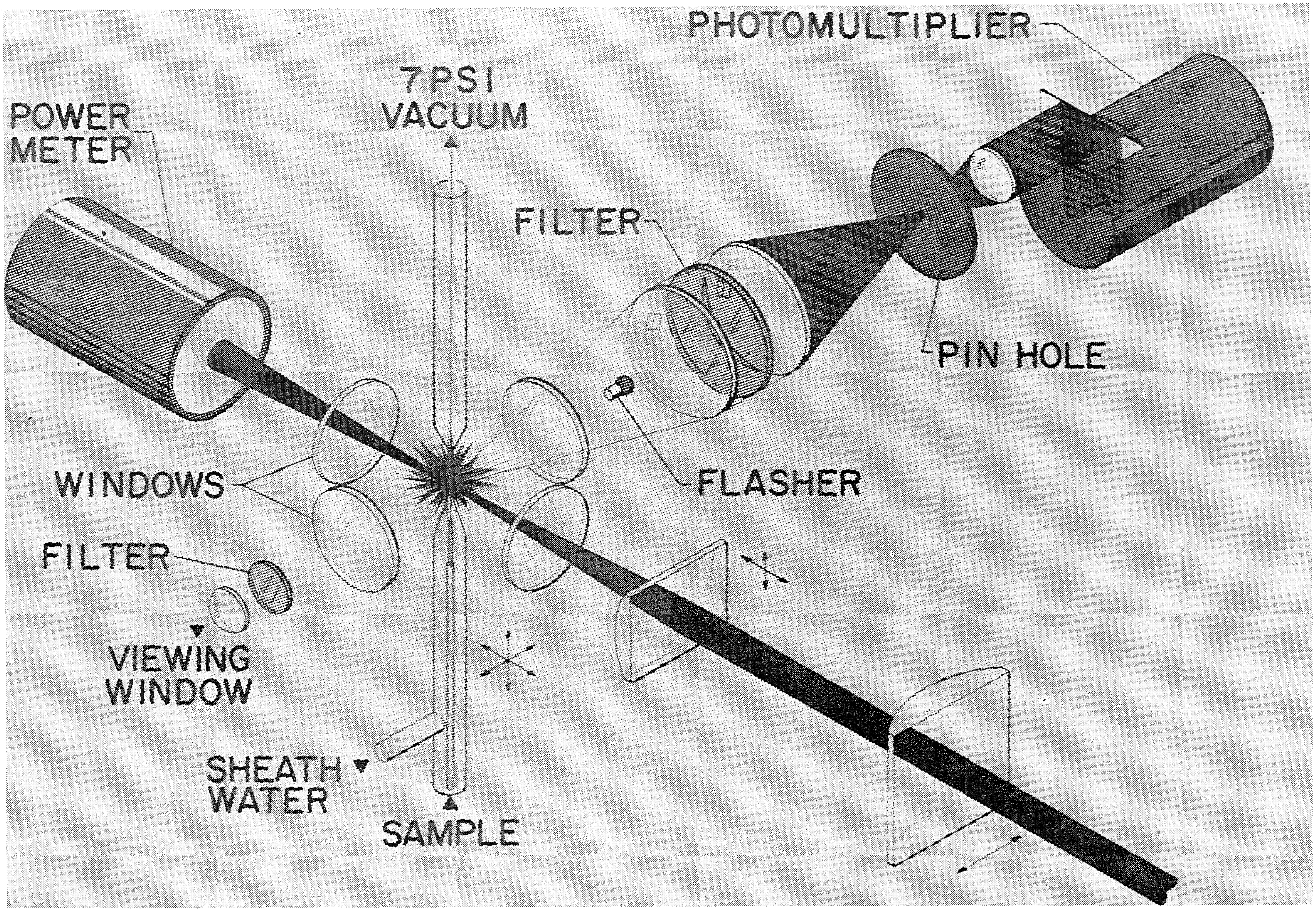
15. Van Dilla, M.A., Gledhill, B.L., Lake, S., Dean, P.N., Gray, J.W., Kachel, V., Barlogie, B., and Goehde, W. Measurement of mammalian sperm deoxyribonucleic acid by flow cytometry. Problems and approaches. J Histochem Cytochem 25(7):763, 1977.

16. Kerker, M. Microstructure effects upon emitted fluorescent intensity from stained biological cells. (presentation at the Sixth Engineering Foundation Conference on Automated Cytology, Schloss Elmau, W. Germany, April, 1978.)

17. Alabaster, O., Tannenbaum, E., Habersett, M.C., Magrath, I, Herman, C. Drug-Induced Changes in DNA Fluorescence Intensity Detected by Flow Microfluorometry: Implications for Analysis of DNA Content Distributions. Can Res (in press, August, 1978).

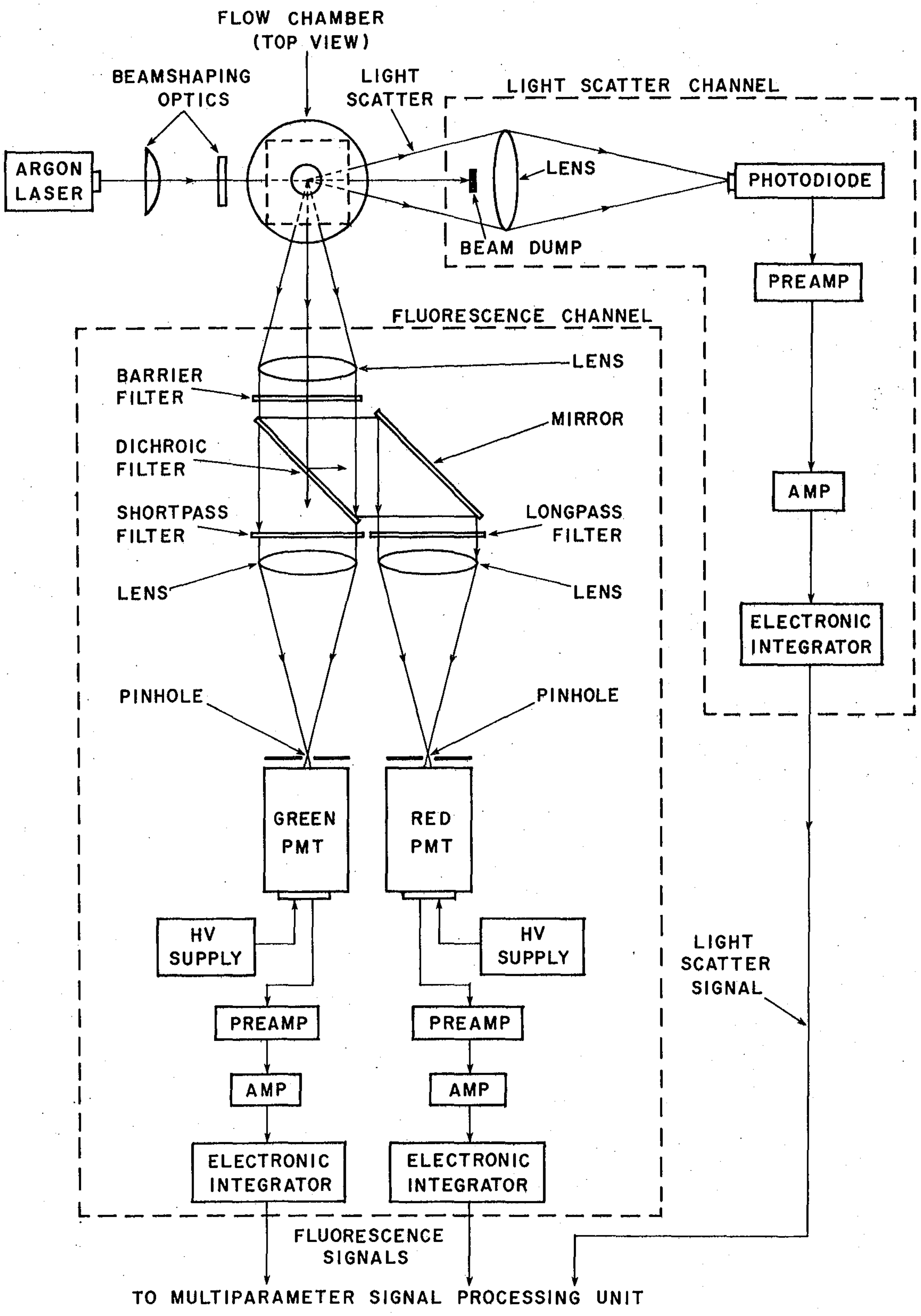
18. Crissman, H.A., and Tobey, R.A. Cell Cycle Analysis in Twenty Minutes. Science 184:1297 (1974).

Figure 1



XBB 765-3975

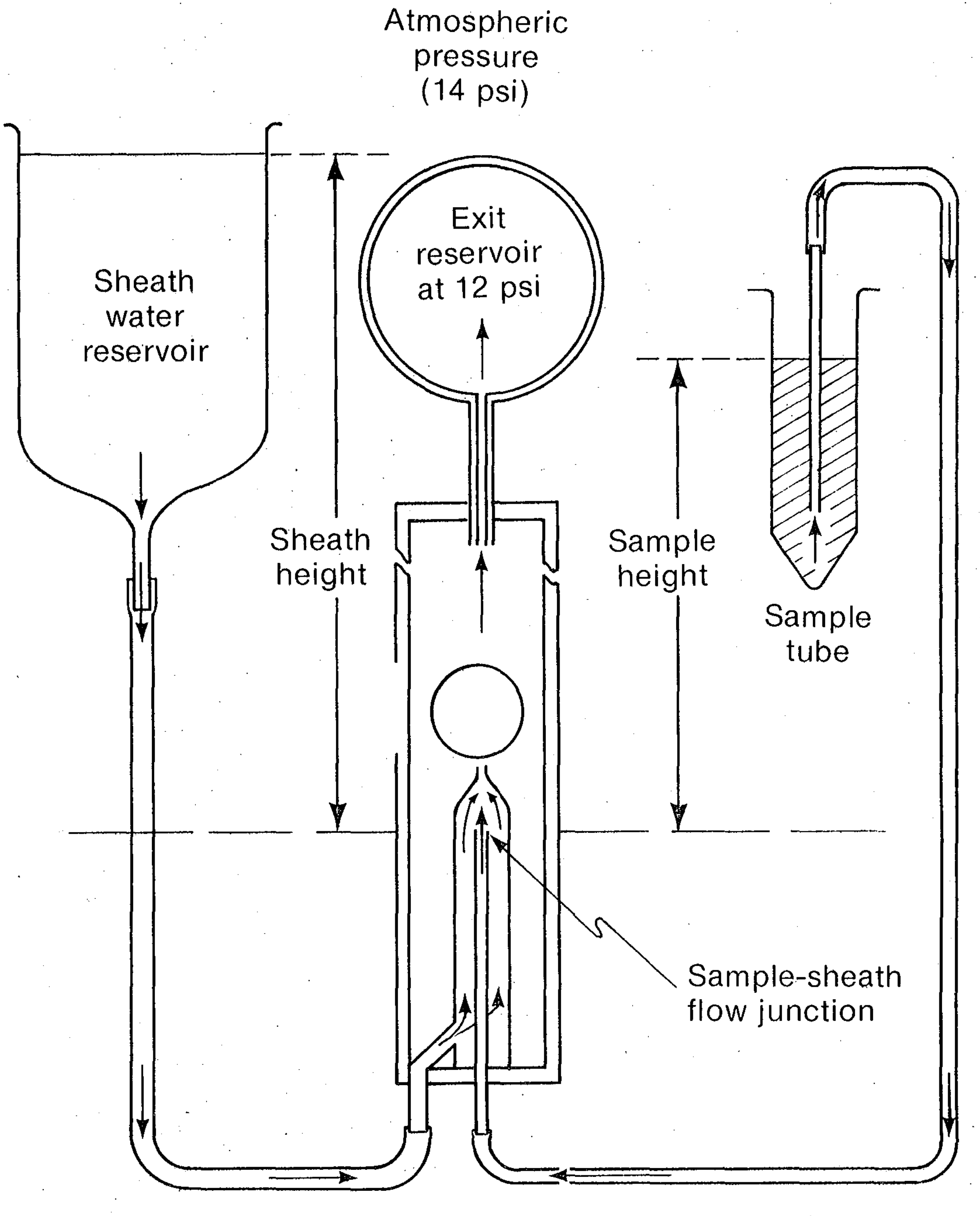
Figure 2



XBL 764-5869

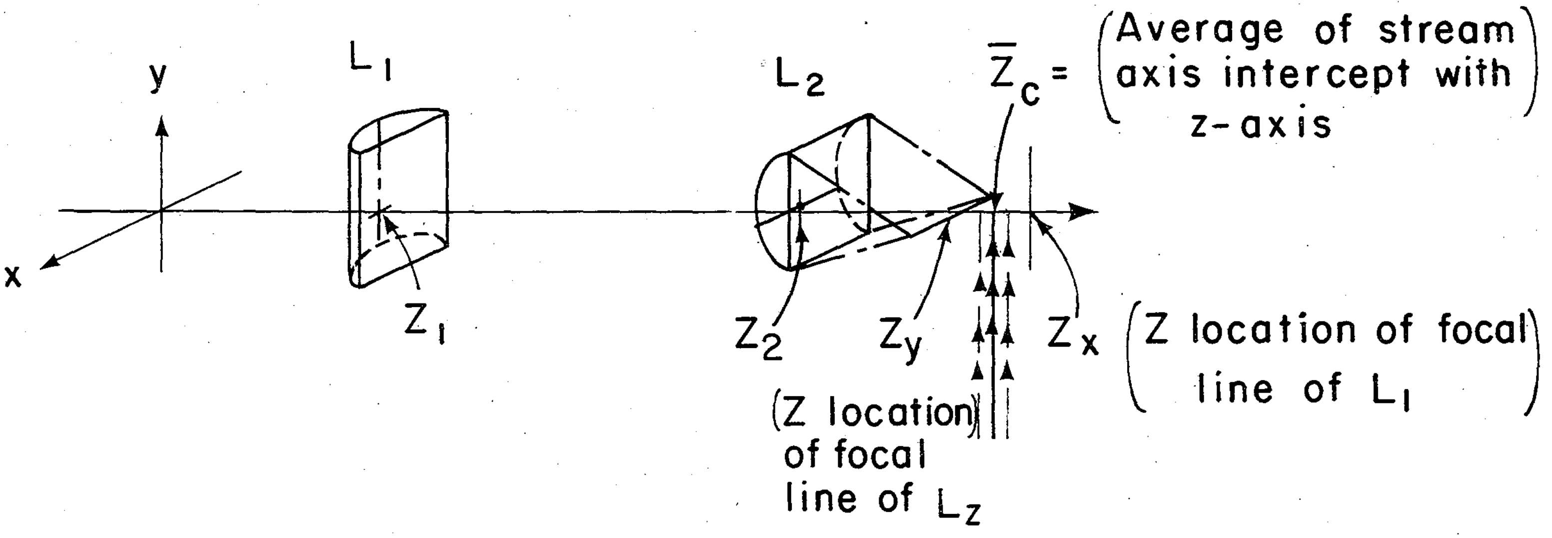
Figure 3

SAMPLE FLOW CELL: OPERATIONAL CONFIGURATION



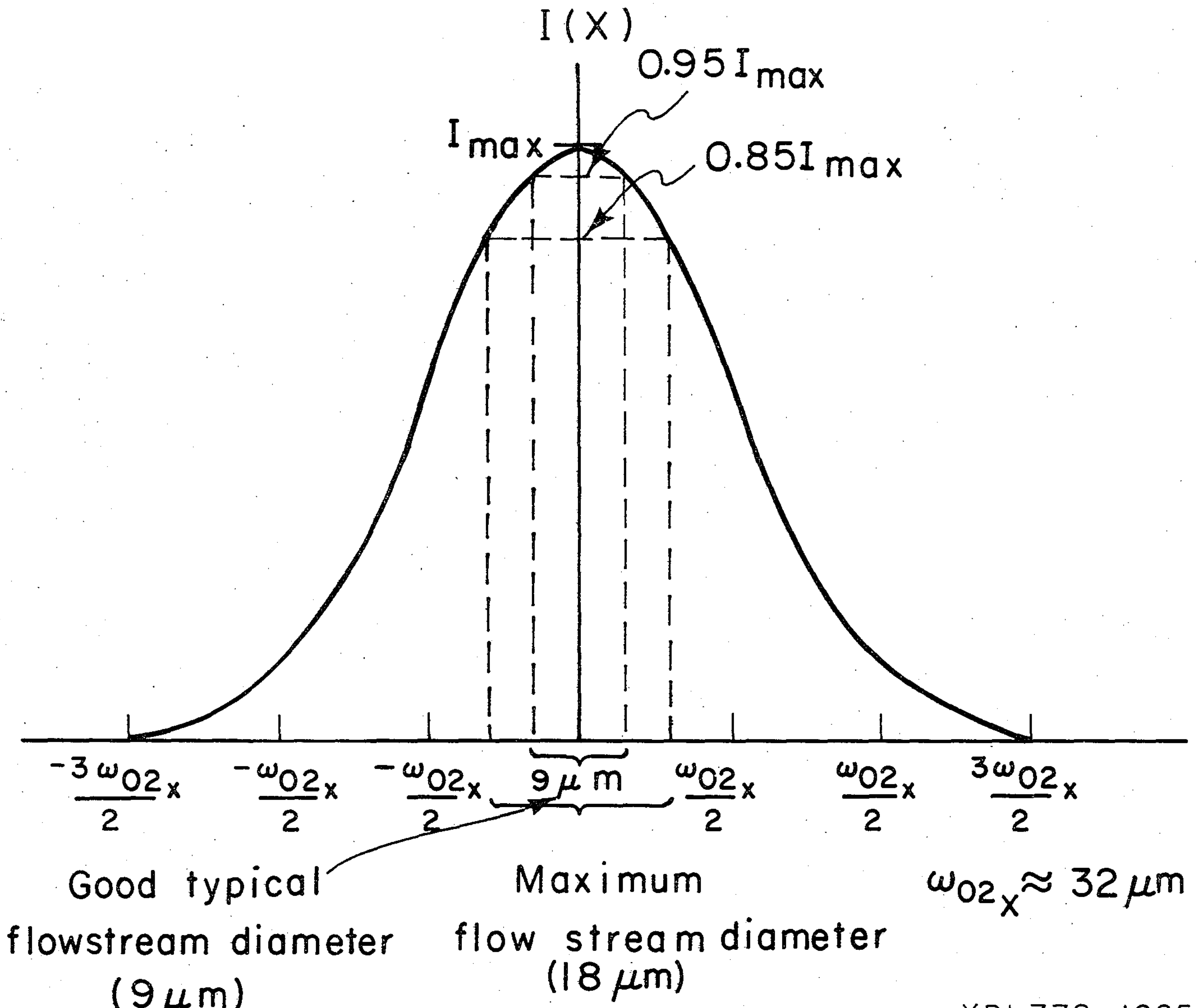
XBL 788-4173

FCM Excitation optical arrangement



XBL778-1664

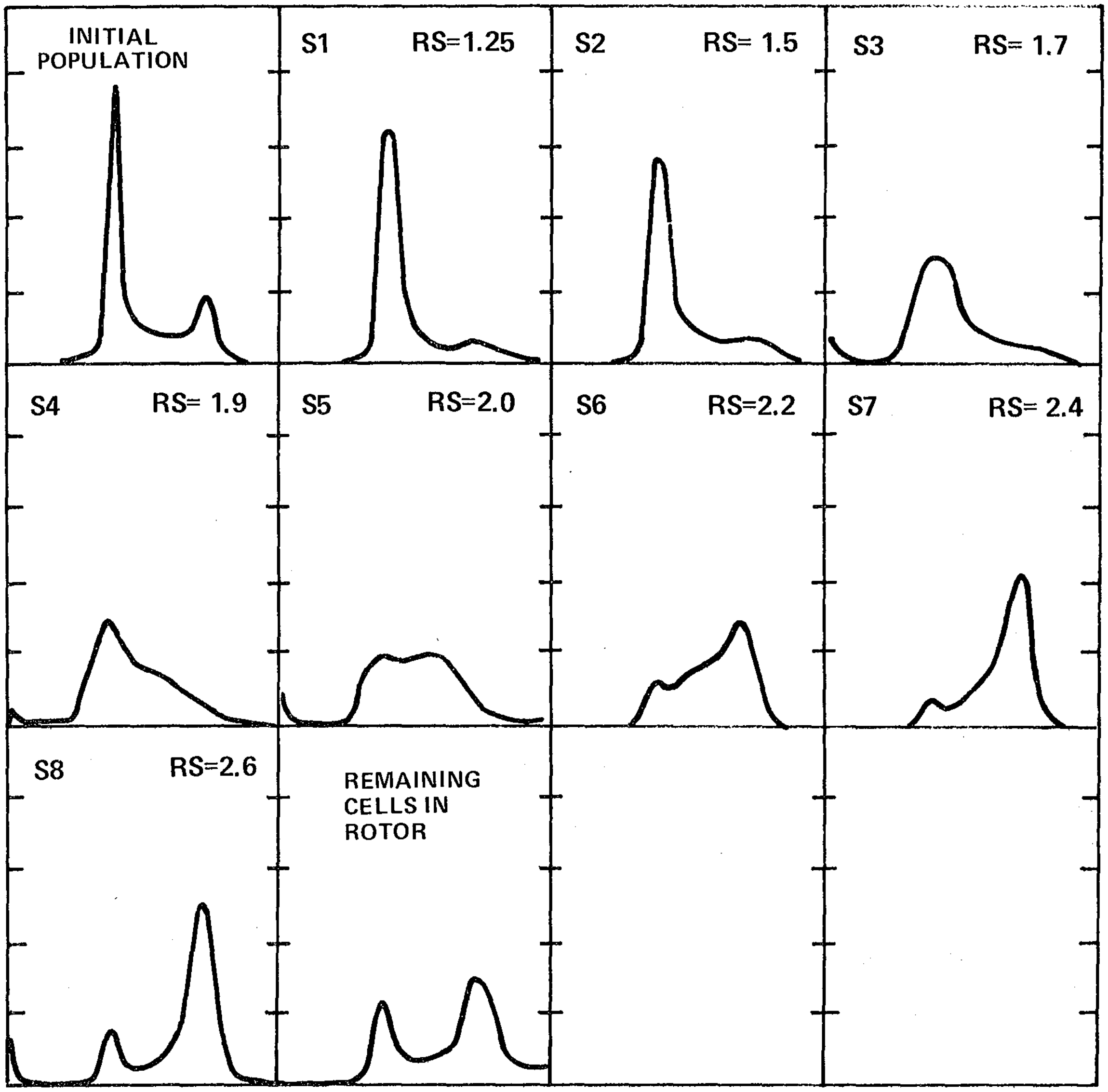
Figure 4



XBL778-1665

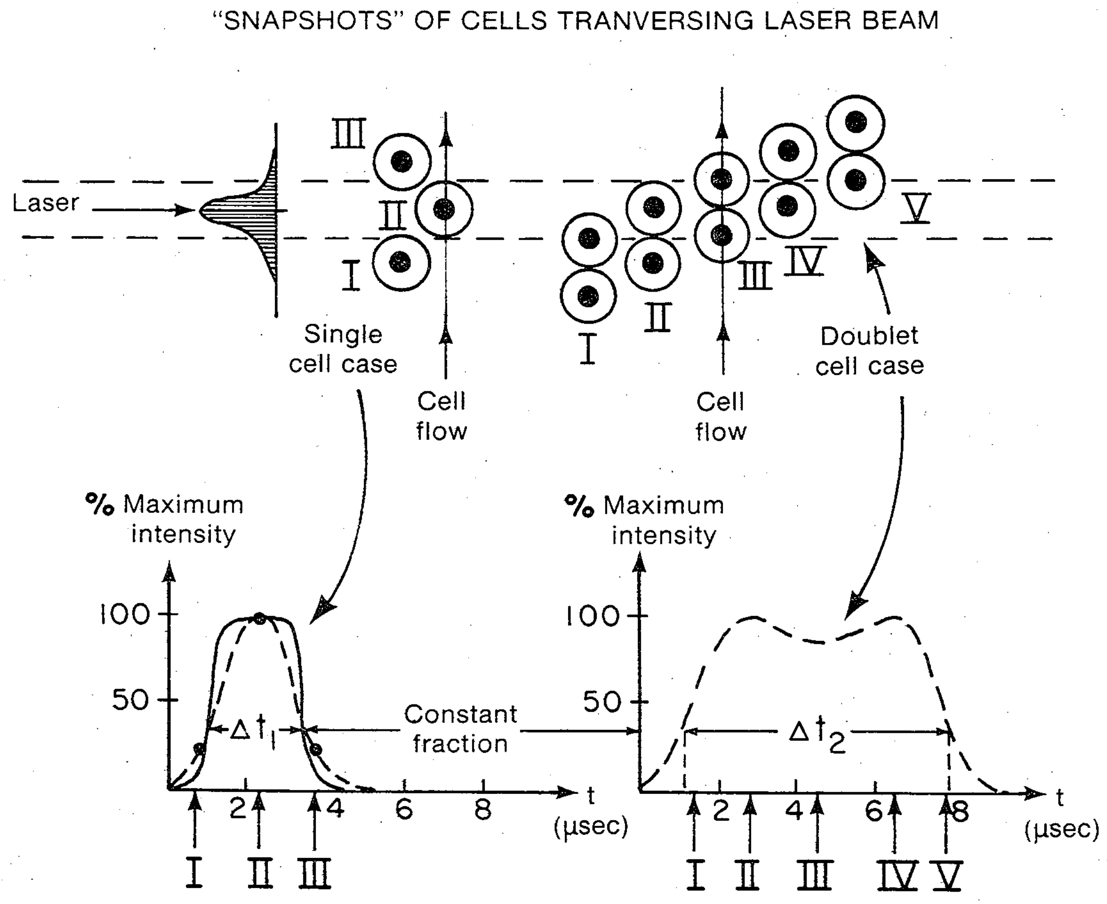
Figure 5

Elutriation fractions from NMuli epithelial cells normalized DNA histograms



XBL 788-4176

PATTERN RECOGNITION OF DETECTED SIGNAL SHAPE: CONSTANT FRACTION DISCRIMINATOR



XBL 788-1684

Figure 7

Chapter II

Cell Cycle Synchronization of Mouse Liver Epithelial Cells by Elutriation Centrifugation

SUMMARY

Detailed methods are described for the sorting and cell cycle synchronization, by means of centrifugal elutriation, of an established mouse liver epithelial cell line, NMuLi, whose isolation is described by Owens (4). In a comparison between three different elutriation buffers and between two different temperatures (4 and 20 degrees celsius), the NMuLi cells were found to be most sensitively sorted in cell cycle when run in their usual growth medium in the absence of serum, at the lower temperature. Under these conditions, and using decrements of rotor speed calculated as described below, an initially asynchronous population (38% G_1, 36% S, 26% G_2M) was sorted into fractions among which were populations enriched to 60% G_1, 75% S, and 50% G_2M compositions, respectively. Of the cells loaded into the rotor, 30% were lost in the elutriation process, and about 20% recovered as aggregates, while the remainder appeared in the various synchronized fractions.

Speed decrements were calculated by an algorithm

modeling the relationship between rotor speed, counterflow rate (of the suspending medium), and relative sedimentation rate (RS), where an RS value of 1.0 was assigned to the first cells to emerge from the rotor (generally G_1 cells). G_1, S, and G_2M enriched fractions were reproducibly obtained from initially asynchronous NMuLi populations when RS ranges of 1.0-1.2, 1.9-2.1, and 2.4-2.6 were collected, respectively.

Cells sorted in this manner demonstrated no loss of viability, and upon replating showed significant movement in the cell cycle, as judged by DNA histograms, by four hours post elutriation.

INTRODUCTION

The mitotic cycle of cells is the focus for a great variety of studies in cell biology, ranging from the control of cellular proliferation, to mutation and heredity (1). In many of these studies, the investigator would like to focus on growing cells that are synchronized in a particular portion of the cycle, but most such cell populations possess little natural synchrony, being more or less randomly distributed about the cycle. Hence, methods for bulk synchronizing large cell populations have long been sought, and several have been brought into common use (1).

Methods for cell syncronization can be classified into two basic groups: induced synchrony (by chemical or physical means) or physical selection. Chemical induction methods involve treatment of the cell population with compounds known to arrest the cells in specific portions of the cycle, or deprivation of essential nutrients from the medium, to achieve the same effect (2). Examples are addition of hydroxyurea (late G_1 arrest), excessive thymidine (halts cells that are actively synthesizing DNA), colchicine (inhibits mitosis by blocking spindle formation), numerous chemotherapeutic drugs such as bleomycin or vinblastine, and modulation of the concentrations of ubiquitous chemical species in the cellular environment such as hydrogen ions, magnesium and calcium ions, and a great many others (1). Deprivation of serum from the growth medium is a typical method for synchronizing cells in G_1 phase (2). Physical induction methods include hypo- and hyper-thermia, irradiation by various forms of ionizing radiation, and others. All these methods share the drawback that they perturb the biochemistry of the cells, affecting metabolism so that growth proceeds in an unbalanced manner after synchronization (1).

Physical selection methods attempt to exploit cell cycle related shape, size or other physical properties of cells in order to separate cells in different cycle phases from one another, with minimal perturbation of their metabolic state. The mitotic shake-off method (3) is a frequently used example of this approach, where one takes advantage of the fact that mitotic cells are less firmly attached to their culture dish surface than are their interphase counterparts. By administering a mechanical shock to the growth surface, cells in mitosis are shaken from the surface, while the the others remain fastened, and the released cells are recovered by spinning down the aspirated medium. Mitotic shake-off has the disadvantage that the size of the final synchronous population is determined mostly by the proportion of cells in mitosis at the time of synchronization, and that proportion is generally less than 10% in unperturbed, growing cell populations. This limiting factor is often prohibitive for many types of studies, so many investigators resort to the above synchrony induction methods to pre-condition the population prior to mitotic shake-off, in order to increase the yield. The notion that propagation of the synchonized cells can yield other desired phase enrichments is bedeviled by the fact that cell synchrony is gradually lost due to the natural variation in the kinetic parameters of the cell cycle (4). Hence, this method shares the drawback of all the above induction methods in that synchrony in only one phase of the cycle is obtained from a given population, in this case, mitosis.

The fact that cell size increases steadily in an exponential manner with age during the cell cycle (5) is the basis for a group of methods for synchronization by size selection. These methods involve sedimentation of the suspended cell population through a viscous medium, either at unit gravity or under high-g forces in a centrifuge, and rely upon the correlation between sedimentation velocity and cell size to achieve separation. Sedimentation rate separation performed by centrifugation through a gradient (6,7) is limited in the number of cells separable and in resolution because of wall effects in the centrifuge tube (8). The Staput method of velocity sedimentation at unit gravity achieves greater resolution; however, it is time-consuming and requires large gradient volumes when large numbers of cells are necessary (2,9,10). Zonal centrifugation overcomes some of the difficulties of these methods, but is difficult to implement (11).

Elutriation centrifugation, in which the cell population sediments through its suspending medium against a continuous counterflow, was first developed by Lindahl in 1948 (12), and is described in detail below. Elutriation provides several important advantages as a method for bulk synchronization, as compared to any of the above methods. As Meistrich (13) and Mitchell (14) argue, elutriation is an effective cell cycle enrichment method for many cell systems, whether their origin is monolayer tissue culture, suspension culture or sampled directly from intact tissue. This is certainly not the case for methods such as mitotic shake-off, for example. Elutriation also yields enriched population fractions from several different cycle phases from the same initial population, a crucial advantage where behavior of different cycle phases is to be compared. This method is also among the most gentle, in that the cells are usually suspended in their normal growth medium during the entire process. The only insult suffered by the cells is that due to the harvesting procedure (in which monocellular suspension is obtained, and which is not necessary for systems grown in suspension culture), and the maintenance of cold ($5^{o}C$) conditions throughout most elutriation runs. A further advantage is that large populations (up to $700x10^{6}$ cells) can be sorted into phase-enriched fractions by a single elutriation run, making possible experiments requiring large synchronous populations. For these reasons, we have sought to implement an elutriation system for obtaining cycle phase enriched popuations for our replating experiments aimed at determining cycle specificity of chemical carcinogen effects upon cultured epithelial cells.

CELL CYCLE SYNCHRONIZATION BY ELUTRIATION CENTRIFUGATION: THEORY

Elutriation is an elaborate form of velocity sedimentation centrifugation, in which particles separate from one another under the influence of a high-g accelerating field by virtue of differences in their rates of sedimentation through the viscous suspending medium. Underlying the use of this separation method are two key assumptions: that sedimentation rate correlates with increased cell "size" (volume, surface area, radius are typical size descriptors), and that "size" correlates with increased cellular maturity. The former is not easily justified on hydrodynamic grounds for other than spherical or ellipsoidal particles, shapes from which suspended cells may significantly vary, but some correlation of sedimentation rate with average cell volume has been demonstrated (2). That mitotic cells should have approximately twice the volume of each of their two G_1 daughter cells is the basis for the latter assumption, and a general correlation between cycle phase and mean cell volume for some cell systems has been demonstrated (2,7).

By arranging a continuous counterflow of medium against the direction of the accelerating g-forces, judicious design of the separation chamber can result in a balance of acceleration and drag forces (Figure 1), so that particles reach steady state radial positions within the chamber, with each sedimentation rate corresponding to a different radius (figure 2). In the Beckman JE-6 rotor, particles with smaller sedimentation rates tend to be found at smaller radii (i.e., near the chamber exit), so to remove a population of slowly sedimenting particles from the chamber, one needs to decrease the accelerating forces (which point away from the exit) relative to the drag forces. As this can be done by either increasing the counterflow rate or decreasing the rotor speed, we have elected the latter because of the somewhat better control obtainable thereof. Sequential speed decrements should correspond to increasing average particle sedimentation rate for the population fractions removed. Even though the presence of turbulence within the chamber often perturbs the steady state conditions described above, we have found the correlation between removed population order and average particle size to be fairly good in practice.

INSTRUMENTATION

The Beckman JE-6 elutriator rotor and associated J-21 centrifuge form the heart of the system, illustrated in Figure 3. The centrifuge rotation speed control has been modified to use a ten turn potentiometer for greater precision, as we have chosen to use decrements of rotor speed as the means of removing subpopulations from the separation chamber. Buffer is continuously drawn through the system by peristaltic pump (Cole Parmer Masterflex, with 7014 head), and the sample is loaded through a post-pump valve at 10ml/min via a syringe drive (Orion Instruments, Cambridge, Mass.). Fluid from the pump passes through a bubble trap (which also serves to damp out the peristaltic pressure variations) prior to traversing a triflat flow meter (5-60 ml/min range, Manostat, New York, NY), from which it flows directly into the elutriator rotor.

Upon exit from the rotor, fluid passes through a Beckman DB Spectrophotometer equipped with a custom built 4cm path length flow cell (Figure 4) capable of sustaining flow rates in excess of 500 ml/min (for flushing purposes). At illumination wavelength of 600nm, the optical detection system easily senses cells emerging from the rotor in concentrations of 2.0×10^3 cells/ml or greater, by virtue of their light scattering and consequent decrease in relative transmissivity through the flow cell. A valve then enables either collection or disposal of the effluent, the desired subpopulations being aseptically collected via a specially designed sterile transfer assembly (Figure 5).

METHODS

CELL CULTURE TECHNIQUE:

Monolayer cultures of the established epithelial cell line NMuLi, derived from the livers of Namru mice by Owens, et al (4), were seeded in 100mm petri dishes (Falcon plastics, Oxnard, Calif.) in Eagle's Minimal Medium

(GIBCO, Grand Island, NY), denoted here as MEM, with 10% donor calf serum (Flow Laboratories, Rockville, MD), and allowed to reach confluency (about 1.0×10^7 cells per dish or 1.3×10^5 cells per square centimeter). 60 hours prior to elutriation, the cells were transferred to roller bottles (Falcon plastics, Oxnard, Calif.) each possessing ten times the surface area of one 100mm dish, and a total of 8 bottles were seeded at one tenth confluent density (1.0×10^7 cells per bottle).

For elutriation, medium was aspirated from the bottles and the cells were washed once with Saline GM (1.5mM Na2HPO4, 1.1mM KH2PO4, 1.1 mM Glucose, and 0.14M NaCl, at pH 7.4). Cells were then loosened from the roller bottles by trypsinization for 3 min. at 37 degrees Celsius using DISPO (saline GM containing 0.5mM EDTA and 0.1 mg/ml crystalline trypsin--GIBCO, Grand Island, NY), and washed off the surface with soybean trypsin inhibitor solution NEUT (Saline GM containing calcium and magnesium ions, soybean trypsin inhibitor, bovine serum albumin and DNAse -- reference 6) , yielding 6.0×10^8 cells in 40ml suspension for loading into the elutriator. Prior to loading the cells, the elutriation system was flushed with 70% ethanol, followed by rinsing with sterile suspension buffer.

OPERATIONAL TECHNIQUE

Multiple tests on NMuLi cells showed that good cycle enrichment required that the cells be suspended in cold (4 degrees celsius) buffer, and that their normal growth medium (MEM), without serum, served as well as any other buffer tested (figure 6). Poor results were obtained for samples of less than 1.5×10^8 or more than 8.0×10^8 cells, so this range was used in our experiments. Cells were routinely loaded with rotor speed set sufficiently high (typically 3500-4000 revolutions per minute for a counterflow rate of 25 ml/min) to prevent the washing out of any likely whole cells. Speed decrements began from this initial loading condition, which corresponded to 100% relative transmissivity through the spectrophotometer flow cell. The set of speed decrements for obtaining the desired cycle phase enrichments in the fractions was determined by a modeling routine, described below, implemented during the experiment by means of a programmable hand-held calculator.

DNA HISTOGRAMS

Cell cycle distributions of all cell populations, both before and after elutriation, were obtained by staining the cells using the propidium iodide technique described by Crissman, et al (5) and analysed in a flow microfluorometer (Los Alamos Flow System II, Ref. 8). A Spectra Physics Model 171 Argon Ion laser provided a 3 watt excitation beam at 488nm wavelength. Individual histograms were normalized to constant total cell count, and gain shifted to place the G_1 peak mode in channel 100, to facilitate visual comparison.

RESULTS

Of the cells loaded into the rotor, we were able to account for only 70-80% as a rule, including some 10-20% that showed up as aggregates in the last fraction collected. Reconstitution experiments, in which equal volumes of each collected fraction were pooled to reconstruct the initial pre-elutriated population, revealed no cycle selectivity in loss of cells during the elutriation process. Thus, we suspect the cell loss to be a random phenomenon arising from the turbulence caused by the entrance of the counterflow into the separation chamber.

Examination of elutriated fractions by microscope showed them to be virutally free of cell aggregates or debris, indicating yet another advantage of the elutriation technique: "clean up" of the sample populations. DNA histograms of collected fractions consistently showed far less low channel noise (debris) and shoulder on the left side of the G_1 peaks than their pre-elutriation parent populations, and much less evidence of aggregates was apparent in the respective histograms. Cellular debris was detected in large amounts by the spectrophotometer when far UV illumination (260nm) was used, due to molecular absorption within the fragments. The effluence of debris consistently occurred at rotor speeds far too high to permit the exit of whole cells, and subsided within 10 to 15 minutes after all the sample cells had been loaded into the chamber. Microscopic examination of this collected effluent revealed a large assortment of cellular fragments, which go virtually undected at the higher spectrophotometer wavelength (600nm) because the suspended particles are too small to scatter significantly.

From one trial to the next, G_1 cells (the first significant fraction detected) eluted from the rotor at widely varying sedimentation rates (equivalent to 5-9 mm/hr at 1-g), reflecting volumetric or density variations we found difficult to control experimentally. However, the ratio of G_2M cell sedimentation rate to that of G_1 cells proved sufficiently constant to be used in reproducibly obtaining fractions enriched in G_1,G_1-early-S, S, late-S-G_2M, and G_2Mphases, respectively.

Speed decrements were calculated by expressing the relationship between rotor speed (W), counterflow rate (F) and relative sedimentation rate (RS) as

$$W = \text{constant times } (F/RS)^{1/2}$$

where the constant has been determined empirically to be 1.93 for our cells, W is in units of thousands of revolutions per minute (KRPM), F is in ml/min, and RS is in mm/hr equivalent at 1-g. To apply the model, the sedimentation rate of the G_1 cells was determined for fixed flow rate by the highest rotor speed at which the optical percent transmission dropped below 60, corresponding to about 3,000 cells/ml. These cells were assigned a relative sedimentation rate (RS) of 1.0, and subsequent rotor speed settings were calculated from the model by inserting the RS values covering a given desired range, typically 1.0-1.2, 1.9-2.1 and 2.5-2.7 for G_1, S, and G_2M enriched fractions, respectively. These values were found to work best when 600-800 million cells were loaded, and to require adjustment for different sample sizes.

In a test of the stability of this method, parallel batches from the same parent cell population were run at two different counterflow rates (15 ml/min and 25 ml/min). Fractions were collected at the same relative sedimentation rates in each case, and their DNA histograms showed them to be virtually identical in cycle phase enrichment, indicating that the method is robust with respect to flow rate for this cell system.

DNA histograms of the parent and elutriated populations collected in one of the more successful runs are shown in figure 8. These cells were replated upon collection, with no detected loss of viability as a result of the elutriation procedure, and all showed movement in the cell cycle within 4 hours post seeding. The particular replate experiment for which this elutriation was done required a very large G_1 enriched sample population, and we consequently chose to sacrifice some synchrony in order to collect enough cells. Other runs have yielded virtually pure G_1 NMuLi cells from asynchronous initial populations. With our sterile collection system, we have yet to suffer any contamination in our replating experiments.

DISCUSSION

If biological cells were spherical in shape, they would obey Stokes' Law, i.e.

$$sv=(2/9)x(g/n)x(densdif)r^2$$

where sv = sedimentation velocity (absolute), g = gravitational or centrifugal force, n = viscosity of the medium, densdif = (particle density - medium density), and r = particle radius. By this relation, one would predict that cells possessing a given volume (at the same density) would sediment at about 1.6 times the rate of cells possessing half that volume. The fact that we find G_2M cells sedimenting at rates far greater than this would allow can be accounted for by any or all of the following reasons:

1) Cell hydrodynamics differ greatly from those of spheres.

2) Density variations among cells may cause some sedimentation rate variation.
3) Viscosity of the medium in the separation chamber is a function of the number of cells/cc in the chamber, among other factors.
4) Interactions between cells within the chamber may result in different hydrodynamic drag forces acting upon them.

In view of the above complexities, our model of a constant relative sedimentation rate range over the cell cycle is surprisingly useful for our cell line, and we are testing it on other lines. One clear way to improve the model would be to include the effects of sample size, which we are presently attempting to do empirically. Perhaps better methods and models for accommodating the above sources of variability will be developed, so that better reproducibility of synchronization can be achieved.

The persistence of G_1 cells in all cell fractions implies that there is a wide variation in sedimentation rate among G_1 cells, and "pollutes" the otherwise rather successful synchronization. We found that this phenomenon is minimized by the use of cold suspension buffer (perhaps causing cells to assume a rounder configuration), but not completely eliminated.

We use a high (25 ml/min) counterflow rate because of the correspondingly higher rotor speeds necessitated by it, the lowest of which is typically 1.4 to 1.7 KRPM for our single cells. We prefer this because at lower speeds, reproducibility and stability of the Beckman J-21 centrifuge speed control drop considerably, resulting in "hunting" by the servo speed control, with consequent improper removal of cell populations from the separation chamber.

Several improvements in the instrumentation would greatly facilitate the use of elutriation in cell cycle kinetic studies: First, a better servo regulator on counterflow rate would stabilize the flow variations due to resistance differences along the "dispose" versus the "collect" branches of the system, and those arising during the separation due to the removal of cells from the chamber (and consequent viscosity change). A better method for detecting the first eluted cells would remove some of the ambiguity in the assignment of the relative sedimentation rate of 1.0 (the present method is insensitive to the total loaded sample size in setting its detection threshold to 60% transmission through the spectrophotometer). Also, since replating is often involved as a subsequent step, it would be highly desirable to know what rotor speed decrement will result in a given desired number of cells emanating from the rotor, so that plating density requirements can be met. The method could perhaps be based on the spectrophotometer reading, with some form of predictor-corrector scheme used to start decreasing rotor speed, and cease decrementing when the appropriate number of cells has exited.

CONCLUSION

We have developed a fairly reproducible synchronization scheme for the mouse liver epithelial cell line NMu-Li, using centrifugal elutriation in conjunction with a model relating relative sedimentation rate to cell cycle phase. The method produces population fractions enriched in five distinct cycle phases from an initially asynchronous pre-elutriation population, and the fractions show neither loss of viability nor major perturbation in their progression through the cell cycle. Improvements are to be expected in operational technique and in predictive modeling of the relation between cell cycle phase and absolute or relative sedimentation velocity for the cells in question.

ACKNOWLEDGEMENTS

We wish to gratefully acknowledge the generous support and advice extended to us by Dr. Robert Grabske of Lawrence Livermore Laboratory, who was vitally helpful in the initial design and debugging of our elutriation system. In addition, thanks are due to Dr. Benjamin Gordon and to Dr. Wallace Irving of our laboratory for their valuable assistace in plumbing and other practical aspects of this project. Finally, we express our deep appreciation to Mr. Cees DeGroot and to Mr. Richard O'Brien of our laboratory, whose design and construction talents were vital to the sterile transfer assembly and collection elements of our elutriation system.

FIGURE LEGENDS

1. How elutriation works: each cell is driven away from the rotor center by centrifugal forces, dragged back towards center by continuous counterflow of suspending medium. Flow chamber geometry gives nonuniform flow field, allowing particles of different sizes to migrate to different stable radial positions, where the two forces are in balance.
2. Beckman JE-6 rotor separation chamber. Geometry designed so that the larger the cell, the further its steady-state position from rotor center.

3. Schematic diagram of elutriation centrifugation system at our laboratory. (see text for details).

4. Specially designed 4 cm path length spectrophotometer flow cell for detecting cells emerging from elutriation rotor. Two quartz windows (9 mm diameter, 1 mm thick) enclose optical path, and 4 mm O.D. glass tubing serves as input and output terminals to the flow

cell. Windows and terminals are cemented to the lexan body by RTV 732 silicone adhesive (General Electric Corp.). The whole unit can be sterilized in an autoclave.

5. Device for sterile transfer of cells from elutriator (sterile system I) to a collection vessel (sterile system II) in nonsterile atmosphere. The two serum capped units are forced together, then the needle driven through both caps by downward pressure on the inner tube of the sliding barrel unit. Aspirating needles maintain constant pressure during the transfer procedure. After transfer, inner tube is withdrawn while caps are still pressed together, so that no contact occurs at any time betwen needle and atmosphere. (patent applied for)

6. Results from a successful elutriation run. Original saple population (upper left) was sorted into 8 collected fractions at the relative sedimentation (RS) rates indicated. RS = 1.0 was assigned to the first cells leaving the rotor (details given in text).

REFERENCES

1. Mitchison, J.M. Biology of the Cell Cycle. Cambridge University Press. London (1971)
2. Whitmore, G.F. In Vitro 6:276 (1971).
3. Terasima, T., and Tolmach, L.J. Exp Cell Res s 33:424 (1963).
4. Anderson, E.C. and Petersen, D.F. Exp Cell Res 36:423 (1964).
5. Anderson, E.C., Bell, G.I., Petersen, D.F., and Tobey, R.A. Biophys J 9:246 (1969).
6. Sinclair, R., and Bishop, D.H.L. Nature 205:1272 (1965).
7. Morris, N.R., Cramer, J.W., and Reno, D. Exp Cell Res 48:216 (19671967).
8. Boone, C.W., Harrell, G.S., and Bond, H.E. J Cell Biol 36:369 (1968).
9. MacDonald, H.R., ad Miller, R.G. Biophys J 10:834 (1970).
10. Shall, S. Methods in Cell Biology (ed D.M. Prescott) vol 7, P. 269. Academic Press, New York (1973)
11. Probst, H. and Maisenbacher, J. Exp Cell Res 78:335 (1973).

12. Lindahl, P.E. Nature 161:648 (1948).

13. Meistrich, M.L. et al. Synchronization of Mouse L-P59 Cells by Centrifugal Elutriation Separation, Exp Cell Res 105(1977)169-177

14. Mitchell, B.F. and Tupper, J.T.

Exp cell res 106(1977)351-355.

15. Grabske, R.J., Lindl, P.A., Thompson, L.H., Gray, J. J. Cell Biol. 67:142A (1975)

16. Owens, R.B., Smith, H.S., Hackett, A.J.

J. Natl. Can. Inst. 53(1):261-269 (July, 1974)

17. Crissman, H.A. and Steinkamp, J.A. Rapid, Simultaneous J. Cell Biol. 59:766-771 (1973)

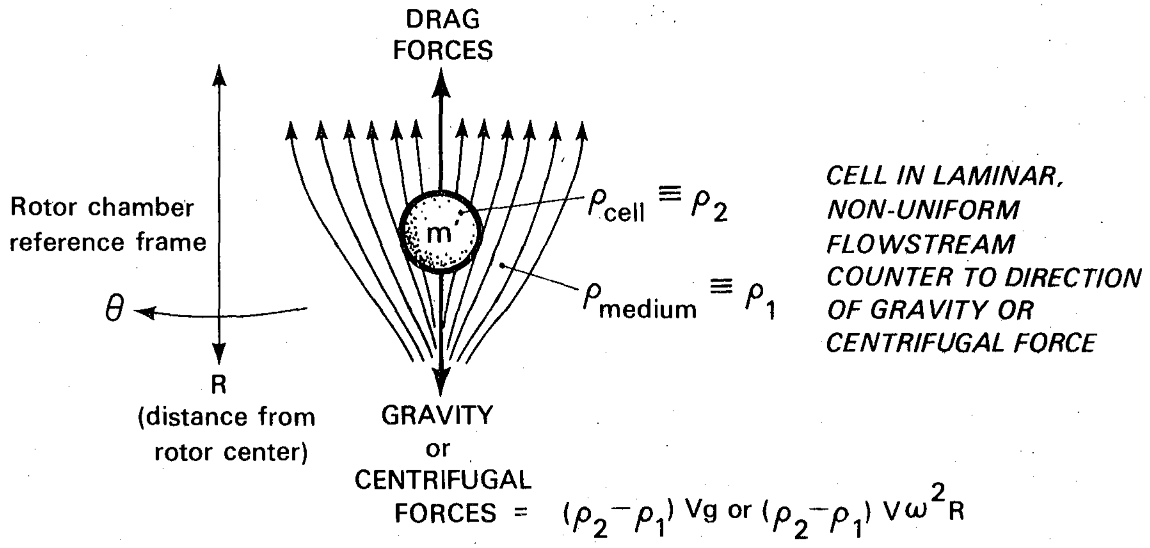
18. Kraemer, P.M., Petersen, D.F., and Van Dilla, M.A. Science 174:714 (1971)

19. Warmsley, A.M.H., and Pasternak, C.A. Biochem. J. 119:493-499 (1970)

20. Holm, D.M., and Cram, L.S. An Improved Flow Microfluorometer for Rapid Measurement of Cell Fluorescence. Exp Cell Res 105-110 (1973).

Figure 1

THEORY OF ELUTRIATION CENTRIFUGATION

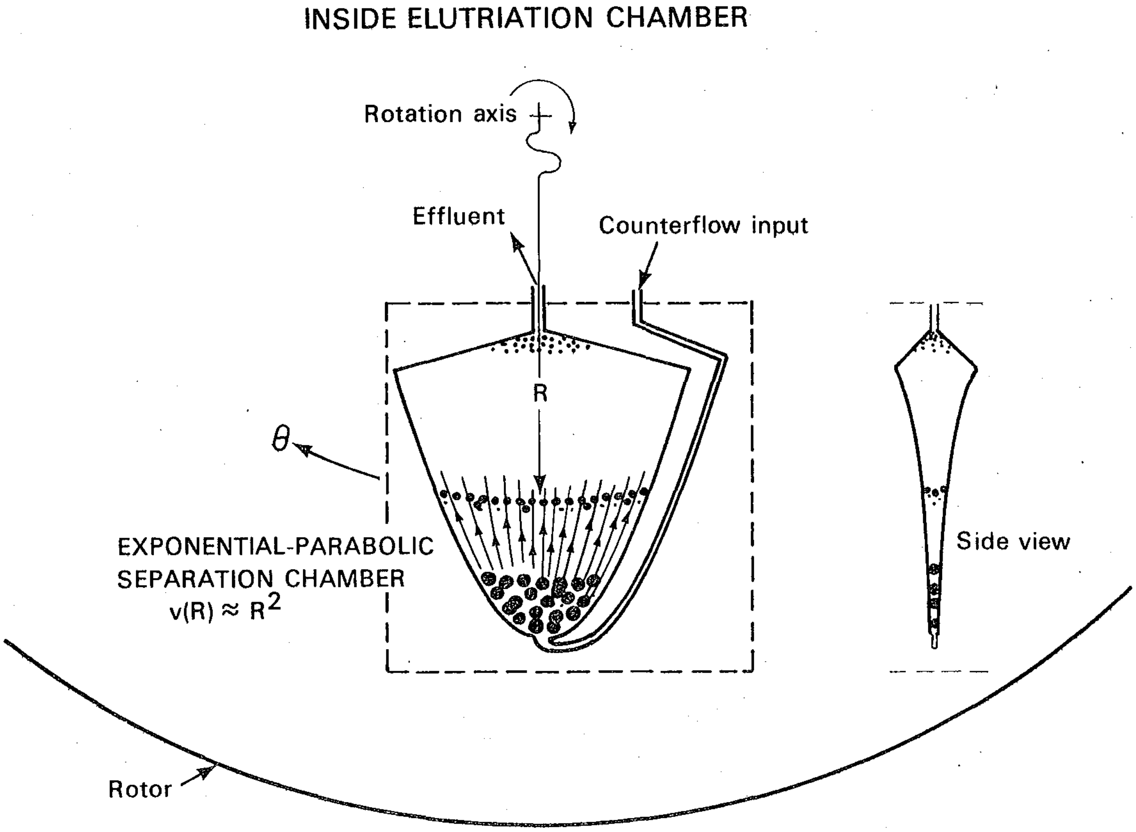


DRAG FORCE: (Friction × flow speed) = $6\pi\eta r v(\vec{R}) \xleftrightarrow[\text{of radius } r]{\text{for spheres}} \frac{4\pi}{3}(\rho_2-\rho_1) r^3\omega^2 R$

WHEN FORCES BALANCE, STEADY STATE REACHED:

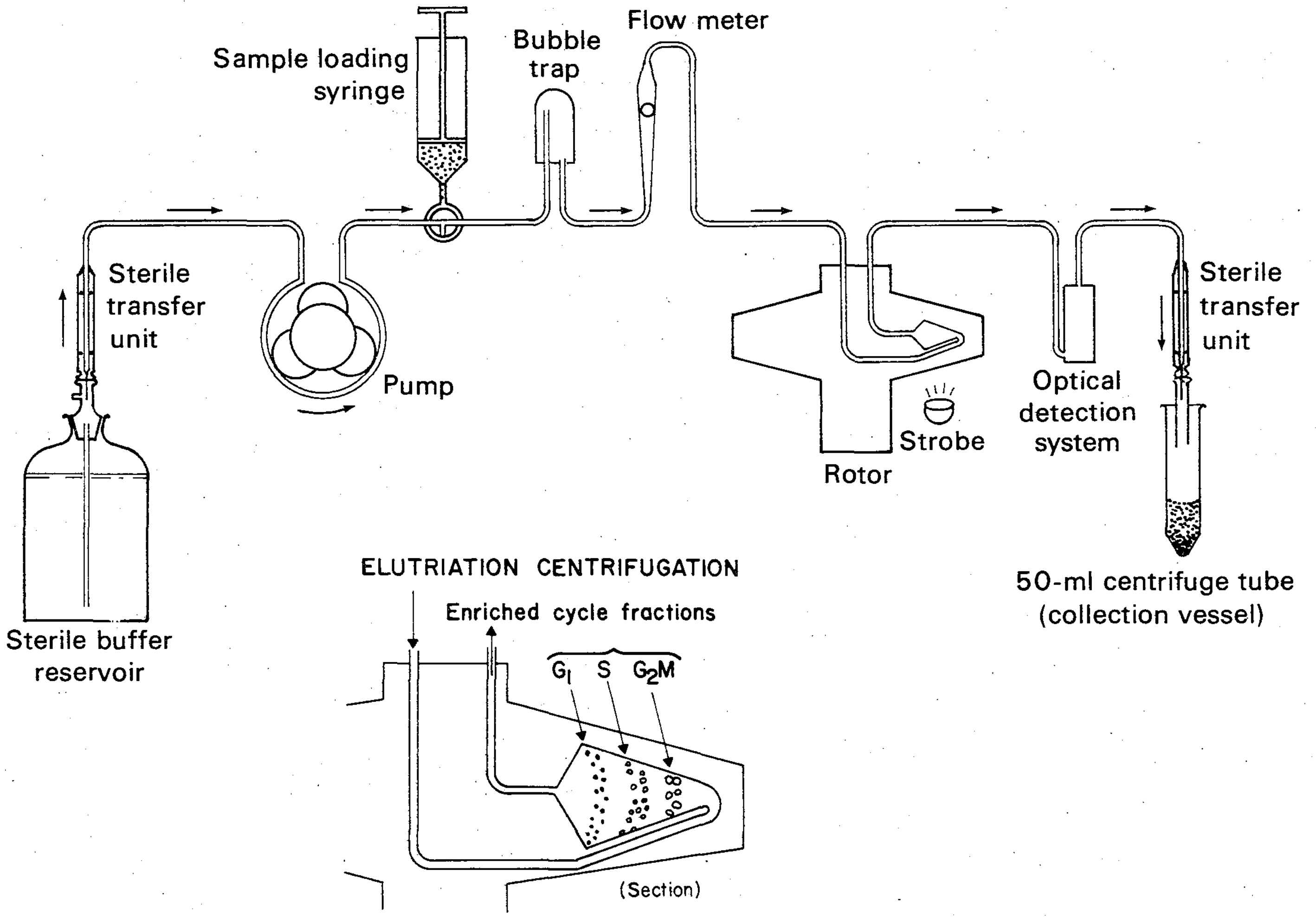
$$r(\vec{R}) = \frac{a}{\omega}\left(\frac{v(\vec{R})}{R}\right)^{1/2}$$

XBL 7712-11467



XBL 7712-11470

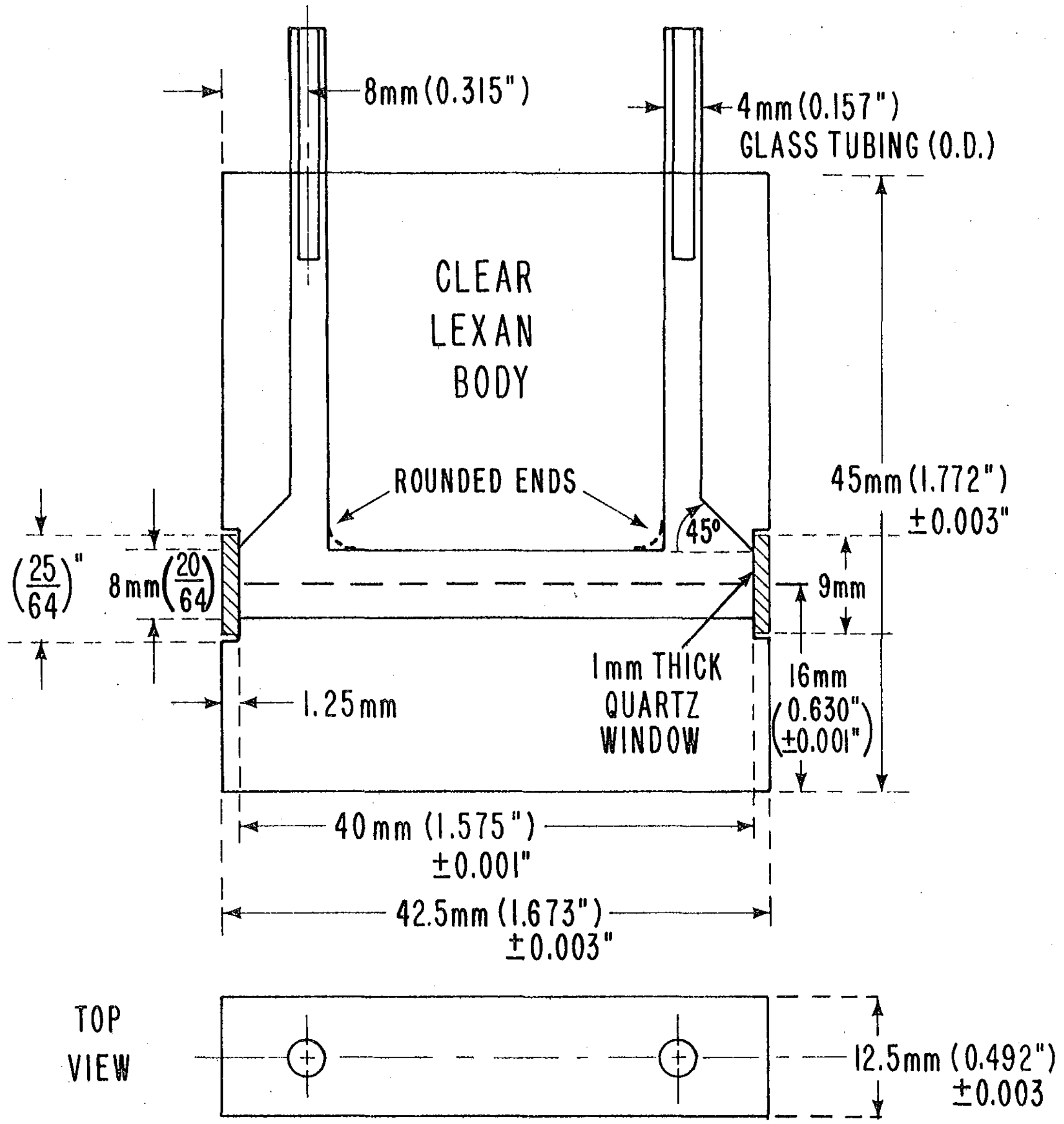
SCHEMATIC DIAGRAM OF LCB ELUTRIATION CENTRIFUGATION SYSTEM



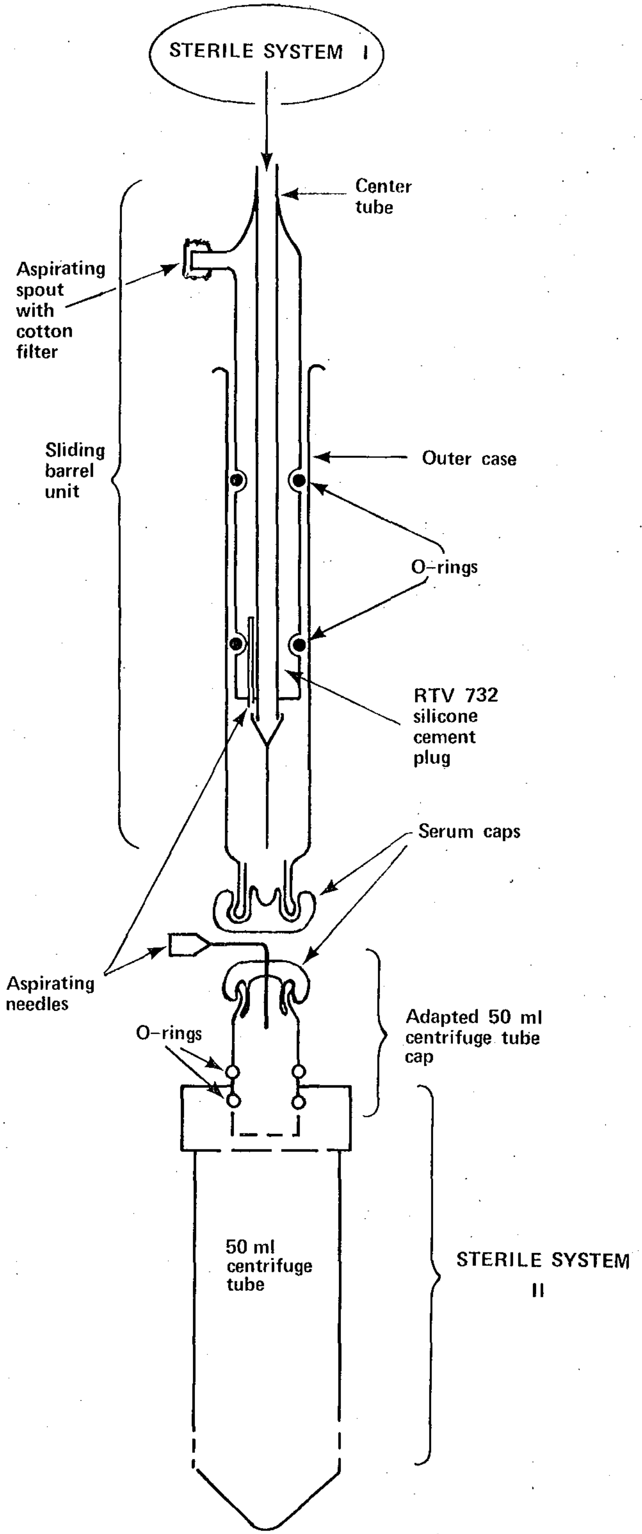
XBL 7712-11468A

Figure 3

4cm PATH LENGTH FLOW CELL FOR ELUTRIATION SYSTEM SPECTROPHOTOMETER

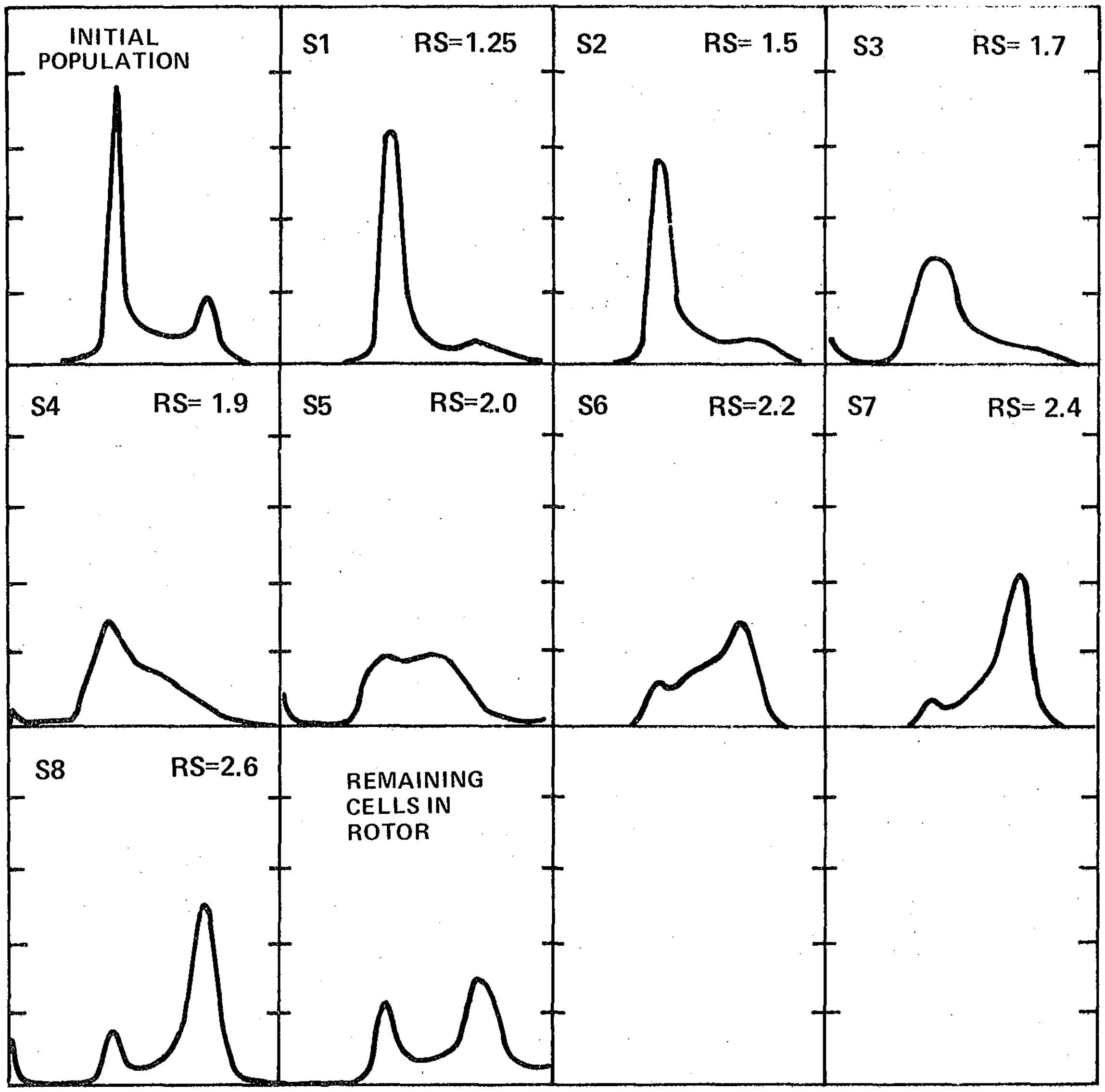


XBL 788-1683


STERILE SYSTEM I
Center tube
Aspirating spout with cotton filter
Sliding barrel unit
Outer case
O-rings
RTV 732 silicone cement plug
Serum caps
Aspirating needles
O-rings
Adapted 50 ml centrifuge tube cap
50 ml centrifuge tube
STERILE SYSTEM II

XBL 788-4171

Figure 6



Elutriation fractions from NMuli epithelial cells normalized DNA histograms

XBL 788-4176

CHAPTER III:

EXTRACTING INFORMATION FROM DNA HISTOGRAMS OBTAINED BY FLOW CYTOMETRY

ABSTRACT

Analysis of DNA histograms obtained by flow cytometry (FCM) is presented from the standpoint of information flow during the FCM preparation and measurement processes, which result in histograms that represent defocused versions of the true DNA distributions of given cell populations. Mathematical analysis leads to a formulation of the problem of recovering maximum information about the true DNA distribution from a given histogram, and it is shown to have properties typical of most inverse problems. The histograms are shown to be well modeled as a convolution of the true DNA distribution with a single overall spread function, when their abscissas have been transformed to represent the logarithm of fluorescent intensity. Specific algorithms for G_1 and G_2M mean estimation, spread function shape extraction, and determination of optimal number of free parameters are presented, along with a discussion of algorithm stability with respect to various sources of error. Results

are shown for various DNA histograms.

INTRODUCTION

Several previous authors (2,6,7) have addressed the problem of fitting or modeling single DNA histograms obtained by flow cytometry (FCM), and a comprehensive review of this subject, as well as its application in the larger problem of cell cycle kinetic studies, has recentely appeared (7). In this paper, we propose a general framework for analysing the problem of recovering information about a given population's DNA distribution from its DNA histogram, with the aim of clarifying the relation between the two. Central to our approch is the notion of a flow of information during the sample preparation, staining and FCM analysis steps leading to the final histogram. The latter is a representation of the original underlying DNA distribution that has been defocused or informationally degraded by the errors introduced at each step. By means of three basic assumptions about the overall errors introduced, we derive a key integral equation relating the histogram to the desired underlying DNA distribution. Given this relation, the problem at hand is seen to be one of inverting the integral operation, and the theory of such inversion methods is used to estimate the limits to which information about the underlying DNA distribution can be recovered from a histogram.

Using the above formulation of the problem, we present an algorithm for analysing virtually any DNA histogram, but preferably one that is part of a time series of histograms. The method assumes no standard shape for the spread function describing the overall errors of FCM, but rather extracts the shape from a reasonably clean G_1 reference peak. We discuss the detailed implementation of this approach, including the determination of the proper number of free parameters in the model of the underlying DNA distribution, estimation of G_1 and G_2M means and of the spread width, for a given histogram. We discuss the stability of the method and propose several ideas for improving upon it.

FLOW OF INFORMATION DURING FCM PROCESSING

Our approach to the analysis of DNA histograms is based on the fundamental notion of information flow from one stage of FCM processing to another, starting with the cells *in vivo* or in culture, and culminating with a stored histogram. A schematic representation of this concept is illustrated in Fig. 1. The information of interest at the first stage is the true population distribution with respect to DNA per cell, henceforth termed the "underlying DNA distribution". This is converted into a fluorescent-label-per-cell distribution by the sample preparation and staining steps, as indicated in Fig. 1, which serves only as an intermediate to the final distribution with respect to measured fluorescent intensity per cell. The observed histogram is a discrete, finite sampling of this final distribution, and as such is subject to the errors of finite counting statistics. We shall not attempt to model the various different sources of error indicated in Fig. 1, but shall rather assume that they can be regarded as a single overall error for each cell analysed by FCM.

MATHEMATICAL FORMULATION

Let $F_1(x)$ represent the underlying (continuous) DNA distribution, and $F_2(x)$ represent the fluorescent intensity distribution of which the histogram is a finite discrete sample.

Let X_1 be the true DNA content of a cell chosen at random from distribution F_1, and let X_2 be the fluorescent intensity assigned to X_1 by FCM analysis.

$$Z = \frac{X_2-X_1}{X_1} \qquad (2)$$

Our analysis is based on four principal assumptions:

1. Measured fluorescent intensity is never negative.

2. The distribution of the relative error Z is invariant with true DNA content X_1. This corresponds to Z and X_1 being presumed statistically independent of one another, and implies that the coefficient of variation is constant throughout the histogram. (This is motivated by the frequent observation that the G_2M peak is about twice as wide as the G_1 peak.)

3. The final histogram displays local scatter, due to finite counting statistics.

We make a fourth assumption, applicable only to histograms recorded as a group: for histograms recorded under tightly controled preparation, staining and FCM analysis conditions within a single experiment of relatively short duration (a few days), the shape of the relative error distribution changes only in width (CV) from one histogram to another. By this assumption, one can use the spread function of one histogram to fit another histogram from the same group; a useful notion, as discussed below.

Using assumptions 1 to 3, we now proceed to show that a given DNA histogram can be well modeled by the convolution of the underlying DNA distribution with a single overall spread function, providing that all are expressed with respect to a logarithmic abscissa.

Rearranging equation (2), we have

$$X_2 = X_1(1+Z) \tag{3}$$

which for small values of Z asymptotically approaches* (by Taylor expansion)

$$X_2 = X_1e^Z \tag{4}$$

Z can be reasonably presumed small enough for this approximation to hold well, since for most DNA histograms, the relative errors (indicated by the CV) are on the order of 0.03 to 0.10 . By a logarithmic change of variables, (4) is transformed into the key relation,

$$Y_2 = Y_1+Z' \tag{5}$$

where $Y_1 = \log_b(X_1)$, $Y_2 = \log_b(X_2)$, $Z' = Z\ln(b)$, and "b" is an arbitrary base.

Defining a new variable (abscissa)

$$y = \log_b(x),$$

we let $P_1(y)$ be the underlying logarithmic DNA distribution, and $P_2(y)$ be the logarithmic fluorescent intensity distribution, noting that P_1 and P_2 are merely semilog plots of F_1 and F_2, respectively. In equation (5), the variables Y_1 and Z' are independent of one another, because they are simply related to X_1 and Z, respec-

---

* As pointed out to the author by Dr. Albert Grunbaum, Dept. of Mathematics, University of California, Berkeley, CA 94704, to whom we are indebted

tively, which are presumed mutually independent by assumption #2.

Thus, the joint distribution of Y_2 is the convolution of the distribution of Y_1, (i.e. P_1), with the distribution of the relative error, (i.e. $A_Z(z)$), expressed mathematically as

$$P_2(y) = \int_{-\infty}^{+\infty} P_1(z')A_Z(y-z')\, dz' \qquad (6)$$

The DNA histogram, with logarithmic abscissa, is a sampled, discrete approximation to P_2.

RECOVERING THE INITIAL DNA DISTRIBUTION: AN INVERSE PROBLEM

The discrete equivalent of equation (6), for unknown underyling log-DNA distribution vector $\overline{P}_1$, log-histogram (vector) $\overline{P}_2$, and convolution spread matrix $\underline{A}$, is

$$\overline{P}_2 = \underline{A}\overline{P}_1 + \overline{e} \qquad (7)$$

where $\overline{e}$ represents the error in the histogram due to finite counting statistics. Recovery of the unknow vector $\overline{P}_1$ amounts to inverting the matrix $\underline{A}$, hence the name "inverse problem." Ignoring the problems of singularity for the moment, let us consider the possible pitfalls to be encountered in attempting to recover $\overline{P}_1$. We shall approach the problem via Fourier analysis.

Let $\hat{p}_1, \hat{p}_2$ and $\hat{a}$ represent the Fourier transforms of P_1, P_2, and A_Z, respectively, let $\hat{y}$ be the Fourier transform variable, and let $\hat{e}$ be the transform of the counting statistical noise distribution, represented as $\overline{e}$ in equation (7). Since convolution of two distributions in real space corresponds to multiplication of their transforms in the Fourier domain, equation (7), including the noise term, becomes

$$\hat{p}_2 = \hat{a} \times \hat{p}_1 + \hat{e} \qquad (8)$$

That is, the Fourier representation of the observed histogram is the transform of the underlying representation, filtered by the transform of the spread function (which attenuates the high frequencies), and corrupted by the addition of counting statistical noise (which should be "white" , i.e., have a constant value throughout the Fourier domain, out to the highest spatial frequency). The generality of the conclusions of this analysis is not reduced by assuming a Gaussian shape for the spread function, whose Fourier transform is also a Gaussian function, but of the variable $\hat{y}$.

In attempting to recover the transform of the underlying distribution, $\hat{p}_1$, we are inverting the convolution, which in the Fourier domain corresponds to dividing both sides of equation (8) by the spread function's transform. Under the present assumption, this amounts to multiplying both sides by the reciprocal of a Gaussian function, which assumes enormous values in the high frequency region of the Fourier domain. This high frequency amplification would probably work, restoring $\hat{p}_1$ to its full frequency representation, were it not for the presence of $\hat{e}$, whose corrupting influence renders the high frequency components of $\hat{p}_2$ highly unreliable. The deconvolution, if carried out to the maximum spatial frequency in the Fourier domain, would result in a reconstructed P_1 made unintelligible by the superposition of overwhelming amounts of high frequency noise, appearing as if finite counting scatter were the main signal, and the function , P_1, were the negligible noise!

As a consequence of this situation, there is a maximum "safe" frequency beyond which the observed transform, $\hat{p}_2$, should not be amplified in the deconvolution, lest high frequency noise corrupt the reconstruction. One possible choice for such a frequency limit would be that point where the amplitude of the "white" noise of the counting statistics is approximately equal to that of the filtering spread function transform, since at that point, the ratio of signal to noise approaches unity. One way to view the message of this analysis is that the extent to which the underlying distribution can be recovered is governed by the width of the spread function, and by the number of cells counted, since these two values will determine the maximum "safe" frequency in the deconvolution. A broader and even more significant conclusion to be drawn from this is that the statistical reliability of the reconstruction is roughly inversely proportional to its resolution (i.e., the maximum frequency used).

The above results, obtained for the case of Fourier deconvolution methods, have been generalized by Backus and Gilbert (1) to apply to any inversion method, as a consequence of the fundamental nature of the inverse problem. Stated in our terminology, their theorem holds that regardless of what model is used for the underlying DNA distribution, the product of the number of free model parameters used and the statistical reliability of their solution values is constant, and that constant is a measure of the information content of the data. That is to say, a given histogram, with a given CV and total cell count, can only justify a certain number of free parameters in the model, if their accuracy is to be held within a tolerance demanded by the user.

As a further consequence of the above discussion, it can be seen that in the presence of counting statistical noise, $\overline{P}_1$ cannot be recovered uniquely. Even more disturbing is the fact that significant changes in the model of $\overline{P}_1$ are barely reflected in the fit to the data points, meaning that there is a large set of models that fit the data equally well to within any given statistical tolerance. Stated conversely, small uncertainties in the data are magnified to large uncertainties in the solution set, with the magnification depending upon the number of free parameters in the model, all else being equal. As the Fourier analysis showed, the use of a limited number of free model parameters assures a certain statistical confidence in the values obtained for them. This fact, along with any a-priori knowledge about P_1, makes inversion a viable option in the analysis of DNA histograms. A thorough analysis of the recovery of the DNA distribution from its FCM histogram, from the standpoint of inverse theory, would be of much value to the practical use of these methods, since it should be able to quantitate the relation between number of parameters and statistical reliability.

Returning to the recovery problem, the desired quantity is the unknown vector $\overline{P}_1$, which, as a rule, has fewer elements in it that the histogram $\overline{P}_2$. The set of equations defined by (7) is thus overdetermined, and the matrix $\underline{A}$ has no inverse. This is of no consequence to the conclusions reached above, however, since direct inversion is not the method of choice in any case. The greater the degree to which the equations are overdetermined, however, the more reliable the values obtained for $\overline{P}_1$, as this corresponds to using a restricted number of free parameters relative to the number of observations.

Heuristically speaking, it is as though each free parameter enjoyed the advantage of several independent observations in the determination of its value, which for noisy observations reduces the resulting uncertainty.

One final important point on the inverse nature of the problem at hand. Not only do all the above difficulties plague us, but even the spread matrix $\underline{A}$ is subject to error, since our extracted model of the spread function is only an approximation thereto. We have not investigated the extent to which this further muddles the theory, but we have shown how variations in the assumed spread shape lead to significantly different solution sets, all other factors being held constant. (discussed below, in Implementation Details) This formulation of the relation between the logarithmic underlying DNA distribution, the relative error distribution (spread function), and the resultant log(fluorescent intensity) distribution suggests the following approach for obtaining information about $\overline{P}_1$ (details discussed below):

1. Convert the histogram to logarithmic abscissa ("log-histogram")
2. Estimate the spread function shape
3. Establish a parametric model of $\overline{P}_1$
4. Generate an approximation to $\overline{P}_2$, given the

latest parametrization of the model for $\overline{P}_1$, by convoluting the model with the spread function

5. Optimize the parameters of the model for $\overline{P}_1$ by least squares fitting the log-histogram with the approximation to $\overline{P}_2$.

Other authors have suggested similar approaches (2), but the principal difference here is the logarithmic transformation of the histogram abscissa, and its implications. The principal advantage of this formulation is the potential ability, given a reasonably clean G_1 cell population, of obtaining a direct determination of the spread function shape, since the underlying DNA distribution of such a population would be essentially a delta function (that is, possess a variance that is below the resolution of any FCM system available to date). This eliminates the need to assume a Gaussian, or any other analytical shape for the overall spread function, and we feel that this keeps the analysis somewhat closer to the actual situation.

IMPLEMENTATION OF APPROACH

LOGARITHMIC TRANSFORMATION

For a given histogram, an initial estimate of the G_1 peak mean of the raw data is obtained by maximization of the area within a "window" of 7 channels width, swept over the region surrounding the apparent peak mode. The window eliminates the effects of local scatter. We generally choose to then allign this estimated mean in channel 100, by a linear transformation of the abscissa, simulating a gain adjustment of the amplification electronics during the experiment to achieve the same result.

The logarithmic transformation of the histogram is done by choosing the arbitrary base "b", mentioned above, so that the change of variables leaves two abscissa points, a_1 and a_2 , in the same channels after the transformation as those which they previously occupied. The appropriate relation is:

$$Y_2 = a_1 + \left[\frac{\left[a_2 - a_1\right]}{\ln\left[\frac{a_2}{a_1}\right]}\right] \ln\left[\frac{X_2}{a_1}\right] \tag{9}$$

where Y_2 and X_2 are the log and linear observed intensities, respectively. We let a_1 and a_2 be the estimated G_1 and G_2M means, respectively, so as to maintain the same number of channels in the S-phase region in log domain as in linear. Using equation (9), a new abscissa is defined, and the former boundaries of the bins in the linear domain are mapped into new boundary values, generating a new grid on which to digitize the histogram. Since the new boundaries will virtually never coincide with the old, some form of interpolation is needed for the redigitization, and we employ Simpson's rule for this purpose.

ESTIMATION OF SPREAD FUNCTION SHAPE

By assumption #4, above, one can estimate the general spread shape for a set of histograms from the cleanest G_1 peak available in the set, the "reference G_1 peak," and use it to fit the rest of the set. Once extracted from the reference G_1 peak, the shape is expanded or contracted for the particular histogram under analysis, to match its estimated CV. If no peak is sufficiently clean to serve as a reference, one can either use the most recent previous reference peak, or employ a simple Gaussian model in the logarithmic abscissa (log-normal in the linear abscissa).

Extraction of the spread function from a typical reference peak is illustrated in Fig. 2. The wings of the G_1 peak are presumed to be corrupted by the presence of debris and early-S-phase cells, and the entire peak is subject to the usual counting statistical scatter. We first convert the histogram to log-form, as described above. We then smoothe the peak by convolution with a Gaussian distribution of standard deviation equal to one channel (our histograms have 256 channels), then find the mode of the peak. Left and right "standard deviations", SDL and SDR, are estimated by finding the first points to the left and right of the mode, respectively, at which the height is 60% of modal height. To compensate for non-G_1 content in the peak wings, heights are determined at 3xSIGL and 3xSIGR to the left and right of mode, respectively, and a threshold is set to twice the larger of these two sampled values. The reference histogram is thus divided into three parts: the central domain about the G_1 peak in which the threshold is exceeded; channels to the left of the domain, and those to the right of it. The entire smoothed histogram is then multiplied by a window filter whose central portion, corresponding to the domain described above, has a value of unity, and which trails off to the left and right of the domain as Gaussian functions with standard deviations SIGL and SIGR, respectively, and which are centered on the respective endpoints of the domain. In this manner, the wings of the G_1 peak are smoothly suppressed to zero, while the main portion of the peak is used directly as the spread shape. Finally, the resulting shape is normalized to unit area.

MODELING THE UNDERLYING DNA DISTRIBUTION

If the notion is accepted that there exists a separate phase of the cell cycle during which DNA replication is accomplished, Fig. 3a, the relative amount of DNA per cell during the cycle should behave as illustrated in Fig. 3b, for homogeneous, euploid cell populations. We shall assume, for the present, that we are dealing only with such populations, though the same principles would apply to heterogeneous and/or aneuploid populations, albeit with greater difficulty. For a population arbitrarily distributed in the cycle as shown in Fig. 3c, the corresponding histogram of relative cell number versus relative DNA per cell would appear as in Fig. 3d, with the log-transformed equivalent shown in Fig. 3e. Thus, the model of the underlying DNA distribution which most closely matches the above scheme would have delta functions of unknown height at the G_1 and G_2M means, bounding a region characterized by a rather smooth, probably continuous function which can assume virtually any shape. One good representation for the latter would be a piecewise polynomial spline. However, we wish to emphasize that no matter what the model, the total number of free parameters will be limited by the above considerations regarding the inverse nature of this problem.

If the suggested spline representation is employed, the problem of optimization (discussed below) is a nonlinear one, and likely to be rather expensive to implement. Other authors have attempted such nonlinear fits to DNA histograms (7), though they have employed a single polynomial over the entire S-region. The attempt to span so large a domain with a single polynomial almost inevitably results in having to use a low order, since higher orders result in increasingly larger random oscillatory behavior by the function, as proven by a theorem of De la Vallee-Poussin (3). With only a third or fourth degree polynomial to describe all of S-phase, these approaches are limited to those histograms that do not possess a complicated S-phase structure. The spline would, of course, allow for virtually any conceivable structure, and would thus be far more generally applicable. However, for the reasons outlined below, we have elected to employ a discontinuous model for S-phase, rather than the spline.

In his analytical model, Fried (2) approximates the observed histogram by a sum of Gaussian functions, which are the assumed spread shape. Their means comprise a uniformly spaced lattice between, and including, the G_1 and G_2M peak means, and their standard deviations increase in proportion to their means. This formulation is meant to correspond to a subdivision of S-phase into compartments of equal width (the same as the Gaussian lattice spacing), between delta functions at G_1 and G_2M. The overall intended model should thus be a "bar graph" of the underlying S-phase, with spikes at G_1 and G_2M. A given Gaussian is intended to simulate the allocation , throughout the histogram, of cells whose DNA content lies in the compartment centered on the Gaussian in question. Fig. 4 illustrates this "spikes & bars " model, for logarithmically transformed histograms.

In reality, the sum of Gaussians approach corresponds to a "comb function" model of the underlying DNA distribution, that is a lattice of uniformly spaced delta functions, rather than a "bar graph" model. These are illustrated in Fig. 4. This follows from the fact that the approximation to the data is essentially a sum of weighted spread functions. There are two key differences between these models, though both have spikes at G_1 and G_2M. The more important of the discrepancies has to do with the representation of the most critical regions of the histogram: G_1/early-s, and late-S/G_2M.

If the equipartition of S-phase were implemented as intended by Fried, the left boundary of the first compartment would be the G_1 mean, and thus the center of the compartment would be one half of a width to the right of G_1 mean. This situation is illustrated in Fig. 4b. The centers of succeeding compartments would each be one width to the right of the preceeding one, including the last compartment, whose mean would be one half-width to the left of the G_2M mean. Since the ultimate distribution of cells arising from each compartment would be centered on the compartment mean, the Gaussians intended to model this process should also be thus centered. Placement of the first S-phase Gaussian a whole compartment width away from the G_1 mean leaves the early part of S-phase under-represented, and the analgous situation pertains to the late-S/G_2M region as well, as seen in Fig. 5b.

The second discrepancy between models is the use of the same functional shape (that of the presumed spread function), to describe the ultimate distribution of cells arising from compartments (or bars) and from spikes (as in G_1 and G_2M). Rigorously speaking, the former distribution should be the convolution of the spread function with a bar, and the difference between this and the spread function itself is entirely dependent upon the width of the bar, a subject addressed below. However, for bars whose width is less than about twice the standard deviation of the Gaussians being used (as in Fried's case), the difference is not enough to warrant the rather extensive additional calculation necessary to implement the precise distribution in the fitting routine, as opposed to using the spread functions themselves.

There is a further difficulty in the use of uniform lattices over any portion of linear histograms, even if the half-widths are properly implemented, and that is the fact that the spreading at the G_2M side is about twice as bad as that near G_1. Consequently, one should not expect to be able to resolve the right side of the histogram as well as the left, and attempts to do so will result in larger statistical uncertainty in the values for the right side Gaussians than for those on the left, making it difficult to compare them on an equal basis. This problem is eliminated by the log transform, since in the log-histogram, the spreading is presumed to be the same everywhere (and this appears to stand up fairly well in the face of extensive testing), and so uniform logarithmic spacing is statistically justified. One could approximate this solution in the linear domain by positioning the spikes at unequal intervals of exponentially increasing width as one proceeds from G_1 to G_2M, while stretching each spread function so that its standard deviation is proportional to its mean.

In our model, we use the lattice spacings of the bar graph approach, illustrated in Fig. 5. but approximate the bar contributions by the spread functions themselves (for the time being). Since we use log-histograms, the uniform-sized bars in the model correspond to compartments of exponentially increasing width in the linear representation for S-phase. The details for locating the G_1 and G_2M means, estimating the proper spread width, and choosing the proper number of spikes for a given histogram are disucussed in the section "IMPLEMENTATION: DETAILS".

GENERATING AN APPROXIMATION TO THE DATA

The discrete form of equation (6), given as equation

(7) is

$$\overline{P}_2 = \underline{A}\overline{P}_1 + \overline{e}$$

where $\underline{A}$ is the spread matrix that represents the discrete convolution of P_1 with the spread function. If $\overline{P}_1$ and $\overline{P}_2$ are column vectors, of lengths N and M, respectively, N > M, the j^{th} column of matrix $\underline{A}$ is filled with the spread function, centered on the same channel as the j^{th} spike location in the model lattice. This implements the desired effect: a spike of amplitude $P_1(j)$ is spread about the j^{th} spike location in the model to the histogram, with the final approximation summing up the contributions of all spikes. The unknown parameters in the model are the spike amplitudes or weights, which enter the approximation in a linear fashion, thus requiring only a linear optimization routine to solve for the best spike set, in the least squares sense discussed earlier.

OPTIMIZING THE MODEL PARAMETERS

Once the problem has been cast in the matrix equation form of equation (7), the only remaining difficulty is the avoidance of solutions containing negative spike amplitudes. We overcome this aspect by using the Non-Negative Least Squares (NNLS) routine published by Lawson and Hanson (4), which finds the optimal least squares solution set subject to the constraint that all its elements are non-negative.

Nonlinear models, such as the spline mentioned earlier, would require vastly more costly optimization algorithms, but if desired, the National Physical Laboratory of Great Britain has an excellent nonlinear optimization routine, including non-negativity constraints, that would be quite adequate for the purpose.

IMPLEMENTATION: DETAILS

As mentioned above, the principle missing quantities in the implementation of our model for a given histogram are estimates for the G_1 and G_2M means and appropriate standard deviation of the spread process, and the proper number of spikes. A given set of these input values corresponds to a unique best least squares fit and solution set for the log-histogram in question, with significantly different solution sets for different input values. Since our principal use for these analytical methods is for the reduction and comparison of series of DNA histograms, the dominant concern has been that whatever methods are employed for the estimation of the above input values, they should be as unbiased and reproducible as possible, while giving results that are believable to the investigator.

We shall first address the question of the standard deviation estimate for a given histogram. We made a serious attempt to model the G_1 peak of the log-histogram using a Gaussian model, fitting it to the left slope of the peak (where corruption from S-phase cells is least felt) using a nonlinear least squares routine developed by Golub, et al (5). However, the resuls rarely appeared credible to the user, and the addition of the capability to choose different portions of the peak to be fit improved the situation only slightly. One clear message from this experience is that, at least for our data from many different cell systems, the spread process is not very well modeled by a Gaussian distribution in the log domain (or a log-normal distribution in the linear domain). Thus we have, for the meanwhile, resorted to simple estimates of the standard deviation (SD) by finding the apparent peak mean, and letting the estimate be the distance to the left at which the peak falls to .606 of the maximum value, using linear interpolation. In general, we are confident that these estimates are within ±10%, of the proper values, at the worst. This uncertainty can significantly affect the fits, and to a lesser extent, the estimated cycle phase fractions.

Examples of the effects of varied SD estimation upon the resulting fits are shown in Figs. 5a,b,c. We have found that increasing the estimated SD from 8.0 channels to 10 channels, for example, has a significant effect upon the quality of the fit, and though the general properties of the solution sets remain intact, the cycle fraction estimates change by about 10 to 20% of their previous values. Fig. 5b illustrates typical effects of such an increase. We should note that, as discussed below, a changed SD estimate requires a changed number of spikes in the model, and the effects cited here are despite this accommodation. Shown in Fig. 5c are the effects of reducing the estimate of SD from 8.0 to 6.0 on the same histogram, and Fig. 8d summarizes these effects. Though the effects of SD uncertainty vary from one histogram to the next, one can say, as a crude rule, that a large increase of N% in the SD will yield a roughly equivalent change of N% in the cycle fraction values, with G_1 and G_2M increasing at the expense of S-phase. An equivalent reduction in SD will produce similar effects, but with S-phase increasing at the expense of G_1 and G_2M fractions. We use SD values in the large end of the "acceptable range" indicated in 8a, to give spike height estimates with the maximum statistical reliability commensurate with yielding a good fit. The latter property is indicated by a relative chisquare value of between 0.1 and 0.3, as is seen in 8a.

In the estimation of G_1 and G_2M means, it has been pointed out that one must not rely on the apparent peak mode as an estimator, but must compensate for its having been shifted towards histogram center by the presence of S-phase cells. We have simulated this process by computer, and have found that the amount of shift depends not only on the amount of S cells in close proximity to the peak in question, but also on the detailed shape of the spread function (not only its width). This was seen from the use of a Gaussian spread function in simulating the effects of various amounts of early S-phase cells, which produced a model relating the estimated shift to the ratio of early-S histogram height to G_1 peak height. When applied to our data, this model predicted unreasonably large shifts for virtually every histogram, which we attribute to significant discrepancies between the true spread shape and the Gaussian model thereof. We have yet to try to develop a shift model based on extracted spread functions.

We have tested the effects of changing the estimated G_1 and G_2M means by fractions of a channel, when the G_1 is originally in channel 100, and found them to be negligible. Hence, our estimates of means are aimed at determining the values to the nearest channel, which reduces the set of possibilities to that of a few different channel shifts away from the peak mode. Therefore, we employ a simple classification scheme, in which we estimate the influence of early S-phase cells upon the G_1 peak, and that of late-S upon G_2M, as follows. After log transformation, the modes of the peaks are found and the SD is estimated. Stepping a distance of 2 SD inward from the respective peak, we determine the ratio of histogram heights at the early- or late-S point to that of the G_1

or G_2M peak, respectively. For each peak, we classify the ratio into four empirically determined categories:

0 - 30% set mean = mode

30 - 50% set mean = mode ±1 channel, for the G_1 or G_2M means, respectively

50 - 80% set mean = mode ±2 channels (" " ")

>80% set mean = mode ±3 channels (" " ")

Mode is estimated by the 7 channel window maximization method described in the spread function estimation section, above. We find that this mean-estimation scheme works very well, producing solution sets that are quite reasonable on a wide variety of data.

The number of spikes to be used, as discussed in the mathematical formulation section, above, depends on the estimated SD of the histogram. We find that when the inter-spike spacing in S-phase is less than about 1.3 times the estimated SD, the spikes show strong correlation with one another, resulting in oscillatory behavior as shown in Fig. 7a. When the spacing exceeds about 1.7xSD, the fit becomes poor, slowly oscillating about the histogram points, as seen in Fig. 7b. Inter-spike spacings within these limits yield acceptable solution sets and fits, with 1.5xSD apparently the best, as shown in Fig. 7c. However, since the spikes must lie on a lattice that includes the G_1 and G_2M means, the best one can do is to determine that number of spikes, M, which when arranged in the lattice yields inter-spike spacing closest to 1.5xSD.

RESULTS AND DISCUSSION

Results of a typical fitting session are shown in figures 9 and 10. 9 a,b,c show the spread function extraction process. In 10a, the spread function is used to fit the reference histogram from which it was extracted. 10 b,c,d are other histograms from the same experiment as the reference, fit using its spread function.

One method for improving the set of input parameters estimates might be to recursively fit the histogram and re-estimate the parameters. This notion is well illustrated by its application in extracting the spread function from the reference histogram, which we are currently implementing: Extraction proceeds as described above, resulting in a proposed spread function shape. This shape is then used to fit the reference histogram, and the resulting model of early S-phase is used to modify the estimated shape. The latter step involves subtraction of the estimated early-S-phase and sub-G_1 debris from the reference histogram, followed by re-application of the isolating window previously used, but now to the modified reference histogram. This approach is currently being tested, and initial results are encouraging. It is not clear whether repeated recursion, using this method, would be of advantage, since there is no obvious way to tell when the true spread shape has been reached, and the recursion is likely to result in significant modification at each iteration. Whether such a scheme would converge at all, and if so, to anything meaningful, remains to be studied. However, single iterations of such recursive approaches appear to be worthwhile, and could be used for input parameter estimation in an analgous manner to that just described.

FIGURES

1. Flow of Information during FCM Analysis. Cells in vivo or in culture possess a given distribution with respect to DNA content per cell, prior to FCM processing. The DNA/cell distribution is converted to a fluorescent stain content per cell distribution by the harvesting, fixation and staining steps, which involve errors from various sources indicated in the figure. The process of rapid measurement of stain content, as determined by the emitted fluorescent intensity per cell, involves further errors, resulting in a final distribution with respect to measured fluorescent intensity per cell. The final distribution is a degraded version of the original DNA/cell distribution, with the extent of spreading proportional to the original amount of DNA in a given cell.

2. Extraction of Spread Function Shape from Reference G1 Peak. A histogram possessing a G_1 peak and lacking significant early-S cells is chosen as reference and redigitized on a logarithmic abscissa. An "acceptance window" is determined from the wings of the G_1 peak, as described in the text, and the mean, the left and right standard deviations of the peak are estimated. The entire histogram is multiplied, point for point, by a filter with height of one within the window, and dropping off to the left and right of the window as Gaussian functions with left and right standard deviations determined above, respectively. The result is an extracted estimate of the spread function whose wings tend smoothly to zero, and prior to use it is normalized to unit area. Only histograms from the same cell line, FCM processed the same day as the reference, should be analysed using this extracted spread function.

3. How DNA histograms relate to the cell cycle.
   a) Basic cell cycle model
   b) DNA per cell versus Cellular Maturity based on cell cycle model of 3a.
   c) Cell No. versus Age (arbitrary distribution)
   d) Corresponding distribution of Cell No. versus DNA/cell
   e) Corresponding distribution of Cell No. versus log(DNA/cell) By the model shown in 3a, DNA content

per cell should correlate with cellular maturity as shown in 3b. Cells distributed in the mitotic cycle as shown in 3c (arbitrary distribution), should yield the distribution with respect to DNA per cell shown in 3d, and with respect to logarithm of DNA content per cell as shown in 3e.

4. Spikes and Bars Model of Underlying DNA Distribution. The underlying DNA distribution is modeled as a step function (bar graph) in S-phase, bounded by two delta functions (spikes) at G_1 and G_2M, respectively. In the approximation to the observed histogram generated from this model, the spikes become weighted spread functions, while the bars become weighted versions of the convolution of the spread function with a bar of unit height. The separate functions are summed in the final approximation to the histogram data.

5. Comparison of "Comb" and "Spikes and Bars" models of underlying DNA distribution. In the "comb" model, equally spaced spikes become weighted spread functions, also equally spaced, in the approximation to the histogram. In the "spikes and bars" model and resulting approximation to the histogram, the contributions arising from the first and last bars are spaced only half a bar width toward histogram center from the respective G_1 or G_2M peak. Hence, "spikes and bars" accommodates S-phase cells everywhere

between G_1 and G_2M in the underlying DNA distribution model, so that G_1, S, and G_2M phases are properly represented. The corresponding contributions in the "comb" model are spaced a full bar width away from the respective peaks, so that S-phase cells with DNA content lying between G_1 and the first spike, or between the last S-spike and G_2M, are not represented in the model. This results in a low estimate of population S-phase fraction with the "comb" model.

6. Effects of Varying Estimated Standard Deviation (SD) upon Solution Set and Fit to Histogram.

   a) Fit using original estimate of SD = 8.0

   b) Fit using SD = 10.0 (increased by 25% over original)

   c) Fit using SD = 6.0 (decreased by 25% under original) In all three cases, inter-spike spacing was maintained constant

   Fit seen in 6b is clearly inferior to that of 6a or 6c, while solution spikeset of 6c oscillates erratically in early-S. Hence, solution spike set of 6a judged best, because spike heights in mid-S are even, and drop off monotonically near the G_1 and G_2M spikes, while fit to histogram is statistically good (reduced chisquare = 0.14).

7. Effects of Varying Inter-Spike Spacing in DNA Distribution Model.

a) Spacing too small: <1.2 SD

b) Best Spacing: 1.5 SD

c) Spacing too large: >2.0 SD SD maintained constant at 7.5 channels in all three cases. Spikes in solution set of 7a oscillate erratically due to insuf- ficient spacing, while the approximation in 7c oscillates about the histogram because of excessive space between spikes. 7b shows good fit to histogram with spikes of even height in mid-S phase, using a spacing of about 1.5 SD.

8. Summary of effects of input parameter variation upon model and fit.

a) Summarized effects of SD variation upon model and fit. Effects of varied SD estimate upon reduced chisquare of fit, and upon cycle phase fractions FS, FG1, and FG2M, for the same histogram, with inter-spike spacing held 1.5 SD. All observed parameters were constant over the range 6.5 < SD < 8.0. The larger the SD used, the fewer the number of spikes, and hence the simpler the model, so the largest acceptable SD is the one of choice.

b) Summarized effects of Inter-Spike Spacing variation. Effects of varied spacing between model spikes, for the same histogram as in 8a, upon cycle phase fractions FG1, FS, and FG2M, and upon reduced chisquare of the fit, with SD held constant at 7.5 channels. Spacing variation has little effect on cycle phase fractions, but changes the quality of the fit. From the histograms, the range of spacing which produces acceptable models and fits is 1.4 SD < spacing < 1.75 SD.

9. Start of a fitting session: extraction of spread function.

a) reference histogram: few cells seen in early-S phase, clean G_1 peak.

b) smoothed, log-transform of reference histogram.

c) normalized spread function extracted from 9b.

10. Fits to assorted histograms using extracted spread function of figure 9c:

a) Fit of reference histogram using its own spread function. Note that G_1 peak is perfectly fitted using only one spike, while G_2M peak is slightly narrower that the fitting approxi- mation, showing that the width of the spread function is nearly, but not precisely, constant throughout the histogram.

b) Histogram possessing large mid-S-phase cohort. Note sub-G_1 spikes necessary to accommodate apparent debris in left shoulder of G_1 peak. These are not counted in the cycle phase fraction calculations.

c) Histogram with large early-S-phase cohort. Note additional post-G_2M spike attempting to account for presence of aggregates seen in the histogram.

d) Histogram with large late-S-phase cohort

REFERENCES

1. Backus, G.E., Gilbert, F. Uniqueness in the Inversion of Inaccurate Gross Earth Data. Phil Trans Roy Soc London A 266:123. (1970)

2. Fried, J. Analysis of Deoxyribonucleic Acid Histograms from Flow Cytofluorometry; Estimation of the Distribution of Cells Within S-Phase. J Histochem Cytochem 7 (25):942-951 (1977)

3. Isaacson, E., and Keller, H.P. De la Vallee-Poussin Theorem. in Analysis of Numerical Methods, p. 223. John Wiley & Sons, Inc. New York, 1966.

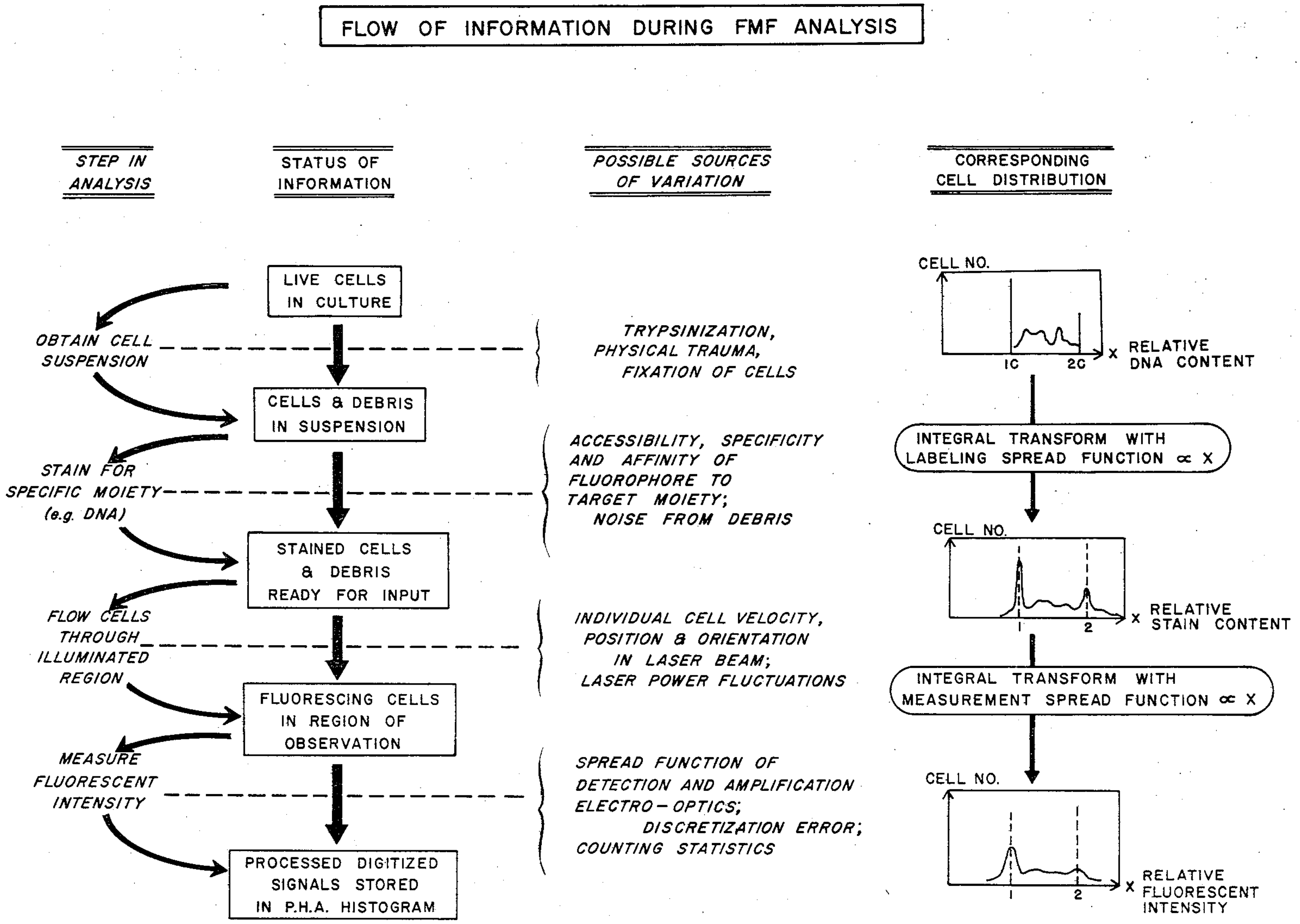
4. Lawson, C., Hanson, R. Problem NNLS. in Solving Least Squares Problems, Prentice Hall, Chapter 23. (1974)

5. Golub, G. Pereyra, V. The Differentiation of Pseudo Inverses and and Nonlinear Least Squares Problems whose Variables Separate. SIAM J Numer Anal (10) 1973

6. Dean, P.N., Jett, J.H. Mathematical Analysis of DNA Distributions Derived from Flow Microfluorometry. J Cell Biol 60:523 (1974)

7. Zietz, S., and Nicolini, C. Flow Microfluorometry and Cell Kinetics: A Review. in Biomathematics and Cell Kinetics, ed. A.J. Valleron. Elsevier/North

Holland (1978)

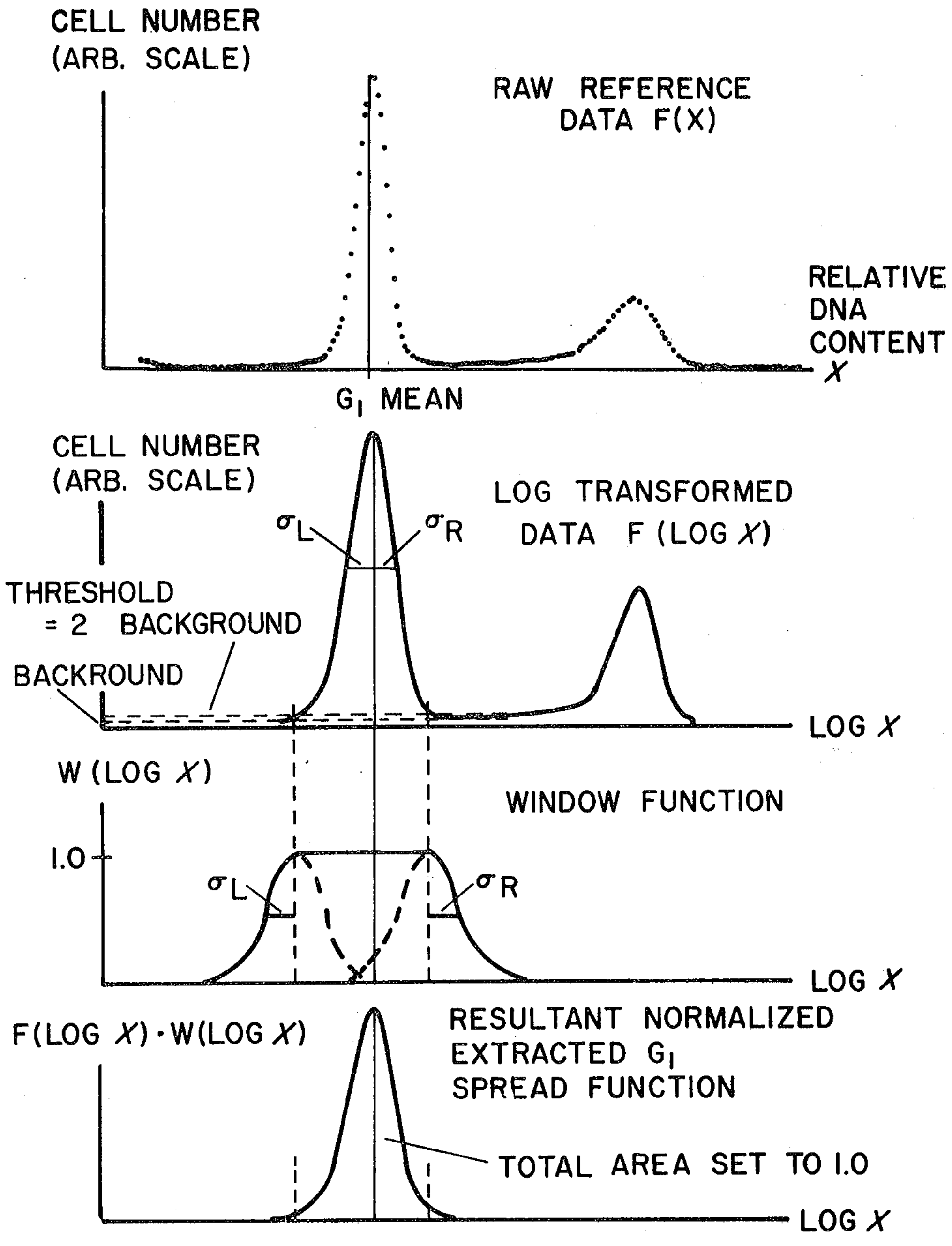


XBL 7612-10786

Figure 1

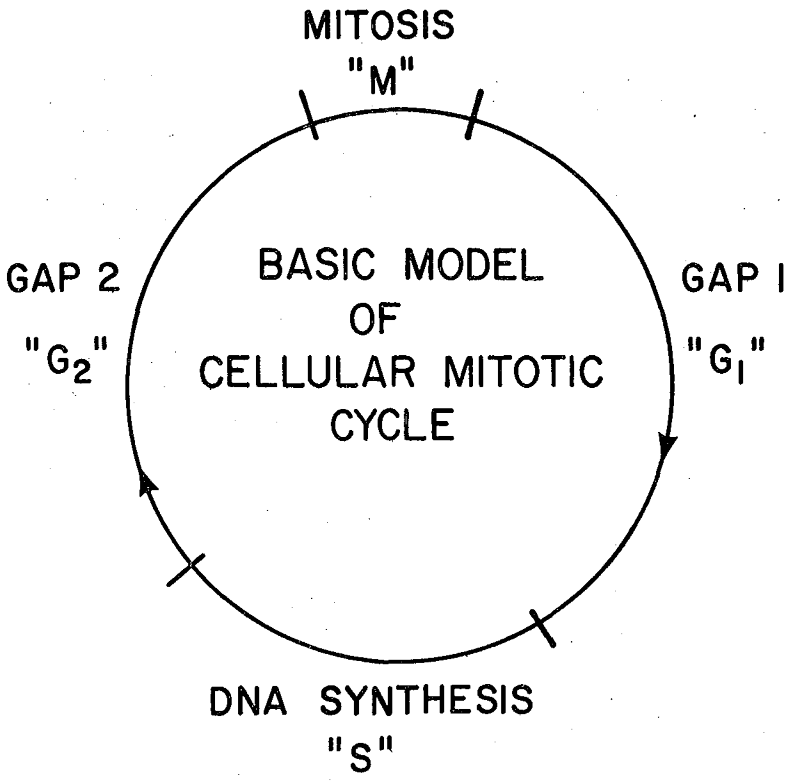
Figure 2

EXTRACTION OF NORMALIZED G_1 SPREAD FUNCTION FROM REFERENCE HISTOGRAM

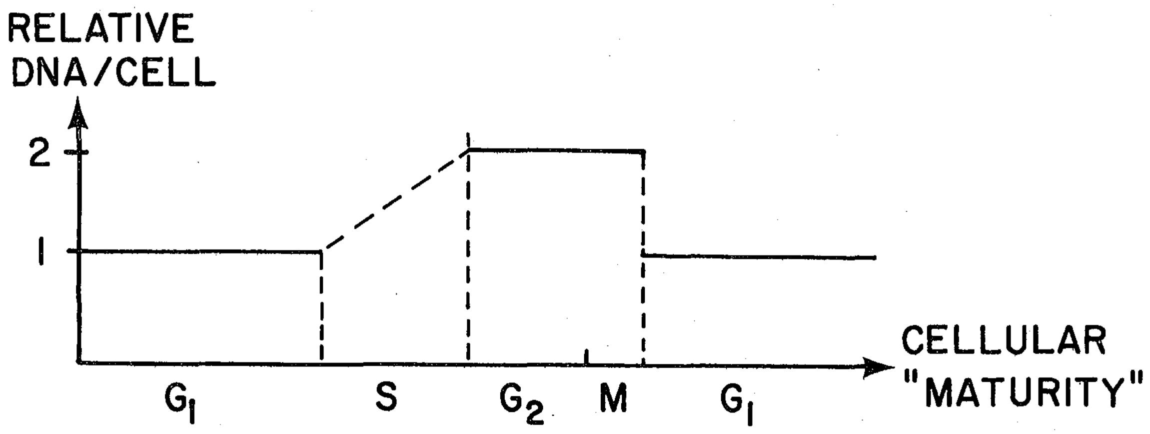


XBL7712-4787

Figures 3a,b

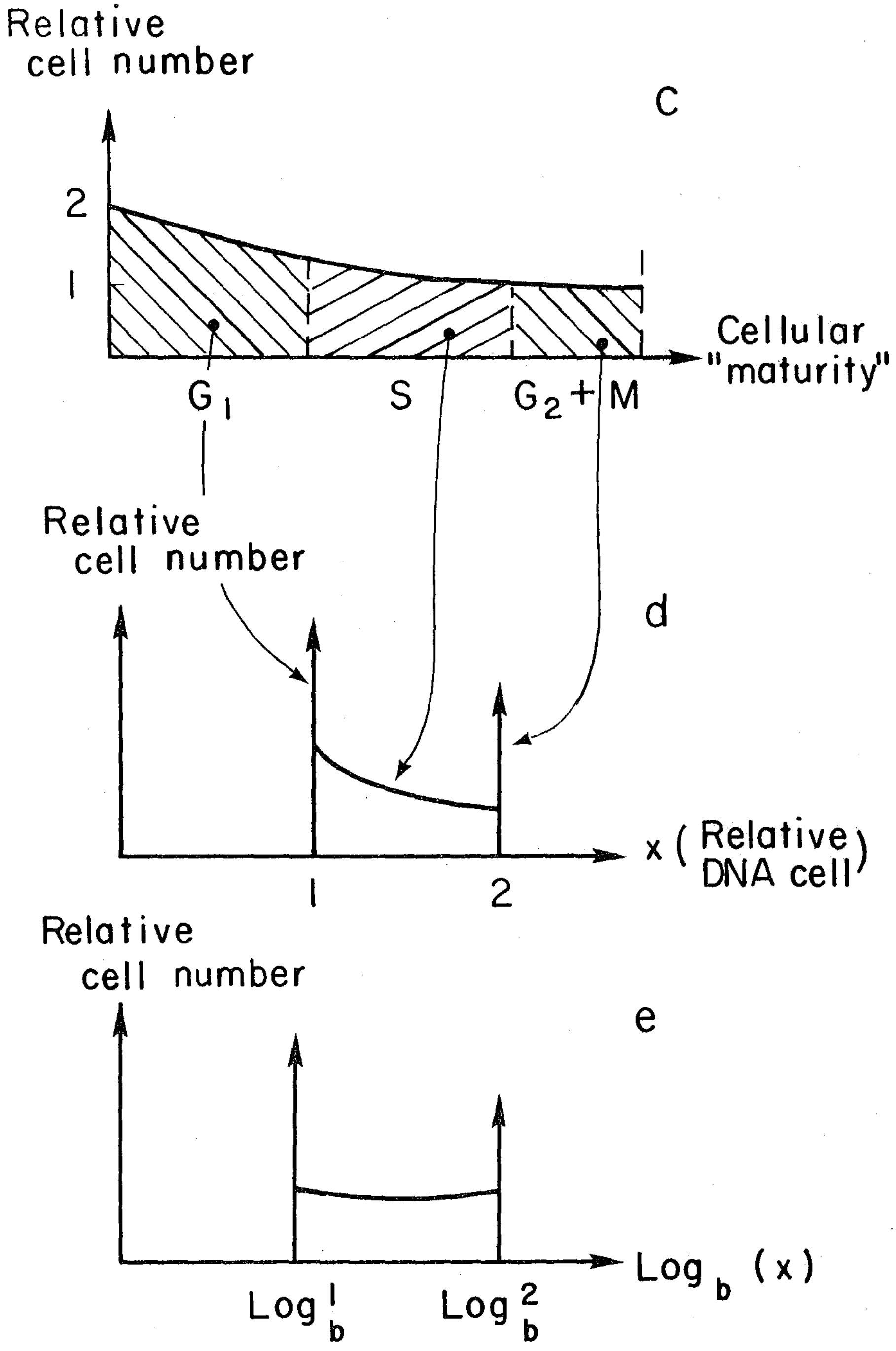


RELATIVE DNA PER CELL AS FUNCTION OF CELLULAR "MATURITY"



XBL 7712-4790

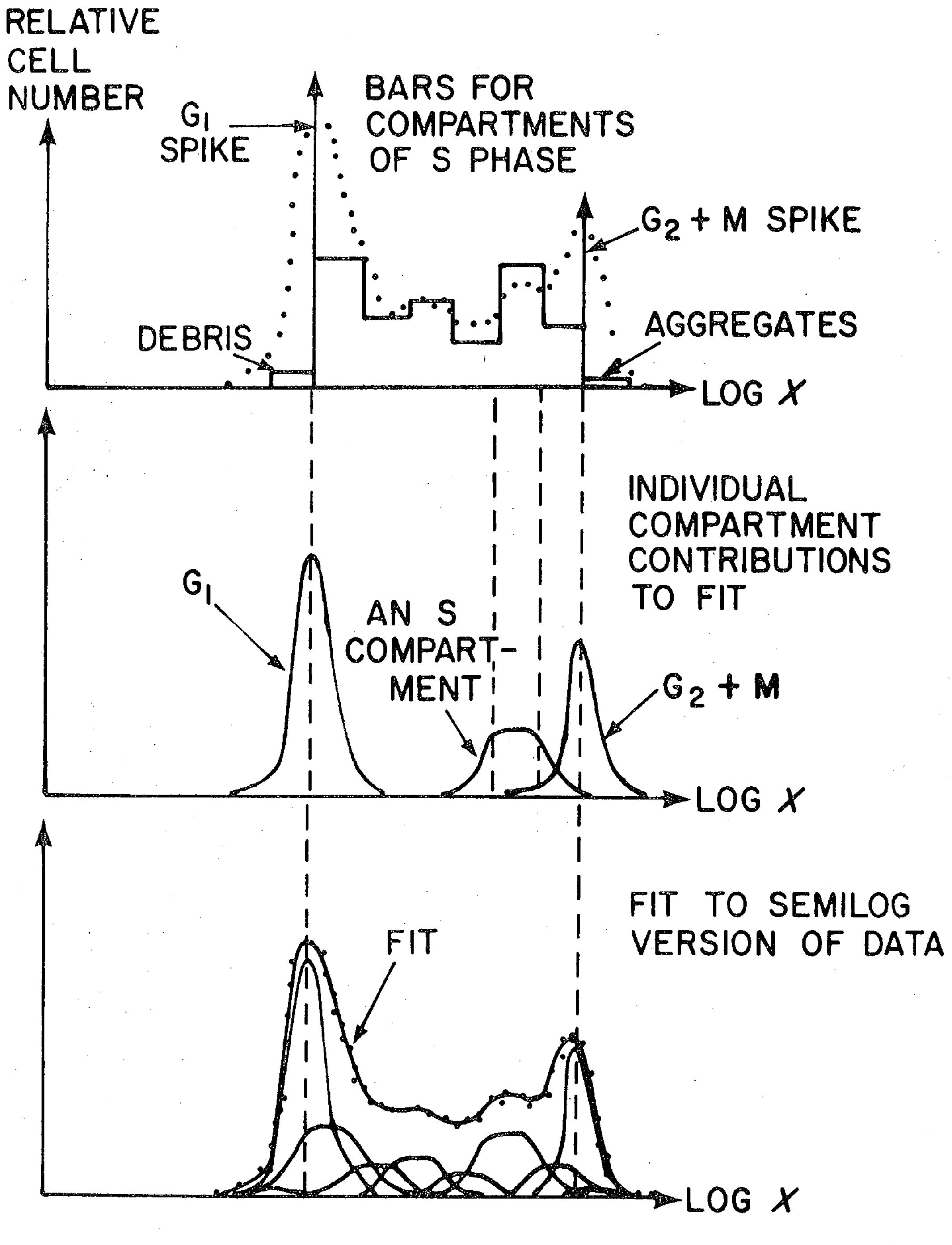
Figures 3c,d,e



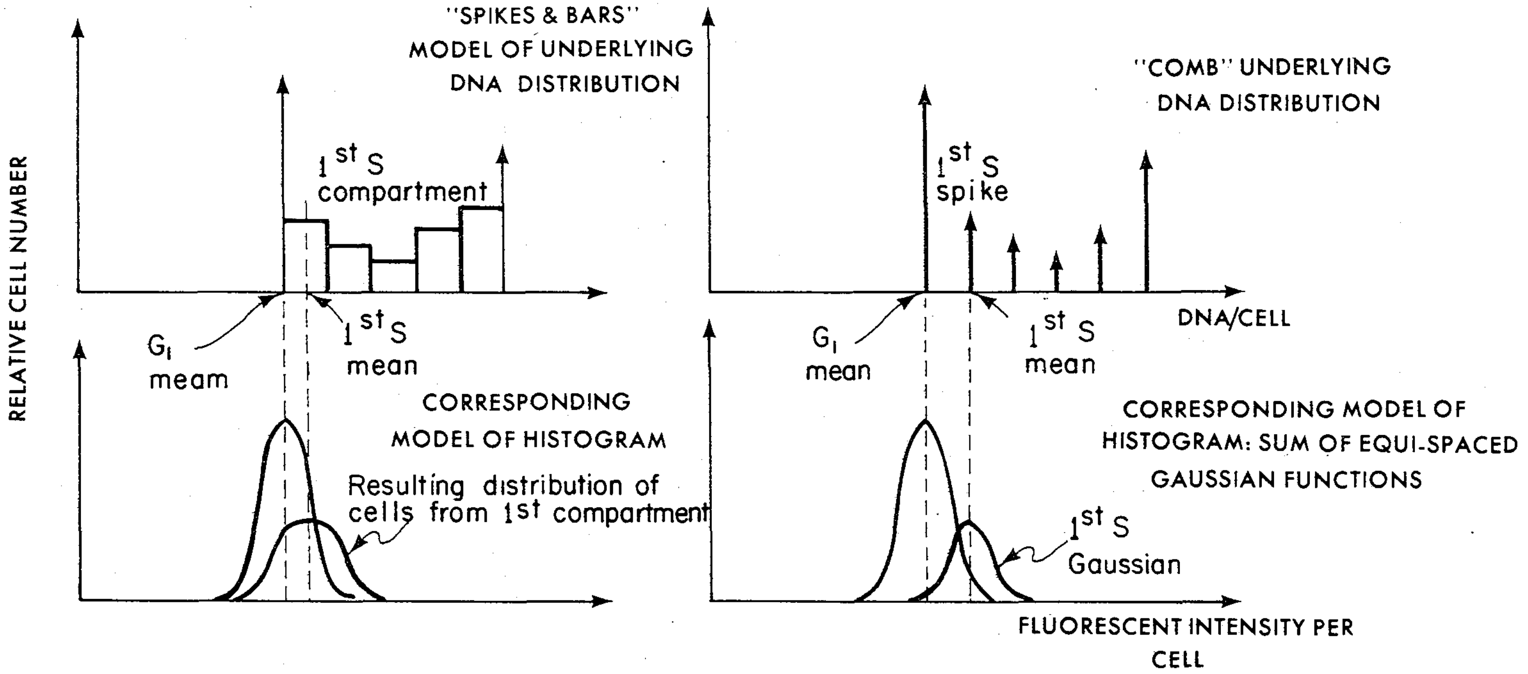
XBL778-1663

Figure 4

SPIKES AND BARS MODEL OF SEMILOG DATA

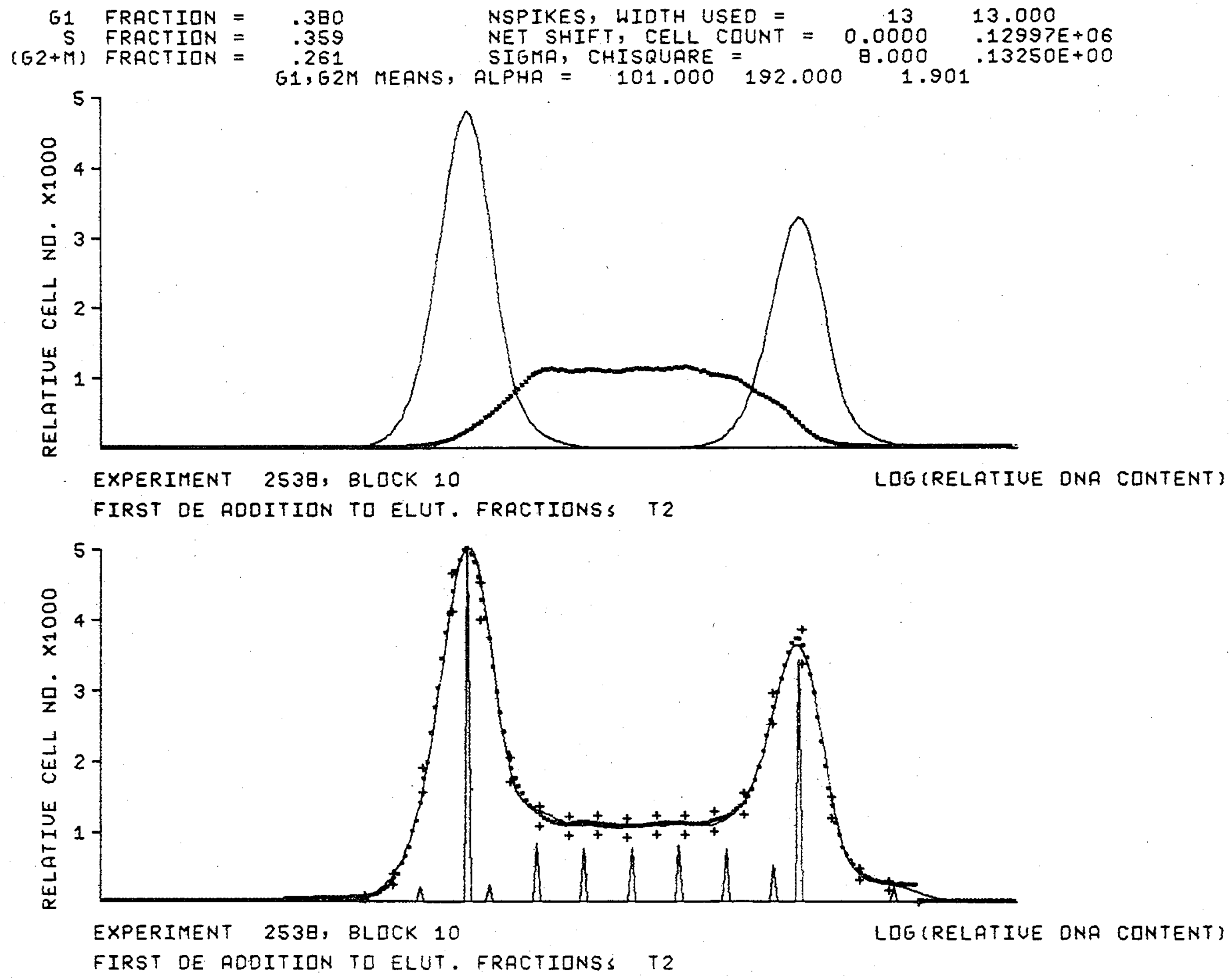


Comparison of two DNA distribution models



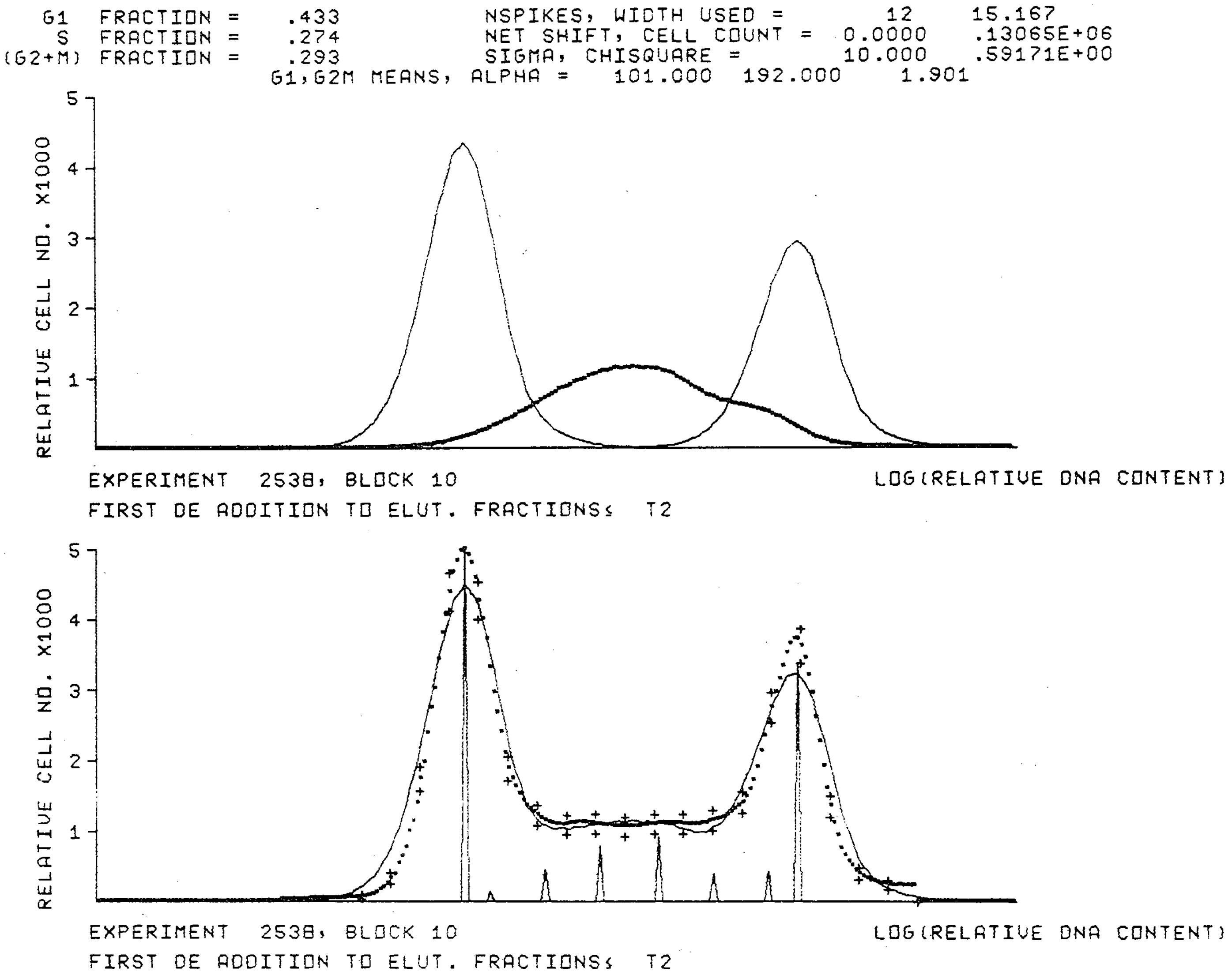
XBL788-1661

Figure 5



XBL 788 - 4107

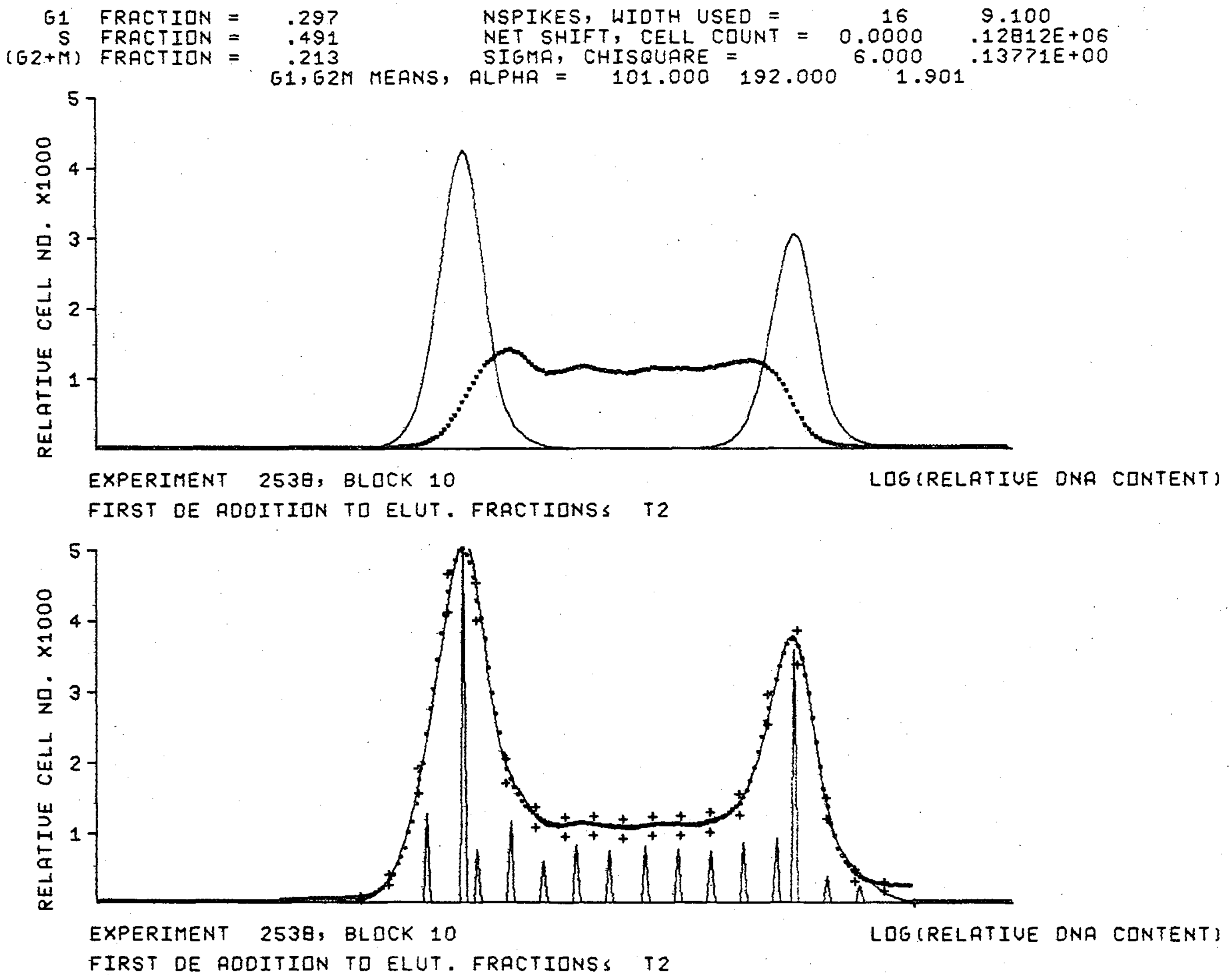
Figure 6a



```
   G1  FRACTION =    .433          NSPIKES, WIDTH USED =        12     15.167
    S  FRACTION =    .274          NET SHIFT, CELL COUNT =  0.0000    .13065E+06
(G2+M) FRACTION =    .293          SIGMA, CHISQUARE =        10.000    .59171E+00
                   G1,G2M MEANS, ALPHA =    101.000  192.000     1.901
```

XBL 788-4108

Figure 6b

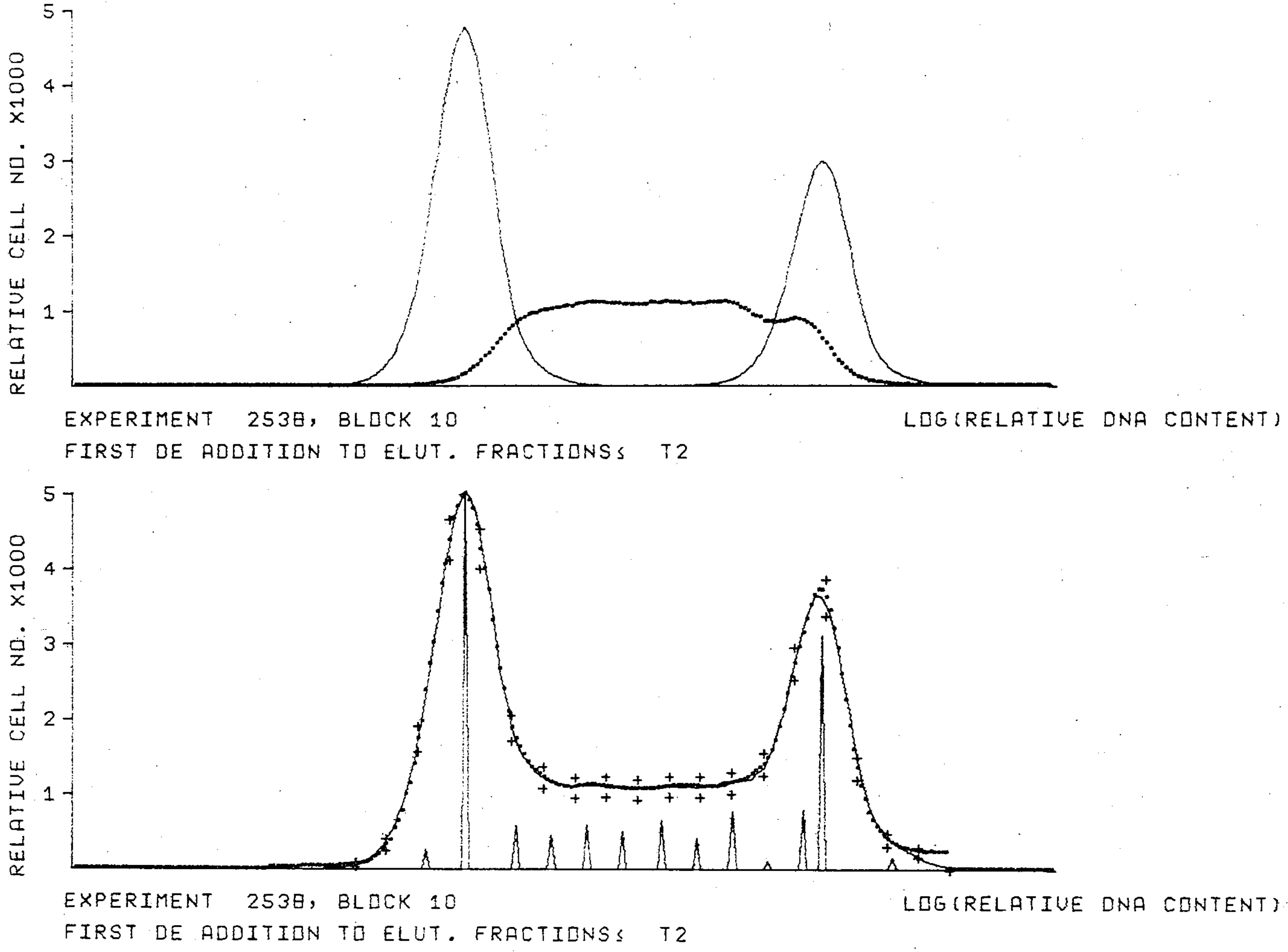


```
   G1  FRACTION =    .297          NSPIKES, WIDTH USED =       16     9.100
    S  FRACTION =    .491          NET SHIFT, CELL COUNT =  0.0000    .12812E+06
(G2+M) FRACTION =    .213          SIGMA, CHISQUARE =        6.000    .13771E+00
             G1,G2M MEANS, ALPHA =    101.000  192.000    1.901
```

XBL788-4109

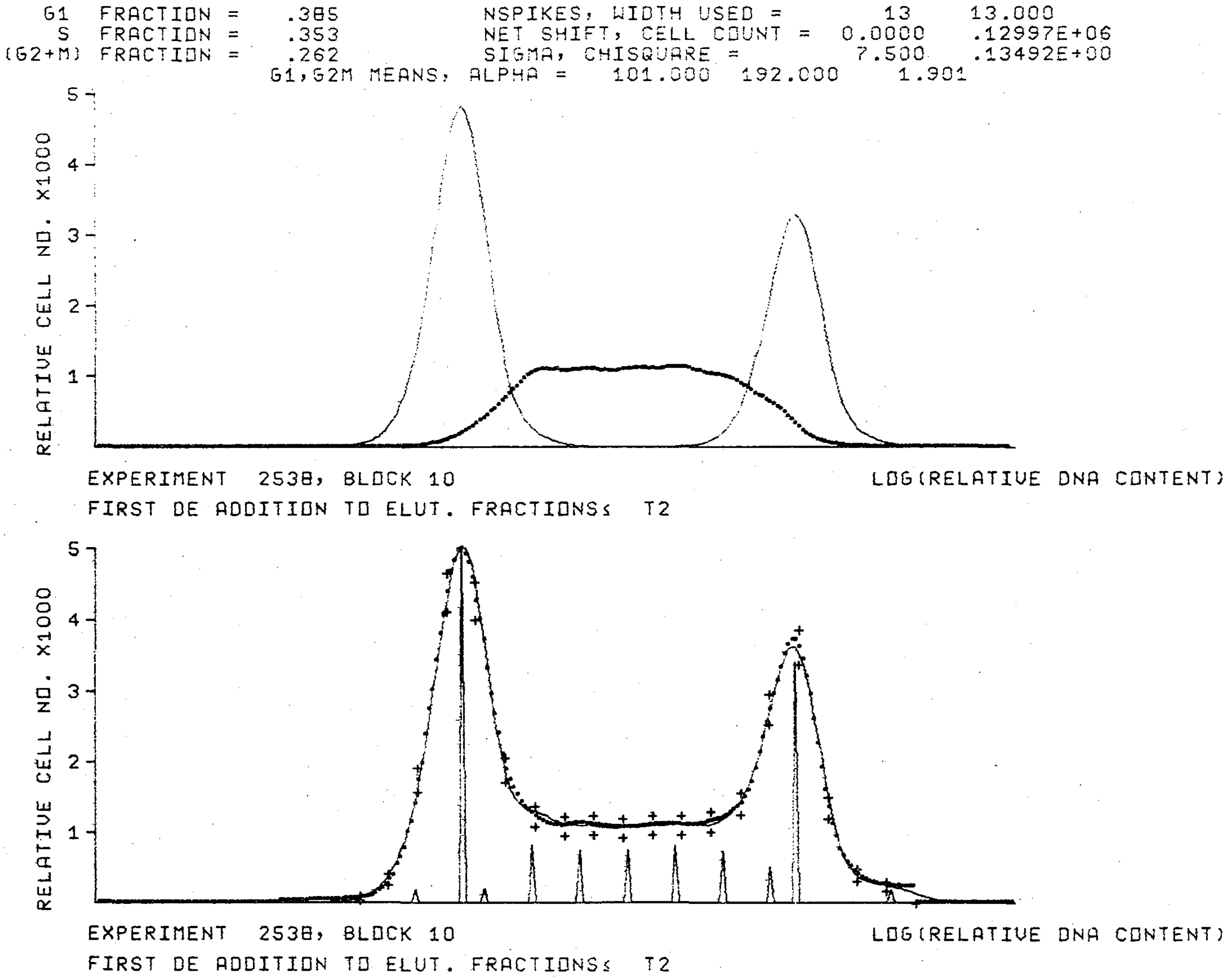
Figure 6c

```
     G1  FRACTION =    .383          NSPIKES, WIDTH USED =        16      9.100
      S  FRACTION =    .378          NET SHIFT, CELL COUNT =  0.0000    .12812E+06
 (G2+M)  FRACTION =    .239          SIGMA, CHISQUARE =          7.500    .11216E+00
                   G1,G2M MEANS, ALPHA =    101.000  192.000     1.901
```



XBL 788-4114

Figure 7a

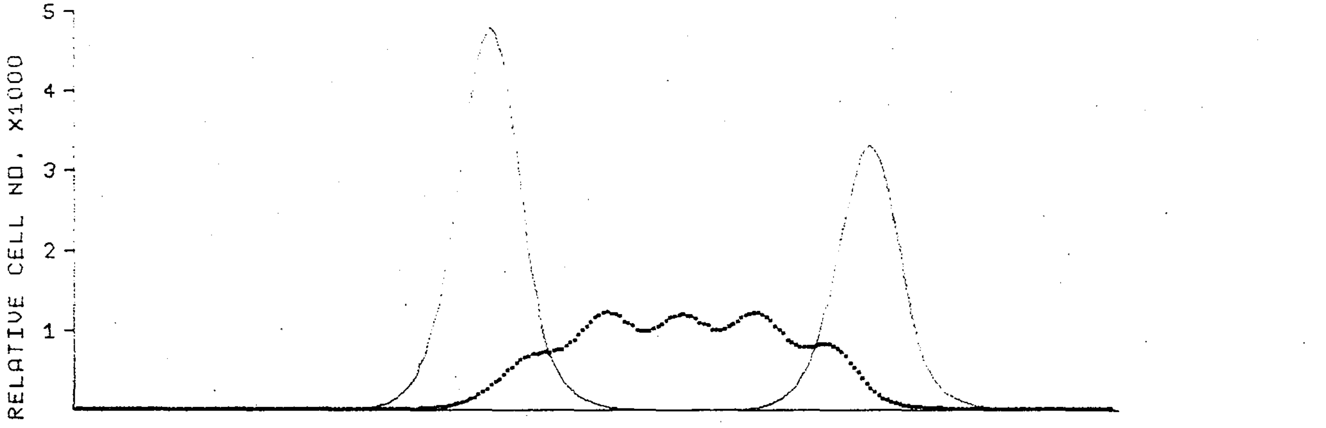


```
   G1  FRACTION =    .385          NSPIKES, WIDTH USED =        13    13.000
    S  FRACTION =    .353          NET SHIFT, CELL COUNT =  0.0000    .12997E+06
(G2+M) FRACTION =    .262          SIGMA, CHISQUARE =        7.500    .13492E+00
                   G1,G2M MEANS, ALPHA =   101.000  192.000     1.901
```

XBL 788-4115

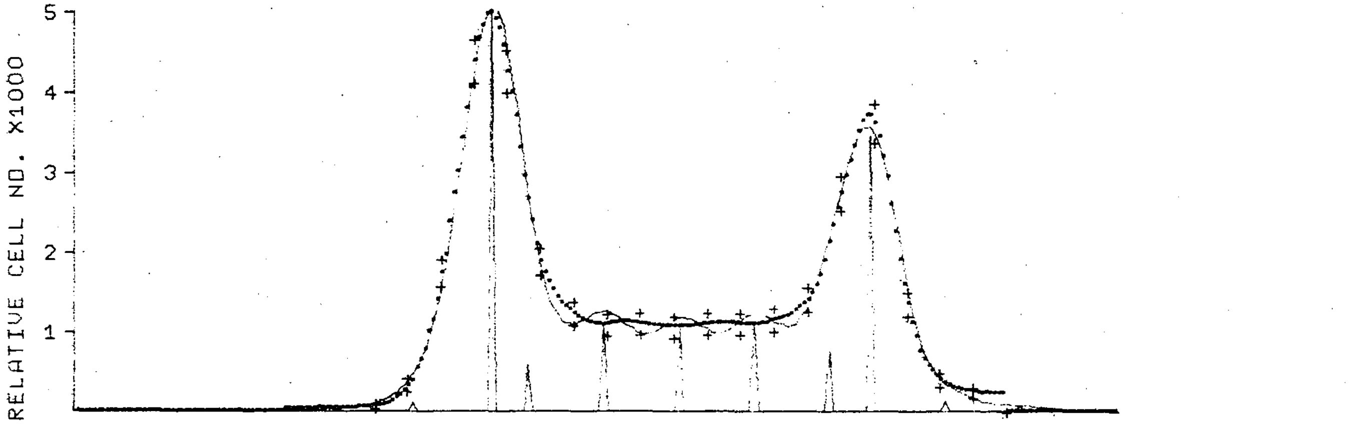
Figure 7b

```
   G1  FRACTION =   .383          NSPIKES, WIDTH USED =        11    18.200
    S  FRACTION =   .354          NET SHIFT, CELL COUNT =  0.0000   .13104E+06
(G2+M) FRACTION =   .263          SIGMA, CHISQUARE =         7.500   .24103E+00
                 G1,G2M MEANS, ALPHA =   101.000  192.000    1.901
```



EXPERIMENT 2538, BLOCK 10

FIRST DE ADDITION TO ELUT. FRACTIONS: T2



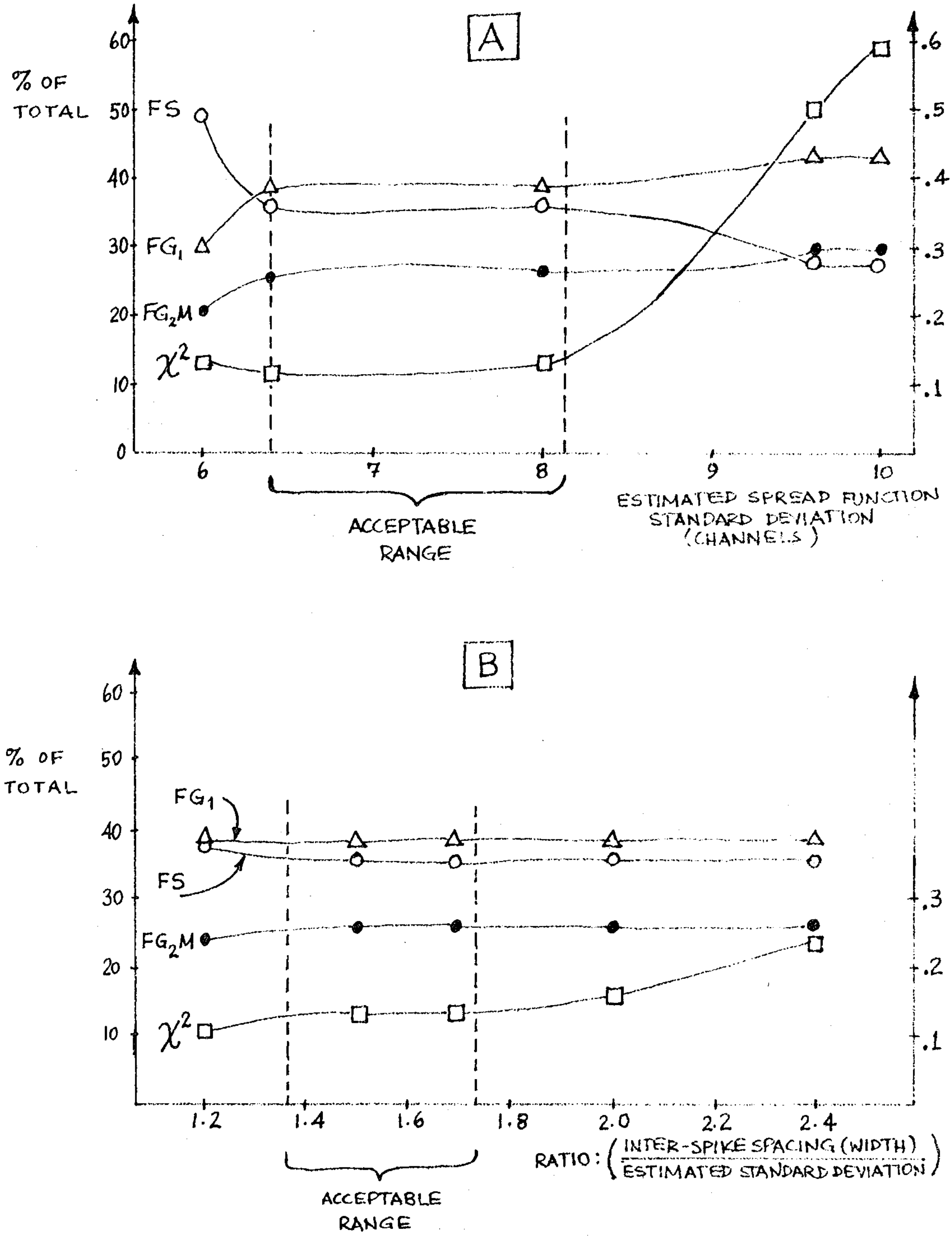
EXPERIMENT 2538, BLOCK 10

FIRST DE ADDITION TO ELUT. FRACTIONS: T2

XBL 788-4113

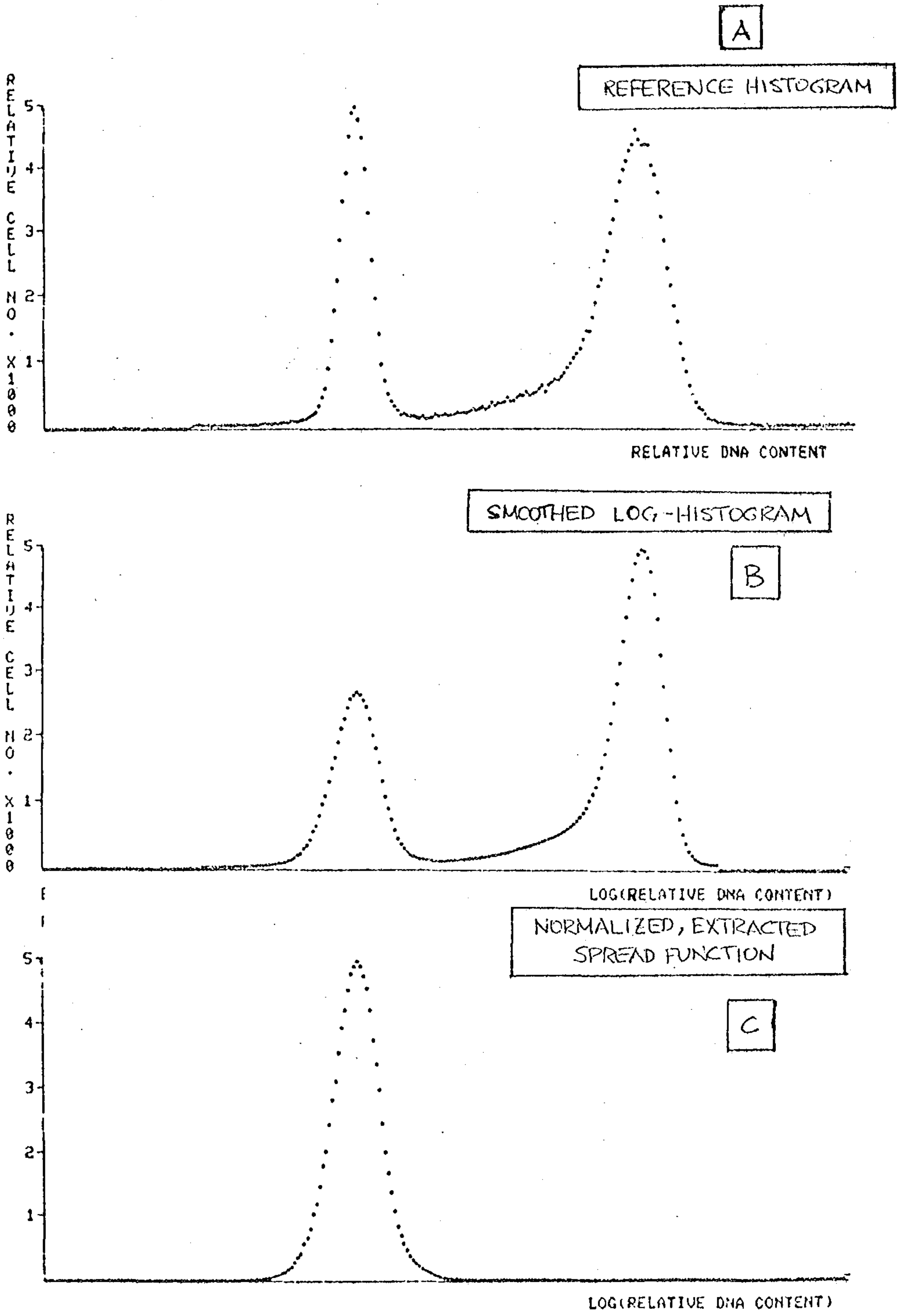
Figure 7c

Figures 8a,b



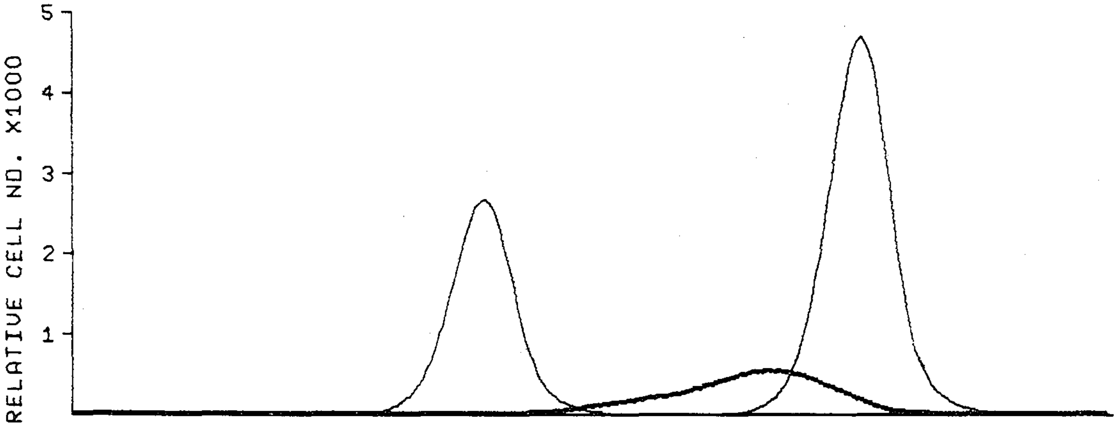
XBL 788-10614

Figures 9a,b,c

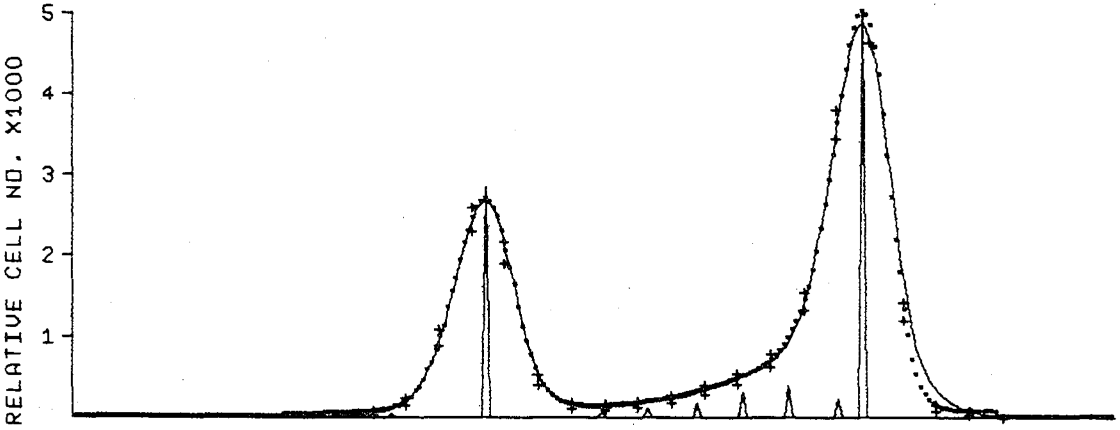


XBL 788-10617

```
    G1  FRACTION =    .313          NSPIKES, WIDTH USED =        14    11.375
     S  FRACTION =    .140          NET SHIFT, CELL COUNT =  0.0000    .15742E+06
(G2+M)  FRACTION =    .548          SIGMA, CHISQUARE =        7.371    .23698E+00
                 G1;G2M MEANS, ALPHA =    100.000  191.000     1.910
```



EXPERIMENT 2538, BLOCK 9
FIRST DE ADDITION TO ELUT. FRACTIONS: T2

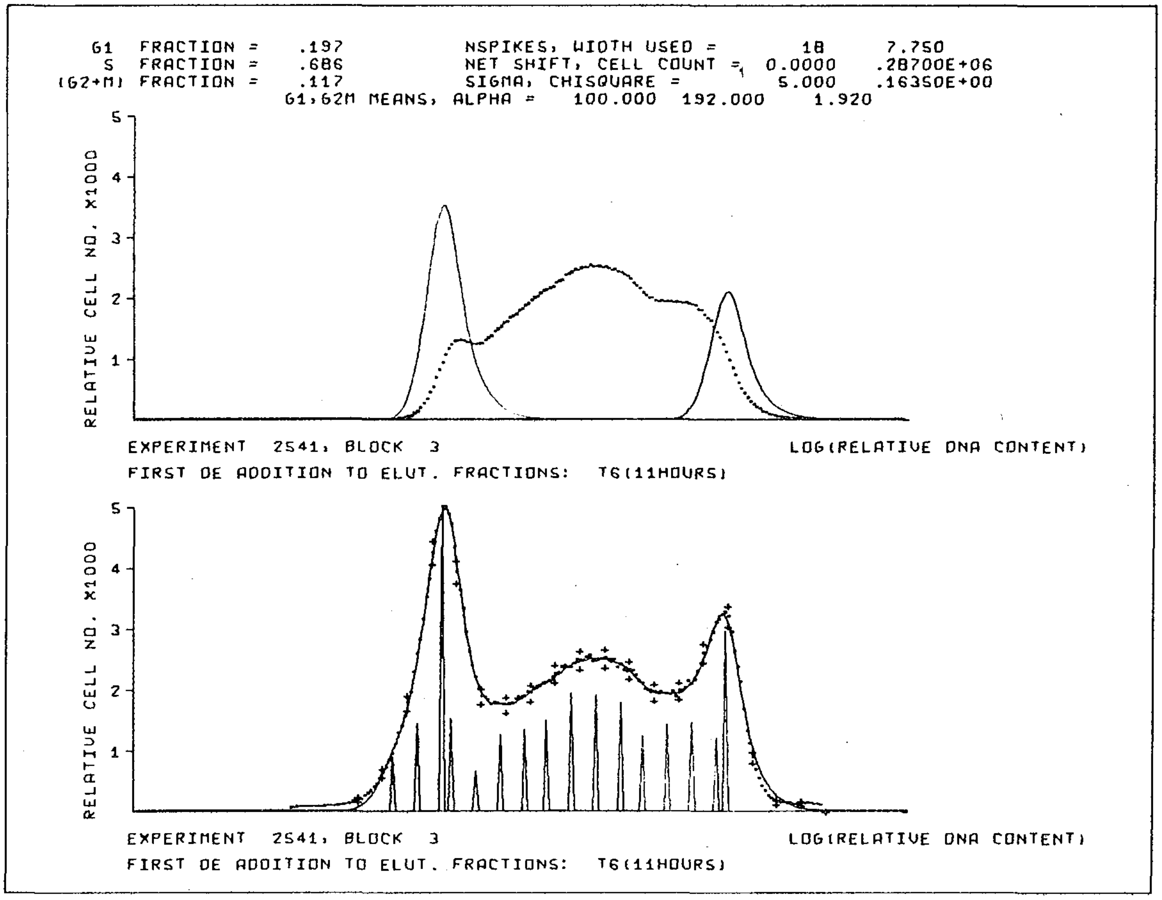


EXPERIMENT 2538, BLOCK 9
FIRST DE ADDITION TO ELUT. FRACTIONS: T2

XBL 788-4112

Figure 10a

Figure 10b



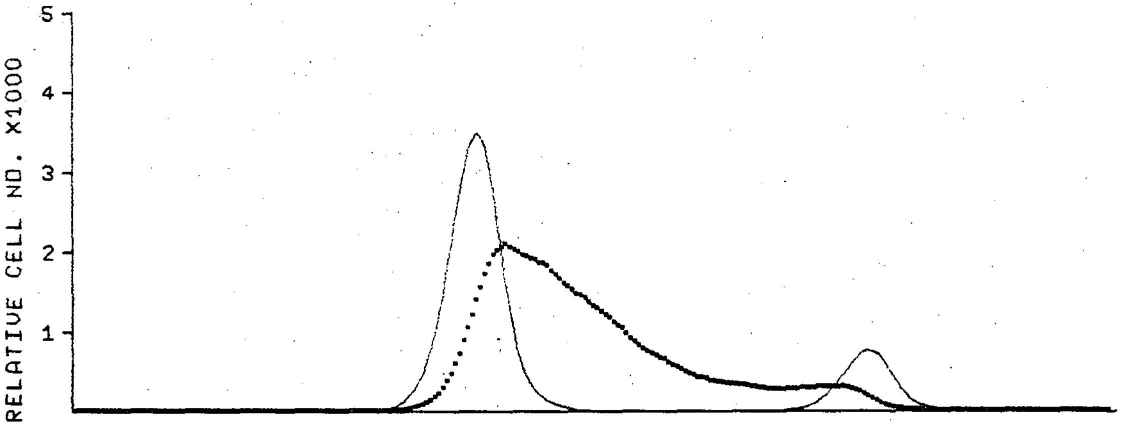
```
   G1  FRACTION =    .197          NSPIKES, WIDTH USED =       18      7.750
    S  FRACTION =    .686          NET SHIFT, CELL COUNT =   0.0000    .28700E+06
(G2+M) FRACTION =    .117          SIGMA, CHISQUARE =          5.000   .16350E+00
                   G1,G2M MEANS, ALPHA =    100.000  192.000    1.920
```

```
EXPERIMENT  2541, BLOCK  3                                LOG(RELATIVE DNA CONTENT)
FIRST DE ADDITION TO ELUT. FRACTIONS:   T6(11HOURS)
```

```
EXPERIMENT  2541, BLOCK  3                                LOG(RELATIVE DNA CONTENT)
FIRST DE ADDITION TO ELUT. FRACTIONS:   T6(11HOURS)
```

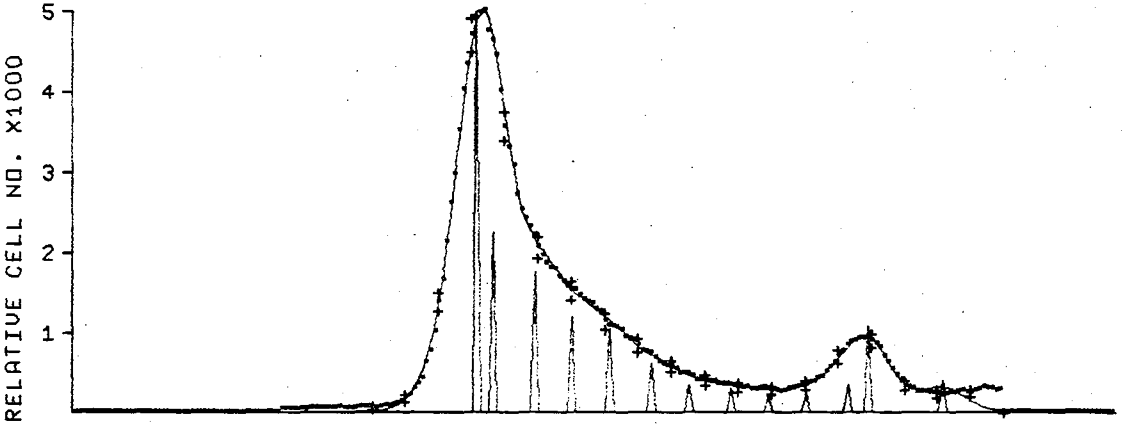
XBL 784-3930

```
   G1  FRACTION =    .345          NSPIKES, WIDTH USED =        16     9.400
    S  FRACTION =    .580          NET SHIFT, CELL COUNT =  0.0000   .15565E+06
(G2+M) FRACTION =    .075          SIGMA, CHISQUARE =         6.000   .15434E+00
                    G1,G2M MEANS, ALPHA =    98.000  192.000    1.959
```



EXPERIMENT 2541, BLOCK 9

FIRST DE ADDITION TO ELUT. FRACTIONS; T6(11HOURS)

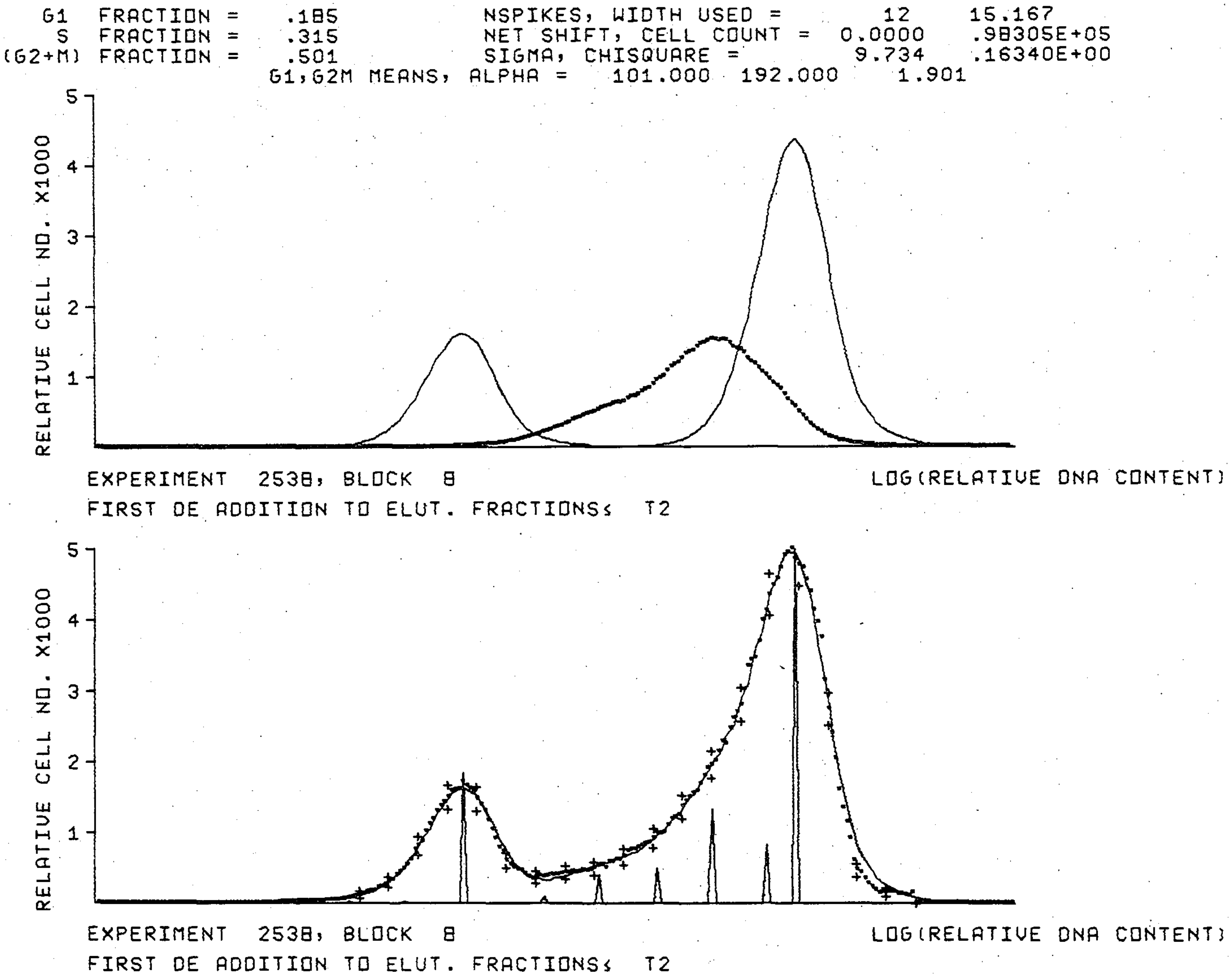


EXPERIMENT 2541, BLOCK 9

FIRST DE ADDITION TO ELUT. FRACTIONS; T6(11HOURS)

XBL 788-4110

Figure 10c



```
     G1  FRACTION =    .185          NSPIKES, WIDTH USED =        12    15.167
      S  FRACTION =    .315          NET SHIFT, CELL COUNT =  0.0000    .98305E+05
 (G2+M)  FRACTION =    .501          SIGMA, CHISQUARE =          9.734   .16340E+00
                G1,G2M MEANS, ALPHA =   101.000  192.000    1.901
```

XBL 788-4111

Figure 10d

CHAPTER IV

PRINCIPLES FOR ANALYSIS OF TIME SERIES OF DNA HISTOGRAMS OBTAINED BY FLOW CYTOMETRY

INTRODUCTION

With the increasing availability of flow cytometric devices capable of rapid analysis of large cell populations, there has grown a corresponding need for appropriate methods of reducing, analysing, and storing the data produced by them. In particular, much attention has been focused of late on the use of time sequences of DNA histograms in the study of mitotic cycle perturbations by various agents (drugs, carcinogens, viruses, heat, etc.). Such studies typically result in dozens of histograms, often varying in resolution and generally not well alligned with one another, making the task of interpreting the data very cumbersome.

We present here several general notions, as well as some specific algorithms, for attempting to extract information from time series of DNA histograms. In our approach, we have focused on two issues: how can the cytokinetic questions of interest to the investigator be translated into forms that are addressable in terms of DNA histogram time series?; how can the information contained in such time series be extracted, prepared and displayed so as to facilitate interpretation by the investigator? We also address the problem of storing both raw and analysed histogram data, and present some solutions that we have found useful.

DISPLAYING RAW DATA FOR OPTIMAL VISUAL STUDY

It is our conviction that much of the useful information contained in a series of DNA histograms can be obtained by direct visual inspection, if the data are properly presented. However, many FCM devices analyse cells either until the sample is exhausted or until one of the channels reaches a predetermined stopping count. This usually results in histograms displayed in a manner that obscures their fundamental nature: they are samples of the cell population's probability distribution, with respect to fluorescent intensity per cell. Being distributions, each histogram should be normalized to the same standard area as all others, so that the distributional properties of one histogram can be directly compared to those of any other. Further, few systems employ any gain stabilizing feature that would allign one histogram with the others in a series. These shortcomings are relatively easy to overcome, not requiring sophistocated modeling programs, and are perhaps the most cost-effective data manipulation done on DNA histograms.

Normalization of the histograms to a single standard area is generally a straightforwad operation: each histogram is integrated to yield its area, call it A1, then each of its bin values is multiplied by the ratio (A2/A1), where A2 is the desired standard area. A complication can arise when only a portion of the histogram is to be used in further study, and needs to be isolated from the larger histogram. Such is the case when, for example, cellular debris or aggregates give rise to unwanted shoulders or extra peaks in a histogram. One way to handle this case is to set the bin values to zero outside the desired region, then proceed as above. This is a simplistic approach, but is sufficient for most purposes.

Allignment of histograms involves two steps: locate some marker on the histogram that indicates the present gain setting; and perform a linear transformation of the abscissa to gain-shift the marker, and hence the histogram, to the desired standard position. A typical marker is the mode of the G1 peak, which is easily detected by a simple search procedure in most histograms, and in almost any case can be interactively estimated by the user.

The second step in the allignment procedure is to linearly transform the abscissa so that the marker is moved to a standard position. That is, to redigitize the historam on a new set of bins that are either broader or narrower than the original set, corresponding to gain-shifting the marker downstream or upstream, respectively. This situation is illustrated in Fig. 1. "Bins" are merely segments of the continuous abscissa, delimited by defined boundaries, one bin set differing from another only by its corresponding set of boundaries. If x represents the abscissa on which the old bin set is defined, and y is that of the new set, a gain-shift is represented by the simple linear transformation,

$$y = bx$$

where b is the ratio of the new desired marker position to that of the old marker.

The bin boundaries, being merely points on these two abscissas, obey the same linear relationship, leading to the algorithm for establishing a new bin set for a given ratio b:

For each old boundary x(i), i=1 to N+1, where N is the total number of bins, the new corresponding boundary is y(i), given by

$$y(i) = \underline{b}x(i)$$

In general, there will be few values "j" for which y(j) =

x(i), as indicated in Fig. 1, and so the histogram must be redigitized by some form of interpolation. Since all interpolation involves passing a functional model through the data points, the principal difference between methods is the choice of the function. A simple linear interpolation is probably satisfactory for most cases, but Simpson's method, employing quadratic fits to every three points, is more accurate and involves little more effort than the linear method. In logarithmic transformations of the abscissa, discussed below, the higher order interpolation is more needed, as there is a greater discrepancy in bin widths to be overcome.

When there is a large sample size variation within a histogram series, the scattering due to counting statistics may be far more serious in some histograms than in others, making it difficult to directly compare them. The user may wish to employ a smoothing scheme to systematically eliminate this often distracting scatter, and thus further facilitate visual comparison. One way to smooth the data is to sweep an "averaging window" across the histogram, replacing the value in bin "i" by some weighted average of the values in "i" and surrounding bins. The width and weights used in the window have definite implications if further analysis is to be done on the histograms, but a simple uniform weight, 5 channel window centered on "i" works well in most instances for the present purposes. However, when smoothing as part of the fitting procedure, we choose to fold the histogram with a Gaussian distribution of 1 channel standard deviation and unit area.

Once the series has been normalized, alligned, and optionally smoothed, it can be displayed in several ways: separate histograms arranged in horizontal or vertical sequence, as in Fig. 2; plotted as a function of two variables, using perspective plots, as in Fig. 3, or as a two dimensional halftone or gray level plot, (not shown). (though not shown here, color coded plots are also quite useful for visual display). The purpose of any of these display modes is to visually enhance those aspects of the histogram information that most directly impinge on the user's interpretory abilities. Once the user has drawn whatever conclusions are evident from well displayed data, and if further information is desired, the next step is to employ modeling approaches to extract features that are not readily seen in the raw data itself. In this light, modeling is seen as a tool for probing the data, to be used, together with optimal visual examination of it, in order to extract the maximum amount of useful information justified by the quality of the data.

GARBAGE IN, GARBAGE OUT: INFORMATION CONTENT OF THE DATA

As is the case in most experimental schemes, where results are to be concluded from modeling the data, the reliability and quality of the results reflect those of the data far more than the properties of the mathematical methods used to model it. For inverse problems of the kind outlined below, Backus and Gilbert (1) have shown there to be a conservation of total information contained in different models of the same data, and the conserved amount is determined by the resolution of the original data. That is, if one model has N free parameters while the other has M, M < N, the statistical reliabilty with which the N parameters can be determined is less than that for the M parameter model. The product of free parameter number and the resolution to which the parameters can be determined in solving the inverse problem is constant, that constant being determined by the information content of the data.

The message in this is an old one: one should be sure the experimental technique is optimized before embarking on mathematical modeling expeditions to interpret it. For DNA histograms, the lesson is to take all reasonable precautions in the sample preparation, staining and FCM analysis steps to minimize error and noise, before seeking ever more sophistocated methods to fit the histograms. Furthermore, this means that histograms possessing a coefficient of variation (CV) of 6% (for example) have roughly half the information content of those with a CV of 3%, and thus should be fit with about half the number of free parameters as the former, if the reliability of the parameter values is to be maintained the same for the two histograms.

MODELING INDIVIDUAL HISTOGRAMS: AN INVERSE PROBLEM

If FCM analysis were error free, the measured fluorescent intensity of a given cell would precisely reflect its exact DNA content, and there would be no ambiguity in deriving the cycle phase fractions or detailed S-phase structure from the histograms. However, due to sample preparation, staining and FCM analysis errors, the measured quantity deviates by a generally small amount from the apparent proper fluorescent intensity for a given cell, and the histogram is thus blurred. A detailed discussion of this subject is found in chapter III, but briefly, we showed that the relation between observed intensity X_2 and correct intensity X_1 (i.e., that which would be obtained in an error free FCM procedure) can be represented as

$$X_2 = X_1(1 + z) \simeq X_1 e^z \qquad (1)$$

where z is the relative random error for true DNA value X_1, and is distributed with mean = 0, by distribution function $A_Z(z)$.

By a logarithmic change of variables, the relation between the observed logarithmic fluorescent intensity distribution $P_2(y)$ and the true underlying logarithmic DNA distribution $P_1(y)$, is given by

$$P_2(y) = \int_{-\infty}^{+\infty} P_1(z) A_Z(y - z)\,dz \qquad (2)$$

The observed histogram, after logarithmic transformation of its abscissa, is a discrete version of the distribution $P_2(y)$, represented here as $\overline{P}_2$. When written in matrix form, the above relation becomes

$$\overline{P}_2 = \underline{A}\overline{P}_1 + \overline{E} \qquad (3)$$

where the elements of vector $\overline{P}_2$ are the bin values of the logarithmically transformed histogram, the vector $\overline{P}_1$ is the unknown underlying discrete logarthmic DNA distribution, the columns of spread matrix $\underline{A}$ are filled with the spread distribution function $A_Z(z)$, and vector $\overline{E}$ represents error due to counting statistics.
In this formulation, with the logarithmically transformed abscissa, $A_Z(z)$ is presumed to be invariant with y.

Recovering the unknown vector $\overline{P}_1$ is equivalent to inverting the matrix $\underline{A}$, yielding

$$\overline{P}_1 \approx \underline{A}^{-1}\overline{P}_2 - \underline{A}^{-1}\overline{E} \qquad (4)$$

where "$\underline{A}^{-1}$" is the inverse of $\underline{A}$, if indeed one exists. As discussed in chapter III, the problem is rather ill-conditioned resulting in a matrix $\underline{A}^{-1}$ (if it exists) that amplifies the magnitude of high frequency components of vectors multiplied by it, in particular the noise vector $\overline{E}$. Thus, small, random errors in the data vector $\overline{P}_2$ will be magnified by $\underline{A}^{-1}$, resulting in large uncertainties in the estimate of $\overline{P}_1$. Conversely, large differences in estimated underlying distribution $\overline{P}_1$ are reflected in relatively small changes in the model of the data vector $\overline{P}_2$.

In general, matrix $\underline{A}$ is singular, and thus possesses no inverse, however, the consequences of its ill-conditioned nature still hold. As a result, the problem assumes the optimization form of finding the best choices for the entries of unknown vector $\overline{P}_1$. "Best" is usually taken to be that set of $\overline{P}_1$ entries which when multiplied by spread matrix $\underline{A}$ yields the smallest squared deviation from the observed vector $\overline{P}_2$.

There are two important conclusions to be reached as a result of the above analysis. First, the severity of the error magnification properties of matrix $\underline{A}^{-1}$ decreases with improved measurement resolution, that is with decreased CV. Also, the error due to counting statistics falls off with increasing sample size, so that the magnitude of error vector E can be reduced by counting more cells. These points imply that the better the quality of the raw data, the greater the extent to which the unknown underlying distribution can be determined. This should not be surprising in view of the principle given above: that the results to be extracted from data are only as good as the data itself.

Second, as discussed above, the estimate of underlying distribution $\overline{P}_1$ is sensitive to small variations in observed vector $\overline{P}_2$. That is, many significantly different estimates of the unknown $\overline{P}_1$ may result in fits that are, from a statistical standpoint, equally good fits to the data $\overline{P}_2$. Thus, there is no uniquely "best" estimate of $\overline{P}_1$, but rather a whole class of estimates that yield equally valid fits to $\overline{P}_2$.

The above discussion leads to the unfortunate conclusion that, as Fried (2) and others have pointed out, there are many ways to fit a given DNA histogram to within a given statistical tolerance. The set of possible fits to the such data may allow for significant variation in estimated cycle fractions (3), as well as details of the S-phase structure, and it is unclear which is, the true solution. One way to restore some confidence to this situation would be to have a library of "well known" histograms, whose composition has been generally agreed upon, on which a given fitting program can be evaluated. Unfortunately, no such library currently exists. The only other standards against which a given fit can be judged are those employing autoradiography, whose low resolution was one of the reasons for the development of the present FCM technology, and is thus capable of setting only broad limits for cycle phase fractions.

SUBJECTIVITY AND EXPERIENCE VERSUS OBJECTIVITY AND REPRODUCIBILITY

Thus, we can only be assured that the estimates from a fitting program are accurate to, say, 10% of their values, and as long as this level of error can be tolerated for the purposes at hand, there does not appear to be a problem. However, if the experiment is designed to detect the time at which a given effect begins, and timepoints are taken only at 2 hour intervals, the above 10% error may mean the difference between assigning the starting time to 8 hours or to 10 hours; a discrepancy that may be quite significant to the investigator. In many such cases, an investigator expecting the outcome to be the 10 hour value will chose to censor the finding by the program that the effect was already significant by 8 hours.

In the face of such uncertainties, the only apparent approach to take is that whatever fitting method is employed, it should treat all histograms in the same manner, to the maximum extent possible. The limitation on this policy is the all too well known fact that despite the best precautions, real data are often perverse, and along with the many histograms that can be "objectively" or "automatically" fit are several "pathological" cases that defy even the smartest algorithms. Such difficult data require interaction with the investigator or whomever is analysing them, and should be isolated from a series of histograms prior to their analysis, so that the tractable cases can be analysed automatically. Still, in the end, since DNA histogram analysis is a developing tool, the evaluation of the investigator will remain the final judge of the success or failure of a given fit, and provision for such interaction should be maintained, if only after the automatic analysis mode has had a chance to fit the data.

THREE PASS ANALYSIS

The above discussion leads to the suggestion that a series of histograms be analysed in three separate passes:

1. Identify and isolate potentially troublesome histograms and analyse them separately.
2. Submit the remaining histograms to an automatic fitting routine.
3. Combine the results of 1 and 2 as input to a kinetic analysis package or other data reduction scheme.

This approach is illustrated by flow chart in Fig. 4. In our implementation of this method, the single histogram analysis program used is that described in chapter III.

Other authors (3,4) have pointed out the desirability of incorporating into the analysis of each histogram the fact that a given set of histograms represents the time evolution of a cell population. In the case of Gray's approach, the simulation of the whole series incorporates the modeling of the individual histograms, converting the latest estimate of the underlying true cell cycle distribution into a model DNA histogram, by smoothing with a Gaussian spread function. In this manner, analysis of each histogram is related to preceding and succeeding histograms via the kinetic model. However, as is discussed below, the kinetic modeling problem is also inverse in nature, and is thus subject to the same kinds of uncertainty as the fitting program, if not worse. Furthermore, it is our understanding that this approach consumes large amounts of computer time and space that place it beyond the capabilities of most investigators. For these reasons, we have based our approach on the separation of the problems of individual histogram analysis and kinetic modeling of the time series of DNA histograms.

KINETIC MODELS: IDENTIFICATION OF USEFUL PARAMETERS

In studies where the investigator is seeking to determine the effects of an agent upon the cell cycle kinetics of a given population, the vocabulary of terms descriptive of the agent's action is limited to the following:

a. Slowing progress through the cycle
b. Recruitment of noncycling cells
c. Blocking cells at some point in the cycle, or causing them to "decycle" (as in G_0)
d. Accelerated progress through the cycle
e. Cell killing

These descriptors, combined with the specific portion of the cycle affected, e.g. early-S-phase, constitute the overwhelming majority of possible kinetic effects that are of interest to most investigators. Hence, the goal of a kinetic analysis approach should be to determine which of the above features are indicated by the data, where and to what extent are they seen in the cycle. The problem of characterizing the effects of a given agent upon the cell population is thus reduced to that of obtaining parametric values for the descriptive properties listed above. This is to be done by modeling the time series of DNA histograms obtained in a cycle perturbation experiment, and thus is a kind of dynamic systems modeling problem.

If the cell cycle is modeled as a sequence of compartments, with time evolution of the population represented as transition from compartment to compartment, the equation governing the number of cells in compartment "i" is given by Zietz as

$$FP_i(t) = \int_0^t FP_{i-1}(t-a) f(a) \exp\left[-\int_{t-a}^{t} R(t)dt\right] da \qquad (5)$$

WHERE $R(t) = \frac{1}{N(t)} \frac{dN(t)}{dt}$ ("RELATIVE MITOTIC RATE"), AND
where $f_{i-1}(a)$ is the distribution of growth rates "a" among the cells in compartment "i-1". The distribution f(a) incorporates all of the kinetic effects described above, in the most general case, and is the quantity to be sought in deriving the kinetic characterization of the population and of the agent's effects upon it. As discussed by Zietz (3), this is equivalent to inversion of the integral operation in equation (5), and is thus yet another inverse problem, generally even more ill-posed than that of unfolding the underlying DNA distributions from individual histograms.

In the approach implemented by Gray (4), the above cell cycle model is invoked, along with local cell conservation laws, to produce a system of ordinary differential equations (ODE's) that describe the time evolution of the population. The various kinetic effects listed above are incorporated into the model by way of the transition probability terms at each compartment, with each route (growth, death, decycling, etc.) having its assigned unknown probability of occurence per unit time. Rather than ask the inverse question, "what set of probability terms, when inserted into these equations, gives the best model of the data," Gray nicely turns the approach into a tool for checking the validity of user-generated kinetic hypotheses. That is, the user provides a (hopefully educated) guess at the numerous transition probabilities, and a few other parameters, and the program solves the resulting system of ODE's, generating a time series of model histograms that the user can compare visually with the data. The approach has the potential to directly test for the presence and extent of the various of kinetic effects discussed above, thus addressing questions of direct relevance to the investigator. Unfortunately, in its present form, the program allows for so many free parameters that it can take 5 minutes of core time on a CDC 7600 computer* to adequately fit a typical set of 7 or 8 histograms. Perhaps a less ambitious version of Gray's approach, involving fewer parameters and using the estimates for $\overline{P}_1$ (the solution vector for the fitting problem, described above) as data to be modeled, rather than the raw data, would be sufficiently inexpensive to be within the budgets and computing capabilities of most investigators.

Lastly, a useful combination of the inverse and direct approaches might be to use an initial guess of the parameters from the user, and conduct an optimization search in a small user-defined vicinity about the intial values. If this proves to be prohibitively large and/or expensive to implement, perhaps the parameters could be optimized one at a time within the restricted ranges given, with the option to quit whenever the model produces fits that are acceptable to the user.

[Appendix B of this thesis deals with my attempts to implement an inverse approach to a modest (11 parameter) kinetic model of the solution sets to fitted histograms] (private communication from Gray)

OTHER WAYS TO PROBE AND REDUCE DATA

The third pass of the analysis scheme outlined above uses the results of fitting the time series of histograms as input to various data reduction and kinetic modeling programs. Short of kinetic models which, as discussed above, tend to be rather complcated affairs, other tools have been developed for probing features of the data that are not seen even in well displayed data.

One approach is to isolate and observe the time behavior in a small region of the abscissa of the DNA histograms. Scherr (5) implements this for normalized and alligned histograms, without fitting them first, by simply establishing observation window at various fixed positions along the abscissa, and counting the fraction of the population in each window, or FPW. The FPW behavior over time is used to estimate the durations of the cycle phases. This is done by comparing the arrival time of a synchronous cohort of cells at judiciously positioned windows in G1, early-S, mid-S, late-S, and G2M, if such a cohort is produced as part of the experiment. An example is shown in Fig. 5. The early and late-S windows are subject to somewhat greater uncertainty than the others, in this scheme, because of their overlap with the G1 and G2M peaks, respectively. Nonetheless, this approach offers consideratble potential as a method for extracting important kinetic information from a histogram time series, with minimal computing necessary. As Scherr shows (5), the FPW curves can be further modeled to give estimates of the kinetic rate distribution described above, $f_i(a)$.

A further adaptation of the FPW approach is to obtain the window values from fitted histograms. That is, after isolating the G1 and G2M peaks from the S-phase contributions, to redigitize S-phase on a new standard bin set, and use the G1, G2M and new bin values as the FPW. We have implemented this approach using the fitting program described in chapter III, and 9 bins in S-phase, for a total of 11 windows through the cycle. Examples of resulting FPW curves are shown in Fig. 6. In experiments where parallel histogram time series from treated and control populations are obtained, differences between their FPW curves may be directly calculated, giving a kind of "net effect" time plot for each of the windows. We have implemented this approach for our **11 FPW windows** derived as described above, Fig. 7, and examples of the resulting plots are shown in Fig. 8. The FPW and net effect approaches, when combined with properly displayed raw data provide a basis for visual assessment of the kinetic effects described above, but are not intended to yield quantitative kinetic parameter values, as would be obtained by a formal kinetic modeling approach.

STORING RAW AND ANALYSED DATA

The implementation of schemes for storing and retrieving hard earned data, and perharps even dearer analysed and reduced forms of data, is a vital step in the FCM process that has received relatively little attention. We present here the procedures and data structures imlemented at the Laboratory of Chemical Biodynamics for this purpose.

Our flow cytometer employs a Northern Scientific multichannel pulse height analyser (PHA) capable of storing 4096 channels in its temporary memory. Since we generally record 256-channel histograms, the PHA memory limits the number of histograms that can be held in temporary storage to 16. (However, the data structure can accommodate other than 256 channel sizes as well) Thus, our data structure is generally built around the assumption that up to 16 256-point histograms are to be contained in one experiment file, to be transmitted from the

FCM device to our in-house computer (a Sigma-2). Other information encoded in the file includes the date and time of transfer, experimenter(s) name(s), an 80 character title of experiment, critical FCM operating parameters (e.g. laser power, wavelength, etc.), and a general description of the experiment, of virtually unrestricted length. In addition, each histogram has an associated comment block of variable length associated with the data itself. The file and record structure just described is illustrated in Fig. 9. The total time necessary to encode this information for a typical set of 16 histograms is about 10 minutes, and can usually be done while the samples are being analysed. Each such set of 4096 total points, and associated commentary, is assigned an experiment file number, and a given biological experiment may comprise several such files.

The storage and retrieval of reduced or analysed data in our laboratory was addressed in detail by Mr. Bruce Navsky (6), who created an operating system for storing, maintaining and updating an archival file, with associated directory, to be accessed and implemented on the CDC 6600 computer used in our histogram analysis routines. The basic file structure for the archive is shown in Fig. 10, and is generally similar to that for the raw data. A "header" identifies the origin of the analysed data, the type of analysis done, the number of blocks of data produced by the analysis, and the date on which stored. The storage routine is usually run upon completing the analysis of a group of DNA histograms. A directory is updated upon each writing of the archival file, and it contains the address in the archive of each file of analysed data. New files are added at the end of the current archive.

Retrieval of analysed files is done by providing the proper header identifier for the desired file(s) to a search routine, are finds the archival address(es) in the directory. The desired files, once found in the archive, are copied onto local disc files for access by the user or for use as input to further analysis or graphics routines.

FIGURES

1. Gain-Shifting: upstream and downstream shifts shown for given bin set, indicating mismatch of bin boundaries and the need for interpolation. Downstream shift results in larger bins, with the modal channel smaller than before the shift; the reverse holds for upstream shift.
2. Time Sequence of DNA histograms: vertical sequence of normalized, aligned histograms.
3. Time sequence of DNA histograms: perspective plot of normalized, aligned histograms.

4. Analysing a sequence of DNA histograms

a) overall approach Histograms are first examined by user to categorize them according to the various analysis options; pathalogical data is set aside for individual analysis. The prepared series of histograms is submitted to the individual histogram modeling routine for non-interactive analysis. The fitted data are then examined by the user, who has the option to submit any of the histograms for reanalysis.

b) Detail of approach shown in a). User determines which arrays are to be used as reference, from which the spread function for fitting subsequent historams is to be extracted. Pathological data are identified for isolation. Details are given in text.

5. FPW ("fractional percent in window") curve (courtesy of Dr. Lawrence Scherr, Lawrence Livermore Laboratory). The fraction of cells in the indicated channel is traced through the succeeding timepoints, yielding a wavey curve. Analysis of FPW curve can yield estimates of average traversal times for the whole cycle and for its sub-phases.

6. "window contents": a form of FPW curve derived from fitted histogram time series. Fraction of total cell population in channels 20, 35 and 50% through S-phase (S_2,S_4,andS_6, respectively), is calculated

from the fitted S-phase of each histogram. For channels near the G_1 or G_2M peaks, use of fitted data enables separation of the contribution of peak cells from those of S-phase cells.

7. "Net Effect" Data reduction and analysis method: schematic illustration. The normalized difference plot between two histograms (test and control) is calculated by direct subtraction at eleven points: at the G_1 and G_2M means, and at nine points in S-phase. The calculation is repeated for each timepoint, and the collection of values at each difference site forms a "net effect" curve. The method is intended to reveal kinetic differences between test and control populations. Sample "net effect" curves at G_1, S_2 (20% through S-phase) and G_2M are shown.

8. "Net Effect" example curves. The data shown were taken from the experiments of this dissertation, and represent the differences at G_1, all of S-phase, and G_2M.

9. File structure for storage of raw DNA histogram data. Each file (up to 16 256-point histograms) consists of a header, a general commentary (usually a description of the experiment), and up to 16 blocks (each a histogram, with accompanying commentary). Information entered into computer program by user, as cell samples are being analysed by FCM.

10. Record structure of archival file for storage of analysed DNA histogram data. Each file consists of N blocks of up to M data types each, all arising from analysis of the data file indicated in the HEADER I.D. Individual block commentaries are appended at the bottom of each file. An archival directory keeps track of the address of each HEADER I.D., and new files are appended at the bottom of the tape.

REFERENCES

1. Backus, G.E., and Gilbert, F. Uniqueness in the Inversion of Inaccurate Gross Earth Data. Phil Trans Roy Soc London A 266:123 (1970)

2. Fried, J. Analysis of Deoxyribonucleic Acid Histograms from Flow Cytofluorometry, Estimation of the Distribution of Cells Within S-Phase. J Histoch Cytoch 25(7):942-951 (1977)

3. Zietz, S. and Nicolini, C. Flow Microfluorometry and Cell Kinetics: A Review. in Biomathematics and Cell Kinetics, ed A.J.Valleron Elsevier/North Holland (1978).

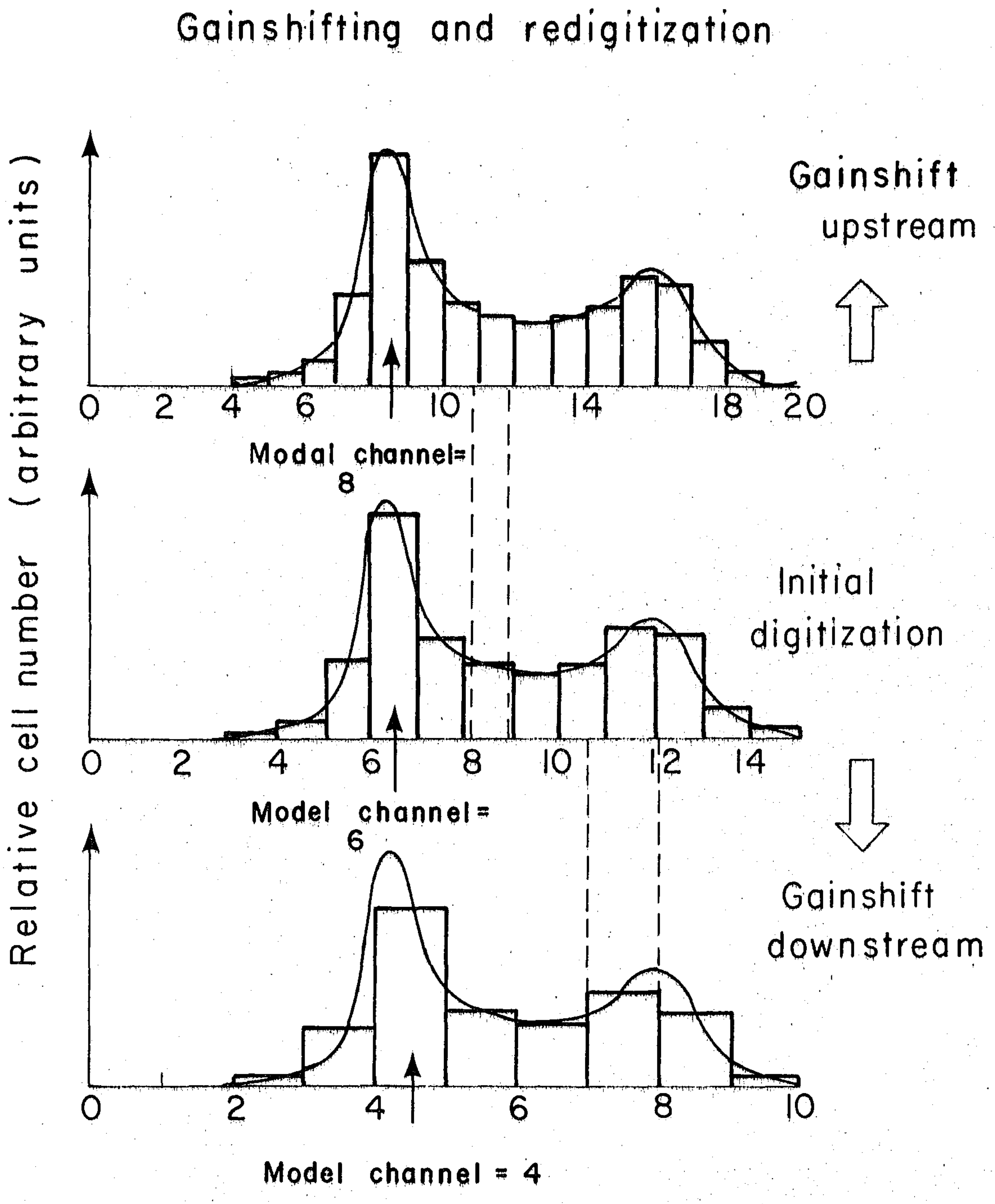
4. Gray, J.W. Cell-Cycle Analysis of Perturbed Cell Populations: Computer Simulation of Sequential DNA Distributions. Cell Tissue Kinet 9:499-516 (1976)

5. Scherr, L. FPW curves yield estimates of cell cycle

phase transit times. Lawrence Livermore Laboratory internal report, April 1978. (submitted 4/78 for publication in Cell and Tissue Kinetics).

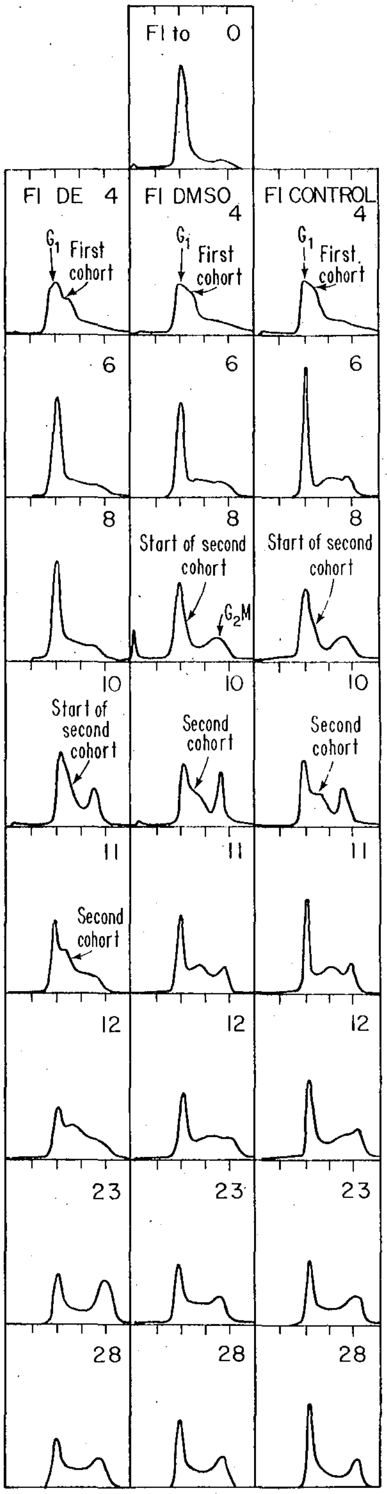
6. Navsky, B.N. OSFCM: An Operating System for the Flow Cytometer. Documented computer program, internal to Laboratory of Chemical Biodynamics, Lawrence Berkeley Laboratory, January, 1978.

Figure 1



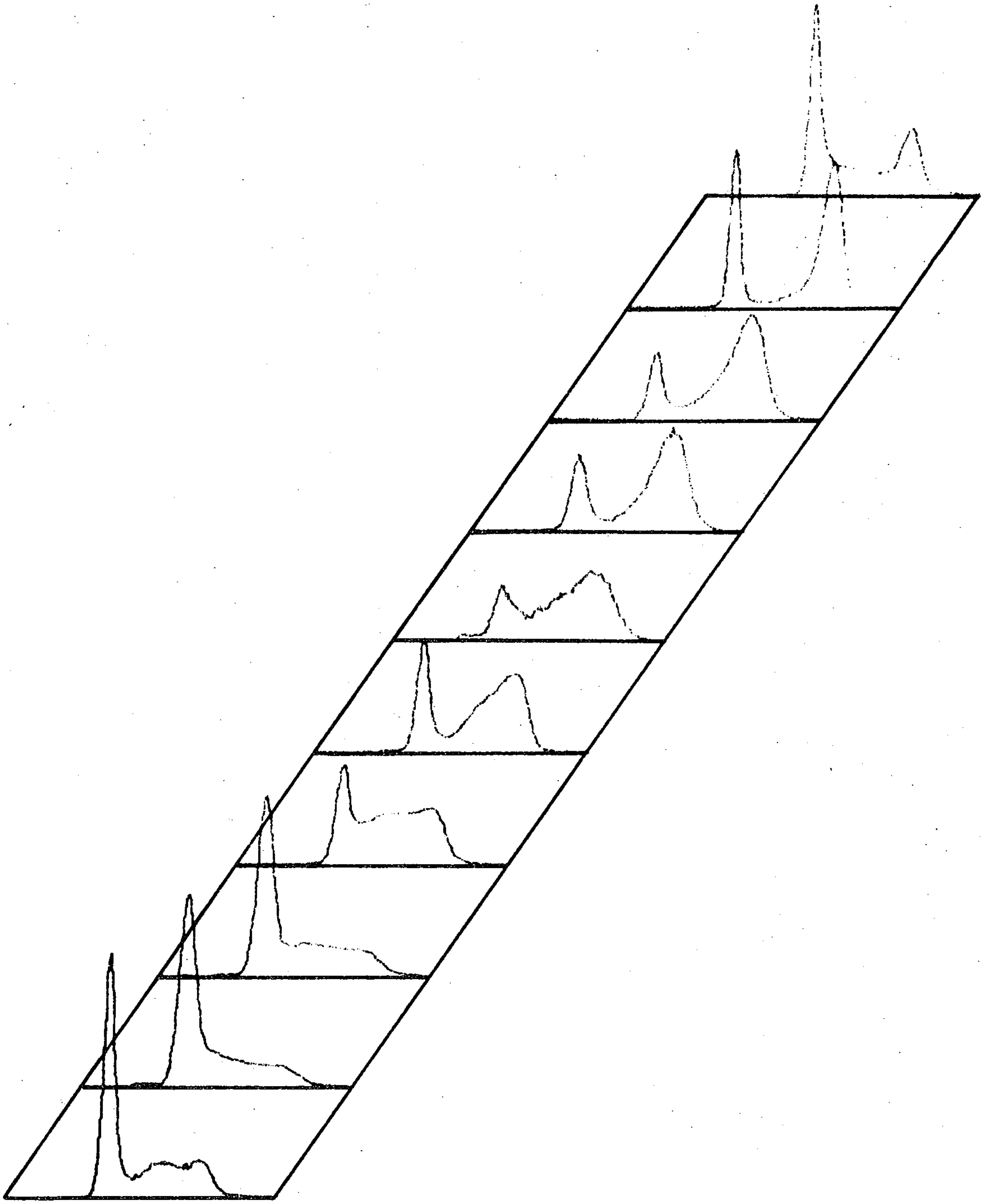
XBL788-1662

Figure 2



XBL 788-4169

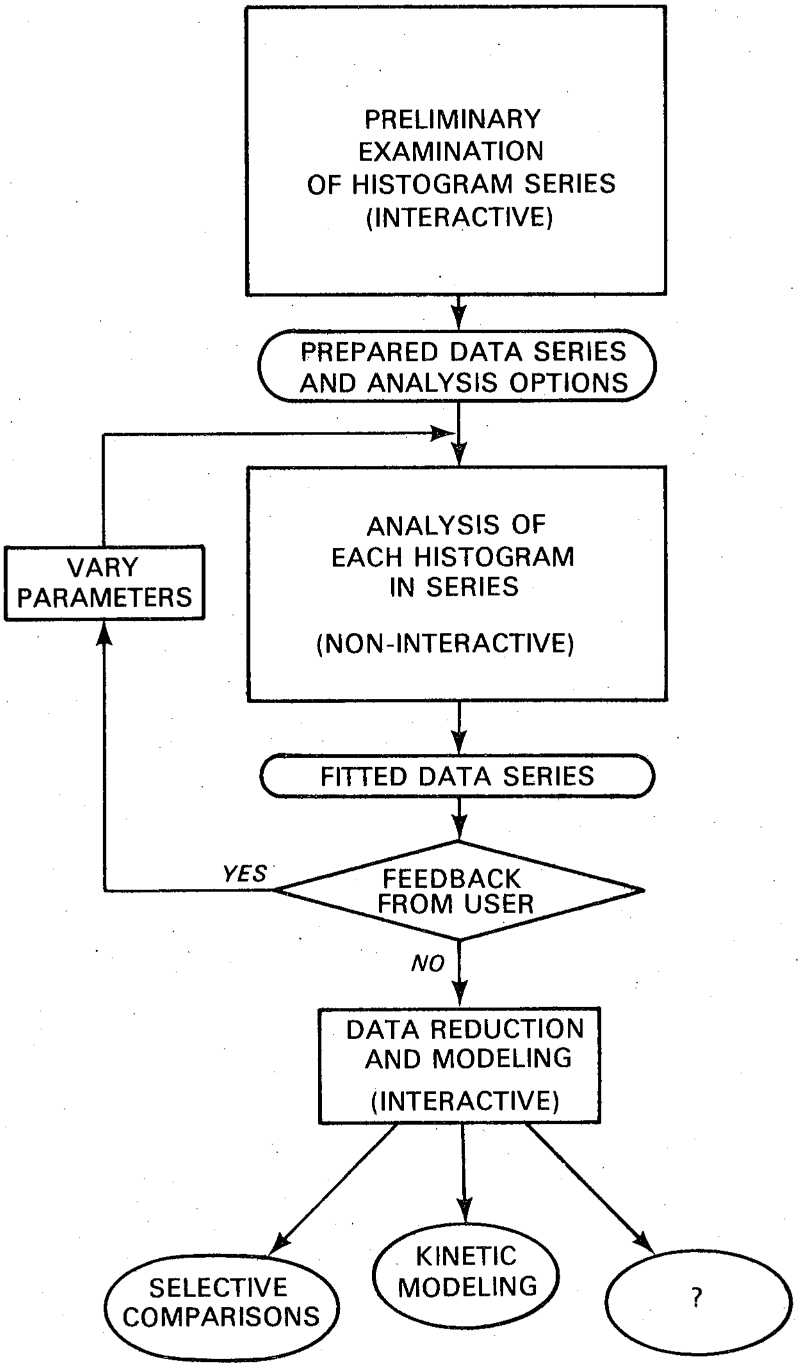
Figure 3



XBL788-1660

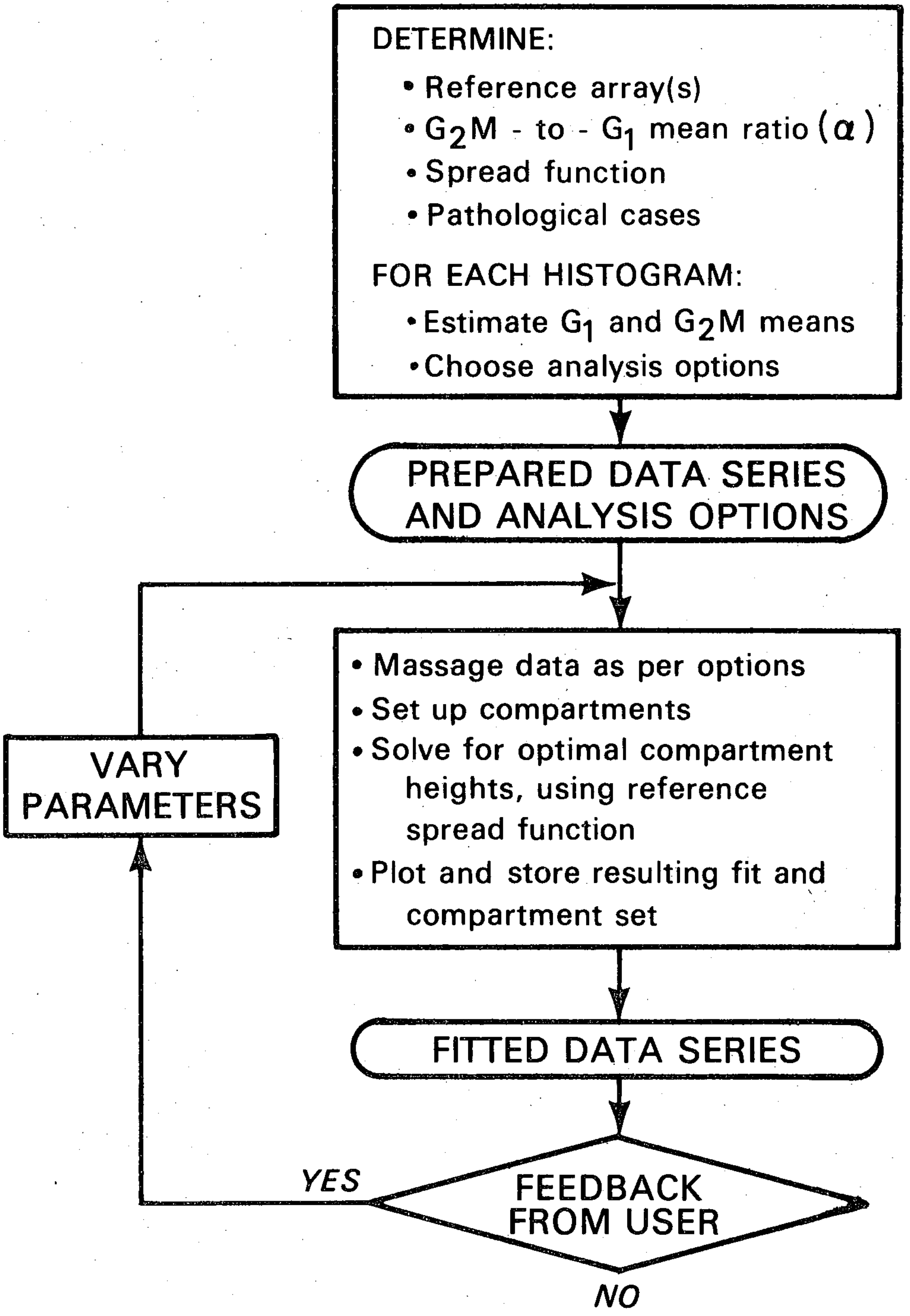
Figure 4a

ANALYSIS OF A SERIES OF DNA HISTOGRAMS
Overall approach



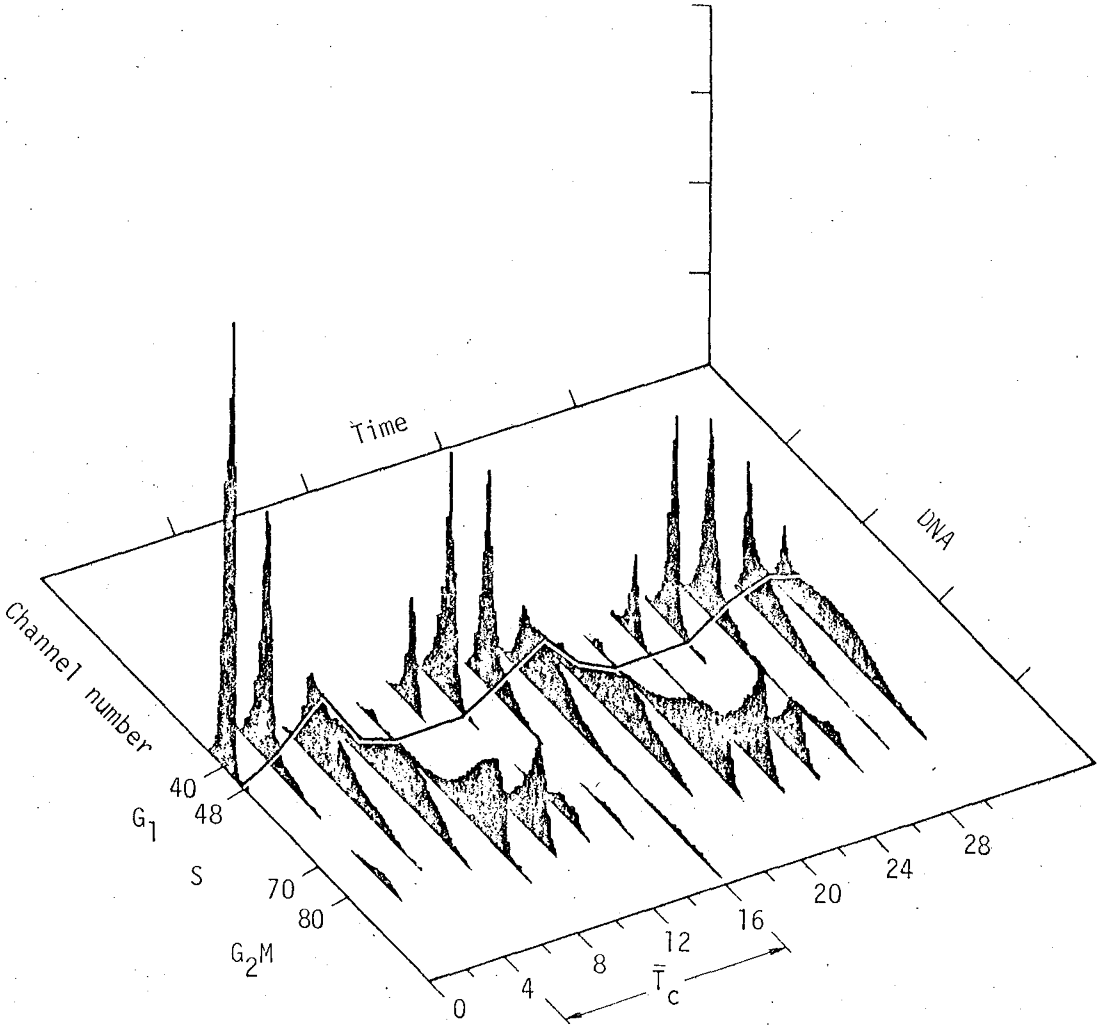
XBL 7712-11469

Figure 4b



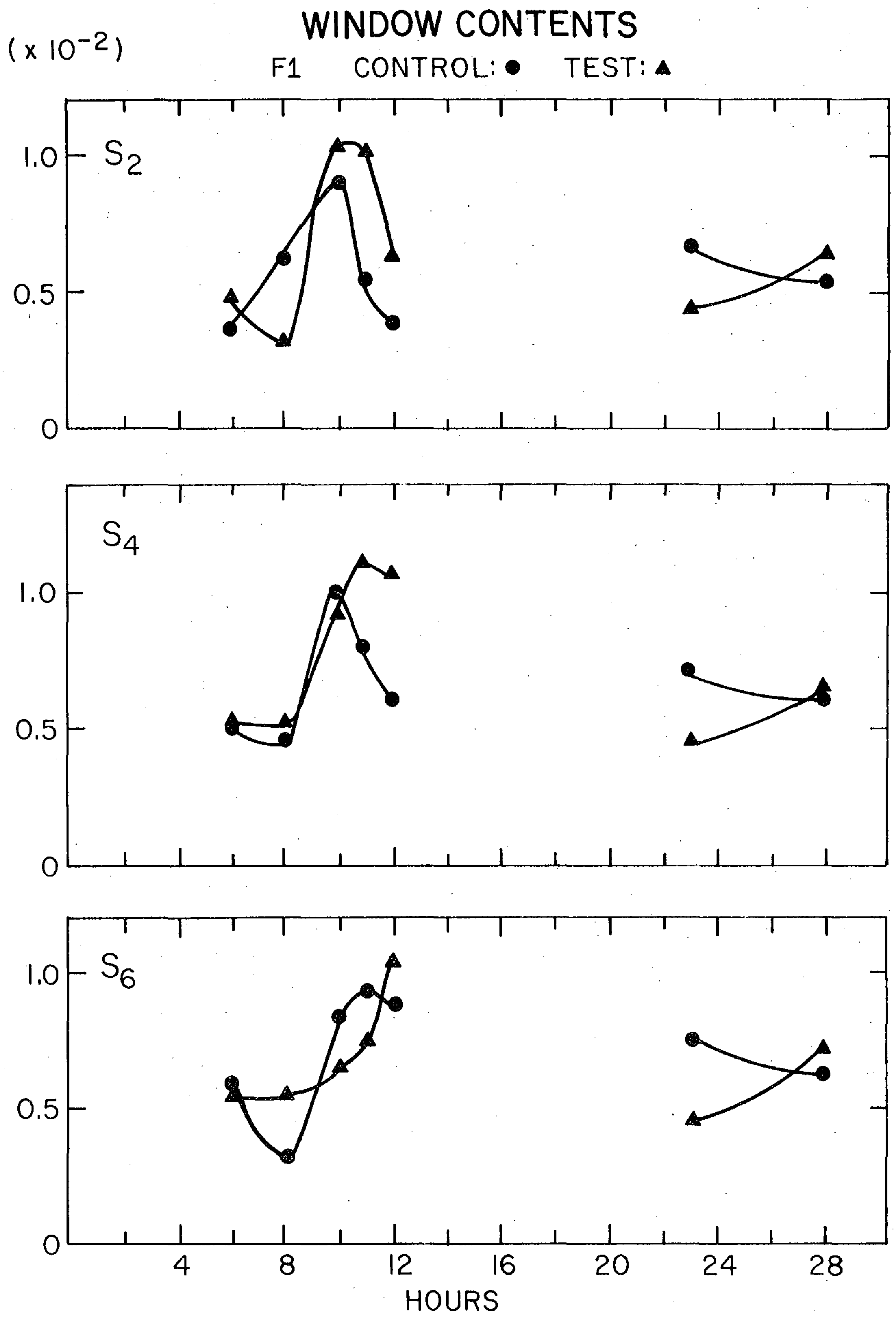
XBL 7712-11466

Figure 5



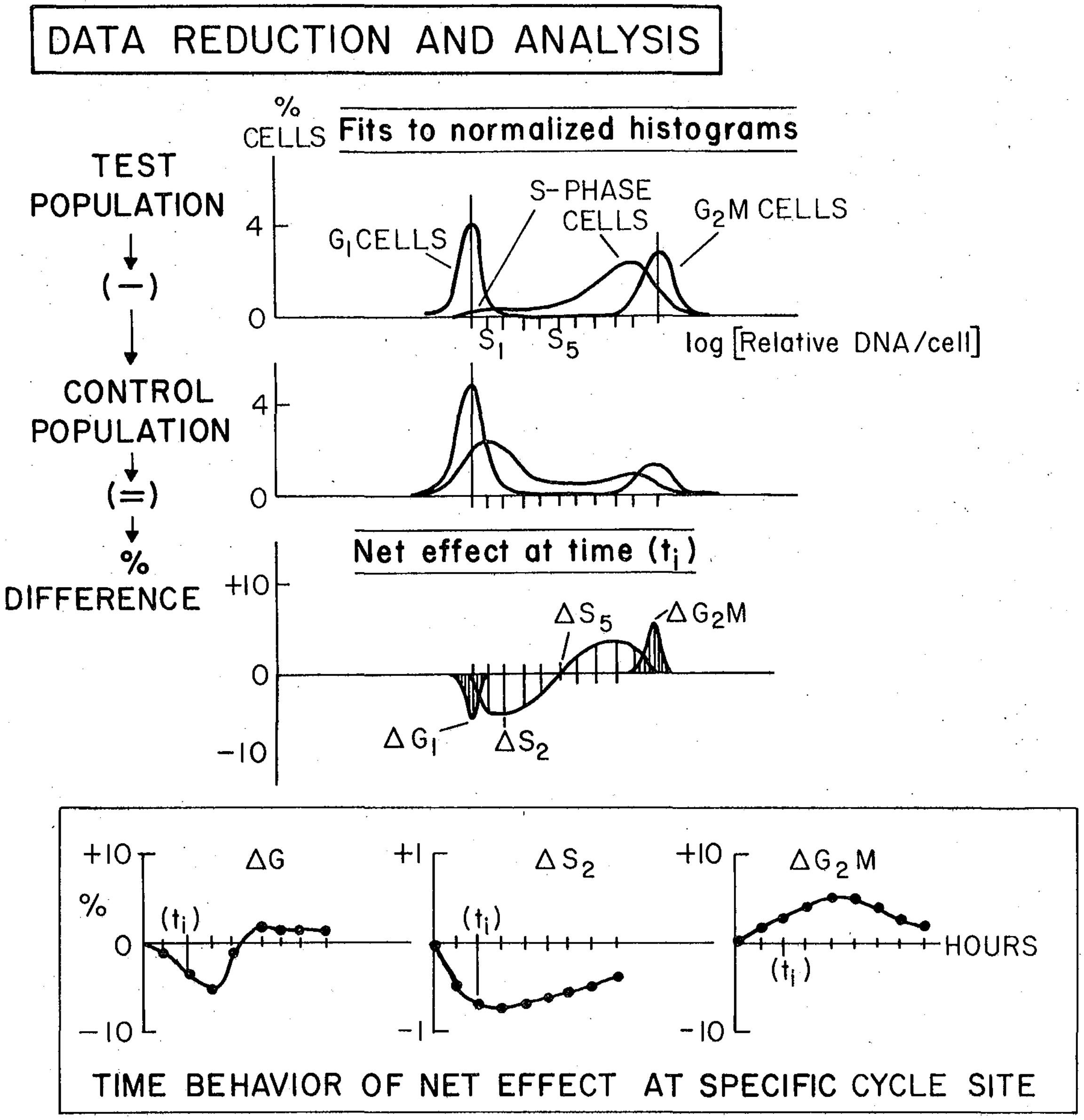
XBL788-1659

Figure 6



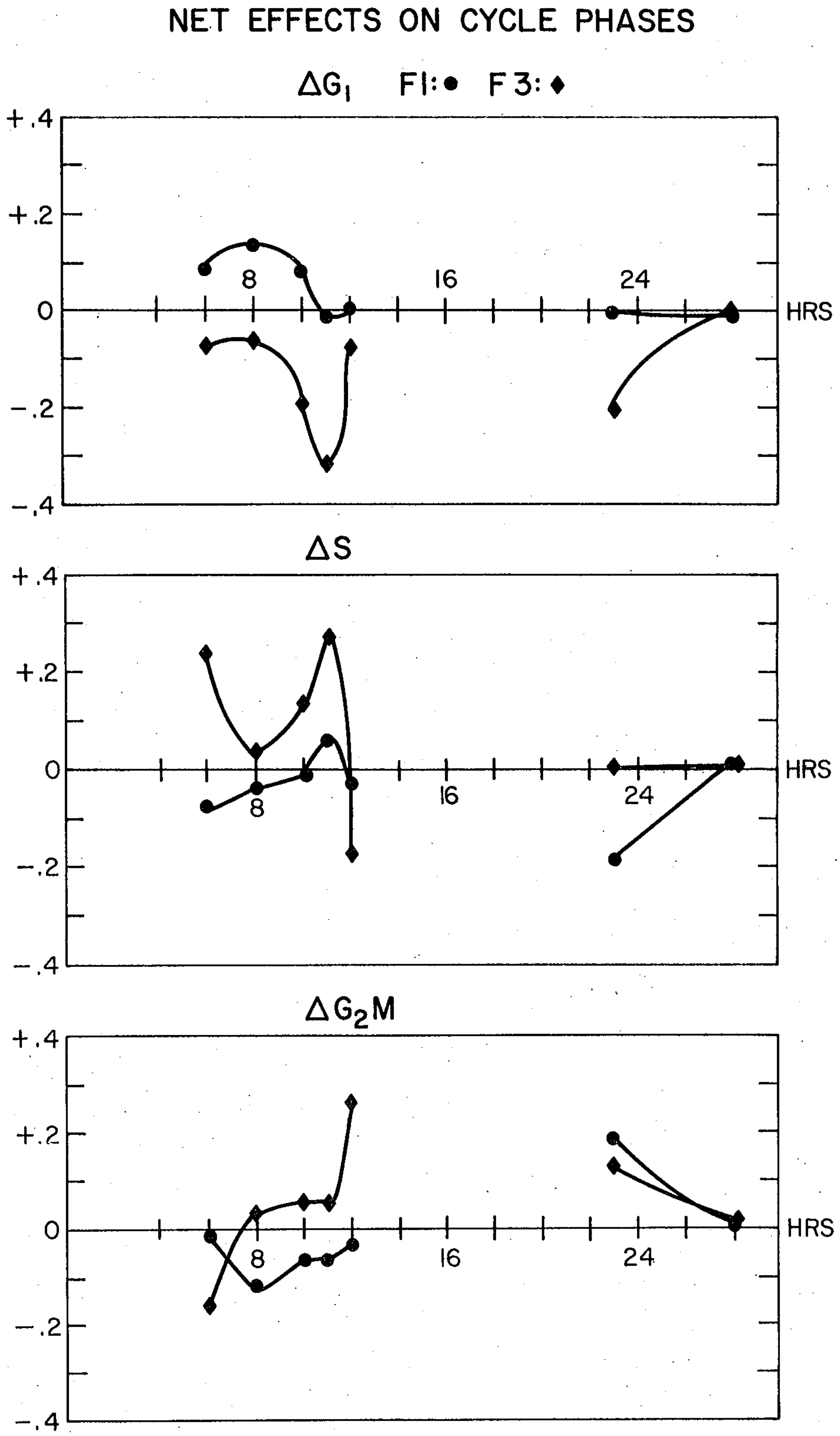
XBL 784-3935

Figure 7



XBL 784-3931

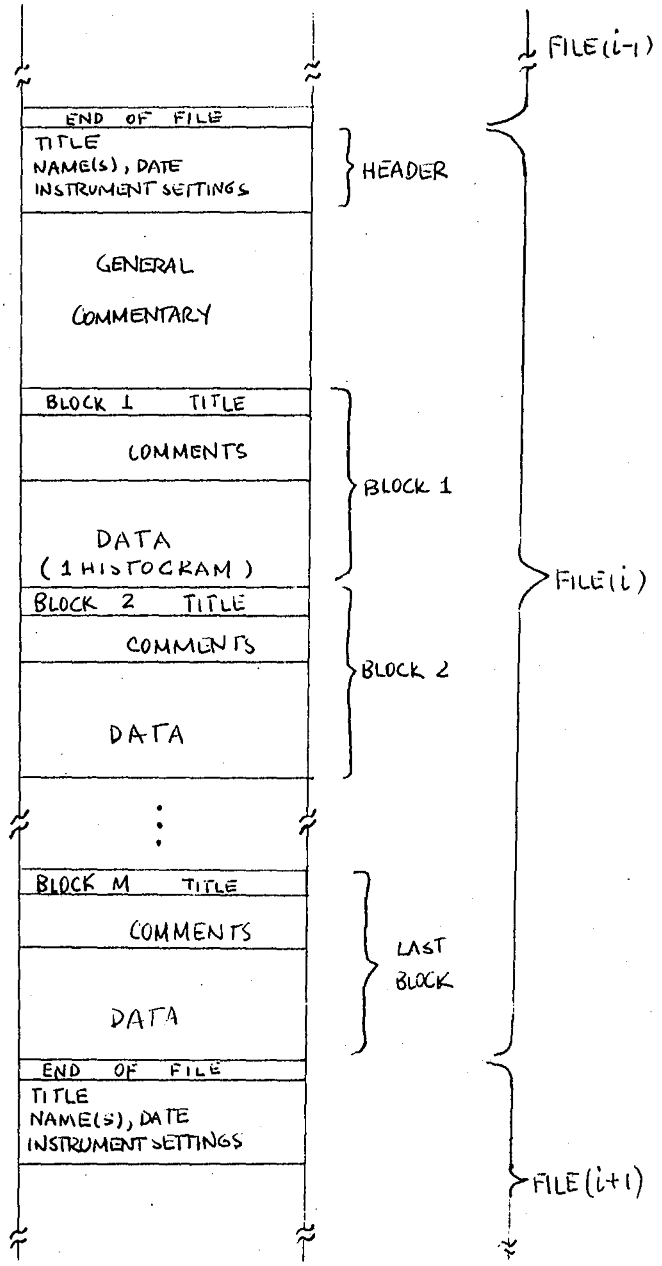
Figure 8



XBL 786-4042

Figure 9

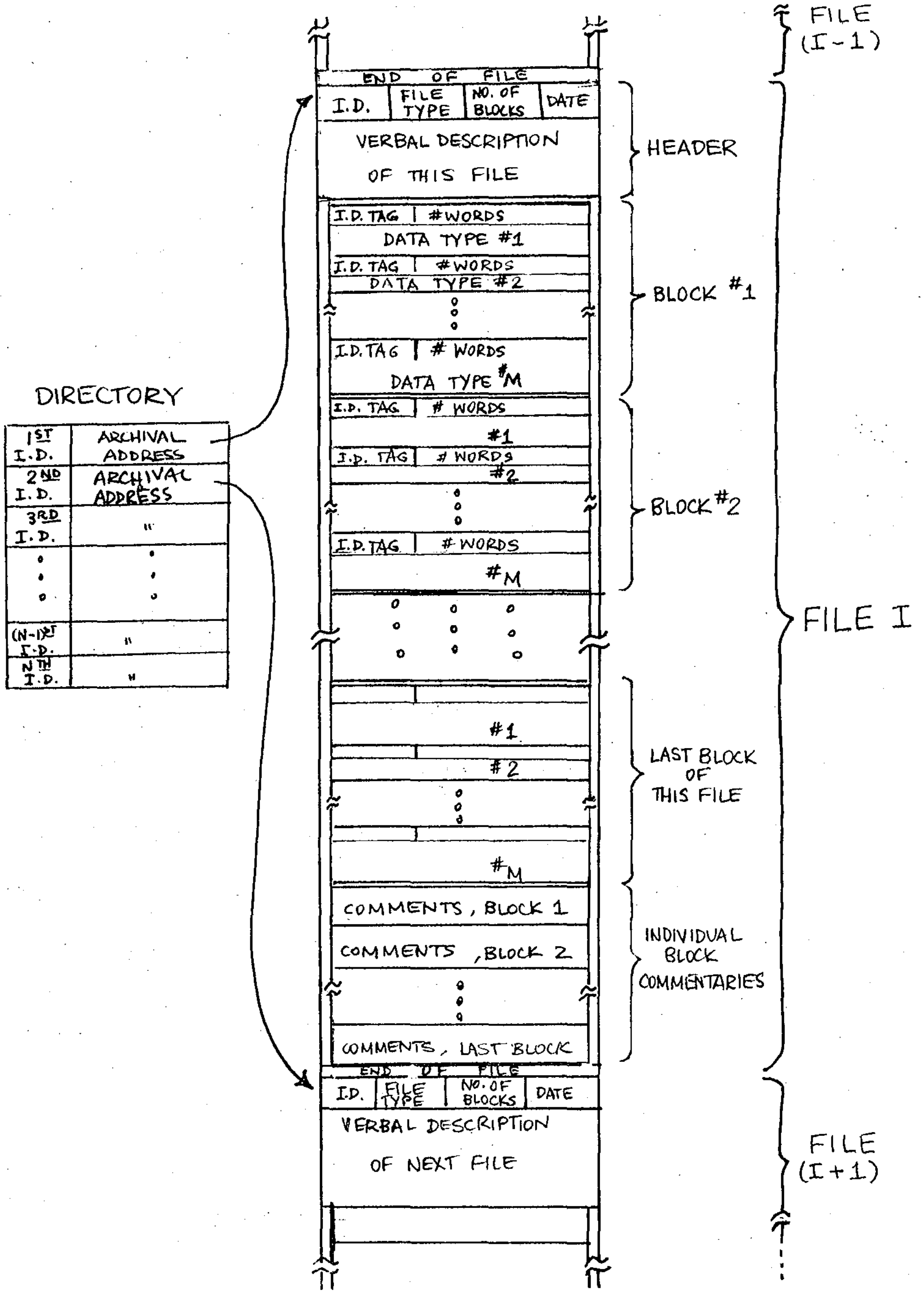
FILE STRUCTURE
FOR DNA HISTOGRAM STORAGE



XBL 788-10615

Figure 10

ARCHIVAL RECORD STRUCTURE



XBL 788-10616

CHAPTER V

BACKGROUND TO THE CELL CYCLE STUDIES ON EFFECTS OF B[a]P DIOL EPOXIDE UPON NMuLi CELLS

A wealth of literature has accumulated over the past several years on the metabolism (1,3), mutagenicity (2), interaction with DNA (3), potential for malignant transformation (4), and tumorigenicity (5) of a whole class of compounds known as polycyclic aromatic hydrocarbons (PAH). The great interest and concern regarding PAH stems largely from the fact that they are produced in metric ton quantities daily, as unwanted biproducts of incomplete fossil fuel combustion in power plants, automobiles, cigarettes, barbeques, and a large variety of other sources. This concern is further heightened by the expected increase in the use of coal as an energy source, combustion of which produces more PAH per gram than any other fossil fuel (6).

Most PAH are rather inert chemical species, requiring enzymatic conversion to active forms in order to cause mutation or malignant transformation (7). It is perhaps ironic that some of the same enzyme systems that normally detoxify harmful ingested chemicals are involved in the activation of PAH, in particular the aryl hydrocarbon hydroxylase enzyme (AHH) of the cytochrome monooxidase system P-450. The irony is further compounded by the fact that many PAH have been shown to induce biosynthesis of AHH (8), which then metabolises the compounds to numerous harmful forms.

WHY BENZO[a]PYRENE DERIVATIVES?

Among the PAH, the five-ring compound Benzo[a]pyrene (BAP), Fig. 1, has become a focus for chemical carcinogenesis studies, and extensive work has resulted in considerable elucidation of its metabolic conversion by cells, Fig. 1, and interaction with subcellular components and macromolecules. Recent studies have identified what is believed to be the ultimate carcinogenic metabolite of BAP: the (+-)-trans-7β,8α-Dihydroxy-9α,10α, -epoxy-7,8,9,10-tetrahydrobenzo(a)pyrene henceforth denoted as DE2[1], and illustrated in Fig. 1. This derivative has been shown to be the most mutagenic (9) and the most carcinogenic (5) of the BAP metabolites, and has been shown to bind directly to DNA bases (10), as illustrated in Fig. 2. Furthermore, the formation of small numbers of DE2-DNA adducts (1 adduct per 5000 bases) has been shown to halt DNA synthesis in viruses (11), while DE2 slows DNA synthesis in cultured mouse fibroblasts (12), and in mouse epitheial cell cultures (13).

---

[1] DE1 is the (±)-cis- version of DE2.

These observations have been seen by some as evidence for the hypotheses that interference with normal DNA synthesis is a vital step in the malignant transformation of cells (13). The strong (though not complete) correlation between mutagenicity and carcinogenicity of most tested chemical compounds is taken as further testimony to the role of direct carcinogen-DNA interaction in carcinogenesis. In addition, cell cycle specificity in the susceptibility of cells to the action of chemical carcinogens has been suggested (13,15), with the implication that those cells actively synthesizing DNA are the most easliy transformed by such agents. The above reasons, as well as the fact that malignant cells usually display different growth kinetics from those of their pre-malignant counterparts, have generated much interest in the nature of cell cycle perturbation by carcinogens.

CHOICE OF CELL SYSTEM

Against this background, I have applied the cell cycle kinetic analysis system described in this thesis to the study of the effects and specificity of DE2 on the cell cycle of an epithelial cell line, NMuLi. The goal was not to investigate chemical carcinogenesis itself, but rather a related phenomenon of importance in the effort to understand carcinogenesis. Choice of the line was dictated largely by practicality, as much work with BAP derivatives had already been done on these cells in our laboratory, but the line had originally been chosen for deeper reasons by my colleagues. Their interest in the study of carcinogen action on epithelial cells had its origin largely in the fact that the overwhelming majority of tumors arise from epithelial tissues (15), and the related fact that it is epithelium that is most directly exposed to the environment and to ingested material, including the carcinogenic compounds in them. We were fortunate to obtain one of the relatively few lines of epithelial cells available, NMuLi, derived from the livers of Namru mice by Owens, et al (16). The scarcity of such lines is due to the difficulty of isolating pure epithelial from fibroblastic cells, the latter of which tend to overgrow the former when seeded together; a problem which Owens was able to solve by clever use of sedimentation separation methods.

In working with the NMuLi cells, my colleagues found them to possess an unusually high inducibility for AHH biosynthesis (17), which has led to extensive studies on the metabolism of BAP by these cells. However, since I have been interested in the effects of pre-synthesized DE2, there was no need for metabolic activation of the agent, and therefore little direct use has been made of the BAP metabolism studies, in this work.

When grown in plastic dishes in a 5% CO_2, humidified environment at $37^{\circ}C$, in minimal essential medium (MEM) supplanted with 10% calf serum and 10% insulin, the NMuLi cells have retained their epithelial morphology in monolayer culture for more than 50 passages (about one year). Under these conditions, in logarithmic growth phase, the doubling time of these cells is about 15 hours (±1hr.), and based upon analysis of their DNA distributions obtained by flow cytometry, the approximate durations of the G_1, S, and G_2M cycle phases are 4.5 hrs., 5.5 hrs., and 5.0 hrs., respectively (14). Cell proliferation ceases at a confluent density of about 2.0×10^5 cells/cm^2, and upon replating at lower density and in medium supplanted with 20% calf serum, cells are first detected entering synthesis phase after about 8 hours.

Though derived from liver, the NMuLi line has maintained a largely diploid karyotype. However, as with many cell lines, a high rate of spontaneous transformation has characterized the NMuLi line, with the result that beyond the early passages, the cells must be considered transformed, even without treatment by carcinogens. Hence, the NMuLi line has served as a convenient target cell population for studies of cycle perturbation by DE2, somewhat decoupled from the question of actual malignant transformation by the agent. A useful next step to these studies would be to repeat them on early passage NMuLi cells (believed to be non-transformed), for comparison.

REFERENCES

1. Heidelberger, C. Chemical Carcinogenesis. Ann. Rev Biochem 44:79-121 (1975)

2. Wood, A.W., Wislocki, P.G., Chang, R.L., Levin W., Lu, A.Y.H., Yagi, H., Hernandez, O., Jerina, D.M., Conney, A.H. Mutagenicity and Cytotoxicity of Benzo-Ring Epoxides. Can Res 36:3358-3366 (Sept., 1977)

3. Yang, S.K., Gelboin, H.V., Trum, BF, Autrup, H., Harris, C.C. Metabolic Activation of Benzo[a]pyrene and Binding to DNA in Cultured Human Bronchus. Can Res 37:1210-1215 (April, 1977)

4. Bresnick, E., McDonald, T.F., Yagi, H., Jerina, D.M., Levin, W., Wood, A.W, Conney, A.H. Epidermal Hyperplasia after Topical Application of Benzo[a]pyrene, Benzo[a]pyrene Diol Epoxides, and other Metabolites. Can Res 37:974-980 (April, 1977)

5. Levin, W., Wood, A.W., Wislocki, P.G., Kapitulnik, J., Yagi, H., Jerina, D.M., Conney, A.H. Carcinogenicity of Benzo-Ring Derivatives of Benzo[a]pyrene on Mouse Skin. Can Res 37:3356-3361 (Sept., 1977)

6. Calvin, M. (private communication)

7. Miller, J.A. Can Res 30:559-576 (1971)

8. Landolph, J., Bartholomew, J., Calvin, M. Quantitative Studies of the Toxicity of Benzo[a]pyrene to a Mouse Liver Epithelial CellS Strain in Culture. Can Res 36:4143-4151 (Nov., 1976)

9. ibid, ref. 5

10. Meehan, T., Straub, K. and Calvin, M. Benzo[a]pyrene Diol Epoxide Covalently Binds to Deoxyguanosine and Deoxyadenosine in DNA. Nature 269(5630) 725-727 (Oct., 1977)

11. Hsu, W., Lin, E.J.S., Harvey, R., and Weiss, S. Mechanism of Phage phiX-174 DNA Inactivation by benzo[a]pyrene-7,8-dihydrodiol-9,10-epoxide. Proc Natl Acad Sci USA (74)8:3335-3339 (Aug., 1977)

12. Petersen, A.R., Bertram, J.S., and Heidelberger, C. Cell Cycle Dependency of DNA Damage and Repair in Transformable Mouse Fibroblasts Treated with N-methyl-N'-nitro-N-nitrosoguanidine. Can Res 34:1600-1607 (1974)

13. Bartholomew, J.C., Pearlman, A.L., Landolph, J.R., and Straub, K.M. Modulation of Cell Cycle Parameters of Mouse Epithelial Cells by Benzo[a]pyrene and Some of its Derivatives. (manuscript in preparation)

14. Marquardt, H. Cell Cycle Dependency of Chemically Induced Malignant Trasformation in Vitro. Can Res

34:1612-1615. (1974)

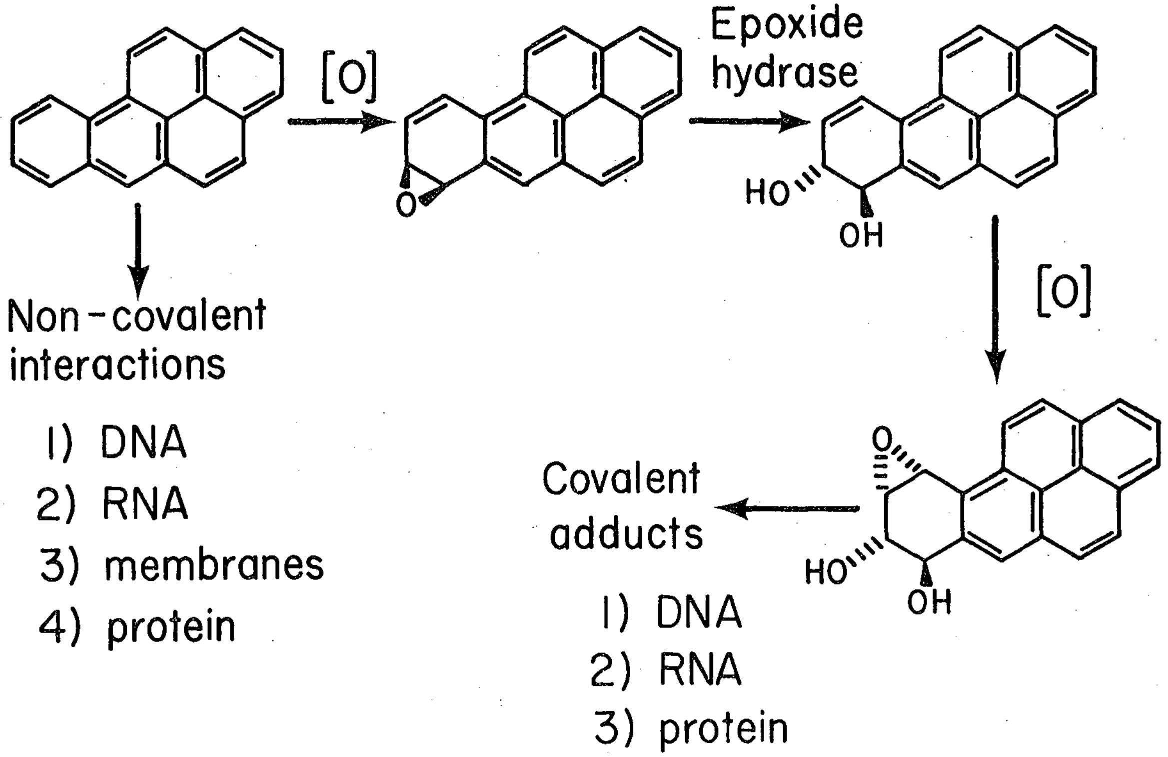
15. Higginson, J. Environment and Cancer, 24th Symp. Fund. Canc. Res. 69-92, Baltimore, Williams and Wilkins (1972)

16. Owens, R. B., Smith, H.S., and hacket, A.J. Epithelial Cells from Normal Glandular Tissue of Mice. J Natl Cancer Inst 53:261-269, 1974.

17. Landolph, J.R. The Interaction of Benzo[a]pyrene with Mouse Liver Epithelial Cells. Doctoral dissertation, Chem. Dept., University of California at Berkeley, 1977.

Figure 1

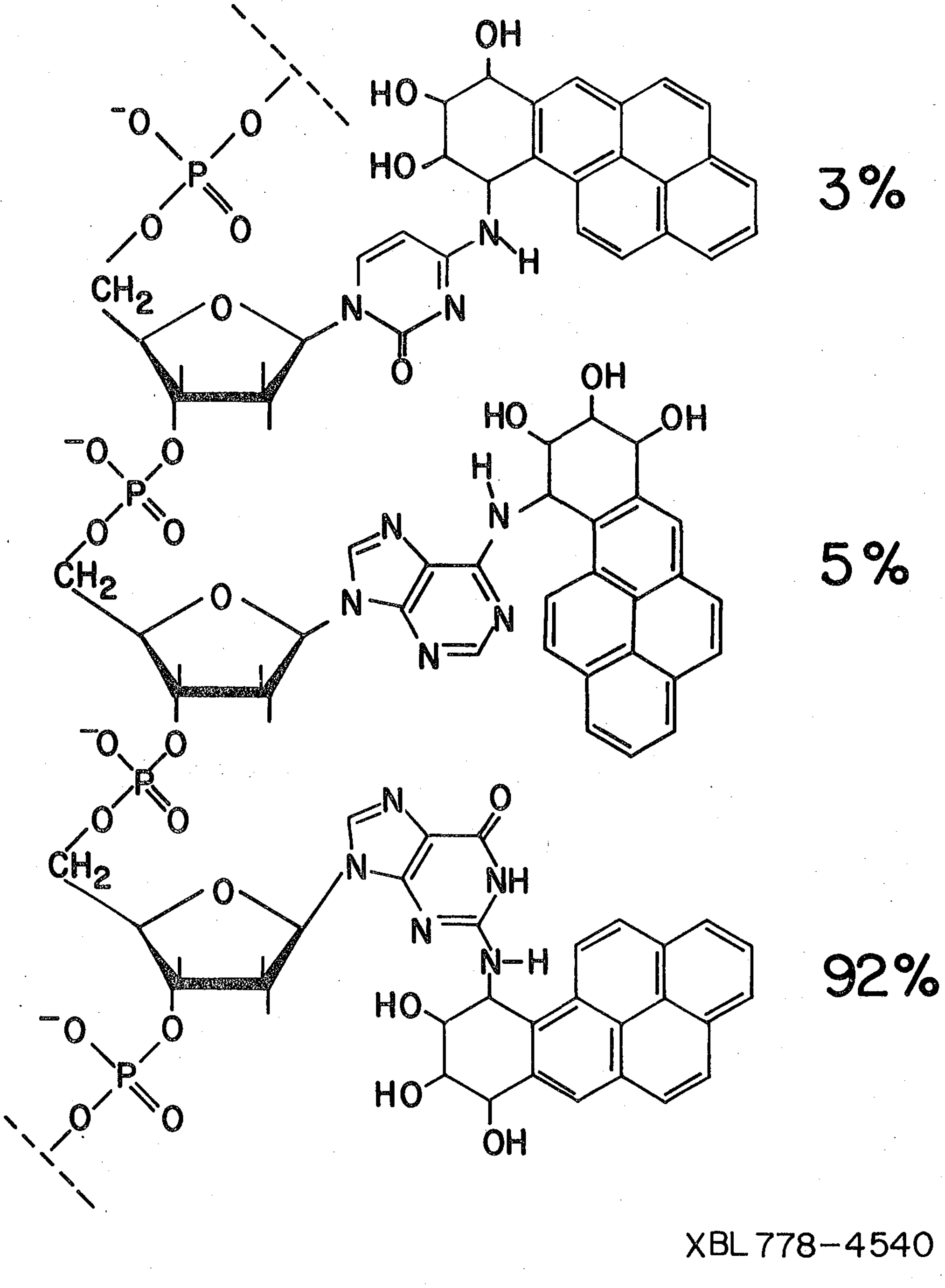
ACTIVATION OF BENZO[a]PYRENE



XBL776 - 4444

Figure 2

STRUCTURE BaP-DNA ADDUCTS AND DISTRIBUTION OF BaP AMONG BASES



XBL 778-4540

CHAPTER VI

Cell Cycle Effects and Dependency of a Carcinogenic Metabolite of Benzo[a]pyrene

SUMMARY

Subtoxic amounts of the 7,8-dihydroxy-9,10-epoxide (DE) derivative of benzo[a]pyrene were added to replated cultures of a mouse liver epithelial cell line, NMuLi, that had been synchronized in various stages of the cell cycle by centrifugal elutriation. The elutriation samples chosen for replating included populations enriched in G_1 (>60%), in S (>70%) and in G_2M (>50%) phases, obtained from an initially asynchronous NMuLi population (38% G_1, 36% S, 26% G_2M).

DNA histograms of the replated samples were obtained by flow cytometry of the samples at several timepoints, analysed by a computer modeling routine, and reduced to a few charts illustrating the "net effect" of DE relative to controls.

DE slowed S-phase traversal by about 40% in all samples, relative to their respective controls, as judged by the transit time of cohorts from early S-phase to G_2M. Traversal through G_2M was also slowed by at least 50% in several of the fractions. DE did not appear to affect G_1 traversal by cycling cells.

Cell cycle dependency in susceptibility to the agent was demonstrated. When added to cycling G_1 cells, DE delayed the apparent recruitment of noncycling G_1 (G_0) cells by about 2 hrs. When added to cycling S-enriched samples, DE elicited the strongest "net effect", indicating that S-phase cells were the most sensitive to DE action. The effects of DE were not transient in nature, but persisted for at least two doubling times, reflecting altered growth kinetics as well as initial cycle perturbation.

INTRODUCTION

Many recent articles (1,2,3) argue that chemical substances account for the vast majority of environmental carcinogens, and that they are primarily industrial in origin. In particular, effluents released as biproducts of incomplete fossil fuel combustion include a wide variety of polycyclic aromatic hydrocarbons (PAH) which have been shown to be both mutagenic and carcinogenic (1). In view of the expected increase in the use of fossil fuels as energy sources, much concern has been raised in regard to their potential carcinogenic risk, which has motivated, in turn, much study of the biochemistry of PAH and their effects, on biological systems.

Like many other PAH, benzo[a]pyrene (BAP) has been shown to require enzymatic conversion to an active form, which is then capable of interaction with critical cellular macromolecules (8). The metabolic activation of BAP has been the subject of intensive investigation during the past several years (8), from which has emerged the concensus that the ultimate carcinogenic derivative of BAP produced by the aryl hydrocarbon hydroxylase metabolism is the 7,8-dihydro-9,10-epoxide derivative "DE" , and its (-)anti- stereoisomer in particular (8).

DE has been shown to form adducts with DNA bases (6,9), and there is evidence that such adduct formation slows or halts DNA replication (21). In radioactive tracer studies, carcinogens have been shown to exhibit cell cycle dependency in malignant transformation (5), and in DNA damage and repair (4). These findings, and others, have motivated the hypothesis that direct interference in normal DNA synthesis is a primary mechanism in malignant transformation of cells (9).

In probing this hypothesis, Bartholomew, et al[1] investigated the cell cycle effects of various metabolites of BAP, including DE, in asynchronously growing and in stationary cultures of mouse liver epithelial cells, whose metabolism of BAP had been the subject of previous study (7). Of all metabolites tested, DE produced the most dramatic cell cycle perturbations, which involved a general increase in the fraction in S-phase.

In this study, we have employed several new methods to determine the effects of DE upon subpopulations of cells enriched in different cycle phases, to test whether the various phases respond differently to DE. Our overall approach was to obtain parallel time series of DNA histograms (by flow cytometry) of DE-treated and control cells, replated after cycle enrichment by centrifugal elutriation and sampled at several time points thereafter. We then sought to extract from the large set of histograms a few "net effect" curves, which depict the calculated difference between treated and control histograms at several distinct sites in the cycle, for every timepoint. These, together with the original time series, were used to discern and compare trends in the kinetic behavior of the two parallel populations.

MATERIALS AND METHODS

CELL LINE AND CULTURE TECHNIQUE

Mouse liver epithelial cells NMuLi, derived by Owens (10), were propagated in monolayer cultures for 40 passages prior to isolating a clone demonstrating particularly high AHH levels (clone 8). Clone 8 cells were seeded in 100mm plastic dishes (Falcon plastics, Oxnard, CA) in Eagle's Minimal Medium, "MEM", (GIBCO, Grand Island, NY) supplemented with 10% donor calf serum (Flow Laboratories, Rockville, MD), and 10 µg/ml insulin (Calbiochem, San Diego, CA) and grown in a humidified atmosphere containing 5% CO_2 at 37°C. Under these conditions, cells showed a doubling time of 14-16 hrs and ceased proliferating at densities of about

1.3×10^5 cells per square centimeter.

48 hrs prior to harvesting for elutriation, 10 roller bottles (Corning) were seeded at 2×10^7 cells per bottle (about 20% of the confluent density) in MEM supplemented with 10% donor calf serum and 10 µg/ml insulin. Transfers from dishes to bottles were done by aspirating the medium, washing once with saline GM (1.5mM Na_2HPO_4, 1.1mM KH_2PO_4, 1.1mM Glucose, and 0.14M NaCl, at pH 7.4), loosening the cells from the surface by 3 min trypsinization at 37°C. with DISPO (saline GM containing 0.5mM EDTA and 0.1 mg/ml crystalline trypsin -- GIBCO, Grand Island, NY) and washing the cells from the plates with NEUT (Saline GM containing soybean trypsin inhibitor, bovine serum albumin, DNAse, calcium and magnesium ions; reference (18)) The resulting suspension was spun down and the pellet resuspended in MEM with serum, which was subsequently diluted and divided among the roller bottles. Harvesting from the bottles followed the same general procedure, without the pelleting step, yielding 8.4×10^8 cells for loading into the elutriator rotor in NEUT suspension. After elutriation, cells were seeded in 100mm plastic dishes (Falcon) for subsequent timepoints.

ELUTRIATION CENTRIFUGATION

Centrifugal elutriation was chosen as the method of cycle phase enrichment because of its ability to gently produce large samples of cells enriched in different cycle phases from the same parent population (19). Details of the elutriation apparatus and technique employed are given elsewhere[2], but briefly, cells were elutriated at a fixed counterflow rate of 25 ml/min, removing subpopulations from the separation chamber by decrementing the rotor speed. Speed decrements necessary to produce the desired cycle enrichments were calculated from a model relating relative sedimentation rate (RS), counterflow rate and rotor speed, where the first cells to emerge from the rotor (generally G_1 cells) were assigned an RS of 1.0. With this approach, G_1, S, and G_2M enriched samples were obtained fairly reproducibly at RS values of 1.0-1.2, 1.9-2.1, and 2.5-2.7, respectively. (detailed determination of these values is discussed in reference 11) In this experiment, a minimum of 20-30 million cells per collected sample was necessary to provide sufficient cells for subsequent timepoints. Of the 7 speed decrements used, 4 yielded sufficient cells for replating, and they are shown in Chart 1.

CHEMICAL CARCINOGEN

A racemic mixture of (+) and (-) anti-DE was synthesized in our laboratory by Kenneth Straub, and generously made available for our use. The carcinogen was administered to the cell samples by first dissolving it in DMSO, then adding the solution to the cell suspensions ready for replating. The final DE concentration was 0.2 μgrams/ml, with the DMSO concentration at 0.2% in the replate suspension. Control cells received the same amount of DMSO, without DE.

Estimates of the chemical half-life of the anti-form of DE range from 1 to 20 min (12)), so care was exercised in the preparation of the DE working solution to allow minimum time from its dissolution in the DMSO to its addition to the medium. This minimized the probability of the DE being hydrolysed to a tetrol form prior to uptake by the cells. When administered in this manner, DE has been previously shown to enter the NMuLi cells (18). Toxicity studies have shown that at this dose level, no cell death is detected for at least 48 hrs post administration, though by 5 days the survival fraction drops to 20-25% of the original cell number[1].

DNA HISTOGRAMS

Cells were harvested as described above, spun out of NEUT, and resuspended in fixative (15mM $MgCl_2$ in 25% EtOH) containing 50 μg/ml RNAse (Calbiochem, San Diego, CA) for 60 min at 37°C. Samples were then spun down and washed once in saline GM, and resuspended in propidium iodide (PI) staining solution (50 μg/ml in saline GM) at room temperature for 30 mins, from which the cells were spun down and finally centrifuged and resuspended in saline GM for input to the flow cytometer. Typical sample sizes were 3.0×10^5 to 2.0×10^6 cells in 2 ml saline GM.

Samples were analysed by a Los Alamos Flow System II flow cytometer (microfluorimeter) described by Holm and Cram

(15). A 2 watt, 514 nm excitation line from an argon ion laser (Spectra Physics Model 171, Mountain View, CA) was focused to 17 μ diameter in the direction traversed by the cells, which were analysed at a rate of 2-3,000 per second, flowing through the exciting beam at 4-5 meters/second. Emitted fluorescence from each cell was measured, digitized and stored in one of 256 channels of a pulse height analyser (Northern Scientific), building up the resultant histogram of cell number versus measured fluorescent intensity per cell. DNA histograms are presented here after some manipulation to facilitate visual examination: computer simulated gain-shifting (linear transformation of the abscissa) to allign the G_1 peaks of each histogram in channel 100, and normalizing to a constant total cell count per histogram (yielding population probability distributions).

ANALYSIS AND REDUCTION OF HISTOGRAM DATA

Individual histograms were analysed by a deconvolution modeling program developed by Pearlman[3], to yield estimated population percent in G_1, S, and G_2M cycle phases, as well as in 9 S-phase subcompartments. These 12-value data sets from each 256 point histogram were stored for later point-by-point comparison between treated and control histograms[4]. Calculated difference spectra between treated and control data sets were generated for each time point, resulting in time series of "net effects" on each of the 12 cycle subdivisions.

RESULTS

Chart 1 shows the comparative effects of DE+DMSO, DMSO, and no additive, upon the F1 replate fraction. DMSO is seen to cause only minor perturbations in comparison to those of DE, and will henceforth in this chapter be referred to as "control." The histograms reveal DE inhibition of S-phase traversal, as judged by the transit time of cohorts from early S-phase to G_2M. as seen in Chart 2, a cohort of cells (indicated by arrow in chart) leaves G_1-phase at 4 hrs post replating, in both control and DE-treated samples. While the cohort reaches G_2M by 8 hrs in the control sample, a significant G_2M peak is not detected until 10 hrs in the DE-treated sample. This delayed arrival time reflects a 2 hr lengthening, or about a 40% slowing of traversal through S-phase. Similar findings were seen in the other replated fractions , seen in Chart 3.

Progress through G_2M appears to have been slowed by DE, as seen, for example, in the histograms for sample F4 of Chart 3. The absense of early S-phase cells after 6 hours in these histograms makes it possible to infer G_2M length from the behavior of G_1, since any change in G_1 is due solely to the influx of cells from G_2M. For sample F4, it is seen that by 6 hours, the control G_1 peak has grown considerably, whereas the G_1 peak of the DE-treated population does not increase significantly until 8 hours. Since both populations had large G_2M fractions at time 0, this implies that DE has effectively lengthened G_2M by at least 2 hours, relative to controls. Similar patterns for G_1 and G_2M are seen in the histograms for samples F2 and F3.

As can be further seen from chart 2, two cohorts of cells leave G_1 from both control and treated populations. In the control, one cohort moves into early-S by 4 hrs and the second at 8 hrs. The first cohort was apparently actively growing in G_1 phase when replated, as it had already moved considerably into S-phase by the 4 hr timepoint. The second cohort (8 hours post replating) was apparently recruited from G_1 cells that were not actively growing at replate time, since it emerged 4 hrs after the first cohort, and since G_2M or S-phase cells present at time 0 were not likely to have cycled around to early-S by 8 hrs post replating. This is supported by the time series for sample F4, seen in Chart 3, where the late-S and G_2M cells seen at time 0 do not appear in early-S until 11 hrs post replating.

DE had little if any effect on the time of emergence of the first cohort in F1, indicating that traversal of G_1 was unaltered by DE in actively growing G_1 cells.

However, DE delayed apparent G_0 recruitment, as indicated by the delay in the appearance of the second cohort in the DE-treated sample. That cohort emerged from G_1 in the control sample at 8 hrs, as evidenced by the large shoulder on the right side of the G_1 peak, but not from the DE-treated sample until 10 hrs. Furthermore, the cohort from the DE-treated G_1 cells was less synchronous than its control counterpart, as seen in the "window contents" curves of Chart 5. These show the proportion of the population at each of several observation "windows" in early- and mid-S-phase, at the various time points. The passage of a cohort through a given window is indicated by the rise and then fall of the window height, and the synchrony of the cohort is seen from the time elapsed during the passage. In Chart 5, windows S4 and S6 show the control cohort to be moving through the window with sharper synchrony than the DE-treated counterpart. We are currently investigating the implications of this finding.

As is also seen in Charts 2 and 3, the DE effects were not transient, but resulted in significantly different distributions at all times, including the last timepoint at 28 hrs (roughly two doubling times) post replating. Since the histograms of the DE-treated populations do not appear to be just delayed versions of the controls, we infer that DE has more than an initial perturbing effect on these cells, altering their growth kinetics as well.

Cells initially synchronized in different cycle phases showed differing sensitivity to DE, as indicated by a great discrepancy in "net effect," calculated as described above. S-phase cells were apparently more sensitive to DE action than cells elsewhere in the cycle, as indicated in Chart 6.

The maximum net effect on each cycle phase, whether positive or negative, is seen to be generally greater for F3 than for F1. The same criterion, applied to the other samples, showed that F2 and F3 (the S-enriched samples) were more affected by DE than F1 or F4.

DISCUSSION

The hypothesis that chemical carcinogens act by perturbing normal DNA synthesis cannot account for all the observed effects of DE in this study. Progress through S-phase is indeed perturbed by DE in all samples, and S-phase cells are shown to be more sensitive to DE action than those in other parts of the cycle. However, the hypothesis does not address the evident effects on phases other than S, particularly the lengthening of G_2M and the apparent delay in the recruitment of noncylcling G_1 cells. The latter observations imply that besides DNA, other important cellular components may be affected by DE and thus be involved in exerting its perturbation of the mitotic cycle of these cells. The effects upon apparent recruitment could be further tested using stationary cultures analysed by the acridine orange and bromodeoxyuridine protocol of Darzynkiewicz (20) to detect cycling versus non-cycling G_1 cells in the DE-treated replates.

The effects of DE upon these cells are far more complex than is the case with familiar cycle perturbing agents such as hydroxyurea or colchicine in other cell systems. The DE action cannot be modeled by a single induced delay or block, rather, a complicated combination of effects is indicated.

Hydrolysed DE (tetrol) is apparently not responsible for the observed DE effects, since experiments analgous to these, using asynchronous NMuLi cultures, have shown that tetrol perturbs the cycle little more than DMSO, and far less than DE, at comparable doses[1]. The long lasting nature of the effects suggests incomplete repair of any damage caused by the DE. That the DE actually binds to the DNA has been demonstrated in our laboratory[5] by the use of radioactively labeled DE, and detection of labeled chromatin from treated cells. However, it is not clear whether the effects are due to direct interactions with DNA or are cytoplasmic in origin, or both. This question is currently under investigation in our laboratory.

The limiting factor in reproducing these experiments is the elutriation step, as even with identical initial populations loaded in the same number, it is virtually impossible to precisely duplicate the cycle enrichments obtained in any one run. It is for this reason that we have employed differences between test and control, where each has originated from the same elutriated sample.

One advantage of the present approach is the ability to simultaneously observe the effects of various agents on differently synchronized samples derived from the same parent cell population. However, a major factor limiting the extent of our observations, and consequent conclusions, has been the number of cells obtained in any one elutriated sample for replating. Our requirement that each histogram contain at least 300,000 cells limits the number of timepoints obtainable from a given sample. Improved elutriation methods should yield at least twice the present cell numbers per sample, with no loss of synchrony, enabling more replates and hence better time series.

The use of differences between fitted test and control histograms has enabled direct comparison of the response of cell populations possessing vastly different cycle distributions at the time of addition of the agent.

The use of fitted and reduced data has aided us in identifying many effects not readily noticed in the raw data, though it has not yielded (nor is it intended to yield) quantitative kinetic characterization of DE action. Perhaps sophistocated kinetic modeling programs such as Gray's (16) would be able to more precisely characterize the kinetic effects of DE from these histogram time series.

CHART LEGENDS

1. DNA histograms of elutriated samples used in this experiment. Ordinates: relative cell number. Abscissae:

relaive fluorescent intensity per cell (proportional to relative DNA content per cell.) (histograms normalized and alligned for direct visual comparison) Top: pre-elutriation growing NMuLi cell population; Bottom: samples F1, F2, F3, F4 collected at indicated relative sedimentation velocities (RS). Cycle phase fractions, listed next to each plot, were determined from computer fits to the histograms.

2. Histogram time series from replated sample F1 (mostly G_1 cells). Left column: DE-treated cells. Center column: DMSO-treated cells (controls). Right column: Control cells (no DMSO or DE added). Times indicated are hrs after replating in treated medium.

3. Histogram time series for fractions F2, F3 and F4. Left column: DE-treated cells. Right column: DMSO-treated cells. Notation is the same as for Chart 2.

4. Fractions in G_1, S, and G_2M cycle phases derived by computer modeling the histograms shown in Chart 2, for sample F1.

5. Time behavior of subcompartments in early and mid-S-phase for sample F1. Fraction of cells at sites S2 (15%), S4 (30%) and S6 (50%) through S-phase, derived from computer fits to histograms.

6. "Net effects" on cycle phase fractions for samples F1 and F3 . Calculated differences between DE-treated and

control cycle phase percentages for samples F1 and F3, derived from computer fits to histograms. Calculated differences between DE-treated and control subcompartments. Values derived from computer fits to histograms.

ACKNOWLEDGEMENTS

We gratefully acknowledge the valuable input of Mr. Hisao Yokota, of our laboratory, who helped us spot several important aspects of the data cited above. We also acknowledge the much needed assistance rendered us by Ms. Jean Lawson and Mrs. Maria Decosta, who made the cell culture aspects of this work possible. We also thank Mr. Sheldon Wong for implementing the simulated gain-shifting and area normalization programs on our computer, as well as the crucial data handling and storage software for our flow cytometer system. All of the above are members of the Laboratory of Chemical Biodynamics, Lawrence Berkeley Laboratory.

FOOTNOTES

1. Bartholomew, J.C., Pearlman, A.L., Landolph, J.R., and Straub, K.M. Modulation of Cell Cycle Parameters of Mouse Epithelial Cells by Benzo[a]pyrene and Some of its Derivatives. (manuscript in prep.)

2. Pearlman, A.L. Cell Cycle Synchronization of Epithelial

Cells by Elutriation Centrifugation (manuscript in preparation).

3. Pearlman, A.L., Navsky, B.N. Extraction of Information from Single DNA Histograms (manuscript in preparation).

4. Pearlman, A.L. Extracting Information from Sets of DNA Histograms (manuscript in preparation).

REFERENCES

1. Heidelberger, C. Chemical Carcinogenesis. Ann. Rev Biochem 44: 79-121 (1975)

2. Higginson, J. (1972) Environment and Cancer, 24th Symp. Fund. Canc. Res. 69-92, Baltimore, Williams and Wilkins

3. Higginson, J., Muir, C.S., (1973) Cancer Medicine, Ed. J.F. Holland, E. Frei III, 241-306, Philadelphia, Lea and Febiger

4. Petersen, A.R., Bertram, J.S. and Heidelberger, C. Cell Cycle Dependency of DNA Damage and Repair in Transformable Mouse Fibroblasts Treated with N-methyl-N'-nitro-N-nitrosoguanidine. Cancer Res. 34: 1600-1607 (1974)

5. Marquardt, H. Cell Cycle Dependency of Chemically Induced Malignant Transformation in Vitro. Cancer Res. 34: 1612-1615 (1974)

6. Bartholomew, J.C., Salmon, A.G., Gamper, H.B., and Cal-

vin, M. Benzo[a]pyrene Effects on Mouse Epithelial Cells in Culture. Cancer Res. 35: 851-856 (1975)

7. Pezzuto, J.M., Yang, C.S., Yang, S.K., McCourt, D.W., and Gelboin, H.V. Metabolism of Benzo[a]pyrene and (-)trans-7,8-Dihydroxy-7,8- dihydrobenzo(a)pyrene by Rat Liver Nuclei and Microsomes. Cancer Res 38: 1241-1245 (1978)

8. Peterson, A.R., Mondal, S., Brankow, D.W., Thon, W., and Heidelberger, C. Effects of Promoters on DNA Synthesis in C3H/10t1/2 Mouse Fibroblasts Cancer Res 37: 3223-3227 (1977)

9. Owens, R.B., Smith, H.S., Hackett, A.J. Epithelial Cell Cultures from Normal Glandular Tissue of Mice. J Natl Can Inst 53(1): 261-269 (1974).

10. Yang, S.K., McCourt, D.W., and Gelboin, H.V. The Mechanism of Hydrolysis of the Non-K-Region Benzo[a]pyrene Diol Epoxide r-7,t-8-Dihydroxy- t-9,10-oxy-7,8,9,10-tetrahydrobenzo[a]pyrene. J Am Chem Soc 99(15): 5130-5134 (1977).

11. Holm, D.M., and Cram, L.S. An Improved Flow Microfluorometer for Rapid Measurement of Cell Fluorescence. Exp Cell Res 80:105-110 (1973).

12. Gray, J.W. Cell Cycle Analysis o Perturbed Cell Populations: Computer Simulatin of Sequential DNA Distribu-

tions. Cell Tissue Kinet 9: 499-516 (1976).

13. Meehan, T., Straub, K. and Calvin, M. Benzo[a]pyrene Diol Epoxide Covalently Binds to Deoxyguanosine and Deoxyadenosine in DNA. Nature 269(5630) 725-727 (1977)

14. Horan, P.K., Romero, A., Steinkamp, J.A., and Petersen, D.F. Detection of Heteroploid Tumor Cells. J Natl Cancer Inst 52:843-848 (1974)

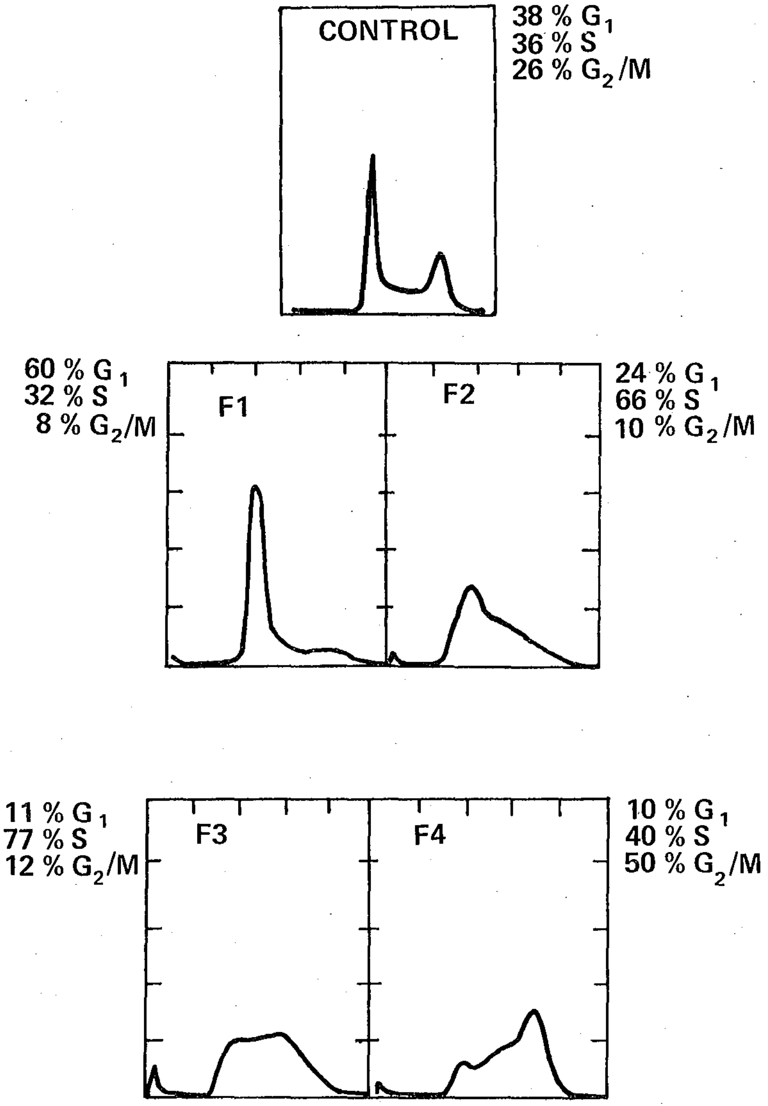
15. Meistrich, M.L., Meyn, R.E., and Barlogie, B. Synchronization of Mouse L-P59 Cells by Centrifugal Elutriation. Exp Cell Res 105:169-177 (1977).

16. Darzynkiewicz, Z., Andreeff, M., Traganos, F., Sharpless, T., and Melamed, M.R. Discrimination of Cycling and Noncycling Lymphocytes by Bromodeoxyuridine-suppressed Acridine Orange Fluorescence in a Flow Cytometric System. Exp Cell Res (in press, June, 1978).

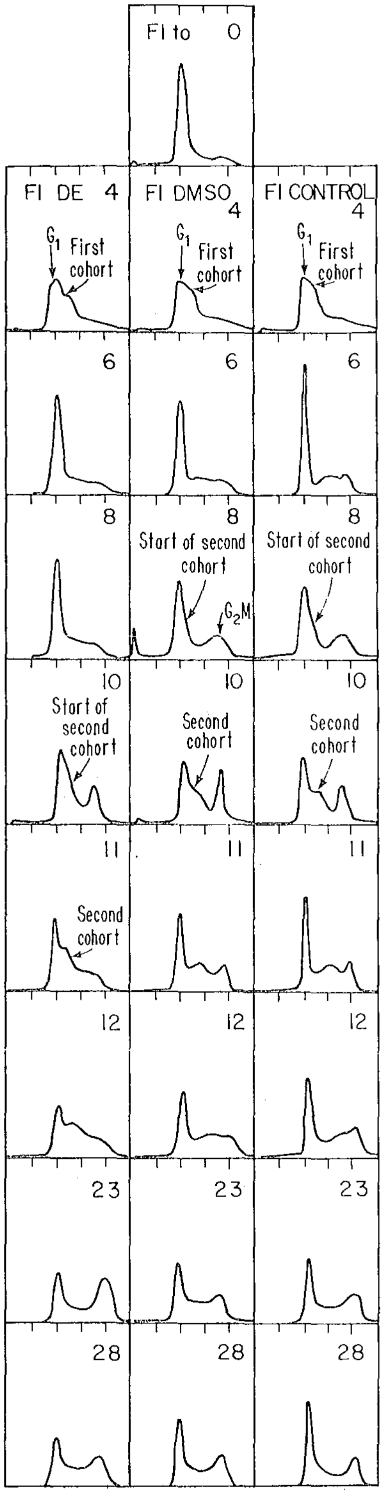
17. Hsu, W.-T., Lin, E.J.S., Harvey, R.G., and Weiss, S.B. Mechanism of Phage phi-X174 DNA Inactivation by Benzo[a]pyrene-7,8- dihydrodiol-9,10-epoxide. Proc Natl Acad Sci USA 74:3335-3339 (1977)

18. Wade, C.G., Baker, D.E., and Bartholomew, J.C. Selective Fluorescence Quenching of Benzo[a]pyrene and a Mutagenic Diol Epoxide Derivative in Mouse Cells. Biochemistry. In Press.

Chart 1



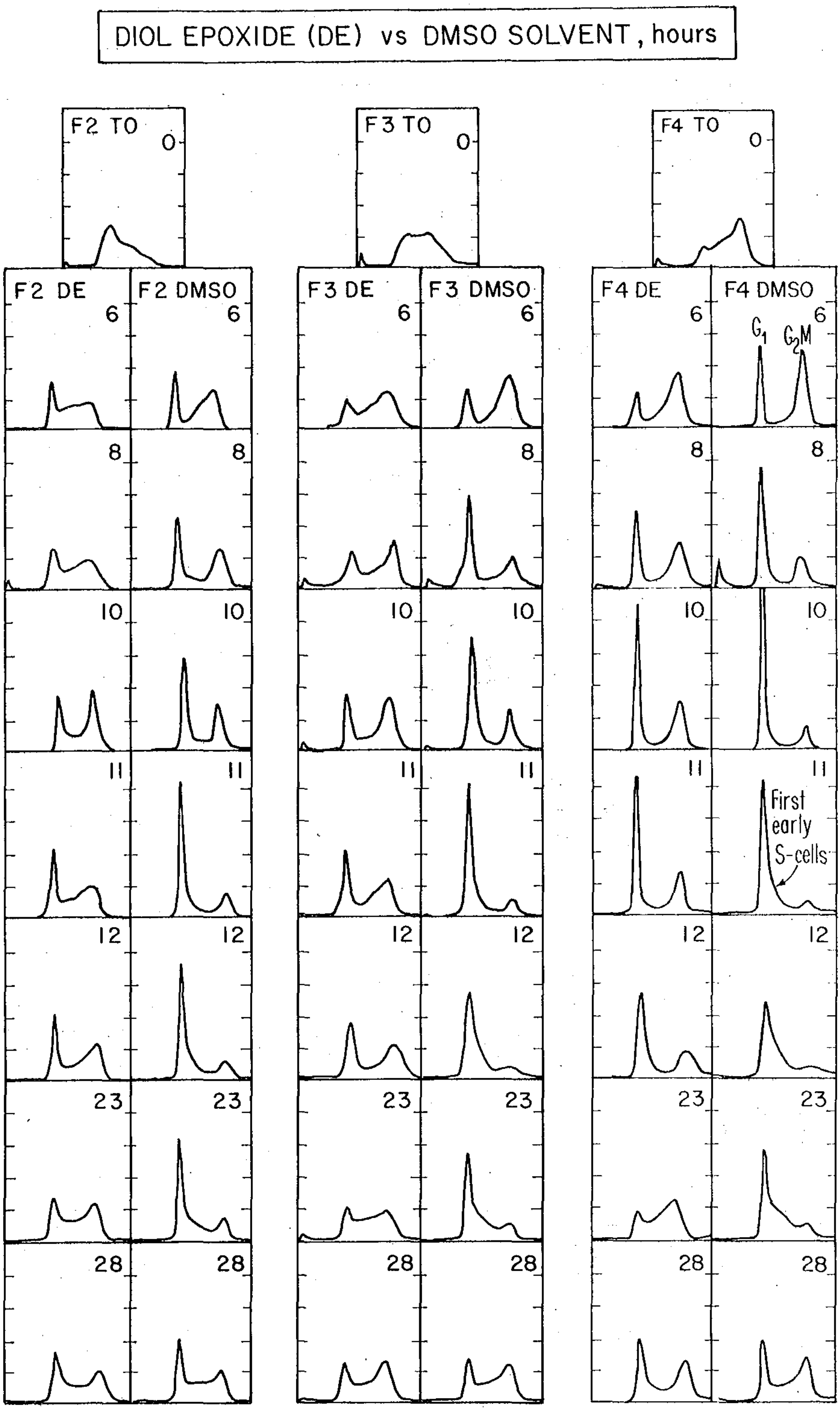
XBL 788-4174



FI to 0
FI DE 4
FI DMSO 4
FI CONTROL 4
G1
First cohort
6
8
Start of second cohort
G2M
10
Start of second cohort
Second cohort
11
12
23
28

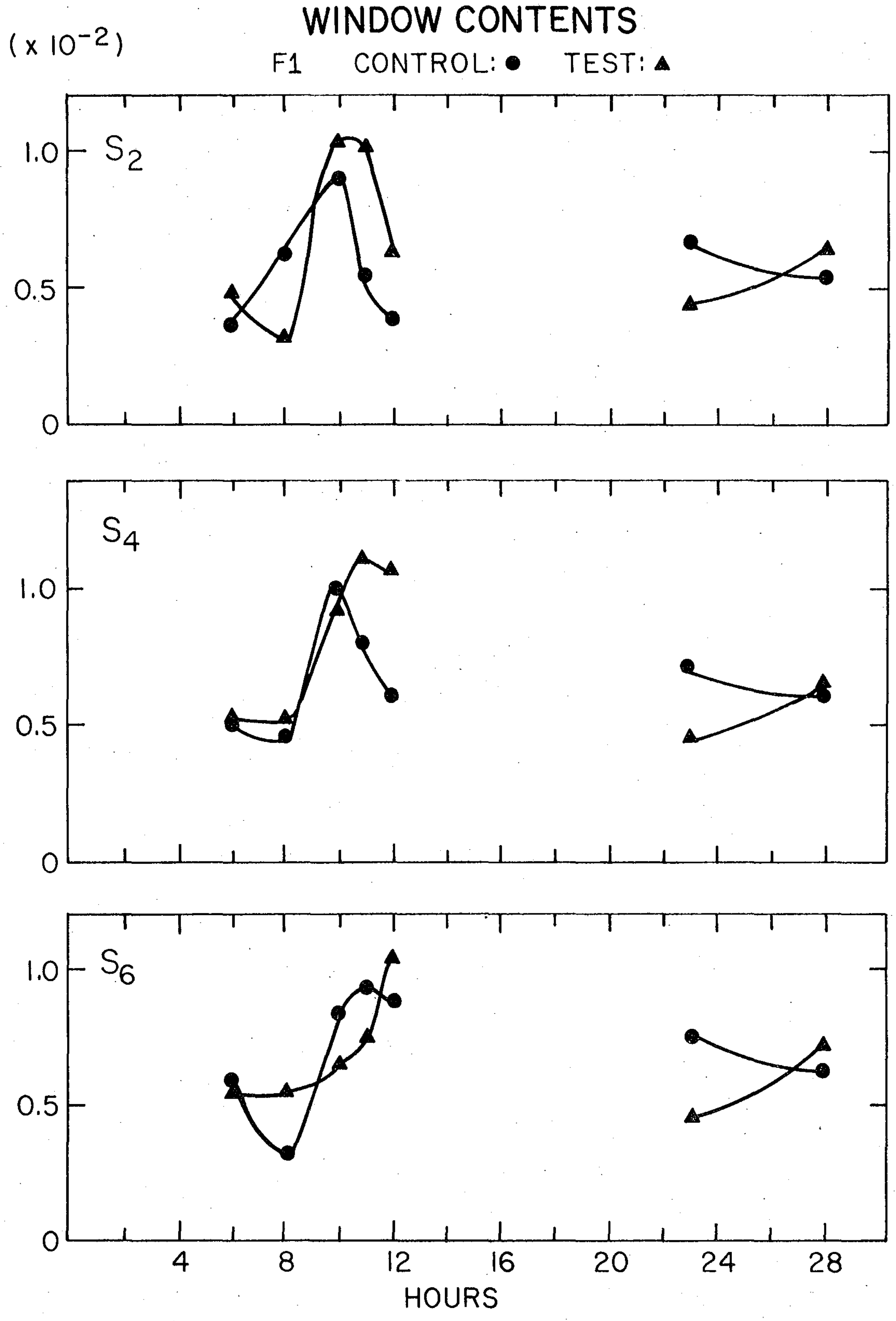
XBL 788-4169

Chart 3



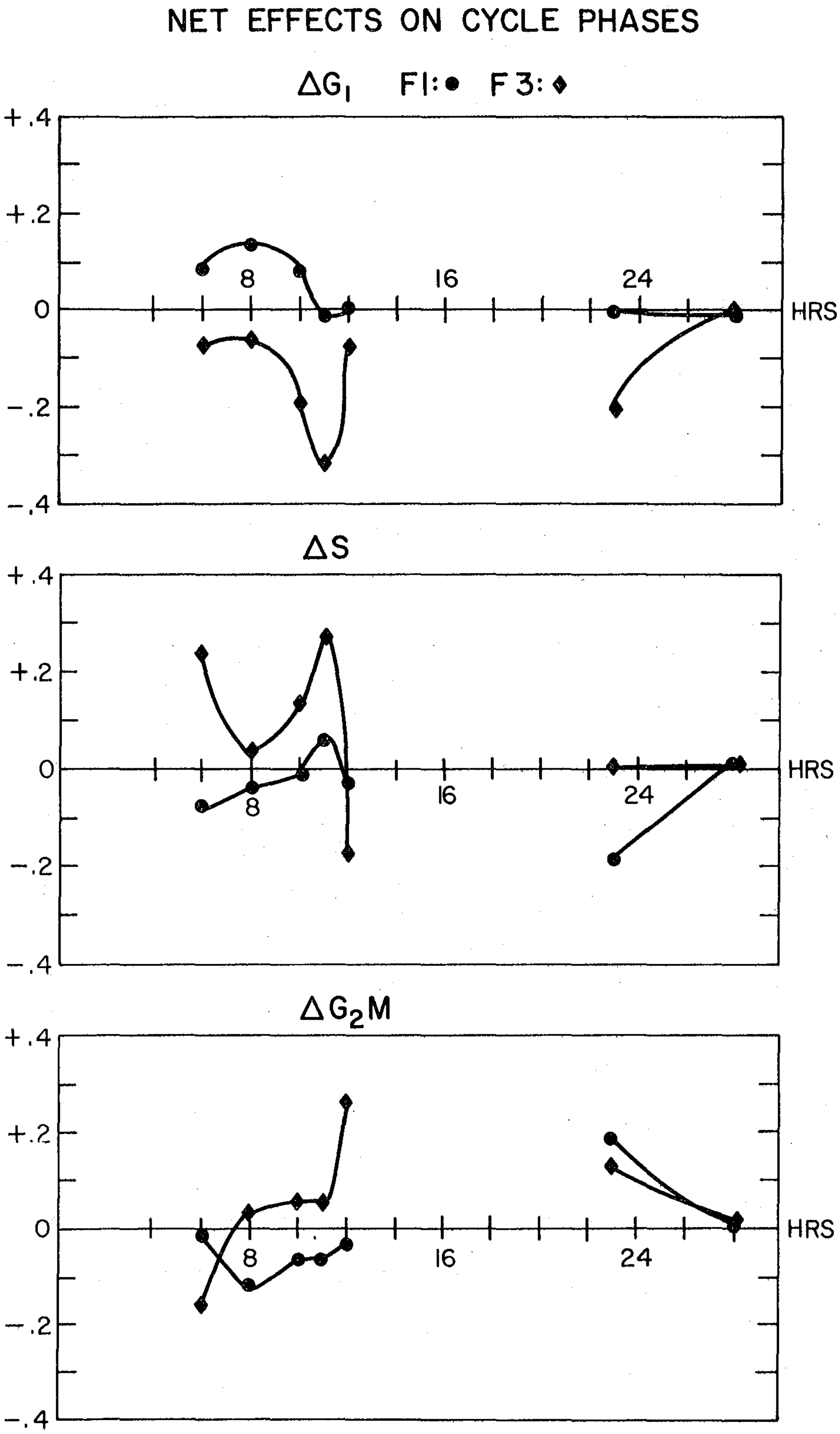
XBL 788-4170

Chart 4



XBL 784-3935

Chart 5



XBL 786-4042

APPENDIX A:

ANALYSIS OF THE EXCITATION OPTICS OF THE FLOW CYTOMETER AT THE LABORATORY OF CHEMICAL BIODYNAMICS

INTRODUCTION

As discussed in the body of the thesis, apart from variations in the binding of fluorescent dye label, the principal source of error in the FCM analysis procedure variation in velocity and position of cells as they traverse the region of excitation. These variations are due to imperfections in the flow device, and contribute to spreading of the resulting histograms. The purpose of this work is to characterize the FCM excitation optics in detail, arriving at a mathematical description of the irradiance distribution in the region of excitation. This is then to be used in estimating the nature and extent of the contribution of illumination variations to the overall spread of the FCM process.

DETAILED CHARACTERIZATION OF THE EXCITATION OPTICS

A Spectra Physics Model 170 Argon-ion laser is operated at the 488nm line for typical FCM usage, and in TEM_{00} mode operation produces a 3 watt continuous beam for excitation illumination. Referring to Fig. 1, we presume the initial beam to possess a Gaussian cross-sectional intensity profile in both x- and y- directions, with cylindrical symmetry about the propagation axis (z-axis). At point z_1, the beam enters the first cylindrical lens, whose focal length in the x-z plane is f_x, and the beam proceeds to the second cylindrical lens at $z = z_2$, whose y-z focal length is f_y. In the FCM of our laboratory, f_x=23cm, f_y=3cm. The refracted rays next encounter the flow cell, whose entrance windows are flat quartz plate of about 3mm thickness, and then traverse the ~7mm of water surrounding the cell stream, whose diameter is 8-30 μm.

The effects of the quartz window and water surround shall be ignored, on the assumption that their main result is to slow down the propagation of the exciting beam. Once their presence has been compensated for by adjustment of the $z_1$and z_2 lens positions, they are not expected to affect the final irradiance distribution in the region of excitation.

A good formulation of the problem at hand is presented in Gerrard and Burch (1), pp. 116-131 and 146-147. Following their analysis, I describe the complex amplitude near the optic axis by

$$A(r,z) = A_0\left(\frac{\omega_0}{\omega}\right)\exp\left[i\left(\frac{2\pi z}{\lambda} + \phi\right) - r^2\left(\frac{1}{\omega^2} - \frac{i\pi}{\lambda R}\right)\right] \qquad (1)$$

where $w(z)$ = "spot radius" at which amplitude A drops by 1/e, and intensity I drops by $1/e^2$, at point z along the optic axis

w_0 = minimum "spot radius", occuring at the "beam neck" (z=0)

$R(z)$ = radius of curvature of equiphase surfaces

$\phi = \tan^{-1}\left(\frac{\lambda z}{\pi \omega_0^2}\right)$

and $z_0 = \frac{\pi w_0^2}{\lambda}$ = which is the "confocal beam parameter", the radius in the z-direction about the beam neck over which $w(z)$ is essentially constant

Gerrard and Burch show, from the wave equation, that for a Gaussian beam such as this, the spot radius and curvature dependence are:

$$\omega^2(z) = \omega_0^2\left[1 + \left(\frac{\lambda z}{\pi \omega_0^2}\right)^2\right] \qquad (2)$$

$$R(z) = z\left[1 + \left(\frac{\pi \omega_0^2}{\lambda z}\right)^2\right] \qquad (3)$$

Now, since intensity is inversely proportional to the square of spot radius, the expression for I(z) is:

$$I(z) = \frac{\beta}{\omega_0^2\left[1 + \left(\frac{\lambda z}{\pi \omega_0^2}\right)^2\right]} \qquad (4)$$

where β is the constant of proportionality. Hence, the intensity dependence in the z-direction is Lorentzian, with parameter w_0 to be determined.

In the above formulation, the two principal parameters are w(z) and R(z), and Gerard and Burch combine them into one "complex curvature parameter", q(z), as follows:

$$\frac{1}{q} = \frac{1}{R} + \frac{i\lambda}{\pi\omega^2} \tag{5}$$

from which the r^2 coefficient in equation (1) is identified as $\left[\frac{-i\pi}{\lambda q}\right]$

so,

$$A(r,z) = A_0\left(\frac{\omega_0}{\omega}\right)\exp\left[i\left(\frac{2\pi z}{\lambda} + \phi\right) + \frac{i\pi r^2}{\lambda q}\right]$$

The real part of 1/q represents the divergence of the constant phase surfaces, while the imaginary part corresponds to the degree of irradiance concentration in the axial region of the beam. $q(0) \equiv q_0$ is the value at the beam neck, where

$$q_0 = \left(\frac{i\lambda}{\pi\omega_0{}^2}\right)^{-1}$$

In order to proceed, the following important preliminaries are needed.

Let RP1 and RP2 represent reference planes perpendicular to the optic axis and situated on the entrance (RP1) and exit (RP2) sides of a given optical system, as depicted in Fig. 2. Let y_1 and v_1 be the vertical distance above, and angle to the optic axis, respectively, of an incoming paraxial (non-skew) ray at RP1. Let y_2 and v_2 be the corresponding quantities at RP2. The optical system may be characterized by the matrix that relates the vector (y_1, V_1) to (y_2, V_2), where $V_1 = n_1 v_1$, and $V_2 = n_2 v_2$ (n_1 and n_2 are the refractive indices at RP1 and RP2, respectively). Thus,

$$\begin{bmatrix} y_2 \\ V_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_1 \\ V_1 \end{bmatrix} \qquad (6)$$

Further, the radius of curvature of the wavefront at RP1 is $R_1 = \frac{y_1}{V_1}$, and similarly, $R_2 = \frac{y_2}{V_2}$, and from equation 6, we have the relation

$$\frac{y_2}{V_2} = \frac{Ay_1 + BV_1}{Cy_1 + DV_1}$$

$$\text{or} \quad R_2 = \frac{AR_1 + B}{CR_1 + D} \qquad (7)$$

By substitution of Eqn. 5 into Eqn. 7, one can show that

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D} \qquad (8)$$

which is the central relation in the present study.

To apply this to the FCM excitation optical arrangement, I regard the 2 crossed cylindrical lenses as being mutually independent, and treat each as a one-dimensional lens whose equivalent matrix for rays in the appropriate directions is:

$$M_1 = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \quad \text{for } L_1 \text{ (x-z plane)},$$

$$M_2 = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \quad \text{for } L_2 \text{ (y-z plane)}.$$

Following the treatment of Gerrard and Burch (Pp. 146-147) we presume the input to the FCM to be a Gaussian laser beam of neck radius w_{01}, and corresponding confocal beam parameter

$$z_{01} = \left| \frac{\pi w_{01}^2}{\lambda} \right|$$

and let z_1 and z_2 be the respective distances from beam neck to lens along the optic axis, as illustrated in Fig. 3. Thus, for lens L_1, we have:

$$q_2 = \frac{A_1 q_1 + B_1}{C_1 q_1 + D_1} \quad \text{where } q_1 = q_{01} + z_1,$$

with $$q_{01} = \left(\frac{i\lambda}{\pi \omega_{01}^2} \right)^{-1} = -iz_{01}$$

thus, $q_1 = z_1 - iz_{01}$. Further, let $Z_{02} = \left(\frac{\pi\omega_{02}^2}{\lambda}\right)$

so that $q_2 = \frac{1}{R_2} + \frac{i\lambda}{\pi\omega_{02}^2} = \frac{1}{R_2} - iZ_{02}$.

Since we are interested in the spot size parameter, we should concentrate on the imaginary part of q_2, so after algebra and substituting for q_1 and q_2 we have:

$$\mathrm{Im}(q_2) = -iz_{02} = \frac{-iz_{01}}{\left(cz_1 + D\right)^2 + c^2 z_{01}^2}$$

or

$$\omega_{02}^2 = \frac{\omega_{01}^2}{\left(cz_1 + D\right)^2 + \left[\frac{c^2\pi^2\omega_{01}^4}{\lambda^2}\right]} \tag{10}$$

Referring to Fig. 3, we place RP1 at the first surface of L_1, and RP2 in the second focal plane of L_1. Since all rays entering RP1 are parallel to the optic axis in the x-z plane, they are representable as ray vectors

$$\begin{vmatrix} x_1 \\ 0 \end{vmatrix}$$

At RP2, all rays have x = 0, regardless of angle V_2, so rays leaving RP2 are represented by

$$\begin{vmatrix} 0 \\ V_2 \end{vmatrix}$$

where

$V_2 = -\frac{x_1}{f_x}$. Thus, we seek the elements of matrix M_1 such that

$$\begin{bmatrix} 0 \\ -\left(\frac{x_1}{f_x}\right) \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

and this, along with other considerations, requires that

$$M_1 = \begin{bmatrix} 0 & 0 \\ \frac{-1}{f_x} & 0 \end{bmatrix}$$

implying that A1=B1=D1 = 0, and C1 = $\frac{-1}{f_x}$. Similarly,

$$M_2 = \begin{bmatrix} 0 & 0 \\ \frac{-1}{f_y} & 0 \end{bmatrix}$$

Thus, upon substitution into Eqn. 10, we have:

$$\omega_{02(x,y)}{}^2 = \frac{\omega_{01(x,y)}{}^2 f_{x,y}{}^2}{z_{(1,2)}{}^2 + \left[\frac{\pi \omega_{01(x,y)}{}^4}{\lambda^2}\right]} \tag{11}$$

where subscripts (1, or (x, refer to lens L_1, and ,2) or ,y) refer to lens L_2. For the LCB device, $w_{01(x,y)} \simeq 2$mm, $f_x = 23$cm, $f_y = 3$cm, $z_1 = 20$cm, $z_2 \simeq 40$cm, and $\lambda = 488$nm.

Thus, the minimum diffraction limited semi-axis of the beam in the x-z plane is 31.66 µm, and in the y-z plane is 4.13 µm. The confocal beam parameters for these two cases, $z_{02(x,y)}$ are thus:

$$z_{02x} = \frac{\pi \omega_{02x}{}^2}{\lambda} = \frac{\pi \left(31.66 \times 10^{-4}\right)^2}{4.88 \times 10^{-5}} = 0.645 \text{ cm}$$

and

$$z_{02y} = \frac{\pi \omega_{02y}{}^2}{\lambda} = \frac{\pi \left(4.13 \times 10^{-4}\right)^2}{4.88 \times 10^{-5}} = 110 \mu\text{m}$$

Thus, when the FCM excitation optics are alligned, the equi-radiant contours at points z near z_f are ellipses, and at z_f the major axis of the elliptical contour for $I = I_{max}e^{-2}$ is 31.66 µm (x-direction), while the minor axis is 4.13 µm (y-direction).

Furthermore, applying equation 2 to the z-dependence of the focal region of L_2, we find

where z is in µm.

Thus, while deviations in the z-component of cell position on the order of 5 to 10 µm are truly negligible from the standpoint of the irradiance variation in the focal region of L_1, such deviations can yield up to 1% variation due to lens L_2.

Positional deviations in the x-direction shall now be considered. The intensity dependence in the x-direction, for fixed z near z_f, is Gaussian, with $1/e^2$ parameter w_{02_x}, i.e.

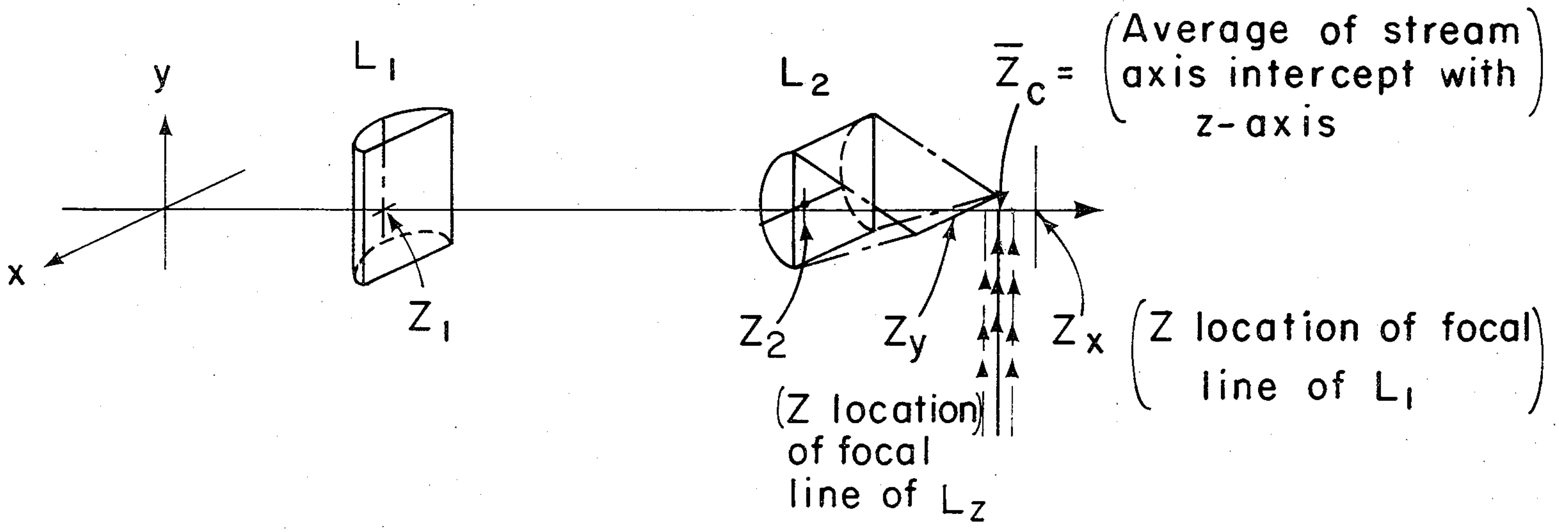
$$I_{\left(\substack{z\\y=0}\right)}(x) = \frac{I_{\max}}{\sqrt{2\pi}\left(\frac{\omega_{02x}}{2}\right)} \exp-\left[\frac{2x^2}{\omega_{02x}{}^2}\right] \qquad (12)$$

Figure 4 illustrates I(x) for w_{02_x} as determined above, with the dimensions of the flow stream superimposed. From Fig. 4, it is apparent that variations much larger than those of the z-direction are induced by the flow positional deviations in x. For a flow stream radius of 5 µm (fairly common), the illumination variation, as calculated here, could account for about 1% error. Since resolution is rarely better than 2%, this would imply that most of the observed coefficient of variation in histograms is due to staining variations, rather than flow cell imperfections.

REFERENCES

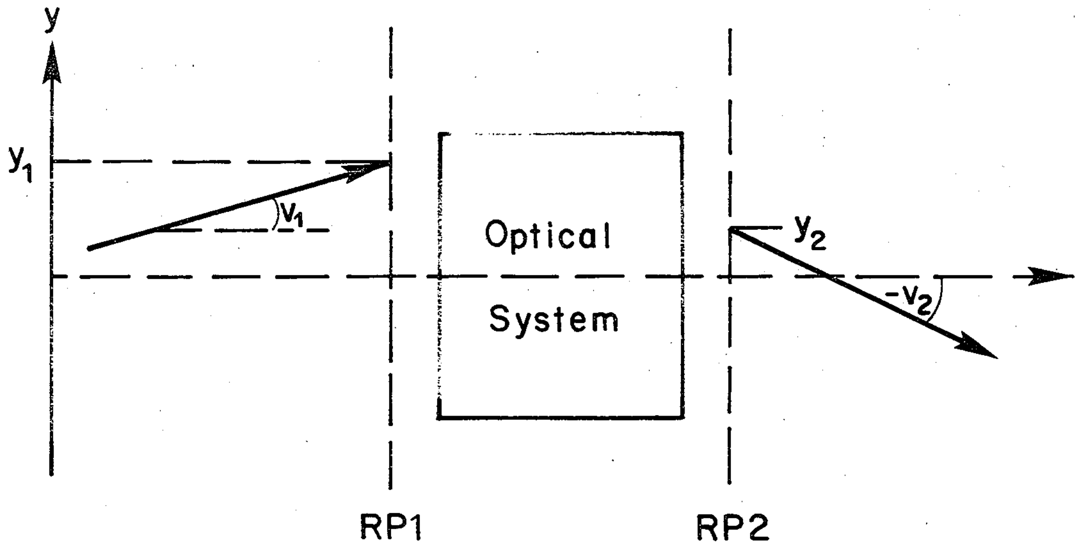
1. Gerrard, A., and Burch, J.M. Introduction to Matrix Methods in Optics. John Wiley, 1975.

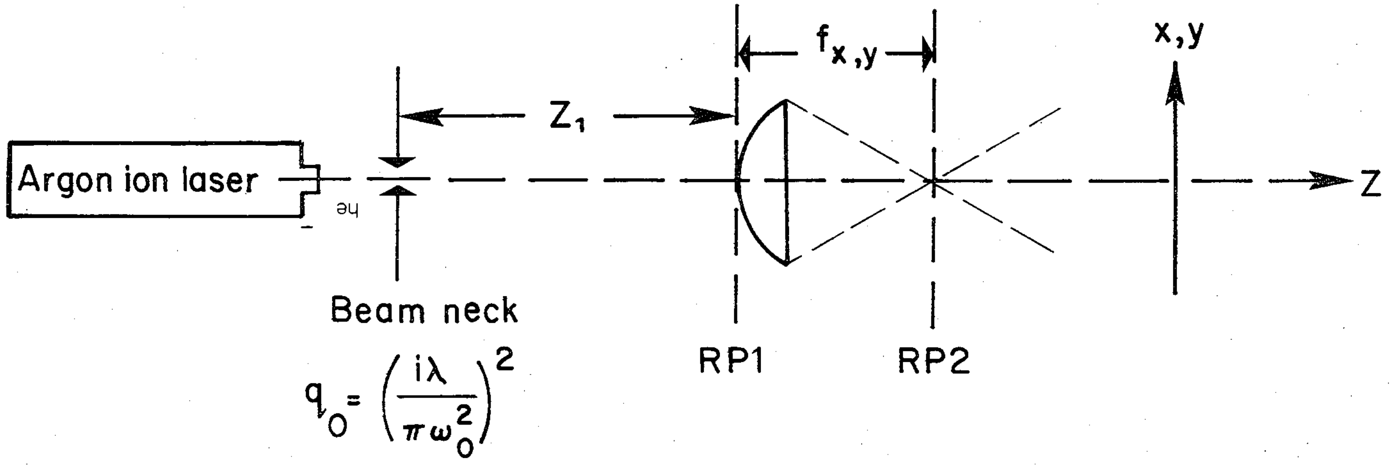
FCM Excitation optical arrangement



XBL778-1664

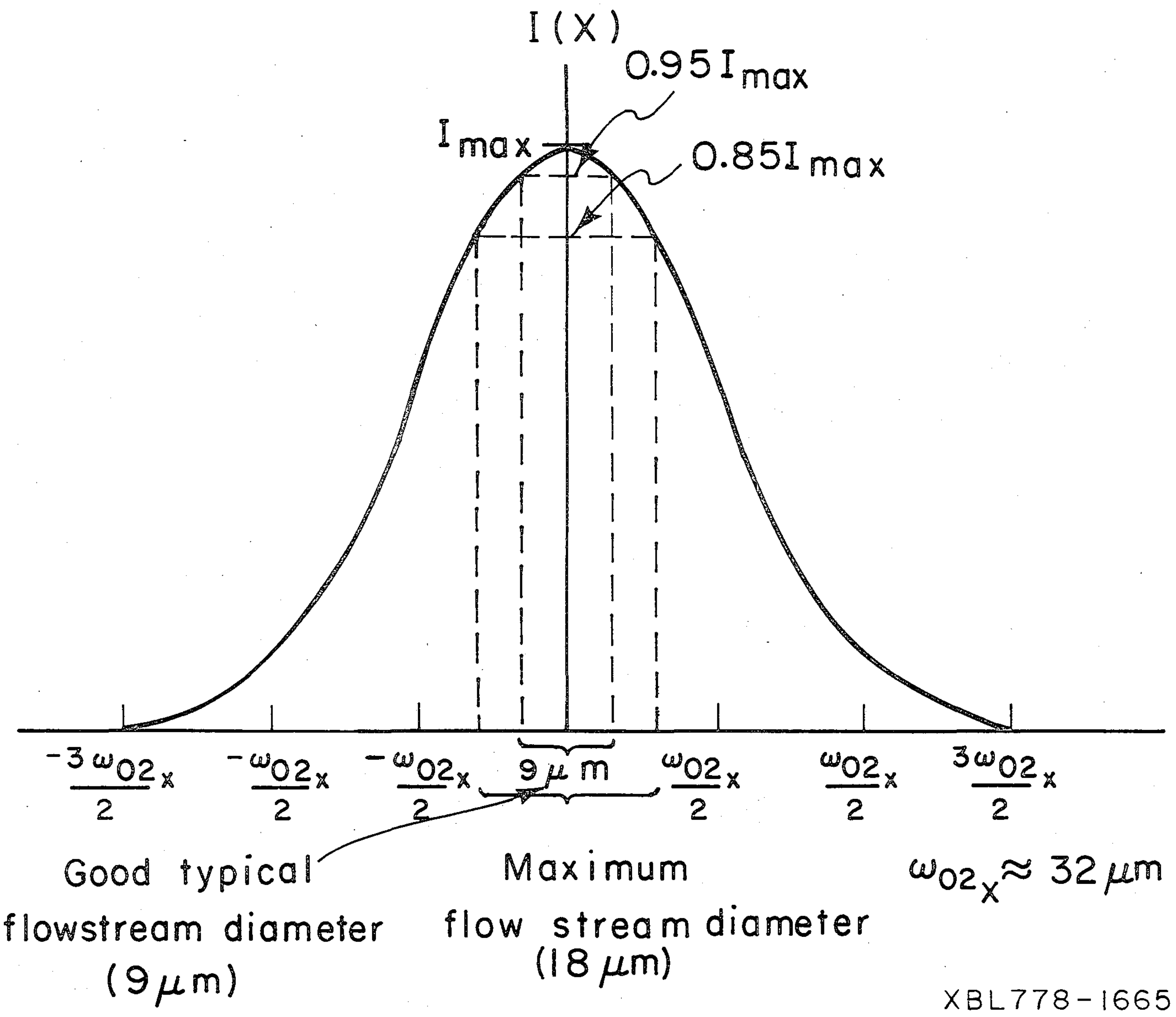
Figure 1





XBL 788-4172

Figures 2 & 3



I(X)
0.95 I_{max}
I_{max}
0.85 I_{max}
$-\frac{3\omega_{02x}}{2}$
$-\frac{\omega_{02x}}{2}$
$-\frac{\omega_{02x}}{2}$
9 μm
$\frac{\omega_{02x}}{2}$
$\frac{\omega_{02x}}{2}$
$\frac{3\omega_{02x}}{2}$
Good typical flowstream diameter (9 μm)
Maximum flow stream diameter (18 μm)
$\omega_{02x} \approx 32\,\mu m$
XBL778-1665

Figure 4

APPENDIX B:

KINETIC MODELING OF REDUCED DATA FROM DNA HISTOGRAM TIME SERIES

INTRODUCTION

The attempt to analyse simultaneously the underlying DNA distributions and the cell cycle kinetics represented in a time series of DNA histograms has proven somewhat successful (1), but is rather expensive in terms of computer time and space. In an effort to reduce the cost, while still obtaining the desired cell cycle kinetic information, we have devised an approach for the analysis of a reduced data set from fitted histograms. The hope is, that given a small set of cycle compartments in the model, and a time series of observed values in those compartments, coupled with some simplifying assumptions, one could obtain estimates of the average local kinetic rate at several points in the cell cycle. Furthermore, by doing the analysis for two parallel time series, one for cells treated with an agent, the other for control cells, the method could lead to direct comparison of the local growth rates between the two series, and hence to determination (at least in general terms) of the kinetic effect of the agent, relative to controls.

As described in reference (2), our fitting program reduces a DNA histogram containing 256 points to 11 values: population fractions in G_1, G_2M, and 9 S-phase subcompartments. The cell cycle model corresponding to this parametrization is illustrated in Fig. 1, and consists of two large compartments (G_1 and G_2M) and nine small compartments through S-phase. The unequal size of the compartments must have some implications for the analysis of the kinetics from the observed data and the given model, but aside from the discussion below, I shall only remark that the alternative for DNA histograms is to subdivide G_1 and G_2M into subcompartments whose contents cannot be measured directly, but only inferred. While allowing for more realistic representation of the cell cycle, this alternative approach contains additional uncertainties, and shall not be addressed further in this thesis.

To generate the kinetic equations governing the model, we invoke conservation of cells at each compartment, i.e., the change in the number of cells in compartment (i), over time interval dt, is just the number entering (i) from (i-1), less the number leaving (i). Included is the provision for mitosis, i.e., two cells enter G_1 for each cell that successfully completes G_2M. Implicit in this scheme is the assumption that cell death is either negligible or is not distinguishable from long term cycle blocking.

To model the cycle kinetics of the system, I use the notion of transition probability suggested by Gray (1), in which cells proceed from one compartment to the next in a random, stochastic manner, under the assumption that all cells in a given compartment have the same probability of transition, and are independent of one another. The number of cells expected to transit from (i) over the next time interval is simply $k_i^t f_i^t$, where f_i^t is the number of cells in (i) at time t, and k_i^t is the probability that a cell in (i) will transfer to (i+1) during the next time interval. Decycling or long-lasting cycle blocking is represented by small relative values for k_i^t, while recruitment of cells from quiescent to active proliferative state is reflected in an increased value over time for the k_i^t corresponding to the site of recruitment. The probabilities k_i^t are assumed to be constant over the time interval between one time point and the next, effectively separating the time series into several independent propagation steps. The shortcomings of this model are discussed below, but it should be emphasized that it was pursued only to obtain order-of-magnitude estimates for the kinetic effects of the agents studied, rather than highly precise kinetic characterization of such effects.

With all the above assumpions, the propagation of the reduced compartment set $\{f_i^t\}$ obeys the set of ordinary differential equations (ODE's):

$$\dot{f}_1{}^j = 2k_{11}{}^j f_{11}{}^j - k_1{}^j f_1{}^j$$

and for the i^{th} compartment, $i = 2$ to 11,

$$\dot{f}_i^j = k_{i-1}^j f_{i-1}^j - k_i^j f_i^j$$

The first equation has a coefficient of 2 for the contribution from compartment 11 because each cell that completes mitosis, i.e. passes from compartment 11, becomes two cells in compartment 1, by virtue of the assumption that there is no cell death.

The probabilities k_i^t are allowed to vary from one timepoint to the next, due to whatever conditions may prevail in the environment of the cells, including the addition of various agents to the growth medium, and the interactions of cells with one another. When the optimal values for the constants k_i^t are obtained, the ratios of the respective rates for treated versus control time series are to be calculated, and compartments where treated differs from control by more than a order of magnitude are to be identified.

When cast in matrix form, these equations can be represented in two ways, either

$$d\frac{\overline{f}}{dt} = \underline{H}\overline{f} \tag{2}$$

or, alternatively, as

$$d\frac{\overline{f}}{dt} = \underline{B}\overline{k} \qquad (3)$$

The latter form is addressed below. The former representation has the familiar form of a set of first order, linear differential equations, in which the constants {k} are incorporated into the elements of matrix $\underline{B}$, and are allowed to vary with time. The set of constants {k} at a given time completely determines the propagation of the population distribution forward to the next timepoint. Hence, heuristically speaking, the quality of the estimated values for the kinetic constants can be judged by comparing the propagated distribution with the observed data. An optimal set of constants can be obtained by iteratively correcting their estimated values until a good match is achieved between model and data.

The task of finding the best set of kinetic constants to describe the system is inverse in nature, as is seen from the integrated form of equation 1, for the i^{th} compartment:

$$f_i^{t+1} = f_i^{t} + \int_t^{t+1} \left| k_{i-1}^{t} f_{i-1}^{t} - k_i^{t} f_i^{t} \right| dt \qquad (4)$$

The desired quantities, the unknown constants {k}, are part of the integrand, and thus to obtain them one must "invert the integral", as is the case for analysis of single DNA histograms (Chapter). The drawback of this situation is that it forces a tradeoff between the number of free parameters used in a given model of the data (in this case, 11 parameters) and the statistical accuracy with which they can be determined. For the present approach, eleven parameters would seem a small enough set to be determinable to within 20 to 30% relative error (my present goal), though this is difficult to ascertain.

The principal difficulty in attempting to implement the approach suggested by equation 3 is that of propagating the data from one time to another, once given the latest set of constants. If the constants could be assumed to be of the same order of magnitude, this calculation could be accomplished by a simple one-step integration routine covering the entire time interval between timepoints. However, since cycle blocks, decycling, and other phenomena are included in the set of possible kinetic effects along with normal progression through the cycle, such an assumption appears totally unjustified, and hence far more expensive integration methods (i.e. multistep methods) must be employed. One of the most powerful and efficient of these methods was developed by Gear (3), and implemented on the CDC 7600 system at our laboratory by Risk,et al (4). I have attempted to use this package to solve the above system of 11 ODE's, using values ranging over 2 orders of magnitude among the equations, and found the average cost to be between 10 and 20 CPU's on a CDC 7600 computer. This integration, constituting one iteration of the above scheme, would have to be performed after each step in the 11-dimensional parameter space, and would likely be repeated on the order of 100 times before optimum is reached, unless large residual error tolerances were employed.

As an alternative approach, the matrix formulation of equation 4 seemed to present some hope of obtaining estimates at an affordable price. In this scheme, one ignores the fact that the equations are meant to describe a dynamical system involving time derivatives, and regards the problem as that of obtaining the unknown elements of vector $\overline{k}$ from approximate values for df_i^t/dt. The latter are estimated from the reduced data, $f_i^t(t)$ by one of several numerical methods, and the observed values f_i^t are used in matrix $\underline{B}$. Since the elements of $\underline{B}$ and the values for df_i^t/dt contain significant error, there is no reason to expect that direct inversion of $\underline{B}$ would yield reliable results. Hence, once all the elements have been provided, the solution is obtained by pseudo-inversion of matrix $\underline{B}$, using a linear least squares routine with non-negativity constraints on the entries of solution vector $\overline{k}$.

To obtain the approximate values for df_i^t/dt, I first attempted to use a simple linear interpolation method, where the time derivative at compartment (i) and time (j), i.e., f_i^j, was set to:

$$d\frac{f_i^j}{dt} = \frac{f_{i-1}^j - f_i^j}{t^{j+1} - t^j} \qquad (5)$$

With this scheme, the derivative at the last time point cannot be determined, since no such difference can be calculated for it. Nonetheless, the values for all but the last time point were calculated, substituted into the above equations, and solutions found as described above. The only criterion for judging the acceptibility of the solutions thus determined arose from the assumptions regarding the growth properties of the control samples. Since they were presumed to be growing asynchronously and exponentially, and underwent no further perturbation following replating, the controls were not expected to demonstrate significant time variation in the rate constant for any given compartment. Furthermore, the constants for all compartments should be comparable, for any given timepoint. By this criterion, this method fared poorly, as variations of up to 4 orders of magnitude were seen in the control samples thus analysed, and it appeared that the source of the error was the crude estimate of the derivatives given by the simple linear interpolatory approach.

As a result, alternative methods for estimating the derivative of a sampled function, at the sampled points, were investigated. It was decided to use a least squares spline-fitting routine developed at the Los Alamos Scientific Laboratory, and published under the name of "Smooth," which fits a piecewise, smooth polynomial to the data points provided by the user. The routine takes into account the estimated error in the sample points (provided by the user) in determining how tightly the spline must fit the data, and the resulting fitted smooth curve can be used to determine the interpolated functional value, first and second derivatives at any point within the observed domain of the abscissa (time, in this case). I estimated the relative error of the values in the reduced data set to be 10%, primarily due to counting statistics and to variations in the parameter values used in fitting the separate histograms from which the reduced data sets were obtained.

The relative error in the derivative estimates provided by "smooth" is about twice that of the data supplied by the user, or in this case, about 20%. Using these estimated values for the derivatives, the problem defined by equation 4 was solved for the same data set as in the linear interpolation case, with the result that kinetic rates for the control series varied by only 2 orders of magnitude over time; a hundred-fold improvement with respect to the above criterion.

To test for consistency in the estimated kinetic parameters, the reduced data set at a given time point was propagated by a Gear ODE solving routine to yield an estimate of the data set for the next time point. Upon comparison with the actual data, the propagated solution was found to be unacceptible, deviating by a factor of 2 or 3 from the data in half the compartments, though fitting the others to within 10-20% of their values. The conclusion to be reached from this effort is that there is little validity to the solutions obtained by the approach of equation 4, at least as far as I have been able to implement it.

DISCUSSION

The next step would appear to be the development of less expensive ways of implementing the approach of equation 3. The most important ingredient in solving this problem would be a fast algorithm for integrating the set of 11 differential equations. Such an algorithm might be gotten through modifying the existing Gear- Hindmarsh package now available (at LBL), making use of all information known about these particular ODE's and their solutions. One might also use fast optimization schemes, involving accelerated approach to the minimum, and relaxed residual tolerance to avoid extra iterations once an acceptible solution has been found. Such methods as the random search routine of Bremermann, et al (5) or a less elaborate version of the BCQNDF routine from the National Physical Laboratory of Great Britain (6) might work. It is necessary to use optimization routines that do not require the user to supply the partial derivatives with respect to the free parameters, as their dependence is neither analytical nor known.

The approach that appears most likely to bear fruit is a combination of limited inversion with user interaction. In such a scheme, the user would provide a first guess as to the set of kinetic parameters, which would then be optimized within a limited neighborhood of the initial values by a minimization program. The interaction would continue until the data is fit to within its statistical accuracy, and could be done by varying one parameter at a time, if so desired.

The most important question to be addressed is whether the above model of the cell cycle, being as restricted and simplified as it is, can justify such elaborate and potentially expensive numerical methods necessary to solve for the model parameters. The assumptions regarding cell death, decycling and recruitment would seem safe for most cycle perturbation studies with non-toxic doses of agent added. However, the assumption of equal probability for all cells within any given compartment is clearly unjustified for at least the G_1 and G_2M compartments, since cells that have just entered either one are far less likely to exit from it during the next time interval than those that have already spent a few hours in the compartment. Hence, G_1 and G_2M could clearly be further subdivided. From the estimates of cycle phase durations available through analysis of exponentially growing cell populations, one could partition the entire cell cycle in analgous manner to that which I use in S-phase (9 subcompartments cover the phase), along the lines proposed by Gray (1). In this fashion, the cycle would be divided into 20 to 30 stages, whose size would correspond to the time resolution of the DNA histogram time series, determined by the CV of the histograms and the time reproducibility of the growth behavior of the cell populations.

The cells in any one stage would then be presumed to have equal transition probability, with several stages required to complete G_1 and G_2M. This further assumption of equal transition probability can be justified if the stages are at the limit of the resolving power of the FCM system, meaning that no evidence is seen to suggest that they can be further subdivided into distinguishable subcompartments. When this holds, the transition into and out of a given compartment is seen as a binary phenomenon, and leads to the Poisson stochastic formulation of the time evolution of the population given in Takahashi's work (7).

REFERENCES

1. Gray, J. Cell-Cycle Analysis of Perturbed Cell Populations: Computer Simulation of Sequential DNA Distributions. Cell Tiss Kinet 9:499-516 (1976)

2. Pearlman, A.L., Navsky, B.N. Extracting Information from Single DNA Histograms Obtained by Flow Cytometry. (manuscript in prep.)

3. Gear (reference on multi-step integrating routine for stiff ODE's)

4. Risk, C., Concus, P. Routine Gear-Hindmarsh. (writeup for the computer center of Lawrence Berkeley Laboratory)

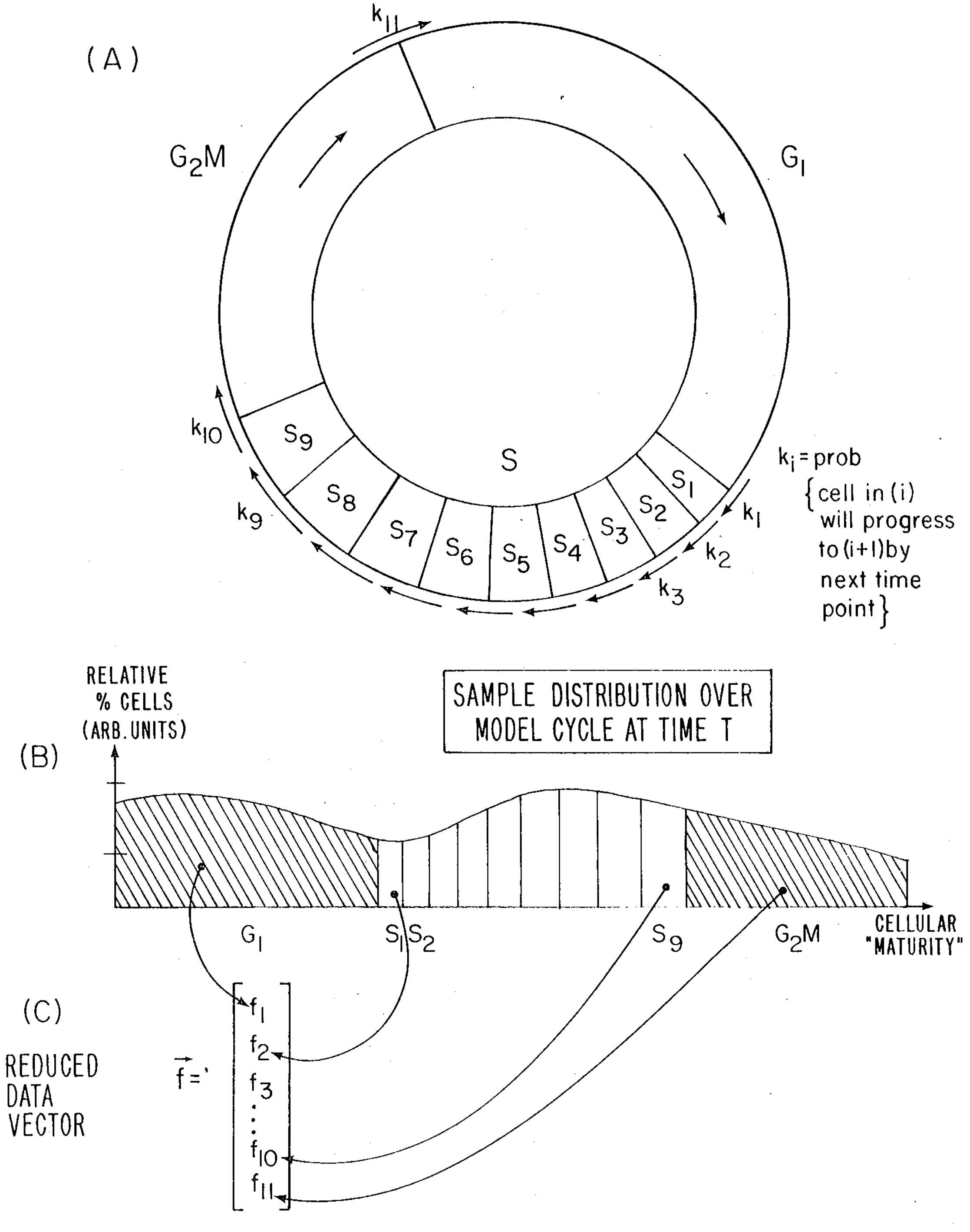
5. Bremermann, H., et al. Global Optimization Routine (private communication)

6. Gill, P.E., and Murray, W. Minimization of Functions Subject to Simple Bounds on the Variables. National Physical Laboratory Report No. NAC 74 (1976)

7. Takahashi, M. Theoretical basis for cell cycle analysis J. Theoret. Biol. 18: 195.(1968)

Figure 1

SIMPLIFIED CELL CYCLE MODEL SUITED TO REDUCED DNA HISTOGRAMS



XBL788-1682

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720