Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

Mathematical Biosciences and Engineering, 5(x): xxx–xxx
DOI:
Received: December 2018
Accepted: September 2019
Published: December 2019

Research article

Nonspecific probe binding and automatic gating of cell identification in flow cytometry

Bhaven A. Mistry[1]* **and Tom Chou**[1,2]*

[1] Department of Biomathematics, UCLA, Los Angeles, CA, 90095-1766, USA
[2] Department of Mathematics, UCLA, Los Angeles, CA, 90095-1555

* **Correspondence:** bamistry@ucla.edu; tomchou@ucla.edu; Tel: +1-310-206-2787; Fax: +1-310-825-8685.

Abstract: Flow cytometry is extensively used in cell biology to differentiate cells of interest (mutants) from control cells (wild-types). For mutant cells characterized by expression of a distinct membrane surface structure, fluorescent marker probes can be designed to bind specifically to these structures while the cells are in suspension, resulting in a sufficiently high fluorescence intensity measurement by the cytometer to identify a mutant cell. However, cell membranes may have relatively weak, nonspecific binding affinity to the probes, resulting in false positive results. Furthermore, the same effect would be present on mutant cells, allowing both specific and nonspecific binding to a single cell. We derive and analyze a kinetic model of fluorescent probe binding dynamics by tracking populations of mutant and wild-type cells with differing numbers of probes bound specifically and nonspecifically. By assuming the suspension is in chemical equilibrium prior to cytometry, we use a two-species Langmuir adsorption model to analyze the confounding effects of non-specific binding on the assay. Furthermore, we analytically derive an expectation maximization method to infer an appropriate estimate of the total number of mutant cells as an alternative to existing, heuristic methods. Lastly, using our model, we propose a new method to infer physical and experimental parameters from existing protocols. Our results provide improved ways to quantitatively analyze flow cytometry data.

Keywords: FACS, cytometry, gating, fluorescence, Langmuir, inference, serial dilution

1. Introduction

A common problem in cell biology research is the desire to differentiate cells into categorical populations based on some defining molecular characteristic. Some examples include the presence or absence of a particular gene transcript [1, 2], cells presenting viral epitopes to indicate infection [3, 4], or expression of particular membrane proteins [5, 6]. Flow cytometry is an effective tool to count the number of cells exhibiting the given characteristic. The process involves suspending cells of interest in a sheath fluid that is pressurized and extruded single file past a laser beam [7, 8, 9]. Each cell will scatter the laser's light towards optical sensors positioned around the stream. Sensors directly in front of the laser measure the forward scattering about a single cell, and is used to quantify the cell's surface area, volume, and shape. Alternatively, side scattering sensors measure photons emitted by fluorescent markers and dyes excited by the laser. The fluorescent probes are designed *a priori* to bind specifically to cell surface proteins and structures that characterize the cell species. Thus, a sufficiently high fluorescence intensity is an indication the cell is of the desired type.

However, details of the protocol arise that can confound the final count of cells. If we focus on the example of a population of "mutant" cells, characterized by the expression of a particular membrane surface receptor, mixed in with a population of "wild-type" cells, we can design a fluorescing probe that binds specifically to the receptor. If we suspend all cells in a solution containing an excess of probes, we expect all probes to bind to free receptors. However, each probe-receptor binding event is a reversible process, allowing some expected proportion of receptors to remain unbound at equilibrium. Furthermore, variation may exist in the number of receptors expressed, increasing the ways in which a mutant may escape binding to any probe [4, 5]. To combat this measurement of false negatives, one can increase the probe concentration in the suspension, increasing its excess and driving the equilibrium towards more bound receptors. However, although the probes are designed to bind specifically to the receptors, they will have a relatively small, but non-zero binding affinity to the rest of the cell membrane and its other embedded structures [10]. Increasing the probe concentration will result in a higher nonspecific binding to wild-type cell membranes, allowing false positive counts of mutants. The equilibrium configuration of probe bindings to all cells, whether specifically or nonspecifically, will produce a distribution of fluorescence data over a range of intensities. Cells that exhibit levels of fluorescence below a threshold intensity are ignored in a process known as "gating." Setting the gating threshold is typically a heuristic procedure, though some methods for automatic gating based on data clustering have been developed [9, 11]. However, these methods do not incorporate the underlying chemical kinetics of probe binding and largely ignore the effects of nonspecific binding.

We develop a kinetic model for both specific and nonspecific binding of probes to cells. We employ a variant on the Langmuir adsorption model [12] with two competing types of binding sites: receptors and a discretization of the cell membrane. Here, the concentration of initially added probes applies the "partial pressure" driving probe binding to the cell surfaces. We will show the isotherm of fractional binding site occupancy to exhibit two regimes in which the receptor and membrane binding sites become saturated at different rates. We discuss how the interface between the two regimes is the ideal concentration of probes to include in the assay and how the model can inform optimal experimental design. We then present a probabilistic model for the expected number density of cells over possible numbers of probe bindings. Employing this model, we develop a variant on the expectation maximization (EM) mixture model [13] to estimate the total number of mutant cells without heuristic gating. Futhemore, we propose a method for inferring the probe binding affinity and the receptor number distribution using a serial dilution protocol. Finally, we discuss potential applications and problems of using our method for the fluorescence activated cell sorting (FACS) assay. It should be noted that throughout this paper we continue using the example of "mutant cells" expressing surface receptors, but our models and analyses extend to all physiological applications of flow cytometry with fluorescing surface markers.

2. Materials and method

2.1. Kinetic model

Let C_T be the total number of cells in a suspension also containing N_T probe molecules. We assume C_T is precisely counted by the forward scattering measurement to differentiate cells from free probes or debris. Let M and $W = C_T - M$ be the number of mutant cells and wild-type cells, respectively, and N be the number of free fluorescing probes unbound to any cells. Mutants are the only cells that express the surface receptors the probes specifically target with association and dissociation rates p_+ and p_-. Alternatively, for both mutants and wild-types, probes can bind and dissociate non-specifically to the cell membrane itself with rates q_+ and q_-. Though we expect the on rates p_+ and q_+ to be comparable, probes bound nonspecifically to the membrane will dissociate significantly more rapidly so that typically $q_- \gg p_-$.

We expect the total number of receptors on a mutant cell I to vary across cells with distribution $f(I)$ and mean $\langle I \rangle$. The exact distribution f will depend on the details of the receptor and its transcription/translation pathways. Furthermore, we consider the the total surface area A of the cell membrane and partition the binding region around a single receptor as A_s. We define this region as that which a probe fated to adsorb to the cell surface is more likely to bind to the associated receptor than directly to the membrane. We can partition the remaining cell surface into $J = \frac{A}{A_s} - I$ discrete effective binding sites for the membrane. We expect the binding region of a receptor to be relatively small compared to the total surface area, making $J \gg I$. Finally, we denote the total number of mutant cells that carry I receptors as $M^I = f(I)M$.

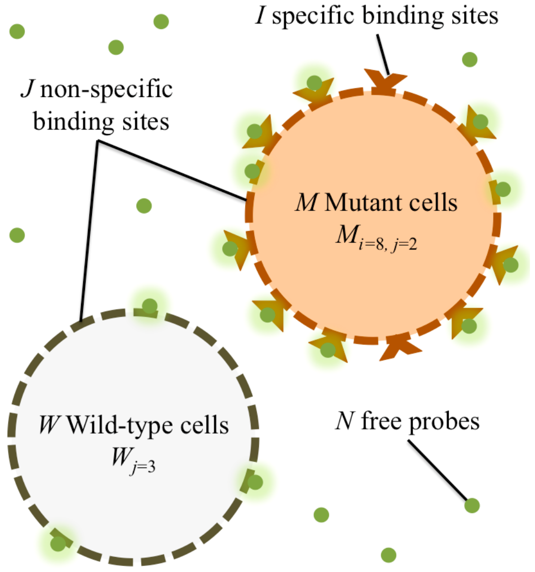

	Table of Variables
C_T	Total cells
M	Total mutants
W	Total wild-types
N	Total unbound probes
I	Specific binding sites per cell
J	Nonspecific binding sites per cell
$M_{i,j}$	Mutants with i specific and j nonspecific bindings
K_s	Specific dissociation constant
K_n	Nonspecific dissociation constant

Figure 1. Cartoon of probe molecules binding to wild-type W and mutant M cells used in a typical flow cytometry assay. Wild-type cells are assumed to bind probes only nonspecifically while each mutant cell expresses I receptors to which probes specifically bind. The variables defining all quantities in the kinetic mass-action model analyzed in this paper are given in the table.

To accurately model the kinetic flows from one bound state of a mutant to another, it becomes necessary to track populations of cells indexed by both the number of probes bound specifically and nonspecifically. Thus, we define $M^I_{i,j}$ as the number of mutant cells with exactly i probes bound to a maximum of I specific binding sites and exactly j probes attached nonspecifically. For wild-type cells, probes can only attach nonspecifically, so we define W_j as the number of wild-type cells with exactly j adsorbed probes. We thus have the chemical rate equations

$$M^I_{i,j} + N \underset{(I-i)p_+}{\overset{(i+1)p_-}{\rightleftharpoons}} M^I_{i+1,j}, \quad M^I_{i,j} + N \underset{(J-j)q_+}{\overset{(j+1)q_-}{\rightleftharpoons}} M^I_{i,j+1}, \quad W_j + N \underset{(J-j)q_+}{\overset{(j+1)q_-}{\rightleftharpoons}} W_{j+1}. \tag{2.1}$$

Let v be the volume of the cell suspension containing all cells and probes. Normalizing the cell and probe counts by v, we have the relevant concentrations $[M^I_{i,j}]$, $[W_j]$, and $[N]$. We can now derive the mass-action equations as

$$\begin{aligned}
\frac{\mathrm{d}[M^I_{i,j}]}{\mathrm{d}t} =& -(I-i)p_+[M^I_{i,j}][N] - (J-j)q_+[M^I_{i,j}][N] + (i+1)p_-[M^I_{i+1,j}] + (j+1)q_-[M^I_{i,j+1}] \\
& + (I-i+1)p_+[M^I_{i-1,j}][N] + (J-j+1)q_+[M^I_{i,j-1}][N], \\
\frac{\mathrm{d}[W_j]}{\mathrm{d}t} =& -(J-j)q_+[W_i][N] + (j+1)q_-[W_{j+1}] + (J-j+1)q_+[W_{j-1}][N], \\
\frac{\mathrm{d}[N]}{\mathrm{d}t} =& q_-\sum_{j=1}^{J} j[M^I_{i,j}] + p_-\sum_{i=1}^{I} i[M^I_{i,j}] - q_+\sum_{j=0}^{J-1}(J-j)[M^I_{i,j}] - p_+\sum_{i=0}^{I-1}(I-i)[M^I_{i,j}].
\end{aligned} \tag{2.2}$$

At chemical equilibrium, we expect detailed balance in each of Eqs. 2.1. For specific and nonspecific binding respectively, we can define the dissociation constants

$$\frac{[N][M_{i,j}]}{[M_{i,j+1}]} = \frac{q_-}{q_+} = K_\mathrm{n} \quad \text{and} \quad \frac{[N][M_{i,j}]}{[M_{i+1,j}]} = \frac{p_-}{p_+} = K_\mathrm{s}. \tag{2.3}$$

One might interpret these constants as the probe concentration's resistance to binding and are parameters that will shape the entire dynamics of the model. Using inductive reasoning, we can characterize the mutant populations solely with the concentration of unbound cells:

$$[M^I_{i,j}] = [M^I_{0,0}]\binom{I}{i}\left(\frac{[N]}{K_\mathrm{s}}\right)^i \binom{J}{j}\left(\frac{[N]}{K_\mathrm{n}}\right)^j. \tag{2.4}$$

By the conservation of total mutant cells with I binding sites, we use the binomial expansion to derive

$$[M^I] = \sum_{i=0}^{I}\sum_{j=0}^{J}[M^I_{i,j}] = [M^I_{0,0}]\left(1+\frac{[N]}{K_\mathrm{s}}\right)^I \left(1+\frac{[N]}{K_\mathrm{n}}\right)^J. \tag{2.5}$$

Using very similar arguments, we derive the concentration of unbound wild-types as

$$[W_0] = [W]\left(1+\frac{[N]}{K_\mathrm{n}}\right)^{-J}. \tag{2.6}$$

Next, we find the following result:

$$\begin{aligned}
\sum_{i=0}^{I}\sum_{j=0}^{J}(i+j)[M^I_{i,j}] &= \sum_{i=0}^{I}\sum_{j=0}^{J}(i+j)[M^I_{0,0}]\binom{I}{i}\left(\frac{[N]}{K_\mathrm{s}}\right)^i\binom{J}{j}\left(\frac{[N]}{K_\mathrm{n}}\right)^j \\
&= [M^I_{0,0}]\left[I\left(\frac{[N]}{K_\mathrm{s}}\right)\left(1+\frac{[N]}{K_\mathrm{s}}\right)^{I-1}\left(1+\frac{[N]}{K_\mathrm{n}}\right)^J + J\left(\frac{[N]}{K_\mathrm{n}}\right)\left(1+\frac{[N]}{K_\mathrm{s}}\right)^I\left(1+\frac{[N]}{K_\mathrm{n}}\right)^{J-1}\right] \\
&= [M^I_{0,0}]\left(1+\frac{[N]}{K_\mathrm{s}}\right)^I\left(1+\frac{[N]}{K_\mathrm{n}}\right)^J\left[\frac{I[N](K_\mathrm{n}+[N])+J[N](K_\mathrm{s}+[N])}{(K_\mathrm{s}+[N])(K_\mathrm{n}+[N])}\right] \\
&= [M^I]\left[\frac{I[N](K_\mathrm{n}+[N])+J[N](K_\mathrm{s}+[N])}{(K_\mathrm{s}+[N])(K_\mathrm{n}+[N])}\right]. \qquad (2.7)
\end{aligned}$$

Using the conservation of the total concentration of initial probes $[N_\mathrm{T}]$, we derive

$$\begin{aligned}
[N_\mathrm{T}] &= [N] + \sum_{j=0}^{J} j[W_j] + \sum_{I=0}^{J}\sum_{i=0}^{I}\sum_{j=0}^{J}(i+j)[M^I_{i,j}] \\
&= [N] + \sum_{j=0}^{J} j[W_0]\binom{J}{j}\left(\frac{[N]}{K_\mathrm{n}}\right)^j + \sum_{I=0}^{J}[M^I]\left[\frac{I[N](K_\mathrm{n}+[N])+J[N](K_\mathrm{s}+[N])}{(K_\mathrm{s}+[N])(K_\mathrm{n}+[N])}\right] \\
&= [N] + \frac{J[N][W]}{K_\mathrm{n}+[N]} + \frac{[N][M]}{(K_\mathrm{s}+[N])(K_\mathrm{n}+[N])}\left[J(K_\mathrm{s}+[N]) + (K_\mathrm{n}+[N])\sum_{I=0}^{J} I f(I)\right]. \qquad (2.8)
\end{aligned}$$

Noting that $\langle I\rangle = \sum_{I=0}^{J} I f(I)$, we can solve Eq. 2.8 for $[N]$ analytically as the roots of a cubic polynomial with $[M]$, $[W]$, $[N_\mathrm{T}]$, $[K_\mathrm{s}]$, $[K_\mathrm{n}]$, $[J]$, and $\langle I\rangle$ as parameters. We will show how this kinetic model can be used to quantify aspects of the assay for optimal experimental design, automatic gating, and parameter inference.

2.2. *Two species Langmuir adsorption model*

If we define θ_T as the fractional occupancy of total binding sites across all cells, using Eq. 2.8 we have

$$\theta_\mathrm{T} = \frac{[N_\mathrm{T}]-[N]}{J[C_\mathrm{T}]} = \left(\frac{\langle I\rangle}{J}\right)\left(\frac{[M]}{[C_\mathrm{T}]}\right)\frac{\frac{[N]}{K_\mathrm{s}}}{1+\frac{[N]}{K_\mathrm{s}}} + \frac{\frac{[N]}{K_\mathrm{n}}}{1+\frac{[N]}{K_\mathrm{n}}} \qquad (2.9)$$

Note that each of the two terms resemble a Langmuir isotherm which measures the fractional occupancy of binding sites on a surface substrate [12]. Framed in the Langmuir adsorption picture, the concentration of free probes $[N]$ is directly analogous to the partial pressure of adsorbing gas. As shown in Fig. 2(a), the fraction of occupied binding sites grows as you add more free probes, but eventually saturates. Note that the saturation is normalized according to the number of non-specific binding sites J as we expect it to be much larger than $\langle I\rangle$. Also note that the rate of adsorption is attenuated by the non-specific dissociation constant K_n. For small $[N]$, when the cell membrane is far from saturation, we see the dynamics of the receptor binding site saturation in Fig 2(b). As $K_\mathrm{s} \gg K_\mathrm{n}$, the total occupancy increases rapidly to saturation relative to that of the more dominant non-specific isotherm. Thus, there is a range of $[N]$ sufficiently large to reach the receptor binding saturation, but low enough to be far from the full membrane saturation. This would be an ideal range to operate one's assay to increase the likelihood that a probe bound to a cell is due to it being bound specifically to receptor as opposed to the membrane itself.

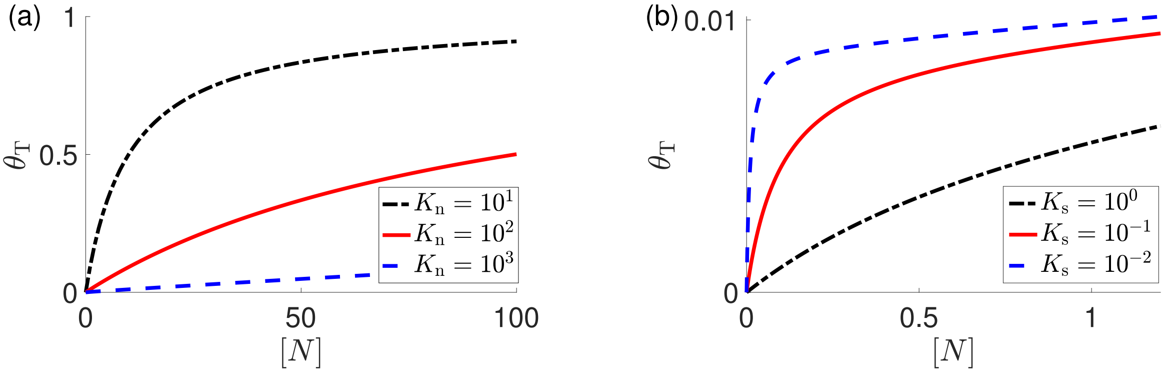

Figure 2. Fractional occupancy θ_T of available binding sites as a function of the free probe concentration $[N]$. (a) The isotherms for large $[N]$ with $[C_\mathrm{T}] = 100$, $[M] = 10$, $K_\mathrm{s} = 10^{-1}$, $J = 10^3$, $\langle I \rangle = 10$, and $K_\mathrm{n} = 10^1$, 10^2, and 10^3. The cell membrane reaches saturation of bound probes at a rate dictated by K_n. (b) The isotherms for small $[N]$ values with $[C_\mathrm{T}] = 100$, $[M] = 10$, $K_\mathrm{n} = 10^3$, $J = 10^3$, $\langle I \rangle = 10$, and $K_\mathrm{s} = 10^{-1}$, 10^{-2}, and 10^{-3}. Due to small K_s, the occupancy reaches saturation for all available receptors quickly, then resumes the slower saturation of non-specific binding to membrane. Optimal assay conditions would be in free-probe ranges just above the specific binding saturation.

2.3. *Cell population mass density*

In order to establish a connection between the equilibrium kinetic model and a typical output of a flow cytometry assay, we define the concentration of cells $[C_r]$ with exactly r probes bound, regardless if they are bound specifically or nonspecifically, as

$$\begin{aligned}[C_r] =&[W_r] + \sum_{I=0}^{J} \sum_{k=0}^{\min(r,I)} [M_{k,r-k}] \\ =&([C_\mathrm{T}] - [M])\left(1 + \frac{[N]}{K_\mathrm{n}}\right)^{-J} \binom{J}{r} \left(\frac{[N]}{K_\mathrm{n}}\right)^{r} \\ &+ [M] \sum_{I=0}^{J} \sum_{k=0}^{\min(r,I)} f(I) \left(1 + \frac{[N]}{K_\mathrm{s}}\right)^{-I} \left(1 + \frac{[N]}{K_\mathrm{n}}\right)^{-J} \binom{I}{k} \left(\frac{[N]}{K_\mathrm{s}}\right)^{k} \binom{J}{r-k} \left(\frac{[N]}{K_\mathrm{n}}\right)^{r-k}. \end{aligned} \tag{2.10}$$

In order to compute Eq. 2.10, we must consider a functional form for the distribution $f(I)$ of the number of receptors I found on a given mutant cell. The receptor number is likely due to transcriptional activity and other cellular processes that result in varying numbers of functional proteins on the membrane. Then a reasonable and simplifying assumption is that I is Poisson-like such that

$$f(I) = \frac{1}{I!} \langle I \rangle^{I} \exp\left(- \langle I \rangle\right) \left(\frac{J - I}{J - \langle I \rangle}\right), \tag{2.11}$$

where the mean $\langle I \rangle$ encompasses all the physiological processes involved in expressing the receptors. Note the extra correction term forcing J to be the carrying capacity for receptors on the membrane is sufficiently close to 1 that it can be ignored for most applications, making the distribution completely Poisson.

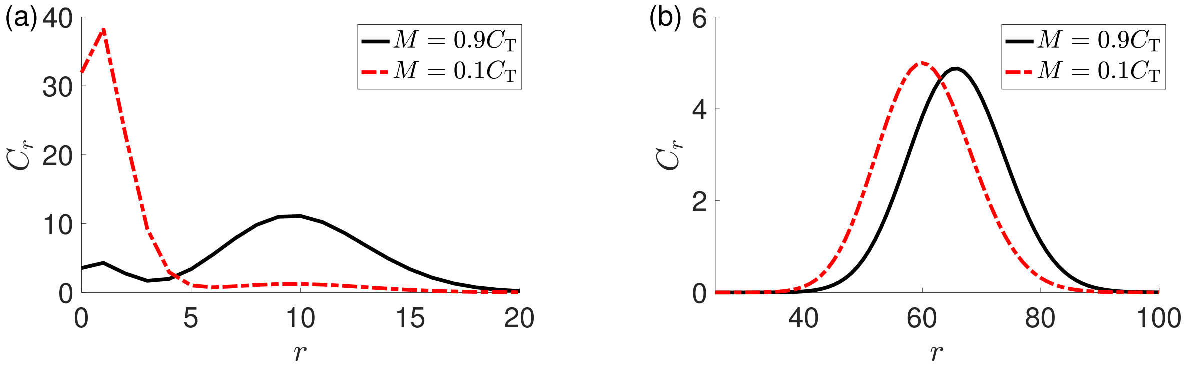

Figure 3. Expected population densities of cells C_r with exactly r probes bound. (a) Low concentrations of free probe $[N] = 1.2$ with $C_{\mathrm{T}} = 100$, $\langle I \rangle = 10$, $J = 10^3$, $K_{\mathrm{s}} = 10^{-1}$, and $K_{\mathrm{n}} = 10^3$ for $M = 10$ and 90 cells. The density will cause clustering of wild-type cells close to $r = 0$ and mutants close to $r = \langle I \rangle$, though the non-specific binding allows some of the density associated with the mutants to contribute to the lower r values of C_r. A clear boundary exists between the two densities and heuristic gating can partition the populations sufficiently. (b) Large concentrations of free probe $[N] = 60$ with $C_{\mathrm{T}} = 100$, $\langle I \rangle = 10$, $J = 10^3$, $K_{\mathrm{s}} = 10^{-1}$, and $K_{\mathrm{n}} = 10^3$. The population densities of wild-types and mutants are now found in similar values of r and overlap extensively, causing difficulty in differentiating the two clusters as probes saturate the membrane.

Eq. 2.10 informs us of how the distribution of cell data will cluster, as illustrated in Fig. 3(a). At relatively low concentrations of free probe $[N]$, the binding of receptors can saturate, but leave the wild-type cells with only nonspecific binding to have significantly lower probe bindings. This effectively makes the two clusters qualitatively separable and imposing a gating threshold is straightforward. However, at high levels of free probe, the clusterings overlap and are thus difficult to differentiate heuristically, as demonstrated in Fig. 3(b). Furthermore, these distributions are taken over the probe binding number r which is not directly measurable. We next show how r and $[C_r]$ relates to the measurable fluorescence intensity distribution.

2.4. Fluorescence intensity

As each cell passes through the cytometer, any bound probes will fluoresce with some strictly positive light intensity F_{s}. However, some variation in the fluorescence signal arises from molecular variability and instrumentation noise, so we assume the intensity is lognormal distributed with the shape parameter σ_0^2 [14]. We also expect each cell to have some relatively small amount of background side scattering with intensity F_0. Then if we define x as the total fluorescence intensity of a given cell and r as its corresponding number of bound probes, then we expect the probability density of x to follow

$$\Pr(x|r) = \frac{1}{x\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(\ln(x) - \ln(F_0 + rF_\mathrm{s}))^2}{2\sigma_0^2}\right]. \tag{2.12}$$

In a typical flow cytometry assay, however, we do not know how many probes are bound to each cell. Furthermore, we do not know the species $b \in \{0, 1\}$ of the cell, where 0 and 1 denote wild-type and mutant respectively. Using Eqs. 2.10 and 2.12, we derive the marginal density of fluorescence intensity as

$$\begin{aligned}\Pr(x) &= \Pr(b=0)\sum_{r=0}^{J}\Pr(x|r)\Pr(r|b=0) + \Pr(b=1)\sum_{r=0}^{J}\Pr(x|r)\Pr(r=j|b=1)\\ &= \frac{1}{[C_\mathrm{T}]}\sum_{r=0}^{J}\frac{[C_r]}{x\sqrt{2\pi\sigma_0^2}}\exp\left[-\frac{(\ln(x)-\ln(F_0+rF_\mathrm{s}))^2}{2\sigma_0^2}\right]. \end{aligned}\tag{2.13}$$

Considering total number of probes r bound to a cell regardless if they are bound specifically or non-specifically is sufficient if each probe fluoresces with an intensity independent of binding. However, for some probes, the fluorescence may be a product of a conformation change when binding to the designed target. This means that for non-specifically bound probes, their conformation change may be partial and can result in a lower mean fluorescence intensity F_n. We must now consider how many probes are bound specifically and nonspecifically, making Eq. 2.10 insufficient for computing the marginal density of x. Thus we derive the conditional densities

$$\Pr(x|b=0) = \left(1 - \frac{[M]}{[C_\mathrm{T}]}\right)\left(1+\frac{[N]}{K_\mathrm{n}}\right)^{-J}\sum_{j=0}^{J}\binom{J}{j}\left(\frac{[N]}{K_\mathrm{n}}\right)^{j}\frac{1}{x\sqrt{2\pi\sigma_0^2}}\exp\left[-\frac{(\ln(x)-\ln(F_0+jF_\mathrm{n}))^2}{2\sigma_0^2}\right], \tag{2.14}$$

and

$$\begin{aligned}\Pr(x|b=1) =& \frac{[M]}{[C_\mathrm{T}]}\left(1+\frac{[N]}{K_\mathrm{n}}\right)^{-J}\sum_{I}f(I)\left(1+\frac{[N]}{K_\mathrm{s}}\right)^{-I}\sum_{i=0}^{I}\sum_{j=0}^{J}\binom{I}{i}\left(\frac{[N]}{K_\mathrm{s}}\right)^{i}\binom{J}{j}\left(\frac{[N]}{K_\mathrm{n}}\right)^{j}\\ &\times\frac{1}{x\sqrt{2\pi\sigma_0^2}}\exp\left[-\frac{(\ln(x)-\ln(F_0+iF_\mathrm{s}+jF_\mathrm{n}))^2}{2\sigma_0^2}\right].\end{aligned}\tag{2.15}$$

In the next section we will show how these mathematical results can be used to iteratively estimate the size of the mutant population $[M]$ from data and infer physical parameters such as K_s and $\langle I\rangle$.

3. Results and Discussion

3.1. EM mixture model estimation of mutant population

Using Eqs. 2.14 and 2.15, we propose an iterative algorithm to automatically infer the concentration of mutant cells $[M]$ without heuristic gating. Let $\vec{x}$ be a set of data, where x_k is the fluorescence intensity measured for cell k and let $b_k \in \{0, 1\}$ be its corresponding species assignment. Because this is an iterative method, we will index each numerical step with $t \in \{0, 1, 2, \cdots\}$. Using Bayes rule, we can compute the probability cell k is a mutant as

$$\Pr(b_k = 1|x_k, [M]^{(t)}) = \frac{[M]^{(t)}\Pr(x_k|b_k = 1, [M]^{(t)})}{([C_\mathrm{T}] - [M]^{(t)})\Pr(x_k|b_k = 0, [M]^{(t)}) + [M]^{(t)}\Pr(x_k|b_k = 1, [M]^{(t)})}, \tag{3.1}$$

where $[M]^{(t)}$ is the current mutant concentration estimate. The iterative procedure starts with an initial guess at the concentration of mutant cells $[M]^{(0)}$ which is used to calculate the probability in Eq. 3.1. The next estimate $[M]^{(t+1)}$ is then given by

$$[M]^{(t+1)} = \sum_{k=1}^{C_\mathrm{T}} \Pr(b_k = 1|x_k, [M]^{(t)}). \tag{3.2}$$

This process is repeated until $[M]^{(t)}$ converges. Note that, though calculating Eq. 3.1 for a single x_k can technically be a $O(J^3)$ computational operation, some asymptotic arguments can be made to concatenate summations to terms that are sufficiently close to zero. More importantly, the only value that changes over all iterations is $[M]^{(t)}$. Thus, the more computationally heavy summations in Eqs. 2.14 and 2.15 can be done once and stored in a matrix, making all subsequent iterations compute linearly with the number of cells C_T.

An example using our algorithm for estimating the mutant cell count from simulated data is shown in Fig. 4(a) where $C_\mathrm{T} = 100$ and $M = W = 50$. Immediately evident is the wild-type cells' propensity to be clustered close to the mutants when as little as one probe is bound. When a reasonable gate threshold is drawn as demonstrated, the 87 cells to the right are counted as mutants, resulting in 27 false positives. Our algorithm accounts for the probability of wild-types having high fluorescence, resulting in the closer estimate of $M = 51$. Even for parameter regimes where probe binding to wild-types are rare, for large numbers of cells C_T, the occasional nonspecific binding event will result in the gating process invariably over-counting mutants.

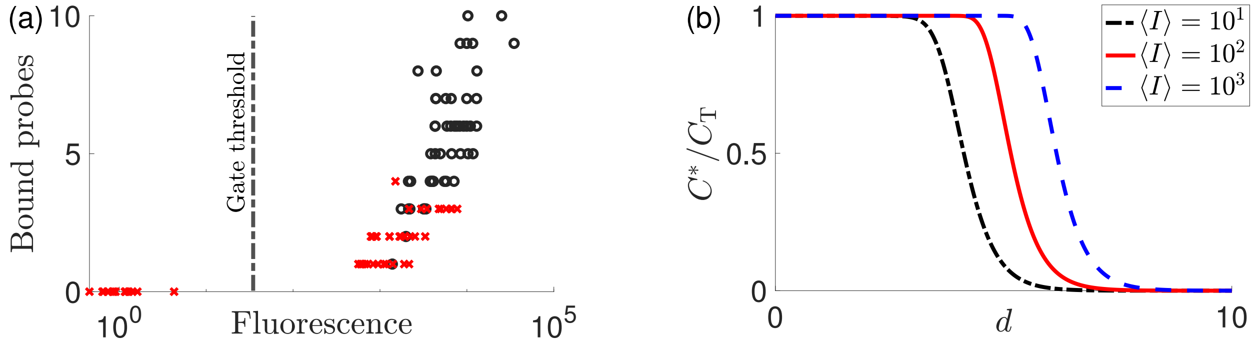

Figure 4. (a) Simulated fluorescence data using parameters $C_T = 100$, $[N] = 1.2$, $K_s = 10^{-1}$, $K_n = 10^3$, $J = 10^3$, $\langle I \rangle = 5$, $F_0 = 1$, $F_s = F_n = 10^3$, and $\sigma_0 = 0.5$. We set $M = 0.5C_T$ and assign each cell k with a number of bound probes r_k using Eq. 2.10 and subsequent fluorescence intensity x_k using Eq. 2.12. At this particular parameter regime, 27 of the 50 wild-type cells managed to bind with at least one probe, increasing their relative fluorescence and clustering them with the mutants. A typical gating threshold, shown above, would separate the two clusters and count all 87 cells on the right hand side as mutants; far larger than the true count of 50. The iterative estimate using Eq. 3.2 returns $M = 51$, relatively close to the actual count. (b) Probability that a given cell has one or more probes bound as a function of the dilution number d as we vary the receptor distribution mean $\langle I \rangle$. Here $[N] = 1$, $K_n = 10^3$, $K_s = 10^{-3}$, $J = 10^3$, and dilution factor $D = 10$.

3.2. *Parameter inference using serial dilution*

In typical flow cytometry assays, probes designed to bind specifically to the receptors of interest are often prepared elsewhere. Thus it is not uncommon for an experimentalist to test the affinity of a probe prior to an assay in order to insure it is sufficiently effective for the planned experiment [5]. This is typically done by preparing a homogeneous suspension of mutant cells with the probes, so that $[M] = [C_T]$. The experimentalist will then perform cytometry with a sufficiently high concentration of free probes $[N]$ and quantify the number of cells that contain any fluorescing probes. The solution of probes is subsequently diluted by some factor D and the assay is repeated $d_{\max}$ number of times. This process, known as serial dilution, arises in many applications from testing antibacterial agents [15] to quantifying viral infectivity [16]. In this context, it is used to find the characteristic dilution number d_c such that all cells are still bound to at least one probe. The experimentalist can then use the corresponding probe concentration for the flow cytometry assay. However, having provided a kinetic model, we can employ this process to infer physical parameters of interest. To do so, we start by using Eq. 2.10 to derive the concentration of the number of cells C^* with one or more probes attached as

$$\begin{aligned}[C^*] &= [C_T] - [C_0] \\ &= [C_T] - [C_T] \sum_I \frac{\langle I \rangle^I \exp(-\langle I \rangle)}{I!} \left(1 + \frac{[N]D^{-d}}{K_s}\right)^{-I} \left(1 + \frac{[N]D^{-d}}{K_n}\right)^{-J} \\ &= [C_T]\left[1 - \left(1 + \frac{[N]D^{-d}}{K_n}\right)^{-J} \exp\left(\frac{-\langle I \rangle \frac{[N]D^{-d}}{K_s}}{1 + \frac{[N]D^{-d}}{K_s}}\right)\right], \qquad (3.3)\end{aligned}$$

where d is the dilution number. If we normalize $[C^*]$ by the total concentration of cells $[C_{\mathrm{T}}]$, then we can treat the expression as a probability that a given cell will fluoresce as a function of d, as shown in Fig. 4(b). The placement of the characteristic drop in probability is dictated by the parameters K_{s} and $\langle I \rangle$. If we consider the data C^*_d as the number of cells that fluoresce at dilution d, then we expect its value to be binomial distributed and we can derive the likelihood function

$$\mathcal{L}\left(\vec{C}^*_d\right) = \prod_{d=d_{\min}}^{d_{\max}} \binom{C_{\mathrm{T}}}{C^*_d} \left[1 - \left(1 + \frac{[N]D^{-d}}{K_{\mathrm{n}}}\right)^{-J} \exp\left(\frac{-\langle I \rangle \frac{[N]}{K_{\mathrm{s}}}}{D^d + \frac{[N]}{K_{\mathrm{s}}}}\right)\right]^{C^*_d} \left[\left(1 + \frac{[N]D^{-d}}{K_{\mathrm{n}}}\right)^{-J} \exp\left(\frac{-\langle I \rangle \frac{[N]}{K_{\mathrm{s}}}}{D^d + \frac{[N]}{K_{\mathrm{s}}}}\right)\right]^{C_{\mathrm{T}}-C^*_d} . \tag{3.4}$$

For a given set of a data $\vec{C}^*_d$, the log of the likelihood is a function of the parameters and can be maximized to solve for maximum likelihood estimates (MLE) of these parameters. As the original intent of the serial dilution procedure is to quantify the affinity of specific probe binding, K_{s} would be the desired inferred parameter. However, depending on the underlying experiment, one can envision estimating the expression of surface receptors $\langle I \rangle$ and its change under differing experimental environments.

3.3. Applications in FACS

A very common use of flow cytometry is in fluorescence activated cell sorting (FACS) in which cells are physically sorted into bins based on their species type [8, 17]. As each cell is sent past the laser, the intensity measurement informs the computer in real-time which category the cell falls into. The droplet containing the cell exits an electrically charged ring that induces an electric charge in the droplet. An electric field controlled by the computer is then used to propel the extruded cell into the appropriate bin based on the fluorescence measurement. However, the confounding factors previously discussed can cause incorrect sorting of cells due to non-specific binding and other background fluorescence. If all parameters are *a priori* known, then using Eq. 3.1 can technically be used to determine the cell species as the expression quantifies the probability that a cell is a mutant over a wild-type cell given its fluorescence. A resulting probability larger than 0.5 will indicate a mutant, making Eq. 3.1 a decision function. However, there are two complications. One is that the evaluations of Eq. 3.1 are relatively computationally intensive, especially if the expected number of receptors $\langle I \rangle$ is large. The real-time nature of the physical process of FACS requires rapid evaluation, though increasing computational resources can alleviate the problem. The second, more pertinent issue is that, though we are assuming all parameters are known, it is unlikely that the concentration of mutants $[M]$ is *a priori* known. Biologists typically use cytometry assays after some experiment and the quantification of $[M]$ is often the primary desired quantity still undetermined. Furthermore, our method of estimating $[M]$ is a model-based clustering technique that leverages all data collectively, making real-time analysis problematic.

One potential solution for both problems is to use a two-pass cytometry method. One pass through the cytometer would be used to quantify the concentration of mutants $[M]$ while also storing the evaluations of Eq. 3.1. All cells would be collected together and reintroduced to the sheath fluid for a second pass for the FACS step. Though it would be improbable to exactly match each cell with their stored evaluation in the first pass, this extra data will act as a prior for more informed statistical sorting of the cells. Though previously discussed applications of our model use protocols already practiced by biologists, the potential overhead of using a two-pass cytometry process would be subject to the specific requirements of each experiment employing the method.

4. Conclusions

In this paper, we have created a full kinetic model of the specific and nonspecific binding dynamics of a cell/probe suspension at chemical equilibrium. Using a mass-action approach, we derived expressions for important equilibrium quantities as functions of physical and experimental parameters of the flow cytometry assay. The total number of afflicted cells, which we refer to as mutants, is often the primary desired quantity of the protocol as the probes are assumed to attach only to those cells. However, we show quantitatively how the nonspecific binding of probes to the membranes of both mutants and wild-type cells can confound the results. Furthermore, using the analogous Langmuir adsorption isotherm, we demonstrated how to choose probe concentration that will minimize these confounding effects. For the estimation of the total number of mutants in flow cytometry output, which is often subject to heuristic gating, we provided an iterative algorithm to obtain this number without input from the experimentalist. We claim that having a fundamental model for which the algorithm is based will increase the accuracy over other clustering attempts. Furthermore, we extract further utility from a serial dilution process often employed to measure the affinity of probes to infer true physical parameters of the cells. Lastly, we discuss the potential applications and issues with using our method for fluorescence activated cell sorting (FACS) while proposing a two-pass cytometry process to alleviate some of the problems.

Our model and analysis approach can be readily extended to include multiple probes, multiple specifically binding receptors, and more general distribution functions for receptor expression by the mutant cells. We expect that in such more complex, higher dimensional discrimination assays, our more systematic and quantitative analysis methods should provide more accurate results. Finally, we are developing a web based tool that fully implements our flow cytometry analysis procedure so that it can be applied to experimentally measured data. This will increase the accessibility of our model and enable quantitative comparisons with existing methods, including heuristic gating.

Acknowledgments

This research was supported by the National Science Foundation through grants DMS-1516675 and DMS-1814364. We are especially grateful to Anupama Dimashkie and Dr. Saravanan Karumbayaram for insightful discussions.

References

1. Ignacio A. Zuleta, Andrés Aranda-Díaz, Hao Li and Hana El-Samad, Dynamic characterization of growth and gene expression, *Nat. Meth.*, **11** (2014), 443–450.
2. Hiroshi Abe and Eric T. Kool, Flow cytometric detection of specific RNAs in native human cells with quenched autoligating FRET probes, *Proc. Nat. Acad. Sci.*, **103** (2006), 263–268
3. Sarah B. Joseph, Kathryn T. Arrildt, Adrienne E. Swanstrom, Gretja Schnell, Benhur Lee, James A. Hoxie and Ronald Swanstrom, Quantification of entry phenotypes of macrophage-tropic HIV-1 across a wide range of CD4 densities, *J. Virol*, **88** (2014), 1858–1869.
4. Nicholas E. Webb and Benhur Lee, Quantifying CD4/CCR5 usage efficiency of HIV-1 Env using the Affinofile system, *HIV Prot.*, 1^{st} edition, Springer, New York, 2016, 3–20.

5. Thomas Myles Ashhurst, Adrian Lloyd Smith and Nicholas Jonathan Cole King, High-dimensional fluorescence cytometry, *Curr. Prot. Immun.*, **119** (2017), 1–38.

6. Benoîte Bourdin, Emilie Segura, Marie-Philippe Téreault, Sylvie Lesage and Lucie Parent, Determination of the relative cell surface and total expression of recombinant ion channels using flow cytometry, *J. Vis. Exp.*, **115** (2016), 54732.

7. Aysun Adan, Gunel Alizada, Yagmur Kiraz, Yusuf Baran and Ayten Nalbant, Flow cytometry: basic principles and applications, *Crit. Rev. Biotech.*, **37** (2017), 163–176.

8. Leonard A. Herzenberg, David Parks, Bita Sahaf, Omar Perez, Mario Roederer and Leonore A. Herzenberg, The history and future of the fluorescence activated cell sorter and flow cytometry: A view from Stanford, *Clin. Chem.*, **48** (2002), 1819–1827

9. Kenneth Lo, Ryan Remy Brinkman and Raphael Gottardo, Automated gating of flow cytometry data via robust model-based clustering, *J. Intl. Soc. Anal. Cyt.*, **73** (2008), 321–332.

10. J.G. Kenna, G.N. Major and R.S. Williams, Methods for reducing non-specific antibody binding in enzyme-linked immunosorbent assays, *J. Immun. Meth.*, **85** (1985), 409–419.

11. Chris P. Verschoor, Alina Lelic, Jonathan L. Bramson and Dawn M. E. Bowdish, An introduction to automated flow cytometry gating tools and their implementation, *Front. Immun.*, **6** (2015), 380

12. R. A. Burns, M. Y. El-Sayed and M. F. Roberts, Kinetic model for surface-active enzymes based on the Langmuir adsorption isotherm: phospholipase C (bacillus cereus) activity toward dimyristoyl phosphatidylcholine/detergent micelles, *Natl. Acad. Sci. USA*, **79** (1982), 4902–4906.

13. Kenneth Lange, *Mathematical and statistical methods for genetic analysis*, 1^{st} edition, Springer-Verlag, New York, 1997.

14. Sarah A. Mutch, Bryant S. Fujimoto, Christopher L. Kuyper, Jason S. Kuo, Sandra M. Bajjalieh and Daniel T. Chiu, Deconvolving single-molecule intensity distributions for quantitative microscopy measurements, *Biophys J.*, **92** (2007), 2926–2943.

15. European Committee for Antimicrobial Susceptibility Testing of the European Society of Clinical Microbiology and Infectious Diseases, Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution, *Clin. Microbio. Infect.*, **9** (2003), 9–15

16. Bhaven A. Mistry, Maria R. D'Orsogna and Tom Chou, The effects of statistical multiplicity of infection on virus quantification and infectivity assays, *Biophysi J.*, **114** (2018), 2974–2985

17. Jason P. Awe, Patrick C. Lee, Cyril Ramathal, Agustin Vega-Crespo, Jens Durruthy-Durruthy, Aaron Cooper, Saravanan Karumbayaram, William E. Lowry, Amander T. Clark, Jerome A Zack, Vittorio Sebastiano, Donald B. Kohn, April D. Pyle, Martin G. Martin, Gerald S. Lipshutz, Patricia E. Phelps, Renee A. Reijo Pera and James A. Byrne, Generation and characterization of transgene-free human induced pluripotent stem cells and conversion to putative clinical-grade status, *Stem Cell Res. Ther.*, **4** (2013), 87

AIMS Press