ELSEVIER

Journal of Immunological Methods 243 (2000) 211–227

JIM
Journal of
Immunological Methods

www.elsevier.nl/locate/jim

Flow cytometric screening of cell-based libraries

Patrick S. Daugherty[a,*], Brent L. Iverson[b,d], George Georgiou[c,d]

[a]*Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109, USA*
[b]*Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, TX 78712, USA*
[c]*Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, USA*
[d]*Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA*

Abstract

Flow cytometry is a powerful, high-throughput library screening tool in numerous applications including the isolation of bioactive molecules from synthetic combinatorial libraries, the identification of virulence genes in microorganisms, and the study and engineering of protein functions. Using flow cytometry, large libraries of protein mutants expressed in microorganisms can be screened quantitatively for desired functions, including ligand binding, catalysis, expression level, and stability. Rare target cells, occurring at frequencies below 10^{-6}, can be detected and isolated from heterogeneous library populations using one or more cycles of cell sorting and amplification by growth. Flow cytometry is particularly powerful because it provides the unique opportunity to observe and quantitatively optimize the screening process. However, the ability to isolate cells occurring at such low frequencies within a population requires consideration and optimization of screening parameters. With this aim, an analysis of the various parameters involved in screening cell-based libraries for rare target cells possessing a desired trait is presented.

Keywords: Flow cytometry; FACS; Combinatorial libraries; High-throughput screening; Rare cell sorting

1. Introduction

Flow cytometric (FCM) analysis of microbial populations is a well-established technique that continues to gain momentum (Davey and Kell, 1996). Although early studies demonstrated that FCM can be employed as a tool in cloning and mutant isolation (Minas et al., 1988), only during the last decade has FCM and cell sorting been applied in the areas of bacterial genetics and protein engineering. The expanded use of FCM as a high-throughput screening tool has been stimulated by the development of fluorescent protein probes to monitor gene expression and the growing applications of combinatorial libraries. In turn, FCM has fostered the development of powerful cell-based library technologies.

The development of green fluorescent protein (GFP) and other fluorescence-based protein reporters of gene expression (Griffin et al., 1998; Tsien, 1998) has opened new avenues for the genetic analysis of transcriptional regulation in microorganisms. Fusions

Abbreviations: FCM, flow cytometry; GFP, green fluorescent protein; scFv, single-chain antibody fragment; scTCR, single-chain T-cell receptor; FRET, fluorescence resonance energy transfer; PI, propidium iodide; FACS, fluorescence-activated cell sorting

*Corresponding author. Tel.: +1-206-667-3672; fax: +1-206-667-6523.

E-mail address: pdaugher@fhcrc.org (P.S. Daugherty).

0022-1759/00/$ – see front matter
PII: S0022-1759(00)00236-2

to reporter proteins whose level can be easily determined by enzymatic assays have been used extensively to monitor the transcription of bacterial genes. The most commonly used reporter protein is β-galactosidase, whose activity can be readily monitored on agar plates using the chromogenic substrate X-gal (Silhavy and Beckwith, 1985). However, plate-based, chromogenic assays have two limitations: first, even when plating cells at high densities, the number of clones that can be analyzed is usually less than 10^5. Second, the colonies expressing functional β-galactosidase fusions frequently range from pale blue to deep blue in color. Thus, the decision on which colonies to select for further analysis is subjective and does not rely on a quantitative measurement of reporter protein expression. The use of fluorescent reporter proteins in conjunction with FCM circumvents both problems. In a recent application of FCM to analyze gene regulation, Falkow and coworkers used FCM to isolate *S. typhimurium* promoters from a library, as upstream fusions to GFP, that are upregulated specifically when the bacteria colonize an animal model (Valdivia and Falkow, 1996, 1997). GFP is becoming the preferred reporter protein for monitoring gene expression and for in vivo fluorescent labeling of bacteria for detection (Vazquez-Torres et al., 1999). However, the isolation of blue, cyan, and yellow mutants of GFP as well as the recent discovery of red fluorescent proteins (Matz et al., 1999) suggests that fluorescent proteins suitable as reporters in FCM assays will soon be available spanning the visible spectrum.

In recent years, the screening of synthetic combinatorial libraries has emerged as one of the most powerful methods for identifying small molecule effectors of biological function. Typically, libraries of molecules are synthesized, usually by solid phase chemistry, and screened for biological activity either in vitro or in vivo. In vitro assays rely on binding to a protein target or inhibition of an enzyme, whereas in vivo assays test for the ability of a molecule to modulate a specific signal transduction pathway (Wipf et al., 1997). Fluorescence-based FCM assays are increasingly used to screen for molecules that bind to a target protein in vitro or that exhibit activity in a cell-based assay. In the latter case, the effect of a small molecule library on signal transduction is analyzed using GFP gene fusions. Bioactive molecules that increase or decrease gene expression result in a distinct fluorescence profile and are identified directly or from library pools by deconvolution. However, for high-throughput compound identification after screening, a physical link between each molecule in the library and a corresponding encoded tag is generally required to identify definitively the molecule of interest. In one of the first examples of encoded chemical libraries (Needels et al., 1993), a peptide library was synthesized on 10 μm polystyrene beads, and each bead was tagged with a corresponding oligonucleotide. The bead library was screened for binding to a fluorescently-tagged ligand using FCM. Highly-fluorescent beads were isolated and the sequence of the oligonucleotide tag was determined, revealing the identity of the corresponding ligand binding peptide.

FCM also offers numerous advantages in the screening of biosynthetic protein libraries displayed on the surface of microorganisms (Georgiou et al., 1997). Protein libraries displayed on the surface of bacteria, fungi, or higher cells provide an economical, self-amplifying and renewable source of encoded libraries. Here, the cell provides the physical link between the displayed protein and encoding gene within the cell. A library of cells, each displaying a unique protein mutant, is incubated with a fluorescently-tagged ligand in solution. Under stringent conditions (i.e. after prolonged washing in the presence of an excess of unlabelled ligand), only cells that display proteins with the highest affinity remain fluorescent and are isolated using FCM. Plasmid DNA is then isolated from individual clones and sequenced to identify the selected proteins. This technique has been used primarily to isolate single-chain Fv (scFv) antibodies exhibiting increased affinity for ligands (Boder and Wittrup, 1997b; Daugherty et al., 1998, 2000). However, the isolation of high affinity mutants of single-chain T-cell receptors (scTCRs) and protease inhibitors has also been reported (Christmann et al., 1999; Wentzel et al., 1999; Kieke et al., 1999). Cell surface display in conjunction with FCM has also been used successfully to isolate protein mutants that exhibit improved expression characteristics in heterologous hosts (Shusta et al., 1999a). Finally, we recently developed a technology for the isolation of enzymes from large

libraries displayed on the surface of *E. coli*, in which enzyme-displaying cells are screened directly for catalytic turnover by FCM (Olsen et al., 2000). As shown in Fig. 1, cells displaying a protein mutant library are incubated with a fluorescence resonance energy transfer (FRET) substrate carrying a +3 positive charge that allows it to associate with the negatively-charged cell surface. In the absence of cleavage, the substrate's red fluorophore is sequestered into a hydrophobic environment causing increased red fluorescence. Enzymatic cleavage of a designed scissile bond in the substrate results in the release of the uncharged portion of the molecule containing the quenching fluorophore, Q. Consequently, FRET is disrupted and the fluorescent moiety Fl accumulates on the cell surface in proportion to the catalytic activity of the displayed protein.

The screening of protein libraries by FCM was predicated by the development of expression systems for protein surface display (Georgiou et al., 1997; Stahl and Uhlen, 1997). Early attempts to construct fusions for targeting non-native proteins to the cell surface were largely unsuccessful either because the fusion proteins failed to localize properly or because they were overly toxic to the host cell. Fusions to membrane proteins cause damage to the cell wall and can result in reduced viability. Although cell wall damage renders the fusion proteins accessible to fluorescently-tagged ligand, since the cells are not viable they cannot be grown and used for further studies. The first useful system for displaying full-length heterologous proteins on Gram-negative bacteria employed a chimeric Lpp-OmpA fusion to target passenger proteins to the cell surface (Francisco et al., 1992). Subsequently, it was shown that cells displaying an scFv antibody could be enriched from a 10^5 excess of cells displaying an unrelated protein using FCM (Francisco et al., 1993). Since then, Lpp-OmpA fusions have been used to screen libraries of antibodies and other proteins (Christ-

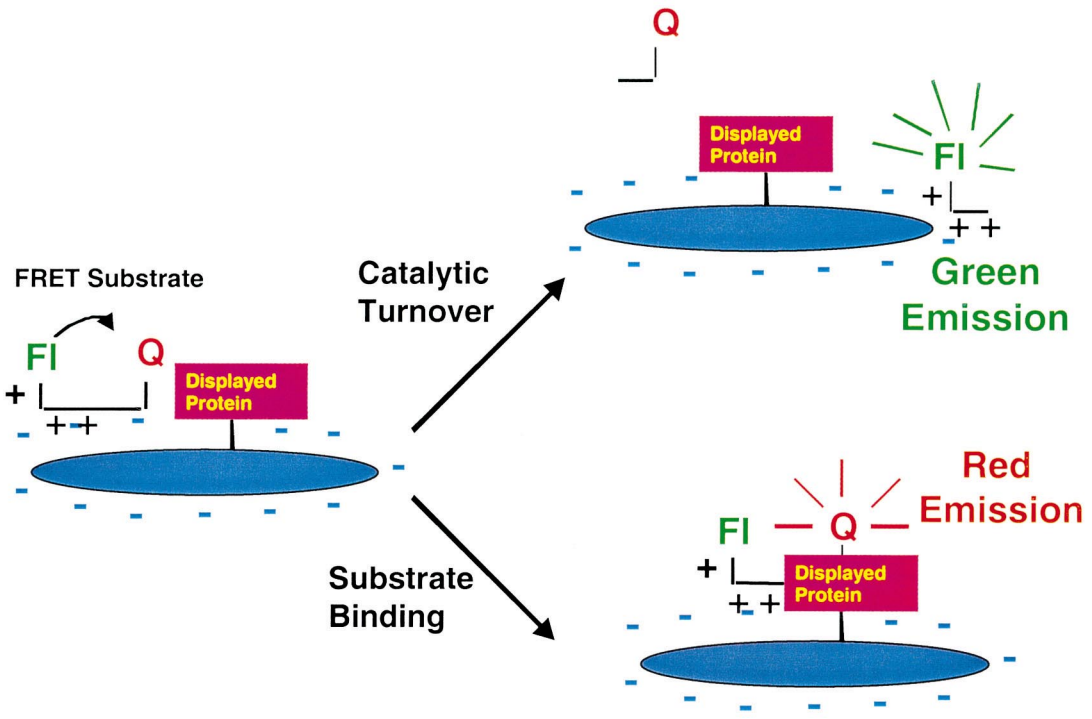

Fig. 1. FCM screening of cell surface displayed enzyme libraries using FRET substrates. Enzymatic cleavage of a designed scissile bond in the substrate results in the release of the uncharged portion of the molecule containing the quenching fluorophore, Q. Consequently, FRET is disrupted and the fluorescent moiety Fl accumulates on the cell surface in proportion to the catalytic activity of the displayed protein. In the absence of cleavage, the substrate's red fluorophore is sequestered into a hydrophobic environment causing increased red fluorescence.

mann et al., 1999; Daugherty et al., 1998). Other fusion systems for protein display in *E. coli* have been developed in recent years (Chang et al., 1999; Jung et al., 1999; Klauser et al., 1993). However, aside from Lpp-OmpA fusions, only the Iga_{β} autotransporter has been used in conjunction with FCM for screening protein libraries (Wentzel et al., 1999). Techniques for protein display in Gram-positive bacteria such as *Streptococci* and *Staphylococci* have also been reported (Fischetti et al., 1996; Gunneriusson et al., 1996; Stahl and Uhlen, 1997), but, thus far, have not been applied in library screening applications.

Wittrup and coworkers have successfully used fusions to the *S. cerevisiae* surface protein Aga2 for protein display in yeast (Boder and Wittrup, 1997b; Kieke et al., 1999), a system commercially available from Invitrogen (Carlsbad, CA, USA). They generated libraries of scFv antibodies and scTCR receptors by random mutagenesis and isolated, respectively, protein mutants with higher ligand affinity or better expression characteristics. Using yeast display and FCM, Wittrup and co-workers reported the evolution of the first sub-picomolar antibody (Shusta et al., 1999b). Yeast, being eukaryotic organisms, may be more suitable for the functional expression of some complex proteins (i.e. glycoproteins) that generally fail to fold correctly in bacteria. On the other hand, DNA transformation of *E. coli* is orders of magnitude more efficient than that of yeast, making *E. coli* the preferred host for screening large libraries. Library display technologies using eukaryotic cells such as insect cells have also been described (Ernst et al., 1998).

The utility of FCM as a high-throughput screening tool stems from the ability to perform quantitative biological assays on large populations, with single cell resolution. With advanced commercial cytometers, such as the MoFlo™ from Cytomation Inc., it is now possible to sort large populations at rates as high as 75,000 cells per second. During the screening process, multiple quantitative parameters for each cell can be analyzed simultaneously, including various fluorescence signals and forward and 90° light scattering.

The combination of high-throughput and quantitative multi-parameter population analysis is particularly well-suited for protein engineering applications. FCM analysis provides the opportunity to examine a heterogeneous distribution of protein functions within a library and to determine quantitatively the fraction of clones having an activity of interest. This information is valuable for understanding the effect of mutations on protein function and, ultimately, for engineering proteins with desired binding or catalytic properties. However, screening combinatorial cell-based libraries presents new technical challenges relative to traditional FCM applications. First, unlike other rare cell applications, library experiments typically require definition of the target cells using a single fluorescence parameter, which may overlap with that of non-target cells. For example, to isolate cell surface-displayed proteins with affinity for a ligand, a sorting window must be set based on the fluorescence due to the binding of the tagged ligand. However, the fluorescence of high affinity clones is only slightly greater than that of clones with moderate affinity, producing a low signal-to-background ratio. This situation contrasts with the sorting of rare cells from biological samples, in which the desired population can be gated more precisely using a panel of fluorescently-tagged antibodies specific to surface markers on only the target cells. Second, microorganisms are smaller than eukaryotic cells and therefore the magnitudes of the fluorescent signals used for sorting are generally smaller. Third, the complex chemical composition of the cell wall of bacteria and yeast often results in high non-specific binding of fluorescent dyes, reducing signal-to-noise and increasing the frequency of false-positive cells. Finally, most FCM facilities have little experience with bacteria or yeast. In fact, many FCM facilities do not permit the analysis of microbial cells for fear of contamination. This fear is largely unjustified, however, since proper sterilization of the instrument after use prevents contamination.

This review addresses the issues encountered in the screening of cell-based libraries using FCM. The analysis presented here is motivated by the screening of polypeptide libraries displayed on the cell surface for high affinity ligands (Georgiou et al., 1997; Boder and Wittrup, 1997a,b; Daugherty et al., 1998, 1999, 2000; Ernst et al., 1998; Wentzel et al., 1999). However, this analysis is also relevant to intracellular libraries, using fusions to fluorescent reporter proteins (Dunn and Handelsman, 1999; Valdivia and

Falkow, 1996), the identification of bio-active ligands (Cho et al., 1998), and the engineering of enzymes and metabolic pathways (Olsen et al., 2000; Douglas and Ballou, 1980). Though the specific examples being discussed here arise from our experience with protein libraries displayed on the surface of *E. coli*, the conclusions are generally applicable to most microbial library screening experiments. The increasing availability of flow cytometers should make cell-based library approaches available to researchers from a wide range of backgrounds. Basic molecular biology experience and a working knowledge of FCM should be sufficient to design a successful library screening experiment for a given application.

2. Experimental design for library screening

Despite the outstanding potential of cell sorting in library screening applications, relatively few experimental protocols have been reported (Cormack et al., 1996; Boder and Wittrup, 1997a,b; Daugherty et al., 1998, 1999; Ernst et al., 1998; Christmann et al., 1999). In these studies, it has proven possible to perform a few simple measurements and statistical calculations and thereby determine whether rare target cells with the desired properties are represented in the library and whether they will likely be isolated. Many factors contribute to the overall probability of isolating rare target cells, including cytometer design, signal-to-noise ratio(s), cell sorting speed, sorting modes, library size, cell fragility, and target cell frequency. Each of these factors must be considered carefully to maximize the probability of isolating rare target cells. In addition, the considerations important to the sorting of rare mammalian cells from biological sources also apply to the sorting of microbial libraries (Leary, 1994; Radbruch and Recktenwald, 1995; Rosenblatt et al., 1997).

The general methodology for the isolation of target cells from libraries is shown in Fig. 2. It is important to note that whereas the isolation of rare mammalian cells is often a single-step process, the screening of large microbial libraries involves multiple rounds of enrichment through sorting and regrowth. Multiple rounds of enrichment sorting are necessary for low frequency ($<10^{-6}$) target cells since (1) the fluorescence distributions of target and non-target cells often overlap and (2) the occurrence of false positives directly limits the enrichment.

2.1. Target cell frequency

The frequency of the desired target events in a microbial library dictates the overall sorting strategy and should be estimated prior to screening. For protein libraries created using cassette mutagenesis (Wells et al., 1985), the expected frequency of the most rare target event in the library can be precisely determined. For example, when the codons corresponding to six amino acids are randomized using an (NNS) scheme (N=G,A,T,C; S=G,C) resulting in a total of $32^6 \approx 10^9$ unique DNA sequences, the most rare polypeptide will occur at a frequency of 10^{-9}. Similarly, for libraries generated by random mutagenesis the frequency of the most rare target cell should be assumed to be the reciprocal of the number of independent transformants. The library size that can be exhaustively screened is determined by the maximum sorting rate afforded by the instrument. Currently, commercially available FCM instrumentation is capable of sorting rates between 10^7 and 10^8 cells/h. Thus, the maximum library size that can be screened exhaustively with 95% confidence is roughly 10^8 members as predicted by the Poisson approximation (Lowman et al., 1991). While this may seem like an impressive number, it must be kept in mind that the protein sequence space is vast: a throughput of 10^8 will allow the analysis of only one-tenth of the six residue library discussed above. Fortunately, library design and not simply size determines whether gain-of-function protein mutants can be isolated (Crameri et al., 1998; Daugherty et al., 2000; Bogarad and Deem, 1999). If necessary, larger libraries, containing $>10^8$ members, can be processed using magnetic cell sorting to pre-enrich target cells immediately prior to FCM sorting (Christmann et al., 1999; Wentzel et al., 1999). Though magnetic sorting has proven useful as an enrichment tool, only FCM enables the quantitative real-time analysis of multiple parameters needed for isolating subtly different cells with high purity.

In some cases, it is possible to estimate the frequency of target cells in a library by FCM analysis of a subset of the library population. About

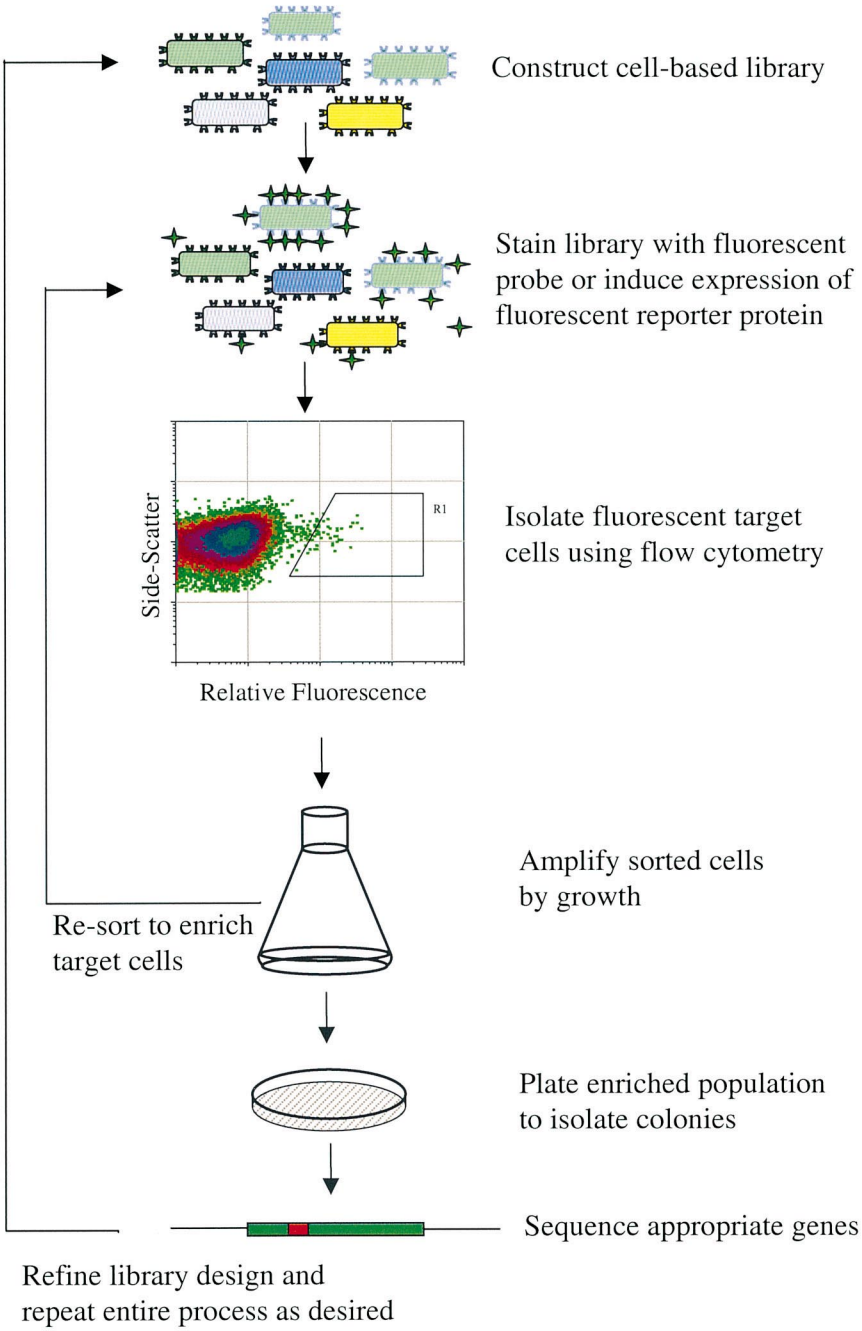

Fig. 2. The screening of cell-based libraries using FCM. After construction of a cell-based library, the cell population is amplified by growth and cells are labeled with appropriate fluorescence probes, or reporter gene expression is induced. Target cells are defined using multiple parameters and enriched from the population using multiple rounds of sorting and amplification by growth. When enrichment is evident by FCM analysis, cells are plated and the protein sequences of individual clones are determined by DNA sequencing. If further optimization is desired, a new library is designed using one or more individuals isolated from the initial library as a template and the entire process is repeated.

10^6 cells can be analyzed by most common flow cytometers in less than 30 min to estimate the frequency of rare target cells prior to screening (Fig. 3). In addition to providing target cell frequency information, such analysis may assist the estimation of maximal enrichment factors for single-step sorting. For example, by performing an identical frequency analysis on a cell population that does not contain target cells, the frequency of background or false-positive cells can be estimated (Fig. 3). Estimating the frequency of false positives is important because it dictates the maximum obtainable enrichment per round of sorting. In surface-displayed protein libraries, the frequency of false positives is determined by analyzing a population of non-mutated, wild-type cells or a mixed population of cells displaying non-binding proteins after incubation with the fluorescent probe that will later be used to screen the library. Analogous control cell populations should be analyzed for other cell-based library screening experiments.

2.2. Minimizing false positives

The frequency of false-positive events is often the limiting factor in the detection and enrichment of rare target cells (Gross et al., 1993; Leary, 1994; Radbruch and Recktenwald, 1995). False-positive events arise from two main sources: (1) non-specifically stained or background cells and debris and (2) instrument-related background, caused by residual cells and particles from previous experiments,

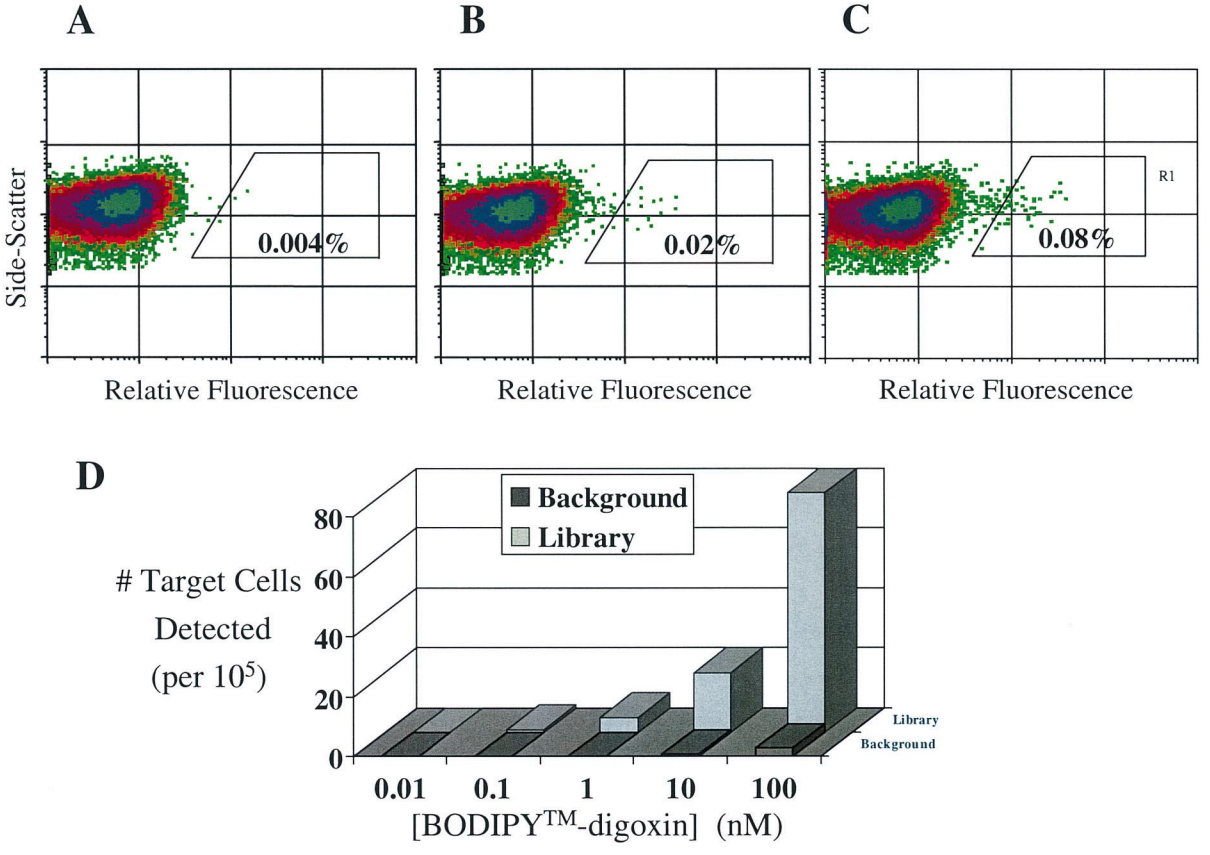

Fig. 3. FCM analysis of target cell frequency. A library of single-chain antibody mutants containing $>10^6$ unique clones was analyzed by FCM after labeling with a BODIPY™-tagged antigen at a concentration of (A) 1 nM, (B) 10 nM, or (C) 100 nM. (D) 10^5 cells from the library and a background control population were analyzed by FCM to determine the frequency of target cells as a function of probe concentration, relative to the frequency of false positives in a control population (Daugherty et al., 1998).

fluidic bursts, and electronic noise (Gross et al., 1995). The latter false-positive source is generally a technical limitation and has been addressed previously with success (Gross et al., 1995). False positives arising from non-specific staining can be greatly reduced by a suitable choice of fluorescent probes and sorting parameters. The ideal fluorescent probe for target cell labeling should exhibit low levels of non-specific staining of the non-target cells. Not surprisingly, the concentration and physicochemical properties of the fluorescent probe can strongly influence the level of non-specific staining. For example, strongly hydrophobic or highly positively-charged probes generally exhibit high levels of non-specific binding to bacteria. As the concentration of fluorescent probe increases, the noise due to non-specific adsorption of the probe increases, whereas the signal due to binding to a specific receptor reaches saturation. In addition to non-specific staining, the free fluorescent probe increases the background fluorescence. As a rule of thumb, the signal-to-noise becomes unacceptable in the micromolar range of probe concentration, even with well-designed fluorescent probes.

The use of multiple parameters for target cell detection and non-target cell exclusion is an effective strategy to minimize false positives. In particular, staining the library cell population with one or more additional 'exclusion' or negative selection fluorescent probes of a different color can be used to exclude false positives such as dead cells and aggregates of the fluorescent probe(s) (Leary, 1994), thereby reducing background staining by orders of magnitude (Gross et al., 1995; Radbruch and Recktenwald, 1995). Cells that become labeled by both the primary and the secondary probes can be excluded from the target cell sort gate using the cytometer software. Propidium iodide (PI) and other DNA intercalators are useful as effective non-specific staining controls and, simultaneously, as viability indicators (Davey and Kell, 1996). Using PI for negative staining of *E. coli* libraries, we have observed greater than 10-fold reduction of false-positive events. Fluorescently-tagged antibodies that do not specifically label the target cells may also be effective negative selection probes. For example, if a green BODIPY-FL™ probe (Molecular Probes (Eugene, OR, USA)) is used for selection, one or more PE (red)-labeled, non-specific antibodies could be used in conjunction with PI (red) for negative selection. Additionally, the use of both forward and 90° light scatter parameters to define target cells and exclude non-target cells can reduce the frequency of false positives, since dead or damaged cells often produce irregular light scatter signals.

2.3. Resolution of target cells from non-target cells

Rare target cells are most often resolved from a library population using a single parameter that provides a measure of the protein or cellular function of interest. Often, the fluorescence distributions of rare target cells overlap with those of non-target cells. When overlap occurs, it is crucial to maximize the fluorescence signal of the target cells relative to non-target cells. The fluorescence ratio of target to non-target cells (F_r) can be increased by:

(a) Increasing the number of cellular binding sites for the primary fluorescent probe. The surface density or intracellular concentration a given receptor, enzyme, or fluorescent reporter protein can be increased by altering the expression level. However, as will be discussed later, high expression levels can adversely effect cell viability and thus the ability to grow and amplify the sorted cell population.

(b) A brighter fluorescent probe. For example, the widely used probe, fluorescein, exhibits maximum fluorescence in alkaline solutions (pH 8.8), while emission is reduced in solutions maintained at near physiological pH. We have found that the signal due to specific binding of fluorescein conjugated to the cardiac glycoside digoxin was not sufficient to allow the discrimination of *E. coli* displaying anti-digoxin antibodies from the non-target cells in a library. However, the same library could be effectively screened for clones expressing high-affinity antibodies using a digoxin-BODIPY™ conjugate (Daugherty et al., 1998), since BODIPY dyes are about two-fold brighter than fluorescein at physiological conditions and give adequate resolution of target and non-target cells.

(c) Signal amplification. Signal amplification has been used successfully in the screening of cell surface-displayed libraries for the isolation of clones that express proteins exhibiting high affinity for a

ligand (Boder and Wittrup, 1997b; Wentzel et al., 1999). The most common amplification scheme involves staining the library with a biotinylated-ligand and subsequently with streptavidin-*R*-phycoerythrin.

2.3.1. Optimization of sorting parameters

The ability to quantitatively optimize the isolation of target cells using FCM provides significant advantage in library screening applications. In the screening of cell surface-displayed libraries for ligand binding, conditions can be optimized to isolate clones expressing proteins exhibiting a desired range of ligand dissociation (or association) kinetics. The optimum probe concentration, washing conditions, and sort gate definition for such a screening experiment can be determined in advance to maximize the probability that clones with the desired properties will be isolated. If necessary, the sorting parameters can be fine-tuned in real time during sorting.

Wittrup and coworkers have developed a simple mathematical analysis for delineating the important parameters for the isolation of ligand binding proteins from cell surface-displayed libraries (Boder and Wittrup, 1997a). The equilibrium dissociation constant (K_D) of the interaction of a surface-displayed protein with a fluorescent ligand is defined as the ratio of the dissociation and association rate constants (k_{diss}/k_{assoc}). For K_D values less than approximately 10 nM, target cells can be sorted based on the dissociation rate of a fluorescent target ligand from the cell surface. Interactions with K_D values greater than 10 nM have relatively fast dissociation rates ($k_{diss} \geq 0.01\ s^{-1}$). As a result, half of the total bound ligand will dissociate from the cell surface in a few minutes or less at 25°C. In this case, unless stopped flow cytometric techniques are used (Nolan and Sklar, 1998), the library can be screened only under conditions where the fluorescently-labeled ligand is in equilibrium with the cells. For equilibrium conditions, cells are incubated with the fluorescent probe and analyzed directly by FCM without removing the unbound probe. Under equilibrium conditions, only moderate ($K_D < 10^{-6}$) to very high affinity interactions are detectable due to high levels of background fluorescence.

In general, the affinity, or equilibrium dissociation constant (K_D), correlates directly with the dissociation rate constant k_{diss} (Schier and Marks, 1996), since association rate constants, k_{assoc}, are similar for related antibodies. To screen for clones having decreased dissociation rate constants, the library population is incubated with an excess of fluorescent ligand to saturate all ligand-binding sites. The cells are then washed to remove excess fluorescent ligand and resuspended in buffer containing a 100–1000-fold excess of a non-fluorescent ligand as a competitor. In the presence of a large excess of competitor, the re-association of the fluorescent ligand to a receptor is negligible. Under these conditions, the dissociation rate constant (k_{diss}) governs the mean fluorescence of a given cell (F) as a function of time (t):

$$F = (F_0 - F_\infty)\exp(-k_{diss}t) + F_\infty, \quad (1)$$

where F_0 and F_∞ are the maximum initial fluorescence after labeling and the fluorescence after complete dissociation of the fluorescent ligand. The experimentally determined time course of fluorescence decay for a clonal population of scFv-displaying cells labeled with a hapten-BODIPY conjugate was found to agree well with that predicted by Eq. (1) (PD, BI, GG, unpublished data). However, for $t \geq 1$ h, the observed fluorescence is higher than that predicted by (1), most likely due to rebinding effects. In other words, the probability that free fluorescent ligand will rebind to a protein on the cell surface increases as the fraction of unoccupied binding sites increases. This effect is particularly significant when the density of binding sites is high.

Proteins with improved affinity for a ligand are often obtained by random mutagenesis of the encoding gene and subsequent screening of the resulting library of protein mutants. Typically, the improvement in ligand affinity that can be obtained in such an experiment is between two- and five-fold. Such gain of function protein mutants occur at low frequencies (10^{-3}–10^{-7}) in random mutant libraries. These rare cells must be sorted from the remainder of the population that includes a large fraction of cells expressing proteins with wild-type ligand binding affinity. Thus, the rare target cells must be sorted from an excess of non-target, wild-type cells that have a fluorescent signal only slightly lower (Fig. 4). The isolation of target cells displaying protein mu-

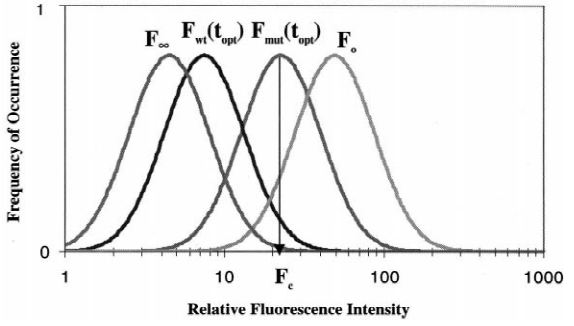

Fig. 4. The predicted fluorescence distributions for the wild-type and a three-fold improved clone, assuming a normal distribution. The fluorescence distributions prior to (F_0) and after complete dissociation (F_∞) are indicated. After the optimal time for dissociation, t_{opt}, the expected wild-type and mutant fluorescence distributions are given by $F_{wt}(t_{opt})$ and $F_{mut}(t_{opt})$, respectively. The probability that a wild-type clone (P_{wt}) or three-fold improved mutant (P_{mut}) will occur within a sort gate with a fluorescence cut-off of F_c is given by the area under the respective curves $F_{wt}(t_{opt})$ and $F_{mut}(t_{opt})$, for $F > F_c$. The expected enrichment of the mutant is then given by the ratio of P_{mut} to P_{wt}.

tants with improved affinity is dependent upon maximizing the ratio of mean fluorescence intensity of the mutant (F_{mut}) to that of the wild-type (F_{wt}):

$$F_r = \frac{F_{mut}}{F_{wt}} = \frac{(S_r - 1)\exp[-k_{diss,wt}t/k_r)]}{(S_r - 1)\exp(-k_{diss,wt}t)}, \quad (2)$$

where $S_r = F_0/F_\infty$, and a dimensionless affinity improvement (k_r) is defined:

$$k_r = \frac{k_{diss,wt}}{k_{diss,mut}}.$$

Differentiating Eq. (2) with respect to $k_{diss,wt}t$ yields the optimal dimensionless time for the dissociation reaction for the isolation of clones improved by a factor k_r (Boder and Wittrup, 1997a). The optimal time (t_{opt}) for the dissociation reaction can be predicted as

$$k_{off,wt}t_{opt} = 0.293 + 2.05\log k_r + \left(2.30 - 0.759\frac{1}{k_r}\right)\log S_r. \quad (3)$$

When the desired proteins are expected to have an equilibrium dissociation constant greater than approximately 10 nM, equilibrium labeling conditions are required to isolate the corresponding target cell. Otherwise, rapid dissociation rates will cause the cellular fluorescence to decay significantly before the entire library can be screened. The fluorescence of a cell displaying a polypeptide with an equilibrium dissociation constant K_D is given by

$$F = \frac{F_{max}([L]/K_D) + F_{bg}}{([L]/K_D) + 1}, \quad (4)$$

where F_{max} is the maximum fluorescent signal upon saturation, F_{bg} is the background autofluorescence and $[L]$ is the equilibrium ligand concentration.

It can be shown that the optimal ligand concentration to achieve the maximal fluorescence ratio between the desired affinity improved mutant and wild-type clones is given by the expression (Boder and Wittrup, 1997a)

$$\frac{[L]_{opt}}{K_{D,wt}} = \frac{1}{\sqrt{S_r(K_{D,wt}/K_{D,mut})}}. \quad (5)$$

For example, to isolate clones expressing proteins having a two- to three-fold greater ligand binding affinity than the wild-type (Daugherty et al., 1998; Low et al., 1996; Yang et al., 1995), the optimal ligand concentration given by Eq. (5) will be $0.18\cdot K_{D,wt}$ or $0.15\cdot K_{D,wt}$ for typical S_r values of 10 and 15, respectively.

3. Defining and sorting target cells

Cytometer calibration is essential for rare cell isolations. For optimal sorting performance, the cell sorting accuracy should be tested immediately prior to library sorting using a mixture of polystyrene fluorescent beads comparable in size and fluorescence intensity to the cells to be sorted. Sorting fluorescent beads in the size range of 1–2 μm in diameter with only slightly differing fluorescence intensities provides an effective evaluation of the fluorescence discrimination and sorting accuracy that may be expected. The beads are mixed at non-target/target ratio of between 1:1 and 10:1 and sorted at a rate comparable to the desired experimental sort rate. If greater ratios of non-target to target beads are used, the bead yield in the sorted sample will be too low for accurate calibration of the sorting accuracy.

After sorting, the target beads are concentrated by centrifugation. The supernatent is then aspirated and the beads are resuspended in about 1 ml of buffer. After carefully cleaning the sample injection port, to prevent carryover of the pre-sort population, the sorted beads are analyzed by FCM for purity. Typically, a sort purity of greater than 98% should be obtainable using a high purity sorting mode.

Before use, the instrument must be sterilized with 70% ethanol. The instrument is usually rinsed with 70% ethanol, bypassing the sheath filter, for about 30 min and subsequently with sterile water or buffered sheath fluid compatible with the cells to be sorted. A rinse period of 30 min or longer may be required to remove all remaining ethanol. Cleaning and sterilization with a 10% bleach solution is more likely to completely inactivate any bacteria adhering to the tubing. However, bleach can degrade some polymeric tubing and should not be used without first consulting the instrument manufacturer. For micron-sized particles with relatively low fluorescence signals, instrument noise and debris can give rise to ghost events and should be measured before sorting. After the instrument has been rinsed thoroughly, an appropriate sort gate is defined to encompass the parameter space to be used in sorting. The instrument is rinsed further with sterile buffer or water until the frequency of events falling into the defined parameter space, resulting from air bubbles and debris, reaches a minimum.

In cell sorting applications, a sort gate specifies a region of parameter space encompassing the target cells. The definition of a sort gate is particularly important when the target cell fluorescence distribution overlaps, even to a small degree, with the fluorescence distribution of non-target cells. In this situation, a very stringent gate will prevent the recovery of some rare target cells, while a relaxed gate will provide little or no enrichment. Thus, it is worthwhile to estimate the percentage of the population that should be recovered to maximize the probability that a particular rare cell will be collected. The probability that a rare target cell will be recovered (P_{rec}) during the first round of screening can be estimated by

$$P_{\mathrm{rec}} = P_{\mathrm{encounter}} \cdot P_{\mathrm{detection}} \cdot P_{\mathrm{gate}} \cdot P_{\mathrm{sort}} \cdot P_{\mathrm{viable}}, \qquad (6)$$

where $P_{\mathrm{encounter}}$ is the probability that the target cell will be in the library, $P_{\mathrm{detection}}$ the probability that the target cell will be detected by the instrument, P_{gate} the probability that the target cell will fall into the sort gate, P_{sort} the probability that the target cell will be captured and P_{viable} the probability that the target cell will grow in liquid culture after sorting

The probability of encountering a given clone is given by the Poisson distribution (Lowman et al., 1991). Sorting three-fold more cells than the expected diversity of the library, i.e. the number of independent transformants, is sufficient to give a $P_{\mathrm{encounter}}$ of 0.95. The probability of detection, $P_{\mathrm{detection}}$, is dependent upon the level of instrument noise and the electronic dead-time (Rosenblatt et al., 1997; Shapiro, 1995). For example, at sorting rates below 2000 s^{-1} with the FACSort™ or FACSCaliber™ from Becton Dickinson (San Jose, CA, USA) the probability of detection is $>90\%$ (Shapiro, 1995), and can be assumed to be unity. The probability of physically capturing a cell, P_{sort}, is dependent upon the particular sorting mode employed (i.e. Recovery/Enrichment, Non-target cell Exclusion, or Purity/Single-cell). In Recovery/Enrichment mode, P_{sort} is unity since all target clones are collected regardless of the presence of coincident non-target events. In general, it is desirable to maintain the viability of the culture above 50% ($P_{\mathrm{viable}} > 0.5$). When culture viability is low, the capture of non-viable cells in Recovery/Enrichment mode at high sort rates results in the coincident collection of viable non-target cells. As a result, when P_{viable} is low, the enrichment of target cells is lower than would be expected from the fluorescence distributions of the target and non-target cells. P_{viable} can be estimated by measuring the viability of the entire library population by counting colony forming units. Alternatively, viability can be conveniently estimated by FCM using a suitable fluorescent probe. The fluorescent probe Di(BAC_4)3 provides a reasonable estimate of cellular viability for many strains of *E. coli* (Davey and Kell, 1996). Unfortunately, this dye has a broad emission spectrum and cannot be used simultaneously with most other fluorescent probes in a multi-color sorting experiment. Propidium iodide (PI) is another commonly used viability probe. Although PI is compatible with many commonly used probes in multi-color FCM, PI does not normally penetrate the outer membrane of Gram-

negative bacteria and thus provides only a relative viability measure for some strains of *E. coli*. Nonetheless, as described above, exclusion of PI-labeled non-viable cells from the sort gate can be useful for preventing the collection of non-target cells to improve the enrichment in Recovery/Enrichment mode.

The probability that a particular target cell will fall within a specified sort gate (P_{gate}) is dependent upon the region of parameter space defined by the sort gate. The sorting parameters typically include:

1. Forward and 90° light scatter, to ensure proper cell morphology.
2. Fluorescence due to the target cell function of interest.
3. Fluorescence due to a viability indicator stain.
4. Fluorescence due to non-specific staining control probes.

If the sort gate encompasses a large range of forward and 90° light-scatter values, the deviation in fluorescence intensity becomes the primary factor in determining the probability that a given clone will fall into the sort gate. In this case, the sort gate can be approximated as a fluorescence cut-off (F_c) (Fig. 4). Assuming a normal distribution for the fluorescence of a monoclonal cell population, the fluorescence distribution of the wild-type and three-fold affinity improved cell populations will have the appearance represented in Fig. 4, after allowing the dissociation reaction to occur for the optimal time predicted by Eq. (3).

To a first approximation, the maximum possible enrichment in a single round of sorting can be calculated by taking the ratio of the area under the mutant ($F_{mut}(t_{opt})$) curve for $F > F_c$ to that under the wild-type ($F_{wt}(t_{opt})$) curve for $F > F_c$ (Fig. 4). In a library containing 10^6 unique individuals, a sufficient number of cells can be sorted such that four or more copies of each individual are encountered. Under these conditions, a sort gate that includes 50% of the fluorescence distribution corresponding to an improved mutant will give a 95% probability of recovering at least one copy of each improved mutant (Boder and Wittrup, 1997a). If the standard deviation of the fluorescence distribution can be estimated, the maximum possible enrichment in a single round of sorting can be predicted by integrating the fluorescence distributions for the mutant and wild-type populations (Fig. 4). For bacterial and yeast display systems, with $S_r = 10$, the maximum enrichment of clones having three-fold improvement in the first round of sorting will be approximately 60-fold, if only the most fluorescent 1% of the population is collected.

In practice, the actual enrichment ratio is reduced by the inadvertent capture of coincident non-target cells during Recovery/Enrichment mode sorting. For libraries containing more than 10^5 unique members, recovery mode sorting must be used during the first round of screening. As discussed above, at least three-fold more cells than the estimated number of unique clones should be processed in the initial round of screening. Since each clone may be encountered only three times during the entire sort process, it is critical to sort the library first in Recovery/Enrichment mode to ensure that at least one viable representative of every desirable clone is collected. While sorting at relatively high rates, non-target cells are co-captured with target cells. When the sorted population is re-grown and sorted again in the second round, the enriched target cells are present at higher frequencies. The sort mode then can be adjusted to Purity/Exclusion to obtain improved enrichment ratios. Under optimal sorting conditions, each target cell initially at a frequency of 10^{-6} will comprise at least 2% of the population after the second round of sorting. Additional rounds of sorting can be performed in exclusion mode, which rejects coincident cells, to achieve high purity. After three or four rounds of sorting, it should be possible to enrich target clones, having even subtly improved function, to greater than 95% purity (Daugherty et al., 1998).

Phosphate buffered saline of physiological pH can be used as a general purpose sorting buffer and sample resuspension fluid. Since sheath fluid composition and pH can affect cell size, viability, and probe fluorescence, compatibility should be verified experimentally. Even with compatible sheath fluids, the stress arising from heterologous protein expression, labeling, and sorting may reduce viability. In this case, sorting directly into 50 ml conical tubes containing 15 ml of growth medium may improve viability (Shapiro, 1995). If cell viability is extreme-

ly low after sorting, the genes of interest can be amplified from the sorted cell population by PCR and reinserted into the expression vector for an additional round of screening (Wentzel et al., 1999). Sorted cells can be isolated directly by filtration onto suitable membranes immediately after sorting and subsequently incubated on solid growth medium to allow colony formation.

4. Controlling and accounting for protein expression levels

Rare target cells expressing proteins exhibiting a desired trait such as ligand binding affinity or catalytic activity, occasionally possess secondary phenotypes that adversely affect their enrichment by FCM. For example, mutations leading to improved binding or catalysis may also reduce protein expression, host cell growth rate, or viability. At the same time, non-target cells may exhibit elevated growth rates and become over-represented in the sorted population despite the fact that they are not specifically enriched by the screening process (Fig. 5) (Boder and Wittrup, 1997a; Daugherty et al., 1999). Therefore, the ability to mask the mutant phenotype during cell growth, through tight regulation of protein expression within the host cell, is an essential feature in cell-based libraries.

The issue of how expression of a certain protein affects library screening has only been investigated in bacteria, though the insights from these experiments should be applicable to other systems. The ideal expression system for library screening purposes should possess the following characteristics:

1. Tight on/off regulation. Preferably, protein expression should be induced only prior to sorting and repressed immediately after. Protein expression must be strongly repressed when the cells are grown for many generations to maintain proportionate representation of target cells, despite a possible growth advantage for some non-target cells.
2. Moderate protein expression under conditions of full induction. Otherwise, low level expression can result in weak fluorescence signals whereas high level expression can result in cell death.

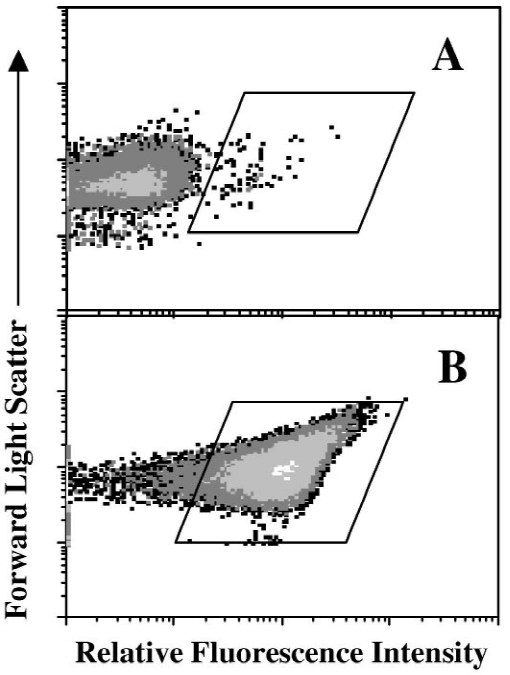

Fig. 5. The effects of (A) constitutive or (B) tightly-regulated protein expression upon the maintenance of library diversity. (A) A library of scFv mutants with a low mutation rate was grown with constitutive scFv expression from the hybrid *lpp-lac* promoter (Daugherty et al., 1999). After overnight culture, an unexpectedly low frequency of target cells was observed. (B) However, when the same library was grown under tightly repressed conditions using the arabinose inducible promoter P_{BAD}, and induced shortly before FCM analysis, the frequency of target cells remained high.

3. Rapid inducible protein expression. In other words, a desired level of protein synthesis should be obtainable shortly after induction. Rapid induction is desirable for two reasons. First, a short induction period reduces the time required for each round of enrichment. Second, because the cells are kept in an induced state for a shorter period, selection of faster growing non-target clones is minimized.
4. The ability to fine-tune expression level by adjusting the inducer concentration. Typically, a clonal cell population induced with a concentration lower than that required for full induction gives rise to two sub-populations, one consisting of fully induced cells and the other consisting of uninduced cells. The appearance of two popula-

tions for each clone presents a significant complication in library screening.

In *E. coli*, the commonly employed *tac* and *trc* promoters yield high levels of transcription after induction with IPTG and may result in cell death. Several available promoters satisfy criteria (1)–(3), including the *ara*BAD and *tet*A promoters. The arabinose inducible promoter P_{BAD} possesses the properties of tight on/off regulation and a moderate strength compatible with library screening in bacteria (Daugherty et al., 1999, 2000). The tightly regulated *tet*A promoter/operator has also been used in the screening of *E. coli* surface-displayed libraries. However, the expression level obtained upon induction was found to result in a loss of culture viability (Christmann et al., 1999; Daugherty et al., 1999). Expression could be reduced to a suitable level by introducing an amber codon prior to the displayed fusion protein sequence and using an amber suppressor strain (Christmann et al., 1999). The expression level may also be controlled using plasmids maintained at a desired copy number within the cell (Daugherty et al., 1999). None of the expression systems that have been employed for library screening applications allow tunable control of expression level as a function of the inducer concentration (Daugherty et al., 1999; Siegele and Hu, 1997). Fortunately, new expression systems that allow tunable control of expression for the entire population are being developed (J. Kiesling, personal communication; Lutz and Bujard, 1997). In yeast, the GAL1 promoter has been shown to enable tight repression in the presence of glucose and has been used successfully for screening surface-displayed libraries (Boder and Wittrup, 1997b; Kieke et al., 1999). However, expression from the GAL1 promoter was found to result in mixed populations of uninduced and fully induced cells (Boder and Wittrup, 1997b).

Generally, the amount of protein accumulating within unique clones in a library will vary widely. Even when protein synthesis is tightly regulated, mutated proteins that exhibit altered folding, aggregation, or stability characteristics will arise. As a result, there will be some cells in a library population in which protein accumulation is substantially increased. This fact was illustrated in the isolation of 'FACS optimized' GFP mutants (Cormack et al., 1996). After mutagenesis and screening by FACS, bacteria expressing GFP mutants that fluoresced more strongly than cells expressing wild-type GFP were isolated. However, the isolated GFP mutants were of two general classes: (1) mutants that exhibited dramatically improved specific fluorescence and unchanged expression level and (2) mutants that exhibited moderately improved specific fluorescence and increased solubility, or decreased aggregation (Cormack et al., 1996). Mutants with improved expression are desirable in many applications (Knappik and Pluckthun, 1995; Stemmer et al., 1993). More often, improvement of a trait other than expression is desired, and thus it may be beneficial to disfavor the isolation of mutants with altered expression properties.

An effective method to disfavor the isolation of expression mutants is to normalize the expression of the gene of interest using a peptide epitope tag (Boder and Wittrup, 1997b). In this approach, a secondary peptide epitope that can be recognized by a fluorescently-tagged monoclonal antibody is appended to the *C*-terminus of the expressed protein using an appropriate gene fusion. After induction of protein expression, the cells are labeled with a fluorescently-tagged antibody specific to the *C*-terminal epitope tag. Cells with altered expression levels will also have altered epitope labeling. A suitable sorting gate can then be used to disfavor the selection of expression mutants.

In the absence of an epitope tag for detecting protein expression levels, the selection of expression mutants can be disfavored using appropriate gating and sorting strategies. It might be assumed that a gate including only the most fluorescent events will capture clones having improved function. However, such a gate definition is also likely to favor the isolation of mutants having improved expression properties. Expression mutants generally arise more frequently than true gain-of-function mutants and therefore are more likely to predominate when a very stringent gate is defined. Instead, a sort gate should be set to include the region of parameter space in which $>50\%$ of target events are expected to occur, despite the inclusion of some non-target cells. Though enrichment factors will be reduced, a less stringent gate will disfavor expression mutants and

increase the probability of recovering rare target clones.

5. Conclusions

FCM is the most powerful and quantitative high-throughput technology currently available for the screening of protein libraries for desirable characteristics such as high affinity binding, catalysis, expression level, and in vivo stability. Significant advantages of FCM relative to alternative methods for library screening include a quantitative multi-parameter analysis of every library member and the ability to carry out real-time statistical analysis of heterogeneous populations. Importantly, the screening process can be monitored and optimized, an important feature that enabled the evolution of the first antibody reported to have sub-picomolar affinity (Shusta et al., 1999b). Cell-based libraries also provide a new method for the identification of in vivo protein–protein interactions and effectors of biological pathways (Cho et al., 1998), and thus may be especially useful in metabolic engineering.

While cell-based libraries hold great potential for the isolation and production of proteins with novel functions for scientific and commercial applications, some limitations should be mentioned. First, these approaches are limited to proteins than can be expressed at high levels in a functional form within or on the surface of the chosen host cell. For example, proteins requiring extensive disulphide bond formation or post-translational modifications like glycosylation may not be functional when expressed in bacteria. Proteins exhibiting even low level host toxicity may be difficult to isolate from a cell-based library. Thus, the selection of a library host cell should be made with the characteristics of the desired protein in mind. Second, given that cytometer throughput rates are currently limited to about 10^8 cells/h, exhaustive screening of large libraries may not be possible. However, as stated earlier, given the vast size of protein sequence space (i.e. $\approx 20^{300}$), library design strategy is generally more important than size.

In order to exploit fully the power of FCM, however, all the parameters must be considered and optimized. As discussed above, the library size and expected frequency of target cells will determine an appropriate sort rate and mode. However, the observed false-positive frequency will directly limit the maximum possible enrichment in a single round of sorting. The signal-to-noise ratio(s) for each of the parameters used for selection must also be maximized by adjusting protein expression levels or by using an alternative fluorescent probe or signal amplification strategy. Cytometer design may also be important in some applications that require high fluorescence sensitivity, high-throughput rates, or sophisticated gating algorithms. At the biological level, protein expression level should be tightly regulated to prevent clonal competition effects. If possible, expression level should be measured independently of function, allowing improvements in a primary trait to be discriminated from that of a secondary trait.

References

Boder, E.T., Wittrup, K.D., 1997a. Optimal screening of surface-displayed polypeptide libraries. Biotechnol. Prog. 14, 55.

Boder, E.T., Wittrup, K.D., 1997b. Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 15, 553.

Bogarad, L.D., Deem, M.W., 1999. A hierarchical approach to protein molecular evolution. Proc. Natl. Acad. Sci. USA 96, 2591.

Chang, H.J., Sheu, S.Y., Lo, S.J., 1999. Expression of foreign antigens on the surface of *Escherichia coli* by fusion to the outer membrane protein traT. J. Biomed. Sci. 6, 64.

Cho, B.K., Kieke, M.C., Boder, E.T., Wittrup, K.D., Kranz, D.M., 1998. A yeast surface display system for the discovery of ligands that trigger cell activation. J. Immunol. Methods 220, 179.

Christmann, A., Walter, K., Wentzel, A., Kratzner, R., Kolmar, H., 1999. The cystine knot of a squash-type protease inhibitor as a structural scaffold for *Escherichia coli* cell surface display of conformationally constrained peptides. Protein Eng. 12, 797.

Cormack, B.P., Valdivia, R.H., Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173, 33.

Crameri, A., Raillard, S.-A., Bermudez, E., Stemmer, W.P.C., 1998. DNA shuffling of family of genes from diverse species accelerates directed evolution. Nature 391, 288.

Daugherty, P.S., Chen, G., Iverson, B.L., Georgiou, G., 2000. Quantitative analysis of the effect of the mutation frequency on the affinity maturation of scFv antibodies. Proc. Natl. Acad. Sci. USA 97, 2029.

Daugherty, P.S., Chen, G., Olsen, M.J., Iverson, B.L., Georgiou, G., 1998. Antibody affinity maturation using bacterial surface display. Protein Eng. 11, 101.

Daugherty, P.S., Olsen, M.J., Iverson, B.L., Georgiou, G., 1999. Development of an optimized expression system for the screening of antibody libraries displayed on the *E. coli* surface. Protein Eng. 12, 613.

Davey, H.M., Kell, D.B., 1996. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. Microbiol. Rev. 60, 641.

Douglas, R.H., Ballou, C.E., 1980. Isolation of *Kluyveromyces lactis* mannoprotein mutants by fluorescence-activated cell sorting. J. Biol. Chem. 255, 5979.

Dunn, A.K., Handelsman, J., 1999. A vector for promoter trapping in *Bacillus cereus*. Gene 226, 297.

Ernst, W., Grabherr, R., Wegner, D., Borth, N., Grassauer, A., Katinger, H., 1998. Baculovirus surface display: construction and screening of a eukaryotic epitope library. Nucleic Acids Res. 26, 1718.

Fischetti, V.A., Medaglini, D.A., Pozzi, G., 1996. Gram-positive commensal bacteria for mucosal vaccine delivery. Curr. Opin. Biotechnol. 7, 659.

Francisco, J.A., Campbell, R., Iverson, B.L., Georgiou, G., 1993. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. Proc. Natl. Acad. Sci. USA 90, 10444.

Francisco, J.A., Earhart, C.F., Georgiou, G., 1992. Transport and anchoring of beta-lactamase to the external surface of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 89, 2713.

Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., Curtiss, R.r., 1997. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nat. Biotechnol. 15, 29.

Griffin, B.A., Adams, S.R., Tsien, R.Y., 1998. Specific covalent labeling of recombinant protein molecules inside live cells. Science 281, 269.

Gross, H.J., Verwer, B., Houck, D., Hoffman, R.A., Recktenwald, D., 1995. Model study detecting breast cancer cells in peripheral blood mononuclear cells at frequencies as low as 10(−7). Proc. Natl. Acad. Sci. USA 92, 537.

Gross, H.J., Verwer, B., Houck, D., Recktenwald, D., 1993. Detection of rare cell at a frequency of one per million by flow cytometry. Cytometry 14, 519.

Gunneriusson, E., Samuelson, P., Uhlen, M., Nygren, P.A., Stahl, S., 1996. Surface display of a functional single-chain Fv antibody on *Staphylococci*. J. Bacteriol. 178, 1341.

Jung, H.C., Lebeault, J.M., Pan, J.G., 1999. Surface display of *Zymomonas mobilis* levansucrase by using the ice-nucleation protein of *Pseudomonas syringae*. Nat. Biotechnol. 16, 576.

Kieke, M.C., Shusta, E.V., Boder, E.T., Teyton, L., Wittrup, K.D., Kranz, D.M., 1999. Selection of functional T cell receptor mutants from a yeast surface-display library. Proc. Natl. Acad. Sci. USA 96, 5651.

Klauser, T., Pohlner, J., Meyer, T.F., 1993. The secretion pathway of IgA protease-type proteins in gram-negative bacteria. Bioessays 15, 799.

Knappik, A., Pluckthun, A., 1995. Engineered turns of a recombinant antibody improve its in vivo folding. Protein Eng. 8, 81.

Leary, J.F., 1994. Strategies for rare cell detection and isolation. Methods Cell Biol. 42, 331.

Low, N.M., Holliger, P.H., Winter, G., 1996. Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. J. Mol. Biol. 260, 359.

Lowman, H.B., Bass, S.H., Simpson, N., Wells, J.A., 1991. Selecting high-affinity binding proteins by monovalent phage display. Biochemistry 30, 10832.

Lutz, R., Bujard, H., 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, and TetR/O and AraC/I_1-I_2 regulatory elements. Nucleic Acids Res. 25, 1203.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., Lukyanov, S.A., 1999. Fluorescent proteins from nonbioluminescent *Anthozoa* species. Nat. Biotechnol. 17, 969.

Minas, W., Sahar, E., Gutnick, D., 1988. Flow cytometric screening and isolation of *Escherichia coli* clones which express surface antigens of the oil-degrading microorganism *Acinetobacter calcoaceticus* RAG-1. Arch. Microbiol. 150, 432.

Needels, M.C., Jones, D.G., Tate, E.H., Heinkel, G.L., Kochersperger, L.M., Dower, W.J., Barrett, R.W., Gallop, M.A., 1993. Generation and screening of an oligonucleotide-encoded synthetic peptide library. Proc. Natl. Acad. Sci. USA 90, 10700.

Nolan, J.P., Sklar, L.A., 1998. The emergence of flow cytometry for sensitive, real-time measurements of molecular interactions. Nat. Biotechnol. 16, 633.

Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P.S., Georgiou, G., Iverson, B.L., 2000. Isolation of novel enzymes from large libraries (submitted for publication).

Radbruch, A., Recktenwald, D., 1995. Detection and isolation of rare cells. Curr. Opin. Immunol. 7, 270.

Rosenblatt, J.I., Hokanson, J.A., McLaughlin, S.R., Leary, J.F., 1997. Theoretical basis for sampling statistics useful for detecting and isolating rare cells using flow cytometry and cell sorting. Cytometry 27, 233.

Schier, R., Marks, J.D., 1996. Efficient in vitro affinity maturation of phage antibodies using BIAcore guided selections. Hum. Antibodies Hybridomas 7, 97.

Shapiro, H.M., 1995. Practical Flow Cytometry, 3rd Edition. Wiley-Liss, New York.

Shusta, E.V., Kieke, M.C.E.P., Kranz, D.M., Wittrup, K.D., 1999a. Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. J. Mol. Biol. 292, 949.

Shusta, E.V., VanAntwerp, J., Witttrup, K.D., 1999b. Biosynthetic polypeptide libraries. Curr. Opin. Biotechnol. 10, 117.

Siegele, D.A., Hu, J.C., 1997. Gene expression from plasmids containing the *araBAD* promoter at subsaturating inducer concentrations represents mixed populations. Proc. Natl. Acad. Sci. USA 94, 8168.

Silhavy, T.J., Beckwith, J.R., 1985. Uses of *lac* fusions for the study of biological problems. Microbiol. Rev. 49, 398.

Stahl, S., Uhlen, M., 1997. Bacterial surface display: trends and progress. Trends Biotechnol. 15, 185.

Stemmer, W.P., Morris, S.K., Kautzer, C.R., Wilson, B.S., 1993. Increased antibody expression from *Escherichia coli* through

wobble-base library mutagenesis by enzymatic inverse PCR. Gene 123, 1.

Tsien, R.Y., 1998. The green fluorescent protein. Annu. Rev. Biochem. 67, 509.

Valdivia, R.H., Falkow, S., 1996. Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. Mol. Microbiol. 22, 367.

Valdivia, R.H., Falkow, S., 1997. Fluorescence-based isolation of bacterial genes expressed within host cells. Science 277, 2007.

Vazquez-Torres, A., Jones-Carson, J., Baumler, A.J., Falkow, S., Valdivia, R., Brown, W., Le, M., Berggren, R., Parks, W.T., Fang, F.C., 1999. Extraintestinal dissemination of *Salmonella* by CD18-expressing phagocytes. Nature 401, 804.

Wells, J.A., Vasser, M., Powers, D.B., 1985. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. Gene 34, 315.

Wentzel, A., Christmann, A., Kratzner, R., Kolmar, H., 1999. Sequence requirements of the GPNG beta-turn of the *Ecballium elaterium* trypsin inhibitor II explored by combinatorial library screening. J. Biol. Chem. 274, 21037.

Wipf, P., Cunningham, A., Rice, R.L., Lazo, J.S., 1997. Combinatorial synthesis and biological evaluation of library of small-molecule Ser/Thr-protein phosphatase inhibitors. Bioorg. Med. Chem. 5, 165.

Yang, W.P., Green, K., Pinz-Sweeney, S., Briones, A.T., Burton, D.R., Barbas, C.F.r., 1995. CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. J. Mol. Biol. 254, 392.