Methods 60 (2013) 91–98

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Arrayed antibody library technology for therapeutic biologic discovery

Cornelia A. Bentley [a], Omar A. Bazirgan [a], James J. Graziano [a], Evan M. Holmes [a], Vaughn V. Smider [a,b,*]

[a] Fabrus Inc., La Jolla, CA, United States
[b] Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, United States

ARTICLE INFO

Article history:
Available online 26 February 2013

Keywords:
Arrayed protein library
Spatially addressed library
Antibody
Protein engineering
Screening high-throughput flow cytometry

ABSTRACT

Traditional immunization and display antibody discovery methods rely on competitive selection amongst a pool of antibodies to identify a lead. While this approach has led to many successful therapeutic antibodies, targets have been limited to proteins which are easily purified. In addition, selection driven discovery has produced a narrow range of antibody functionalities focused on high affinity antagonism. We review the current progress in developing arrayed protein libraries for screening-based, rather than selection-based, discovery. These single molecule per microtiter well libraries have been screened in multiplex formats against both purified antigens and directly against targets expressed on the cell surface. This facilitates the discovery of antibodies against therapeutically interesting targets (GPCRs, ion channels, and other multispanning membrane proteins) and epitopes that have been considered poorly accessible to conventional discovery methods.

1. Introduction

The first monoclonal antibody, OKT3 [1], was approved for human clinical use in 1985. It was derived from a mouse hybridoma and in some patients produced immune reactions that limited its use, hindering further development of monoclonal antibodies. Humanization technology reduced these cross-species immune reactions and resulted in the approval of the first recombinant "chimeric" antibody rituximab in 1997 and the first "humanized" antibody trastuzumab in 1998. The approval of these molecules opened the floodgates for antibody development, as over 30 antibodies are currently approved by the FDA for therapeutic use and hundreds are in clinical development [2]. The need for humanized recombinant antibodies spurred the development of several technologies for their discovery and engineering.

1.1. Antibody discovery and optimization technologies: hybridoma and display

Kohler and Milstein invented the hybridoma technique in 1975 [3]. Since antibodies are made from multiple gene segments that rearrange at the DNA level during B-cell development (Fig. 1), cloning an individual antibody gene was not easily accomplished. However, genes could be cloned through cDNA library techniques developed in the early 1980s [4,5]. The polymerase chain reaction was developed in 1985, and application of this method to hybridoma-derived antibody genes was accomplished in 1989 using many primers covering the V-region repertoire [6,7]. Thus, hybridoma methods could be combined with rapid molecular biology techniques by the 1990s. Since mouse variable regions are highly similar to related human variants, the first generation "chimeric" antibodies involved replacing the mouse constant regions with their human counterparts for both the heavy and light chains [8–10]. Shortly thereafter, molecular and homology modeling allowed replacement of murine V-region sequences with human frameworks, leaving just CDR sequences derived from mouse [11]. Efforts in the 1990s generated transgenic mice that contained the human immunoglobulin loci, such that human antibodies could be induced directly by immunization [12,13].

In 1988 phage display library technology was developed, allowing very large libraries to be constructed in a single vessel where the genotype (gene) for each particle was linked to the target-binding phenotype (protein) [14]. These methods were applied to antibody gene fragments [15,16]. There are now multiple formats for display technology employing mRNA [17], ribosome [18], phage, eukaryotic viruses [19], yeast [20], bacteria [21], or mammalian cells [22–24]. The number of unique variants a given library technology can support varies with the size of the particle: libraries of 10^6 are feasible with mammalian cell display, while mRNA or ribosome display libraries of 10^{14} can be constructed and selected. For particle based discovery using phage, mRNA, or ribosomes, the

Abbreviations: CDR, complementary determining region; mAb, monoclonal antibody; V, variable; D, diversity; J, joining; GPCR, G-protein coupled receptor; CVB, culture vessel block; LIMS, laboratory information management system; Fab, fragment antigen-binding; SAR, structure activity relationship; IMAC, immobilized metal chelate chromatography; FACS, fluorescence activated cell sorting; BLAST, basic local alignment search tool; MHC, major histocompatibility complex.

* Corresponding author at: Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA, United States.
E-mail address: vvsmider@scripps.edu (V.V. Smider).

1046-2023/$ - see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ymeth.2013.02.003

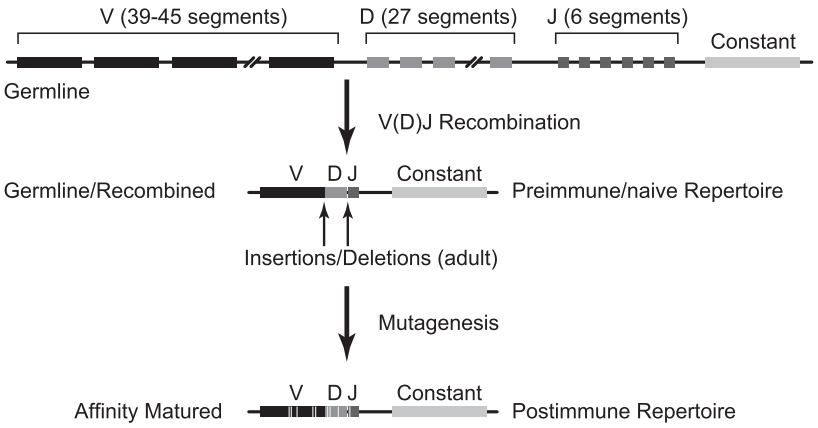

Fig. 1. Genetic mechanisms for generating antibody diversity. In human pre-B cells 39–45 V, 27 D, and 6 J segments can recombine at the DNA level to produce a functional V-D-J variable region at the heavy chain locus. A similar V-J process occurs for the kappa or lambda light chains (where there are no D regions). During recombination, nucleotides can be inserted or deleted at the V-D-J junctions producing greater diversity in the CDR3 region. These processes produce the preimmune repertoire. Following exposure to antigen, somatic hypermutation generates base pair changes throughout the V_H (variable heavy chain) or V_L (variable light chain), allowing enhanced affinity to be selected for during the immune response.

library is typically allowed to contact an adsorbed or biotinylated antigen, several washing steps are performed, and the particles that bind the antigen are eluted. Multiple rounds of this process with subsequent growth and amplification are used to enrich the binding clones. For cell-based display libraries, fluorescence activated cell sorting (FACS) is used with fluorescently labeled antigen to select binders from the library population.

1.2. Selection versus screening

In hybridoma and display technologies, the discovery paradigm is selection; a *pool* of molecules provides the starting material and individual members are selected from the pool based on their affinity to the target. The *selection* process involves *competition* between individual library members for binding to the target antigen. For display libraries, members with lower K_d will have more members bound at equilibrium than those with a higher K_d. Antibodies with slower off-rates will survive washing steps and subsequent incubations. Moreover, antibodies may exist in the library that bind to different but overlapping epitopes on the antigen. Because they are in competition with one another, a low K_d antibody may displace other antibodies, preventing their selection. Additionally, differences in the propagation of the particles post-selection, such as variations in the growth rate of the cells expressing individual library members, can affect the population selected from the library. For immunization, those B-cells with higher affinity outcompete those with lower affinity via signaling through the B-cell receptor (BCR) and expansion of the higher affinity B-cells. In each case, the discovery mechanism involves competitive selection.

Array-based libraries use a *screening* paradigm. Screening evaluates each library member independently. This is accomplished by applying an arrayed library in microtiter plates to a binding or functional assay. The ability to evaluate each member of the library independently facilitates the discovery of less dominant epitopes, potentially more diverse mechanisms of action, structure activity relationships for rational lead development, and a more diverse set of targets not easily addressable by traditional antibody selection technologies. However, an arrayed protein library is substantially more expensive and has more limited diversity than a pooled library. Affinity-based selection approaches have been very successful in creating high affinity antibodies against certain classes of extracellular targets. A screening approach becomes justified to address high value targets or epitopes that do not lend themselves to selection, such as multispanning membrane proteins. The following sections evaluate the potential of array-based antibody libraries to address some of the unmet needs in antibody discovery.

2. Unmet needs in antibody discovery technology

2.1. Membrane protein targets

Enormous progress has been achieved in the past 20 years in the antibody discovery and engineering field. Discovery using hybridoma or display-based techniques has become routine, and further engineering using display methods for affinity maturation or humanization are commonplace. However, all of these techniques have been employed mostly for soluble protein antigens. Entire classes of targets are not easily addressed by antibody discovery technology because they are not readily purified. For example, over 30% of the druggable genome is comprised of G-protein coupled receptors (GPCRs) or ion channels [25]. It is difficult to produce antibodies against GPCR targets because (i) they are not easily purified in active form, and (ii) they are often refractive to overexpression on the cell surface. While cell surface panning with complicated competitive selection schemes could theoretically overcome these issues, there have been few examples of phage display derived antibodies discovered against these difficult targets [26]. Enabling technologies that can address these receptors in the context of the cell would be very useful [26–28]. Using a screening approach, it becomes possible to identify lead antibodies directly on the cell surface as described below.

2.2. Non-protein targets

Current antibody technology is not easily amenable to discovery against many non-protein targets. Carbohydrates, and even certain glycoproteins, are important in the pathogenesis of many diseases. Immunization of these compounds is often not effective in illiciting IgG because peptide display in the MHC is required for T-cell dependent responses and class switching. Additionally, these targets are not easily selectable using display systems because of very low affinity interactions and the affinity threshold problem of phage display (described below). Addressed libraries can potentially be used to identify low affinity antibodies against non-protein targets using highly sensitive homogenous binding and detection methods such as electrochemiluminescence from Meso Scale Discovery (MSD, Gaithersburg, MD), or flow cytometry

of cells. Electrochemiluminescene has been successful employed to identify specific binders of >50 μM K_d [29].

2.3. Novel epitope discovery

Proteins specifically expressed in diseased cells are ideal targets for therapeutic antibodies. In contrast, some proteins may be present in both diseased and healthy cells, but have distinct epitopes exposed specifically in the disease state. Although antibody technology is particularly mature for some target epitopes, many others are elusive to standard techniques. It would be useful to identify antibodies that are specific for a disease-state epitope or an altered form of a receptor due to splice variation, post-translational modifications, changes to multiprotein complex assemblies, or "activated" forms of a receptor. Some unique disease state antibodies have been reported, but their discovery was not straightforward or routine. For example, the mAb EGFR806 binds the *activated* form of EGFR. It was raised by immunization to cells transfected with EGFRvIII, an activated form of EGFR with deletional mutation of exons 2–7, often found in glioblastoma and other cancers. EGFR806 reacts with both EGFRvIII as well as overexpressed, ligand-activated wtEGFR [30]. Such modified receptors may only be found on the surface of cells and not engineerable in the purified proteins used for immunization or display. Other unique antibodies have been found to be present naturally and in germline configuration, one inducing remyelination [31], and another having anaphylactogenic activity [32]. These antibodies would also be difficult to discover in the absence of purified target protein.

2.4. Mechanisms of action

Nearly all antibodies on the market are high affinity antagonists. Hybridoma and display selection techniques rely on affinity as the dominant selection parameter. Some evidence suggests antagonism and affinity are correlated. A low affinity antibody against the erythropoietin receptor has agonist activity but the affinity matured antibody is antagonistic [33]. In contrast to antibodies, small molecule drugs exhibit a variety of mechanisms of action including antagonism, agonism, partial agonism or antagonism, and allosteric modulation. There is no *a priori* reason why antibodies could not also comprise these activities. In small molecule discovery, cell-based screening of arrayed libraries has been an effective paradigm for achieving diversity of mechanisms from lead molecules. Cell-based screening of an arrayed antibody library also has the potential to discover novel mechanisms of action by allowing for selection of less dominant paratopes.

3. Arrayed protein libraries

Small molecule combinatorial chemistry libraries are arrayed or "spatially addressed" in that each individual compound is physically separated from every other compound, usually in microtiter wells. A database of information about each compound is maintained and can include biophysical properties, synthesis procedures, characterization, and cumulative data on screening results (Fig. 2). Such methodology has not been accomplished for antibodies for a variety of reasons. First, until recently the recombinant production of antibodies in a high-throughput parallel fashion was not technically feasible. This issue has been solved through the advent of technologies developed for structural genomics projects and parallel production of highly pure proteins for crystallography [34,35]. Second, the detailed understanding of antibody gene rearrangement and the complete sequences of the human antibody loci were only recently accomplished [36,37]. Such understanding is useful in the rational design of a biologic library in an arrayed format. Lastly, the cost of *de novo* DNA synthesis was prohibitive until recent technology improvements and competition amongst DNA suppliers drove the cost down to affordable levels.

3.1. Examples of arrayed protein libraries

With the human genome sequencing efforts complete, it is possible to bioinformatically predict the human "secretome" (all the proteins that are secreted) by the presence of protein sorting signals. Two groups, Five-Prime Therapeutics and The Genomics Institute of the Novartis Research Foundation (GNF), have both recently developed non-antibody arrayed protein libraries based on this principle, and have used them to discover novel targets or drug candidates in cell-based assays. Five-Prime's library reportedly contains over 4,500 cDNAs derived from multiple tissue sources [38]. This library was produced in transfected HEK293T cells, and conditioned supernatants were used to discover the novel cytokine IL-34 and its receptor in a cell based assay. GNF produced a collection of 806 proteins of mouse and human origin through transfection and purification via engineered Fc tag. The library was screened in a high-content assay for embryonic stem cell self-renewal and identified the pigment epithelium derived factor (PEDF) as a molecule containing the ability to promote pluripotency [39]. Both of these approaches illustrate the ability to identify protein "leads" in cell based screening programs.

Recently, Fabrus Inc. developed a methodology to produce human germline Fab antibody fragments [29] or full length IgG in a high-throughput fashion. In a parallel process, 576 (six 96-well plates) Fabs could be expressed in *Escherichia coli* then purified by IMAC and protein G chromatography. The resultant collection of over 10,000 Fabs was screened in a multiplexed manner on soluble antigens printed in an array at the bottom of microtiter plates. It was demonstrated that hits could be obtained to seven out of nine antigens. Hits to the cancer target DLL4 were further affinity matured using the same process; libraries were made with single point mutations and screened for enhanced binding activity on the microtiter plates. Iterative rounds of mutagenesis and screening for improved binding produced low nanomolar affinities through affinity maturation. Antibodies to unique epitopes were discovered that had both antagonistic as well as partial antagonist activities in cell-based assays.

The Fabrus human germline full length IgG library was produced by transfection of HEK293T cells and then collection of conditioned media. This high-throughput process resulted in a full length IgG library of over 10,000 members. This library was screened against several antigens directly on the cell surface using multiplexed high-throughput flow cytometry. Hits to the cancer target ErbB3 were functionally matured to enhance inhibition of this receptor tyrosine kinase. Herein, we describe the design, production, and screening principles of arrayed antibody libraries.

4. Applications

4.1. Antibody repertoires and natural mechanisms of diversification

In humans the naïve antibody repertoire results from V(D)J recombination of antibody heavy (V_H) and light (V_L) variable gene segments (Fig. 1). Each V_H-V_L pair contains six complementary determining regions (CDRs) at the antigen binding surface. The third CDR of the heavy chain (V_H CDR3) is longer than the others and often accounts for the majority of antigen binding [40]. The V_H CDR3 is encoded by the D and J regions in V(D)J recombination, as well as the junctional diversity at the V-D and D-J joint [37]. The diversity of the V_L CDR3 is achieved by the V-J joint. All of the human germline CDR1 and CDR2 loops can be classified into defined "canonical structures" [40] based on analysis by Chothia and Lesk of sequence and key framework residues. The CDR3 of V_H, due to

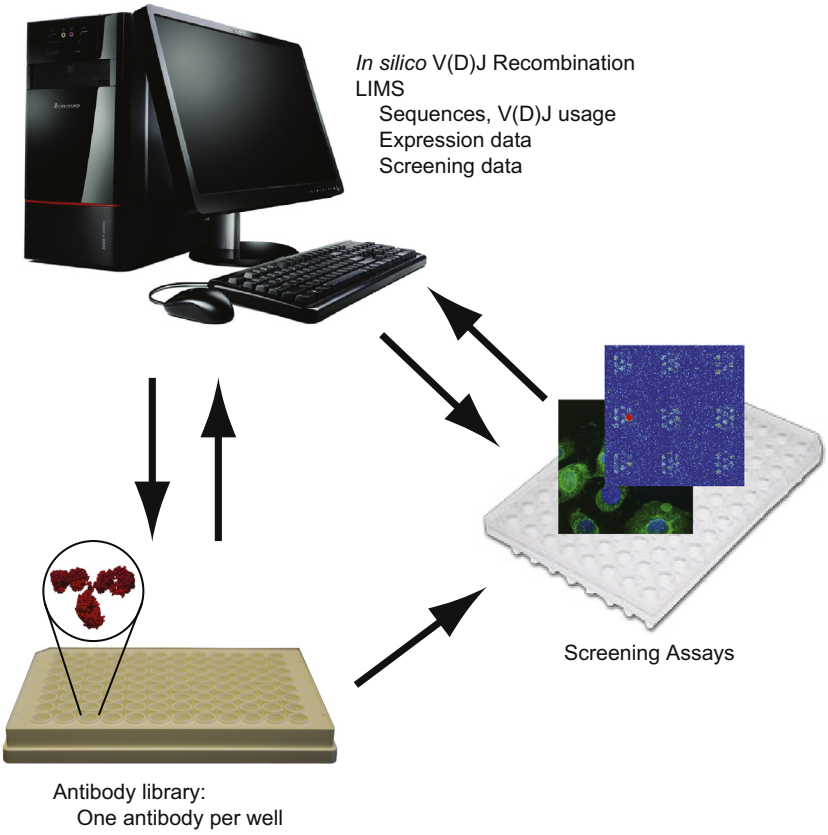

Fig. 2. Arrayed antibody library discovery system. The arrayed library utilized *in silico* V(D)J recombination as well as algorithms to select the most diverse sequences from a theoretical pool of $>10^7$. A laboratory informatics management system (LIMS) allows for storage of sequence, expression, and screening data for all members of the library. Relationships between hits and non-hits can be easily retrieved through searching parameters. The data is added cumulatively as the library is constructed and screened.

its longer length and sequence diversity, is not so easily classified. Different combinations of canonical structures of CDR1 and CDR2 and V(D)J combinations of CDR3 account for repertoire diversity. These structural components were considered in designing the arrayed library described below.

There are at least three mechanisms by which diversity is generated in humans (Fig. 1). First, combinatorial diversity, which defines the fetal or neonatal repertoire, is generated by V(D)J recombination of the V_H regions and VJ recombination of V_L regions. Although VDDJ recombination has been reported, the use of multiple D regions is disputed [41]. Second, junctional diversity can be created by nucleotide addition, or deletion at the V-D, D-J, or V-J junctions [37]. These nucleotide insertions and deletions at the recombination junctions can shift the reading frame of downstream D or J regions, leading to orders of magnitude more diversity in the expressed antibody repertoire. Combinatorial and junctional diversity together yield the germline adult repertoire. Third, following exposure to antigen, an antibody producing cell proliferates and undergoes somatic hypermutation of the recombined antibody gene, primarily via point mutations, though insertions and deletions may occur. The hypervariable regions coincide with the regions encoding the CDRs, producing orders of magnitude more diversity in the repertoire. The three mechanisms together produce the post-immune repertoire, which develops after exposure to antigen.

4.2. Repertoire size

It has been widely thought that antibody libraries of over 10^8 are required for efficient antibody lead discovery. This is largely based on the experience with phage display, where the required repertoire size is inflated due to biochemical characteristics specific to display: hits are only found if they survive multiple washing steps and exceed a high affinity detection limit caused by nonspecific "sticky" phage [42]. Studies on the repertoire *in vivo*, however, illustrate that antibody libraries of 10^4 may be sufficient [43,44]. Indeed, the diversity of the fetal/neonatal repertoire is relatively small due to the rarity of nucleotide insertions and deletions [45]. Because of this limited junctional diversity, the fetal/neonatal repertoire size can be estimated based on the product of the number of V, D, and J gene segments for the heavy chain, combined with the product of V and J for the light chain. The combinatorial diversity of the heavy chain can be calculated as $V_H \times D_H \times J_H$, or 6,318 (for 39 V_H regions and assuming one reading frame for D_H), and for the light chain as $V_\kappa \times J_\kappa + V_\lambda \times J_\lambda$, or 290. Since most D-segments can be used in two reading frames, the potential V_H repertoire is 12,636. This number is an estimate since some D-segments can actually be used in three reading frames and increase diversity; however, some D-segments encode amino acid sequences that are identical to other related D family members and thus do not add additional functional diversity. If every heavy chain can pair with every light chain, then $12{,}636 \times 290 = 3.7 \times 10^6$ different combinations could result. Thus, the potential fetal/neonatal repertoire is at least 3.7×10^6. These calculations do not account for biases in the repertoire or the distribution of variants produced; it is simply a calculation of the number of different antibodies that theoretically could be produced using the underlying molecular mechanisms of V(D)J recombination. It is appreciated that the light chain is often redundant in binding specificity [46] and some species like camelids lack light

chains altogether. Therefore, the total number of V_H recombinants may be considered a "lowest limit" of library size to be tested at a repertoire level.

4.3. Library design, construction and expression

Fabrus' *E. coli*-produced Fab arrayed library was designed to maximize structural diversity using a combinatorial approach. The theoretical combinatorial space was evaluated by recombining *in silico* all known human V, D, and J gene sequences ($\sim 10^6$ sequence combinations). Junctional diversity was not included, so the recombined collection was representative of the neonatal repertoire. At the V-D, D-J, or V-J joints a single nucleotide was inserted or deleted, as necessary, to maintain an in-frame sequence. Since 10^6 is a number too large to produce recombinantly in parallel, a library design algorithm was used to maximize structural diversity by using many different variable frameworks and V_H-CDR3 regions. This "reverse BLAST" alignment chose the most diverse (as opposed to most similar) sequences amongst the theoretically possible germline recombinants. In addition, every VDJ combination for V_H3-23 was included because this heavy chain framework is more highly represented *in vivo* [47–49] and has been used as a framework scaffold in several display-based library designs [50–52]. Thus 690 sequences containing V_H3-23 were produced, allowing complete coverage of the germline neonatal repertoire for this framework in an arrayed "sublibrary". Co-transformation of *E. coli* with V_H and V_L genes on separate compatible plasmids [53] was accomplished in 96-well format. Typically, 96 different heavy chains were co-transformed with a single light chain per plate. Replica heavy chain plates could then be paired with different light chains, in a process that scaled diversity from a relatively small number of antibody V_H and V_L genes. This scalability allowed small numbers of cloning events to produce much larger numbers of Fabs due to the combinatorial nature of V_H-V_L pairing. It became clear early in library production that some V_H and V_L frameworks conferred very poor expression properties, and these frameworks were eliminated as the library grew. In our experience, about 10% of Fabs were poor producers/expressors [29]. The final arrayed library contained 11,000 individual purified Fab species in microtiter plates.

Fabrus' IgG arrayed library was produced in mammalian cells and designed to represent a diversity of V_H and V_L frameworks. Similar to the Fab library, two plasmids encoding heavy chain and light chain were co-transfected into HEK293T cells. To construct the heavy and light chain variable domains, multiple families of human germline V regions were combined with D and J. Again, replica heavy chain plates were paired with different light chains to achieve scalable combinatorial association of separate heavy and light chain genes. Currently the full length library contains 10 V_H frameworks for a total of 1344 heavy chain genes and 4 kappa and 11 lambda light chain genes. The final arrayed IgG library contains 11,000 individual IgG species in conditioned media supernatants.

Information on the Fab or IgG concentration was determined for each individual member of the respective libraries using high-throughput quantification techniques. The performance of each member in each subsequent screen is also tracked. The maintenance of these large data sets requires some bioinformatic infrastructure, and a Laboratory Information Management System (LIMS) is required to truly maximize the value of the library and the data from each screen (Fig. 2). Each screen cumulatively adds to the data of each molecule in the database. Much like a small molecule library, this allows for the relationships among interesting molecules to be rapidly ascertained for SAR analysis.

5. Screening arrayed antibody libraries

The format of the antibody library is the fundamental differentiating factor for discovery between screening arrayed antibody libraries compared to selection from hybridoma or display based systems. The arrayed nature allows each antibody to be evaluated independently, eliminating competition from other antibodies and enabling true screening. The arrayed format also permits the use of high concentrations of protein in screening assays, similar to the use of combinatorial chemistry libraries. Unlike typical antibody discovery mechanisms like display or hybridoma, arrayed libraries do not rely on purified antigen material. With proper controls, arrayed libraries can be used directly on cells for either binding assays or assays based on cellular functions.

5.1. Multiplexed discovery

Since the arrayed antibody libraries are relatively small and comprised of germline antibody sequences, it is expected that the monovalent affinity for each hit would be low (micromolar range). Washing steps can remove specific binders with low affinity because they have fast off-rates. This issue can be overcome using either the Meso Scale Discovery (MSD) detection system or high-throughput flow cytometry for homogenous antibody binding. Each of these systems is amenable to multiplexed analysis, so multiple antigens can be screened against the antibody library simultaneously.

The arrayed Fab library was screened using the MSD system [29]. Purified protein antigen targets were printed on microtiter plates containing carbon electrodes in the bottom of each well. When a current is applied to the plate, electron transfer can occur between the plate and a ruthenium-labeled secondary antibody if the latter is very near the bottom of the plate. Upon application of a current, the ruthenium tag emits light in an electrochemiluminescent reaction. Importantly, *unbound* secondary antibody in solution does *not* give electrochemiluminescent signal. This multiplexed system allows each antibody to be evaluated against several targets simultaneously. This is important as specificity information is obtained at the earliest stage of discovery. Because the electrochemiluminescent platform does not require washing steps, Fabs could be discovered against multiple targets that ranged in affinity between 0.5 and >50 μM K_d. When many of these germline antibodies were evaluated for target binding by standard ELISAs, they could not reproducibly be detected (unpublished data). These hits would not likely be identified by other discovery methods like phage display due to their low affinity. Furthermore, such low affinity binders would likely be outcompeted by higher affinity binders in display systems that might bind alternative, immunodominant epitopes. When a "hit" is identified it can be confirmed by titrating newly expressed Fab in an MSD binding assay, or converted to IgG for evaluation on cells overproducing the target of interest. The discovery rate was high; hits were obtained against seven of nine targets, and between 4 and 30 hits were found per target.

Discovery using recently developed high-throughput flow cytometry systems [52] allows for multiplexed screening directly on cells using the arrayed Fabrus IgG library. This again facilitates detection of low affinity binders owing to the added avidity of the double armed IgG and the further stabilizing effect of the secondary detection antibody. After incubating target cells with library IgG, secondary antibody is added, cells are pelleted and resuspended and directly run on the flow cytometer. Flow cytometry is also amenable to multiplex screening by the labeling of different target cell populations with varying amounts of the fluorescent dye, calcein AM.

The flow cytometry screen described here was carried out against 3 cell populations transiently transfected with ErbB3, Tim3, or TPOR. Cells expressing different antigens were then stained with different doses of calcein AM violet. The transfected and differentially stained cells were mixed, and the IgG library screened using a three laser CyAn ADP flow cytometer. The IgG binding to the cells was detected using a fluorescent secondary antibody and was evaluated for each uniquely-stained population. The Hypercyt Autosampler and Hyperview Analysis software (IntelliCyt, Inc.) provide a high-throughput solution to library screening. Fig. 3A shows a representative multiplex screening of one library plate of 96 IgGs against three antigens, with each panel showing one antigen (using the violet channel to define each population). In each panel, fluorescence of the detecting secondary antibody (*Y*-axis) is plotted over time (*X*-axis), which reveals the sequential delivery of 96 samples to the flow-cytometer. This data provides immediate information about the specificity of a hit.

The target cells were co-transfected with GFP and the target of interest. Taking advantage of the correlation between GFP expression and expression of the co-transfected target antigen, hits could be identified by binding of the library antibodies to GFP-high cells compared to the GFP-low cells within a given target population as well as GFP-high cells between target populations. This strategy provided a robust signature for a true hit. Fig. 3B shows an example of a TPOR hit from newly transfected supernatants and tested singly to confirm binding.

Once hits are confirmed, they are characterized and prioritized by a variety of cellular assays. For ErbB3-binding antibodies, interesting characteristics included ability to bind endogenous ErbB3 on cancer cell lines, competitive effects of ligand on the antibody binding to ErbB3, domain mapping of the antibody binding site on ErbB3 (all assessed by flow cytometry), and functional inhibition of ErbB3 phosphorylation (pErbB3) in BT474 cells. Based on the characterization of each hit by the described assays, two hits were selected for further development and affinity maturation, FE3003 and FE3013. Fig. 3C illustrates the functional inhibition of neuregulin-stimulated pErbB3 by affinity matured mutants of the germline hits. In this assay, affinity maturation was required to achieve functionality. We could apply the same screening and confirmation evaluations used on the original hits to our affinity maturation candidates in a high-throughput manner. This approach allows for the functional assessment of each mutation of a given hit and creates the opportunity to do functional maturation, rather than just affinity maturation, and proceed with candidate development using a rational stepwise paradigm.

6. Considerations in arrayed antibody libraries

6.1. Hit optimization

Most antibody optimization technologies rely on display techniques and involve producing libraries of randomized amino acids

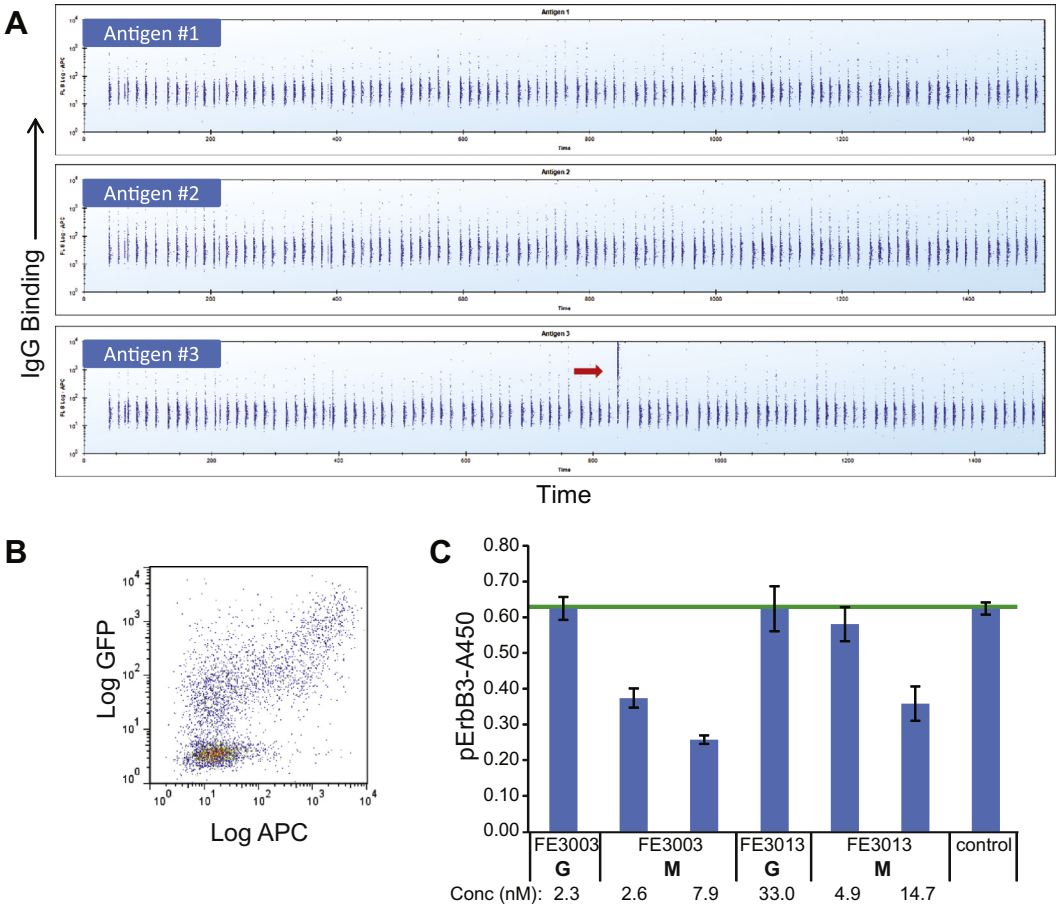

Fig. 3. IgG arrayed library screening and development. (A) Binding of the arrayed IgG library by flow cytometry. A representative 96-well library plate is screened against three antigen-expressing cell populations. Each cell population is shown separately with detection of the secondary antibody on the *Y*-axis and time on the *X*-axis, revealing the sequential delivery of 96 samples to the flow-cytometer. Red arrow indicates strong binding of an IgG to the antigen #3 population, but not antigen #1 or #2. (B) TPOR hit binding is proportional to expression level in TPOR and GFP co-transfected cells. More efficiently transfected cells will express more of both TPOR and GFP, and here APC-conjugated detection antibody binding correlates with GFP signal. (C) Inhibition of phosphorylated ErbB3 (pErbB3) in neuregulin treated BT474 cells (10 ng/ml). Supernatants containing mutants (M) of germline (G) IgG hits FE3003 and FE3013 were able to inhibit pErbB3 in a concentration dependent manner (final assay concentrations shown) compared to controls as determined by sandwich ELISA (R&D, cat# DYC1769). The green line indicates pErbB3 levels in the presence of 10 ng/mL neuregulin.

in the CDRs, then selection for the higher affinity variants. Multiple reviews have discussed affinity maturation techniques and strategies [54–58]. During the discovery process with an arrayed library, multiple "hits" for a given target can be analyzed for similarities or differences in their sequence characteristics, and can also be compared to non-binding but related sequences. In this way, true sequence-activity-relationships (SARs) can be achieved in a manner analogous to combinatorial and medicinal chemistry optimization. Such relationships are not possible in display screening, since it is impossible to know which sequences in the original library were true negatives. Once SARs are determined, regions of interest for optimization can be identified. Often, a clear CDR or even amino acid pattern can be recognized as being important in antigen interaction based on the SAR [29]. If patterns are not recognizable for all six CDRs, an alanine scan can be done to determine the importance of each CDR, and also the role of each individual amino acid. Once residues are selected as being involved in antigen binding, small sublibraries can be created based on the hit sequence. Most often, a single residue is mutated to multiple other residues then each individual mutant is screened.

As in the screening of the arrayed antibody library, the arrayed format for optimization of early leads has great benefits. The magnitude of improvement of a single mutation can be compared to all other mutants, and improvements to other parameters such as expression yield or function can be weighed in concert. The format itself, as in discovery, is amenable to optimization using functional cell based screens, where affinity *per se* may not be the optimized parameter. Antibody hits with a K_d as low as >10 μM have not been a barrier to antibody development and affinity maturation. Often one or two mutations can improve the off-rate, producing binders in the nM range [29]. Interestingly, substantial improvements in affinity are often seen with a simple alanine scan. It appears that replacing bulky residues with small residues can often "remove a negative" and enhance antigen binding. Iterative processes of library creation, mutation screening, combining useful mutations, and further screening have achieved subnanomolar K_ds.

6.2. Potential pitfalls

Some pitfalls encountered in arrayed protein screening stem from the state of the purified antigen and cell surface complexity. Purified antigen can be difficult to characterize to ensure that epitopes are biologically available. Additionally, purified antigen has the potential to have alternative or aggregated states. Screening on cells obviates the need for purified protein. Multiplexed screening on cells using high-throughput flow cytometry has the potential to find "hits" that bind the cell type of interest, but not the target of interest. This artifact may result from (i) genetic drift of the stable line from the parental line, (ii) secondary effects of expression of the target on the cell, (iii) expression of the target at levels that cause biologically irrelevant conformations, or (iv) in the case of transient transfection, intersection of transfection-mediated signaling with the specific target expressed. A robust confirmation and validation strategy can identify true and developable lead antibodies. Validation options include siRNA knockdown of the target, target ligand competition with the antibody hit, or reduced hit binding due to ligand-induced internalization.

6.3. Costs

The costs associated with producing display based libraries are variable and dependent on the library design goals. Arrayed protein libraries are more costly. The first Fab library constructed by Fabrus used total DNA synthesis of each individual V-region (approximately 300 base pairs). Currently, a cost of $0.30/bp would allow a single V-region to be synthesized for less than $100. Thus a 10,000 member antibody gene library consisting of 1000 heavy chains and 10 light chains would have DNA material costs of $110,000. The cost of the library is obviously determined by the ratio of heavy chains to light chains. For example a heavy chain library of 10,000 paired with only a single light chain would cost over $1 M. These costs can be significantly decreased by engineering restriction sites in multiple framework vectors such that the cost to expand the gene library is determined by short oligonucleotides as opposed to larger tracts of gene synthesis. Using Piccolo, a fully automated system for recombinant protein production, costs included plasticware, media, inducer, and purification resin. Overall, the material cost of the 11,000 member Fabrus Fab library was approximately $15 per purified antibody fragment. Similarly, the material cost of the 11,000 member Fabrus IgG library was approximately $13 per full length IgG supernatant. While these libraries are completely scalable, production, storage and maintenance increase proportionally to the size of the library. This is substantially more expensive than a typical phage display library. In contrast, a combinatorial chemistry small molecule library may cost well over $300 per compound to produce. As the arrayed format is more comparable to combinatorial chemistry small molecule libraries, $15/mAb is inexpensive. The reason for this substantial difference is that the synthesis and purification for every individual antibody is standardized, as opposed to chemistry libraries where synthesis and purification methods can vary widely.

Despite the increased cost of library production compared to display libraries, the ultimate efficiency of the arrayed approach will be determined when the advantages are clearly defined over several projects. For example, it is clear that initial discovery is far more rapid than display systems. Attrition of hits that are false positives or undevelopable is immediate. Thus, the costs associated with determining specificity, manufacturability, etc., after several rounds of immunization or panning in an individual project may off-set the initial higher costs of arrayed library construction. Furthermore, the utility of finding antigens against unique and difficult targets like GPCRs, or novel epitopes, could be the difference between success and failure of a discovery campaign. These latter enabling features remain to be proven.

7. Conclusions

Arrayed libraries for protein drug discovery are now a reality. Although the construction and maintenance of these libraries may take considerable resources, the ability to achieve rapid, multiplexed discovery in cell-based systems could open new areas for biologic discovery. Extremely important targets like GPCRs, ion channels, and other multispanning membrane proteins may be accessible through cell-based assays screened against an arrayed library. Similarly, mechanisms of action not typically attributed to antibodies could be discovered and optimized. The future may hold antibody drugs with modulator, partial agonist, or partial antagonist functions, all of which could be tunable for optimal drug function against traditionally difficult targets.

Acknowledgements

We thank the late William D. Huse, Richard A. Lerner, Peter G. Schultz, Helen Mao, and Alex Polinsky for helpful scientific discussions.

References

[1] P. Kung, G. Goldstein, E.L. Reinherz, S.F. Schlossman, Science 206 (1979) 347–349.
[2] A.L. Nelson, E. Dhimolea, J.M. Reichert, Nat. Rev. Drug Discov. 9 (2010) 767–774.
[3] G. Kohler, C. Milstein, Nature 256 (1975) 495–497.

[4] U. Gubler, B.J. Hoffman, Gene 25 (1983) 263–269.
[5] H. Okayama, P. Berg, Mol. Cell. Biol. 2 (1982) 161–170.
[6] Y.L. Chiang, R. Sheng-Dong, M.A. Brow, J.W. Larrick, Biotechniques 7 (1989) 360–366.
[7] J.W. Larrick et al., Biochem. Biophys. Res. Commun. 160 (1989) 1250–1256.
[8] G.L. Boulianne, N. Hozumi, M.J. Shulman, Nature 312 (1984) 643–646.
[9] S.L. Morrison, M.J. Johnson, L.A. Herzenberg, V.T. Oi, Proc. Natl. Acad. Sci. 81 (1984) 6851–6855.
[10] M.S. Neuberger et al., Nature 314 (1985) 268–270.
[11] C. Queen et al., Proc. Natl. Acad. Sci. 86 (1989) 10029–10033.
[12] L.L. Green et al., Nat. Genet. 7 (1994) 13–21.
[13] N. Lonberg et al., Nature 368 (1994) 856–859.
[14] G.P. Smith, Science 228 (1985) 1315–1317.
[15] W.D. Huse et al., Science 246 (1989) 1275–1281.
[16] J. McCafferty, A.D. Griffiths, G. Winter, D.J. Chiswell, Nature 348 (1990) 552–554.
[17] R.W. Roberts, J.W. Szostak, Proc. Natl. Acad. Sci. U.S.A 94 (1997) 12297–12302.
[18] J. Hanes, L. Jermutus, S. Weber-Bornhauser, H.R. Bosshard, A. Pluckthun, Proc. Natl. Acad. Sci. U.S.A 95 (1998) 14130–14135.
[19] E.S. Smith, S. Shi, M. Zauderer, Methods Mol. Biol. 269 (2004) 65–76.
[20] E.T. Boder, K.D. Wittrup, Nat. Biotechnol. 15 (1997) 553–557.
[21] G. Georgiou et al., Nat. Biotechnol. 15 (1997) 29–34.
[22] R.R. Beerli et al., Proc. Natl. Acad. Sci. U.S.A 105 (2008) 14336–14341.
[23] M. Ho, S. Nagata, I. Pastan, Proc. Natl. Acad. Sci. U.S.A 103 (2006) 9637–9642.
[24] C. Zhou, F.W. Jacobsen, L. Cai, Q. Chen, W.D. Shen, MAbs 2 (2010) 508–518.
[25] J.P. Overington, B. Al-Lazikani, A.L. Hopkins, Nat. Rev. Drug Discov. 5 (2006) 993–996.
[26] I. Hötzel et al., Prot. Eng. Des. Sel. 24 (2011) 679–689.
[27] C.J. Hutchings, M. Koglin, F.H. Marshall, MAbs 2 (2010) 594–606.
[28] D.R. Webb, T.M. Handel, A. Kretz-Rommel, R.C. Stevens, Biochem. Pharmacol. 85 (2013) 147–152.
[29] H. Mao et al., Nat. Biotechnol. 28 (2010) 1195–1202.
[30] A.A. Jungbluth et al., Proc. Natl. Acad. Sci. 100 (2003) 639–644.
[31] D.J. Miller, M. Rodriguez, J. Immunol. 154 (1995) 2460–2469.
[32] T. Bobrzynski et al., J. Immunol. 175 (2005) 6589–6596.
[33] S.E. Lacy et al., J. Immunol. 181 (2008) 1282–1287.
[34] J.M. Chandonia, S.E. Brenner, Science 311 (2006) 347–351.
[35] A. Joachimiak, Curr. Opin. Struct. Biol. 19 (2009) 573–584.
[36] F. Matsuda et al., J. Exp. Med. 188 (1998) 2151–2162.
[37] V. Smider, G. Chu, Semin. Immunol. 9 (1997) 189–197.
[38] H. Lin et al., Science 320 (2008) 807–811.
[39] R. Gonzalez et al., Proc. Natl. Acad. Sci. 107 (2010) 3552–3557.
[40] C. Chothia, A.M. Lesk, J. Mol. Biol. 196 (1987) 901–917.
[41] S.J. Corbett et al., J. Mol. Biol. 270 (1997) 587–597.
[42] B. Levitan, J. Mol. Biol. 277 (1998) 893–916.
[43] M.F. Bachmann, T.M. Kundig, C.P. Kalberer, H. Hengartner, R.M. Zinkernagel, J. Immunol. 152 (1994) 4235–4241.
[44] M. Cohn, R.E. Langman, Immunol. Rev. 115 (1990) 11–147.
[45] S. Lewis, Adv. Immunol. 56 (1994) 27–150.
[46] M. Czerwinski, D. Siemaszko, D.L. Siegel, S.L. Spitalnik, J. Immunol. 160 (1998) 4406–4417.
[47] H.-P. Brezinscheck, S.J. Foster, T. Dorner, R.I. Brezinschek, P.E. Lipsky, J. Immunol. 160 (1998) 4762–4767.
[48] H.P. Brezinschek et al., J. Clin. Invest. 99 (1997) 2488–2501.
[49] R.M.T. de Wildt, R.M.A. Hoet, W.J. van Venrooij, I.M. Tomlinson, G. Winter, J. Mol. Biol. 285 (1999) 895–901.
[50] A. Pini et al., J. Biol. Chem. 273 (1998) 21769–21776.
[51] S. Schoonbroodt et al., J. Immunol. 181 (2008) 6213–6221.
[52] E. Soderlind et al., Nat. Biotechnol. 18 (2000) 852–856.
[53] B. Leonard, V. Sharma, V. Smider, J. Immunol. Methods 317 (2006) 56–63.
[54] A.R. Bradbury, S. Sidhu, S. Dubel, J. McCafferty, Nat. Biotechnol. 29 (2011) 245–254.
[55] D. Filpula, Biomol. Eng. 24 (2007) 201–215.
[56] A.M. Levin, G.A. Weiss, Mol. Biosyst. 2 (2006) 49–57.
[57] N. Lonberg, Curr. Opin. Immunol. 20 (2008) 450–459.
[58] A. Rothe, R.J. Hosse, B.E. Power, Expert Opin. Biol. Ther. 6 (2006) 177–187.