Biotechnology and Genetic Engineering Reviews - Vol. 28, 1-16 (2011)

Arrayed protein library technology for therapeutic biologic discovery

HONGYUAN MAO, JAMES GRAZIANO, AND VAUGHN V. SMIDER[1*]

Fabrus LLC, La Jolla, CA and 2: Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA

Introduction

The first monoclonal antibody was appproved for human clinical use in 1985. This antibody, OKT3 (Kung *et al*., 1979), was derived from a mouse hybridoma and produced anti-mouse reactions that limited its use, and hindered further development of monoclonal antibodies. Humanization technology eliminated these cross-species issues and resulted in the approval of the first recombinant "chimeric" antibody rituximab and "humanized" antibody trastuzumab in 1997 and 1998, respectively. The approval of these molecules opened the floodgates for antibody development, as over 20 antibodies are currently approved by the FDA for therapeutic use and hundreds are in clinical development (Nelson *et al*., YEAR?). The need for humanized recombinant antibodies spurred the develoment of several technologies for their discovery and engineering.

TECHNOLOGIES USED FOR DISCOVERY AND OPTIMIZATION OF BIOLOGICS.

Hybridoma technology. Kohler and Milstein invented the hybridoma technique in 1975 (Kohler and Milstein, 1975). Since antibodies are made from a multitude of gene segments that rearrange at the DNA level during B-cell development (Figure 1), cloning an individual antibody gene was not straightforward. However, genes could be cloned through cDNA library techniques developed in the early 1980s

*To whom correspondence may be addressed (wsmider@scripps.edu)

Abbreviations: CDR : complementary determining region; mAb: monoclonal antibody; O.D.: optical density; TdT: terminal deoxynucleotidyl transferase; V: variable; D: diversity; J: joining; GPCR: G-protein coupled receptor; CVB: culture vessel block; LIMS: laboratory information management system; Fab: fragment antigen-binding

(Gubler and Hoffman, 1983; Okayama and Berg, 1982). 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 (Chiang *et al.*, 1989; Larrick *et al*., 1989). Thus, hybridoma methods could be combined with rapid molecular biology techniques by the 1990s. Since mouse variable regions were often 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 chain (Boulianne *et al.*, 1984; Morrison *et al*., 1984; Neuberger *et al*., 1985). Shortly thereafter, molecular and homology modeling allowed replacement of murine V-region sequences with human frameworks, leaving just CDR sequences derived from mouse origin (Queen *et al*., 1989). Efforts in the 1990's generated transgenic mice that contained the human immunoglobulin loci, such that human antibodies could be induced directly by immunization (Green *et al*., 1994; Lonberg *et al.*, 1994).

Display technologies. 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) (Smith, 1985). Shortly thereafter, these methods were applied to antibody gene fragments (Huse *et al*., 1989; McCafferty *et al*., 1990). There are now multiple formats for display technology including mRNA (Roberts and Szostak, 1997), ribosome (Hanes *et al*., 1998), phage, eukaryotic viruses (Smith *et al*., 2004), yeast (Boder and Wittrup, 1997), bacterial (Georgiou *et al.*, 1997), and mammalian cells (Beerli *et al.*, 2008; Ho *et al*., 2006; Zhou *et al*., 2010). The library size depends on the size of the particle, where libraries of 10^{14} in mRNA or ribosome display to 10^{6} for mammalian cell display can be constructed and selected. For particle based discovery using phage, mRNA, or ribosomes, the 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 is used with fluorescently labeled antigen to select binders in the library population.

Selection versus screening. In hybridoma and display technologies the discovery paradigm is selection; a *pool* of molecules is the starting point and individual members are selected from the pool based on their affinity to the target. In this regard the *selection* process involves *competition* between individual library members for binding the target antigen. For display libraries those members with lower K_d will have more members bound at equilibrium than those with a higher K_d. Similarly, those with longer off-rates will survive washing steps and incubation times. Additionally, other factors like differences in growth rate of the cells expressing individual library members can affect selection by allowing non-binders to efficiently propagate. For immunization experiments, those B-cells with higher affinity out compete those with lower affinity based on signaling through the B-cell receptor (BCR) and expansion of the higher affinity B-cells. In each case, however, the discovery mechanism is competitive selection. Arrayed based libraries use a *screening* paradigm, which could have advantages and disadvantages compared to competitive selection.

UNMET NEEDS FOR BIOLOGIC DISCOVERY TECHNOLOGY

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 is 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. Thus, 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. It is extraordinarily difficult to produce antibodies against these targets because (i) they are not easily purified in active form, and (ii) they are often refractive to overexpression on the cell surface. Enabling technologies able to address these receptors in the context of the cell would be very useful.

Non-protein targets. Importantly, current antibody technology is not easily amenable to discovery against many non-protein targets. Carbohydrates, and even certain glycoproteins, are extremely important in the pathogenesis of many diseases. Immunization is often not effective in illiciting IgG due to the requirement of peptide display in MHC 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 the highly sensitive MSD system described below which can identify specific binders of >50 μM K_d.

Novel epitopes. Lastly, it would be useful to identify antibodies that are specific for a particular form of a receptor. Current antibody discovery efforts are aimed at finding specific proteins that are expressed, or at least significantly upregulated, in the disease state compared to normal tissue. It is clear that disease-specific proteins like Her2 or CD20 are relatively rare. *Epitopes* on some of these proteins, however, may be specifically present in the disease state. Such epitopes could be the basis for target specific antibody therapy. For example splice variants, post-translational modifications, multiprotein complexes, and "activated" forms of the receptor would be very useful species to target. Some unique 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 with cells transfected with EGFRvIII, an activated form of EGFR with deletional mutation of exons 2-7 that is often found in glioblastoma and other cancers. The resulting antibody reacted with both EGFRvIII as well as overexpressed, ligand-activated wtEGFR (Jungbluth *et al*., 2003). Such modified receptors may only be found on the surface of cells and not engineerable in the purified proteins used for immunization or display. Interestingly, other very unique antibodies that induce remyelination (Miller and Rodriguez, 1995), or have anaphylactogenic activity (Bobrzynski *et al*., 2005) have been found to be present naturally and in germline configuration. These antibodies would also be difficult to discover in the absence of purified target protein. Thus, although antibody technology is particularly mature for some targets, many others are elusive to standard techniques.

Mechanisms of action. Nearly all antibodies on the market are high affinity antagonists. Antagonism and affinity appear to be correlated. For example, a low affinity antibody against the erythropoietin receptor had agonist activity but the affinity matured antibody was antagonistic (Lacy *et al*., 2008). Thus, the ability to modulate affinity could have substantial impact on the mechanism of action of an antibody drug. Since hybridoma and display techniques rely on inherent affinity as the dominant parameter by which to identify an antibody candidate, these methods could miss leads that have novel mechanisms of action that are only observable at lower affinities. Small molecule drugs can have many mechanisms of action including antagonist, agonist, partial agonist or antagonist, and modulator. There is no *a priori* reason why antibodies could not also have these activities, however discovery systems like hybridoma and display methods are not built for identifying them. Cell based screening assays could be used to identify such activities in spatially addressed libraries as they do for small molecules.

ARRAYED PROTEIN LIBRARIES

Small molecule combinatorial chemistry libraries are "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. 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 (Chandonia and Brenner, 2006; Joachimiak, 2009). Second, the detailed understanding of antibody gene rearrangement and the complete sequences of the human antibody loci were only recently accomplished (Matsuda *et al*., 1998; Smider and Chu, 1997). Such understanding is useful in the rational design of a biologic library in a spatially addressed format. Third, the cost of *de novo* DNA synthesis was prohibitive until recent competition amongst DNA suppliers drove the cost down to affordable levels.

Examples of spatially addressed protein libraries. Two groups have recently developed non-antibody spatially addressed protein libraries based on the human "secretome". With the human genome sequencing efforts complete, it is possible to bioinformatically predict all proteins that are secreted by the presence of a leader peptide. Five-Prime Therapeutics and The Genomics Institute of the Novartis Research Foundation (GNF) have both recently developed 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. This library was produced in transfected HEK293T cells, and conditioned supernatents 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 through an 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 (Gonzalez *et al*., 2010). Both of these approaches illustrate the ability to identify protein "leads" in cell based screening programs.

Recently Fabrus LLC has developed methodology to produce human germline Fab antibody fragments in a high-throughput fashion (Mao *et al*., 2010). In a parallel process, 576 (six 96-well plates) Fabs could be expressed in *E.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. We demonstrated that hits could be obtained to 7 out of 9 antigens. Hits to the cancer target DLL4 were further 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 could produce 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. Herein we describe the design, production, and screening principles of spatially addressed antibody libraries.

Applications

ANTIBODY REPERTOIRES AND LIBRARY DESIGN

In humans the naïve antibody repertoire results from V(D)J recombination of antibody heavy (V_H) and light (V_L) variable gene segments (Figure 1). The diversity of the repertoire results from V(D)J recombination as well as from structural plasticity present within individual antibodies (*e.g.* different folding patterns of the same sequence). For Fabs, each chain contains one constant and one variable region. The variable regions participate in antigen binding as V_H-V_L dimers. Each V_H-V_L pair contains six complementary determining regions (CDRs) at the antigen binding surface, which are loops that join beta strands in the immunoglobulin fold. The third hypervariable region of the heavy chain (V_H CDR3), is longer than the others and often accounts for the majority of antigen binding (Zu and Davis, 2000). The V_H CDR3 is encoded by the D region in V(D)J recombination, as well as the junctional diversity at the V-D and D-J joint (Smider and Chu, 1997). The other CDRs of V_H are entirely encoded by the V-region genes. The V-J joint of light chains can produce diversity within CDR3 of V_L, whereas CDR1 and CDR2 of V_L is entirely encoded within the V-regions. Although considerable diversity can be introduced into CDR3 during recombination, the number of V-regions allow significant structural diversity of CDR1 and CDR2 to be present in the repertoire as well. Chothia and Lesk analyzed the loop structures in antibody CDRs, and classified them into several groups. These classifications are based on both loop sequence and key framework residues. All of the human germline CDR1 and CDR2 loops can be classified into one of these "canonical structures" (Chothia and Lesk, 1987). CDR3 of V_H, due to its longer length and sequence diversity, is not easily classified. There are seven human V_H families (Matsuda *et al*., 1998), six V-kappa, and 10 V-lambda. Within V-region families, different combinations of canonical structures account for structural diversity of the repertoire. These structural components were considered in designing the spatially addressed repertoire described below.

Natural mechanisms of antibody diversification. There are at least three mechanisms by which diversity is generated in humans (Figure 1). First, V(D)J recombination

allows one of 39-45 V_H regions to recombine with one of 27 D_H regions, and one of 6 J_H regions. Although VDDJ recombination has been reported, the use of multiple D regions is controversial(Corbett et al., 1997). Similarly for the V_L, one of 34 V_κ regions recombines with one of 5 J_κ regions, or one of 30 V_λ regions recombines with one of 4 J_λ regions. Second, diversity can be created by templated (*e.g.* P-nucleotides) or non-templated (N-nucleotides) addition, as well as nucleotide deletion, at the V-D, D-J, or V-J junctions (Smider and Chu, 1997). In fact these insertions and deletions at the recombination junctions can lead to orders of magnitude more diversity in the repertoire. Third, following exposure to antigen, the antibody genes become somatically mutated, wherein both nucleotide base changes as well as insertions and deletions can occur. These mutations again produce orders of magnitude more diversity in the repertoire. The first process, combinatorial diversity, produces the fetal or neonatal repertoire. The first and second produce the germline adult repertoire. The three together produce the post-immune repertoire, which develops after exposure to antigen.

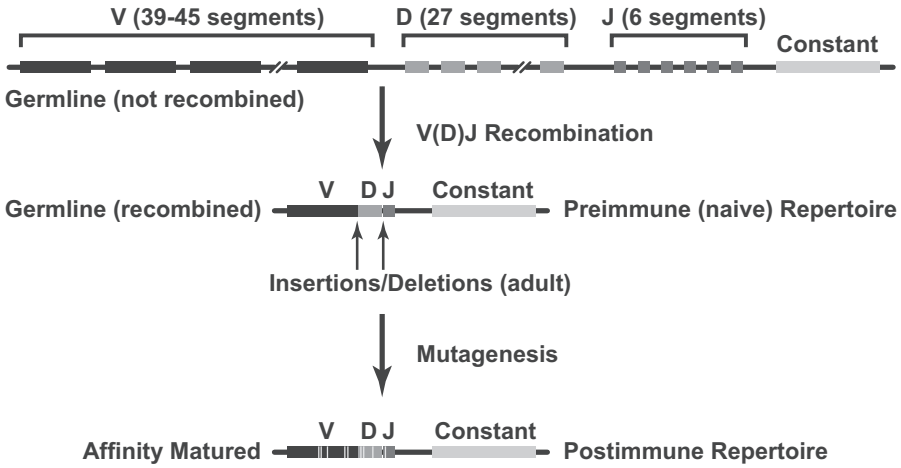

Figure 1. Antibody diversity generating mechanisms. In human pre-B cells 39-45 V, 27 D, and 6 J segments can recombine at the DNA level to produce a functional V-D-J variable region at the heavy chain locus. A similar V-J process occurs for the kappa or lambda light chains (where there are no D regions).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

Repertoire sizes. It has been widely accepted 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, namely multiple washing steps and a high affinity detection limit caused by nonspecific "sticky" phage (Levitan, 1998). Studies on the repertoire *in vivo*, however, illustrate that this number may be as low as 10^4 (Bachmann *et al*., 1994; Cohn and Langman, 1990). Indeed, the diversity of the fetal/neonatal repertoire is relatively small. The fetal/neonatal repertoire does not use TdT to add N-nucleotides, however there is still the potential for templated (*e.g.* P-nucleotides) or deletions to occur at the recombination junctions. P-nucleotides are rare (Lewis, 1994b), and the fetal/neonatal repertoire has limited deletional junctional diversity, so

we estimated the fetal/neonatal repertoire as being the product of the number of V, D, and J gene segments for the heavy chain, or the product of V and J for the light chain. The combinatorial diversity of the heavy chain can be calculated as $V_H x D_H x J_H$, or 6,318 (for 39 V_H regions and assuming one reading frame for D_H), and for the light chain as $V_k x J_\kappa + V_l x J_\lambda$, 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; but also because some D segments encode the exact amino acid sequence as related D family members and thus do not add additional functional diversity). If every heavy chain can pair with every light chain, then $12{,}636 \times 290 = 3.6 \times 10^6$ different combinations could result. Thus, the potential fetal/neonatal repertoire is at least 3.6×10^6. Of course 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, and some species like camelids even lack light chains, therefore the total number of V_H recombinants may be considered a "lowest limit" of library size to be tested at a repertoire level.

Library Design and Construction. After obtaining all known V, D, and J gene sequences, all of these were recombined *in silico* to produce over 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, while minimizing the total diversity. This "reverse blast" chose the most diverse (as opposed to most similar) sequences amongst the theoretically possible germline recombinants. A very common heavy chain framework, V_H3-23, is overutilized *in vivo* and has been used as a framework scaffold in several display-based library designs. We used this V_H framework and combined it with every possible D_H and J_H, utilizing every possible D_H reading frame that did not have a stop codon. Thus 690 sequences containing V_H3-23 were produced, allowing complete coverage of the germline neonatal repertoire for this framework in a spatially addressed "sublibrary". The V_H and V_L genes were cloned in separate compatible plasmids (Leonard *et al*., 2006). Thus, the genetic component of the library was segregated as separate heavy and light chains, which were combined for Fab expression. It became quite clear early in library construction that some V_H and V_L frameworks conferred very poor expression properties, and these frameworks were eliminated as the library grew; thus they are relatively underrepresented.

Library expression. Fabrus utilized *E.coli* to secrete Fab fragments in the periplasm. Co-transformation of the separate V_H and V_L genes was accomplished in 96-well format. Typically 96 different heavy chains were cotransformed 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.

Deep well 96-well blocks were used as inocula for the Piccolo automated parallel expression unit (Wollerton *et al.*, 2006). 12-well culture vessel blocks (CVBs) held 12 ml of media/well and each individual well was monitored for growth by $O.D._{600}$

readings at programmed intervals. The Piccolo unit allows for automated induction at a predetermined $O.D._{600}$ so growth and expression parameters were fairly uniform for all members of the library, except in cases where specific antibody fragments impact growth and/or expression. Final $O.D._{600}$ readings were often greater than 20 in this parallel fermentation system. A C-terminal 6xHis tag allowed a first purification step by IMAC within the Piccolo unit. Following reformatting back to 96-well blocks, a second purification on protein G columns was accomplished with an AKTA FPLC unit containing a 96-well autosampler. A single 1 ml column (Hi-trap) was used per plate, with regeneration between each library member after acid elution. $O.D._{280}$ readings were automated so a calculation of purified yield and sample concentration could be obtained for each library member. The purified protein libraries were stored at 4° or -80° C in neutralized elution buffer (Tris-glycine) at 1 ml per sample containing between 0-400 mg of Fab, depending on the yield.

Bioinformatics requirements. A laboratory informatics management system (LIMS) was used to track samples, as well as maintain a database of information about each individual sample (Figure 2). Barcoded blocks, plates, and 2-D barcoded tubes (Matrix/Thermo Scientific, Hudson, NH) were employed for ease of sample tracking and use. The LIMS system employed by Fabrus was developed by CoreLIMS (Branford, CT), and customized to store sequence, V, D, J gene usages, sample concentrations, and screening data. The data is searchable by keyword, sequence, ID numbers, or genetic composition (*e.g.* V, D, J usage).

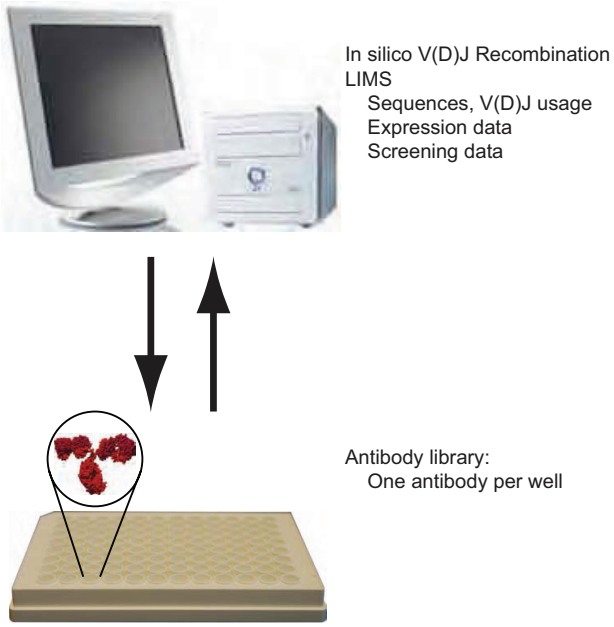

Figure 2. Addressed antibody libraries and informatics components. The original design of the spatially addressed library utilized *in silico* V(D)J recombination as well as algorithms to select the most diverse sequences from a theoretical pool of $>10^7$. A laboratory informatics management system (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.

SCREENING ARRAYED PROTEIN LIBRARIES

The format of the antibody library is the fundamental differentiating factor for discovery using this system compared to hybridoma or displayed based systems. The addressed nature allows every antibody to be evaluated independently of one another, eliminating competition as a mechanism for discovery, and enabling true screening. Furthermore, data on each individual library member is maintained in the LIMS, allowing relationships amongst interesting molecules to be rapidly ascertained. In this regard, each screening program cumulatively adds to the data of each molecule in the database. The arrayed format allows high concentrations of protein to be used in screening assays, similar to the use of combinatorial chemistry libraries. Although typical antibody discovery mechanisms like display or hybridoma methods rely on purified target protein, the arrayed library can be used directly on cells for either binding assays or assays based on cellular functions.

Multiplexed discovery. The Mesoscale Discovery (MSD, Gaithersburg, MD) detection system is comprised of printed proteins in wells of a microtiter plate that contains electrodes in the bottom of each well (Figures 3 and 4). When a current is applied to the plate, electron transfer can occur between the plate and a ruthenium-labeled secondary antibody if it is bound to 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. The MSD system is useful for discovery using arrayed antibody libraries for two reasons. First, it allows printing of protein in spots in the bottom of each well in a miniarray format, allowing each antibody to be evaluated against several targets simultaneously. This is important as specificity information is obtained at the earliest stage of discovery. Second, the electrochemiluminescent platform does not require washing steps. Since the arrayed antibody library is 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 but low affinity binders because they have fast off-rates. Indeed we could discover Fabs against multiple targets that ranged in affinity between 0.5 and >50 μM. When many of these germline antibodies were evaluated for target binding by standard ELISA, they could not reproducibly be detected (HM and VS, unpublished data). These hits would not likely be identified by other discovery methods like phage display due to this low affinity. Furthermore, such low affinity binders would also likely be outcompeted by higher affinity binders in display systems that might bind alternative and 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 and evaluated for binding activity to cells transfected to overproduce the target of interest. Of course the latter strategy requires high enough affinity and avidity of the IgG to detect the interaction by flow cytometry. The discovery rate was high; hits were obtained against seven of nine targets, with from 4 to 30 hits found per target.

Optimization. Most antibody optimization technologies rely on display techniques and involve producing libraries of randomized amino acids in the CDRs, then selection for the higher affinity variants. Multiple reviews have discussed affinity maturation techniques and strategies (Bradbury *et al*., 2001; Filpula, 2007; Levin and Weiss, 2006; Lonberg, 2008; Rothe *et al*., 2006). The addressed format allows each individual

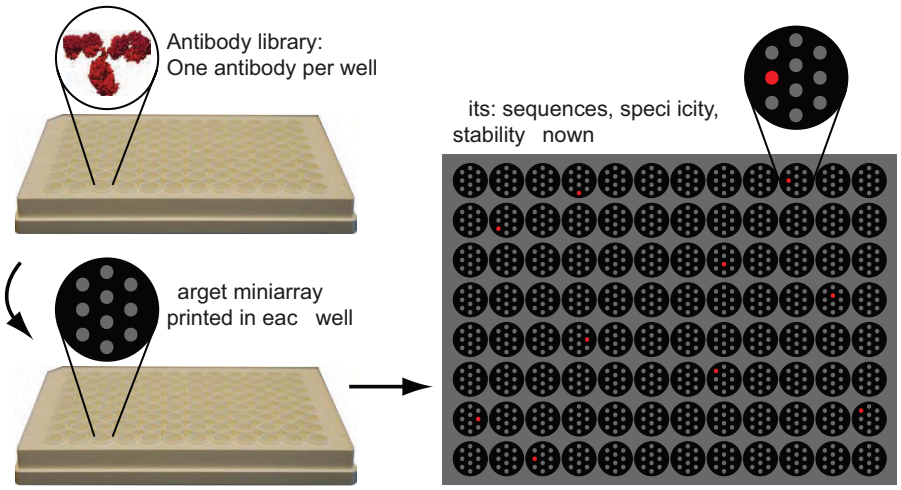

Figure 3. Scheme for addressed library screening using the MSD system. The antibody (or Fab library) is incubated with MSD plates that have target antigens printed in a miniarray format in the bottom of each well. Each well is identical to one another except that each will be incubated with a different library member. A ruthenium tagged secondary antibody is used for detection through electrochemiluminescence. This reaction will induce light output at an individual spot if it is bound by an antibody. The binding reaction is rapid (minutes to hours) and provides immediate information on specificity. "Hits" can be identified and biochemical and genetic properties retrieved from the LIMS system (Figure 2) for SAR.

mutation to be evaluated independently; thus "up" mutations can be distinguished from "down" mutations, and the magnitude of improvement is revealed compared to all other mutants. 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. With regards to affinity maturation, the ability to achieve significant improvements is quite substantial when the starting antibody has a $K_d > 1$ μM. Often one or two mutations can improve the off-rate, producing binders in the nM range. Interestingly, substantial improvements in affinity are often seen with a simple alanine scan. In this regard, it appears that replacing bulky residues with small residues can often "remove a negative" and enhance antigen binding. During the discovery process multiple "hits" for a given target can be analyzed for similarities or differences in their sequence characteristics, and can also be searched for related sequences in the LIMS. 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 (Mao *et al*., 2010). If patterns are not recognizable for all 6 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 residues, then each individual mutant screened. Iterative processes of library creation, mutation screening, combining useful mutations, and further screening can achieve subnanomolar K_ds.

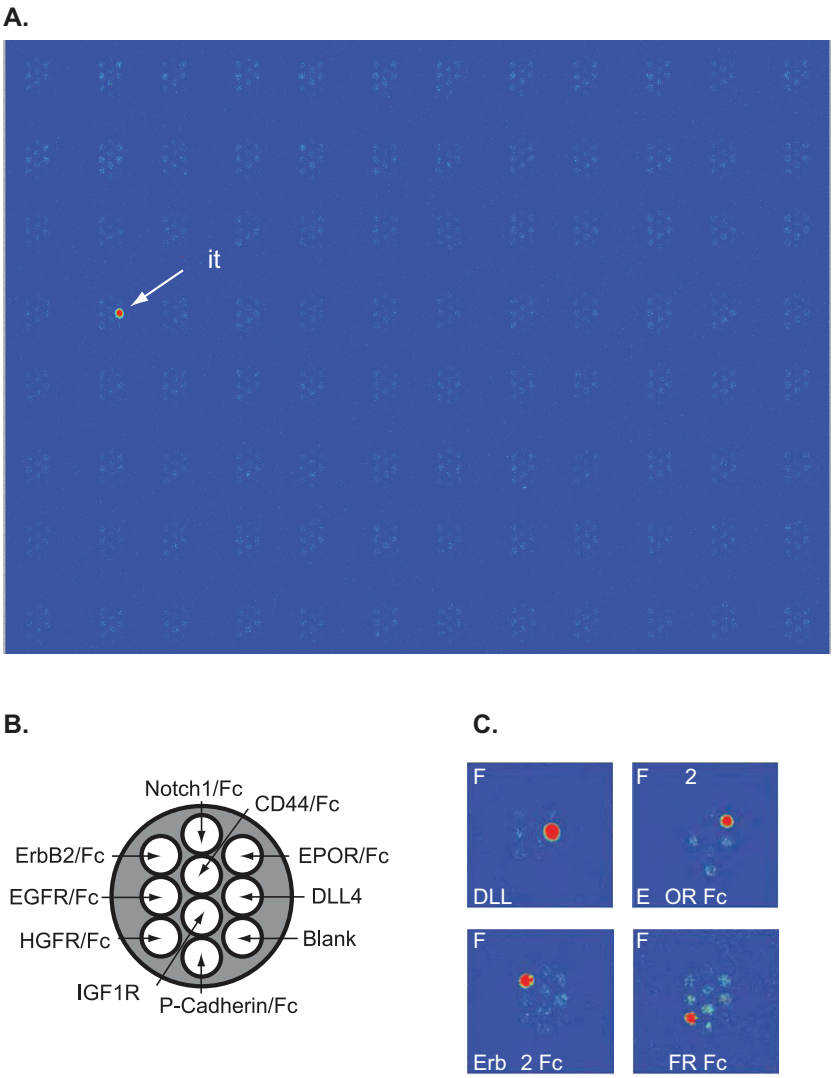

Figure 4. Data from screening an Fab library. (A) A heat map of a 96-well plate is shown with a single hit identified in well B4 (arrow). (B) Schematic of the antigens printed in the bottom of the plate. The "hit" from A is against DLL4, with no detectible binding to the other targets printed in the same well. (C) Examples of hits against four other targets.

Potential Pitfalls. Some pitfalls we have encountered in arrayed protein screening relate to the state of the antigen. For multiplexed screening on protein, purified antigen should be well characterized and not have alternative or aggregated states. We have also found "hits" to affinity tags such as Fc receptors that were used to purify the antigen. Additionally, a target that is naturally a multimer like a trimer (for example TNF-receptor superfamily members) may not exist in a native state when forced to dimerize through Fc receptors. Such molecules can give false positive signals upon MSD screening. These issues are not specific to arrayed libraries and would be issues

for any discovery technology. Of course screening on cells could obviate the need for purified protein, and discovery using recently developed high throughput flow cytometry systems (Sklar *et al*., 2007) could potentially allow this type of screen in a multiplexed format. Fabrus has now developed cell based screens and has identified multiple hits to three out of three targets screened thus far (unpublished data).

Costs. The costs associated with producing display based libraries are variable, and dependent on the library design goals. Addressed 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 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 $1M. These costs can be significantly decreased, for example, 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. The costs of production in the Piccolo unit were substantial and included plasticware, media, inducer, and purification resin costs. Overall the total cost of the 10,000 member Fabrus Fab library was approximately $15 per antibody fragment. 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 spatially addressed 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 spatially addressed 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. Because of this, 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 or failure of a discovery campaign. These latter enabling features remain to be proven.

Conclusions

Spatially addressed libraries for protein drug discovery are now a reality. Although the construction and maintainance of these libraries may take considerable resources, the ability to achieve extremely rapid and multiplexed discovery and the potential for discovery in cell based systems in the absence of purified target could open up new areas for biologic discovery. Extremely important targets like GPCRs, ion channels, and other multispanning membrane proteins could be accessible through cell based assays screened against a spatially addressed library. Similarly, mechanisms of action

not typically attributed to antibodies could be discovered and optimized. The future could potentially hold antibody drugs with modulator, partial agonist or antagonist functions, all of which could be tunable for optimal drug function against traditionally difficult targets.

Acknowledgements

This article is dedicated to William D. Huse, MD, PhD. (1954-2011), a Fabrus Board Member, founder of Applied Molecular Evolution (acquired by Lilly), and a pioneer in antibody discovery and engineering technology.

References

Bachmann, M.F., Kundig, T.M., Kalberer, C.P., Hengartner, H. & Zinkernagel, R.M. How many specific B cells are needed to protect against a virus? *J Immunol* **152**, 4235-4241 (1994).

Beerli, R.R. et al. Isolation of human monoclonal antibodies by mammalian cell display. *Proc Natl Acad Sci U S A* **105**, 14336-14341 (2008).

Bobrzynski, T. et al. A high-affinity natural autoantibody from human cord blood defines a physiologically relevant epitope on the FcepsilonRIalpha. *J Immunol* **175**, 6589-6596 (2005).

Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **15**, 553-557 (1997).

Boulianne, G.L., Hozumi, N. & Shulman, M.J. Production of functional chimaeric mouse/human antibody. *Nature* **312**, 643-646 (1984).

Bradbury, A.R., Sidhu, S., Dubel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nat Biotechnol* **29**, 245-254 (2011).

Chandonia, J.M. & Brenner, S.E. The impact of structural genomics: expectations and outcomes. *Science* **311**, 347-351 (2006).

Chiang, Y.L., Sheng-Dong, R., Brow, M.A. & Larrick, J.W. Direct cDNA cloning of the rearranged immunoglobulin variable region. *Biotechniques* **7**, 360-366 (1989).

Chothia, C. & Lesk, A.M. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* **196**, 901-917 (1987).

Cohn, M. & Langman, R.E. The protection: the unit of humoral immunity selected by evolution. *Immunol Rev* **115**, 11-147 (1990).

Corbett, S.J., Tomlinson, I.M., Sonnhammer, E.L., Buck, D. & Winter, G. Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. *J Mol Biol* **270**, 587-597 (1997).

Filpula, D. Antibody engineering and modification technologies. *Biomol Eng* **24**, 201-215 (2007).

Georgiou, G. et al. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol* **15**, 29-34 (1997).

Green, L.L. et al. Antigen-specific human monoclonal antibodies from mice engineered with human Ig heavy and light chain YACs. *Nat Genet* **7**, 13-21 (1994).

Gonzalez, R. et al. Screening the mammalian extracellular proteome for regulators of embryonic human stem cell pluripotency. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 3552-3557 (2010).

Gubler, U. & Hoffman, B.J. A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-269 (1983).

Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H.R. & Pluckthun, A. Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proc Natl Acad Sci U S A* **95**, 14130-14135 (1998).

Ho, M., Nagata, S. & Pastan, I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc Natl Acad Sci U S A* **103**, 9637-9642 (2006).

Huse, W.D. et al. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **246**, 1275-1281 (1989).

Joachimiak, A. High-throughput crystallography for structural genomics. *Curr Opin Struct Biol* **19**, 573-584 (2009).

Jungbluth, A.A. et al. A monoclonal antibody recognizing human cancers with amplification/overexpression of the human epidermal growth factor receptor. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 639-644 (2003).

Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497 (1975).

Kung, P., Goldstein, G., Reinherz, E.L. & Schlossman, S.F. Monoclonal antibodies defining distinctive human T cell surface antigens. *Science* **206**, 347-349 (1979).

Lacy, S.E. et al. The potency of erythropoietin-mimic antibodies correlates inversely with affinity. *J Immunol* **181**, 1282-1287 (2008).

Larrick, J.W. et al. Rapid cloning of rearranged immunoglobulin genes from human hybridoma cells using mixed primers and the polymerase chain reaction. *Biochem Biophys Res Commun* **160**, 1250-1256 (1989).

Leonard, B., Sharma, V. & Smider, V. Co-expression of antibody fab heavy and light chain genes from separate evolved compatible replicons in E. coli. *J Immunol Methods* **317**, 56-63 (2006).

Levin, A.M. & Weiss, G.A. Optimizing the affinity and specificity of proteins with molecular display. *Mol Biosyst* **2**, 49-57 (2006).

Levitan, B. Stochastic modeling and optimization of phage display. *J Mol Biol* **277**, 893-916 (1998).

Lewis, S. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv. Immunol.* **56**, 27-150 (1994b).

Lonberg, N. Fully human antibodies from transgenic mouse and phage display platforms. *Curr Opin Immunol* **20**, 450-459 (2008).

Lonberg, N. et al. Antigen-specific human antibodies from mice comprising four distinct genetic modifications. *Nature* **368**, 856-859 (1994).

Mao, H. et al. Spatially addressed combinatorial protein libraries for recombinant antibody discovery and optimization. *Nature biotechnology* **28**, 1195-1202 (2010).

Matsuda, F. et al. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med* **188**, 2151-2162 (1998).

McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554 (1990).

Miller, D.J. & Rodriguez, M. A monoclonal autoantibody that promotes central nervous system remyelination in a model of multiple sclerosis is a natural autoantibody encoded by germline immunoglobulin genes. *J Immunol* **154**, 2460-2469 (1995).

Morrison, S.L., Johnson, M.J., Herzenberg, L.A. & Oi, V.T. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6851-6855 (1984).

Nelson, A.L., Dhimolea, E. & Reichert, J.M. Development trends for human monoclonal antibody therapeutics. *Nat Rev Drug Discov* **9**, 767-774.

Neuberger, M.S. et al. A hapten-specific chimaeric IgE antibody with human physiological effector function. *Nature* **314**, 268-270 (1985).

Okayama, H. & Berg, P. High-efficiency cloning of full-length cDNA. *Molecular and cellular biology* **2**, 161-170 (1982).

Queen, C. et al. A humanized antibody that binds to the interleukin 2 receptor. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 10029-10033 (1989).

Roberts, R.W. & Szostak, J.W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A* **94**, 12297-12302 (1997).

Rothe, A., Hosse, R.J. & Power, B.E. Ribosome display for improved biotherapeutic molecules. *Expert Opin Biol Ther* **6**, 177-187 (2006).

Sklar, L.A., Carter, M.B. & Edwards, B.S. Flow cytometry for drug discovery, receptor pharmacology and high-throughput screening. *Curr Opin Pharmacol* **7**, 527-534 (2007).

Smider, V. & Chu, G. The end-joining reaction in V(D)J recombination. *Semin Immunol* **9**, 189-197 (1997).

Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317 (1985).

Smith, E.S., Shi, S. & Zauderer, M. Construction of cDNA libraries in vaccinia virus. *Methods Mol Biol* **269**, 65-76 (2004).

Wollerton, M.C., Wales, R., Bullock, J.A., Hudson, I.R. & Beggs, M. Automation and optimization of protein expression and purification on a novel robotic platform. *JALA* **11**, 291-303 (2006).

Xu, J.L. & Davis, M.M. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13**, 37-45 (2000).

Zhou, C., Jacobsen, F.W., Cai, L., Chen, Q. & Shen, W.D. Development of a novel mammalian cell surface antibody display platform. *MAbs* **2** (2010).