



# Mating antibody phage display with proteomics

Michael Hust and Stefan Dübel

Institute for Biochemistry and Biotechnology, Department Biotechnology, Technical University of Braunschweig, Spielmannstrasse 7, 38106 Braunschweig, Germany

**Antibodies are central to mastering the next main task of 'postgenomic' research: that is, to understand gene function by analysing the proteome. This analysis of more than 90 000 human gene products requires high-throughput methods for antibody generation. In contrast to animal-based methods, *in vitro* selection systems using recombinant antibodies offer this capability. Antibody phage display, the most commonly used contemporary system, has yielded hundreds of antibodies for therapy, research and diagnostics. Numerous cloning and mutagenesis strategies have been developed to create, preserve and exploit maximal antibody diversity, with libraries of different antibody formats constructed from different genetic sources. Evaluating the spectrum of phage-display antibody libraries indicates that the current focus on generating human therapeutic antibodies has yielded methods that are not readily applicable to postgenomic research. A streamlined process must be defined by closely integrating antigen logistics with library and antibody formats, panning procedures and robotic handling. Only by optimizing the interplay of biology, bioinformatics and technology can an affordable high-throughput process be created to meet the demands of proteome analysis and microarray technology.**

With the sequencing of the human genome close to completion [1,2], the research focus has shifted to the analysis of gene products. The human genome encodes about 31 000 genes [1] or more [3]. Through alternative mRNA splicing and post-translational modifications, including glycosylation and phosphorylation among others, the number of different proteins of the human proteome is thought to easily exceed 90 000 [4]. Through different interpretations of the same genetic content, marked morphological variations such as those between caterpillar and butterfly are achieved. For each gene, therefore, tools are needed to investigate the amount, localization and function of its products, because whereas the genotype defines the species of an organism, the proteome defines the phenotype.

Two-dimensional electrophoresis coupled with mass spectroscopy has provided the first glimpse of proteome variations either for different phenotypes, tissues and cells, or in response to external factors or differentiation

processes. The throughput of this serial method is limited, however, and the parallel analysis of large numbers of proteins will be desirable to handle the foreseeable demand. In genomics, microarrays are well established for this task, enabling the parallel analysis of whole genomes [5]. The current bottleneck in developing microarrays for proteome analysis stems from the fact that generating a molecule that binds specifically to a given protein is far more laborious than is designing an oligomer to identify its encoding mRNA. Antibodies are the protein-binding agents of choice because the antibody pool in vertebrates has been designed by evolution to bind to any possible protein with high affinity and specificity.

Numerous methods have been developed to obtain antibodies, as recently reviewed by Bradbury *et al.* [6,7]; however, only the *in vitro* methods have the potential to meet the requirements of the highly parallel processes needed for proteomics [8]. Furthermore, in contrast to animal-based methods, the biochemical milieu during the selection step can be controlled completely, allowing the *in vitro* selection of conformation-specific antibodies [9] or selection in the presence of competitive agents to fine-tune specificity. In addition, because no immune system is involved, antibodies to toxic compounds, lethal pathogens and highly conserved antigens can be obtained by *in vitro* methods.

Among the various strategies developed for the *in vitro* selection of antibodies, phage display has proved to be a robust technology and has yielded most of the recombinant antibodies generated so far. In the first decade of the development of antibody phage display, technology has been driven, mainly by companies, towards methods to obtain therapeutic products. As a consequence, the process has been optimized for factors that are not really relevant to generating research agents (Table 1). In this article, we look at the available technologies from the viewpoint of their applicability as tools for proteome research. A comprehensive analysis of the libraries, cloning strategies to generate library diversity, molecular formats and panning methods shows that a new model for both antibody library formats and the selection process is needed to meet the requirements of proteomics.

## Antibody phage display

Many different methods have been described for antibody selection, including yeast, bacterial or ribosomal display, and picking from arrays (reviewed in Ref. [6]). Of all of

Corresponding author: Stefan Dübel (s.duebel@tu-bs.de).

**Table 1. Comparison of the requirements and properties of phage-display antibody selection systems for developing therapeutic antibodies and research applications**

| Application        | Therapeutics                                                                                                                           | Proteomics                                                                                                                                           |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| Target proteins    | Few<br>Known and available<br>Well characterized<br>Production cost not an issue                                                       | Large number<br>Usually not known and available<br>Properties unknown<br>Production cost crucial                                                     |
| Selection process  | Step number not an issue<br>Selection cost not crucial<br>Optimized for lowest failure rate<br>Re-engineering after selection required | Minimal step number<br>Optimized for low selection cost<br>Several percent failure rate not crucial<br>Re-engineering after selection not affordable |
| Product (antibody) | Optimized for low immunogenicity<br>Optimized for good pharmacokinetics<br>Optimized for binding its antigen                           | Optimized for robustness<br>Optimized for compatibility to established assays<br>No individual optimization possible                                 |

these methods, however, antibody phage display has developed into the most robust, versatile and thus widespread selection method during the past decade. In consequence, the obvious choice for a proteome-wide antibody project is to use such a well-established method that has already proved its adaptability to automation.

The groundbreaking work on peptides by G. P. Smith [10] paved the way for antibody phage display [11–14]. In this method, the genotype and phenotype of a polypeptide were linked by fusing antibody gene fragments to a gene encoding the minor coat protein of the filamentous bacteriophage M13. This resulted in the expression of an antibody fusion protein on the surface of phage, thereby allowing the affinity purification of antibody-coding genes by antigen binding. To date, ‘single-pot’ antibody libraries (see below) with calculated diversities of up to  $10^{11}$  independent clones have been assembled as molecular repertoires for phage-display selections [15], and various methods have been established for ‘panning’ the desired specificity from vast repertoires.

Various types of phage-display library have been constructed. The amplification of variable (V) genes isolated from immunoglobulin- $\gamma$  (IgG)-secreting plasma cells from immunized donors results in immune libraries [16]. Selections from immune libraries have yielded antibody fragments with monovalent dissociation constants in the nanomolar range. Immune libraries are typically created and used in medical research to obtain an antibody specific for a specific antigen, such as an infectious pathogen, and therefore would not be the source of choice for selecting a larger number of antibodies with different specificities. Naive, semi-synthetic and synthetic libraries have been subsumed as single-pot libraries, because they are designed to isolate antibody fragments that bind to every possible antigen.

A clear correlation is seen between the size of the repertoire and the affinities of the antibodies isolated from it. Antibodies with a micromolar affinity have been isolated from a single-pot library consisting of about  $10^7$  clones, whereas antibody fragments with nanomolar affinities have been obtained from a repertoire of  $10^9$  clones [17].

According to the source of antibody genes, single-pot libraries (Table 2) can be naive, semi-synthetic, synthetic and fully synthetic. Naive libraries are constructed from rearranged V genes from B cells (IgM) of non-immunized donors. An example of this library type is the naive human

Fab library constructed by de Haardt *et al.* [18], which yielded antibodies with affinities of up to  $2.7 \times 10^{-9}$  M. Semi-synthetic libraries are derived from unrearranged V genes from pre-B cells (germline cells) or from a single antibody framework with genetically randomized complementarity-determining region 3 (CDR3) regions, such as that described by Pini *et al.* [19]. The antibodies obtained from this library had affinities of  $10^{-8}$  M to  $10^{-9}$  M, and one single-chain antibody variable region fragment (scFv) had a dissociation constant of  $5 \times 10^{-11}$  M. The fully synthetic libraries have a human framework with randomly integrated CDR cassettes [20]. Antibodies selected from this type of library have affinities of  $10^{-6}$  M to  $10^{-11}$  M [21]. It is evident that naive or synthetic libraries and their intermediates are useful sources for selecting antibodies for proteome research.

### Generation of phage-display libraries

Various methods have been used to clone the genetic diversity of antibody repertoires. Immune libraries are usually constructed by a two-step cloning or assembly PCR method (see below) after mRNA isolation and cDNA preparation from the desired cell type. Naive libraries are constructed by two or three cloning steps. In the two-step strategy, the amplified repertoire of light chain genes is cloned into the phage-display vector in the first step, because the heavy chain contributes more to diversity owing to its highly variable CDRH3 region. In the second step, the heavy chain repertoire is cloned into the amplified light chain library [22–24]. In the three-step strategy, separate heavy and light chain libraries are engineered, and the repertoire of the  $V_H$  library is then excised and cloned into the phage-display vector containing the  $V_L$  library [18].

Another common method that has been used to clone naive [25,26], immune [16] or hybridoma [27] scFv phage-display libraries is assembly PCR.  $V_H$  and  $V_L$  genes are amplified separately, connected in a subsequent PCR step and then cloned into the vector. Assembly PCR is frequently combined with randomization of the CDR3 regions, leading to semi-synthetic libraries. To achieve this, oligonucleotides encoding various CDR3 and joining (J) gene segments have been used to amplify V gene segments of human germ lines [28]. CDRH3 is a main source of sequence variety [29]. Hoogenboom and Winter [30] and Nissim *et al.* [31] used degenerated CDRH3 primers to produce a semi-synthetic heavy chain repertoire derived

Table 2. Human and mouse 'single-pot' phage-display libraries

| Library vector            | Library type                                                                                 | Antibody type                    | Library cloning strategy                                        | Library size                                    | Refs                                             |
|---------------------------|----------------------------------------------------------------------------------------------|----------------------------------|-----------------------------------------------------------------|-------------------------------------------------|--------------------------------------------------|
| pHEN1                     | Naive                                                                                        | scFv                             | Assembly PCR                                                    | $2 \times 10^5/2 \times 10^6$                   | [69]                                             |
| pCANTAB3his               | Naive <sup>a</sup>                                                                           | scFv                             | Assembly PCR                                                    | $1.2 \times 10^6$                               | [25]                                             |
| pSEX81                    | Naive                                                                                        | ScFv <sup>b</sup>                | Four-step cloning                                               | $1.6 \times 10^7/1.8 \times 10^7/4 \times 10^7$ | [70]                                             |
| pSEX81                    | Naive                                                                                        | scFv <sup>b</sup>                | Two-step cloning                                                | $4 \times 10^7$                                 | [71]                                             |
| pCANTAB 5                 | Naive <sup>a</sup>                                                                           | scFv                             | Assembly PCR                                                    | $4.5 \times 10^7$                               | [72]                                             |
| pHEN1                     | Naive                                                                                        | scFv                             | Assembly PCR                                                    | $10^7-10^8$                                     | [55]                                             |
| pFAB60                    | Naive <sup>a</sup>                                                                           | Fab                              | Two-step cloning                                                | $> 10^8$                                        | [22]                                             |
| fdTet                     | Naive                                                                                        | scFv                             | Recloning of a naive library <sup>c</sup>                       | $5 \times 10^8$                                 | [73]                                             |
| pSEX81                    | Naive                                                                                        | scFv <sup>b</sup>                | Two-step cloning                                                | $4 \times 10^9$                                 | [24]                                             |
| pHEN1                     | Naive                                                                                        | scFv                             | Assembly PCR                                                    | $6.7 \times 10^9$                               | [36]                                             |
| pCANTAB 6                 | Naive                                                                                        | scFv                             | Assembly PCR                                                    | $1.4 \times 10^{10}$                            | [26]                                             |
| pCES1                     | Naive                                                                                        | Fab                              | Three-step cloning (L chain, V <sub>H</sub> )                   | $3.7 \times 10^{10}$                            | [18]                                             |
| pDAN5                     | Naive                                                                                        | scFv                             | Cre- <i>lox</i>                                                 | $3 \times 10^{11}$                              | [15]                                             |
| pAP-III <sub>6</sub> scFv | Naive                                                                                        | scFv                             | Assembly PCR                                                    | ND <sup>d</sup>                                 | [74]                                             |
| pEXmide5                  | Semi-synthetic (germline V <sub>H</sub> DP47 and V <sub>L</sub> DPL3 framework)              | scFv                             | Assembly PCR, CDR shuffling                                     | $9 \times 10^6$                                 | [34]                                             |
| pHEN1-V <sub>λ</sub> 3    | Semi-synthetic (V <sub>λ</sub> 3 anti-BSA Ab light chain)                                    | scFv                             | PCR with random CDR H3 primers                                  | $10^7$                                          | [30]                                             |
| pScUAGDcp3                | Semi-synthetic                                                                               | scFv connected to C <sub>K</sub> | Three-step cloning with random CDR3 primers                     | $1.7 \times 10^7$                               | [28]                                             |
| pDN322                    | Semi-synthetic (anti-AMCV CP ab framework)                                                   | scFv                             | Random CDR3 primer, assembly PCR                                | $3.75 \times 10^7$                              | [33]                                             |
| pComb3                    | Semi-synthetic (anti-tetanus Ab framework <sup>e</sup> )                                     | Fab                              | PCR with random CDR H3 primers                                  | $5 \times 10^7$                                 | [32]                                             |
| pComb3                    | Semi-synthetic (anti-tetanus Ab framework <sup>e</sup> )                                     | Fab                              | PCR with random CDR H3 primers                                  | $> 10^8$                                        | [75]                                             |
| pHEN1-V <sub>λ</sub> 3    | Semi-synthetic (V <sub>λ</sub> 3 anti-BSA Ab light chain)                                    | scFv                             | PCR with random CDR H3 primers                                  | $> 10^8$                                        | [31]                                             |
| pIT2 (Tomlinson I/J)      | Semi-synthetic (3 V <sub>H</sub> and 4 V <sub>K</sub> genes)                                 | scFv                             | PCR with random CDR2 and CDR3 primers                           | $1.47 \times 10^8/1.37 \times 10^8$             | [76]                                             |
| pAALFab                   | Semi-synthetic (anti-hen egg white lysozyme Ab framework)                                    | Fab                              | PCR with random CDR primers, assembly PCR                       | $2 \times 10^8$                                 | [77]                                             |
| pLG18                     | Semi-synthetic (anti-HER2 Ab framework)                                                      | Fab                              | PCR with random CDR primers, two-step cloning                   | $2-3 \times 10^8$                               | [78]                                             |
| pDN322                    | Semi-synthetic (V <sub>H</sub> DP47 and V <sub>K</sub> DPK22 V genes)                        | scFv                             | Random CDR3 primer, assembly PCR                                | $3 \times 10^8$                                 | [19]                                             |
| pHEN1-V <sub>K</sub> 3    | Semi-synthetic (V <sub>H</sub> )/naive (V <sub>L</sub> )                                     | scFv                             | Three-step cloning, PCR with random CDR H3 primers              | $3.6 \times 10^8$                               | [79]                                             |
| Pdnek (ETH2)              | Semi-synthetic (V <sub>H</sub> DP47, V <sub>λ</sub> DPL16 and V <sub>K</sub> DPK 22 V genes) | scFv                             | Random CDR3 primer, assembly PCR                                | $5 \times 10^8$                                 | [80]                                             |
| pHEN2 (Griffin 1)         | Semi-synthetic                                                                               | scFv                             | Recloning of the <i>lox</i> library in scFv format <sup>f</sup> | $1.2 \times 10^9$                               | Centre for Protein Engineering <sup>g</sup> [35] |
| pFAB5c-His (n-CoDeR)      | Semi-synthetic (germline V <sub>H</sub> DP47 and V <sub>λ</sub> DPL3 framework)              | scFv                             | Assembly PCR, CDR shuffling                                     | $2 \times 10^9$                                 |                                                  |
| fdDOG-2lox, pUC19-2lox    | Semi-synthetic                                                                               | Fab                              | PCR with random CDR3 primers, Cre- <i>lox</i>                   | $6.5 \times 10^{10}$                            | [38]                                             |
| pMorph series (HuCAL)     | Synthetic                                                                                    | scFv                             | Two-step cloning plus two-step CDR3 replacement                 | $2 \times 10^9$                                 | [21]                                             |

<sup>a</sup>Mouse library.<sup>b</sup>with N-terminus of C<sub>H</sub>1 and C<sub>L</sub>.<sup>c</sup>Sheets *et al.* [36].<sup>d</sup>Not determined.<sup>e</sup>Persson *et al.* [81].<sup>f</sup>Griffiths *et al.* [38].<sup>g</sup><http://www.mrc-cpe.cam.ac.uk>.

from human V gene germline segments and combined it with a bovine serum albumin (BSA) antibody light chain.

In some cases, the framework of a well-known antibody has been used as a scaffold on which to integrate new, randomly created CDRH3 and CDRL3 regions [32,33]. Jirholt *et al.* [34] and Söderlind *et al.* [35] amplified all CDRs derived from B cells and then shuffled them into one antibody framework in an assembly PCR step. As an example of a completely synthetic library, Knappik *et al.* [21] used seven different V<sub>H</sub> and V<sub>L</sub> germline frameworks combined with six synthetically created CDR cassettes.

The construction of large naive and semi-synthetic libraries [18,24,26,35,36] requires considerable effort to funnel the created genetic diversity through the bottleneck of *Escherichia coli* transformation. The Cre-*lox* or λ phage recombination system has been used to move the

diversity-potentiating step of random V<sub>H</sub> and V<sub>L</sub> combination beyond the transformation bottleneck [35,37–39]; however, library sizes above  $10^{10}$  have now been accomplished by conventional transformation, rendering most of these complicated methods unnecessary, especially as they can result in decreased genetic stability. A notable exception is the use of a genomically integrated *CRE* recombinase gene [15], which is expected to solve the instability issue and to aid the generation of libraries with complexities above the limit achievable by conventional assembly.

In summary, antibodies with nanomolar affinities can be selected from both naive libraries and synthetic libraries (which can be assembled by either cloning or PCR) if the preservation of molecular complexity is carefully controlled at every step of library construction and if library sizes of around or above  $10^{10}$  are reached.

## Antibody formats

An essential requirement of applications to proteome research is a robust molecular format that is compatible with current assays and that does not require major adaptations of established protocols. Ideally, a reagent should be fully interchangeable with commercially available IgG used for western blotting, immunoprecipitation, immunohistochemistry, enzyme-linked immunosorbent assay (ELISA) and antibody microarrays. This issue primarily concerns stability and affinity, but it also concerns compatibility with detection systems, which usually are secondary antisera labelled with an enzyme or a fluorescent dye. As throughput is crucial, the optimal format should be produced directly by the library selection process without the need for subsequent cloning steps.

A straightforward use of IgG is not possible, because the expression of complete IgG molecules in *E. coli* has been achieved only in exceptional cases [40]. Consequently, phage-display systems use scFv or Fab fragments. In Fabs, the light chain and the  $V_H/CH_1$  fragment of the heavy chain are connected by a disulfide bond. By contrast, scFvs are composed of  $V_H$  and  $V_L$  connected by a 15–25 amino acid linker [41–43].

### scFv fragments

Small antibody fragments such as Fv and scFv can be easily produced in *E. coli*. The yield of functional Fv fragments expressed in *E. coli* is higher than that of the corresponding Fabs, because Fabs have a slower folding rate [44,45]. In one study, fragment stability during long-term storage was found to be much better for Fabs than for scFvs. After six months, the functionality of scFvs stored at 4 °C was reduced by 50%, whereas Fabs showed no significant loss of functionality after 1 year [46].

Soluble scFvs rely on the 'kinetic' stabilization by the linker, which does not prevent  $V_H-V_L$  assembly.  $V_H-V_L$  dissociation constants vary between  $10^{-5}$  and  $10^{-8}$  M [47]; as a result, although some scFv fragments have a stability comparable to IgG, many scFvs are significantly less-rigid molecules. In our laboratory, we have observed that scFv half-lives in solution at 4 °C vary from a few days to more than a year, illustrating the large individual variability of this format. Furthermore, different scFvs showed large variations in stability on microarray surfaces [48,49]. scFvs also need a peptide tag for their detection, which usually results in additional incubation steps as compared with IgG, requiring additional adaptive work.

In summary, despite being a commonly used format for the selection process and therefore being used in most of the current libraries (Table 1), scFv fragments are far from being the optimal agent for proteome research.

### Fab fragments

In contrast to scFvs, Fab fragments are structurally stabilized by a disulfide bond between the fd fragment and the light chain. The use of Fab fragments, which can be generated enzymatically from IgG, has been established for decades. Because they contain the whole light chain and one constant domain of the heavy chain, Fabs can be detected by a wide range of available secondary antibodies. Most significantly, these reagents are, in most

cases, identical to those used for IgG detection, rendering major adaptations of assay protocols unnecessary.

The slight disadvantages of the Fab fragments with respect to library creation and protein assembly are counterbalanced by superior stability [6] and by direct compatibility with the end-users assays without the need for a re-engineering step. In consequence, Fabs would be the antibody format of choice for research and proteome analysis.

### Fragment oligomerization

Both scFv and Fab fragments are monovalent molecules, in contrast to bivalent IgGs. As is the case in many assays, such as western blot, ELISA and immunohistochemistry among others, antigen is used at a density enabling an IgG molecule to bind by both arms, thereby leading to a significant increase in apparent affinity. This 'avidity' effect can result in an increase in affinity of several orders of magnitude [50,51]. Animal-derived IgGs are bivalent for antigen, and thus studies using such IgGs benefit from this effect. To give comparable performance, therefore, antibodies generated *in vitro* have to be adapted.

Numerous methods have been generated to provide oligomers of antibody fragments. Most of them use additional protein domains with dimerization or tetramerization properties (reviewed in Refs [52,53]) and/or are based on the scFv format [54]. Both strategies require re-engineering of the antibody fragments after selection, thereby limiting their use in a high-throughput setup. Thus, a system is needed that allows dimerization in a step that is separate from selection and production.

The method of choice is the addition of a short peptide tag that does not interfere with phage display or production in *E. coli*, but that facilitates binding to an oligovalent protein to form complexes consisting of 2–4 Fab fragments. Well-established tags are available, including the c-Myc epitope [55] and the Flag tag [19]. To avoid the necessity of adding different tags for different tasks, however, it would be preferable to use a peptide tag that facilitates purification. Variants such as the Strep-Tag II [56] or biotinylation acceptor peptides [57] that bind to streptavidin offer both purification and detection. Ideally, by using a properly designed Fab fragment with a peptide tag, both a bivalent reagent can be created and a solid-phase immobilization (e.g., on a microarray) can be achieved.

### Antigen choice

Many antibodies have been derived by immunization or by panning on purified or recombinantly expressed antigen. For large-scale proteomic research, the large number of antigens – each requiring individual characterization and purification methods – creates logistic and cost barriers. From a proteomics perspective, therefore, despite being superior with respect to antigen quality (e.g. folding and posttranslational modifications), the use of antigens purified from naive material is limited to niche applications.

Two other possibilities for producing antigens remain: synthesis from cDNA libraries and the use of synthetic peptides. Both methods have been thoroughly validated by the production of countless monoclonal and polyclonal

antibodies in animals. For generating recombinant antigen, large cDNA collections [58,59] and rapid *in vitro* translation and transcription systems that produce milligrams of protein within 2 days [60] are available. Even so, with respect to throughput the input logistics needed to generate recombinant antigens easily equals or exceeds the efforts required for the antibody selection process itself. In this case, panning on a set of peptides predicted to be accessible at the surface of the protein [61] offers a shortcut, enabling the design of a robotic panning process requiring nothing more than the input of sequence data. Without doubt, not all antibodies directed against peptides will bind to the native form of the protein; however, countless examples of antibodies generated against peptides in rabbits and mice have been described and prove the validity of this approach.

The observed limitations of the peptide approach might well be overcome by the possibility of using many different peptides per antigen in a high-throughput screening (HTS) approach. In addition, current developments in bioinformatics might improve the rate of epitope prediction. Furthermore, by using, for example, phosphorylated variants, peptides can facilitate the search for posttranslational modifications. Finally, by using the recently developed methods of synthesizing peptides in array formats on solid phase [62] for phage panning [63], a significant reduction in numbers of experimental steps might yield an improvement of throughput that by far exceeds current capabilities.

### High throughput

Considerable progress has been made in adapting phage-display selections to robotics (reviewed in Refs [7,64,65]). In particular, multipin selection methods have been shown to combine efficiency with low consumption of the precious library. Methods for automated analysis by phage ELISA, clone picking and sequencing are established.

From a proteomics perspective, a minimal number of process steps should lead from a selected phage to a soluble antibody molecule. Vector systems that allow a switch from pIII fusions to soluble fragments without the necessity for DNA ligation would be preferable and can be achieved, for example, by using supE technology [55]. Recent developments in antibody phage display will reduce the number of experimental steps needed for the panning process itself. For example, by using 'hyperphage' instead of the standard M13 helperphage, antibody display on phage is improved by two orders of magnitude, resulting in improved panning efficiency [66].

The validation of specificity requires considerable effort in any antibody generation project. In an affordable HTS setup optimized for proteomics, obtaining a control for specificity that is independent of circular reasoning (i.e., that does not use the antigens used for panning) within the process is close to impossible. Because the resulting sets of antibodies will be used in different assays, including immunofluorescence, fluorescence-activated cell sorting, microarrays and western blots among others, the proof that they bind naive or fixed forms of the target proteins can be obtained only from their use. This problem is independent of antibody format and the selection process,

and affects conventional monoclonal and polyclonal technologies. It could be tackled by establishing thoroughly defined validation standards and a well-structured data exchange network among producers and users. Parallelized binding analysis on microarrays of cDNA-derived expression products could be employed to control the specificities of antibodies generated on peptides, and to check for cross-reactivities.

In the near future, antibody fragments might well be replaced by one of the many alternative scaffolds [49,67]. For example, an automated process using a combination of puromycin display with a robust single domain fibronectin scaffold library [68] has already been established. So far, however, the alternative scaffold systems have not reached the maturity of antibody phage display. In addition, in contrast to the 'conservative' antibody technology that every laboratory has used for years, these alternative systems require additional efforts on the end-user's side. Further work is necessary – in particular, a comparison of the properties of various scaffolds with those of IgG, scFv and Fab fragments with respect to microarray production, storage and application.

### Conclusion

In summary, the separate technological building blocks for antibody generation on a genomic scale are established, but they remain to be integrated into a streamlined process that takes advantage of every possible reduction in the number of experimental steps to meet the needs of proteome research at a reasonable cost.

### References

- 1 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- 2 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
- 3 Wright, F.A. *et al.* A draft annotation and overview of the human genome. *Genome Biol.* (in press).
- 4 Harrison, P.M. *et al.* (2002) A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* 30, 1083–1090
- 5 Lockhart, D.J. and Winzler, E.A. (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836
- 6 Bradbury, A. *et al.* (2003) Antibody in proteomics I: generating antibodies. *Trends Biotechnol.* 21, 275–281
- 7 Bradbury, A. *et al.* (2003) Antibodies in proteomics II: screening, high-throughput characterization and downstream applications. *Trends Biotechnol.* 21, 312–317
- 8 Wingren, C. *et al.* (2003) Recombinant antibody microarrays – a viable option? *Nat. Biotechnol.* 21, 223
- 9 Nizak, C. *et al.* (2003) Recombinant antibodies to the small GTPase Rab6 as conformation sensors. *Science* 300, 984–987
- 10 Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317
- 11 McCafferty, J. *et al.* (1990) Phage antibodies: filamentous phage displaying antibody variable domain. *Nature* 348, 552–554
- 12 Barbas, C.F. III *et al.* (1991) Assembly of combinatorial antibody libraries on phages surfaces: the gene III site. *Proc. Natl. Acad. Sci. U. S. A.* 88, 7978–7982
- 13 Breitling, F. *et al.* (1991) A surface expression vector for antibody screening. *Gene* 104, 147–153
- 14 Hoogenboom, H.R. *et al.* (1991) Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* 19, 4133–4137
- 15 Sblattero, D. and Bradbury, A. (2000) Exploiting recombination in single bacteria to make large phage antibody libraries. *Nat. Biotechnol.* 18, 75–80

- 16 Clackson, T. *et al.* (1991) Making antibody fragments using phage display libraries. *Nature* 352, 624–628
- 17 Hoogenboom, H.R. (1997) Designing and optimizing library selection strategies for generating high-affinity antibodies. *Trends Biotechnol.* 15, 62–70
- 18 De Haardt, H.J. *et al.* (1999) A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J. Biol. Chem.* 274, 18218–18230
- 19 Pini, A. *et al.* (1998) Design and use of a phage display library. *J. Biol. Chem.* 273, 21769–21776
- 20 Hayashi, N. *et al.* (1994) Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques* 17, 310–316
- 21 Knappik, A. *et al.* (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus framework and CDRs randomized with trinucleotides. *J. Mol. Biol.* 296, 57–86
- 22 Johansen, L.K. *et al.* (1995) pFab60: a new, efficient vector for expression of antibody Fab fragments displayed on phage. *Protein Eng.* 8, 1063–1067
- 23 Welschof, M. *et al.* (1997) The antigen binding domain of a human IgG-anti-F(ab')<sub>2</sub> autoantibody. *Proc. Natl. Acad. Sci. U. S. A.* 94, 1902–1907
- 24 Little, M. *et al.* (1999) Generation of a large complex antibody library from multiple donors. *J. Immunol. Methods* 231, 3–9
- 25 McCafferty, J. *et al.* (1994) Selection and rapid purification of murine antibody fragments that bind a transition-state analog by phage-display. *Appl. Biochem. Biotechnol.* 47, 157–173
- 26 Vaughan, T.J. *et al.* (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat. Biotechnol.* 14, 309–314
- 27 Krebber, A. *et al.* (1997) Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods* 201, 35–55
- 28 Akamatsu, Y. *et al.* (1993) Construction of a human Ig combinatorial library from genomic V segments and synthetic CDR3 fragments. *J. Immunol.* 151, 4651–4659
- 29 Shirai, H. *et al.* (1999) H3-rules: identification of CDR3-H3 structures in antibodies. *FEBS Lett.* 455, 188–197
- 30 Hoogenboom, H.R. and Winter, G. (1992) By-passing immunisation: human antibodies from synthetic repertoires of germline V<sub>H</sub> gene segments rearranged *in vitro*. *J. Mol. Biol.* 227, 381–388
- 31 Nissim, A. *et al.* (1994) Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO J.* 13, 692–698
- 32 Barbas, C.F. III *et al.* (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problems. *Proc. Natl. Acad. Sci. U. S. A.* 89, 4457–4461
- 33 Desiderio, A. *et al.* (2001) A semi-synthetic repertoire of intrinsically stable antibody fragments derived from a single-framework scaffold. *J. Mol. Biol.* 310, 603–615
- 34 Jirholt, P. *et al.* (1998) Exploiting sequences space: shuffling *in vivo* formed complementary determining regions into a master framework. *Gene* 215, 471–476
- 35 Söderlind, E. *et al.* (2000) Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat. Biotechnol.* 18, 852–856
- 36 Sheets, M.D. *et al.* (1998) Efficient construction of a large nonimmune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6157–6162
- 37 Waterhouse, P. *et al.* (1993) Combinatorial infection and *in vivo* recombination: a strategy for making large phage antibody repertoires. *Nucleic Acids Res.* 21, 2265–2266
- 38 Griffiths, A.D. *et al.* (1994) Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* 13, 3245–3260
- 39 Geoffroy, F. *et al.* (1994) A new phage display system to construct multicombinatorial libraries of very large antibody repertoires. *Gene* 151, 109–113
- 40 Simmons, L.C. *et al.* (2002) Expression of full-length immunoglobulins in *Escherichia coli*: rapid and efficient production of aglycosylated antibodies. *J. Immunol. Methods* 263, 133–147
- 41 Bird, R.E. *et al.* (1988) Single-chain antigen-binding proteins. *Science* 242, 423–426
- 42 Huston, J.S. *et al.* (1988) Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 85, 5879–5883
- 43 Bird, R.E. and Walker, B.W. (1991) Single chain variable regions. *Trends Biotechnol.* 9, 132–137
- 44 Plückthun, A. (1990) Antibodies from *Escherichia coli*. *Nature* 347, 497–498
- 45 Plückthun, A. (1991) Antibody engineering: advances from the use of *Escherichia coli* expression systems. *Biotechnology* 9, 545–551
- 46 Kramer, K. *et al.* (2002) A generic strategy for subcloning antibody variable regions from the scFv phage display vector pCANTAB 5E into pASK85 permits the economical production of Fab fragments and leads to improved recombinant immunoglobulin stability. *Biosens. Bioelectron.* 17, 305–313
- 47 Glockshuber, R. *et al.* (1990) A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry* 29, 1362–1367
- 48 Steinhauer, C. *et al.* (2002) Single framework recombinant antibody fragments designed for protein chip applications. *Biotechniques* 33, 538–545
- 49 Bradbury, A. (2003) scFvs and beyond. *Drug Discov. Today* 8, 737–739
- 50 Roggenbuck, D. *et al.* (1994) Real-time biospecific interaction analysis of a natural human polyreactive monoclonal IgM antibody and its Fab and scFv fragments with several antigens. *Scand. J. Immunol.* 40, 64–70
- 51 Ciric, B. *et al.* (1995) Effect of valency on binding properties of the antihuman IgM monoclonal antibody 202. *Hybridoma* 14, 537–544
- 52 Plückthun, A. and Pack, P. (1997) New protein engineering approaches to multivalent and bispecific antibody fragments. *Immunotechnology* 3, 83–105
- 53 Kriangkum, J. *et al.* (2001) Bispecific and bifunctional single chain recombinant antibodies. *Biomol. Eng.* 18, 31–40
- 54 Kortt, A.A. *et al.* (2001) Dimeric and trimeric antibodies: high avidity scFvs for cancer targeting. *Biomol. Eng.* 18, 95–108
- 55 Marks, J.D. *et al.* (1991) By-passing immunization: human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* 222, 581–597
- 56 Skerra, A. and Schmidt, T.G.M. (1999) Applications of a peptide ligand for streptavidin: the Strep-tag. *Biomol. Eng.* 16, 79–86
- 57 Cull, M.G. and Schatz, P.J. (2000) Biotinylation of proteins *in vivo* and *in vitro* using small peptide tags. *Methods Enzymol.* 326, 430–440
- 58 Albala, J.S. *et al.* (2000) From genes to proteins: high-throughput expression and purification of the human proteome. *J. Cell. Biochem.* 80, 187–191
- 59 Holz, C. *et al.* (2001) A human cDNA expression library in yeast enriched for open reading frames. *Genome Res.* 11, 1730–1735
- 60 Sawasaki, T. *et al.* (2002) A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14652–14657
- 61 Janin, J. *et al.* (1978) Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* 125, 357–386
- 62 Frank, R. (2002) High-density synthetic Peptide microarrays: emerging tools for functional genomics and proteomics. *Comb. Chem. High Throughput Screen.* 5, 429–440
- 63 Bialek, K. *et al.* (2003) Epitope-targeted proteome analysis: towards a large-scale automated protein-protein-interaction mapping utilizing synthetic peptide arrays. *Anal. Bioanal. Chem.* 376, 1006–1013
- 64 Hallborn, J. and Carlsson, R. (2002) Automated screening procedure for high-throughput generation of antibody fragments. *Biotechniques* 33, 30–37
- 65 Konthur, Z. and Walter, G. (2002) Automation of phage display for high-throughput antibody development. *Targets* 1, 30–36
- 66 Rondot, S. *et al.* (2001) A helperphage to improve single chain antibody presentation in phage display. *Nat. Biotechnol.* 19, 75–78
- 67 Skerra, A. (2000) Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* 13, 167–187
- 68 Xu, L. *et al.* (2002) Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* 9, 933–942
- 69 Marks, J.D. *et al.* (1992) By-passing immunization: building high affinity human antibodies by chain shuffling. *Biotechnology* 10, 779–783
- 70 Schmidt, A. *et al.* (2001) Generation of human high-affinity antibodies specific for the fibroblast activation protein by guided selection. *Eur. J. Biochem.* 268, 1730–1738

- 71 Dörsam, H. *et al.* (1997) Antibodies to steroids from a small human naive IgM library. *FEBS Lett.* 414, 7–13
- 72 Meulemans, E.V. *et al.* (1994) Selection of phage-displayed antibodies specific for a cytoskeletal antigen by competitive elution with a monoclonal antibody. *J. Mol. Biol.* 244, 353–360
- 73 O'Connel, D. *et al.* (2002) Phage versus phagemid libraries for generation of human monoclonal antibodies. *J. Mol. Biol.* 321, 49–56
- 74 Haidaris, C.G. *et al.* (2001) Recombinant human antibody single chain variable fragments reactive with *Candida albicans* surface antigens. *J. Immunol. Methods* 257, 185–202
- 75 Barbas, C.F. III *et al.* (1993) Selection of human anti-hapten antibodies from semisynthetic libraries. *Gene* 137, 57–62
- 76 Goletz, A. *et al.* (2002) Selection of large diversities of antiidiotypic antibody fragments by phage display. *J. Mol. Biol.* 315, 1087–1097
- 77 Iba, Y. *et al.* (1997) Expression vectors for the introduction of highly diverged sequences into the six complementary-determining regions of an antibody. *Gene* 194, 35–46
- 78 Garrard, L.J. and Henner, D.J. (1993) Selection of an anti-IGF-1 Fab from a Fab phage library created by mutagenesis of multiple CDR loops. *Gene* 128, 103–109
- 79 De Kruif, J. *et al.* (1995) Selection and application of human single chain Fv antibody fragments from a semi-synthetic phage antibody display library with designed CDR3 regions. *J. Mol. Biol.* 248, 97–105
- 80 Viti, F. *et al.* (2000) Design and use of phage display libraries for the selection of antibodies and enzymes. *Methods Enzymol.* 326, 480–497
- 81 Persson, M.A.A. *et al.* (1991) Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning. *Proc. Natl. Acad. Sci. U. S. A.* 88, 2432–2436

### Articles of Interest

Articles of interest in other *Trends* and *Current Opinion* journals

#### Transgenic animals in inflammatory disease models

Carrie M Brodmerkel and Krishna Vaddi *Current Opinion in Biotechnology* 10.1016/j.copbio.2003.10.001

#### Engineering proteins that bind, move, make and break DNA

Cynthia H Collins, Yohei Yokobayashi, Daisuke Umeno and Frances H Arnold  
*Current Opinion in Biotechnology* 10.1016/j.copbio.2003.10.002

#### Biochemical engineering in the 21st century

Jeremy S Edwards and Kenneth J Kauffman *Current Opinion in Biotechnology* 5, 451–453

#### Enabling inverse metabolic engineering through genomics

Ryan T Gill *Current Opinion in Biotechnology* 5, 484–490

#### GABA signaling: a conserved and ubiquitous mechanism

Nicolas Bouché, Benoît Lacombe and Hillel Fromm *Trends in Cell Biology* 10.1016/j.tcb.2003.10.001

#### Aberrant trafficking of transmembrane proteins in human disease

Christian Cobbold, Anthony P. Monaco, Asipu Sivaprasadarao and Sreenivasan Ponnambalam  
*Trends in Cell Biology* 10.1016/j.tcb.2003.10.008

#### Sweet new world: glycoproteins in bacterial pathogens

M. Alexander Schmidt, Lee W. Riley and Inga Benz *Trends in Microbiology* 10.1016/j.tim.2003.10.004