

Applications of display technology in protein analysis

Min Li

Department of Physiology and Department of Neuroscience, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205 (minli@jhmi.edu).

Received 26 June 2000; accepted 31 October 2000

Display technology refers to a collection of methods for creating libraries of modularly coded biomolecules that can be screened for desired properties. It has become a routine tool for enriching molecular diversity and producing novel types of proteins. The combination of an ever-increasing variety of libraries of modularly coded protein complexes with the development of innovative approaches to select a wide array of desired properties has facilitated large-scale analyses of protein-protein/protein-substrate interactions, rapid isolation of antibodies (or antibody mimetics) without immunization, and function-based protein analysis. Several practical and theoretical challenges remain to be addressed before display technology can be readily applied to proteomic studies.

Keywords: phage display, functional genomics, proteomic technologies, protein-protein interactions

Estimates of the total complement of human genes vary from as low as 40,000 to 160,000. The human genome is thought to encode anywhere between 300,000 and 7.5 million protein and peptide forms¹. But these numbers may be significantly underestimated because even for the immunoglobulin gene—one single-gene locus containing hundreds of various V, D, and J segments—the immune system is capable of producing more than 10 million different proteins simply through gene rearrangement.

Efforts to analyze cellular protein content and function on a global scale thus require technologies that can cope with the tremendous diversity of proteins (at least a majority of the protein forms) in a high-throughput format. Biomolecular display technologies, which allow the construction of a large pool of modularly coded biomolecules, their display for property selection, and rapid characterization (decoding) of their structures, are particularly useful for accessing and analyzing protein diversity on a large scale.

To date, display technologies have comprised two major groups: biological display systems that employ a biological host/biological reactions, and nonbiological display systems that use chemical and engineering techniques. Regardless of the format, a display library consists of modularly coded molecules, each of which contains three components: displayed entities, a common linker, and the corresponding individualized codes (Fig. 1). Over the past decade, many display formats have been developed and applied in biological and pharmaceutical research. These technologies use different types of displayed entities, linkage formats, and coding strategies.

Several excellent review articles cover various aspects of phage display, including structural basis, uses, and evolution (e.g., see refs 2–4). In this review, I summarize several of the most common display formats used specifically for biomolecules, including, but not limited to, phage display. By focusing on their unified conceptual framework, emphasis is placed on present and possible future uses of biological display technologies in protein and proteomics research. Chemical display systems are not discussed.

Building blocks, linkers, and codes

Display libraries vary widely in size and complexity. On the basis of theoretical calculation as well as experimental results, these param-

eters essentially determine the probability and quality of identified biomolecules⁵. Although increasing both the size and the complexity of a display library is an important objective (as in medicinal chemistry), the fundamental aim is to optimize the assembly of building blocks that potentially lead to novel/more diverse properties.

Biological display exploits the cellular biosynthesis machinery to assemble biopolymers, the sequence of which ultimately specifies structure and distinct properties. Although nucleotide polymers, such as RNA/DNA aptamers, have yielded interesting molecules⁶, the most commonly exploited for biological display is the nucleic acid-coded synthesis of L-amino acid polymers (proteins).

Most biological display systems use the 20 natural L-amino acids as building blocks and take advantage of enzymatic protein synthesis. It is now also possible to achieve template-based incorporation of unnatural building blocks, such as synthetic amino acid derivatives⁷. One of the key issues is whether the somewhat limited number of building blocks, such as D- or L-amino acids, is sufficient to confer a large enough structural “space” for the desired properties. In contrast, for nonbiological displays, there is virtually no specific limit in terms of what building blocks may be chosen for the generation of diverse molecules. Many unique building blocks have been chosen and tested, ranging from linear polymers of D-amino acids to a variety of spatially modified organic moieties⁸. Despite significant progress, the establishment of a robust system that can combine the advantages of biological and nonbiological displays continues to be a key challenge.

One of the most important characteristics of display technologies is the ability to determine the structure of a desired compound rapidly after initial screening. Structural (or sequence) characterization is often accomplished by a process commonly known as coding and decoding, which can be achieved via a coupled amplification and purification process. In a biological display, chemical entities include polymers of amino acids or nucleotides. These are “linked” to codes that have chemical and physical properties that can be readily determined (e.g., the sequence of nucleic acids⁹, chromatographic properties¹⁰, or spatially addressed locations¹¹). A “linker” is used to establish the modularly coded biomolecular units, each of which possesses a unique property for either detection or deconvolution

REVIEW

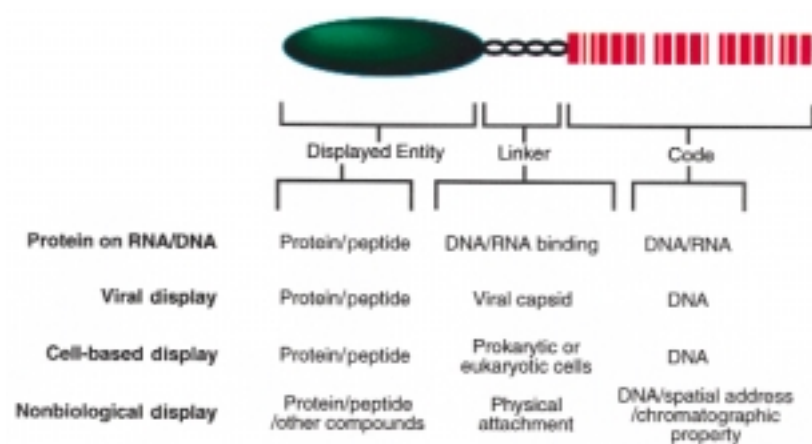


Figure 1. A schematic diagram of a typical display module and a list of four major display systems.

because of the attached codes. In most cases, the linkage between the displayed entity and the corresponding code is achieved by physical connections via either covalent or noncovalent chemical binding.

There are two requirements for coding strategies: the availability of adequate information space to cover the diverse entities to be displayed, and the presence of a highly sensitive and/or amplifiable property that permits the required rapid deconvolution or decoding. In a biological display, the distinctions between different linking strategies are often the basis for classification of various display formats.

Coding formats

Three types of coding formats are commonly used: peptide-on-DNA/RNA display, viral (phage) display, and cell-based display. The first of these formats uses protein–DNA/RNA complexes as its foundation. By expressing the peptide in a form that is capable of binding to its coding DNA/RNA, one can screen a large pool of complexes and identify bound peptides by the isolation and sequencing of nucleotide sequences of either DNA or RNA. Two main strategies are employed, which differ depending on whether DNA-coding¹² or RNA-coding^{13,14} formats are used.

Peptide-on-plasmid (DNA coding) is a system in which a library of plasmids expresses random peptides fused to individual identical DNA-binding proteins. Because the peptide–DNA binding protein fusion in a given bacterium is encoded by the corresponding plasmid, the formation of modularly coded peptide–plasmid complexes allows the selection of individual peptides of interest through the recovery of the bound plasmid (see Fig. 2A).

In contrast to the peptide-on-plasmid display, the polysome display (RNA coding) starts with a pool of RNA, typically synthesized by transcription *in vitro*. After *in vitro* translation under appropriate conditions, the synthesized peptides form a complex with the coding RNA via either noncovalent¹³ or covalent chemical bonding¹⁴. In the noncovalent polysome complex, the interaction of ribosome with tethered nascent peptides on mRNA can be stabilized at a high divalent Mg^{2+} concentration. Covalent chemical bonding is usually achieved by performing *in vitro* translation using an RNA–DNA fusion template where each peptide-coding RNA molecule in the library is individually fused to a common single-stranded DNA fragment to which puromycin is chemically attached. Under these conditions, the *in vitro* translation stalls at the end of the RNA open reading frame (ORF) and the RNA–DNA junction. This allows the *cis*-tethered puromycin to enter the A-site of a ribosome and form a covalent linkage with the C terminus of the corresponding nascent peptide. The modular protein–DNA (or RNA) complexes in these systems have several advantages, including freedom from the selec-

tive pressures of biological hosts (e.g., see phage coat assembly discussion below) and capacity for subsequent biochemical or chemical modifications of the peptides.

The second type of display format, viral (or phage) display, is one of the most commonly used¹⁵. This system takes advantage of our understanding of the coat proteins of viruses, mostly bacterial viruses (phages). By cloning, for example, random oligonucleotides into the coding sequence of viral coat proteins, one can produce a library of viruses in which each carries a distinct peptide sequence as part of the coat protein. The coding information is the corresponding DNA sequence embedded in the viral genome that is packed inside the viral capsid, thereby establishing the link between displayed entities and the code (see Fig. 2B). To achieve certain desired display advantages, several different viral systems have been used to display peptides, including lysogenic filamentous phages¹⁵ and lytic lambda phage^{16–19}, T7 bacteriophage²⁰, and T4 bacteriophage^{21,22}.

Lysogenic filamentous phage remains the most commonly used phage display system. For cases in which displayed proteins may be toxic to filamentous phage assembly or incompatible with the bacterial secretion pathway, lytic phages can be used that allow displayed sequences to minimize negative selection. In addition, some proteins may be preferably displayed as a C-terminal fusion, a configuration that is more amenable to lytic phage-based display systems.

The last type of display format is cell-based systems, which are essentially identical to the expression cloning approach, in which a complementary DNA (cDNA) library encoding various proteins is recombinantly expressed in tissue culture cells and selected for binding to a specific ligand on the cell surface²³. Cell-based display is divided into two classes: extracellular display^{24–26} and intracellular display^{27–29}.

A cell-based display of large cDNA libraries in mammalian cells requires efficient gene transfer, for example, via a viral vector to introduce DNA. In yeast and other prokaryotic systems, this is less of a problem because transformation is more efficient. For example, *Escherichia coli* flagellin 26 has been used to display moieties (Fig. 2C). In both cases, the cellular host is used to establish the modular link between the coding DNA and the displayed peptides/proteins.

The value of using cellular hosts may very well be underestimated, because it is becoming increasingly evident that post-translational modifications of biomolecules could play an essential role in coding for both new and native properties. For functional studies of unknown proteins, the native host provides unparalleled advantages that would not be possible from the combinatorial assembly of 20 amino acid building blocks via a heterologous expression system either *in vitro* or *in vivo*.

Applications exploiting linear sequence diversity

One of the earliest applications of display technology was to search for new peptides that bind cell-surface receptors, primarily protein-binding cytokine receptors (see review in ref. 3). Targets used in such studies have been expanded to include enzymes, intracellular signaling proteins, ion channels, cell surface architectures, and nonprotein targets, such as special surface material. Many of these experiments are geared to discover therapeutic peptides or understand the mechanisms behind receptor signaling mechanisms, protein folding, and enzymatic catalysis.

One example in which the display approach was particularly effective is the discovery of peptides agonistic to the erythropoietin receptor (EPOR)³⁰. EPO mimetic peptides (EMPs) have several interesting properties, including the ability to bind EPORs and the formation of a peptide dimer, which induces dimerization of recep-

tor and confers agonistic activity³¹. Detailed studies of the affinity and transduction efficacy of EMPs and EPO have revealed mechanistic coupling among the different steps of ligand–receptor interactions, including receptor dimerization, receptor contact geometry, and signaling efficacy³².

Display technologies have been especially useful for elucidating the complex network through which proteins interact with one another. Many signaling and structural proteins contain protein modules designed to mediate protein–protein interactions (<http://www.ebi.ac.uk/interpro/>). Proteins that possess such interaction modules provide the molecular scaffold on which to organize macromolecular protein complexes. These modules include Src homology (SH)3 domain (which interacts with proline-rich peptides) and PDZ domains (which normally bind the C terminus of target proteins). Because these domains bind to small linear peptide sequences, peptide display has been a useful tool for identifying optimal binders and orphan ligands, for revealing the binding preferences of protein modules³³, and in some cases for providing direct proof of a biological model, such as the diversity of PDZ domain binding³⁴ and binding preference of enabled/vasodilator-stimulated phosphoprotein (VASP) homology (EVH) domains³⁵.

Identification of the optimal binding substrate of orphan protein or protein interaction domains may also provide essential information for identification of their physiologically interacting substrates^{36,37}. Considering the large number of different protein interaction domains, a high-throughput screen using display technology combined with bioinformatics tools could provide an important data set of protein interactions.

Other applications of peptide display include screens for peptides that direct binding of phage particles to specific cell or tissue surfaces, where the specific target proteins are not preselected^{38,39}. Although the original objective is to use resultant peptides as targeting signals for specific tissue delivery of gene therapy vectors, it is now conceivable that this approach can be broadened to profile cell surface receptors for interactions with proteins or peptides. It is also possible to array a library of proteins/peptides (e.g., immunoglobulins) on a chip. By indexing cell attachment and/or morphological changes, one may learn important information concerning cellular surface landscape and provide indexing profiles for cells in different physiological or pathological states.

Applications exploiting spatial diversity

With 20 amino acids as building blocks, it is possible to create libraries that contain all possible sequences for relatively short peptides. However, most biological reactions are carried out by large, well-folded proteins. A folded protein usually has one or several protein modules (or domains). For example, peptide-binding PDZ domains usually have an average length of 85–100 amino acids. To create a library of 100 residues, one needs to generate a total of $(20)^{100}$ different sequences. If such a library consisting of one copy of each sequence were possible, it would have a total mass exceeding that of the earth. Thus, the sheer mass renders experimental approaches impractical for exploiting conformational diversity space. In silico approaches provide an alternative and have already shown very promising results⁴⁰.

The display of large, three-dimensionally folded proteins began in 1990 when the antibody coding fragments were engineered and presented as part of a bacteriophage coat protein^{41–43}. The broad use of antibodies for diagnostics, industrial enzymes, and more recently as protein therapeutics has added impetus to the development of antibody display technologies that enable: first, the development of single-chain antibodies and antibody mimetics; second, heterologous expression and screening of the complete repertoire of human antibody genes in both mammalian and prokaryotic expression systems; and third, the expansion of the natural antibody gene reper-

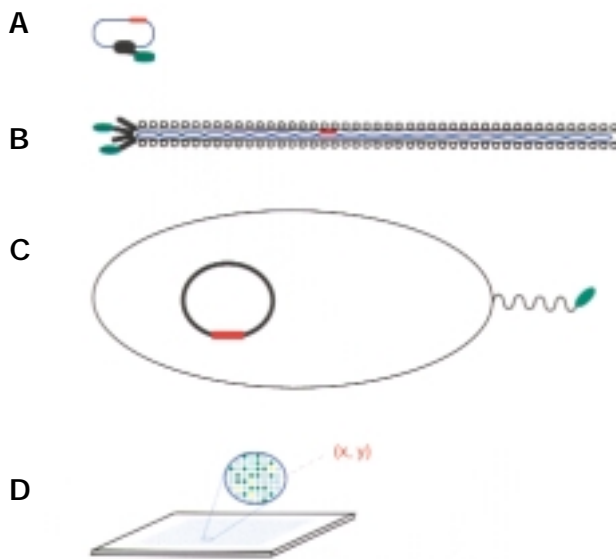


Figure 2. Examples of four major display systems. The displayed entities are shown in green and corresponding codes are in red. The cell-based display system is not to scale. (A) Protein on RNA/DNA. (B) Viral display. (C) Cell-based display. (D) Nonbiological display.

toire of existing organisms, allowing the improvement of library qualities such as size, diversity, and functionality. Because of the modular nature of display systems, clonal display of immunoglobulin also permits unique selection schemes for biological effects or readouts, including signal transmission and membrane internalization and trafficking^{44,45}.

With the development of humanized antibodies, various display formats have been exploited to identify optimal binders^{46,47}. The development of single-chain antibodies (scFv) has significantly simplified the display of immunoglobulin binding on a phage surface. Molecular manipulations of various phage display libraries allow the combinatorial selection of scFvs produced through coexpression of light-chain and heavy-chain segments, reminiscent of immunoglobulin gene rearrangements during immune differentiation^{48,49}. It is significant that the immune system uses a natural naïve repertoire of as few as 10^8 molecules to generate binders for essentially any antigen. Perhaps future improvements could be made to display immunoglobulin folds in ways that allow loop movement, induced fit, and V_H-V_L domain reorientation, the three key features that have been learned from studying the antibody–antigen interaction.

The display of relatively large protein molecules, such as the variable regions of immunoglobulins, has met with some success, but it is important to recognize that the display of one family of identically (or very similarly) folded proteins is a very different prospect from the display of the entire complement of proteins encoded by the estimated 40,000–160,000 genes in the human genome. Increasing numbers of non-immunoglobulin proteins have been functionally displayed on phage⁵⁰. Similar approaches have now been applied to genomic proteins derived from total cDNAs^{51,52,54,55}.

Many strategies have been taken to minimize the incompatibility between the prokaryotic expression system and eukaryotic proteins. For example, phage coat proteins have been mutated to make them more amenable to displaying larger proteins and proteins that are functional only if they are oligomerized properly⁵³. In addition, lytic phages such as lambda, T4, and T7 have been used to provide an alternative that is independent of the *Escherichia coli* secretion machinery. These lytic viral display systems have provided promising results as proteomic tools to study protein–protein interactions in signaling pathways^{54,55}.

REVIEW

Table 1. Comparison of display systems applicable to mammalian cDNAs

Property	Polysome display	Peptide-on-plasmid	Filamentous phage display	Lytic phage display	Prokaryotic cell-based display	Mammalian cell-based display
Theoretical upper limit of library size	>10 ¹¹	<10 ¹¹	<10 ¹¹	<10 ¹¹	~10 ⁹	~10 ⁸
Host expression systems	In vitro	Prokaryote	Prokaryote	Prokaryote	Prokaryote	Mammalian cell
Linkers	Non-covalent or covalent	Non-covalent	Viral capsid	Viral capsid	Cell	Cell
cDNA size restriction	Yes	No	Yes	Yes	No	No
Folding machinery	No	Nonnative	Nonnative	Nonnative	Nonnative	Native
High-throughput readiness	High	High	Low	Low	Low	Low
Post-translational modifications	No	No	No	No	No	Native

As the demand for proteomic display systems increases, it is necessary to validate these bacteriophage-based systems in different experimental settings. In particular, one needs to consider the mode of interaction and the effect of negative selection. If a target receptor (molecule) is a large protein with a binding pocket, it is more likely to identify its binding partners because this interaction makes less demands on the overall folding quality of the displayed peptide/proteins in a library. In contrast, if the target is a small molecule, such as a peptide or organic compound, it is expected that the interaction will require the displayed proteins to be well folded. Whereas the phage systems exert significant negative selection pressure on the functional expression of mammalian proteins, recent results suggest that at least in some cases expressed mammalian proteins are functional and can actually interact with small protein targets⁵⁵ and small organic targets^{56,57}.

The ultimate goal of biological display is to ensure that the selection criteria optimize the likelihood of identifying binders with a specific physiological function, such as protein-protein interaction, catalytic activity, or activation of a signaling pathway. The genomic yeast two-hybrid system takes advantage of spatial address coding by seeding cells on an array combined with a mating step⁵⁸. As a result, cells with different combinations of bait and prey can be selected genetically on a conditional medium.

The quality of molecular diversity

One of the key considerations for applying display technologies to the global analysis of protein function (proteomics) will be to determine and define the criteria for a high-quality library. Can all the possible protein sequences be included and covered in one library? Should constraint or biased libraries be generated to provide rigidity or preference for recognition?

A complete random display approach is not practical for protein display. Thus, a balanced consideration of library size, source of coding sequences, clonal representation, and property diversity and selection should be used to maximize the quality and diversity of the roughly 1 billion (10⁹) library members, a level of complexity that can be generated and screened through conventional molecular biology techniques.

The most common protein display libraries are made from either total cDNAs of a given tissue or coding segments of immunoglobulin genes. Despite the similarity, the two libraries require rather distinctive considerations. An antibody library is sensitive to disulfide bond formation, potential toxicity to the host (resulting in negative selection), and possibly degenerate binding specificity. In contrast, the display of cDNAs that encode diversely folded cellular proteins requires a robust folding machinery, native post-translational modifications, and compatibility with the property selection scheme. Table 1 summarizes pros and cons of several existing systems for displaying mammalian cDNA libraries for property selection.

Limitations and opportunities

Despite the rapid progress in many areas of display technologies, several limitations remain both in theory and in practice. These limitations include system-specific negative selections, incompatibility with the expression host, and restricted scope of selection parameters (e.g., binding).

One important question in display technology concerns the nature of screening experiments and the interpretation of screening results. Because failure to identify an interacting ligand from a given screen is essentially a negative result, it is unclear whether this is due to the complexity of the library, the procedures used to retrieve binding sequences, or the format of the receptor target presentation. Even in cases in which interacting candidates have been identified, it is still questionable whether the clones with the most desirable properties or of physiological significance either are present in the library or have been isolated from the screen.

One way of addressing such problems, for example, is to allow the contact between the library of compounds with a target to reach chemical equilibrium, and to reduce binding valency/include a binding competitor, thereby increasing the binding affinity threshold. However, in many cases, there is no reliable experimental means or theoretical basis to evaluate the practical complexity and specific usefulness of a given library. Future development of a profiling index for display libraries would have many benefits, including assess-

ments of “diversity space” and suitability of a library, and evaluation of effectiveness for a given screen. Although development of such a profiling index may be particularly difficult for a synthetic naïve library, the profiling of natural naïve libraries appears possible.

In display systems, the current selection criteria are primarily based on the “contacting” property. In almost all reported cases, screening is accomplished using a specific contacting probe (i.e., the target). Whereas the contacting properties are often the ultimately desired features, it is increasingly recognized that folding diversity can be essential for achieving desired contacting features. For example, the intrinsic dimerization property of the EPO peptide is necessary for its agonistic activity. In this case, the ability of peptides to dimerize was not planned during library construction or specifically selected for during screening^{30,31}. The development of screening strategies that no longer rely on the binding affinity alone, such as conditional growth, may be particularly useful to retrieve molecular entities with the desired, noncontacting physical, chemical, and structural properties.

Display technologies may also be useful for studying other biological processes, for example, understanding peptide sequences that mediate membrane uptake, endocytosis, and nuclear transportation, processes that are both conceptually fundamental and of enormous practical application. Some exciting progress has been made in this area using genetic phenotypes to select peptides/proteins that interfere with certain signaling pathways—a method called transdominant genetic analysis that can be achieved using short peptides^{27–29} or immunoglobulins⁵⁹. Such a system does not need to have a predetermined receptor, but still enables rapid deconvolution of the peptide(s) that induces the phenotypic changes.

Several other challenges also face the application of phage display in proteomics. Considering that display is a search for a matching key/lock interaction, current technologies have particular strengths in the area of finding keys for a lock (target). The converse application of finding locks for a key (target) has met with only limited success. It is also increasingly recognized that a system that could accommodate two variables would allow the selection of a property via coevolution of two variables. This notion has already been tested by combining phage display with cell-based display⁶⁰. Because both interacting partners are in coded systems, one can readily change either or both interaction partners to achieve a variety of goals, including the cloning of receptors, identification of interaction residues on one or both partners, and coevolution of the desired properties. More recently, the library versus library screen strategy has been developed further using the oligomerization-dependent dihydrofolate reductase (mDHFR) activity, in which two interacting partners reconstitute a functional enzyme thereby allowing growth selection in a prokaryotic host system^{61,62}.

Expression systems for display approaches also need to be optimized. The most commonly used host systems, such as *E. coli*, lack the native translational machinery, folding apparatus, and mammalian enzymes/proteins that carry out the ~100 different post-translational modifications. In addition, biological display is inherently under certain selection pressure independent from that imposed by *in vitro* systems. This negative selection pressure eliminates or reduces proteins that are toxic to host, disruptive in folding, sensitive to proteolytic degradation, poorly translated, and processing/secretion-defective.

Because biological displays almost always involve repeated steps for amplification, the only limited positive selections are sequences that have expression advantages and/or host growth stimulation. In some synthetic naïve library screens, such as isolation of single-chain antibody molecules, the expression advantages have been seen as positive selection, because the subsequent steps often involve large-scale protein production and purification. However, for most proteomic applications, the lasting objectives are to functionally dis-

play all proteins and to minimize any selection pressure or expression bias. Improvements in display technologies in these areas could potentially provide an essential interface between genomic proteins and physiologically significant biomolecules and therapeutic molecules.

Conclusions

The combination of extremely high detection sensitivity of amplification and the ability to screen whole genomic proteins suggests that biological display will be useful for addressing biochemical and pharmacological questions at the genomic level. If the discovery of immunoglobulin gene rearrangement marks our understanding that a collection of genes that a person inherits is not necessarily a fixed endowment, the integration of directed evolution technologies and biological display may offer unprecedented opportunity for both understanding functional space and expanding diversity space.

Acknowledgment

I thank William J. Dower for many insightful discussions and Rachel Derin for assistance in manuscript preparation. The work in the author's laboratory is supported by grants from National Institute of Health.

1. Frost and Sullivan. *World Proteomic Market Report*. [AU:CITY OF PUB? initials of authors?] (2000).
2. Katz, B.A. Structural and mechanistic determinants of affinity and specificity of ligands discovered or engineered by phage display. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 27–45 (1997).
3. Smith, G.P. & Petrenko, V.A. Phage display. *Chem. Rev.* **97**, 391–410 (1997).
4. Bradbury, A. Display technologies expand their horizons. *Trends Biotechnol.* **17**, 137–138 (1999).
5. Perelson, A.S. & Oster, G.F. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J. Theor. Biol.* **81**, 645–670 (1979).
6. Patel, D.J. & Suri, A.K. Structure, recognition and discrimination in RNA aptamer complexes with cofactors, amino acids, drugs and aminoglycoside antibiotics. *J. Biotechnol.* **74**, 39–60 (2000).
7. Cornish, V.W. et al. Site-specific incorporation of biophysical probes into proteins. *Proc. Natl. Acad. Sci. USA* **91**, 2910–2914 (1994).
8. Gordon, E.M., Barrett, R.W., Dower, W.J., Fodor, S.P. & Gallop, M.A. Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J. Med. Chem.* **37**, 1385–1401 (1994).
9. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381–5383 (1992).
10. Ohlmeyer, M.H. et al. Complex synthetic chemical libraries indexed with molecular tags. *Proc. Natl. Acad. Sci. USA* **90**, 10922–10926 (1993).
11. Fodor, S.P. et al. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
12. Cull, M.G., Miller, J.F. & Schatz, P.J. Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc. Natl. Acad. Sci. USA* **89**, 1865–1869 (1992).
13. Mattheakis, L.C., Bhatt, R.R. & Dower, W.J. An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA* **91**, 9022–9026 (1994).
14. Roberts, R.W. & Szostak, J.W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* **94**, 12297–12302 (1997).
15. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).
16. Santini, C. et al. Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. *J. Mol. Biol.* **282**, 125–135 (1998).
17. Sternberg, N. & Hoess, R.H. Display of peptides and proteins on the surface of bacteriophage lambda. *Proc. Natl. Acad. Sci. USA* **92**, 1609–1613 (1995).
18. Maruyama, I.N., Maruyama, H.I. & Brenner, S. Lambda foo: a lambda phage vector for the expression of foreign proteins. *Proc. Natl. Acad. Sci. USA* **91**, 8273–8277 (1994).
19. Dunn, I.S. Assembly of functional bacteriophage lambda virions incorporating C-terminal peptide or protein fusions with the major tail protein. *J. Mol. Biol.* **248**, 497–506 (1995).
20. Rosenberg, A. et al. T select phage display system: a powerful new protein display system based on the bacteriophage T7. *Innovations* **6**, 1–6 (1996).
21. Ren, Z.J. et al. Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4. *Protein Sci.* **5**, 1833–1843 (1996).
22. Efimov, V.P., Nepluev, I.V. & Mesyanzhinov, V.V. Bacteriophage T4 as a surface display vector. *Virus Genes* **10**, 173–177 (1995).
23. Seed, B. & Aruffo, A. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA* **84**, 3365–3369 (1987).
24. Lu, Z. et al. Expression of thioredoxin random peptide libraries on the *Escherichia coli* cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. *Biotechnology* **13**, 366–372 (1995).

REVIEW

25. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
26. Little, M., Breittling, F., Dubel, S., Fuchs, P. & Braunagel, M. Human antibody libraries in *Escherichia coli*. *J. Biotechnol.* **41**, 187–195 (1995).
27. Caponigro, G. et al. Transdominant genetic analysis of a growth control pathway. *Proc. Natl. Acad. Sci. USA* **95**, 7508–7513 (1998).
28. Colas, P. et al. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550 (1996).
29. Norman, T.C. et al. Genetic selection of peptide inhibitors of biological pathways. *Science* **285**, 591–595 (1999).
30. Wrighton, N.C. et al. Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273**, 458–464 (1996).
31. Livnah, O. et al. Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **273**, 464–471 (1996).
32. Livnah, O. et al. Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **283**, 987–990 (1999).
33. Sparks, A.B. et al. Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc. Natl. Acad. Sci. USA* **93**, 1540–1544 (1996).
34. Stricker, N. et al. PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. *Nat. Biotechnol.* **15**, 336–342 (1997).
35. Tu, J. C. et al. Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron* **21**, 717–726 (1998).
36. van Huizen, R. et al. Two distantly positioned PDZ domains mediate multivalent INAD- phospholipase C interactions essential for G protein-coupled signaling. *EMBO J.* **17**, 2285–2297 (1998).
37. Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. & Kay, B.K. Cloning of ligand targets: systematic isolation of SH3 domain-containing proteins. *Nat. Biotechnol.* **14**, 741–744 (1996).
38. Barry, M.A., Dower, W.J. & Johnston, S.A. Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. *Nat. Med.* **2**, 299–305 (1996).
39. Pasqualini, R. & Ruoslahti, E. Organ targeting in vivo using phage display peptide libraries. *Nature* **380**, 364–366 (1996).
40. Dahiyat, B.I. & Mayo, S.L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
41. McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).
42. Griffiths, A.D. et al. Human anti-self antibodies with high specificity from phage display libraries. *EMBO J.* **12**, 725–734 (1993).
43. Barbas, C.F., Kang, A.S., Lerner, R.A. & Benkovic, S.J. Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl. Acad. Sci. USA* **88**, 7978–7982 (1991).
44. Becerril, B., Poul, M.A. & Marks, J.D. Toward selection of internalizing antibodies from phage libraries. *Biochem. Biophys. Res. Commun.* **255**, 386–393 (1999).
45. Kragler, F., Monzer, J., Xoconostle-Cazares, B. & Lucas, W.J. Peptide antagonists of the plasmodesmal macromolecular trafficking pathway. *EMBO J.* **19**, 2856–2868 (2000).
46. Griffiths, A.D. & Duncan, A.R. Strategies for selection of antibodies by phage display. *Curr. Opin. Biotechnol.* **9**, 102–108 (1998).
47. Winter, G., Griffiths, A.D., Hawkins, R.E. & Hoogenboom, H.R. Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455 (1994).
48. Sblattero, D. & Bradbury, A. Exploiting recombination in single bacteria to make large phage antibody libraries. *Nat. Biotechnol.* **18**, 75–80 (2000).
49. Sheets, M.D. et al. 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. USA* **95**, 6157–6162 (1998). (Published erratum appears in *Proc. Natl. Acad. Sci. USA* **96**, 795, 1999.)
50. Clackson, T. & Wells, J.A. In vitro selection from protein and peptide libraries. *Trends Biotechnol.* **12**, 173–184 (1994).
51. Cramer, R. & Suter, M. Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene* **137**, 69–75 (1993).
52. Jespers, L.S. et al. Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology* **13**, 378–382 (1995).
53. Sidhu, S.S., Weiss, G.A. & Wells, J.A. High copy display of large proteins on phage for functional selections. *J. Mol. Biol.* **296**, 487–495 (2000).
54. Santi, E. et al. Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics. *J. Mol. Biol.* **296**, 497–508 (2000).
55. Zozulya, S., Lioubin, M., Hill, R.J., Abram, C. & Gishizky, M.L. Mapping signal transduction pathways by phage display. *Nat. Biotechnol.* **17**, 1193–1198 (1999).
56. Zavala, A.G., Lancaster, T., Groopman, J.D., Strickland, P.T. & Chandrasegaran, S. Phage display of scFv peptides recognizing the thymidine(6-4)thymidine photoproduct. *Nucleic Acids Res.* **28**, E24 (2000).
57. Sche, P.P., McKenzie, K.M., White, J.D. & Austin, D.J. Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem. Biol.* **6**, 707–716 (1999).
58. Uetz, P. et al. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
59. Persic, L. et al. Targeting vectors for intracellular immunisation. *Gene* **187**, 1–8 (1997).
60. Li, M. Use of a modified bacteriophage to probe the interactions between peptides and ion channel receptors in mammalian cells. *Nat. Biotechnol.* **15**, 559–563 (1997).
61. Arndt, K.M. et al. A heterodimeric coiled-coil peptide pair selected *in vivo* from a designed library-versus-library ensemble. *J. Mol. Biol.* **295**, 627–639 (2000).
62. Pelletier, J.N., Arndt, K.M., Pluckthun, A. & Michnick, S.W. An *in vivo* library-versus-library selection of optimized protein–protein interactions. *Nat. Biotechnol.* **17**, 683–690 (1999).