In vitro display technologies – new tools for drug discovery

Kevin FitzGerald

Over the past decade, several ligand discovery techniques have been developed that mimic the process of natural evolution. Phage display technology is the most established of these methods and has been applied to numerous technological problems including the discovery of novel drugs. More recently, some new display technologies have emerged which, unlike phage display, operate entirely *in vitro* and have concomitant advantages. This review describes this new generation of display technologies and indicates how they might fit into the modern drug discovery process.

The process by which new pharmaceutical compounds are being discovered and developed is undergoing a dramatic transformation. The traditional process of drug discovery has depended largely on an empirical approach to screening compounds for novel pharmacological activity. To date, this approach has generated pharmaceutical compounds that collectively act on only a limited number of *in vivo* targets – 417 in total[1]. With the combination of rising R&D costs, short product life cycles, and the public's ever increasing demand for new, high-quality treatments, pharmaceutical companies are facing increasing pressures to develop new drugs faster, more cost-effectively and in greater numbers than ever before. These pressures have motivated the pharmaceutical companies to foster links with smaller biotechnology companies that could provide technological solutions to the need to enhance and accelerate the drug discovery process.

To this end, biotechnological developments in the areas of genomics and proteomics, bioinformatics and molecular library systems have attracted great interest from pharmaceutical companies. Genomics and proteomics technologies should satisfy the need to identify large numbers of novel drug targets by identifying disease correlations among the 100,000 or so genes in the human genome, as well as the considerably larger pool of potential targets that could arise through post-transcriptional and post-translational processing. The vast quantities of data that are generated by genomics and proteomics studies need to be organized in an accessible and meaningful way and this has led to an explosion in the adoption of sophisticated bioinformatics systems. Finally, the ability to generate and interrogate large, complex libraries of potential drug compounds provides the means to act on genomics and bioinformatics data to validate new drug targets and to produce the bioactive compounds required for therapeutic intervention.

Library technologies can be broadly divided into two categories: those that are based on entirely synthetic, low-MW compounds that can be generated by, for example, combinatorial chemistry techniques (reviewed in Ref. 2) and those that are based on biological compounds such as nucleic acids[3,4], peptides[5] or proteins[6]. They all consist of large, diverse collections of molecules and associated mechanisms for identifying individual compounds that exhibit certain desired properties (most typically, the ability to bind with sufficient levels of affinity and specificity to a given target molecule).

Among the various biological libraries that have emerged, most interest has focussed on those that enable the identification of bioactive proteins or peptides, principally because proteins are natural modulators of phenotype and, thus, represent obvious drug candidates. This article profiles some of the key protein and peptide library systems, focussing in particular on a new generation of *in vitro*

Kevin FitzGerald, Head of Library Technologies, Actinova Ltd, 185 Cambridge Science Park, Cambridge, UK CB4 0GA. tel: +44 1223 728458, fax: +44 1223 728448, e-mail: kevfitz1@hotmail.com

 1359-6446/00/$ – see front matter PII: S1359-6446(00)01501-4

display technologies that should enable the construction and exploitation of exceptionally large and complex libraries.

Display libraries

The earliest gene expression libraries were created using non-display systems that were typically generated by ligating DNA fragments into λ phage or plasmid expression vectors for the subsequent transformation of *Escherichia coli* (e.g. Refs 7,8). Potential genes of interest were identified by screening large numbers of individual colonies or plaques for the production of bioactive proteins. This screening approach limits the size of this type of expression library to $\approx 10^6$ recombinants, simply because of the prohibitive number of bacterial culture plates that would be required to array a larger number of spatially distinct clones.

The unifying feature of display systems that sets them apart from non-display systems is that, in all instances, the cloned gene and its encoded protein are in some way physically connected. The value of this linkage is that display libraries can be interrogated by a process of selection rather than screening. The library exists as a soluble pool from which target-binding members are enriched by affinity selection against an immobilized target molecule. Because ligands and associated coding sequences are co-selected, useful genes are separated from non-useful genes and can be subsequently recovered for storage or further manipulation. The physical constraint of arraying individual members of the entire library on a solid support for screening is therefore removed with the consequence that much larger libraries can be exploited, potentially enabling the facile isolation of much higher affinity ligands.

Phage display

Phage display[9] is the most widely adopted molecular display technique. It uses a bacterial virus (the M13-related filamentous phage, fd) to provide the physical coupling of gene and gene product. The diverse population of DNA fragments that comprise the library are simultaneously ligated into many copies of the phage genome in juxtaposition with the gene for one of the viral coat proteins. The recombinant phage DNA is introduced into *E. coli* cells resulting in large numbers of bacterial clones, each containing the coding sequence for a unique library member. When recombinant phage particles assemble during extrusion from the host cell, the cloned library DNA is packaged within the virion and the encoded gene product is incorporated into the viral coat. In many cases, the incorporated polypeptide is both properly folded and able to interact with cognate ligands. Large, diverse pools of recombinant phage particles can therefore be prepared and incubated with an immobilized target molecule. Non-binding particles are washed away and those rare library members that bind specifically to the target are eluted and amplified by infection into fresh *E. coli* cultures. In this way, potential genes of interest are enriched from very large libraries simply as a function of the affinity of the encoded protein for a given target molecule.

Phage display has been successfully applied to the selection of ligands from linear peptide libraries[10–12] and to the display of functional protein domains such as antibody fragments[13], enzymes[14], hormones[15] and DNA-binding proteins[16]. There are, however, limitations to the technique. The reliance of the method on the relatively low efficiency with which DNA can be introduced into *E. coli* cells (the transformation efficiency) typically imposes a ceiling on library sizes of $\approx 10^9$ recombinants (although exceptional libraries of $>10^{10}$ members have been described[17]). Although this greatly extends the complexity of gene expression libraries compared with those possible without a mechanism for display, the ability to generate even larger libraries might be expected to greatly enhance the utility of the molecular display concept. This expectation derives from the proposed direct correlation that exists between library size and the binding affinities of isolated ligands for a given target[18]. Indeed, belief in this correlation has been reinforced by the technological advances in phage antibody display that have occurred over the past decade: the first non-immune phage antibody libraries contained $\approx 10^7$ members and usually yielded antibodies with binding affinities in the 100 nM range[19], whereas those from a 10^{10} library often have affinities in the subnanomolar range[17].

Alternatives to phage display

Several groups have considered alternative display methods such as display on bacterial surfaces[20], yeast surfaces[21], eukaryotic viruses[22], or directly on the encoding plasmid DNA (Ref. 23). These systems still require transformation of a cellular host and have therefore not succeeded in increasing library sizes.

In vitro display technologies

In recent years, several *in vitro* display technologies have emerged that, by circumventing the need to introduce DNA into a cellular host, promise to provide access to libraries that are limited only by the quantity of DNA that can be physically added to cell-free protein synthesis systems (potentially up to 10^{14}–10^{15} molecules). This should enable the construction of libraries that are several orders

of magnitude larger than those possible when using whole cells or viruses for molecular display.

Covalent Display Technology

Covalent Display Technology (CDT; Actinova Ltd, Cambridge, UK and David Andrews and Yun Li, McMaster University, Hamilton, Canada) is a proprietary *in vitro* library platform that has the potential to avoid many of the problems that limit other display systems. CDT exploits the extraordinary properties of a replication initiator protein from the *E. coli* bacteriophage P2. The protein is the product of the viral *A*-gene (P2A) and is an endonuclease that initiates a rolling circle replication process by binding to the viral origin (*ori*) and introducing a single strand discontinuity (nick) in the DNA. The 3′-OH group that is exposed by the action of P2A is used to prime progeny DNA synthesis using the host replication machinery[24–26]. The nicking event also exposes a 5′-phosphate and, crucially for the development of CDT, this becomes covalently attached to a tyrosine residue in the active site of P2A (Refs 27,28). One further property of P2A that is exploited in the CDT system is that P2A exclusively attaches to the same molecule of DNA from which it has been expressed[27]. The mechanism of this *cis*-activity is not well understood. However, the high fidelity of the *cis*-activity and the fact that the recognition sequence for the covalent attachment, *ori*, occurs within P2A's own coding sequence[24,29] enables pools of polypeptides that are genetically fused to P2A to be synthesized *in vitro* such that they also become covalently attached to their own coding sequences (Fig. 1).

To operate CDT, a pool of DNA molecules is prepared, each containing the coding sequence of P2A fused to the coding sequence for one of a diverse population of potential binding moieties (linear peptides or protein domains). The DNA pool is transcribed and translated concurrently *in vitro* using an *E. coli* S30 lysate and, because of the *cis*-activity of P2A, each DNA molecule becomes covalently tagged with its own expressed gene product. The protein–DNA complexes are then subjected to affinity selection strategies that are analogous to those that are used for phage selections: binding complexes are retained by an immobilized target, non-binders washed away and specific complexes eluted and propagated [e.g. by amplification using polymerase chain reaction (PCR)] for subsequent rounds of selection or for cloning into a bacterial host to generate a clonal archive of retained genes (Fig. 2).

As previosuly discussed, because these libraries can be created without cellular transformation, they should be able to contain $>10^{12}$ peptide sequences, thereby creating a richer source of structural diversity from which to select ligands than is currently possible with phage display.

To date, it has been possible to use this technique to specifically enrich for covalent complexes displaying target-binding protein domains or linear peptides from excess backgrounds of non-binding complexes. Libraries consisting of linear peptides of random sequence fused to either the N- or C-terminus of P2A have also been generated.

Polysome display

Two companies, Cambridge Antibody Technology (Melbourn, UK) and MorphoSys AG (Martinsried, Germany),

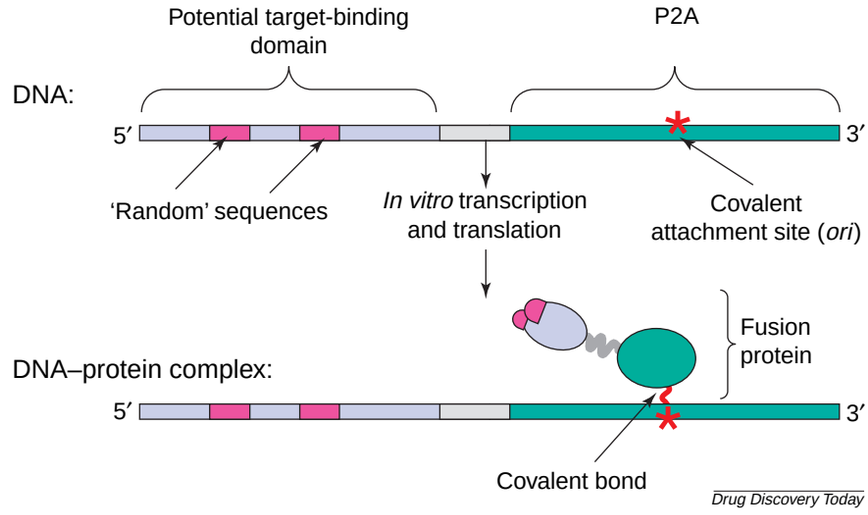

Figure 1. *Generation of covalent DNA–polypeptide complexes in CDT. P2A is a DNA-binding protein that attaches covalently to its own coding sequence. Concurrent in vitro transcription and translation of gene fusions comprising the coding sequence for P2A and the coding sequences for potential target-binding protein domains results in the formation of covalently joined DNA–protein complexes. The cis-activity of P2A facilitates the simultaneous production of a pool of DNA–protein complexes in which the faithful linkage of each gene and its own expressed product is achieved. The DNA–protein complexes can then be subjected to affinity selection against a target molecule to identify those rare library members that can bind specifically to the target.*

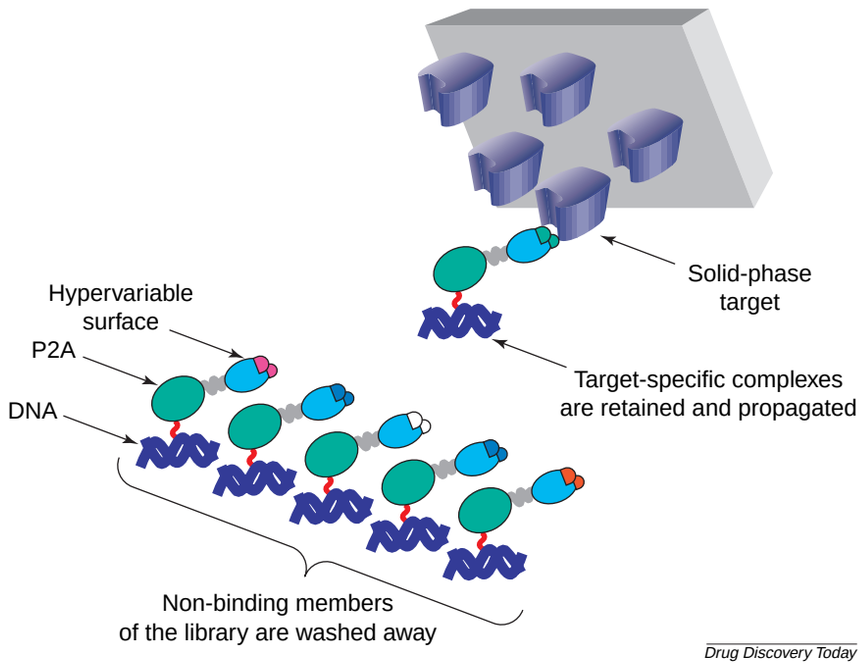

Figure 2. *Selection of CDT complexes against a solid-phase receptor. Affinity selection techniques involve the incubation of pools of covalent DNA–protein complexes with an immobilized target molecule. Binding moieties are retained on the solid surface while non-specific complexes are washed away. The physical link between the gene and the gene product results in the coincident retention of both target-specific polypeptide and its encoding DNA. The retained DNA is then amplified by PCR for further rounds of selection or used to generate a clonal archive of the binding domain's gene.*

are developing an *in vitro* technology, polysome display (reviewed in Ref. 30), for the discovery of novel therapeutic antibodies. The system works by transcribing and translating DNA templates *in vitro* under conditions that enable the isolation of stable mRNA–ribosome–nascent polypeptide complexes (Fig. 3). This is achieved by controlling the concentration of magnesium ions (to stabilize the ribosome particle) and by either terminating polypeptide elongation by the addition of chloramphenicol[31] or cooling down the translation products of mRNA templates that lack stop codons[32]. Target-specific polysome complexes are retained on an appropriately derivatized solid surface and the co-selected mRNAs released by dissociation of ribosomes using ethylene diamine tetraacetate (EDTA). These are then recovered by reverse transcription (RT) and PCR for further rounds of selection.

Ligands specific to an anti-dynorphin B monoclonal antibody have been successfully isolated using an *E. coli* S30 lysate to generate polysome complexes from an input library consisting of 10^{12} DNA molecules encoding random decapeptides[31]. Furthermore, by separating the steps of transcription and translation, the selection of polysomes displaying complex protein domains has been demonstrated. Single-chain antibodies that were specific for a monomeric variant of the yeast transcription factor GCN4 were selected from a library derived from mRNA that was prepared from 10^6 spleen cells taken from mice that had been immunized with this antigen[33].

The use of eukaryotic *in vitro* transcription/translation systems to generate functional polysome complexes has also been described[34]. Here, polysomes displaying an anti-progesterone antibody fragment were generated using a rabbit reticulocyte lysate and were successfully isolated from a mixture containing an excess of non-binding polysomes by selection against immobilized progesterone. Enrichments of 10^4–10^5-fold per round of selection were reported. However, a recent report has suggested that while the separation of the *in vitro* transcription and translation reactions facilitates the optimization of the redox conditions necessary for correct protein folding, neither eukaryotic nor prokaryotic translation systems appeared to provide any intrinsic advantage to polysome production[35].

RNA–peptide fusions

Phylos (Boston, MA, USA) is developing an *in vitro* display system in which a puromycin molecule is used to provide a covalent linkage between mRNA molecules and their encoded polypeptides[36]. Puromycin is an antibiotic that mimics the aminoacyl end of tRNA and functions by entering the ribosomal A-site and forming an amide linkage with nascent polypeptide through the peptidyl transferase activity of the ribosome[37,38].

In the RNA-peptide fusion system, the puromycin is attached to the 3′ end of a single-stranded DNA linker that is in turn ligated to the 3′ end of the library-encoding mRNA. When the mRNA is translated *in vitro*, a ribosome reaches the junction between the mRNA and the DNA linker and stalls. The puromycin can then enter the ribosomal A-site and form a stable amide linkage with the

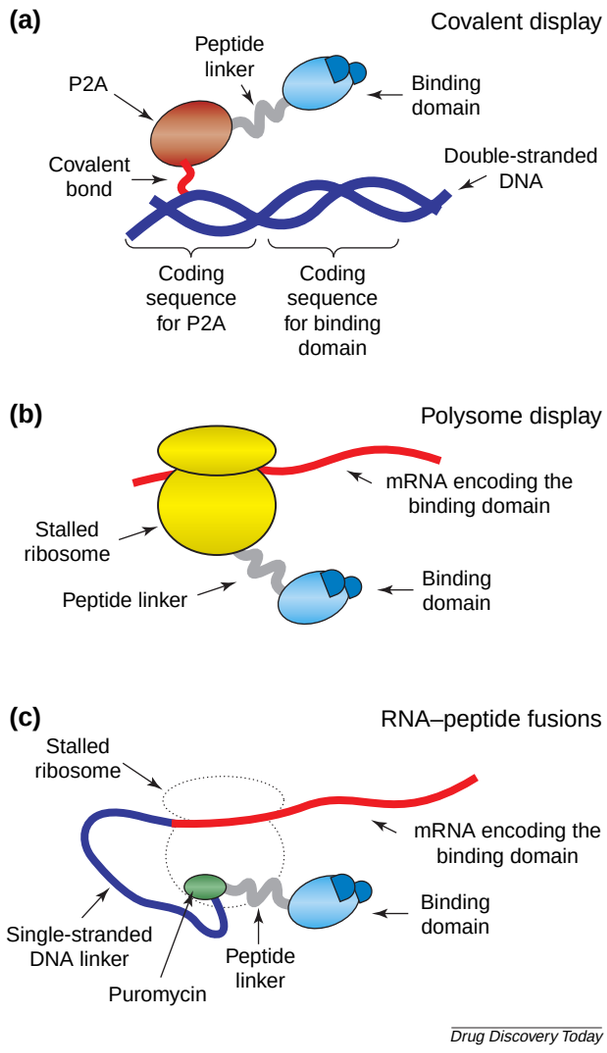

Figure 3. *In vitro-produced selectable complexes. In covalent display technology, the link between the gene and the gene product is provided by the covalent linkage that forms between a cis-acting DNA-binding protein (P2A) and its own gene sequence (a). In polysome display, the linkage is achieved by producing stable complexes of mRNA, ribosomes and associated nascent polypeptide (b). The RNA–peptides fusion system generates linkage between the gene and the gene product through a puromycin molecule that is attached via a single-stranded DNA linker to the 3′ end of the mRNA. The puromycin enters a ribosome that stalls at the DNA linker–mRNA junction and becomes covalently attached to the nascent polypeptide (c).*

encoded peptide (Fig. 3). A library pool of mRNA-DNA-puromycin molecules can therefore be translated *in vitro* and purified RNA–peptide complexes incubated with a solid-phase target molecule. As with the polysome display system, retained complexes are recovered for further manipulation by RT-PCR. The successful selection of complexes displaying the myc epitope when mixed with an excess of complexes displaying random peptide sequences has been achieved using immobilized anti-myc antibody as the target. Enrichments of 20–40-fold per round of selection were reported[36].

Conclusion

Molecular display technologies have great potential in providing the target-specific ligands that are continually required by companies engaged in the discovery of novel therapeutics, diagnostics, medical imaging compounds and research reagents. They also have the potential to provide the crucial interface between modern therapeutic target discovery (genomics, proteomics and bioinformatics) and the development of new and potent drugs. For example, the ability to generate molecular probes to the products of genes that have been identified by genomics companies as having a disease association should aid in the rapid validation of those gene products as viable therapeutic targets.

The *in vitro* technologies should be particularly well suited to this type of endeavour: the lack of a transformation or infection step during cycles of selection should make them more amenable to automation than is the case with phage display, therefore enabling ligand discovery to be conducted on a genomics scale. The non-dependence on *E. coli* protein expression could also make the *in vitro* display technologies more compatible with the evolution of novel antibiotics, an area of drug discovery receiving renewed interest as antibiotic resistance continues to spread rapidly among human pathogenic bacteria.

It remains to be seen which of the new *in vitro* display technologies achieves the widest adoption and yields the best returns on R&D investment. It is possible that the different systems will find independent niches, each providing some advantage in the solution of distinct technological problems. However, in both the polysome display and the puromycin systems, the genetic information is carried by mRNA molecules, which are notably difficult to manipulate because of their susceptibility to ubiquitous ribonucleases. Ribosome display also suffers from the potential dissociation of the tripartite RNA–ribosome–polypeptide complexes and the puromycin technology requires complex synthesis and attachment of the puromycin–DNA linker moiety to mRNA. Because of its user-friendliness, together with the highly stable nature of the protein–DNA complexes that are formed, CDT might therefore become the *in vitro* display technology of choice for many applications.

REFERENCES

1 Drews, J. (1996) Genomic sciences and the medicine of tomorrow. *Nat. Biotechnol.* 14, 1516–1518
2 Gordon, E.M. *et al.* (1994) Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J. Med. Chem.* 37, 1385–1401
3 Ellington, A.D. and Szostak, J.W. (1990) *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 346, 818–822
4 Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510
5 Geyson, H.M. *et al.* (1986) *A priori* delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* 23, 709–715
6 Huse, W.D. *et al.* (1989) Generation of a large combinatorial library of the immunoglobulin repertoire in phage λ. *Science* 246, 1275–1281
7 Young, R.A. *et al.* (1985) Dissection of *Mycobacterium tuberculosis* antigens using recombinant DNA. *Proc. Natl. Acad. Sci. U. S. A.* 82, 2583–2587
8 Ruther, U. and Muller-Hill, B. (1983) Easy identification of cDNA clones. *EMBO J.* 2, 1791–1794
9 Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317
10 Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. *Science* 249, 386–390
11 Cwirla, S.E. *et al.* (1990) Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. U. S. A.* 87, 6378–6382
12 Devlin, J.J. *et al.* (1990) Random peptide libraries: a source of specific protein binding molecules. *Science* 249, 404–406
13 McCafferty, J. *et al.* (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348, 552–554
14 McCafferty, J. *et al.* (1991) Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. *Protein Eng.* 8, 955–961
15 Bass, S. *et al.* (1990) Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* 8, 309–314
16 Rebar, E.J. and Pabo, C.O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 263, 671–673
17 Vaughan, T. *et al.* (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat. Biotechnol.* 14, 309–314
18 Perelson, A.S. and Oster, G.F. (1979) Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J. Theor. Biol.* 81, 645–670
19 Marks, J.D. *et al.* (1991) By-passing immunization: human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* 222, 581–597
20 Fuchs, P. *et al.* (1991) Targeting recombinant antibodies to the surface of *Escherichia coli*: fusion to a peptidoglycan associated lipoprotein. *Biotechnology* 9, 1369–1372
21 Boder, E.T. and Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* 15, 553–557
22 Kasahara, N. *et al.* (1994) Tissue-specific targeting of retroviral vectors through ligand–receptor interactions. *Science* 266, 1373–1376
23 Cull, M.G. *et al.* (1992) Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1865–1869
24 Schnos, M. and Inman, R.B. (1971) Starting point and direction of replication in P2 DNA. *J. Mol. Biol.* 55, 31–38
25 Geisselsoder, J. (1976) Strand-specific discontinuity in republicating P2 DNA. *J. Mol. Biol.* 100, 13–22
26 Chattoraj, D.K. (1978) Strand-specific break near the origin of bacteriophage P2 DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* 75, 1685–1689
27 Lindahl, G. (1970) Bacteriophage P2: replication of the chromosome requires a protein which acts only on the genome that coded for it. *Virology* 42, 522–533
28 Liu, Y. and Haggard-Ljungquist, E. (1994) Studies of bacteriophage P2 DNA replication: localization of the cleavage site of the A protein. *Nucleic Acids Res.* 22, 5204–5210
29 Liu, Y. *et al.* (1993) Studies of bacteriophage P2 DNA replication. The DNA sequence of the *cis*-acting gene A and ori region and construction of a P2 mini-chromosome. *J. Mol. Biol.* 231, 361–374
30 Schaffitzel, C. *et al.* (1999) Ribosome display: an *in vitro* method for selection and evolution of antibodies from libraries. *J. Immunol. Methods* 231, 119–135
31 Mattheakis, L.C. *et al.* (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. U. S. A.* 91, 9022–9026
32 Hanes, J. and Pluckthun, A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4937–4942
33 Hanes, J. *et al.* (1998) Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14130–14135
34 He, M. and Taussig, M.J. (1997) Antibody–ribosome–mRNA (ARM) complexes as efficient selection particles for *in vitro* display and evolution of antibody combining sites. *Nucleic Acids Res.* 25, 5132–5134
35 Hanes, J. *et al.* (1999) Comparison of *Escherichia coli* and rabbit reticulocyte ribosome display systems. *FEBS Lett.* 450, 105–110
36 Roberts, R.W. and Szostak, J.W. (1997) RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* 94, 12297–12302
37 Monro, R.E. and Marcker, K.A. (1967) Ribosome-catalysed reaction of puromycin with a formylmethionine-containing oligonucleotide. *J. Mol. Biol.* 25, 347–350
38 Monro, R.E. and Vazquez, D. (1967). Ribosome-catalysed peptidyl transfer: effects of some inhibitors of protein synthesis. *J. Mol. Biol.* 28, 161–165