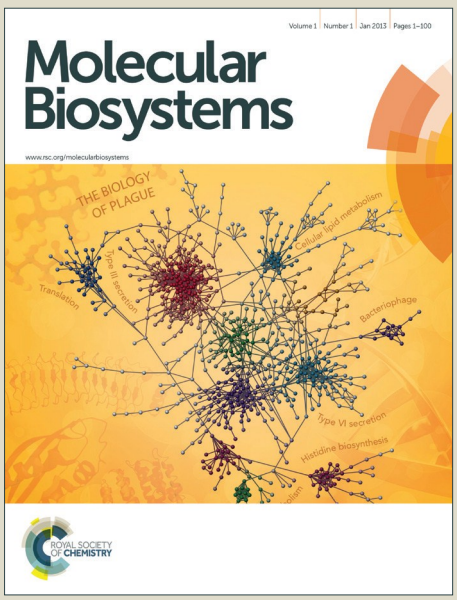
Molecular BioSystems

Accepted Manuscript

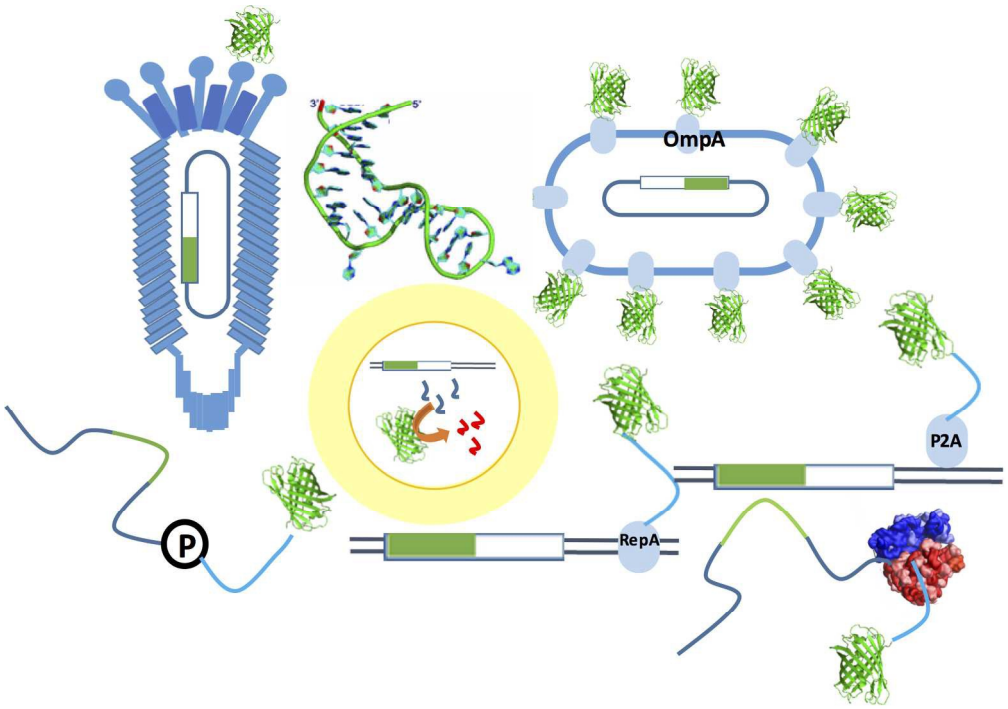
This article can be cited before page numbers have been issued, to do this please use: A. Galan, L. Comor, A. Horvati, N. Guillemin, J. Kuleš, V. Mrljak and M. Bhide, *Mol. BioSyst.*, 2016, DOI: 10.1039/C6MB00219F.



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

View Article Online
DOI: 10.1039/C6MB00219F



177x125mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6MB00219F

Library-based display technologies: where do we stand?

Authors

Asier Galán[1], Luboš Comor[2,3] , Anita Horvatić[1], Josipa Kuleš[1], Nicolas Guillemin[1], Vladimir Mrljak[1], Mangesh Bhide[1,2,3]

Affiliations

[1] ERA Chair FP7, Internal diseases, Faculty of Veterinary Medicine, University of Zagreb; Heinzelova 55, 10 000 Zagreb, Croatia

[2] Laboratory of Biomedical Microbiology and Immunology - University of Veterinary Medicine and Pharmacy, Kosice, Slovakia,

[3]Institute of Neuroimmunology, Slovakia Academy of Sciences, Bratislava, Slovakia Kosice , SK

Corresponding author

Asier Galán, ERA Chair FP7, Internal Diseases, Faculty of Veterinary Medicine, University of Zagreb, Heinzelova 55, 10 000 Zagreb, Croatia;

E-mail : agcousillas@vef.hr

Running title

Library based technologies: an overview

Key words

Library display, phage display, mRNA display, CIS display, Covalent Antibody display (CAD), ribosome display, directed evolution, affinity maturation, biopanning

View Article Online
DOI: 10.1039/C6MB00219F

Abstract

Over the past two decades, library-based display technologies have been staggeringly optimized since their appearence in order to mimick the process of natural molecular evolution. Display technologies are essential for the isolation of specific high-affinity binding molecules (proteins, polypeptides, nucleic acids and others) for diagnostic and therapeutic applications in cancer, infectious diseases, autoimmune, neurodegenerative, inflammatory pathologies etc. Applications extend into other fields as antibody and enzyme engineering, cell-free protein synthesis and discovery of protein-protein interactions. Phage display technology is the most established of these methods but more recent fully *in vitro* alternatives such as ribosome display, mRNA display, cis-activity based (CIS) display and covalent antibody display (CAD), as well as aptamer display and *in vitro* compartmentalization, offer advantages over phage in library size, speed and the display of unnatural amino acids and nucleotides. Altogether, they have produced several molecules currently approved or in diverse stages of clinical or preclinical testing and have provided researchers with tools to address some of the disadvantages of peptides and nucleotides such as their low affinity, low stability, high immunogenicity and difficulty to cross membranes. In this review we assess the fundamental technological features and point out some recent advances and applications of display technologies.

List of abbreviations

AIDA-I Adhesin Involved in Diffuse Adherence-I
BG Benzylguanidine
CAD Covalent Antibody Display
cDNA Complementary DNA
CFPS Cell Free Protein Synthesis
CIS display *Cis*-activity based display

View Article Online
DOI: 10.1039/C6MB00219F

CotG Spore coat protein G

GFP Green Fluorescent Protein
GPI Glycophosphatidylinositol
GVNP Gas Vesicle NanoParticle
InaK Ice nucleation protein K
IVC *In vitro* compartmentalization
Lpp-ompA Lipoma prefered partner-outer membrane protein A hybrid
MACS Magentic activated cell sorting
NGS Next Generation Sequencing
PCR Polymerase Chain Reaction
PRM Protein– Ribosome–mRNA
RBS Ribosome Binding Site
SELEX Systematic Evolution of Ligands by EXponential enrichment
SOMAmer Slow Off-rate Modified Aptamer

Introduction

Mimicking nature by performing directed evolution in the laboratory has shown very promising perspectives[1]. Natural evolution as well as directed evolution require compartmentalization and/or physical linkages between genotype and phenotype[2]. *In vitro,* that linkage is achieved through the physical binding of polypeptides/proteins of interest to their cognate genes, or non-coding DNA or RNA, or by using the inherent ability of oligonucleotides to bind other molecules in the case of aptamer libraries. Expression of proteins on the surface of bacteria[3] or yeast[4] has been used successfully, however the display of peptides[5] and polypeptides[6] on the surface of filamentous phages (Phage display) is still the most widely used display technology. Many types of acellular approaches have also been developed in which the presence of cells is not required. For example, ribosome display, mRNA display, aptamers, *in vitro* compartmentalization (IVC), cis-activity based (CIS) display and covalent antibody display (CAD) are, among others, leading *in vitro* technologies. They are not restricted by the effort spent on the cell transformation step. This advantage is apparent in the primary library, which can be of bigger size, as all DNA (or mRNA) molecules present can in principle give rise to proteins that take part in

View Article Online
DOI: 10.1039/C6MB00219F

the selection. The most important advantage of full *in vitro* technologies is the easy combination with PCR-based randomization techniques, and thus the utilisation of a Darwinian evolution process – in contradistinction to a mere selection from an existing "constant" library, *in vitro* methods give the user full control over where mutations should occur in the sequence (by using, *e.g.,* a randomized cassette), which residue types are to be introduced, or how many random mutations should occur on average[7], a level of control not yet within reach in cellular systems.

In vitro display and selection approaches involve three main steps: (i) the generation of a large collections of variants (a library); (ii) multiple rounds of enrichment of variants (biopaning) displaying the desired properties via the genotype–phenotype linkage provided by the display system used; and (iii) functional screening and characterization of selected variants using appropriate assays (Fig 1). At each of these steps, analysis of variants via sequencing is commonly used to control the process. To generate libraries with high sequence diversity, recombination-based methods are used (Fig 1). The combination of error-prone PCR and gene shuffling *in vitro* in subsequent cycles can yield an increased population of affinity-improved variants, as compared with error-prone PCR only[8]. Additionally, fully *in vitro* displays offer the possibility of modifying the nucleic acid libraries with aminoacids and non-natural mononucleotides to improve affinity and specificity or other features (as biocompatibility) of the selected molecules. Both cellular and acellular approaches have produced a plethora of antibodies, proteins and peptides with a wide spectrum of therapeutic and diagnostic (theranostic) applications (Table 1).

In recent years, the development of next generation sequencing (NGS) technologies has profoundly impacted multiple aspects of biological research[9], among them the display and selection process of proteins with desired properties. At the

View Article Online
DOI: 10.1039/C6MB00219F

library generation stage, it is crucial to cover as much sequence and structural diversity as possible to increase the likelihood of including protein variants with desired properties. The diversity of phage display libraries typically lies between 10^7 and 10^{12} and it reaches $10^{13\text{-}15}$ in acellular display technologies (Table 2). A major limitation using Sanger sequencing is that only a minute fraction of the library is actually sampled (ten to few hundreds out of $> 10^7$ clones). The large number of sequencing reads delivered by NGS technologies ($>10^6$) offers the capacity to obtain sequence information on far more if not all clones at each round, producing a comprehensive analysis of the selection process and potentially boosting the reach of display technologies in next future.

In summary, all these streamlined procedures for identification of ligand-receptor pairs that could be used as targets in disease diagnosis, therapy, profiling, imaging and other applications have relied on the display technologies, in which polypeptides or nucleic acids with desired binding properties can be iteratively selected by biopanning. In this review only display occurring outside of the cell will be considered and technological features, advantages, limitations and projection of cellular and acellular displays will be addressed.

1. Cellular approaches

1.1 Phage display

Bacteriophages are viruses that infect bacterial cells widely used for protein/peptide engineering[10]. In such systems genotype-phenotype linkage results from their ability to express peptides or proteins fused to their coat proteins while genes coding for those proteins are encapsulated inside the phage. All phage coat proteins have been adapted to display foreign peptide and protein fusions. While the major coat protein pVIII, most abundant protein on the phage coat (2700 copies), can display multiple copies of small peptides, pIII, which is at the tip of the phage and is

View Article Online
DOI: 10.1039/C6MB00219F

essential for infection, has been shown to be more suitable for the display of larger protein fusions[11].

Originally whole phages were used for iterative selection (biopanning), mainly bacteriophage M13 and fd filamentous phages[12], but due to higher transfection efficiency phagemids (plasmids containing an f1 origin of replication from a f1 phage) are more widespread[10]. If phagemid vectors are used, the process is similar, though a helper phage (*e.g.* M13KO7) that provides all structural proteins is necessary[13]. This is because phagemids consist only of replication origin of a plasmid, selective marker, intergenic region, segments of DNA encoding coat protein and a packaging signal utilized by helper phage[14]. Need for helper phage is avoided if cell lines containing phage packaging proteins are used[15]. Additionally, T7 lytic phage systems have been developed that circumvent host cell synthesis and secretion mechanisms preventing protein folding problems[16] .

In phage library selections biopanning (Fig 2) applies specific pressure to select a subpopulation of peptides from the library that bind a target molecule, which is usually immobilized onto a surface. Diversity (library size) is important as the majority of the population is lost from the system under stringent washing conditions, leaving bound peptides and their associated phage particles to be enriched by infection and re-growth in bacteria. Phage display has revolutionized the ability to engineer larger proteins, in particular antibodies[17] used for therapeutics and countless other applications (Table 1).

Construction of libraries

View Article Online
DOI: 10.1039/C6MB00219F

Phage libraries usually contain billions of phage particles, some of which have as high as 10^{12} diverse particles. Not all protein or peptide libraries can be selected by phage display because some sequences may be sensitive to bacterial enzymes or toxic to *E. coli* and cause interference with the assembly of phage. Depending on the goal of the studies, two types of libraries are now extensively used for novel drug discovery, theranostic applications, biosensing, bioimaging...

Peptide libraries

Random phage peptide libraries are one of the most common types of phage display constructs. Using degenerate oligonucleotides introduced into the phage genome, the linear random peptide library can be obtained. The random oligonucleotides are cloned between the coding sequence of the signal peptide and the N-terminus of the protein pIII. Random peptide libraries varying in length from 6 to 43 amino acids have been successfully cloned and expressed as peptide-capsid fusion proteins. One of the most common approaches to construct random peptides is to use (NNK)n codon degeneracy, where N indicates an equimolar mixture of all four nucleotides (A, G, C and T), and K indicates a 1:1 mixture of G and T. By adding an amber stop codon (TAG) at the beginning of gene III of phage genome, monovalent peptide libraries can be generated. Although the phage library may contain as many as 10^{12} phage particles, the real diversity of the library is much lower. This is because *E coli* transformation efficiency has a limit of 10^9 per microgram of DNA so the diversity of transformed clones is typically around 10^9 and only 1-2% of phage particles display the protein target in the M13 phage system.

Antibody libraries

View Article Online
DOI: 10.1039/C6MB00219F

The conventional antibody is comprised of two identical heavy (H) and light (L) chains inter-connected by disulfide bridges. The chains composed of constant (C) and variable (V) domains, combine to form one interaction site for the antigen. The Fab consists of the L chain and two domains of the H chain located at its N terminus (V_H and C_{H1}). The single chain fragment variable (scFv) is a small engineered fragment composed of two variable domains with a short flexible glycine-rich linker peptide between V_H and V_L domains. Due to higher structural stability and diminished aggregation, Fab fragments, instead of the scFv, have been used for library construction in many of the more recent phage display systems.

In early 90s, naturally occurring antibodies showing no light chains were discovered in *Camelidae*[18] and later in *Chimaeridae*[19]. These antibodies are named heavy-chain antibodies (HcAbs) or nanobodies, they consist only of C_{H2}, C_{H3} and V_H regions and they can constitute up to 75% of all immunoglobulins in *Camelidae*[18]. In the HcAbs higher frequency than in conventional antibodies of the hotspots for mutation exists[20] and they have been developed for diverse therapeutic application (Table 1).

Production of antibody libraries

In case of active immunization, the animal is inoculated with antigen, after 5-6 weeks the memory B-cells and plasma cells are isolated and mRNA is extracted, and reversely transcribed to cDNA. From this cDNA, the encoding sequence of V_H domains is later amplified and used to construct the antibody library. Another method, naïve library construction, uses naïve B-cells from non-immunized animals as source of mRNA. Naïve libraries show advantages in the construction of antibodies for all antigens with a single library and antibodies to self, non-immunogenic or toxic substances.

View Article Online
DOI: 10.1039/C6MB00219F

Synthetic libraries are based on *in silico* construction of variable domains with already known frameworks used as a scaffold[21], while semi-synthetic libraries combine naïve and synthetic libraries. Transcribed cDNA is used as a template for *in silico* randomization.

1.2 Cell surface display

Cell-surface display is the expression of peptides and proteins on the surface of living cells or their organelles by fusing them to functional components of cells which are exposed to the environment. This can be performed using different surface proteins of cells as anchoring motifs and different proteins from different sources as a passenger protein. The principle behind cell display technology is analogous to phage display[22], with the recombinant protein anchored on the surface and the encoding DNA encapsulated inside the cell.

Cell display for combinatorial protein engineering purposes takes advantage of particle size and the multivalent surface expression of recombinant proteins[4]. Unlike in phage and ribosome display, in which selection from libraries is generally based on a capture and elution procedure, selections from cell-displayed libraries are typically performed using fluorescence-activated cell sorting (FACS) by labeling the antigen with a fluorophore and subsequently incubating it with cell-displayed protein library in solution. Quantification of the affinity during selection is thus an important feature of cell- based display systems, making the process more similar to the screening of micro-well plates than the biopanning in phage and ribosome display[23].

Cell surface display has been developed for the surfaces of bacteria, yeast, spores and even gas vesicles. Peptide or protein to be displayed is fused with a protein or signal peptide that ensures the sorting to the corresponding surface. In the case of gram-positive bacteria (Fig 3, A), surface proteins are synthesized as precursors with

View Article Online
DOI: 10.1039/C6MB00219F

an N-terminal signal peptide and a C-terminal LPXTG motif sorting signal. Following initiation into the secretory pathway, the signal peptide is cleaved, while sortase A scans translocated precursors for LPXTG motif sequences. Following cleavage of the precursor by sortase A, an acyl-enzyme intermediate is formed between the active site cysteine (C) of sortase A and the carboxyl-group of threonine at the C-terminal end of the surface protein. The amino group of cell wall crossbridges within lipid II, pentaglycine (Gly5) in *S. aureus*, forming an amide (isopeptide) bond between the C-terminal threonine and lipid II[24].

In the case of Gram-negative bacteria (Fig 3, B) many surface proteins, for example, Lpp-ompA, InaK and AIDA-I from *Escherichia coli*, have been used for displaying target proteins, such as antigens, enzymes and bioadsorbents. Autotransporters have emerged as a convenient tool for bacterial surface display. These proteins are composed of an N-terminal signal peptide, followed by a passenger domain and a translocator domain that mediates the outer membrane translocation of the passenger. The natural passenger domain of autotransporters can be replaced by heterologous proteins that become displayed at the bacterial surface[25].

In the yeast *Saccharomyces cerevisiae* (Fig 3, C), α-agglutinin (Agα1), a-agglutinin (Aga1p-Aga2p), or flocculin have been used as an anchor protein. These proteins exist on the yeast cell surface and have glycosylphosphatidylinositol (GPI) anchors that play important roles in the surface localization of proteins. The GPI-anchored proteins translocate to the cell surface through the secretory pathway of *S. cerevisiae.* The GPI attachment signal is assigned to the C-terminus of GPI-anchored protein[26]. MACS (Magentic Activated Cell Sorting) technology makes use of magnetic beads coated with streptavidin for the selection of antibody libraries

View Article Online
DOI: 10.1039/C6MB00219F

displayed on the yeast cell. In such case a biotin conjugated antigen is used for the selection [27].

Bacterial spores have also been used in surface display technology. Formed in response to nutrient starvation, spores are robust and can withstand extremes of heat, desiccation, and chemicals. The spores of *Bacillus subtilis* are encased in a protective coat consisting of an inner and an outer layer, which are composed of >70 different proteins. Thus, generating fusion proteins with coat proteins is an easy display method. CotG spore coat protein has been used to display *E. coli* b-galactosidase (b-Gal) and other proteins including GFP, carboxymethylcellulase and lipase. Highly expressed proteins may also be adsorbed to the spore surface via hydrophobic or electrostatic protein–protein interactions circumventing the need of generating fusion proteins[28].

For display on gas vesicles, gas nanoparticles are produced by microorganisms like the halophilic archaeon *Halobacterium sp*. NRC-1 as gas-filled buoyant organelles, which are easily purified as gas vesicle nanoparticles (GVNPs). GVNPs are non-toxic, exceptionally stable, bioengineerable, and self-adjuvanting. One protein, GvpC, found on the exterior surface of the nanoparticles, can accommodate insertions near the C terminal region and results in GVNPs displaying the inserted sequences on the surface of the nanoparticles[29].

The main advantage of cell-surface display is that the cell is large enough to be analyzed and sorted using flow cytometry. In addition, the high polyvalency, with expression levels from a few hundred to several hundred thousand proteins displayed per cell, allows for sorting in a truly quantitative manner. Advantages of cell surface display in the case of eukaryotic cells over most other display systems are its ability

View Article Online
DOI: 10.1039/C6MB00219F

to express mammalian proteins with high fidelity and with some level of post-translational modification.

Disadvantages include smaller mutant library sizes than other displays and differential glycosylation in yeast compared with mammalian cells. One of the highest monovalent ligand-binding affinity reported to date for an engineered protein has been obtained through yeast surface display[30]. Surface-engineered cells and spores have many potential uses ranging from therapeutics, including vaccines[31, 32], precise antibody epitope mapping[33], production of aglycosylated therapeutic antibodies[34] and others (Table 1) to novel enzyme assays[35] and environmental applications[36].

2. Acellular approaches

2.1 Ribosome display

Ribosome display is carried out fully *in vitro*, which overcomes some of the limitations of cell-based display systems. Ribosome display was originally developed for *in vitro* selection and optimization of antibody combining sites[37] and extended for a wider range of proteins.

Ribosome display connects nascent proteins to their encoding mRNA through the generation of stable PRM (protein– ribosome–mRNA) complexes. This link enables the isolation of a specific protein, together with its encoding mRNA, through the selection by ligand binding and affinity maturation[38] (Fig 4). A ribosome display construct designed for use with *E. coli* extracts, typically consists of a T7 promoter that allows mRNA synthesis, followed by a ribosome binding site (RBS) that can base-pair with ribosomal RNA thus recruiting the ribosome to the downstream start codon where protein synthesis is initiated. Depending on the origin of the cell-free extract, a Shine-Dalgarno (prokaryote) or Kozak (eukaryote) sequence is used to

View Article Online
DOI: 10.1039/C6MB00219F

initiate translation. The open reading frame frequently starts with a protein detection tag such as the RGS-His6 tag or the FLAG tag, followed by the library of binding proteins and a spacer protein. The spacer should provide the displayed library the flexibility to fold outside of the ribosome tunnel (at least 23-30 amino acids)[39]. The usual method for coupling nascent polypeptide with its encoding mRNA is by stalling the translating ribosome at the 3'-end of the mRNA through deletion of the stop codon. Alternative methods for forming PRM complexes include the use of antibiotics or the fusion of the gene encoding the protein of interest to a C-terminal ricin A chain which inactivates, consequently stalling, the translating ribosome as it is translated[40]. Being a cell-free display, ribosome display has the potential to screen very large populations with library sizes of 10^{12}–10^{14} members in a few PCR reactions (Table 2), whereas at least 1000 transformations would be required to achieve a comparable library size by cell-based methods[41].

Transcription and translation in ribosome display can be performed simultaneously in a one-step process, with coupled transcription and translation in the same conditions, or separately in two steps uncoupled transcription and translation. The one-step system usually consists of a eukaryotic rabbit reticulocyte or a prokaryotic *Escherichia coli* S30 extract[42] for transcription and translation. This strategy avoids problems of mRNA degradation through the use of prepared extracts based on the Rnase I deficient strain *E. coli* MRE600[43]. Exo- and endonucleases originating from natural extracts rapidly degrade library mRNA during the *in vitro* translation procedure at 37°C during ribosome display, and PRM complexes need to be kept cold in the presence of RNase inhibitors. Additionally, *E. coli* shows an SsrA RNA-mediated mechanism to rescue stalled ribosomes and tag the translated protein

View Article Online
DOI: 10.1039/C6MB00219F

for proteolytic degradation. To neutralize this reaction, an oligonucleotide complementary to SsrA RNA can be added to the *E. coli* extract.

A widespread alternative is the use of *E. coli*-based system reconstituted from recombinantly produced and purified components (initiation, elongation and release factors, aminoacyl tRNA synthetases, T7 RNA polymerase and tRNAs) and *E. Coli* ribosomes. This systems are practically RNase free.

Proteins containing disulfide bridges, such as immunoglobulins, in general, have difficulties to fold correctly under the reducing conditions of *E.coli* extracts and a two-step procedure with uncoupled transcription and translation in eukaryotic extracts or reconstituded systems is preferred. Firstly, the DNA is transcribed *in vitro* and is then purified. The mRNA can then be translated in a rabbit reticulocyte lysate or a wheat germ extract [44]. As transcription and translation reactions are performed separately, each can be optimised for the amount of DNA and RNA as well as buffer compositions.

Novel in vitro yeast ribosome display methods based on reconstituted cell-free protein synthesis (CFPS) have been successful in resolving folding and low yield problems[45].

Reconstituted translation reagents may increase the functional library size versus use of traditional cell extracts for translation and display by the omission of peptide release and ribosome recycling factors that may further stabilise ribosome complexes and increase functional library size[46]. Additional favourable conditions over crude extracts would be the reduced abortive energy consumption (spent in incomplete translation) and the lack of unknown inhibitors present in the S30 cell extract. Disadvantages of reconstituted transcription and translation systems are often lower efficiency due to absence of cellular factors capable of enhancing protein translation

View Article Online
DOI: 10.1039/C6MB00219F

and folding. In general, eukaryotic systems, reconstituted or extracts, are able of expressing more proteins with more complex folding than prokaryotic ones and to produce a broader variety of post-translational modification [47].

In vitro selection techniques in ribosome display

In the case of ribosome–mRNA–antibody complexes, selection and maturation are performed by incubation at 4°C with decreasing concentrations of Fc-tagged, purified antigen before a capture step with Protein G coated magnetic beads or microtitre plates[41]. Another possibility is to use the streptavidin/biotin capture but it can lead to random biotinylation and loss of protein function. Ribosome display technology for single chain variable fragment (scFv) optimisation[37] has one of the most powerful display ranges (typically 10^{12} molecules) and the advantage of being performed entirely *in vitro*.

2.2 mRNA/cDNA display

mRNA-display (or cDNA after reverse transcription) is an *in vitro* selection technology based on the physical linkage of a peptide to a nucleic acid tag (the mRNA that encoded it) that can be amplified by PCR and read by DNA sequencing[48]. It is a fully *in vitro* selection technique that allows for the identification of polypeptide sequences with desired properties from a natural protein library or a combinatorial peptide library. The key to this technology is the antimicrobial natural product puromycin, which inhibits translation by mimicking the substrate of the ribosome (the 3'- end of an aminoacyl-tRNA). The structure of puromycin resembles

View Article Online
DOI: 10.1039/C6MB00219F

the amino acid tyrosine linked via a stable amide bond to the 3' carbon (Fig 5) of a modified adenine nucleoside. In mRNA-display, an mRNA pool is modified with puromycin on its 3' end, and then translated in an *in vitro* translation reaction. As ribosomes complete the translation of individual mRNAs to the corresponding peptides they encounter the 3' puromycin. Because puromycin is chemically similar to the 3' end of aminoacyl-tRNA, it is recognized by the peptidyl transfer center of the ribosome, which catalyzes the transfer of the nascent peptide to the modified tyrosine of puromycin. The mRNA is then covalently attached to the corresponding translated peptide via the puromycin, and the ribosomes are stalled. The initial pool or library mRNAs have been translated and linked via puromycin to the peptides that they encode in a stable molecular conjugate referred to as an mRNA–peptide fusion. In two separate steps, the peptide and the mRNA component of these fusion molecules are further modified before being subject to *in vitro* selection. The mRNA is reverse transcribed to cDNA for amplification by PCR during *in vitro* selection. The amplified cDNA, which has been enriched in sequences that at the peptide level bind the target, can then be used as the template for transcription to regenerate mRNAs. This cycle, from mRNA transcription, through translation, peptide cyclization and reverse transcription, *in vitro* selection against an immobilized target, to PCR, is commonly referred to as one round of selection. Even a relatively stringent selection might require between five and seven rounds to converge sufficiently on the tightest binders from libraries of this size.

As a peptide *in vitro* selection technology, mRNA-display is conceptually similar to phage display, however, there are three main differences: 1. All of the steps of mRNA-display are entirely *in vitro*, therefore, library size is not limited by the need to transform bacteria. Transformation steps impose an upper working limit on

View Article Online
DOI: 10.1039/C6MB00219F

the diversity of phage display libraries of $10^{9,49}$. By contrast, peptide mRNA-display libraries can be as large as 10^{13} (Table 2). By accessing larger libraries, extremely rare sequences (such as long, discontinuous epitopes or peptides with better functional properties) can be selected and amplified[50]. 2. mRNA-display is monovalent, with only one copy of the peptide displayed on a given mRNA, allowing the displayed peptides to be enriched based solely on their intrinsic target affinity and circumventing avidity effects observed in phage display due to the presence of multiple copies of the peptide on the phage surface[51]; 3. The synthesis of the peptide–mRNA fusion molecules can be done in a modified *in vitro* translation reaction that can be reconfigured with a modified set of amino acids[52].

The lability of mRNA, however, in both ribosome and mRNA displays has restricted the experimental selection conditions and cDNA display method has been developed to improve the stability by converting mRNA to cDNA with a novel puromycin-linker. Although the cDNA display method was useful for *in vitro* peptide and protein selection, its productivity was hindered by the low generation of mRNA/cDNA-protein fusion complexes, only around 0.1% of the initial mRNAs[53].

2.3 Aptamer library display

In the past 25 years, aptamers technology has increased in both published works and number of applications[54]. The word “aptamer” was derived from the Latin *aptus*, meaning “to fit”, and Greek *meros*, meaning “region”[55]. Aptamers are essentially short RNA or single-stranded DNA oligonucleotides (usually 20–80 nucleotides with 6–25 kDa molecular weights) that can fold into unique three-dimensional conformations. Similar to conformational antibody–antigen interactions, aptamers, often referred to as 'chemical antibodies', bind to their targets with high-

View Article Online
DOI: 10.1039/C6MB00219F

specificity and affinity through van der Waals forces, hydrogen bonding, electrostatic interactions, stacking of flat moieties and shape complementarity[56], with dissociation constants (Kd) usually ranging from pico- to nanomolar[57] (Table 2).

Aptamers have superior applicability for clinical practice over antibodies in several aspects: (i) Aptamers are almost completely nonimmunogenic and nontoxic *in vivo*[58]; (ii) Aptamers do not possess redundant Fc regions that can bind to Fc receptors leading to unpredictable side effects[59]; (iii) Due to their smaller size, aptamers can penetrate into tissue barriers and be internalized by their target cells, enhancing their therapeutic potential[60]; (iv) Aptamers can be developed against a virtually unlimited range of targets, including small inorganic ions, small organic molecules, peptides, proteins, enzymes, organelles, complex cells and tissues. Furthermore, aptamers have important properties that simplify their industrialization; (v) They are thermally stable, so they can be stored and transported easily; (vi) Based on well-established chemical synthesis and modifications a given aptamer can be produced and modified in large scale, with minimal batch-to-batch variation and in a short time (hours); (vii) Chemical production process of an aptamer is not affected by viral of bacterial contamination[58]; (viii) The lack of large hydrophobic cores of proteins avoids aggregation of aptamers; (ix) Unlike antibodies, aptamers do not show cross-reactivity and/or false positive signal issues and (x) aptamers can be readily synthesized in large quantities at relatively low cost.

On the other hand, aptamers present variable pharmakokinetic properties often difficult to predict[61], because of their small size they are susceptible to renal filtration and unmodified aptamers are more likely to be degraded *in vivo* by DNases and RNases. To overcome those limitations aptamers are often optimized for activity and persistence under physiological conditions during or after selection by means of

View Article Online
DOI: 10.1039/C6MB00219F

conjugation with polyethylene glycol, cholesterol or nanostructures (e.g. gold and silver nanoparticles)[62].

So far, only Macugen (pegaptanib sodium), an RNA aptamer specific against vascular endothelial growth factor (VEGF), has been approved by the US Food and Drug Administration (FDA). Several other therapeutic aptamers are being evaluated in clinical trials for their effectiveness and safety in treating this and other disorders involving coagulation, oncology, and inflammation[63] **(**Table 1**).**

In Vitro selection of aptamers, SELEX

SELEX (Systematic Evolution of Ligands by EXponential enrichment) is the gold-standard methodology to develop aptamers. The conventional SELEX process includes multiple rounds of exponential amplification and enrichment, which allows directed evolution of aptamers with high target specific affinity starting from a random oligonucleotide pool[64] (Fig 6). Although the conventional SELEX technique and dozens of variations have been used successfully to develop aptamers against hundreds of targets, including cutting edge microfluidic-based techniques[65] , aptamers developed through SELEX techniques are not always effective for *in vivo* applications because target molecular conformations are usually different *in vitro* and *in vivo*. To close this gap, a modified SELEX technology, using whole living cells as targets, Cell-SELEX[66], and others using tissues and whole organisms have been developed recently[67].

Most of the earlier SELEX studies used the nitrocellulose filter binding method for aptamer selection due to its non-specific affinity for amino acids. The unbound aptamers pass through the filter and are removed from the library while

View Article Online
DOI: 10.1039/C6MB00219F

proteins, most frequent ligands in earliest SELEX developments, remain attached to the membrane. However, the efficiency of protein capture varies from protein to protein as well as between different experimental conditions, other target molecules (*e.g.*, small molecules) cannot be captured by the filter and certain aptamers are known to bind the nitrocellulose filter (aptamers with a multi-G motif)[68].

Affinity chromatography was also used successfully for *in-vitro* selection of aptamers. Many different affinity tags and affinity resins for each of them (6xHis-, GST-, or MBP- tags and Ni-NTA, Glutathione, Amylose resins, respectively) and coupling chemistries (Amine-, Thiol-, Carboxyl- or EDC, 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide[69]) make this strategy attractive and applicable to various target molecules. During the binding and elution steps of SELEX, target molecules that are pre-immobilized on a resin are incubated with the aptamer library. Alternatively, target molecules in solution can be incubated with the aptamer library and later be captured by the affinity resin. Microcolumns provide a version of affinity chromatography-based SELEX that minimizes the amount of resin and aptamer and can be easily multiplexed[70].

The conventional SELEX techniques for aptamer isolation are often manually manipulated iterative processes and require multiple successive cycles of selection and amplification. Typically it would take weeks to months to complete one selection. To overcome this problem, microfluidic SELEX (M-SELEX) was developed, which exerts stringent selection pressure by employing low amounts of target molecules and continuous washing to isolate high affinity aptamers in fewer rounds of selection[71].

Another rapid method developed is capillary electrophoresis SELEX. This SELEX enables affinity maturation of aptamers within 2–4 rounds, unlike many other screening methods. In contrast to other SELEX methods, the random library is freely

View Article Online
DOI: 10.1039/C6MB00219F

incubated with the target molecule in the binding buffer, facilitating the capture of the targets in their native three-dimensional structures, without using labels, separating columns or resins. The incubation medium is later loaded into a capillary electrophoresis system in which the target-bound aptamer candidates and the free oligonucleotides separate. Library diversity is though limited by the injection volume in capillary systems (usually 5–50 μl).

The major breakthrough in the SELEX technology was the development of cell- and tissue-SELEX, which enabled the selection of aptamers against molecules in their real biological state. The targets of cell- and tissue-SELEX are extracellular proteins on the cell surface or unique structures of the cell. In cell-SELEX, washing or centrifugation steps and counter selections (with control cells as target) are necessary during the separation of aptamers to avoid selection of aptamers that non-specifically recognize the cell surface. However, the resulting aptamers, are powerful for cell-specific diagnosis, cell targeted drug delivery, and cell-specific therapy (Table 1). The advantages offered by cell-SELEX are: 1. Prior knowledge about the molecular features of target cells is unnecessary since cell SELEX generates aptamers that can recognize and differentiate the molecular signatures of cells; 2. In cell-SELEX each molecule is a potential target and successful selections will generate a panel of aptamers for many different targets providing data for accurate disease diagnosis as well as new opportunities for personalized medicine; and 3. Unlike in other SELEX, libraries bind to the native state of target molecules, making it possible for aptamer probes to directly recognize their cognate target, creating a true molecular profile of diseased cells.

The SELEX with chemically-modified nucleotides has been successfully developed in last years. The first innovation in developing the 'slow off-rate modified

View Article Online
DOI: 10.1039/C6MB00219F

aptamer' (SOMAmer) was motivated by the idea that aptamers can be endowed with protein-like properties by adding functional groups that mimick amino acid side-chains to expand their chemical diversity[61]. SOMAmers interact with more hydrophobic surfaces compared with conventional aptamers, increasing the range of epitopes that are available for binding. Recently SOMAmers against over 3,000 human proteins (growth factors, cytokines, enzymes, hormones, and receptors) and additional SOMAmers aimed at pathogen and rodent proteins have been developed with applications in diagnostics and therapeutics[72].

Alternative non-SELEX selection of aptamers has also been described as a process that involves repetitive steps of partitioning through 'non-equilibrium capillary electrophoresis of equilibrium mixtures' (NECEEM) with no amplification between them. The process shows resulting affinities comparable with SELEX and shorter time requirements. Drug candidates from libraries of small molecules, which cannot be PCR-amplified and thus are not approachable by SELEX[73] can also be selected using this procedure.

2.4 Cis activity-based display (CIS display) and covalent antibody display (CAD)

A common limitation with ribosome display and mRNA display technologies is that mRNA is used as the library encoding nucleic acid, which may be prone to rapid degradation. Most of the DNA-based *in vitro* selection systems are dependent on emulsion encapsulation of DNA and are limited to libraries of 10^9 to 10^{10} per ml. An alternate DNA-based approach that does not require any compartmentalization of the library-encoding nucleic acid was developed by Odegrip and colleagues[74] (Fig 7). CIS display, harnesses the ability of a bacterial replication initiator protein, RepA, which exclusively binds back to its encoding DNA (termed cis-activity) of the R1

View Article Online
DOI: 10.1039/C6MB00219F

plasmid. This activity is dependant upon two non-coding regions (CIS and ori) of the repA sequence that are essential for cis-activity.

CIS display was the first published example of recovery of a specific binder from a pool of nonbinding members present at a ratio of 1 in 10^{10}, therefore demonstrating potential for unprecedented library sizes[75]. The technique has been used commercially, for the selection of high affinity peptides and folded protein domains, including antibody fragments (Table 1). In case of peptides, technique has been adapted to select peptides containing L-amino acids which have greater stability by subjecting the libraries to proteolysis and isolating peptides that were resistant to degradation. Interestingly, these peptides had increased half life *in vivo* thus demonstrating a wider spectrum of resistance.

Another display system, covalent antibody display (CAD) (Fig 7), uses a cis-acting DNA binding protein (bacteriophage endonuclease P2A) that covalently links to its own coding strand through the activity of a catalytic tyrosine within its sequence. This system has been used to select antibodies against tetanus toxin from an immune human library with enrichment rates between 14- and 300-fold[76].

2.5 Emulsion compartmentalization and SNAP-tag display

SNAP display is based on the covalent reaction of the DNA repair protein AGT (O(6)-alkylguanine DNA alkyltransferase), the "SNAP-tag", with its substrate benzylguanine (BG)[77] (Fig 8). Linear, BG-labelled template DNA is encapsulated in water-in-oil emulsion droplets with a diameter of a few micrometres (*i.e.* 1 mL of emulsion contains ~10^{10} compartments). Each droplet contains only a single DNA copy, which is transcribed and translated *in vitro*. The expressed AGT fusion proteins attach to their coding DNA via the BG label inside the droplet, which ensures that a

View Article Online
DOI: 10.1039/C6MB00219F

specific genotype-phenotype linkage is established. Subsequently, the emulsion is broken and protein-DNA conjugates, which constitute a DNA-tagged protein library, are selected via affinity panning. This method complements the array of *in vitro* display systems, distinguished by the stability of DNA as the coding nucleic acid and the covalent link between gene and protein.

Conclusions

With display technologies new era of biomolecule discovery has commenced. These technologies permit the *de novo* discovery of highly active, high affinity peptides from libraries, improvement of the performance of natural peptide sequences and modification of binding properties of natural peptide structures. These natural peptide folds can be sampled through semi-rational engineering or such structures can be mined from genomic libraries. Furthermore, *in vitro* display technologies allow the exploration of chemical and structural space which can be expanded through the use of unnatural side chain substitutions to potentially adjust the flexibility in peptide structures [78]. Further optimizations through synthetic chemistry and rational design that may be complementary to such recombinant methods are improving the applicability of display technologies. Moreover, recently, the development of *in vivo* aptamer SELEX and *in vivo* display of phage libraries that can impart delivery to peripheral tissues and circulating immune cells, where they act as ligand mimicks (or can be modified to carry payloads) is opening new and very promising avenues in therapeutics and other applications[67, 79]. Many of these technologies are still in their infancy and further improvements to the methodologies, target presentation and

View Article Online
DOI: 10.1039/C6MB00219F

panning of libraries against cells, tissues and difficult receptor targets, will expand the list of targets that have so far been intractable to therapeutic intervention. The combination of display methodologies with next-generation sequencing platforms and mining of protein–protein interactions in the proteome will also facilitate the understanding of functional proteomics, the validation of therapeutic targets and disease mechanisms.

Table 1

View Article Online
DOI: 10.1039/C6MB00219F

Disease (pathogen) or name	Target / application	Display format	Library type	Reference
PHAGE DISPLAY				
AIDS (HIV)	HIV1 gp41 protein / Therapeutics against HIV1	M13/ pIII Oligopeptide antigen	HIV1 Env-tailored library	[80]
Contagious Bovine Pleuropneumonia, CBP (Mycoplasma mycoides)	Immunogenicity of oligopeptides/ Therapeutics against CBP	M13/ pIII Oliopeptide antigen	Genomic DNA restriction fragments of the MmmSC African field strain B237 in phagemid vector pHORF3	[81]
Pneumonia (Streptococcus pneumoniae)	Immunogenicity of oligopeptides/ Therapeutics against pneumonia	Lambda Oligopeptide antigen	Genomic DNA library of *Streptococcus pneumoniae*	[82]
Cancer	Active site (CA domain) of carbonic anhydrase IX/ Therapeutics against cancer	Phagemid/ M13K07 helper phage/ pIII anti-CAIX Nanobody under research	V_{HH} fragments from immunized camel heavy-chain antibodies	[83]
RIBOSOME DISPLAY				
Cetuximab	Epidermal growth factor receptor (EGFR)/ Treatment of several cancer types	Ribosome display Chimeric (mouse/human) IgG1k monoclonal antibody anti-Epidermal growth factor receptor	Recombinant combinatorial antibody library	12/02/2004 (US) 29/06/2004 (EU)
Natalizumab	α4-integrin/ Treatment of multiple sclerosis and Crohn's disease	Ribosome display Humanized, IgG4k monoclonal antibody	Recombinant combinatorial antibody library	23/11/2004 (US)
Tralokinumab (CAT-354)	IL-13/ Treatment of asthma and other	Ribosome display Phase III for asthsma,	Conventional antibody with fully randomized	[84]

View Article Online
DOI: 10.1039/C6MB00219F

	inflammatory diseases	phase IIb for topic dermatitis	library at 3 regions of 6 amino acids of V_H CDR3	
MP0112/AGN-150998 DARPin (designed ankyrin repeat proteins)	VEGF-A/ Ophtalmic intravitreal therapies	Ribosome display Phase II	Naïve DARPin library	[85]
Anti-VEEV single scFv ToR67-3B4	VEEV(Venezuelan equine encephalitis virus)/ Treatment of VEEV infection	Ribosome display scFv-Fc	Non-human primate (NHP) immunized antibody gene library	[86]
TNF-TeAb	TNFa/ Treatment of inflammatory diseases	Ribosome display Tetramerization domain of human p53 fused to the C-terminus of an anti-TNF-scFv via a linking peptide derived from human serum albumin	Chimera with TNF-scFv gene	[87]
APTAMER DIPLAY				
Macugen : Pegaptanib sodium, only approved aptamer for therapeutics in human	Vascular endothelial growth factor (VEGF)-165/Treatment of age-related macular degeneration	RNA aptamer	Randomized 40-nucleotide library	[88]
RIG-I aptamer	RIG1 (cytosolic receptor for non-self RNA)/ Antiviral activity via modulating IFNα/β production	RNA aptamer with fluorinated nucleotides	Randomized 40-nucleotide library	[89]
MDA aptamer M1	MDA (4,4'-methylenedianiline)/ Blocking of MDA, a carcinogenesis and DNA damaging Agent	RNA aptamer	Randomized 40-nucleotide library	[90]

View Article Online
DOI: 10.1039/C6MB00219F

S66A-C6, S68BC5 **S69A-C15**	S66A-C6 and S68BC5: V3 loop of gp120; S69A-C15: Reverse transcriptase of HIV-1/ Therapy against HIV	RNA aptamer	Randomized 50-nucleotide RNA library and 2'-NH2-pyrimidine RNA library	[91]
SYL3	EpCAM (epithelial cell adhesion molecule)/ Treatment of Breast cancer (MDA-MB-231) and gastric cancer (Kato III)	DNA aptamer	Randomized 40-nucleotide RNA library flanked by primer hybridization sites	[92]
CELL SURFACE DISPLAY				
Engineered *B.subtilis* **Glucose dehydrogenase**	Engineering of Glucose Dehydrogenase / Property engineered: Substrate specificity and stability	Gram positive bacterial surface display of enzyme fused to ice nucleation protein *E.coli*	Random replacing of 4 amino acids from *Bacillus subtilis* by using site-directed mutagenesis	[93]
Engineered lipase B from *C. Antarctica*	Engineering of Lipase B from *C. Antarctica* / Property engineered: Thermostability	Yeast cell surface display of enzyme	Random replacement of 6 aminoacids by multi-site saturation mutagenesis	[94]
Engineered DhbE adenylation domain of NRPS	Engineering of the DhbE adenylation domains of nonribosomal peptide synthetase (NRPS)/ Property engineered:Sustrate specificity	Yeast cell surface display of expressing enzyme fused to the yeast agglutinin protein Aga2p	Random replacement of 4 aminoacids	[95]
Displayed of cytotoxic-associated gene A: CagA	Display of Antigen from H.pilori (CagA) / Therapeutic applications	Spore surface display fused to C-terminal end of CgeA *B.subtilis* spore coat protein.	Use of different vectors for spore surface display	[96]
mRNA DISPLAY				
Serum stable cycGiBP (cyclic peptide binding to	Engineering of cyclic peptide binders to G-alpha i1 protein/	mRNA display	Trillion-member mRNA display cyclic peptide	[97]

View Article Online
DOI: 10.1039/C6MB00219F

	Phage display	Eukaryotic display	Prokaryotic display	Ribosome display	mRNA/ cDNA display	Aptamers	CAD	CIS	IVC*
Host organism	Filamentous phages, M13, T4, T7, lambda,	*S. cerevisiae, P. pastoris*	*E.coli, B. Subtilis, L.bacillus, S. camosus*	*In vitro*	*In vitro*	*In vitro*	*In vitro*	*In vitro*	*In vitro*

Table 2

View Article Online
DOI: 10.1039/C6MB00219F

	phagemid								
Library size	10^{9} *	10^{7}	10^{8-10}	10^{13-14}	10^{13-14}	10^{13-15}	10^{7}	10^{12}	10^{10}/ml
Highest affinity kd (M)*	10^{-11}	10^{-14}	10^{-13}	10^{-12}	10^{-10}	10^{-12}	10^{-9}	10^{-10}	10^{-12}
Typical enrichment factor per round	10^{2-4}	10^{2-3}	10^{2-3}	10^{1-3}	10^{1-3}	10^{1-2}	10^{1-2}	10^{3}	10^{1-3}
Nucleic acid selected	DNA	DNA	DNA	mRNA	mRNA/cDNA	RNA, DNA, ssDNA	DNA	DNA	DNA
Transformation required	Yes	Yes	Yes	No	No	No	No	No	No
Library form	Plasmid	Plasmid	Plasmid	PCR fragment or mRNA	mRNA/cDNA, plasmid	RNA,ssDNA	plasmid	Plasmid	plasmid
Proteins to be displayed	Soluble, nontoxic, compatible with crossing membranes	Soluble and membrane, nontoxic, compatible with crossing membranes	Soluble and membrane, nontoxic, compatible with crossing membranes	Most proteins including cytotoxic, chemically modified and membrane proteins	Soluble, including cytotoxic, chemically modified	Nucleic acids	Developed for antibody display	Soluble, including cytotoxic, chemically modified	Soluble, including cytotoxic, chemicallymodified
Covalent link	No	No	No	No	Yes (synthetic)	No	Yes (Inherent)	Yes	No
Surface anchorage	Capsid proteins	Agglutination proteins, flocculation proteins	Lpp-OmpA, autotransporter proteins, ice nucleation proteins	Ribosome	In vitro	In vitro	In vitro	In vitro	In vitro
Post translational	Simple	Sophisticated	Moderate	Moderate	Simple	No protein	Simple	Simple	Simple

View Article Online
DOI: 10.1039/C6MB00219F

machinery									

*In phage display, initial library diversity can reach 10^{12} but limitations in transformation efficiency and level of expression on phage surface reduce considerably library size

Acknowledgment

Authors acknowledge the European commission for funding the ERA chair team VetMedZg (ERA Chair Iniciative) and HRZZ 4135. LC and MB would like to acknowledge APVV-14-218; VEGA1/0258/15 and VEGA 1/0261/15.

Conflict of interest

The authors declare that they have no conflicts of interest concerning this article.

View Article Online
DOI: 10.1039/C6MB00219F

References

1. M. Li, *Nature biotechnology*, 2000, **18**, 1251-1256.
2. A. R. M. Bradbury, S. Sidhu, S. Duebel and J. McCafferty, *Nature Biotechnology*, 2011, **29**, 245-254.
3. E. P. Hudson, M. Uhlen and J. Rockberg, *Scientific Reports*, 2012, **2**.
4. E. T. Boder and K. D. Wittrup, *Nature Biotechnology*, 1997, **15**, 553-557.
5. J. J. Devlin, L. C. Panganiban and P. E. Devlin, *Science*, 1990, **249**, 404-406.
6. T. T. Takahashi and R. W. Roberts, *Methods in molecular biology (Clifton, N.J.)*, 2009, **535**, 293-314.
7. M. Zaccolo, D. M. Williams, D. M. Brown and E. Gherardi, *Journal of Molecular Biology*, 1996, **255**, 589-603.
8. W. P. C. Stemmer, *Nature*, 1994, **370**, 389-391.
9. M. L. Metzker, *Nature Reviews Genetics*, 2010, **11**, 31-46.
10. H. R. Hoogenboom, A. D. Griffiths, K. S. Johnson, D. J. Chiswell, P. Hudson and G. Winter, *Nucleic acids research*, 1991, **19**, 4133-4137.
11. E. Brunet, C. Chauvin, V. Choumet and J. L. Jestin, *Nucleic Acids Research*, 2002, **30**.
12. B. Kotlan and M. C. Glassy, *Methods Mol Biol*, 2009, **562**, 1-15.
13. H. R. Hoogenboom, A. P. de Bruine, S. E. Hufton, R. M. Hoet, J. W. Arends and R. C. Roovers, *Immunotechnology : an international journal of immunological engineering*, 1998, **4**, 1-20.
14. H. Qi, H. Lu, H. J. Qiu, V. Petrenko and A. Liu, *Journal of molecular biology*, 2012, **417**, 129-143.
15. L. Chasteen, J. Ayriss, P. Pavlik and A. R. M. Bradbury, *Nucleic Acids Research*, 2006, **34**.
16. L. R. H. Krumpe, A. J. Atkinson, G. W. Smythers, A. Kandel, K. M. Schumacher, J. B. McMahon, L. Makowski and T. Mori, *Proteomics*, 2006, **6**, 4210-4222.
17. R. J. Hosse, A. Rothe and B. E. Power, *Protein Science*, 2006, **15**, 14-27.
18. C. Hamerscasterman, T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman and R. Hamers, *Nature*, 1993, **363**, 446-448.
19. J. P. Rast, C. T. Amemiya, R. T. Litman, S. J. Strong and G. W. Litman, *Immunogenetics*, 1998, **47**, 415-415.
20. S. Muyldermans, *Annu Rev Biochem*, 2013, **82**, 775-797.
21. C. V. Lee, W. C. Liang, M. S. Dennis, C. Eigenbrot, S. S. Sidhu and G. Fuh, *Journal of molecular biology*, 2004, **340**, 1073-1093.
22. S. Y. Lee, J. H. Choi and Z. H. Xu, *Trends in Biotechnology*, 2003, **21**, 45-52.
23. C. Garcia-Rodriguez, R. Levy, J. W. Arndt, C. M. Forsyth, A. Razai, J. Lou, I. Geren, R. C. Stevens and J. D. Marks, *Nature Biotechnology*, 2007, **25**, 107-116.
24. O. Schneewind and D. M. Missiakas, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 2012, **367**, 1123-1139.
25. N. Rutherford and M. Mourez, *Microbial Cell Factories*, 2006, **5**.
26. H. Matsuoka, K. Hashimoto, A. Saijo, Y. Takada, A. Kondo, M. Ueda, H. Ooshima, T. Tachibana and M. Azuma, *Yeast*, 2014, **31**, 67-76.
27. A. Gruetzkau and A. Radbruch, *Cytometry Part A*, 2010, **77A**, 643-647.
28. J. G. Pan, E. J. Kim and C. H. Yun, *Trends in Biotechnology*, 2012, **30**, 610-612.
29. S. DasSarma and P. DasSarma, *Vaccines*, 2015, **3**, 686-702.
30. E. T. Boder, K. S. Midelfort and K. D. Wittrup, *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 10701-10705.
31. A. T. V. Nguyen, C. K. Pham, H. T. T. Pham, H. L. Pham, A. H. Nguyen, L. T. Dang, H. A. Huynh, S. M. Cutting and P. Tuan-Nghia, *Fems Microbiology Letters*, 2014, **358**, 202-208.

View Article Online
DOI: 10.1039/C6MB00219F

32. F.-Y. Huang, C.-C. Wang, Y.-H. Huang, H.-G. Zhao, J.-L. Guo, S.-L. Zhou, H. Wang, Y.-Y. Lin and G.-H. Tan, *Immunology*, 2014, **143**, 230-240.
33. T. Van Blarcom, A. Rossi, D. Foletti, P. Sundar, S. Pitts, C. Bee, J. M. Witt, Z. Melton, A. Hasa-Moreno, L. Shaughnessy, D. Telman, L. Zhao, W. L. Cheung, J. Berka, W. Zhai, P. Strop, J. Chaparro-Riggers, D. L. Shelton, J. Pons and A. Rajpal, *Journal of Molecular Biology*, 2015, **427**, 1513-1534.
34. M.-S. Ju and S. T. Jung, *Current Opinion in Biotechnology*, 2014, **30**, 128-139.
35. S. T. Henriques, L. Thorstholm, Y.-H. Huang, J. A. Getz, P. S. Daugherty and D. J. Craik, *Plos One*, 2013, **8**.
36. M. K. Maruthamuthu, S. P. Nadarajan, I. Ganesh, S. Ravikumar, H. Yun, I.-k. Yoo and S. H. Hong, *Bioprocess and Biosystems Engineering*, 2015, **38**, 2077-2084.
37. J. Hanes and A. Pluckthun, *Proceedings of the National Academy of Sciences of the United States of America*, 1997, **94**, 4937-4942.
38. E. Bencurova, L. Pulzova, Z. Flachbartova and M. Bhide, *Molecular Biosystems*, 2015, **11**, 1515-1524.
39. E. V. Makeyev, V. A. Kolb and A. S. Spirin, *Febs Letters*, 1996, **378**, 166-170.
40. M. J. May, M. R. Hartley, L. M. Roberts, P. A. Krieg, R. W. Osborn and J. M. Lord, *Embo Journal*, 1989, **8**, 301-308.
41. H. R. Hoogenboom, *Nature Biotechnology*, 2005, **23**, 1105-1116.
42. M. He and M. J. Taussig, *Nature Methods*, 2007, **4**, 281-288.
43. C. Schaffitzel, J. Hanes, L. Jermutus and A. Pluckthun, *Journal of Immunological Methods*, 1999, **231**, 119-135.
44. T. Kawasaki, M. D. Gouda, T. Sawasaki, K. Takai and Y. Endo, *European Journal of Biochemistry*, 2003, **270**, 4780-4786.
45. M. G. Casteleijn, A. Urtti and S. Sarkhel, *International Journal of Pharmaceutics*, 2013, **440**, 39-47.
46. D. Villemagne, R. Jackson and J. A. Douthwaite, *Journal of Immunological Methods*, 2006, **313**, 140-148.
47. J. R. Hillebrecht and S. Chong, *Bmc Biotechnology*, 2008, **8**.
48. R. W. Roberts and J. W. Szostak, *Proceedings of the National Academy of Sciences of the United States of America*, 1997, **94**, 12297-12302.
49. D. Ponsel, J. Neugebauer, K. Ladetzki-Baehs and K. Tissot, *Molecules*, 2011, **16**, 3675-3700.
50. T. T. Takahashi, R. J. Austin and R. W. Roberts, *Trends in Biochemical Sciences*, 2003, **28**, 159-165.
51. S. S. Sidhu, H. B. Lowman, B. C. Cunningham and J. A. Wells, *Applications of Chimeric Genes and Hybrid Proteins, Pt C*, 2000, **328**, 333-363.
52. K. Josephson, M. C. T. Hartman and J. W. Szostak, *Biopolymers*, 2005, **80**, 508-509.
53. Y. Mochizuki, S. Kumachi, K. Nishigaki and N. Nemoto, *Biological Procedures Online*, 2013, **15**.
54. G. T. Rozenblum, V. G. Lopez, A. D. Vitullo and M. Radrizzani, *Expert opinion on drug discovery*, 2016, **11**, 127-135.
55. L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas and J. J. Toole, *Nature*, 1992, **355**, 564-566.
56. M. Saberian-Borujeni, M. Johari-Ahar, H. Hamzeiy, J. Barar and Y. Omidi, *BioImpacts : BI*, 2014, **4**, 205-215.
57. S. M. Nimjee, C. P. Rusconi and B. A. Sullenger, in *Annual Review of Medicine*, 2005, vol. 56, pp. 555-+.
58. A. D. Keefe, S. Pai and A. Ellington, *Nature Reviews Drug Discovery*, 2010, **9**, 537-550.
59. Y.-H. Lao, K. K. L. Phua and K. W. Leong, *Acs Nano*, 2015, **9**, 2235-2254.

View Article Online
DOI: 10.1039/C6MB00219F

60. D. Xiang, S. Shigdar, G. Qiao, T. Wang, A. Z. Kouzani, S.-F. Zhou, L. Kong, Y. Li, C. Pu and W. Duan, *Theranostics*, 2015, **5**, 23-42.
61. B. E. Eaton, L. Gold, B. J. Hicke, N. Janjic, F. M. Jucker, D. P. Sebesta, T. M. Tarasow, M. C. Willis and D. A. Zichi, *Bioorganic & Medicinal Chemistry*, 1997, **5**, 1087-1096.
62. J. G. Bruno, *Pharmaceuticals (Basel, Switzerland)*, 2013, **6**, 340-357.
63. H. Sun, X. Zhu, P. Y. Lu, R. R. Rosato, W. Tan and Y. Zu, *Molecular Therapy-Nucleic Acids*, 2014, **3**.
64. M. Rajendran and A. D. Ellington, *Nucleic Acids Research*, 2003, **31**, 5700-5713.
65. H. Lin, W. Zhang, S. Jia, Z. Guan, C. J. Yang and Z. Zhu, *Biomicrofluidics*, 2014, **8**.
66. D. A. Daniels, H. Chen, B. J. Hicke, K. M. Swiderek and L. Gold, *Proceedings of the National Academy of Sciences of the United States of America*, 2003, **100**, 15416-15421.
67. C. Cheng, Y. H. Chen, K. A. Lennox, M. A. Behlke and B. L. Davidson, *Molecular Therapy-Nucleic Acids*, 2013, **2**.
68. T. Wang, J. A. Hoy, M. H. Lamm and M. Nilsen-Hamilton, *Journal of the American Chemical Society*, 2009, **131**, 14747-14755.
69. K.-M. Song, E. Jeong, W. Jeon, H. Jo and C. Ban, *Biosensors & Bioelectronics*, 2012, **33**, 113-119.
70. K. Szeto, S. J. Reinholt, F. M. Duarte, J. M. Pagano, A. Ozer, L. Yao, J. T. Lis and H. G. Craighead, *Analytical and Bioanalytical Chemistry*, 2014, **406**, 2727-2732.
71. M. Cho, Y. Xiao, J. Nie, R. Stewart, A. T. Csordas, S. S. Oh, J. A. Thomson and H. T. Soh, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 15373-15378.
72. J. C. Rohloff, A. D. Gelinas, T. C. Jarvis, U. A. Ochsner, D. J. Schneider, L. Gold and N. Janjic, *Molecular Therapy-Nucleic Acids*, 2014, **3**.
73. M. V. Berezovski, M. U. Musheev, A. P. Drabovich, J. V. Jitkova and S. N. Krylov, *Nature Protocols*, 2006, **1**, 1359-1369.
74. R. Odegrip, D. Coomber, B. Eldridge, R. Hederer, P. A. Kuhlman, C. Ullman, K. FitzGerald and D. McGregor, *Proceedings of the National Academy of Sciences of the United States of America*, 2004, **101**, 2806-2810.
75. B. Eldridge, R. N. Cooley, R. Odegrip, D. P. McGregor, K. J. FitzGerald and C. G. Ullman, *Protein Engineering Design & Selection*, 2009, **22**, 691-698.
76. H. Reiersen, I. Lobersli, G. A. Loset, E. Hvattum, B. Simonsen, J. E. Stacy, D. McGregor, K. FitzGerald, M. Welschof, O. H. Brekke and O. J. Marvik, *Nucleic Acids Research*, 2005, **33**.
77. M. Kaltenbach and F. Hollfelder, *Methods in molecular biology (Clifton, N.J.)*, 2012, **805**, 101-111.
78. Y. Sako, Y. Goto, H. Murakami and H. Suga, *Acs Chemical Biology*, 2008, **3**, 241-249.
79. B. Bakhshinejad, M. Karimi and M. Khalaj-Kondori, *Neural Regeneration Research*, 2015, **10**, 862-865.
80. M. Zhou, T. Meyer, S. Koch, J. Koch, H. von Briesen, J. M. Benito, V. Soriano, A. Haberl, M. Bickel, S. Duebel, M. Hust and U. Dietrich, *European Journal of Immunology*, 2013, **43**, 499-509.
81. S. Naseem, J. Meens, J. Jores, M. Heller, S. Duebel, M. Hust and G.-F. Gerlach, *Veterinary Microbiology*, 2010, **142**, 285-292.
82. E. Beghetto and N. Gargano, *Methods in molecular biology (Clifton, N.J.)*, 2013, **1061**, 79-95.
83. F. Araste, W. Ebrahimizadeh, I. Rasooli, M. Rajabibazl and S. L. M. Gargari, *Biotechnology Letters*, 2014, **36**, 21-28.

View Article Online
DOI: 10.1039/C6MB00219F

84. G. Thom, A. C. Cockroft, A. G. Buchanan, C. J. Canclotti, E. S. Cohen, D. Lowne, P. Monk, C. P. Shorrock-Hart, L. Jermutus and R. R. Minter, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 7619-7624.
85. A. Stahl, M. T. Stumpp, A. Schlegel, S. Ekawardhani, C. Lehrling, G. Martin, M. Gulotti-Georgieva, D. Villemagne, P. Forrer, H. T. Agostini and H. K. Binz, *Angiogenesis*, 2013, **16**, 101-111.
86. T. Ruelker, L. Voss, P. Thullier, L. M. O' Brien, T. Pelat, S. D. Perkins, C. Langermann, T. Schirrmann, S. Duebel, H.-J. Marschall, M. Hust and B. Huelseweh, *Plos One*, 2012, **7**.
87. M. Liu, X. Wang, C. Yin, Z. Zhang, Q. Lin, Y. Zhen and H. Huang, *Biochemical Journal*, 2007, **406**, 237-246.
88. E. W. M. Ng and A. P. Adamis, in *Oligonucleotide Therapeutics*, eds. T. Tuschl and J. Rossi, 2006, vol. 1082, pp. 151-171.
89. S.-Y. Hwang, H.-Y. Sun, K.-H. Lee, B.-H. Oh, Y. J. Cha, B. H. Kim and J.-Y. Yoo, *Nucleic Acids Research*, 2012, **40**, 2724-2733.
90. U. Brockstedt, A. Uzarowska, A. Montpetit, W. Pfau and D. Labuda, *Biochemical and Biophysical Research Communications*, 2004, **313**, 1004-1008.
91. T. M. A. Gronewold, A. Baumgartner, J. Hierer, S. Sierra, M. Blind, F. Schaefer, J. Bluemer, T. Tillmann, A. Kiwitz, R. Kaiser, M. Zabe-Kuehn, E. Quandt and M. Famulok, *Journal of Proteome Research*, 2009, **8**, 3568-3577.
92. Y. Song, Z. Zhu, Y. An, W. Zhang, H. Zhang, D. Liu, C. Yu, W. Duan and C. J. Yang, *Analytical Chemistry*, 2013, **85**, 4141-4149.
93. B. Liang, Q. Lang, X. Tang and A. Liu, *Bioresource Technology*, 2013, **147**, 492-498.
94. X.-Q. Peng, *Applied Biochemistry and Biotechnology*, 2013, **169**, 351-358.
95. K. Zhang, K. M. Nelson, K. Bhuripanyo, K. D. Grimes, B. Zhao, C. C. Aldrich and J. Yin, *Chemistry & Biology*, 2013, **20**, 92-101.
96. A. Iwanicki, I. Piatek, M. Stasilojc, A. Grela, T. Lega, M. Obuchowski and K. Hinc, *Microbial Cell Factories*, 2014, **13**.
97. S. M. Howell, S. V. Fiacco, T. T. Takahashi, F. Jalali-Yazdi, S. W. Millward, B. Hu, P. Wang and R. W. Roberts, *Scientific Reports*, 2014, **4**.

View Article Online
DOI: 10.1039/C6MB00219F

Fig legends

Fig. 1 Different procedures to generate DNA libraries for display technologies and their directed evolution. Left, **A**, Error prone PCR introduces point mutations during amplification. **B**, DNA shuffling combines random fragments of different DNA sequences and **C**, Digestion and ligation adds random sequences coding for small peptides to the ORFs to be displayed. **Right**, schematic diagram representing concept of directed molecular evolution behind all display technologies: Conformationally and chemically different elements (libraries with genotype-phenotype linkage) are allowed to interact with a specific target and non-interacting elements are discarded. Interacting elements are replicated and submitted to next round of selection (biopanning). After several rounds of biopanning or during biopanning the selected elements with highest affinities are often modified to improve specificity, biocompatibility, stability or any other desired feature.

Fig 2. Phage display biopanning. At first, phage library is screened with targeted antigen. Then, unbound phages are washed away, bound phages are infected into bacteria and are later amplified. This process is repeated 5 times on average. **Top right,** schematic example of phagemid or phasmid library (pHORF3). Abbreviations: M13 ori: origin of M13 phage, bla: ampicillin resistance, colE1: colE1 plasmid origin, lacZ promoter: promoter of the bacterial lac operon; RBS: ribosome binding site; pelB: signal peptide sequence of bacterial pectate lyase Erwinia caratovora, mediating secretion into the periplasmic space; gIII: gene coding for the phage protein III; amber: amber stop codon; his: six histidine residues; ochre: ochre stop codon. The elements of the inserts are not drawn to scale.

View Article Online
DOI: 10.1039/C6MB00219F

Fig 3. Examples of cell surface display and associated vectors harboring libraries. **A** Schematic representation of the different types of surface proteins found **in Gram-positive bacteria.** Membrane-associated proteins can either (i) possess transmembrane domain(s) or (ii) be lipoproteins, and thus be covalently attached to long chain fatty acids of the cytoplasmic membrane. Cell wall-associated proteins can either (i) possess C-terminal LPXTG motif (or related motifs) and then be covalently anchored to peptidoglycan via sortase activity, or (ii) possess cell wall binding domain(s) (CWBDs) and thus remain attached to the cell wall; four CWBDs are currently characterized, i.e. CWBD1, CWBD2, LysM, GWmodules and SLHD. Importantly, since cell wall of Gram-positive bacteria is permeable, proteins can interact with their environment without ever having a domain protruding out of the cell wall. **B,** Scheme of the autotransporter system used for display on the surface of **Gram-negative bacteria.** Following translation, the AT polypeptide is transported across the inner membrane by the Sec machinery. Hereto, most ATs have a characteristic signal peptide (dark green) although some of them possess an extended signal peptide which possibly extends the attachment to the Sec translocon to avoid periplasmic misfolding Upon entry in the periplasm, chaperone proteins interact with the AT polypeptide. These chaperones bind the passenger domain (red), the TU (yellow, a-helix; orange, b-barrel) or both domains. After signal peptide cleavage the AT is released in the periplasm. The chaperone-AT complex is targeted to the b-barrel assembly machinery (Bam). The Bam complex is involved in the binding, folding and insertion of the b-barrel into the outer membrane (OM). Prior to the OM insertion, the a-helix is already incorporated into the b-barrel. After translocation, the passenger domain can be processed and released into the extracellular environment or remain attached to the cell surface. **C,** Scheme of Top 7 scaffold displayed on the surface of

View Article Online
DOI: 10.1039/C6MB00219F

S. cerevisiae. The Top7 scaffold is expressed as a protein fusion to the Aga2p mating agglutinin protein. The fusion protein is tethered to the yeast cell wall via disulfide bridges between the Aga2p protein and the Aga1p protein (which is covalently attached to the cell wall). The fusion protein also contains N-terminal HA and C-terminal c-myc epitope tags for detection of the displayed fusion protein by means of fluorescent antibodies for fluorescent activated cell sorting (FACS).

Fig 4. Ribosome-display selection cycle. The DNA of the library of interest, fused in frame to a spacer carrying no stop codon, is transcribed *in vitro*. After transcription, the resulting mRNA is used for *in vitro* translation. After a short time of translation (a few minutes) the ribosomes have probably run to the end of the mRNA and synthesized the encoded protein, but because of the absence of the stop codon, the protein remains connected to the tRNA. The mRNA-ribosome-protein ternary complexes are used for affinity selection on an immobilized target. After washing, the mRNA of the selected complexes is released by complex dissociation. The genetic information of binders is rescued by RT-PCR yielding a PCR product ready to go for the next cycle. Several of these cycles generate highly specific and pure binders which can be used to detect the protein of interest.

Fig 5. Overview of mRNA-display *in vitro* selection process. Schematic view of the structure of the library and the selection process. The DNA library includes, from 5‘ to 3 ‘, a T7RNA polymerase promoter (T7), a tobacco mosaic virus translation enhancer (TMV), a start codon (ATG), degenerate sequence coding 88 random amino

View Article Online
DOI: 10.1039/C6MB00219F

acids, a hexahistidine tag (H6), and a 3' constant region (spacer). This library is transcribed by using T7 RNA polymerase, after which the puromycin-containing linker is ligated onto the 3' end of the mRNA. When this template is translated *in vitro*, the nascent peptide forms a covalent bond to the puromycin moiety. The resulting mRNA- peptide covalent fusion molecules are then purified on oligo-dT-cellulose (which anneals to the oligo-dA sequence in the puromycin-containing linker) and Ni-NTA agarose. The mRNA portion of this display construct is then reverse transcribed. The cDNA -RNA-peptide species (with streptavidin in RNA) are then incubated with the immobilized target protein and unbound library members are washed off. Target-bound peptides are then displaced with biotin. The eluted molecules are then amplified by PCR, thus completing the first round of selection and amplification. (Top right) Chemical structure of the 30 end of Tyr-tRNATyr (left) is compared with puromycin (right).

Fig 6. Schematic representation of SELEX procedure to produce DNA or RNA aptamers (peptide aptamers are not shown) for specific target molecules (proteins, peptides, small organic molecules, cells, tissues, whole organisms...) immobilized on a surface, magnetic beads, or in the case of cell SELEX the surface of a cell (center, top). The initial DNA or RNA (as shown, T7 RNA polymerase and reverse transcriptase, RT, ar used to produce either oligonucleotide before or after SELEX) combinatorial library consisting of oligonucleotides with a centrally randomized or degenerate region of n nucleotides (usually $N_{40\text{-}50}$) flanked by two primer hybridization regions is incubated with the immobilized target for binding. Unbound oligonucleotides are washed away and the bound oligonucleotides are eluted from the target by several treatments. The selected DNAs are amplified by PCR using specially modified primers and RNAs are reverse transcribed, amplified by PCR, and

View Article Online
DOI: 10.1039/C6MB00219F

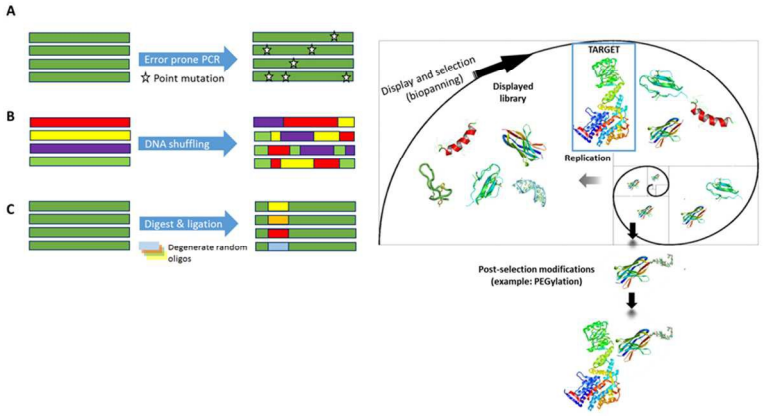
transcribed for the next round of selection. The fluorescein (F) labelling of the selected oligonucleotides after the first SELEX round enables their quantification in further rounds and thus monitoring of the enrichment of target-specific aptamers. In the last round the selected aptamers are cloned and several monoclonal aptamers were characterized.

Fig. 7 Principle of the CIS and CAD display technology. In CIS display libraries template DNA encoding an N-terminal library peptide is ligated to the RepA gene. *In vitro* transcription is initiated at the promoter and pauses when the RNA polymerase reaches the CIS element. Concurrent translation produces the RepA protein, which transiently interacts with the CIS element, thereby forcing its subsequent binding to the adjacent *ori* sequence. This process establishes a stable linkage between a template DNA and the expressed polypeptide that it encodes. CAD (Covalent Antibody Display), mainly developed for the display of antobodies, is based on the ability of endonuclease P2A to become covalently attached (via Y454) to the 5′ phosphate of its own DNA at Ori of replication (CCT CGG, *****) at position 1860. CIS display and CAD selections begin with the construction of a peptide-encoding DNA library followed by *in vitro* transcription/translation to form a pool of protein–DNA complexes (only shown for CIS display). The library pool is incubated with an immobilized target, nonbinding peptides are washed away, and the retained DNA that encodes the target-binding peptides is eluted and amplified by PCR, to form a DNA library ready for the next round of selection. After three to five rounds of selection, recovered DNA is cloned into an appropriate expression vector for the identification of individual target-binding peptide sequences.

View Article Online
DOI: 10.1039/C6MB00219F

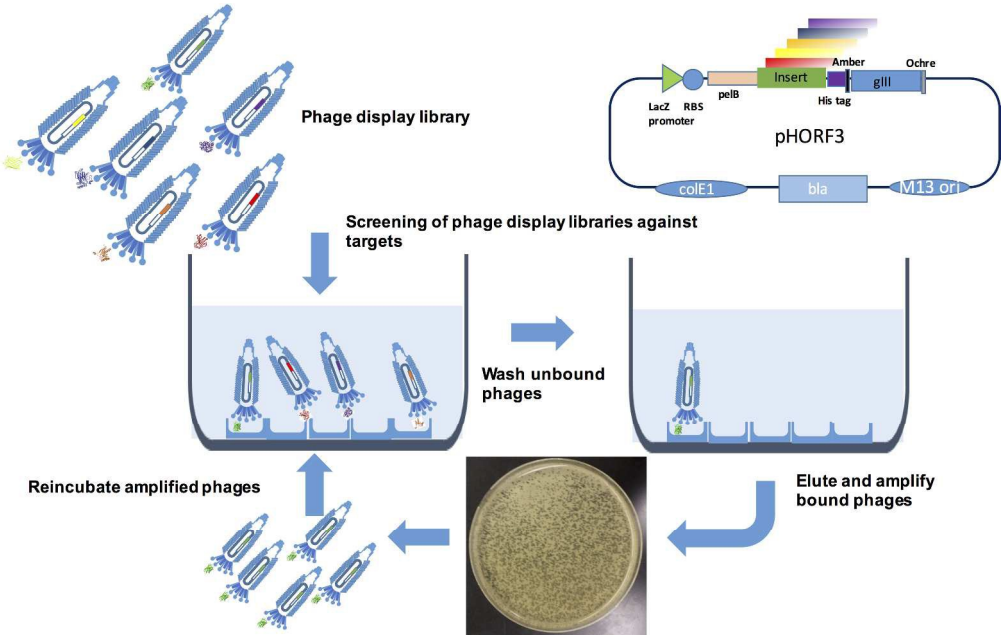
Fig 8. Formats for artificial covalent genotype–phenotype linkages based on droplet compartmentalization. The key initial step of both display methods is that a DNA library (coding for SNAP-tag-fused variants of the protein of interest (POI) is compartmentalized in water-in-oil emulsion droplets, so that each compartment contains no more than one DNA template (Poisson distribution). (*a*) In SNAP display, the POI is *in vitro* expressed from a single gene in fusion to the SNAP-tag (1). The SNAP-tag of the expressed fusion protein then reacts with its substrate (BG) that has been covalently linked to the DNA template. As a result, the SNAP-tag connects genotype and the displayed protein (responsible for the phenotype). (2) SNAP-tagged proteins are de-emulsified and challenged for binding against an antigen by affinity panning. (3) After non-binders are washed away, binders are eluted together with their encoding genes that can feed the next round of selection.

View Article Online
DOI: 10.1039/C6MB00219F



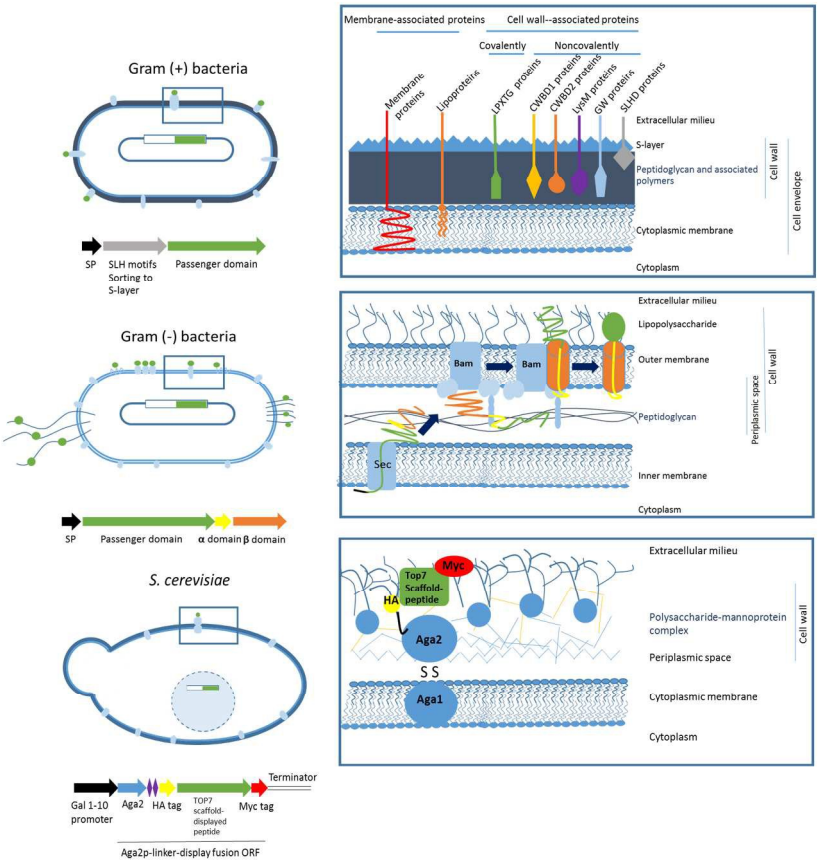
324x142mm (96 x 96 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



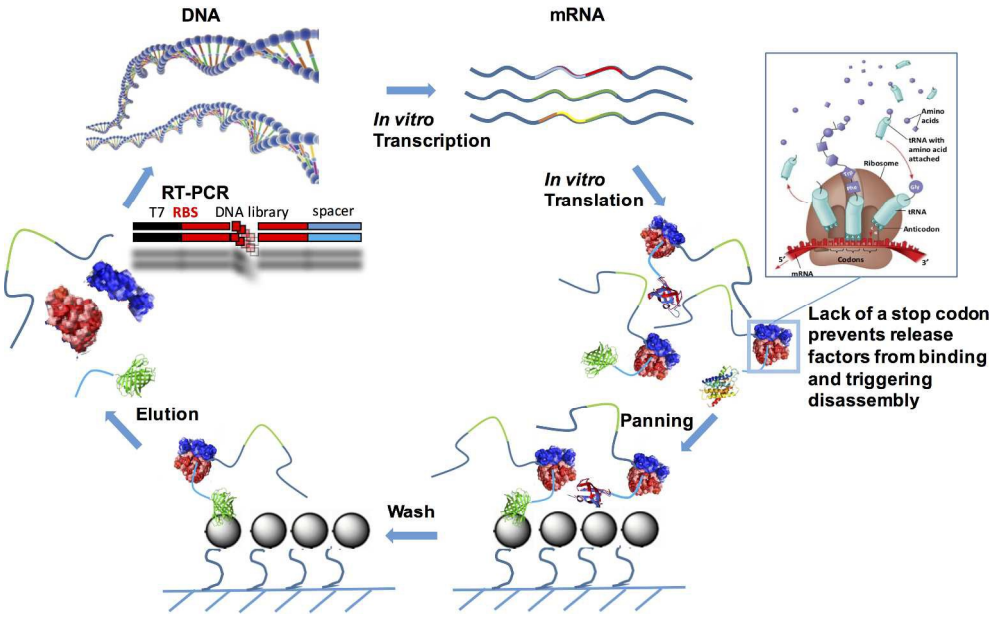
297x187mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



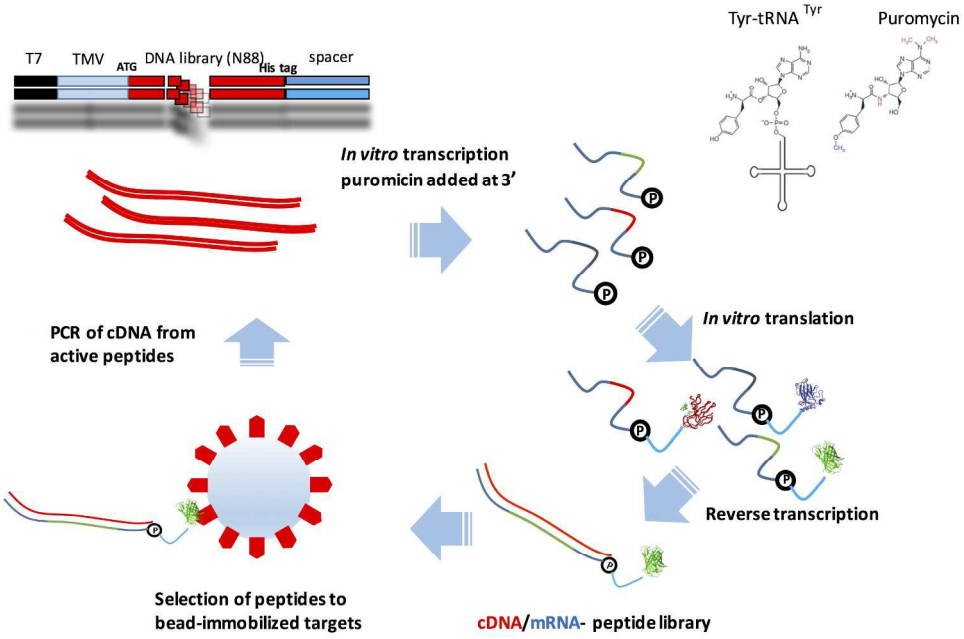
776x881mm (72 x 72 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



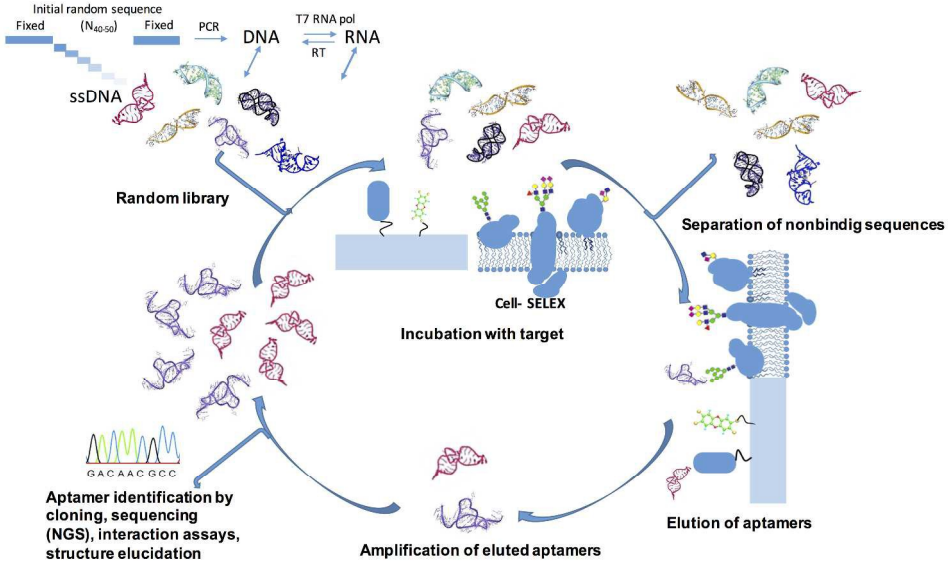
271x167mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



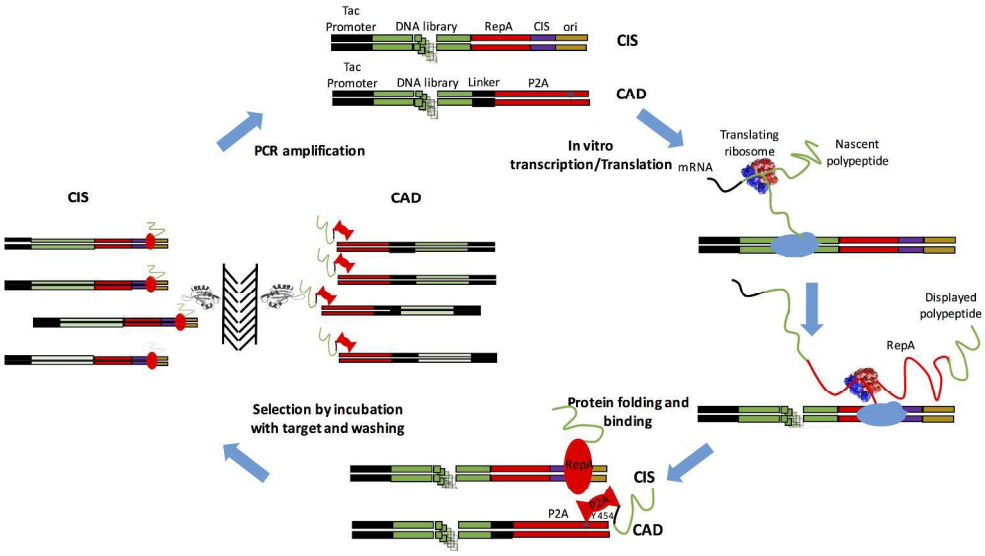
255x166mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



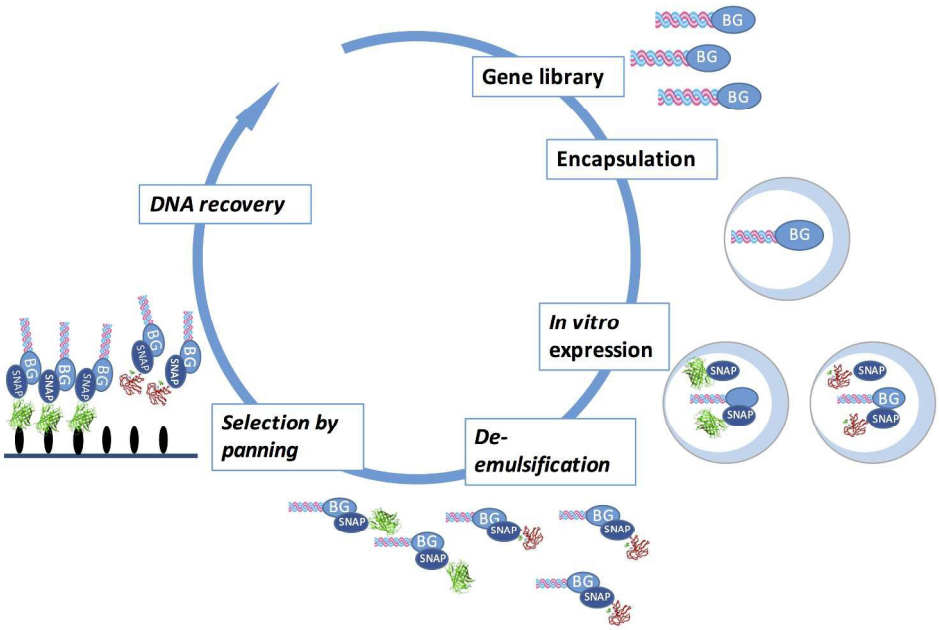
320x183mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



323x190mm (300 x 300 DPI)

View Article Online
DOI: 10.1039/C6MB00219F



242x153mm (300 x 300 DPI)