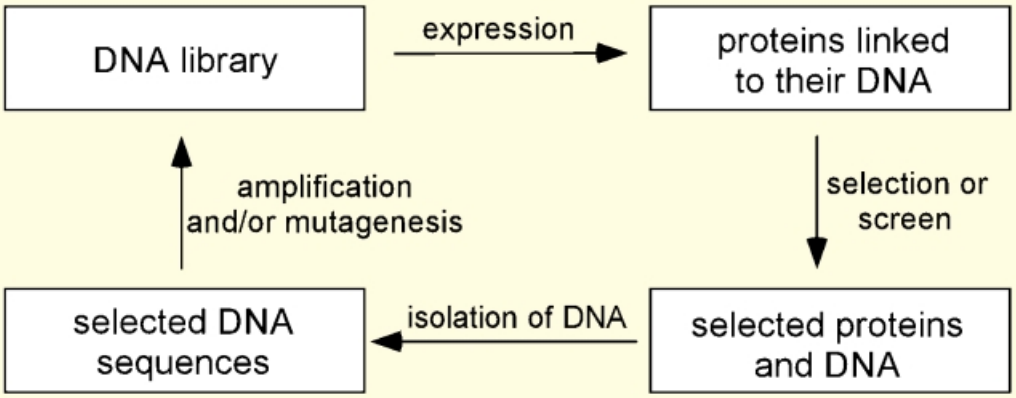
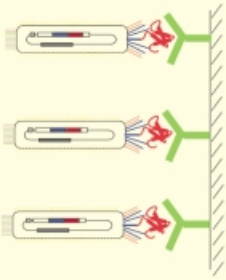
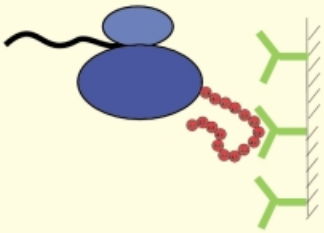
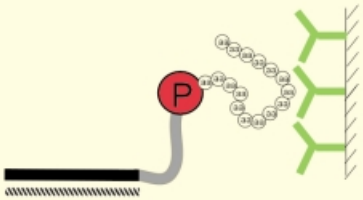


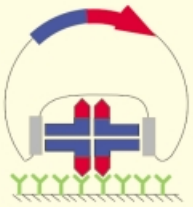
phage display



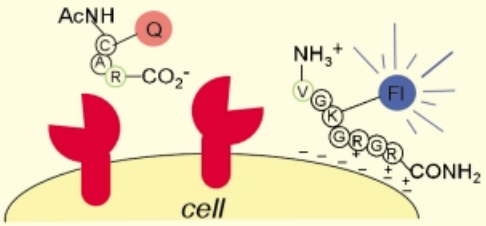
ribosome display



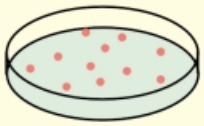

mRNA–peptide fusion



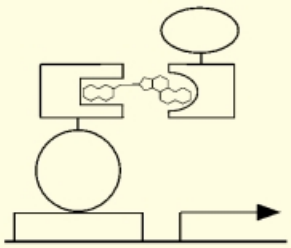
plasmid display



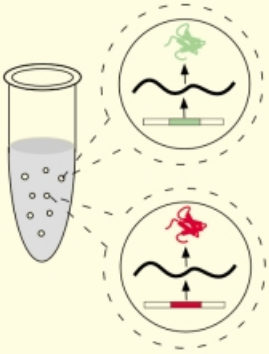

cell-surface display



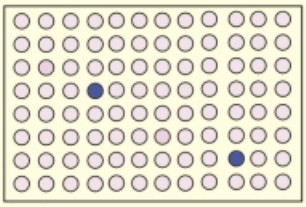
genetics



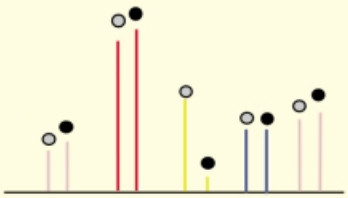
*n*-hybrid systems



in vitro compartmentalization



spatial address



mass spectrometry

ANGEWANDTE CHEMIE

# Screening and Selection Methods for Large-Scale Analysis of Protein Function

**Hening Lin and Virginia W. Cornish***

High-throughput assays hold tremendous promise for protein engineering and proteomics. With powerful assays it should be possible to evolve, for example, a stereoselective esterase for the chemical synthesis or a site-specific endonuclease for biomedical research. Entire cDNA libraries, which encode all of the proteins expressed in a given organism or cell line, should simply be passed through a battery of biochemical assays to determine the function of each individual protein. Herein we look at the types of assays that have been developed and how close we are to our goals of engineering proteins with new activities as well as rapidly assigning function to the thousands of proteins that make up each genome.

**Keywords:** directed evolution · high-throughput screening · protein engineering · proteomics · selection

## 1. Introduction

High-throughput assays are beginning to show promise for the evolution of proteins with new properties. Because of the complexity of protein function, it has proven difficult to engineer proteins with new specificities, let alone new functions, simply by introducing directed mutations in the active site of the proteins. Thus, attention has turned to generating millions of mutants simultaneously and then testing these mutants for the desired activity. Advances in molecular biology have made it fairly straightforward to generate the protein variants at the DNA level. Because the manipulations are at the DNA level, the same methods can be used to generate the proteins, no matter what the protein fold or desired function is. This is in contrast to small molecule libraries in which new chemistry has to be developed for each new scaffold. It has proven more difficult, however, to design high-throughput assays that allow millions of proteins to be tested at once. Thus, in this Review we have chosen to focus on advances made during the last two decades in high-throughput assays for protein function.

In addition to applications in protein engineering, high-throughput assays for protein function have long been used for enzymology and drug discovery. More recently, high-throughput assays have proved critical to proteomics. The ability to generate and compare the activity of millions of protein variants not only allows us to engineer proteins with new properties, but also provides a powerful new tool for understanding protein structure and function. The pharmaceutical industry, of course, has relied on high-throughput assays for decades to identify small molecules that inhibit protein function. More recently, with the completion of several genome sequencing projects, we are left with the challenge of figuring out what these proteins do. The genome sequencing project is of limited use if we still have to determine protein function one at a time. Whereas some classes of proteins can be identified based on sequence homology, high-throughput assays still are needed to deconvolute the substrate specificities of these proteins. High-throughput assays are essential for determining the activity of the large percentage of all proteins whose functions are not known.

To develop high-throughput assays for protein function, there are two challenges. First, because DNA is much easier to amplify and sequence than proteins, methods are needed for tagging each protein with its unique DNA sequence (Figure 1). Just like a tag is used to encode each member of a small

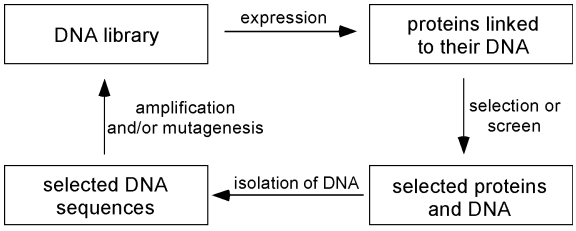



Figure 1. General strategy for large-scale analysis of protein function.

[*] Prof. Dr. V. W. Cornish, H. Lin
Department of Chemistry, Columbia University
New York, NY 10027 (USA)
Fax: (+1) 212-932-1289
E-mail: vc114@columbia.edu

  1433-7851/02/4123-4403 $ 20.00+.50/0 

Table 1. The link between genotype and phenotype in different assay methods

| Entry | Method | Genotype–phenotype link | Assay (selection or screening) method |
|---|---|---|---|
| 1 | Phage display | Phage particle | Binding to affinity matrix |
| 2 | Ribosome display | Ribosome complex | Binding to affinity matrix |
| 3 | mRNA peptide fusion | Peptide–mRNA fusion | Binding to affinity matrix |
| 4 | Peptide on plasmid | Peptide–plasmid complex | Binding to affinity matrix |
| 5 | Cell surface display | Cell | Fluorescence-activated cell sorting (FACS) |
| 6 | In vitro compartmentalization | Water-in-oil droplet | Polymerase chain reaction (PCR) |
| 7 | Genetic assay | Cell | Complementation, colormetric assay |
| 8 | Microtiter plates and protein chips | Spatial address | Radiometric, UV/Vis absorption, or fluorescence assay |

molecule library, the DNA will serve as a tag for the protein sequence. DNA is an ideal tag for protein libraries because DNA naturally serves as the code for protein synthesis in vivo. Second, a high-throughput assay is needed that is compatible with this connection between the protein and the DNA (Figure 1).

Most of the advances in high-throughput assays in the last two decades have been in the first area—designing methods for tagging each protein with its unique DNA sequence. Several refined formats are now well developed. In approaches such as phage display, it is literally a physical linkage between the protein and the DNA. For cell-based assays the linkage is cellular compartmentalization because the protein is only made if a piece of DNA that encodes the protein is present in the cell. For microtiter plate assays, the protein and DNA are given a common "spatial address". Put more simply, the protein in column F, row 11 is encoded by the DNA on a second plate in column F, row 11. Although several methods for tagging protein libraries with their DNA sequence are now available, there may be advantages to completely new approaches. For example, it is interesting to think about the possibilities with methods such as mass spectrometry that might eliminate the need for a DNA tag.

A method to tag a protein with its DNA sequence, however, is only the first step. The actual assays still have to be developed (Table 1). Throughout this paper, "selection" is used to refer to methods in which only proteins with the desired function are carried through, "screening" refers to methods in which all the proteins must be examined, and "assay" is a general term that encompasses both selections and screens. Several methods, such as ribosome display, lend themselves naturally to binding assays. The protein–DNA complex is passed over an affinity matrix, and proteins that bind to the matrix are then eluted with a concentrated salt

V. W. Cornish

H. Lin

*Virginia W. Cornish was born in Louisville, KY in 1969. She completed her undergraduate studies at Columbia University. As an undergraduate, Virginia carried out research in Professor Ronald Breslow's laboratory aimed at identifying the receptor for a new class of anticancer compounds being developed in the laboratory. She graduated from Columbia summa cum laude with a B.A. in Biochemistry in 1991. She then moved west to do research with Professor Peter Schultz in the Chemistry Department at the University of California, Berkeley. In Professor Schultz's laboratory she helped develop a new method for incorporating synthetic amino acids into proteins by using the protein biosynthetic machinery. At Berkeley she was a National Science Foundation Predoctoral Fellow and a recipient of fellowships from the American Chemical Society Division of Organic Chemistry and the Howard Hughes Medical Institute. In 1996, she became an NSF Postdoctoral Fellow in the Biology Department at M.I.T. under the guidance of Professor Bob Sauer. At M.I.T. she initiated an independent project which is the basis of her research at Columbia. Virginia joined the Columbia faculty in 1999. Her group uses a combination of chemistry and molecular genetics to develop a cell-based assay for screening large collections of molecules simultaneously based on function. She is the recipient of a Beckman Young Investigator Award, a Burroughs-Wellcome Fund New Investigator Award in the Toxological Sciences, a Camille and Henry Dreyfus New Faculty Award, and a National Science Foundation Career Award.*

*Hening Lin was born in Tsingdao, China in 1975. He completed his undergraduate study at Tsinghua University in Beijing, China and received his B.A. in Chemistry in 1998. He then went to Columbia University and joined the research group of Virginia W. Cornish. He is one of V.W.C.'s first students and is now a fifth-year graduate student, working on the development of a cell-based assay for screening large collections of molecules simultaneously based on function. He is the Arun Guthikonda Memorial Fellow of 2001.*

solution or a solution of the ligand. The same assay can be used, regardless of the particular protein–molecule interaction. Enzymatic turnover is more of a challenge. For the most part, researchers have focused on reactions in which the chemistry provides an intrinsic read-out—when the product is fluorescent, the substrate is a peptide or the product is an essential metabolite. With recent success in this area, attention has turned to the design of assays that are reaction-independent and compatible with a broad range of chemical reactions.

In this Review, rather than focus on a particular method, we have tried to compare and contrast these different methods and to ask which methods are best suited to which problems. For high-throughput assays, the key question, of course, is one of numbers. For studies of protein function or to test all the proteins in a given cell line, we know what these numbers are and the problem is clear. For example, *Sacchromyces cerevisiae* encodes approximately 6000 different proteins, and the human genome compromises about 30000 proteins. It is more difficult to evaluate the number of tests required for the evolution of proteins with new properties, because the question is tied to the answer. We do not know how many variants need to be tested to change enzyme specificity or to create an enzyme that catalyzes new reactions. It is worth reemphasizing that even for a small, monomeric protein, the potential diversity is staggering. Given that 32 DNA codons, half of the genetic code, are sufficient to encode all 20 amino acids found in proteins, to test all possible combinations of all 20 amino acids at every position in a 200-residue protein represents a library of $32^{200}$ variants. If we consider the necessity of insertions and deletions, the complexity of the problem increases. Although, interestingly, if all 20 amino acids at each position in a 200-residue protein are tested independently of one another, the library size is only $32 \times 200$. As a point of reference, 1 mg of plasmid DNA, a tractable amount of DNA to prepare routinely on a large scale, represents about $10^{14}$ plasmid DNA molecules. Thus, if the DNA is the limiting reagent, approximately $10^{12}$ positions can be randomized independently of one another, but only nine positions can be randomized if the positions are all varied simultaneously. A few methods are beginning to approach the limit at which the DNA is the limiting reagent, but for the most part other steps in the process limit the numbers. For each method for tagging a protein with its DNA sequence, we look at the steps that limit the numbers, what the numbers are, and what the possibilities are for improving the limiting steps. We are only now beginning to have a sense of how many residues must be varied and whether or not these variations can be made independently of one another to change protein specificity; we do not yet have a feel for this problem when the function of the protein is changed. The examples in this Review are presented with the aim of addressing the question of what numbers have been needed to answer different questions.

In addition to the numbers, we focus on the generality of these methods and the technical ease with which they can be carried out. There are two issues with respect to generality: First, is the method compatible with a broad range of proteins with different structures? Small, monomeric, cytosolic proteins tend to be easy to handle and can be displayed readily in a variety of formats. Oligomeric, membrane-bound, and more complex proteins are often difficult to express outside the cell. For protein engineering this is less of an issue because one can just work with a protein that is easy to handle. To test all the proteins expressed in a given cell line, however, one must be able to display a variety of different types of proteins. The other issue of generality is in the chemistry and the assay. Ideally one wants methods that are compatible with a broad range of molecular interactions and reactions. By considering the applications carried out with different methods, we hope to show which methods are better suited to binding assays and which to assays for enzyme catalysis, and for which methods one can develop assays that are reaction-independent. Finally, as with any chemical technique, the technical ease with which a method can be used often will dictate its utility. Ideally, methods should be accessible to laboratories that do not specialize in the methods themselves. By presenting selected examples in detail we hope to draw out issues such as whether or not specialized equipment is needed to display the proteins or carry out the assay, what the frequency of false positive results is and how easy they are to eliminate with a secondary assay, and finally how straightforward it is to amplify the DNA at the end of the assay for further rounds of mutagenesis or for sequencing.

This Review is organized around the method used to tag the protein with its DNA sequence. We have divided these methods into those based on a physical linkage, compartmentalization, or spatial separation. A brief overview of new approaches is presented at the end of each section. For each method, we describe how the method works, key technical improvements, and selected applications. The applications have been chosen to address the questions of the numbers, generality, and technical ease. There have been impressive technical improvements and achievements in each of these areas, which we cannot even begin to do justice in this Review, and so references are given at the beginning of each section for more detailed reviews in each area.

## 2. Physical Linkage Method

Perhaps the most straightforward way to tag a protein with its DNA sequence is simply to create a physical link between each protein and the DNA that encodes it. This link can be direct (as in the case of mRNA display or peptide on plasmid) or indirect (phage and cell-surface display). A variety of approaches have been developed to create this physical link. These strategies lend themselves naturally to assays based on binding, and there have been several impressive applications in the engineering of proteins with new specificities. Recently, several groups have reported variations that allow them to measure catalytic activities. The library sizes obtained by using these methods are closest to the DNA limit of $10^{14}$ different molecules. The technical issues to be considered in these methods are the stability of the physical linkage and whether it is compatible with a wide range of proteins.

### 2.1. Phage Display

Phage display is one of the first methods introduced for physically linking a protein with its DNA sequence. Not surprisingly, some of the most exciting advances have been made with this technique, which has been reviewed extensively.[1–4] The phage particle provides the physical connection between each protein and its unique DNA sequence. The M13 phage particles commonly used in phage display consist of a single-stranded DNA molecule (the phage genome), which is surrounded by thousands of copies of the phage coat proteins (Figure 2A). These coat proteins are encoded by the

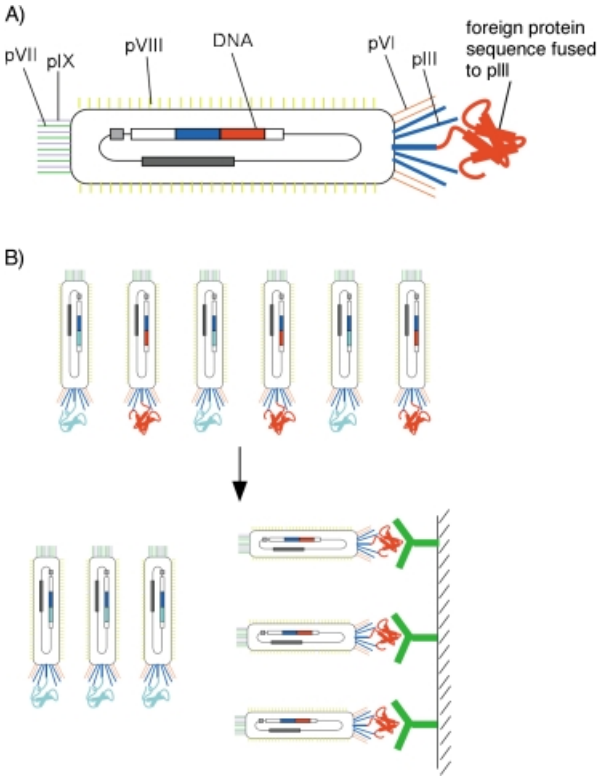



Figure 2. A) The structure of the M13 filamentous phage used in phage display. A phage particle consists of a single stranded DNA molecule surrounded by several coat proteins—pVIII, the major coat protein, pVII and pIX on one end, and pIII and pVI on the other end. A foreign protein sequence can be fused to any of the coat proteins. Shown here is a protein (in red) fused to the N-terminus of pIII (in blue). B) Enrichment of phage by affinity purification (panning).

single-stranded DNA genome that they encapsulate. In phage display, the DNA that encodes the protein of interest is fused to one of the coat protein genes such that the DNA sequence is part of the phage genome and the protein is expressed on the phage surface as a fusion with the coat protein. The coat proteins include thousands of copies of pVIII (the major coat protein), five copies of pIII and pVI at one end of the phage particle, and five copies of pVII and pIX at the other end (Figure 2A). In 1985, Smith introduced the idea of using phage particles to enrich protein libraries by using affinity chromatography to purify the phage particles that express foreign peptide sequences, EcoRI endonuclease fragments in this example, on their surface.[5] EcoRI endonuclease fragments were inserted into coat protein pIII by cloning the EcoRI endonuclease gene fragments in frame with the pIII gene in the phage genome. These modified phage were shown to retain their viability and infectivity. Further, preincubation with anti-EcoRI antibodies significantly reduced infectivity. As a whole, this work showed that foreign protein sequences can be displayed on the phage surface in an immunologically accessible form. Finally, anti-EcoRI antibodies were adsorbed on a polystyrene petri dish and were used to purify selectively phage that display the EcoRI endonuclease fragments from a background of wild-type phage.

#### *2.1.1. Technical Considerations*

Phage display libraries are limited in size by the transformation efficiency of bacteria. In a typical phage display experiment, a phagemid DNA library is first constructed in vitro, and then transformed into competent *Escherichia coli* (*E. coli*) cells. Assuming that 10 μg of DNA ($10^{12}$ DNA molecules) is transferred into 200 μL of competent cells ($10^{11}$ cells), by using electroporation, $10^{10}$ cells can be obtained, each of which contains a unique DNA molecule. In practice, libraries with approximately $10^8$ elements are usually prepared. Recently, Sblattero and Bradbury used recombination in bacteria to produce a phage scFv library of $3 \times 10^{11}$.[6] In theory, the electroporation can be optimized with 1 mg of DNA and 20 mL of competent cells to give a library of $10^{12}$ members.[7] In the transformants, proteins will be expressed from the phagemid, and phage particles will assemble if the fusion protein has no effect on this process. Phage particles are then harvested from the transfected bacteria and selected for proteins with specific binding properties by passing them through an affinity matrix in a process called "panning". Non-binding phage will be washed away, and binding phage can then be eluted with a high salt or low pH solution or a solution of the ligand (Figure 2B). The selected phage can be used to infect bacteria, amplified and subjected to further rounds of panning with more stringent selection conditions. Finally, the selected clones will be analyzed by sequencing and in vitro binding assays.

Since the initial report by Smith, there have been several important technical improvements to phage display. Co-infection with helper phage has been used to control the valency of display.[8] Besides pIII, other coat proteins, including pVIII,[9] pVII, and pIX,[10] have also been used to control the valency, orientation, or allow multiple components to be displayed on the phage surface. A particularly clever derivation, selectively infective phage (SIP)[11, 12] takes advantage of the modular structure of coat protein pIII, which consists of three domains: N1 (68 amino acids), N2 (132 amino acids), and CT (149 amino acids) (Figure 3A). N1 is required for *E. coli* infection. N2 specifically recognizes the F-pilus and helps to improve infectivity but is not essential. The CT domain forms part of the coat and is required for phage morphogenesis. Duenas and Borrebaeck were the first to demonstrate this approach:[11] In their initial report, anti-hen-egg white lysozyme (HEL) antibody fragments were displayed on the phage surface as a fusion to pIII proteins that lack N1. These phage were not infective since they lacked the

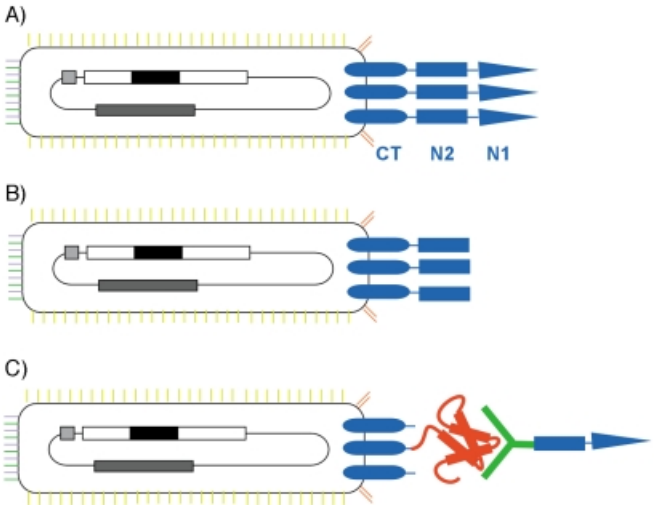



Figure 3. Selectively infective phage. A) In wild-type phage, coat protein pIII consist of three domains: the N-terminal domains N1 and N2 and the C-terminal domain CT. B) Phage particles that lack N1 are not infective; phage particles that lack N1 and N2 (not shown) are also not infective. C) The infectivity can be restored if a protein fused to CT can interact with a protein fused to N1 – N2.

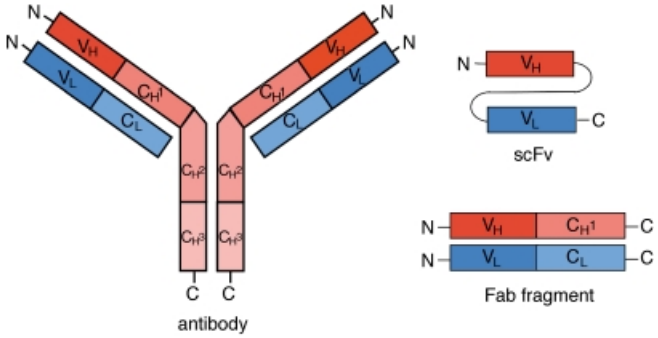



Figure 4. Antibody fragments displayed on phage. An antibody consists of two polypeptide chains, a light (L) chain and a heavy (H) chain. The subunit composition is $L_2H_2$. Each chain has a variable region ($V_L$, $V_H$) and a constant region ($C_L$, $C_{H1}$, $C_{H2}$, $C_{H3}$). The light chain and the heavy chain are connected to each other by disulfide bonds, and the two heavy chains are also connected by disulfide bonds. Antibodies can be displayed on phage as an scFv (single chain Fv fragment) where the heavy chain variable domain ($V_H$) is fused to the light chain variable domain ($V_L$), or as a Fab fragment where the light chain ($V_L - C_L$) and the N-terminal half of the heavy chain ($V_H - C_{H1}$) are displayed as a heterodimer.

N1 domain (Figure 3B). But when these phage were assembled in *E. coli* cells that express an HEL – N1 fusion protein, the binding interaction between the anti-HEL antibody fragment and HEL restored pIII function and phage infectivity (Figure 3C). Thus, phage that display the anti-HEL antibody were enriched $10^5$-fold when used to infect bacteria. SIP has two significant advantages over traditional phage display. First, because it circumvents the affinity-panning step, it can be carried out more rapidly than conventional phage display. Second, in theory, it overcomes a significant technical limitation of phage display—the inability of many proteins to be expressed as fusions with pIII on the phage surface.

### 2.1.2. Protein – Protein Interactions

Short peptide sequences were among the first "proteins" to be displayed on phage. Phage-displayed peptide libraries have been used to search for optimal epitope sequences,[13] peptide ligands,[14] enzyme inhibitors,[15] and even to create peptides that can discriminate semiconductor monolayers with subtle differences in structure.[16] Phage peptide libraries are now commercially available. Phage display has been used most extensively in the field of antibody engineering in the hope that it could replace time-consuming hybridoma technology.[17, 18] A major hurdle was the successful display of antibodies on the phage surface (the different antibody fragments employed are shown in Figure 4). It was first shown in 1990 that the heavy and light chain variable domains of an antibody can be displayed as a single chain Fv fragment on the surface of fd phage.[19] Soon after, phage display of Fab fragments was also reported.[20, 21] Once this hurdle was overcome, many successful applications were reported. Different phage antibody libraries, including immunized, nonimmunized, and synthetic libraries, have been used to isolate new antibodies and improve the affinity of existing antibodies for a given antigen. Catalytic antibodies have even been selected by using phage display.[22–24]

Phage display is clearly well suited to applications built around a common protein fold, such as antibodies and peptides. It is less clear, however, how well the technique will work with problems such as assigning function to proteins expressed in a cDNA library with thousands of different proteins. Shanmugavelu et al. recently reported the isolation of a novel protein that binds to juvenile hormone esterase (JHE) from an insect tissue cDNA library by using phage display.[25] Juvenile hormone (JH) acts in conjunction with ecdysteroids to control gene expression in insects. JHE is critical to insect development since it degrades JH. The authors were interested in identifying proteins that interact with JHE and might regulate its activity. In this application, the library size was not an issue since the cDNA library only includes approximately $10^4$ proteins. The challenge instead was to express the thousands of different proteins included in the insect tissue cDNA library. Because of the presence of a high percentage of stop codons in the cDNA library, the cDNA library could not be expressed as an N-terminal fusion to pIII or pVIII. Instead, the cDNAs were expressed as C-terminal fusions to the Fos leucine zipper domain, whereas the Jun leucine zipper domain was fused to the N-terminus of pIII. The heterodimerization of Fos and Jun effectively displayed proteins on the phage surface. Four rounds of panning were used to enrich proteins that bound to JHE immobilized on one well of a polystyrene 24-well microtiter plate. In the first three rounds of the selection, acidic buffer was used to elute bound phage. In the fourth round, purified JHE was used. The eluted phage was then used to infect *E. coli*. Phage particles isolated from individual colonies were assayed for JHE-binding activity in a 96-well plate format by using an enzyme-linked immunosorbent assay. Notably, multiple rounds of panning were required, not to test more proteins, but because the signal-to-noise in the panning step is quite low. Nine positive clones were sequenced, and one of

them was found to encode a protein, P29, which had not been characterized previously.

### 2.1.3. Protein – DNA Interactions

After the antibodies, DNA-binding proteins were one of the first classes of proteins whose binding specificity was engineered by using phage display.[26, 27] Pioneering work in this field was carried out by Rebar and Pabo.[28] They displayed the three zinc fingers of the Zif268 protein on the surface of filamentous phage as a pIII fusion. The four amino acids in the first zinc finger were chosen for randomization because they had been shown to be most important for DNA recognition, which resulted in a $10^6$-membered library that was readily covered by a pool of $10^{10}$ phage particles. These phage fingers were then selected against three biotinylated duplex DNA sequences immobilized on streptavidin-coated microtiter wells. The DNA sequences (GACC, GCAC, or CCTG) were different from the wild-type DNA sequence (GCGC) in the region recognized by the first finger. Zinc fingers that could bind to the modified DNA sequences with $K_D$ values in the low picomolar range and with 10- to 100-fold preference for the modified over the wild-type DNA sequence were identified for the GACC and GCAC, but not the CCTG, sequences. Again, five rounds of selection were required because of the high background of nonspecific phage particles carried through in the panning step (Figure 13B). Interestingly, the authors note that during the course of panning with the Zif library, retention efficiencies began at 0.01% and rose to ca. 1% in cases where the selection was successful. These and other similar results show that for a natural protein, whose mode of recognition is well understood, targeted mutations to produce libraries on the order of $10^6$ are sufficient to modify protein specificity.

### 2.1.4. Protein-Small Molecule Interactions

The next question, then, is how many more variants need to be tested to go beyond changes in specificity to the de novo evolution of recognition. While antibodies, of course, have been routinely engineered to bind different molecules, it is generally assumed that the antibody scaffold is uniquely malleable. There have been several successes with protein-protein interactions,[29] but a major breakthrough recently was proof that a lipocalin protein, the bilin-binding protein (BBP) from the butterfly *Pieris brassicae*, could be evolved to bind small molecules de novo.[30, 31] The lipocalins have an eight-stranded antiparallel β-barrel structure, and the four loops on one end of the barrel form the entrance to the binding pocket. The sixteen residues that directly contact bilin, which lie along these four loops, were completely randomized. Since only $10^8$ individual transformants were obtained, only a very small portion of all possible $10^{24}$ lipocalin variants was tested. The mutant BBPs were fused to pIII and monovalently displayed on phage. Selection was achieved by panning the phage fusion library against BSA or RNase conjugated fluorescein. After six rounds of panning to increase the percentage of specific phage eluted from the fluorescein conjugate, mutant BBPs with $10^{-7}$ M affinity and high specificity for fluorescein were isolated.[30] Similarly, mutant BBPs with high affinity and specificity for the steroid digoxigenin were obtained.[31] Further experiments are needed to test whether the lipocalin fold, like the IgG fold, is a privileged scaffold, or in fact libraries on the order of $10^8$ are sufficient to engineer molecular recognition de novo.

### 2.1.5. Enzyme Catalysis

While phage display seems naturally suited to selections based on binding, a more difficult challenge is to adapt this powerful technique to catalysis. Catalytic antibodies that catalyze a broad range of transformations have been generated by evolving catalytic antibodies that bind to transition state analogues. In theory, it should not be difficult to adapt this method to phage display since phage display selection is naturally based on binding. For example, catalytic antibody 17E8, obtained by immunization with a norleucine phosphate transition state analogue, catalyzes the hydrolysis of amino acid phenyl esters.[32] Baca et al.[23] grafted the antigen binding loop of 17E8 onto a human antibody framework to improve its recombinant expression efficiency and displayed it on pIII as a Fab fragment. Site directed mutagenesis was used to generate humanized 17E8 (hu17E8) libraries. Five to six rounds of panning were used to select for mutants that bind to the norleucine phosphate transition state analogue. A mutant with two-fold increased catalytic activity was identified. However, this mutant bound the transition state analogue less tightly than the original hu17E8, while all the mutants isolated with increased binding affinity gave lower catalytic efficiency. This example illustrates that binding to a transition state analogue and catalytic activity are not the same. Surprisingly, despite widespread use in the catalytic antibody field, there are few reports of using transition state analogues to generate catalysts with phage display. Attention seems to have turned to direct selections for catalytic activity.

Suicide substrates have also been used in the affinity panning step to select for enzymes based on catalytic activity. Soumillion[33] and coworkers synthesized a $\beta$-lactamase suicide inhibitor linked to biotin through a disulfide bond linker. $\beta$-Lactamase was fused to pIII with a proteolytic cleavage site in between them. Phage particles displaying active and inactive $\beta$-lactamases were incubated with the suicide inhibitor, after which the active enzymes were labeled with biotin. Panning through streptavidin beads then selectively enriched phage particles displaying active enzymes. A similar approach that selects enzymes by linking phage infectivity to catalytic activity was also reported.[34] While suicide substrates have been used to isolate the wild type enzyme, it is not clear that they can be used to discriminate enzymes with different catalytic efficiencies.

What is really wanted is a direct assay for catalytic activity, ideally one that is reaction independent. Recently several related methods,[35–39] where the basis for selection is the formation or cleavage of a bond from solid support or differential recognition of substrate and product, have been reported. In one such example, a non-specific nuclease, staphylococcal nuclease (SNase), was enriched from a phage library that contained unrelated proteins.[35] As shown in

Figure 5, both SNase and the DNA substrate were displayed simultaneously on the phage surface. SNase, an enzyme that hydrolyzes RNA and DNA, was displayed as a pIII fusion using a phage vector. A Fos–Jun leusine zipper heterodimer

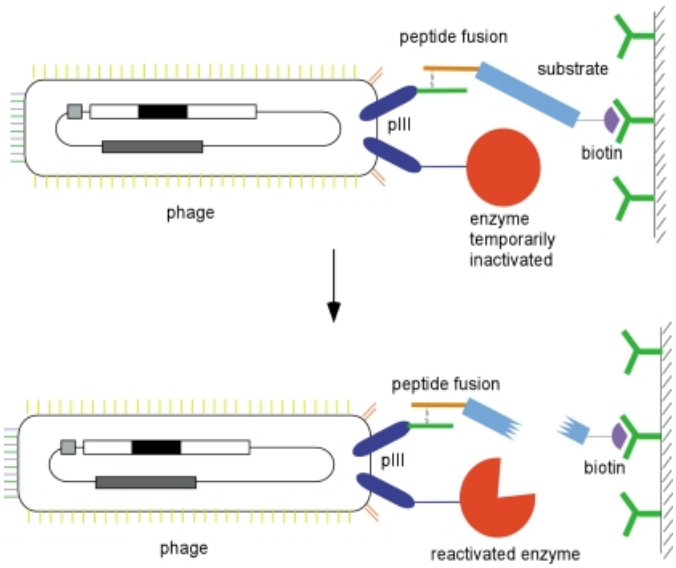



Figure 5. A selection for enzymatic activity by using phage display.[35] The substrate, an oligonucleotide, was attached to the phage particle using a peptide heterodimer. The pIII–peptide fusion was provided by a helper phage. The enzyme, staphylococcal nuclease (SNase), was displayed as a pIII fusion by using a phagemid vector. SNase was first displayed in an inactive form by leaving out $Ca^{2+}$ (top). Then the phage were incubated with the complex of peptide substrate and biotin and absorbed to streptavidin beads. After the unbound phage were washed away, the SNase was reactivated by adding back $Ca^{2+}$. Phage that displayed active SNase enzymes cleaved the DNA substrate, thus releasing the phage from the solid support (bottom).

was used to display the DNA substrate. Helper phage provide the pIII–Fos fusion, and the biotin–DNA substrate–Jun fusion was prepared using synthetic chemistry. Phage displayed SNase was first displayed in an inactive form by leaving out $Ca^{2+}$ from the buffer, incubated with Jun–substrate–biotin, then absorbed to the strepavidin beads. After the unbound phage were washed away, SNase was activated by adding $Ca^{2+}$ to the buffer. Phage displaying active SNase cleaved the substrate, releasing the phage particle from the solid support. It was shown that phage displaying SNase could be enriched 100-fold after one round of panning from a 1:100 mixture of phage displaying SNase and control phage displaying a Fab fragment. Very recently, this approach was used to convert *Thermus aquaticus* DNA polymerase into an RNA polymerase.[88] The use of the leucine zipper peptides to incorporate the substrate should allow the chemistry to be varied readily. As currently designed, this assay is limited to metal-dependent enzymes in which enzymatic activity can be turned on and off during different steps in the selection procedure. It is not clear that this switch is fast enough to distinguish between enzymes with different levels of catalytic activity.

## 2.2. Ribosome display and mRNA-peptide fusion

In this method, the linkage between phenotype and genotype is provided by the non-covalent or covalent complex formed between the translated peptide and the mRNA that encodes it, as shown in Figure 6.[40] This method was first

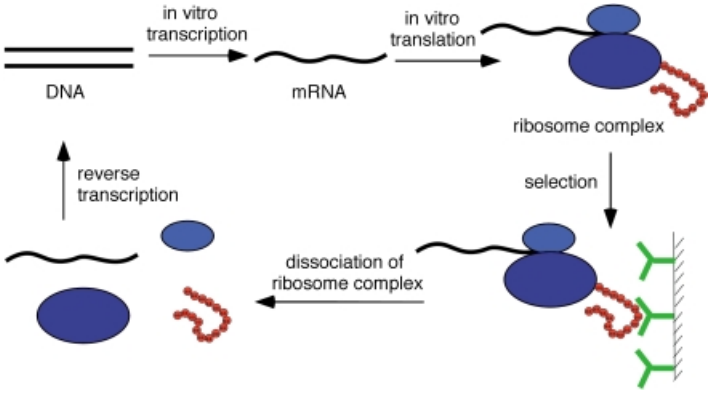



Figure 6. Ribosome display.[41, 42]

demonstrated by Mattheakis et al.[41] in 1994 and then optimized by Hanes and Plückthun.[42] Using an *E. coli* S30 coupled transcription/translation system, Mattheakis et al. in vitro transcribed and translated a library of DNA sequences that encode short peptides. When the translation reaction was stopped by the addition of chloramphenicol, the translated peptide and mRNA were found to remain associated with the ribosome complex. The ribosome complexes were selected based on binding of the peptides to an immobilized antibody. The mRNA was then dissociated from the complex, reverse transcribed to cDNA and amplified, enriching genes encoding peptides that recognized the antibody target. These selected DNA sequences can be subjected to further rounds of transcription/translation and selection (Figure 6).

### *2.2.1. Technical Considerations*

In Mattheakis's experiment,[41] 400 ng DNA ($10^{12}$ molecules) was transcribed to $3 \times 10^{12}$ mRNA molecules in a 50 μL transcription/translation reaction containing about $3 \times 10^{12}$ ribosomes (the ribosome concentration is ca. 0.1 μM). At the end of the transcription/translation reaction, about 27 % of the mRNA molecules were found to be complexed with the ribosomes. However, the yield of ribosome complex formation is dependent on the length of the transcribed peptide. For whole proteins, even under optimized conditions, the yield is only 0.2 %.[42] Thus for short peptides, 1 mg of DNA would give a library size of $10^{15}$, but for large proteins only $10^{12}$.

A problem with ribosome display is the instability of the ribosome–mRNA–peptide ternary complex. It is difficult to keep the complex intact during the selection steps. A major technical advance was the introduction of a covalent linkage between the peptide and the mRNA.[43, 44] The covalent linkage between the peptide and the mRNA was introduced using puromycin, an antibiotic that mimics the aminoacyl end of tRNA (Figure 7 A). As shown in Figure 7 B, puromycin was

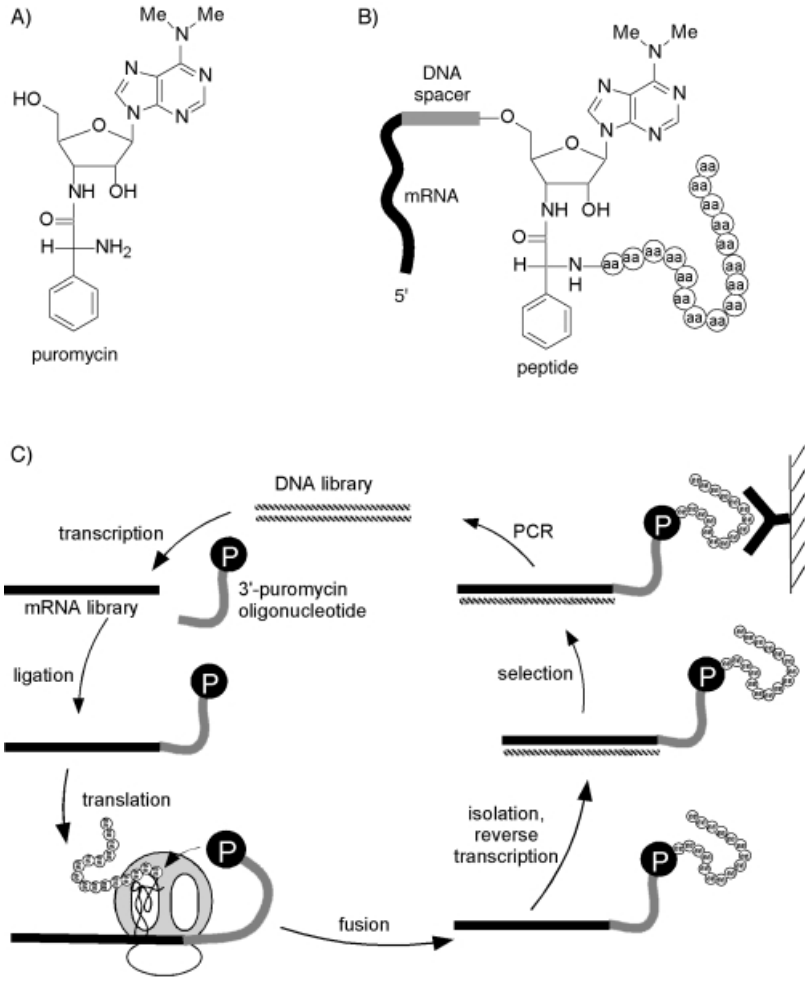



Figure 7. A) Structure of the amino acyl–tRNA analogue puromycin. B) The covalent complex formed between the mRNA and the peptide through puromycin. C) Typical mRNA–peptide fusion experiment procedure. Figure 7C reproduced from reference [46], Copyright (2000), with permission from Elsevier Science.

attached to the 3′ end of mRNA lacking a stop codon via a DNA spacer. When the mRNA was translated in vitro, the ribosome paused at the DNA spacer, allowing the puromycin molecule at the end of the message to enter the ribosome A site and react with the nascent polypeptide. This method is also called mRNA display. Optimized conditions for the puromycin coupling reaction that increase the efficiency with which puromycin adds to the stalled polypeptide recently have been reported.[45, 46]

A typical mRNA display experiment is shown in Figure 7C. A DNA library is first in vitro transcribed to give the mRNA library. Then the total mRNA is ligated to a chemically synthesized 30-bp DNA sequence with puromycin at the 3′ end. An in vitro translation reaction is then carried out for these mRNA-DNA-puromycin hybrid molecules. Covalent complexes between the mRNA and the peptides are formed via puromycin. The complexes are isolated from the reaction mixture, reverse transcribed, and selected against an immobilized antigen or ligand. Selected complexes can then be PCR amplified for sequencing or further rounds of mutation and selection.

The mRNA-peptide fusion method also increases the library size. In the initial paper by Roberts and Szostak[43], 5 picomoles of mRNA molecules were used in a 25 μL translation reaction, and 0.7% of the mRNA were incorporated into mRNA-peptide fusions. Thus the library size was only $2.1 \times 10^{10}$. Since this initial paper, the coupling conditions have been optimized and coupling yields as high as 40% for short peptides have been reported.[46] The potential disadvantage to this method is that all steps are carried out in vitro, making it difficult to express membrane proteins or proteins that require post-translational modifications.

#### *2.2.2. Protein-Protein Interactions*

As with phage display, the first applications of ribosome display have been in the field of antibody engineering. For example, Hanes et al. evolved picomolar affinity scFvs against insulin from a synthetic antibody library.[47] While there are few reports of using ribosome display for evolving proteins with new specificities or activities, presumably this is simply because this method is relatively new and these types of applications are just yet to be undertaken. In theory, ribosome display should be well suited to the same types of problems as phage display.

Szostak and co-workers recently took advantage of the diversity of mRNA display to isolate high-affinity binding proteins from completely random peptide libraries.[48] They constructed a random peptide library 88 amino acids in length using the optimized puromycin-coupling method.[45, 46] With an estimated library size of $7 \times 10^{12}$, only a small percentage of all possible 88-mers were examined. After seven rounds of affinity panning against immobilized streptavidin, peptides with $K_D$s for streptavidin as low as 5 nM were isolated. Similarly, nanomolar ATP-binding polypeptides were selected from a random peptide library.[49] As in phage display, multiple rounds of panning are being used not to test more variants, but to enrich the RNA-peptide fusions for true hits because a high percentage of false positives are carried through at each round of panning. These results are surprising and raise the interesting possibility that the evolution of protein structure and function may be a much easier problem than one would assume. Based on the panning percentages in the ATP-binding selection, the authors argue that the frequency of occurrence of functional polypeptides was on the order of $10^{-11}$ and that these peptides could only have been isolated with ribosome display, which allows libraries on the order of $10^{12}$ to be tested.

### 2.3. Peptide on plasmid

In this method, the noncovalent interaction between a DNA-binding protein and the plasmid DNA is used to link the protein to its DNA sequence. The first DNA-binding

protein used for this approach was lac repressor.[50] Lac repressor is encoded by the *lacI* gene and binds very tightly to the specific DNA sequence *lacO*. As shown in Figure 8, a library of $10^8$ dodecapeptides was fused to the C-terminus of lac repressor by cloning synthetic oligonucleotides downstream of the *lacI* gene in plasmid pMC5, which also contained tandem *lacO* sequences. Inside a cell, the plasmid directed the synthesis of the lac repressor-dodecapeptide fusions, which subsequently formed tetramers and bound to the lacO sites on the plasmid. After lysing the cells, the peptide-plasmid complexes were selected based on the binding interaction between the peptides and an immobilized antibody.

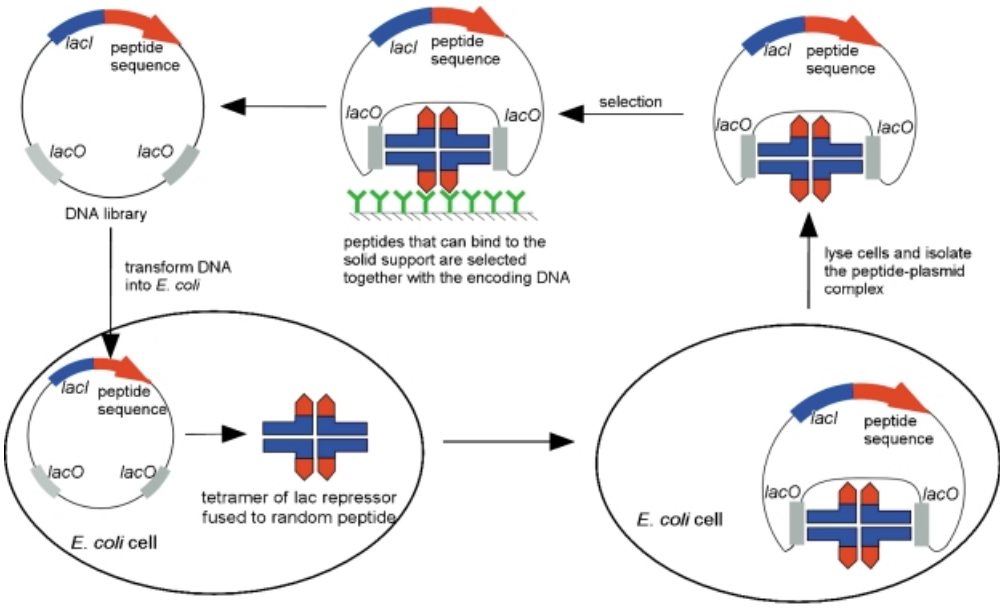



Figure 8. Peptide on plasmid.[50] Random DNA sequences are fused to the *lacI* gene in a plasmid that also contains *lacO* sites. This DNA library is then transformed into *E. coli*. Random peptides (red) fused to lac repressor (blue) are then synthesized and bind to the *lacO* sites on the plasmid as tetramers.

### 2.3.1 Technical Considerations

In this method, the peptides were displayed in a multivalent format since the lac repressor binds DNA as a tetramer, and each plasmid can bind up to two tetramers. Therefore this method is good for selecting low to moderate affinity binders. In order to make this method suitable to select for high affinity binders, Gates et al.[51] developed a lac repressor "headpiece dimer" as the DNA-binding domain. The headpiece dimer, which presented fewer copies of the peptide library, was shown to be suitable for selecting higher affinity binders, although the DNA binding mechanism of the headpiece dimer was not clear. The library size of this method is similar to phage display because it is also limited by the DNA transformation efficiency of bacteria.

### 2.3.2. Protein – Protein Interactions

While in theory this method could be used for the same variety of applications as with phage display, to date this method has been used primarily for screening peptide libraries. In 1997 Cwirla et al. reported the successful isolation of a nanomolar peptide agonist for the thrombopoietin receptor.[52] Of interest as a point of comparison, the authors not only used the peptide-on-plasmid method, but also phage display and ribosome display. They first used pVIII phage display and lac repressor, both multivalent display formats, to select for peptides with $IC_{50}$s ranging from 20 mM to 60 nM. Then they used headpiece dimer and ribosome display, which allow monovalent display, to further enrich for mutant peptides with $IC_{50}$s ranging from 20 to 60 nM. Finally, the headpiece dimer method was employed to select a high affinity peptide, AF12505, which had an $IC_{50}$ of 2 nM. The dimer of AF12505 showed an $IC_{50}$ of 0.5 nM and was as potent as recombinant human thrombopoietin. Although not clear why to this reviewer, the authors conclude that the peptide-on-plasmid selection is the most effective of all three methods for generating high affinity peptide binders. This application illustrates the importance of being able to control the valency of display. High valency is needed early on to detect low affinity binders, while low valency is needed in the final rounds to discriminate the highest affinity binders.

## 2.4. Cell Surface Display

Methods for displaying proteins on the surface of living cells have long been available, although these methods were not initially introduced as a means to screen protein libraries.[53] For example, the most common application of bacteria display is the development of live-bacteria-vaccine delivery systems. The principle of this method is relatively simple. A protein, or protein library, is fused to a membrane protein, which serves as an anchor to present the proteins on the cell surface. Selection or screening, normally fluorescence-activated cell sorting (FACS), can then be carried out using the cells displaying the proteins. In bacteria, the most frequently used anchor protein is an Lpp-OmpA hybrid, lipoprotein fused to amino acids 46 – 159 of the outer membrane protein A.[54] More recently, a yeast system has been designed that uses the yeast cell-surface receptor a-agglutinin, which is involved in the yeast mating pathway, as an anchor.[55] Mammalian systems have also been developed.[56]

### 2.4.1. Protein-Protein Interactions

To date, cell surface display has primarily been used for the display of peptide,[57] antibody,[58] and T cell receptor libraries,[59] although again this method should be compatible with

the same variety of applications as phage display. One of the most impressive examples is a recent report by Boder and coworkers in which they evolved scFvs that bind to fluorescein with femtomolar affinity.[60] The fluorescein binding antibody scFv 4-4-20 was mutagenized using error-prone DNA shuffling and displayed on the yeast surface via a-agglutinin. Cells were incubated with fluorescein-biotin, washed, and then incubated with 5-aminofluorescein competitor. Cells that remained bound to fluorescein-biotin were then labeled with streptavidin-*R*-phycoerythrin and isolated by fluorescence-activated cell sorting (FACS). After three additional rounds of mutagenesis and screening, an scFv with a $K_D$ of 48 fM, a dissociation constant even lower that that for the interaction between biotin and streptavidin, was isolated. Presumably because FACS has a low background, the library can be further mutagenized and screened with FACS, allowing antibodies with increased affinity to be generated.

### 2.4.2. *Enzyme Catalysis*

Recently, bacterial cell surface display has been used to screen for enzymatic activity. Olsen et al. displayed a mutagenized *E. coli* OmpT protein library on the surface of *E. coli* cells.[61] In this case, no anchor protein was necessary since OmpT itself is a membrane-associated protein. OmpT is a serine protease implicated in microbial pathogenicity and has a strong preference for the basic residues Arg and Lys in the P1 and P1′ subsites. The OmpT library was screened for proteases with modified substrate specificity that would hydrolyze substrates with Arg and Val in the P1 and P1′ subsites. The substrate used, shown in Figure 9, was a peptide derivatized with a fluorophore (Fl) and quenching fluorophore (Q), which acted as an intramolecular fluorescence resonance energy transfer (FRET) partner. The substrate was designed to bear a +3 charge on the Fl side, so that when the peptide was incubated with *E. coli* cells, whose surface is negatively charged, it would bind to the cell surface. Cells displaying mutants of OmpT that were active toward the Arg-Val substrate would cleave the peptide, releasing the quenching fluorophore Q and changing the fluorescence profile of the cells. Thus, cells displaying active enzyme could be isolated by FACS. Using this method, the authors screened $1.9 \times 10^6$ cells and isolated 352 successful clones. The 352 hits were further analyzed for protease activity in a 96-well plate assay. The most active OmpT mutant was shown to have a 60-fold improvement in catalytic efficiency for the Arg-Val substrate. This approach is similar to that reported with phage display, except the assay is based on FRET rather than cleavage/formation of a bond to solid support. This approach is general and should be compatible with a broad range of bond formation and cleavage reactions. Presumably the chemistry will be somewhat limited by the reliance on FRET. FRET has a marked distance dependence and is influenced by the orientation of the donor and acceptor molecules. Additionally, to use this method to detect a range of catalytic activities, the ratio of the enzyme molecules to the substrate molecules on the cell surface as well as their diffusion rates must be further characterized.

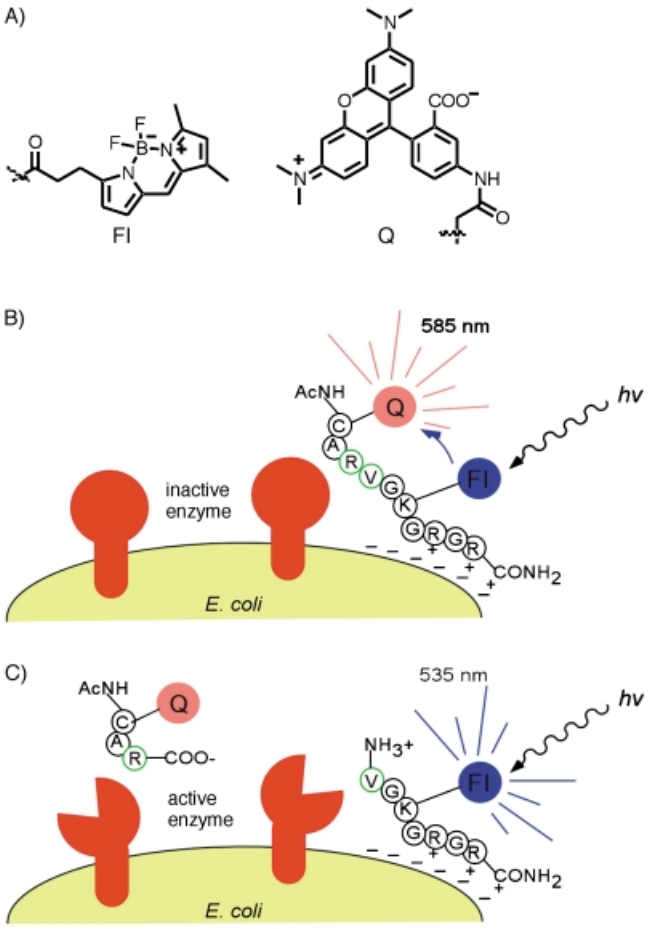



Figure 9. Screening for protease activity by bacteria display.[61] A) Structures of the fluorophore (Fl) and quenching fluorophore (Q) attached to the peptide substrate. B) and C) The enzyme library is displayed on the surface of *E. coli* cells, while the substrate with Fl and Q is attracted to the cell surface through the ionic interactions between the negatively charged cell surface and the positively charged substrate. When an inactive enzyme is displayed, the quenching fluorophore Q efficiently quenches the fluorescence of Fl, which results in fluorescence at 585 nm (B). When an active enzyme is displayed, the enzyme cleaves the substrate, releasing Q from the substrate and resulting in fluorescence at 535 nm (C). By detecting fluorescence of the cells, active enzymes can be isolated.

## 2.5. Summary

Phage display, as well as newer methods for physically linking a protein to its DNA sequence, provides a robust method for re-engineering the binding specificity of an existing protein scaffold. While the DNA itself limits the number of protein variants that can be tested to ca. $10^{14}$, phage display, peptide on plasmid, and cell surface display are all further limited by transformation efficiency. While in theory libraries on the order of $10^{10}$ can be achieved if the transformations are done with *E. coli*, in practice libraries on the order of $10^8$ usually are reported. Ribosome display shows the most promise for reaching the limit where DNA itself is the limiting reagent, and the puromycin coupling method can already easily produce libraries of $10^{12}$. Surprisingly these library sizes seem sufficient to modify protein function. Starting with proteins whose mode of action is well under-

stood, libraries on the order of $10^6 - 10^8$ have proven sufficient to generate proteins with new DNA or small molecule specificities. Further work is needed to ensure that these results are general and that engineering proteins with new specificities is in fact a tractable problem. It still remains unclear, however, how many protein variants need to be tested to evolve new function. All of these methods place constraints on the protein being displayed, and these limitations are best documented in the case of phage display. While not a problem for protein engineering, where presumably one would just begin with a well behaved scaffold, more work is needed to establish what percentage of proteins expressed in a given cDNA library are in fact displayed in an active form with these different methods. These methods are naturally inclined toward assays based on molecular recognition, and the same panning procedure can be used no matter what the protein-molecule interaction is. With the exception of peptide on plasmid, the linkage between protein and DNA is quite robust and places few constraints on the panning conditions. It will be a challenge, however, to adapt these methods to detect catalysis. Some clever approaches based on breaking or forming a bond to solid support or FRET recently have been reported. The difficulty will be to use these on/off assays to discriminate enzymes with a range of catalytic efficiencies. While effort must go into the preparation of these libraries, whether it be the phage particles or the in vitro transcription/translation system, the methods seem to be robust and are carried out by many laboratories. There clearly is poor signal to noise in the panning step, and multiple rounds of library amplification and panning are needed to identify true interactions. The fact that multiple panning steps are required makes it impractical to carry out multiple rounds of mutagenesis, limiting the number of protein variants ultimately tested. For most of these methods, the DNA can be amplified for sequencing or further rounds of mutagenesis easily, whether it be infecting bacteria with the phage particle, retransforming plasmid, or growing the cells. The only exception is ribosome display, where the RNA must be reverse transcribed before it can be amplified.

## 3. Compartmentalization Methods

A related approach to physically linking a protein to its DNA sequence is to restrict each protein and its DNA sequence to a distinct compartment. This compartmentalization is achieved naturally by introducing plasmid DNA encoding the protein into a cell. More recently, water-in-oil emulsions have been introduced as an artificial compartment that in a way is a stripped-down cell. With cell-based assays, the advantage is the ease of molecular biology and genetic techniques, although the library size is limited to the transformation efficiency of the cell. As with mRNA display, water-in-oil emulsions have the potential to reach the limit where DNA is the limiting reagent for library size. Both approaches seem well suited to assays based on enzyme catalysis. Arguably the most advances in evolving enzymes with new properties have been with cell-based assays.

### 3.1. Cell-Based Assays

Geneticists, essentially, have always been doing high-throughput screening, and much can be learned from the approaches taken in this field. With cell-based assays, the protein is linked to the DNA that encodes it because the DNA directs the synthesis of the protein inside the cell. The cell then provides a compartment that links each protein to its DNA sequence. These assays are powerful because millions of plasmid DNA molecules, each encoding a different protein or protein variant, can be transferred into cells en masse. The statistics of DNA transfer and the use of selectable markers ensure that each cell ends up with a single plasmid encoding a single protein variant. The cells can be physically isolated from one another simply by growing them at an appropriate dilution on solid media. At the end of the experiment, the DNA encoding the protein variant can be readily extracted from the cell for sequencing or further rounds of mutagenesis using standard molecular biology techniques. In a now classic experiment, Beadle and Tatum used genetic analysis in the early 1940s to identify enzymes responsible for the biosynthesis of essential metabolites (Figure 10).[62] They introduced random mutations throughout the chromosomal DNA of the ascomycete *Neuospora crassa* by treating the *Neurospora* with X-rays. Approximately 2000 individual mutated strains were then screened for growth in both rich and minimal medium, where the minimal medium contained only nutrients such as

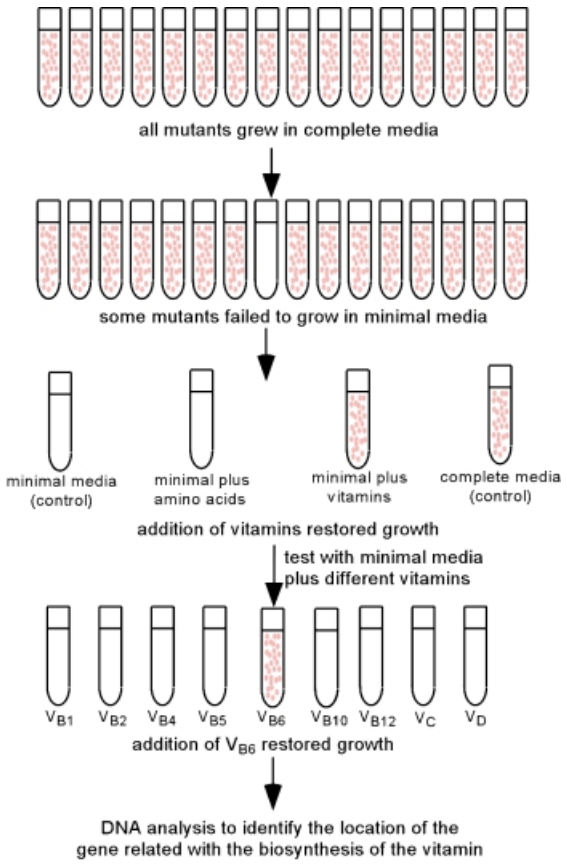



Figure 10. A genetic method to identify genes responsible for the biosynthesis of essential metabolites.[62]

biotin that *Neurospora* cannot make themselves. Three mutant strains were identified that had growth rates indistinguishable from the wild-type strain in rich media, but grew poorly or not at all in minimal medium. By adding back individual metabolites to the minimal media and then measuring the cell growth rates, the three strains were shown to be unable to synthesize vitamin $B_6$, vitamin $B_1$, and *p*-aminobenzoic acid, respectively. This basic approach of mutating the DNA encoding the proteins and then assaying for some detectable phenotype such as cell growth is the basis for genetic analysis. Most of the genes whose function is known have been identified through this type of method, and the functions of thousands of proteins have been determined in this manner. The contribution of Beadle and Tatum was to point out that the mutant strains could be screened based on enzymatic activity, as opposed to just cell death. As we discuss in this section, the challenge in fact continues to be developing clever assays that target the activity of interest.

#### *3.1.1. Technical Considerations*

Based on advances in molecular biology, a number of original genetic tricks have been developed that increase the ease with which cell-based assays can be carried out. One of the biggest hurdles in genetic analysis is "backing-out" which gene has the mutation of interest. For example, in the experiment by Beadle and Tatum described above, several additional years of experiments were required to determine which genes in the *Neurospora* genome had the mutations that affected cofactor biosynthesis and hence encoded the biosynthetic enzymes. A refined twist, introduced by Holden and co-workers in 1995,[63] is to replace every open reading frame in an organism with a unique 20 bp tag rather than simply mutating the chromosomal DNA. This way the gene responsible for a given mutant phenotype can be assigned rapidly based on its unique 20 bp tag. In the 1950s, Benzer introduced the strategy of complementation, which provided a general solution to this problem.[64] In genetic complementation, a pool of plasmid DNA encoding fragments of wild type chromosomal DNA is introduced into the mutant cell line and then the DNA fragment that complements the mutation is identified based on the recovery of the wild-type phenotype. In the *Neurospora* example above, plasmid DNA that encoded fragments of the wild-type *Neurospora* genome would be introduced into the mutant *Neurospora* deficient in vitamin $B_6$ biosynthesis. Only mutant *Neurospora* with a copy of the wild type vitamin $B_6$ biosynthetic gene would be able to grow in the absence of vitamin $B_6$, allowing the plasmid DNA encoding the vitamin $B_6$ biosynthetic gene to be distinguished readily. The recent availability of several complete genome sequences further simplifies genetic complementation because the plasmid DNA sequence can be compared to the complete genome sequence. In addition, it has allowed several labs to construct libraries of plasmid DNA not just encoding random fragments of chromosomal DNA, but instead explicitly encoding every open-reading frame (ORF) expressed in a given organism or cell line.[65]

A major limitation to genetic analysis is that it is limited to activities naturally found in the cell that are screenable or selectable. "Two-hybrid" assays provide a way to open these powerful genetic assays up to a wide variety of chemistry. The two-hybrid assay, which was first introduced as a method for testing potential protein-protein interactions in vivo, has now been adapted to detect protein-DNA, protein-RNA and protein-small molecule interactions, and, most recently, even catalytic activity.[66–72] The two-hybrid assay is based on reconstitution of eukaryotic transcriptional activators from dimerization of the DNA-binding (DBD) and transcription activation (AD) domains. As outlined in Figure 11 A, if the two proteins of interest interact, they effectively dimerize the DBD and the AD, recruiting the transcription machinery and activating transcription of a downstream reporter gene. The assay was demonstrated initially using two yeast proteins, the serine-threonine protein kinase SNF1 and the SNF1 activator protein SNF4, known to be physically associated in vivo.[73] SNF1 was fused to the GAL4 DNA-binding domain (DBD), and SNF4 was fused to the GAL4 activation domain (AD). A GAL4 binding sequence was placed upstream of a $\beta$-galactosidase reporter gene. Plasmids encoding the protein fusions and the reporter gene were introduced into yeast, and $\beta$-galactosidase synthesis levels were quantified using standard biochemical assays. Control experiments established that neither the DBD nor the AD domain on their own nor the individual protein chimeras induced $\beta$-galactosidase synthesis above background levels. $\beta$-Galactosidase synthesis levels were increased 200-fold when the DBD-SNF1 and SNF4-AD fusion proteins were introduced together. By comparison, the direct DBD-AD fusion protein activated $\beta$-galactosidase synthesis levels 4000-fold.

#### *3.1.2. Protein-Protein Interactions*

Traditional genetic assays and the yeast two-hybrid assay have primarily been used to identify natural protein-protein interactions. A particularly impressive recent application is the use of automation techniques to identify all possible protein-protein interactions in *S. cerevisiae*.[65] Every open-reading frame encoding a protein—there are ca. 6000 in *S. cerevisiae*—was fused both to the Gal4 DBD and to the Gal4 AD, and the two fusion libraries were screened against one another. The major challenge here is how to transform all combinations of the 6000 DBD and 6000 AD fusions into yeast and then how to assay so many cells. A library of $10^7$ is at the limit of the transformation efficiency of yeast and so in theory is achievable. Uetz and Glot and co-workers compared two approaches. One, they explicitly mated haploid MATa cells containing 192 DBD fusions with haploid MAT$\alpha$ cells containing the 6000 AD fusions in a spatially addressable format and assayed each well using a *HIS3* growth selection. Two, MATa cells containing the 6000 DBD fusions were mated with MAT$\alpha$ cells containing the 6000 AD fusions, and only diploids that survived in a *LEU2* growth selection were arrayed and analyzed individually. Interestingly, there were significantly more "hits" in the spatially addressable format, underscoring the importance of parameterizing new methods for high-throughput screening and the problem of distinguishing false positives and negatives in genetic assays.

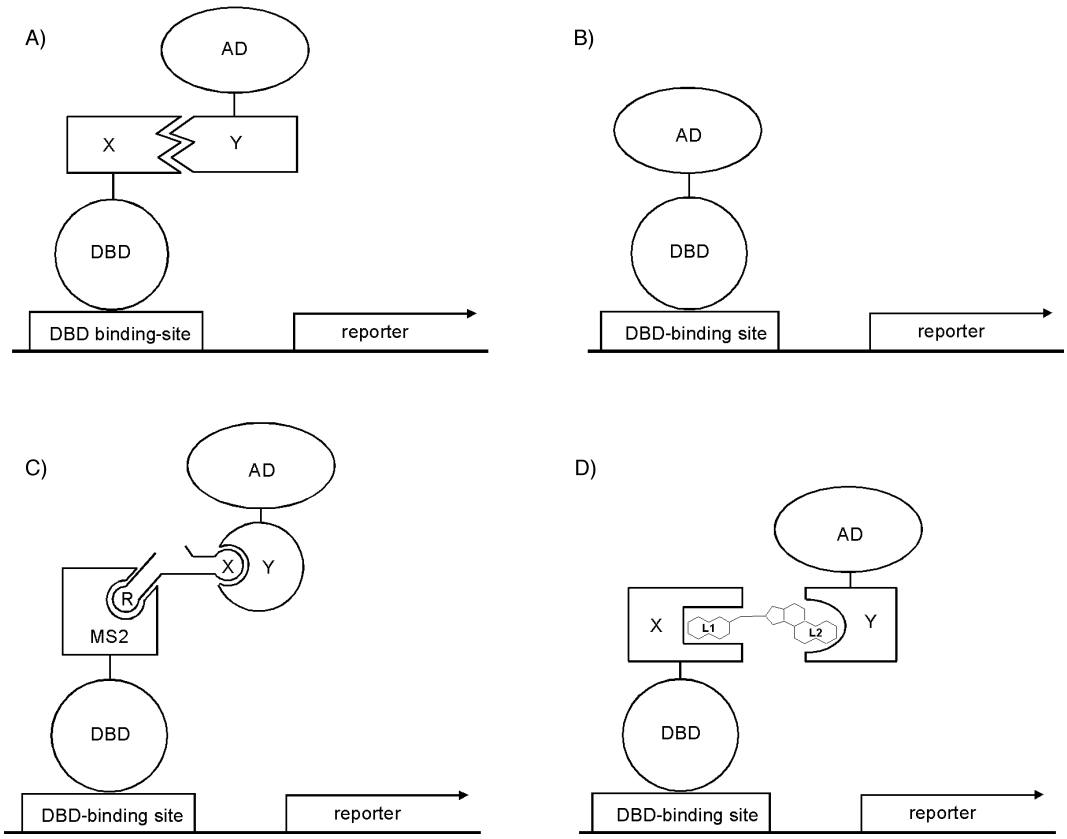



Figure 11. Different yeast *n*-hybrid systems that have been developed to study protein–protein, protein–DNA, protein—RNA, and protein–small molecule interactions. A) In the original version of the yeast two-hybrid system, transcriptional activation of the reporter gene is reconstituted by recruitment of the activation domain (AD) to the promoter region through direct interaction of protein X and Y, since protein X is fused to a DNA-binding domain (DBD) and protein Y is fused to the AD. B) In the one-hybrid assay, the AD is fused directly to the DBD. This assay can be used to assay either DBDs that can bind to a specific DNA sequence or the in vivo binding site for a given DBD. C) The three-hybrid system used to detect RNA-protein interactions has one more component than the yeast two-hybrid system: a hybrid RNA molecule. One half of the hybrid RNA is a known RNA (R) that binds to the MS2 coat protein (MS2) with high affinity and serves as an anchor. The other half is RNA X, whose interaction with protein Y is being tested. D) Another version of the yeast three-hybrid system can be used to detect small molecule–protein interactions. Ligand L1 which interacts with protein X is covalently linked to ligand L2. Thus if L2 interacts with Y, transcriptional activation of the reporter gene will be reconstituted.

While the two-hybrid method has been used primarily to detect natural protein-protein interactions, it should also be well suited to protein evolution. Brent and co-workers[74] demonstrated that the two-hybrid assay could be used to identify peptide aptamers that inhibit cyclin-dependent kinase 2 (Cdk2) from a library of random peptide sequences (Figure 12). The 20 residue peptide library was displayed, not as a simple extension to the AD, but rather in the active site of *E. coli* thioredoxin (TrxA). The TrxA loop library was fused to the B42 AD, and Cdk2 was fused to the LexA DBD. In a single round of assay, $6 \times 10^6$ B42-TrxA transformants, a very small percentage of the $10^{27}$ 20mers possible, were tested for binding to LexA-Cdk2. From this assay, they isolated 66 colonies that activated transcription of both a *LEU2* and a *lacZ* reporter gene. Impressively, these colonies converged on 14 different peptide sequences that bound Cdk2 with high affinity. Using surface plasmon resonance, the peptide aptamers were shown to bind Cdk2 with $K_D$s of 30–120 nM (Table 2). In kinase inhibition assays, the peptide aptamers had $IC_{50}$s for the Cdk2/cyclin E kinase complex of 1–100 nM.

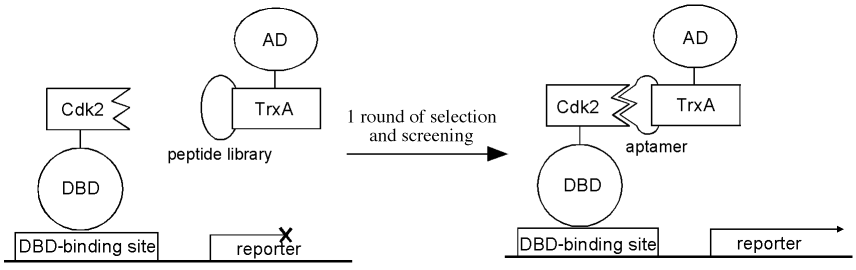



Figure 12. The yeast two-hybrid system can be used to evolve peptides that bind target proteins with high affinities.

What is particularly impressive about this experiment is that nanomolar affinity ligands are being isolated in a single round of selection. As in the phage display applications presented earlier in this Review, this example suggests that libraries on the order of $10^6$–$10^8$ may be sufficient to modify protein specificity. In contrast to phage display, however, only one round of selection is needed to distinguish true high affinity binders. Similar results have been obtained using peptide aptamers in a traditional genetic selection.[75]

Table 2. The sequences and binding affinities of the 14 different aptamers isolated in a yeast two-hybrid assay.[74]

| Aptamer | $K_D$ [nM] | Amino Acid Sequence |
|---|---|---|
| pep1 | ND[a] | ELRHRLGRAL SEDMVRGLAW GPTSHCATVP GRSDLWRVIR FL |
| pep2 | $64 \pm 16$ | LVCKSYRLDW EAGALFRSLF |
| pep3 | $112 \pm 1$ | YRWQQGVVPS NMASCSFRCQ |
| pep4 | ND | SSFSLWLLMV KSIKRAAWEL GPSSAWNTSG WASLSDFY |
| pep5 | $52 \pm 3$ | SVRMRYGIDA FFDLGGLLHG |
| pep6 | ND | RVKLGYSFWA QSLLRCISVG |
| pep7 | ND | QLYAGCYLGV VIASSLSIRV |
| pep8 | $38 \pm 5$ | YSFVHHGFFN FRVSWREMLA |
| pep9 | ND | QQRFVFSPSW FTCAGTSDFW GPEPLFDWTR D |
| pep10 | $105 \pm 10$ | QVWSLWALGW RWLRRYGWNM |
| pep11 | $87 \pm 7$ | WRRMELDAEI RWVKPISPLE |
| pep12 | ND | RPLTGRWVVW GRRHEECGLT |
| pep13 | ND | PVCCMMYGHR TAPHSVFNVD |
| pep14 | ND | WSPELLRAMV AFRWLLERRP |

[a] ND: not determined.

### 3.1.3. Protein-DNA Interactions

Early on, it was realized that, just as the yeast two-hybrid assay could be used to detect protein-protein interactions, transcriptional activators could be used directly, in a "one-hybrid" assay, to detect DNA-protein interactions (Figure 11B). In truth, this type of experiment was done before the one-hybrid assay was conceptualized as such. For example, as early as 1983 a His6 $\rightarrow$Pro Mnt variant was generated that preferentially binds a mutant Mnt operator using a transcription-based selection[76]. More recently, Pabo and coworkers adapted a bacterial two-hybrid assay[77] to evolve zinc finger variants with defined DNA-binding specificities (Figure 13).[78] In this assay, protein-protein interactions are detected based on dimerization of a DNA-binding protein and the $\alpha$-domain of RNA Polymerase (RNAP) and activation of RNAP-dependent transcription. To create a read-out for protein-DNA interactions, three tandem zinc-fingers were fused to Gal11, and Gal4 (which binds with high affinity to Gal11), was fused to the $\alpha$-domain of RNAP. In addition, a reporter system was engineered, such that if the zinc fingers bound with high affinity to the desired DNA sequence, they would activate transcription of a *HIS3* reporter gene. As with the peptide aptamers, zinc fingers with new DNA-binding specificities could be isolated from a library of circa $2 \times 10^8$ zinc finger variants in a single round of selection. The authors, who have used phage display routinely in the past,[28, 79] note that the ease of selection is in contrast to phage display, where multiple rounds of selection and amplification were required in their lab for the identical evolution experiment (Figure 13, Table 3).

### 3.1.4. Protein-Small Molecule Interactions

The two-hybrid assay has been extended to protein-small molecule and protein-RNA interactions by adding a bridging small-molecule or RNA to provide a yeast "three-hybrid" system (Figure 11C,D).[80–82] One half of the bridging molecule is used as an anchor and the other half is the molecule of interest. Licitra and Liu built a small-molecule yeast "three-

Table 3. Comparison of the zinc finger sequences (which bind to the target DNA sequence AAA), isolated by using a bacteria two-hybrid method and a phage-display method.[78, 79]

| | Bacteria two-hybrid system (1 round of selection) | | | | | | Phage display (5–8 Rounds of selection) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **-1** | **1** | **2** | **3** | **5** | **6** | **-1** | **1** | **2** | **3** | **5** | **6** |
| | Q | R | G | N | L | V | Q | R | T | N | I | T |
| | Q | K | T | N | M | V | Q | Q | H | N | K | L |
| | Q | K | Y | N | I | L | Q | R | N | N | L | L |
| | Q | R | Y | N | V | V | Q | A | N | N | R | T |
| | Q | K | G | N | M | V | Q | K | T | N | L | N |
| | Q | K | G | N | H | V | Q | H | G | N | V | A |
| | Q | K | G | N | M | V | Q | K | T | N | L | T |
| | Q | R | G | N | K | V | Q | K | T | N | D | T |
| | Q | R | G | N | K | T | Q | K | H | N | Q | V |
| | Q | L | G | N | M | V | Q | P | G | N | Q | T |
| | Q | K | G | N | K | V | Q | K | T | N | E | H |
| | Q | L | G | N | K | V | | | | | | |
| **Consensus:[a]** | **Q** | + | **g** | **N** | _ | **v** | **Q** | _ | _ | **N** | _ | _ |
| | N | S | G | A | Y | N | N | S | G | N | H | T |
| | N | S | G | A | W | N | N | S | G | A | S | N |
| | N | S | G | A | F | N | N | S | G | A | A | N |
| | N | S | G | T | H | N | N | S | G | A | T | N |
| | N | T | G | A | Y | N | N | S | G | A | T | N |
| **Consensus:[a]** | **N** | **s** | **G** | **a** | _ | **N** | **N** | **S** | **G** | **a** | _ | **n** |
| | G | S | G | A | Y | N | A | T | G | A | H | N |

[a] Capital letters indicate that all sequences showed this amino acid; lower case letters indicate a strong preference for this amino acid; "+" represents a preference for a positively charged amino acid; while "_" indicates that no preference could be determined.

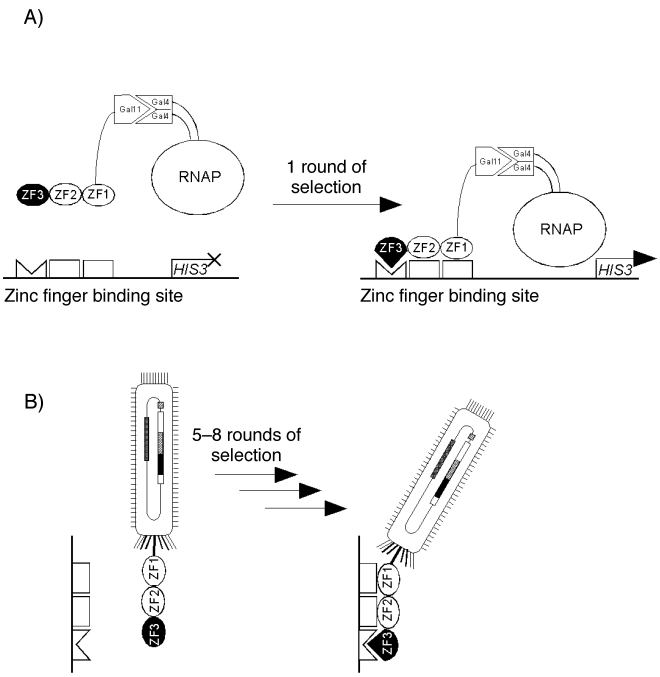



Figure 13. Development of zinc fingers specific for the DNA sequence AAA using a two-hybrid assay[78] and a phage display selection[26, 79]. A) The bacterial RNA Polymerase (RNAP) system was used for the two-hybrid assay. Zinc fingers (ZF) 1, 2, and 3 from the Zif268 protein were fused to the Gal11 protein. The Gal4 protein, which binds Gal11 with high affinity, was fused to the N-terminal domain of the $\alpha$-subunit of RNAP. A subsite sequence AAA was engineered at the binding site for ZF3, and six residues in ZF3 were randomized. Thus, if ZF3 bound to the AAA site with high affinity, the RNAP complex would be recruited, activating transcription of a *HIS3* reporter gene. Significantly, in just one round of assay, several proteins were identified that bound specifically to the target DNA sequence. These sequences, which fall into two distinct sequence categories, are shown in Table 3. B) A similar experiment was carried out by phage display, in which ZF 1, 2, and 3 from the Zif268 were displayed on the phage surface as a fusion to coat protein pIII.

hybrid" assay that uses the steroid dexamethasone as an anchor and demonstrated that it could be used to find the in vivo target of a drug. As a proof of principle, they showed that FKBP12 could be pulled out from a Jurkat cDNA library using Dex-FK506 as the bridging small molecule. Specifically, Dex-FK506 dimerized a LexA-Glucocorticoid Receptor (GR) fusion protein and a cDNA-B42 AD library. First a *LEU2* growth selection was used to select for proteins that bound to FK506. Then hits were confirmed by comparing the levels of transcription of a *lacZ* reporter gene with and without the Dex-FK506 small molecule. In the three-hybrid assays, the bridging molecule provides a convenient tool for ruling out false positives. The authors noted that while several FKBPs were present in the Jurkat cDNA library, only FKBP12 was identified. FKBP12 is the highest affinity FK506 binding protein and binds FK506 with a $K_D$ of 0.4 nM. Presumably the key to the success of the three-hybrid assays is the anchor. Anchors with high affinity to the DBD-receptor fusion are needed if small molecule and RNA-protein interactions with a range of affinities are to be detected. For example, our lab recently developed methotrexate (Mtx), which binds to dihydrofolate reductase (DHFR) with picomolar affinity, for use in this system.[83]

### *3.1.5. Enzyme Catalysis*

Genetic methods seem to provide particularly powerful screens and selections for enzymatic activity. The trick here is finding an activity that is screenable or selectable. One approach is to choose a reaction where the product is fluorescent. For example, Arnold and co-workers generated P450 variants that use hydrogen peroxide, rather than dioxygen and NADH, as a cofactor using a genetic screen (Figure 14).[84] Briefly, random mutations were introduced throughout the gene encoding cytochrome P450 using error-prone PCR. Naphthalene was used as the substrate in an *E. coli* strain that had been engineered such that any hydroxylated naphthalene products would be oxidatively coupled to produce fluorescent dimers. Cells containing P450 variants with the desired activity could then be identified by fluorescence digital imaging of cells grown on a petri plate containing the naphthalene substrate. Because a screen rather than a selection was used, the experiment was limited to a smaller number of protein variants, about 20000 per plate, than would be the case in a growth selection. In the first round, where 200000 colonies were screened, a large number of clones showed enhanced fluorescence. Three mutants from the group showing the highest fluorescence were character-

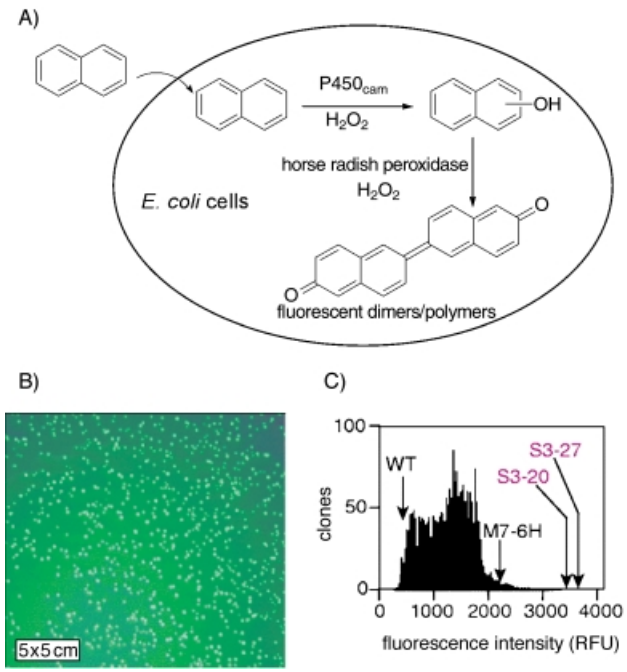



Figure 14. A genetic screen for evolving cytochrome P450 variants that use hydrogen peroxide as a cofactor.[84] A) The screening strategy: The screen is based on the detection of fluorescent products, produced by first hydroxylation of naphthalene catalyzed by $P450_{cam}$ variants and then oxidative coupling catalyzed by horseradish peroxidase. B) Fluorescence digital imaging of *E. coli* cells harboring $P450_{cam}$ variants. Cells with $P450_{cam}$ mutants were grown on plates that contained naphthalene and hydrogen peroxide. Fluorescence digital imaging was then carried out. The cells that contained an active $P450_{cam}$ mutant and therefore displayed higher fluorescence intensities were subsequently analyzed. C) The histogram showing the fluorescence values of ~70000 clones from the second generation of mutants. The fluorescence intensity of the wild-type $P450_{cam}$, the first generation mutant M7–6H, and the second generation mutants S3–20 and S3–27 are indicated. Reproduced from reference [84] with permission from Nature, Copyright (1999), Macmillan Magazines Limited.

ized and found to be about 10-fold more active than wt P450. In a second round, five active mutants were recombined using Staggered extension process (StEP), and the progeny were screened. Several variants were isolated from this step that showed about a 20-fold improvement in activity compared to the wild type enzyme. It is interesting to speculate if the activity might have been improved further if a larger number of mutants could have been tested. Here the problem may have been distinguishing proteins with a small increase in activity from those with a more significant change.

In addition to its utility in genetic analysis, complementation seems a natural choice for evolving proteins with new functions. The idea here is that, rather than testing fragments of the chromosome for complementation of a mutant phenotype, a protein is mutated and "evolved" to complement a given phenotype.[85, 86] A recent example by Yano and co-workers, in which they used this strategy to evolve an aspartate aminotransferase into a branched-chain amino transferase, illustrates the power of this approach.[87] First, they engineered an *E. coli* strain with a knock-out in the wild type branched-chain aminotransferase gene and showed that this strain was not viable unless the growth media was supplemented with Val, Ile, and Leu. The wild type aspartate aminotransferase gene was mutated and recombined using DNA shuffling and then introduced into the *E. coli* selection strain. The authors point out that DNA shuffling, a PCR-based method, could be used both to introduce point mutations and to generate recombined gene products. Five rounds of mutagenesis and selection were carried out. At each round, $10^6 - 10^7$ mutants were examined, and the stringency of the selection could be readily increased by changing the concentration of 2-oxovaline in the growth media, the expression level of the enzyme, or the incubation time. After the final round, a mutant enzyme with 13 point mutations was isolated that had a $10^5$-fold increase in $k_{cat}/K_M$ for $\beta$-branched amino acids and a 30-fold decrease in $k_{cat}/K_M$ for aspartate. The improvement in $k_{cat}/K_M$ here is impressive and likely reflects the importance of being able to modulate the stringency of a selection—first to detect poor catalysts and finally to demand high catalytic efficiency. The authors also address the question of how hard it is to alter specificity. By testing each of the 13 mutations individually, they showed that in fact only 6 point mutations are required for a $10^5$-fold increase in activity toward a new substrate. Interestingly, their results also showed that all 6 mutations are beneficial on their own—raising the possibility that each position in a protein can be randomized independently.

### 3.2. Liposome-Based Assays

Natural selection, the driving force of Darwinian evolution, happens within the compartmentalization of cells. Tawfik and Griffiths[91] sought to mimic cellular compartmentalization in a simpler in vitro system of water-in-oil emulsions. By adding an in vitro transcription/translation reaction mixture to a stirred suspension of mineral oil that contained surfactants, an emulsion with mean droplet diameter similar to that of bacterial cells was obtained. DNA could then be transcribed and translated within the aqueous compartments of these emulsions. The conditions were controlled so that each water-in-oil droplet contained only a single gene on average. A mixture of two genes, the *M. HaeIII* gene (which encodes DNA methyltransferase HaeIII) and the *folA* gene (which encodes DHFR), were transcribed and translated in the aqueous compartments. Then the emulsion was broken, and the DNA in the aqueous phase was subjected to cleavage by HaeIII endonulease followed by amplification by PCR. Because the DNA methyltransferase methylated HaeIII-cleavage sites in droplets that contained the *M. HaeIII* gene, only the *M. HaeIII* gene would survive the cleavage by HaeIII and be amplified by PCR. After one round of selection, a mixture of *M. HaeIII* to *folA* gene with a ratio of 1:1000 was enriched to a ratio of 1:1 (Figure 15).[91] A similar approach to select Ni-binding proteins was also reported.[92]

Since all of the steps in in vitro compartmentalization are, indeed, carried out in vitro, this method has the potential to reach the limit at which the DNA is the limiting reagent in library size. In the initial paper by Tawfik and Griffiths, 50 μL of an in vitro transcription/translation reaction mixture was added to 950 μL of mineral oil that contained surfactants. The mean diameter of the droplets formed was 2.6 μm. Therefore, there were approximately $10^{10}$ droplets per mL of emulsion, which means that $10^{10}$ unique proteins could be tested on this scale. In vitro compartmentalization, as opposed to methods like RNA display, provides a suitable format for the detection

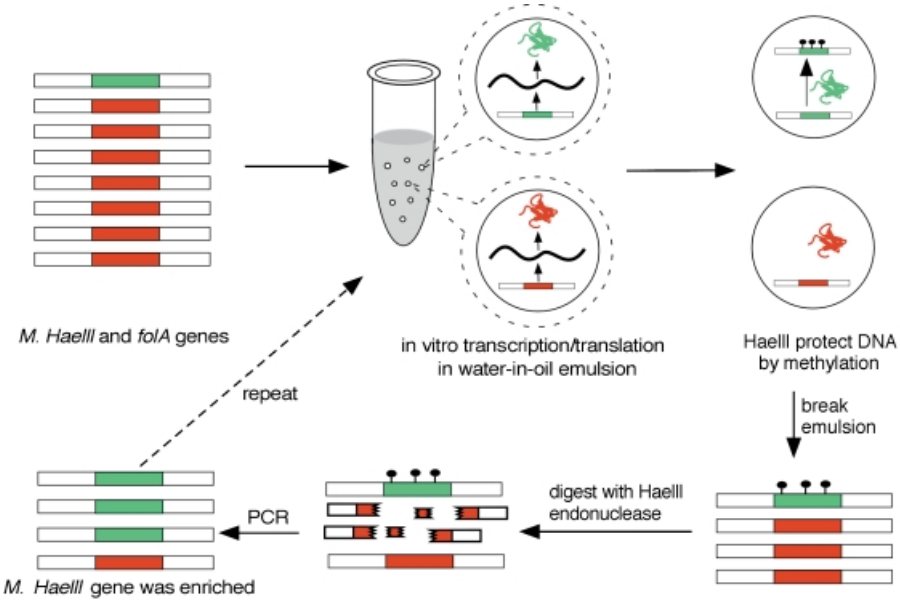



Figure 15. In vitro compartmentalization.[91] A mixture of two genes, *M. HaeIII* and *folA* was transcribed and translated in the aqueous compartments of water-in-oil emulsions. The *M. HaeIII* gene product, the DNA methyltransferase HaeIII, methylated the DNA and prevented cleavage by HaeIII endonulease. After disrupting the emulsion and cleavage by HaeIII endonulease, only M. HaeIII survived and was amplified by PCR. Therefore, the M. HaeIII gene was selectively enriched.

of catalytic reactions with multiple turnovers. Thus far, the assays have been limited to DNA modification or amplification enzymes.[93] The challenge will be to extend this method to chemical reactions that don't involve DNA.

### 3.3. Summary

Genetic complementation and in vitro compartmentalization seem to provide the most powerful assays for enzyme catalysis. The number of protein variants that can be tested with genetic assays, as with phage display, are limited by DNA transformation efficiencies, whereas in vitro compartmentalization, like RNA display, is limited only by the DNA. Again, $10^6$ – $10^8$-membered libraries seem to be sufficient to engineer enzymes with new specificities. There are several impressive applications for the modification of enzyme specificities with in vivo complementation. The advantage of carrying out the assays in vivo, particularly for cDNA libraries, is that the protein is expressed in its native environment. While initially applied to the complementation of natural enzyme function, one-, two-, and three-hybrid assays have extended these powerful assays to protein – DNA, protein – protein, protein – RNA, and protein – small molecule interactions. The next challenge is to modify these assays to detect reactions that are not natural to the cell. A recent twist is in vitro compartmentalization. To be general, this method will have to be adapted to detect functions other than DNA chemistry. Genetic assays are technically straightforward because they build from advances in molecular biology. The DNA from positive cells can simply be extracted and carried on to further rounds of mutagenesis or sequencing. The major challenge with genetic assays is discriminating false positive and negative hits. Here the key seems to be a powerful secondary screen. In vitro compartmentalization is similar technically to RNA display. The question is how robust the emulsion technology will prove.

## 4. Spatially Addressable Methods

Spatially addressable methods link the identity of a peptide or protein to a unique address in space such as a well in a microtiter plate or a patch on a solid support, so that any protein hits can be traced back to their DNA sequence.[94–97] Microtiter or "96-well" plates have long been used for this purpose. More recently, progress has been made in derivatizing a solid support with proteins to give a "protein chip".

### 4.1. Microtiter-Plate Assay

The use of 96-well plates in biochemical research dates back to the 1950s.[98] Microtiter plates have been routinely used in the enzyme-linked immunosorbent assay (ELISA) since this assay was first introduced in the 1970s.[99] ELISA is a sensitive method that detects antibody binding of an antigen, the substance used to immunize mice, by detecting the catalytic activity of an enzyme fused to the antibody. The antigen is adsorbed to the wells of a 96-well plate and then incubated with the enzyme-linked antibody. After washing away the unbound antibody, bound antibody is detected by monitoring the chromogenic reaction catalyzed by the enzyme. The screening of peptide or protein libraries based on a biochemical function in a microtiter plate is very similar to carrying out an ELISA. Peptide or protein solutions can be dispensed into the wells of a microtiter plate, and suitable biochemical assays can then be carried out in each well of the plate.

#### *4.1.1. Technical Considerations*

In microtiter-plate assays, thousands of protein samples, either cell cultures, crude cell extracts or purified proteins, must first be prepared and then transferred to the thousands of wells of the microtiter plates. Then biochemical assays can be carried out in each well of the plate to identify proteins with the desired function. Therefore one of the major challenges is the rapid and cost-effective purification of thousands of proteins. If the desired activity can be assayed using crude cell extracts, this collapses to an engineering problem. The range of activities that could be tested, however, would be greatly increased if methods could be developed for purifying proteins en masse. Martzen et al. solved this problem by purifying the proteins as GST-fusion proteins. Specifically, they purified all 6000 yeast open-reading frames (ORFs) expressed in *S. cerevisiae* by using a GST tag and screened these ORF libraries for several enzymatic activities.[100] The trick was how to construct the GST fusion library. The ORFs were amplified by a two-step PCR procedure so that all the ORFs could be amplified by a universal primer in the second PCR step. This approach allows the ORF library to be cloned readily into a new expression format in the future. In this example, the ORFs were placed under control of an inducible copper promoter. The resulting 6144 yeast strains, each of which contains a yeast ORF fused to GST, were arrayed into 64 96-well plates. First, these yeast strains were grown in 64 pools with each 96-well plate as a pool, and the GST – ORF fusion proteins were purified and assayed in these 64 pools. Then the positive pool was deconvoluted by doing the same assay with each row in the 96-well plate as a pool and each column in the 96-well plate as a pool to identify the GST – ORF that was responsible for the enzyme activity. The proteins were assayed for tRNA ligase, 2′-phosphotransferase, and cyclic phosphodiesterase activity by using thin-layer-chromatography assays with radioactive substrates. In this way, five enzymes were cloned, three of which were not previously known.

It is important to remember that while microtiter plates provide a tool to link the protein to its DNA sequence, they do not provide the screen or selection. The advantage, however, is that many traditional biochemical assays can be readily adapted to this format and automated.[101–103] Popular assays include scintillation proximity assays,[104] UV/Vis absorbance assays, chemiluminescent assays, and fluorescence assays.[105] Fluorescence assays are particularly powerful, and variations such as FRET and fluorescence polarization are quite general. The trend in microtiter plate assays is assay miniaturization, and high-density microtiter plates with 384 or 1536 wells are now standard. Fluorescence assay methods, because of their high sensitivity and analysis speed, are gaining more and more

attention in miniaturized high-throughput assays. For example, Eigen and co-workers[106] applied dual-color fluorescence cross-correlation spectroscopy (FCS) to detect catalytic turnovers by restriction endonucleases. They termed the method RAPID FCS (rapid assay processing by integration of dual-color FCS). A 66-bp oligonucleotide with endonuclease cleavage sites was labeled with the fluorescent dye Cy5 on one end and Rhodamine Green on the other end. Cross-correlation analysis was carried out to give distinguishable cleaved and uncleaved oligonucleotide samples. Analysis time of approximately 1 second was sufficient to give distinguishable cross-correlation values. The sample volume required for the measurement was only a few microliters. The short analysis time and small assay volume make RAPID FCS well suited for high-throughput screening.

### *4.1.2. Enzyme Catalysis*

Microtiter plate assays are particularly well suited to assays for enzyme catalysis because each reaction is kept spatially isolated. Hilvert and co-workers used a 96-well plate assay as a secondary screen in their efforts to engineer a catalytic antibody that catalyzes a rearrangement reaction. They screened thousands of monoclonal antibodies generated against a transition-state analogue and were able to isolate an antibody that catalyzed the decomposition of substituted benzisoxazoles into 2-cyanophenols with a rate acceleration of $10^8$.[107] First, monoclonal antibodies were generated against a transition-state analogue by using standard hybridoma technology. Then, about 1000 antibodies were transferred to 96-well plates and screened for the desired activity based on an increase in the absorption at 340 nm upon formation of the 2-cyanophenol product (Figure 16). Two of these antibodies

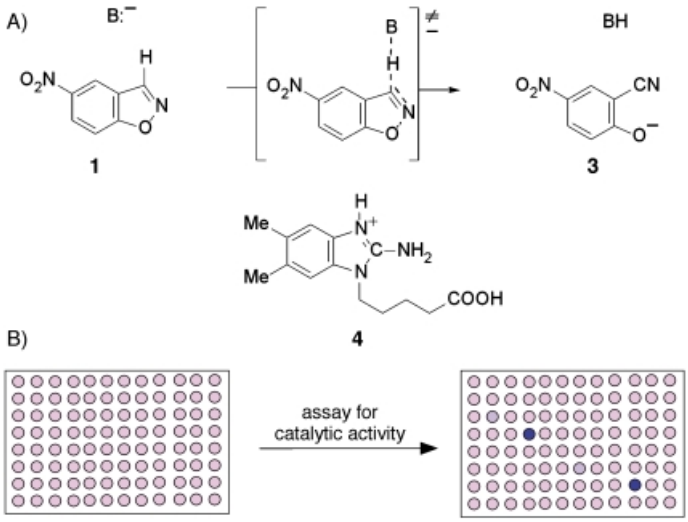



Figure 16. High efficiency catalytic antibodies generated by using transition state analogue and microtiter plate screening.[107] A) The structure of the transition state analogue **4** used to generate the catalytic antibodies that catalyze the transformation from **1** to **3**. B) The antibodies generated were transferred to 96-well plates and screened for catalytic activity by monitoring the absorption change at 340 nm upon conversion of **1** into **3**. Two highly active antibodies were isolated. Figure 16A and Table 4 reprinted from reference [107] with permission from Nature, Copyright (1995), Macmillan Magazines Limited.

were chosen for further analysis and showed rate accelerations of $3.4 \times 10^8$ and $3.5 \times 10^7$ compared with the non-catalyzed reaction (Table 4). Typically, only 10–100 monoclonal antibodies are screened for the desired activity, and each antibody is analyzed in a separate kinetic assay. The $10^8$-fold rate acceleration obtained in this example is much greater than that normally achieved for catalytic antibodies and demonstrates the benefit of such a large-scale secondary screen and of assaying directly for catalytic activity.

Table 4. Kinetic parameters of the catalytic antibodies obtained by Hilvert and co-workers.[107]

| Catalyst | $k_{cat}$ [$s^{-1}$] | $K_m$ [μM] | $k_{cat}/K_m$ [$s^{-1}M^{-1}$] | $\frac{k_{cat}/K_m}{k_{AcO^-}}$ | p*K*a | $k_{cat}/k_{AcO^-}$ [M] |
|---|---|---|---|---|---|---|
| 34E4 | 0.66 | 120 | $5.5 \times 10^3$ | $3.4 \times 10^8$ | 6.0 | 41000 |
| 35F10 | 0.35 | 630 | $5.6 \times 10^2$ | $3.5 \times 10^7$ | 5.5 | 22000 |

Microtiter plate assays have been used, not only for antibody engineering, but also for protein engineering generally. For example, Giver et al. screened a mutagenized *Bacillus subtilis p*-nitrobenzyl esterase (pNBE) library to evolve thermostable pNBEs and to elucidate the relationship between thermostability and catalytic activity.[108] The protein library shown in Figure 17 was generated by mutagenic PCR and DNA shuffling and then expressed in *E. coli*. Between

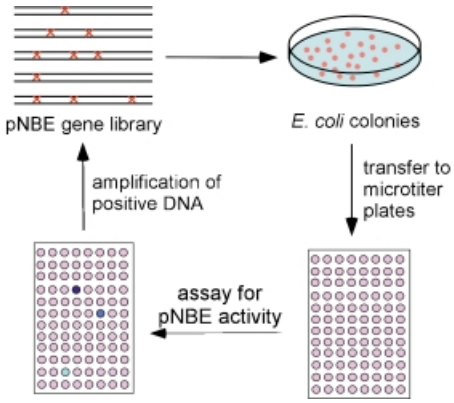



Figure 17. The evolution of thermostable pNBE.[108] The protein library, generated from the wild-type pNBE gene by mutagenic PCR and DNA shuffling, was expressed in *E. coli*. Single colonies were picked and transferred to 96-well plates in which the activities and thermostabilities of the proteins were assayed by using a standard *p*-nitrobenzyl acetate assay. The DNA from the clones that showed increased thermostability were isolated and subjected to further rounds of mutagenesis and screening. After six rounds, the $T_m$ of the enzyme was increased by 14 °C, whereas its activity was unchanged at room temperature.

500 and 2000 colonies in each generation were picked and transferred to 96-well plates. The pNBE activities and thermostabilities of the proteins were measured in a standard *p*-nitrobenzyl acetate assay, which monitors enzymatic turnover based on the formation of a chromogenic product by using crude cell extracts. Six rounds of mutagenesis and screening stabilized pNBE significantly, with a $T_m$ increase of 14 °C. Interestingly, the activity of the mutant enzyme was unchanged at room temperature. This application was successful, probably because a tractable problem—increasing the

thermostability of a known enzyme—that could be addressed with a $10^4$-membered library was chosen and because a convenient cell assay was available for the esterase activity.

Spatially addressable libraries seem particularly well suited to screening cDNA libraries in which fewer proteins need to be tested than when evolving a protein with new activity. Haushalter et al. used in vitro expression cloning to identify new DNA glycosylases from a *Xenopus* embryo cDNA library.[109] The cDNA library was divided into 120 pools, each containing roughly 100 cDNAs. These cDNA pools were transcribed and translated in vitro in the wells of microtiter plates. The translated proteins were incubated with $^{32}$P-labeled mechanism-based glycosylase inhibitors and then subjected to native polyacrylamide gel electrophoresis. The pool that showed distinct inhibitor binding was progressively subdivided and retested until a single active clone was identified. This clone was characterized and shown to be a new single-strand-selective monofunctional uracil-DNA glycosylase.

## 4.2. Protein Chips

Recently, inspired by successes with DNA microchips, effort has turned to developing protein microchips; in this case proteins are immobilized on a solid support with a high density to study protein function on a genomic scale. For example, proteins have been immobilized on a polyvinylidene difluoride membrane,[110] a nitrocellulose membrane,[111] glass slides,[112, 113] and in a polyacrylamide gel.[114] Macbeath and Schreiber[112] covalently attached several proteins to a glass slide in a high density format by using a high-precision contact-printing robot. The glass slides were first derivatized with an aldehyde-containing silane reagent. Then nanoliter volumes of protein samples were printed to the slides. The aldehyde groups on the glass slides reacted with primary amines on the proteins to form a Schiff base link, thus attaching the proteins covalently to the slides. The density of the protein chips thus made is about 1600 spots per square centimeter. The utility of these protein chips was also demonstrated by screening for known protein–protein and protein–small molecule interactions and identifying protein substrates of protein kinases.

### *4.2.1. Technical Considerations*

In contrast to microtiter plate assays in which crude cell lysates can be used, in protein microchip assays the proteins must be purified and then printed onto the chips. Therefore, the purification of thousands of proteins is a key limiting step for this method. A higher throughput is found with protein chips than with microtiter plate assays, and only nanoliters of purified protein solution are needed for each spot on the chip. Another advantage of protein chips over microtiter plates is that wash steps do not disrupt the protein location. On the other hand, as shown in these proof of principle studies,[111, 114, 115] immobilized protein libraries are more suitable to binding assays than catalytic assays. A clever solution that combines the advantages of the two methods is to immobilize proteins in nanowells.[116] A particularly interesting recent approach taken by Ziauddin and Sabatini is to immobilize cDNAs on a solid support and then transfect cells en masse.[117]

### *4.2.2. Applications*

Most reports to date have focused on establishing the methodology for attaching a few proteins to solid supports. An impressive recent advance is the printing and application of a chip displaying the complete yeast proteome. Zhu et al.[118] cloned 5800 ORFs of yeast into a yeast high-copy expression vector as GST–His6 fusion proteins under control of the *GAL1* promoter. The proteins were purified by using glutathione–agarose beads in a 96-well plate format. These purified proteins were then printed to glass slides coated with either aldehydes or nickel. The aldehyde groups react with the primary amines on the proteins, or the nickel forms complexes with the His6 tag on the protein termini, therefore linking the proteins to the glass slide. Detection of the GST tag on the protein chip showed that more than 90% of the proteins were expressed, purified, and printed. Calmodulin-binding proteins were identified by incubating the protein chip with biotinylated calmodulin, followed by detection of bound biotinylated calmodulin with Cy3-conjugated streptavidin after washing away unbound biotinylated calmodulin (Figure 18). This experiment identified six known calmodulin-binding proteins and 33 unknown potential calmodulin-binding proteins. There were six known calmodulin-binding proteins that were not detected. Similarly, phospholipid-binding proteins were also detected. Importantly, by probing for the GST tag and camodulin-binding proteins, the authors establish that the chip successfully presents a high percentage of the cDNA library.

## 4.3. Summary

The advantage of spatially addressable methods seems to be the ease with which they can be adapted to traditional biochemical assays for protein function. Particularly in applications in which the proteins must be purified, spatially addressable libraries realistically are limited to approximately $10^4$ protein variants. Although there are applications in protein engineering, spatially addressable methods seem best suited to testing cDNA libraries to determine protein function. As with any in vitro method, these assays are limited to proteins that are stable outside the cell. In addition, it is not yet clear how the solid support may influence protein function. A wide variety of chemical reactions can be detected with these methods since many traditional assays can be used. Proteins printed in microwells are compatible with both assays for binding and for enzyme catalysis. The drawback of spatially addressable methods is that they are technically demanding. Purifying and printing proteins is labor-intensive. It is reasonable to assume that protein chips will not be as robust for multiple assays as DNA chips. Also at the end of the assay each protein must be traced back to its unique DNA tag.

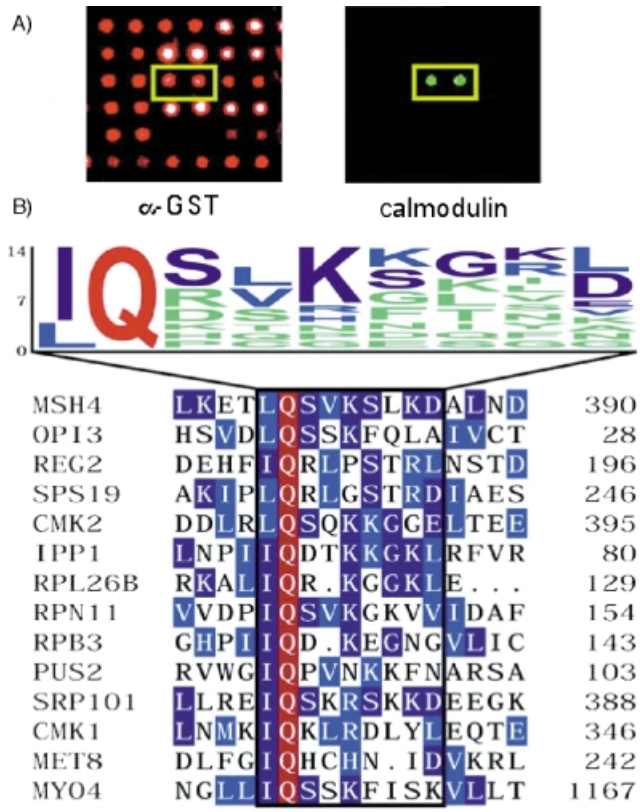



Figure 18. Assay for camodulin-binding proteins on proteome chips.[118] A) On the left, a control proteome chip probed with anti-GST (red) is shown. This control experiment demonstrated that more than 90% of the proteins were successfully expressed, purified, and spotted on the glass slides. The second panel shows positive signals in duplicate (green) detected by a biotinylated calmodulin probe. B) Positive proteins (39) were identified and 14 of them share a motif whose consensus is (I/L)QXK(K/X)GB,. X is any residue and B is a basic residue. The size of the letter indicates the relative frequency of the amino acid indicated. Reprinted with permission from reference [118]. Copyright (2001), American Association for the Advancement of Science.

## 5. Direct Protein-Sequencing Methods

Another possibility is that methods for sequencing proteins will advance to the point where a tag is unnecessary. In this section we look at recent advances in mass spectrometry that begin to achieve this goal.

### 5.1. Technical Considerations

Mass spectrometry (MS) is emerging as a powerful tool for proteomics. The driving force comes from advances in both MS and genome sequencing. Advances in MS include development of novel ionization techniques such ESI and MALDI.[119–121] MS can be used to determine protein sequence because the protein can be fragmented into small peptides whose masses or sequences can be obtained by MS, and then the protein sequence can be deduced from the masses or sequences of the fragmented peptides with the aid of the genome sequence. Therefore, a known genome sequence is a prerequisite for these MS methods. Henzel et al. first introduced this approach in the late 1980s.[122] They purified *E. coli* protein extracts by 2D gel electrophoresis. The proteins were then electroblotted onto poly(vinylidene difluoride) (PVDF) membrane and stained with Coomassie Blue. Individual protein spots were excised, reduced in situ with dithiothreitol, alkylated with iodoacetic acid, and digested with trypsin. Peptide fragments thus generated were then subjected to MALDI analysis. The masses of the peptide fragments were analyzed by using the computer program FRAGFIT, which searched for peptide sequences that matched the obtained masses from a protein sequence database. In this way, the identity of the proteins separated by 2D gel electrophoresis can be inferred. Of the 10 protein spots chosen for analysis, the identity of 9 spots was confirmed by N-terminal sequence analysis. Methods that improve the sensitivity and certainty of identification have been reported.[123]

### 5.2. Applications

MS methods can be used to identify the proteins that make up a protein complex, that is, proteins that associate with each other. An advantage to this approach is that the entire complex rather than one protein–protein interaction at a time, can be assigned in a single step. Link et al. identified 75 of 78 predicted yeast ribosomal proteins in the 80S ribosome complex by using MS.[124] The protein complex was purified, denatured, reduced, and digested to peptide fragments. These peptide fragments were then separated by 2D column chromatography and eluted into a mass spectrometer. Peptide sequences were obtained by using tandem mass spectrometry (MS/MS) and used to search translated genomic databases. By using 120 μg of the purified protein complex, 95 polypeptides were identified, 90 of which were from distinct ribosomal genes and represent 75 of the 78 proteins (some proteins are encoded by two separate genes with >98% amino acid sequence identity). A protein that had not been previously identified as part of the yeast ribosome complex, YMR116C (BEL1), was also identified.

Quantitative MS methods can be used to measure protein expression levels at different cellular states. This approach is similar to DNA microarrays, except that the actual protein levels, as opposed to mRNA levels, can be monitored. Global profiling of gene expression levels at different cellular states can help to identify proteins and genes involved in a given biological pathway. Gygi et al. developed a method that used isotope-coded affinity tags (ICAT) to measure relative protein expression levels at different cellular states.[125] As shown in Figure 19, the ICAT consists of a biotin molecule, a linker that can incorporate isotopes, and a thiol-specific reactive group. Soluble proteins from yeast cells grown in either 2% galactose or 2% ethanol as the carbon source were isolated and reduced. Two ICAT reagents, one with a heavy isotope and one with a light isotope, were used to label the two protein samples, respectively. Then the two protein samples were mixed and digested. Cysteine-containing peptides were isolated by avidin-affinity chromatography and analyzed by using microcapillary liquid chromatography–MS/MS. Protein sequences were obtained by searching the sequence database using the identified peptide sequences, and

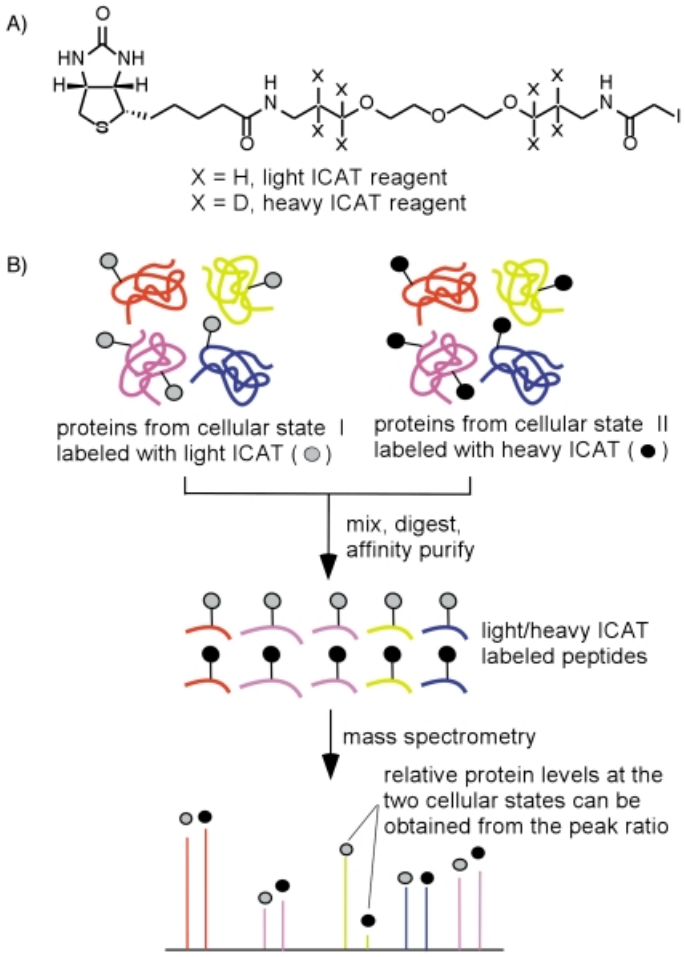



Figure 19. A mass spectrometry method to measure relative protein expression levels in different cellular states.[125] A) The structure of the isotope-coded affinity tags (ICAT), which consist of a biotin molecule, a linker that can incorporate isotopes, and a thiol-specific reactive group. B) Soluble proteins from two different cellular states are isolated and reduced. Two ICAT reagents, one with heavy isotope, one with light isotope are used to label the two protein samples respectively. Then the two protein samples are mixed and digested. Cysteine-containing peptides, which are labeled with the ICAT reagents, are isolated by avidin affinity chromatography. After separation by microcapillary liquid chromatography, these labeled peptides are analyzed by MS. The relative protein expression levels at the two cellular states can be obtained by measuring the ratio of the two peaks for the same peptide sequence from the two different cellular states. The identity of the protein can be obtained by first obtaining the peptide sequence via tandem MS and then searching the peptide sequence in the database.

the relative protein expression levels of the two different cellular states were obtained by comparing the intensity of the pairs of peptide ions of identical sequences that were labeled with isotopically differentiated ICAT reagents. The expression levels of 34 genes for yeast grown in either 2% galactose or 2% ethanol as the carbon source were compared and the results correlate well with what is known about yeast metabolism.

As with phage display and ribosome display, the challenge with MS methods is how to assay enzyme catalysis as well as binding. Cravatt and co-workers recently used an inhibitor of serine hydrolases in a method they termed "activity-based protein profiling" to detect serine hydrolyases from crude tissue extracts.[89] In their specific case, they used a biotinylated fluorophosphonate (FP–biotin). Fluorophosphonate derivatives are selective and irreversible inhibitors for serine hydrolases, and they only react with enzymes that are in a catalytically active state. Crude rat brain extracts were incubated with FP–biotin, after which the proteins were subjected to standard SDS/PAGE-Western blotting procedures and detected by using an avidin–horseradish peroxidase conjugate. More than ten proteins were detected, two of them, a 75-kDa protein and an 85-kDa protein, were subjected to MS analysis. Genome database searching identified the 75-kDa protein as the rat orthologue of a human protein sequence KIAA0436 and the 85-kDa protein as acylpeptide hydrolase. Chemical probes for cysteine proteases have also been described.[90]

### 5.3. Summary

MS methods are more recent and much may change technically in the next several years. In theory, MS assays are limited only by the mass of the proteins, thus allowing libraries even larger than those obtainable with RNA display. However, only relatively small libraries have been tested to date. MS methods place no inherent limits on the type of proteins that can be tested and are compatible with both in vitro and in vivo assays. It is not yet clear what type of assays will be applied to MS methods, and there is much opportunity here. Likewise, many technical issues are yet to be resolved, such as how to detect proteins with very low cellular concentrations or how to regenerate the DNA that encodes a protein hit for further rounds of mutagenesis. In addition to MS, there may be other methods for the direct sequencing of proteins that are being overlooked.

## 6. Summary and Outlook

The development of robust methods for assaying proteins based on function over the past two decades seems to promise a breakthrough in protein engineering and proteomics in the near term. Several methods for tying a protein to its DNA sequence are now well developed, ranging from methods such as phage display (in which a physical link is created between the protein and its DNA sequence) to microtiter plate assays (in which each protein is simply assayed independently) to cell-based assays (in which each cell contains a single plasmid that encodes a unique protein variant). While these methods are already impressive, recent advances in protein chips and MS emphasize that new methods can still be advantageous. Researchers are building on these methods and are beginning to modify the specificity of existing proteins successfully. It should be noted that these successes are partly a result of the correct choice of reactions that provide a convenient selection or screen, for example, a reaction in which the product is fluorescent or an essential cellular metabolite. The next challenge is to develop general assays for enzymatic activity that allow us to engineer new chemistry and to apply these methods not only to protein engineering, but also to proteomics.

*We would like to thank S. Licht, C. Noren, T. Yano, and the members of our lab for their helpful comments as we were preparing this manuscript. V.W.C. is a recipient of a Beckman Young Investigator Award, a Burroughs-Wellcome Fund New Investigator Award in the Toxological Sciences, a Camille and Henry Dreyfus New Faculty Award, and a National Science Foundation Career Award. H.L. is a Arun Guthikonda Memorial Fellow.*

Received: June 6, 2001 [A 478]

[1] G. P. Smith, V. A. Petrenko, *Chem. Rev.* **1997**, *97*, 391.
[2] S. S. Sidhu, *Curr. Opin. Biotechnol.* **2000**, *11*, 610.
[3] D. J. Rodi, L. Makowski, *Curr. Opin. Biotechnol.* **1999**, *10*, 87.
[4] P. Forrer, S. Jung, A. Plückthun, *Curr. Opin. Struct. Biol.* **1999**, *9*, 514.
[5] G. P. Smith, *Science* **1985**, *228*, 1315.
[6] D. Sblattero, A. Bradbury, *Nat. Biotechnol.* **2000**, *18*, 75.
[7] S. S. Sidhu, H. B. Lowman, B. C. Cunningham, J. A. Wells, *Methods Enzymol.* **2000**, *328*, 333.
[8] S. Bass, R. Green, J. A. Wells, *Proteins: Struct. Funct. Genet.* **1990**, *8*, 309.
[9] J. Greenwood, A. E. Willis, R. N. Perham, *J. Mol. Biol.* **1991**, *220*, 821.
[10] C. Gao, S. Mao, C. L. Lo, P. Wirsching, R. A. Lerner, K. D. Janda, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6025.
[11] M. Duenas, C. A. Borrebaeck, *Bio/Technology* **1994**, *12*, 999.
[12] K. Gramatikoff, O. Georgiev, W. Schaffner, *Nucleic Acids Res.* **1994**, *22*, 5761.
[13] J. K. Scott, G. P. Smith, *Science* **1990**, *249*, 386.
[14] N. C. Wrighton, F. X. Farrell, R. Chang, A. K. Kashyap, F. P. Barbone, L. S. Mulcahy, D. L. Johnson, R. W. Barrett, L. K. Jolliffe, W. J. Dower, *Science* **1996**, *273*, 458.
[15] R. Hyde-DeRuyscher, L. A. Paige, D. J. Christensen, N. Hyde-DeRuyscher, A. Lim, Z. L. Fredericks, J. Kranz, P. Gallant, J. Zhang, S. M. Rocklage, D. M. Fowlkes, P. A. Wendler, P. T. Hamilton, *Chem. Biol.* **2000**, *7*, 17.
[16] S. R. Whaley, D. S. English, E. Hu, P. F. Barbara, A. M. Belcher, *Nature* **2000**, *405*, 665.
[17] W. Dall′Acqua, P. Carter, *Curr. Opin. Struct. Biol.* **1998**, *8*, 443.
[18] H. R. Hoogenboom, P. Chames, *Immunol. Today* **2000**, *21*, 371.
[19] J. McCafferty, A. D. Griffiths, G. Winter, D. J. Chiswell, *Nature* **1990**, *348*, 552.
[20] A. S. Kang, C. F. Barbas III, K. D. Janda, S. J. Benkovic, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4363.
[21] C. F. Barbas, III, A. S. Kang, R. A. Lerner, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7978.
[22] J. Yu, S. Y. Choi, K.-D. Moon, H.-H. Chung, H. J. Youn, S. Jeong, H. Park, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2880.
[23] M. Baca, T. S. Scanlan, R. C. Stephenson, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10063.
[24] K. D. Janda, L.-C. Lo, C.-H. L. Lo, M.-M. Sim, R. Wang, C.-H. Wong, R. A. Lerner, *Science* **1997**, *275*, 945.
[25] M. Shanmugavelu, A. R. Baytan, J. D. Chesnut, B. C. Bonning, *J. Biol. Chem.* **2000**, *275*, 1802.
[26] Y. Choo, A. Klug, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11163.
[27] A. C. Jamieson, S.-H. Kim, J. A. Wells, *Biochemistry* **1994**, *33*, 5689.
[28] E. J. Rebar, C. O. Pabo, *Science* **1994**, *263*, 671.
[29] G. A. Weiss, H. B. Lowman, *Chem. Biol.* **2000**, *7*, 547.
[30] G. Beste, F. S. Schmidt, T. Stibora, A. Skerra, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1898.
[31] S. Schlehuber, G. Beste, A. Skerra, *J. Mol. Biol.* **2000**, *297*, 1105.
[32] J. C. Guo, W. Huang, T. S. Scanlan, *J. Am. Chem. Soc.* **1994**, *116*, 6062.
[33] P. Soumillion, L. Jespers, M. Bouchet, J. Marchand-Brynaert, G. Winter, J. Fastrez, *J. Mol. Biol.* **1994**, *237*, 415.
[34] C. Gao, C.-H. Lin, C.-H. L. Lo, S. Mao, P. Wirsching, R. A. Lerner, K. D. Janda, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11777.
[35] H. Pedersen, S. Holder, D. P. Sutherlin, U. Schwitter, D. S. King, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10523.
[36] S. Demartis, A. Huber, F. Viti, L. Lozzi, L. Giovannoni, P. Neri, G. Winter, D. Neri, *J. Mol. Biol.* **1999**, *286*, 617.
[37] J.-L. Jestin, P. Kristensen, G. Winter, *Angew. Chem.* **1999**, 111, 1196; Angew. *Chem. Int. Ed.* **1999**, *38*, 1124.
[38] S. Atwell, J. A. Wells, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9497.
[39] I. Ponsard, M. Galleni, P. Soumillion, J. Fastrez, *ChemBioChem* **2001**, *2*, 253.
[40] R. W. Roberts, *Curr. Opin. Chem. Biol.* **1999**, *3*, 268.
[41] L. C. Mattheakis, R. R. Bhatt, W. J. Dower, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 9022.
[42] J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4937.
[43] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297.
[44] N. Nemoto, E. Miyamoto-Sato, Y. Husimi, H. Yanagawa, *FEBS Lett.* **1997**, *414*, 405.
[45] G. Cho, A. D. Keefe, R. Liu, D. S. Wilson, J. W. Szostak, *J. Mol. Biol.* **2000**, *297*, 309.
[46] R. Liu, J. E. Barrick, J. W. Szostak, R. W. Roberts, *Methods Enzymol.* **2000**, *318*, 268.
[47] J. Hanes, C. Schaffitzel, A. Knappik, A. Plükthun, *Nat. Biotechnol.* **2000**, *18*, 1287.
[48] D. S. Wilson, A. D. Keefe, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3750.
[49] A. D. Keefe, J. W. Szostak, *Nature* **2001**, *410*, 715.
[50] M. G. Cull, J. F. Miller, P. J. Schatz, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1865.
[51] C. M. Gates, W. P. C. Stemmer, R. Kaptein, P. J. Schatz, *J. Mol. Biol.* **1996**, *255*, 373.
[52] S. E. Cwirla, P. Balasubramanian, D. J. Duffin, C. R. Wagstrom, C. M. Gates, S. C. Singer, A. M. Davis, R. L. Tansik, L. C. Mattheakis, C. M. Boytos, P. J. Schatz, D. P. Baccanari, N. C. Wrighton, R. W. Barrett, W. J. Dower, *Science* **1997**, *276*, 1696.
[53] S. Ståhl, M. Uhlén, *Trends Biotechnol.* **1997**, *15*, 185.
[54] J. A. Francisco, C. F. Earhart, G. Georgiou, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2713.
[55] E. T. Boder, K. D. Wittrup, *Nat. Biotechnol.* **1997**, *15*, 553.
[56] W. Ernst, R. Grabherr, D. Wegner, N. Borth, A. Grassauer, H. Katinger, *Nucleic Acids Res.* **1998**, *26*, 1718.
[57] A. Christmann, K. Walter, A. Wentzel, R. Kratzner, H. Kolmar, *Protein Eng.* **1999**, *12*, 797.
[58] P. S. Daugherty, G. Chen, M. J. Olsen, B. L. Iverson, G. Georgiou, *Protein Eng.* **1998**, *11*, 825.
[59] M. C. Kieke, E. V. Shusta, E. T. Boder, L. Teyton, K. D. Wittrup, D. M. Kranz, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5651.
[60] E. T. Boder, K. S. Midelfort, K. D. Wittrup, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10701.
[61] M. J. Olsen, D. Stephens, D. Griffiths, P. Daugherty, G. Georgiou, B. L. Iverson, *Nat. Biotechnol.* **2000**, *18*, 1071.
[62] G. W. Beadle, E. L. Tatum, *Proc. Natl. Acad. Sci. USA* **1941**, *27*, 499.
[63] M. Hensel, J. E. Shea, C. Gleeson, M. D. Jones, E. Dalton, D. W. Holden, *Science* **1995**, *269*, 400.
[64] S. Benzer, *Sci. Am.* **1962**, 72.
[65] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg, *Nature* **2000**, *403*, 623.
[66] B. L. Drees, *Curr. Opin. Chem. Biol.* **1999**, *3*, 64.
[67] D. S. McNabb, L. Guarente, *Curr. Opin. Biotechnol.* **1996**, *7*, 554.
[68] E. Warbrick, *Structure* **1997**, *5*, 13.
[69] H. Lin, V. W. Cornish, *Angew. Chem.* **2001**, *113*, 895; *Angew. Chem. Int. Ed.* **2001**, *40*, 871.
[70] R. K. Brachmann, J. D. Boeke, *Curr. Opin. Biotechnol.* **1997**, *8*, 561.
[71] M. Vidal, P. Legrain, *Nucleic Acids Res.* **1999**, *27*, 919.
[72] P. Colas, R. Brent, *Trends Biotechnol.* **1998**, *16*, 355.
[73] S. Fields, O. Song, *Nature* **1989**, *340*, 245.
[74] P. Colas, B. Cohen, T. Jessen, I. Grishina, J. McCoy, R. Brent, *Nature* **1996**, *380*, 548.
[75] T. C. Norman, D. L. Smith, P. K. Sorger, B. L. Drees, S. M. O′Rourke, T. R. Hughes, C. J. Roberts, S. H. Friend, S. Fields, A. W. Murray, *Science* **1999**, *285*, 591.
[76] P. Youderian, A. Vershon, S. Bouvier, R. T. Sauer, M. M. Susskind, *Cell* **1983**, *35*, 777.
[77] S. L. Dove, J. K. Joung, A. Hochschild, *Nature* **1997**, *386*, 627.

[78] J. K. Joung, E. I. Ramm, C. O. Pabo, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7382.
[79] H. A. Greisman, C. O. Pabo, *Science* **1997**, *275*, 657.
[80] D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **1993**, *262*, 1019.
[81] E. J. Licitra, J. O. Liu, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12817.
[82] E. C. Griffith, E. J. Licitra, J. O. Liu, *Methods Enzymol.* **2000**, *328*, 89.
[83] H. Lin, W. M. Abida, R. T. Sauer, V. W. Cornish, *J. Am. Chem. Soc.* **2000**, *122*, 4247.
[84] H. Joo, Z. Lin, F. H. Arnold, *Nature* **1999**, *399*, 670.
[85] B. Hall in *Evolutionary Biology*, *Vol. 15* (Eds.: M. Hecht, B. Wallace, G. Pierce ), Plenum Press, New York, **1982**, p. 85.
[86] J. D. Hermes, S. C. Blacklow, J. R. Knowles, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 696.
[87] T. Yano, S. Oue, H. Kagamiyama, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5511.
[88] G. Xia, L. Chen, T. Sera, M. Fa, P. G. Schultz, F. E. Romesberg, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6597.
[89] Y. Liu, M. P. Patricelli, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14694.
[90] D. Greenbaum, K. F. Medzihradszky, A. Burlingame, M. Bogyo, *Chem. Biol.* **2000**, *7*, 569.
[91] D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652.
[92] N. Doi, H. Yanagawa, *FEBS Lett.* **1999**, *457*, 227.
[93] F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4552.
[94] A. Q. Emili, G. Cagney, *Nat. Biotechnol.* **2000**, *18*, 393.
[95] G. Walter, K. Büssow, D. Cahill, A. Lueking, H. Lehrach, *Curr. Opin. Microbiol.* **2000**, *3*, 298.
[96] E. J. Grayhack, E. M. Phizicky, *Curr. Opin. Chem. Biol.* **2001**, *5*, 34.
[97] H. Zhu, M. Snyder, *Curr. Opin. Chem. Biol.* **2001**, *5*, 40.
[98] R. Manns, http://www.messebasel.ch/miptec/mission-history.htm, **1999**.
[99] A. Voller, D. Bidwell, G. Huldt, E. Engvall, *Bull. WHO* **1974**, *51*, 209.
[100] M. R. Martzen, S. M. McCraith, S. L. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, E. M. Phizicky, *Science* **1999**, *286*, 1153.
[101] S. A. Sundberg, *Curr. Opin. Biotechnol.* **2000**, *11*, 47.
[102] R. P. Hertzberg, A. J. Pope, *Curr. Opin. Chem. Biol.* **2000**, *4*, 445.
[103] J. E. González, P. A. Negulescu, *Curr. Opin. Biotechnol.* **1998**, *9*, 624.
[104] N. Bosworth, P. Towers, *Nature* **1989**, *341*, 167.
[105] A. J. Pope, U. Haupts, K. J. Moore, *Drug Discovery Today* **1999**, *4*, 350.
[106] A. Koltermann, U. Kettling, J. Bieschke, T. Winkler, M. Eigen, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1421.
[107] S. N. Thorn, R. G. Daniels, M. M. Auditor, D. Hilvert, *Nature* **1995**, *373*, 228.
[108] L. Giver, A. Gershenson, P.-O. Freskgard, F. H. Arnold, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12809.
[109] K. A. Haushalter, P. T. Stukenberg, M. W. Kirschner, G. L. Verdine, *Curr. Biol.* **1999**, *9*, 174.
[110] A. Lueking, M. Horn, H. Eickhoff, K. Büssow, H. Lehrach, G. Walter, *Anal. Biochem.* **1999**, *270*, 103.
[111] H. Ge, *Nucleic Acids Res.* **2000**, *28*, e3.
[112] G. MacBeath, S. L. Schreiber, *Science* **2000**, *289*, 1760.
[113] C. A. Rowe, L. M. Tender, M. J. Feldstein, J. P. Golden, S. B. Scruggs, B. D. MacCraith, J. J. Cras, F. S. Ligler, *Anal. Chem.* **1999**, *71*, 3846.
[114] P. Arenkov, A. Kukhtin, A. Gemmell, S. Voloshchuk, V. Chupeeva, A. Mirzabekov, *Anal. Biochem.* **2000**, *278*, 123.
[115] B. Schweitzer, S. Wiltshire, J. Lambert, S. O'Malley, K. Kukanskis, Z. Zhu, S. F. Kingsmore, P. M. Lizardi, D. C. Ward, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10113.
[116] H. Zhu, J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. Reed, M. Snyder, *Nat. Genet.* **2000**, *26*, 283.
[117] J. Ziauddin, D. M. Sabatini, *Nature* **2001**, *411*, 107.
[118] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, *Science* **2001**, *293*, 2101.
[119] J. S. Andersen, M. Mann, *FEBS Lett.* **2000**, *480*, 25.
[120] M. J. Chalmers, S. J. Gaskell, *Curr. Opin. Biotechnol.* **2000**, *11*, 384.
[121] J. Godovac-Zimmermann, L. R. Brown, *Mass Spectrom. Rev.* **2001**, *20*, 1.
[122] W. J. Henzel, T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley, C. Watanabe, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5011.
[123] A. Shevchenko, O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14440.
[124] A. J. Link, J. Eng, D. M. Schieltz, E. Carmack, G. J. Mize, D. R. Morris, B. M. Garvik, I. J R Yates, *Nat. Biotechnol.* **1999**, *17*, 676.
[125] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, R. Aebersold, *Nat. Biotechnol.* **1999**, *17*, 994.