35

Protein Engineering

Thomas Willemsen*, Urs B. Hagemann*, Eva M. Jouaux*, Sabine C. Stebel, Jody M. Mason, Kristian M. Müller, and Katja M. Arndt

1. Introduction to Protein Engineering

Relating primary sequence to three-dimensional structure has long been the holy grail of structural biology and appears to be far from achievement. Within grasp however, is the use of intuitive or unintuitive methodology to modify existing known protein structures to achieve the desired effect. We use *protein engineering* as a general term for the design of proteins with useful or valuable properties. The technique has become possible due to our increasing knowledge of detailed protein structures, which in turn highlights potential for improving key facets of protein structure; for example, the mutation of specific residues with a view to improving binding or catalysis. This *rational design* (Section 2.1) requires the scientist to have a detailed prior knowledge of the protein to attempt to make specific informed changes to the sequence to exert the desired effect. The technique is quite straightforward, involving mutation at the genetic level followed by expression and characterization. This site-directed mutagenesis approach is discussed in Section 2.1.1. However, rational mutations do not always generate the desired effect. This has invariably led to computer-based approaches for protein design. These are designed to save time in identifying mutations that generate the desired effect of low energy structures, and aim for lower the sequence conformation space that is required in the search. To simplify the procedure, these algorithms are based on approximations that require less processing time. Unfortunately, approximations can also lead to false positives which do not yield the predicted desired effect at the protein level. Computer aided protein engineering strategies are discussed in Section 2.1.2.

The second protein engineering approach, known as *directed evolution* relies on a selection system to pick from a range of variants. This involves the construction of protein libraries that contain a wealth of randomized positions. The generation of libraries is discussed extensively in the chapter "*Directed Protein Evolution*" in this book. Many of these residues will be intuitively predicted to have the desired result, while for other residues the outcome of the change may not be known. By screening these mutations at the protein

*TW, UBB and EMJ contributed equally.

From: *Molecular Biomethods Handbook, 2nd Edition.*
Edited by: J. M. Walker and R. Rapley © Humana Press, Totowa, NJ

level for their desired function, sequences conforming to the best molecule for the desired role can be screened. This has the advantage over site-directed mutagenesis or computer-based design that you obtain exactly what you select for. Theory of library-based design strategies is discussed in Section 2.2, and includes a discussion as well as published examples of the phage display (2.2.1.), ribosome display (2.2.2.), and yeast two-hybrid systems (2.2.3) that have been used to screen protein libraries. Also discussed are the advantages and pitfalls of working with any one of these techniques. Protein-fragment complementation assay (PCA) systems are discussed (2.2.4) along with several examples of the screening system in action, as well as methods of cell surface display (2.2.5). Finally, *in vitro* compartmentalization methods are discussed (2.2.6). The chapter closes (Section 3) with a range of examples for each of the techniques highlighted.

2. Methods

2.1. Rational Design Strategies

Rational design is one of the strategies for protein engineering in which a detailed knowledge of the structure and function of the protein is used to predict beneficial changes. These changes are introduced into the protein by site-directed mutagenesis techniques. As an increasing number of high-resolution protein structures is available the creation and application of computational methods to identify amino acid sequences that have low energies for the target structure is used more and more. These two principles can complement each other or be used alone. The major drawback is the need of detailed structural knowledge of a protein, and depending on the design, it can be extremely difficult to predict the effects of various mutations, especially long-range effects.

A variety of strategies have emerged for modulating protein properties, such as stability, specificity, solubility, conformational state, binding affinity, oligomerization state, substrate selectivity for enzymes, protease susceptibility, immunogenicity, and pharmacokinetics (the last three for therapeutical approaches) (see **Fig. 35.1**). Mechanisms for altering these properties include manipulation of the primary structure, incorporation of chemical and post-translational modifications and utilization of fusion-partners ***(1)***. There are many rational strategies to change protein characteristics. One simple stabilization strategy is to replace free cysteines, thereby preventing the formation of unwanted intermolecular and intramolecular disulphide bonds. Substituting exposed nonpolar residues with polar residues can enable soluble expression and improve the solubility of the protein. Alteration of the net charge and isoelectric point (pl) of a protein can also affect its solubility. In some cases, increasing the binding affinity for a target protein can produce an increase in biological activity. In other cases, it is possible to reduce undesired biological activities by decreasing the affinity for nontarget molecules. Many proteins undergo conformational changes that are central to their function. In such cases, the conformational equilibrium can be driven towards the desired state.

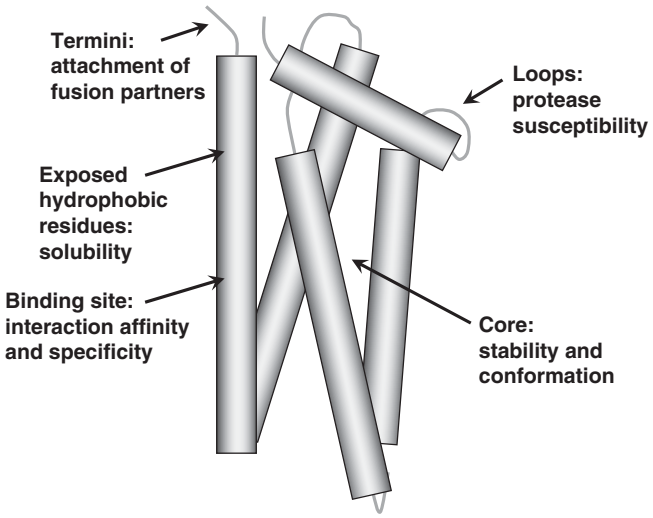

Fig. 35.1. Different strategies for rational protein design

The starting point for rational design is the development of a molecular model, based on the protein structure and function, often in combination with an algorithm. This is followed by experimental construction and analysis of the properties of the designed protein. If the experimental outcome is failure or partial success, then a next round of the design cycle is started ***(2)***. Sometimes new mutants based on initial information are developed which leads to a repetition of design steps until a variant is found that meets all the requirements. This iterative process, where theory and experiments alters, is often referred to as a "design cycle."

A possible design strategy procedure is described below, based on the availability of a 3-D structure and sequence of the protein ***(3)***.

1. Collect available information from the literature as well as from experimental analysis on the protein of interest, its homologues and other family members.
2. Find as many amino acid sequences as possible of homologues sequences and make a multiple sequence alignment. Take the secondary structure of the protein into account (see, e.g., www.expasy.org for databases and tools).
3. Compare the structures of the protein, homologues and family members by structural alignment to see, whether there is anything remarkable and whether all residues are in an optimal structural environment. Also check whether any of the homologues structures are more stable and if so, why. Examine if the structures possess additional interactions in the form of salt bridges, disulphide bridges, etc. Verify the difference in packing i.e., by looking for any cavities or steric clashes. Take variations in loop length or other conspicuous differences into account. Helpful programs are PyMol and Swiss PDB Viewer.
4. Try and apply the design concepts as described above. Also, apply programs that can predict mutations. You can find a variety of such tools on www.expasy.org/tools/.

5. Simulate the protein under folding conditions and identify the regions that appear to be the least stable. Try to design mutants that counteract the early unfolding processes.
6. Model the mutants you have and check whether the structure has improved and whether the mutation causes other problems, like less favorable torsion angles in the side chain or less optimal packing.
7. Produce the mutants experimentally and analyze their properties.

Nowadays, the first steps of rational design are more and more computer based but historically, site-directed mutagenesis was first. Here, we kept this chronological order.

2.1.1. Site-Directed Mutagenesis

Since the late 1980s protein molecules were altered by site-directed or site-specific mutagenesis of their genes ***(4–6)***. In this technique, a mutation is created at a defined site in the DNA leading to a change in the amino acid of the corresponding protein. This method requires the wild-type sequence to be known. The change itself is made by PCR methods where primers containing the desired mutation are used. In the first cycle, there is a priming mismatch for the primers binding the template DNA strand, but after the first cycle, the primer-based strand, containing the mutation, will be at about equal concentration to the original template. After successive cycles, its number will increase exponentially and outnumber the original, unmutated strand, resulting in a nearly homogeneous solution of mutated amplified fragments. For this PCR it is necessary to design primers that are suitable for the desired changes, considering also their annealing temperature.

Two techniques are commonly used to introduce specific amino acid replacements into a target gene. The first of these is termed the overlap extension method (**Fig. 35.2**). In this method, four primers are used in the

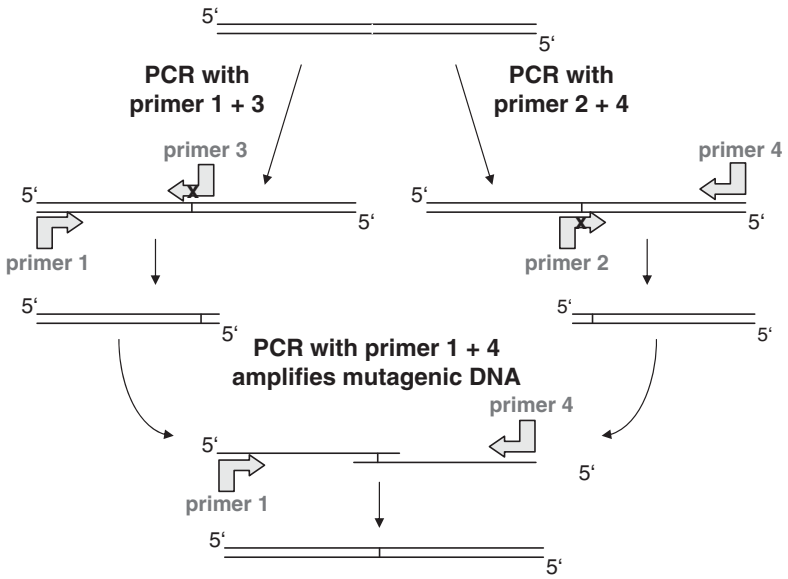

Fig. 35.2. Principle of site-directed mutagenesis by overlap extension. Mutations introduced by primers 2 and 3 are marked with an "x". Further explanations are given in the text

first polymerase chain reaction (PCR) step with two separate PCRs being performed. The primer pairs for these PCRs are 1 and 3 as well as 2 and 4, respectively, with primers 2 and 3 containing the mutant codon with a mismatched sequence. Two double-stranded DNA products containing the desired mutagenic codon are obtained over several PCR cycles. In the second PCR step these two dsDNA products are amplified using primers 1 and 4 resulting in the mutated DNA. A useful variant of the overlap extension method is the megaprimer method *(7)*. In this procedure, two rounds of PCR are performed employing two flanking primers and one internal mutagenic primer that contains the desired base substitutions. A benefit of this method is that mutations can be inserted into the flanking primers so that multiple codons relatively far from each other can be replaced at the once. The second method for performing site-directed mutagenesis is referred to as whole plasmid, single-round PCR (**Fig. 35.3**). In this protocol, two oligonucleotide primers containing the desired mutation(s) are extended with DNA polymerase. In this PCR step, both strands of the template are replicated without displacing the primers to obtain a mutated plasmid containing breaks that do not overlap. As the original wild type plasmid originates from *Eschenichia coli* and is thus methylated on various A and C residues, it may then be selectively digested using Dpnl methylase endonuclease resulting in a circular, nicked vector containing the mutant gene. When this nicked vector is transformed into competent cells, the nick in the DNA is repaired by the cell machinery to give a mutated, circular plasmid. The advantages of the whole plasmid, single-round PCR are that only one PCR needs to be performed and only two primers are required. The disadvantages of this technique relative to overlap extension are that it does not work well with large plasmids (>10 kB) and typically only two nucleotides can be replaced at a time *(8)*. Several companies offer kits for performing these methods.

After the PCR step and the cloning and/or transformation, expression and purification of the recombinant protein mutants must be performed for testing and evaluation.

2.1.2. Computational Protein Design

During the past two decades, computer simulations of the dynamics of proteins has become a widely used tool to deepen our understanding of these molecules. Computer simulations can be used to understand the properties of a molecular system in terms of interactions at the atomic level. One of the main challenges is the development of algorithms that can deal directly with structural and

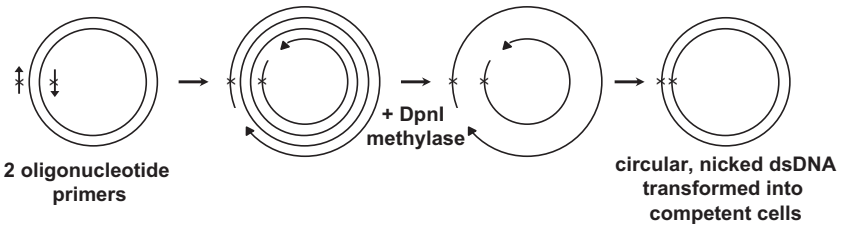

Fig. 35.3. Principle of site-directed mutagenesis by whole plasmid, single-round PCR. Explanations are given in the text

functional specificity. An excellent overview of the strengths and weaknesses of various search algorithms is reported ***(9)***, and implementations of these algorithms were quantitatively evaluated ***(10)***.

Computational protein design methods seek to identify amino acid sequences that generate low-energy interactions of a specified target protein structure by employing a variety of optimization techniques. These fall into two broad categories: stochastic algorithms, including Monte Carlo, and deterministic algorithms, including dead-end elimination. Stochastic algorithms semi-randomly sample sequence-structure space and move toward lower energy solutions whereas deterministic algorithms perform semi-exhaustive searches.

The advantage of stochastic methods is that they can deal with problems of significant combinatorial complexity because they do not require an exhaustive search. The disadvantage is that there is no guarantee that these methods converge to the global minimum energy solution or even the same solution when run multiple times ***(10)***. In contrast, deterministic methods always converge on the same solution.

2.1.2.1. Monte Carlo (MC) Method: These simplest stochastic methods are a widely used class of computational algorithms for simulating the behavior of various physical and mathematical systems. They are distinguished from other simulation methods (such as Molecular Dynamics, see Section 2.1.2.3.) in that they are nondeterministic in some manner, usually by using random numbers. In the context of design, a starting structure is perturbed by a random change in residue type or rotamer at some position. If the change decreases the energy of the structure, it is accepted. Otherwise, the Metropolis criterion, including a Bolzmann weighted probability, is used to accept or reject the change. This permits energetically unfavored uphill moves and escape from local minima. MC methods are especially useful in studying systems with a large number of coupled degrees of freedom, such as liquids, disordered materials, and strongly coupled solids ***(11,12)***.

2.1.2.2. Dead-End Elimination (DEE): The DEE algorithm is a method for minimizing a function over a discrete set of independent variables. The basic idea is to identify "dead ends," i.e., "bad" combinations of variables that cannot possibly yield the global minimum and to refrain from searching such combinations further. Hence, good combinations are identified and explored further. The method itself has been developed and applied mainly to the problems of predicting and designing the structures of proteins ***(13)***. The basic requirements for DEE are a well-defined finite set of discrete independent variables, a precomputed numerical value, the energy, associated with each element in the set of variables, a criterion or criteria for determining when an element is a "dead end," and an objective function, the energy function, to be minimized. DEE has been used efficiently to predict the structure of side chains on a given protein backbone structure by minimizing an energy function. A large-scale benchmark of DEE compared to alternative methods of protein structure prediction and design is that DEE reliably converges to the optimal solution for a given protein length, and it runs in a reasonable amount of time ***(13)***. However, other methods are significantly faster than DEE and thus can be applied to larger and more complex problems. DEE is guaranteed to converge to the global minimum energy solution ***(13)***. The effectiveness of DEE for a combinatorial search is due to the systematic elimination or pruning

of high-energy rotamers or rotamer-combination. A requirement is that the energy function must be written as the sum of individual and pairwise terms. Additionally, for extremely complex problems, DEE may fail to converge, but due to some large improvements DEE currently seems to be the most powerful method for finding the global minimum energy solution ***(10)***.

2.1.2.3. Molecular Dynamics (MD) Simulation: Molecular modeling tools are used in protein engineering studies to indicate which amino acid substitutions or mutations have a high probability of success and should be tested experimentally. Molecular dynamics (MD) are able to correlate the increase in protein stabilization with the conformational and structural changes caused by (single) amino acid replacements. It represents an interface between laboratory experiments and theory. MD also serves as a tool in protein structure determination and refinement using experimental tools such as X-ray crystallography and NMR. Additionally, MD has been applied as a method of redefining protein structure prediction.

The computer simulation method of MD is based on an extremely simple principle: given the coordinates of all atoms in a molecular system and an accurate description of the total potential interaction energy as a function of the atomic coordinates, the force on each atom can be calculated. Describing the interactions accurately in a protein is a key element to protein design and probably the most difficult. The energy functions must be fast and accurate, yet not oversensitive to the fixed backbone approximations and discreteness of the rotamer library (reviewed in **ref. *14)***. In chemistry and biophysics, the interaction between the objects can be described by a force field. Molecular mechanics force fields for proteins, such as AMBER, GROMOS, and CHARMM, usually include van der Waals, electrostatics, dihedral angle (torsion), bond angle, and bond stretching (length) terms. These parameters are further adjusted by simulations that attempt to reproduce experimental data, such as small molecular crystal structures. For protein design calculations, considerable modifications are required. Energies must be adjusted to reduce artifacts resulting from the use of discrete rotamers and fixed backbones. Energy terms that describe solvation must be added. Secondary structure propensities have also been used as constraints for sequence design. A reference state needs to be defined, since the relevant value for protein design is the difference in energy between the probed and reference state. Finally, all these terms must be weighted appropriately. For molecular dynamics simulations, the individual energy terms are typically added and must be appropriately parameterized and scaled with respect to one another ***(15)***.

Considerations for computational protein design ***(16)***:

1. Energy expression or force field used to rank the desirability of each amino acid sequence for a particular backbone.
2. Energy minimization of the target backbone must be determined in order to experimentally test the energy expression. (Published algorithms include MC techniques and DEE).
3. Discrete side chain conformations must be made to restrict the complexity to a reasonable limit. The allowed side chain conformations are typically chosen from a library of discrete possibilities, known as rotamers.
4. Classification of residue position to reduce the size of the design problem. Protein cores are typically composed of hydrophobic amino acids, and

protein surfaces are largely composed of hydrophilic amino acids, but the boundary residues must be selected from the full range of amino acids as these positions are observed to be both.
5. Modeling of backbone flexibility by using a softer van der Waals potential, which means, giving the modeled atoms a fuzzy edge.

Available computer power must be considered when designing MD simulations. Simulation size (number of particles, typically up to 10^5 atoms), time-steps and total time duration must be selected so that the calculation can finish within reasonable time. However, the simulations should be long enough to be relevant to the time scale of the natural processes being studied. Most scientific publications about the dynamics of proteins and DNA use data from simulations spanning nanoseconds to microseconds. To obtain such simulations, several CPU-days to CPU-years are needed. Another factor that impacts total CPU requirement by a simulation is the size of the integration time-step. This is the time length between evaluations of the potential. The time-step must be chosen small enough to avoid discrete errors. Typical time-steps for classical MD are in the order of one femtosecond. Furthermore, a choice should be made between explicit solvent and implicit solvent. Explicit solvent particles (like water) must be calculated extensively by the force field. The impact of explicit solvents on CPU-time can be 10-fold or more. In simulations with explicit solvent molecules, the simulation box must be large enough to avoid boundary condition artifacts.

Limitations must not only be kept in mind when setting up simulations but also when drawing conclusions from such simulations. Consequently, the results of simulations must be critically evaluated and, whenever possible, validated through experiments. When applied in an appropriate way, MD is a tool complementary to experimental methods, which can be used to access atomic details inaccessible to experimental probes ***(17)***.

2.2. Library-Based Design Strategies

Library-based design strategies have the advantage that they do not rely on structural information. Various methods for designing libraries exists which are described in detail in the chapter *Directed Protein Evolution*. The success of libraries, however, strongly depends on the selection or screening method. This section introduces the most prominent techniques and discusses advantages and disadvantages of each system.

2.2.1. Phage Display

Phage display is a reliable and widely used selection technique. It enables the rapid screening of peptide libraries or proteins against virtually any desired target both of biological and synthetic origin. This could be either of biological interest or for technical or medical applications. The benefits of phage display rely on the fact that the phenotype is directly linked to the genotype. This is because the peptides to be screened are expressed as fusion proteins of a phage coating protein, and genetic information is packaged into the phage ***(18)***.

Typically, phage display as well as ribosome display (see Section 2.2.2) selection rounds are carried out *in vitro* where incubation with the target takes place. Strongest binders remain bound to the target and nonbinders or only weakly interacting binders are removed from the pool upon increasing

stringency. The selected pool of binders is amplified, either in *E. coli* for phage display or by RT-PCR in the case of ribosome display. Enriched phages then enter the next round of selection. This procedure is repeated for three to five times and leads to the enrichment of binders dominating the pool. This procedure is called "panning" and without it would be akin to searching for a needle in a haystack.

Phage display was the first display technology shown to physically couple the phenotype with genotype ***(19)***. It was originally used to map antibody epitope binding sites by screening peptide-phage libraries against immobilized immunoglobulins. Filamentous phages use its bacterial host to replicate and to assemble the phage particles. For phage display, the phage genome can be modified to incorporate the gene of interest to be displayed in fusion to a surface protein. The most commonly used phage protein for displaying peptides of interest is the minor coat protein 3, which is presented three to five times on the M13 particle (**Fig. 35.4**) ***(20)***. The coat protein of gene 3 consists of three domains, a C-terminal constant region which anchors the protein to the phage particle, and two N-terminal domains, N1 and N2, mediating infectivity. N1 binds to the TolA receptor and N2 binds to the F-pilus of *E. coli*. Proteins of interest are usually fused to the N-terminus of the gene 3 protein. During the assembly process, resulting fusion proteins are transported through the inner cell membrane to the bacterial periplasm and incorporated into the phage particle,

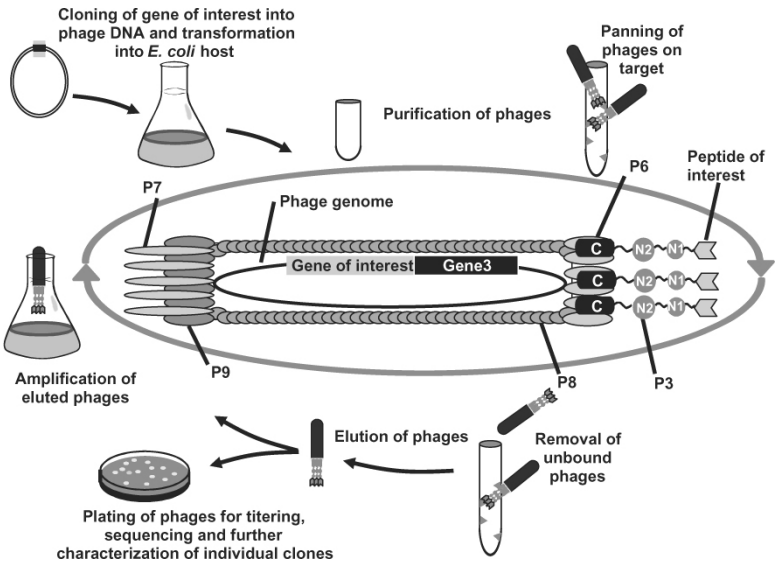

Fig. 35.4. Schematic presentation of an Fd-bacteriophage (middle) and the different steps of a phage panning round. The gene of interest is genetically linked to the N-terminal domain of the phage surface protein 3 and thus is incorporated up to five times in the phage particle (peptide of interest). Another surface protein, which can be used for multivalent display, is the protein P8. After cloning and transformation into the *E. coli* host, phages are purified via polyethylene-glycol precipitation (PEG/NaCl) and incubated with the immobilized target. Unbound phages are removed by increasing washing steps for each selection round. Binders are eluted from the target by acidic pH shift or tryptic digest and amplified upon host *E. coli* infection

while their respective single-stranded DNA (ssDNA) gets packaged in the phage ***(20)*** thereby coupling phenotype and genotype.

Typically, a selection (or panning) round can be divided into several distinct steps (**Fig. 35.4**) ***(21)***. To start, the gene of interest, which can be a library or a single protein, is fused to the gene 3. This modified phage genome is transformed into an *E. coli* host strain (e.g., XL1-blue or ER2738). Upon phage production, the protein of interest is displayed on the phage surface as fusion to protein 3. For the selection, phages are incubated with the target protein, immobilized either in an ELISA well or an immuno test tube. Simple washing steps remove unspecific or weak binding phages. Stringency can be increased from round to round by adding more washing steps and harsher conditions. Phages are well tolerable against heat and denaturing agents ***(22)***. Binders are eluted by an acidic pH-shift or by a tryptic digest. These phages are then amplified in an *E. coli* host strain, purified, and enter the next round of selection.

Phage display classically means multivalent display, as the gene 3 protein is modified in the phage genome. This technique is well suited for short peptides that do not influence infectivity of the protein 3. If selection of longer proteins is desired, a trypsin cleavage site should be incorporated between the protein of interest and the protein 3. If phages are eluted by trypsin digest, the protein of interest is cleaved off and the free phages display a wild-type like protein 3. Furthermore, multivalent display is advantageous for low affinity binders. Alternatively, monovalent display can be achieved using a phagemid system ***(21)***. In this case, the gene of interest is cloned to a truncated version of gene 3 in a phagemid vector. A phagemid carries in addition to an *E. coli* origin of replication for plasmids and an antibiotic resistance gene an origin of replication for phages, which is only used after cells are super-infected with helper phages. Helper phages provide the full phage genome, including the protein 3. Thus, cells transformed with the phagemid and infected with helper phages express a mixture of wild type and fusion protein 3. Consequently, phages show the same ratio of wild type and fusion protein 3, which ideally is one fusion protein per phage.

Phage display allows for the rapid selection of target-specific binders in three to five panning rounds. Identification of the selected clones occurs via sequencing of the DNA of phage pools and single clones and hence yields directly the primary structure of the selected peptide. Typically, selected peptides harbor affinities in the μM- to the nM-range. Owing to the avidity effect, multivalent display is more sensitive and therefore detects lower-affinity binding.

Beside the protein 3, the major coat protein P8, which is represented up to ~2700 times, can also be used as fusion proteins ***(18)***. The high number of displayed peptides in this case was recently shown to have advantages for imaging applications ***(23)***. However, the protein 3-based display system is the major method of choice as it enables the screening of large proteins or protein domains. Also, the display efficiency can be increased by choosing different signal sequence domains N-terminally of the protein 3 which are necessary for periplasmic transport during phage assembly ***(24)***. Other systems using a split version of the protein 3, so-called “selectively-infective phages” (SIP) have been tested as well but were found to be more susceptible to mutation or recombination events ***(25,26)***. The filamentous phage system is limited to proteins which correctly fold in the periplasm of *E. coli*. Other proteins can be

screened using lytic phage systems such as T4 ***(27)*** and T7 ***(28)***. The phage assembly and hence incorporation of fusion proteins occurs in the cytoplasm and virions are released upon cell lyses.

2.2.2. Ribosome Display

The first cell-free *in vitro* display described in 1994 was ribosome display ***(29)***. The basis of ribosome display is the linkage of the mRNA with the protein of interest. This can be via a stabilized complex on the ribosome (ribosome display) or via a covalent protein-mRNA complex by means of a DNA–puromycin linker (mRNA display).

Typically, a selection round consists of the following steps (**Fig. 35.5**): First, the DNA encoding the gene of interest to be selected needs to be transcribed into mRNA, which is next translated using either a bacterial or a eukaryotic *in vitro* translation system. The stabilization of complexes between the expressed protein, ribosome, and mRNA upon termination of elongation is achieved by a terminator sequence forming a hairpin structure combined with low temperature or chloramphenicol. This is where mRNA-display differs from ribosome display ***(30)***. The selected protein is covalently linked to its mRNA via incorporation of puromycin, which has been previously attached to the 3'-end of the mRNA via a short oligonucleotide. Thus, the large complex of ribosome, mRNA and protein in ribosome display is missing in mRNA-display, and unspecific interactions between the selected protein and the ribosomes are circumvented.

Once expressed, the selection rounds itself can be performed. The target of interest is immobilized in an ELISA well or test tube via adsorption, comparable to phage display. Unbound target is removed in washing steps. In mRNA display,

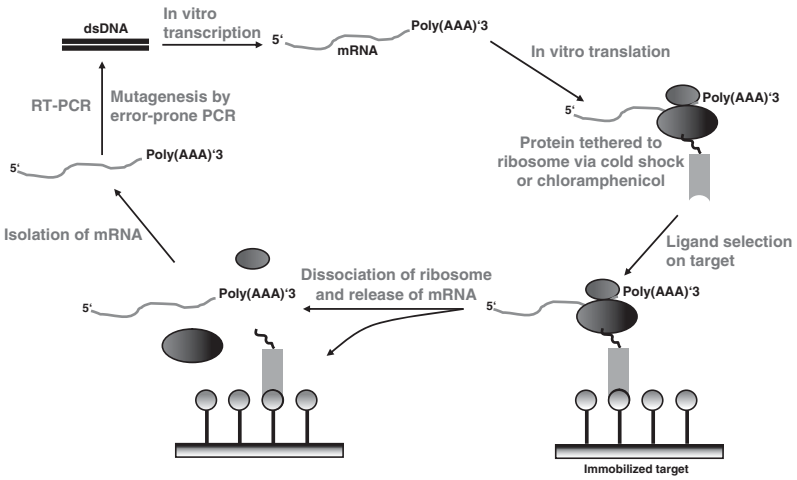

Fig. 35.5. Schematic presentation of a ribosome display round. The gene of interest is transcribed from dsDNA into mRNA and translated into proteins by *in vitro* techniques. The ribosomes remain tethered to the mRNA by either cold shock or chloramphenicol. This step ensures that the genotype remains coupled to the phenotype. The proteins are incubated on the target, and the mRNA of the strongest binding interaction partner is captured after selective washing steps. Using RT-PCR, the mRNA is reverse transcribed into DNA. Using error-prone PCR, defined mutations can be inserted which further increase the binding affinity and specificity

binders are eluted together with their immobilized mRNA, whereas in ribosome display, the mRNA is freed by destroying the protein-ribosome-mRNA complex. In both cases, the mRNA is then amplified by RT-PCR, and the resulting cDNA matrices are transcribed again into mRNA and enter the next round of selection.

One big advantage of ribosome or mRNA display is that it is a cell-free system where expression of toxic proteins or poorly folded proteins can be tolerated. Moreover, the expression and selection of the proteins of interest is not influenced by any growth stress originating from the bacterial host. Importantly, the library size is not limited by transformation efficiencies. Instead, the DNA encoding the library members is directly transcribed into mRNA, and immediately enters the selection process and in this way is only limited by the enzyme reaction. The stringency conditions during the selection rounds are similar to those performed in phage display. However, it is notable that phage particle are very robust. They remain functional even under elevated temperatures or in presence of a chemical denaturant like guanidine. An advantage of mRNA or ribosome display is the potential for affinity maturation through recursive mutagenesis, in which selectants can be further mutated after each round of selection ***(31)***. This is faster in comparison to cell-based selection as the encoding DNA does not need to be retransformed into *E. coli* host cells.

2.2.3. Yeast Two-Hybrid System

The "two-hybrid" or "interaction trap" method enables to identify, characterize and even to manipulate protein–protein interactions. It was invented in the early 1990s by Stanke and Fields ***(32)***. The yeast-two hybrid system exploits the fact that many eukaryotic transcription factors have at least two distinct functional domains, one that drives DNA-binding to a promoter region and one that activates transcription. It has been shown that DNA-binding and activation domains of one transcription factor can be exchanged from one to another while retaining its function. The basis for this method is the use of the yeast transcription factor GAL4, which is incapable of activating transcription without physical linkage to an activating domain ***(33)***. This linkage, which can be mediated by two interacting proteins, is the key to the successful use of the "two-hybrid" method. Only interaction between these proteins connects the DNA-binding domain to the activator domain, resulting in the expression of a reporter gene and thus leading to the identification of interacting partner proteins. The most extensively used vectors are based on GAL4. An alternative system makes use of the DNA-binding domain of the LexA protein and the activator domain of the viral protein 16 (VP16).

In general, in any two-hybrid experiment, a protein of interest is fused to a DNA-binding domain and transfected into a yeast host cell bearing a reporter gene controlled by this DNA-binding domain. This fusion protein, which can not activate transcription on its own, can be used as "bait" or as "target" to screen a library of cDNA clones (prey) that are fused to an activation domain. The cDNA clones capable of forming a protein–protein interaction with the bait protein are identified by their ability to cause activation of the reporter gene (**Fig 35.6A**). The DNA-binding (DBD) domain and the activator domain proteins (AD) can be transformed separately into two different strains, resulting in an AD- and a DBD-strain. In this way, a haploid DBD strain can be mated

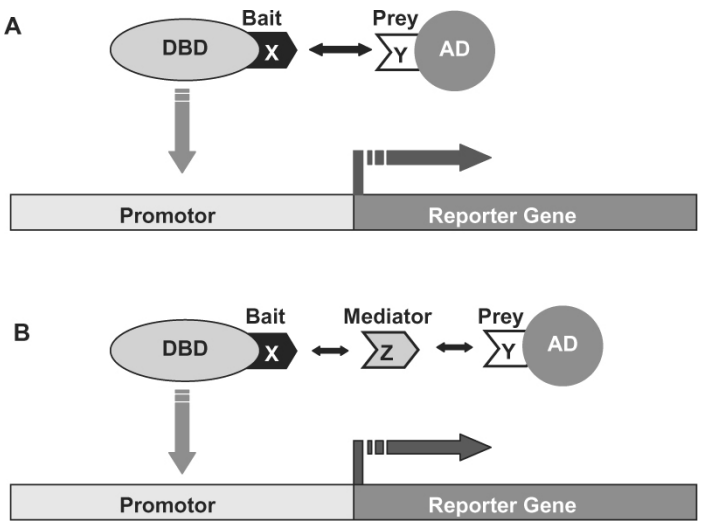

Fig. 35.6. Schematic representation of the yeast two-hybrid **(A)** and yeast three-hybrid system **(B)**. **(A)** The bait protein X is genetically fused to a DNA-binding domain (DBD) of a transcription factor, missing the transactivation domain. Upon interaction with the prey protein Y, which is fused to the transactivator domain (AD) of the transcription factor, the transcription of a reporter gene is initiated, and in this manner, an interaction between the bait protein X and the prey protein Y is mapped. **(B)** In the yeast three-hybrid system, the interaction between bait X and prey Y is mediated by a third protein Z

to the haploid AD array to identify individual interacting AD fusions. Another approach would be to mate individual DBD strains with libraries of AD strains ***(34)***. Reporter gene activation leads to the identification of the selected AD fusion. Thus, this method enables the screening of proteins that interact *in vivo* and is therefore a well-suited method to create a protein–protein interaction map of a cell or an organism.

After transformation and expression of the fusion protein, the first test is to check whether the target protein with the DNA-binding domain exerts autotranscriptional activity. If this should be the case, the experiment needs to be redefined. After testing the autoactivity of the fusion protein, the library of choice in fusion to the activator domain can be transformed. Upon screening on selection marker plates, positive clones are identified via reporter gene assays, e.g., LacZ, and the DNA of the selected clones is prepared and sequenced. After the identification of a selected clone it is necessary to test the specificity again in the two-hybrid system and also in a different system. This can include *in vitro* pull-down assays or co-immunoprecipitations, both evaluating the biological relevance of the interaction.

2.2.3.1. Advantages: One big advantage of the two-hybrid system over classical biochemical and genetic approaches is its use as an *in vivo* assay, with yeast as a live test tube, exhibiting similar conditions to higher eukaryotes. Compared to biochemical approaches that need huge amounts of purified protein or good quality antibodies, the two-hybrid system requires only the cloning of the full-length or even partial cDNA of interest to start the screening.

The genetic reporter gene strategy results in a significant amplification of the read out. This facilitates also detection of weak or transient interactions, which are often the most interesting in signaling cascades. Besides the screening of new interaction partners, the two-hybrid system allows also for mapping of residues crucial for an interaction. The most convincing argument in favor of the two-hybrid is the number and the speed in which many signaling cascades have been resolved in molecular detail.

2.2.3.2. Disadvantages: As mentioned above, the key to the method relies on the fact that the DNA-binding and transcriptional activation are separated. Thus, if the protein of interest exhibits transcriptional function on its own, the use of this protein in a two-hybrid system may not be successful and could be a limiting factor. Furthermore, as the bait and the prey proteins are expressed in fusion to the DNA- and to the activating domains, the resulting chimeras might have different conformations which could result in altered function, resulting in lower activity or even in the inaccessibility of binding sites. If this is the case, it might be worth trying to switch the fusion proteins of bait and prey.

Moreover, the protein of interest needs to be stably expressed and folded in yeast. This can be seen as an advantage rather than a disadvantage, since yeast is closer to higher eukaryotes than *in vitro* experiments or those based on bacterial hosts. Folding problems in yeast can also be accompanied by post-translational modifications that either do not occur or are yeast specific. However, this can possibly be circumvented by co-expressing the enzyme responsible for the posttranslational modification. In addition, it should be noted that the system needs the fusion proteins to be targeted to the nucleus, which could be a limiting factor for, e.g., extracellular proteins. Another problem could be a toxic effect upon expression of the fusion proteins, which has been shown for cyclins and homeobox proteins. Usage of an inducible promoter might circumvent the problem. It has to be noted that after successful identification of two interacting partners, the biological relevance of this interaction remains to be determined to prevent the identification of artificially interacting partners. Even if identified in the assay, it could be possible that these proteins are never in close proximity to each other in the cell. A good representation of the library is necessary to screen successfully. Therefore, it has to be considered that only one out of six fused cDNAs is in the correct frame, which increases the number of clones to be investigated.

2.2.3.3. Reverse Hybrid System: The two-hybrid system does not allow genetic selection of events responsible for dissociation of particular interactions. However, a reverse two-hybrid system makes use of the expression of a counter selectable marker that is toxic and hence leads to a growth arrest. Thus, the dissociation of an interaction provides a selective advantage. One example given here is the "split-hybrid" system, which is based on the *E. coli* TN10-encoded tet repressor/operator system. Upon interaction of the target protein with its prey protein, the transcription of the TetR is initiated. The TetR protein represses then the expression of the HIS3 gene, leading to a growth phenotype on plates without histidine in the growth medium ***(35)***. Abrogation of the interaction, either by mutating one of the proteins or by introducing a dissociator protein shuts down the TetR expression and enables again HIS3 expression and thus growth on selective plates. This method can be used in

screening large libraries of peptides or compounds that inhibit selectively a protein interaction ***(36,37)***.

2.2.3.4. Sos-Recruitment System (SRS): The mammalian GDP-GTP exchange factor hSos (*human son of sevenless*) can only activate Ras when hSos is localized to the plasma membrane in close proximity to Ras. In yeast, functional Ras signaling pathway is required for cell viability. This fact has been exploited in the hSos-recruitment system and similarly in the Ras-recruitment system (RRS). Both systems benefit from the fact that a yeast strain, mutated in the Ras guanyl nucleotide exchange factor cdc25-2, shows temperature sensitive growth. For the screening assay, the target protein is fused to hSos, and the prey protein to be screened is fused to a membrane localization signal. Coexpression of these proteins in a cdc25-2 yeast strain leads to a temperature dependent growth phenotype if the fusion proteins interact and allow hSos recruitment to the membrane ***(38–40)***.

2.2.3.5. Yeast Three-Hybrid System: A limitation in the two-hybrid system is the lack of the detection of post-translational modifications, e.g., tyrosine-phosphorylation, which do not occur in *Saccharomyces cerevisiae*. In the so called kinase three-hybrid system, a cytosolic tyrosine kinase has been introduced into the yeast cell, phosphorylating specific substrates ***(41)***.

In the yeast three-hybrid system, the target protein activates only transcription via the activating prey protein if a third protein is present. This third protein either mediates the interaction or induces a conformational change thereby promoting interaction (**Fig 35.6B**). In this way it has been shown that the interaction between Sos and the cytoplasmic domain of the EGFR is Grb2-mediated ***(42)***. The three-hybrid system can also be extended to the use of a heterodimer of small organic ligands, incorporated into the media plates, which induce dimerization of, e.g., the glucocorticoid receptor and in this way activate the transcription after diffusion into the yeast cell ***(43)***. This system is of great interest in pharmacological approaches since small-ligand receptor interactions are the basis for many signaling cascades and misregulation is the cause of many diseases. Hence, the screening of a library of small ligand compounds with the three-hybrid system could identify new drug lead compounds.

Together, these advances have led to a variety of hybrid screening systems each with its own limit and suffering from the fact that each strategy is capable of detecting only a subset of interactions. This argues for the use of multiple systems to maximize coverage.

2.2.4. Protein-Fragment Complementation Assay (PCA)

Protein-Fragment Complementation Assays (PCA) are a powerful tool for studying protein-protein interactions and are used e.g. in protein engineering for selecting tightest binding partners from peptide libraries. For PCA selection, a peptide library and the target protein or a domain thereof are fused to a reporter protein which is dissected into two non-functional fragments, sometimes referred to as $\Delta\alpha$ and $\Delta\omega$ or fragment 1 and 2. Interactions of the studied proteins or domains are demonstrated by restoration of the reporter proteins functionality (**Fig. 35.7**). The reporter protein must monitor the association of the test proteins without promoting it. Interaction must be mediated by the interaction under investigation. A combinatorial approach for generating a reporter protein for PCAs was introduced by Tafelmeyer et al. ***(44)***.

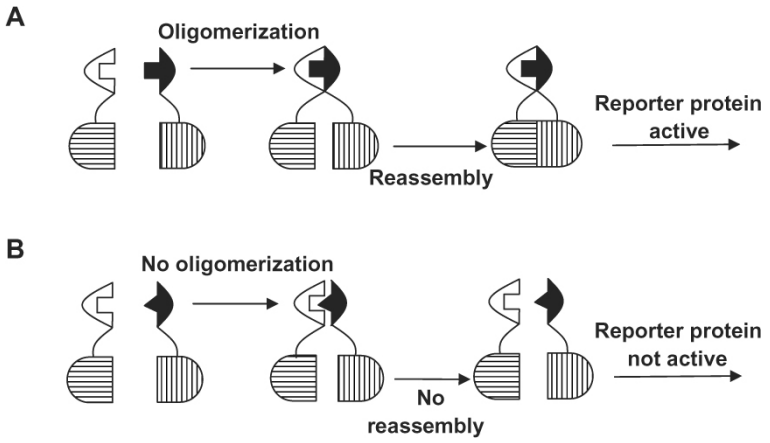

Fig. 35.7. General principle of protein-fragment complementation assays (PCA). Oligomerization domains (black and white) are fused via linkers to the reporter protein fragments (striped). Further explanation are given in the text

In contrast to two-hybrid techniques PCAs are not generally limited to the nucleus, where the proteins lack the appropriate cellular context. Also two-hybrid assays cannot be used to test temporal aspects of protein interactions (***45***).

An interesting point of the PCA system is that, using known interaction domains, it can also be used to study the reporter protein, e.g., mutate residues in the binding interface of the two fragments.

Reporter proteins for PCA have to fulfill several requirements:

- small and monomeric
- overexpression possible in eukaryotic, prokaryotic or both cell types
- the two fragments must be stable and soluble to enable reassembly
- the cleavage site must not be in a functional position
- cleavage site ideally close to the N- or C-terminus to permit different orientation in the fusion protein
- only reassembly of both fragments restores activity to avoid false positives
- miminal auto-reassembly to prevent background (false positive)
- easy discrimination of active and inactive reporters for selection or screening of interacting partners
- no endogenous reporter protein of same activity present in the host or host protein can be efficiently inhibited

Different systems have been developed each with inherent advantages and disadvantages. Here we provide a short overview over the different reporter systems used for PCA.

2.2.4.1. Murine Dihydrofolate Reductase (mDHFR): The dihydrofolate reductase (DHFR) ***(46)*** is an enzyme in the nucleotide synthesizing pathway, which fulfills the requirements of a reporter protein very well. It is a small (21 kD), monomeric protein of known structure ***(47)*** and its folding properties and kinetics are well characterized ***(48)***. In nucleotide-free media, DHFR is essential for cell growth. The endogeneous DHFR of *E. coli* can be inhibited by the substrate analogue trimethoprim to which it has a 12,000 fold higher affinity

compared to mammalian DHFR ***(49)***. Consequently, a murine DHFR can be used for simple survival assays in *E. coli*. It has been shown that mDHFR can be disrupted in a loop-region formed by the amino acids 101–108 ***(50)*** which is the loop between domain two and three, in close proximity to the N-terminus of the enzyme. This permits fusion of either the N-terminus of both fragments to the dimerization domains or alternatively one N- terminal and one C-terminal fusion. Pelletier et al. chose to fragment mDHFR between residue 107 and 108 and fuse interacting proteins N-terminally to the resulting DHFR fragments. The DHFR fragments are stably expressed and reassemble only when fused to a dimerization domains (in this case the leucine zipper of GCN4). After its assisted reassembly mDHFR becomes active and allows *E. coli* to grow on trimethoprim-containing minimal media (**Fig. 35.8**). Without interaction of the dimerization domain no growth will occur. The speed of growth is related to the strength of interaction of the dimerization domain. In addition to cell growth, DHFR activity can be monitored *in vitro* by fluorimetry following the appearance of tetrahydrofolate (THF) (excitation at 310 nm; emission at 360 nm) using the inhibitor methotrextate (MTX) as a control.

The system has two minor disadvantages: As the DHFR needs NADPH as a cofactor, the assay does not function in the periplasm of *E. coli*. Another drawback is the ability of *E. coli* to eventually overcome inhibition of the endogenous DHFR by mutation after some rounds of selection ***(51)***. However, this is rare and easily to detect by controls on plates without IPTG. Without induction, no DHFR-fragments are expressed and thus no growth occurs.

The most obvious advantage of the DHFR-PCA is the easy screening of positive interactions by the survival assay, which makes this assay very valuable especially for screening large libraries. The survival assay can be followed by growth competition in liquid culture under selective conditions to enrich the best binding sequence. Another advantage is the control of stringency of the

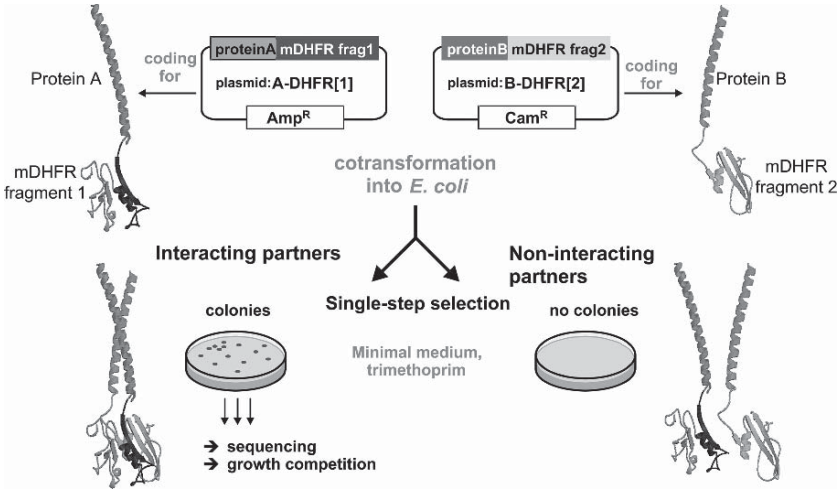

Fig. 35.8. Proteins of interest are genetically fused to mDHFR-fragments and cotransformed in *E. coli* and assayed on minimal medium plates containing trimethoprim. Only interacting proteins enable reassembly of mDHFR, resulting in growths of colonies. Colonies are pooled and best interacting proteins are enriched in growth competitions. Library pools and single clones can be analyzed by sequencing

assay by mutating the binding interface of the two fragments to alter the stability of their reconstituted state, e.g., exchanging wild type isoleucine 114 for alanine or valine ***(46)***.

2.2.4.2. Ubiquitin-Based Split-Protein Sensor (USPS): This PCA was developed in 1994 by Johnsson and Varshavsky ***(45)***. Ubiquitin acts here as split protein, but for detection of ubiquitin-reassembly an additional protein is needed as reporter. For this, the property of eukaryotic cells is used to cleave ubiquitin-fused proteins by the specific protease UBP. This process is strongly dependent on the correct folding of ub. For USPS ubiquitin is dissected into its two domains and the N-terminal fragment is mutated to inhibit autoreassembly. The reporter protein is fused to the C-terminal fragment. The reporter protein is cleaved on reassembly and correct folding of ubiquitin. Detection depends on the chosen reporter, for example an domain detectable by antibodies or an enzyme which becomes active only after cleavage. The proteins of interest are fused to the ubiquitin fragments which when reassamble permit ubiquitin to fold correctly and the reporter protein is cleaved. However, the signal can be the result of both fragments binding to a common ligand, because USPS detects the proximity of proteins, which does not necessarily means direct interaction. USPS relies on constitutive expression of the host cells ubiquitinase, which had been shown only for cytosol and nucleus.

USPS can also be used *in vitro* with purified fragments and purified proteases such as Ubp1 from yeast.

2.2.4.3. β-*Galactosidase*: Intracistronic β-galactosidase complementation has long been observed ***(52–54)***, and shows that bacterial β-galactosidase activity can be restored when two variants with inactivating mutations in different crucial domains of the enzyme share their intact domains. This is largely efficient but depends on the nature of the mutations ***(55)*** and holds also true when transferring to mammalian cells ***(56)***. Developed in the Blau group, the β-galactosidase-based PCA utilizes this well-known property of β-galactosidase complementation for the first time for time-dependent *in situ* studies of protein–protein interactions in living eukaryotic cells. Rossi et al. ***(52)*** chose β-gal mutants which have been shown to be unable to complement each other by themselves. The first one lacks only amino acids 11–41 of the wild type and is a naturally occurring mutant described earlier as M15 ***(57,58)***. The second contains the first 788 residues of β-gal ***(56)***. The proteins of interest are fused to these fragments. Activity of β-gal measured at different time points by either biochemical assays or FACS (summarized in ***(53)***) shows the interaction of the fused protein fragments and gives a quantitative readout. As an enzyme, β-gal amplifies the resulting signal allowing monitoring of physiological interaction without overexpression. In contrast to other methods, this PCA can be used to analyze protein–protein interactions in different subcellular compartments of eukaryotic cells. However, because its active form is a large tetramer some interactions might be sterically hindered.

2.2.4.4. β-*Lactamase*: The bacterial enzyme β-lactamase ***(59,60)***, which confers resistance to the antibiotic ampicillin, has long served as a model in protein engineering with well understood properties. Its structure does not suit fragmentation, but this enzyme has significant advantages for use in mammalian cells over other systems. This is because the fragments are small and there

is no endogenous β-lactamase activity. In addition in 1998 a cell permeable, fluorescent substrate, called CCF2/AM, was developed ***(61)*** which detects β-lactamase activity with high sensitivity in eukaryotic cells. A drawback is the high price of these substrates.

The Blau group split β-lactamase after residue 197 ***(60,62)***, but found a high background activity in *E. coli*. They were able to reduce the signal-to-noise ratio by stabilizing the α-fragment by adding the empirically found tri-peptide NRG after aa197. The Michnick group established the assay in the cytosol ***(59)*** using the M182T mutant, which has been reported earlier to stabilize the structure of β-lactamase ***(63)***.

2.2.4.5. Luciferase: The enzyme luciferase ***(64)*** emits light when reacting with a specific substrate in the presence of cofactors. In nature, different types of organisms use this bioluminescence, for example by attracting prey or to scare off predators. In the laboratory, luciferases are used as reporter enzyme for studying the properties of regulatory elements in living organisms ***(65)***. Paulmurugan and colleagues used the firefly luciferase to show for the first time a protein-fragment complementation assay that monitors protein interactions in living subjects. The luciferase was split after residue 437 and both fragments ($\Delta\alpha = 437$aa, $\Delta\omega = 117$aa) were fused to the strongly interacting proteins MyoD and Id. For comparison, the fragments were also fused to inteins, resulting in a reconstituted luciferase by protein splicing. Both approaches, complementation or reconstitution, showed the same high level of luciferase activity after transfection in eukaryotic cells. Best results have been achieved in 293T-cells. In a follow up project Paulmurugan and Gambhir split successfully the synthetic humanized *Renilla* luciferase (hRLUC) for usage in PCA ***(66)***. Recently, Remy and Michnick used *Gaussia princeps* luciferase as a PCA reporter protein and found a higher activity compared to the *Renilla* luciferase ***(67)***.

Compared to other PCAs the luciferase signal is relatively short living and must be recorded immediately, making the technique somewhat laborious.

2.2.4.6. Green Fluorescent Protein (GFP): A widely used reporter protein for PCA in recent years has been GFP ***(68)*** and its derivates YFP, CFP and BFP plus the enhanced variants. This method is often referred to as BiFC short for bimolecular fluorescent complementation. Jellyfish *Aequria victoria* GFP is a 238 residue protein that forms an 11-stranded β-barrel with a coaxial helix. The chromophore ρ-hydroxybezylideneimidazolidinon is located with the helix at the center of the barrel ***(69)***. Ghosh dissected GFP at a surface loop at residue 157, which has also been shown to accept a 20-residue amino acid insertion ***(70)*** and fused the resulting fragments with leucine zipper domains for oligomerization. After protein reassembly the fluorophore formes and the protein shows its characteristic fluorescence. In contrast to other PCAs, the complex is stable and does not dissociate, thus allows capturing of transient interactions ***(71)***.

2.2.5. Cell-Surface Display

2.2.5.1. Bacterial Display: Bacterial surface display was described in 1986 for the Gram-negative bacteria *E. coli* and *Salmonella* ssp. Short gene fragments were inserted into the genes for the outer membrane proteins LamB, OmpA, and PhoE and were displayed on the cell surface. In 1992 followed

the first examples of display techniques using Gram-positive bacteria like *Staphylococcus xylosus* or *S. carnosus* and *Streptococcus gordinii*. The displayed protein mimics a receptor protein like SpaA and M6 respectively and is covalent bound to the outer membrane surface. For staphylococcal display, plasmids are used, while in *S. gordinii* the target genes are incorporated in the genome be homologous recombination. These and other bacterial display systems are reviewed by Lee et al. ***(72)***. The main application for bacterial display has been the presentation of antigens, applied to test animals by oral delivery to stimulate the production of specific antibodies by the immune system. Comparison with intracellular expressed or secreted antigens resulted in a more effective immunization using the surface display technique.

Direct comparison with phage display gives better results for the phage-based systems ***(73)***, but bacterial display has some advantages. Bacteria are easy to cultivate and can be kept free from contamination by using antibiotic selection markers. Phages in contrast need a host organism, which is susceptible to contamination with wild-type bacteriophages.

2.2.5.2. Yeast Surface Display: *Saccharomyces cerevisae* is generally a good system for expressing heterologous proteins, and transformations are possible both, by plasmid and stable integration of new genes into the genome of yeast cells ***(74,75)***. It is generally regarded as safe ("GRAS") and can therefore be used for food and drug production. A big advantage is that yeast as a eukaryotic organism is able to glycosylate and process proteins in its ER, even if it is not fully identical to mammal cells. The first targeting of a heterologous protein to the yeast cell wall was accomplished by Schreuder et al. ***(76)***. They fused the protein of interest to the C-terminus of the Aga2p subunit of α-agglutinin. This subunit is connected by two disulfide bonds to the second subunit Aga1p, which is covalently linked to the yeast cell wall (**Fig. 35.9**) The wild-type Aga2p mediates cell-cell contacts during yeast cell mating.

The display of peptide libraries on the surface of yeast was first published by Boder & Wittrup in 1997 ***(77)***. In this first attempt to display a fully functional antibody fragment and improve it by random mutation the authors enriched clones with a tenfold higher binding capacity to their target than the wild-type scFv.

Yeast surface display has some advantages over other display techniques. The covalent linkage to the yeast cell wall results in a more stable display of proteins than in other eukaryotic surface expression systems. The displayed proteins can easily be released from the cell surface for further characterization by reduction of the disulfide bonds. The cell wall also gives yeast a higher life time in industrial applications. The culture conditions are well know, thus biomass can be produced in high concentration. Choosing yeast as display system avoids unpredictable bias against expression of some eukaryotic proteins in *E. coli* and of course, this bias also affects phage display. The expression of mammalian proteins in yeast does not work in every case as has been shown for T-cell receptors ***(78)***. A limiting factor for library selection by yeast surface display is a smaller achievable library size than in *E. coli*.

Yeast cells can easily be used for quantitative screening by FACS. Alternatively, if it is difficult to obtain a purified ligand or no FACS is available, a ligand can be expressed on mammalian cell surface and binding cells can be selected by density centrifugation ***(79)***.

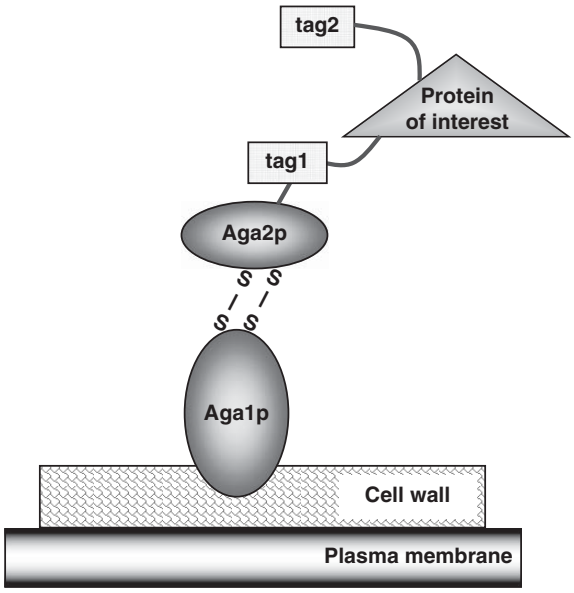

Fig. 35.9. Yeast display. Schematic view of the surface of yeast. The C-terminus of the subunit of the receptor α-agglutinin is covalently anchored in the cell wall, which is located outside the plasma membrane. The second subunit Aga2p is linked via two disulfide bonds. The protein of interest together with several tags, depending on application, is fused to the C-terminus of Aga2p

2.2.5.3. Viral Display: Similar to phage display, proteins can be displayed on the surface of viruses that infect eukaryotic cells. Because the viruses are propagated in the eukaryotic host, the proteins are fully processed by the cellular machinery thus avoiding a main disadvantage of phage display. One widely used host system is the baculovirus ***(80)***, which propagates mainly in insect cells, but is transcriptionally silent in mammalian cells, making it relatively safe. It has already been used for expression of recombinant proteins ***(81)*** and is well established (reviewed ***(82)***). Boublik et al. displayed a heterologous protein on the virus surface by fusing it to the surface glycoprotein gp64 ***(83)***. Ernst et al. demonstrated the possibility to use a baculovirus system for library selections ***(84)***. They expressed HIV-1 glycoprotein gp41 containing a randomized region in Ac-omega and selected for higher affinity to the human antibody 2F5.

Also used for viral display of libraries are retroviruses, originally shown with the murine leukemia virus (MLV) in the human cell line HT 1080 ***(85)***. The library was fused to the envelope spike glycoprotein (Env). In contrast to other viruses, retroviruses permit display of large peptides on their surface without reducing infectivity ***(86)***. They can be rescued from the target by adding permissive mammalian cells with low loss of high affinity binders. The achievable library size is with 10^6–10^8 ***(86)*** lower than for phage display but still sufficient for a range of applications.

2.2.5.4. Mammalian Cell Display: In 2005, the first example for library selection in a mammalian cell display system was published. A library with up to 13 residues was displayed on the surface of mammalian cells ***(87)***. The library was fused to the chemokine receptor CCR5 and transferred into cells using a retrovirus-based vector. After integration into the genome, the peptide library was constitutively expressed und displayed on the cell surface. As a proof of principle a peptide mimicking the FLAG epitope was successfully enriched after three rounds of selection. Ho et al. fused a small scFV library to the human platelet-derived growth factor receptor PDFGR ***(88)***, displayed it on mammalian HEK293 cells, and selected a fully functional scFV against CD22.

2.2.6. In Vitro Compartmentalization (IVC) Methods

In vitro compartmentalization (IVC) uses a water-in-oil (w/o) emulsion with some special surfactants to physically link genotype to phenotype. This is possible because most droplets contain only one gene of a library and the machinery necessary for replication. The surfactants are composed of hydrophilic and lipophilic compounds. Hydrophilic groups associate with the aqueous phase and the lipophilic groups with the oil phase thus forming stable droplets ***(89)***. It is thought that such tiny vesicles were part of the primordial soup which enabled the emergence of life ***(90)***.

By this emulsification a large reaction volume is divided in many microscopic compartments (up to 10^{10} in 50 µl reaction volume ***(89)***) thus increasing the effective concentration of all components used and at the same time reducing diffusion distances ***(91)***. These tiny w/o droplets are like the wells of a micro titer plate. If they are re-emulsified in water these w/o droplets are enveloped in water to form w/o/w droplets that can be analyzed by FACS in a much higher efficiency than it would be possible in micro titer or 96 well format if there is a fluorescence-based screening method available (***92***). The transcription and translation apparatus is provided either by bacterial cell extracts in case of prokaryotic targets and wheat germ or rabbit reticulocyte (RRL) for eukaryotic targets ***(89)***.

IVC has several advantages over methods such as phage display, ribosome display or cell surface display in that it can select for properties other than binding, such as sequence specificity, intermolecular catalysis *in trans* (substrate not linked to the catalyst) and regulatory characteristics of proteins and RNA ***(92)***. IVC is a highly flexible method with potential for totally new approaches in screening for desired properties.

2.2.6.1. Compartmentalized Self-Replication (CSR): The most simplistic IVC variant is *compartmentalized self-replication (CSR)* (**Fig. 35.10**). CSR was developed for the evolution of enzymes, especially polymerases. PCR is performed in which the individual variants of a polymerase and their respective genes are separated into compartments of an w/o emulsion ***(93)***. First the different polymerase variants are cloned into a bacterial host. These bacteria are suspended with appropriate flanking primers and nucleoside triphosphates in a heat stable w/o emulsion. Ideally, each compartment contains only one polymerase variant with its respective gene. During PCR the cells are disrupted and the polymerase is freed. Due to compartmentalization, each polymerase replicates only its own encoding gene and thus only genes that encode for active polymerases are replicated. The more active the variant is the more DNA that is produced. Consequently, there is an increased probability

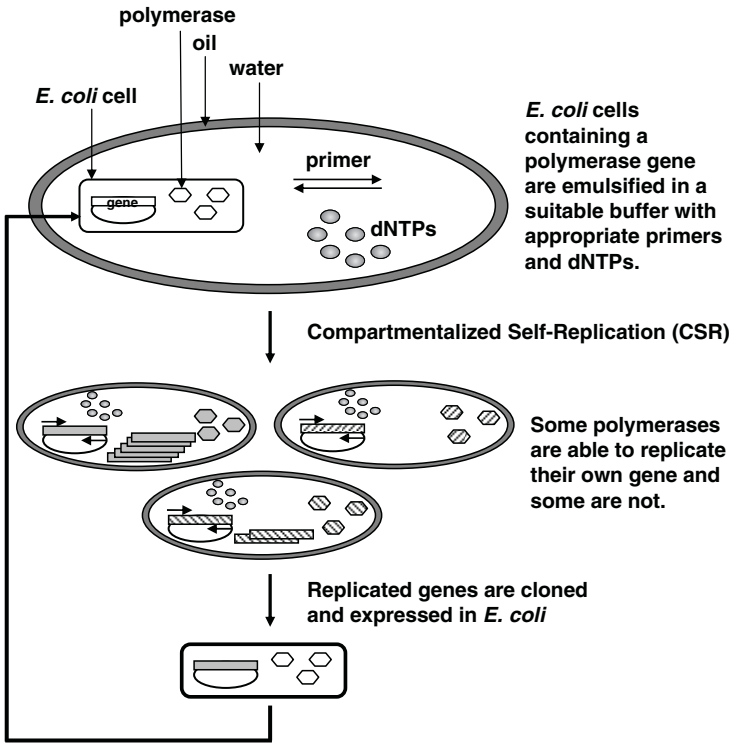

Fig. 35.10. Compartmentalized self-replication

of this variant generating clones for the next round of CSR. Inactive variants fail to amplify their own gene and are thus eliminated from the gene pool. The method bears the potential to select for enzymatic activity under a wide range of conditions.

A modification of this method is short-patch compartmentalized self-replication—spCSR ***(94)***. SpCSR is based on CSR but in SpCSR only a short region a so-called "patch" of the gene of interest is diversified and replicated. This variation allows for selection of polymerases under conditions where catalytic activity and processivity are compromised resulting in an inefficient full self-replication.

DNA-modifying enzymes like DNA-methyltransferases can be screened in a manner similar to CSR. Instead of amplifying its own DNA the enzyme can modify its own DNA thus cannot be digested after breaking the emulsion (***89***).

Doi et al. ***(95)*** adjusted IVC to select for endonucleases with altered restriction sites. The DNA coding for the endonuclease is emulsified and translated *in vitro*. An active enzyme cuts its own DNA resulting in sticky ends. In compartments with inactive enzyme, the DNA stays intact. After breaking the emulsion, a biotinylated dNTP is incorporated into the cohesive ends of the cleaved DNA by DNA polymerase, and biotinylated genes are recovered from the mixture using streptavidin coated beads and amplified using PCR. Using this method the coding gene can only be mutated in front of the restriction site of the enzyme as mutations after the restriction site are lost. If a special cleavage site is to be selected, the biotinylated tag could be added to a special oligonucleotide

representing the restrictions site to be selected. Thus, only genes coding for enzymes with the correct restriction site will be selected.

2.2.6.2. Microbead Display: The coupling of genotype with phenotype can be achieved by different approaches: In *microbead display* proteins are linked to DNA via microbeads (**Fig. 35.11**). A library of genes coding for a protein with a common tag are labeled with biotin and coupled to streptavidin-coated beads so that every bead carries approximately one gene. These beads additionally carry antibodies against the common tag. The beads are compartmentalized in w/o emulsion and the protein is translated *in vitro*. In each droplet, the transcribed proteins become attached to the antibodies on the bread and thus are linked to the gene encoding them. The emulsion is broken and the beads are incubated with horseradish peroxidase (HRP) coupled ligand or substrate. HRP converts fluorescein tyramide into intermediates that react with the protein which thus becomes labeled with multiple fluorescein molecules. These fluorescent beads can be afterwards sorted by FACS, and the DNA can be amplified and subjected to a new round of selection ***(96)***.

If the protein to be screened has enzymatic activity by which a fluorogenic substrate directly is transformed into a fluorescent product, water-oil-water (w/o/w) droplets containing active enzyme can be sorted by FACS ***(97)***.

Selection of Diels-Alderase ribozymes can be achieved by coupling a DNA library via a PEG linker to anthracene. These genes are compartmentalized in w/o emulsion and the genes are transcribed to RNA. Mg^{2+} and biotin-maleimide are added to the emulsion and allowed to diffuse into the compartments. If compartments contain active Diels-Alderase ribozymes a cycloadduct of biotin-maleimide is generated, thereby biotinylating the ribozyme coding gene. After

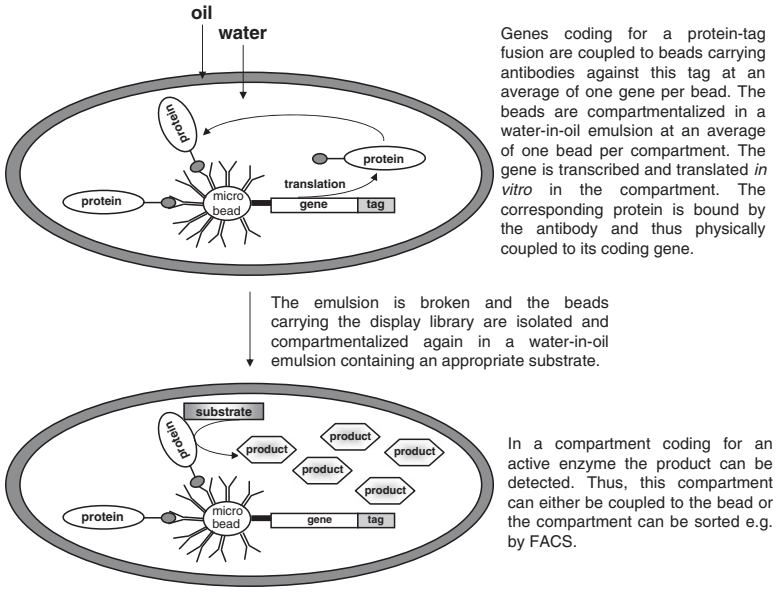

Fig. 35.11. Microbead display

breaking the emulsion, genes coding for active Diels-Alderase ribozymes are bound to streptavidin coated magnetic beads and amplified by PCR *(98)*.

3. Applications

3.1. Applications of Rational Design Approaches

Rational design strategies have successfully been used in the field of protein therapeutics by improving existing products and enabling the development of novel therapeutics. Several designed protein therapeutics are currently on the market (**Table 35.1**, from *(1)*).

Some of the most visible and successful applications of rational biotherapeutic engineering methods have occurred in the field of antibodies. Monoclonal antibodies are widely used as a treatment for a variety of conditions from arthritis to cancer. Some antibody products are already available on the market (**Table 35.2** from *(1)*). Antibody variable domains suffer from stability issues like all proteins. However, because antibodies share a common structural scaffold, rational engineering studies have been able to dissect some of the sequencial and structural determinants of variable region solubility and stability *(99)*.

For example, the best success for immunogenicity reduction has been the humanization of murine antibodies, which was made possible by the high regularity of antibody sequence and structure and close proximity to the human sequence.

Table 35.1. Engineered protein therapeutics.

Name	Family	Company	Indication	Modification
Proleukin® (aldesleukin)	IL-2	Chiron	Cancer	Mutated free cysteine
Betaseron® (interferon beta-1b)	IFN-β	Berlex/Chiron	Multiple sclerosis	Mutated free cysteine
Humalog® (insulin lispro)	Insulin	Eli Lilly	Diabetes	Monomer not hexamer
NovoLog® (insulin aspart)	Insulin	Novo Nordisk	Diabetes	Monomer not hexamer
Lantus® (insulin glargine)	Insulin	Aventis	Diabetes	Precipitates in dermis
Enbrel® (etanercept)	TNF receptor	Immunex/Amgen /Wyeth	Rheumatoid arthritis	Fc fusion
Ontak® (denileukin diftitox)	Diptheria toxin -IL-2	Seragen/Ligand	Cancer	Fusion
PEG-Intron® (peginterferon alfa-2b)	IFN-a	Schering-Plough	Hepatitis	PEGylation
PEGasys® (peginterferon alfa-2a)	IFN-a	Roche	Hepatitis	PEGylation
Neulasta™ (pegfilgrastim)	G-CSF	Amgen	Leukopenia	PEGylation
Oncaspar® (pegaspargase)	Asparaginase	Enzon	Cancer	PEGylation
Aranesp® (darbepoetin α)	Epo	Amgen	Anemia	Additional glycosylation sites
Somavert® (pegvisomant)	Growth hormone	Genentech/ Seragen/ Pharmacia	Acromegaly	PEGylation; binding site mutations

Table 35.2. Engineered antibodies.

Name	Company	Target	Indication	Type
Orthoclone OKT3® (muromonab-CD3)	Ortho Biotech/Johnson & Johnson	CD3	Transplant rejection	Murine
ReoPro® (abciximab)	Centocor/Lilly	GPIIb/IIIa	Restenosis	Chimeric
Rituxan® (rituximab)	IDEC/Genentech	CD20	B-cell non-Hodgkins lymphoma	Chimeric
Simulect® (basiliximab)	Novartis	IL-2R	Transplant rejection	Chimeric
Remicade® (infliximab)	Centocor	TNF-a	Crohn's disease, rheumatoid arthritis	Chimeric
Zevalin® (ibritumomab tiuxetan)	IDEC/Schering AG	CD20	B-cell non-Hodgkins lymphoma	Chimeric
Zenapax® (daclizumab)	PDL/Roche	IL-2R	Transplant rejection	Humanized
Synagis® (palivizumab)	Medlmmune	RSVF protein	Respiratory syncitial virus	Humanized
Herceptin® (trastuzumab)	Genentech	HER2/neu	Breast cancer	Humanized
Mylotarg® (gemtuzumab ozogamicin)	Celltech/Wyeth	CD33	Acute myeloid leukemia	Humanized
Campath® (alemtuzumab)	Millenium/ILEX	CD52	B-cell chronic lymphocytic leukemia	Humanized

3.1.1. Site-Directed Mutagenesis

To increase protein stability the replacement of free cysteines into serines have been introduced into several therapeutic proteins, including granulocyte colony-stimulating factor (G-CSF) and interferon (IFN) β1b, resulting in a longer half-life ***(100,101)***.

The replacement of exposed non-polar for polar residues was applied successfully to the A1 domain of cholera toxin. Of the six variants produced, one retained full biological activity, stability and displayed significant improvement in solubility ***(102)***.

A single chain antibody targeting renal cell carcinoma was altered to increase solubility by adding 5 glutamic acid residues to the C-terminus, thus lowering the pl from 7.5 to 6.1 ***(103)***.

An example of affinity enhancement is the generation of superagonist variants of human thyrotropin (hTSH) by altering the net charge of the protein. The hTSH receptor has a negative charge, and mutations that introduce positively charged residues or replace negatively charged residues in the peripheral loops of hTSH increase activity ***(104,105)***.

4-helix bundle cytokines, including vascular endothelial growth factor (VEGF), hGH and interleukin-6 (IL-6), have been engineered to function as receptor antagonists. Antagonistic VEGF variants were designed as heterodimers, which contain one functional binding site per dimer ***(106)***. An IL-6 superantagonist was generated by selecting mutations that disrupt binding to gp130 and incorporated mutations that resulted in increased affinity for the IL-6 coreceptor ***(107)***. An especially interesting example of a designed cytokine antagonist is Somavert® (pegvisomant, Genentech/Pharmacia), a hGH variant that has recently successfully completed clinical trials for treatment of acromegaly.

Somavert® contains a point mutation at the second of the two receptor binding sites that blocks receptor dimerization upon binding ***(108)***. Eight additional mutations, identified by phage display, that increase the receptor-binding affinity of the first receptor binding site were introduced ***(109)***.

A notable example is the design of constitutively active and inactive integrin I domain variants. The integrin I domains can populate two dominant conformations: an "open" conformation, which can bind intracellular adhesion molecule-1 (ICAM-1), and a "closed" conformation, which has very low affinity for ICAM-1. Springer and coworkers introduced pairs of cysteines that form disulfide bonds compatible with either the closed or open conformation ***(110,111)***, and they designed mutations in the core of the domain that were computationally selected to stabilize the open conformation and disallow the closed state ***(112)***.

Wong and coworkers switched the substrate preference of the 2-deoxyribose-5-phosphate aldolase (DERA) from phosphorylated to nonphosphorylated substrates. The kcat/K_M value for the nonphosphorylated substrate increased 2.5 times for a variant with a single point mutation relative to wild type ***(113,114)***.

Lim and coworkers engineered the active site of magnesium-dependent ribonulclease H to form an active metal-independent enzyme. Replacement of an aspartate and a glutamate residue that interact with the metal ion yields an enzyme that is active in the absence of Mg^{2+}. As a result the pH activity profile is dramatically altered ***(115)***.

3.1.2. Computational Protein Design

Computation interface design was used to fuse two domains of distantly related homing endonucleases (Dmol and Crel), each carrying a recognition site for a specific DNA target half site ***(116)***. The resulting functional chimeric protein combines the two different binding specificities of the parent proteins. The crystal structure of the designed interface confirms the accuracy of the design algorithm. Extending this approach, computational design offers the possibility to create novel interfaces that would go beyond the interaction capabilities of independent modules.

Optimizing the fairly promiscuous calmodulin interface for one of its ligands using a successful computational protein design method, by Shifman and Mayo ***(117)***, resulted in a stable interaction in the nanomolar range that is more specific for the selected ligand. This is the first study showing that computational interface redesign is capable of enhancing the specificity of an interaction.

The study of Havranek and Harbury ***(118)*** describes the development and experimental verification of a novel computational protocol that automatically selects for sequences that prefer desired cognate interaction over alternative partners and conformations (negative design). The experimental results of the formation of homodimeric or heterodimeric coiled coil interfaces verified the predicted specificities in all instances.

Dwyer and Hellinga used computational design for the enzymatic activity in a protein scaffold of known structure. They demonstrate the feasibility of creating new enzymatic activities by introducing mutations at or near the substrate-binding site ***(119)***.

Several authors have compared the behavior of enzymes from thermophilic and mesophilic organisms using MD. The difference in the thermostability

was explained by reduced backbone flexibility of the thermostable enzyme for thioredoxin ***(120)*** and rubredoxin ***(121)***.

The automated design of a novel sequence onto a given protein backbone by computational screening of a combinatorial library was achieved by Dahiyat and Mayo ***(122)***. NMR spectroscopy showed that the resulting structure, a short zinc-finger protein fold, is in excellent agreement with the designed target structure ***(122)***.

Hellinga and coworkers have developed a powerful computational tool DENZYMER to assist in reprogramming the specificities and properties of proteins. This computational technique has been applied to the design of novel variants of *E. coli* periplasmic binding proteins to bind the nonnatural ligands trinitrotoluene, L-lactate, and serotonin with high affinities ***(123–126)***.

The success of the computational design process used in these studies strongly suggests that such techniques will play an increasingly important role in protein engineering, especially when paired with experimental data.

3.2. Applications of Library-Based Design Approaches

3.2.1. Phage Display

Phage display is largely used to screen peptide or antibody libraries for ligands using purified and immobilized molecules *in vitro*, with the aim of stabilizing protein–protein interactions or identifying protein–protein interaction domains ***(127)***. One example of an FDA approved antibody generated by phage display is Adalimumab (HUMIRA), which is used against rheumatoid arthritis ***(128)***. When nonhuman antibodies are used for this approach, immunogenicity can limit application; therefore, the epitope-binding region can be transferred onto the framework of a human IgG antibody. This process is called "humanization."

Another phage display approach is the so called Proside (protein stability increased by directed evolution) approach which directly links thermodynamic stability of a protein with the infectivity of the filamentous phage ***(129)***. In this case, the protein is inserted between two domains of the gene3 protein. Upon incubation of phages with either trypsin or chymotrypsin, only phages harboring well-folded guest-proteins inaccessible to proteolysis remain infective and enter the next round of selection after *E. coli* host amplification. Thus, Proside is independent of interactions with ligands or any specific enzymatic catalysis. Rather, it selects proteins that remain folded upon treatment with proteases, and it is therefore useful for selecting thermodynamic stability.

Moreover, phage display can also be used for *in vivo* screening, so called biopanning. In this approach, phage libraries can be incubated on whole cells targeting a specific receptor or can be used to select for cell-targeting gene therapy vectors ***(130)***. Selection remains more specific *in vivo* than *in vitro* as the target protein remains in its "native" conditions and the ligand is challenged by degradation from cellular proteases and competed by native ligands, both improving stability and specificity. This *in vivo* biopanning can be even expanded to select for peptides that are home to receptors differentially expressed on vasculature organs. These selection procedures identify ligands that target specific vascular beds. In this case phages are intravenously injected and circulate in the blood for a certain time period. Nonspecifically bound phages are removed by washing off the tissue cells. Specific clones that bind to selective vascular beds are recovered by host *E. coli* infection and are amplified for further

selection rounds ***(131)***. Molecular profiles of different diseases can be exploited and lead to the identification of marker genes. Highlighted here are the selection of peptides that "home" to receptors of the lung ***(132)*** or the breast ***(133)***.

3.2.2. Ribosome Display

As for phage display, the list of examples of different approaches is long and applications for ribosome display in biomarker identification, imaging and targeting are likely to evolve further ***(134)***. Examples given here include the identification of semi-synthetic factors that have the potential to exhibit transcriptional activity ***(135)***; DNA-binding proteins were selected out of a zinc finger protein library which now could be used as novel transcription factors. Another example is the selection of MAP-kinase binders. A combinatorial library of ankyrin repeat proteins (DARPins) displayed with ribosomal display, lead to the isolation of binders displaying nanomolar affinities to JNK1, JNK2, and p38 ***(136)***.

3.2.3. Yeast Two-Hybrid System

Aside from the molecular dissection of known interactions and the identification of new potential interacting partners, the evolution of the technique, also mentioned above, enables solving several new problems. The yeast two-hybrid system is the method of choice to study signaling cascades e.g., the Ras/Raf-pathway ***(137)***. In the post-genomic era, efforts are now being made to analyze all known genes and proteins and the way they interact in a whole cell/organism with the aim to establish so-called protein linkage maps. These maps consist of all possible protein interactions that occur in a cell and give insight into the overall cell complexity, serving as a starting point for studies at the systems level. The first genome-wide interaction map was generated for the bacteriophage T7 ***(138)***. In the same manner, a protein interaction map of the yeast strain *S. cervisiae* was established, comprising 69% of the whole proteome ***(139)***. Another screen identified over 5400 interactions of *C. elegans* covering 12% of the genome ***(140)***. Other studies have integrated these data with functional genomics data to derive models for genetic pathways. In addition, two-hybrid screens have recently been used to analyze the human proteome, screening ~7200 full-length Open Reading Frames (ORFs) which identified 2754 protein interactions ***(141)***.

The identified interaction data from *S. cerevisiae, C. elegans, Drosophila*, and from human permits comparison of these interaction maps and helps to predict additional interactions, missing in one system but found in others. Together with a computer assisted confidence score that relates the interaction to a biological significance, it is also possible to lower the amount of false-positive interactions. Hence, statistical scoring systems facilitate integration of datasets.

3.2.4. Protein-Fragment Complementation Assay

PCAs have been used in many different ways for finding, improving and studying protein–protein interactions. Below are examples given for the presented systems.

3.2.4.1. Murine Dihydrofolate Reductase (mDHFR): Pelletier et al. and Arndt et al. (***51,142,143***) used the DHFR assay to study interactions of coiled coil domains, which are naturally abundant oligomerization domains. The aim was to generate stable heterodimeric artificial coiled coils, which can be used as heterodimerization modules for a variety of protein engineering applications ***(144)***. For library design,

outer positions were taken from Jun and Fos and core positions from GCN4. The core-flanking residues were randomized with polar and charged residues to create complementary libraries. These libraries were fused to the DHFR fragments and co-transformed in *E. coli* to select for the best heterodimer. In a further study, Amdt and colleagues selected peptides binding natural targets such as C-Jun, C-Fos and C-Myc (***145–148***).

Mossner et al. ***(149)*** fused a single-chain antibody (scFv) and its antigen to the DHFR-fragments and optimized the linker length and orientation of this system. Replacing the antibody with a library permits use of this assay to select for high affinity antibodies in a robust and easy way.

The DHFR-assay can also be used in mammalian cells when a DHFR-deficient strain is available. Remy et al. studied conformational changes of the erythropoietin receptor upon ligand binding (***150***) and effects of linker length in the assay system. Dimerization was detected by fluorescent-labeled methotrextate, which binds only to the reassembled DHFR. This study demonstrated applicability of the DHFR assay for membrane proteins.

3.2.4.2. Ubiquitin-Based Split-Protein Sensor (USPS): To find new players in the regulation of the galactose pathway in *Saccharomyces cerevisae*, Laser et al. ***(151)*** partially digested the genome of *S. cerevisae* with the restriction enzyme Sau3A and fused the resulted DNA fragments in all three frames to the gene of the α_{ub}-fragment. The ω-fragment of ubiquitin was fused to Gal4p or Tup1p which are known to bind to the Gal1-operon and were used as bait for the library. Laser successfully identified Nhp6 as new interaction partner to both bait-proteins.

Stagljar et al. used USPS to detect interactions between membrane proteins *in vivo* ***(148)***. The cleavage of the fusion protein leads to the release of a transcription factor, which activates a reporter gene in the nucleus.

3.2.4.3. β-Lactamase: The developers of the technique SEER (sequence-enabled reassembly) used first GFP but finally the β-lactamase complementation assay ***(153,154)*** and modules of zinc finger domains to detect specific sequences of DNA. Six or more single zinc fingers were chosen for their combined ability to bind to the DNA sequence of interest. Zinc fingers recognizing the 3′-half of the target were fused to the first fragment of the reporter, and the other fingers to the second fragment. Only if all zinc fingers bind to the DNA in the correct orientation, the β-lactamase can reassemble and become active. β-lactamase proved superior to GFP for this application, because it reassembled and folded much faster than GFP and its enzymatic signal amplification allowed detection of fewer target sites. With this method, the authors were able to specifically detect a target DNA sequence in a complex mixture.

3.2.4.4. Luciferase: Massoud et al. applied the luciferase PCA to study homodimeric protein–protein interactions in mammalian cells and living mice ***(155)***. They used a split synthetic humanized renilla luciferase (hRLUC) to visualize and quantify the dimerization of herpes simplex virus type 1 thymidine kinase (TK1). Splitting hRLUC resulted in two fragments $\Delta\alpha = 229$ residues and $\Delta\omega = 82$ residues ***(66)***. 293T cells expressing the TK chimeras were implanted in mice and mock-transfected cells as negative controls at another site of the same mice. The luciferase substrate coelenterazine was injected into the mouse tails and the bioluminescence signal recorded by a cooled charged coupled device (CCD) camera. They also tested which order and orientation the luciferase fragments and the TK monomers resulted in the

highest bioluminescence. It was possible to locate and quantify luciferase activity with high sensivity in a living subject, which makes this system a valuable tool for studying protein–protein interactions in animals.

3.2.4.5. Green Fluorescent Protein: Hu and Kerppola visualized differential protein interactions in the same cell by multicolored BiFc ***(156)***. For this, they performed PCAs with various combininations of four GFP variants (GFP, YFP, CFP, BFP) with two fragmentation sites at aa155 or aa173. The protein fragments were fused to the bZIP domains of Fos and Jun (bFos and bJun) and screened for fluorescence in mammalian cells. The study characterized 12 bimolecular fluorescent complexes with 7 spectral classes, thus providing an impressive set to analyse complex protein interaction networks in living mammalian cells. In the successful complex forming combinations YFP was most prominent.

In 2004, Bracha-Drori et al. and Walter et al. ***(157,158)*** adapted BiFc for monitoring protein interactions in the nucleus and the cytoplasm of living plants. Using YFP they could visualize protein interactions and show that BiFc occurred in the correct compartement of the plant cells.

3.2.5. Cell-Surface Display

Most examples using cell surface display are from the *de novo* selection and improvement of antibodies.

3.2.5.1. Bacterial Display: Christman et al. applied a bacterial display system for epitope mapping of monospecific antibodies ***(159)***. A random library of gene fragments of the classical swine fever virus (CSFV) envelope protein E^{rns} was generated by DNase I digestion. For bacterial surface display, the fragments were fused to a carboxyterminal truncated intimin ***(160)***, an *E. coli* adhesin, which is located far enough from the outer membrane' lipopolysaccharide layer to be sufficiently accessible to the tested antibodies. The epitope-presenting *E. coli* cells were incubated with specific antibodies produced in mice. A biotinylated anti-mouse antibody was used as a secondary antibody and detected by streptavidin conjugated to the fluorescent dye R-phycoerythrin. Cells were sorted by fluorescence-activated cell sorting (FACS), and FACS-positive clones analyzed for their epitope sequence. Eight of eleven clones presented a carboxy-terminal fragment of E^{rns} on their surface, three displayed other regions.

Metal-binding peptides could become a powerful tool in cleaning the ecosystem from heavy metals and radionuclides. Kjaergaard et al. ***(161)*** screened a library of approximate 4×10^6 clones for the ability to bind Zn^{2+}. The library was fused to the adhesin FimH, a component of the fimbrial organelle of *E. coli.* After several rounds of selection of peptide-displaying cells against Zn^{2+}-nitrilotriacetic acid beads, binding clones were analyzed. From those 15 clones, no consensus sequence could be derived but all carried at least one histidine. Data bank research revealed no noteworthy sequence similarities, suggesting that novel Zn^{2+}-binding peptides were selected.

3.2.5.2. Yeast Surface Display: Calmodulin is a highly conserved protein in mammals that is part of a variety of signaling pathways (reviewed in ***(162)***). It contains four Ca^{2+}-binding sites and undergoes structural changes upon binding Ca^{2+} ions. There are only few reports about monoclonal antibodies against calmodulin; due to its high conservation it was difficult to deal with the self tolerance mechanism of the immunsystem of the antibody-producing

animal system. Yeast display offered an alternative way to the classic method. Feldhaus et al. ***(163)*** selected from a human nonimmune scFv library displayed on the surface of yeast new antibodies against calmodulin. Sequences were further improved by error-prone PCR to yield specific high affinity binders to the two different conformations of calmodulin. The antibody optained for Ca^{2+}-calmodulin was a scFc with an equilibrium dissociation constant (K_d) of 800 pM and more than 1,000-fold higher specificity for this conformation relative to the Ca^{2+}-free form of the protein. For the latter a single-domain antibody (dAb) was selected with a K_d of 1 nM and more than 300-fold higher specificity relativ to Ca^{2+}-calmodulin.

Red Sea Bream Iridovirus (RSIV) infects, amongst others, cultured and ornamental fishes in Japan ***(164)*** and can thus lead to severe damage to the economy. Tamaru et al. ***(165)*** successfully expressed the antigen 380R on the surface of yeast. This may lead to production of an oral vaccine against RSIV.

Finding a binding protein is generally not enough to stimulate the desired cellular response. This response is often the result of multiple amplification events following receptor activation. High valency of receptor interactions are needed, also with complementary molecules on other cells. For example, T-cell activation requires a high number of interactions between T-cell receptors (TCR) and antigen presenting cells. Cho et al. ***(166)*** presented high levels of a ligand on the surface of yeast to target T cells. This lead to the necessary clustering of TCRs on the surface and to activation of T cells, as demonstrated by increased levels of CD25 and CD69 and a decreased number of TCRs on the surface. The authors also demonstrated the ability to activate T cells in the presence of high concentrations of nonpresenting yeast, suggesting that the system is applicable to library based approaches. More applications for yeast display are reviewed by Kondo and Ueda (***167***).

3.2.5.3. Viral Display: Buchholz et al. ***(85)*** applied a viral display system to the selection of protease cleavage sites. They expressed the epidermal growth factor (EGF) on the surface of murine leukemia viruses, linked via a seven-residue linker to the envelope glycoprotein. The virus was propagated on EGF receptor-poor cells without loss of the displayed EGF. In contrast, it did not replicate on EGF receptor-rich cells, because the EGF-displaying viruses were intercepted by the EGF receptors. The authors randomized the seven-residue linker and let the viruses propagate on EGF receptor-rich HT1080 cells. Only viruses whose EGF was proteolytically cleaved from the viral surface were able to infect cells and to replicate. After three passages of selection resulting sequences were all arginine-rich and matched the consensus sequence for furin-like proteases.

To enhance gene delivery to target cells, Raty et al. ***(168)*** altered the surface of baculo viruses to display avidin, the constructed virus was named Baavi. Avidin is highly positively charged and was therefore expected to improve cell transfection. In this study, Baavi achieved a five-fold increase in transduction efficiency in rat malignant glioma cells and a 26-fold increase in rabbit aortic smooth muscle cells. Even higher transduction efficiency was shown for biotinylated cells.

3.2.5.4. Mammalian Cell Display: Riddle et al. ***(169)*** mimicked the natural binding of antibodies to tumor cells by displaying the Fc portion of the murine IgG2a heavy chain (IgFc) on the surface of tumor cells in an orientation where

its C-terminus pointed away from the surface. In this way they hoped to activate an immune response against the tumor cells equivalent to an antibody-based approach, which showed some inherent problems like poor penetration of the antibody in the tumor and the need for tumor-specific antigens. In a first *in vitro* experiment, they displayed IgFc on the surface of B16 melanoma cells. Indeed, these cells were specifically recognized and rapidly lysed by natural killer cells. Subsequent *in vivo* data demonstrated that tumor formation was severely delayed. Direct intratumoral injection of adenoviral vectors expressing IgFc led to total clearance of the tumor cells but did not prevent metastasis or led to antitumor immunization. For this, an additional immunostimulatory signal was needed, achieved here by coexpression of heat shock protein 70 (hsp70).

For many years, antibodies have been successfully selected and matured by phage display ***(170–173)***, bacterial display ***(174–176)***, yeast display ***(177,178)*** and ribosome display ***(31,179,180)***, but these techniques are limited by problems with protein folding, posttranslational modification and codon usage. Ho et al. ***(88)*** for the first time used a mammalian display system for the purpose of antibody maturation. They fused the anti-CD33 scFv and the high-affinity derivative HA22 scFv to the transmembrane domain of human platelet-derived growth factor receptor (PDGFR) and displayed the chimeric protein on human embryonic kidney (HEK) 293T cells. They were able to achieve a 240-fold enrichment of the high-affinity variant relative to the wt scFv. Furthermore, Ho ***(88)*** selected an antibody with even higher affinity from a scFv library with randomized intrinsic antibody hot spots.

3.2.6. In Vitro Compartmentalization Methods

DNA polymerase from *Thermus aquaticus* (*Taq*) is one of the most important enzymes in modern biotechnology. The various DNA amplification and modification techniques that are used often have requirements that are difficult to achieve with polymerases on the market. Thus, it is important and profitable to adjust polymerases to the conditions needed for special applications till the aim of a "gold standard" polymerase is achieved ***(181)***.

Ghadessy et al. ***(93)*** used three cycles of CSR to select for *Taq* DNA polymerases with 11-fold higher thermostability than wild-type *Taq* and increased resistance to the inhibitor heparin. A few years later Ghadessy et al. ***(182)*** started from ramdomly mutated *Taq* clones and selected these by CSR for efficient mismatch extension. In three cycles of CSR they enriched *Taq* polymerase with the general ability to extend 3′ mispaired termini. This "unfussy" *Taq* promiscuously extended mismatches and was able to incorporate noncanonical substrates with high turnover, processivity, and fidelity.

Bacterial phosphotriesterase (PTE) has the ability to degrade pesticides and nerve agents like soman, sarin, and VX and thus is very interesting for bioremediation or disarmament of chemical weaponry. Griffiths et al. ***(183)*** used six rounds the microbead display IVC to generate an extremely fast phophotriesterase with 63 times higher k_{cat} than the wild-type enzyme. For selection of active enzymes the substrate was coupled to caged biotin which was afterwards uncaged by UV light. Thus, the product is coupled to the straptavidin coated beads and thus is linked to the coding gene. After breaking the emulsion, the product was detected by an antiproduct antibody which could be detected by a fluorescence labeled secondary antibody. Sorting was done by FACS.

Sequence recognition of enzymes is poorly understood and thus extremely challenging. Methylases as well as endonucleases are valuable tools in biotechnology. M.HaeIII methytransferase methylates the first cytosine after the second guanine of the canonical sequence 5′ GGCC 3′ but it is known that there is a promiscuous methylation at other sites like AGCC at lower rates ***(184)***. Cohen et al. ***(184)*** altered the sequence preference of HaeIII methyltransferase by use of IVC from GGCC to AGCC and additionally this mutant also methylates at a low rate three other sites (AGCC, CGCC and GGCC) but discriminates as efficiently as the wild type enzyme against other sites. A library of mutated HaeIII genes was translated *in vitro* using IVC. Active enzymes methylated their genes and unmethylated genes were digested with a suitable enzyme NheI. The undigested genes were amplified and subjected to new rounds of IVC.

Ribozymes are catalytically active RNAs which ligate two RNAs which are aligned to a template by a reaction similar to enzymes which synthesize RNA ***(138)***. Levy et al. (***184***) selected by microbead display IVC a ligase ribozyme capable to act *trans* on oligonucleotide substrates after two rounds of IVC. The ribozyme coding DNA was coupled to the beads together with an RNA oligonucleotide serving as substrate. DNA coding for functional ribozymes are able to ligate a tagged RNA to the coupled substrate RNA molecule can be selected by antiproduct antibodies. These primary antibodies were likewise detected by fluorescence labeled secondary antibodies, and the beads with DNA coding for active *trans*-acting ligase ribozymes were sorted by FACS.

References

1. Marshall SA, Lazar GA, Chirino AJ, Desjarlais JR (2003) Rational design and engineering of therapeutic proteins. Drug Discov Today 8:212–221
2. Hellinga HW (1997) Rational protein design: combining theory and experiment. Proc Natl Acad Sci USA 94:10015–10017
3. Creveld LD (2001) Molecular dynamics simulations in rational protein design: stabilization of fusarium solani pisi cutinase against anionic surfactants. University of Groningen
4. Balland A, Courtney M, Jallat S, Tessier LH, Sondermeyer P, de la Salle H, Harvey R, Degryse E, Tolstoshev P (1985) Use of synthetic oligonucleotides in gene isolation and manipulation. Biochimie 67:725–736
5. Garvey EP, Matthews CR (1990) Site-directed mutagenesis and its application to protein folding. Biotechnology 14:37–63
6. Wagner CR, Benkovic SJ (1990) Site directed mutagenesis: a tool for enzyme mechanism dissection. Trends Biotechnol 8:263–270
7. Kammann M, Laufs J, Schell J, Gronenborn B (1989) Rapid insertional mutagenesis of DNA by polymerase chain reaction (PCR). Nucleic Acids Res 17:5404
8. Antikainen NM, Martin SF (2005) Altering protein specificity: techniques and applications. Bioorg Med Chem 13:2701–2716
9. Desjarlais JR, Clarke ND (1998) Computer search algorithms in protein modification and design. Curr Opin Struct Biol 8:471–475
10. Voigt CA, Gordon DB, Mayo SL (2000) Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. J Mol Biol 299:789–803
11. Fishman GS (1995) Monte Carlo: concepts, algorithms, and applications. Springer Verlag, New York
12. Metropolis NaU, S (1949) The Monte Carlo method. J Am Stat As 44:335

13. Desmet J, de Maeyer M, Hazes B, Lasters I (1992) The dead-end elimination theorem and its use in protein side-chain positioning. Nature 356:539–542
14. Gordon DB, Marshall SA, Mayo SL (1999) Energy functions for protein design. Curr Opin Struct Biol 9:509–513
15. Pokala N, Handel TM (2001) Review: protein design – where we were, where we are, where we're going. J Struct Biol 134:269–281
16. Street AG, Mayo SL (1999) Computational protein design. Structure 7:R105–109
17. van Gunsteren WF, Mark AE (1992) On the interpretation of biochemical data by molecular dynamics computer simulation. Fur J Biochem 204:947–961
18. Smith GP, Petrenko VA (1997) Phage Display. Chem Rev 97:391–410
19. Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228:1315–1317
20. Kehoe JW, Kay BK (2005) Filamentous phage display in the new millennium. Chem Rev 105:4056–4072
21. Willats WG (2002) Phage display: practicalities and prospects. Plant Mol Biol 50:837–854
22. Jung S, Honegger A, Plückthun A (1999) Selection for improved protein stability by phage display. J Mol Biol 294:163–180
23. Jaye DL, Geigerman CM, Fuller RE, Akyildiz A, Parkos CA (2004) Direct fluorochrome labeling of phage display library clones for studying binding specificities: applications in flow cytometry and fluorescence microscopy. J Immunol Methods 295:119–127
24. Steiner D, Forrer P, Stumpp MT, Pluckthun A (2006) Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. Nat Biotechnol 24:823–831
25. Arndt KM, Jung S, Krebber C, Plückthun A (2000) Selectively infective phage technology. Methods Enzymol 328:364–388
26. Jung S, Arndt KM, Müller KM, Plückthun A (1999) Selectively infective phage (SIP) technology: scope and limitations. J Immunol Methods 231:93–104
27. Jiang J, Abu-Shilbayeh L, Rao VB (1997) Display of a PorA peptide from Neisseria meningitidis on the bacteriophage T4 capsid surface. Infect Immun 65:4770–4777
28. Danner S, Belasco JG (2001) T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc Natl Acad Sci USA 98:12954–12959
29. Mattheakis LC, Bhatt RR, Dower WJ (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. Proc Natl Acad Sci USA 91:9022–9026
30. Wilson DS, Keefe AD, Szostak JW (2001) The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci USA 98:3750–3755
31. Hanes J, Jermutus L, Weber-Bornhauser S, Bosshard HR, Plückthun A (1998) Ribosome display efficiently selects and evolves high-affinity antibodies *in vitro* from immune libraries. Proc Natl Acad Sci USA 95:14130–14135
32. Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature 340:245–246
33. Ma J, Ptashne M (1988) Converting a eukaryotic transcriptional inhibitor into an activator. Cell 55:443–446
34. Walhout AJ, Vidal M (2001) High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods 24:297–306
35. Shih HM, Goldman PS, DeMaggio AJ, Hollenberg SM, Goodman RH, Hoekstra MF (1996) A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP. Proc Natl Acad Sci USA 93:13896–13901
36. Vidal M, Brachmann RK, Fattaey A, Harlow E, Boeke JD (1996) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. Proc Natl Acad Sci USA 93:10315–10320

37. Vidal M, Braun P, Chen E, Boeke JD, Harlow E (1996) Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. Proc Natl Acad Sci USA 93:10321–10326
38. Aronheim A (1997) Improved efficiency sos recruitment system: expression of the mammalian GAP reduces isolation of Ras GTPase false positives. Nucleic Acids Res 25:3373–3374
39. Aronheim A, Zandi E, Hennemann H, Elledge SJ, Karin M (1997) Isolation of an AP-1 repressor by a novel method for detecting protein/protein interactions. Mol Cell Biol 17:3094–3102
40. Köhler F, Müller KM (2003) Adaptation of the Ras-recruitment system to the analysis of interactions between membrane-associated proteins. Nucleic Acids Res 31:e28
41. Osborne MA, Dalton S, Kochan JP (1995) The yeast tribrid system-genetic detection of trans-phosphorylated ITAM-SH2-interactions. Biotechnology (NY) 13:1474–1478
42. Zhang J, Lautar S (1996) A yeast three-hybrid method to clone ternary protein complex components. Anal Biochem 242:68–72
43. Licitra EJ, Liu JO (1996) A three-hybrid system for detecting small lig-and-protein receptor interactions. Proc Natl Acad Sci USA 93:12817–12821
44. Tafelmeyer P, Johnsson N, Johnsson K (2004) Transforming a (beta/alpha)8-barrel enzyme into a split-protein sensor through directed evolution. Chem Biol 11:681–689
45. Johnsson N, Varshavsky A (1994) Split ubiquitin as a sensor of protein interactions *in vivo*. Proc Natl Acad Sci USA 91:10340–10344
46. Pelletier JN, Campbell-Valois FX, Michnick SW (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. Proc Natl Acad Sci USA 95:12141–12146
47. Stammers DK, Champness JN, Beddell CR, Dann JG, Eliopoulos E, Geddes AJ, Ogg D, North AC (1987) The structure of mouse L1210 dihydrofolate reductase-drug complexes and the construction of a model of human enzyme. FEBS Lett 218:178–184
48. Blakley RL (1984). Folates and pterins: chemistry and biochemistry of Folates. In: Blakley RL, Cenkovic S (eds), Folates and pterins: chemistry and biochemistry of folates, Vol. 1. John Wiley & Sons, New York, pp 191–253
49. Appleman JR, Prendergast N, Delcamp TJ, Freisheim JH, Blakley RL (1988) Kinetics of the formation and isomerization of methotrexate complexes of recombinant human dihydrofolate reductase. J Biol Chem 263:10304–10313
50. Buchwalder A, Szadkowski H, Kirschner K (1992) A fully active variant of dihydrofolate reductase with a circularly permuted sequence. Biochemistry 31:1621–1630
51. Pelletier JN, Arndt KM, Plückthun A, Michnick SW (1999) An *in vivo* library-versus-library selection of optimized protein–protein interactions. Nat Biotechnol 17:683–690
52. Rossi F, Charlton CA, Blau HM (1997) Monitoring protein–protein interactions in intact eukaryotic cells by beta-galactosidase complementation. Proc Natl Acad Sci USA 94:8405–8410
53. Rossi FM, Blakely BT, Blau HM (2000) Interaction blues: protein interactions monitored in live mammalian cells by beta-galactosidase complementation. Trends Cell Biol 10:119–122
54. Ullmann A, Perrin D, Jacob F, Monod J (1965) [Identification, by *in vitro* complementation and purification, of a peptide fraction of *Escherichia coli* beta-galactosidase]. J Mol Biol 12:918–923
55. Villarejo M, Zamenhof PJ, Zabin I (1972) Beta-galactosidase. *In vivo* - complementation. J Biol Chem 247:2212–2216
56. Mohler WA, Blau HM (1996) Gene expression and cell fusion analyzed by lacZ complementation in mammalian cells. Proc Natl Acad Sci USA 93:12423–12427

57. Beckwith JR (1964) A deletion analysis of the Lac operator region in *Escherichia coli*. J Mol Biol 78:427–430
58. Prentki P (1992) Nucleotide sequence of the classical lacZ deletion delta M15. Gene 122:231–232
59. Galarneau A, Primeau M, Trudeau LE, Michnick SW (2002) Beta-lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions. Nat Biotechnol 20:619–622
60. Wehrman T, Kleaveland B, Her JH, Balint RF, Blau HM (2002) Protein–protein interactions monitored in mammalian cells via complementation of beta - lactamase enzyme fragments. Proc Natl Acad Sci USA 99:3469–3474
61. Zlokarnik G, Negulescu PA, Knapp TE, Mere L, Burres N, Feng L, Whitney M, Roemer K, Tsien RY (1998) Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. Science 279:84–88
62. Ambler RP, Coulson AF, Frere JM, Ghuysen JM, Joris B, Forsman M, Levesque RC, Tiraby G, Waley SG (1991) A standard numbering scheme for the class A beta-lactamases. Biochem J 276 (Pt 1) 269–270
63. Farzaneh S, Chaibi EB, Peduzzi J, Barthelemy M, Labia R, Blazquez J, Baquero F (1996) Implication of lle-69 and Thr-182 residues in kinetic characteristics of IRT-3 (TEM-32) beta-lactamase. Antimicrob Agents Chemother 40:2434–2436
64. Paulmurugan R, Umezawa Y, Gambhir SS (2002) Noninvasive imaging of protein–protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. Proc Natl Acad Sci USA 99:15608–15613
65. Massoud TF, Gambhir SS (2003) Molecular imaging in living subjects: seeing fundamental biological processes in a new light. Genes Dev 17:545–580
66. Paulmurugan R, Gambhir SS (2003) Monitoring protein–protein interactions using split synthetic renilla luciferase protein-fragment-assisted complementation. Anal Chem 75:1584–1589
67. Remy I, Michnick SW (2006) A highly sensitive protein–protein interaction assay based on Gaussia luciferase. Nat Methods 3:977–979
68. Ghosh I, AD, H, Regan L (2000) Antiparallel Leucine zipper-directed protein reassembly: application to the green fluorescent protein. J Am Chem Soc 122:5658–5659
69. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ (1996) Crystal structure of the Aequorea victoria green fluorescent protein. Science 273:1392–1395
70. Abedi MR, Caponigro G, Kamb A (1998) Green fluorescent protein as a scaffold for intracellular presentation of peptides. Nucleic Acids Res 26:623–630
71. Hu CD, Chinenov Y, Kerppola TK (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell 9:789–798
72. Lee SY, Choi JH, Xu Z (2003) Microbial cell-surface display. Trends Biotechnol 21:45–52
73. Lunder M, Bratkovio T, Doljak B, Kreft S, Urleb U, Strukelj B, Plazar N (2005) Comparison of bacterial and phage display peptide libraries in search of target-binding motif. Appl Biochem Biotechnol 127:125–131
74. Romanos MA, Scorer CA, Clare JJ (1992) Foreign gene expression in yeast: a review. Yeast 8:423–488
75. Gellissen G, Melber K, Janowicz ZA, Dahlems UM, Weydemann U, Piontek M, Strasser AW, Hollenberg CP (1992) Heterologous protein production in yeast. Antonie Van Leeuwenhoek, 62:79–93
76. Schreuder MP, Brekelmans S, van den Ende H, Klis FM (1993) Targeting of a heterologous protein to the cell wall of Saccharomyces cerevisiae. Yeast, 9:399–409
77. Boder ET, Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol 15:553–557

78. Kieke MC, Shusta EV, Boder ET, Teyton L, Wittrup KD, Kranz DM (1999) Selection of functional T cell receptor mutants from a yeast surface-display library. Proc Natl Acad Sci USA 96:5651–5656
79. Richman SA, Healan SJ, Weber KS, Donermeyer DL, Dossett ML, Greenberg PD, Allen PM, Kranz DM (2006) Development of a novel strategy for engineering high-affinity proteins by yeast display. Protein Eng Des Sel 19:255–264
80. Oker-Blom C, Airenne KJ, Grabherr R (2003) Baculovirus display strategies: Emerging tools for eukaryotic libraries and gene delivery. Brief Funct Genomic Proteomic 2:244–253
81. Smith GE, Summers MD, Fraser MJ (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. Mol Cell Biol 3:2156–2165
82. O'Reilly DR, Miller LK, Luckow VA (1992). Baculovirus expression vectors: a laboratory manual 1st ed, W.H. Freeman and Company, New York
83. Boublik Y, Di Bonito P, Jones IM (1995) Eukaryotic virus display: engineering the major surface glycoprotein of the Autographa californica nuclear polyhedrosis virus (AcNPV) for the presentation of foreign proteins on the virus surface. Biotechnology (NY) 13:1079–1084
84. Ernst W, Grabherr R, Wegner D, Borth N, Grassauer A, Katinger H (1998) Baculovirus surface display: construction and screening of a eukaryotic epitope library. Nucleic Acids Res 26:1718–1723
85. Buchholz CJ, Peng KW, Morling FJ, Zhang J, Cosset FL, Russell SJ (1998) *In vivo* selection of protease cleavage sites from retrovirus display libraries. Nat Biotechnol 16:951–954
86. Urban JH, Schneider RM, Compte M, Finger C, Cichutek K, Alvarez-Vallina L, Buchholz CJ (2005) Selection of functional human antibodies from retroviral display libraries. Nucleic Acids Res 33:e35
87. Wolkowicz R, Jager GC, Nolan GR (2005) A random peptide library fused to CCR5 for selection of mimetopes expressed on the mammalian cell surface via retroviral vectors. J Biol Chem 280:15195–15201
88. Ho M, Nagata S, Pastan I (2006) Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. Proc Natl Acad Sci USA 103:9637–9642
89. Rothe A, Surjadi RN, Power BE (2006) Novel proteins in emulsions using *in vitro* compartmentalization. Trends Biotechnol 24:587–592
90. Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387–390
91. Leamon JH, Link DK, Egholm M, Rothberg JM (2006) Overview: methods and applications for droplet compartmentalization of biology. Nat Methods 3:541–543
92. Miller OJ, Bernath K, Agresti JJ, Amitai G, Kelly BT, Mastrobattista E, Taly V, Magdassi S, Tawfik DS, Griffiths AD (2006) Directed evolution by *in vitro* compartmentalization. Nat Methods 3:561–570
93. Ghadessy FJ, Ong JL, Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. Proc Natl Acad Sci USA 98:4552–4557
94. Ong JL, Loakes D, Jaroslawski S, Too K, Holliger P (2006) Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. J Mol Biol 361:537–550
95. Doi N, Kumadaki S, Oishi Y, Matsumura N, Yanagawa H (2004) *In vitro* selection of restriction endonucleases by *in vitro* compartmentalization. Nucleic Acids Res 32:e95
96. Sepp A, Tawfik DS, Griffiths AD (2002) Microbead display by *in vitro* compartmentalisation: selection for binding using flow cytometry. FEBS Lett 532:455–458
97. Bernath K, Hai M, Mastrobattista E, Griffiths AD, Magdassi S, Tawfik DS (2004) *In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. Anal Biochem 325:151–157

98. Agresti JJ, Kelly BT, Jaschke A, Griffiths AD (2005) Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using *in vitro* compartmentalization. Proc Natl Acad Sci USA 102:16170–16175
99. Worn A, Pluckthun A (2001) Stability engineering of antibody single-chain Fv fragments. J Mol Biol 305:989–1010
100. Lin L (1998) Betaseron. Dev Biol Stand 96:97–104
101. Arakawa T, Prestrelski SJ, Narhi LO, Boone TC, Kenney WC (1993) Cysteine 17 of recombinant human granulocyte-colony stimulating factor is partially solvent-exposed. J Protein Chem 12:525–531
102. Culajay JF, Blaber SI, Khurana A, Blaber M (2000) Thermodynamic characterization of mutants of human fibroblast growth factor 1 with an increased physiological half-life. Biochemistry 39:7153–7158
103. Tan PH, Chu V, Stray JE, Hamlin DK, Pettit D, Wilbur DS, Vessella RL, Stayton PS (1998) Engineering the isoelectric point of a renal cell carcinoma targeting antibody greatly enhances scFv solubility. Immunotechnology 4:107–114
104. Weintraub BD, Szkudlinski MW (1999) Development and *in vitro* characterization of human recombinant thyrotropin. Thyroid 9:447–450
105. Grossmann M, Leitolf H, Weintraub BD, Szkudlinski MW (1998) A rational design strategy for protein hormone superagonists. Nat Biotechnol 16:871–875
106. Siemeister G, Schimer M, Reusch P, Barleon B, Marme D, Martiny-Baron G (1998) An antagonistic vascular endothelial growth factor (VEGF) variant inhibits VEGF-stimulated receptor autophosphorylation and proliferation of human endothelial cells. Proc Natl Acad Sci USA 95:4625–4629
107. Savino R, Ciapponi L, Lahm A, Demartis A, Cabibbo A, Toniatti C, Delmastro P, Altamura S, Ciliberto G (1994) Rational design of a receptor super-antagonist of human interleukin-6. Embo J 13:5863–5870
108. Fub G, Cunningham BC, Fukunaga R, Nagata S, Goeddel DV, Wells JA (1992) Rational design of potent antagonists to the human growth hormone receptor. Science 256:1677–1680
109. Cunningham BC, Lowman HB, Wells JA, Clark RG, Olson K, Fuh GG (1988) Human Growth Hormone Variants, USA
110. Lu C, Shimaoka M, Ferzly M, Oxvig C, Takagi J, Springer TA (2001) An isolated, surface-expressed I domain of the integrin alphaLbeta2 is sufficient for strong adhesive function when locked in the open conformation with a disulfide bond. Proc Natl Acad Sci USA 98:2387–2392
111. Shimaoka M, Lu C, Palframan RT, von Andrian UH, McCormack A, Takagi J, Springer TA (2001) Reversibly locking a protein fold in an active conformation with a disulfide bond: integrin alphaL I domains with high affinity and antagonist activity *in vivo*. Proc Natl Acad Sci USA 98:6009–6014
112. Shimaoka M, Shifman JM, Jing H, Takagi J, Mayo SL, Springer TA (2000) Computational design of an integrin I domain stabilized in the open high affinity conformation. Nat Struct Biol 7:674–678
113. DeSantis G, Liu J, Clark DP, Heine A, Wilson IA, Wong CH (2003) Structure-based mutagenesis approaches toward expanding the substrate specificity of D-2-deoxyribose-5-phosphate aldolase. Bioorg Med Chem 11:43–52
114. Heine A, DeSantis G, Luz JG, Mitchell M, Wong CH, Wilson IA (2001) Observation of covalent intermediates in an enzyme mechanism at atomic resolution. Science 294:369–374
115. Babu CS, Dudev T, Casareno R, Cowan JA, Lim C (2003) A combined experimental and theoretical study of divalent metal ion selectivity and function in proteins: application to E. coli ribonuclease H1. J Am Chem Soc 125:9318–9328
116. Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL (2002) Design, activity, and structure of a highly specific artificial endonuclease. Mol Cell 10:895–905

117. Shifman JM, Mayo SL (2002) Modulating calmodulin binding specificity through computational protein design. J Mol Biol 323:417–423
118. Havranek JJ, Harbury PB (2003) Automated design of specificity in molecular recognition. Nat Struct Biol 10:45–52
119. Dwyer MA, Looger LL, Hellinga HW (2004) Computational design of a biologically active enzyme. Science 304:1967–1971
120. Pedone EM, Bartolucci S, Rossi M, Saviano M (1998) Computational analysis of the thermal stability in thioredoxins: a molecular dynamics approach. J Biomol Struct Dyn 16:437–446
121. Lazaridis T, Lee I, Karplus M (1997) Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. Protein Sci 6:2589–2605
122. Dahiyat BI, Mayo SL (1997) De novo protein design: fully automated sequence selection. Science 278:82–87
123. Marvin JS, Hellinga HW (2001) Conversion of a maltose receptor into a zinc biosensor by computational design. Proc Natl Acad Sci USA 98:4955–4960
124. de Lorimier RM, Smith JJ, Dwyer MA, Looger LL, Sali KM, Paavola CD, Rizk SS, Sadigov S, Conrad DW, Loew L, Hellinga HW (2002) Construction of a fluorescent biosensor family. Protein Sci 11:2655–2675
125. Yang W, Jones LM, Isley L, Ye Y, Lee HW, Wilkins A, Liu ZR, Hellinga HW, Malchow R, Ghazi M, Yang JJ (2003) Rational design of a calcium-binding protein. J Am Chem Soc 125:6165–6171
126. Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. Nature 423:185–190
127. Samoylova TI, Morrison NE, Globa LP, Cox NR (2006) Peptide phage display: opportunities for development of personalized anti-cancer strategies. Anticancer Agents Med Chem 6:9–17
128. Bongartz T, Sutton AJ, Sweeting MJ, Buchan I, Matteson EL, Montori V (2006) Anti-TNF antibody therapy in rheumatoid arthritis and the risk of serious infections and malignancies: systematic review and meta-analysis of rare harmful effects in randomized controlled trials. Jama 295:2275–2285
129. Sieber V, Pluckthun A, Schmid (1998) Selecting proteins with improved stability by a phage-based method. Nat Biotechnol 16:955–960
130. Barry MA, Dower WJ, Johnston SA (1996) Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. Nat Med 2:299–305
131. Sergeeva A, Kolonin MG, Molldrem JJ, Pasqualini R, Arap W (2006) Display technologies: application for the discovery of drug and gene delivery agents. Adv Drug Deliv Rev 58:1622–1654
132. Oh Y, Mohiuddin I, Sun Y, Putnam JB, Jr, Hong WK, Arap W, Pasqualini R (2005) Phenotypic diversity of the lung vasculature in experimental models of metastases. Chest 128:596S–600S
133. Essler M, Ruoslahti E (2002) Molecular specialization of breast vasculature: a breast-homing phage-displayed peptide binds to aminopeptidase P in breast vasculature. Proc Natl Acad Sci USA 99:2252–2257
134. Rothe A, Hosse RJ, Power BE (2006) Ribosome display for improved biotherapeutic molecules. Expert Opin Biol Ther 6:177–187
135. Ihara H, Mie M, Funabashi H, Takahashi F, Sawasaki T, Endo Y, Kobatake E (2006) *In vitro* selection of zinc finger DNA-binding proteins through ribosome display. Biochem Biophys Res Commun 345:1149–1154
136. Amstutz P, Koch H, Binz HK, Deuber SA, Pluckthun A (2006) Rapid selection of specific MAP kinase-binders from designed ankyrin repeat protein libraries. Protein Eng Des Sel 19:219–229
137. Khosravi-Far R, White MA, Westwick JK, Solski PA, Chrzanowska-Wodnicka M, Van Aelst L, Wigler MH, Der CJ (1996) Oncogenic Ras activation of Raf/

mitogen-activated protein kinase-independent pathways is sufficient to cause tumorigenic transformation. Mol Cell Biol 16:3923–3933
138. Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. Science 261:1411–1418
139. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403:623–627
140. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M (2004) A map of the interactome network of the metazoan C. elegans. Science 303:540–543
141. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lin J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437:1173–1178
142. Arndt KM, Pelletier JN, Müller KM, Alber T, Michnick SW, Plückthun A (2000) A heterodimeric coiled-coil peptide pair selected *in vivo* from a designed library-versus-library ensemble. J Mol Biol 295:627–639
143. Arndt KM, Pelletier JN, Müller KM, Plückthun A, Alber T (2002) Comparison of *in vivo* selection and rational design of heterodimeric coiled coils. Structure 10:1235–1248
144. Arndt KM, Müller KM, Plückthun A (2001) Helix-stabilized Fv (hsFv) antibody fragments: substituting the constant domains of a Fab fragment for a heterodimeric coiled-coil domain. J Mol Biol 312:221–228
145. Mason JM, Schmitz MA, Müller KM, Arndt KM (2006) Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. Proc Natl Acad Sci USA 103:8989–8994
146. Mason JM, Müller KM, Arndt KM (2007) Positive Aspects of Negative Design: Simultaneous Selection of Specificity and Interaction Stability. Biochemistry 46:4804–4814
147. Mason JM, Hagemann UB, Arndt KM (2007) Improved Stability of the Jun-Fos Activator Protein-1 Coiled Coil Motif: A Stopped-flow Circluar Dichroism Kinetic Analysis. J Biol Chem 282:23015–23024
148. Jouaux, EM, Schmidkunz K, Müller KM, Arndt KM (2008) Targeting the c-Myc coiled coil with interfering peptides to inhibit DNA binding. J Pept Sci (in press)
149. Mossner E, Koch H, Plückthun A (2001) Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs. J Mol Biol 308:115–122
150. Remy I, Wilson IA, Michnick SW (1999) Erythropoietin receptor activation by a ligand-induced conformation change. Science 283:990–993
151. Laser H, Bongards C, Schuller J, Heck S, Johnsson N, Lehming N (2000) A new screen for protein interactions reveals that the Saccharomyces cerevisiae high mobility group proteins Nhp6A/B are involved in the regulation of the GAL1 promoter. Proc Natl Acad Sci USA 97:13732–13737

152. Stagljar I, Korostensky C, Johnsson N, te Heesen S (1998) A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. Proc Natl Acad Sci USA 95:5187–5192
153. Ooi AT, Stains CI, Ghosh I, Segal DJ (2006) Sequence-enabled reassembly of beta-lactamase (SEER-LAC): a sensitive method for the detection of double-stranded DNA. Biochemistry 45:3620–3625
154. Stains CI, Porter JR, Ooi AT, Segal DJ, Ghosh I (2005) DNA sequence-enabled reassembly of the green fluorescent protein. J Am Chem Soc 127:10782–10783
155. Massoud TF, Paulmurugan R, Gambhir SS (2004) Molecular imaging of homodimeric protein–protein interactions in living subjects. Faseb J 18:1105–1107
156. Hu CD, Kerppola TK (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. Nat Biotechnol 21:539–545
157. Bracha-Drori K, Shichrur K, Katz A, Oliva M, Angelovici R, Yalovsky S, Ohad N (2004) Detection of protein–protein interactions in plants using bimolecular fluorescence complementation. Plant J 40:419–427
158. Walter M, Chaban C, Schutze K, Batistio O, Weckermann K, Nake C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J 40:428–438
159. Christmann A, Wentzel A, Meyer C, Meyers G, Kolmar H (2001) Epitope mapping and affinity purification of monospecific antibodies by *Escherichia coli* cell surface display of gene-derived random peptide libraries. J Immunol Methods 257:163–173
160. Wentzel A, Christmann A, Kratzner R, Kolman H (1999) Sequence requirements of the GPNG beta-turn of the *Ecballium elaterium* trypsin inhibitor II explored by combinatorial library screening. J Biol Chem 274:21037–21043
161. Kjaergaard K, Schembri MA, Klemm P (2001) Novel Zn(2+)-chelating peptides selected from a fimbria-displayed random peptide library. Appl Environ Microbiol 67:5467–5473
162. James P, Vorherr T, Carafoli E (1995) Calmodulin-binding domains: just two faced or multi-faceted? Trends Biochem Sci 20:38–42
163. Weaver-Feldhaus JM, Miller KD, Feldhaus MJ, Siegel RW (2005) Directed evolution for the development of conformation-specific affinity reagents using yeast display. Protein Eng Des Sel 18:527–536
164. Williams T (1996) The iridoviruses. Adv Virus Res 46:345–412
165. Tamaru Y, Ohtsuka M, Kato K, Manabe S, Kuroda K, Sanada M, Ueda M (2006) Application of the arming system for the expression of the 380R antigen from red sea bream iridovirus (RSIV) on the surface of yeast cells: a first step for the development of an oral vaccine. Biotechnol Prog 22:949–953
166. Cho BK, Kieke MC, Boder ET, Wittrup KD, Kranz DM (1998) A yeast surface display system for the discovery of ligands that trigger cell activation. J Immunol Methods 220:179–188
167. Kondo A, Ueda M (2004) Yeast cell-surface display–applications of molecular display. Appl Microbiol Biotechnol 64:28–40
168. Raty JK, Airenne KJ, Marttila AT, Marjomaki V, Hytonen VF, Lehtolainen P, Laitinen OH, Mahonen AJ, Kulomaa MS, Yla-Herttuala S (2004) Enhanced gene delivery by avidin-displaying baculovirus. Mol Ther 9:282–291
169. Riddle DS, Sanz L, Chong H, Thompson J, Vile RG (2005) Tumor cell surface display of immunoglobulin heavy chain Fc by gene transfer as a means to mimic antibody therapy. Hum Gene Ther 16:830–844
170. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR (1994) Making antibodies by phage display technology. Annu Rev Immunol 12:433–455
171. Low NM, Holliger PH, Winter G (1996) Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. J Mol Biol 260:359–368

172. de Bruin R, Spelt K, Mol J, Koes R, Quattrocchio F (1999) Selection of high-affinity phage antibodies from phage display libraries. Nat Biotechnol 17:397–399
173. Chowdhury PS, Pastan I (1999) Improving antibody affinity by mimicking somatic hypermutation *in vitro*. Nat Biotechnol 17:568–572
174. Francisco JA, Campbell R, Iverson BL, Georgiou G (1993) Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci USA 90:10444–10448
175. Francisco JA, Georgiou G (1994) The expression of recombinant proteins on the external surface of *Escherichia coli*. Biotechnological applications. Ann NY Acad Sci 745:372–382
176. Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss R, 3rd (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nat Biotechnol 15:29–34
177. Boder ET, Midelfort KS, Wittrup KD (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc Natl Acad Sci USA 97:10701–10705
178. Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JM, Yeung YA, Cochran JR, Heinzelman P, Colby D, Swers J, Graff C, Wiley HS, Wittrup KD (2003) Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. Nat Biotechnol 21:163–170
179. Hanes J, Plückthun A (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci USA 94:4937–4942
180. Gold L (2001) mRNA display: diversity matters during *in vitro* selection. Proc Natl Acad Sci USA 98:4825–4826
181. Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AL (2004) Recent developments in the optimization of thermostable DNA polymerases for efficient applications. Trends Biotechnol 22:253–260
182. Ghadessy FJ, Ramsay N, Boudsocq F, Loakes D, Brown A, Iwai S, Vaisman A, Woodgate R, Holliger P (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. Nat Biotechnol 22:755–759
183. Griffiths AD, Tawfik DS (2003) Directed evolution of an extremely fast phosphotriesterase by *in vitro* compartmentalization. Embo J 22:24–35
184. Cohen HM, Tawfik DS, Griffiths AD (2004) Altering the sequence specificity of Haelll methyltransferase by directed evolution using *in vitro* compartmentalization. Protein Eng Des Sel 17:3–11
185. Levy M, Griswold KE, Ellington AL (2005) Direct selection of transacting ligase ribozymes by *in vitro* compartmentalization. Rna 11:1555–1562