Evolutionary Approaches to Protein Engineering

B. Steipe

Genzentrum der Ludwig-Maximilians-Universität, Feodor-Lynen-Str. 25, D-81377 Munich, Germany
e-mail: steipe@lmb.uni-muenchen.de

1 Targets and Tasks for Protein Engineering

The very term "protein engineering" remains something of an oxymoron, at least as far as engineering implies the rational application of well understood principles towards achieving a prespecified goal. Designed novel functions of proteins remain largely beyond our capabilities, despite intense efforts of numerous research groups in academia and industry. But we are trying, and the last few years have seen a rapid growth in the number of reports describing the successful application of a novel, and at the same time ancient, principle to the problem: evolutionary protein engineering. While this sophisticated trial-and-error approach may at first appear less rational than crystal structure gazing, it is undisputedly more successful – and there is nothing irrational about experimental success. This chapter will focus on the principles, concepts and methods of this field.

The targets for protein engineering have shifted over the last several years, away from medical applications and towards protein biotechnology, partly from concerns about the potential immunogenicity of engineered proteins, partly anticipating superior pharmacokinetic properties of small-molecule pharmacophores. Thus enzymes like proteases for washing powder formulations, amylases for food processing, cellulases and xylanases for pulp and paper processing are today's paradigms for industrial protein engineering (Rubingh 1997), while landmark applications in chemical synthesis (Moore et al. 1997), biosensors (Miesenböck et al. 1998) or bioremediation (Kumamaru et al. 1998) are appearing on the horizon. The situation may change, as protein engineering becomes an increasingly mature science and the first clinical trials of designed immunotoxins afford a view of the many things that are yet to come (Pai et al. 1996).

But what are the goals of engineering in the first place? A protein's role can be loosely divided into two aspects: folding and function, i.e. the intrinsic, structural aspects of the protein and its extrinsic interactions with its surroundings and both are targets for engineering (Table 1).

1.1 Folding

In general the prediction of the folded structure from sequence alone has been as elusive as the rational, targeted change of the sequence to produce novel structures or functions. It is only very recently that progress has been made. The protein folding problem results from the fact that proteins are context-sensitive complex systems, in which the precise effect of any sequence change is highly dependent on the interactions of the altered residue with its surroundings, including the solvent shell, and on the effects on the unfolded state. Predictions that would be based on a precise knowledge of these structures are accordingly difficult to make. Nevertheless, we can measure equilibrium and rate constants for the folding reaction and thus quantify the driving forces behind the phenomenon of self-organization of a polypeptide chain. Thus stability can be regarded as the metric of the protein folding problem.

Table 1. Examples of the diversity of properties that have been improved by evolutionary engineering

Protein	Altered function	Reference
Barley α-amylase	Thermostability: ten-fold increase of half-life at 90°C	Joyet et al. 1992
Subtilisin	Alkaline stability: doubling the autolytic half-time at pH 12	Cunningham and Wells (1987)
Subtilisin	Tolerates loss of stabilizing divalent cations	Strausberg et al. (1995)
Subtilisin E	Active in 60% DMF	You and Arnold (1996)
Streptomyces griseus protease B	Broadened substrate specificity	Sidhu and Borgford (1996)
Green fluorescent protein	40-fold brighter fluorescing bacterial colonies	Crameri et al. (1996)
Immunoglobulin constant domain	Preferential formation of heterodimers	Atwell et al. (1997)
Immunoglobulin variable domain	Tolerates loss of structural disulfide bridge	Martineau et al. (1998)

1.1.1 Thermodynamic Stability

The concept of thermodynamic stability applies to an equilibrium between the native and the unfolded state. If, and only if, the folding reaction is completely reversible and satisfies the two-state approximation – no intermediate is more stable than either the folded or the unfolded state – we can simply count molecules in the folded state F and the unfolded state U and calculate the free energy difference ΔG between the two states (see, e.g. Steipe et al. 1994 for an experimental protocol).

$$\Delta G = -RT \ln \frac{F}{U} \tag{1}$$

This is the purist's definition of stability. But for purposes of engineering we may be more interested in properties such as expression levels or inactivation rates and these will not have to correlate with thermodynamic stability in all cases.

1.1.2 Thermal and Environmental Stability

Thermostability is a desirable property in biotechnological applications for a number of reasons. Substrate solubility may be increased, the risk of microbial contamination may be minimized and the reaction rates may not only be increased in general, but may favor some side-reactions over others (Cowan 1997). Biotechnologically important processes may require extremes of pH, or the presence of chelators, proteases and detergents. Stability in aprotic environments would make protein catalysts interesting for a wide range of chemical transformations for which stereo- or regioselective catalysis is required. Unfortunately, most proteins denature only a few degrees above the physiological temperature and this is frequently an irreversible process that rapidly draws folded protein out of the equilibrium into the unfolded state. In general, the reason for irreversible inactivation at high temperatures or under other adverse environmental conditions is aggregation of the unfolded state. This process will be governed by the concentration and the unfolding rate k_{unfold} which is itself related to equilibrium stability.

$$\Delta G = -RT \ln \frac{k_{\text{fold}}}{k_{\text{unfold}}} \tag{2}$$

Thus in any comparison of thermostability, care must be taken to consider the exact experimental conditions under which the measurements were performed. As illustrated in Fig. 1, mutations that lead to increased thermostability will affect the unfolding activation energy and need not necessarily increase the thermodynamic stability.

Solid experimental evidence exists today that the effect of point mutations on folding stability can be well approximated as additive, distributed, largely independent interactions. This situation is ideal for engineering, since the combination of stabilizing mutations can sum up to quite significant stabilization (Zhang et al. 1995). Various methods have been described that allow the design of stabilizing mutations, such as the stabilization of α-helix macrodipoles (Walter et al. 1995), the engineering of structural motifs like helix N-caps (Aurora and Rose 1998) or β-turns (Ohage et al. 1997), or the introduction of residues with higher intrinsic propensities for their respective conformational state (Zhang et al. 1992), the introduction of disulfide bridges (Johnson et al. 1997), the reduction of the unfolded state entropy with X → Pro mutations (Nicholson et al. 1992) or the analysis of aligned sequence distributions (Steipe et al. 1994). Engineering of protein stability is the one area in which rational engineering is competitive with evolutionary protocols.

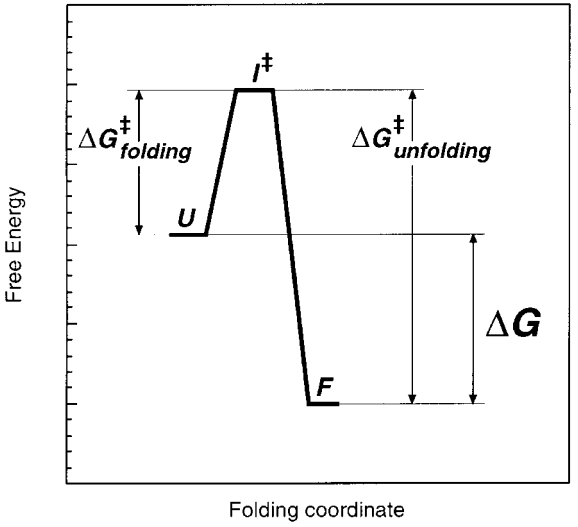

Fig. 1. Energy levels of the folding reaction. While the folding equilibrium is governed by the free energy difference, ΔG, between the unfolded and the folded state, U and F, the unfolding rate is determined by the free energy difference between the folded and the transition state, $\Delta G^{\ddagger}_{\text{unfolding}}$. Mutations can significantly affect either or all states, for instance a mutation that raises exclusively the transition state free energy will decrease the unfolding rate without changing the protein's thermodynamic stability

1.1.3 Other Folding Considerations

Properties other than thermodynamic or kinetic stability may need to be optimized for technological applications. The formation of structural disulfide bonds may be yield limiting, especially during in vitro refolding when free thiols may oxidize statistically, since non-native disulfide bonds will be favored. Unfortunately, the engineered removal of disulfide bonds carries a heavy energetic penalty, but with attention to refolding protocols very good yields can be obtained for many proteins (Rudolph and Lilie 1996). For proteins that posess a *cis*-peptidyl-prolyl bond in the native structure, in general the *trans-cis* isomerization during folding will be the rate-limiting step (Schmid et al. 1996). Some of these *cis*-peptide bonds may be dispensable and engineering may be possible (Kiefhaber et al. 1990). Aggregation of intermediates during folding is the third major source of folding problems (Kiefhaber et al. 1991). Cellular chaperones have evolved to prevent aggregation in vivo, but engineering may also significantly improve the efficiency of folding. For example, the comparison of sequences of well-expressed immunoglobulin V_H domains identified two residues that independently reduce domain aggregation in the periplasm and cell lysis during expression (Knappik and Plückthun 1995).

1.2 Function

Engineering a protein's function is significantly more challenging than stabilizing its structure. While stability can be improved with a number of independent, distributed point mutations – each of which may be a crude approximation to an optimal solution – the active site of a protein is typically localized to a unique discontinuous epitope, and function is exquisitely sensitive to the precise orientation and interactions of the participating residues. In this setting, mutations are no longer independent and their combined effects will deviate strongly from simple additivity. As a consequence, an evolutionary trajectory to a novel function may require crossing significant barriers of reduced activity. This has been demonstrated in an analysis of the catalytic triad of the serine protease subtilisin (Carter and Wells 1988). Both the substitution of the catalytic serine and histidine reduce the turnover number by a factor of approximately 10^6; the combined mutations, as well as the substitution of the aspartic acid to alanine, have no additional deleterious effect. As a consequence, three coordinated amino acid changes would be needed to generate the catalytic triad and intermediates confer no selective advantage on the protein.

1.2.1 Binding

The most elementary interaction of a protein with its surroundings is binding another molecule. Highly complementary molecular surfaces have evolved to perform any biologically required task of binding and discrimination. How many epitopes will be required to bind any molecular shape with high affinity? This

question bears directly on the design of evolutionary experiments. Experience with the immune system suggests that the number may be surprisingly small. While on the order of 10^7 different combining sites may be generated in a primary immune response, as little as 10^2–10^4 different B cells are sufficient to provide viral immunity (Bachmann et al. 1994). Indeed, from phage display libraries with diversities of $>10^8$, domains with subnanomolar dissociation constants and off-rates of 10^{-3}/s can apparently be isolated almost routinely (Vaughan et al. 1996; Pini et al. 1998).

1.2.2 Catalysis

To a significant degree, catalysis is just a different form of binding – binding the transition state of a chemical reaction and thus lowering its free energy (Jencks 1969). This concept is well borne out by the catalytic activity of antibodies that have been raised against transition state analogs (Lerner et al. 1991). On the other hand, the catalytic function of antibodies has been consistently inferior to that of "true" enzymes that catalyze the same reaction. This is commonly believed to stem from the absence of functional groups that might take an active part in the reaction, but which do not increase affinity to the transition state analogue (Wentworth and Janda 1998). Indeed, it can be shown that increased binding to the transition state analogue need not lead to increased catalytic rates (Baca et al. 1997). The lesson, not only for antibodies, is that successful evolutionary engineering of catalysts should involve direct selection for activity, and, in case this is difficult, more attention should be invested in the design of the selection process. Two strategies have been devised for catalytic antibodies that address this problem: reactive immunization (Wirsching et al. 1995; Barbas et al. 1997) and mechanism-based panning of phage display libraries (Fig. 2) (Janda et al. 1997).

In summary, function is governed by localized, highly cooperative interactions. Since even small differences in geometry may have dramatic effects on function, traditional structure-based engineering approaches have consistently failed to improve enzymes. For example, even something seemingly as straightforward as the reengineering of trypsin towards the substrate specificity of its close relative chymotrypsin has required major remodeling and transplantation of entire loops, comprising the substrate recognition subdomain (Hedstrom 1996).

2 Concepts for Rational and Evolutionary Engineering Approaches

Protein engineering, whether rational or evolutionary, is the modification of an existing sequence for a new purpose. It is not trivial that protein engineering is possible at all.

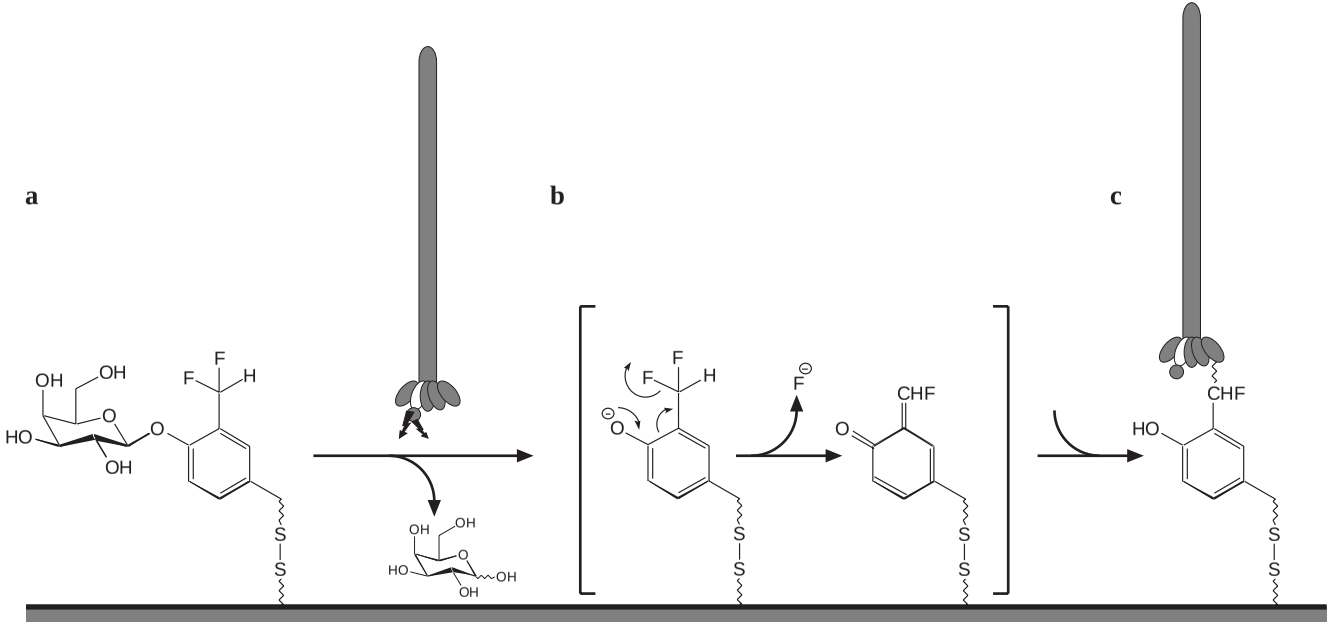

Fig. 2a–c. Principle of mechanism based panning of a phage display library for glycosidic bond cleavage. **a** An ortho-(difluoromethyl)-phenyl-galactoside inhibitor is coupled to a matrix via a disulfide spacer. Catalytic antibody domains in a phage display library hydrolyse the glycosidic bond. **b** The difluoromethylphenyl moiety generates the reactive quinone methide. **c** Any nearby nucleophile can be alkylated, covalently trapping the catalytically active phage. Subsequently the matrix is washed and bound phage are eluted by reducing the disulfide spacer with DTT. (After JANDA et al. 1997)

2.1 Theoretical Considerations

The basic assumption of protein engineering is that a natural sequence can be modified to improve a certain function. This implies: (1) that the protein is not already at an optimum for that function, otherwise it could not be improved; (2) that the required sequence changes can be accommodated without disrupting the structure, otherwise it would not fold; and (3) that the new sequence is not too different from the native sequence, otherwise it could not be found. None of these three observations is trivial. The first highlights the fact that evolution cannot generate proteins that are optimal for any given task, it can only generate proteins whose performance confers a selective advantage on the host organism. Thus optimization will cease when no more selective advantage can be gained from further improvement. In this view, proteins are not optimal but sufficient for their task and there is significant potential for improvement under guided selection. The second observation highlights the plasticity in protein structure and the redundancy inherent in a protein sequence. Most point mutations have only local, medium range effects and even though a large majority of mutations will be destabilizing, not all are completely disruptive. In fact there is a good reason that protein sequences should be optimized by evolution to be error-tolerant: the adaptability of a fold to random mutations is a factor determining the height of barriers on its evolutionary landscape. The third observation relates to the density of new optima in sequence space.

2.1.1 Sequence, Structure, Function Spaces and Landscapes

A sequence space is a very useful concept for the discussion of the evolution of proteins even though from a biochemist's perspective a sequence space has some very odd properties (see Kauffman 1993). Every sequence is considered to be represented by a point in space. The dimensions of the space correspond to the positions in the sequence, they at least equal the sequence length, but there may be more when gaps are needed. The dimensions are ordered, with an index increasing from the NH_2- to the COOH-terminal, and every dimension – the positions in the sequence – can take at least 20 different discrete values that have no obvious inherent ordering. Sequence spaces are large: a protein of 230 amino acids (233 amino acids is the median length of a protein chain in a library of 635 unrelated sequences in the structural database) spans a sequence space of 20^{230} or 10^{300} points. The relationship between two sequences can be described as their distance in sequence space; the size of a molecular library can be related to a volume surrounding its progenitor sequence.

Based on this concept, an evolutionary landscape can be defined by associating a function value, commonly called a fitness function, with every point in a sequence space. This can be a Boolean value, like "survival", a discrete value, like "number of oligomers", or, most frequently, a continuous value, like "melting point", "k_{cat}", or "racemic excess". Obviously, the function value and thus the shape of the landscape depends on the fitness function that is considered, or embodied in the experimental

design. For example, on a landscape representing the catalytic rate towards the natural substrate, all the wild-type sequence neighbors in space are expected to perform worse than the wild-type, which is accordingly in a (local) sequence optimum. But the activity towards a different substrate implies a different fitness function and some direct sequence neighbors may improve on the wild-type in this case.

2.1.2 Evolutionary Trajectories

A walk in sequence space is a series of sequences, each derived from its predecessor in single steps or jumps. Such a walk can be random or adaptive, guided by increasing fitness function values. An evolutionary trajectory is the path between the initial and the final sequence. What elevation profile a trajectory will have will depend on the fitness function considered. The length of the trajectory is equal to the number of single point mutations in the evolutionary process.

As is evident from Table 2, the distance in evolutionary space that has been traversed in real, successful experiments is generally less than half a dozen steps. This can be interpreted in two ways: on the one hand, sequences for improved properties apparently lie close to their progenitors; they are not sparsely scattered in sequence space. On the other hand, the capacity of our present methods to generate functional sequences, more than a few point mutations away from a starting point, appears rather limited. Additionally, mutations are generated in a highly biased fashion.

Even if sequence space is to be sampled to a distance of only a few mutations, a strategy that relies on finding a successful variant by exhaustive search of a single, large sequence pool is likely to fail. The reason is a combinatorics problem: the codon dilemma. Amino acid sequence changes are encoded on a nucleotide level,

Table 2. Representative experiments using successive cycles of variation and selection

Protein	Property	Number of cycles for success	Number of nucleotide changes required	Number of amino acid changes required	Reference
β-Lactamase	Increased activity	3	4	4	Stemmer (1994b)
GFP	Improved folding and expression	3	3	3	Crameri et al. (1996)
Subtilisin E	Stability in aqueous DMF	2	3	3	You and Arnold (1996)
Arsenite membrane pump	Increased activity	3	3	3	Crameri et al. (1997)
FLP-recombinase	Thermostability	8	3–4	3–4	Buchholz et al. (1998)

In almost all cases a single nucleotide change leading to a single amino acid change was sufficient per cycle, the number of silent mutations was approximately the same. No amino acid change was reported that would have required more than one nucleotide change. Thus current protocols appear to sample sequence space in a biased fashion, in single mutation steps.

and a single amino acid change may require up to three coordinated changes of the coding sequence. The consequences can be dramatic: from the perspective of the amino acid sequence, the probability for a specific change in a sequence of length 230 is:

$$p = \frac{1}{\begin{matrix}\text{sequence}\\ \text{length}\end{matrix}} \cdot \frac{1}{\begin{matrix}\text{amino acid}\\ \text{alternatives}\end{matrix}} = \frac{1}{230} \cdot \frac{1}{19} = 2.3 \cdot 10^{-4} \tag{3}$$

But from the genetic perspective, the average probability for encoding a specific amino acid change through random nucleotide changes depends strongly on the number of required nucleotide changes:

$$p \approx \left(\frac{1}{\begin{matrix}\text{gene}\\ \text{length}\end{matrix}} \cdot \frac{1}{\begin{matrix}\text{nucleotide}\\ \text{alternatives}\end{matrix}} \right)^{\left(\begin{matrix}\text{number of}\\ \text{changes}\end{matrix}\right)} = \left(\frac{1}{230 \cdot 3} \cdot \frac{1}{3} \right)^{d}. \tag{4}$$

$d = 1$ in 40% of mutations, e.g. Tyr(TAC) → Phe(TTC);

$$p = \left(\frac{1}{230 \cdot 3} \cdot \frac{1}{3} \right) = 4.8 \cdot 10^{-4}.$$

$d = 2$ in 53% of mutations, e.g. Tyr(TAC) → Trp(TGG);

$$p \approx \left(\frac{1}{230 \cdot 3} \cdot \frac{1}{3} \right)^{2} = 2.3 \cdot 10^{-7}.$$

$d = 3$ in 7% of mutations, e.g. Tyr(TAC) → Met(ATG);

$$p \approx \left(\frac{1}{230 \cdot 3} \cdot \frac{1}{3} \right)^{3} = 1.1 \cdot 10^{-10}.$$

Surprisingly, even a large library by common laboratory standards, say 10^8 sequences, will not exhaustively encode all single point mutations! The consequences are: since one can only expect to densely sample sequence space to a distance of one, at best two, mutations, the successful application of evolutionary engineering requires that an evolutionary path exists that will yield a detectably improved function for every single evolutionary step.

Indeed, natural evolution works fundamentally along the principle of achieving results against impossibly small odds by arriving at the target sequence in stepwise improvements. The true power of evolutionary engineering lies in devising methods to iterate variation and selection.

If the improved function requires three or more cooperatively interacting sequence changes – every individual mutation being deleterious – then the chances of traversing such a barrier become vanishingly small. When this must be suspected, efforts should be focused on reducing the volume of sequence space that is to be searched, e.g. by developing some hypothesis on which region of the protein should be targeted, or by employing some scheme of site-directed random mutagenesis.

2.1.3 Search in Sequence Space

A model well suited for a theoretical investigation of the structure of molecular fitness landscapes and search trajectories was introduced by S. Kauffman (Kauffman 1993). His *NK* model considers sequence spaces for sequences of a length of N sites. Each site can take A states and makes a contribution to the overall fitness of the sequence that depends on its own state and that of K other sites. When $K = 0$, the sites contribute independently and additively to the global fitness, when K is maximal, i.e. $K = N - 1$, each site is influenced by every site. While the value for K in natural proteins or even peptides is different for every site and has not been well determined experimentally, computer models that vary K can shed some light on the ruggedness of the evolutionary landscape and suggest efficient ways to locate minima. In a comparison of pooling, recombination and mutation strategies for an *NK* model of a random hexapeptide library, the available experimental data apparently support a value of K around $0.5 \times N$ – intermediate between being random and fully correlated (Kauffman and Macready 1995). The landscape for $K = 0$ is smooth, possessing a single peak which can be readily found. For small K, sequences in a local optimum will be fitter than most one- or two-mutant neighbors. The larger K is, the more likely it is that an evolutionary trajectory will become trapped in a local optimum and the probability for finding improved sequences becomes independent of search distance – the landscape is then uncorrelated. Conversely, for small K, i.e. correlated landscapes, the probability of finding an improved sequence decreases with search distance. How does this translate into the vocabulary of molecular biology? Properties that require a significant number of cooperative interactions before an improvement in fitness is observed cannot be found by any strategy that is currently practical. Properties that can be improved with independent or quasi-independent point mutations have a good chance to be selectable in iterated evolutionary cycles. In this case, single or double mutations per cycle search sequence space more efficiently than more radical changes.

The most important conclusion is the importance of investing more effort in the design of the experimental protocol: being able to detect even slight advantages in the desired function and running the evolutionary optimization through a large number of cycles, rather than constructing ever larger libraries.

2.2 Complementing Evolutionary Approaches with Rational Concepts

Rational engineering designs solutions top down: it is an attempt to divine the location of the desired optimum and to design experiments according to this insight. Rational engineering requires knowledge of the sequence and preferably the structure of the protein, delineation of the active site, understanding of the mechanism, identification of cofactors, etc. Most importantly, it requires a hypothesis about the limiting step for the desired function. In well characterized systems, the performance of rational engineering can be quite remarkable. As an example, the thermolysin-like protease (TLP) has been engineered with eight point

mutations for thermostability to resist boiling temperatures (Van den burg et al. 1998). Remarkably, the mutant enzyme is as active at room temperature as the wild-type. Individual mutations were contributed from an analysis of sequence differences to thermolysin, from an increase in the number of residues that lower the entropy of the unfolded state and from a designed disulfide bridge – tried and proven approaches to rational protein engineering.

State of the art protein engineering and design applies some computational algorithm, an objective function, to a novel sequence and then attempts to find an improved sequence through methods of combinatorial optimization. That this process is becoming practical is evident from the successful de novo design of a protein G-β1 domain that is 18kJ/mol more stable than the wild-type (Malakauskas and Mayo 1998). Thus modern protein design in its application of combinatorial optimization principles frequently is itself in silico evolutionary engineering.

Two main benefits of rational design for evolutionary engineering can be identified: the first is the possibility to construct stable structural frameworks for the display of combinatorial libraries, the second is the generation of hypotheses that allow limiting the required size of the library, such as constraining diversity to spatially adjacent residues or conserving hydrophobicity profiles.

3 Evolutionary Engineering Methods

Evolution implies iteration, and the practical application of evolutionary principles to protein engineering involves repeating cycles that can be divided into three parts: the generation of genetic diversity, the coupling of genotype and phenotype and the identification of successful variants.

3.1 Generating Diversity

The exhaustive mutation of a limited number of sites is a fundamentally different experiment from the stepwise optimization of entire genes. The former case can avoid the codon dilemma: sequence space can well be sampled exhaustively to five or six positions. The disadvantage is that only a subset of the entire gene can be targeted. For this reason, degenerate oligonucleotides are commonly used in the construction of epitope libraries, while diversity in libraries of entire proteins is commonly generated with some PCR-based procedure.

3.1.1 Oligonucleotide Directed Mutagenesis: Circumventing Genetic Code Degeneracy

If only short regions of the protein are to be targeted, various methods of directed mutagenesis with degenerate oligonucleotides can be employed. The simplest and oldest approach is to use equimolar mixtures of all four nucleotides, (N)(N)(N), for

the codons that are to be changed (Oliphant et al. 1986), but this may not be the best strategy. In procedures that involve the synthesis of a complementary strand, a bias for incorporating the original nucleotide will arise from the preferential hybridization of oligonucleotides that form larger numbers of Watson-Crick base pairs. This bias can be eliminated by reducing the concentration of the wild-type nucleotide during synthesis (Airaksinen and Hovi 1998). But more importantly, an (N)(N)(N) codon mixture is biased in favor of those amino acids with more entries in the genetic code table; for instance, it will contain six times more leucine than methionine and it will contain 4.7% stop codons in every position. Thus, the chance of arriving at a randomized sequence of length N that can be translated without stop codons is:

$$p = \left(1 - \frac{3}{64}\right)^N \tag{5}$$

e.g. $p \approx 0.6$ for ten residues and $p \approx 0.4$ for 20 residues. If the library is large enough to contain every variant sequence and the selection process can pick out individual sequences, these shortcomings will not be relevant. But if the library can sample sequence space only sparsely, more intelligent strategies are needed to improve its diversity and quality. A useful alternative is the codon mixture (N)(N)(C,G,T). This mixture not only encodes a more even distribution of amino acids, but also reduces stop codon frequency to 2% – improving the chances for a translatable sequence to $p \approx 0.8$ for ten residues and $p \approx 0.67$ for 20 residues. Alternate schemes have been published that exploit the error-tolerance features inherent in the genetic code. Mixtures can be biased towards residues with common physicochemical properties such as size, hydrophobicity or charge while at the same time the redundancy is reduced (Balint and Larrick 1993) (Table 3). Biasing amino acid distributions requires a hypothesis about which choices are advantageous. This may be based on sequence alignments of homologous genes, on conserving the physicochemical properties of the mutated residues (e.g., Table 4), or, as in a procedure termed "recursive ensemble mutagenesis" (Delagrave et al. 1993), on compiling the distributions from the sequence pool of the preceding evolutionary cycle.

Table 3. A codon mixture for charged amino acids (Balint and Larrick 1993)

Charged (R)(R)(K)			
Position	1	2	3
A	50%	50%	50%
C	–	–	–
G	50%	50%	–
T	–	–	50%

Resulting amino acid spectrum (probability): acidic, E(0.125) D(0.125); basic, R(0.125) K(0.125) H(–); hydrophilic, Q(–) N(0.125) T(–) S(0.125); hydrophobic, V(–) L(–) M(–) I(–) Y(–) W(–) F(–); small, A(–) G(0.250); problems, P(–) C(–) Stop(–).

Note that all encoded amino acids except glycine are present in the mixture with equal probabilities. Stop codons are excluded. If a charged residue is required with certainty at the targeted position, this mixture is nearly optimal, as far as simplicity of synthesis and complexity is concerned.

Table 4. A codon mixture centered on glutamate for the construction of evolutionary libraries

Glu (at $d = 0.5$ of average distance in BLOSUM 62 matrix)			
Position	1	2	3
A	38%	41%	–
C	27%	18%	30%
G	31%	21%	53%
T	4%	20%	17%

Resulting amino acid spectrum (probability): acidic, E(0.291) D(0.137); basic, R(0.028) K(0.112) H(0.048); hydrophilic, Q(0.102) N(0.053) T(0.029) S(0.007); hydrophobic, V(0.023) L(0.008) M(0.006) I(0.003) Y(0.002) W(0.0005) F(0.0001); small, A(0.074) G(0.046); problems, P(0.026) C(0.0003) Stop(0.005).
Note the good correspondence of the amino acid frequencies with physicochemical measures of similarity – charge, hydrophobicity and volume. The frequency for nonsense mutations is reduced by a factor of ten relative to its occurrence in a random nucleotide mixture. No amino acid is completely excluded. The mixtures are adjusted to compensate for unequal reactivity of nucleotides during synthesis. Such mixtures can be optimized individually for every amino acid and synthesized on standard oligonucleotide synthesizers that allow independent control of reagent concentrations.

Under specific circumstances it may be desirable not to approximate amino acid distributions with degenerate codons, but to specify them explicitly by synthesizing random libraries directly from building blocks of trinucleotides. That such an approach is indeed feasible, after careful optimization of the synthesis strategy, has now been reported by a number of groups (Virnekäs et al. 1994; Lyttle et al. 1995; Ono et al. 1995; Kayushin et al. 1996; Gaytan et al. 1998). The downsides of this method are that it is not commercially available and it does not solve the problem of deciding which mixture of amino acids may be desirable at any given position.

In order to combine ease of synthesis with balanced mixtures of amino acids, redundancies and symmetries in the genetic code can be favorably exploited (Steipe and Bruhn, in preparation). A typical application would be the synthesis of conformationally constrained epitopes, such as antibody combining sites, for which some preference for each position can be defined, e.g. conserving the hydrophobicity profile, yet no amino acid should be rigorously excluded, since it might be just the one critical for the desired function. In effect, amino acid properties such as hydrophobicity or size should be statistically constrained to limit the destabilizing effect of the new sequence on the framework structure. Thus similar residues should be more frequent in the mixture than dissimilar residues. To achieve this, nucleotide mixtures can be simultaneously optimized in every codon position, to maximize the resulting codons' complexity – defined as the information-theoretical information content – and to limit their dissimilarity, measured empirically in terms of exchange probabilities from a mutation data matrix (Table 4).

3.1.2 Chemical Mutagenesis, Mutator Strains and UV Irradiation

While these methods were among the earliest used in evolutionary engineering (Singer and Kusmierek 1982), they have been largely superseded by the more modern techniques described below. The main disadvantage of all three methods is their indiscriminate targeting of the entire genome (or at least an entire plasmid),

and this makes it rather likely that the screen for function will be influenced by non-specific effects, like altered expression rates, or even the spontaneous modification of cellular enzymes to perform the task that is being screened or selected for.

3.1.3 Error-Prone PCR

For most purposes, the introduction of nucleotide changes via error-prone PCR will be the method of choice: it is simple, efficient, restricted to the region of interest and well characterized. The protocol devised by Leung et al. (1989) was subsequently improved (Cadwell and Joyce 1994) to reduce the inherent bias of nucleotide transitions over transversions (Shafikhani et al. 1997). Both methods allow tuning the mutation rate by varying the concentrations of Mn^{2+} and dNTPs, or the number of PCR cycles. Since beneficial mutations are rare and the combination with a disruptive mutation will produce an inactive protein, it is best to keep the mutation rate to a level of one or two sequence changes per gene. Since some amino acid changes will require three concerted nucleotide substitutions, in practice the ideal number of nucleotide changes is between two and six over the length of the gene. Very high mutation rates for the randomization of short epitopes can be achieved with the inclusion of synthetic nucleoside analogues, that can base-pair ambiguously (Zaccolo and Gherardi 1996).

3.1.4 DNA Shuffling

By far the most successful approach to molecular evolution appears to be the DNA shuffling method, pioneered by W.P. Stemmer to address the question of how an efficient walk on an evolutionary landscape can be generated experimentally (Stemmer 1994a) (Fig. 3). This protocol allows successful mutations to be passed among sequences by recombination. The power of this "sexual PCR" process stems from the possibility of preserving locally optimal solutions, which may improve the desired property synergistically when combined.

In a first step, the gene of interest is cleaved into many short, random fragments with DNAse I. These fragments of 10–50 base pairs are then purified and recombined in a PCR-like process without exogenous primers. Terminal primers are added to the last step of extension and full length sequences are amplified and cloned. Since the melting and annealing steps will cause fragments from different strands to hybridize, an efficient recombination of strands takes place. In principle, this process would simply regenerate the native sequence, but variation can be introduced into the pool by various processes:

1. By initially amplifying the wild-type gene before fragmentation under mutagenic PCR conditions
2. By initially using a pool of genes, such as homologous genes from different organisms (Crameri et al. 1998)
3. Intrinsically, by the process of extension and recombination itself, which has an intrinsic, tunable error-rate (Zhao and Arnold 1997)

4. By the addition of mutagenic primers to the mixture, and
5. By pooling the evolved genes from the most successful mutants of each evolutionary cycle

After the successful isolation of mutants, further cycles of recombination can be performed with an excess of the wild-type sequence under stringent selection conditions. This process of back-crossing will revert nonessential mutations to the wild-type sequence and thus give some insight into the essential sequence changes.

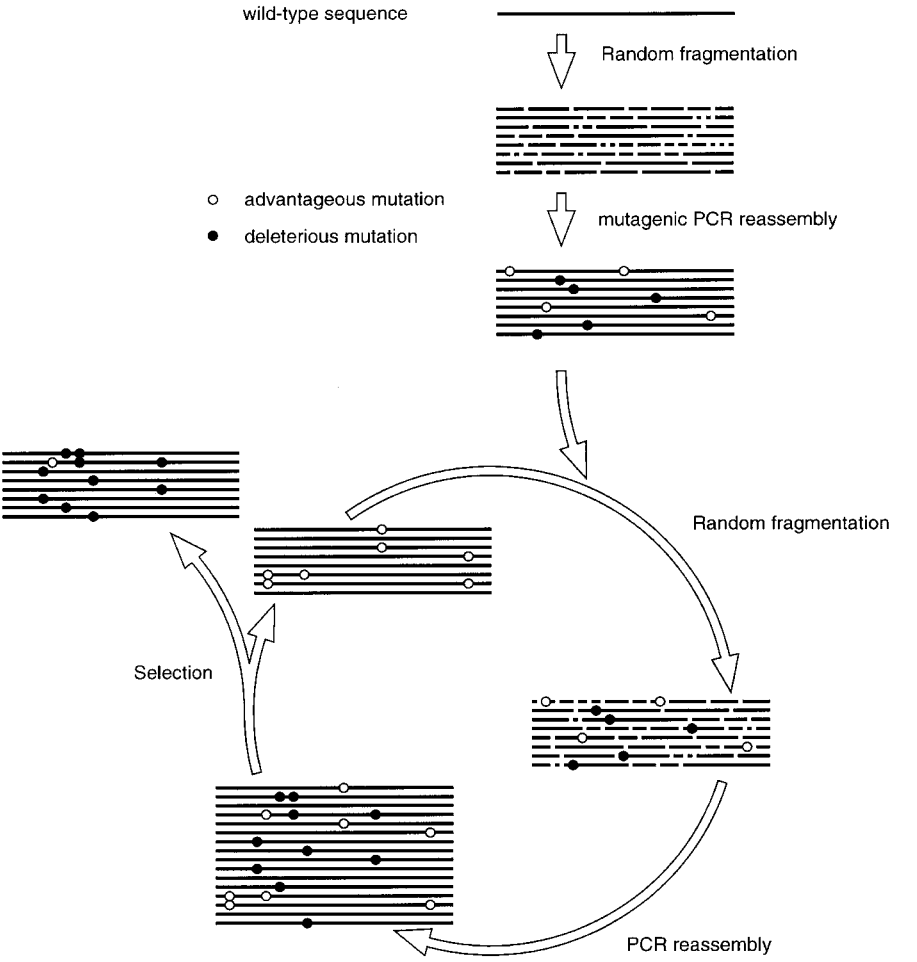

Fig. 3. The DNA shuffling method for molecular evolution. After random fragmentation, a pool of genes is reassembled with a PCR protocol that at the same time is mutagenic and generates multiple recombination events. From the recombined library, functional sequences are selected and the new resulting pool is improved by further iterations of the protocol

3.1.5 Recombination In Vivo and In Vitro

DNA shuffling is modeled along natural recombination, but can natural recombination itself be used for the generation of large molecular libraries? Three site-specific recombination systems have been analyzed in some detail in *E. coli*: phage lambda Int, transposon Tn3 and the Cre recombinase of bacteriophage P1. Of these, the Cre-loxP system appears ideally suited for engineering purposes (Hoess et al. 1984). It is simple, requiring only 34 bp of DNA binding site and the recombinase, and it appears to work independently of local DNA conformation.

In a particularly elegant application, Fisch et al. (1996) reported the generation of a large, combinatorial peptide library. Two artificial exons, each encoding ten randomized amino acids, were joined via a five residue spacer and fused to the pIII phage coat protein for phage display. Exon shuffling was achieved via the lox recombination site cloned into a self-splicing group I intron (Cech 1990), which automatically excises itself after transcription. The authors report a library size in excess of 10^{11} peptides and note the potential of the system for the de novo evolution of small peptides and proteins.

Recombining a set of highly homologous genes in vitro can also be achieved in a procedure called staggered extension process (StEP) recombination (Zhao et al. 1998). StEP involves a PCR procedure with a low concentration of terminal primers or random-sequence primers (Shao et al. 1998) and very short extension cycles at reduced temperature, which will only extend primers over 5–20 base pairs per cycle. These abbreviated fragments will switch templates during the denaturation/annealing cycles and the final, full-length sequence will have been synthesized from a number of different templates.

Whether in vitro recombination is achieved by template switching or by DNA shuffling, its capacity to accelerate the search process makes it the core of modern, efficient evolutionary protocols. Initial point mutations sample local new optima of the evolutionary landscape. Successful variants can subsequently be combined, removing silent and deleterious mutations and further increasing activity (Moore et al. 1997). The combined mutations put the protein into a more distant region of sequence space, one that would not previously have been accessible with a library of practical size. At this new optimum, the process can be repeated until the limiting factor is the sensitivity of the experimental setup to identify further improvement.

3.2 Coupling Genotype and Phenotype

In order to identify desired sequences, some strategy needs to be devised that will ensure that the desired function will be in some way physically associated with its gene. The alternative of direct sequencing of the improved protein is currently not technically feasible, even though it has been successfully applied to the analysis of peptide libraries on beads (Lam et al. 1991). Similarily, encoding schemes have been developed for non-genetic combinatorial libraries (Czarnik 1997). Yet another

similar concept is embodied in the synthesis of peptide libraries on beads, together with a synthetic oligonucleotide encoding the sequence (Needels et al. 1993). The advantage, as in all procedures based on combinatorial chemistry, is the possibility to incorporate non-proteinogenic amino acids; but the chief disadvantage is the limited library size and the added difficulty of decoding as compared to genetically based methods. An array of methods to couple information and function are described below, ranging from the binding of expressed peptides to their genes to the association of gene and protein in living cells.

3.2.1 RNA-Peptide Fusions

Covalent fusions of an mRNA and its encoded peptide can be achieved when a pool of mRNAs is synthesized with the peptidyl-acceptor antibiotic puromycin attached to the 3′ end. The mRNAs are in vitro translated, the 3′ puromycin end – an analog to a charged tRNA – will bind to the ribosomal A site at some time during the translation and its free amino group will be transferred to the carboxylate end of the nascent peptide chain (Nemoto et al. 1997). Once this has happened, the mRNA is covalently bound to the peptide and the adduct will dissociate from the ribosome. These adducts can then be screened for the desired function, the mRNA of successful sequences reverse-transcribed, amplified and cloned for analysis (Roberts and Szostak 1997). This procedure has tremendous potential, since library sizes of 10^{12} have been achieved and 10^{15} should be attainable with some optimization and scale-up; they are thus far larger than those obtainable with other methods. We are certain to see reports of refinements soon, such as the use of longer sequences, or even obviating the need for in vitro mRNA-puromycin synthesis, perhaps through the use of a ribozyme sequence. The only downside appears to be, in principle, the requirement for single molecule detection efficiency.

3.2.2 Ribosome Display

An alternative to the chemical coupling of mRNA and peptide is to preserve their association on the ribosome. This procedure has been developed for peptide libraries (Mattheakis et al. 1996) and for functional proteins (Hanes and Plückthun 1997; He and Taussig 1997). While the procedure requires some biochemical sophistication, the large library sizes of $>10^{12}$ individual molecules and the possibility to use full-length proteins makes it very attractive. No additional transformation steps are required and PCR amplification between cycles of enrichment allows the introduction of random mutations – evolutionary engineering, entirely in vitro.

3.2.3 Peptide on Plasmid

One of the simplest in vivo embodiments of the coupling of information and structure is the direct, physical association of the target molecule with its gene via a DNA-binding domain. For peptide libraries, this has been achieved with the fusion of a library to the C-terminus of the *lac*-repressor (Cull et al. 1992; Schatz et al.

1996). After isolation of the repressor-plasmid complex from the cell, ligand binding candidates can be retained on an affinity column. The plasmid can be eluted either by denaturation or by adding the inducer and the eluate used to transform cells. Note that even though the number of molecules participating in the experiment can be large, the actual diversity of the library will equal the number of cells into which the initial library has been transformed.

While the procedure is conceptually extremely simple and powerful, requiring no additional decoding or cloning steps, it requires a gentle, yet quantitative procedure of lysing cells that will not interfere with tight physical association of the protein with the DNA, and it requires a slow off-rate of the DNA-binding domain to prevent exchange of the binders. The procedure is also sensitive to interference from intrinsic DNA-binding properties of the target or the matrix.

Another potential problem arises from the fact that LacI dimerizes via its 300-amino acid COOH-terminal domain. Dimerization is a disadvantage for screening, because avidity effects can result in the selection of intermediate- to low-affinity interactors which outnumber high-affinity binders. Thus a monomer domain would be desirable. Such a protein has been constructed by evolutionary engineering of a synthetic linker peptide that fuses two 60-amino acid DNA binding 'headpiece' domains of LacI. This monomeric protein binds DNA stably and can be used for panning and enrichment of high-affinity binding peptides (Gates et al. 1996).

3.2.4 Phage Display

The most widely used system for screening libraries today is phage display (Burton 1995). Peptides or protein domains are fused (most commonly) to the gene III protein (gIIIp) of filamentous phage. After the host cell is infected by helper phage, the fusion protein is incorporated into newly made phages together with its coding gene. Functional sequences are subsequently enriched from the pool by binding a ligand matrix and eluted phages can be directly transformed into host cells for amplification and analysis. The system is very versatile: besides peptide libraries, successful fusions have been reported for the engineering of enzymes such as alkaline phosphatase (Mccafferty et al. 1991), β-lactamase (Soumillion et al. 1994), staphylococcal nuclease (Light and Lerner 1995) or even trypsin (Wang et al. 1996). Protease inhibitors have been engineered (Markland et al. 1996) as successfully as cytokines (Vispo et al. 1997), growth hormone (Chien et al. 1991), and zinc-finger domains (Rebar and Pabo 1994). A particularly interesting new development is the recruitment of a lipocalin framework for the display of a large epitope library (Beste et al. 1999). These novel proteins have been aptly called anticalins by the authors. The largest area of application, however, has been immunoglobulin domains, particularly single-chain Fv fragments (scFvs) (see Hoogenboom et al. 1998 and Griffiths and Duncan 1998 for recent reviews).

In general, stable cytoplasmic proteins appear to cause problems in this system. The assembly of filamentous phage takes place in specific assembly sites where inner and outer membranes come in close contact and the proteins involved are stored as integral proteins of the inner membrane until they are incorporated into

the growing phage. Both overexpression of the fusion protein as well as fusions with proteins that cannot unfold for secretion will lead to toxicity and ultimately to plasmid instability, degrading library diversity or leading to complete loss of the molecule to be displayed. Two factors have been identified that appear to alleviate the problem. First, it is important to repress the background expression as completely as possible, before induction of phage assembly. For example, this can be achieved through introduction of a transcriptional terminator upstream of the *lac* promoter (Krebber et al. 1996); other tightly regulated promoters (Skerra 1994) may serve the same purpose. Second, efficient secretion of the fused protein appears to be crucial for efficient display and this includes translocation across the inner membrane as well as folding in the periplasmic space. By panning a library of *E. coli* proteins, coexpressed with a poorly folding scFv-gIIIp fusion, for high expression levels, Bothman and Plückthun (1998) were able to identify a protein that improves the expression of a wide range of scFv fragments by increasing the amount of displayed protein. Significantly, the protein indeed is a periplasmic chaperone, the *skp* or *ompH* gene product (Chen and Henning 1996), reminiscent of earlier reports that coexpression of the cytoplasmic chaperone GroE would increase phage titers by two orders of magnitude (Soderlind et al. 1993).

Fusions to gIIIp have the advantage of monovalency, since gIIIp is present in only five copies and the native gIIIp is supplied in excess by the helper phage, but this may not be desirable in all cases. Alternatives for multivalent display have been described, such as fusions to gene VIII protein, the major coat protein of filamentous phage (Makowski 1994), or fusions to the D protein of the phage lambda capsid (Sternberg and Hoess 1995).

Selectively infectious phages (SIP) have been constructed, which obviate the panning step (Spada et al. 1997). The NH_2-terminal domain of gIIIp is replaced by the protein library, e.g. an scFv, while the ligand is chemically coupled or genetically fused to NH_2-terminal domains (Fig. 4). While the phage itself is non-infectious, the interaction of a protein from the library with the ligand restores infectivity. Based on a similar concept, ligand epitopes have been expressed as fusions to the tip of the bacterial F pilus. While this abolishes infectivity of wild-type phage, phages displaying an scFv against the peptide epitope became selectively infectious (Malmborg et al. 1997).

3.2.5 Cell-Surface Display

Many thousands of copies of protein or peptide libraries can be displayed on the surface of cells. Thus such libraries can be targeted with a fluorescent labeled ligand, the cells sorted by FACS, and grown, obviating amplification or transformation steps (Fuchs et al. 1996). In contrast to phage display, which may only recover less than 10^{-3} of library elements, cell-based systems can almost guarantee quantitative recovery of library elements. With the variety of proteins now available that can accept extensions or insertions, cell-based screens can be predicted to rapidly grow in importance (Georgiou et al. 1997).

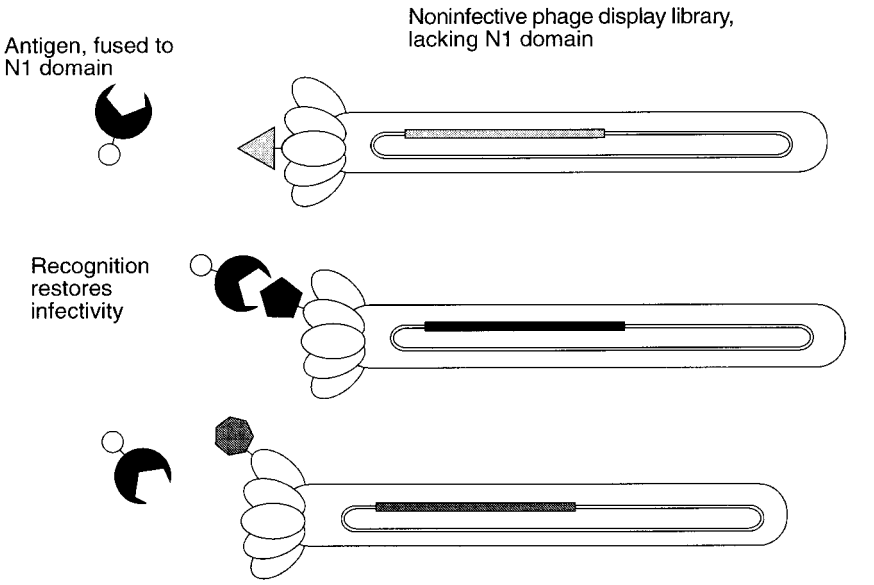

Fig. 4. Selectively infectious phage display

3.2.6 Micro-compartmentalization

It may be attractive to abstract the principle of micro-compartmentalization from cells to artificial systems. Beyond coupling information and function, entire pathways may be spatially isolated. For instance, using a simple spray-gun, droplets of 50–200nl volume can be generated that may contain substrates, cells and even synthesis beads (Borchardt et al. 1997). Stable preparations can be achieved by encapsulating the desired molecules and reactions in liposomes or in oil–water emulsions. Liposomes are the closest artificial models of cells and may even withstand the elevated temperatures required for PCR (Oberholzer et al. 1995), thus they may have potential for allowing multiple evolutionary cycles to be performed in situ. But water-in-oil emulsions may be even simpler to prepare and their use for molecular evolution has already been demonstrated by an in vitro enrichment of DNA methyltransferase genes from a 10^7-fold excess of DHFR genes (Tawfik and Griffiths 1998). Under the conditions reported by the authors, the mean droplet diameter was 2.6μm – on the order of a bacterial cell – and there were approximately 10^{10} compartments formed per milliliter.

3.3 Screening and Selection

Screening is the identification of active variants by comparing them with all other elements in a molecular library. Selection is the enrichment of active variants in a molecular library. Since screening, in principle, requires assaying every single

individual, the size of screenable libraries will be limited in practice to 10^5–10^7 sequences. Selection procedures may be used on much larger libraries and they may be more sensitive than screens, frequently requiring less than 1% of background activity for success. However, they require that the desired activity can be linked somehow to a significant growth advantage. Finally, selective pressure on living cells will induce a variety of responses, and the appearance of random phenotypic variants displaying the desired property must not be significantly more probable then its emergence as the result of library evolution.

3.3.1 Screening

Since the size of screenable libraries is limited (i.e. one can grow on the order of 10^5 isolated colonies on a large petri dish), to screen large libraries either the evolutionary process has to be broken down into more cycles, successful variants have to be enriched by a preselection step or individuals have to be pooled and singled out in successive steps.

Screening commonly relies on visual detection and much ingenuity has gone into the design of protocols that couple some function to a visual signal, commonly via activation of a reporter gene. The three most commonly used reporter enzymes are β-galactosidase, chloramphenicol acetyl transferase and luciferase (Groskreutz and Schenborn 1997). Various substrates are available for chromogenic enzymatic reactions; they are most frequently based on color changes of a nitrophenol leaving group which is released by hydrolysis of a substrate, or the precipitation of an insoluble, blue indigo dye (e.g. X-gal or BCIP/NBT). Protease activity can frequently be directly visualized by the formation of halos around colonies grown on casein- or skim milk-agar. This principle has been used successfully to screen variants of subtilisin E with increased activity (You and Arnold 1996).

Fluorescence-based screening methods are rapidly gaining importance, since they provide very high sensitivity, down to single-molecule detection, together with low background (Eigen and Rigler 1994). Fluorophore binding proteins are a common model system for molecular evolution since affinities can be well determined (Hennecke et al. 1998; Beste et al. 1999). Fluorogenic enzyme substrates have been in use for a while, with 4-methylumbelliferone being a common fluorophore. An interesting new development is the use of intramolecular fluorescent resonant energy transfer (FRET) for the detection of catalysis (Zlokarnik et al. 1998). In this work, a β-lactam-based fluorogenic substrate with a large emission wavelength shift after hydrolysis was synthesized. The expression of as few as 100 β-lactamase molecules per single cell can be detected, making this an extremely sensitive and versatile system to monitor gene expression.

Undoubtedly the most important contribution to fluorescence based screening has come from green fluorescent protein (GFP) (Tsien 1998). In only 4 years, this protein has become a standard component of the tool kits of cell biologists and protein engineers alike, and the ready visual identification of variants has made it one of the important models of evolutionary engineering. For example, an error-prone PCR amplification will produce the mutation Tyr66His with a frequency of

$\approx 10^{-4}$, a blue fluorescent mutant (BFP), which can be well distinguished from the wild-type with a hand-held UV-lamp. Wavelength shifted mutants of GFP can be used as the basis for genetically expressible intracellular sensors. For example Roger Tsien's group has fused GFP and BFP to calmodulin, constructing a sensitive FRET-based calcium sensor (Miyawaki et al. 1997), while Miesenböck et al. (1998) have used evolutionary methods and microtiter plate screens to evolve intracellular pH sensors.

Bioluminescence assays complement fluorescence as reporter systems for gene expression. Their substrates can be synthesized by the host after supplying the necessary genes in *trans* on a separate plasmid (Manen et al. 1997), and a wide variety of commercially available cloning vectors exist (Groskreutz and Schenborn 1997).

Screening has traditionally been an analysis of single bacterial colonies. The colonies can either be directly visualized, or lysed, blotted and a variety of immunochemical methods applied. A particularly elegant example utilizing colony blots, is the optimization of streptavidin to bind a peptide tag, by randomization of a surface loop and subsequent screens (Voss and Skerra 1997). For reactions that need to be quantitated, 96-well microtiter formats can be used. There is currently a vigorous effort underway to increase the number of wells and reduce the required volumes, for pharmaceutical high-throughput screens. Whether this investment in dedicated hardware will be productive, or whether ultimately modular cellular selection systems will supersede screening efforts remains to be seen.

3.3.2 Panning

Panning is most frequently employed in the selection of ligand binding molecules, by their enrichment in a pool after binding to a matrix. Thus it can be considered an in vitro selection protocol.

The question of what exactly is being selected for, e.g. by panning a phage display library, deserves some consideration. Depending on the experimental protocol, either equilibrium or dissociation rate constant govern the amount of phage retained on the binding matrix which can subsequently be eluted (Mandecki et al. 1995). Dissociation constants between nanomolar and micromolar can be determined for interactions that have been enriched (Dyson et al. 1995). Obviously, at the upper level, this is far larger than the concentration of phage in the medium. Either multivalence or rebinding affects must be invoked to explain the observed binding of intermediate affinity proteins to the matrix, over the 10–20 washing steps suggested in current protocols (Mccafferty and Johnson 1996), or the selection is in fact for slow dissociation rates. The latter interpretation is corroborated by the finding that the affinity of eluted phages correlates with the time points of collecting them, which can be monitored directly on a surface plasmon resonance chip (Malmborg et al. 1996).

As an alternative to matrix-based panning, fluorescently labeled cells can be sorted directly in a FACS. Again, GFP provides an elegant and modular access to read out a large spectrum of signals that modulate its expression. This was dem-

onstrated with the optimization of GFP itself for FACS sorting. A library of 20 randomized residues flanking the GFP fluorophore yielded variants with 100-fold increased brightness in the cell (Cormack et al. 1996).

Yet another interesting alternative involves exploiting bacterial chemotaxis to select desired functions. This has already been used in a mutational analysis of the *E. coli* chemotaxis receptor Trg (Baumgartner and Hazelbauer 1996). An improved understanding of the molecular mechanisms of bacterial responses to physical and chemical stimuli (Grebe and Stock 1998) can be expected to significantly contribute to the tool kit of available selection systems that can be functionalized.

3.3.3 Selecting for Growth

Classically, a system for selection confers a growth advantage on the cell carrying a library molecule with the desired properties. Such systems are conceptually simple and have been widely used; unfortunately they have the highest chance of false positives since a cell generally will have a number of options to cope with selective pressure. This is a rather general phenomenon, for example, an experiment with a T4-lysozyme mutant library under selective pressure to complement a β-galactosidase deficiency was not successful in altering the enzyme's substrate specificity but uncovered a novel *E. coli* locus that weakly complements the defect (Patten et al. 1996).

3.3.3.1 Functional Complementation

Obviously, functional complementation of a genetic defect will confer a selective advantage in a suitable host strain. This principle has been frequently applied: a recent example is the construction of an active dihydrofolate reductase that is formed from two fragments of the polypeptide when they are brought into proximity by two interacting proteins (Pelletier et al. 1998). A powerful extension of this principle selects proteins in host cells living in extreme environments. This strategy was first used to isolate thermostable variants of kanamycin nucleotidyltransferase, generated in an *E. coli* mutator strain, by transforming a shuttle vector into *Bacillus stearothermophilus* and selecting for growth at elevated temperatures up to 70°C (Liao et al. 1986). Subsequently, further variants were identified and combined and it could be shown that these evolved, thermostable enzymes were at the same time more resistant to protease, urea, detergents and organic solvents (Liao 1993). The same procedure has been applied to chloramphenicol acetyltransferase at 58°C (Turner et al. 1992). An obvious further improvement of this concept would be to use hyperthermophile Archaebacteria- or eubacteria. Unfortunately, molecular biology is a lot more difficult in these cells – the absence of transformable genetic elements requires chromosomal integration for recombinant expression. Nevertheless, the group of T. Oshima has recently developed a shuttle integration vector system for this purpose (Tamakoshi et al. 1997) and successfully used it to stabilize *B. subtilis* isopropylmalate dehydrogenase in *Thermus thermophilus* by gradual adaptation of the integrated gene to growth at up to 70°C in a leuB-deficient strain (Akanuma et al. 1998). Transformation of

hyperthermophiles is an active area of research, and progress is under way (Noll and Vargas 1997).

3.3.3.2 Modular Systems Based on Reporter Genes

The most universal strategy for selecting novel functions is to couple the function to the expression of some selectable reporter gene, such as antibiotic resistance. A variety of systems has been used in evolutionary engineering projects for this purpose. In eukaryotic cells, the two-hybrid system has made a profound impact with a large number of variations to screen and select for protein–protein interactions (Brachmann and Boeke 1997; Colas and Brent 1998). Prokaryotes have received less attention in this respect – partly because the motivation behind developing the two-hybrid system came from questions in cell biology, partly because eukaryotic transcriptional activation is a particularly intensely studied area. Nonetheless, there are still significant advantages to work with prokaryotes in engineering: transformation numbers are much higher, molecular biology is simpler and growth is faster. A widespread family of bacterial transcriptional regulators is the AraC/XylS protein family of "winged-helix-turn-helix" transcription factors (Gallegos et al. 1997; Martinez and Stock 1997). One member of this family, ToxR of *Vibrio cholerae*, is activated by periplasmic dimerization. Chimeras of the cytoplasmic and transmembrane segment with a periplasmic immunoglobulin domain are functional after dimerization and activate transcription of a reporter gene from the *ctx* promoter (Kolmar et al. 1994, 1995b). This system was successfully used for the screening of stabilized immunoglobulin domains (Kolmar et al. 1995a) and mutational analysis of a dimerizing transmembrane segment (Langosch et al. 1996; Brosig and Langosch 1998) and has recently been further optimized (Jappelli and Brenner 1998).

A different approach to a genetic screen has been developed, based on the phage lambda N protein which induces the modification of *E. coli* RNA polymerase to a termination-resistant form. This anti-termination screen was originally used for the identification and optimization of RNA binding peptides (Harada et al. 1996, 1997), but a generalization appears straightforward.

3.3.4 Screening and Selecting Second Site Suppressors

A powerful alternative to searching for mutations that improve a protein may be the search for second site suppressors of a previously introduced deleterious mutation. Whenever a desired property arises from additive effects, a mutation in one site may be compensated for by a sequence change in a different site. The combination of wild-type sequence and second-site suppressor can be expected to improve the protein over and above the wild-type. The advantage of this approach is that baseline activity in the screening experiment can be reduced. This greatly simplifies the detection of successful variants. While the method is general, e.g. it may allow further improvement of enzymes that already function at a level in which a further increase of activity or stability may not be readily detectable, the downside is that not all second site mutations must also improve the wild-type. An early

success with this approach was reported for ribonuclease HI, which had been previously destabilized through COOH-terminal deletions (Haruki et al. 1994). Of 11 second-site suppressor mutations that were identified, eight were also found to improve the wild-type protein. Even thermostable enzymes may be further improved with this strategy, as demonstrated for isopropylmalate dehydrogenase with a chimeric, destabilized enzyme that was subjected to random mutagenesis and selected in an auxotrophic variant of *Thermus thermophilus* at high temperature (Kotsuka et al. 1996).

In another example, a monomeric variant of chorismate mutase has been engineered by introducing point mutations at the dimer interface. The resulting monomer has almost no detectable enzymatic activity and is significantly destabilized relative to the wild-type. A library of sequences of an interhelical turn was screened for activity, resulting in a variant with almost native catalytic rates (Macbeath et al. 1998).

3.3.5 You Get (Exactly!) What You Ask For

One last caveat may be in order. Evolutionary procedures optimize a fitness function which is not completely under control of the experimenter. For instance, a careful investigation of binding determinants in antibody CDRs of a phage-displayed scFv against fluorescein, using the SIP method, demonstrated that the selection is influenced by a composite fitness function, including affinity, stability and efficient folding (Pedrazzi et al. 1997). Examples of surprising results exist, like high affinity binders to the column matrix, or enzymes with reduced activities but higher expression levels. The importance of careful experimental design must be emphasized.

4 Outlook

The speed and quality of evolutionary solutions to protein engineering problems is truly impressive, ever less knowledge is required about the system that is being optimized. One of the most pointed applications of this principle is the simultaneous engineering of a multigene operon, the arsenate resistance operon of *Staphylococcus aureus* (Crameri et al. 1997). While the wild-type plasmid conferred resistance to *E. coli* at a level of 4–10 mM arsenate, after three rounds of DNA shuffling and selection, operons were recovered that conferred resistance up to 400 mM arsenate to the host cells. In addition to ten silent mutations, only three missense mutations in the arsenite membrane pump gene, *arsB*, were sufficient for the increased resistance. Besides improving expression levels and specific activity, apparently an improved functional coupling of the proteins to each other had occurred. This impressive improvement in function, in the absence of a structural model or even a precise understanding of the molecular details of the protein's interactions or the rate-limiting step, is a good indication that evolutionary protein engineering is rapidly moving biotechnology into a new phase.

References

Airaksinen A, Hovi T (1998) Modified base compositions at degenerate positions of a mutagenic oligonucleotide enhance randomness in site-saturation mutagenesis. Nucleic Acids Res 26:576–581

Akanuma S, Yamagishi A, Tanaka N, Oshima T (1998) Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. Protein Sci 7:698–705

Atwell S, Ridgway JB, Wells JA, Carter P (1997) Stable heterodimers from remodeling the domain interface of a homodimer using a phage display library. J Mol Biol 270:26–35

Aurora R, Rose GD (1998) Helix capping. Protein Sci 7:21–38

Baca M, Scanlan TS, Stephenson RC, Wells JA (1997) Phage display of a catalytic antibody to optimize affinity for transition-state analog binding. Proc Natl Acad Sci USA 94:10063–10068

Bachmann MF, Kundig TM, Kalberer CP, Hengartner H, Zinkernagel RM (1994) How many specific B cells are needed to protect against a virus? J Immunol 152:4235–4241

Balint RF, Larrick JW (1993) Antibody engineering by parsimonious mutagenesis. Gene 137:109–118

Barbas C, Heine A, Zhong G, Hoffmann T, Gramatikova S, Bjornestedt R, List B, Anderson J, Stura EA, Wilson IA, Lerner RA (1997) Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. Science 278:2085–2092

Baumgartner JW, Hazelbauer GL (1996) Mutational analysis of a transmembrane segment in a bacterial chemoreceptor. J Bacteriol 178:4651–4660

Beste G, Schmidt FS, Stibora T, Skerra A (1999) Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold. Proc Natl Acad Sci USA 96:1898–1903

Borchardt A, Liberles SD, Biggar SR, Crabtree GR, Schreiber SL (1997) Small molecule-dependent genetic selection in stochastic nanodroplets as a means of detecting protein-ligand interactions on a large scale. Chem Biol 4:961–968

Bothmann H, Plückthun A (1998) Selection for a periplasmic factor improving phage display and functional periplasmic expression. Nature Biotech 16:376–380

Brachmann RK, Boeke JD (1997) Tag games in yeast: the two-hybrid system and beyond. Curr Op Biotech 8:561–568

Brosig B, Langosch D (1998) The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. Protein Sci 7:1052–1056

Buchholz F, Angrand PO, Stewart AF (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. Nature Biotech 16:657–662

Burton DR (1995) Phage display. Immunotech 1:87–94

Cadwell RC, Joyce GF (1994) Mutagenic PCR. PCR Methods Appl 3:S136–140

Carter P, Wells JA (1988) Dissecting the catalytic triad of a serine protease. Nature 332:564–568

Cech TR (1990) Self-splicing of group I introns. Annu Rev Biochem 59:543–568

Chen R, Henning U (1996) A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. Mol Microbiol 19:1287–1294

Chien CT, Bartel PL, Sternglanz R, Fields S (1991) The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. Proc Natl Acad Sci USA 88:9578–9582

Colas P, Brent R (1998) The impact of two-hybrid and related methods on biotechnology. Trends Biotech 16:355–363

Cormack BP, Valdivia RH, Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene 33–38

Cowan DA (1997) Thermophilic proteins: stability and function in aqueous and organic solvents. Comp Biochem Physiol A Physiol 118:429–438

Crameri A, Dawes G, Rodriguez EJ, Silver S, Stemmer WP (1997) Molecular evolution of an arsenate detoxification pathway by DNA shuffling. Nature Biotech 15:436–438

Crameri A, Raillard SA, Bermudez E, Stemmer WP (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391:288–291

Crameri A, Whitehorn EA, Tate E, Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. Nature Biotech 14:315–319

Cull MG, Miller JF, Schatz PJ (1992) Screening for receptor ligands using large libraries of peptides linked to the C terminus of the *lac* repressor. Proc Natl Acad Sci USA 89:1865–1869

Cunningham BC, Wells JA (1987) Improvement in the alkaline stability of subtilisin using an efficient random mutagenesis and screening procedure. Protein Eng 1:319–325

Czarnik AW (1997) Encoding strategies in combinatorial chemistry. Proc Natl Acad Sci USA 94: 12738–12739
Delagrave S, Goldman ER, Youvan DC (1993) Recursive ensemble mutagenesis. Protein Eng 6:327–331
Dyson MR, Germaschewski V, Murray K (1995) Direct measurement via phage titre of the dissociation constants in solution of fusion phage-substrate complexes. Nucleic Acids Res 23:1531–1535
Eigen M, Rigler R (1994) Sorting single molecules: application to diagnostics and evolutionary biotechnology. Proc Natl Acad Sci USA 91:5740–5747
Fisch I, Kontermann RE, Finnern R, Hartley O, Soler GA, Griffiths AD, Winter G (1996) A strategy of exon shuffling for making large peptide repertoires displayed on filamentous bacteriophage. Proc Natl Acad Sci USA 93:7761–7766
Fuchs P, Weichel W, Dübel S, Breitling F, Little M (1996) Separation of *E. coli* expressing functional cell-wall bound antibody fragments by FACS. Immunotech 2:97–102
Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL (1997) Arac/XylS family of transcriptional regulators. Microbiol Mol Biol Rev 61:393–410
Gates CM, Stemmer WP, Kaptein R, Schatz PJ (1996) Affinity selective isolation of ligands from peptide libraries through display on a *lac* repressor "headpiece dimer". J Mol Biol 255:373–386
Gaytan P, Yanez J, Sanchez F, Mackie H, Soberon X (1998) Combination of DMT-mononucleotide and Fmoc-trinucleotide phosphoramidites in oligonucleotide synthesis affords an automatable codon-level mutagenesis method. Chem Biol 5:519–527
Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss Rr (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nature Biotech 15:29–34
Grebe TW, Stock J (1998) Bacterial chemotaxis: the five sensors of a bacterium. Curr Biol 8:R154–R157
Griffiths AD, Duncan AR (1998) Strategies for selection of antibodies by phage display. Curr Op Biotech 9:102–108
Groskreutz D, Schenborn ET (1997) Reporter systems. Meth Mol Biol 63:11–30
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
Harada K, Martin SS, Frankel AD (1996) Selection of RNA-binding peptides in vivo. Nature 380: 175–179
Harada K, Martin SS, Tan R, Frankel AD (1997) Molding a peptide into an RNA site by in vivo peptide evolution. Proc Natl Acad Sci USA 94:11887–11892
Haruki M, Noguchi E, Akasako A, Oobatake M, Itaya M, Kanaya S (1994) A novel strategy for stabilization of *Escherichia coli* ribonuclease HI involving a screen for an intragenic suppressor of carboxyl-terminal deletions. J Biol Chem 269:26904–26911
He M, Taussig MJ (1997) Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites. Nucleic Acids Res 25:5132–5134
Hedstrom L (1996) Trypsin: a case study in the structural determinants of enzyme specificity. Biol Chem 377:465–470
Hennecke F, Krebber C, Plückthun A (1998) Non-repetitive single-chain Fv linkers selected by selectively infective phage (SIP) technology. Protein Eng 11:405–410
Hoess R, Abremski K, Sternberg N (1984) The nature of the interaction of the P1 recombinase Cre with the recombining site *loxP*. Cold Spring Harb Symp Quant Biol 49:761–768
Hoogenboom HR, de BA, Hufton SE, Hoet RM, Arends JW, Roovers RC (1998) Antibody phage display technology and its applications. Immunotech 4:1–20
Janda KD, Lo LC, Lo C, Sim MM, Wang R, Wong CH, Lerner RA (1997) Chemical selection for catalysis in combinatorial antibody libraries. Science 275:945–948
Jappelli R, Brenner S (1998) Changes in the periplasmic linker and in the expression level affect the activity of ToxR and lambda-ToxR fusion proteins in *Escherichia coli*. FEBS Lett 423:371–375
Jencks WP (1969) Catalysis in Chemistry and Enzymology, Dover, Mineola, NY
Johnson CM, Oliveberg M, Clarke J, Fersht AR (1997) Thermodynamics of denaturation of mutants of barnase with disulfide crosslinks. J Mol Biol 268:198–208
Joyet P, Declerck N, Gaillardin C (1992) Hyperthermostable variants of a highly thermostable alpha-amylase. Biotechnology (N Y) 10:1579–1583
Kauffman SA (1993) The Origins of Order. Self-organization and Selection in Evolution, Oxford University Press, New York, Oxford
Kauffman SA, Macready WG (1995) Search strategies for applied molecular evolution. J Theor Biol 173:427–440

Kayushin AL, Korosteleva MD, Miroshnikov AI, Kosch W, Zubov D, Piel N (1996) A convenient approach to the synthesis of trinucleotide phosphoramidites-synthons for the generation of oligonucleotide/peptide libraries. Nucleic Acids Res 24:3748–3755

Kiefhaber T, Grunert HP, Hahn U, Schmid FX (1990) Replacement of a *cis* proline simplifies the mechanism of ribonuclease T1 folding. Biochemistry 29:6475–6480

Kiefhaber T, Rudolph R, Kohler HH, Buchner J (1991) Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. Biotechnology (N Y) 9:825–829

Knappik A, Plückthun A (1995) Engineered turns of a recombinant antibody improve its in vivo folding. Protein Eng 8:81–89

Kolmar H, Frisch C, Gotze K, Fritz HJ (1995a) Immunoglobulin mutant library genetically screened for folding stability exploiting bacterial signal transduction. J Mol Biol 251:471–476

Kolmar H, Frisch C, Kleemann G, Gotze K, Stevens FJ, Fritz HJ (1994) Dimerization of Bence Jones proteins: linking the rate of transcription from an *Escherichia coli* promoter to the association constant of REIV. Biol Chem Hoppe Seyler 375:61–70

Kolmar H, Hennecke F, Gotze K, Janzer B, Vogt B, Mayer F, Fritz HJ (1995b) Membrane insertion of the bacterial signal transduction protein ToxR and requirements of transcription activation studied by modular replacement of different protein substructures. EMBO J 14:3895–3904

Kotsuka T, Akanuma S, Tomuro M, Yamagishi A, Oshima T (1996) Further stabilization of 3-isopropylmalate dehydrogenase of an extreme thermophile, *Thermus thermophilus*, by a suppressor mutation method. J Bacteriol 178:723–727

Krebber A, Burmester J, Plückthun A (1996) Inclusion of an upstream transcriptional terminator in phage display vectors abolishes background expression of toxic fusions with coat protein g3p. Gene 178:71–74

Kumamaru T, Suenaga H, Mitsuoka M, Watanabe T, Furukawa K (1998) Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nature Biotech 16:663–666

Lam KS, Salmon SE, Hersh EM, Hruby VJ, Kazmierski WM, Knapp RJ (1991) A new type of synthetic peptide library for identifying ligand-binding activity. Nature 354:82–84

Langosch D, Brosig B, Kolmar H, Fritz HJ (1996) Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. J Mol Biol 263:525–530

Lerner RA, Benkovic SJ, Schultz PG (1991) At the crossroads of chemistry and immunology: catalytic antibodies. Science 252:659–657

Leung D, Chen E, Goeddel D (1989) A method for Random Mutagenesis of a defined DNA segment using a Modified Polymerase Chain Reaction. Technique 1:11–15

Liao H, McKenzie T, Hageman R (1986) Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. Proc Natl Acad Sci USA 83:576–580

Liao HH (1993) Thermostable mutants of kanamycin nucleotidyltransferase are also more stable to proteinase K, urea, detergents, and water-miscible organic solvents. Enzyme Microb Technol 15:286–292

Light J, Lerner RA (1995) Random mutagenesis of staphylococcal nuclease and phage display selection. Bioorg Med Chem 3:955–967

Lyttle MH, Napolitano EW, Calio BL, Kauvar LM (1995) Mutagenesis using trinucleotide beta-cyanoethyl phosphoramidites. Biotechniques 19:274–281

MacBeath G, Kast P, Hilvert D (1998) Redesigning enzyme topology by directed evolution. Science 279:1958–1961

Makowski L (1994) Phage display: structure, assembly and engineering of filamentous bacteriophage M13. Curr Opin Struct Biol 4:225–230

Malakauskas SM, Mayo SL (1998) Design, structure and stability of a hyperthermophilic protein variant. Nature Struct Biol 5:470–475

Malmborg AC, Duenas M, Ohlin M, Soderlind E, Borrebaeck CA (1996) Selection of binders from phage displayed antibody libraries using the BIAcore biosensor. J Immunol Methods 198:51–57

Malmborg AC, Soderlind E, Frost L, Borrebaeck CA (1997) Selective phage infection mediated by epitope expression on F pilus. J Mol Biol 273:544–551

Mandecki W, Chen YC, Grihalde N (1995) A mathematical model for biopanning (affinity selection) using peptide libraries on filamentous phage. J Theor Biol 176:523–530

Manen D, Pougeon M, Damay P, Geiselmann J (1997) A sensitive reporter gene system using bacterial luciferase based on a series of plasmid cloning vectors compatible with derivatives of pBR322. Gene 186:197–200

Markland W, Ley AC, Lee SW, Ladner RC (1996) Iterative optimization of high-affinity proteases inhibitors using phage display. 1. Plasmin. Biochemistry 35:8045–8057
Martineau P, Jones P, Winter G (1998) Expression of an antibody fragment at high levels in the bacterial cytoplasm. J Mol Biol 280:117–127
Martinez HE, Stock AM (1997) Structural relationships in the OmpR family of winged-helix transcription factors. J Mol Biol 269:301–312
Mattheakis LC, Dias JM, Dower WJ (1996) Cell-free synthesis of peptide libraries displayed on polysomes. Meth Enz 267:195–207
McCafferty J, Jackson RH, Chiswell DJ (1991) Phage-enzymes: expression and affinity chromatography of functional alkaline phosphatase on the surface of bacteriophage. Protein Eng 4:955–961
McCafferty J, Johnson KS (1996) Construction and screening of antibody display libraries. Phage display of peptides and proteins. A laboratory manual (Kay, B. K., Winter, J. & McCafferty, J., Eds.), Academic Press, San Diego.
Miesenböck G, DeAngelis AD, Rothman JE (1998) Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. Nature 394:192–195
Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, Tsien RY (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature 388:882–887
Moore JC, Jin HM, Kuchner O, Arnold FH (1997) Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. J Mol Biol 272:336–347
Needels MC, Jones DG, Tate EH, Heinkel GL, Kochersperger LM, Dower WJ, Barrett RW, Gallop MA (1993) Generation and screening of an oligonucleotide-encoded synthetic peptide library. Proc Natl Acad Sci USA 90:10700–10704
Nemoto N, Miyamoto SE, Husimi Y, Yanagawa H (1997) In vitro virus: bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. FEBS Lett 414:405–408
Nicholson H, Tronrud DE, Becktel WJ, Matthews BW (1992) Analysis of the Effectiveness of Proline Substitutions and Glycine Replacements in Increasing the Stability of Phage T4 Lysozyme. Biopolymers 32:1431–1441
Noll KM, Vargas M (1997) Recent advances in genetic analyses of hyperthermophilic archaea and bacteria. Arch Microbiol 168:73–80
Oberholzer T, Albrizio M, Luisi PL (1995) Polymerase chain reaction in liposomes. Chem Biol 2:677–682
Ohage EC, Graml W, Walter MM, Steinbacher S, Steipe B (1997) b-Turn propensities as paradigms for the analysis of structural motifs to engineer protein stability. Protein Sci 6:233–241
Oliphant AR, Nussbaum AL, Struhl K (1986) Cloning of random-sequence oligodeoxynucleotides. Gene 44:177–183
Ono A, Matsuda A, Zhao J, Santi DV (1995) The synthesis of blocked triplet-phosphoramidites and their use in mutagenesis. Nucleic Acids Res 23:4677–4682
Pai LH, Wittes R, Setser A, Willingham MC, Pastan I (1996) Treatment of advanced solid tumors with immunotoxin LMB-1: an antibody linked to Pseudomonas exotoxin. Nature Med 2:350–353
Patten PA, Sonoda T, Davis MM (1996) Directed evolution studies with combinatorial libraries of T4 lysozyme mutants. Mol Divers 1:97–108
Pedrazzi G, Schwesinger F, Honegger A, Krebber C, Plückthun A (1997) Affinity and folding properties both influence the selection of antibodies with the selectively infective phage (SIP) methodology. FEBS Lett 415:289–293
Pelletier JN, Campbell VF, Michnick SW (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. Proc Natl Acad Sci USA 95:12141–12146
Pini A, Viti F, Santucci A, Carnemolla B, Zardi L, Neri P, Neri D (1998) Design and use of a phage display library. Human antibodies with subnanomolar affinity against a marker of angiogenesis eluted from a two-dimensional gel. J Biol Chem 273:21769–21776
Rebar EJ, Pabo CO (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science 263:671–673
Roberts RW, Szostak JW (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci USA 94:12297–12302
Rubingh DN (1997) Protein engineering from a bioindustrial point of view. Curr Op Biotech 8:417–422
Rudolph R, Lilie H (1996) In vitro folding of inclusion body proteins. FASEB J 10:49–56
Schatz PJ, Cull MG, Martin EL, Gates CM (1996) Screening of peptide libraries linked to lac repressor. Meth Enz 267:171–191
Schmid FX, Frech C, Scholz C, Walter S (1996) Catalyzed and assisted protein folding of ribonuclease T1. Biol Chem 377:417–424

Shafikhani S, Siegel RA, Ferrari E, Schellenberger V (1997) Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. Biotechniques 23:304–310
Shao Z, Zhao H, Giver L, Arnold FH (1998) Random-priming in vitro recombination: an effective tool for directed evolution. Nucleic Acids Res 26:681–683
Sidhu SS, Borgford TJ (1996) Selection of *Streptomyces griseus* protease B mutants with desired alterations in primary specificity using a library screening strategy. J Mol Biol 257:233–245
Singer B, Kusmierek JT (1982) Chemical mutagenesis. Annu Rev Biochem 51:655–693
Skerra A (1994) Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. Gene 151:131–135
Soderlind E, Lagerkvist AC, Duenas M, Malmborg AC, Ayala M, Danielsson L, Borrebaeck CA (1993) Chaperonin assisted phage display of antibody fragments on filamentous bacteriophages. Biotechnology (N Y) 11:503–507
Soumillion P, Jespers L, Bouchet M, Marchand BJ, Winter G, Fastrez J (1994) Selection of β-lactamase on filamentous bacteriophage by catalytic activity. J Mol Biol 237:415–422
Spada S, Krebber C, Plückthun A (1997) Selectively infective phages (SIP). Biol Chem 378:445–456
Steipe B, Schiller B, Plückthun A, Steinbacher S (1994) Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. J Mol Biol 240:188–192
Stemmer WP (1994a) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci USA 91:10747–10751
Stemmer WP (1994b) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391
Sternberg N, Hoess RH (1995) Display of peptides and proteins on the surface of bacteriophage lambda. Proc Natl Acad Sci USA 92:1609–1613
Strausberg SL, Alexander PA, Gallagher DT, Gilliland GL, Barnett BL, Bryan PN (1995) Directed evolution of a subtilisin with calcium-independent stability. Biotechnology (N Y) 13:669–673
Tamakoshi M, Uchida M, Tanabe K, Fukuyama S, Yamagishi A, Oshima T (1997) A new *Thermus-Escherichia coli* shuttle integration vector system. J Bacteriol 179:4811–4814
Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. Nature Biotech 16:652–656
Tsien RY (1998) The green fluorescent protein. Annu Rev Biochem 67:509–544
Turner SL, Ford GC, Mountain A, Moir A (1992) Selection of a thermostable variant of chloramphenicol acetyltransferase (Cat-86). Protein Eng 5:535–541
Van den Burg B, Vriend G, Veltman OR, Venema G, Eijsink VG (1998) Engineering an enzyme to resist boiling. Proc Natl Acad Sci USA 95:2056–2060
Vaughan TJ, Williams AJ, Pritchard K, Osbourn JK, Pope AR, Earnshaw JC, McCafferty J, Hodits RA, Wilton J, Johnson KS (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. Nature Biotech 14:309–314
Virnekäs B, Ge L, Plückthun A, Schneider KC, Wellnhofer G, Moroney SE (1994) Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. Nucleic Acids Res 22:5600–5607
Vispo NS, Callejo M, Ojalvo AG, Santos A, Chinea G, Gavilondo JV, Arana MJ (1997) Displaying human interleukin-2 on the surface of bacteriophage. Immunotech 3:185–193
Voss S, Skerra A (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. Protein Eng 10:975–982
Walter S, Hubner B, Hahn U, Schmid FX (1995) Destabilization of a protein helix by electrostatic interactions. J Mol Biol 252:133–143
Wang CI, Yang Q, Craik CS (1996) Phage display of proteases and macromolecular inhibitors. Meth Enz 267:52–68
Wentworth P, Janda KD (1998) Catalytic antibodies. Curr Op Chem Biol 2:138–144
Wirsching P, Ashley JA, Lo CH, Janda KD, Lerner RA (1995) Reactive immunization. Science 270:1775–1782
You L, Arnold FH (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. Protein Eng 9:77–83
Zhang X-J, Baase WA, Matthews BW (1992) Multiple alanine replacements within a-helix 126–134 of T4 lysozyme have independent, additive effects on both structure and stability. Protein Sci 1:761–776
Zhang X-j, Baase WA, Shoichet BK, Wilson KP, Matthews BW (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. Protein Eng 8:1017–1022
Zhao H, Arnold FH (1997) Optimization of DNA shuffling for high fidelity recombination. Nucleic Acids Res 25:1307–1308

Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature Biotech 16:258–261

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 β-lactamase as reporter. Science 279:84–88