Biochemistry

 Volume 40, Number 44 November 6, 2001

Current Topics

Directed Evolution of Single Proteins, Metabolic Pathways, and Viruses

Claudia Schmidt-Dannert*

Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Avenue, St. Paul, Minnesota 55126

Received June 22, 2001; Revised Manuscript Received August 21, 2001

The rapid expansion of the field of biotechnology during the past decade has resulted in an increase in the use of enzymes in biocatalytic processes. However, limited stability as well as limited catalytic activity of many biocatalysts under process conditions, which differ greatly from conditions found in biological environments, demands design of improved enzyme catalysts. Rational design strategies were therefore applied to biocatalyst optimization. Our still limited ability to precisely predict structural requirements of macromolecules such as proteins for desired functional changes, however, makes rational engineering of biomolecules a time-consuming and cumbersome process.

The idea of *directed* evolution of biomolecules in vitro on a molecular level was first introduced by pioneering work of Spiegleman and co-workers (*1*) for nucleic acids, and later, Eigen (*2*) and Kauffman (*3*) proposed a theory of molecular evolution. Arnold's group was among the first to apply the principles of molecular evolution to the creation of improved enzyme biocatalysts (*4*). They demonstrated that iterative rounds of random mutagenesis followed by screening of the generated mutant libraries could produce more stable and active enzyme variants on a time scale of months, weeks, and even days. Later, Stemmer developed a method called DNA shuffling (*5*), which allows in vitro recombination of mutants to eliminate deleterious mutations and combine beneficial mutations. The method of DNA shuffling was then extended to the in vitro recombination of homologous genes (*6*), called family shuffling, that allowed access to a larger sequence space. Directed evolution has become a widely used tool for protein design, which is now also increasingly applied to investigate the structure−function relationships of proteins.

Over the past two years, new studies on directed evolution have been published at an amazing pace. Because earlier work on directed evolution has been covered in a number of reviews (*7*−*9*), this paper will focus on the most recent examples and developments in the rapidly expanding field of directed evolution. Following a description of current strategies and methods for in vitro protein evolution, directed evolution of enzymes, binding proteins, and more complex biological systems such as multienzyme pathways and viruses will be discussed. Finally, present technologies for screening of large protein variant libraries for rare desired variants will be described. The application of molecular evolution to the design of nucleic acids, which indeed was the first application of in vitro evolution, will not be the subject of this article as it has been reviewed elsewhere (*10*, *11*).

Strategies and Methods of Directed Evolution

Although laboratory evolution aims to mimic natural evolution processes of mutagenesis, recombination, and selection, fundamental differences between both exist. First, laboratory evolution is a guided ("directed") process toward a final goal, resulting mostly in the accumulation of adaptive mutations in a protein, while natural evolution accumulates both adaptive mutations as a result of the selection pressures imposed on an organism and neutral mutations that do not affect the fitness (also termed drift) (*12*). Second, properties targeted by laboratory evolution often go beyond requirements that would make biological sense. The absence of traits

* To whom correspondence should be addressed: Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Ave., St. Paul, MN 55126. Phone: (612) 625-5782. Fax: (612) 625-5780. E-mail: schmi232@tc.umn.edu.

10.1021/bi011310c CCC: $20.00
Published on Web 10/10/2001

and combination of properties in any natural occurring protein does therefore not necessarily imply the existence of inherent physical limits but may rather reflect biological constraints.

When a strategy or problem for laboratory evolution is chosen, it must be considered that the present occupation by a protein of a specific location in a fitness landscape (biological fitness is defined as the ability of an organism to survive and reproduce) is the result of an evolutionary pathway spanning millions of years. The entire fitness landscape of a protein is made up of the fitness, or performance of a given trait in the case of laboratory evolution, of all possible amino acid sequences (X^{20} sequences for a protein with a chain length of X amino acids), which make up the sequence space for a protein (*7*). For laboratory evolution, the "performance landscape" consists of global and local minima and maxima, which will change with each property or combination of properties that is investigated. The current position of a protein in a performance landscape determines which maxima are in reach by exploring the surrounding protein sequence space using in vitro evolution strategies (discussed in ref *13*).

The simplest way to create sequence diversity is by random point mutagenesis using, for example, error-prone PCR, which is the method of choice if starting from a single protein sequence (*14*). The mutation rate is typically adjusted to a low rate of one or two amino acid changes per protein that allows for an exhaustive screen of the created library in which most of the variants are functional (*7*). If the number of amino acid substitutions is increased, the number of possible variants reaches numbers outside our current high-throughput screening abilities (libraries with 10^4-10^6 members for enzyme screens) and a large fraction of the library would contain nonfunctional variants because beneficial mutations are rare and combinations of beneficial mutations are extremely rare (*15*). In a few studies, libraries generated with higher mutation frequencies resulted in the isolation of functional variants (*16*, *17*). However, to obtain functional variants, libraries of 10^5-10^6 members were generated and screened using very high-throughput methods such as cell-based selection or FACS (fluorescence-activated cell sorting)[1] of antibody displaying cells.

Improved variants selected in a first round of random mutagenesis are typically subjected to additional iterative rounds of random mutagenesis and selection so more beneficial mutations would accumulate (Figure 1A). Recombination of improved variants using DNA shuffling or other methods (*14*, *18*) allows the combination of additive beneficial mutation and elimination of deleterious mutations for larger steps in improvement (Figure 1B).

Random mutagenesis using error-prone PCR proved to be very successful for problems such as improving the stability or activity of an enzyme (discussed below), where most likely several sequence solutions exist and mutations obtained in iterative rounds or when recombined from different improved variants are additive. However, random point mutagenesis fails if a new protein function requires the simultaneous introduction of more than one amino acid change.

Apart from the observed inherent bias of the DNA polymerases used in error-prone PCR toward certain nucleotide exchanges, another drawback of random point mutagenesis is based on the conservative nature of the genetic code. The genetic code allows from any given codon only access by single-base mutation in average codons for approximately seven different amino acid residues, typically with similar physicochemical properties. Comparison of homologous proteins suggests that conservative mutations are the driving force of natural evolution (*19*). However, Miyazaki and Arnold (*20*) showed that the introduction of nonconservative substitutions in laboratory evolution allows access to sequence space for large improvements in protein function as demonstrated by saturation mutagenesis of a hot spot identified in laboratory-evolved subtilisin S41 variants for increased thermostability.

In contrast to random mutagenesis and recombination of single-gene variants (Figure 1A,B), a much wider sequence space, including the combination of nonadditive mutations, can be explored for the discovery of novel functions, when creating chimeras of homologous genes by recombination (Figure 1C). Unlike random mutagenesis, the amino acid changes that accumulated in homologous genes are the result of mostly neutral mutations (*21*) and most importantly have already been selected for protein folding and function and, thus, result in mostly functional variants when recombined. The combination of neutral mutation into chimeric proteins, however, can create new functions, which has been described as "constructive neutral evolution" (*22*) or "selective neutrality" (*23*).

At present, "family shuffling" (*6*) is the most widely used technique for the recombination of homologous gene sequences of different origin, which has been derived from Stemmer's DNA shuffling method (*5*), and variations thereof (*24*−*26*), originally used for the recombination of single-gene sequence variants. Family shuffling has been applied to the creation of highly functional chimeric libraries from more than 20 homologous genes (*27*, *28*). In one study, shuffling of more than 20 isolated human interferon-α genes generated a chimeric library containing variants with increased antiviral and antiproliferation activities in murine cells (*28*). The best chimera of this first library was derived from six parental gene sequences and exhibited 87-fold higher antiviral activity than the best parent. Subsequent rounds of DNA shuffling yielded a chimera with activity improved 185-fold over that of the best human interferon parent. In another study, a large family of 26 subtilisin sequences with levels of DNA sequence identity as low as 56.4% (pairwise gene shuffling requires ~70% sequence identity) was shuffled and screened for four industrial

[1] Abbreviations: FACS, fluorescence-activated cell sorting; RACHITT, random chimeragenesis on transient templates; dszC, dibenzothiophene monooxygenase; ITCHY, incremental truncation for the creation of hybrid enzymes; GAR, glycinamide ribonucleotide formyltransferase; SHIPREC, sequence homology-independent protein recombination; AZT, zidovudine; TK, thymidine kinase; *p*NCA, *ω*-*p*-nitrophenoxycarboxylic acid; Fab, antibody fragment; scF, single-chain antibody; CDR, complementarity-determining region; scTCR, single-chain T-cell receptor; IGPS, indole-3-glycerol phosphate synthase; PRAI or TrpF, phosphoribosyl anthranilate isomerase; HisA, *N*′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase; TIM, triosephosphate isomerase; MLV, murine leukemia virus; CHOK1, Chinese hamster ovary cells; FRET, fluorescent resonance energy transfer; BODIPY, (*E*,*E*)-3,5-bis(4-phenyl-1,3-butadienyl)-4,4-difluoro-4-bora-3a,4a-diaza-s-indacene; PECS, periplasmic expression with cytometric screening; GFP, green fluorescent protein; IR, infrared; HPLC, high-performance liquid chromatography; GC, gas chromatography; HT, high-throughput.

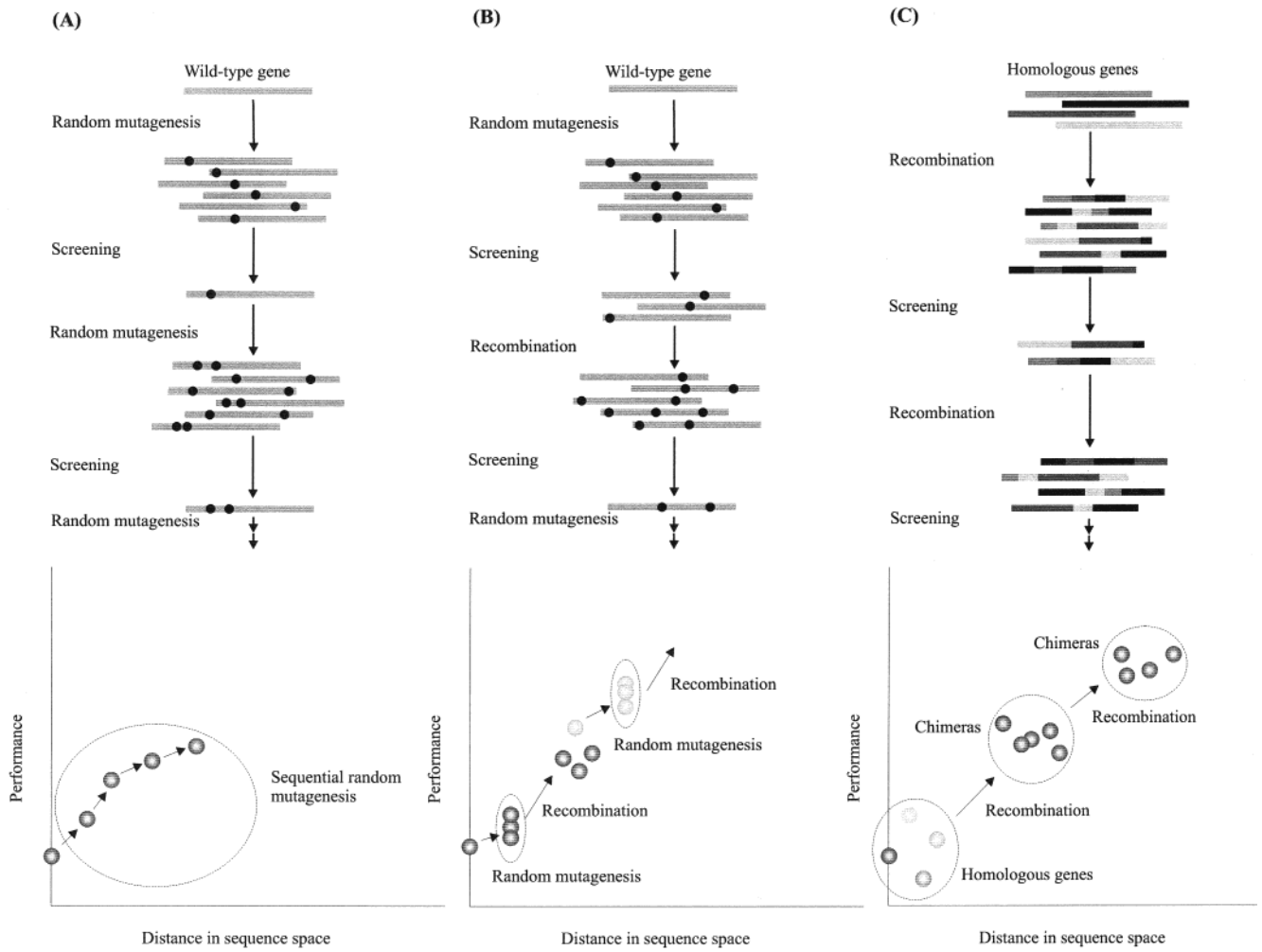

FIGURE 1: Strategies for in vitro evolution of proteins. (A) Sequential random mutagenesis followed by screening for the best improved single variant leads to the accumulation of 1−2 useful mutations per cycle. (B) Random mutagenesis followed by screening and recombination of two or more improved variants allows combination of useful mutations and elimination of deleterious mutations. (C) Recombination of homologous genes sequences (family shuffling) creates sequence chimeras for protein function with already preselected neutral mutations and allows combination of many useful mutations in one step, including nonadditive mutations.

important enzymatic properties (*27*). Although only 20% of the chimeras of the library were active, this small library contained chimeras with enzymatic properties that outperformed by far the best parental gene. In addition, the progeny chimeras displayed a wide variety of combinations of properties, e.g., thermostability and low-temperature activity, not exhibited by any of the parents. Interestingly, for some of the evolved properties, the best parent was not the closest in sequence to the best chimera, suggesting that the best parent is not always the best starting point for a single-gene evolution.

Although family shuffling proved to be a highly efficient strategy with which to address a complex directed evolution problem, a major limitation of this method is its reliance on PCR-based reassembly of short random fragments generated from homologous genes demands a high level of sequence identity of >70% and 10−15 bp stretches of continuous sequence identity between the sequences to be recombined. As a result, only highly homologous sequences can be recombined with typically only 1−4 crossovers per chimeric gene variant. In addition, some gene sequences, although highly homologous, are sometimes difficult to recombine, and mainly parental gene sequences are obtained in the library. To improve the results of a gene shuffling experiment, a "crossover allocation estimator" (*29*) has been developed, which allows in silico prediction of the probability distributions of crossover events depending on the parent genes and experimental conditions such as fragmentation length and annealing temperature.

To increase the efficiency of recombination between homologous genes, Coco et al. (*30*) described a new method for gene shuffling, RACHITT. This method is based on the hybridization of random DNA fragments on a single-stranded transient DNA scaffold, followed by trimming of overlapping flaps, filling of gaps, and ligation of nicks. Digestion of the scaffold DNA strand and generation of double-stranded DNA then creates chimeric DNA fragments. An average of 14 crossovers per gene variant were obtained with RACHITT compared to 1−4 crossovers reported in previous "shuffling" publications. Moreover, this method allowed crossovers to occur in regions with ≤5 bases of continuous sequence identity, which is important for the recombination of genes with lower levels of sequence identity. However, this method is technically more demanding than Stemmer's DNA shuffling protocol. The authors used RACHIT to create chimeras of two microbial dszC genes for improved biodesulfurization and conversion of nonnatural substrates, which they were unable to obtain using DNA shuffling (*30*).

Despite the development of a number of new methods for laboratory evolution over the past few years, our present set of methods is limited to only two mutational events, base substitutions and homologous recombination, compared to nature's manifold evolutionary toolbox for the creation of protein diversity (*31*, *32*). In particular, new mutational methods for nonhomologous recombination would be highly useful for directed evolution as it would allow exploration of the presently inaccessible sequence and, thus, function space that could be created by shuffling of structural fold homologues sharing little sequence identity as well as structural motifs and domains for new tertiary folds (*33*). The diversity of catalytic functions that can be performed by one structural fold, for example, is reflected in contemporary protein superfamilies where the members have been evolved through gene duplication and incremental mutations (*34*, *35*).

Two methods have been reported for nonhomologous recombination of genes through the creation of head-to-tail fused enzyme chimera libraries (*36*−*38*). Ostermeier et al. (*36*, *37*) described an approach, ITCHY, for the generation of a library of chimeric *Escherichia coli* and human GAR (50% DNA sequence identity) as a model system. On the basis of the introduction of nested deletions at the 5′-end of the *E. coli* and 3′-end of the human GAR followed by ligation of the truncated fragments, chimeras were created. However, the frequency of functional chimeras (e.g., in one library 111 among 1.2×10^6 clones) was very small, and the fusion points occurred predominantly in the central region of the proteins (*36*). Sieber et al. (*38*) reported on a different strategy, SHIPREC, for sequence homology-independent protein recombination that creates single-crossover chimeras with retained sequence alignment and, hence, structural fold integrity. A chimeric library of the membrane-associated mammalian cytochrome P450 1A2 (C-terminal part) and the soluble bacterial P450 BM-3 (N-terminal part) with a level of amino acid sequence identity of only ~16% was generated by intramolecular circularization of randomly truncated dimers of the two genes, which are separated by a short linker containing a restriction site, followed by linearization in the linker sequence and cloning of fragments with a size similar to that of the intact P450 genes. To select for soluble expressed chimeras with a correct reading frame, the fragments were fused in-frame with the chloramphenicol acetyltransferase gene that confers resistance to chloramphenicol when functionally expressed and is known to be a marker for soluble protein expression. Using SHIPREC, two chimeras were isolated that exhibited the catalytic activity of the mammalian P450 1A2 but were much more soluble expressed in *E. coli* because of the P450 BM-3 N-terminal part of the chimeras. The authors suggest that iterative SHIPREC could generate libraries of chimeras containing multiple crossovers.

Although the first steps toward the development of nonhomologous recombination methods have been undertaken, a method for the creation of randomly distributed crossovers between sequences with little homology at a high frequency is yet to be developed.

Improving Enzymes

Applications of in vitro evolution strategies toward the improvement of enzyme stability and activity clearly dominate. One reason for this is certainly the importance of enzyme stability and activity for biocatalytic application, but another reason may be that both properties are very good targets for directed evolution and many mutations that are in most cases additive lead to improved variants. Only recently have more complex enzyme design problems such as substrate specificity and enantioselectivity, which may require the simultaneous introduction of nonadditive mutations and the development of new screening assays, been addressed by in vitro evolution.

For biocatalytic applications, a biocatalyst must be stable under process conditions but must also be active at moderate temperatures at which many biocatalytic applications are performed. Evolutionary protein design has been successfully used either to increase the theromostability of mesophilic or psychrophilic enzymes while maintaining of increasing its activity at moderate temperatures (see, for example, refs *39* and *40*) or to increase the low-temperature activity of a thermophilic enzyme without affecting its thermostability (see, for example, refs *41*−*43*). For example, five rounds of random mutagenesis, recombination, and screening could convert a mesophilic subtilisin E to an enzyme equivalent to its thermophilic homologue thermitase (43% amino acid identity) (*39*). The best variant exhibited only eight amino acid changes that conferred a T_{opt} increase of 17 °C and a 200-fold increased half-life at 65 °C while retaining activity.

In another study, Hirano et al. (*41*) isolated from a randomly mutagenized library of ribonuclease HI from *Thermus thermophilus* three single-point mutants exhibiting increased activity at 30 °C and retained thermostability. Combination of the mutations resulted in a thermostable variant with a 40-fold increased k_{cat}/K_m value at 40 °C. However, the enzyme activity of the variant was increased at both low and higher temperatures, which was also observed with evolved low-temperature active variants of a hyperthermostable β-glucosidase (*42*) and a thermophilic indoleglycerol phosphate synthase (*43*). Low-temperature activity, however, was frequently accompanied by a significantly reduced thermostability of the variants (*42*, *43*).

These studies prompted the question of whether low-temperature activity and high-temperature stability may be incompatible. However, Wintrode et al. (*44*) showed that both properties can be evolved in vitro independent from each other, and the observed correlation of both enzyme properties in other studies may rather be the result of a screen that selected for only one of the properties. In their study, random mutagenesis of a mesophilic subtilisin (SSII) yielded in the first generation a variant with a >2-fold increased k_{cat} at 10 °C and a 1.4-fold increased half-life at 70 °C compared to those of the wild-type enzyme. The observation that both K_m and k_{cat} at 10 °C of the obtained more thermostable mutants increased suggested that the conformational flexibility required for low-temperature catalytic turnover is restricted to the active site and, hence, results in a reduced level of substrate binding, while the global stability or flexibility of the protein is not decreased (*44*). The implications of enzyme flexibility on stability and activity have recently been investigated in detail for laboratory-evolved thermostable *p*-nitrobenzyl esterase variants (*45*). The application of selective pressure for both increased thermostability and low-temperature activity during in vitro evolution

of the psychrophilic subtilisin S41 yielded variants with both increased catalytic activity and themostability (*46*).

Recently, substrate specificity has become the subject of an increasing number of in vitro evolution studies (*47*−*52*), demonstrating that laboratory evolution is an effective method for obtaining remarkable improvements of catalytic activities for new substrates requiring the remodeling of the active site. However, devising an appropriate screen that assesses the direct conversion of either the new substrate or a close substrate analogue makes directed evolution of substrate specificity challenging. The in vitro evolution of an aspartate aminotransferase with a 1 million-fold increased catalytic efficiency for the non-native substrate valine represents an impressive example for directed evolution of substrate specificity. Most interestingly, only one of the 17 mutations that accumulated in this variant appeared to be in contact with the substrate. X-ray analysis of this mutant complexed with valine showed a remodeled active site by cumulative effects of the mutations resulting in conformational changes, which would have been difficult or impossible to predict rationally (*47*). In another example, two herpesvirus TK genes with a level of DNA sequence identity of 78% were shuffled to obtain chimeras with increased phosphorylation activity of the prodrug AZT (*51*). The relatively low level of sequence diversity among the clones was sufficient to obtain after four rounds of directed evolution a best variant that sensitizes *E. coli* at 32- and 16000-fold less AZT compared to the parental TK genes due to a reduced K_m for AZT which was accompanied by a decreased specificity for thymidine. Molecular modeling suggests that the best chimeras contained contributions of both parents at important positions for substrate binding that result in active sites better suited to accommodate the azido group of AZT (*51*). In both examples, the active sites were gradually reshaped over several rounds of directed evolution through the accumulation of mutations.

A class of enzymes that have attracted particular interest for their potential in biotransformation applications are P450 monooxygenases. These enzymes catalyze the selective oxidation of an enormous range of substrates. However, low turnover numbers, cofactor requirements, low stability, and often very low expression levels of the recombinant proteins make practical application of P450s currently impossible. Rational design attempts to improve the catalytic properties of P450s have so far not yielded significantly improved biocatalysts. Two groups, however, recently demonstrated that in vitro protein evolution can be used to obtain improved P450 biocatalysts provided that the enzyme can be functionally expressed and a suitable screen is available (*53*−*55*). Joo et al. (*53*) devised a clever fluorescent whole cell screen that uses horseradish peroxidase coexpressed with the oxygenase in *E. coli* cells to measure the level of hydroxylation of an aromatic substrate, such as naphthalene, by $P450_{cam}$ from *Pseudomonas putida*. The hydroxy product is oxidized by the peroxidase to a highly fluorescent dimer/polymer enabling rapid digital imaging screening of large libraries of *E. coli* cells expressing mutated P450 (as well as dioxygenases) enzymes for improved catalytic functions (*53*, *56*). A spectrophotometrical screen suitable for high-throughput screening was also developed for P450 BM-3 (*57*). The introduction of a single-point mutation into P450 BM-3 conferred to the enzyme the ability to hydroxylate *p*NCA, which results in the release of the photometrically easy to detect nitrophenol. Because random mutagenesis of the entire P450 BM-3 gene did not yield improved variants, molecular modeling was used to identify eight positions in the P450 BM-3 structure that possibly determine the hydroxylation activity toward different chain length substrates. Random mutagenesis of these rationally identified sites resulted in mutants with improved activities toward shorter chain length substrates, which then were recombined to yield a best mutant containing five mutations that efficiently hydroxylates 8-*p*NCA, while hydroxylating 10-*p*NCA and 12-*p*NCA at a rate comparable to that of the wild-type enzyme (*55*). Surprisingly, during these directed evolution studies of P450 BM-3, several mutants were identified that hydroxylate the aromatic indole to indigo and indirubin resulting in blue *E. coli* cells (*54*). Maves and Sligar (*58*) recently reported on the laboratory evolution of less thermostable variants of a recently described and crystallized thermostable P450 monooxygenase from *Sulfolobus solfactoriys* to gain insight into the interactions responsible for its stabilizations that may be important for the stabilization of other P450s.

Enantioselectivity is another important function for the commercial application of a biocatalyst, which is extremely difficult to predict rationally (*59*−*61*). Recently, in vitro evolution proved to be a successful strategy for changing the enantioselectivities of enzymes (*62*−*65*). A lipase from *Pseudomonas aeruginosa* with almost no enantioselectivity toward the hydrolysis of the racemic 2-methyl decanoate *p*-nitrophenol ester was evolved in vitro to one with an *ee* value of 81% for the *S*-enantiomer (*62*). A combination of random mutagenesis and site-specific saturation mutagenesis with a selected set of lipase variants could increase the enantionselectivity further (*ee* > 90%), and by changing the selective pressure to the *R*-enantiomer, the evolutionary process could even be reversed and *R*-selective mutants obtained (*63*). May et al. (*64*) inverted the enantioselectivity of a hydantoinase from *Athrobacter* sp. for improved L-methionine production in a commercial whole-cell process. After two rounds of error-prone PCR and screening, one variant was obtained that was 4-fold more active than the wild-type enzyme and exhibited a reduced activity toward D-5-(2-methylthioethyl)hydantoin. Saturation mutagenesis at one site, which has been shown to relax enantioselectivity, then yielded the 5-fold more active L-selective mutant (*ee* of 20% at ∼30% conversion). This mutant increased the productivity of the whole-cell process 5-fold for >90% conversion.

Other enzyme functions, which were often resistant to prior rational design approaches but could be improved by laboratory evolution, include improved heterologous expression of functional proteins (*66*−*68*), cofactor and activator requirements (*69*, *70*), and resistance to oxidizing conditions (*71*) or chemical modifications (*72*). So far, in vitro evolution studies have been published that successfully targeted all major enzyme functions.

Optimization of Binding Proteins

Immunological diversity is created by gene rearrangements, mutations, and selection on a time scale of weeks instead of millions of years as for most proteins. This

principle of rapid in vivo evolution and the combinatorial structure of the binding proteins of the immune system make them a natural target for evolutionary protein design methods. With the development of molecular biology tools for the production of recombinant Fabs or scFvs in *E. coli* and yeasts, the exploitation of antibodies for clinical and analytical applications became rapidly possible (*73*). Another key discovery was the development of the phage-display technology that allows display of antibody fragments on the surface of a phage to screen of antibody libraries for specific binders (*74*). Subsequently, bacterial and yeast surface display technologies were developed (*75*, *76*), and more recently, in vitro ribosome display of peptides and proteins, where the polypeptide and its encoded RNA message are linked through the ribosomal complex, was applied to the display of antibodies (*77*). Alternatively, an in vitro selection method based on the covalent fusion of the synthesized polypeptide chain to its encoding RNA after in vitro translation has been described (*78*). These display technologies in combination with a binding assay, for example, the biotin−streptavidin system (*79*, *80*) or selective fluorescent labeling of antibody presenting cells followed by FACS (*75*, *81*), allow rapid screening of very large natural, synthetic, or engineered antibody libraries, which contributed greatly to the success of in vitro evolution strategies for antibody engineering. Libraries of more than 10^9 binding protein variants can rapidly be screened, compared to the much smaller size of enzyme variant libraries with up to 10^4 members that can presently be screened using well-plate enzyme activity assays. The following examples show that all important properties of binding proteins, such as specificity, affinity, stability, and folding, could be addressed by in vitro evolution using both random mutagenesis and DNA shuffling techniques. The evolved affinities go well beyond those found in nature, and frequently, different sequence solutions for a selected function were obtained. The limits for affinities that can currently be obtained by in vitro evolution appear to be defined by the sensitivity of the available selection methods.

Different combinatorial strategies have been employed to generate highly diverse antibody libraries for the selection of antibodies with high affinities and novel specificities. In one recent strategy, highly variable CDR sequences were amplified by PCR from cDNA of different human B-cell sources and randomly recombined into the appropriate six variable regions of a single master framework of human immunoglobulin genes known to be well expressed in bacteria and displayed on phages (*82*). Using this library, a number of specific scFvs against a variety of different protein, peptide, and small-molecule antigens could be selected.

Recently, Knappik et al. (*83*) synthesized a completely synthetic naïve human scFv library (2×10^9 members), covering more than 95% of the human antibody diversity, by combining the consensus sequences of seven human heavy and light chain frameworks in all different combinations and then introducing fully randomized CDR3 cassettes. After six rounds of error-prone PCR to further increase the diversity of this library followed by in vitro ribosome display and selection, high-affinity binders against bovine insulin could be enriched with affinities 40-fold higher ($K_d = 82$ pM) than those of binders selected in the synthetic library. The selected scFvs exhibited 2−9 point mutations introduced by error-prone PCR and were composed of different frameworks.

Ribosome display of a fluorescein-binding scFv antibody library generated by one round of error-prone PCR followed by several rounds of DNA shuffling and off-rate selection could improve the affinity (K_d) 30-fold to 0.04 nM (*84*). In a second experiment, five rounds of selection performed under increasingly reducing conditions during the in vitro translation and protein folding step yielded mutant scFvs (scFvs contain two conserved disulfide bridges) that not only folded more efficiently under reducing in vitro conditions but also showed a 4-fold higher level of functional periplasmic expression in *E. coli* and furthermore could be functionally expressed in the reducing cytoplasm (*84*). In both experiments, the mutations that accumulated in the selected variants were distributed over the entire sequence, resulting in very different sequence solutions for the selected functions.

The stability of a poorly folding anti-fluorescein binding antibody was initially improved by grafting its binding loops onto the framework of a well-folding human antibody (*85*). To further improve its stability, a combination of site-directed and random mutagenesis using several rounds of DNA shuffling and selection of the phage-displayed library under increasing temperature or guanidine hydrochloride stress was carried out. The best mutant with a 4 cal/mol improved thermodynamic stability exhibited only two point mutations, one of which was located in one of the CDRs and increased the binding constant 20-fold (*85*).

Therapeutical application of antibodies in, for example, cancer treatment demands antibodies that preferably bind essentially irreversibly. The mammalian immune system generates high-affinity antibodies through somatic mutation in a process called affinity maturation. However, an inherit ceiling to in vivo affinity maturation prevents the generation of antibodies with dissociation constants K_d of <0.1 nM (*86*). By evolving in vitro a scFv's fluorescin antibody with monovalent femtomolar antigen binding affinity ($K_d = 48$ fM), Boder et al. (*81*) demonstrated that there appear to be no physical constraints for affinities that go far beyond those generated in vivo. In this study, only four rounds of error-prone PCR and DNA shuffling of the entire scFv sequence, followed by yeast-surface display and optimized kinetic screening of the displayed mutant libraries, yielded a best-improved variant with a >1000-fold decreased dissociation rate. Interestingly, a number of consensus mutations identified in different selected high-affinity variants occurred in regions that are inaccessible by the in vivo somatic hypermutation process.

Recently, T-cell receptors, another group of immunological binding proteins with structures and mechanisms for the creation of diversity (however, no somatic mutation is involved) similar to those of antibodies, became amenable to in vitro biological design methods (*87*). Compared to scFv antibodies, recombinant scTCRs are less stable and poorly soluble when expressed in *E. coli* and, hence, exhibit poor surface-display capabilities. Random mutagenesis of a scTCR followed by selection of yeast surface-displayed variants could identify variants that are well-displayed while retaining antigen affinity (*87*). Interestingly, many of the identified mutations changed specific variable site residues at the α- and β-chain interface to residues that are naturally present in many variable regions of the more stable antibodies. Three

of the best-displayed variants together with the wild-type scTCR were then mixed and subjected to error-prone PCR (*88*). The library generated thusly was displayed on yeast and split for parallel FACS-assisted screening of enhanced thermoresistance and improved secretion in yeast. After three rounds of enrichment, clones with greatly improved stability were obtained with both strategies. Subsequent recombination of the selected mutations by site-directed mutagenesis created a scTCR scaffold that could resist incubation for 1 h at 65 °C, compared to 45 °C for the starting mutant, and exhibited a 20-fold increased secretion level (2000 mg/L). Unlike antibodies, TCRs do not undergo somatic mutation during in vivo affinity maturation, and the affinities reported for TCRs (10^5-10^7 M^{-1}) are lower than those for antibodies (10^5-10^{10} M^{-1}). However, in vitro evolution of the evolved stable scTCR (*88*) could increase the affinity (K_d) greatly ~100-fold to ~9 nM (*89*), demonstrating, as previously for antibodies, the absence of inherent structural limitations for high affinity.

An evolutionary design strategy has recently also been employed to increase the catalytic activity (k_{cat}) of a catalytic antibody 6−20-fold (*90*). Improved variants could be isolated from phage-displayed libraries by panning against a transition-state analogue of the antibody-catalyzed chloramphenicol prodrug activation.

Breeding of Pathways and Viruses

Metabolic pathways produce an amazing diversity of compounds with different chemical structures and biological activities. Likewise, an enormous number of pathways exist, predominantly in microorganisms, for the degradation of small molecules. It is estimated that plants alone produce hundreds of thousands of low-molecular weight compounds, ~50000 of which have been identified (*91*). Natural products provide therefore a tremendous resource for the discovery of novel useful compounds.

To access natural product diversity, heterologous pathways have been engineered into recombinant organisms such as *E. coli* for the production of small molecules. Combinatorial biosynthetic approaches, which make use of the relaxed substrate specificity often observed with enzymes involved in secondary metabolism (*92*−*94*), emerged over the past few years for the production of additional natural compounds. This approach has been particularly successful for the synthesis of novel polyketide structures because of the inherent biosynthetic flexibility of the modular polyketide synthases (*95*). However, for many metabolic pathways, combinatorial assembly of metabolic genes does not yield functional pathways.

The diversity of contemporary metabolic pathways points to the high evolutionary potential of metabolic genes to acquire new functions. In fact, the increasing availability of genome sequences and structural and biochemical data for enzymes of primary and secondary metabolism provides strong evidence that pathways mainly evolved by recruiting existing enzymes of different pathways by gene duplication (*91*, *96*−*98*). For example, the TIM barrel [$(\beta\alpha)_8$-barrel] fold is the most abundant protein fold found in many enzyme families of the central metabolism catalyzing completely different reactions (*99*), but which possibly share one common ancestor (*97*). Enzymes of secondary metabolism (including degradation pathways), unlike most enzymes of the primary metabolism, frequently are catalytically promiscuous because of relaxed substrate specificity and in some cases can produce more than one product (*91*, *96*, *98*). The plasticity of secondary metabolism has been proposed to be the result of selection for traits that enhance the generation and retention of chemical diversity without increasing the fitness costs (*96*).

Directed evolution strategies together with those of metabolic engineering should therefore allow mimicking of natural metabolic evolution by recruitment of metabolic genes from different pathways and matching and changing of their catalytic functions to create novel pathways for the production and discovery of natural and unnatural molecular diversity as well for the degradation of xenobiotic compounds. First studies have recently been published that demonstrate the evolvability in vitro of metabolic enzymes and pathways.

Two recent studies used in vitro evolution to mimic the evolution of the $(\alpha\beta)_8$-barrel scaffold by transplanting either a new catalytic activity on a scaffold with an existing binding site (*100*) or a new binding site on a scaffold with an existing catalytic activity (*101*). In one study, Altamiro et al. (*100*) used the IGPS scaffold to graft onto its binding site the catalytic activity of PRAI by first engineering the catalytic loop system of PRAI into IGPS followed by two rounds of in vitro recombination and in vivo selection in *E. coli*. The selected best clone exhibited a 6-fold higher k_{cat}/K_m than the wild-type PRAI, but shared a level of sequence identity with PRAI of only 28% while sharing a level of identity of 90% with IGPS. In a second study, Juergens et al. (*101*) sought to convert HisA involved in histidine biosynthesis into an enzyme that catalyzes a similar reaction in the tryptophane biosynthetic pathway catalyzed by TrpF (synonym for PRAI). Although TrpF and HisA share the same fold, they lack amino acid similarity. In vitro evolution of HisA generated two variants with TrpF activity, and one of the variants retained significant HisA activity. More recently, combinatorial mutagenesis and functional selection has been applied to investigate the sequence elements that specify the $(\alpha\beta)_8$-barrel fold using the prototypical TIM. Surprisingly, at 142 of 182 positions that were investigated, mutation to at least one of the seven residues that was tested (FVLAKEQ) resulted in a functional TIM (*102*).

We have extended the concept of molecular breeding to the creation of novel multienzyme pathways for the production of novel carotenoids in *E. coli* by combining carotenoid biosynthetic genes from different microorganisms into pathways and evolving new catalytic functions (*103*) (Figure 2). In a first step, two genes were introduced in *E. coli* to produce phytoene, the first carotenoid precursor. This pathway was then extended with a library of two shuffled desaturase genes from Erwinia strains, which normally catalyze the desaturation of phytoene to lycopene. Among the thousands of colored variants that were screened, several yellow variants were found that introduced fewer desaturations into phytoene and one pink clone was identified that produced the fully conjugated linear carotenoid tetradehydrolycopene. Further extension of the evolved tetradehydrolycopene pathway with a library of shuffled cyclase genes from Erwinia strains, which usually cyclize lycopene to form β-carotene, generated a kaleidoscope of different colored

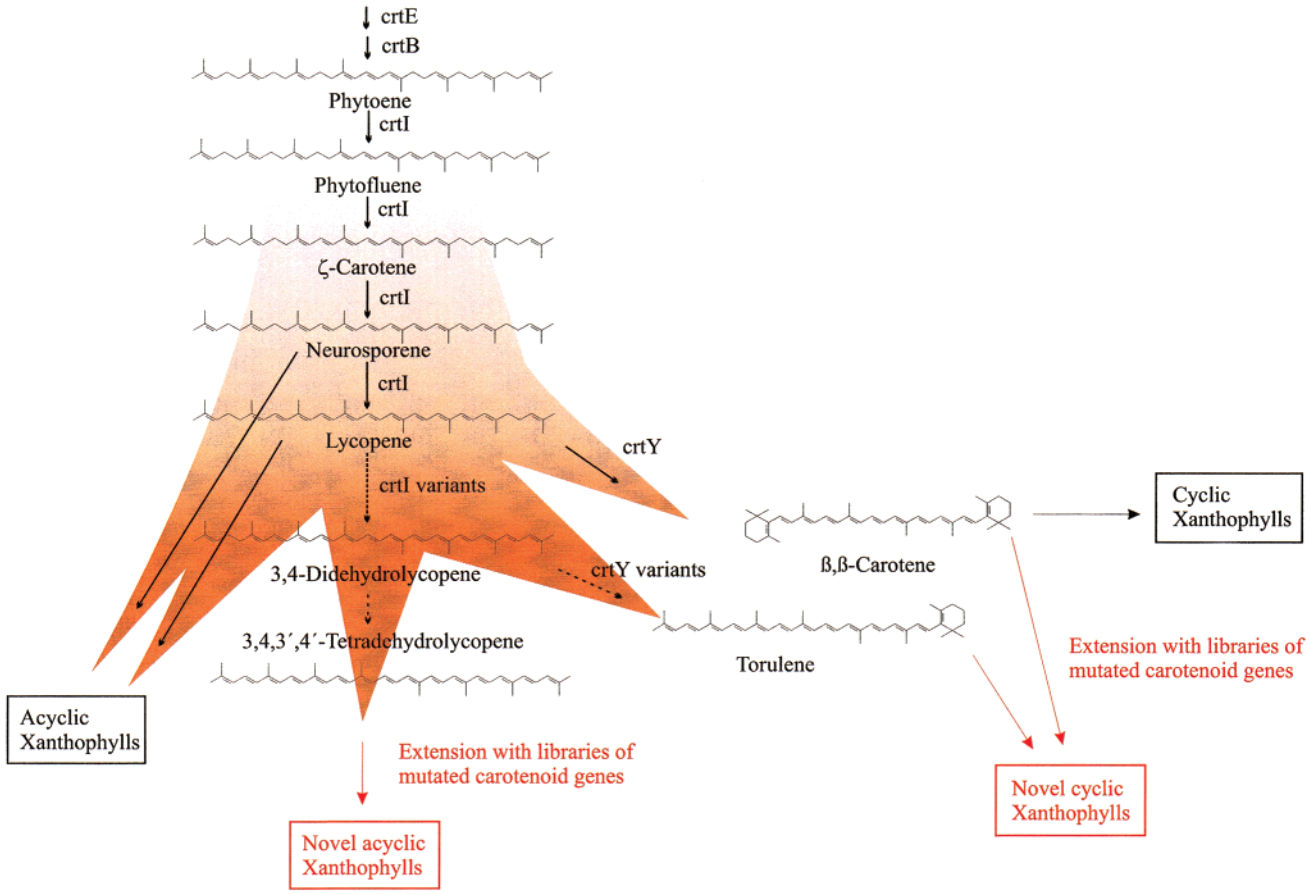

FIGURE 2: Directed evolution of carotenoid biosynthetic pathways (*103*). Carotenoid biosynthesis was engineered into *E. coli* by the introduction of two biosynthetic genes (crt E, geranylgeranyl diphosphate synthase, and crtB, phytoene synthase) required for the synthesis of the first carotenoid phytoene. This pathway was extended with a library of shuffled phytoene desaturase (crtI) genes, creating a novel pathway in *E. coli* for the production of the fully conjugated tetradehydrolycopene. Extension of this pathway with additional evolved biosynthetic genes can create a kaleidoscope of novel cyclic and acyclic carotenoid structures as shown by the production of the novel cyclic carotenoid torulene by a variant cyclase (crtY) in *E. coli*. The torulene-producing *E. coli* has been isolated among a number of colored *E. coli* variants expressing different evolved carotenoid pathways.

variants. Many of the new variants accumulated different ratios and quantities of acyclic and cyclic carotenoid intermediates, but one selected bright-red variant contained a novel pathway for the production of torulene in *E. coli*, a carotenoid not found in bacteria but produced in some red yeast by a different pathway. The evolved as well as the known wild-type carotenoid biosynthetic pathways can be extended with additional libraries of mutated carotenoid biosynthetic genes such as hydroxylase, oxygenases, and glycosylases to create a variety of novel pathways for the production of new carotenoids in *E. coli*.

Not only can the products of metabolic pathways be altered by directed evolution, but also the metabolic flux can be changed as has be shown for an engineered carotenoid biosynthetic pathway (*104*). Wang et al. (*104*) subjected geranylgeranyl diphosphate synthase from *Archaeglobus fulgidus*, which is a rate-controlling enzyme of the engineered carotenoid pathway in *E. coli*, to one round of random mutagenesis followed by screening of the resulting library for colonies with brighter orange colors indicating improved astaxanthin production. The selected variants improved lycopene production in *E. coli* up to 2-fold.

The recent studies have shown that the principles of directed evolution can be successfully applied to the design of complex biological systems such as metabolic pathways. The next step would be to apply in vitro evolution to the design of systems of an even higher level of complexity, to the laboratory evolution of viruses and even complete genomes. Although directed evolution of whole genomes has not yet been reported (studies, however, are in progress), two recent studies report on the laboratory evolution of retroviruses (*105*, *106*). Retroviruses have been shown to evolve very rapidly through recombination of their genomes at high frequencies. The studies showed that the inherent high evolutionary potential of retroviruses could be used to breed optimized viruses for gene therapy and vaccine applications. In contrast to natural recombination of retroviruses, DNA shuffling allowed recombination of sequences from several retroviruses and therefore moving beyond the diversity created by nature. In a first example, Soong et al. (*105*) shuffled the envelope sequence of six MLVs and selected for chimeras with a new tropism for CHOK1 cells. After several passages of infection and selection, one dominant chimeric virus sequence containing sequence elements from four of the parents was obtained that efficiently transfected CHOK1 cells not transfected by any of the parental retroviruses. In a subsequent study carried out by Powell et al. (*106*), the envelope sequences of the same six MLV viruses were shuffled to select for chimeras with improved stabilities under downstream process conditions

used for the purification of high-titer preparations for gene therapy applications. The selected chimeras, recombinants from several different parents, were resistant to centrifugation-stress conditions that reduced the titers of the parental viruses by 30−100-fold.

Searching Protein Libraries

Because in vitro-directed evolution experiments depend on the availability of appropriate screens for the identification of the desired rare variants in large libraries, the development of a broad range of screening technologies is critical to the issue of which design problems can be solved by using evolutionary strategies. Typically, each evolutionary design problem requires its own individual screening solution, and often, a considerable amount of time needs to be spent on devising a screening method.

Enzyme activities in libraries are most commonly screened by means of either direct assays performed in microtiter plates that detect the formation of a colored or fluorescing reaction product (*42*, *46*, *64*, *65*, *71*) or solid-phase assays (e.g., agar plates or filter plates) (*69*, *72*, *107*). In some cases, the desired new catalytic activity can be directly linked to the phenotype of the expression host by, for example, complementation of a host auxotrophy or conferring resistance to an antibiotic or other cytoxic agent, which than allows screening of large libraries of $>10^6$ members by in vivo selection (*41*, *47*, *51*, *108*). However, the applied selective pressures on a microbial host are often bypassed through occurring adaptive events, because of the high adaptability of microbial metabolism to changed environmental conditions (*109*).

Solid-phase screening of colonies on agar plates or on filters after colony transfer allows rapid screening of large libraries. Typically, variants are identified on the basis of the cleavage of a substrate present in the solid phase and detection of product release by a local change in pH using a pH indicator (*110*), formation of a colored or fluorescent product (*72*), development of a clearing zone or precipitation caused by the altered solubility of the released product in the solid phase (*52*), or monitoring the growth rate depending on the cleavage of an insoluble substrate (*111*). At present, most solid-phase screening methods lack sensitive quantification of catalytic activity and therefore are typically qualitative screens suitable for a first round of directed evolution. Recently, however, a sensitive fluorescent high-throughput digital imaging screen has been developed for P450 monooxygenase expressed in *E. coli* (*56*). Digital imaging is a powerful technique for quantitative solid-phase screening of large libraries if a suitable fluorescent or chromogenic substrate is available. For example, KAIROS, a California-based company, developed HT-solid-phase digital imaging spectroscopy screening methods and technologies for the directed evolution of catalytic activities.

Quantitative colorimetric or fluorimetric direct assays using plate readers are performed on either cells (whole or lysed) or culture supernatant depending on the location of the expressed enzyme variants and the substrate accessibility (*42*, *46*, *64*, *65*, *71*). Although more time-consuming, these assays offer the advantage of tailoring the assay conditions according to the desired functions to be evolved. Because assays are often performed on substrate analogues, the activity for other substrates, including that of actual practical interest, is subject to genetic drift, and the substrate profile of the identified variants must therefore be verified to ensure activity for the desired substrate (*43*, *61*). Other important properties such as thermostability are also subject to drift when no selection pressure is applied.

On the basis of the technologies developed for the display of binding proteins on phages and microbial cells (*112*), similar technologies emerged for the display of enzyme libraries (*113*−*116*). Surface display of protein libraries is a powerful strategy as it directly links the genetic information to the phenotype of the displayed proteins, which are fully accessible for substrates and ligands. However, in contrast to binding proteins, screening of displayed enzyme libraries requires coupling of catalytic activity to a specific phage or microbial cell, because the formed reaction products diffuse away. In one strategy, the reaction compounds of an enzymatic conversion are therefore anchored to the surface of a phage in the proximity of the displayed enzyme (*117*−*119*), allowing their use for the identification of the phage displaying the active enzyme. In another strategy, enzyme catalysts are selected on the basis of binding to an immobilized transition-state analogue (*49*). Yet two other approaches make use of the particular catalytic properties of the phage-displayed enzymes (*120*, *121*). Ponsard et al. (*120*) used catalytic elution of phage-displayed metallo-β-lactamase to enrich the activity of a mutant library 60-fold. The peptide bond formation activity of subtiligase has been employed by Atwell and Wells (*121*) to bind a biotinylated peptide substrate to an amino-terminal extension of the enzyme for enrichment of displayed subtiligase variants on a biotin-binding matrix.

FACS has become a highly efficient and rapid screening technology for large surface-displayed binding protein libraries (*89*, *122*, *123*). Flow cytometry of microbial cells (*124*) expressing enzyme activities, however, is more difficult because either the enzyme substrate must be transported into the cell and then immediately specifically converted to a fluorescent product that is retained inside the cell without affecting its viability or, if the enzyme is surface-displayed, the generated fluorescent product must be physically linked to the displayed enzyme. Moreover, the generated fluorescence signal must correlate to the enzyme activity.

Two new FACS screening strategies for surface-displayed enzymes (*125*) and periplasmically expressed binding proteins (*126*) have recently been developed in the laboratories of Iverson and Georgiou. The surface display strategy (*125*) makes use of the negatively charged surface of Gram-negative bacteria such as *E. coli*, which allows binding of a positively charged FRET substrate. The FRET substrate consists of a neutral FRET quenching partner released upon enzymatic conversion, and a positively charged fluorophore bound to the cell surface. A FRET substrate for proteolytic cleavage by the displayed OmpT protein was designed that contains the BODIPY (positively charged C-terminal cleavage product) and tetramethylfluorophore (uncharged N-terminal cleavage product) on either side of an Arg−Arg peptide bond. Using this substrate, cells displaying no OmpT activity or different levels of OmpT activity could be discriminated by FACS. OmpT variants with 60-fold enhanced catalytic activity for Arg−Val cleavage could be isolated from a randomly mutagenized library with a FRET

substrate containing a nonsubstrate Arg−Val peptide bond. The second strategy, PECS (*126*), is based on selective fluorescent labeling of periplasmically expressed proteins. The secretion of proteins into the periplasmic space of *E. coli* allows production of proteins that cannot be functionally and stably displayed on a cell or phage surface. Using PECS, digoxin binding scFvs with sub-nanomolar K_d values were isolated from libraries with a low-throughput benchtop flow cytometer. The authors found that under certain conditions, large ligands (up to 10 kDa) can diffuse into the periplasmic space, making PECS very useful for the screening of a number of binding proteins, including nucleic acid binding proteins. A modified PECS method may even be applicable to the screening of enzyme libraries.

Other new screening technologies for directed in vitro evolution that recently emerged include directed evolution of a polymerase by compartmentalized self-replication where polymerase variants replicate only their own genes (*127*) and the development of an araC-based three-hybrid system (named QUEST) (*128*). QUEST has been developed for the intracellular detection of scytalone dehydratase activity, but may be modified for the detection of other enzyme activities as well.

Two protein folding assays have been reported to be suitable for the directed evolution of expression of soluble recombinant proteins (*129*, *130*). Waldo et al. (*130*) constructed a "folding reporter" vector where the protein of interest is N-terminally fused to GFP as a folding reporter. The fluorescence of *E. coli* cells directly correlated with the production of folded, soluble proteins fused to GFP. More soluble variants of a bacteriophage F1 mutant gene V protein, and the H subunit of ferritin could be isolated after several rounds of directed evolution using this assay. A different strategy based on the structural complementation of the α- and ω-fragments of β-galactosidase was employed by Wigley et al. (*129*). Here, the protein of interest is N-terminally fused to the α-fragment of β-galactosidase. Only soluble proteins can interact with the ω-fragment to form functional β-galactosidase that can be assayed in vivo or in vitro using chromogenic substrates.

High-throughput assays for direct sensitive detection of substrate conversion are required for directed evolution of many catalytic functions. At present, a number of direct methods such as fluorescence-based, IR-thermographic, and circular dichroism assays (HPLC, GC, and capillary electrophoresis methods) are being exploited for HT screening of enzyme activities (*61*). Without a doubt, many new methods and devices developed for the ever increasing HT needs of combinatorial chemistry and drug discovery as well as genome sequencing will benefit the development of new HT methods, e.g., assay miniaturization and automation, for directed protein evolution.

Conclusions

Evolutionary strategies proved to be successful for many design problems that oftentimes were resistant to rational approaches. With methods for laboratory evolution in place that allow the generation of desired improved protein variants on a time scale of weeks, researchers focus now on the development of new screening and sequence-independent recombination technologies and on the expansion of in vitro evolution to complex biological system such as pathways, viruses, and even complete genomes. The enormous number of new gene sequences that are derived from genome sequencing projects will provide a tremendous toolbox from which to select genes for directed evolution of new biocatalysts and metabolic pathways.

ACKNOWLEDGMENT

I thank Dr. Arjo de Boer for critically reading the manuscript.

REFERENCES

1. Mills, D. R., Peterson, R. L., and Spiegelman, S. (1967) *Proc. Natl. Acad. Sci. U.S.A. 58*, 217−224.
2. Eigen, M. (1984) *Adv. Chem. Phys. 55*, 119−137.
3. Kauffman, S. (1993) *The origins of order: self-organization and selection in evolution*, Oxford University Press, New York.
4. Chen, K., and Arnold, F. H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5618−5622.
5. Stemmer, W. P. C. (1994) *Nature 370*, 389−391.
6. Crameri, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. C. (1998) *Nature 391*, 288−291.
7. Arnold, F. H., and Wintrode, P. L. (1999) in *Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis and Bioseparation* (Flickinger, M. C., and Drew, S. W., Eds.) pp 971−987, John Wiley & Sons, New York.
8. Tobin, M. B., Gustafsson, C., and Huisman, G. W. (2000) *Curr. Opin. Struct. Biol. 10*, 421−427.
9. Liu, D. R., and Schultz, P. G. (1999) *Angew. Chem., Int. Ed. 38*, 36−54.
10. Landweber, L. F. (1999) *Trends Ecol. Evol. 14*, 353−358.
11. Schimmel, P., and Kelley, S. O. (2000) *Nat. Struct. Biol. 7*, 5−7.
12. Benner, S. A., and Ellington, A. D. (1990) *Bioorg. Chem. Front. 1*, 1−70.
13. Voigt, C. A., Kauffman, S., and Wang, Z. G. (2001) *Adv. Protein Chem.*, 79−160.
14. Zhao, C., Moore, J. C., Volkov, A. A., and Arnold, F. H. A. (1999) in *Manual of Industrial Microbiology and Biotechnology* (Demain, A. L., and Davies, J. D., Eds.) pp 597−604, ASM Press, Washington, DC.
15. Arnold, F. H. A. (1996) *Chem. Eng. Sci. 51*, 5091−5102.
16. Christians, F. C., and Loeb, L. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 6124−6128.
17. Daugherty, P. S., Chen, G., Iverson, B. L., and Georgiou, G. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2029−2034.
18. Volkov, A. A., and Arnold, F. H. (2000) *Method. Enzymol. 328*, 447−456.
19. Go, M., and Miyazawa, S. (1980) *Int. J. Pept. Protein Res. 15*, 211−224.
20. Miyazaki, K., and Arnold, F. H. (1999) *J. Mol. Evol. 49*, 716−720.
21. Kimura, M. (1968) *Nature 217*, 624−626.
22. Stoltzfus, A. (1999) *J. Mol. Evol. 49*, 169−181.
23. Demetrius, L. (1997) *J. Mol. Evol. 45*, 370−377.
24. Kikuchi, M., Ohnishi, K., and Harayama, S. (2000) *Gene 243*, 133−137.
25. Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) *Gene 236*, 159−167.
26. Zhao, H. M., Giver, L., Shao, Z. X., Affholter, J. A., and Arnold, F. H. (1998) *Nat. Biotechnol. 16*, 258−261.
27. Ness, J., Welch, M., Giver, L., Bueno, M., Cherry, J., Borchert, T., Stemmer, W., and Minshull, J. (1999) *Nat. Biotechnol. 17*, 893−896.
28. Chang, C. C. J., Chen, T. T., Cox, B. W., Dawes, G. N., Stemmer, W. P. C., Punnonen, J., and Patten, P. A. (1999) *Nat. Biotechnol. 17*, 793−797.
29. Moore, G. L., Maranas, C. D., Lutz, S., and Benkovic, S. J. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 3226−3231.

30. Coco, W. M., Levinson, W. E., Crist, M. J., Hektor, H. J., Darzins, A., Pienkos, P. T., Squires, C. H., and Monticello, D. J. (2001) *Nat. Biotechnol. 19*, 354−359.
31. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) *Science 278*, 631−637.
32. Henikoff, S., Greene, E. A., Pietrokovski, S., Bork, P., Attwood, T. K., and Hood, L. (1997) *Science 278*, 609−614.
33. Bogarad, L. D., and Deem, M. W. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 2591−2595.
34. O'Brien, P. J., and Herschlag, D. (1999) *Chem. Biol. 6*, R91−R105.
35. Todd, A. E., Orengo, C. A., and Thornton, J. M. (2001) *J. Mol. Biol. 307*, 1113−1143.
36. Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) *Nat. Biotechnol. 17*, 1205−1209.
37. Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 3562−3567.
38. Sieber, V., Martinez, C. A., and Arnold, F. H. (2001) *Nat. Biotechnol. 19*, 456−460.
39. Zhao, H. M., and Arnold, F. H. (1999) *Protein Eng. 12*, 47−53.
40. Gonzalez-Blasco, G., Sanz-Aparicio, J., Gonzalez, B., Hermoso, J. A., and Polaina, J. (2000) *J. Biol. Chem. 275*, 13708−13712.
41. Hirano, N., Haruki, M., Morikawa, M., and Kanaya, S. (2000) *Biochemistry 39*, 13285−13294.
42. Lebbink, J. H. G., Kaper, T., Bron, P., van der Oost, J., and de Vos, W. M. (2000) *Biochemistry 39*, 3656−3665.
43. Merz, A., Yee, M. C., Szadkowski, H., Pappenberger, G., Crameri, A., Stemmer, W. P. C., Yanofsky, C., and Kirschner, K. (2000) *Biochemistry 39*, 880−889.
44. Wintrode, P. L., Miyazaki, K., and Arnold, F. H. (2000) *J. Biol. Chem. 275*, 31635−31640.
45. Gershenson, A., Schauerte, J. A., Giver, L., and Arnold, F. H. (2000) *Biochemistry 39*, 4658−4665.
46. Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., and Arnold, F. H. (2000) *J. Mol. Biol. 297*, 1015−1026.
47. Oue, S., Okamoto, A., Yano, T., and Kagamiyama, H. (1999) *J. Biol. Chem. 274*, 2344−2349.
48. Doyle, S. A., Fung, S. Y. F., and Koshland, D. E. (2000) *Biochemistry 39*, 14348−14355.
49. Hansson, L. O., Widersten, M., and Mannervik, B. (1999) *Biochem. J. 344*, 93−100.
50. Hansson, L. O., Bolton-Grob, R., Massoud, T., and Mannervik, B. (1999) *J. Mol. Biol. 287*, 265−276.
51. Christians, F. C., Scapozza, L., Crameri, A., Folkers, G., and Stemmer, W. P. C. (1999) *Nat. Biotechnol. 17*, 259−264.
52. Kauffman, I., and Schmidt-Dannert, C. (2001) *Protein Eng.* (in press).
53. Joo, H., Lin, Z. L., and Arnold, F. H. (1999) *Nature 399*, 670−673.
54. Li, Q. S., Schwaneberg, U., Fischer, P., and Schmid, R. D. (2000) *Chemistry 6*, 1531−1536.
55. Li, Q. S., Schwaneberg, U., Fischer, M., Schmitt, J., Pleiss, J., Lutz-Wahl, S., and Schmid, R. D. (2001) *Biochim. Biophys. Acta 1545*, 114−121.
56. Joo, H., Arisawa, A., Lin, Z. L., and Arnold, F. H. (1999) *Chem. Biol. 6*, 699−706.
57. Schwaneberg, U., Schmidt-Dannert, C., Schmitt, J., and Schmid, R. D. (1999) *Anal. Biochem. 269*, 359−366.
58. Maves, S. A., and Sligar, S. G. (2001) *Protein Sci. 10*, 161−168.
59. Patel, R. N., Ed. (1999) Stereoselective Biocatalysis, Marcel Dekker, New York.
60. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) *Nat. Biotechnol. 409*, 258−268.
61. Reetz, M. T. (2001) *Angew. Chem., Int. Ed. 40*, 284−310.
62. Reetz, M. T., Zonta, A., Schimossek, K., Liebeton, K., and Jaeger, K. E. (1997) *Angew. Chem., Int. Ed. 36*, 2830−2832.
63. Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B. W., Reetz, M. T., and Jaeger, K. E. (2000) *Chem. Biol. 7*, 709−718.
64. May, O., Nguyen, P. T., and Arnold, F. H. (2000) *Nat. Biotechnol. 18*, 317−320.
65. Fong, S., Machajewski, T. D., Mak, C. C., and Wong, C. H. (2000) *Chem. Biol. 7*, 873−883.
66. Lin, Z. L., Thorsen, T., and Arnold, F. H. (1999) *Biotechnol. Prog. 15*, 467−471.
67. Morawski, B., Lin, Z. L., Cirino, P., Joo, H., Bandara, G., and Arnold, F. H. (2000) *Protein Eng. 13*, 377−384.
68. Kim, G. J., Cheon, Y. H., and Kim, H. S. (2000) *Biotechnol. Bioeng. 68*, 211−217.
69. Allen, S. J., and Holbrook, J. J. (2000) *Protein Eng. 13*, 5−7.
70. Levine, K., Kiang, L., Jacobson, M. D., Fisher, R. P., and Cross, F. R. (1999) *Mol. Cell 4*, 353−363.
71. Cherry, J. R., Lamsa, M. H., Schneider, P., Vind, J., Svendsen, A., Jones, A., and Pedersen, A. H. (1999) *Nat. Biotechnol. 17*, 379−384.
72. Matsumura, I., Wallingford, J., Surana, N., Vize, P., and Ellington, A. (1999) *Nat. Biotechnol. 17*, 696−701.
73. Verma, R., Boleti, E., and George, A. J. T. (1998) *J. Immunol. Methods 216*, 165−181.
74. Hoogenboom, H. R., and Chames, P. (2000) *Immunol. Today 21*, 371−378.
75. Daugherty, P. S., Olsen, M. J., Iverson, B. L., and Georgiou, G. (1999) *Protein Eng. 12*, 613−621.
76. Boder, E. T., and Wittrup, K. D. (2000) *Method. Enzymol. 328*, 430−444.
77. Schaffitzel, C., Hanes, J., Jermutus, L., and Plückthun, A. (1999) *J. Immunol. Methods 231*, 119−135.
78. Roberts, R. W., and Szostak, J. W. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 12297−12302.
79. Hanes, J., Schaffitzel, C., Knappik, A., and Plückthun, A. (2000) *Nat. Biotechnol. 18*, 1287−1292.
80. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R., and Plückthun, A. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 14130−14135.
81. Boder, E. T., Midelfort, K. S., and Wittrupt, K. D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 10701−10705.
82. Soderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Aberg, A. M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C., Danielsson, L., Carlsson, R., and Borrebaeck, C. A. K. (2000) *Nat. Biotechnol. 18*, 852−856.
83. Knappik, A., Ge, L. M., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Plückthun, A., and Virnekas, B. (2000) *J. Mol. Biol. 296*, 57−86.
84. Jermutus, L., Honegger, A., Schwesinger, F., Hanes, J., and Plückthun, A. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 75−80.
85. Jung, S., Honegger, A., and Plückthun, A. (1999) *J. Mol. Biol. 294*, 163−180.
86. Batista, F. D., and Neuberger, M. S. (1998) *Immunology 8*, 751−759.
87. Kieke, M. C., Shusta, E. V., Boder, E. T., Teyton, L., Wittrup, K. D., and Kranz, D. M. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 5651−5656.
88. Shusta, E. V., Holler, P. D., Kieke, M. C., Kranz, D. M., and Wittrup, K. D. (2000) *Nat. Biotechnol. 18*, 754−759.
89. Holler, P. D., Holman, P. O., Shusta, E. V., O'Herrin, S., Wittrup, K. D., and Kranz, D. M. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5387−5392.
90. Takahashi, N., Kakinuma, H., Liu, L., Nishi, Y., and Fujii, I. (2001) *Nat. Biotechnol. 19*, 563−567.
91. Pichersky, E., and Gang, D. R. (2000) *Trends Plant Sci. 5*, 439−445.
92. Bentley, R., and Bennett, J. W. (1999) *Annu. Rev. Microbiol. 53*, 411−446.
93. Albrecht, M., Takaichi, S., Steiger, S., Wang, Z. Y., and Sandmann, G. (2000) *Nat. Biotechnol. 18*, 843−846.
94. Schmidt-Dannert, C. (2000) *Curr. Opin. Biotechnol. 11*, 255−261.
95. Khosla, C., and Harbury, P. B. (2001) *Nature 409*, 247−252.
96. Firn, R. D., and Jones, C. G. (2000) *Mol. Microbiol. 37*, 989−994.
97. Copley, R. R., and Bork, P. (2000) *J. Mol. Biol. 303*, 627−640.
98. Copley, S. D. (2000) *Trends Biochem. Sci. 25*, 261−265.
99. Wierenga, R. K. (2001) *FEBS Lett. 492*, 193−198.

100. Altamirano, M. M., Blackburn, J. M., Aguayo, C., and Fersht, A. R. (2000) *Nature 403*, 617−622.
101. Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M., and Sterner, R. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 9925−9930.
102. Silverman, J. A., Balakrishnan, R., and Harbury, P. B. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 3092−3097.
103. Schmidt-Dannert, C., Umeno, D., and Arnold, F. H. (2000) *Nat. Biotechnol. 18*, 750−753.
104. Wang, C. W., Oh, M. K., and Liao, J. C. (2000) *Biotechnol. Prog. 16*, 922−926.
105. Soong, N. W., Nomura, L., Pekrun, K., Reed, M., Sheppard, L., Dawes, G., and Stemmer, W. P. C. (2000) *Nat. Genet. 25*, 436−439.
106. Powell, S. K., Kaloss, M. A., Pinkstaff, A., McKee, R., Burimski, I., Pensiero, M., Otto, E., Stemmer, W. P. C., and Soong, N. W. (2000) *Nat. Biotechnol. 18*, 1279−1282.
107. Stefan, A., Radeghieri, A., Rodriguez, A., and Hochkoeppler, A. (2001) *FEBS Lett. 493*, 139−143.
108. Patel, P. H., and Loeb, L. A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 5095−5100.
109. Sauer, U., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wuthrich, K., and Bailey, J. E. (1999) *J. Bacteriol. 181*, 6679−6688.
110. Bornscheuer, U. T., Altenbuchner, J., and Meyer, H. H. (1999) *Bioorg. Med. Chem. 7*, 2169−2173.
111. Kim, Y. S., Jung, H. C., and Pan, J. G. (2000) *Appl. Environ. Microbiol. 66*, 788−793.
112. Benhar, I. (2001) *Biotechnol. Adv. 19*, 1−33.
113. Vanwetswinkel, S., Avalle, B., and Fastrez, J. (2000) *J. Mol. Biol. 295*, 527−540.
114. Forrer, P., Jung, S., and Plückthun, A. (1999) *Curr. Opin. Struct. Biol. 9*, 514−520.
115. Fuh, G., and Sidhu, S. S. (2000) *FEBS Lett. 480*, 231−234.
116. Sidhu, S. S., Weiss, G. A., and Wells, J. A. (2000) *J. Mol. Biol. 296*, 487−495.
117. Pedersen, H., Holder, S., Sutherlin, D. P., Schwitter, U., King, D. S., and Schultz, P. G. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 10523−10528.
118. Demartis, S., Huber, A., Viti, F., Lozzi, L., Giovannoni, L., Neri, P., Winter, G., and Neri, D. (1999) *J. Mol. Biol. 286*, 617−633.
119. Jestin, J. L., Kristensen, P., and Winter, G. (1999) *Angew. Chem., Int. Ed. 38*, 1124−1127.
120. Ponsard, I., Galleni, M., Soumillion, P., and Fastrez, J. (2001) *ChemBioChem 2*, 253−259.
121. Atwell, S., and Wells, J. A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 9497−9502.
122. Daugherty, P. S., Iverson, B. L., and Georgiou, G. (2000) *J. Immunol. Methods 243*, 211−227.
123. VanAntwerp, J. J., and Wittrup, K. D. (2000) *Biotechnol. Prog. 16*, 31−37.
124. Winson, M. K., and Davey, H. M. (2000) *Methods 21*, 231−240.
125. Olsen, M. J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., and Iverson, B. L. (2000) *Nat. Biotechnol. 18*, 1071−1074.
126. Chen, G., Hayhurst, A., Thomas, J. G., Harvey, B. R., Iverson, B. L., and Georgiou, G. (2001) *Nat. Biotechnol. 19*, 537−542.
127. Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 4552−4557.
128. Firestine, S. M., Salinas, F., Nixon, A. E., Baker, S. J., and Benkovic, S. J. (2000) *Nat. Biotechnol. 18*, 544−547.
129. Wigley, W. C., Stidham, R. D., Smith, N. M., Hunt, J. F., and Thomas, P. J. (2001) *Nat. Biotechnol. 19*, 131−136.
130. Waldo, G., Standish, B., Berendzen, J., and Terwilliger, T. (1999) *Nat. Biotechnol. 17*, 691−695.

BI011310C