The Diversity Challenge in Directed Protein Evolution

Tuck Seng Wong, Daria Zhurina and Ulrich Schwaneberg*

International University Bremen (IUB), Campus Ring 1, D-28759 Bremen, Germany

Abstract: Over the past decade, we have witnessed a bloom in the field of evolutive protein engineering which is fueled by advances in molecular biology techniques and high-throughput screening technology. Directed protein evolution is a powerful algorithm using iterative cycles of random mutagenesis and screening for tailoring protein properties to our needs in industrial applications and for elucidating proteins' structure function relationships.

This review summarizes, categorizes and discusses advantages and disadvantages of random mutagenesis methods used for generating genetic diversity. These random mutagenesis methods have been classified into four main categories depending on the method employed for nucleotide substitutions: enzyme based methods (Category I), synthetic chemistry based methods (Category II), whole cell methods (Category III) and combined methods (Category I-II, I-III and II-III). The basic principle of each method is discussed and varied mutagenic conditions are summarized in Tables and compared (benchmarked) to each other in terms of: mutational bias, controllable mutation frequency, ability to generate consecutive nucleotide substitutions and subset diversity, dependency on gene length, technical simplicity/robustness and cost-effectiveness. The latter comparison shows how highly-biased and limited current diversity creating methods are. Based on these limitations, strategies for generating diverse mutant libraries are proposed and discussed (RaMuS-Flowchart; KISS principle).

We hope that this review provides, especially for researchers just entering the field of directed evolution, a guide for developing successful directed evolution strategies by selecting complementary methods for generating diverse mutant libraries.

Keywords: Review, protein engineering, random mutagenesis, directed evolution, epPCR.

INTRODUCTION

Directed protein evolution represents the most versatile method for tailoring protein properties to the needs in industrial applications and to elucidate structure-function relationships. Inspired by an experiment from Spiegelman and coworkers [1] who introduced the concept of "evolution in a test tube", we have seen a blossom of the evolutive protein engineering field [2-4]. Success stories include the field of biotransformations for chemical syntheses [5-7], biosensors [8], bioremediation [9], vaccines [10], therapeutic proteins [11], protein structure-function relationships [12] and many others. Progress has been fueled by rapid advancements in molecular biology techniques as well as high-throughput screening technology.

Today, the field of protein engineering by directed evolution is confronted with two fundamental challenges: *a) the diversity challenge*: how to create unbiased diversity and how to scan through a statistically meaningful fraction of the generated sequence space, and *b) the recombination challenge*: how to shuffle DNA with sequence identity below 60 %. State of the art DNA recombination methods are capable of combining two parental genes with 1-3 crossovers per hybrid gene [13-16]. Recombination methods [11,17] and advances in screening technologies [18-20] have been extensively reviewed and will not be discussed in this review.

*Address correspondence to this author at the International University Bremen, Campus Ring 1, D-28759 Bremen, Germany; Tel: +49 (421) 200 3543; Fax: +49 (421) 200 3229; E-mail: u.schwaneberg@iu-bremen.de

The purpose of this review is to summarize, categorize and discuss current methods for generating genetic diversity. We hope to provide, especially for researchers just entering the field of directed evolution, a guide for choosing mutagenic methods wisely and for designing own directed evolution strategies. In addition, we aim to create awareness that our biased diversity methods hamper the progress in understanding structure-function relationships; especially when screened library sizes are small and affinity based methods cannot be used for identifying improved protein mutants.

PERFORMANCE CRITERIA FOR A RANDOM MUTAGENESIS METHOD

The Ideal Mutagenesis Method

For objectively comparing advantages and limitations of mutagenesis methods, it is important to define criteria for benchmarking. The ultimate goal of any mutagenesis method is to replace any amino acid of a polypeptide chain by the other 19 amino acids in a statistical manner without limiting protein expression in the host organism. A superb method to randomize the gene of interest should therefore receive full marks in seven categories: 1) Unbiased mutational spectrum, 2) Controllable mutation frequency, 3) Consecutive nucleotide substitutions or codon-based substitutions, 4) Enable subset mutagenesis (e.g. introducing mainly positively or negatively charged amino acids), 5) Independent of gene length, 6) Technically simple and reproducible, and 7) Economical.

1386-2073/06 $50.00+.00

Non-Bias Mutational Spectrum: Transition and Transversion Indicators

A widely accepted way of assessing bias in mutational spectra is to analyze the ratio of transitions (Ts; nt substitutions: purine to purine or pyrimidine to pyrimidine) to transversions (Tv; nt substitutions: purine to pyrimidine or pyrimidine to purine). Since there are four possible transitions and eight possible transversions, a non-bias mutational spectrum will have a Ts/Tv ratio of 0.5 and an AT→GC/GC→AT ratio of 1 (Table **1**).

Table 1. Definitions and Criteria for the Quality of a Mutant Library (Bias Indicator, Transition and Trans-Version)

		Value
Bias indicator		
Ts/Tv		0.5
AT→GC/GC→AT		1
A→N, T→N		50 %
G→N, C→N		50 %
Transitions		
A→G, T→C	(AT→GC)	16.67 %
G→A, C→T	(GC→AT)	16.67 %
Transversions		
A→T, T→A	(AT→TA)	16.67 %
A→C, T→G	(AT→CG)	16.67 %
G→C, C→G	(GC→CG)	16.67 %
G→T, C→A	(GC→TA)	16.67 %

Some changes produced in random mutagenesis methods are equivalent. For example, a mutation in the G:C base pair during error-prone PCR (epPCR) can occur at either the G in the upper strand (sense strand) or the C in the lower strand (anti-sense strand). A C→A mutation and a G→T mutation would lead after subsequent PCR cycles, to an identical T:A pair. It is therefore reasonable to combine mutational combinations leading to equivalent mutations (Table **1**). Each of the possible six combinations has a statistical occurrence of 16.67 %. Experimental results confirm these conclusions for large mutant libraries within statistical errors [21,22]. However, mutating the upper or lower nt of a pair causes a different amino acid substitution profile due to codon's usage nature. For example, C→A mutations in the first nucleotide of a codon in sense strand lead to amino acid changes from (Leu, Pro, His, Gln, Arg) to (Arg, Ser, Lys, Asn, Thr, Met and Ile) whereas G→T mutations in sense strand prefer changes from (Gly, Glu, Asp, Ala, Val) to (Phe, Leu, Ser, Tyr, Cys, Trp, Stop codon).

Mutation Frequency

For a successful directed evolution experiment, it is important to fine tune mutation frequencies to the tolerance of protein structure and the protein property aimed to be improved. Being able to adjust the mutation frequency in directed evolution experiments is therefore important. For directly understanding protein structure-function relationships, a mutation frequency inducing one amino acid change per gene seems most desirable. Most directed evolution studies aiming to improve activity or stability of enzymes employ however higher mutation frequencies inducing 1-4 amino acid changes (2-7 nucleotide substitutions) per gene [23]. Those conditions are often chosen and limited by practical considerations. For instance many university labs can handle or afford only to screen in liquid culture small libraries of 2000-20000 clones. In order to have a sufficient number of active mutated clones mutation frequencies are empirically adjusted to 40-60 % of active clones.

As a result, there is still insufficient experimental data to draw conclusions about which strategies are best in terms of mutation frequency, library size and property of interest. Encouragingly, proteins with improved performance have also been isolated from highly mutagenized libraries exhibiting up to 20 mutations per gene [24]. Moreover, affinity based ultra-high throughput screening/selection systems are well elaborated and have successful been used at high mutation frequencies [25]. Control over the mutation frequency in the whole gene of interest is thus an important feature of a versatile random mutagenesis method.

Subsequent Mutations and Subset Mutations

Having high mutation frequencies of 2-7 nucleotides per 1000 base pairs is not sufficient to explore sequence space efficiently due to the redundancy of genetic code. Three subsequent nucleotides, one codon, encode one amino acid; for saturating one amino acid position it is often required to change two or all three subsequent nucleotides of a codon. High mutation frequencies of 2-7 nucleotides per 1000 base pairs cause in most cases only single nucleotide substitution in the mutated codons. With a chance of 7 in 1000 per nucleotide position, it is an unlikely event to substitute two or three subsequent codon positions, especially in small libraries (2000-20000 clones). Being able to perform consecutive nucleotide substitutions or even codon-based substitutions is an important feature for generating diverse libraries. Single nucleotide substitutions particularly T→C or A→G transitions in the third nucleotide position of a codon do not usually change the encoded amino acid [26]. The genetic code is structured such that a single nucleotide substitution in a codon causes predominantly an amino acid substitution between amino acids with similar properties. This "structural fidelity" is believed to be an adaptation of genetic code during evolution [27].

Being able to manipulate the mutational spectrum by introducing desired mutational subsets would enable novel directed evolution strategies and generate functionally enriched mutant libraries. Examples of desired subset comprise: a) introducing mainly charged amino acid substitutions for increasing solubility of proteins or crystallization, b) bypassing stop codons, proline as well as glycine residues, or c) avoiding codon substitutions that limit protein expression.

Practical Considerations

It is trivial but worth a mention that an important benchmark of a mutagenesis method is its reproducibility,

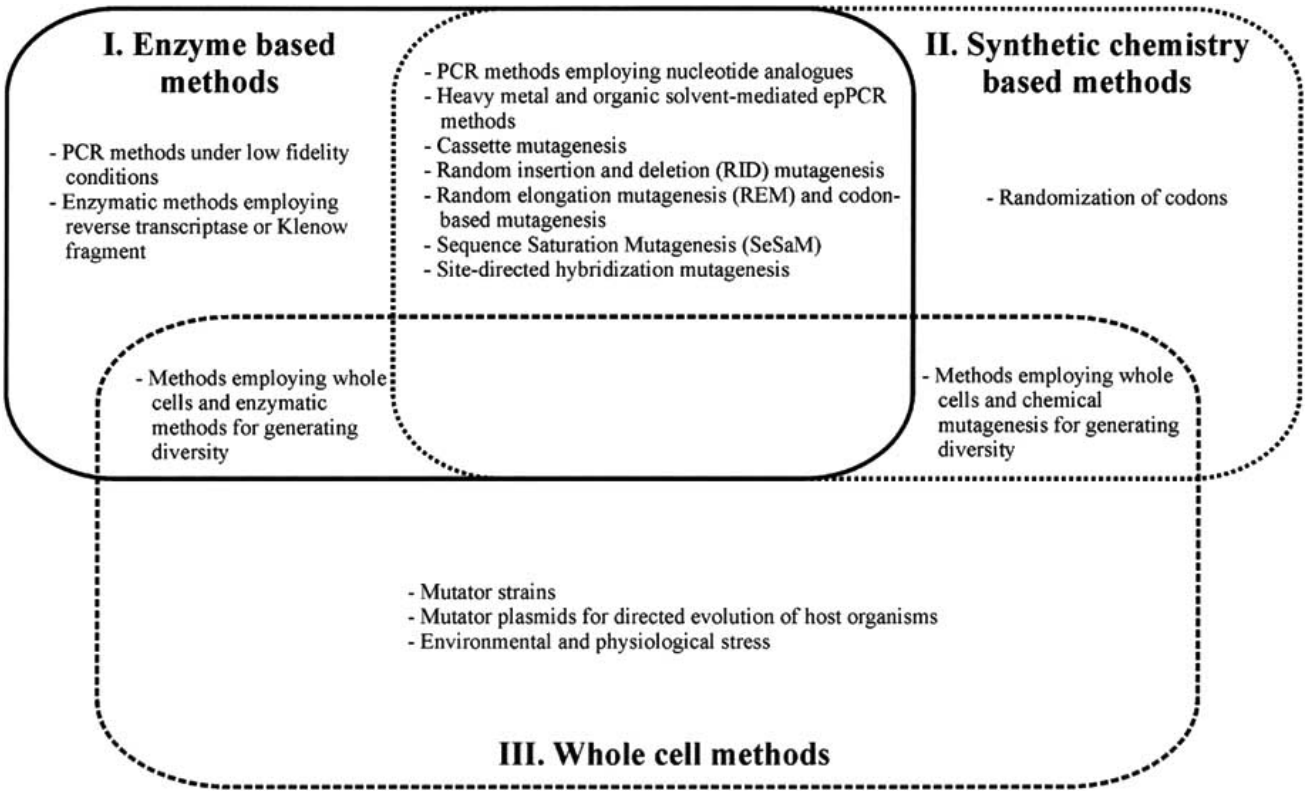

Fig. (1). Classification of random mutagenesis methods.

technical simplicity and cost-effectiveness measured by time-consumption and consumable expenses.

CLASSIFICATION & SUMMARY OF RANDOM MUTAGENESIS METHODS

Classification

Methods for generating randomized genes can be divided into three Categories (I-III) depending on the method employed for nucleotide substitutions (Fig. **1**); Category I comprises the enzyme based methods, Category II the synthetic chemistry methods, and Category III whole cells methods.

Enzyme based methods (Category I) encompass *in vitro* techniques for mutagenesis that use DNA polymerases or DNA modifying enzymes such as ligases, restriction enzymes and nucleases. For example, error-prone polymerase chain reaction (epPCR) protocols fall into this category. Synthetic chemistry based methods (Category II) generate mutations *via* oligonucleotide-based synthesis, methods employing nucleotide analogues or DNA-modifying chemicals. *In vivo* random mutagenesis methods (Category III) use mutator strains or "mutator proteins" for generating mutations in host organisms.

In Fig. **1** diversity methods are arranged according to Categories I-III. Only a few whole cell and purely chemical methods have been developed; most diversity methods use enzymatic or combined chemo-enzymatic approaches for mutagenesis. Sequence Saturation Mutagenesis (SeSaM) [28] is, for example, a chemo-enzymatic method that saturates every single nucleotide position in a target gene through a chemical nucleotide analog (degenerate or universal bases) that will, through a polymerase chain reaction (enzymatic step), be replaced by a standard nucleotide.

I. ENZYME BASED METHODS

PCR Methods Under Low Fidelity Conditions

A technically simple and elegant method to introduce mutations in DNA is employing thermostable DNA polymerases exhibiting low fidelity during amplification of DNA in PCR. The workhorse in directed protein evolution experiments is *Taq* DNA polymerase from *Thermus aquaticus* that lacks a 3'-5' proofreading exonuclease activity [21,29,30]. *Taq* polymerase is a robust polymerase with an intrinsic high error rate (8.0×10^{-6} mutation/bp/duplication, [31]). An inaccuracy of 1 substitution per 125000 bp per PCR cycle is however still insufficient for randomizing genes that seldom exceed 5000 bp in length. Most prominent methods to increase error rates are a) Addition of $MnCl_2$ [29], and b) Unbalanced nucleotide concentrations [29]. Further possibilities with minor mutagenic effects include c) Increased concentration of *Taq* DNA polymerase [29], d) Increased extension time [29] and e) Increased concentration of $MgCl_2$ [29]. A further important parameter in adjusting mutation rates is the number of PCR cycles. The influence of unbalanced concentrations of dNTPs and Mn^{2+} ions on the *Taq* DNA polymerase fidelity are summarized in Table **2**. The mutational spectra of *Taq* DNA polymerase remains under error-prone conditions highly biased to transitions.

Instead of varying parameters (a-e), it is possible by protein engineering to increase the inaccuracy of *Taq* DNA

Table 2. Mutational Spectra of Random Mutagenesis Methods: The Mutational Spectra of a) "Bringer" Strategy, b) Enzymatic Method Using Klenow Fragment, c) *E. coli* XL1-Red Mutator Strain, d) Mutator Plasmid, e) Heavy Metal- [e.g. Co^{2+}] and Organic Solvent- [e.g. Propan-1-ol] Mediated epPCR have not been Reported and are Therefore not Included in the Table 2

Category	I	I	I	I	I	I	I	I	I	I
Mutational Spectra	***Taq*-Pol (Unbalanced dNTPs)[a*]**	***Taq*-Pol (Mn^{2+}/ Balanced dNTPs)[b] [21]**	***Taq*-Pol (Mn^{2+}/Un balanced dNTPs)[c*]**	***Taq*-Pol (Mn^{2+}/Un balanced dNTPs)[d*]**	***Taq*-Pol (Mn^{2+}/ Unbalanced dNTPs)[e] [136]**	***Taq*-Pol (Mn^{2+}/ Unbalanced dNTPs)[f] [22]**	***Taq*-Pol I614K [32]**	**Mutazyme I [33]**	**Muta-zyme II [35]**	***Pfu*-Pol (exo^-) D473G [34]**
Transitions										
A→G, T→C	33.3	63.2	42.7	74.0	27.6	60.9	41.5	10.3	17.5	19.2
G→A, C→T	8.3	9.5	11.5	4.9	13.6	18.1	23.6	43.7	25.5	22.0
Transversions										
A→T, T→A	16.7	16.1	26.0	13.8	40.9	11.4	24.5	11.1	28.5	28.0
A→C, T→G	27.8	4.0	8.3	4.1	7.3	3.3	2.8	4.2	4.7	7.4
G→C, C→G	0.0	1.4	0.0	1.6	1.4	4.3	3.8	8.8	4.1	7.3
G→T, C→A	0.0	3.3	6.3	0.8	4.5	1.8	2.8	20.0	14.1	10.3
Insertion	2.8	Not reported	2.1	0.0	0.3	Not reported	Not reported	0.8	0.7	2.9
Deletion	11.1	1.5	3.1	0.8	4.2	Not reported	0.9	1.1	4.8	2.9

[a]*Taq*-Pol and [dATP] = [dCTP] = [dTTP] = 0.2 mM, [dGTP] = 0.24 mM.
[b]*Taq*-Pol in the presence of $[Mn^{2+}]$ = 0.25 mM and four standard nucleotides (each 0.02 mM).
[c]*Taq*-Pol in the presence of $[Mn^{2+}]$ = 0.64 mM and [dATP] = [dCTP] = [dTTP] = 0.2 mM, [dGTP] = 0.24 mM.
[d]*Taq*-Pol in the presence of $[Mn^{2+}]$ = 0.64 mM and [dATP] = [dCTP] = [dTTP] = 0.2 mM, [dGTP] = 0.40 mM.
[e]*Taq*-Pol in the presence of Mn^{2+} and [dATP] = [dGTP] = 0.2 mM, [dCTP] = [dTTP] = 1 mM.
[f]*Taq*-Pol in the presence of $[Mn^{2+}]$ = 0.5 mM and [dCTP] = [dATP] = 0.03 mM, [dGTP] = [dTTP] = 1 mM.
[*]Values adapted from the user manual of the Diversify PCR Random Mutagenesis Kit (Clontech).

(Table 2) contd…..

Category	I	III	I & II	I & II	I & III	II & III	II & III	II & III	II & III
Mutational Spectra	**Enzymatic Method Employing Reverse Transcriptase [38]**	***E. coli* Expressing mutA Allele of *glyV* gene [62]**	***Taq*-Pol (Nucleotide Analogues dPTP and 8-oxodGTP)[g] [137]**	**Error-prone Rolling Circle Amplification [79]**	**Pol I Method [107]**	**Chemical Mutagen (Nitrous Acid) [109]**	**Chemical Mutagen (Formic Acid) [109]**	**Chemical Mutagen (Hydrazine) [109]**	**Chemical Mutagen (Ethyl Methane Sulfonate) [111]**
Transitions									
A→G, T→C	30.83	5	78.3	7.0	34.7	29.1	21.7	60.4	34.2
G→A, C→T	36.25	23	13.2	60.6	44.9	65.5	8.7	22.9	42.1
Transversions									
A→T, T→A	19.17	21	0.2	7.0	14.3	1.8	13.0	4.2	0
A→C, T→G	8.34	13	8.4	1.4	2.0	0	6.5	8.3	0
G→C, C→G	0.42	5	0.7	5.6	0.0	3.6	13.0	2.1	23.7
G→T, C→A	5	20	0.0	9.9	4.1	0	37.0	2.1	0
Insertion	Not reported	5	~0	8.5	Not reported	Not reported	Not reported	Not reported	Not reported
Deletion	Not reported	8	~0	0.0	Not reported	Not reported	Not reported	Not reported	Not reported

[g]*Taq*-Pol in the presence of Mg^{2+}, four standard nucleotides (0.5 mM) and two non-natural mutagenic bases 8-oxodGTP and dPTP.

polymerase (*Taq*-Pol I614K) [32] or to use other polymerases with high-error rates such as the Mutazyme polymerase (introduced by Stratagene in the GeneMorph Random Mutagenesis Kit, [33]) or an engineered *Pfu* polymerase (*Pfu*-Pol (exo^-) D473G) reported by Connolly and coworkers [34]. Table **2** summarizes the mutational spectra of these polymerases under different mutagenic conditions. All polymerases are highly biased and especially the transversions ((A→C, T→G); (G→C, C→G)) are highly underrepresented substitutions. A blend of both Mutazyme

and *Taq* mutant named Mutazyme II is used in the GeneMorph II Kit to improve the transition to transversion ratio, the transversions ((A→C, T→G); (G→C, C→G)) are however still strongly underrepresented in a highly biased mutational spectra [35] (Table **2**). *Pfu* DNA polymerase mutant (*Pfu*-Pol (exo$^-$) D473G) [34] is in its mutational spectra more balanced than *Taq* polymerase or Mutazyme. The occurrence of the rarest transversion pair (G→C, C→G) is 7.3 % compared to 0.0-2.8 % of various *Taq* polymerases and 4.2 % (A→C, T→G) of Mutazyme (Table **2**).

The "Bringer" strategy can be regarded as a variation of Stratagene's QuikChange method [36] in which the number of amplification cycles is increased from 12-18 cycles to 29 cycles [37]. Under these conditions, the high-fidelity *PfuTurbo* DNA polymerase (error rate 1.3 x 10^{-6} mutation/bp/duplication, [31]) accumulates mutations during amplification of ring-closed plasmids. The whole-plasmids amplification is advantageous and disadvantageous at the same time. On the one hand, the method is fast and no restriction digest and subsequent ligation step limit the library generation; on the other hand the whole plasmid is subjected to mutagenesis and unfavorable mutations might occur in other parts of the plasmid such as promoter or selection marker.

Enzymatic Methods Employing Reverse Transcriptase or Klenow Fragment

In the 1980s, Knowles and coworkers [38] and Shimura and coworkers [39] developed independently two conceptually similar random mutagenesis methods. Mutagenesis is initiated by generating a pool of target DNA fragments with random length distribution by limited primer elongation [38] or time-controlled random exonuclease III digestion [39]. At the 3'-termini, a "mutagenic" nucleotide is misincorporated deliberately by reverse transcriptase [38] or Klenow fragment [39]. After tailing the 3'-termini, the fragments are elongated to full-length genes by employing a single-stranded template and cloning into an appropriate vector. These conceptually appealing techniques aim to generate single base substitutions at every single nucleotide position in a target gene; a feature which is hardly attainable in epPCR or GeneMorph Random Mutagenesis due to the polymerase bias. The general use of reverse transcriptase or Klenow fragment is hampered by a non-even length distribution of the DNA fragments after exonuclease III digestion [39], a low percentage of mutated to non-mutated genes (40-60 % [38], 37.5 % [39]), and a significant mutational bias (T→G, T→A, C→G and G→T substitutions are underrepresented with occurrence of lower than 5 %, [38]) caused by the template assisted elongation that favors transitions due to base pairing properties.

II. SYNTHETIC CHEMISTRY BASED METHODS

Chemical synthesis of DNA enables us to synthesize complete genes. Gene synthesis offers potentially freedom over genetic diversity including control over mutation frequency, mutational spectrum, and compensation for codon redundancy. The ability to synthesize routinely, in fully automated manner, single-stranded DNA fragments has enabled numerous applications in molecular biology and biotechnology and has been a main driver for progress in research. Companies such as BioSpring (http://www.biospring.de/) or Geneart (http://www.geneart. com/) offer to produce synthetic genes of kb length based on customers' demand within weeks. Main applications of synthetic gene synthesis have been in enhancing protein expression levels [40,41] in heterologous hosts by optimizing codon usage [42], increasing mRNA stability and translation efficiency [43-45] (e.g. introducing mRNA stem loop, [46]), and introducing restrictions sites to facilitate further gene manipulations for example by cassette mutagenesis [45].

The use of synthetic chemistry for generating random mutant libraries over a whole gene has to our knowledge not been reported yet. We assume that the use of synthetic chemistry methods is limited by: a) expenses in synthesis, b) yield and time required for synthesizing a whole gene, and c) technological challenges including how to generate a random and average of 1 to 5 amino acid changes per gene.

Randomization of Codons

Randomization of a codon to encode all 20 amino acids using a base by base synthesis for the three subsequent nucleotides (NNN) results in a statistical equal amount of A, T, G and C at each positions and 64 possible codons. Due to the codon usage of host organisms, some amino acids will be overrepresented such as Ser and Arg (6 codons) whereas amino acids such as Met and Trp will be strongly underrepresented (1 codon). Additionally, three stop codons increase the number of inactive clones in a mutant library and hamper the total number of positions that should simultaneously be saturated with limited screening throughput. The latter NNN-strategy was improved in terms of bias and number of stop codons by making use of the redundancy in the third position of many codons. Using all four nucleosides in the first two codon positions, but only G and C or A and T in the third position (NNG/C or NNA/T) results in 32 triplets encoding all 20 amino acids with only 11 redundant codons and 1 stop codon [47].

In 1991, Knapps and coworkers described a "one-bead, one-peptide" approach [48] in which 20 codons were synthesized in separate columns. After synthesis the beads were mixed, and split again into 20 columns to synthesize the next codon. In this iterative process, codon redundancy and stop codons were eliminated. Monaci and coworkers [49] achieved a randomized codon-based synthesis using 4 parallel columns and seven dimer-phosphoramitides. More recent methods such as the Fmoc-protected trinucleotide phosphoramidites [50] is based on the use of 20 pre-synthesized codons as monomeric units for each amino acid circumventing mutational bias, stop codons and low expression mutations. Using the Fmoc-protected trinucleotide phosphoramidites, a region encoding for nine amino acids of the bacterial enzyme TEM-1 β-lactamase was mutated and structurally interesting mutants in the omega loop were obtained [50]. Further mutagenesis methods employing trinucleotides phosphoramidites used automated DNA synthesis methodologies [51,52]. An alternative approach was introduced by Hine and coworkers who generate a synthetic gene template that is NNN-randomised at specific sites [53]. For each NNN-saturated position, a pool of 20 oligos is synthesized that hybridizes at the 3'-end "mainly" with the corresponding codon that is complementary and highly expressible. Three DNA-

contacting residues of a synthetic zinc finger protein were randomised using this "selectional hybridization approach", NNN and NNG/T randomization [53]. This approach is promising for encoding functional subset of amino acids (e.g. acidic, basic and amide-containing side-groups) but differ for certain amino acids significantly from the ideal 5 % occurrence per amino acid [53]. The number of simultaneously mutated amino acid positions is limited by a ligation step [53].

In summary, chemical synthesis for generating diverse gene libraries is, despite its huge potential, still in an infant stage of development. Chemical synthesis methods are however a valuable alternative to chemo-enzymatic saturation mutagenesis methods, especially for saturating "longer" peptides [54-56].

III. WHOLE CELLS BASED METHODS

Cells maintain during DNA replication, a high fidelity with error rates as low as one substitution/10^{10} bp/generation/cell [57]. Fidelity can be reduced by mutations in genes required for DNA replication or repair (mutator strains), in response to "mutator proteins", and environmental as well as physiological stress.

Mutator Strains

A vector harboring the gene of interest is transformed into a mutator strain that is deficient in major DNA repair pathways. During cell cultivation, mutations are accumulated in the genomic DNA of the mutator strain, the vector, and the gene of interest. Mutagenized vectors are subsequently isolated, re-transformed into a standard expression host with low mutation frequency and screened for beneficial variants.

E. coli XL1-Red, a popular commercialized mutator strain from Stratagene, has been mutated in three DNA repair pathways (*mutS* (error-prone mismatch repair), *mutD* (deficient in 3'-5' exonuclease activity of DNA polymerase III) and *mutT* (prevents removal of 8-oxodGTP in mismatches involving A:G, resulting in mainly AT→CG transversions [58])). As a result, the mutation rate was increased by ~5000 fold compared to the unaltered *E. coli* strain (one mutation/10^{10} bp/generation/cell) [59]. After 24 generations of growth, a spontaneous mutation frequency of 5 x 10^{-6} mutation/bp was reported for *E. coli* XL1-Red [59]. Another report indicates after 30 generations of growth an average mutation frequency of 5 x 10^{-4} [60]. Mutations in *mutS, mutD* and *mutT* result in a transition (85 %) biased mutational spectrum with few transversions (10 %) and frameshifts (5 %) [61].

A further target to boost mutation rates is the *mutA* allele that differs from the wild-type *glyV* gene by a base substitution in the anticodon such that the resulting tRNA misreads certain amino acid codons. The base substitutions in *E. coli* cells expressing the *mutA* allele of *glyV* under stress-induced conditions are elevated by ~17-fold including all mutations and 23-fold for base substitutions [62]. An overall relatively low error rate of 4.8 x 10^{-7} was achieved; interestingly transversions increased significantly higher than transitions (35-fold compared to 13-fold) [62] and become predominant with a ratio of 2.1 to 1.

E. coli mutator strains have been the preferred hosts in directed evolution experiments due to rapid growth and established molecular biology protocols. Mutator strains have been established for a variety of organisms. Yeast mutator strains for example with mutant alleles of POL1, POL2, and POL3 encoding replicative DNA polymerases (Polα, Polδ and Polε) are attractive candidates [63,64].

Mutator Plasmids for Directed Evolution of Host Organisms

Mutator plasmids allow us to temporarily convert a bacterial strain into a mutator strain by inducing temporarily the recombinant expression of a "mutator protein" [65,66]. The mutator plasmid carries for example a temperature-sensitive origin of replication that allows, after diversity generation, to remove the mutator plasmid during cultivation yielding a genetically stable and "evolved" strain that can directly be used for protein expression and screening.

The system from Schellenberger and coworkers [66] employs such a temperature-sensitive mutator plasmid based on the *mutD* gene (two amino acid substitutions) that encodes the proofreading function of DNA polymerase III. Induction exhibits in three investigated bacterial strains (*E. coli* MM294, *E. coli* W1485 and *E. blattae* EB33429) a strong mutator phenotype with a significantly increased mutation frequency of 0.46-9.3 x 10^{-5} compared to 0.8-4.2 x 10^{-8} in the corresponding wild-type strains [62,66]. After growing at elevated temperature, the bacteria lose mutator plasmids and the mutation frequencies decrease to a value of 1.5-2.5 x 10^{-8} which is close to mutation frequencies prior to evolution. Using this approach an *E. coli* MM294 strain was evolved that grows in DMF/water mixture containing up to 80 g/l DMF [66].

Environmental As Well As Physiological Stress

When a population of microorganisms is subjected to nonlethal selections (environmental or physiological stress for instance fluoroquinolone resistance of *E. coli* [67], heavy metal resistance of *Candida albicans* [68], organic solvent resistance of *E. coli* [69] and many others), more tolerant variants with better growth are isolated. This adaptive process allows evolution of microorganisms that tolerate non-natural environments [70] and are useful for industrial biotransformation.

Advantages and Limitations of *In Vivo* System for Generating Diversity

The beauty of mutator strains is their simplicity in handling that requires little experience in molecular biology techniques and makes these systems attractive for researchers entering the field of directed protein evolution. Handicaps limiting the versatile use of *in vivo* systems are: a) a highly biased mutational spectrum [62], b) a non-specific mutagenesis accumulating mutations in the host organism as well as the plasmid [59], and c) a reduced growth rate and transformation efficiency of the host organism [59] due to inactivation of multiple major DNA repair pathways. In addition, it is difficult with mutator strains to obtain multiple or subsequent mutations in one gene and to control the mutation frequency that is dependent on the target gene length, plasmid copy number, number of generations and origin of replication [59]. Mutator plasmids have, in contrast

to mutator strains, mainly been used to improve host organisms and not properties of enzymes. In order to mutagenize mainly the gene of interest, *in vivo* systems have been developed that employ for example Pol I polymerase. These systems are discussed in the section: combined enzymatic and whole cell methods.

MIXED METHODS

CATEGORY I & II

Mutagenesis methods combining enzymatic and synthetic chemistry steps represent in terms of number the major population. Chemo-enzymatic methods can be categorized into seven subcategories: 1) PCR with nucleotide analogues, 2) Chemical mediated epPCR, 3) Cassette mutagenesis, 4) Random elongation, insertion and deletion mutagenesis, 5) Sequence Saturation Mutagenesis, 6) Site-directed hybridization mutagenesis and 7) Codon-based mutagenesis.

PCR Methods Employing Nucleotide Analogues

Loakes has summarized in an excellent review [71] the use of universal DNA base analogues in hybridization experiments. Universal base analogs can base-pair with all four standard nucleotides and are used for sequencing and in PCRs as probes and in form of primers. For generating unbiased diversity, universal base analogs should: a) pair "equally" with all four natural bases when opposite them in an oligonucleotide duplex, b) be a substrate for polymerases as 5'-triphosphate, c) be elongated at the 3'end after incorporation by polymerase using dNTPs as substrates, d) allow incorporation of dNTPs opposite them in a template-based polymerase synthesis, and e) be amenable to nuclease digest and cloning.

For the use of universal bases in epPCR, polymerases have to exhibit both "read" and "write" activities during PCRs [72]. "Read" is defined as a polymerase catalysed and template-directed extension of DNA fragments containing a universal base at 3'-termini. "Write" is defined as polymerase catalysed incorporation of standard nucleotides opposite to the template site harboring a universal base. Reports using universal bases in epPCRs have been rare due to polymerase performances that are limited by write and/or read activities [72]. Limited write and/or read activities result often in meager PCR product yields and low mutation frequencies. 5-Nitroindole is a prominent example of a truly universal base that cannot be amplified efficiently (write activity) by polymerases [73]. An exception to which polymerases exhibited an acceptable read and write activity is the universal base deoxyinosine (dI) [72] that has been used in epPCR [74]. By limiting one nucleotide to 7 % of the dITP concentration, a mutation frequency of 3.75×10^{-3} per bp was achieved by Kuipers and coworkers [74]. The mutational spectrum was, as expected from base pairing properties [75], biased to (AT→GC) transitions.

In Table **3**, we have summarized base-pairing properties of nucleotide analogues on which polymerases exhibit read/write activities and have been used in random mutagenesis experiments. Recent development of engineered DNA polymerases capable of using base analogues such as 5-nitroindole [73] likely open opportunities for epPCR methods employing universal bases. The lack of subsequent mutations will however hardly be solved by incorporating single nucleotide analogue.

Heavy Metal and Organic Solvent-Mediated epPCR Methods

The influence of low fidelity conditions in PCR has been summarized in Table **2** and described in section I. Enzyme based methods, section PCR methods under low fidelity conditions. Valuable supplements are the reports from Loeb and coworkers who studied intensively the effect of the heavy metal ions (Mn^{2+} and Co^{2+}) and Mg^{2+} ions on the fidelity of PCR during DNA replication [76-78]. Mutation frequencies higher than 5×10^{-4} were achieved with heavy metal ion addition [76-78] and 4.4×10^{-4} with elevated concentrations of Mg^{2+} ions [76-78]. Combining unbalanced dNTP concentrations with low mM concentrations of heavy metal ions boosted mutation frequencies up to 1×10^{-2} and provides awesome control over mutation frequencies [76-78]. In a commercialized kit from Clontech (Diversify PCR Random Mutagenesis Kit), mutation frequencies are adjusted similarly by varying Mn^{2+} and dGTP concentrations.

Recently, Hayashi and coworkers [79] combined epPCR employing 0.5-2 mM $MnCl_2$ with the Rolling Circle Amplification (RCA) technology [80] that has been commercialized as TempliPhi DNA Amplification kit [81] by Amersham Biosciences. Rolling circle amplification (RCA) uses bacteriophage ϕ29 DNA polymerase to exponentially amplify single- or double-stranded circular DNA templates [80]. This isothermal amplification produces microgram quantities of DNA within 4-6 h and does not require a thermocycler. The generated DNA can directly be transformed or used for sequencing. Hayashi and coworkers [79] achieved a mutation frequency of 3.5×10^{-3} after 24 hrs of reaction time and in the presence of 2 mM $MnCl_2$. Extent of template bias was low even without any purification step and mutation frequencies can be adjusted by varying reaction times [79]. epRCA-method does not require a restriction digest or ligation step making the epRCA-method a valuable alternative to mutator strains. The mutational spectrum is highly biased (transitions 67.6 % *vs* transversions 23.9 % [79], Table **2**) and subsequent nucleotide substitutions will be difficult to obtain.

The influence of urea, isopropanol, propan-1-ol, and butan-1-ol on PCR fidelity and yield was investigated for three DNA polymerases by Beauregard and coworkers [82]. A mutagenic effect was observed with Vent (exo^-) DNA polymerase in the presence of 7-8 % (v/v) propan-1-ol, achieving error rate up to 9.8×10^{-3} mutation/bp/PCR. Under alcohol-mediated condition, a bias towards replacements of Gs and Cs was observed (86 % G/C→N, 14 % A/T→N) [82]. This bias differs interestingly from Mn^{2+}-mediated epPCR that favors substitution of As and Ts (75 % A/T→N, 25 % G/C→N) [82].

Cassette Mutagenesis

Completely or partially randomized oligonucleotides that have been synthesized chemically (see synthetic chemistry based methods) can be introduced by ligation into the gene of interest that has been digested with restriction enzymes [83]. The randomized region has in general been selected by sequence alignments, crystallographic information and/or

Table 3. Summary of Nucleotide Analogues on which Polymerases Exhibit Read/Write Activities

Nucleotide Analogues	Symbol	Purine/Pyrimidine Analogue	Behavior
2'-deoxyinosine	I	Purine analogue	Behaves as G (Bias not reported) [74,138]
N6-methoxy-2,6-diaminopurine	K	Purine analogue	Behaves as A (87 %) and G (13 %) [139,140]
N6-methoxyaminopurine	Z	Purine analogue	Weakly mutagenic (Bias not reported) [140]
6-(2-deoxy-β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one	P	Pyrimidine analogue	Behaves as T (60 %) and C (40 %) [137,139,141,142]
N4-aminodeoxycytidine	C^{am}	Pyrimidine analogue	Behaves effectively as C, but less effectively as T (Bias not reported) [143]
5,6,7,8-tetrahydropyrimido[4,5-c]pyridazine	H	Pyrimidine analogue	Behaves effectively as C, but less effectively as T (Bias not reported) [144]
8-oxo-2'-deoxyguanosine	8-oxoG	Capable of base-pairing with A and C	Behaves as T and G equally [137,142]
5-bromo-2'-deoxyuridine	BrU	Not reported	Bias not reported [138]
2'-deoxyribosyl pyrrole 3-monocarboxamide	M	Capable of base-pairing with T and G	Behaves as A and C (Bias not reported) [145]
2'-deoxyribosyl pyrrole 3,4-dicarboxamide	D	Capable of base-pairing with T and G	Behaves as A and C (Bias not reported) [145]
2-hydroxy-2'-deoxyadenosine	2-oxoA	Capable of base-pairing with T and G	Behaves as A and C (Bias not reported) [146,147]

previous mutagenesis studies [84]. Cassette mutagenesis allows scanning of all possible amino acid substitutions at any position within the randomized region in "relatively" small libraries. It permits rapid screening of catalytically and structurally important residues for beneficial variants [85]. Randomization of 6 to 8 amino acids have been performed successfully [84]. Other success stories include active site engineering of the phosphotriesterase for enhanced degradation of chemical warfare agents [86], reshaping a ligand pocket of bilin-binding protein for fluorescein binding [87] and replacing the omega loop of a class A PSE-4 β-lactamase with loops from TEM-1, SHV-1 and *Streptomyces albus* enzymes [88]. Cassette mutagenesis is limited by the distance between amino acids selected for randomization in the protein of interest and the synthetic oligos have to be designed carefully to minimize secondary structure formation. Furthermore, cassette mutagenesis requires a hypothesis or an educated "guess" for property-determining residues and it is often necessary to introduce restriction site in the gene of interest. Low time consumption and high diversity of mutant libraries make cassette mutagenesis a valuable method for gene randomization, especially if structure-function relationships have been elucidated.

Random Insertion and Deletion (RID) Mutagenesis

In 2000, a general method named random insertion and deletion (RID) mutagenesis was introduced by Sisido and coworkers [89-91] for inserting and/or deleting sequences of defined length. RID enables deletion of an arbitrary number of consecutive bases (up to 16 bases) at random positions and, at the same time, insertion of a specific sequence or random sequences of an arbitrary number to the same position. RID can be regarded as advanced version of cassette mutagenesis over the whole gene of interest. The single-stranded gene of interest is first fragmented by Ce(IV)-EDTA treatment [92], followed by a low efficiency single-stranded ligation introducing two anchor sequences with restriction sites that will subsequently be used for insertion and deletion of an arbitrary number of consecutive bases. RID mutagenesis has successfully been used to generate novel GFP variants, a yellow fluorescent protein and an enhanced green fluorescent protein, as well as several DsRed variants with altered fluorescence properties [89,90]. RID is one of the most advanced gene randomization methods; its universal use is hampered by its complexity and high demand on handling skills: a) RID comprises five ligation steps, which can be problematic and of low efficiency especially in the case of single-stranded DNA, and b) good handling practice is required to have, after multiple purifications, sufficient DNA left in the 9^{th} RID step for library generation and to control homogeneity and preferably single random cleavage of single-stranded gene of interest by Ce(IV)-EDTA [89]. In summary, RID is a powerful and technically demanding method.

Random Elongation Mutagenesis (REM) and Codon-Based Mutagenesis

Urabe and coworkers reported a random mutagenesis method named random elongation mutagenesis (REM) for adding peptide tails with random sequences to the C-termini of enzyme molecules [93,94]. As in cassette mutagenesis, chemically synthesized random sequences are ligated to

restrictive digested plasmids that encode the C-termini of the truncated enzymes. In their studies, Urabe and coworkers subjected a triple mutant of catalase I (I108T/D130N/I222T) to REM yielding 44 elongation mutants that regain or exceed the thermostability of wild-type catalase [93]. Though REM represents a novel concept, it is technically simple and evolutionary interesting (has nature evolved proteins by random elongation?). From a structural point of view, it is limited to adding amino acids at either the N- or C-terminus of a protein. How important can these amino acids and the relayed interactions from these amino acids be for the stabilization of a protein or its catalytic activity [95]? Structural limitations might likely impede the general use of REM, as indicated by a lack of further REM success stories.

In a method termed "Codon Shuffling" [96], T4 DNA ligase is used to ligate a pool of 6 bp-DNA duplexes of random sequences forming a library of DNA fragments with varied length. The DNA fragments were subsequently ligated to a truncated DNA sequence of a β-lactamase. The truncated β-lactamase was inactive and regained activity after "Codon Shuffling" [96]. The mutations found differed in length and the corresponding gene sequences were highly diverse [96]. "Codon Shuffling" uses random libraries of high quality and is limited by ligation efficiencies and to selection based directed evolution experiments. Codon-based random deletion (COBARDE) is a complementary method to "Codon Shuffling" and REM and permits deleting complete codons in a random and combinatorial mode [97].

Sequence Saturation Mutagenesis (SeSaM)

Sequence Saturation Mutagenesis (SeSaM) [28] was introduced in 2004 as a conceptually novel and technically simple method that randomizes a target sequence at every single nucleotide position. A SeSaM experiment can be accomplished within 2-3 days and comprises four main steps: 1) Generating a pool of DNA fragments with random length; 2) "Tailing" the DNA fragments with universal bases using terminal transferase at 3'-termini; 3) Elongating DNA fragments in a PCR to the full-length genes using a single-stranded template; and 4) Replacing the universal bases by standard nucleotides. Random mutations are created at universal sites due to the ambivalent base-pairing property of universal bases.

In order to prove that the generated mutations can be attributed to the SeSaM method, adenosine positions were targeted with a phosphorothioate nucleotide dATPαS and deoxyinosine was selected as universal base due to write and read activity of *Taq* DNA polymerase [74,98]. The pool of random fragments (SeSaM step 1) can be created using each of the four dNTPαS separately and mixed in equal amount after quantification in order to target every single nucleotide position of a gene of interest (unpublished results). SeSaM circumvents thereby the mutational bias of DNA polymerases. Mutational spectrum of SeSaM ("write" mode) can be adjusted *via* the use of different universal bases or degenerate bases with distinctive base-pairing properties (Table **3**). Read and write activities of *Taq* polymerase for universal bases can be improved by directed evolution as shown for 5-nitroindole [73]. Moreover, the mutation frequency in SeSaM experiments can be tuned by varying the concentration of sodium chloride and/or sodium hydroxide in the DNA melting solution (SeSaM step 1) [99].

SeSaM is capable of fully randomizing each nucleotide of a gene of interest. Improvements in read and write activities of polymerases for universal bases offer opportunities to further develop the method and thereby eliminating mutational bias and introducing consecutive nucleotide substitutions (SeSaM step 2).

Site-Directed Hybridization Mutagenesis

For site-directed mutagenesis various rapid and technically simple methods have been developed. Site-directed mutagenesis has become an invaluable tool for studying protein structure-function relationships. These techniques allow, due to the synthetic oligos employed, a site-directed mutagenesis of a selected position or region.

Starting in the 80ies, M13-based bacteriophage vectors have been used in site-directed hybridization mutagenesis studies. One of the most comprehensive procedure was outlined by Kunkel [100], in which the gene of interested is cloned into a M13 vector and transformed into an *E. coli dut*$^-$ *ung*$^-$ strain. This strain is deficient in the dUTPase (product of *ung* gene), resulting in accumulation of intracellular dUTP and subsequent incorporation into the newly synthesized single-stranded DNA (template DNA). A synthetic oligonucleotide containing the desired mutation(s) hybridizes to uridinylated-template DNA, will be elongated by T7 DNA polymerase and subsequently ligated to form circular DNA. The DNA-product is introduced into proficient uracil-glycosylase strain that removes uracil bases and creates abasic site in the original DNA-strand. The DNA-strand synthesized by T7 DNA polymerase does not contain uracil and is predominantly amplified. Based on this principle, Bio-Rad introduced a commercial kit (Muta-Gene M13 *in vitro* Mutagenesis Kit) and guaranties mutation frequencies higher than 50 % (BioRad instruction manual). The efficiency of the protocol has been widely proven [101], however it is more time consuming compared to the QuikChange Site-Directed Mutagenesis Kit from Stratagene.

The QuikChange kit is to our knowledge the most commonly used site-directed mutagenesis protocol [102]. QuikChange procedure utilizes a double-stranded DNA vector containing the gene of interest and a pair of complementary synthetic oligonucleotides with mutated positions in the middle of the primer pair. Each oligonucleotide amplifies one template strand of the vector resulting in nicked DNA products. The PCR product is not methylated or hemi-methylated in contrast to the parental vector template. By DpnI digest, the methylated or hemi-methylated parental DNA is degraded and the remaining mutated nicked DNA is transformed into *E. coli*. The whole procedure including PCR and transformation can be completed within one day. Key for success is to reduce the template concentration to values below 50 ng for efficient DpnI digest. The principle was extended to create simultaneously point mutations at different sites in a gene of interest (Stratagene QuikChange Multi Site-Directed Mutagenesis Kit) [103]. Efficiencies >50 % have been reported for randomizing three different positions simultaneously [103].

The QuikChange protocol has also been used in costly studies to saturate one by one all codons of a gene. It proves thereby to be very useful in identifying structurally important residues [104] and to broaden our knowledge about evolution strategies [105].

CATEGORY I & III

Methods Employing Whole Cells and Enzymatic Methods for Generating Diversity

A major disadvantage of mutator strains employing a low fidelity DNA polymerase III (Pol III) is the accumulation of mutations in chromosomal DNA leading to genetic instability of these *E. coli* cells [106]. Polymerase I (Pol I) is a repair polymerase in *E. coli* and Pol I variants with low fidelity have been reported [106,107]. Mutations can be directed into the gene of interest by overexpressing a low fidelity Pol I polymerase and introducing the gene of interest on a vector after a preferred Pol I binding region [107].

Wild-type Pol I has a relatively high mutation frequency of 1.4-1.9 x 10^{-7} [107]. The mutation frequency was further increased to 2.1 x 10^{-3} in a triple mutant (D424A/I709N/A759R) under optimized conditions [107]. Control over the mutation frequency is achieved through cultivation time. A major drawback of Pol I systems was the relatively short gene length that could be amplified and mutagenized (~500 bp-1 kbp downstream of the ori then a switch to the more accurate Pol III occurs). Loeb and coworker were able to amplify and mutagenize fragments up to 3 kbp, an important prerequisite for a general use of Pol I systems [107]. The distribution of mutations over the gene remains however uneven, especially in the low fidelity variants of Pol I with a mutation frequency decrease of at least one order of magnitude after the first 650 nt [107]. Pol I methods are, as other epPCR methods, highly biased {56 % transitions (39 % A→G; 17 % T→C), 9 % transversions} with an unusually high number of frameshifts (35 %) [106]. Overall, simple and robust protocols (only plasmid transformations [107]) make the Pol I based mutagenesis methods appealing alternatives to epPCR and chemical mutagens.

CATEGORY II & III

Methods Employing Whole Cells and Chemical Mutagenesis for Generating Diversity

Before PCR based methods for random mutagenesis became popular in the early 90ies, a myriad of chemical methods had been developed from the 70ies to early 90ies for chemical mutagenesis of gene sequences.

In Table **4**, selected examples are summarized by employed compounds and generated mutational spectra. Most chemical mutagens were deaminating agents (e.g. sodium bisulfite [108], nitrous acid [109] and hydroxylamine [110]) and alkylating agents (e.g. ethyl methane sulfonate [111]). Mutations are generated by incubation of the single- and/or double-stranded DNA with the DNA-modifying mutagen followed by the synthesis of complementary strands, e.g. by reverse transcriptase [108,109]. The mutation frequency can be controlled by exposition time of DNA to a chemical mutagen and the mutagens' concentration. Protocols resulting in uracil formation require an *E. coli* strain deficient in uracil-DNA glycosylase [108].

A valuable feature of chemical mutagenesis methods and UV-induced deaminations of cytosine [112] is the generation of subsequent mutations (e.g. CC→TT transitions). A main drawback of chemical modification is the often strong bias that originates from the specific base-targeting nature of the chemical mutagen. Methoxylamine, for example, reacts specifically with cytosine [108] resulting in mainly GC→AT transitions [108]. Overall, chemical mutagenesis of DNA is simple and well established [113]; it is therefore surprising that chemical mutagenesis methods have barely been used for directed protein evolution. One exception is a recent publication from Wu and coworkers [111] who randomized the gene of a serine protease with the chemical mutagen ethyl methane sulfonate, yielding 76.3 % transitions ((GC→AT); (AT→GC)) and 23.7 % transversions ((G→C); (C→G)). The treatment of DNA with the ethyl methane sulfonate reduced the transformation efficiency by only 40 % [111].

SUMMARY

Performance of Current Random Mutagenesis Methods

An ideal mutagenesis method replaces any amino acid of a polypeptide chain by 19 different amino acids in a statistical manner without limiting protein expression in the host organism. We stated seven important benchmarks for a valuable random mutagenesis method: unbiased mutational spectrum, controllable mutation frequency, consecutive nucleotide substitutions or codon-based substitutions, enables subset mutagenesis (e.g. introducing mainly positively or negatively charged amino acids), is independent of gene length, technically simple and reproducible, and economical.

Table 4. Summary of Chemical Mutagens Used in Random Mutagenesis

Chemical Mutagens	Mode of Action	Mutations Observed
Sodium bisulfite [108]	Deaminates cytosine to uracil	Mostly G→A, C→T
Nitrous acid [109]	Deaminates cytosine to uracil, adenosine to hypoxanthine and guanosine to xanthine	See Table **2**
Hydroxylamine [110]	Deaminating agent	Mostly G→A, C→T
Ethyl methane sulfonate [111]	Alkylating agent	See Table **2**
Methoxylamine [108]	Reacts with cytosine to give N4-methoxycytosine	Mostly G→A, C→T
Formic acid [109]	Depurines DNA by breaking N-glycosyl bonds of purine bases	See Table **2**
Hydrazine [109]	Opens pyrimidine rings	See Table **2**

Table 5. Performance of Random Mutagenesis Methods in Benchmark Form

Category	Method	Unbiased Mutational Spectrum	Controllable Mutation Frequency	Consecutive Nucleotide Substitution or Codon-Based Substitution	Subset Mutations	Independent of Gene Length	Technically Simple and Robust	Economical (1-inexpensive 2-fair 3-costly)
I	PCR methods under low fidelity conditions	No	Easy	No	No	Yes	Yes	1
I	Enzymatic methods employing reverse transcriptase or Klenow fragment	No	Difficult	No	No	Yes	Yes	1
II	Randomization of codons	Yes	Easy	Yes	Yes	No	Yes	3
III	Mutator strains	No	Difficult	No	No	Yes	Yes	2
III	Mutator plasmids for directed evolution of host organisms	Likely no*	Difficult	No	No	Yes	Yes	1
III	Environmental as well as physiological stress	Likely no*	Difficult	No	No	Yes	Yes	1
I & II	PCR methods employing nucleotide analogues	No	Easy	No	No	Yes	Yes	2
I & II	Heavy metal and organic solvent-mediated epPCR methods	No	Easy	No	No	Yes	Yes	1
I & II	Cassette mutagenesis	Yes	Easy	Yes	Yes	No	Yes	2
I & II	Random insertion and deletion (RID) mutagenesis	Yes	Not required	Yes	Yes	Yes	No	2
I & II	Random elongation mutagenesis (REM) and codon-based mutagenesis	Yes	Not required	Yes	Yes	Yes	Yes	2
I & II	Sequence Saturation Mutagenesis (SeSaM)	Yes	Easy	Rare event	No	Yes	Yes	2
I & II	Site-directed hybridization mutagenesis	Yes	Not required	Yes	Yes	Yes	Yes	2
I & III	Methods employing whole cells and enzymatic methods for generating diversity	No	Difficult	No	No	No	Yes	1
II & III	Methods employing whole cells and chemical mutagenesis for generating diversity	No	Difficult	Rare event	No	Yes	Yes	1

* Not reported.

We summarized the pro and cons of each method in Table **5** (overview form) and in Table **6** (detailed outline). None of the current methods scores high in all benchmark criteria. Diverse libraries of high quality can for example be created by RID or synthetic gene synthesis, but they lack either the technical simplicity of epPCR and *in vivo* methods or the cost-effectiveness of Pol I methods.

Strategies for Generating Genetic Diversity

Depending on the property of interest, various strategies for generating diversity at the gene level can be applied [114]. Current success stories do not allow us to rank different diversity strategies according to enzyme fold, enzyme property or enzyme reaction. The main reason lies in a lack of reports that screened statistically significant numbers of clones, in a manner that diversity methods can be compared objectively. We have therefore tried, in Fig. **2** the RaMuS flowchart (Random Mutagenesis Strategy flowchart), to provide a guide that helps especially researchers entering the field of evolutive protein design to achieve success and avoid pitfalls. We will further discuss in a broader prospective which requirements have to be

Table 6. Performance of Random Mutagenesis Methods in Pros and Cons Form

Category	Method		Principle	Main Advantages	Main Disadvantages	Comments
I	PCR methods under low fidelity conditions	Mutagenic PCR (Mn^{2+} and/or unbalanced dNTPs)	Reducing fidelity of DNA polymerases during DNA amplification in PCRs	- Technically simple and robust - Tunable mutation frequency by varying concentration of Mn^{2+}	- Transitions favored	Most commonly used method for directed evolution
		GeneMorph Random Mutagenesis Kit	Employing low fidelity polymerases (Mutazyme I and Mutazyme II) for DNA amplification in PCR	- Technically simple and robust - Tunable mutation frequency by varying initial amount of target DNA	- Biased mutational spectra; especially (AT→CG) and (GC→CG) are underrepresented	Alternative to *Taq* DNA polymerase in epPCR
		"Bringer" strategy	Variation of Stratagene QuikChange method by increasing number of PCR cycles	- Technically simple - No subcloning step	- Mutations not targeted to gene of interest	
I	Enzymatic methods employing reverse transcriptase and Klenow fragment	Enzymatic method employing reverse transcriptase	Using reverse transcriptase to incorporate a "mutagenic" nucleotide to a random fragment pool created by limited primer elongation		- Low mutation efficiency (40-60 %) - Transitions favored	
		Enzymatic method employing Klenow fragment	Using Klenow fragment to incorporate a "mutagenic" nucleotide to a random fragment pool created by time-controlled random exonuclease III digestion		- Low mutation efficiency (37.5 %) - Non-even length distribution of random DNA fragments after exonuclease III digest	
II	Randomization of codons	Synthetic gene synthesis	Chemical synthesis of DNA	- Allows optimizing codon usage for higher expression - Allows introducing restriction sites - Allows introducing affinity tags or to remove inverted repeats	- Costly	Application mainly in industry
III	Mutator strains	XL1-Red	Transforming plasmids into *E. coli* strain deficient in major DNA repair pathways	- Technically simple - No subcloning step	- Mutations not limited to gene of interest - Host is genetically not stable	Suitable for researchers entering the field of protein engineering
III	Mutator plasmids		Temporarily inducing recombinant expression of "mutator protein"	- Technically simple - Plasmid curing allows to restore a genetically stable and improved strain		Suitable for directed evolution of host organisms
I & II	PCR methods employing nucleotide analogues		Taking advantages of nucleotide analogues exhibiting promiscuous base-pairing properties in PCR	- Technically simple - Tunable mutation frequency by adjusting nucleotide analogue concentrations	- Biased mutational spectrum depending on base-pairing properties of nucleotide analogues	
I & II	Heavy metal and organic solvent-mediated epPCR methods	Error-prone rolling circle amplification	Reducing fidelity of bacteriophage ϕ29 DNA polymerase during rolling circle amplification of DNA using Mn^{2+}	- Technically simple - No subcloning step - Tunable mutation frequency by adjusting reaction time	- Transitions favored	
I & II		Alcohol-mediated epPCR	Reducing fidelity of polymerases during DNA amplification in PCR by supplementing alcohols	- Technically simple	- Biased to replacement of Gs and Cs	Mutational bias opposite to *Taq* DNA polymerase that favors replacements of As and Ts

(Table 6) contd…..

Category	Method		Principle	Main Advantages	Main Disadvantages	Comments
I & II	Cassette mutagenesis		Ligating completely or partially randomized oligos into digested plasmids	- Allows scanning of amino acid substitutions in a selected protein region	- Requires hypothesis of structure-function relationship	Useful for screening of structurally or catalytically important residues
I & II	Random insertion and deletion (RID) mutagenesis		Insertion and/or deletion of sequences of defined length at random positions	- Highly diverse mutant libraries	- Technically demanding (many ligation steps)	Suitable for advanced users
I & II	Random elongation mutagenesis (REM) and codon-based mutagenesis	Random elongation mutagenesis (REM)	Elongating random peptide tails to C-termini of truncated enzymes	- Technically simple	- Limited by elongation at either C-termini or N-termini	
		"Codon Shuffling"	A pool of 6 bp-DNA duplexes of random sequences are ligated to longer fragments and subsequently fused to truncated gene of interest	- Technically simple		
		Codon-based random deletion (COBARDE)	Random or combinatorial deletion of codons in-frame within a target region	- Technically simple		
I & II	Sequence Saturation Mutagenesis (SeSaM)		Saturating every single nucleotide position of a target gene with all 4 standard nucleotides by using universal bases	- Every single nucleotide position is subjected to mutagenesis - No polymerase bias - Consecutive mutations have been observed		Suitable for creating highly diverse libraries
I & II	Site-directed hybridization mutagenesis		Employing synthetic oligos carrying desired mutations to synthesize circular DNAs	- Technically simple	- Requires hypothesis of structure-function relationship	Useful for investigating structure-function relationship, modifying vectors and introducing restriction sites
I & III	Methods employing whole cells and enzymatic methods for generating diversity	Pol I method	Generating random mutations in the gene of interest using DNA polymerase I mutant(s)	- Technically simple - Cost-effective method	- Transitions favored - More mutations are accumulated at position < 650 nt	
II & III	Methods employing whole cells and chemical mutagenesis for generating diversity		Exposing gene of interest to DNA-modifying chemical mutagens	- Technically simple	- Transitions favored - Difficult to control mutation rate - Nucleobases are targeted differently	

considered and have to be fulfilled for a promising directed evolution experiment (Fig. **3**).

The RaMuS Flowchart

Before starting a directed evolution experiment, it is important to ensure that a suitable starting point has been selected in order to minimize the number of properties that have to be improved by directed evolution. Many researchers were hit hard by Murphy during directed evolution since improving the property of interest (e.g. activity) came at the expense of reduced thermal resistance or reduced total turnover number making the catalyst unsuitable for industrial applications. Having a good starting point is a very important consideration before initiating a time-consuming and costly

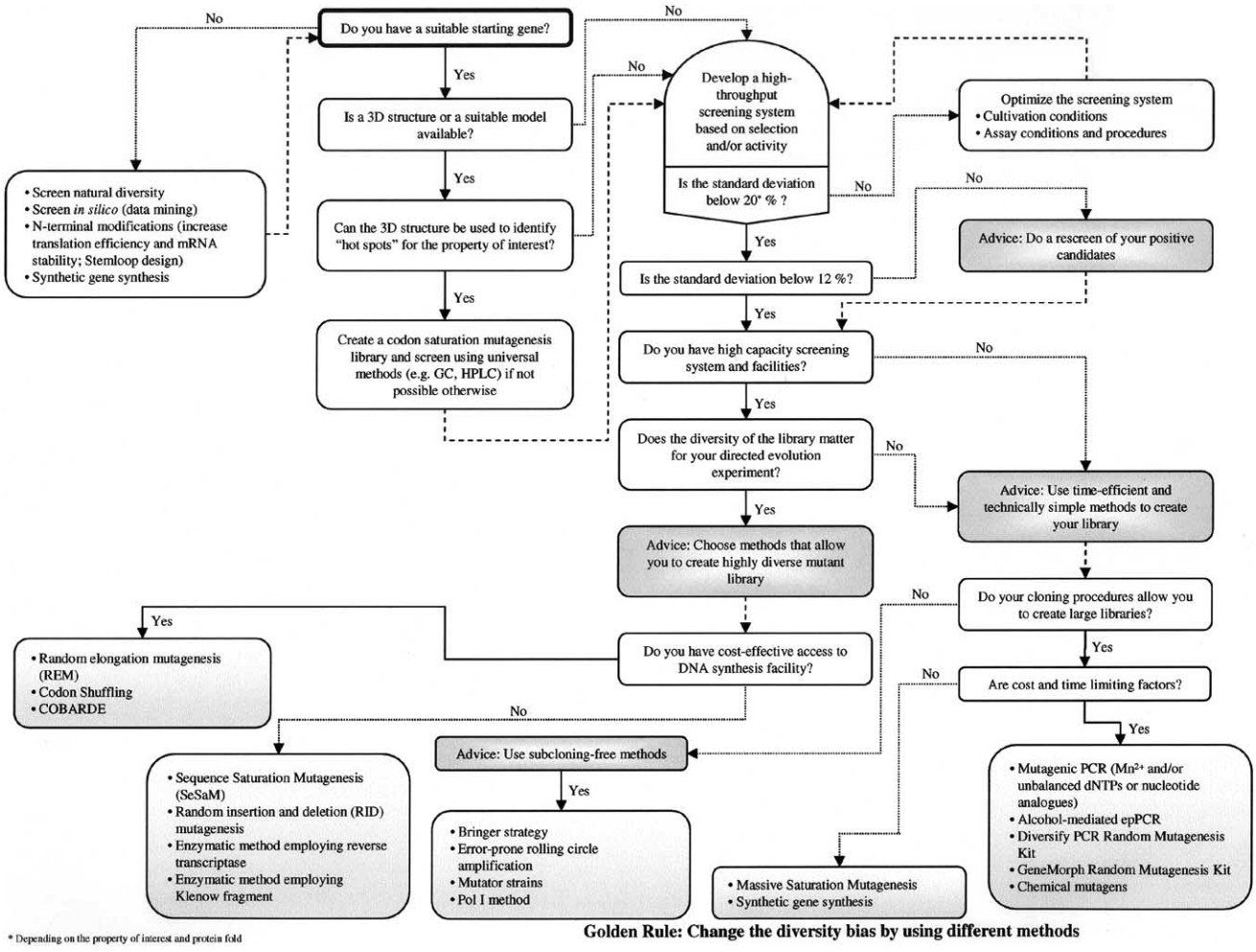

Fig. (2). RaMuS (Random Mutagenesis Strategy) flowchart: A guide for picking a suitable random mutagenesis method.

directed evolution experiment. It should further be noted that by introducing only 1 to 3 amino acid changes per round of directed evolution, it is unlikely to improve more that one property at once. The few amino acid changes would have to be responsible for multiple properties for example improving activity and stability at once; this is unlikely.

After deciding on a good starting, it should be investigated whether structural information and reports identifying important residues are available for the property of interest. If so, this information should be exploited using site-directed mutagenesis that allows us to rapidly saturating these positions [115,116]. Screening can in the worst scenario be performed with universal low-throughput systems such as GC, HPLC or TLC. This approach can lead within a few weeks to valuable results. It is much faster than developing and validating a screening system for directed evolution that takes often 6 to 8 months.

The Screening System

The next step in the RaMuS flowchart represents the heart of every directed evolution experiment: the screening system. The ultimate goal should be a reliable and high-throughput screening (HTS) system that eliminates rate limiting steps such as centrifugation or removal of cell debris by filtration [117]. For affinity screens, powerful HTS systems have been developed and summarized in reviews [25,118] as well as book chapters [119]. For activity measuring screening systems in liquid culture, a reliable protocol has to be established by a) optimizing growth and expression conditions, b) minimizing evaporation at the rim of microtiter plates, and c) improving activity as well as detection limits for example by permeabilizing cells with polymyxin B [23,120] and dissolving cell debris with NaOH/SDS [23]. In one of our directed evolution experiment improving the organic solvent resistance of a monooxygenase [23], we achieved without any re-screen hit rates of 65 %. The pNCA-screening system has standard deviations below 12 % [120,121]. Directed evolution on a monooxygenase and on glucose oxidizing enzyme using screening systems with standard deviations up to 20 % resulted in our cases in improved variants after screening small libraries (2000-8000 clones). A rescreen or double rescreen of improved variants was in all cases highly advisable in order to eliminate the majority of false positives.

Standard deviations for reliable screening systems might differ from enzyme to enzyme class or property of interest

and should only be regarded as thumb values and not as strict rules. A rescreen is often advisable to compare the putatively beneficial variants in a single 96-well plate. Developing a reliable screen requires persistence in reducing standard deviations and control experiments to stay within the linear detection range of the screening system. It takes additionally self-discipline to withstand the temptation of rushing screen development and hunting for improved variants before the screen is validated. The latter usually results only in wasting time and resources. Screening systems should constantly be adjusted to improved properties. Activity improvement can for example lead to substrate limitations in the detection system or to product formation that exceeds the linear detection range of the screening systems. Achieved improvements per round of mutagenesis would as consequence diminish.

The Diversity Strategy

When selecting a strategy to construct a mutant library, various considerations have to be balanced: the mutant library size should fit the screening capacity, reflect the required diversity of the library, stay within the available time frame and budget, and match the molecular biology experience of the researcher.

Subcloning-free methods such as Pol I methods or mutator strains are simple in handling and well used for validating screening system and initial rounds of directed evolution. Beneficial positions should in general be randomized in a subsequent step using saturation or cassette mutagenesis to overcome the mutational bias. In many cases (e.g. improvement of the organic solvent resistance of P450 BM-3 [23]), saturation mutagenesis of positions identified in epPCR leads often to a further improvement in the property of interest that is comparable to the results of a full round of directed evolution [23]. In addition an efficient *in vitro* recombination method such as multiplex-PCR-based recombination (MUPREC) could be used for combining point mutations [122].

Often 3 to 5 rounds of directed evolution are performed to improve the property of interest significantly; thereby a larger number of mutations (8-10 mutations per gene) are accumulated in the gene of interest. Beneficial mutations can in these cases be distinguished from neutral or non-contributing substitutions with the Staggered Extension Process (StEP) [123,124] by "breeding" the mutagenized gene(s) with the wild-type gene. During the StEP reaction, the average number of mutations per gene is reduced and important substitutions can be identified by sequencing of clones maintaining their improvement. Such an approach is much faster than changing positions one by one with site-directed mutagenesis.

After a few rounds of directed evolution using the same method for mutant library preparation, observed improvements per round can diminish. It is fruitful to revalidate the screening system as described and to change the diversity generating method (e.g. from *Taq* to Mutazyme in epPCR or a combination of them [125]); thereby counteracting the mutational bias.

DESIGNING THE RIGHT STRATEGY FOR DIRECTED EVOLUTION EXPERIMENTS

Directed protein evolution is from a protein engineer's point of view the most universal method for tailoring properties to industrial demands or to elucidate structure-function relationships. Directed evolution eliminates deficiencies of enzymes by altering directly the biocatalyst performance and not by developing technologies to cope with poor enzyme performance or insufficient catalyst properties. Being so universal comes however with a price tag: directed evolution is, compared to enzyme immobilization, rational design and other methods, a time-consuming and expensive method. Results can often not be transferred from one biocatalyst to even a structurally or functionally similar one. The high investment in time and costly screening in combination with an often narrow substrate spectrum of enzymes hamper the rapid development of biocatalysis. Before starting a directed evolution approach, one must be aware that development and validation of a novel screening system might take 6-8 months. Depending on the property to be improved, alternative approaches such as enzyme immobilization might result in comparable results within a shorter time frame.

In Fig. **3**, we outlined a generic strategy for directed evolution which demands answers to three decisive aspects: a) Finding a suitable point for directed evolution, b) Deciding on a screening system, and c) Selecting a method for generating diversity at the gene level. The later question has been discussed in detail in the previous chapter and the RaMuS flowchart (Fig. **2**).

Directed evolution is an iterative adaptation process and the prime directive for designing a directed evolution strategy should follow the ***KISS principle*** (*Keep It Simple and Smart*).

Finding a Good Starting Point

Finding a good starting point for directed evolution is imperative. During a directed evolution experiment, one should bare in mind that while improving one property other properties including stabilities can diminish despite all efforts to mimic application conditions. For finding a suitable biocatalysts (starting point) one could perform data mining in databases (*in silico* screening) [126], screening natural diversity (e.g. metagenome screening [127]) or searching the literature. In cases where protein expression is low and the standard deviation of the assay is high, one should consider optimizing the enzyme expression by constructing a synthetic gene [41]. After having a good starting point for directed protein evolution, it is advisable to focus on the screening system first and to use a technically simple method for mutant library generation and screen validation.

Deciding on a Screening System

A valuable screening system should allow a high throughput at moderate expenses, and be reliable as well as reproducible. Eliminating inactive mutants increases the probability of success; screens based on selection should

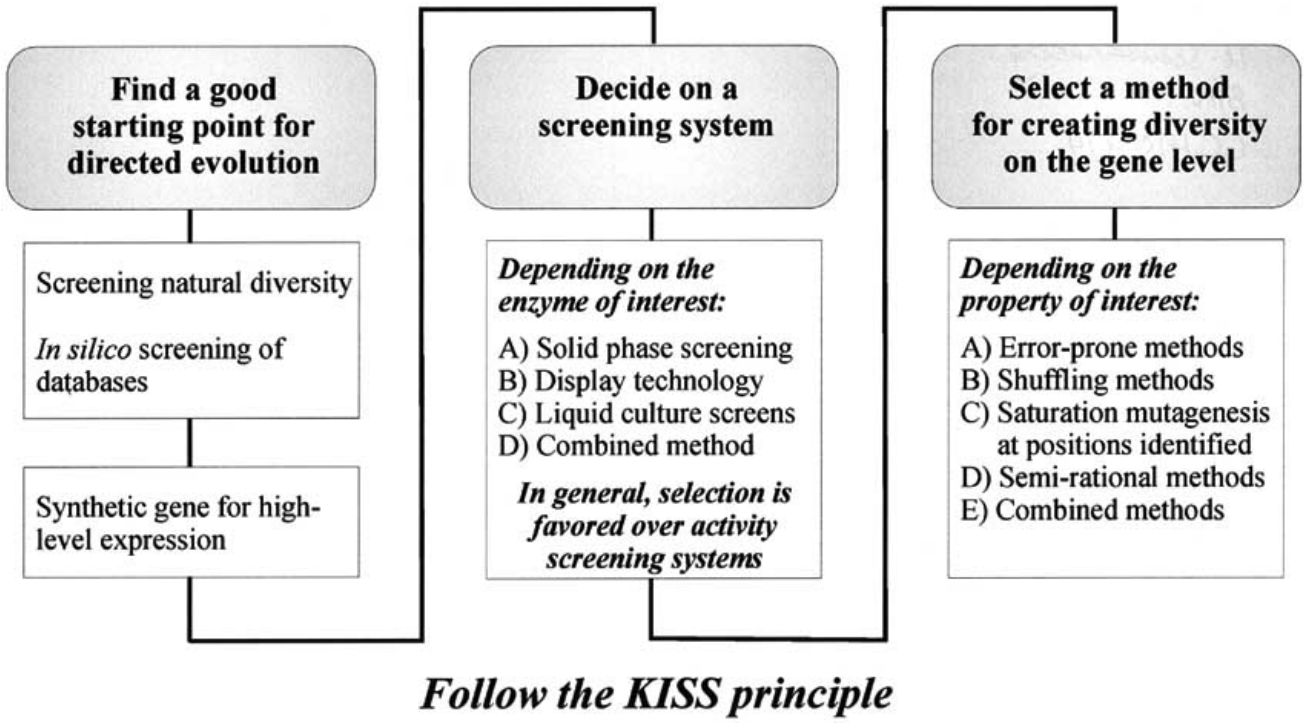

Fig. (3). General strategies for directed evolution.

therefore be favored over non-selective ones. Independently, it should be explored whether FACS [128], display technologies [129,130], *in vitro* evolution systems based on compartmentalization [131,132] or solid phase screening systems [133] could be employed before choosing liquid culture screens with limited throughput in 96- or 384-well format.

OUTLOOK

By summarizing the diversity methods for directed evolution experiments, we aimed to provide researchers a guide to develop their own strategy in randomizing gene sequences in their directed evolution experiments. We hope that further rapid methodological progress in high-throughput screening technology and methods for generating diversity will close the gap between directed evolution success stories and rational understanding of structure-function relationships in evolved mutants. By creating less biased mutant libraries and screening increased numbers of clones, we expect more directed evolution experiments reporting mutations that are not located on a protein surface and can rationally be explained within the time frame of simulations. Semi-rational design by randomizing substrate-binding sites is a valuable complementary approach to understand structure-function relationships as shown by Kazlauskas [134] and others [115,135]. The same effort and dedication that have been fueling directed protein evolution methods and technologies is now required for elucidating structure-function relationships. Bridging this gap will shed the light on principles for evolutive protein design and will lead to superior directed evolution strategies that may depend on the protein fold as well as on the property of interest. Until then, follow in directed evolution experiments the KISS principle (Keep It Simple and Smart).

ACKNOWLEDGEMENTS

We thank Deutsche Forschungsgemeinschaft (DFG, award SCHW 790/2-1) and German Academic Exchange Service (DAAD, Bonn, Germany) for financial support.

REFERENCES

[1] Mills, D.R.; Peterson, R.L.; Spiegelman, S. *Proc. Natl. Acad. Sci. USA,* **1967**, *58*, 217-224.
[2] Brakmann, S.; Johnson, K. Directed molecular evolution of proteins: or how to improve enzymes for biocatalysis, Wiley-VCH: Weinheim, Germany, **2002**.
[3] Brakmann, S.; Schwienhorst, A. Evolutionary methods in biotechnology: Clever tricks for directed evolution, Wiley-VCH: Weinheim, Germany, **2004**.
[4] Arnold, F.H. Evolutionary approaches to protein design (advances in protein chemistry), Academic Press: New York, **2000**.
[5] Jaeger, K.E.; Eggert, T. *Curr. Opin. Biotech.,* **2004**, *15*, 305-313.
[6] Zaks, A. *Curr Opin Chem. Biol.,* **2001**, *5*, 130-136.
[7] Huisman, G.W.; Gray, D. *Curr. Opin. Biotech.,* **2002**, *13*, 352-358.
[8] Doi, N.; Yanagawa, H. *FEBS Lett.,* **1999**, *453*, 305-307.
[9] Furukawa, K. *Trends Biotechnol.,* **2003**, *21*, 187-190.
[10] Locher, C.P.; Soong, N.W.; Whalen, R.G.; Punnonen, J. *Curr. Opin. Mol. Ther.,* **2004**, *6*, 34-39.
[11] Kurtzman, A.L.; Govindarajan, S.; Vahle, K.; Jones, J.T.; Heinrichs, V.; Patten, P.A. *Curr. Opin. Biotech.,* **2001**, *12*, 361-370.
[12] Lee, N.; Bessho, Y.; Wei, K.; Szostak, J.W.; Suga, H. *Nat. Struct. Biol.,* **2000**, *7*, 28-33.
[13] Kawarasaki, Y.; Griswold, K.E.; Stevenson, J.D.; Selzer, T.; Benkovic, S.J.; Iverson, B.L.; Georgiou, G. *Nucleic Acids Res.,* **2003**, *31*, e126.
[14] Lutz, S.; Ostermeier, M.; Moore, G.L.; Maranas, C.D.; Benkovic, S.J. *Proc. Natl. Acad. Sci. USA,* **2001**, *98*, 11248-11253.
[15] Ostermeier, M.; Shim, J.H.; Benkovic, S.J. *Nat. Biotechnol.,* **1999**, *17*, 1205-1209.
[16] Sieber, V.; Martinez, C.A.; Arnold, F.H. *Nat. Biotechnol.,* **2001**, *19*, 456-460.
[17] Wang, P.L. *Dis. Markers,* **2000**, *16*, 3-13.

[18] Cipolla, L. *Comb. Chem. High. Throughput Screen.*, **2004**, *7*, 101-114.
[19] Lin, H.N.; Cornish, V.W. *Angew. Chem. Int. Edit.,* **2002**, *41*, 4403-4425.
[20] Arnold, F.H.; Georgiou, G. Directed enzyme evolution: screening and selection methods, Humana Press: Clifton, NJ, **2003**.
[21] LinGoerke, J.L.; Robbins, D.J.; Burczak, J.D. *Biotechniques,* **1997**, *23*, 409-412.
[22] Vartanian, J.P.; Henry, M.; WainHobson, S. *Nucleic Acids Res.,* **1996**, *24*, 2627-2631.
[23] Wong, T.S.; Arnold, F.H.; Schwaneberg, U. *Biotechnol. Bioeng.,* **2004**, *85*, 351-358.
[24] Daugherty, P.S.; Chen, G.; Iverson, B.L.; Georgiou, G. *Proc. Natl. Acad. Sci. USA,* **2000**, *97*, 2029-2034.
[25] Becker, S.; Schmoldt, H.U.; Adams, T.M.; Wilhelm, S.; Kolmar, H. *Curr. Opin. Biotech.,* **2004**, *15*, 323-329.
[26] Woese, C.R. *Proc. Natl. Acad. Sci. USA,* **1965**, *54*, 71-75.
[27] Zhu, C.T.; Zeng, X.B.; Huang, W.D. *J. Mol. Evol.,* **2003**, *57*, 533-537.
[28] Wong, T.S.; Tee, K.L.; Hauer, B.; Schwaneberg, U. *Nucleic Acids Res.,* **2004**, *32*, e26.
[29] Cadwell, R.C.; Joyce, G.F. *PCR Methods and Applications,* **1992**, *2*, 28-33.
[30] Cirino, P.C.; Mayer, K.M.; Umeno, D. *Methods in Molecular Biology,* **2003**, *231*, 3-9.
[31] Cline, J.; Braman, J.C.; Hogrefe, H.H. *Nucleic Acids Res.,* **1996**, *24*, 3546-3551.
[32] Patel, P.H.; Kawate, H.; Adman, E.; Ashbach, M.; Loeb, L.K. *J. Biol. Chem.,* **2001**, *276*, 5044-5051.
[33] Cline, J.; Hogrefe, H. *Strategies,* **2000**, *13*, 157-162.
[34] Biles, B.D.; Connolly, B.A. *Nucleic Acids Res.,* **2004**, *32*, e176.
[35] Stratagene *Strategies,* **2004**, *17*, 20-21.
[36] Stratagene *Strategies,* **2002**, *15*, 8-9.
[37] Bichet, A.; Bureik, M.; Lenz, N.; Bernhardt, R. *Appl. Biochem. Biotech.,* **2004**, *117*, 115-122.
[38] Lehtovaara, P.M.; Koivula, A.K.; Bamford, J.; Knowles, J.K.C. *Protein Eng.,* **1988**, *2*, 63-68.
[39] Shiraishi, H.; Shimura, Y. *Gene,* **1988**, *64*, 313-319.
[40] Bradel-Tretheway, B.G.; Zhen, Z.; Dewhurst, S. *J. Virol. Methods.,* **2003**, *111*, 145-156.
[41] Neves, F.O.; Ho, P.L.; Raw, I.; Pereira, C.A.; Moreira, C.; Nascimento, A.L.T.O. *Protein Expres. Purif.,* **2004**, *35*, 353-359.
[42] Zhou, Z.Y.; Schnake, P.; Xiao, L.H.; Lal, A.A. *Protein Expres. Purif.,* **2004**, *34*, 87-94.
[43] Wu, X.Q.; Jornvall, H.; Berndt, K.D.; Oppermann, U. *Biochem. Bioph. Res. Co.,* **2004**, *313*, 89-96.
[44] Griswold, K.E.; Mahmood, N.A.; Iverson, B.L.; Georgiou, G. *Protein Expres. Purif.,* **2003**, *27*, 134-142.
[45] Gustafsson, C.; Govindarajan, S.; Minshull, J. *Trends Biotechnol.,* **2004**, *22*, 346-353.
[46] Paulus, M.; Haslbeck, M.; Watzele, M. *Nucleic Acids Res.,* **2004**, *32*, -.
[47] Huang, W.D.; Santi, D.V. *Anal. Biochem.,* **1994**, *218*, 454-457.
[48] Lam, K.S.; Salmon, S.E.; Hersh, E.M.; Hruby, V.J.; Kazmierski, W.M.; Knapp, R.J. *Nature,* **1991**, *354*, 82-84.
[49] Neuner, P.; Cortese, R.; Monaci, P. *Nucleic Acids Res.,* **1998**, *26*, 1223-1227.
[50] Yanez, J.; Arguello, M.; Osuna, J.; Soberon, X.; Gaytan, P. *Nucleic Acids Res.,* **2004**, *32*, e158.
[51] Lyttle, M.H.; Napolitano, E.W.; Calio, B.L.; Kauvar, L.M. *Biotechniques,* **1995**, *19*, 274-281.
[52] Ono, A.; Matsuda, A.; Zhao, J.; Santi, D.V. *Nucleic Acids Res.,* **1995**, *23*, 4677-4682.
[53] Hughes, M.D.; Nagel, D.A.; Santos, A.F.; Sutherland, A.J.; Hine, A.V. *J. Mol. Biol.,* **2003**, *331*, 973-979.
[54] Komiyama, T.; VanderLugt, B.; Fugere, M.; Day, R.; Kaufman, R.J.; Fuller, R.S. *Proc. Natl. Acad. Sci. USA,* **2003**, *100*, 8205-8210.
[55] Dube, D.K.; Parker, J.D.; French, D.C.; Cahill, D.S.; Dube, S.; Horwitz, M.S.Z.; Munir, K.M.; Loeb, L.A. *Biochemistry,* **1991**, *30*, 11760-11767.
[56] Palzkill, T.; Botstein, D. *Proteins,* **1992**, *14*, 29-44.
[57] Miller, J.H. Short course in bacterial genetics: a laboratory manual and handbook for Escherichia coli and related bacteria, Cold Spring Harbor Laboratory Press: New York, **1992**.
[58] Yanofsky, C.; Cox, E.C.; Horn, V. *Proc. Natl. Acad. Sci. USA,* **1966**, *55*, 274-281.
[59] Greener, A.; Callahan, M.; Jerpseth, B. *Mol. Biotechnol.,* **1997**, *7*, 189-195.
[60] Greener, A.; Callahan, M. *Strategies,* **1994**, *7*, 32-34.
[61] Nguyen, A.W.; Daugherty, P.S. In *Methods in molecular biology: Directed evolution library creation: Methods and protocols*; Arnold, F.H., Georgiou, G., Eds.; Humana Press: New Jersey, **2003**; Vol. *231*, pp. 39-44.
[62] Balashov, S.; Humayun, M.Z. *Mutat. Res-Fund. Mol. M.,* **2004**, *548*, 9-18.
[63] Sliva, P.; Kluz, J.; Korona, R. *Genetica,* **2004**, *121*, 285-293.
[64] Gutierrez, P.J.A.; Wang, T.S.F. *Genetics,* **2003**, *165*, 65-81.
[65] Selifonova, O.; Schellenberger, V. In *Methods in molecular biology: Directed evolution library creation: Methods and protocols*; Arnold, F.H., Georgiou, G., Eds.; Humana Press: New Jersey, **2003**; Vol. *231*, pp. 45-52.
[66] Selifonova, O.; Valle, F.; Schellenberger, V. *Appl. Environ. Microb.,* **2001**, *67*, 3645-3649.
[67] Kern, W.V.; Oethinger, M.; Jellen-Ritter, A.S.; Levy, S.B. *Antimicrob. Agents Ch.,* **2000**, *44*, 814-820.
[68] Malavasic, M.J.; Cihlar, R.L. *J. Med. Vet. Mycol.,* **1992**, *30*, 421-432.
[69] Clark, D.P.; Beard, J.P. *J. Gen. Microbiol.,* **1979**, *113*, 267-274.
[70] Foster, P.L. *Annu. Rev. Microbiol.,* **1993**, *47*, 467-504.
[71] Loakes, D. *Nucleic Acids Res.,* **2001**, *29*, 2437-2447.
[72] Day, J.P.; Bergstrom, D.; Hammer, R.P.; Barany, F. *Nucleic Acids Res.,* **1999**, *27*, 1810-1818.
[73] Ghadessy, F.J.; Ramsay, N.; Boudsocq, F.; Loakes, D.; Brown, A.; Iwai, S.; Vaisman, A.; Woodgate, R.; Holliger, P. *Nat. Biotechnol.,* **2004**, *22*, 755-759.
[74] Spee, J.H.; Devos, W.M.; Kuipers, O.P. *Nucleic Acids Res.,* **1993**, *21*, 777-778.
[75] Ohtsuka, E.; Matsuki, S.; Ikehara, M.; Takahashi, Y.; Matsubara, K. *J. Biol. Chem.,* **1985**, *260*, 2605-2608.
[76] Beckman, R.A.; Mildvan, A.S.; Loeb, L.A. *Biochemistry,* **1985**, *24*, 5810-5817.
[77] Kunkel, T.A.; Loeb, L.A. *J. Biol. Chem.,* **1979**, *254*, 5718-5725.
[78] Weymouth, L.A.; Loeb, L.A. *Proc. Natl. Acad. Sci. USA,* **1978**, *75*, 1924-1928.
[79] Fujii, R.; Kitaoka, M.; Hayashi, K. *Nucleic Acids Res.,* **2004**, *32*, e145.
[80] Dean, F.B.; Nelson, J.R.; Giesler, T.L.; Lasken, R.S. *Genome Res.,* **2001**, *11*, 1095-1099.
[81] Nelson, J.R.; Cai, Y.C.; Giesler, T.L.; Farchaus, J.W.; Sundaram, S.T.; Ortiz-Rivera, M.; Hosta, L.P.; Hewitt, P.L.; Mamone, J.A.; Palaniappan, C.; Fuller, C.W. *Biotechniques,* **2002**, 44-47.
[82] Claveau, S.; Sasseville, M.; Beauregard, M. *DNA Cell Biol.,* **2004**, *23*, 789-795.
[83] Borrego, B.; Wienecke, A.; Schwienhorst, A. *Nucleic Acids Res.,* **1995**, *23*, 1834-1835.
[84] Saab-Rincon, G.; Juarez, V.R.; Osuna, J.; Sanchez, F.; Soberon, X. *Protein Eng.,* **2001**, *14*, 149-155.
[85] Black, M.E.; Loeb, L.A. *Methods in Molecular Biology,* **1996**, *57*, 335-349.
[86] Hill, C.M.; Li, W.S.; Thoden, J.B.; Holden, H.M.; Raushel, F.M. *J. Am. Chem. Soc.,* **2003**, *125*, 8990-8991.
[87] Beste, G.; Schmidt, F.S.; Stibora, T.; Skerra, A. *Proc. Natl. Acad. Sci. USA,* **1999**, *96*, 1898-1903.
[88] Sanschagrin, F.; Theriault, E.; Sabbagh, Y.; Voyer, N.; Levesque, R.C. *J. Antimicrob. Chemother.,* **2000**, *45*, 517-520.
[89] Murakami, H.; Hohsaka, T.; Sisido, M. *Nat. Biotechnol.,* **2002**, *20*, 76-81.
[90] Murakami, H.; Hohsaka, T.; Sisido, M. *Methods in Molecular Biology,* **2003**, *231*, 53-64.
[91] Murakami, H.; Hohsaka, T.; Sisido, M. *Nucleic. Acids Symp. Ser.,* **2000**, 69-70.
[92] Kitamura, Y.; Komiyama, M. *Nucleic Acids Res.,* **2002**, *30*, e102.
[93] Matsuura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto, K.; Urabe, I. *Nat. Biotechnol.,* **1999**, *17*, 58-61.
[94] Matsuura, T.; Yomo, T.; Urabe, I. *Methods in Molecular Biology,* **2002**, *182*, 221-230.
[95] Scopes, R.K. *Nat. Biotechnol.,* **1999**, *17*, 21-21.
[96] Chopra, S.; Ranganathan, A. *Chem. Biol.,* **2003**, *10*, 917-926.

[97] Osuna, J.; Yanez, J.; Soberon, X.; Gaytan, P. *Nucleic Acids Res.,* **2004**, *32*, e136.
[98] Motz, M.; Paabo, S.; Kilger, C. *Biotechniques,* **2000**, *29*, 268-270.
[99] Wong, T.S.; Tee, K.L.; Hauer, B.; Schwaneberg, U. *Anal. Biochem.,* **2005**, *341*, 197-9.
[100] Kunkel, T.A. *Proc. Natl. Acad. Sci. USA,* **1985**, *82*, 488-492.
[101] Siemers, N.O.; Yelton, D.E.; Bajorath, J.; Senter, P.D. *Biochemistry,* **1996**, *35*, 2104-2111.
[102] Braman, J.; Papworth, C.; Greener, A. *Methods in Molecular Biology* **1996**, *57*, 31-44.
[103] Hogrefe, H.H.; Cline, J.; Youngblood, G.L.; Allen, R.M. *Biotechniques,* **2002**, *33*, 1158-1160, 1162, 1164-1155.
[104] Liebeton, K.; Zonta, A.; Schimossek, K.; Nardini, M.; Lang, D.; Dijkstra, B.W.; Reetz, M.T.; Jaeger, K.E. *Chem. Biol.,* **2000**, *7*, 709-718.
[105] Jaeger, K.E.; Eggert, T.; Eipper, A.; Reetz, M.T. *Appl. Microbiol. Biot.,* **2001**, *55*, 519-530.
[106] Fabret, C.; Poncet, S.; Danielsen, S.; Borchert, T.V.; Ehrlich, S.D.; Janniere, L. *Nucleic Acids Res.,* **2000**, *28*, -.
[107] Camps, M.; Naukkarinen, J.; Johnson, B.P.; Loeb, L.A. *Proc. Natl. Acad. Sci. USA,* **2003**, *100*, 9727-9732.
[108] Kadonaga, J.T.; Knowles, J.R. *Nucleic Acids Res.,* **1985**, *13*, 1733-1745.
[109] Myers, R.M.; Lerman, L.S.; Maniatis, T. *Science,* **1985**, *229*, 242-247.
[110] Kajiyama, N.; Nakano, E. *Biochemistry,* **1993**, *32*, 13795-13799.
[111] Lai, Y.P.; Huang, J.; Wang, L.F.; Li, J.; Wu, Z.R. *Biotechnol. Bioeng.,* **2004**, *86*, 622-627.
[112] Peng, W.; Shaw, B.R. *Biochemistry,* **1996**, *35*, 10172-10181.
[113] Walton, C.R.; Booth, K.; Stockley, P.G. In *Directed mutagenesis: A practical approach*; McPherson, M.J., Ed.; IRL Press: Oxford, **1991**, pp. 135-162.
[114] Arnold, F.H.; Georgiou, G. Directed evolution library creation: methods and protocols, Humana Press: Clifton, NJ, **2003**.
[115] Wong, T.S.; Wu, N.; Roccatano, D.; Zacharias, M.; Schwaneberg, U. *J. Biomol. Screen.,* **2005**, *10*, 246-252.
[116] Li, Q.S.; Schwaneberg, U.; Fischer, P.; Schmid, R.D. *Chem-Eur. J.,* **2000**, *6*, 1531-1536.
[117] Cohen, N.; Abramov, S.; Dror, Y.; Freeman, A. *Trends Biotechnol.,* **2001**, *19*, 507-510.
[118] Olsen, M.; Iverson, B.; Georgiou, G. *Curr. Opin. Biotech.,* **2000**, *11*, 331-337.
[119] Adams, T.M.; Schmoldt, H.U.; Kolmar, H. In *Evolutionary methods in biotechnology: Clever tricks for directed evolution*; Brakmann, S., Schwienhorst, A., Eds.; Wiley-VCH: Weinheim, Germany, **2004**, pp. 31-44.
[120] Schwaneberg, U.; Otey, C.; Cirino, P.C.; Farinas, E.; Arnold, F.H. *J. Biomol. Screen.,* **2001**, *6*, 111-117.
[121] Schwaneberg, U.; Schmidt-Dannert, C.; Schmitt, J.; Schmid, R.D. *Anal. Biochem.,* **1999**, *269*, 359-366.
[122] Eggert, T.; Funke, S.A.; Rao, N.M.; Acharya, P.; Krumm, H.; Reetz, M.T.; Jaeger, K.E. *ChemBioChem.,* **2005**, *6*, 1062-1067.
[123] Zhao, H.M.; Giver, L.; Shao, Z.X.; Affholter, J.A.; Arnold, F.H. *Nat. Biotechnol.,* **1998**, *16*, 258-261.
[124] Zhao, H.M. *Protein Eng.,* **2004**, *388*, 42-49.
[125] Vanhercke, T.; Ampe, C.; Tirry, L.; Denolf, P. *Anal. Biochem.,* **2005**, *339*, 9-14.
[126] Oh, H.J.; Cho, K.W.; Jung, I.S.; Kim, W.H.; Hur, B.K.; Kim, G.J. *J Mol. Catal. B-Enzym.,* **2003**, *26*, 241-250.
[127] Lorenz, P.; Liebeton, K.; Niehaus, F.; Eck, J. *Curr. Opin. Biotech.,* **2002**, *13*, 572-577.
[128] Kwon, S.J.; Petri, R.; deBoer, A.L.; Schmidt-Dannert, C. *ChemBioChem.,* **2004**, *5*, 1069-1074.
[129] Lipovsek, D.; Pluckthun, A. *J. Immunol. Methods,* **2004**, *290*, 51-67.
[130] Becker, S.; Theile, S.; Heppeler, N.; Michalczyk, A.; Wentzel, A.; Wilhelm, S.; Jaeger, K.E.; Kolmar, H. *FEBS Lett.,* **2005**, *579*, 1177-1182.
[131] Tawfik, D.S.; Griffiths, A.D. *Nat. Biotechnol.,* **1998**, *16*, 652-656.
[132] Griffiths, A.D.; Tawfik, D.S. *Curr. Opin. Biotech.,* **2000**, *11*, 338-353.
[133] Delagrave, S.; Murphy, D.J.; Pruss, J.L.R.; Maffia, A.M.; Marrs, B.L.; Bylina, E.J.; Coleman, W.J.; Grek, C.L.; Dilworth, M.R.; Yang, M.M.; Youvan, D.C. *Protein Eng.,* **2001**, *14*, 261-267.
[134] Park, S.; Morley, K.L.; Horsman, G.P.; Holmquist, M.; Hult, K.; Kazlauskas, R.J. *Chem. Biol.,* **2005**, *12*, 45-54.
[135] Urlacher, V.B.; Lutz-Wahl, S.; Schmid, R.D. *Appl. Microbiol. Biot.,* **2004**, *64*, 317-325.
[136] Shafikhani, S.; Siegel, R.A.; Ferrari, E.; Schellenberger, V. *Biotechniques,* **1997**, *23*, 304-310.
[137] Zaccolo, M.; Williams, D.M.; Brown, D.M.; Gherardi, E. *J. Mol. Biol.,* **1996**, *255*, 589-603.
[138] Ueda, H.; Nakanishi, M.; Suzuki, E.; Nagamune, T. *J. Ferment. Bioeng.,* **1995**, *79*, 303-305.
[139] Hill, F.; Loakes, D.; Brown, D.M. *Proc. Natl. Acad. Sci. USA,* **1998**, *95*, 4258-4263.
[140] Hill, F.; Williams, D.M.; Loakes, D.; Brown, D.M. *Nucleic Acids Res.,* **1998**, *26*, 1144-1149.
[141] Moriyama, K.; Otsuka, C.; Loakes, D.; Negishi, K. *Nucleos Nucleot. Nucl.,* **2001**, *20*, 1473-1483.
[142] Nampalli, S.; Kumar, S. *Bioorg. Med. Chem. Lett.,* **2000**, *10*, 1677-1679.
[143] Negishi, K.; Matsumoto, K.; Bessho, T.; Tada, F.; Hayatsu, H. *Nucleic Acids Symposium Series,* **1988**, 33-36.
[144] Loakes, D.; Brown, D.M.; Salisbury, S.A.; McDougall, M.G.; Neagu, C.; Nampalli, S.; Kumar, S. *Helv. Chim. Acta,* **2003**, *86*, 1193-1204.
[145] Loakes, D.; Guo, M.J.; Brown, D.M.; Salisbury, S.A.; Smith, C.L.; Felix, I.; Kumar, S.; Nampalli, S. *Nucleos Nucleot. Nucl.* **2000**, *19*, 1599-1614.
[146] Kamiya, H.; Ito, M.; Harashima, H. *Biol. Pharm. Bull.,* **2004**, *27*, 621-623.
[147] Kamiya, H.; Kasai, H. *Nucleic Acids Res.,* **2000**, *28*, 1640-1646.

Received: April 4, 2005 **Revised: May 9, 2005** **Accepted: May 16, 2005**