Russian Chemical Reviews **74** (3) 285 – 296 (2005) © 2005 Russian Academy of Sciences and Turpion Ltd

DOI 10.1070/RC2005v074n03ABEH000899

Directed evolution as an approach to the design of efficient biocatalysts

S N Zagrebelny

Contents

I.	Introduction	285
II.	Key ways of directed evolution of enzymes	285
III.	Evolution of thermal stability	289
IV.	Evolution of protein solubility and expression	289
V.	Modification of substrate specificity	290
VI.	Stereoselectivity and enantioselectivity	292
VII.	Selection of the modified gene variants	293
VIII.	Other methods of DNA recombination	293
IX.	Rational approaches to enzyme modification	293
X.	Conclusion	294

Abstract. Methodical approaches to the modification of the catalytic properties of enzymes by various means of directed evolution are summarised. A number of particular examples of the design of enzymes with new and enhanced properties, including higher thermal stability, solubility, substrate specificity, stereoselectivity, *etc*., are considered. The prospects for further development of the directed evolution strategy for enzymes and proteins with other functions are discussed. The bibliography includes 67 references.

I. Introduction

Enzymes are used more and more often in industrial engineering processes. The well-known advantages of enzymes over conventional catalysts include reduced energy consumption due to much higher efficiency, lower material consumption due to high specificity, simpler process design due to a smaller number of product isolation and purification stages, and so on. However, in some cases, these advantages turn into drawbacks. Enzymes, like other proteins, are subject to irreversible denaturation with temperature rise, treatment of organic solvents, or at higher or lower pH values. The too high specificity often does not allow the use of enzymes for reactions of the same type with a broad range of substrates belonging to the same class of compounds but having somewhat different structures. This brings about the task of modifying the properties of enzymes in order to extend the scope of their applicability.

The state-of-the-art knowledge of the principles of organisation of protein molecules does not allow one to predict what properties would arise upon targeted replacement of particular amino acid residues, although methods for such replacement have been developed rather thoroughly at the level of modifications of the corresponding genes. In recent years, methods for enzyme modification by introducing random changes into the structure of the protein macromolecule have been proposed. This gives rise to a broad range of structures which may include the optimal ones for the solution of the particular problem. Then molecules with preferred properties are selected. These manipulations are repeated in several consecutive stages to finally obtain a variant that complies with the task. This approach is called 'directed evolution.' This review presents consideration of various types of this method and evaluation of the results, relying on the data published in 1993 – 2003.

S N Zagrebelny Novosibirsk State University, ul. Pirogova 2, 630090 Novosibirsk, Russian Federation. Fax (7-3832) 39 71 01, tel. (7-3832) 30 06 56, e-mail: snz@post.nsu.ru

Received 16 January 2004
Uspekhi Khimii **74** (3) 307 – 320 (2005); translated by Z P Bobkova

II. Key ways of directed evolution of enzymes

The modification of enzymes has been known for long time. Over several decades, this has been done mainly using specific chemical methods and has aimed at elucidating the role of particular amino acid residues in the mechanism of the catalytic action of enzymes.

The accumulation of data on the structures of enzyme-encoding genes and the development of the experimental procedures for work with genetic material provided the possibility of modifying the enzyme molecules by directed mutagenesis. This allowed more detailed investigation of the structure-functional organisation of enzymes. Simultaneously, attempts to modify enzymes in order to improve their operation properties for use in industry have been undertaken. This gave rise to a new line of research in molecular biology, 'protein engineering.'

However, currently it is impossible to predict the influence of one or other amino acid replacement on protein function. This is despite the numerous data on the 3D structures of enzymes obtained by X-ray diffraction and development of computer modelling (for example, molecular dynamics calculations). This is why methods of modification of enzyme properties along a specified direction based on the introduction of random replacements at the gene level followed by selection of the mutant variants according to a specified criterion have experienced vigorous development. In the numerous publications in this direction, one can distinguish several major approaches.

1. Incremental truncation for the creation of hybrid enzymes

Benkovic and co-workers [1] proposed a combinatorial approach in genetic engineering called incremental truncation. For this purpose, the gene encoding this enzyme is split into unequal parts in

two different ways using different restriction endonucleases so that in one fragment pair, the fragment encoding the N-terminal part of the protein is longer, while in the other pair, the longer fragment is that encoding the C-terminal part. The longer fragments are cloned in different plasmids. This gives the overlapping fragments of a gene, which can be modified independently. Then each fragment is truncated 'from the inside' by treatment with exonuclease. This results in generation of variants with deletions in the gene fragments that encode the N- and C-parts of the protein. The plasmids thus obtained are introduced simultaneously into the same cells devoid of the enzyme under study. The combined expression of gene fragments restores the ability of cells to grow on a minimum medium until the catalytic site of the enzyme is damaged due to shortening of the gene fragments. Using this method, the points at which some of the inner sections of the polypeptide chain can be removed without substantial damage of the catalytic properties of the enzyme were identified, and the monomeric enzyme (glycinamide ribonucleotide formyltransferase from *E. coli,* PurN) was converted into its functional heterodimer.

It is evident that the proposed method allows one to study the structure-functional organisation of the enzyme. Subsequently, chimeric enzymes were produced, irrespective of the degree of DNA homology, using the Incremental Truncation for Creation of Hybrid Enzymes (ITCHY) technique. Figure 1 shows a diagram showing the use of this method to create libraries of modified genes.

The possibility of this non-homologous recombination giving rise to interspecies functional hybrids was demonstrated [2] in relation to PurN and its analogues, the human GART enzyme (hGART), which have only a 50% homology at the DNA. ITCHY libraries between the amino end of PurN and the carboxy end of GART were obtained by successive truncation of the encoding sequences with exonuclease III and ligation of products. The most active PurN-GART hybrid was about 500 times less active than the wild-type PurN.

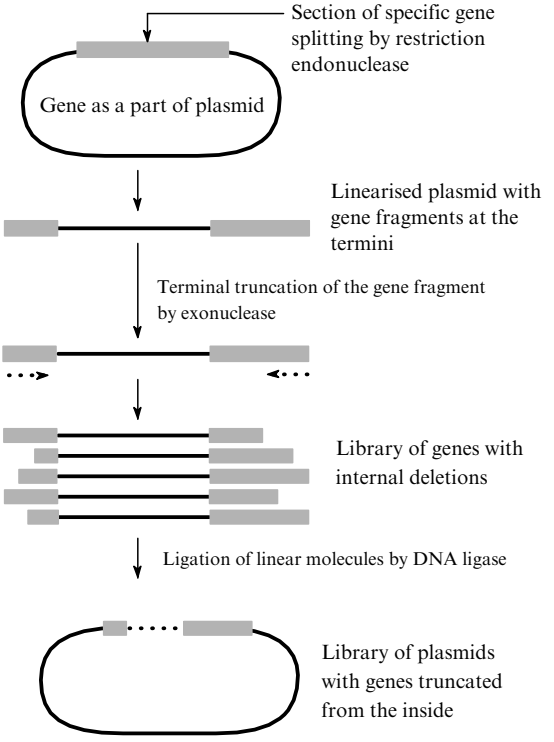


Figure 1. Block diagram for the creation of the library of modified genes by stepwise truncation of enzyme genes with non-homologous recombination.

Subsequently,[3] more general approaches to the compilation of libraries of shortened gene fragments have been proposed based on the possibility of assembling enzyme molecules from large numbers of molecular building units. To this end, when conducting the polymerase chain reaction (PCR), α-thionucleoside triphosphates are added to the reaction mixture in order to insert α-thiophosphate units into the chain being formed. During the subsequent cutting of gene fragments, exonuclease 'stops' at these units. The positions in which the thiophosphate residues occur are generated randomly, thus giving a set of fragments that compose the library. Its components can further undergo recombination *in vitro* in any combinations. Using *E. coli* PurN and its analogue, hGART, functional hybrids with recombination points in the regions with minimum homology were obtained.

The advantages of molecular evolution with non-homologous recombination *in vitro* require further investigation.

2. Evolution of enzymes by DNA family shuffling

The term 'DNA shuffling' has been proposed by Stemmer.[4, 5] Figure 2 shows the pattern of creation of gene libraries by this method. The key idea of the approach is to generate short fragments of the nucleotide sequence of the gene, to multiply fragments using error-prone PCR (as a method for introducing random mutations), to assemble full-size gene sequences by means of PCR and finally to select the variants that meet the conditions imposed by the task set. The fragmentation is accomplished by hydrolysis of a large gene (the researcher cited used a $\sim$1 kb long gene) with DNAse I up to a pool of random fragments with a size of 10 – 50 bp. These fragments can further be assembled into a full-size gene by repeated hybridisation cycles in the presence of a DNA polymerase. The fragments themselves serve as primers due to the presence of homology, recombination taking place when

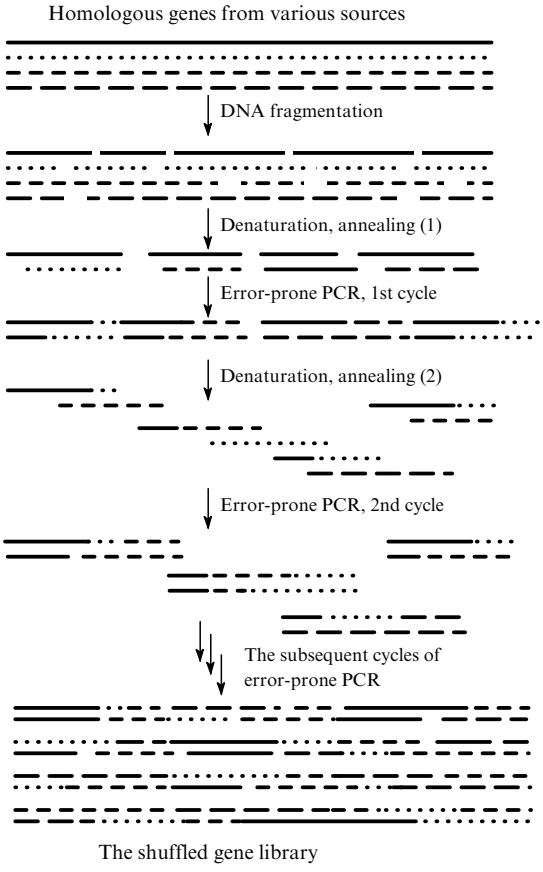


Figure 2. Block diagram for the creation of the library of modified genes by DNA family shuffling.

fragments from one copy of the gene initiate the synthesis on the other copy (which results in template switching). At the end of this round of procedures, selection is performed the pattern of which is chosen in conformity with the project goal. The selected best variant (or several promising variants) is subjected to one more set of the same procedures with the subsequent selection. Several DNA shuffling and selection cycles result in a satisfactory variant, which is subsequently cloned in vectors ensuring a high expression level. Stemmer demonstrated the potential of this approach in relation to *E. coli* β-lactamase and β-galactosidase and to interleukin 1. Shuffling can be performed for one gene to obtain its multiple variants or for families of homologous genes from different sources including those encoding enzymes that catalyse transformations of substrates with related structures. In the latter case, the probability of obtaining improved variants is higher; however, if the number of variants is too great, the selection process would be hampered.

It should be noted that each DNA shuffling gives rise to point mutations. This is undesirable when working with large genes, because the number of variants, among which selection is to be performed substantially increases. A new DNA shuffling protocol increases the accuracy of the procedure at the stages of fragmentation of the modified gene by DNAse I, PCR and the assembly of the full-size gene.[6] This protocol includes, in particular, the replacement of Mg^{2+} ions by Mn^{2+} ions, the limited use of PCR (whenever possible) and the application of more accurate DNA polymerases. The frequency of mutations also decreases when a mixture of *Taq* and *Pfu* (or *Pwo*) in a 1 : 1 ratio is used. Finally, the mutation frequency was decreased to 0.05% and even below; in the case of large genes, this provides the possibility of obtaining a sufficient number of active variants. Further improvement of the protocol can be directed at decreasing the number of PCR cycles and using only *Pfu* polymerase, characterised by higher accuracy than *Taq* polymerase, in the last cycle.

The DNA family shuffling or evolution by recombination *in vitro* of natural homologous genes has been used to improve some properties of enzymes. The thymidine kinase of the herpes simplex virus (HSV TK) is used to obtain thymidine analogues suitable as antiviral agents. The homologous human enzyme has a high specificity towards thymidine, and its use for the synthesis of virus replication inhibitors is of low utility. In a study,[7] shuffling was performed for the two most closely related thymidine kinase genes from different strains of this virus (78% of DNA homology) in order to select chimeras with an enhanced ability to phosphorylate zidovudine (3′-azidothymidine, AzT). Four shuffling cycles resulted in two clones, which increased the sensitivity of the *E. coli* TK to AzT by a factor of 500 relative to the more active HSV-1 enzyme and by a factor of 4000 relative to HSV-2. The shuffling involving HSV-2 increased the yield of sensitive clones, although this enzyme is more stable with respect to AzT.

Under real conditions, an improvement of combinations of enzyme properties is as important as the enhancement of particular features. The shuffling of 26 subtilisin genes from various sources (56% – 99% of DNA identity) has been reported.[8] Subtilisins are proteinases of the bacterial origin, which are used in food industry (to upgrade tough meat), in leather industry (to remove hair from the skin of animals), in medical industry (to prepare protein hydrolysates for parenteral feeding), *etc*., are of substantial practical interest. Genes from various bacteria were used to generate chimeric libraries, which were evaluated using four parameters: activity at 23 °C, thermal stability, resistance against organic solvents and the pH dependence of the activity. Chimeras possessing all combinations of properties of the individual parent enzymes were identified, and structures with markedly improved particular properties have been found.

Controlled evolution methods have been used[9] to study the relationships between the stability and the activity of the mesophilic esterase. Previous studies of these properties in homologous enzymes have shown that the rise of the inactivation temperature under thermal treatment is incompatible with retention of a high catalytic activity at low temperatures. One reason is the different conformation states (in particular, flexibility) responsible for stability and activity. Six generations of statistical mutagenesis, recombination and selection have substantially stabilised the *Bacillus subtilis p*-nitrobenzyl esterase (the inactivation temperature increased by more than 14 °C) without deterioration of the catalytic activity at lower temperatures. Analysis of the stability and the activity for a large number of random mutants has shown that these properties are not inversely correlated; nevertheless, mutations that increase the thermal stability while maintaining the low-temperature activity are very rare. Thus, evolution of one feature through accumulation of single amino acid replacements usually takes place at the expense of another feature, irrespective of whether these features obey an inverse correlation or do not correlate at all.

Directed evolution[10] has been proposed for modification of properties of diversified enzymes. This enables fast optimisation of enzymes as regards one or several properties even when data on the structure or the mechanism of action are missing. Directed evolution experiments make use of a simple Darwinian optimisation algorithm with successive approximations. The molecular diversity is brought about by random mutagenesis and/or recombination of the target gene or a family of connected genes. The modified variants are identified by selection, which allows one to choose the required variants. The genes encoding these improved enzymes are used as the initial ones for the subsequent evolution round.

During shuffling of DNA families, the DNA fragmentation by means of restriction endonucleaes (instead of DNAse I) increases the yield of recombinants[11] due to an increase in the chimerisation frequency. For example, an increase in the chimerisation frequency from 1% to 100% has been observed between two catechol-2,3-dioxygenase genes (*xylE* and *nahH*) from different sources having a 84% identity in the DNA sequence. Hydrolysis of each of the parent genes by several sets of resriction endonucleases resulted in the formation of more diverse chimeric structures, and the hybrid variants thus obtained were 8 – 13 times more stable at 50 °C than the parent enzymes.

A DNA shuffling protocol called degenerate oligonucleotide gene shuffling (DOGS) with an enhanced yield of target products in the molecular evolution of the gene has been described.[12] The advent of this new procedure was due to the fact that the proportion of parent (unshuffled) gene variants in the case of DNA fragmentation by DNAse I was too high. This stipulated the necessity of labour-consuming screening for the selection of the chimeric genes, especially in the case of shuffling of loosely related DNA sequences. The new approach was tested using a family of xylanase genes from various sources. These enzymes are of interest as regards the development of advanced processes for the treatment of wastes from the wood processing industry: for destroying hemicelluloses and subsequent use of the hydrolysis products (*e.g*., in the industrial production of ethanol). On the basis of homology of amino acid sequences, fragments were identified in the gene structures and degenerate primers used subsequently for amplification of these gene fragments were constructed. The amplification products were mixed and subjected to joint amplification to obtain full sequences of these genes. This gave a library of chimeric variants. The degree of mosaicism could be varied by specifying the ratio of the initial amplified fragments. The extensive library thus obtained contained variants that differed appreciably in the optimal temperature. Analysis of the nucleotide sequences of these variants identified the origin of the fragments contributing to new properties of enzymes. However, no variants with a changed optimal pH were encountered in this case. This was attributed to the fact that the enzyme active site incorporates two glutamic acid residues, which act as a nucleophile and an acid catalyst, respectively. The replacement of any of them results in the loss of a catalytic activity. This situation is apparently insurmountable in enzyme modification by this method.

A convenient supplement to DNA shuffling is heteroduplex recombination.[13] This method allows one to create libraries of chimeric DNA sequences derived from two homologous parent sequences. First, heteroduplexes are obtained *in vitro* from the parent sequences and then they are used to transform cells of an appropriate host. During the intracellular reparation, chimeric gene libraries are formed. No physical exchange by fragments of the parent genes takes place. The heteroduplex recombination generates numerous chiasmas and can be especially useful for recombination of large genes or whole operons. The method, however, has limitations: the more remote the mutation sites from each other, the higher the recombination frequency. The minimum distance between the mutated units that still allows recombination is 24 bp.

3. Staggered extension process (StEP)

The StEP procedure has been proposed by Zhao *et al.*[14] (Fig. 3). A polymerase chain reaction on a target enzyme gene is initiated by one pair of primers with a short-term stage of chain growth. The PCR products are then subjected to denaturation followed by annealing (heat-induced denaturation – renaturation) during which the synthesised fragments are switched to other chains of the template as all fragments possess a sufficient degree of homology. The process continues until a full-size gene is formed. Owing to switching of the growing chains from one template to the other, gene libraries for recombination containing information from different parent genes are formed. This approach has been used to obtain a thermally hyperstable variant of subtilisin E. The wild-type enzyme from the mesophilic strain *B. subtilis* has a half-inactivation time of ~ 5 min at 65 °C. The mutagenesis using the error-prone PCR and the subsequent selection gave rise to five variants with a 3 – 8-fold increase in the half-inactivation time. Equal amounts of genes of these five variants were subjected to StEP recombination. About 8000 clones expressing the resulting variants were tested for stability at 75 °C and three variants with half-inactivation times (at 65 °C) 25 – 50 times longer than those of the wild type were selected.

The efficiency of StEP recombination is close to that of other *in vitro* recombination methods; however, StEP can be used in one pot, as it does not require separation of the parent sequences and recombination products.

4. Random priming recombination (RPR)

An even simpler method for the compilation of mutant gene libraries has been proposed [15] (Fig. 4). For this purpose, a copy of the gene subject to modification (obtained by PCR amplification) is used as a template in the PCR, while a random set of oligonucleotides containing six nucleotide residues (dpN_6) serve as primers. This PCR using an error-prone DNA polymerase generates a large number of short DNA fragments, which prove to be complementary to different sections of the matrix sequence. These fragments contain point mutations due to both the erroneous incorporations of nucleotides by the polymerase and imperfect hybridisation of primers with the template chain. In these complexes, short DNA fragments can act as primers for one another and are assembled into the full sequence by multiple repetition of these rounds with subsequent screening and selection. This method has a number of advantages compared to the DNA shuffling.

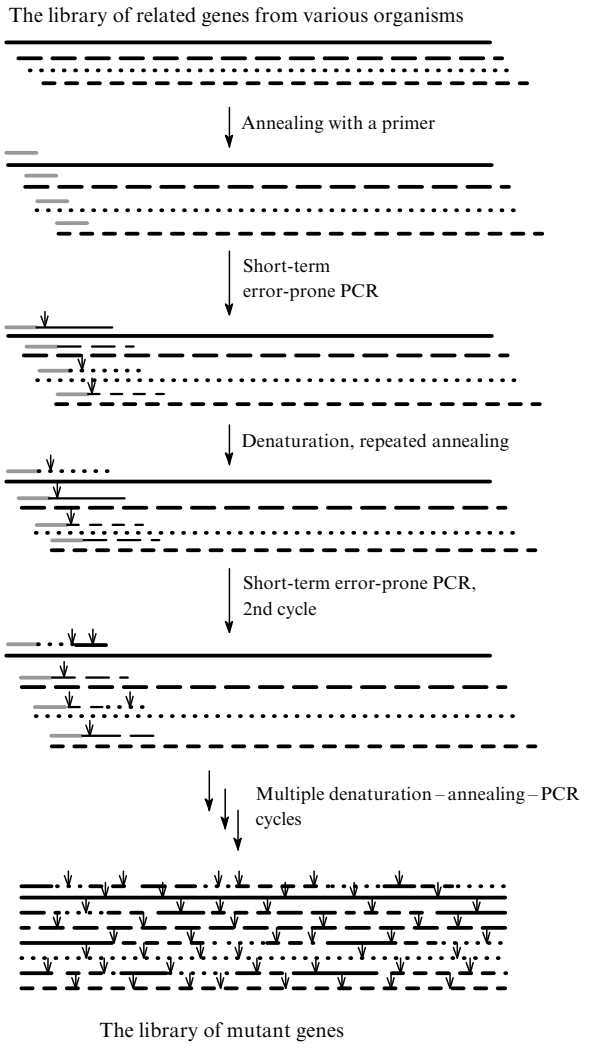


Figure 3. Block diagram for the creation of a library of modified genes by the staggered extension technique;
↓ are the positions of random mutations.

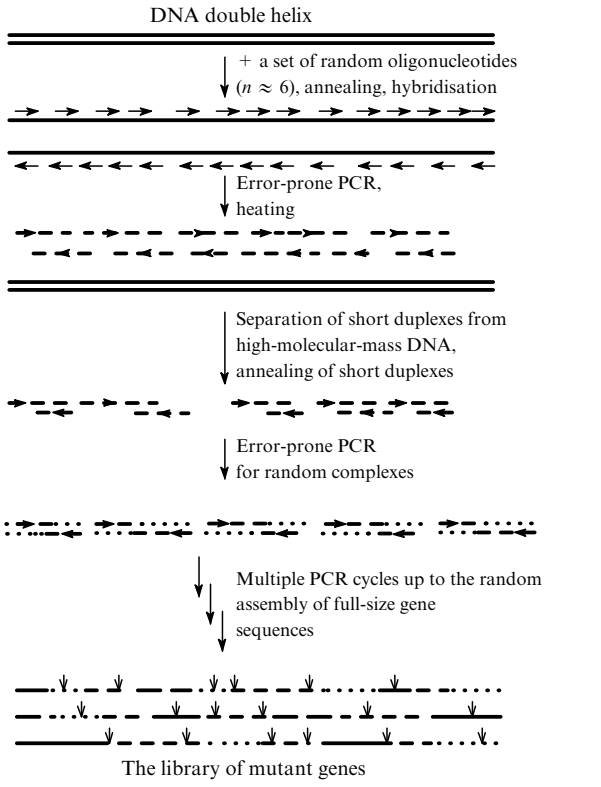


Figure 4. Block diagram for the creation of a library of modified genes by random priming recombination;
↓ are the mutation points induced by the error-prone polymerase reaction.

1. For RPR recombination, both single- and double-stranded DNA can be used as the template without intermediate synthesis of the full-size second strand; potential mutations arise both at the level of single-stranded DNA using error-prone DNA polymerases and at the RNA level using RNA-dependent DNA polymerases.

2. DNA shuffling requires DNA fragmentation with the DNAse I, which must be thoroughly removed prior to the assembly of a full-size mutant gene variant.

3. The random synthetic primers have equal lengths and do not induce predominant synthesis of any gene fragments. The heterogeneity of the sequence of primers allows them to form hybrids with template chains in many positions, so that each nucleotide of the template can be either copied or mutated, with the same frequency, during the chain growth. The random distributions of short fragments of the growing DNA chain along the template (or templates) and point mutations in the DNA fragment ensure the random character of distribution of the chiasmas and mutations in the progeny full-size genes.

4. The synthesis of DNA with random primers does not depend on the template length. Even on short DNA fragments (200 nucleotides long), the synthesis can be initiated equally efficiently as on large molecules (for example, on the linearised DNA of the λ phage). This provides unique opportunities for the engineering of short peptides.

5. Since the parent DNA chain is used only as the template for the synthesis of growing single strands, the whole cycle requires a 10–20 times smaller amount of the initial DNA than the DNA shuffling technique.

III. Evolution of thermal stability

The enhancement of thermal stability is the widely used approach in the modification of enzyme properties. The modern concepts in changing thermal stability are discussed in reviews,[16,17] which report the structural features of thermally stable enzymes providing grounds for the development of rational approaches to their design and application in industry. Note the high level of similarity of primary structures of thermophilic † and mesophilic enzymes, which demonstrates once again imperfection in the state-of-the-art views on structural organisation of protein molecules and the limited nature of the prognostic conclusions.

The modification of a psychrophilic protease, subtilisin S41, has been reported.[18] Previously, this enzyme was subjected to three rounds of mutagenesis/recombination followed by screening. This gave a product with a half-inactivation time (60 °C) 500 times as great as that of the wild-type enzyme. Five more rounds of laboratory evolution gave a product 1200 times as stable as the wild-type enzyme and 20 times as stable as the homologous mesophilic subtilisins SSII and BPNP. About a half of 13 amino acid replacements in the resulting enzyme variants are mutations at serine residues, which occur in thermophilic enzymes less frequently than in mesophilic ones. The variant obtained after the eighth (the last) round of directed evolution contained the Asp216Glu replacement, which is not encountered in either mesophilic or thermophilic enzymes. A nearly neutral replacement resulted, nevertheless, in one more twofold increase in the subtilisin thermal stability with respect to the preceding variant.

Enzymes of thermophilic organisms are often not active enough at low temperatures. An attempt has been made[19] to increase the catalytic activity of the hyperthermophilic indole glycerophosphate synthetase, an enzyme involved in tryptophan biosynthesis, from *Sulfolobus solfatarius*, at moderate temperatures. Using DNA shuffling, *Sulfolobus solfatarius* mutated gene variants able to restore the growth of the *E. coli* culture deficient in the proper enzyme were obtained. It was found that the low activity of the thermophilic enzyme at low temperatures was due to the slow dissociation of the reaction product. In modified enzyme variants, this stage is substantially accelerated, which results in a higher catalytic activity with only a minor decrease in the thermal stability. Thus, in some cases, molecular evolution methods also allow one to elucidate the structural factors responsible for the modified properties of enzymes.

The thermophilic xylose isomerase enzyme from *Thermus thermophilus* has been adapted to normal temperatures in a similar way.[20]

The molecular evolution of the fructosylamino acid oxidase from *Corynebacterium sp.* (strain 2-4-1) has been studied.[21] This enzyme is considered as a potential analytical reagent for determination of non-enzymic 'glycation' of proteins in clinical biochemical analysis.‡ The researchers used mutagenesis *in vivo* by transforming the *E. coli* XL1-Red strain with the plasmid containing the fructosylamino acid oxidase gene.§

After selection of the clones with the modified enzyme, four rounds of mutagenesis and selection were carried out; finally, this gave an enzyme-expressing clone, which retained 94% activity at 47 °C (incubation for 10 min) against 1% for the wild-type enzyme. The mutagenesis *in vivo* used has a number of advantages over the error-prone PCR; however, this approach does not cover all the possibilities provided by the creation and use of mutant gene libraries.

An increase in thermal stability with the simultaneous increase in catalytic activity has been described[22] for the D-amino acid amidase from the soil bacterium *Ochrobactrum anthropi* SV3, which facilitates the preparation of optically pure amino acids during the synthesis of pharmaceuticals, agrochemicals and food additives. Unlike the analogues from other sources, this enzyme catalyses the transformations of a broad range of substrates (amides of amino acids), including those with a bulky side chain. The enzyme gene was subjected to molecular evolution, two rounds of PCR mutagenesis and selection gave a variant with a 5 °C higher denaturation temperature and with a higher catalytic efficiency.

IV. Evolution of protein solubility and expression

The use of enzymes in organic synthesis brings about the problem related to their denaturation in organic solvents. Similar difficulties are faced by the development of cleansing agents that contain proteases and lipases susceptible to denaturation in the presence of detergents. Therefore, attempts have been made to develop enzymes stable against organic solvents. In particular, subtilisin has been modified by molecular evolution methods.[23] This was done using random mutagenesis with a polymerase chain reaction used as the generator of mutations. The selected variant of the PC3 enzyme was 130 times more active than the wild-type enzyme (subtilisin E) in a medium containing 40% DMF. The activity was also retained in a medium with 85% DMF. A study of the primary structure of the selected variant has shown that all 10 amino acid replacements were located near the active and substrate-binding sites of the enzyme. Seven of these can be found in other variants of subtilisin from different strains of the producer. Subsequently, researchers were able to enhance the subtilisin resistance against the organic environment and the thermal stability to an even greater extent.[24,25] Strains producing enzymes with the optimal

† The terms 'thermophilic,' 'mesophilic,' and 'psychrophilic' mean that higher, normal and lower temperatures, respectively, are preferred for enzyme functions; the normal temperature is about 37 °C.

‡ The term 'glycation' is introduced to mean the non-enzymic modification of proteins (unlike glycosylation). This phenomenon takes place in the case of diabetic processes in the body and has a great diagnostic value for evaluation of the patient's condition during the therapy.

§ *E. coli* XL1-Red is a mutant strain devoid of *mutS*, *mutD* and *mutT* genes, which participate in the DNA reparation. The rate of mutations in this strain is 5000 times as high as that of the wild-type strain (see the STRATAGENE catalogue).

pH more convenient for technological purposes and simultaneously with a high stability in the presence of solvents (DMF) have been found [8] in the same library of subtilisin clones.

Attempt have been made [26] to improve the performance properties of the *Bacillus licheniformis* glucose dehydrogenase, which serves as an alternative reagent for glucose determination and ensures regeneration of the cofactor (NAD^+) in the systems of enzymic synthesis using NAD^+-dependent dehydrogenases. Using DNA family shuffling, a glucose dehydrogenase variant free from the proneness to dissociation at high pH values has been obtained. Study of the primary structure of the variant showed that it contained only one amino acid replacement with respect to that prepared previously, which, in turn, also differed from the wild-type enzyme by a single replacement. The thermal stability of the enzyme has also markedly increased.

The creation of superproducers of protein products is faced with problems related to the formation of insoluble grains of these proteins inside the producer cells, so-called 'inclusion bodies'. Their formation is usually due to an improper spatial folding of the protein molecules when present in the cell at high concentration. This results in procedural difficulties in the isolation of the target products and in the subsequent 're-folding' to give the native or a functionally active structure.

The molecular evolution technique can be used to solve this problem. When selecting protein variants able to be folded 'properly', the selection procedure is the most complicated. Two promising approaches to the selection according to solubility and protein expression have been developed. In one of these, the target protein is expressed as a recombinant product with the full sequence of chloramphenicol acetyltransferase (CAT) at the N-end.[27] It was shown that the addition of CAT to insoluble proteins induces a much more pronounced decrease in the stability against chloramphenicol than in the case of addition to soluble proteins. This feature provides the possibility to pick the variants with higher solubility. The soluble HIV integrase mutant was selected in this way from the pool of insoluble wild-type constructs.

Yet another method involves the addition of the target protein to the N-end of the GFP protein ¶ (see Ref. 28). The fluorescence of GFP as a part of these recombinants is related to the proper folding of the attached protein. The utility of GFP as a reporter protein has been confirmed by directed evolution of the folding and demonstrated using different examples. After several rounds of shuffling of DNA with a wild-type sequence, readily soluble proteins were selected [H-subunit of the bullfrog ferritin and the mutant form (Cys33Thr) of V protein of F1 bacteriophage], which usually aggregate during expression in *E. coli* due to improper folding. An important advantage of these approaches over mutagenesis with screening of functional activity is elimination of the necessity of the labour-consuming screening. However, testing of the variants selected in this way for functional activity is still necessary, because the solubility of chimeric protein or a proper folding of the reporter protein does not ensure the retention of activity in the target protein.

The molecular evolution with screening based on GFP has been used [29] for expression of the *Mycobacterium tuberculosis* Rv2002 antigen in the soluble form (usually, this protein is expressed in *E. coli* as 'inclusion bodies'). It was possible to obtain the soluble material in amounts required for detailed structural studies.

The same approach has been used [30] to develop a high-performance producer of gene fragments of HIV1 shell protein. The N-terminal recombinants of a fragment of the gp120 protein of HIV1 containing important immunodominant determinats of the virus with GFP were constructed and subjected to DNA shuffling in combination with PCR. Out of the 300 000 cell colonies, those with continuous fluorescence were selected. Only one such colony was found (the fluorescence protein was distributed over the whole cell volume). The subsequent analysis has shown that under these conditions, the level of synthesis of the target product amounts to $\sim$15% of the total cell protein. This approach led to the desired outcome, despite the lack of a functional test for the target product.

The scope of problems that can be solved by methods of the molecular evolution of enzymes is rather diversified. A study [31] reports an attempt to increase the stability of wild-type β-glucuronidase (GUS) to glutaraldehyde for its use as a reporter in a study of gene expression in the cells of transgene plants. The β-galactosidase system used most often for these purposes is inapplicable, as these cells synthesise their own enzyme. The *E. coli* β-glucuronidase is also not very convenient due to the sensitivity to glutaraldehyde and formaldehyde used to retain the tissue to be studied in order to preserve its structure. These and some other restrictions hampered the use of β-glucuronidase in eukaryotic systems as the reporter enzyme. Using molecular evolution of the enzyme, the sensitivity of β-glucuronidase to aldehydes was reduced. As soon as after one error-prone PCR round and two rounds of DNA shuffling, a mutant variant of GUS stable to higher ($\sim$80 times) concentrations of glutaraldehyde was isolated. Although glutaraldehyde is known to modify proteins through the formation of lateral cross-links at lysine residues; only one of the eight replacements involves this amino acid, while the other seven replacements are mapped on the protein surface near the lysine residue. Apparently, the surface amino acid residues affect appreciably the possibility of modification with glutaraldehyde.

An increase in the expression level of the *Myceliophthora thermophila* laccase in the *Saccharomyces cerevisiae* yeast has been attained by the same method.[32] Laccases present substantial commercial interest as ligninolytic enzymes used to degrade aromatic compounds. This problem often arises in developing industrial processes for bioremediation of areas contaminated by petroleum products. The attempts at elaboration of strains convenient for the industrial cultivation with a high level of laccase production does not always result in the desired outcome. However, using directed evolution, it was possible to increase the activity level of the enzyme in yeast 170-fold (due to increase in the enzyme production and taking into account the 8-fold increase in its catalytic activity). The formation of a site recognised by the Kex2 protease in the C-terminal section of the protein processing was found to be the most efficient (resulting in a 10-fold increase in the overall activity). This is the way the expressed protein accommodates itself to processing in the heterologous host.

V. Modification of substrate specificity

During the design of biocatalysts for industrial use, the problem of adjustment of enzymes to substrates which they do not encounter in nature arises more and more often.

The efficiency of molecular evolution methods in modification of substrate specificity has been convincingly demonstrated [33] using the enhancement of the fucosidase activity of β-galactosidase as an example. The fucose (6-deoxygalactose) glycosides are weakly split by β-galactosidase. The β-galactosidase to β-fucosidase activity ratio of the enzyme attained over eight rounds of DNA shuffling was 1000.

In order to improve the operation characteristics, modification has been performed for xylose isomerase from the thermophilic bacterium *Thermotoga neapolitana*,[34] which also exhibits glucose isomerase activity and, therefore, it is used in the production of glucose – fructose syrups, which is among the largest-scale processes of industrial biotechnology. This enzyme catalyses the glucose to fructose isomerisation much less efficiently than the xylose to xylulose isomerisation. For this reason, attempts at modifying the enzyme for increasing its catalytic activity towards glucose have been undertaken. An increase in the activity of the thermophilic variants of the enzyme at moderate temperatures

¶ GFP — Green Fluorescent Protein — is widely used as a reporter protein in the design of recombinant microbe strains.

and displacement of the optimal pH towards lower values are also improvements of the operating properties of the enzymes. The latter is needed to simplify the process, because the preceding stage, starch conversion, is carried out using glucoamylase operating in a weakly acidic medium. The variant prepared previously was subjected to directed evolution; two rounds of PCR mutagenesis and activity selection (at moderate temperatures and required pH values) gave two new variants, which were much more active at any temperature and pH, the optimal pH being shifted by 0.5 units in the desired direction.

Studies on modification of the substrate specificity of aspartate aminotransferase have been carried out.[35] A combination of DNA shuffling and selection in an auxotrophic *E. coli* strain was used;† this markedly increased the enzyme specificity towards amino and keto acids with chain branching in the β-position. After five rounds of selection, one mutant had a k_{cat}/K_M ratio for unnatural substrates 10^5 times as high as the wild-type enzyme, the catalytic efficiency towards the main substrate being decreased 30-fold. This mutant variant contained 13 replacements of amino acid residues, six of these being responsible for 80% – 90% of the effect. Five of these six replacements were stable, being retained in five mutants that exhibited the highest efficiency with respect to β-branched substrates; one of the six residues was located at a distance from the direct contact with the substrate. Thus, long-range interactions in the protein molecule can contribute significantly to functional properties, which hampers the use of a rational approach to the targeted modification of enzymes.

The mutant variants of cephalosporin acylase obtained from the glutaryl acylase *Pseudomonas* SY-77[36, 37] make it possible to obtain, in one enzymatic stage, 7-aminocephalosporanic (7-ACA) or 7-aminodeacetoxycephalosporanic (7-ADCA) acids used to modify cephalosporin antibiotics. The solution of this problem markedly improves the economy and environmental characteristics of the industrial production of cephalosporins. These acids are prepared by deacylation of cephalosporin C and adipinyl-7-ADCA, respectively (Fig. 5). The removal of the side-chain acyl residue in the antibiotic is catalysed by glutaryl acylase, which, however, exhibits a low activity with respect to adipinic acid residues. A mutant variant of the enzyme with 10-fold increased activity ratio towards the adipinic-to-glutaric acid residues has been obtained; this markedly extends the scope of enzyme application in the modification of penicillin and cephalosporin antibiotics.

The modification of the specificity of the *Rhodobacter sphaeroides* phytoene desaturase has been described.[38] This enzyme catalyses the successive dehydrogenisation of phytoene in the biochemical pathway to isoprenoids, as shown in Fig. 6. In many bacteria, the process stops after the formation of neurosporene; this is also true for *Rhodobacter sphaeroides*. The product of the next dehydrogenisation stage, lycopene, is more valuable, as it has more pronounced antioxidant properties and, hence, it is used as a component of food additives. By using directed evolution technique, the researchers prepared a variant of the *Rhodobacter sphaeroides* enzyme able to synthesise lycopene.

This task was accomplished by two rounds of random mutagenesis with subsequent selection; modification was carried out for an enzyme belonging to the family of neurosporene producers, whereas most of phytoene desaturases (90% of the known enzymes) belong to another family of proteins. The homology in the primary structure between these two families is relatively low (less than 60%). Nevertheless, specificity modification is attained relatively easily: even a single replacement gave a variant with 39% production of lycopene. The final variant (after two more rounds) produced 90% of lycopene. Enzyme variants able to catalyse the synthesis of C_{35} and C_{40} isoprenoids have also been obtained;[39, 40] This opens up the way to the synthesis of new analogues of natural carotene derivatives.[39, 40]

In recent years, substantial progress has been attained in the decoding of biochemical routes of microbial degradation of

† In this case, a strain with a damaged operon of isoleucine biosynthesis was specially designed. This damage could be counterbalanced by providing aspartate aminotransferase with the ability to catalyse transamination of amino acids with a branched side chain.

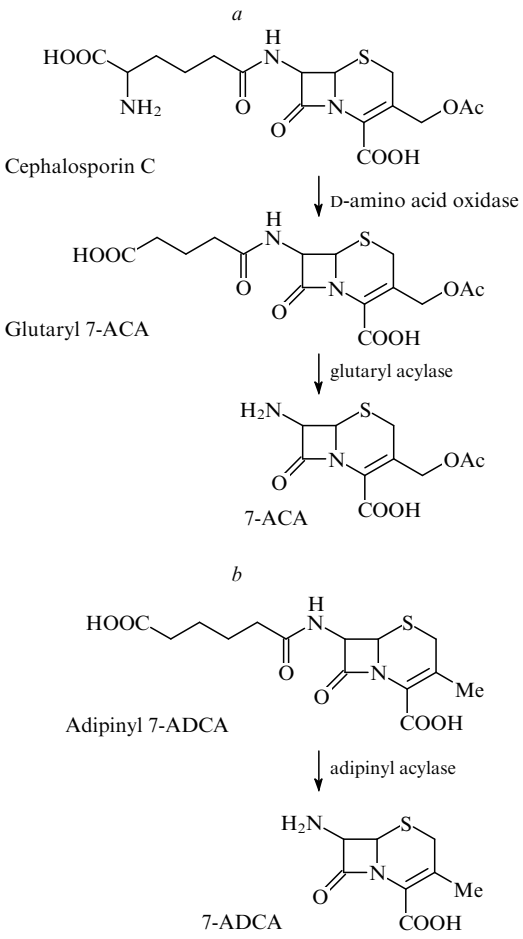


Figure 5. Preparation of deacylated cephalosporins by means of glutaryl acylase (*a*) and adipinyl acylase (*b*).

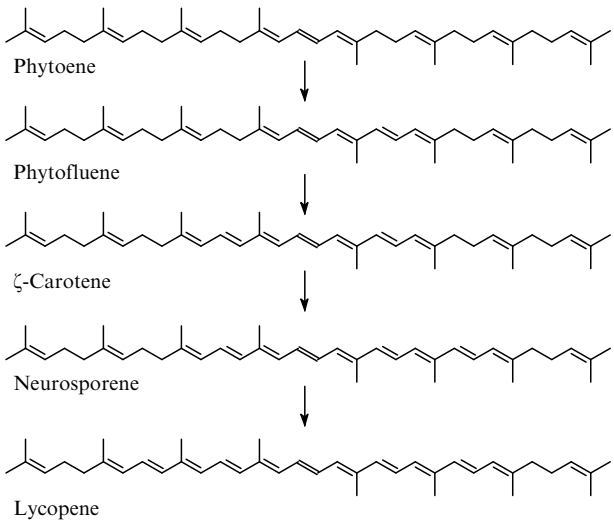


Figure 6. Successive stages of phytoene dehydrogenisation during the formation of lycopene induced by phytoene desaturase.

various aromatic compounds that are frequently encountered pollutants. The attention is focused on the creation of microbe strains able to degrade efficiently these compounds under natural conditions, which can be used for bioremediation of contaminated areas. To this end, attempts have been made [41] to extend the range of substrates for biphenyl dioxygenase by random priming recombination. Two rounds of manipulations with the *bphA1 Pseudomonas pseudoalcaligenes* KF707 gene gave enzyme variants with new properties, in particular, those able to oxidise not only polychlorobiphenyls but also dibenzofuran, dibenzo-*p*-dioxin, dibenzothiophene, fluorene, *i.e.*, compounds hardly utilised by the initial strain.

The obtained [42] 2-hydroxybiphenyl monooxygenase variant from *Pseudomonas azelaica* HBP1 catalyses hydroxylation of indole and its hydroxy and bromo derivatives (wild-type enzymes do not catalyse this reaction). The new variant of the enzyme was found to contain two amino acid replacements separated from each other by 146 residues. The enzymes that oxidise indole and its derivatives can also be used in biotechnological processes to produce indigoid dyes and other products of organic synthesis.

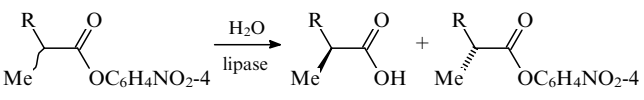
Studies into the development of microorganisms able to biodegrade halogenated organic compounds formed as wastes in the chemical industry or contained in many pesticides, which are degraded with difficulty by the natural populations of microbes in soil and water, are in progress. The activity of the key enzyme for 1,2,3-trichloropropane degradation, haloalkane dehalogenase from *Rhodococcus sp*. (m15-3), has been markedly increased by DNA shuffling.[43] 1,2,3-Trichloropropane is highly toxic and does not occur in nature, but is encountered as a man-made pollutant. The modified gene was introduced into bacteria (*Agrobacterium radiobacter* AD1) that utilise the product made by this enzyme, namely, 2,3-dichloropropan-1-ol. The resulting strain proved to grow on trichloropropane as the only carbon and energy source; thus, it can be used as a bioremediation agent.

VI. Stereoselectivity and enantioselectivity

The production of enantiopure compounds becomes more and more important for chemical and pharmaceutical industries. In 2000, the world sales volume of optically active drugs was above $100 billion. The demand in these products is due to the fact that the molecules of efficient drugs should comply with the asymmetry of the receptors. For example, the US Food and Drug Agency demands strict determination of whether it is possible to produce a drug as an optically pure isomer.

One way to solve this problem is to use enzymes in stereospecific reactions. However, the enantioselectivity of enzymes is often too low, but it can be increased using molecular evolution techniques. The results of such studies have been analysed in detail in a review.[44]

An approach to the development of enantioselective catalysts based on directed evolution *in vitro* has been proposed.[45] Hydrolysis of racemic *p*-nitrophenyl 2-methyldecanoate (**1**) was chosen as the test reaction. Free *S*-acid **2** [the product of hydrolysis of *S*-enantiomer of ester **1**] belongs to a class of compounds in which one finds pharmaceuticals, plant protection compounds and intermediates of organic synthesis. The *Pseudomonas aeruginosa* PAO1 lipase composed of 285 amino acids was used as the catalyst.



1 (*S* + *R*) **2** (*R*)-**1**

The wild-type enzyme exhibits only a 2% enantioselectivity in favour of the *S*-configuration. Using the error-prone PCR, the lipase gene comprising 933 bp was subjected to statistical mutagenesis. The frequency of mutations was varied in such a way as to introduce an average of one or two replacements into each copy of the lipase gene. This was attained by appropriate selection of the PCR conditions (for example, by the change in the concentration of Mg^{2+} ions). The mutated genes were inserted into the expression vector, amplified in *E. coli* and introduced in *P. aeruginosa*. After four selection rounds, the enantioselectivity increased from 2% to 81%.

By inverting the enantioselectivity of D-hydantoinase, which is the key enzyme of the multienzyme route for methionine synthesis, the production of L-methionine in *E. coli* was improved.[46] The production of L-amino acids by means of hydantoinases (Fig. 7) is widely used in industry. For methionine, this reaction proceeds preferentially with the D-isomer of hydantoin, resulting in extremely slow transformation of D-*N*-carbamoylmethionine into the target product (D-selective hydantoinase induces accumulation of D-*N*-carbamoylamino acid, which is converted very slowly into the desired amino acid).

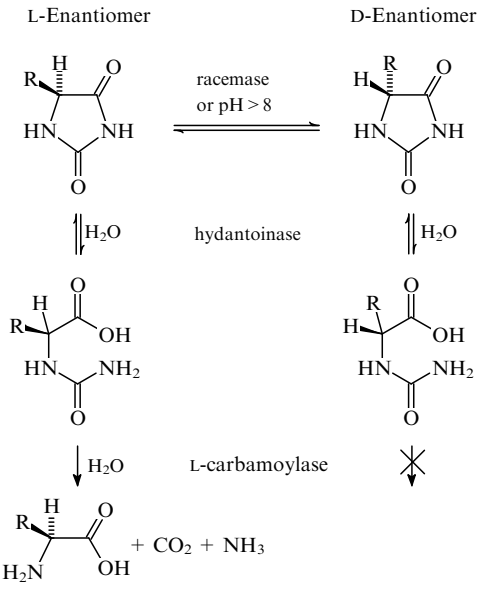


Figure 7. Enzymic synthesis of amino acids from racemic 5-substituted hydantoin.

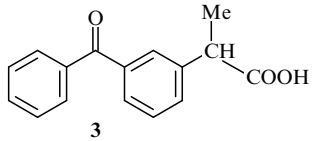
Using successive PCR mutagenesis and screening procedures, the enantioselectivity of D-hydantoinase was inverted [from *ee*(D) 40% to *ee*(L) 20%], the overall activity being increased fivefold. The introduction of the modified L-hydantoinase gene into the recombinant bacterial strain with intact cells as the catalyst increased the productivity with respect to L-Met and decreased the accumulation of undesirable intermediates as compared with wild-type cells. This effect was attained by replacement of only one amino acid. Variants with high D-selectivity [*ee*(D) 90%], which also had a single amino acid replacement, were found. However, due to the lack of efficient screening methods, studies dealing with directed evolution of enantioselective catalysts are few in number.

In recent years, considerable progress has been achieved in the understanding of the role of carbohydrate components in glycosylated proteins; this enhanced the interest in the synthesis of glycoside derivatives of proteins. The chemical synthetic methods in the carbohydrate series are fairly complicated; therefore, enzymic methods become more and more attractive. Commercially justified glycosidases, which exhibit not only hydrolase but also glycosyl transferase activity, are the enzymes of choice. However, the regioselectivity of these enzymes is often inadequate. An attempt to modify the regioselectivity of the *B. stearothermophilus* α-galactosidase has been undertaken.[47] The wild-type enzyme is specific towards 1,6-glycosidic bonds and, to a lesser extent, to 1,3-bonds. Using statistical mutagenesis and the StEP technique for the molecular evolution, enzyme variants that mainly cleaved the 1,3-glycosidic bonds were

obtained. For the majority of enzyme variants obtained, the activity towards 1,6-bonds decreased appreciably as regards both hydrolysis and transglycosylation. Enzyme variants with an (1,3) : (1,6) regiospecificity ratio of 6.9 against 0.3 for the wild-type enzyme have been obtained. However, the transglycosylase activity was markedly lower compared to that of the wild-type enzyme. The researchers believe that they have demonstrated the feasibility of solving this problem; however, further studies are required.

An important application of the stereoselectivity of enzymes is separation of sterically hindered alcohols.[48] Enzymes have already been used for analogous purposes; however, in the case of sterically hindered alcohols, the attempts have often failed. The use of the *Pseudomonas fluorescens* lipase and molecular evolution using the *E. coli* XL1-Red strain with a high frequency of spontaneous mutations resulted in an enzyme variant with an enhanced stereoselectivity towards sterically hindered esters. However, the desired result has not been exactly attained, because mutations are introduced in one variant of the gene. A greater diversity of mutants can be obtained using DNA shuffling.

The esterase from *Pseudomonas fluorescens* KCTC 1767 is used for the preparation of optically pure ketoprofen (**3**) from a racemic mixture of ethyl esters.



3

The modification of the working properties of the enzyme in order to increase the thermal stability, shift the pH optimum towards higher pH and increase the stability against the action of some organic solvents, non-ionic detergents or high substrate concentrations has been carried out [49] with three cycles of molecular evolution. This example illustrates the possibility of simultaneous 'improvement' of several parameters using molecular evolution.

The modification of enzyme stereospecificity is related most often to changing their substrate specificity with respect to a particular enantiomer. Using a tagatose-1,6-diphosphate aldolase-catalysed reaction as an example, a new, in principle, approach has been proposed.[50] The approach implies the change in the reaction stereochemistry in such a way that the enzyme would catalyse the formation of a different product from the same substrates. After six DNA shuffling rounds, the aldolase subjected to the evolution has demonstrated a 80-fold increase in the k_{cat}/K_M ratio in the reaction with an unnatural substrate, fructose 1,6-diphosphate, because the enzyme has switched to the preferred (75%) formation of fructose diphosphate from glyceraldehyde 3-phosphate and dioxyacetone phosphate instead of the predominant (99%) formation of tagatose diphosphate. It is noteworthy that the direct reaction catalysed by this enzyme is cleavage of hexose diphosphates to triose phosphate; therefore, the results obtained could be attributed to the usual modification of the substrate specificity. Nevertheless, the use of this enzyme in the synthesis is of obvious interest.

VII. Selection of the modified gene variants

It has already been noted that the selection procedure is a key stage in directed evolution, which is specific for each particular case. There exist, however, some common techniques that could underlie a selection strategy.

One method is the use of a phage display. This approach implies expression of the target protein as a recombinant with a bacteriophage surface protein, which is exposed on a phage particle and presents the target protein on the surface. The protein is thus brought in contact with the solution and can exhibit the enzymatic activity or present the antigenic determinants of the target protein to antibodies.

An example of using this approach is an attempt to transform the DNA polymerase (the Stoffel fragment of the *Thermus aquaticus* DNA polymerase I) into an RNA polymerase.[51] This was done by mutagenesis of the functional fragments (binding sites of metal ions, polynucleotide duplex, the substrate, and interaction with the matrix) performed separately, which was followed by the assembly of a full-size gene by means of polymerase (two PCR rounds) and by selection using the phage display method. This gave an effective RNA polymerase, which retained the DNA polymerase activity and had a sufficient processivity.

VIII. Other methods of DNA recombination

The methods of directed molecular evolution of enzymes developed during the last decades has proved to be fairly efficient. Meanwhile, accumulation of the data on structural organisation of enzyme molecules is in progress; this allows also the use of rational approaches to modification of the protein structure in the design and perfection of biocatalysts. In some cases, it is useful to combine both approaches.

The possibility of increasing the thermal stability of glowworm luciferase has been demonstrated.[52] The enzyme is used as an analytical reagent in medicine and in the quality control of foodstuffs (in particular, to detect some microorganisms). It can be conjugated to antibodies (specific to particular antigens), and the resulting fluorescence induction or suppression is used to judge the presence or absence of a structure containing this antigen. The insufficient thermal stability of the wild-type enzyme restricts its use; that is why attempts have been made to increase the thermal stability.

Molecular evolution has been used [53] to enhance the thermal stability of phytase, the enzyme that splits phytic acid (inositol hexaphosphate), which is the main form of phosphorus storage in plants. This enzyme, common in microorganisms and in herbivorous animal stock, is missing from the gastrointestinal tract of pigs and birds. The addition of inorganic phosphate to the fodder is uneconomic. To enhance the uptake of the green fodder, phytase tablets are added to the fodder, which requires increasing the natural thermal stability of phytase. The enzyme gene was subjected to site-directed and cassette mutagenesis in order to replace the α-helical fragment of the enzyme from *Aspergillus terreus* by a similar fragment of the *Aspergillus niger* enzyme,[53] which ensured a higher thermal stability of the hybrid.

IX. Rational approaches to enzyme modification

Rational approaches to increasing the thermal stability of enzymes involve either comparison of the amino acid sequences of the target protein and a thermally more stable related protein or detailed analysis of the 3D structure of the target protein in order to predict promising replacements of amino acid residues on the basis of both structural data and the accepted concepts of thermal stabiliy design; the predictions are to be verified by introducing single mutations in the wild-type protein sequence.

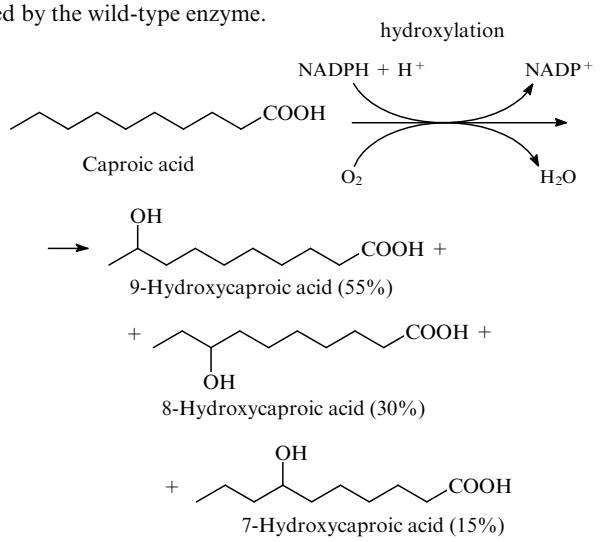
The modern thermostabilisation concepts include, among other issues, the introduction of additional disulfide bridges, improvement of the packing of the hydrophobic core, creation of a network of surface salt bridges or dipole interactions in the α-helix, change in the proneness to form the α-helix and entropy changes (*e.g.*, the X → Pro or Gly → Ala replacements). All rational approaches of this type have been used previously; however, the predictions are not always successful. The reason is that the replacement of the amino acid residue cannot only entail a local effect, but also induces unexpected conformational changes in various (including remote) parts of the molecule. Nevertheless, there are examples where rational methods led to the desired outcome. Thus using site-directed mutagenesis, the contributions of all 17 amino acid residues that were different for barnase and binase (bacterial nucleases with a 85% amino acid homology) to the protein thermal stability have been studied.[54] The effect of one

amino acid replacement on the activation energy of denaturation varied from +1.1 to −1.1 kcal mol^{-1}. The multiple mutant that included six amino acid domains demonstrated an increase in the stability by 3.3 kcal mol^{-1} compared to the wild type. Thus, comparison of related proteins provides useful information regarding the increase in their thermal stability.

The data obtained by various researchers on the molecular evolution of proteases have been described systematically in a review.[55] The site-directed mutagenesis has been used to construct variants of subtilisin with enhanced storage stability and resistance against oxidation. The replacement of Met222 susceptible to oxidation with oxidation-resistant alanine, serine or glutamine residues in *B. amyloliquefaciens* BPN2 subtilisin resulted in prevention of enzyme inactivation by oxygen. Six individual replacements of amino acids (Asn218Ser, Gly169Ala, Tyr217Lys, Met50Phe, Gln206Cys and Asn76Asp) in different positions of the tertiary structure of subtilisin resulted in an increase in the enzyme stability at elevated temperature (65 °C) and in alkaline media (pH 12). Under these denaturing conditions, the inactivation rate of this variant was 300 times lower than that for the wild-type enzyme. Other replacements of amino acids, resulting in a higher enzyme stability, have also been reported.

The rational approach in enzyme reconstruction has been applied[56] to D-amino acid oxidase from the *Rhodotorula gracilis* yeast, which is used to remove the side chain of cephalosporin in the synthesis of antibiotics of this series. Using X-ray diffraction data for the crystalline enzyme, a mutant variant of Met213Arg, capable of catalysing the oxidative deamination of aspartic acid has been obtained (the wild-type enzyme does not use the aspartic or glutaminic acids as substrates). This results in a lower activity with respect to alanine; however, the substrate specificity is extended). It is worth noting that D-amino acid oxidase from *Rhodotorula gracilis* is the most promising enzyme for industrial use owing to the high catalytic activity and strong binding of the cofactor (FAD).

The specificity of *Bacillus megaterium* cytochrome P450 BM-3 has been modified[57] in order to make it catalyse hydroxylation of the hydrocarbon radical of short-chain (less than C_{12}) fatty acids at the pre-terminal units. The rational approach was based on X-ray data for the enzyme. However, the choice of the replacement was determined by inclusion into the sequence of PCR primers of triplets of a random sequence into positions planned for the replacement. Finally, the resulting variants catalysed the oxidation of caproic acid, which cannot be hydroxylated by the wild-type enzyme.



Subsequently, modification of cytochrome has been carried out.[58] Thus it became capable of oxidising polycyclic aromatic hydrocarbons (naphthalene, fluorene, acenaphthene and other). To impart these properties to the enzyme, it was sufficient to modify the same amino acid residues as in the case of oxidation of aliphatic fatty acids, although the replacements were different. The goal was attained in a triple mutant in which the Leu188Gln replacement proved to be the optimal, instead of Leu188Lys.

A concept of increasing the thermal stability of enzymes has been proposed;[16] the concept is based on not only a comparison of amino acid sequences of homologous enzymes from various sources but also on the construction of a consensus sequence from the amino acids found most often. This amino acid sequence is transposed into a nucleotide sequence with allowance for the frequency of codon use (using the library of the amino acid sequences of phytases as an example). Then the gene corresponding to the consensus sequence is synthesised by polymerase assembling, cloned and expressed. The very first variant was much more stable than any of the parent proteins. Several rounds with inclusion of additional genes from new sources into the comparison of sequences have resulted in an enzyme variant with a denaturation temperature of 88.5 °C, which is higher by 8 °C than that for the initial *Aspergillus fumigatus* enzyme. The activity of the consensus enzyme was higher than that of the parent enzymes over the whole pH range. The contribution of each amino acid replacement was evaluated. Later, the denaturation temperature of the consensus phytase was raised additionally by 2 °C.

Most bacterial phytases have the pH optimum in the neutral region, whereas the medium in the gastrointestinal tract of animals is weakly acidic. Therefore, enzyme variants with the pH optimum shifted toward lower pH have been obtained[59,60] by site-directed mutagenesis based on the consensus phytase with replacement of amino acid residues affecting the pH optimum of the enzyme.

Modification of formate dehydrogenases by point replacements of amino acid residues has been reported.[61,62] Formate dehydrogenase is an effective component of systems for regeneration of the NAD^+ cofactor; it is used in the enzymic synthesis of many pharmaceuticals. An increase in the thermal stability of the enzyme due to enhancement of the hydrophobic character of α-helical fragments was induced by the replacement of serine residues in positions 131, 160, 184 and 228 by alanine residues by means of site-directed mutagenesis.[61] Interestingly, the effect of multiple replacements is additive, although these residues are rather remote in the protein chain. An enzyme variant capable of regenerating $NADP^+$ has been prepared[62] on the basis of computer analysis of the structure by replacement of two neighbouring amino acid residues in formate dehydrogenase from bakers' yeast: Asp196Ala and Tyr197Arg. The ratio of the catalytic efficiencies (k_{cat}/K_M) for NADP and NAD has increased by almost 10 orders of magnitude. This replacement of the cofactor increases the commercial value of the enzyme, because NADP serves as the cofactor for many dehydrogenases used in the synthesis and biotransformations of valuable chiral biologically active compounds. This result should be regarded as significant, despite the fact that the replacement has led to a substantial increase in K_M of the enzyme with respect to formate compared with the initial enzyme. This fact confirms once again the inadequacy of X-ray diffraction data on the structure of an enzyme globule for an exact prediction of the role of amino acid residues in its functions.

X. Conclusion

In recent years, the directed evolution of enzymes has become the key technique for the creation of enzymes with new improved properties which are especially important for industrial use. This approach has provided the procedural basis for the design of enzymes with new properties without exact knowledge of the structure and the mechanism of catalytic action.

When choosing the way of solving a particular task, it should be borne in mind, however, that, as versatile as they are, different methods of directed evolution differ in results. In particular, the technique of stepwise truncation of enzyme genes with non-homologous recombination allows the modification of inner

areas of the enzyme, whereas the staggered extension technique results in the formation of hybrids with more homologies in the N-terminal fragment of the gene. The random priming recombination yields an almost uniform distribution of random replacements over the whole enzyme molecule.

The number of variants for the selection of enzymes with 'improved' properties is great with any methods of molecular evolution. Therefore, different ways for optimisation of compilation of gene libraries including computer methods are proposed. Recently, an approach of this type based on accounting of the thermodynamic and kinetic parameters of the DNA shuffling have been proposed and the predictions have been compared with the results.[63]

In some cases, it is necessary to protect an amino acid critical for the protein functions from participation in the mutation events. For this purpose, a procedure has been reported [64] based on introduction into the codon of this amino acid of an oligonucleotide fragment containing the splitting site of some restriction endonuclease. After DNA shuffling, this oligonucleotide is restricted by the restriction endonuclease and ligation is carried out. If mutation events have taken place in the codon, the complementarity of the extending ends of the restriction site is violated, no ligation occurs and these clones are excluded from the subsequent selection.

The molecular evolution techniques can be used not only to 'improve' the properties of particular enzymes but also to solve larger-scale problems, for example, to construct genetic regulatory systems.[65,66] This permits fine tuning of the biochemical parameters of a metabolic circuit done experimentally *in vivo*. By applying the directed evolution to genes that constitute the simple genetic circuit, researchers have demonstrated that a circuit that does not function due to an improper selection of components is rapidly evolved into a functional circuit. A library of genetic tools with different behaviours suitable for the construction of more complicated circuits has been composed.

The methods of directed molecular evolution can also be used successfully in the design of 'nucleozymes,' *i.e.*, catalytically active nucleic acids. A self-phosphorylated DNA has been designed and isolated from a pool of random mutant DNA sequences as a variant able to phosphorylate its own 5′-end.[67] Six selection rounds gave a nucleozyme, which efficiently phosphorylated both RNA and DNA.

The major difficulties faced by directed evolution are related to the development of an efficient strategy for the selection of variants that have enhanced properties compared with the initial material. The solution of this problem mainly rests on researcher's skill. Nevertheless, the information presented in the review demonstrates the exceptionally high potential of molecular evolution in the design of biocatalysts.

This review was prepared with financial support from the Russian Foundation for Basic Research (Project Nos 01-05-64615 and 04-05-64188).

References

1. M Ostermeier, A E Nixon, J H Shim, S J Benkovic *Proc. Natl. Acad. Sci. USA* **96** 3562 (1999)
2. M Ostermeier, J H Shim, S J Benkovic *Nat. Biotechnol.* **17** 1205 (1999)
3. S Lutz, M Ostermeier, S J Benkovic *Nucleic Acids Res.* **29** e16 (2001)
4. W P C Stemmer *Nature (London)* **370** 389 (1994)
5. W P C Stemmer *Proc. Natl. Acad. Sci. USA* **91** 10747 (1994)
6. H Zhao, F H Arnold *Nucleic Acids Res.* **25** 1307 (1997)
7. F C Christians, L Scapozza, A Crameri, G Folkers, W P C Stemmer *Nat. Biotechnol.* **17** 259 (1999)
8. J E Ness, M Welch, L Giver, M Bueno, J R Cherry, T V Borchert, W P C Stemmer, J Minshull *Nat. Biotechnol.* **17** 893 (1999)
9. L Giver, A Gershenson, P-O Freskgard, F H Arnold *Proc. Natl. Acad. Sci. USA* **95** 12809 (1998)
10. I P Petrounia, F H Arnold *Curr. Opin. Biotechnol.* **11** 325 (2000)
11. M Kikuchi, K Ohnishi, S Harayama *Gene* **236** 159 (1999)
12. M D Gibbs, K M H Nevalainen, P L Bergquist *Gene* **271** 13 (2001)
13. A A Volkov, Zh Shao, F H Arnold *Nucleic Acids Res.* **27** e18 (1999)
14. H Zhao, L Giver, Z Shao, J A Affholder, F H Arnold *Nat. Biotechnol.* **16** 258 (1998)
15. Z Shao, H Zhao, L Giver, F H Arnold *Nucleic Acids Res.* **8** 681 (1998)
16. M Lehmann, L Pasamontes, S F Lassen, M Wyss *Biochim. Biophys. Acta* **1543** 408 (2000)
17. C Vieille, G J Zeikus *Microbiol. Mol. Biol. Rev.* **65** 1 (2001)
18. P L Wintrode, K Miyazaki, F H Arnold *Biochim. Biophys. Acta* **1549** 1 (2001)
19. A Merz, M-C Yee, H Szadkovski, G Pappenberger, A Crameri, W P C Stemmer, Ch Yanofsky, K Kirschner *Biochemistry* **39** 880 (2000)
20. A Lölnn, M Gárdonyi, W van Zyl, B Hahn-Hägerdal, R C Otero *Eur. J. Biochem.* **89** 116 (2002)
21. R Sakaue, N Kajiyama *Appl. Environ. Microbiol.* **69** 139 (2003)
22. H Komeda, N Ishikawa, Y Asano *J. Mol. Catal. B: Enzymatic* **21** 283 (2003)
23. K Chen, F H Arnold *Proc. Natl. Acad. Sci. USA* **90** 5618 (1993)
24. L You, F H Arnold *Protein Eng.* **9** 77 (1996)
25. H Zhao, F H Arnold *Protein Eng.* **12** 47 (1999)
26. S-H Baik, T Ide, H Yoshida, O Kagami, S Harayama *Appl. Microbiol. Biotechnol.* **61** 329 (2003)
27. K L Maxwell, A K Mittermaier, J D Forman-Kay, A R Davidson *Protein Sci.* **8** 1908 (1999)
28. G S Waldo, B M Standish, J Berendzen, T C Terwilliger *Nat. Biotechnol.* **17** 691 (1999)
29. J K Yang, M S Park, G S Waldo, S W Suh *Proc. Natl. Acad. Sci. USA* **100** 455 (2003)
30. Y Zhang, Z Zhao, F Li, Ch Cao, Ch Liu, Q Ma *Biotechnol. Lett.* **22** 947 (2000)
31. I Matsumura, J B Wallingford, N K Surana, P D Vize, A D Ellington *Nat. Biotechnol.* **17** 696 (1999)
32. T Bulter, M Alcalde, V Sieber, P Meinhold, Ch Schlachtbauer, F H Arnold *Appl. Environ. Microbiol.* **69** 987 (2003)
33. J -H Zhang, G Dawes, W P C Stemmer *Proc. Natl. Acad. Sci. USA* **94** 4504 (1997)
34. D Sriprapundh, C Vieille, J G Zeikus *Protein Eng.* **16** 683 (2003)
35. T Yano, Sh Oue, H Kagamiyama *Proc. Natl. Acad. Sci. USA* **95** 5511 (1998)
36. L G Otten, Ch F Sio, J Vrielink, R H Cool, W J Quax *J. Biol. Chem.*, **277** 42121 (2002)
37. Ch F Sio, A M Riemens, J-M van der Laan, R M D Verhaert, W J Quax *Eur. J. Biochem.* **269** 4495 (2002)
38. Ch-W Wang, J C Liao *J. Biol. Chem.* **276** 41161 (2001)
39. D Umeno, F H Arnold *Appl. Environ. Microbiol.* **69** 3163 (2003)
40. D Umeno, A V Tobias, F H Arnold *J. Bacteriol.* **184** 6690 (2002)
41. H Suenaga, M Goto, K Furukawa *J. Biol. Chem.* **276** 22500 (2001)
42. A Meyer, M Würsten, A Schmid, H-P E Kohler, B Witholt *J. Biol. Chem.* **277** 34161 (2002)
43. T Bosma, J Dambrovský, G Stucki, D B Janssen *Appl. Environ. Microbiol.* **68** 3582 (2002)
44. K-E Jaeger, T Eggert, A Eipper, M T Reetz *Appl. Microbiol. Biotechnol.* **55** 519 (2001)
45. M T Reetz, A Zonta, K Scimossek, K Liebeton, K-E Jaeger *Angew. Chem., Int. Ed. Engl.* **36** 2830 (1997)
46. O May, P T Nguen, F H Arnold *Nat. Biotechnol.* **18** 317 (2000)
47. M Dion, A Nisole, P Spangenberg, C André, A Glottin-Fleury, R Matte, Ch Tellier, C Rabiller *Glycoconjugate J.* **18** 215 (2001)
48. U T Bornscheuer, J Altenbuchner, H H Meyer *Bioorg. Med. Chem.* **7** 2169 (1999)
49. J H Kim, G-S Choi, S-B Kim, W-H Kim, J-Y Lee, Y-W Ryu, G-J Kim *J. Mol. Catal. B: Enzymatic* **27** 169 (2004)
50. G J Williams, S Domann, A Nelson *Proc. Natl. Acad. Sci. USA* **100** 3145 (2003)
51. G Xia, L Chen, M Fa, P G Schultz, F E Romesberg *Proc. Natl. Acad. Sci. USA* **99** 6597 (2002)

52. L C Tisi, P J White, D J Squirelli, M J Murphy, C R Lowe, J A H Murray *Anal. Chim. Acta* **416** 115 (2002)
53. L Jermutus, M Tessier, L Pasamontes, A P G M van Loon, M Lehmann *J. Biotechnol.* **85** 15 (2001)
54. L Serrano, A G Day, A R Fersht *J. Mol. Biol.* **233** 305 (1993)
55. R Gupta, Q K Beg, P Lorenz *Appl. Microbiol. Biotechnol.* **59** 15 (2002)
56. S Sacchi, S Lorenzi, G Molla, M S Pilone, C Rossetti, L Pollegioni *J. Biol. Chem.* **277** 27510 (2002)
57. O Lentz, Q-Sh Li, U Schwaneberg, S Lutz-Wahl, P Fischer, R D Schmid *J. Mol. Catal. B: Enzymatic* **15** 123 (2001)
58. Q-Sh Li, J Ogawa, R D Schmid, S Shimizu *Appl. Environ. Microbiol.* **67** 1635 (2001)
59. M Lehmann, C Loch, A Middendorf, D Studer, S F Lassen, L Pasamontes, A P G M van Loon, M Wyss *Protein Eng.* **15** 403 (2002)
60. A Tomschy, R Brugger, M Lehmann, A Svendsen, K Vogel, D Kostrewa, S F Lassen, D Burger, A Kronenberger, A P G M van Loon, L Pasamontes, M Wyss *Appl. Environ. Microbiol.* **68** 1907 (2002)
61. A M Rojkova, A G Galkin, L B Kulakova, A E Serov, P A Savitsky, V V Fedorchuk, V I Tishkov *FEBS Lett.* **445** 183 (1999)
62. A E Serov, A S Popova, V V Fedorchuk, V I Tishkov *Biochem J.* **367** 841 (2002)
63. N Maheshri, D Schaffer *Proc. Natl. Acad. Sci. USA* **100** 3071 (2003)
64. D Umeno, K Hiraga, F H Arnold *Nucleic Acids Res.* **31** e91 (2003)
65. Y Yokobayashi, R Weiss, F H Arnold *Proc. Natl. Acad. Sci. USA* **99** 16587 (2002)
66. Y Yokobayashi, C H Collins, J R Leadbetter, R Weiss, F H Arnold *Adv. Complex Syst.* **6** 37 (2003)
67. Y Li, R R Breaker *Proc. Natl. Acad. Sci. USA* **96** 2746 (1999)