ELSEVIER

Biophysical Chemistry 66 (1997) 203–210

Biophysical Chemistry

The Role of Peptide Modules in Protein Evolution

Merle A. Fuchs, Christiane Buta

Max-Planck-Institute for Biophysical Chemistry, D-37077 Göttingen, Germany

Received 24 April 1997; accepted 24 April 1997

Abstract

Protein evolution shows interesting strategies to be used in protein design. During evolution the creation of new proteins has been accomplished by combining different peptide modules, i.e. evolutionary successful stable folding units. Thereby, the evolution of proteins has been greatly enhanced. Today this mechanism of recombining optimized building blocks to design new proteins has been introduced into applied molecular evolution.

Keywords: protein evolution; protein design; applied molecular evolution

1. Mutation, selection and amplification

Since Charles Darwin published his major work *On the Origin of Species by Means of Natural Selection* in 1859 [1], the ideas of evolution have fundamentally influenced the intellectual life of Western civilization. Whereas Darwin's thoughts were very controversial in his days, they are now well accepted in the scientific world. This broad acceptance is a result of the continuously growing knowledge in genetics, physiology and ecology.

Darwinism explains the development of different species by variation, survival of the fittest and their reproduction, or in other words, by **mutation, selection** and **amplification**. Reproduction (amplification) of the genotype and the phenotype is essential for the survival of a species consisting of mortal individuals. But this would not be sufficient for survival under varying conditions without the error-prone reproduction of the genotype which introduces mutations. These variations endow the phenotype with the flexibility to adapt to a changing environment because selection among the mutants will choose suitable individuals that fit the actual demands.

The breathtaking diversity and complexity of present and extinct life forms raises a very interesting question. How could this diversity evolve in the short time since the beginning of life on earth about 3.8 billion years ago? To illustrate the enormous number involved, *Manfred Eigen* has given an impressive example. For a short sequence of 100 nucleotides, there are $4^{100} \sim 10^{60}$ different possibilities for combining the four nucleotides. If the

PII S0301-4622(97) 00067-7

entire earth were covered with a solution of these oligonucleotides at a concentration of one gram per liter and with a depth of one centimeter, and if all these molecules had been in constant turnover with a lifetime of one second per molecule, then only 10^{50} sequences could have existed during a period of one billion years [2].

The interplay of mutation and selection on single individual species alone cannot account for the evolution of today's highly diverse life forms. *Richard W. Hamming* introduced the concept of **sequence space** into information science [3], and *Ingo Rechenberg* suggested applying this idea to evolutionary problems [4]. The abstract sequence space of a polymer represents all possible monomer sequences (e.g., of an oligonucleotide) of a defined length, and it describes the connections between the elements of the space. Neighboring points in that space belong to sequences that differ in only one position.

If the sequences are plotted against their fitness parameter for carrying out a special task, a multi-dimensional **fitness landscape** is obtained. It is often pictured as a three dimensional mountainous region, where the peaks describe sequences with high fitness values for some defined quality (see Fig. 1). *Manfred Eigen* and *Peter Schuster* formulated the concept of **quasispecies** [2,5] and introduced it into the science of evolution; from here it was taken over by protein science. During the replication of molecules and organisms some errors occur, for that reason populations usually consist of a distribution of more or less closely related individuals. By using this idea which is described in detail by *Susanne Brakmann* in this issue, it became possible to formulate a quantitative model of molecular self-organization processes. This new concept of evolutionary selection no longer considers an individual wild-type species to be the exclusive object for selective processes; instead the whole population within the quasispecies is affected by the mutation and the selection processes. By using the models of sequence space and quasispecies, it was possible for the first time to describe the principles of molecular self-organization. However, to understand protein evolution in more detail it is important to incorporate the growing knowledge about protein structure and function.

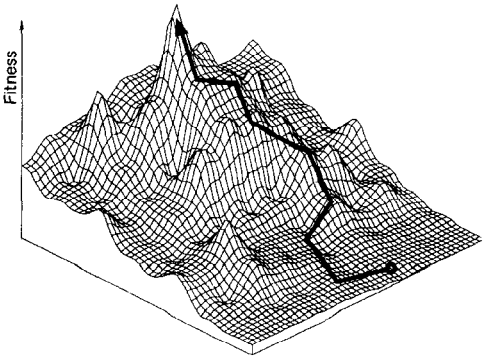

Fig. 1. Illustration of a fitness landscape. Each point represents a certain sequence. During evolution selection pressure leads towards the peaks, i.e. the sequences with the highest fitness (from [6]).

2. Shape and function of proteins

Under physiological conditions the tertiary structure of proteins is determined by their amino acid sequences. Although a remarkable number of protein structures have been solved, it is still impossible to predict the tertiary structure of molecules more complicated than peptides. Interestingly, some proteins show quite similar structures with other proteins even though they exhibit less than 10 % amino acid identity [7]. In T4 lysozyme, less than half of the amino acids seem to be responsible for determining the protein fold. Single amino acid changes may cause disastrous effects at certain positions by preventing the formation of important conformational intermediates. At other positions, extensive changes in the sequence may be tolerated, depending on whether the varied region makes an active or a passive contribution to the folding process. Up to ten contiguous alanine residues can be inserted at certain positions in T4 lysozyme without losing the overall folding pattern [8].

Surveying these results brings one to the conclusion, that the number of seperate elements in **shape space** (i.e. the spatial projection of all possible structures of a given sequence length) must be much smaller than in sequence space. *Peter Schuster et al.* have examined the shape space of

RNA secondary structures [9,10]. The shape space for ribonucleotides with n bases contains 1.849^n positions (indicating distinct secondary structures), far less than the corresponding sequence space which has 4^n elements. Selection experiments with RNA libraries agree with these calculations by showing that different sequences can form similar structures [11].

Besides the distribution of charges and the location of single amino acids that are important for catalysis, the structure of a protein molecule is a primary prerequisite for enzyme function. Consequently, also the **function space** will be much smaller than the sequence space. *Stephen M. Edgington* estimated that only 10^8 enzymes would be sufficient to catalyze all enzymatic reactions [12].

These relations between the sequence space and the function space suggest new ways to design proteins by utilizing these evolution concepts. The entire sequence space does not have to be screened in order to select new peptides with a desired function. A limited number of peptides should be sufficient provided that they represent enough different shapes. The amount of peptides will be large but it will be much smaller than the number of points in sequence space.

Can we create these shape libraries (in contrast to sequence libraries)? In addition to the traditional view of evolution (being the mutation of single base-pairs), nature invented an alternative mechanism to speed up evolution. A new dimension is gained by the recombination of building blocks that had already proved that they were successful.

Many proteins can be constructed from combinations of modular domains, each of which imparts a particular function to the multimeric protein (e.g. [13]). With the help of this mechanism, structures that were developed for special functions can be reused to build new molecules. Thereby, the number of different combinations that have to be tested by evolution in order to create new functional proteins can be reduced by orders of magnitude.

3. The NAD-binding fold as an example for domain-shuffling

Until the early 1970s biologists took the point of view that proteins arose mainly by gene duplication and modification [14]. By analyzing the structure of dogfish *lactate dehydrogenase* (LDH), *Michael G. Rossmann et al.* found significant similarities between nicotinamid adenine dinucleotide (NAD)-binding domains of different dehydrogenases [15,16]. The three-dimensional folds of LDH, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and liver alcohol dehydrogenase (LADH) are similar and well conserved, although no significant sequence homology is detectable, even when looking at only short stretches. By sequence alignment criterion alone they would appear to be unrelated enzymes [17].

NAD-dependent dehydrogenases consist of two structurally and functionally distinct domains. One part of the enzyme binds the coenzyme NAD, the other part determines the specificity by acting as a substrate-binding site and as a catalyst.

The basic elements of secondary structure in the coenzyme-binding domains - six parallel strands of pleated-sheet and four helices - form an open, parallel β-sheet with helices on both sides [17]. Besides the identical topology of the NAD-binding domains in LDH, GAPDH, and LADH, most of the main chain atoms are superimposed within 2 Å in their three-dimensional structures (see Fig. 2).

A striking correlation between the different NAD-binding domains is the sequential order of these elements, which is identical and symmetrical in all

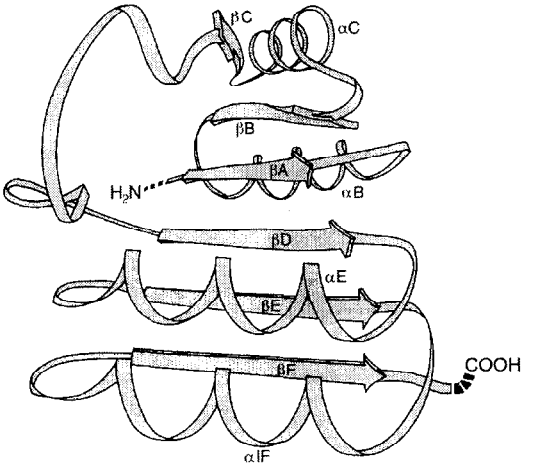

Fig. 2. Schematic representation of the structure of the coenzyme-binding domain in dehydrogenases (modified from [18]).

three dehydrogenases [18]. Each half is built up by a pair of β-α-β motifs, which are called mononucleotide-binding motifs or Rossmann folds. These motifs bind mononucleotides - each nucleotide in the NAD dinucleotide is bound to one of these coenzyme-binding domains.

The NAD-binding domains of the three dehydrogenases discussed above are located at different positions within the primary sequences of these proteins. In GAPDH this motif is located directly at the N-terminus of the polypeptide chain; in LDH it is moved by 22 amino acids downstream. Finally, in LADH the coenzyme-binding domain is positioned near the C-terminal end, giving the impression that the domain is shuffled along the polypeptide chain (see Fig. 3).

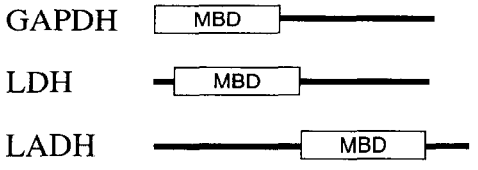

Fig. 3. Location of the mononucleotide-binding domain (MBD) in three different dehydrogenases (modified from [19]).

Interestingly, the Rossmann fold does not only occur in dehydrogenases, but also in many other proteins with nucleotide-binding activities, such as kinases or aminoacyl-tRNA-synthetases [20], both of which bind ATP, and flavodoxin which binds flavine mononucleotide [21]. Therefore, *Mitiko Go* suggested that the NAD-binding domain originated prior to the emergence of the aminoacyl-tRNA synthetases. He considered these coenzyme-binding domains to be one of the oldest building blocks used in primitive proteins [22].

Therefore, **recombination** must be added to the familiar trio mutation, amplification and selection as one of the main mechanisms of evolutive processes. Today many modular proteins are known to have evolved by recombinatory mechanisms. Among them are many cell-surface proteins that mediate specific protein-protein interactions in multicellular organisms,

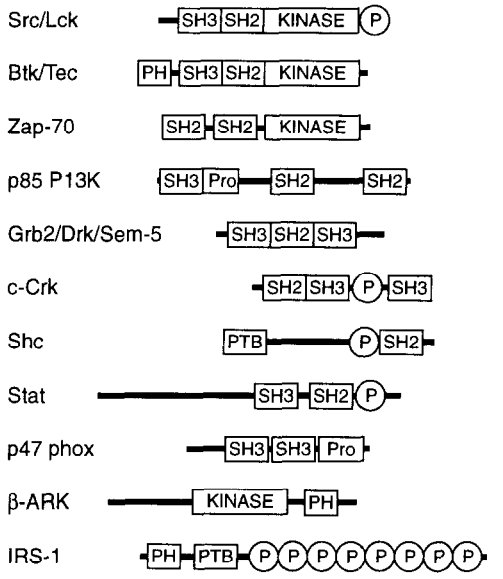

Fig. 4. Organization of different proteins built from a few common domains. SH2, Src-homology-2 domain; SH3, Src-homology-3 domain; P, phosphothyrosine-containing SH2-binding site; PH, pleckstrin homology domain; Pro, proline-rich SH3-binding site; PTB, phosphothyrosine-binding domain (modified from [23]).

as well as intracellular signaling proteins (see Fig. 4; [24]). Of course, attempts have been made to explain how these mechanisms function, that made the spread of mobile elements possible.

4. Exons and introns

As early as 1974, *Ingrid Ohlsson et al.* postulated that dehydrogenases and other complex enzymes emerged by joining together and coding for simple precursor peptides [18]. With the discovery of exon-intron arrangements, *Walter Gilbert* has proposed that the eukaryotic arrangements of genes in exons and introns might enhance recombination, which leads to the evolution of new genes [25]. *Colin C.F. Blake* assumed that exons could be correlated with functional units in proteins [26]. Since then exon-shuffling has been discussed as one of the essential processes in evolution.

Some exon transcripts correspond to folded segments of supersecondary structures [27]. These

results suggested to many biologists that ancient proteins were encoded by minigenes which were spliced together. The so-called 'introns early' theory claims that structures were first encoded as exons, which were then separated from each other by introns. Therefore, recombination between intron regions resulted in new combinations of exons, and thereby new genes and protein structures were created [25,28].

In contrast to this theory, the 'introns late' theory postulates that the ancient genes must have been contiguous once. According to this model introns did not occur before the major endosymbiotic events that led to the appearance of mitochondria in eukaryots [29], possibly because of transposons that interrupted the DNA [30,31].

Both theories still are vigorously discussed (e.g. [32-34] and the arguments for both models will not be repeated here. However, some proteins fit the concept of exon shuffling well, while others do not. On the other hand, most exons do not correspond to mobile domains. Unfortunately, the question of how mobile domains became recombined cannot yet be answered. *Russel F. Doolittle* believed that some proteins may have been assembled by exon shuffling, but the subsequent loss and gain of introns makes it difficult to resolve the relationships now. Other proteins might have been assembled in a modular fashion without the participation of introns [29], although the mechanism for this case is unknown. Certainly for a biologist it is unsatisfactory that the mechansim of domain shuffling is unsolved. But for a biotechnologist the mere knowledge that mobile building blocks exist at least provides new perspectives for the field of evolutive biotechnology; this will be discussed in the last three sections of this article.

5. Domains, foldons and modules

Polypeptides with a minimum of 30 to 40 amino acids in length have sufficient intramolecular interactions to become stabilized, thus forming a favored tertiary structure [14]. However, shorter sequences may also assume stable structures, often stabilized by disulfide bridges, or by complexation by metal ions. Short repetitive units can gain even higher stabilities by packing tightly against each other [29]. For this reason, a peptide of minimum size will always fold into the same tertiary structure under defined conditions [14].

In the scientific literature the designation of these stable folding units is not consistent: "foldon", "folding unit", "domain" and "module" are sometimes used synonymously. In this paper we use the term **domain** to designate a structurally and functionally idependent unit; it does not need to consist of a continuous amino acid sequence. A domain can be composed of several different modules and can fold into a compact globular structure.

The typical size of a domain is 100-200 amino acids, although much larger domains are known. They are the structural and functional building blocks of most globular proteins. In multidomain proteins, they are often spatially well separated from each other. For that reason, the individual domains of a protein can sometimes be identified, and isolated, by limited proteolytic cleavage. Usually each domain exhibits a specific catalytic or binding function. This is why a single domain can be combined with functions from other domains, providing the opportunity for carrying out complex tasks [35].

A given type of domain can sometimes be recognized either in the same or in a different protein. These similar folds are clearly identifiable by their tertiary structure, even if there is no obvious sequence homology at the level of the primary structure.

Folding units and **foldons** are used synonymously. According to *Anna R. Panchenko et al.* foldons are subunits of a protein which can fold in a single, cooperative step [36]. These quasi-independent folding units can be characterized by *Mitiko Go's* [37] geometrical definitions. *Go* assumed that these folding units correlate with exons. But now many examples are known where exons do not correlate with folding units.

Normally, the term **module** designates the smallest stable folding unit - we will use this meaning in the text below. In different proteins a module can exhibit different functions. Modules are always formed by a contiguous primary sequence and normally contain between 40 and 100 amino acids [38]. They fold into a defined arrangement of helices, β-folds and loops [17], and the three-dimensional structure of a module often remains even after it has

been isolated [38]. Often a module's tertiary structure is better conserved than its amino acid sequence [14] because selection aims at the phenotype - not at the genotype. In some articles, "module" is also used synonymously with a peptide which is coded by one single exon.

6. Monomeric and modularly constructed DNA libraries

Bearing in mind the great potential of combining well-tried building blocks in order to evolve proteins with new functions, it is astounding that this methodology, i.e. using natural evolution as a model, did not become a popular field of research in applied molecular evolution.

As discussed in more detail by *Andre Koltermann* and *Ulrich Kettling* in this issue, the creation of complex DNA libraries that code for complex peptide libraries is a prerequisite for many selection experiments. In that context, "complexity" is usually used within the meaning of "many different sequences", for DNA [39,40], for RNA [41], or for (synthetic) peptide libraries [42], and even for banks containing organic molecules used in combinatorial chemistry [43,44].

For example, the most popular currently used methods for creating large oligonucleotide libraries are based on error-prone PCR (e.g. [45,46]), repeated oligonucleotide-directed mutagenesis (e.g. [47]) and the utilization of random oligonucleotide cassettes (e.g. [48,49]). These libraries have one major disadvantage: they are not optimized for searching in shape space [50], because they use monomers instead of preselected building blocks. Most sequences cannot fold into stable structures, and this is the basic requirement for biological function. This statement is not only true for peptides and proteins, but also for other functional molecules, such as ribozymes.

For instance, in the case of random oligonucleotide cassettes only short sequences can be permuted. In order to create all possible sequences of a 30mer nucletide, 4^{30}~10^{18} different molecules are necessary. This is about 2 micromoles of oligonucleotide (which corresponds to the tenfold amount of a typical chemical DNA synthesis). Therefore, only small sequences can be changed by random oligonucleotide cassettes.

Nevertheless, it has been shown that this technique has the potential to mutate a protein's binding site or the catalytic center [51]. Short peptides can be randomized; however, whole proteins cannot be designed by means of applied evolution technology with this method. In order to make large proteins - and not just vary details of well-known molecules- novel strategies must be applied.

Willem P.C. Stemmer has given a nice description of the restrictions for mutation strategies that are based on changing single monomers: "The limitations of point mutagenesis strategies become obvious if one imagines editing a manuscript by changing individual letters rather than by moving blocks of letters, words, and sentences around" [52]. In the context of this picture, peptides, modules and domains are the words and the sentences, respectively, for the construction of new proteins.

A peptide library representing 150mers can show 10^{195} different sequences. Most of these molecules cannot form a compact tertiary structure at all. Entropy considerations show that random sequences which mould a stable fold become less probable as the sequences become longer. Therefore, modular constructions made by a complex but limited set of building blocks, that have been preselected for assembling new proteins, should essentially enhance the evolution of larger structures in sequence space [35].

In addition, "monomer libraries" are also redundant when considering the shapes of the molecules. A limited set of sequences that covers only a restricted region of the sequence space, might not contain a sequence corresponding to a desired function. Even worse, the overwhelming majority of these sequences do not even have any biological function. Consequently, these libraries must be very large, and the selection system must be very sophisticated to enable finding "the needle in a haystack".

In contrast to monomer libraries, the results obtained when using modular libraries are very satisfying. A clever approach is the use of antibody libraries [53,54]. This technique takes advantage of the high potential for antibodies to bind to almost all possible shapes of small molecules. Besides for the selection of binding-peptides, these libraries can even be used for the "evolutive design" of new enzymes.

The goal of using antibody libraries is the selection of molecules that can bind to chemical substances, that show structures similar to the transition state of the reaction to be catalyzed. By stabilizing the transition state, these antibodies can reduce the activation energy of the reaction, thereby catalyzing the conversion.

Combining different heavy- and light-chain variable regions [53] always results only in antigen binding fragments of antibodies. No other class of proteins can be produced by this method. Therefore for constructing new proteins different approaches have to be taken.

A more general way to make modular libraries is a technique called "DNA shuffling" [50,55]. This technique is described in more detail by *Susanne Brakmann* in this issue. In this method genes are reassembled randomly after they have been digested by DNase I. This has been used to create a library of chimeric proteins made from human and murine genes for interleukin 1b [55] and to produce an improved green fluorescent protein [56].

The possibility of experimentally shuffling exons has already been discussed [52]. This technique could create a way for combining all different types of modules for the cases where exons really code for folding units; as mentioned above, this is true only for a minor fraction of exons.

Another strategy for constructing modular libraries is being developed in our laboratory (*Merle A. Fuchs*, in preparation). This method, based on site-directed blunt-end ligation, allows all chemically synthesized desoxyribonucleotides to be combined without any limitations regarding their sequences. With the help of this method, it should be possible to obtain more information concerning the potential and the limitations of using module libraries for developing new proteins.

7. The potential of modular libraries

By extrapolating from known protein sequences and three-dimensional structures, it has been estimated that there are probably only approximately 1000 naturally occurring folding patterns, i.e. combinations of secondary and tertiary structures [8]. The synthesis of 1000 different pairs of oligonucleotides of 100 base-pairs that code for DNA modules is not a problem for present-day oligonucleotide syntheses. The cost of the instruments and the chemicals for synthesizing these DNA modules are reasonable and financing should not be a problem.

Constructs of three to five monomers can be made by ligating several of these modules; this results in approximately 10^{15} different sequences, which is still feasible experimentally. *Wlodeck Mandecki* claims that 10^{16} different clones are the maximum number that can be handled in the laboratory [57]. If this complex library (regarding shape and function of its molecules), is expressed in a suitable vector, it should be possible to carry out many different selection experiments.

8. Conclusive remarks

Sydney Brenner and *Richard A. Lerner* used the term "irrational design" to depict evolution-based statagies for finding functional molecules by searching huge libraries [58]. In contrast to "rational design", these methods invert the traditional sequence of steps used for protein design. First, they select a molecule with a desired function, and then in a second step, they solve the sequence and sometimes even the structure.

Nevertheless, even "irrational design" requires well-founded strategies for choosing the selection systems, the selection experiments and the method for constructing highly complex molecule libraries. Preselected modular libraries may help reduce the disadvantages inherent in randomization strategies, namely the production of large numbers of useless molecules, without losing the advantages of handling a gigantic number of shapes and functions, that cannot be dealt with by the human mind. Before optimal methods for producing new molecules are developed, much work remains to be done.

References

[1] C.D. Darwin, On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. (J. Murray, London, 1859).

[2] M. Eigen, Stufen zum Leben (R. Piper-Verlag, München, 1987).

[3] R.W. Hamming, Bell Syst. Techn. J., 24 (1950) 147-160.

[4] I. Rechenberg, Evolutionsstrategie (Problemata Frommann-Holzboog, Stuttgart - Bad Cannstatt, 1973).

[5] M. Eigen, J.S. McCaskill and P. Schuster, Adv. Chem. Phys., 75 (1987) 149-263.

[6] S. Brakmann, U. Kettling and F. Oehlenschläger, Biologie in unserer Zeit, 25 (1995) 355-366

[7] A.M. Lesk and C. Chothia, J. Mol. Biol., 160 (1982) 325.

[8] M.J. Geisow, TIBTECH, 11 (1993) 111-114.

[9] P. Schuster, Orig. Life Evol. Biosys., 23 (1993) 373-391.

[10] P. Schuster, W. Fontana, P.F. Stadler and I.L. Hofacker, Proc. R. Soc. Lond. B., 255 (1994) 279-284.

[11] A.D. Ellington and J.W. Szostak, Nature, 346 (1990) 818-822.

[12] S.M. Edgington, Bio/Technology, 11 (1993) 285-289.

[13] G. Shaw, BioEssays, 18 (1996) 35-46.

[14] R.F. Doolittle and P. Bork, Sci. American, 269 (1993) 44-40.

[15] M.G. Rossmann, D. Moras and K.W. Olsen, Nature, 250 (1974) 194-199.

[16] M.G. Rossmann, A. Liljas, C.I. Brandén and L.J. Banaszak, Enzymes, 11A (1975) 61-102.

[17] C.I. Brandén, H. Eklund, C. Cambillau and A.J. Pryor, Experientia Supplementum, 36 (1980) 40-84.

[18] I. Ohlsson, B. Nordström and C.I. Brandén, J. Mol. Biol., 89 (1974) 339-354.

[19] C.I. Brandén and J. Tooze, Introduction to protein structure (Garland Publishing Inc., New York, 1991).

[20] S. Clermont, C. Corbier, Y. Mely, D. Gerard, A. Wonacott and G. Branlant, Biochemistry, 32 (1993) 10178-10184.

[21] C.I. Brandén, H. Eklund, C. Cambillau and A.J. Pryor, EMBO Journal, 3 (1984) 1307-1310.

[22] M. Go and M. Mizutani, in: The origin and evolution of the cell, eds. H. Hartmann and K. Matsuno (World Scientific, Singapore, 1992) 359-374

[23] T. Pawson, Nature, 373 (1995) 573-580.

[24] I.D. Campbell and A.K. Dowing, TIBTECH, 12 (1994) 168-172.

[25] W. Gilbert, Nature, 271 (1978) 501-501.

[26] C.C.F. Blake, Nature, 273 (1978) 267-268.

[27] C. Blake, Nature, 306 (1983) 535-537.

[28] W. Gilbert and M. Glynais, Gene, 135 (1993) 137-144.

[29] R.F. Doolittle, Annu. Rev. Biochem., 64 (1995) 287-314.

[30] T. Cavalier-Smith, Nature, 315 (1985) 283-284.

[31] J.D. Palmer and J.M. Logson Jr., Curr. Opin. Gen. Devel., 1 (1991) 470-477.

[32] J.M. Logsdon and J.D. Palmer, Nature, 369 (1994) 526.

[33] G.N. Kumar, Curr. Sci., 66 (1994) 336-339.

[34] L. Patthy, Curr. Opin. Struct. Biol., 1 (1991) 351-361.

[35] D.G. Hardie and J.R. Coggins, Multidomain proteins - structure and evolution. (Elsevier, Amsterdam, 1986) p. 2.

[36] A.R. Panchenko, Z. Luthey-Schulten and P.G. Wolynes, Proc. Natl. Acad. Sci. USA, 93 (1996) 2008-2013.

[37] M. Go, Proc. Natl. Acad. Sci. USA, 80 (1983) 1964-1968.

[38] M. Baron, D.G. Norman and I.D. Campbell, TIBS, 16 (1991) 13-17.

[39] J.J. Devlin, L.C. Panganiban and P.E. Devlin, Science, 249 (1990) 404-406.

[40] J.K. Scott, TIBS, 17 (1992) 241-245.

[41] D. Irvine, C. Tuerk and L. Gold, J. Mol. Biol., 222 (1991) 739-761.

[42] R.A. Houghten, J.R. Appel, S.E. Blondell, J.H. Cuervo, C.T. Dooley and C. Pinilla, BioTechniques, 13 (1992) 412-421.

[43] L.H. Caporale, Proc. Natl. Acad. Sci USA, 92 (1995) 75-82.

[44] M.H.J. Ohlmeyer, R.N. Swanson, L.W. Dillard, J.C. Reader, G. Asouline, R. Kobayashi, M. Wigler and W.C. Still, Proc. Natl. Acad. Sci. USA, 90 (1993) 10922-10926.

[45] R.C. Cadwell and G. Joyce, PCR Meth. Appl., 1 (1992) 1-6.

[46] R.C. Cadwell and G. Joyce, PCR Meth. Appl., 3 (1994) 136-140.

[47] S. Delagrave, E.R. Goldmann and D.C. Youvan, Prot. Eng., 6 (1993) 327-331.

[48] B. Borrego, A. Wienecke and A. Schwienhorst, Nucl. Acids Res. USA, 23 (1995) 1834-1835.

[49] S. Delagrave and D.C. Youvan, Bio/Technology, 11 (1993) 1548-1552.

[50] W.P. Stemmer, Nature, 370 (1994) 389-391.

[51] A. Sättler, S. Kanka, K.H. Maurer and D. Riesner, Electrophoresis, 17 (1996) 784-92.

[52] W.P. Stemmer, Bio/Technology, 13 (1995) 549-553.

[53] C.F. Barbas, W. Amberg, A. Simoncsits, T.M. Jones and R.A. Lerner, Gene, 137 (1993) 57-62.

[54] R.A. Lerner, S.J. Benkovic and P.G. Schultz, Science, 252 (1991) 659-667.

[55] W.P. Stemmer, Proc. Natl. Acad. Sci. USA, 91 (1994) 10747-10751.

[56] A. Crameri, E.A. Whitehorn, E. Tate and W.P.C. Stemmer, Nature Biotech., 14 (1996) 315-319.

[57] W. Mandecki, Prot. Eng., 3 (1990) 221-226.

[58] S. Brenner and R.A. Lerner, Proc. Natl. Acad. Sci. USA, 89 (1992) 5381-5383.