Acta Biochimica Polonica

Vol. 45 No. 3/1998

705–720

QUARTERLY

Review

Phage display of proteins[o]

Katarzyna Kościelska, Liliana Kiczak, Monika Kasztura, Olga Wesołowska and Jacek Otlewski[✉]

Institute of Biochemistry and Molecular Biology, University of Wrocław, Wrocław, Poland

Received: 24 February, 1998

Key words: phage display, proteinase inhibitors, antibody, protein minimization, *in vitro* selection

In recent years the phage display approach has become an increasingly popular method in protein research. This method enables the presentation of large peptide and protein libraries on the surface of phage particles from which molecules of desired functional property(ies) can be rapidly selected. The great advantage of this method is a direct linkage between an observed phenotype and encapsulated genotype, which allows fast determination of selected sequences. The phage display approach is a powerful tool in generating highly potent biomolecules, including: search for specific antibodies, determining enzyme specificity, exploring protein–protein and protein–DNA interactions, minimizing proteins, introducing new functions into different protein scaffolds, and searching sequence space of protein folding. In this article many examples are given to illustrate that this technique can be used in different fields of protein science. The phage display has a potential of the natural evolution and its possibilities are far beyond rational prediction. Assuming that we can design the selection agents and conditions we should be able to engineer any desired protein function or feature.

Application of molecular biology techniques to protein science provided two, basically different, approaches. The older method, the site directed mutagenesis, enables introduction of

---

[o]Supported by grant No. 6 P04A 035 13 from the Polish State Committee for Scientific Research.

[✉]Author to whom all correspondence should be addressed: Jacek Otlewski, Institute of Biochemistry and Molecular Biology, University of Wrocław, Tamka 2, 50-137 Wrocław, Poland; tel: (48 71) 402 824; fax: (48 71) 402 608; e-mail: otlewski@bf.uni.wroc.pl

Abbreviations: ANP, atrial natriuretic peptide; APPI, Alzheimer amyloid β-protein precursor; BPTI, basic pancreatic trypsin inhibitor (Kunitz); EPO, erythropoietin; EPOR, soluble part of the erythropoietin receptor; hGH, human growth hormone; hGHbp, extracellular domain of human growth hormone receptor; IgG, immunoglobulin; Fab, antigen binding fragment of IgG; Fc, constant fragment of IgG; scFv, single chain Fv antibody fragment; TF-VIIa, tissue factor–factor VIIa complex; K_a, association constant.

single or multiple amino-acid substitution(s) into a protein sequence [1-2]. This method is based on the assumption that, at the present level of protein understanding, we can rationally predict the influence of a particular mutation on protein function. It is also usually assumed that effects of single mutations are additive, i.e. a functional feature of a protein (e.g. enzyme activity, protein stability) can be enforced after stepwise introduction of several single mutations [3].

The alternative methods are based on the combinatorial principle [4]. Several (typically 3-6) codons are randomized simultaneously up to 64 possible triplets, yielding a library of protein variants. The number of amino-acid sequences represented in the library depends on the number of randomized codons (amino-acid positions) and the extent of randomization for a given codon.

The most popular version of the combinatorial approach, called the phage display, is presentation of a protein library on the surface of bacteriophage (usually M13 phage) particles [5-6]. The protein is displayed through fusion of its gene to the coat protein gene, what leads to presentation of a natively folded protein variant on a single phage particle. The phage-linked proteins of desired functional property(ies) can be rapidly selected if an appropriate and sufficient selection procedure can be developed [7]. Most typically, selection is achieved through affinity chromatography and should be repeated several times to enrich the pool of phages in genuine high-affinity protein variants presented. A great advantage of this method is a direct linkage between an observed phenotype and encapsulated genotype, which allows fast determination of selected variants. The selection step is the most crucial in the whole procedure, according to the rule: "you get what you select for" [7].

Compared to site directed mutagenesis, combinatorial protein design seems quite an irrational approach, although an intelligent strategy both to generate a library and to search for a desired property of the protein requires substantial invention. Moreover, in view of the enormous size of the protein sequence space [8], iterative strategies of successive libraries, which progressively accumulate positive mutations had to be developed [9-11]. Fast progress can also be observed in development of new selection procedures that more closely reflect those of the natural evolution [12-14].

In the last years application of phage display resulted in spectacular achievements, that have been reviewed elsewhere [15-26], and include: development of new protein drugs, minimization of proteins, engineering of novel binding sites, and many others. In this article we will review some of these most recent accomplishments which provided new proteins with selected functional features. Virtually in all cases a rational design of these macromolecules would be very unlikely.

THE IDEA OF PHAGE DISPLAY

The phage display of peptides was presented for the first time by George Smith in 1985 [5] and some years later the same laboratory created the first library on the surface of a filamentous phage [6]. Since that time a lot of various peptides and proteins have been expressed on the phage capsid.

The idea of phage presentation is based on the engineering of phage particles, which in a place of some coat protein molecules contain a fusion composed of the missing protein (or its part) and a desired peptide/protein linked together through a short spacer. Inside a particle, there is DNA coding for a protein expressed on the phage surface. This provides a simple relation between the sequence of the protein variant expressed on the particle and the enclosed DNA sequence that is far more easier to analyse.

Most usually, the phage chosen for a protein/peptide presentation is the M13 filamentous bacteriophage. It offers a few different

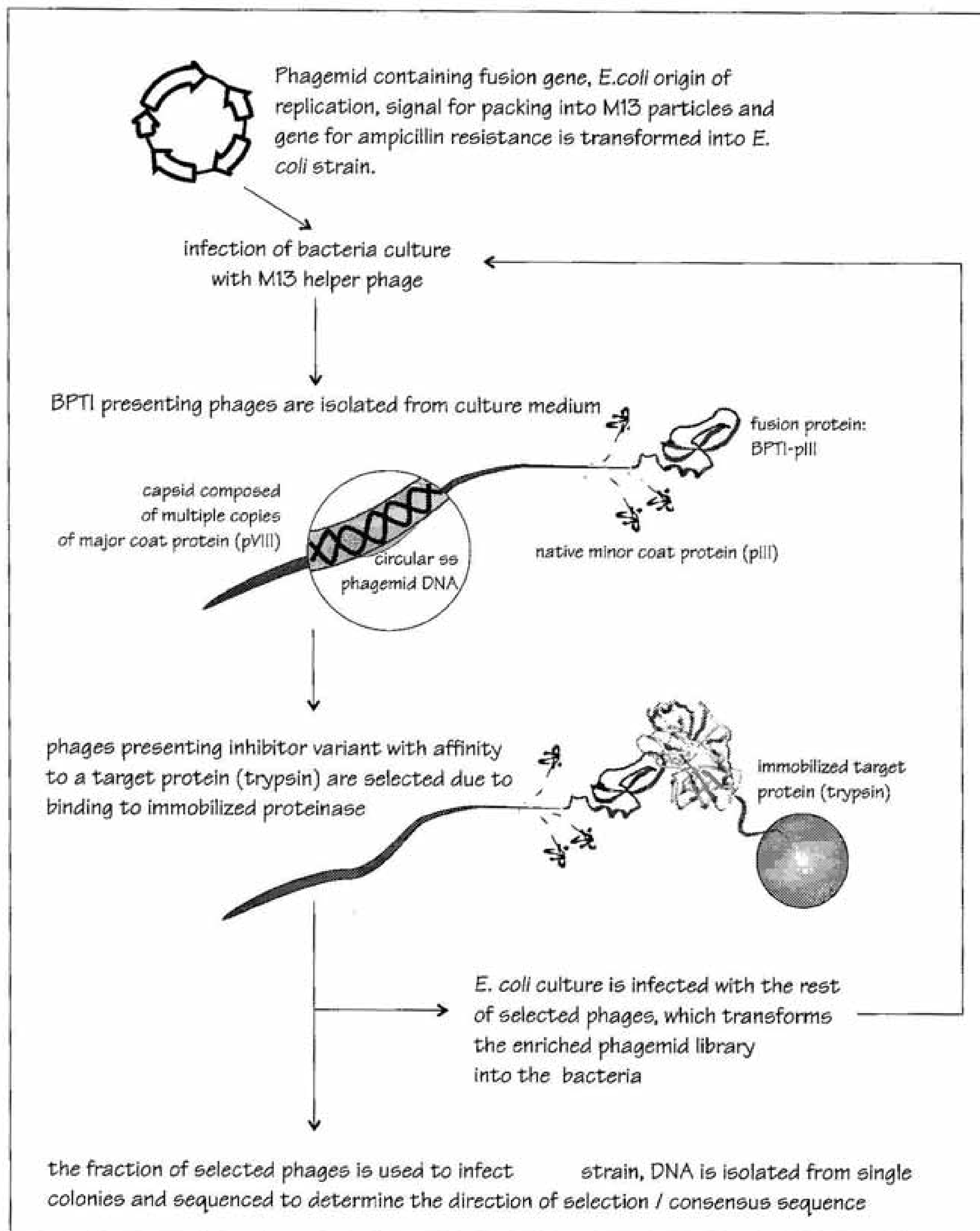

Figure 1. Schematic representation of phage display idea on the example of selection of the library of M13 displayed BPTI on bovine trypsin [57].

ss, single-stranded.

possibilities for expression, two of which are most frequently applied: one through major coat protein (pVIII), which provides a multivalent presentation, and the other through mi-

nor coat protein (pIII) for a monovalent display [27]. Sometimes, peptides are fused also to other coat proteins of filamentous phage, for example pVI [28]. Another, though less popular, bacteriophage used in the phage display approach is λ phage [29–31].

The idea of phage display strategy is summarized in Fig. 1. The first step of this approach is creation of a capsid fused expression system of a desired molecule. The subsequent randomization (full or partial) of a part of the gene, followed by an enrichment procedure, leads to the selection of a subpopulation of the library possessing the desired functional property(ies). Finally, the DNA sequence obtained from selected variants should be determined. The whole procedure, if successful, provides a consensus protein sequence for a selected feature(s).

Display systems are based on two alternative genetic constructs. The first is just a phage genome coding also for a fusion protein. The other is composed of a phagemid vector, which contains the fusion gene and the packing signal for enclosure of this DNA into phage capsid, and the helper phage, which is a source of the phage proteins.

The construction of a library is accomplished by several different methods. The most frequently used are cassette mutagenesis or double PCR. High yields of randomized sequence synthesis, ligation, and transformation are important for sufficient representativeness of the presented protein/peptide library.

An important factor, called the library representativeness (p), defines the probability value that all variants will exist in the pool of phages. The theoretical value of the representativeness, $p(n,t)$, can be calculated from the equation (1) [32]:

$$p(n,t)=1-\sum_{k=1}^{n}(-1)^{k+1}\times\binom{n}{k}\times\left(1-\frac{k}{n}\right)^{t} \quad (1)$$

where n is the number of variants in the library, and t is the number of independent transformants, phage particles, etc. For large n, the equation (1) can be approximated with equation (2) [32]:

$$p(n,t)=e^{-n\times e^{-t/n}} \quad (2).$$

The graphical illustration of the function $p(n,t)$ for different n is given in Fig. 2.

The actual representativeness of the library should be verified by DNA sequencing of a number of individual clones and/or of the entire DNA pool.

When introducing mutations, different randomization schemes can be applied. The full randomization of a given codon can be achieved with either NNN (N = equimolar G, C, A, T) or NNS (S = equimolar G, C), though the second one is more suitable, because it leads to smaller differences in the numbers of codons representing two different amino acids. That is extremely important, because differences in frequencies of different variants in a library have some pressure on selective binding to a target protein. In many particular cases there is no need to fully randomize all amino-acid positions. Some combinations of equimolar nucleotide mixtures (given in

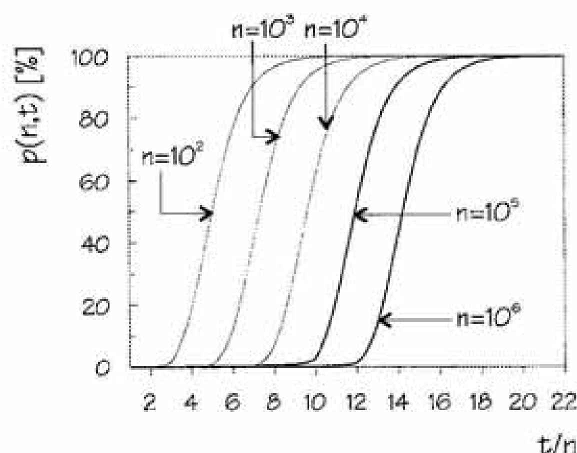

Figure 2. Plots of the library representativeness *p(n,t)* for different numbers of variants (*n*) *versus* the ratio of the number of independent transformants (*t*) to *n* [32].

brackets) and the resultant amino acids are given below:

- GA[G/C] codes for Asp and Glu;
- A[A/G]A codes for Lys and Arg;
- [A/G][A/G][C/G] results in charged amino acids: Asp, Glu, Lys, Arg and also Asn, Ser, and 2 Gly;
- T[T/A/G][G/C] gives large hydrophobic residues: Phe, Trp, Tyr, Leu and Cys, and amber stop codon;
- aliphatic amino acids: Ala, Val, Gly, Leu, Ile, Thr are obtained with [T/A/G][C/G/T]A, which also gives Arg, Ser and amber stop codon.

The nucleotide mixtures do not have to be equimolar and can be optimized to give particular subsets of amino acids [33].

The library selection can be performed in different ways, but almost all of them are based on the affinity chromatography principle. Different approaches use panning in 96-well plates, small affinity columns, slurry of an immobilized target or magnetic beads. In each of these procedures the target protein is bound to some carrier enabling easy selection of the phage presented mutants exhibiting the desired affinity. An efficient selection procedure should usually consists of a number (two to twelve) rounds of enrichment and propagation of selected phages. With the aim to achieve high specificity molecules, the enrichment can be also performed in the presence of negatively selecting agents.

IMMUNOGLOBULINS

Fab fragments of antibodies (Fig. 3) were among the first proteins expressed on the phage surface [15, 16, 34–42]. Application of the phage display approach to presentation of immunoglobulin and creation of their libraries offered a new method for monoclonal antibodies isolation, which did not demand traditional immunization procedures. It also enabled searching for high affinity human or humanized [38] monoclonal antibodies, which could not be easily obtained in any other way.

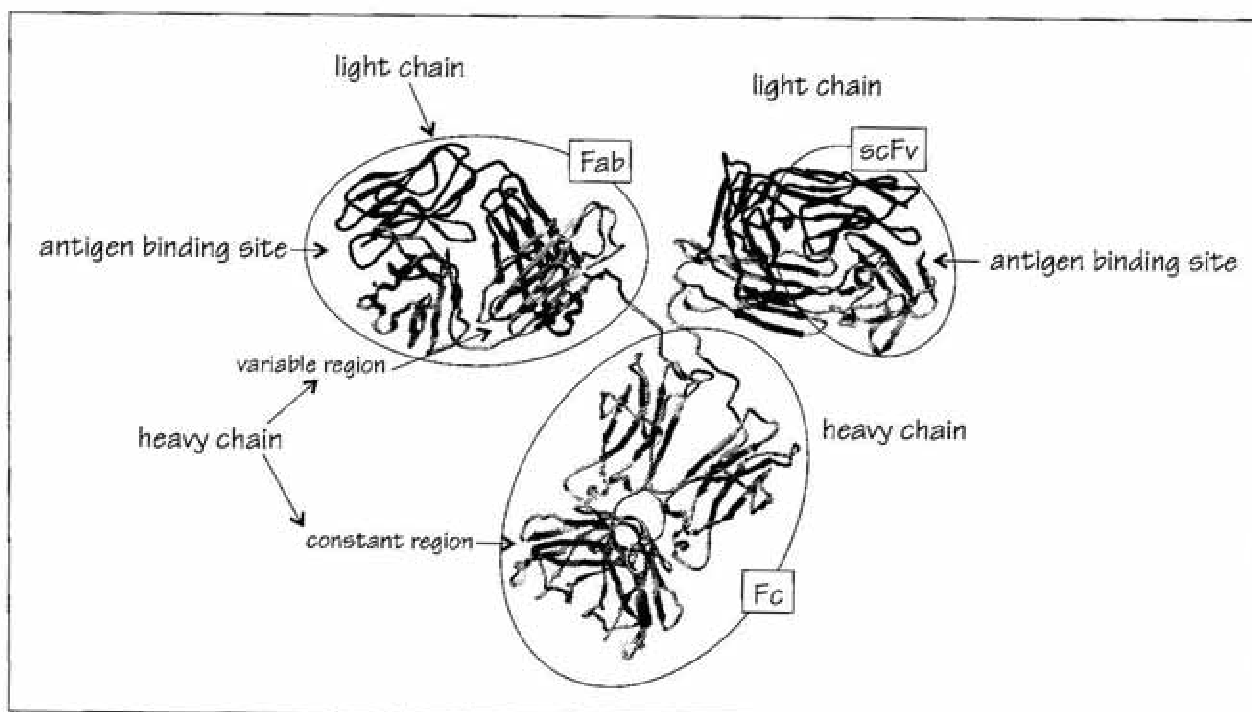

Figure 3. Three dimensional structure of immunoglobulin [70]. Different fragments of the molecule are indicated.

Among antibodies isolated by phage display is human anti-HIV immunoglobulin of picomolar activity [39], directed against HIV envelope glycoprotein gp120. Another example was isolation of the first human IgE of known specificity from a phage displayed library [40]. In this case, the library of IgE Fab fragments from peripheral blood lymphocytes of an allergic patient was expressed on the phage surface and selected for binding to a grass pollen allergen.

The potential use of phage antibodies in tumor targeting has also been demonstrated [41]. Researched on mice antibody fragments specific for an angiogenesis-associated fibronectin isoform demonstrated their ability to mark solid tumors, as markers of newly forming blood which are present in the most aggressive ones and should be reachable intravenously. The ability of these antibody fragments to target tumors can be used for cancer detection, when coupled to some fluorescent or radioactive agent. They can be also used for delivery of drugs, toxins, photosensitizers for immunophotodynamic therapy or enzymes for antibody-directed enzyme prodrug tumor therapy.

Usually, antibodies are phage displayed in the form of Fab or scFv fragments (see Fig. 3), and so they lack glycosylation and the structural region responsible for complement activation and cell-mediated effector functions of native antibodies. It is possible to mimic the function of a constant antibody region by adding a second binding site and creating bispecific immunoglobulins [42]. To demonstrate that such diabodies can activate the immune response, the phage antibody specific for C1q, the first component of the complement system, which binds to immune complexes, and also for hen egg lysozyme has been engineered. The diabody was able to recognize the lysozyme and to trigger complement-mediated effector functions. Thus, it appeared to mimic the binding of C1q to constant regions of complexed immunoglobulins.

Catalytic antibodies demonstrate the chemical potential of immune structures. Since their introduction, numerous antibodies of different catalytic possibilities have been obtained [43]. Usually, antibody catalyzed reactions proceed with rates 10^3–10^6-fold faster than the uncatalyzed ones [43]. Phage display also offers possibilities to select catalytic antibodies with stronger and optimized binding to the substrate [44].

PROTEINASE INHIBITORS

Protein inhibitors of serine proteinases are ideal candidates for drug design using phage display strategies, since they interact with the cognate enzyme *via* a sequential epitope called the proteinase binding loop [45, 46]. This six-amino-acid residue segment determines to a major extent the energetics of interaction with a proteinase. Of utmost importance is the P1 position (nomenclature according to [45]) which is located in the central part of this loop, while flanking positions, particularly in the case of specific enzymes, contribute to the energetics of recognition. Therefore, an applied phage display strategy often includes randomization of several residues in this region.

Basic pancreatic trypsin inhibitor (BPTI) and other inhibitors of similar architecture were frequently phage displayed to direct their specificity toward different proteinases. The P1 residue of BPTI is Lys15 which is responsible for extremely strong binding of the wild type inhibitor to bovine trypsin. In the first attempt, pIII fused BPTI was partially randomized at positions 15 to 19 with the aim to direct its activity toward human leukocyte elastase [47–48]. The variants selected adopted 4–5 mutations and exhibited an over 1×10^6-fold increased affinity towards the enzyme.

Another BPTI-like protein, called the Alzheimer amyloid β-protein precursor (APPI),

was used to select its variants capable of strong inhibition of tissue factor – factor VIIa (TF-VIIa) complex [49-50]. This complex initiates the blood clotting cascade, hence, its inhibitors are of pharmaceutical interest. The selection was performed on immobilized TF that was subsequently complexed with VIIa factor to mimic native conditions. The binding loop region, together with positions 34 and 39, which also contact the proteinase, were randomized to all possible amino acids in three different libraries. A consensus amino-acid sequence was found, which differed from the wild type APPI at four positions and inhibited the TF-VIIa complex 150-fold stronger than did the parent inhibitor. However, it also inhibited four other plasma serine proteinases (factors Xa and XIa, kallikrein and plasmin). In order to find a sequence which binds only to TF-VIIa complex, a selection was performed in the presence of factor Xa, kallikrein and plasmin. The phage-presented sequences were selected which could bind to the immobilized TF-VIIa complex and simultaneously did not interact with the other three enzymes present in soluble form. The variants obtained strongly bound to TF-VIIa complex and up to 10^4 weaker to the remaining three enzymes. This example nicely shows that different selection strategies can be adopted and the selected sequences will rather precisely reflect the conditions which were designed. The same libraries which were used for the TF-VIIa complex were also used to find strong inhibitors of kallikrein [51]. Again, several strong inhibitors could be selected with affinity over 10^4 higher than that of wild type of APPI.

Similar experiments were also performed using the first domain of human tissue factor – another example of BPTI-like protein – and selection was performed toward plasmin, kallikrein and thrombin [52, 53]. For the first two proteinases strong inhibitors could be easily found. However, in the case of thrombin, an enriched consensus variant did not inhibit this enzyme in a purified form. Most likely the variant was selected because it exhibited an affinity not toward the active site but to another region of the enzyme surface.

Unpredictable results were obtained when the library of APPI randomized at five loop positions was selected for affinity toward chymotrypsin [54, 55]. This proteinase is known to prefer large hydrophobic side chains – Trp, Phe, Tyr, Leu and Met – at the P1 position of inhibitor [56] which to a large extent determines energetics of the chymotrypsin-inhibitor interaction. Surprisingly, instead of the amino acids mentioned, high frequencies of Asn and His were selected. Our results on the library of P1 randomized BPTI enriched on chymotrypsin are more in keeping with known preferences of this proteinase, with Leu selected and Trp, Met, Tyr, and Phe also observed after the selection [57].

SUBSTRATE PHAGE DISPLAY

Substrate phage display is a powerful tool which might simplify proteinase specificity determination [58-61]. This technology is based on a phage presentation of a randomized peptide sequence which is inserted between a capsid protein and some protein domain (e.g. growth hormone) that interacts with another protein (e.g. growth hormone receptor) as demonstrated in Fig. 4. The randomized peptide sequence serves as a substrate library and growth hormone enables an attachment of phages to immobilized growth hormone receptor. After treatment of this construct with proteinase, the phages presenting a sequence susceptible to proteolysis are released to the solution and analysis of their DNA sequence allows to determine the proteinase specificity.

MINIMIZED PROTEINS

Although typically only about 5-15 amino-acid residues are involved in protein function, protein domains are composed of at least

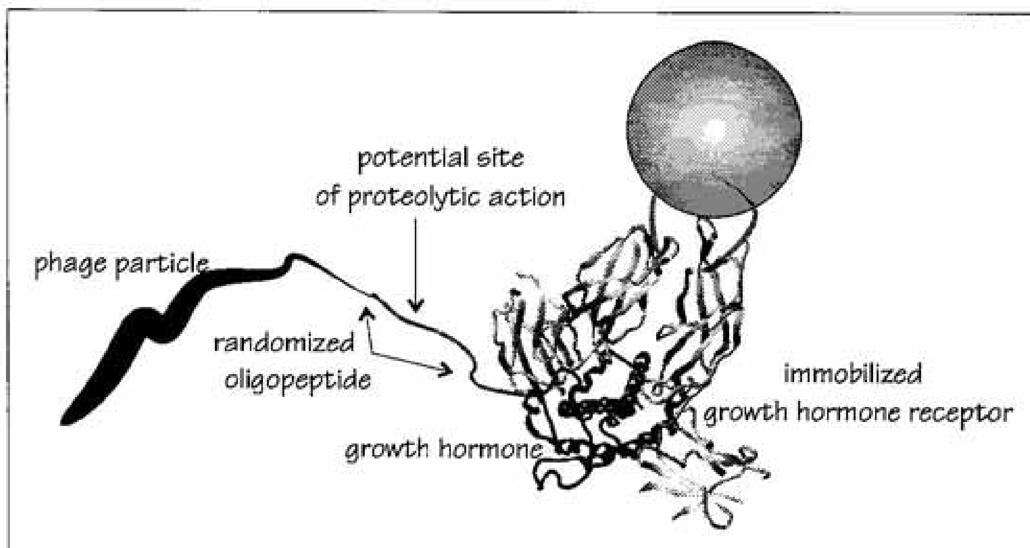

Figure 4. The illustration of substrate phage display method [58 -61]. Details of the procedure are described in the text.

30–60 residues. According to the common opinion, the major part of a protein molecule serves as a scaffold which ensures the optimal conformation of the functional residues. It is of interest whether it would be possible to reduce the size of a protein while still retaining its function. This task seems to be beyond possibilities of directed mutagenesis, as a significant reduction of protein sequence will lead to loss of protein activity and, eventually, only a coincidental set of mutations might lead to function recovery. Thus, the phage display strategy appeared to be a method of choice. Indeed, four last years provided successful examples of generating highly active minimized proteins [23, 26].

The first successfully reduced protein was atrial natriuretic peptide (ANP) which, in natural version, is built of 28 residues and contains one Cys7–Cys23 disulfide bond [10]. The procedure of ANP minimization is summarized in Fig. 5. Alanine scanning mutagenesis of ANP revealed that mutations at seven positions led to at least 10-fold lowering of affinity toward the extracellular domain of ANP receptor. Five of those residues form a sequential functional epitope on one side of peptide surface. In the first step of the minimization procedure an alternative Cys6–Cys17 disulfide bond was introduced to close this functional sequence into a smaller ring. Although relocation of the disulfide resulted in 100-fold reduction in receptor binding, in the next step almost full affinity was recovered through phage display randomization of five non-critical residues combined with selection for receptor binding. Next, the C-terminal part of ANP was shortened by deleting Gln18–Tyr28 segment, which caused again a 300-fold reduction in receptor binding. After this truncation the recovery of the biological function was accomplished through phage display reoptimization at residues following Cys17. Finally, optimization of the four residues flanking the Cys6–Cys17 disulfide combined with removing of the first N-terminal residues yielded the 15-residue peptide that binds only seven times weaker to the receptor than the wild type ANP. Comparative alanine-scanning mutagenesis revealed that basic structural and functional epitopes were similar in full-length and in minimized version of ANP, as judged from similar drops in affinity after mutations at analogous sequence positions.

This example clearly shows that 50% reduction of protein size with preservation of its function is possible. It involves, however, a long systematic multistep optimization of protein sequence, which seems to be possible only *via* combinatorial methods.

The second protein successfully minimized was a 59-residue three-helix bundle, the Z-domain of staphylococcal protein A [11]. This

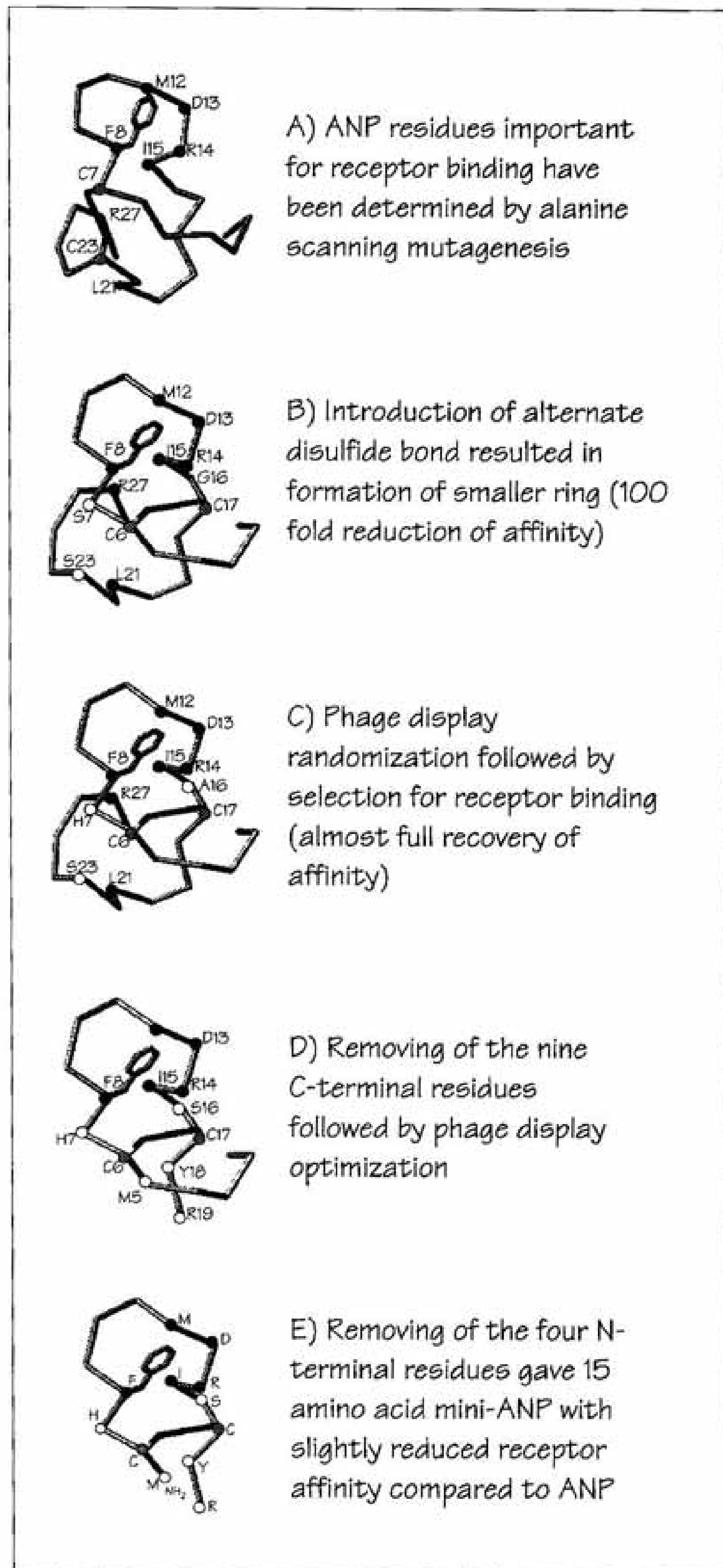

Figure 5. The overview of the procedure for ANP minimization [10]. Residues mutated or selected *via* phage display randomization at particular step are marked with white dots.

protein binds the constant fragment (Fc) of IgG through helix-1 (residues 7–18) and helix-2 (residues 20–38) while the third helix is not involved in contacts with Fc. Unlike in the ANP, the Fc binding determinants are sequentially discontinuous, this making protein minimization substantially more difficult. Truncation of the domain to the 38-residue two-helix peptide resulted in loss of the binding capability due to destabilization of its helices. The full recovery of biological activity could be accomplished in this case again through iterative multistep procedure that included phage display optimization of the hydrophobic residues that interacted with helix-3, optimization of 5 residues that form a helix-1–helix-2 interface and, finally, construction of next 5 libraries comprising 19 residues to improve contacts with Fc. A solution NMR structure of this truncated two-helix Z domain which contains 13 mutations showed that this domain, although it had adopted a spatial structure similar to that of the three-helix domain, was significantly less stable [62]. The stability could be enhanced through introduction of Cys10–Cys39 disulfide bond and truncation of 5 C-terminal residues. This final version of minimized domain Z binds to Fc with the same affinity as the wild type three-helical protein.

In the third example the filamentous phage display method was used to find peptide sequences capable of recognizing the extracellular domain of the human erythropoietin (EPO) receptor (EPOR) [63]. Human erythropoietin is a glycoprotein composed

in half of a 165 amino-acid residues protein and in half of a carbohydrate moiety. Initial studies allowed to isolate an 8-residue cyclic peptide flanked by Cys residues from the library of peptides presented on pVIII, that was capable of competing with EPO for binding to EPOR. Further optimization of this peptide sequence was accomplished by displaying it as a pIII fusion in such a way that three Cys-flanking residues were added at both N- and C-termini. As a result a family of closely related peptides with enhanced binding and EPOR agonist activity were obtained. These peptides have no homology with the sequence of EPO. The most surprising is that despite a huge reduction in size (2 kDa compared to 34 kDa) they can mimic the action of EPO inducing a tyrosine phosphorylation pathway.

MIRROR IMAGE PHAGE DISPLAY

Another surprising possibility of the phage display was described by Schumacher *et al.* [64] who generated a peptide composed of D-amino-acid residues capable of interacting with natural Src homology domain 3 (SH3). In this method phage display is applied to select L-peptides that bind to D-enantiomer of a target protein in order to, by a principle of mirror image, find a sequence of D-peptides that bind to the L-protein (see Fig. 6). At present this approach is limited to relatively small protein targets (up to about 100 residues) which can be chemically synthesized and purified to homogeneity, hence a 60-amino-acid protein was chosen to test the method. Interestingly, when the phage display library was screened with D-SH3 protein, peptides were isolated that were sequentially unrelated to those selected against L-SH3. The selected peptides contained a single disulfide bond that was not present in peptides selected against L-protein, and bound to D-SH3 with an affinity slightly lower than that of optimal ligand for L-SH3. Upon reduction of this disulfide bond the peptide became inactive. Heteronuclear ^{1}H-^{15}N NMR correlation spectra of the L-SH3–D-peptide complex showed that only part of the binding site for D-peptide overlaps the site for L-peptide. It is likely that D-peptides binding even stronger to L-SH3 domain could be obtained through further optimization of the contact area.

Thus, a biological barrier limiting encoded peptide and protein libraries to those composed of L-amino acids has been passed by. This opens possibilities to construct drugs re-

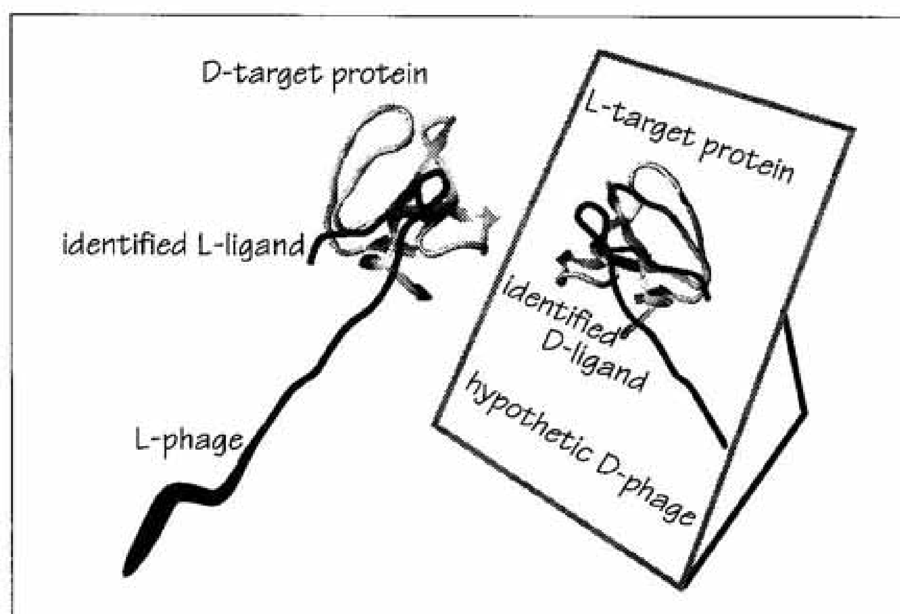

Figure 6. The principle of mirror image phage display with SH3 domain as a target protein [64].

sistant to proteolytic enzymes and, therefore, also of limited humoral immune response through MHC class II system.

SCAFFOLDS FOR INTRODUCTION OF NEW BINDING SITES

A different class of phage display approach is connected with the transplantation of binding sites into new protein frameworks. In this method we ask whether it is possible to adapt another protein structure to generate a high affinity receptor site that natively exists in a completely other protein scaffolding. This idea was tested by introducing an antigen-binding site into a few proteins, thus generating in them antibody-like activity (Fig. 7). For this purpose Ku & Schultz [65] adopted a cytochrome b_{562} four-helix bundle framework and, after intensive randomization of two loops connecting helices 1–2 and 3–4, obtained a cytochrome able to selectively bind to the conjugate of *N*-methyl-*p*-nitrobenzylamine and bovine serum albumin.

In another attempt, 13 residues of Z domain of staphylococcal protein A were randomized and panned against three different proteins: Taq DNA polymerase, human insulin and human apolipoprotein [66]. Degenerated residues are located on two α-helices involved in the binding of the Fc fragment of antibody. The selected clones had binding constants for respective proteins in micromolar range and a secondary structure similar to that of the parent sequence.

FOLDING AND STABILITY

Despite that phage display appears to be the method of choice, there are relatively few papers on applications of combinatorial strategies to protein folding and stability. O'Neil *et al.* [12] presented an IgG binding protein – the B1 domain of bacterial protein G, on M13 phage. The phages presenting the B1 domain exhibited a small plaque phenotype, compared to wild type M13. Interestingly, protein mutants that were significantly destabilized also produced the large plaque phenotype, enabling easy identification of less stable variants. In a systematic approach a library of the B1 domain was created using random muta-

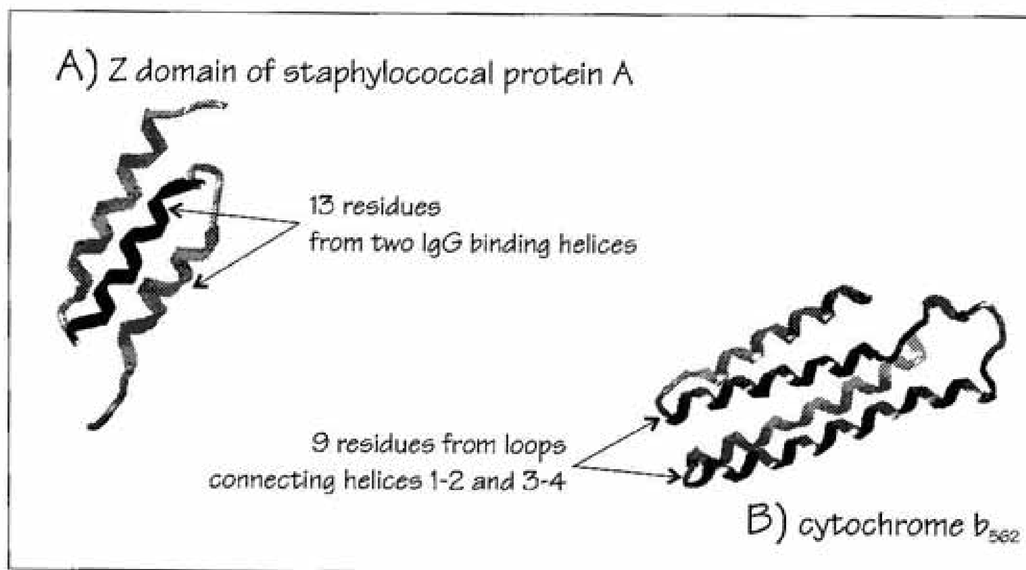

Figure 7. Examples of protein scaffolds for introduction of novel binding activities [65, 66].

A) In Z domain 13 residues from two helices were chosen for randomization to generate binding sites for Taq DNA polymerase, human insulin and human apolipoprotein [66]. B) In cytochrome b_{562} nine residues from loops connecting helices 1–2 and 3–4 were randomized to induce IgG-like activity toward the conjugate of *N*-methyl-*p*-nitrobenzylamine and bovine serum albumin [65].

genesis that comprised two third of the sequence including the helix and the last two β-strands. Many large plaque forming phages were isolated, sequenced and expressed and characterized by biophysical methods. The less stable mutants comprised those with disrupted packing of the hydrophobic core, destabilized secondary structure and altered hydrogen bonding network.

In a more general approach Gu *et al.* [67] searched from a large phage display library for protein variants that fold. As a model the 62-residue IgG-binding domain B1 of bacterial protein L was chosen. It was assumed that the selection procedure based on IgG-binding requires the protein variants to be properly folded. Indeed, the enrichment of the wild type library, compared to the library in which an additional strongly destabilizing Ile4Asp mutation was installed, was at the level of 10^5. It is surprising that both libraries differed only by one mutation placed in the hydrophobic core which did not contact the IgG. In the proper library 14 residues in the first two β-strands of protein L were partly randomized and after a selection procedure checked for their amino-acid sequences. The protein variants which could bind to IgG accumulated up to 9 mutations. The expression and biophysical characterization of 6 chosen mutants with 5–9 substitutions and significantly different sequences showed that they possessed relatively well defined tertiary structure albeit of significantly lowered stability.

In a distinct approach the 57-residue src SH3 domain was tested for the ability to recover its architecture from the simplified 5-amino-acid alphabet [9]. Five amino acids were chosen for this purpose: Lys, Glu, Ile, Ala and Gly. Because of experimental limitations on the size of library, the sequence of the domain was divided into three parts and initially simplified independently. The simplification could be achieved at 38 of the 40 positions that were randomized in these libraries. Next, SH3 variants with the sequence simplified throughout the entire length were obtained by random splicing of the simplified segments. In this manner two highly simplified sequences were generated – in one of them 40 of the 45 simplified residues were represented by one of those 5 amino-acid residues. Biophysical analysis showed that both variants share typical features of a small monomeric globular protein similar to wild type SH3, including: stability and folding, dispersion of NMR signals, gel filtration behaviour, CD and fluorescence spectra.

DNA-BINDING PROTEINS

Proteins responsible for DNA binding are of great importance in cell biology. The selection of DNA binding proteins of new specificities is required for biotechnology, medicine and research [68]. The phage display offers a tremendous tool for generating such novel affinities.

Zinc fingers (Zif) are ideal models to study protein–DNA interaction and to engineer models of different specificities [18]. Proteins known as zinc fingers occur as modular structures composed of a few small domains of similar structure, each stabilized by a zinc atom. Such a domain recognizes three contiguous base pairs, so identification of fingers binding to all 64 possible codons would give a chance to create combinations of domains that could recognize any DNA sequence [69]. Since altering of the specificity often requires changing multiple amino-acid residues, the phage display offers a wonderful methodology for this purpose.

The thorough analysis of sequence specificity was based on the structure of Zif 268, the three zinc finger murine transcription factor which recognizes a nine base pair sequence (Fig. 8) [69]. The phage display of the randomized domains of Zif 268 was applied and tested for binding to different DNA sequences. The engineered structures were able not only to bind to a target sequence and to discriminate between very similar binding sites, but also

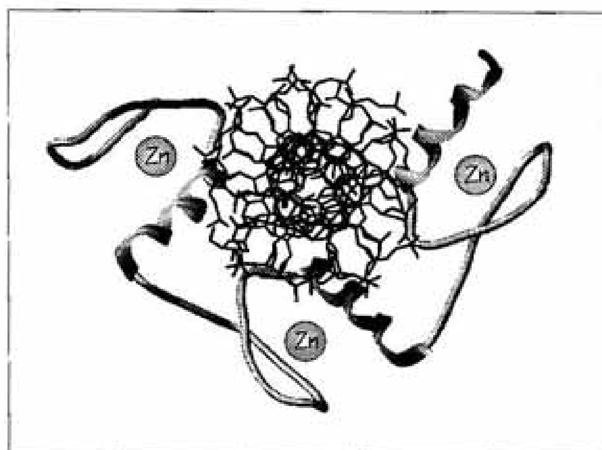

Figure 8. The three dimensional structure of the zinc finger Zif 268 complexed with double DNA helix [71].

The loops contacting DNA strand were targets for randomization followed by selection for novel binding affinities [68-69].

could block in mammalian cells the transcription of DNA containing a given target sequence. The specificity changes of Zif 268 have also been engineered by others [68].

We thank Dr. Antoni Kościelski for helpful discussion on the problem of library representativeness.

REFERENCES

1. Kunkel, T.A., Roberts, J.D. & Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367-382.
2. Shaw, W.V. (1987) Protein engineering. The design, synthesis and characterization of factitious proteins. *Biochem. J.* **246**, 1-17.
3. Wells, J.A. (1990) Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517.
4. Ellman, J., Stoddard, B. & Wells, J. (1997) Combinatorial thinking in chemistry and biology. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2779-2782.
5. Smith, G.P. (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the surface of the virion. *Science* **228**, 1315-1317.
6. Smith, G.P. (1991) Surface presentation of protein epitopes using bacteriophage expression systems. *Curr. Opin. Biotechnol.* **2**, 668-673.
7. Zhao, H. & Arnold, F.H. (1997) Combinatorial protein design: Strategies for screening protein libraries. *Curr. Opin. Struct. Biol.* **7**, 480-485.
8. Stemmer, W.P.C. (1995) Searching sequence space. *Biotechnology* **13**, 549-553.
9. Riddle, D.S., Santiago, J.V., Bray-Hall, S.T., Doshi, N., Grantcharova, V.P., Yi, Q. & Baker, D. (1997) Functional rapidly folding proteins from simplified amino acid sequence. *Nature Struct. Biol.* **4**, 805-809.
10. Li, B., Tom, J.Y.K., Oare, D., Yen, R., Fairbrother, W.J., Wells, J.A. & Cunningham, B.C. (1995) Minimization of a polypeptide hormone. *Science* **270**, 1657-1660.
11. Braisted, A.C. & Wells, J.A. (1996) Minimizing a binding domain from protein A. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5688-5692.
12. O'Neil, K.T., Hoess, R.H., Raleigh, D.P. & deGrado, W.F. (1995) Thermodynamic genetics of the folding of the B1 immunoglobulin-binding domain from streptococcal protein G. *Proteins: Struct. Funct. Genet.* **21**, 11-21.
13. Jucovic, M. & Hartley, R.W. (1996) Protein-protein interaction: A genetic selection for compensating mutations at the barnase-brarstar interface. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2343-2347.
14. Pasqualini, R. & Ruoslahti, E. (1996) Organ targeting *in vivo* using phage display peptide libaries. *Nature* **380**, 364-366.
15. Rader, C. & Barbas, C.F. (1997) Phage display of combinatorial antibody libraries. *Curr. Opin. Struct. Biol.* **8**, 503-508.

16. Hoogenboom, H.R. (1997) Designing and optimizing library selection strategies for generating high-affinity anibodies. *Trends Biotechnol.* **15**, 62–70.

17. Nygren, P.-A. & Uhlén, M. (1997) Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* **7**, 463–469.

18. O'Neil, K.T. & Hoess, R.H. (1995) Phage display: Protein engineering by directed evolution. *Curr. Opin. Struct. Biol.* **5**, 443–449.

19. Katz, B.A. (1997) Structural and mechanistic determinants of affinity and specificity of ligands discovered or engineered by phage display. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 27–45.

20. Lowman, H.B. (1997) Bacteriophage display and discovery of peptide leads for drug development. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 401–424.

21. Barbas, C.F., III (1993) Recent advances in phage display. *Curr. Opin. Biotechnol.* **4**, 526–530.

22. Allen, J.B., Walberg, M.W., Edwards, M.C. & Elledge, S.J. (1995) Finding prospective partners in the library: Two-hybrid system and phage display find a match. *Trends Biochem. Sci.* **20**, 511–516.

23. Nedvidek, M.N. & Hecht, M.H. (1997) Minimized protein structures: A little goes a long way. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10010–10011.

24. Cortese, R., Monaci, P., Luzzago, A., Santini, C., Bartoli, F., Cortese, I., Fortugno, P., Galfrè, G., Nicosia, A. & Felici, F. (1996) Selection of biologically active peptides by phage display of random peptide libraries. *Curr. Opin. Struct. Biol.* **7**, 616–621.

25. Dunn, I.S. (1996) Phage display of proteins. *Curr. Opin. Struct. Biol.* **7**, 547–553.

26. Cunningham, B.C. & Wells, J.A. (1997) Minimized proteins. *Curr. Opin. Struct. Biol.* **7**, 457–462.

27. Kay, B.K., Winter, J. & McCafferty, J. (1996) *Phage Display of Peptides and Proteins, Laboratory Manual.* Academic Press, San Diego.

28. Jespers, L.S., Messens, J.H., De Keyser, A., Eeckhout, D., Van Den Brande, I., Gansemans, Y.G., Lauwereys, M.J., Vlasuk, G.P. & Stanssens, P.E. (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology* **13**, 378–382.

29. Mikawa, Y.Gi., Maruyama, I.N. & Brenner, S. (1996) Surface display of proteins on bacteriophage λ heads. *J. Mol. Biol.* **262**, 21–30.

30. Maruyama, I.N. *et al.* (1997) Efficient epitope mapping by bacteriophage λ surface display. *Nature Biotechnology* **15**, 74–78.

31. Sternberg, N. & Hoess, R.H. (1995) Display of peptides and proteins on the surface of bacteriophage λ. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1609–1613.

32. Feller, W. (1966) *Wstęp do Rachunku Prawdopodobieństwa*; pp. 92–95, PWN, Warszawa (in Polish).

33. Arkin, A.P. & Youvan, D.C. (1992) Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis. *Biotechnology* **10**, 297–300.

34. Barbas, C.F., III & Lerner, R.A. (1991) Combinatorial immunoglobulin libraries on the surface of phage (Phabs): Rapid selection of antigen-specific Fabs. *Methods Companion Methods Enzymol.* **4**, 119–124.

35. Barbas, C.F. & Burton, D.R. (1993) *Manual for 1993 Cold Spring Harbor Laboratory Course on monoclonal antibodies from combinatorial libraries.* Cold Spring Harbor Laboratory.

36. Smith, G.P. & Scott, J.K. (1993) Libraries of peptides and protein displayed on filamentous phage. *Methods Enzymol.* **217**, 228–257.

37. Barbas, C.F., III & Burton, D.R. (1996) Selection and evolution of high-affinity human antiviral antibodies. *Trends Biotech.* **14**, 230–234.

38. Baca, M., Presta, L.G., O'Connor, S.J. & Wells, J.A. (1997) Antibody humanization using monovalent phage display. *J. Biol. Chem.* **272**, 10678–10684.

39. Yang, W.P., Green, K., Pinz-Sweeney, S., Briones, A.T., Burton, D.R. & Barbas, C.F. (1995) CDR walking mutagenesis for the affinity maturation of a potent human anti-HIV-1 antibody into the picomolar range. *J. Mol. Biol.* **254**, 392–403.

40. Steinberger, P., Kraft, D. & Valenta, R. (1996) Construction of a combinatorial IgE library from an allergic patient. *J. Biol. Chem.* **271**, 10967–10972.

41. Neri, D., Carnemolla, B., Nissim, A., Leprini, A., Querzè, G., Balza, E., Pini, A., Tarli, L., Halin, C., Neri, P., Zardi, L. & Winter, G. (1997) Targeting by affinity-matured recombinant antibody fragments of an angiogenesis associated fibronectin isoform. *Nature Biotechnology* **15**, 1271–1275.

42. Kontermann, R.E., Wing, M.G. & Winter, G. (1997) Complement recruitment using bispecific diabodies. *Nature Biotechnology* **15**, 629–631.

43. Lerner, R.A., Benkovic, S.J. & Schultz, P.G. (1991) At the crossroads of chemistry and immunology: Catalytic antibodies. *Science* **252**, 659–667.

44. Baca, M., Scanlan, T.S., Stephenson, R.C. & Wells, J.A. (1997) Phage display of a catalytic antibody to optimize affinity for transition-state analog binding. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10063–10068.

45. Bode, W. & Huber, R. (1992) Natural protein proteinases inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451.

46. Otlewski, J. & Apostoluk, W. (1997) Structural and energetic aspects of protein–protein recognition. *Acta Biochim. Polon.* **44**, 367–388.

47. Roberts, B.L., Markland, W., Ley, A.C., Kent, R.B., White, D.W., Guterman, S.K. & Ladner, R.C. (1992) Directed evolution of a protein: Selection of potent neutrophil elastase inhibitors displayed on M13 fusion phages. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2429–2433.

48. Roberts, B.L., Markland, W., Siranosian, K., Saxena, M.J., Guterman, S.K. & Ladner, R.C. (1992) Protease inhibitor display M13 phage: Selection of high-neutrophil elastase inhibitors. *Gene* **121**, 9–15.

49. Dennis, M.S. & Lazarus, R.A. (1994) Kunitz domain inhibitors of tissue factor – factor VIIa. I. Potent inhibitors selected from libraries by phage display. *J. Biol. Chem.* **35**, 22129–22136.

50. Dennis, M.S. & Lazarus, R.A. (1994) Kunitz domain inhibitors of tissue factor – factor VIIa. II. Potent and specific inhibitors by competitive phage selection. *J. Biol. Chem.* **35**, 22137–22144.

51. Dennis, M.S., Herzka, A. & Lazarus, R.A. (1995) Potent and selective Kunitz domain inhibitors of plasma kallikrein designed by phage display. *J. Biol. Chem.* **43**, 25411–25417.

52. Markland, W., Ley, A.C. & Ladner, R.C. (1996) Iterative optimization of high affinity protease inhibitors using phage display. 1. Plasmin. *Biochemistry* **35**, 8045–8057.

53. Markland, W., Ley, A.C. & Ladner, R.C. (1996) Iterative optimization of high affinity protease inhibitors using phage display. 2. Plasma kallikrein and thrombin. *Biochemistry* **35**, 8058–8067.

54. Kossiakoff, A.A., Hynes, T. & de Vos, A. (1993) Molecular recognition in biological systems: From activation to inhibition. *Biochem. Soc. Trans.* **21**, 614–618.

55. Scheidig, A.J., Hynes, T.R., Pelletier, L.A., Wells, J.A. & Kossiakoff, A.A. (1997) Crystal structure of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid β-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI):

Engineering of inhibitors with altered specificities. *Protein Sci.* **6**, 1806-1824.

56. Lu, W., Apostol, I., Qasim, M.A., Warne, N., Wynn, R., Zhang, W.L., Anderson, S., Chiang, Y.W., Rothberg, I., Ryan, K. & Laskowski, M., Jr. (1997) Binding of amino acid side chains to S_1 cavities of serine proteinases. *J. Mol. Biol.* **266**, 441-461.

57. Otlewski, J., Apostoluk, W., Buczek, O.L., Cholawska, L., Grzesiak, A., Kościelska, K., Krokoszyńska, I., Krowarsch, D., Stachowiak, D. & Dadlez, M. (1997) Structural and energetic aspects of protein inhibitor – serine proteinase recognition; in *Polish-Japanese Seminar. Trends Protein Res.* pp. 49-58.

58. Matthews, D.J., Goodman, L.J., Gorman, C.M. & Wells, J.A. (1994) A survey of furin substrate specificity using substrate phage display. *Protein Sci.* **3**, 1197-1205.

59. Chang, T.K., Jackson, D.Y., Burnier, J.P. & Wells, J.A. (1994) Subtiligase. A tool for semisynthesis of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12544-12548.

60. Ballinger, M.D., Tom, J. & Wells, J.A. (1995) Designing subtilisin BPN' to cleave substrates containing dibasic residues. *Biochemistry* **34**, 13312-13319.

61. Ding, L., Coombs, G.S., Strandberg, L., Navre, M., Corey, D.R. & Madison, E.L. (1995) Origins of specificity of tissue-type plasminogen activator. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7627-7631.

62. Starovasnik, M.A., Braisted, A.C. & Wells, J.A. (1997) Structural mimicry of a native protein by a minimized binding domain. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10080-10085.

63. Wrighton, N.C., Farrell, F.X., Chang, R., Kashyap, A.K., Barbone, F.P., Mulcahy, L.S., Johnson, D.L., Barrett, R.W., Jolliffe, L.K. & Dower, W.J. (1996) Small peptides as potent mimetics of their protein hormone erythropoietin. *Science* **273**, 458-463.

64. Schumacher, T.N.M., Mayr, L.M., Minor, D.L., Milhollen, M.A., Burgess, M.W. & Kim, P.S. (1996) Identification of D-peptide ligands through mirror-image phage display. *Science* **271**, 1854-1857.

65. Ku, J. & Schultz, P.G. (1995) Alternate protein frameworks for molecular recognition. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6552-6556.

66. Nord, K., Gunneriusson, E., Ringdahl, J., Stahl, S., Uhlén, M. & Nygren, P.-A. (1997) Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nature Biotechnology* **15**, 772-777.

67. Gu, H., Yi, Q., Bray, S.T., Riddle, D.S., Shiau, A.K. & Baker, D. (1995) A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* **4**, 1108-1117.

68. Rebar, E.J., Greisman, H.A. & Pabo, C.O. (1996) Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. *Methods Enzymol.* **267**, 129-149.

69. Choo, Y. & Klug, A. (1995) Designing DNA-binding proteins on the surface of filamentous phage. *Curr. Opin. Biotech.* **6**, 431-436.

70. Harris, L.J., Larson, S.B., Hasel, K.W. & McPherson, A. (1997) Refined structure of an intact IgG2A monoclonal antibody. *Biochemistry* **36**, 1581-1589.

71. Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: A model for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180.