

Directed Evolution with Fast and Efficient Selection Technologies

Ekkehard Mössner and Andreas Plückthun*

Abstract: Directed molecular evolution has proven to be a very powerful concept for the generation of proteins with improved properties, such as increased activity, binding affinity, folding efficiency or enhanced chemical and/or thermodynamic stability. We review here advances in the selection of proteins carrying desired mutations from pools of proteins that mostly carry unfavourable alterations. A short overview of the concept of directed evolution with a discussion of randomisation strategies is given first. Two technologies for the selection of proteins, each with its own advantages, are then discussed: In Ribosome Display, all steps are carried out in a cell-free system, which allows one to create very large libraries (diversity > 10¹¹), rapidly introduce mutations and thus obtain an iterative evolution. Examples with antibodies evolved for affinity or stability are discussed. In the Protein Fragment Complementation Assay, a library-versus-library selection is possible, that is, a simultaneous selection of binders against many targets. Examples with peptide and antibody libraries are discussed.

Keywords: Antibody library · *In vitro* selection · *In vivo* selection · Protein fragment complementation assay · Ribosome display

1. Evolution in the Test Tube

Evolution, the gradual adaptation to a selection pressure, is nature's way of responding to different environmental challenges. The evolutionary principle can be described as repetitive cycles of introducing mutations into the genome, selection for beneficial phenotypes and passing on the selected mutations to offspring, either directly (vegetative propagation) or *via* genetic recombination with a partner (sexual propagation). In the test tube, biochemists have been trying to mimic this process by performing mutagenesis on a gene of interest, selecting clones with improved properties and repeating these procedures over several rounds. But how can proteins be selected for their binding properties? A very time-consuming task would be the sequential screening of binders from a library of either a randomised gene or a cDNA library. In order to accelerate this process several technologies have been devel-

oped that allow the simultaneous screening of large protein libraries for the identification and enrichment of specific binders. All of these methods have in common that the genetic information for the protein of interest is physically linked to its phenotype, and the genes of selected proteins can be reamplified and are readily available for further analysis. This spectrum of screening and selection techniques include, among others, phage display [1][2], surface display on bacteria [3] and on yeast [4], the yeast two-hybrid system [5] and two more methods, which the authors believe have some distinct advantages and are described in this review: Ribosome Display [6] and the Protein Fragment Complementation Assay [7][8].

In this review, we will first summarize current methods of creating diversity before describing two of the selection technologies in greater detail. The first one, Ribosome Display (RD) [6], works entirely *in vitro* and allows the facile introduction of as many mutations as desired by the researcher. The second one, the Protein Fragment Complementation Assay (PCA) [7][8] is carried out in the bacterial cell and allows a parallel selection of many interacting pairs to proceed simultaneously.

2. Mutagenesis Strategies

Mutagenesis of a gene can be performed either by introducing mutations that are statistically scattered over the whole sequence or by focusing them only to a particular region. The first strategy can be achieved either by so-called 'error prone PCR' or by 'DNA-shuffling' technology (Fig. 1) [9][10]. In the former, the gene of interest is amplified by a DNA polymerase under conditions where the transcriptional fidelity is low and thus errors are introduced into the newly generated copies. Since most of the mutations are not beneficial, two favourable mutations would accumulate in the same gene only with low probability and many beneficial mutations would be masked by deleterious ones. To combine several desirable ones, the latter method, DNA-shuffling, can be applied, whereby the pool of genes generated by error prone PCR is digested by DNase I, an enzyme that cuts DNA unspecifically, and the fragments are reassembled by PCR (Fig. 1). In principle, each mutation can be recombined and propagated individually and independently of other mutations. This whole procedure resembles the crossing-over process that takes place on the chromosomes in living organisms.

*Correspondence: Prof. Dr. A. Plückthun
Biochemisches Institut der Universität Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Tel.: +41 1 635 55 70
Fax: +41 1 635 57 12
E-Mail: plueckthun@biocfebs.unizh.ch
<http://www.unizh.ch/~pluckth>

To introduce mutations only in a particular region of a gene, PCR cassette mutagenesis can be used (Fig. 1C). The DNA-fragment containing the desired mutational 'hot-spot' is cut out between two flanking restriction sites. Then a PCR product, which can be ligated exactly into these two restriction sites, is generated with a 'degenerate' primer. This primer is chemically synthesized to carry a randomised mixture of nucleotides in one or more codons and the encoded protein therefore carries a randomised set of amino acids at this particular position. Unfortunately, a completely randomised codon will exhibit a strong bias towards some amino acids which are overrepresented in the genetic code (e.g. the amino acid serine is encoded by six different codons, whereas tryptophan is encoded only by one codon), and it is also difficult to prevent the introduction of stop co-

dons. The solution to this problem is to use presynthesised trinucleotides (codons) of all 20 amino acids as the building blocks for DNA synthesis of the desired oligonucleotides instead of using the conventional mononucleotide building blocks [11]. These building blocks can then be mixed in any ratio, and only those trinucleotides need be added during the synthesis of a particular codon that the researcher desires at this position in the protein sequence.

3. Ribosome Display, an *in vitro* Display Technology

Ribosome Display (RD) is a selection technology developed in our lab, which works entirely *in vitro* and thus avoids any transformation steps of DNA into living cells [6][12]. Transformation is

usually the limiting factor in generating functional diversity in a library, and restricts the library size typically to 10^6 in the case of eukaryotic cells, such as yeast, or to 10^8 in the case of bacteria. The basic principle of RD, which gets around this limitation, is depicted in Fig. 2. First, a DNA fragment (or a pool of similar fragments) encoding the gene of interest is transcribed into mRNA by an RNA polymerase. This mRNA is translated into a protein using a ribosomal extract of *E. coli* (the so-called S30 extract). If this mRNA does not contain a stop codon, the ribosome stalls on the mRNA and a ternary complex is formed consisting of the mRNA, the ribosome and the newly synthesized protein which is still connected to its tRNA. This complex accounts for the linkage of genotype and phenotype. In other words, it physically tethers the peptide or protein to be selected to its

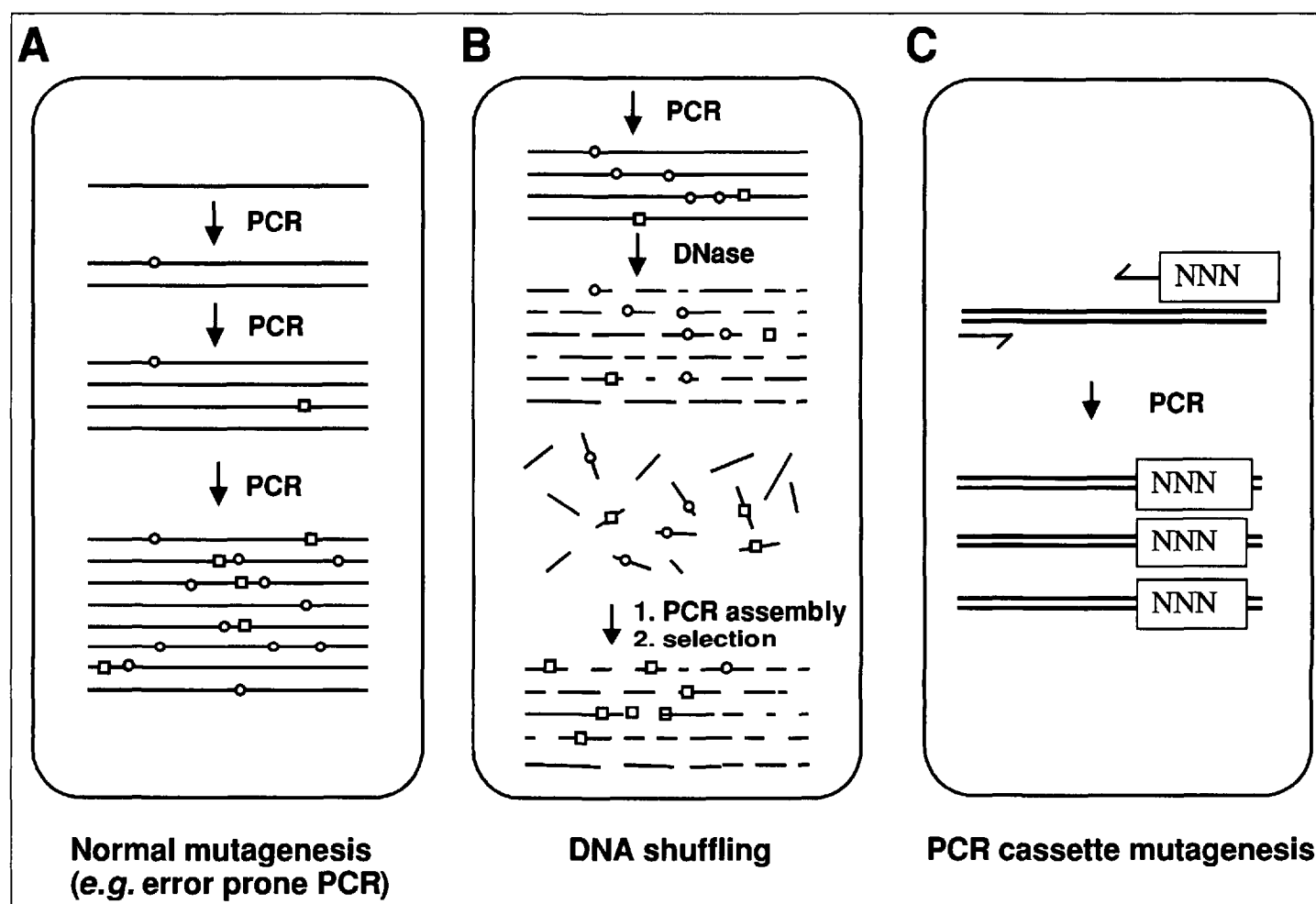


Fig. 1. Mutagenesis strategies for localized and whole-gene mutagenesis. Error prone PCR is depicted schematically in panel (A) as a strategy for the undirected distribution of mutations over a whole gene of interest. Two types of mutations are shown, favourable ones (open squares) and unfavourable ones (open circles). In successive cycles of PCR, more mutations of each are introduced, and usually molecules will contain some of either type. Thus, the beneficial effect of the favourable mutations can be completely obscured by the presence of unfavourable ones. A possible solution to this is shown in panel (B) and is called DNA shuffling [10]. The PCR product is cut into small fragments by the enzyme DNase I and subsequently reassembled by PCR. Mutations are thereby crossed, and genes with mostly favourable mutations can be enriched by selection. Panel (C) shows the principle of PCR using a 'degenerate' primer. This primer contains a mixture of all four possible nucleotides (abbreviated by the letter 'N') at a certain position or, alternatively, a mixture of codons assembled from trinucleotides [11]. The proteins synthesized from this gene will thus carry a randomised set of amino acids at this position.

blueprint. In the next 'panning' step, this complex is incubated with the desired binding partner, itself covalently attached to a solid support. Specific binders can be retained. Others will be washed away. Subsequent dissociation of the remaining complexes, isolation of the mRNA and backtranslating it into DNA by reverse transcription and subsequent PCR yields the genes of the selected binders. During this PCR step transcriptional errors can be introduced by the polymerase, and this process will generate a new library. Many mutations will be deleterious or neutral, but a few will lead to enhanced binding properties compared to the initial one.

Selecting for high affinity binders requires a stringent selection procedure, and an off-rate selection was found to be most successful [13]. The introduction of mutations during the selection process offers the possibility of generating 'diversity on demand', and therefore the sequence space that can be screened with this technology is far greater than that present in the initial library. This newly obtained DNA can then be subjected to the next round of selection, or can be analysed by sequencing, ELISA or RIA.

The crucial prerequisite for this whole process is the stable physical connection of the mRNA and the folded protein during the whole translation and panning procedure. Since only a folded protein will exhibit a phenotype that can be screened for, the protein that is attached to the ribosome must adopt a folded conformation. To allow this protein to fold, a linker sequence of 20 or more amino acids is encoded behind the coding sequence of the protein of interest, which allows the protein to leave the interior of the ribosome (where the translation takes place) without losing the connection to the tRNA, which is trapped inside the ribosome.

In theory, RD should be able to select even a single molecule that shows the desired binding property out of a large pool of unspecific binders. Two examples illustrate the power of this method. In one approach RD was applied to the simultaneous *in vitro* selection and evolution of scFvs from a large synthetic library (Human Combinatorial Antibody Library, HuCAL) against bovine insulin [14][15]. In different independent ribosome display experiments several scFvs were selected, all of which bound the antigen

with high affinity and specificity. All selected scFvs were affinity matured up to 40-fold, compared to their HuCAL progenitors, by accumulating point mutations during the ribosome display cycles. The dissociation constants of the isolated scFvs were as low as 82 pM, which strongly validates the power of this evolutionary method.

In another approach, also carried out in our lab, scFvs were selected after stability maturation [13]. Disulfide bridges are important stability elements in antibodies and normally can be formed only under oxidizing conditions. This observation served as a starting point for the selection of improved stability. One particular scFv, whose stability properties could be considered as 'average', was subjected to several rounds of RD while gradually increasing the selection pressure by increasing the concentration of the reducing agent DTT from 0.5 to 10 mM from the first to the last round. Mutants could only survive the selection pressure if they folded into a stable conformation in the presence of DTT and retained their antigen binding activity. Indeed, several scFvs could be isolated with increased thermodynamic stability

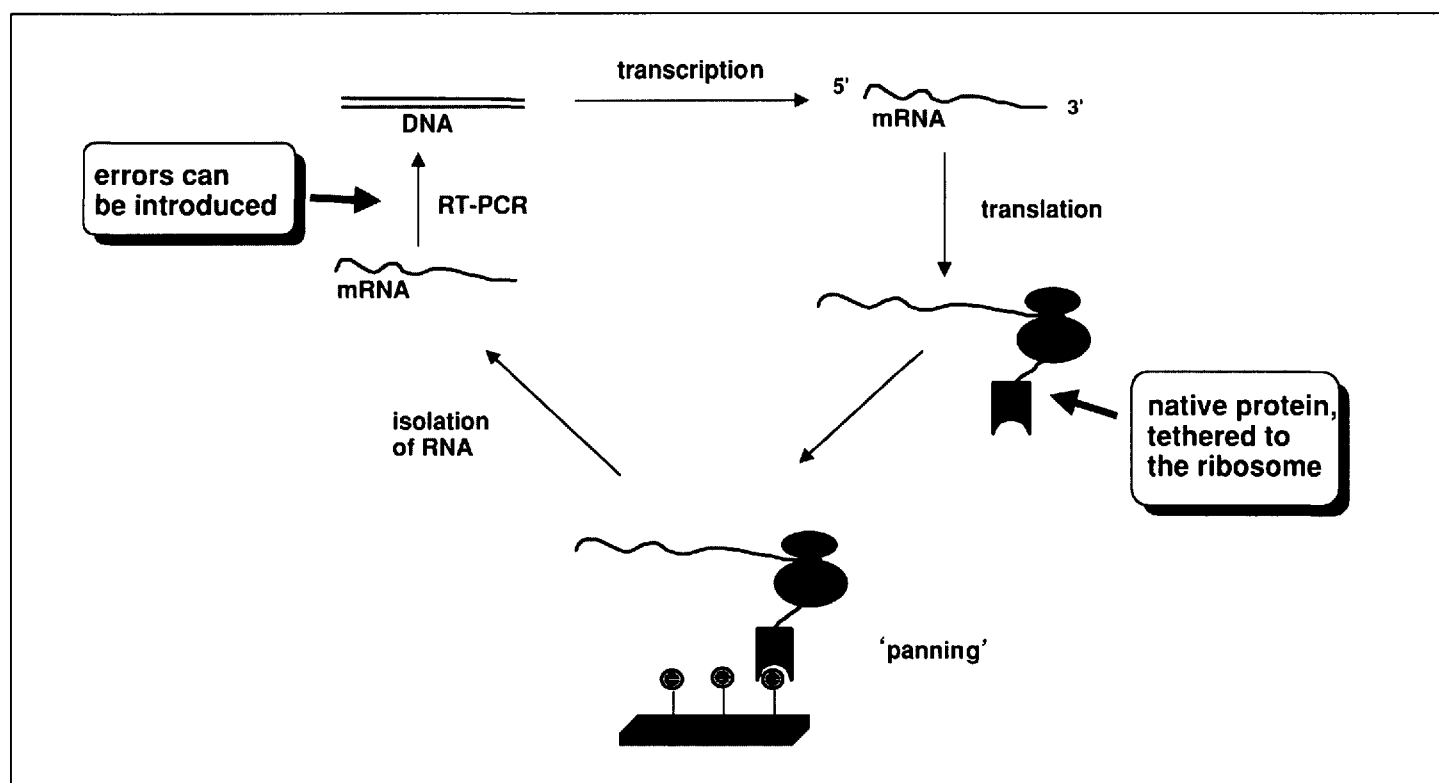


Fig. 2. Principle of screening and selecting a ligand-binding protein from a DNA library using the Ribosome Display technology. A DNA library is transcribed *in vitro* into mRNA by an RNA polymerase. This mRNA library can be translated into a protein library by applying the bacterial translational machinery (using the bacterial S30 extract). After stopping translation by cooling and increasing the magnesium concentration the mRNA, the ribosome and the newly synthesized protein will form a stable ternary complex. The desired ribosome complexes are affinity selected from the translation mixture by binding of the native protein to the immobilized antigen. Unspecific ribosome complexes are removed by intensive washing. The bound ribosome complexes can be eluted with antigen and the mRNA can be recovered and reverse transcribed into cDNA by RT-PCR. This cDNA can either be used for the next cycle of enrichment or can be analysed by sequencing and/or ELISA or RIA.

as measured *in vitro* either under reducing or oxidizing conditions.

These two examples demonstrate the ability of ribosome display to rapidly scan a large sequence space for beneficial mutations and to evolve defined biophysical parameters, provided a stringent selection strategy can be designed. This method is applicable to many protein classes and should have strong impact on the understanding of structure-function relationships as well as on the principles of directed evolution.

4. Protein Fragment Complementation Assay (PCA) for the Simultaneous Selection of Protein/Protein Interactions

Since the particular strength of RD lies in the identification and simultaneous optimisation of binders out of a protein library recognizing one defined target, another system was established which should in principle allow the parallel screening of many protein-protein interactions. In this approach a library of proteins can be screened against another library of proteins (library-*vs.*-library screening). This can be useful for obtaining a map of interacting proteins within one organism but also for identifying binders from an antibody library (see below). A similar approach had been described in the literature in which the yeast two hybrid system was used to screen a cDNA-library of the yeast genome (encoding virtually all of the approximately 6000 genes of *Saccharomyces cerevisiae*) against the same library to select cells that harbour two interacting proteins, one from each library [5]. Despite the wide usage of the yeast two hybrid system, it has been characterized by a high percentage of false positive results [5].

In order to design a system that would be very fast, permit the construction of large libraries and be less sensitive to false positives, a bacterial selection technique was developed in a collaboration between our lab and that of Stephen Michnick (University of Montreal) [7][8]. It is based on the functional complementation of the bacterial enzyme dihydrofolate reductase (DHFR) with two fragments of its murine counterpart (mDHFR). This selection technology has been called Protein Fragment Complementation Assay (PCA) [7][8]. In PCA the gene of mDHFR is dissected into two parts (fragment I and fragment II, see Fig. 3), each of them fused to a protein or peptide that

can form a complex together. When two plasmids encoding these constructs are expressed in *E. coli*, the two interacting partners recognize each other, allowing the two halves of mDHFR to come into close contact, thus restoring its enzymatic activity. The *E. coli* DHFR is inhibited by the antibiotic trimethoprim (TMP), whereas the murine DHFR does not show inhibition. Thus, functional mDHFR confers to *E. coli* the ability to grow on minimal medium in the presence of TMP. The dimerisation domains, which were used in one of the initial experiments, were derived from the leucine zipper domains of the proto-oncogenes c-Jun and c-Fos [16]. In order to find new pairs of stably interacting heterodimerisation domains, each of the c-Jun and c-Fos domains were randomised separately at certain positions and thus two libraries were constructed. Cotransforming both libraries in the same bacteria followed by a library-*vs.*-library selection indeed led to the isolation of a number of new pairs of heterodimerisation domains which permitted the allowed combinations at these positions to be defined. By applying different stringencies of selection, different sets of sequence pairs were obtained, and the most stringent selection was dominated by one pair, which showed a dissociation constant (K_D) as low as 24 nM [7][17].

We have recently adapted this system for the selection of antibodies in the single-chain Fv format (scFv) [18]. One of the leucine zippers was replaced by an antibody in the scFv format which recognizes a GCN4 leucine-zipper (Fig. 3D). In this positive control, reassembly of the two mDHFR fragments was obtained and was nearly as efficient as with the coiled coil helices. As mentioned above, antibodies in the scFv format normally contain two conserved disulfide bridges, one in each domain. Disulfide bridges are important stability elements and their removal causes a significant or even total loss of activity. The DHFR complementation process takes place in the bacterial cytoplasm, whose reducing environment normally prevents the formation of disulfide bonds. It was therefore not clear *a priori* whether this selection would be successful. On the other hand, several antibodies can be expressed in an active form in the cytoplasm of *E. coli* or yeast [19][20]. We used this anti-GCN4 antibody/antigen system to optimise linker length and fragment orientation.

With several model antibodies, specific for peptides and proteins, we found indeed that cognate interactions give rise

to about seven orders of magnitude more colonies than non-specific interactions. When transforming mixtures of plasmids encoding different antigens and/or antibodies, all colonies tested contained plasmids encoding cognate pairs.

Encouraged by these promising initial results, selections with both antibody and antigen libraries are now under way. We believe that this system will be very powerful as a routine system for generating antibodies especially in functional genomics, since, unlike in phage display, it does not require purification and immobilization of the antigen. The identification of an antibody specific for a cDNA- or EST-encoded protein will only require cloning, transformation and plating of bacteria. Furthermore, this technology may allow the simultaneous generation of antibodies against many targets, thereby considerably accelerating this process.

Acknowledgements

The help of Dr. Hideo Iwai with preparing figures and of Dr. Stephen Marino for critical reading of the manuscript is gratefully acknowledged.

Received: March 6, 2001

- [1] G. Winter, A.D. Griffiths, R.E. Hawkins, H.R. Hoogenboom, *Annu. Rev. Immunol.* **1994**, 12, 433.
- [2] G.P. Smith, *Science* **1985**, 228, 1315.
- [3] G. Georgiou, C. Stathopoulos, P.S. Daugherty, A.R. Nayak, B.L. Iverson, R. Curtiss, 3rd, *Nat. Biotechnol.* **1997**, 15, 29.
- [4] E.T. Boder, K.D. Wittrup, *Nat. Biotechnol.* **1997**, 15, 553.
- [5] P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadmodar, M. Yang, M. Johnston, S. Fields, J.M. Rothberg, *Nature* **2000**, 403, 623.
- [6] J. Hanes, L. Jermtus, S. Weber-Bornhauser, H.R. Bosshard, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 14130.
- [7] J.N. Pelletier, K.M. Arndt, A. Plückthun, S.W. Michnick, *Nat. Biotechnol.* **1999**, 17, 683.
- [8] J.N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12141.
- [9] R.C. Cadwell, G.F. Joyce, *PCR Methods Appl.* **1994**, 3, 136.
- [10] W.P. Stemmer, *Nature* **1994**, 370, 389.
- [11] B. Virrekäs, L. Ge, A. Plückthun, K.C. Schneider, G. Wellenhofer, S.E. Moroney, *Nucleic Acids Res.* **1994**, 22, 5600.

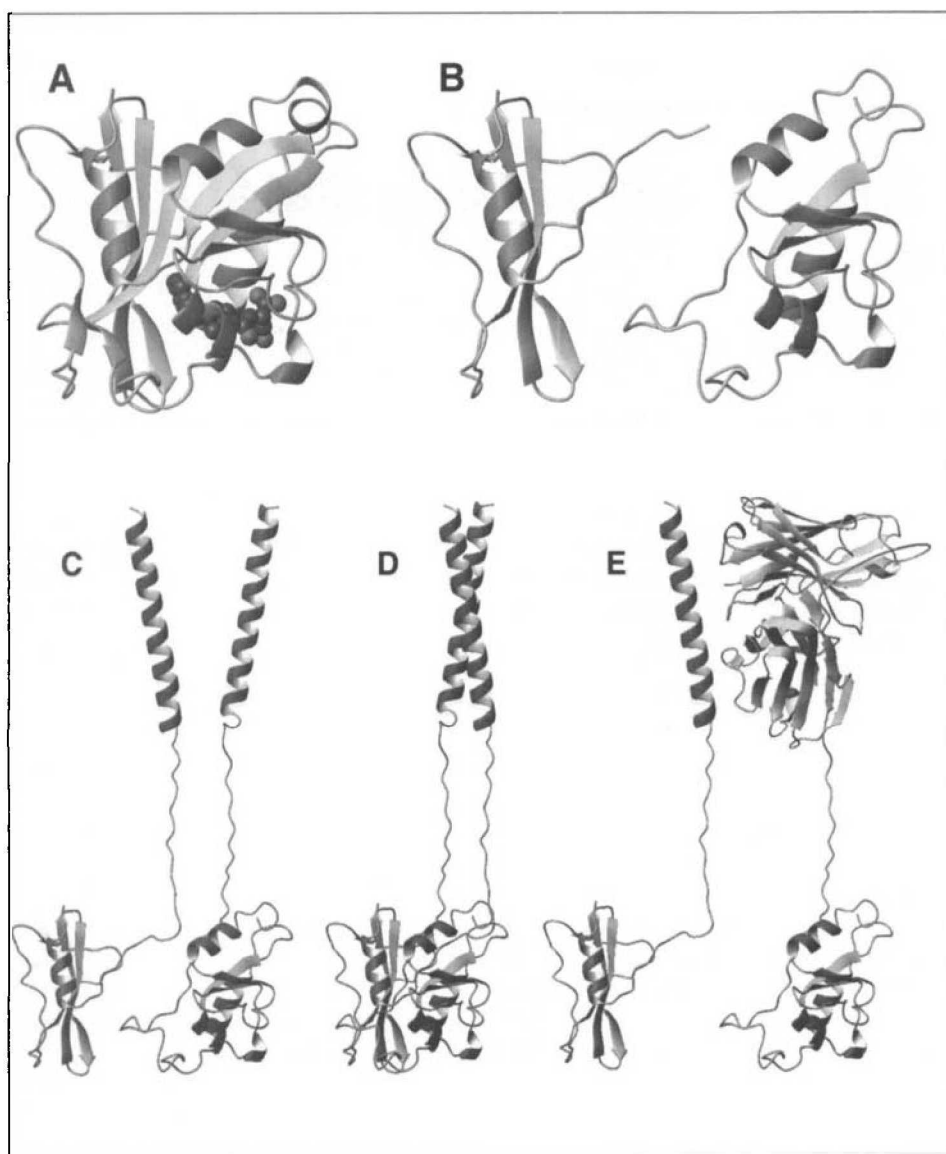


Fig. 3. Principle of the Protein Fragment Complementation Assay. (A) Native murine DHFR, shown as a ribbon diagram. This enzyme is important for the biosynthesis of purines, thymidylate, methionine and pantothenate. In this picture the folate cofactor is also shown as a space filling model. If this enzyme is genetically split into two parts (B), the fragments will not reassemble and activity is lost. Therefore, cell division and growth on minimal medium is prevented. Fusion of the individual fragments to protein domains that form a complex (C) can direct the reassembly of the DHFR fragments and activity is regained (D). This system is now used to select for scFv antibodies out of a DNA library against defined targets fused to one DHFR fragment (E).

- [12] J. Hanes, L. Jermutus, C. Schaffitzel, A. Plückthun, *FEBS Lett.* **1999**, 450, 105.
- [13] L. Jermutus, A. Honegger, F. Schwesinger, J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 75.
- [14] A. Knappik, L. Ge, A. Honegger, P. Pack, M. Fischer, G. Wellenhofer, A. Hoess, J. Wölle, A. Plückthun, B. Virnekäs, *J. Mol. Biol.* **2000**, 296, 57.
- [15] J. Hanes, C. Schaffitzel, A. Knappik, A. Plückthun, *Nat. Biotechnol.* **2000**, 18, 1287.
- [16] T. Curran, B.R. Franza, Jr., *Cell* **1988**, 55, 395.
- [17] K.M. Arndt, J.N. Pelletier, K.M. Müller, T. Alber, S.W. Michnick, A. Plückthun, *J. Mol. Biol.* **2000**, 295, 627.
- [18] E. Mössner, H. Koch, A. Plückthun, *J. Mol. Biol.* **2001**, in press.
- [19] G. De Jaeger, E. Fiers, D. Eeckhout, A. Depicker, *FEBS Lett.* **2000**, 467, 316.
- [20] K. Proba, L. Ge, A. Plückthun, *Gene* **1995**, 159, 203.