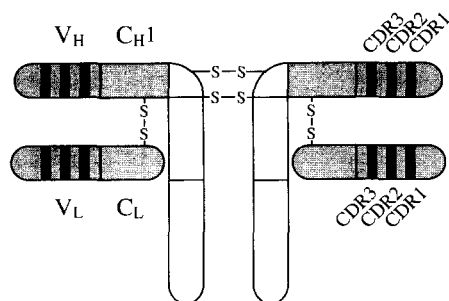


Specific Binding of Antibodies to DNA through Combinatorial Antibody Libraries

Michael Famulok* and Dirk Faulhammer

The immune system of a higher developed organism is probably the most potent system in nature for the highly specific recognition of all different kinds of molecular species. The antibody repertoire in man or mouse has a diversity between 10^7 and 10^{11} . The enormous number of antibody specificities is achieved by the combination of a relatively small number of incomplete genes (see e.g. immunoglobulin G1, Scheme 1). For



Scheme 1. Immunoglobulin G1. The regions shown in gray represent the Fabs. They consist of the variable (V_H) and the constant (C_H1) regions of the heavy chain, as well as the variable (V_L) and constant (C_L) regions of the light chain. The antigen is bound mainly in the complementarity-determining regions (CDR regions) within the V_H and V_L .

the variable region V_H of the heavy chain around 100 V_H genes combine with 30 diversity genes (D genes) and 6 joining genes (J genes) to yield about 18000 different complete V_H chains. Because of inaccuracies in connecting the genes and enzymatic treatment, this number can increase by three orders of magnitude. The variable region V_L of the light chain results from the combination of 100 V_L genes with 5 J genes. An antigen binding site consists of a V_L and a V_H chain; their random fusion increases the number of possible combinations to 10^{11} . Finally, somatic mutations in these gene regions can raise the number of different antibody specificities even further by several orders of magnitude. This diversity, however, is far from fully utilized by the organism. The actually existing repertoire is sufficient for the development of antibodies against an incredible number of

possible antigens, and as a result it effectively protects an organism against pathogenic or otherwise harmful agents.

This natural process, however, has certain limitations. Overcoming them would be of great importance for the more extensive technical, therapeutic, and diagnostic application of antibody technology. For example, the generation of antibodies specific for nucleic acids, such as double-stranded DNA (dsDNA) or RNA, by classical immunization of animals is quite often difficult, because nucleic acids typically lead to an extremely weak immune response. Likewise, the generation of catalytic antibodies by immunization with transition state analogues (TSAs) does not always lead to results desirable for an effective catalyst. In addition, only a fraction of the antibodies isolated from TSA-immunized animals bind the immunogen. On the other hand, only a few binders obtained are catalytically active.

For these and other reasons, extensive effort has been devoted for some time to the generation of synthetic antibodies, whereby the natural process is not only copied, but even improved.^[1] The goal is to use an antibody library with a huge number of binding specificities without relying on the immunization of animals or the employment of eukaryotic cells. A recent publication by Barbas et al. on the specific recognition of double-stranded DNA (dsDNA) by semisynthetic antibodies shows that these efforts are very promising.^[2]

How is it possible to copy nature's combinatorial process of antibody formation in the laboratory? How can it be used successfully for the selection of antibodies with novel functionalities? More importantly, how can the selected functional molecules be traced back genetically and amplified?

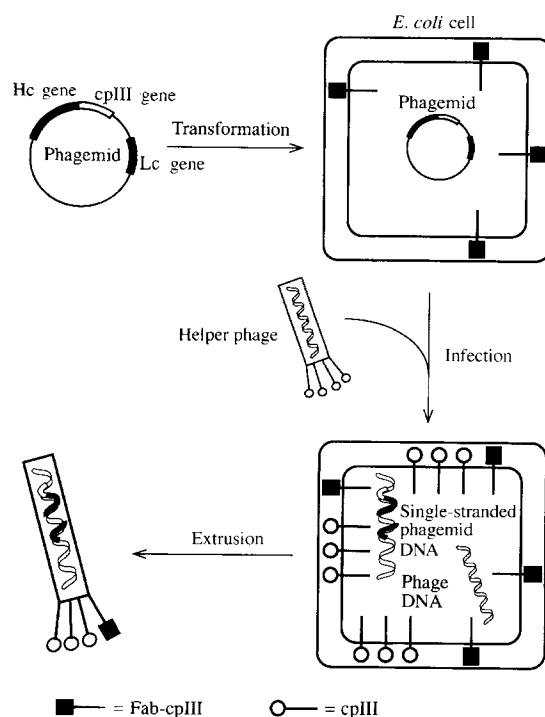
Among the most decisive steps for the generation of antibody diversity in vitro is the expression of single antigen-binding fragments (Fabs) in the bacterium *Escherichia coli*.^[3] The antigen-binding fragment is a heterodimer which consists of a light chain and a fragment of the heavy chain. Another crucial aspect is the amplification of V_H and V_L genes by the polymerase chain reaction (PCR).^[4] This was employed in the straightforward construction of libraries from the genomic cDNA of antibody-producing cells (spleen, bone marrow, peripheral blood cells). Furthermore, in this step semisynthetic antibody genes can be used, in which a part of the complementarity-determining region (CDR) is substituted by a randomized oligonucleotide. Finally, a gene vehicle, a vector, is required which is suitable for cloning the largest possible number of antibody genes and for

[*] Dr. M. Famulok, Dipl.-Chem. D. Faulhammer
Institut für Biochemie der Universität München
Am Klopferspitz 18a, D-82152 Martinsried (FRG)
Telefax: Int. code + (89)8578-2470

multiplying the corresponding proteins in the foreign cells. For this purpose, a bacterial virus (phage) is used, which contains the genes for the heavy and light chain and which is able to specifically infect bacteria with high efficiency.^[5] Another crucial advantage of phage was shown by G. P. Smith in 1985: any peptide can be expressed on the phage surface.^[6] This was achieved by cloning the peptide-encoding DNA into the phage in such a way that the resulting peptide is fused with either its N- or C-terminus to a phage protein called coat protein III (cpIII). The cpIII is responsible for the binding of the phage to the bacterium. For that purpose it is expressed at the phage surface at about four copies per virion. Exactly the same fusion principle could be applied to the antibody genes.^[7] For the construction of the antibody library a specially designed vector (fdCAT1) is used. This vector is constructed in such a way that owing to its fusion with cpIII the heavy chain can be anchored to the inner membrane of the bacterium, whereas the light chain is excreted into the periplasm, the space between the inner and outer bacterial membrane. Hence, the two chains assemble on the membrane to give the complete antibody. By this design it is possible to combine the genetic and the structural information: the interior of the phage contains the antibody genes which are expressed on the outside such that the antibody can interact with its antigen in solution. In addition, it is thus possible to select the phage by binding them to immobilized antigens, for example, by means of affinity chromatography. The phage selected in this way can then be used again to infect bacteria. This selection process can be repeated until a maximal enrichment of functional antigen-binding fragments (Fabs) is achieved.

However, this method still had two disadvantages: the phage requires cpIII on its surface to enter the bacterium. But after the direct cloning of the antibody genes into the phage, all four cpIII molecules on the "tip" of the phage contain a Fab, which leads to a dramatic reduction of infectivity. Furthermore, the four Fabs on the phage tip lead to an increase in binding capacity; in this way, weak binders can simulate low dissociation constants by chelation effects.

This problem was circumvented by using a special cloning vector, the phagemid pComb3, in combination with a helper phage for the cloning of the antibody library (Scheme 2). A phagemid is a hybrid of phage and plasmid, which consists of parts of the genome of a single-stranded DNA phage and of double-stranded components of plasmid DNA. Phagemids can reproduce within a host cell just like double-stranded plasmids. The simultaneous infection of the cells by the helper phage (M13) results in the synthesis of the gene products required for the packaging of phage particles, including the phagemids. In this case the phagemids can reproduce just like single-stranded DNA phages. In addition to the coat proteins, both intact cpIII and the Fab-cpIII fusion protein are produced; their genetic information is contained in the DNA of the helper phage and phagemid, respectively. If the DNA of the helper phage is mutated in a special way, the phagemid DNA is packaged preferentially. When the assembled phage is excreted from the cell, its surface is usually supplied with three cpIIIs and one Fab-cpIII fusion protein. The excreted phage thus contains the genetic information for the antibody genes in the form of a single-stranded copy of the phagemid DNA, together with a Fab molecule at its surface as well as the cpIII molecules required for

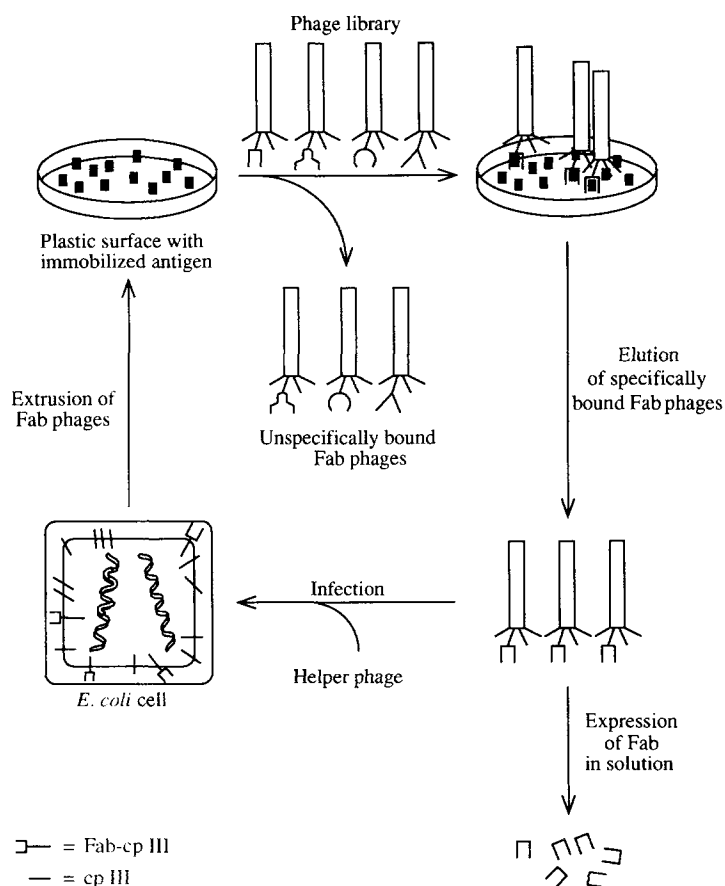


Scheme 2. Phagemid rescue of phagemid DNA by helper phages. The phagemid contains the genes for the light chain (Lc) and for the fusion protein, which consists of the cpIII and the heavy chain (Hc) of the Fab. Transformation into *E. coli* cells leads to the expression of the Fab-cpIII fusion protein in the periplasm (see text for details). The helper phage expresses all the proteins required in the construction of the virus particle, including cpIII. Since the phagemid carries signals required for phage packaging, the phagemid DNA is packaged into the phage particle in single-stranded form.

efficient infection. In this way about 10^5 copies of each phage and of each set of antibody genes are produced.

The selection step, in technical terminology referred to as "panning", is represented in Scheme 3. The whole pool consisting of roughly 10^{12} Fab-containing phage (10^5 copies $\times 10^7$ different phages in the library) is incubated with the antigen, which is immobilized on a plastic surface. Nonbinding phages are eluted with binding buffer, and the binding phages are thus enriched by about 100-fold. Specifically binding phages are then eluted, for example by affinity elution with the antigen or by lowering the pH. The isolated phages are used again to infect bacteria and amplify the phagemid DNA during cell growth. This DNA can then be packaged after addition of the helper phage, and the Fab can be expressed at the surface, as described above. These steps are repeated until most of the phage of the enriched phage-Fabs of the pool bind to the antigen. The DNA is then isolated from the specifically binding phage-Fabs, the cpIII gene is removed by treatment with restriction enzymes, and the phagemid is regenerated by ligation. The phagemid can now be used to infect bacteria that do not express the Fab at their surface, because the cpIII part at the surface is missing. The unfused Fab is produced in soluble form, and the free antibody can be isolated from the bacterial cell extract. The antibodies can now be tested individually for their binding affinity to the antigen.

Barbas et al. used three antibody libraries each consisting of 10^9 different molecules,^[2] in which certain regions in the V_H chain were randomized, to isolate specific DNA-binding anti-



Scheme 3. The panning process for selection of antigen-specific Fabs (see text for details).

bodies. In one library 10 amino acids in the CDR-3 of the heavy chain were randomized, in the other two libraries 16 amino acids. In general the CDR-3 of the antibody contributes most significantly to antigen recognition, because this region is characterized by the highest diversity. For the first two libraries the V_L region was held constant; in the third library the V_L chain contained a short randomized portion. For the panning, dsDNA from human placenta was used, which was immobilized on a plastic surface. After four cycles of panning, individual Fabs were tested for their binding behavior by enzyme linked immunosorbent assay (ELISA). Two Fabs (SD1 and SD2) gave a strongly positive signal. Both Fabs, which were bound to the immobilized DNA, could be eluted with a DNA solution that was only 10^{-9} M, which indicates a dissociation constant between 10 and 100 nM. The antibody SD1 clearly binds preferentially to G-C-rich regions; increasing concentrations of poly(dG)·poly(dC) and poly(dG-dC)·poly(dG-dC) inhibit its binding to DNA, whereas poly(dA)·poly(dT) or poly(dA-dT)·poly(dA-dT) do not show such an effect.

The antibodies selected by Barbas et al. do not fulfil the demands for an immediate technical or therapeutic application. In order to do so, they would have to bind with much better sequence specificity. With their preference for G-C-rich regions, however, the selected Fabs provide the basis for a novel class of sequence-specific DNA-binding molecules. This becomes evident when the 16-mer V_H region of SD1 is compared to the 10-mer V_H region of SD2. The amino acid sequence of both

initially randomized regions show considerable homologies:

SD1: Gly-Arg-Ala-Tyr-Gly-Gly-Asn-Gly-Leu-Lys-Met-Met-Ser-Leu-Asp-Asn

SD2: Gly-Arg-Gly-Trp-Ser-Gly—Ser-Leu-Asp-Ile

However, SD1 contains six additional amino acids. Since this is the most significant difference between the two Fabs, it is thought that this insertion of six amino acids might be responsible for the G-C preference of SD1. The effect of changes in this region on the sequence specificity of the antibody is being tested.

Antibodies tailored for the recognition of specific DNA sequences could be used in many different ways. In genome mapping, for example, specific DNA sequences could be blocked making them inaccessible to other enzymes. Synthetic DNA-binding Fabs could be used to interfere with the control of the expression of specific genes, which could be employed in a range of therapeutic applications. Sequence-specific antibodies might help to identify certain DNA sequences in gene diagnostics. But these are only a few of the potential applications. It is important to note that in principle the range of applications is unlimited whether for the generation of synthetic biocatalysts or for the neutralization of viruses or toxins. One of the most important prospects might be the selection and improvement of catalytic antibodies. The first promising results have already been reported.^[18, 9] For example, Chen et al. showed that an antibody library from mice immunized with a TSA hapten could be transferred to phage.^[18] From this phage library, hapten-binding clones were isolated, some of which were catalytically active. This result was the proof for the principal applicability of combinatorial Fab libraries for the generation of catalytic antibodies.

It can be expected that a broad spectrum of antibody chains with various binding motifs will soon be available. The corresponding genes could be used as building blocks for the construction of designer antibodies. Another emerging possibility is the selection of antibodies that function even under exotic conditions, for example in the presence of organic solvents or toxic compounds.

German version: *Angew. Chem.* **1994**, *106*, 1911

- [1] R. A. Lerner, A. S. Kang, J. D. Bain, D. R. Burton, C. F. Barbas III, *Science* **1992**, *258*, 1313–1314; A. S. Kang, T. M. Jones, D. R. Burton, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 11 120–11 123; C. F. Barbas, III, E. Björling, F. Chiodi, N. Dunlop, D. Chababa, T. M. Jones, S. L. Zebede, M. A. Persson, P. L. Nara, E. Norrby, D. R. Burton, *ibid.* **1992**, *89*, 9339–9343; D. R. Burton, *Acc. Chem. Res.* **1993**, *26*, 405–411.
- [2] S. M. Barbas, P. Ghazal, C. F. Barbas III, D. R. Burton, *J. Am. Chem. Soc.* **1994**, *116*, 2161–2162.
- [3] A. Skerra, A. Plückthun, *Science* **1988**, *240*, 1038–1041; M. Better, C. P. Chang, R. R. Robinson, A. H. Horvitz, *ibid.* **1988**, *240*, 1041–1043.
- [4] J. W. Larrick, Y. L. Chiang, R. Cheng-Dong, G. Senyk, P. Casali, *Prog. Biotechnol.* **1988**, *5*, 231; R. Orlandi, D. H. Güssow, P. T. Jones, G. Winter, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 3833–3837.
- [5] W. D. Huse, L. Sastry, S. A. Iverson, A. S. Kang, M. Alting-Mees, D. R. Burton, S. J. Benkovic, R. A. Lerner, *Science* **1989**, *246*, 1275–1281.
- [6] G. P. Smith, *Science* **1985**, *228*, 1315–1316; J. K. Scott, G. P. Smith, *ibid.* **1990**, *249*, 386–390.
- [7] J. McCafferty, A. D. Griffiths, G. Winter, D. J. Chiswell, *Nature* **1990**, *348*, 552–554; A. S. Kang, C. F. Barbas III, K. D. Janda, S. J. Benkovic, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4363–4366; C. F. Barbas III, A. S. Kang, R. A. Lerner, S. J. Benkovic, *ibid.* **1991**, *88*, 7978–7982; T. Clackson, H. R. Hoogenboom, A. D. Griffiths, G. Winter, *Nature* **1991**, *352*, 624–628.
- [8] Y.-C. J. Chen, T. Danon, L. Sastry, M. Mubarak, K. D. Janda, R. A. Lerner, *J. Am. Chem. Soc.* **1993**, *115*, 357–358.
- [9] B. Posner, J. Smiley, I. Lee, S. Benkovic, *Trends Biochem. Sci.* **1994**, *19*, 145–150; J. D. Stewart, S. J. Benkovic, *Int. Rev. Immunol.* **1993**, *10*, 229–240.