Molecular Evolution of Proteins on Filamentous Phage

MIMICKING THE STRATEGY OF THE IMMUNE SYSTEM

James D. Marks†, Hennie R. Hoogenboom‡, Andrew D. Griffiths§, and Greg Winter$$

From the †Medical Research Centre for Protein Engineering and the ‡Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom

Mutation. Proteins can be built by designing changes in a protein, altering the gene encoding the mutant protein in an organism and screening for new properties. Such protein engineering has been used to dissect the structure and function of proteins, for example to identify the residues involved in enzyme catalysis (1, 2) and protein folding (3) and to build altered proteins with useful properties (4, 5). In nature, proteins are built blindly, evolving by a process of mutation and selection. Sometimes the power of molecular evolution can be harnessed to order, for example bacteria with new or improved catalytic activities have been selected with substrates that are rate-limiting for growth (6, 7). However, it is unusual to be able to focus the selection pressure on a single enzyme since the properties of one enzyme are usually intermixed with others. For example, the $K_m$ values of enzymes for substrates have evolved to be similar to the concentration of substrate in vivo (8), which in turn is fixed by the activities of the other enzymes in the same pathway.

Ideally, we would like to be able to focus the selection pressure on a single protein. One such strategy is suggested by the immune system, which builds antibodies of high affinity and specificity of binding against a vast range of antigens. In the immune system: 1) a diverse repertoire of antibody molecules is generated by gene rearrangement; 2) this repertoire of antibodies is expressed (displayed) on the surface of B-cells, with each cell displaying a single antibody species; 3) single B-cells proliferate on binding of antigen and differentiate to short-lived cells (plasma cells) making soluble antibody or to long-lived cells (memory cells) with surface antibody; and 4) the affinities of these antibodies are increased (affinity maturation) by random point mutation (somatic mutation) of the antibody genes and further antigen selection (Fig. 1). Here we review recent progress and prospects for making human antibodies and evolving other proteins for new binding and catalytic functions by mimicking in bacteria the strategy of the immune system (9, 10).

Mimicking the Display of Antibody on the Surface of the B-cell

In the immune system, the B-cell provides a "genetic display package," with antibody expressed (displayed) on the outside of the cell to encounter and bind to antigen and the genes encoding the antibody contained within. Viruses could also be used as a genetic display package by inserting antibody genes into a gene encoding a viral coat protein. The resulting virus would display antibody on its surface fused to the coat protein and contain the antibody gene within.

Filamentous bacteriophage, viruses which infect bacteria, have been used to display small peptides fused to a minor coat protein (pIII) (11, 12) or to the major coat protein (pVIII) (13, 14) by inserting synthetic DNA encoding the peptide into either gene III or gene VIII. There are three (or possibly five) copies of pIII located at the tip of the phage and about 2500 copies of the pVIII per phage. pIII is responsible for attachment of the bacterial F pilus, and pVIII for coating the single stranded DNA. The pIII protein has two "domains." Fusions can be made to the N terminus of pIII (12, 15-19), or the N-terminal domain can be removed and fusions made to the second domain (20-22); however, phage lacking the N-terminal domain are not infective.

When the gene encoding a single chain Fv (scFv) antibody fragment (defined in the legend to Fig. 1) was inserted into gene III, the scFv was fused to the N terminus of pIII and incorporated into the phage, allowing the phage to bind antigen (15). Heterodimeric Fab antibody fragments have also been displayed by linking the heavy or light chain to a coat protein and secreting the other chain into the bacterial periplasm where the two chains associate (23, 24). Fab fragments have been displayed at the N terminus of pIII (16, 19), the second domain of pIII (21, 22), or at the N terminus of pVIII (25-27).

With filamentous phage, it is also possible to mimic the expression of soluble antibody from the plasma cell. By inserting an amber stop codon between the antibody gene and gene III, when phage is grown in a supE suppressor strain of Escherichia coli, the amber codon is read as glutamine and the antibody fused to pIII is displayed on the surface of the phage. When the phage is grown in non-suppressor strains, the amber codon is read as a stop codon, and soluble protein is secreted from the bacteria (Ref. 16; see also Ref. 28) (Fig. 2).

Mimicking Antigen-driven Selection

Affinity selection of phage with ligand mimics the antigen-driven selection process of the immune system and results in recovery of phage encoding the displayed antibody. Phage displaying peptides have been selected by binding to a monoclonal antibody (29-32), and phage displaying antibodies by binding to antigen (15, 19, 21, 22, 26). Several formats have been used for affinity selection of the phage, for example binding to biotinylated antigen in solution followed by capture on streptavidin-coated beads (33), binding to antigen-coated dishes (34), or binding to antigen on a column matrix (15, 35). Bound phage can be eluted by acid or alkali and enrichment factors of 20-fold (34) to 1000-fold (15) to more than 1,000,000-fold (21) obtained for a single round of affinity selection. However, by infecting bacteria with the eluted phage, more phage can be grown and subjected to another round of selection. In this way an enrichment of 1000-fold in one round of selection can become a factor of 1,000,000-fold over two rounds of selection (15). Thus even when enrichments are low (33, 34), multiple rounds of affinity selection can lead to the isolation of rare phage and the genetic material contained within.

Phage can be selected on the basis of affinity, avidity, or kinetics of binding. For example, phage displaying antibodies have been allowed to bind monovalent biotinylated antigen in solution and then been captured on streptavidin-coated paramagnetic beads. By setting the concentration of biotinylated antigen lower than the equilibrium dissociation constant (but in excess of the phage concentration), phage could be selected by their affinities of binding (33). Using many rounds of selection (up to 12), it was possible to select for antibodies even with small differences (2-4-fold) in affinity (33). Alternatively, phage could be selected by the kinetics of dissociation from the biotinylated antigen by diluting phage into excess unlabeled antigen prior to

1 The abbreviations used are: scFv, single chain Fv; PCR, polymerase chain reaction; CDR, complementarity-determining region.
FIG. 1. The strategy of the immune system in vivo and using phage. Antibody heavy (VH, dark shading) and light (VL, light shading) chain variable domains create the antigen binding site. Each domain consists of a β-sheet framework supporting three antigen binding loops (CDRs) at the tip of each domain. Antibody molecules, by having loops of different lengths, different main chain conformations, and different amino acid side chains can create binding sites of a variety of sizes (400–800 Å) (57–59) and shapes, ranging from flat surfaces to pockets (57–59). In vivo (upper panel) the sequence and structural diversity of the loops is generated by the rearrangement of separate gene elements. The VH domain results from the rearrangement of a germ line V-gene (encoding the first two CDRs) to a diversity (D) gene and joining (J) gene (together encoding the third CDR). The VL domain results from the rearrangement of a germ line V- or Vx-gene, encoding the first two CDRs and most of the third CDR, and a Jx- or Jx-gene encoding the remainder of the third CDR. In mice, the rearrangement of the V-gene segments could give rise to >10^6 VH sequences and >10^4 VL sequences and a primary repertoire of >10^11 VH/VL sequences; however, only 10^6–10^7 B-cell clones are present in the circulation at any time (50, 60). The antibody molecules are displayed on the surface of the B-cell where they function as an antigen receptor binding of antigen results in proliferation and differentiation of the B-cells to short lived plasma cells making soluble antibody or to long lived memory cells. Antibody affinity is increased by mutation of the antibody V-genes followed by further antigen selection. To mimic the immune system in phage (lower panel), rearranged V-genes can be amplified from mRNA using the PCR, as the ends of rearranged heavy and light chain V-genes are sufficiently conserved to design "universal" primers for PCR amplification of both heavy and light chain V-genes from RNA (34, 61-64) (or DNA (65, 66)). "Natural" repertoires can be made from unimmunized donors using IgM mRNA, whereas IgG mRNA can be used to tap the postimmunization repertoire (35, 67, 68). Alternatively, entirely synthetic repertoires of rearranged V-genes can be generated (45). The rearranged heavy and light chain V-genes are assembled together randomly (69) to encode either scFv antibody fragments (VH and VL domains artificially linked together by a flexible polypeptide (70, 71)) or Fab fragments and cloned into a phage or phagemid vector (see Fig. 2) to create a library of millions of different phages. Binding phage are affinity-purified on antigen and used to infect a non-suppressor strain of bacteria to produce soluble antibody fragment. The isolated antibody genes can be mutated and the resulting mutant phage subjected to further affinity selection to produce higher affinity antibodies.

FIG. 2. Display of antibodies on pIII and pVIII using phage and phagemid vectors. For display on gene III, phage vectors give 3 copies of the antibody-pIII fusion protein while phagemid vectors result in 0–3 copies of fusion protein. Other details are explained in the text. Antibody fragments are depicted as dark (heavy chain) and light (light chain) shaded areas; antibody genes are depicted as a dark shaded bar (heavy chain) and a light shaded bar (light chain). For displayed antibody fragments, only infectious phage is illustrated; p gene III, gene 3 promoter; p lacZ, lacZ promoter; p phoA, alkaline phosphatase promoter; AMP, ampicillin resistance; TET, tetracycline resistance. Citations for vectors: fd-tet-DOG1 (35), pHEN1 (16), pDH188 (21), M13XHL (27), pTacCP (28).

capture on magnetic beads. Such "off rate" selection is powerful, as small differences can be greatly amplified over time due to the exponential nature of the dissociation (33). The binding of phage to a solid phase followed by washing also selects for those phages with slower off rates (36).

The avidity of binding of phage to a solid phase coated with antigen depends on the avidity of binding of each molecule of displayed antibody and also on the number of antibody fragments per phage that can engage in binding. The number of antibody fragments per phage is largely determined by the choice of either pIII (16, 19, 21, 22) or pVIII (25–27) coat proteins for fusion, the use of phage (15, 35) or phagemid vectors (16, 19, 20–22), and the use of amber mutations between antibody and coat protein (16, 27, 28) (Fig. 2). Other factors may also contribute, for example any association of antibody fragments as dimers on the phage or the proteolysis of the fusion protein. For binding to a solid phase, the avidity of binding will also depend on the coating density of antigen.
Thus for phage vectors, fusion of antibody to pIII should lead to three copies of the fusion protein on each phage particle allowing the phage to bind multivalently to solid phase antigen (Fig. 2). With pVIII fusions, many more copies of the fusion can be displayed, for example up to 24 antibody molecules/phage (26) (Fig. 2). The higher avidity of binding with multivalent display may help retain lower affinity phage on solid phase antigen during washing, especially for phage with fast off rates, but may hinder the discrimination between phage with different affinities (31).

For phagemid vectors, which require the use of helper phage to make phage, the pIII coat protein from the helper phage competes with pIII fusion protein for incorporation into the phage particle leading to a population of phage with an average of less than three copies of antibody per phage (for pIII fusions). Indeed, with fusions to the second domain of pIII, infective fusion phage must contain at least one pIII from the helper phage and, therefore, a maximum of two fusion proteins per phage (Fig. 2). Although it has been shown that on average there is less than one fusion protein per phage (20, 21), the population may still contain a significant fraction of bivalent phage. “Monovalent” display appears to enhance the discrimination between phage with different affinities (20).

**Mimicking Affinity Maturation**

**In vivo,** the increase in the binding affinity of antibodies results from somatic mutation of V-gene pairings seen after the initial immunization (primary response) and also from the appearance of antibodies that use V-genes not seen in the primary response (repertoire shift) (37-39). The number of somatically mutated B-cells arising from a single selected B-cell is much less than the primary library of about $10^7 - 10^8$ B-cells in mice. In principle, the use of phage display could allow the selection of higher affinity antibodies than is possible in the immune system as much larger numbers of variant phage could be made, for example with $>10^9$ mutants of a single antibody, with more (perhaps $>10^{10}$ mutants) if the phage is grown in mutator strains (see below).

The process of somatic mutation can be mimicked in vitro by scattering the entire V-gene with random mutations. For example, an error prone polymerase was used to introduce random point mutations (33, 40), and in principle mutations could also be introduced at random by spiked oligonucleotides (41) or growth of the phage in mutator strains (42, 43). For more extensive variation, artificial crossovers between V-genes could be induced using the polymerase chain reaction (PCR) (44) or somatically mutated V-genes harvested from B-cells (see below) (36). Alternatively, synthetic oligonucleotides could be used to focus mutation on residues likely to be involved in binding and away from key structural residues, for example in the antibody CDRs. However, the number of residues that can be systematically varied is limited; in order to replace 6 amino acid residues with 32 random codons (NN(T/G)) at each site, we would need to make a repertoire of $>10^8$ nucleotide sequences.

**By-passing Immunization and the Immune System**

It is difficult to make human monoclonal antibodies for therapy, particularly against self-antigens such as cell surface markers (see Ref. 10 for review). Display and selection of antibodies on phage offer a process for making human antibodies of high affinity and specificity entirely in vitro. (The complete strategy (making repertoires of rearranged V-genes, cloning these for display on phage, selecting the phage with antigen, and increasing the binding affinities by further rounds of mutation and selection) is summarized in Fig. 1.) The rearranged V-genes can be harvested from B-cells of unimmunized human donors (34) or can be built directly from unrearranged human (germ line) V-gene segments (45).

Thus, a “natural” repertoire of antibody fragments was built from the rearranged V-genes of peripheral blood lymphocytes of unimmunized blood donors. Alternatively, a repertoire of synthetic antibody fragments (45) was built from a single light chain and 49 cloned human VH gene segments (46) containing a third hypervariable loop of random sequence. Both the “natural” and synthetic antibody repertoires were displayed on the surface of phage as scFv fragments fused to pIII (34). After several rounds of affinity selection, using different antigens with the same repertoire, phages were isolated with binding activity to both hapten and protein antigens, including self-antigens (34, 45). The soluble antibody fragments prepared from the phages were specific in binding to hapten or antigen, with affinities comparable with those of the primary immune response in vivo ($K_d = 10^{-7}-10^{-9}$ M).

In contrast, when a repertoire of mouse Fab fragments taken from the V-genes of the bone marrow was displayed as fusions to pVIII and the phage selected for binding to a hapten, the antibody fragments had poor specificities and affinities (apparent $K_d = 10^{-10}-10^{-12}$ M) (40).

The affinities of antibodies from primary phage libraries can be improved by affinity selection of random mutants. For example, the structure of a human antibody ($K_d = 3 \times 10^{-10}$ M) isolated from the natural phage library and directed against the hapten phenyloxazolone (phOx) was diversified by sequentially replacing the heavy and light chains with repertoires of mutated chains obtained from peripheral blood lymphocytes (chain shuffling). First the light chains were shuffled (while retaining the heavy chain) and an antibody with a 20-fold improved affinity ($K_d = 1.5 \times 10^{-10}$ M) isolated by affinity selection of the mutant phage (36). The heavy chains of the “secondary” antibody were then shuffled (while retaining the third hypervariable loop of the heavy chain and the entire new light chain) and an antibody with a further 15-fold improved affinity ($K_d = 1.1 \times 10^{-9}$ M) isolated by affinity selection of the mutant phage (36). Much of the sequence and structural variation of antigen binding sites is encoded by the third hypervariable loop, which is located at the center of the antigen binding site. By retaining it, while replacing the other loops, the structure was diversified without disrupting the key features of the antigen binding site. Indeed, both heavy and light chains of the shuffled antibodies were derived from the same germ line V-genes as the primary antibody (36) but with different somatic mutations. Thus the repertoire of rearranged V-genes from unimmunized donors thus provides a rich source of genetic diversity.

The affinities of these human antibody fragments made on bacteriophage are similar to hybridoma antibodies derived from mice immunized with the same hapten three times (tertiary immune response) (Fig. 3) and yet were created entirely without immunization. In the future, the use of synthetic repertoires of V-gene segments, rearranged and mutated in vitro rather than in vivo, should allow the construction of high affinity human anti-bodies entirely outside the immune system.

Using phages to make human antibodies should not only allow the construction of anti-self-antibodies but antibodies that are toxic to mammalian cells, for example by binding or activating receptors. However, human antibody fragments made in bacteria would generally need to have effector functions to mediate cell killing. For example, the fragments could be recloned for mammalian expression as complete glycosylated antibodies with native disulfide bonds (36).

**Fig. 3. Affinities of anti-phenyloxazolone antibodies created in vivo or in vitro. Affinity constants ($K_d$) of monoclonal antibodies for the hapten phenyloxazolone are shown grouped according to the number of times the mice were immunized (primary, secondary, and tertiary immune responses) before fusion of the spleen cells (32). The affinities are compared with those of human antibody fragments derived from phage repertoires made without prior immunization (34, 36) and grouped according to the number of repertoires created (primary, primary repertoire, secondary, light chain shuffled repertoire, and tertiary, heavy chain shuffled repertoire). □, hybridoma antibody; ●, phage antibody.**
ural effector functions (47) or as bispecific F(ab)'s; fragments in bacteria, in which one arm of the antibody is directed against antigen and the other arm recruits novel effector functions, such as triggering cytotoxic T-cells (48).

By-passing Antibodies for Binding and Catalysis?

Several other proteins, including human growth hormone (20), alkaline phosphatase (18), and bovine pancreatic trypsin inhibitor (44), are based on a barrel structure, but they are not antibodies. These proteins are known to engage in applications for which antibodies are not well adapted. Thus, the barrel structure seems more realistic since antibodies have not evolved to facilitate catalysis. For example, the catalytic residues of an enzyme could be retained and loop residues mutated to alter specificity (53). Alternatively, "catalytic residues," for example histidine or carboxylate residues, could be introduced into repertoires of binding sites. New enzymes might be selected by binding of substrate, product or transition states (as with catalytic antibodies (54,55)), and also directly by the catalytic mechanism, for example by reaction with suicide inhibitors (56). A technology that can create human antibodies de novo may be capable of creating enzymes de novo.

REFERENCES

1. Winter, G., Foy, J.-R., and Milstein, C. (1982) Nature 299, 578-578.
2. Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1992) J. Mol. Biol. 229, 301-310.
3. Marks, J. D., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1991) Nature 326, 642-645.
4. Marks, J. D., Griffiths, A. D., and Milstein, C. (1985) Nature 316, 412-418.
5. Marks, J. D., and Milstein, C. (1987) Immunol. Rev. 94, 23-41.
6. Marks, J. D., et al. (1991) in Antibody Engineering: A Practical Approach (Bor-}

10. Foy, J.-R., and Milstein, C. (1982) Nature 302, 713-718.
11. Foy, J.-R., Griffiths, A. D., and Milstein, C. (1989) Nature 339, 577-583.
12. Foy, J.-R., and Milstein, C. (1989) in The Enzymes (Sigman, W. S., ed) pp. 237-253.
13. Foy, J.-R., and Milstein, C. (1989) in The Enzymes (Sigman, W. S., ed) pp. 237-253.
14. Foy, J.-R., and Milstein, C. (1989) in The Enzymes (Sigman, W. S., ed) pp. 237-253.