Antibody Engineering to Study Protein-Ligand Interactions and Catalysis: The Phosphorylcholine Binding Antibodies

reprinted from:
A. Plückthun, Antibody Engineering to study protein-ligand interactions and catalysis: The phosphorylcholine binding antibodies. In: Bioorganic Chemistry Frontiers, H. Dugas and F. P. Schmidtchen, Eds. (Springer Verlag, Berlin, 1993), vol. 3, pp. 25-66.

Andreas Plückthun[1]

Max-Planck-Institut für Biochemie, Protein Engineering Group,
Am Klopferspitz, D-8033 Martinsried, FRG

Antibodies binding the small hapten phosphorylcholine are an ideal model system for investigating the structural logic of the selection of only a few genes in the immune response to this antigen, as well as the structural and energetic consequences of point mutations. Using bacterial expression technology, originally developed with this model system, a large number of mutations have been analyzed with F_v and F_{ab} fragments of these antibodies. While the hapten is very small and only few amino acids are in direct contact, residues up to 20 Å away can have a dramatic effect on antigen binding. This is probably due to indirect conformational effects on the residues that do make direct contacts, giving these interactions improved geometries. These antibodies are able to catalyze the hydrolysis of choline carbonate esters, and the structure-function relationship of this reaction can be studied, because of the known crystal structure of one of the antibodies.

1 General Introduction 26
2 Antibody Structure 27
2.1 Overview 27
2.2 Modeling 29
3 Antibody Genes 32
4 Manipulating an Antibody by Genetic Engineering 38
5 The PC-System 41
5.1 General Overview 41
5.2 The Binding Site of M603 46
6 Mutational Analysis of the Anti-PC Antibodies 50
7 Catalysis 56
8 Avidity 59
9 Conclusions and Future Approaches 62
10 References 63

[1] Address after Oct 1993: Biochemisches Institut, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland.

Bioorganic Chemistry Frontiers, Vol. 3

1 General Introduction

The problem of binding a substance selectively ("host-guest chemistry") has been an enormous challenge in organic chemistry. It requires understanding of the principal forces involved, their relative magnitudes, and their dependence on geometry. Furthermore, it requires being able to manipulate the "host" structure with sufficient facility to actually produce a molecule commensurate with the desired specifications. In nature, this problem has been solved very elegantly, and for the bioorganic chemist, it may be worthwhile analyzing nature's strategy at some depth.

The binding molecules (the "hosts") to be discussed are antibodies. These are used by higher animals to neutralize foreign invaders, mostly bacteria, viruses and other parasites. In many instances, the animal will never have seen the antigen (the "guest") before. Essentially the same set of reactions is triggered by artificially "immunizing" an animal. The animal will keep descendents of a few of these antibody producing cells as memory cells and respond much faster the next time the same antigen is recognized.

From the perspective of the bioorganic chemist, several features can be delineated which show why this system works so efficiently:

1. The antibody generation is governed by an automatic selection process. Each B-cell produces only one kind of specificity. A cell that binds an antigen onto its surface then proliferates and is converted into an antibody secreting plasma cell.
2. The selection is done stepwise. To avoid the screening of an astronomic number of different molecules, on the order of perhaps 10^8 primary antibodies (Holmberg et al., 1986; Klinman and Press, 1975) are assembled entirely randomly. Promising ones (as defined by initial binding) are "refined" by random somatic mutation and selection.
3. The primary diversity is generated by combinatorial approaches. Antibodies consist of light and heavy chains, and the binding sites in each chain are again assembled from 2 or 3 parts, respectively (see below). Thus, random combinations of the *sum* of all the parts (perhaps a thousand) leads to the *product* of different molecules (perhaps 10^8–10^{10}).
4. As in all biochemistry, the "host" molecule is economically assembled from prefabricated building blocks in the form of the 20 different amino acids. The variation of the binding site is generated not only by changing the chemical nature of a particular building block but also by changing its position in space by changing any of the other building blocks of the protein structure.

Even so, antigen binding may not reach the conceivable maximum. Most likely, the selection pressure stops when a dissociation constant of perhaps 10^{-11}–10^{-13} M has been reached. Higher binding constants have rarely been observed, and there may not be any gain from binding the antigen any better, although the large contact surface might in principle allow this for a protein antigen. Nevertheless, these numbers are, of course, a steep challenge for the host-guest chemist who traditionally deals with binding affinities 9 orders of magnitude lower. The efficient binding is probably not achieved because every distance and every angle has been optimized. Rather, especially for large protein antigens, there is a sufficiently large number of interactions over a large area and no bad contacts. Some of the interactions may contribute a more significant part

of the binding energy than others, because their geometries of interaction may fit better than other ones.

This article will summarize research aimed at understanding the bioorganic basis of antibody-antigen recognition. New technologies of manipulating antibodies now allow the direct test of these binding contributions. After some introduction of antibody structure and genes, the analysis of one particular immune response will be summarized: that to microorganisms which carry the small molecule phosphorylcholine at their surface.

2 Antibody Structure

2.1 Overview

Antibodies of most classes consist of 4 chains, two identical light chains and two identical heavy chains (Fig. 1, Fig. 2). The antibody molecule is made up from homologous domains, all forming β-barrel structures. The constant domains consist of two sheets with 4 and 3 antiparallel strands, while the variable domains, which make contact with the antigen, contain an insertion, leading to another pair of strands. Thus, the variable domains consist of two sheets of 4 and 5 strands (Fig. 3). The two sheets making up the barrel are connected by a conserved disulfide bond (Fig. 2, Fig. 3).

From the comparison of many variable domain sequences (Wu and Kabat, 1970; Kabat et al., 1991) it became apparent early on that there are three regions of hypervariability in each variable domain, termed the complementarity determining regions (CDR). X-ray structure analysis later showed that these are situated in three loops connecting the framework (for recent reviews see e.g. Alzari et al., 1988; Davis et al., 1990). The antigen binding occurs largely by contacts to residues in these three loops in each chain connecting the strands of the β-sheet. The typical variable domain has a topology as in Fig. 3.

CDR1 arches from one sheet (the one distal to the dimerization interface) to the one proximal to the interface. Because of the β-barrel folding topology of the domain, this type of loop is often referred to as a greek key loop. In the light chain, the whole loop is variable, while in the heavy chain, the hypervariable portion is usually in the extended β-strand. CDR2 connects two adjacent strands, and the hypervariable portion of this is part of the extended strand and the turn in the case of the light chain, but two adjacent pieces of strand connected by a variable turn in the case of the heavy chain. CDR3 also connects two adjacent strands with a very variable loop. This CDR can be very extended-like but also deviate from β-strand geometry, and the variability is highest in the heavy chain. CDR3 of the heavy chain is the greatest challenge to modelling, but unfortunately, because of its central location, it has a pivotal role in antigen

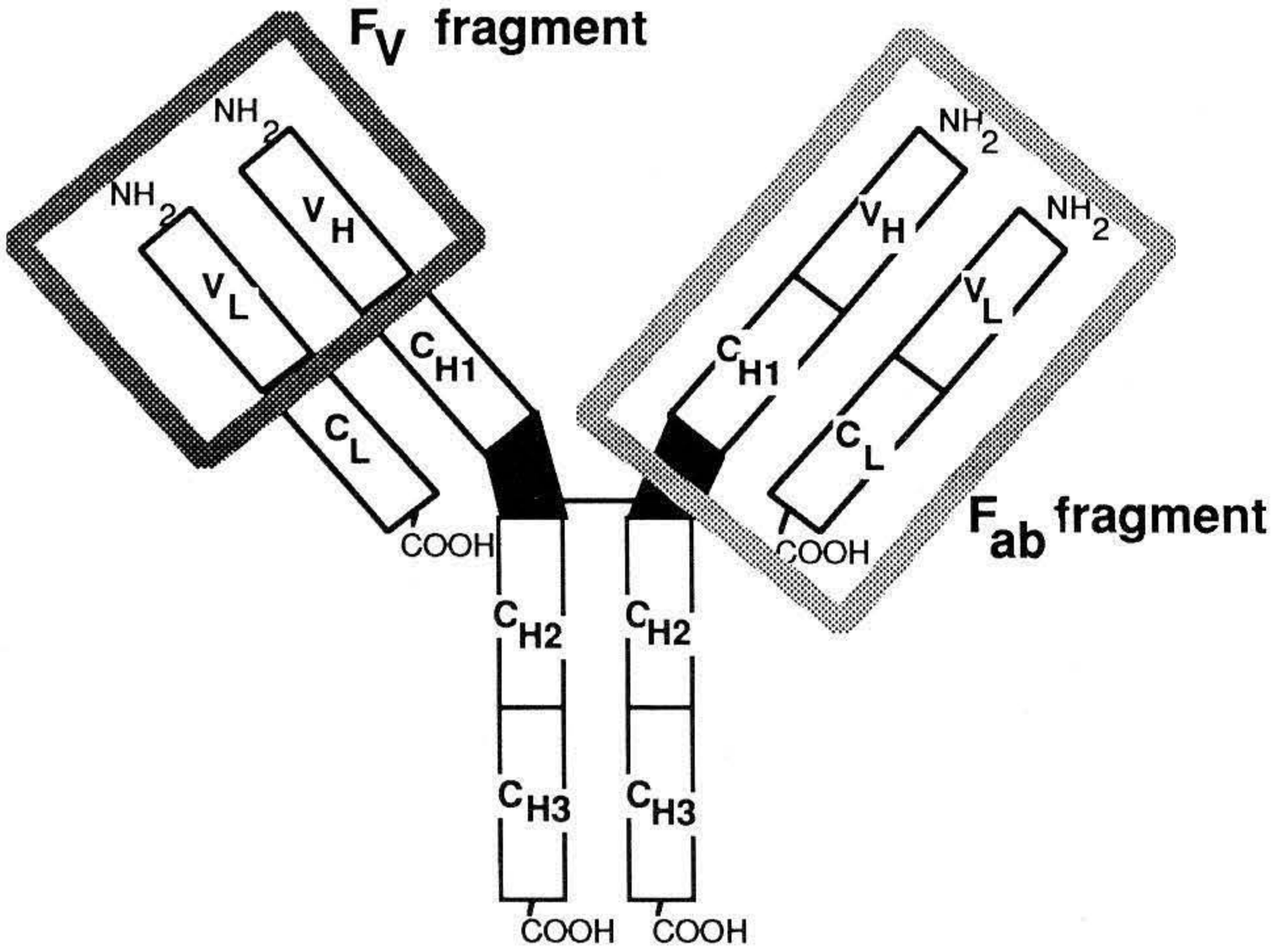


Fig. 1. Schematic structure of an antibody and its antigen binding fragments. The domains are labeled: *V*, variable; *C*, constant; *H*, heavy chain; *L*, light chain. The F_{ab} fragment can usually be obtained by limited proteolysis in the hinge region (*black*), whereas the F_v fragment can usually not be obtained by proteolysis. The F_v fragment is the smallest fragment still containing the complete binding site

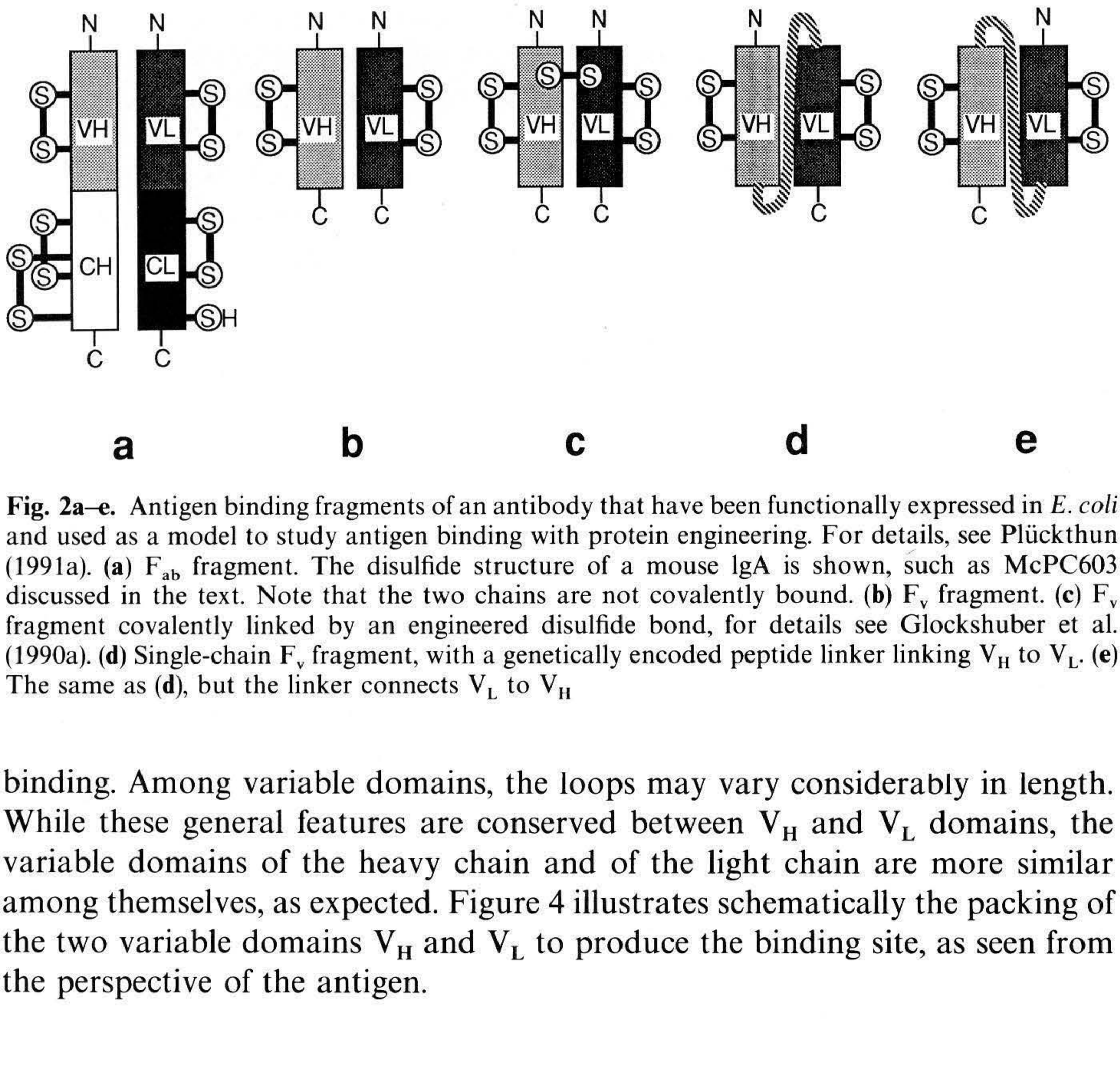


Fig. 2a–e. Antigen binding fragments of an antibody that have been functionally expressed in *E. coli* and used as a model to study antigen binding with protein engineering. For details, see Plückthun (1991a). (**a**) F_{ab} fragment. The disulfide structure of a mouse IgA is shown, such as McPC603 discussed in the text. Note that the two chains are not covalently bound. (**b**) F_v fragment. (**c**) F_v fragment covalently linked by an engineered disulfide bond, for details see Glockshuber et al. (1990a). (**d**) Single-chain F_v fragment, with a genetically encoded peptide linker linking V_H to V_L. (**e**) The same as (**d**), but the linker connects V_L to V_H

binding. Among variable domains, the loops may vary considerably in length. While these general features are conserved between V_H and V_L domains, the variable domains of the heavy chain and of the light chain are more similar among themselves, as expected. Figure 4 illustrates schematically the packing of the two variable domains V_H and V_L to produce the binding site, as seen from the perspective of the antigen.

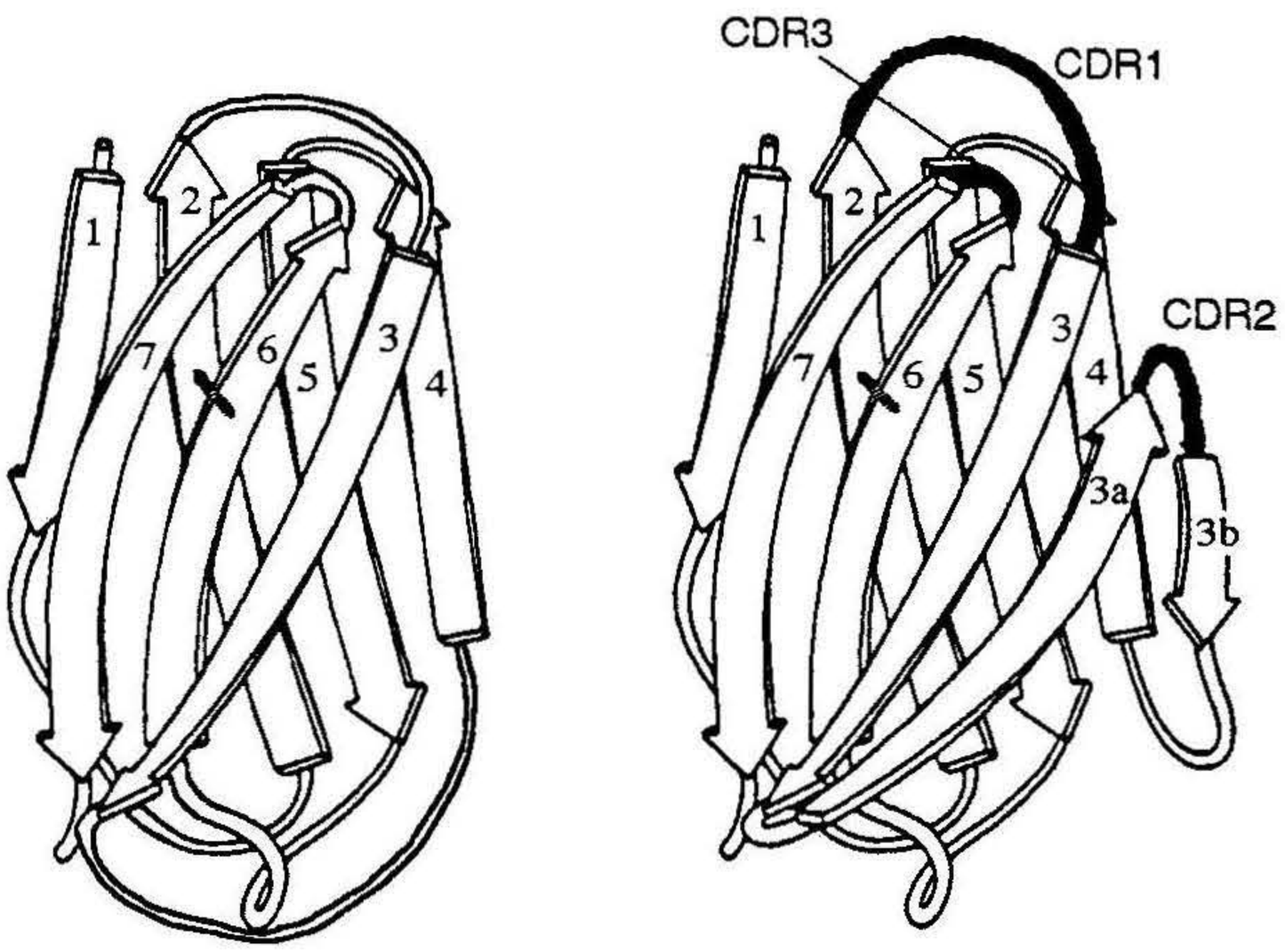


Fig. 3. Topology of antibody constant (left) and variable (right) domains. The two additional strands in the variable domains can clearly be seen. The front face of the domain in this view is the dimerization interface. This figure has been adapted from Branden and Tooze, 1991

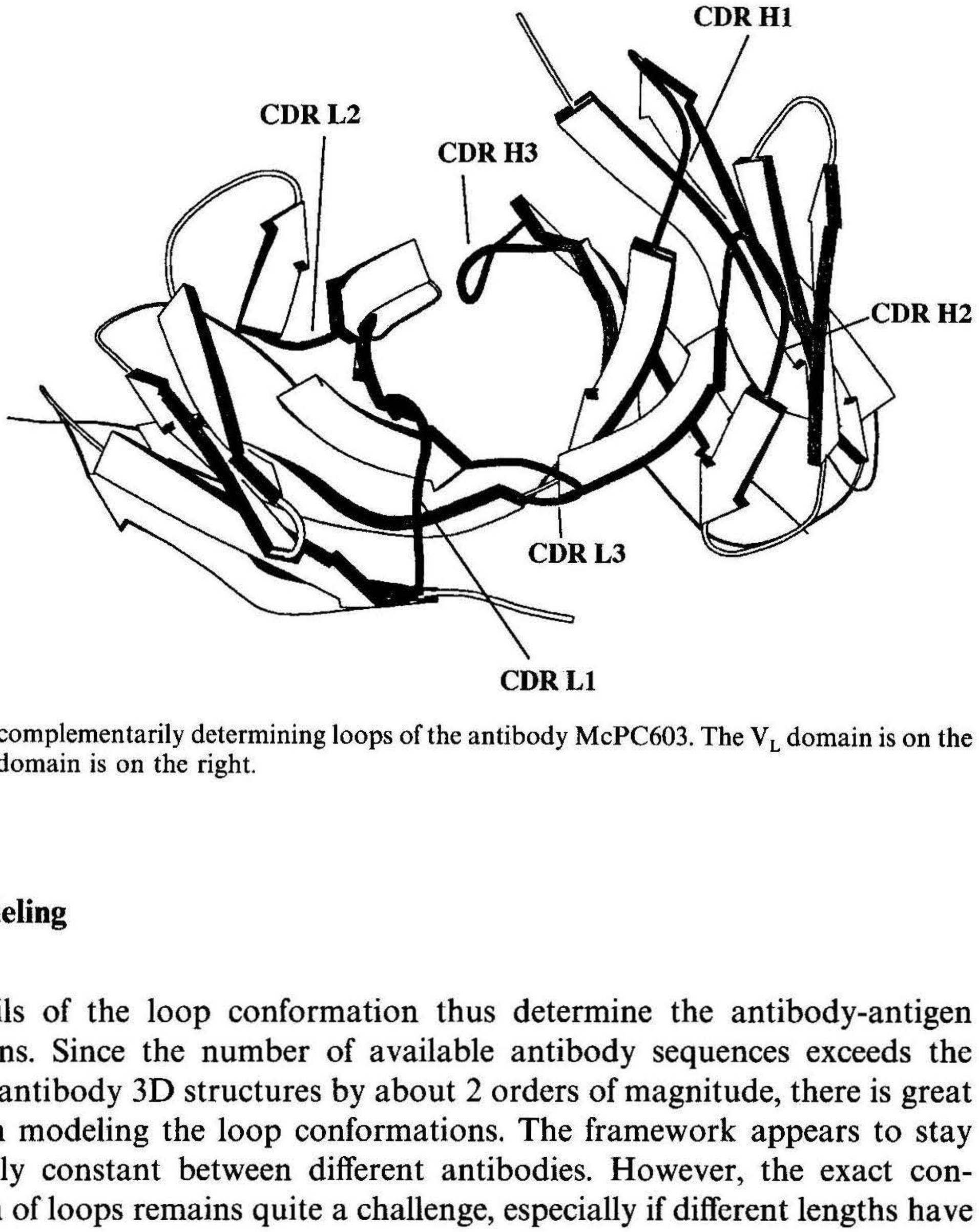


Fig. 4. The complementarily determining loops of the antibody McPC603. The V_L domain is on the left, the V_H domain is on the right.

2.2 Modeling

The details of the loop conformation thus determine the antibody-antigen interactions. Since the number of available antibody sequences exceeds the available antibody 3D structures by about 2 orders of magnitude, there is great interest in modeling the loop conformations. The framework appears to stay remarkably constant between different antibodies. However, the exact conformation of loops remains quite a challenge, especially if different lengths have

to be considered and no similar precedent is available. Furthermore, framework residues are known to interact with the antigen in some cases, and also with the loops, and thus they can determine the loop conformation (see below). Thus, while the framework structure can be predicted with good accuracy, its sequence can influence the binding properties of the antibody directly, and indirectly via the loop conformation.

The first strategy of predicting the loop conformation is to use the known structures as a database. Jones and Thirup (1986) have developed efficient search procedures for loops in the database of known structures. Chothia, Lesk, and coworkers have applied this strategy to antibodies (Chothia and Lesk, 1987, Chothia et al., 1989) and put forward the interesting hypothesis that each of the CDRs can only take on a limited number of conformations, the so-called "canonical structures". Thus, the observed conformations can be clustered, and key residues identified that might determine the membership in a cluster. However, the number of experimental structures is still fairly limited, and one must treat this hypothesis with some degree of caution, because (1) so far, too few examples are available to test the generality of this hypothesis, (2) several lengths of CDRs known from sequences have not yet been observed structurally, (3) some canonical structures may be incorrect due to low resolution data. An example is the originally proposed structure for CDR2 of the light chain in McPC603, different from the other canonical structures. This is probably due to an incorrect orientation of a peptide bond at the resolution available from the original crystals (Steipe et al., 1992).

Nevertheless, the classification of loops by Chothia, Lesk and co-workers has been very useful in practical modeling approaches, and their proposal of only very few critical interactions determining the canonical structure of a loop lends itself to experimental testing. Notably, critical residues have been identified both in the framework and in the loops (Chothia and Lesk, 1987; Chothia et al., 1989, Tramontano et al., 1990) that must be considered in any engineering experiment, such as e.g. the "humanization" of antibodies (Jones et al., 1986; Riechmann et al., 1988; Verhoeyen et al., 1988; Queen et al., 1989; Co et al., 1991; Gorman et al., 1991; Kettleborough et al., 1991; Maeda et al., 1991; Tempest et al., 1991). In experiments of this type, the CDRs from a mouse monoclonal antibody are grafted onto a human framework, with the intention of minimizing the antigenicity of the antibody in therapy, since otherwise there is a human anti-mouse response. While it has been shown for a number of cases that the binding properties can be transplanted using this strategy, it has also been observed in most cases that a number of the human framework residues have to be altered as well to restore the same binding affinity in the "humanized" antibody. This is partly because of direct interactions of framework residues with the antigen and partly because of framework residues influencing the CDR conformations. As the theoretical understanding of these effects is currently rather modest, empirical investigations of framework mutations are necessary. Clearly, structural work on loop-grafted antibodies would be required (Steipe et al., 1992).

The second strategy for predicting antibody conformations involves the use of force fields. The building of loops using either a systematic generation of a family of conformations (Bruccoleri and Karplus, 1987; Bruccoleri et al., 1988) or a random sampling (Fine et al., 1986; Shenkin et al., 1987) has been attempted. The loops are then closed by a known algorithm (Go and Sheraga, 1970) and the potential energy of each conformation is evaluated. Remarkably, the low energy structures of some short loops coincided well with the experimentally observed structure, even when loops only consisting of alanines were used (Fine et al., 1986; Shenkin et al., 1987). It appears that these methods are overall comparable in performance to the canonical structure approach. Recently, both approaches have been combined (Martin et al., 1989).

Molecular dynamics (see e.g. van Gunsteren and Berendsen, 1990) may be another possibility for sampling the conformational space. Extended molecular dynamics calculations have been carried out for the system of phosphorylcholine antibodies to evaluate this methodology (Köhler et al., 1991). The methods of simulated annealing (Kirkpatrick et al., 1983) may be especially appropriate for these conformational search problems. The prediction problem will fail with any energy-based method if the current force fields are inadequate, but the problem of sampling (which should be concentrated to relevant regions of conformational space, but be complete there) appears to be a more severe one. Structural features are usually reproduced adequately in molecular dynamics simulations (van Gunsteren and Berendsen, 1990), when an experimental structure is used as a starting point. This suggests that the quality of the force fields is reasonable, although it remains questionable whether energetic criteria are suitable for evaluating widely diverging loop conformations (Novotny et al., 1988; Holm and Sander, 1992).

Besides the general questions of the peptide backbone conformations in the loops, it is necessary to be able to predict side chain torsional angles, in order to evaluate the interactions between side chains and the antigen, and to evaluate contacts within the antibody structure. If the antigen is a peptide or protein, the same problem arises for the antigen as well. In a number of studies it could be shown that there are preferences for only few side-chain torsion angles, although this preference is not absolute (Janin et al., 1978; James and Sielicki, 1983; Summers et al., 1987; Summers and Karplus, 1989). Clearly, more structural work is necessary to delineate the relative importance of narrow torsional preferences versus additional free energy from side-chain interactions and strain or loss of entropy.

2.3 Binding Site

The binding site of an antibody can vary very much in size and shape. Generally, the binding site for small molecules is more pocket or groove-like, often brought about by a long CDR1 of the light chain. The contact area to protein antigens is

much flatter, and may even have protruding amino-acid side chains pointing into the active site of an enzyme antigen, as seen in the case of the anti-lysozyme antibody HyHel10 (Davies et al., 1988, 1990). The reported surface areas buried by the antigen range from 161 $Å^2$ (for phosphorylcholine) to 886 $Å^2$ (for neuraminidase) (Wilson et al., 1991). The small antigens are largely buried; for instance, phosphorylcholine is 81% covered by the antibody.

In even the smallest antigens, 4 out of the 6 hypervariable loops take part in binding interactions, while with protein antigens, 5 or even all 6 CDRs are used (Davies et al., 1990; Wilson et al., 1991). For the small antigens (haptens), CDR2 of the light chain is apparently not used. Its conformation seems to be conserved in the known mouse kappa chains (Steipe et al., 1992). Perhaps, too little variability can be generated in CDR2 of the light chain by variations in the sequence at constant length to reach into the binding site in order to participate in interactions with small substrates. Both light and heavy chain contribute to binding (as evidenced by the buried surface area), but the contribution of the heavy chain is usually somewhat larger. Yet, while there have been reports of V_H domains binding the antigen by themselves (Ward et al., 1989), it remains doubtful whether this will turn out to be a general phenomenon. The degree of selectivity is smaller than for complete antigen binding sites (Berry and Davies, 1992). Furthermore, it must be established in each individual case that any observed binding by V_H alone is specific for the antigen in question and not due to a nonspecific interaction of many proteins with the V_H–V_L dimerization interface. This interface is where V_L would normally bind, but it becomes accessible in single V_H domains.

3 Antibody Genes

The enormous variability of the immune response, enabling the animal to handle almost any conceivable antigen, is due to a surprisingly small number of genes. The reason why this is possible is that nature uses combinatorial approaches to their fullest extent. The variable domain, while being a single structural unit, is not made from a single gene. Rather, only the major part of the domain is encoded by the "V" gene. In the case of the light chain, the V_L gene is fused to a J gene. Both together code for the complete structural domain. In the case of the heavy chain, the γ segment gene is first fused to a D segment, which in turn is fused to the V_H gene, and only all three together form the complete coding information of the structural domain (Fig. 5). Via splicing of the mRNA, the variable domain is then linked to a constant domain, the type of which can be "switched" in the case of the heavy chain during the development of the B-cell (Esser and Radbruch, 1990).

For the antibody light chains of the mouse, there are on the order of 10^2 V_L genes, and 4 functional J_L segments in the κ-locus (Ponath et al., 1989). The λ-light chains are only rarely used in the mouse. For the heavy chain, there are approximately 10^2–10^3 V_H genes, a total of 12 D segments (Kurosawa and Tonegawa, 1982; Ichihara et al., 1989) and 4 J_H segments (Solin and Kaartinen, 1992) in the genome of the mouse. While the genetic and biological background of the immune response is not the topic of this account (and there are excellent

Fig. 5a, b. Assembly of antibodies from their genes. (**a**) The situation in the mouse light chain κ-locus is shown schematically. Each V gene is preceded by another exon encoding a signal sequence, but this is not shown. (**b**) The situation in the mouse heavy chain locus is shown schematically. In reality, there are 4 Cγ-genes (γ1, γ2a, γ2b, γ3). A further rearrangement ("class switching") can couple the VDJ-gene to C-genes other than Cμ, giving rise to other antibody classes such as IgG (not shown). As in the light chain, each V gene is preceded by a separate exon encoding the signal sequence (not shown)

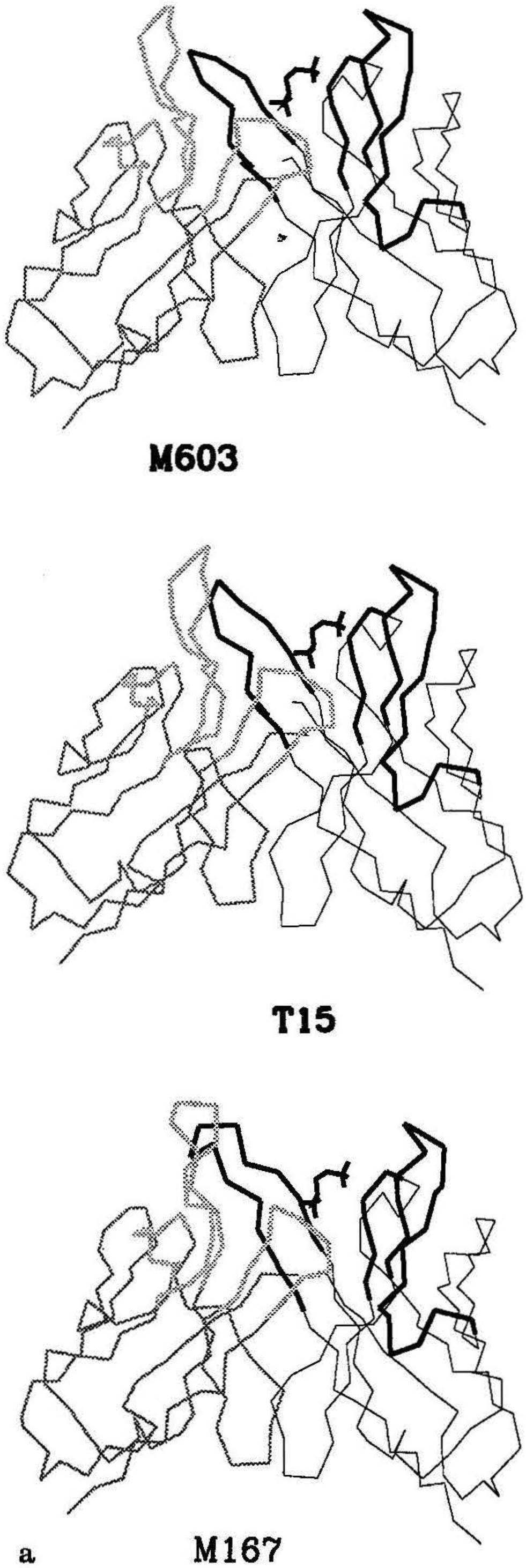
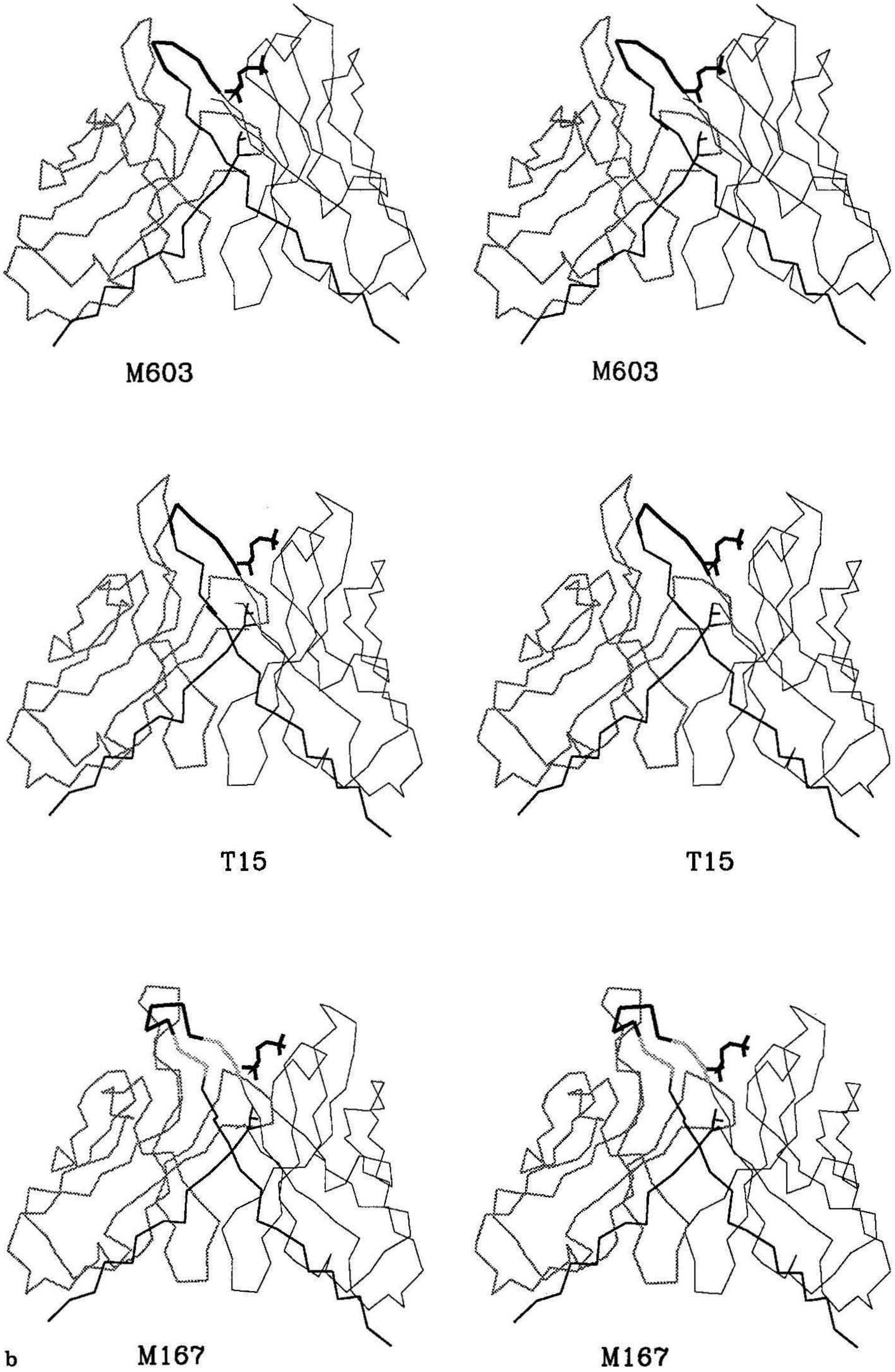
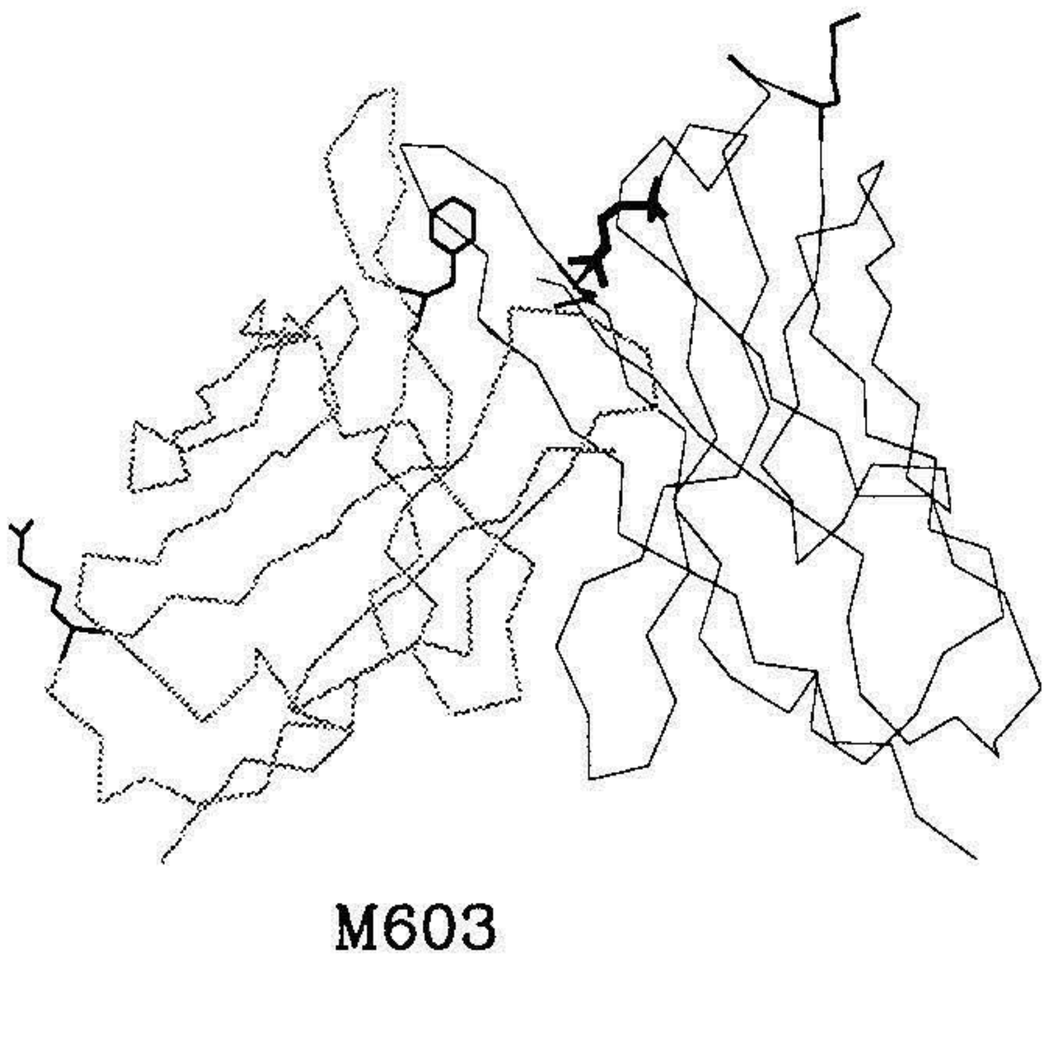


Fig. 6a–c. Stereoviews of the structures of the F_v fragments of M603, T15 and M167. The structure of M603 is derived from the crystal structure, while the others are modelled based on this structure. The heavy chain is shown in *black* (*right*), and the light chain in *grey* (*left*). The antigen phosphorylcholine is shown with *thick lines*. (**a**) Hypervariable loops according to the definition of Kabat et al. (1987), emphasized by thick lines. (**b**) Location of the genetic elements in the structure. The D-element is shown with thick black lines, and non-coded amino acids (N-region) with *thick light-grey lines*. The J-elements are shown with *medium black lines*, for the heavy chain on the right and the light chain on the left. (**c**) Location of the somatic mutations in the three structures

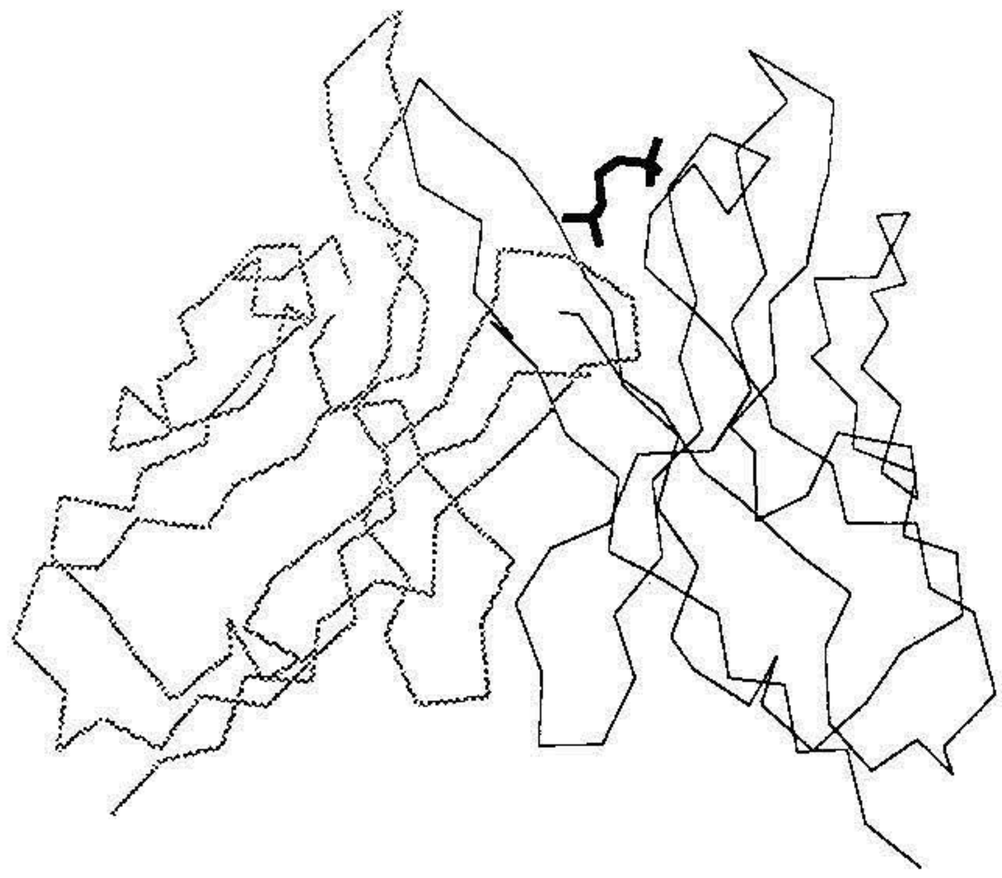


b

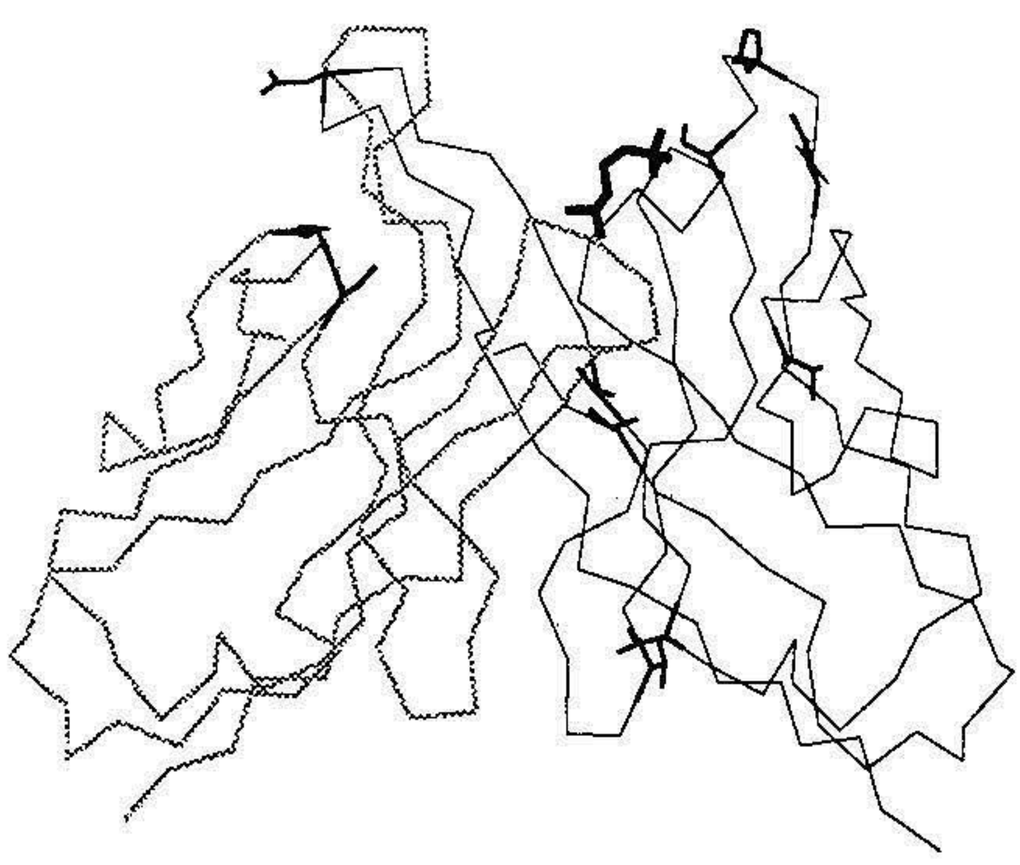
Fig. 6b.



M603

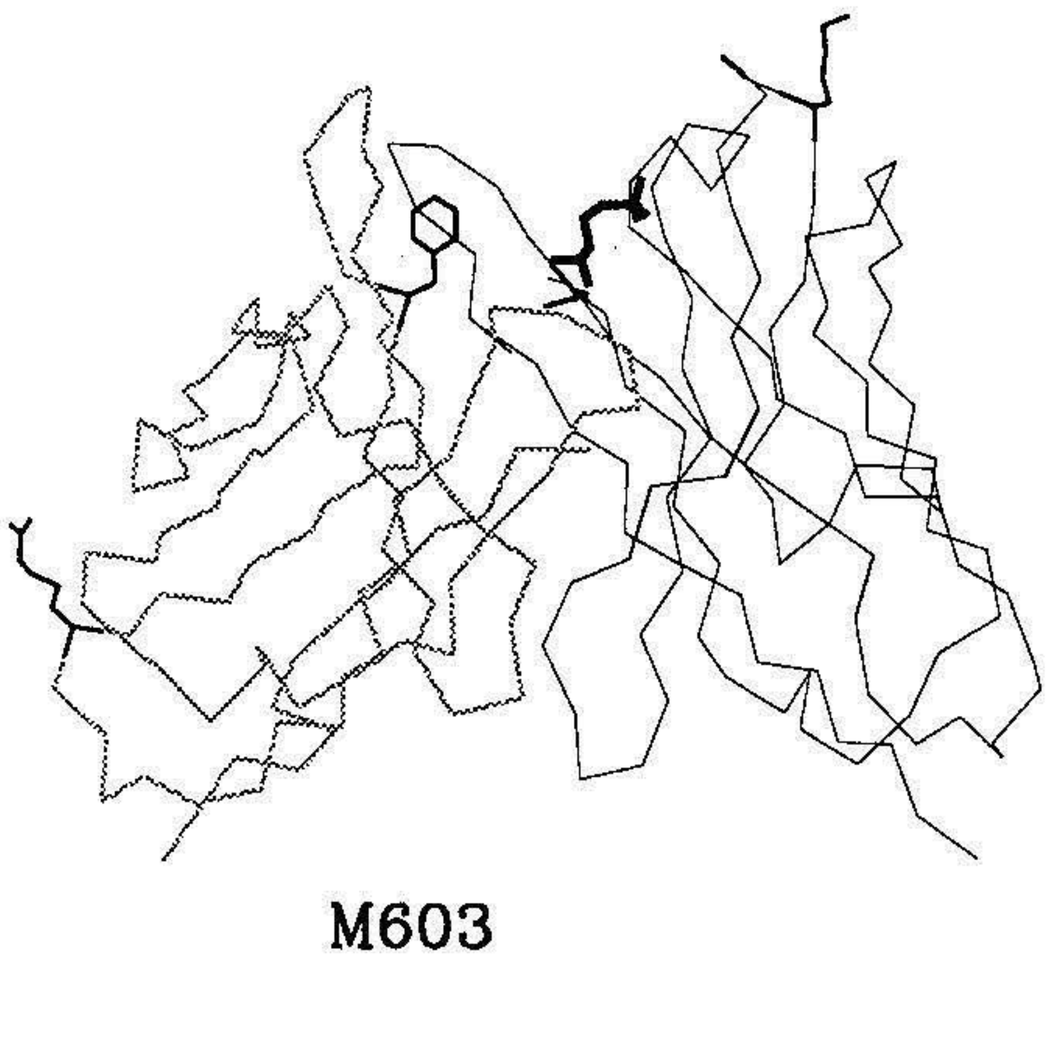


T15

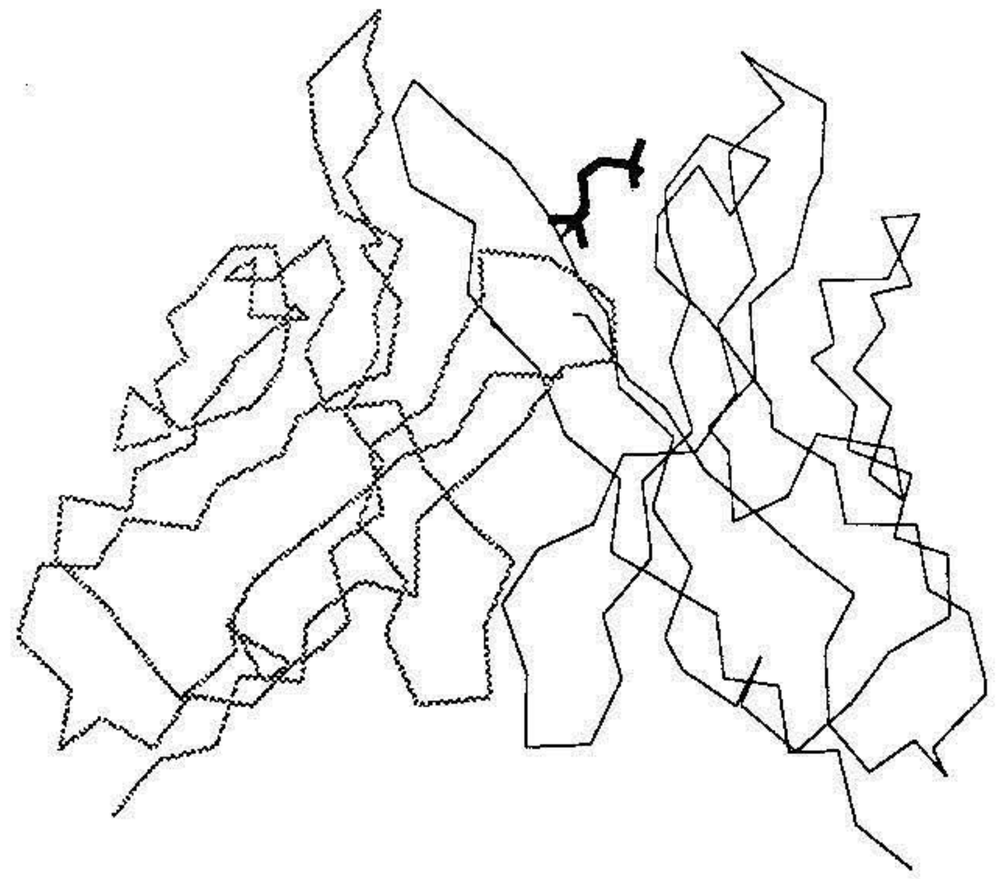


M167

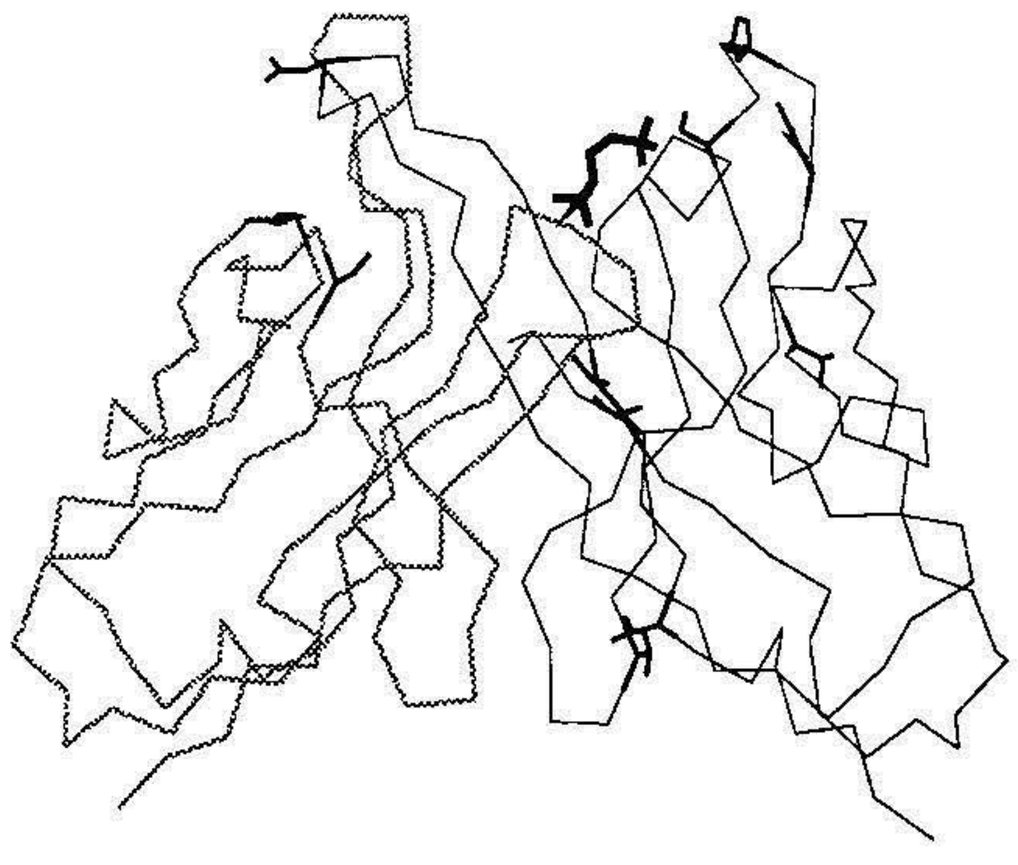
c



M603



T15



M167

Fig. 6c.

overviews [see e.g. Calabi and Neuberger, 1987]), a few points relevant for understanding the structural consequences for antibody combining sites must be mentioned. The line of cells giving rise to antibodies, the B-cells, mature from the so-called pre-B-cells. During this process, the DNA is rearranged, i.e. a particular heavy chain V_H element is joined with a particular D and J element on the genomic level. Similarly, the light chains are rearranged, resulting in a continuous reading frame for a particular V_L and J_L element. Very importantly, the mechanism of joining the D element to J and V in the heavy chain is deliberately imprecise and involves non-encoded nucleotides (the so-called N-region diversity). While the exact mechanism is not yet clear despite intensive research (reviewed e.g. by Reth and Leclerc, 1987; Yancopoulous and Alt, 1986), a terminal transferase adding a few random nucleotides has been implicated. Often, there also seems to be some "nibbling" by a double-stranded exonuclease

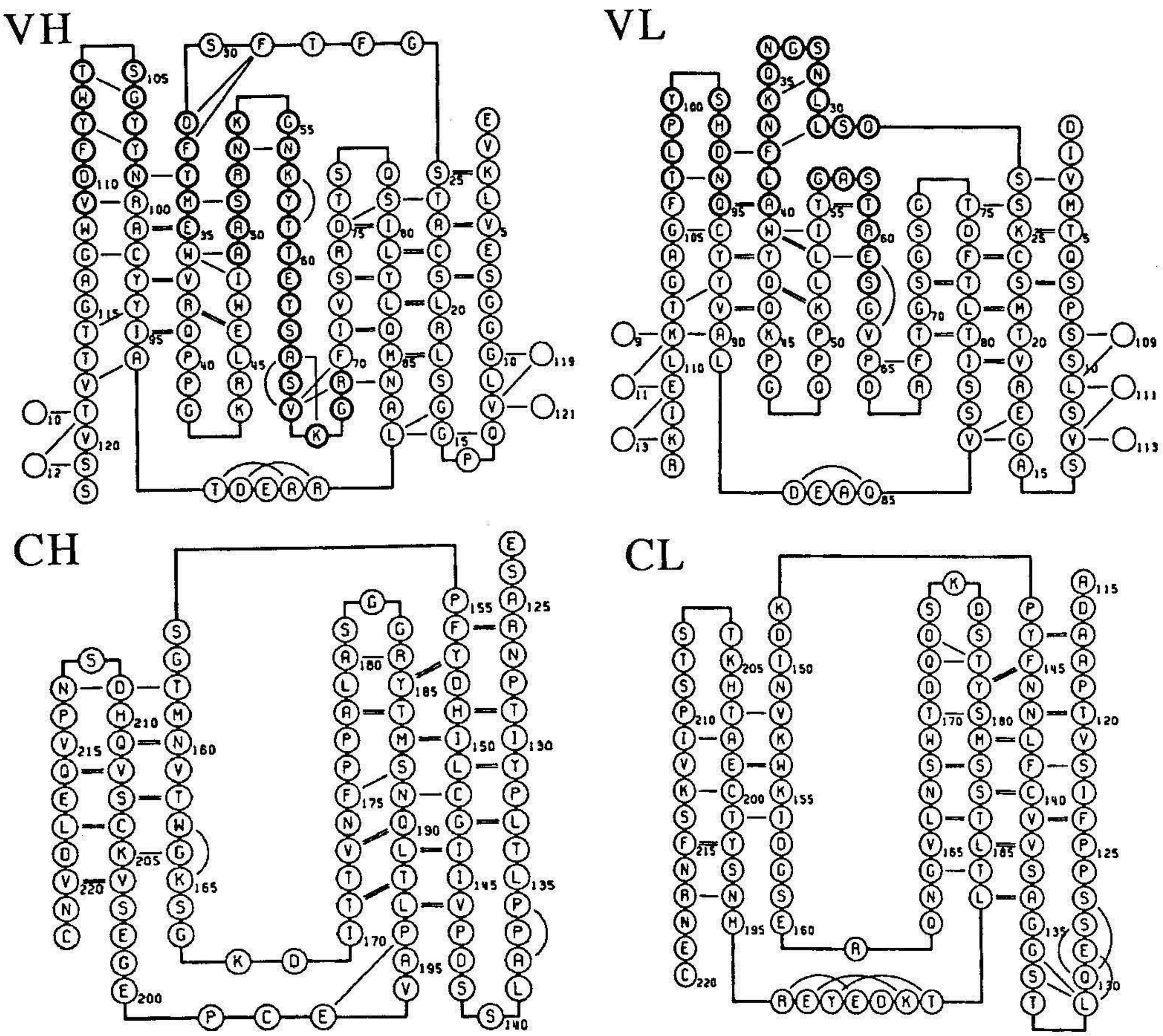


Fig. 7. Secondary structure of the antibody McPC603, according to Satow et al. (1986). The CDR regions have been emphasized by *bold circles. Thick lines* follow the peptide backbone, *thin* (*single or double*) *lines* indicate hydrogen bonds

at either side of the joint. The net effect is very high variability and an enormous variation in *length* at these positions, corresponding to CDR3. The remaining gaps in the reading frame between the signal sequence, the rearranged V(D)J region and the constant domains are removed by splicing of the RNA (Fig. 5).

Figure 6 shows the structural correlation of the three different gene segments for the phosphorylcholine binding antibodies McPC603, TEPC15 and MOPC167 (see below). In the light chain, the J element forms the last β-strand of the domain. This can also be seen in a secondary structure representation of the antibody structure (Fig. 7). In the heavy chain, the D element forms the tip of CDR3, and the J element again forms the last strand of the β-sheet.

What are then the structural consequences of using different D and J elements? Both involve the CDR3 loop. In the heavy chain, the CDR3 loop itself, its length and its sequence is influenced by the end of the V_H gene, the N region diversity between V and D, the sequence of the D segment, the second N-region diversity between D and J and the beginning of the J segment. In V_L, the choice of J may somewhat alter the endpoint of this loop. Since the J regions of both the heavy and the light chain make contacts with the other chain (Fig. 6, Fig. 7), their alteration is likely to adjust the relative orientation of the domains at least slightly. In addition, both J_H and J_L may contribute direct binding interactions to the antigen.

At the present time, it is not clear whether structural restrictions exist between certain combinations of elements. From the available sequence information, a very large number of combinations has been found, but the available sequences do not show whether few or even a significant number of combinations of V, D and J in the heavy chain or V and J in the light chain might be excluded for being structurally unstable. Furthermore, it is not clear yet whether all combinations of light and heavy chains are possible. While the general answer to this question will be difficult to find, the tools of molecular biology now allow the testing of many new or artificial combinations. Notably, the interactions between CDR1 and CDR2, which are not varied independently in nature, as both are encoded on the V gene, can now be tested.

4 Manipulating an Antibody by Genetic Engineering

For engineering an antibody combining site, the use of *E. coli* expression systems (reviewed in Plückthun, 1991a) has initiated many of the recent advances in this field. These advances can be separated into those: (1) that lead to the convenient production and the facile engineering possible by using *E. coli* and (2) those that use the screening and random mutagenesis methods based on using the *E. coli* secretion technology, by which the protein can be produced in the native state (Skerra and Plückthun 1988; Better et al., 1988).

With *E. coli* production systems, the genetic manipulation of antibodies has been very much facilitated. Their handling is much easier, the establishment of a new clone or mutant is much faster and the scale-up can be carried out with simpler bioreactor equipment than in any other type of host organism. The bacterial expression systems are suited for F_v, single-chain F_v and F_{ab} fragments of the antibody and more recently, bivalent molecules have been produced as well (see section 8) (Fig. 2). These fragments contain the complete antigen binding site. Experiments with the recombinant proteins have shown that indeed even the smallest fragments (the F_v fragment) show the same binding affinity to a monomeric hapten as the whole antibody (Skerra and Plückthun, 1988; Glockshuber et al., 1990a). Because of their small size, the F_v fragment or the single-chain F_v fragment are suitable for structural investigations. The F_v fragment is within the size range of a structure determination by NMR (Freund et al., 1993), and may also give rise to better ordered crystals than the larger fragments (Bhat et al., 1990; Glockshuber et al., 1990b). Since at the current time antibody engineering is still largely basic research, feedback about the 3D-structures by experimental structure determination is vital for the further development of this field.

For investigating binding or catalysis, the F_v fragment, the single-chain F_v fragment or the F_{ab} fragment can all be used, and some of the properties will be briefly discussed. The F_v fragments of various antibodies differ in stability. They may reversibly dissociate into V_H and V_L (Glockshuber et al., 1990a). The dissociation constant depends on the particular antibody under study (Hochmann et al., 1976; Horne et al., 1982; Klein et al., 1979; Glockshuber et al., 1990a). For the phosphorylcholine binding antibody McPC603, for instance, it was found to be about 10^{-6} M. This dissociation problem can be counteracted in several ways (Glockshuber et al., 1990a) (Fig. 2). First, the two domains can be covalently linked by chemical means. Second, a disulfide bond can be designed *between* the two variable domains (Huston et al., 1993). Third, a single chain F_v fragment can be used, in which a peptide linker connects the two domains. Forth, the F_{ab} fragment can be used, in which the constant domains C_L and C_H1 contribute association energy between the heavy and light chain (Fig. 2).

The design of the disulfide bond to connect the two chains was carried out using a purely geometrical criterion: Positions were searched by computer that gave a good overlap between the main chain atoms of the two amino acids to be considered and any disulfide bond in the database (Pabo and Suchanek, 1986). Some of the positions thus found were tested experimentally and found to lead to correctly assembled, covalently linked fragments assembled by *E. coli* (Glockshuber et al., 1990a).

A number of different linkers have been reported to connect the two variable domains genetically (reviewed in Plückthun, 1991b; Huston et al., 1993). Also, both the arrangements V_H-linker-V_L and V_L-linker-V_H were found to be successful in the same antibody, and compatible with in vivo and in vitro folding (Glockshuber et al., 1990a; Knappik et al., 1993). Most successful linkers appear to be hydrophilic flexible sequences that can adapt to the protein surface.

The technology of producing antibody fragments in *E. coli* will only briefly be summarized. One may divide the current strategies into those leading directly to folded fragments and those requiring in vitro refolding. The former are based on the secretion of the protein to the periplasm of *E. coli* (the space between the inner and outer membranes), and can be applied to the F_v fragment, the disulfide-linked F_v fragment, the single-chain F_v fragment and the F_{ab} fragment and the bivalent miniantibodies discussed in section 8. The advantage is that these fragments are directly obtained in functional form with this methodology, as they assemble by themselves, and all disulfide bonds can form in vivo (Skerra and Plückthun, 1988; Better et al., 1988). Any strategy aiming at screening of binding activity or even catalytic activity must therefore be based on secretion, as it obviously requires the presence of folded molecules in the bacterial cell, or secretion from the bacterial cell.

An alternative strategy is to refold the recombinant antibody protein from inclusion bodies. While in many instances more total antibody protein per cell can be produced than in the secretory system, the overall success depends largely on the yield of refolding of the recombinant protein (Buchner and Rudolph, 1991).

The expression of recombinant antibodies in eukaryotic cells is also feasible. The production in Chinese Hamster Ovary (CHO) cells has been reported as suitable for large scale fermentation (Page and Sydenham, 1991). The use of non-producing myeloma cells as hosts has also been widely used (Morrison and Oi, 1989). While the yields can be quite high, the establishment of clones takes much longer than in *E. coli*, and large scale production is not trivial. Other developments in recombinant antibody expression include the use of yeast (Horwitz et al., 1988) or baculovirus infected insect cells as hosts (Haseman and Capra, 1990; Putlitz et al., 1990). Both systems have not yet been optimized, and currently offer no particular advantage or spectacular yields. Plants have also been proposed (Hiatt et al., 1989) as hosts, but it is unlikely that they will be suitable for biophysical studies, as their handling is not particularly convenient. While the production of whole antibodies may be carried out better in higher cells, the binding site of the antibody can still be engineered and optimized in bacteria using a fragment and the new sequence then transferred to a whole antibody in a eukaryotic production host.

Bacterial expression of functional antibodies has also permitted the display of functional antibody fragments on filamentous bacteriophages (McCafferty et al., 1990). This is possible since the coat proteins are made as intermediates in the inner membrane, before they coat the phage DNA. Their N-terminus is first exposed to the periplasm and later to the outside of the phage, and a hybrid protein consisting of an N-terminal antibody fragment and the phage coat protein will end up on the phage. The phage thus contains the DNA coding for the antibody inside and the antibody protein on its surface. Phages binding to particular antigens can thus be enriched out of a library and amplified after infection of bacterial cells. This technology (reviewed in Hoogenboom et al., 1992) will be of increasing importance in studying antibodies with random mutagenesis techniques.

5 The PC-System

5.1 General Overview

One of the most extensively studied systems of antibodies has been the immune response to phosphorylcholine. Phosphorylcholine, while a component of the phospholipids phosphatidylcholine and sphingomyelin, normally does not seem to act as an immunogen in this form, as it would lead to an autoimmune response (McNeil et al., 1991; Mercolino et al., 1986). Rather, the natural immunogen related to phosphorylcholine is the surface polysaccharide of a number of rather different microbial species (Potter, 1970; 1971; Potter and Leon, 1968; Potter and Lieberman, 1970, Péry et al., 1974; Claflin et al., 1985). The best studied one is the C-teichoic acid of *Streptococcus pneumoniae*, the structure of which has been elucidated (Jennings et al., 1980) (Fig. 8). It is with this antigen that the immunodominant group has been identified as phosphorylcholine (Leon and Young, 1971). *S. pneunoniae* also possesses a related lipoteichoic acid, linked to diacyl glycerol in the cytoplasmic membrane (Fischer, 1988). Its structure was only recently solved (Behr et al., 1992). The 2 to 8 repeating units appear to be identical as those of the cell-wall polysaccharide (Fig. 8), except that the third ring also carries a PC group, like on the fourth esterified to C6. This lipoteichoic acid is the pneumococcal Forssman antigen (which elicits antibodies in rabbits cross-reacting with sheep erythrocytes), but it is not the PC moiety, which is responsible.

Phosphorylcholine thus occurs esterified to sugar residues in a regular array on the surface of the microbial cell. Phosphorylcholine has been found in a number of other bacteria such as *Proteus morganii* (Claflin et al., 1987), *Lactobacillus acidophilus* (Potter and Lieberman, 1970), and even nematodes (Péry et al., 1974), but the exact chemical structures have not been elucitated. *Lactobacillus* and other PC-carrying microorganisms seem to be normal inhabitants of the murine digestive tract (Potter, 1970; 1971). The murine response to

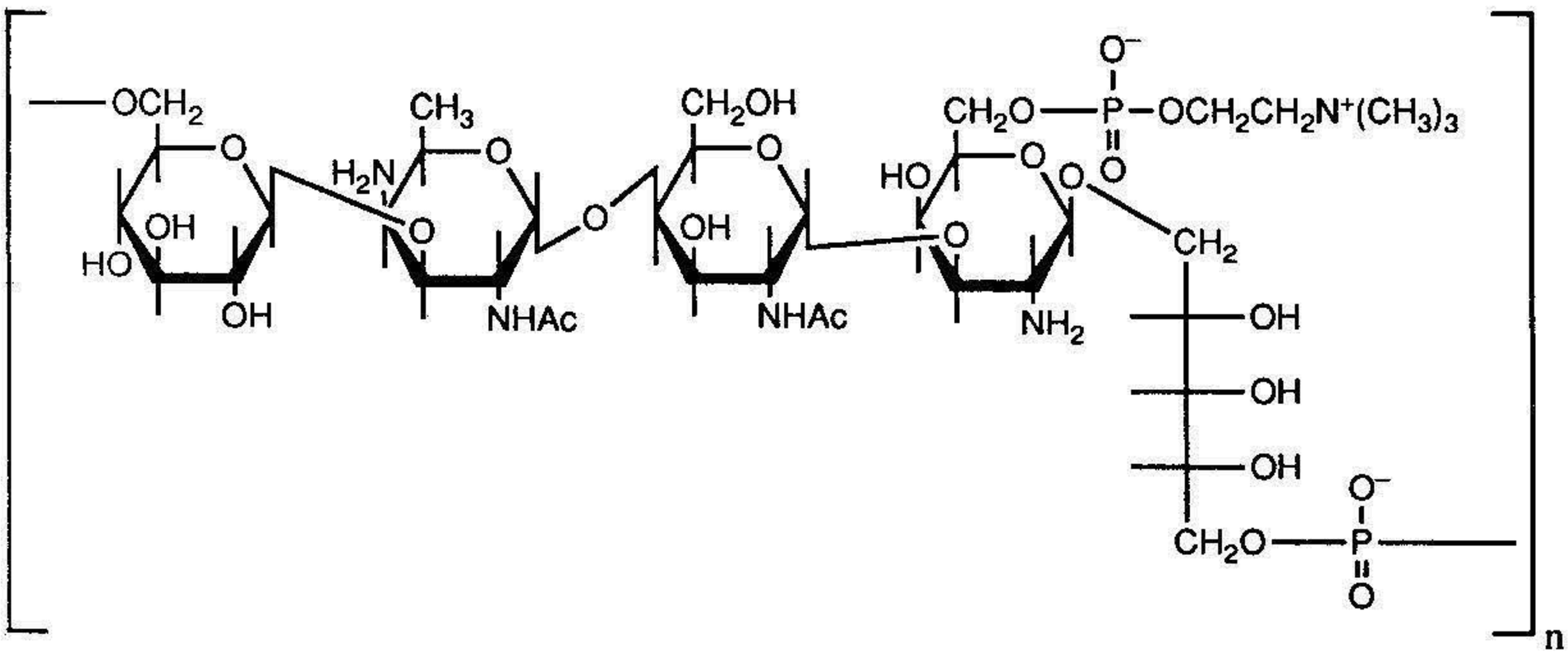


Fig. 8. Structure of the complex cell-wall polysaccharide ("C-substance") from *Streptococcus pneumoniae*

phosphorylcholine indeed appears to be protective against infection by such microorganisms, as antiidiotypic antibodies against the anti-PC antibodies decrease the response to pneumococci (Cosenza and Köhler, 1972).

Most studies on the immune response to phosphorylcholine have used inactivated cells as immunogens (see e.g. Williams and Claflin, 1982; Claflin et al., 1985; 1987; Claflin and Berry, 1988). However, phosphorylcholine has also been used as a hapten derivatized to carrier proteins such as keyhole limpet hemocyanin or other proteins (see e.g. Wolfe and Claflin, 1980; Gearhart et al., 1981; Rodwell et al., 1983). In these cases phosphorylcholine has been esterified as shown in Fig. 9. Both approaches have been compared and were found to give very similar responses, underlining the immunodominant nature of the PC moiety (see e.g. Clarke et al., 1983; Andres et al., 1981).

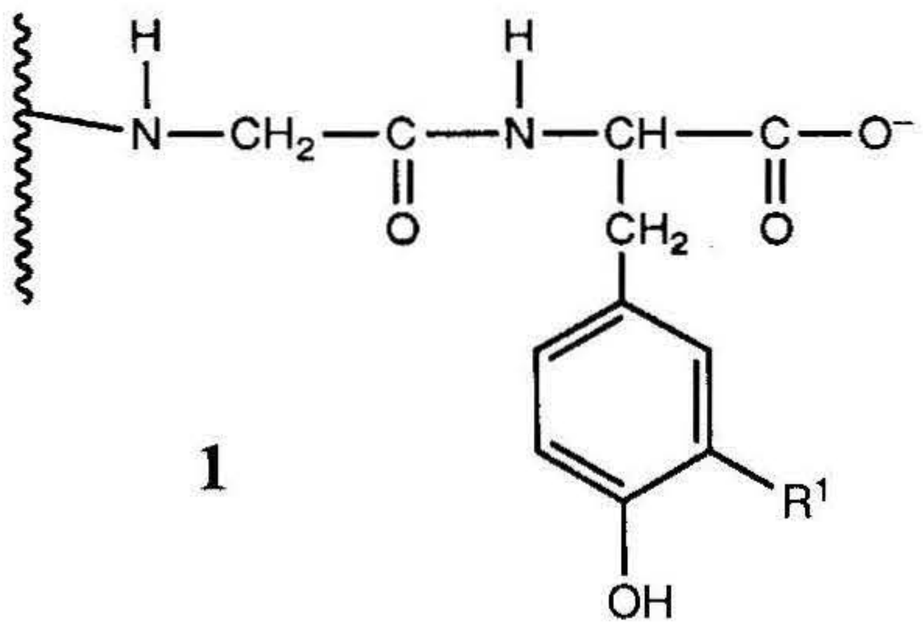
Early advances in understanding the phosphorylcholine response came from the study of experimentally induced myeloma and the isolation of the related proteins (Potter, 1970; 1972). These myeloma proteins have turned out to be regular IgA. It is not completely clear why the anti-phosphorylcholine response was found so frequently in myeloma, since the experimental induction of myeloma does not stimulate the response to any particular antigen, nor is it completely understood why IgA is so prevalent in experimentally induced myeloma. Yet the early availability of chemically homogeneous antibodies with known specificity led to a number of important early biochemical studies, including the determination of protein sequences of these antibodies (Perlmutter et al., 1984; and references therein) and the crystal structure of an antibody combining site (Segal et al., 1974), a scientific milestone.

The murine response to phosphorylcholine is extremely attractive as a model, as it is rather restricted. It is characterized by the use of only one V_H gene (Crews et al., 1981) and the use of three different V_L genes (Claflin et al., 1981). The anti-phosphorylcholine antibodies are exemplified by three well characterized myeloma proteins[1]: McPC603 (or M603), MOPC 167 (or M167) and TEPC 15 (or T15). They use the following three light chain genes $V_{\kappa 8}$ (M603), $V_{\kappa 24}$ (M167) and $V_{\kappa 22}$ (T15). Their aligned sequences are shown in Fig. 10. As the antibody T15 contains the unaltered genomic V_H sequence, it has given the name to the gene, and the whole family.

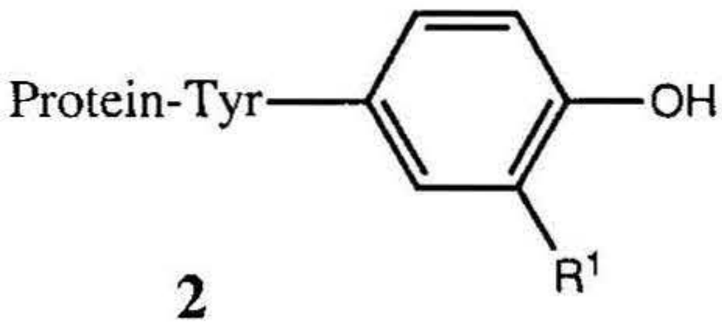
The detailed analysis of the immune response to phosphocholine can now be summarized from a large number of studies (for other reviews see e.g. Perlmutter et al., 1984; Malipiero et al., 1987). The V_H gene used by BALB/c mice is called V_1 or V_{T15}. Mice with alterations in this gene respond poorly to PC (Hilbert and Cancro, 1988). Nevertheless, there are exceptions to the universal use of V_1 (Claflin and Berry, 1988; Hilbert and Cancro, 1988; Stenzel-Poore et al., 1987). Of the 12 D elements known (Fig. 11), $D_{FL16.1}$ is predominantly used, although

---

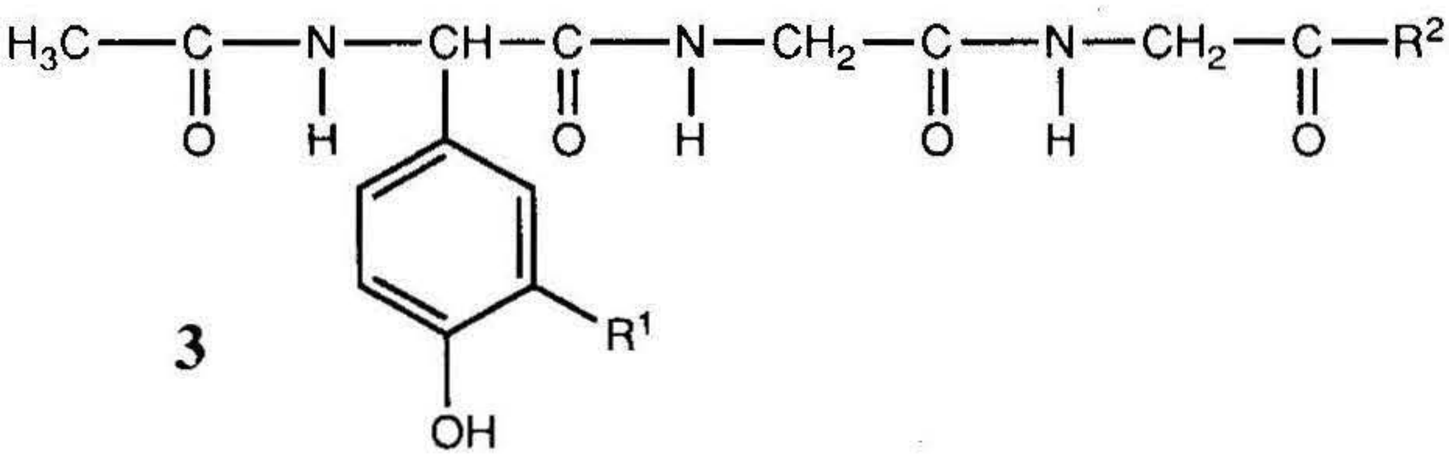
[1] The "PC" in the antibody names signifies plasmacytoma or plasma cell tumor (not phosphorylcholine), and mineral oil ("MO") or 2,6,10,14-tetramethyl-pentadecane (pristane) ("TE") as the method of myeloma induction. Alternatively, a prefix encoding the investigator has been used, e.g. Mc for K.R. McIntire. (Taken from Potter, 1972)



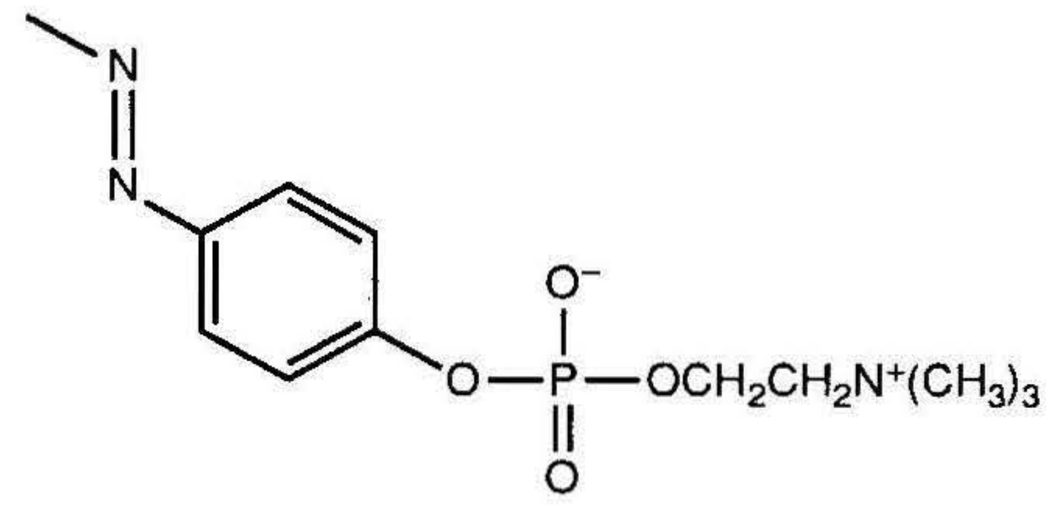
1



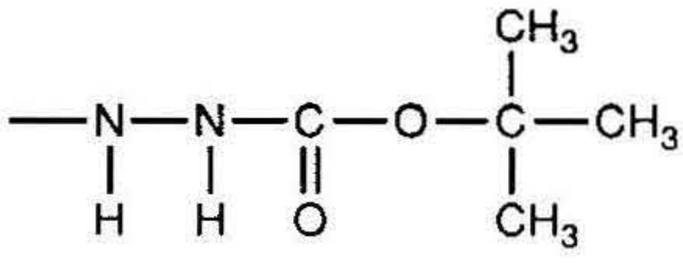
2



3



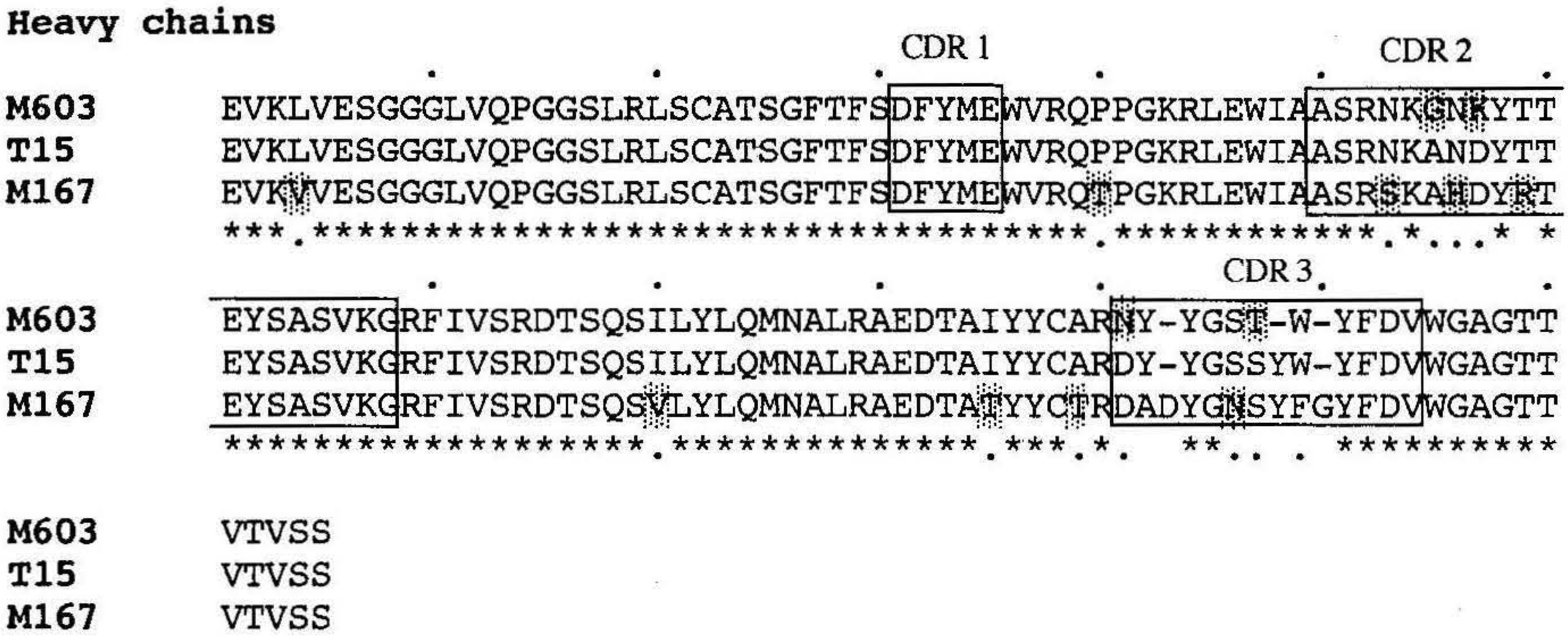
3a $R^2 =$



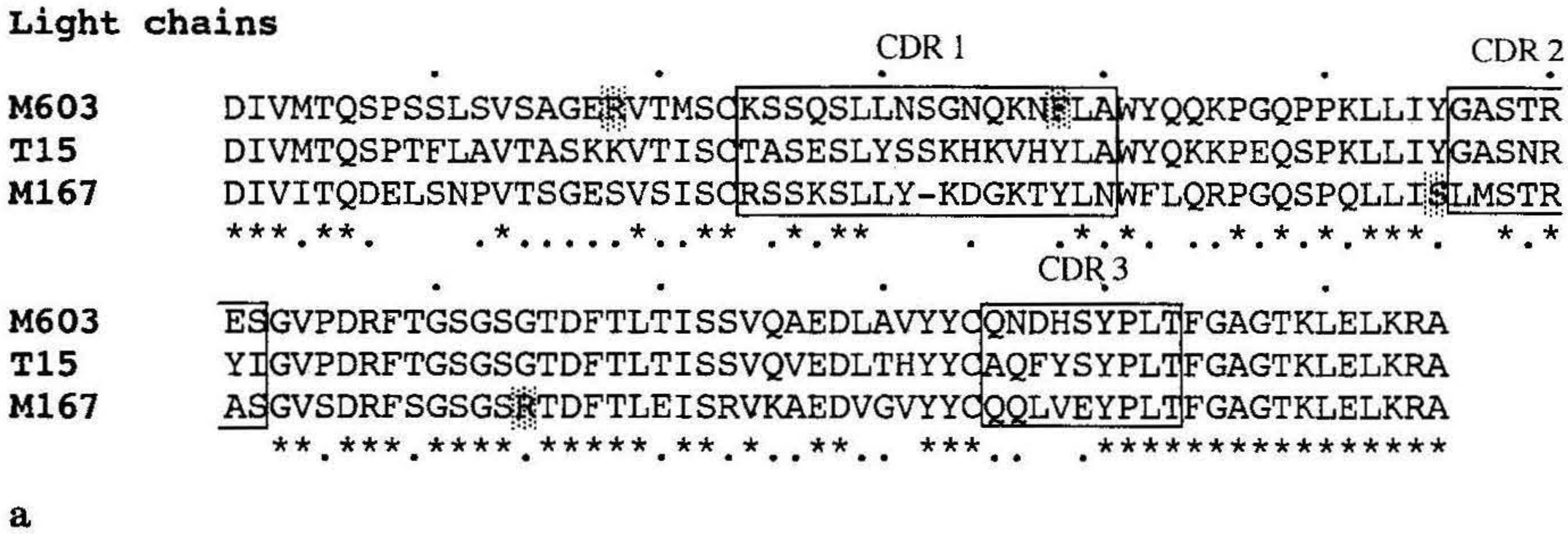
3b $R^2 =$ —NH—Protein

Fig. 9. Structure of phosphorylcholine esterified to carrier molecules. (**1**) Hapten affinity ligand, usually prepared from glycyl-tyrosine linked to cyanogen bromide activated sepharose, and linked to the hapten by azo coupling with *p*-diazonium-phenyl-PC. As immunogen, *p*-diazonium-phenyl-PC has been reacted directly with tyrosyl residues on carrier proteins (**2**), or via a spacer (**3**). In this case, *N*-acetyl-tyr-gly-gly-tBOC-hydrazide (**3a**) is cleaved with acid to remove the tBOC group, oxidized to generate *N*-acetyl-tyr-gly-gly-azide and reacted with amino groups on carrier proteins (**3b**) (Inman et al., 1973a, b)

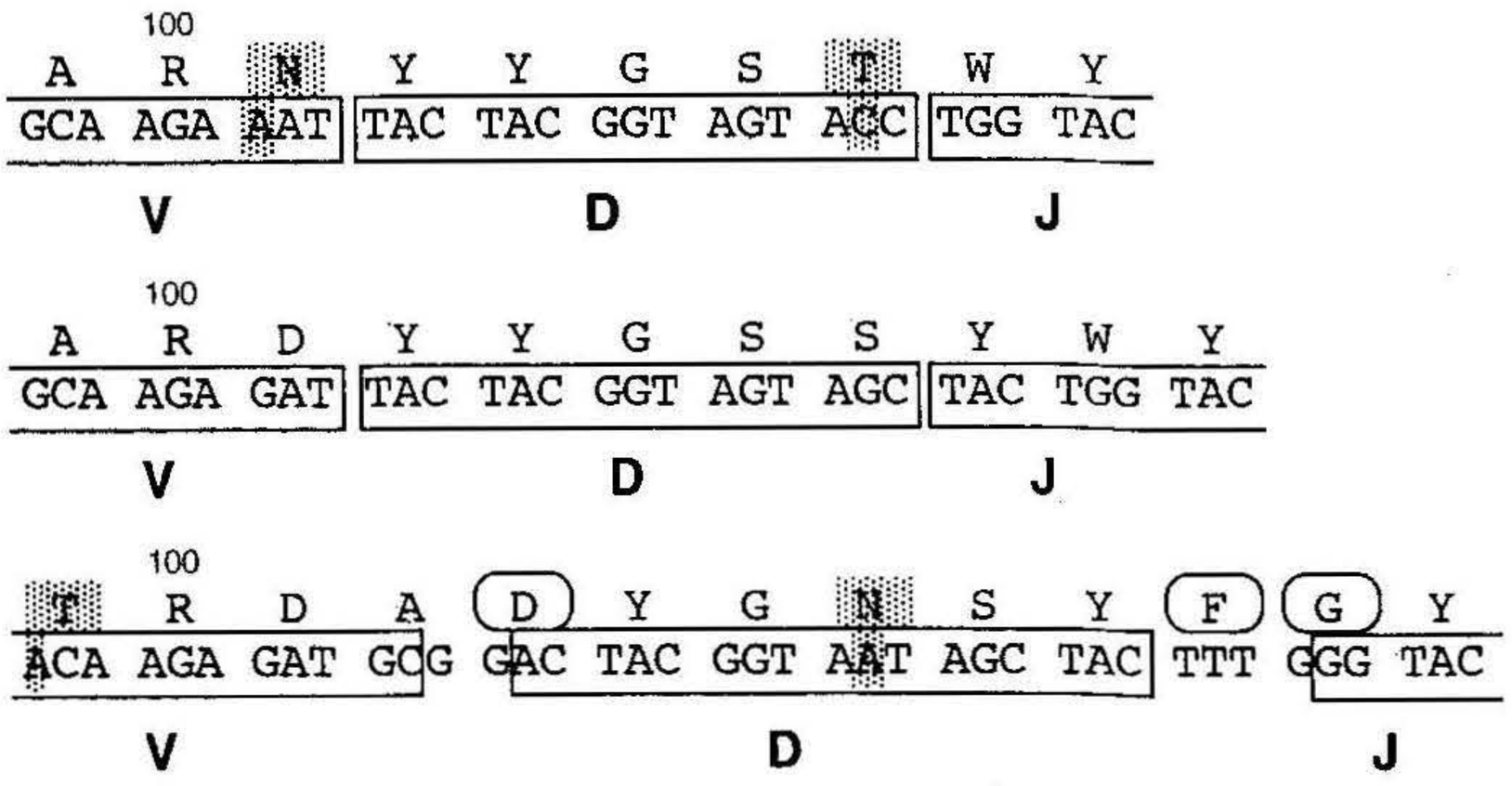
Heavy chains



Light chains



a



b

Fig. 10a. Alignment of the heavy and light chains of the three PC-binding antibodies M603, T15 and M167. The CDRs are boxed and somatic mutations are shaded. (**b**) CDR3 of the heavy chain. The genetic elements are *boxed*, somatic mutations are *shaded* and non-encoded amino acids (N-region) are *framed*

it may be heavily truncated. Of the 4 J elements of the heavy chain of the mouse, only J_H1 appears to be used. Possible structural reasons of this selection will be discussed below. In addition, there is the so-called N-region diversity, which can alter the length of CDR3 by introducing (or nibbling off) nucleotides at either side of the D element (Fig. 10). In the light chain, the diversity of the anti-PC response is larger, as three completely different light chain genes are used. The mouse contains 4 functional J elements for the light κ chain, but only one, $J_{\kappa 5}$ is used in combination with all three V_L genes (Fig. 10).

There had been some uncertainty about the nature of the amino acid in position L112 (Fig. 10a) in M603. It was originally determined from protein sequencing as Leu, then suggested to be Ile (Rudidoff, unpublished, quoted in Satow et al., 1986). Since the $J_\kappa 5$ element encodes Leu (Ponath et al., 1989) at this position, Leu is almost certainly correct, and Leu was used in the synthetic genes for M603, M167 and T15 (Plückthun et al., 1987; Plückthun and Pfitzinger, 1991).

```
Jκ1        WTFGGGTKLEIK          JH1      YWYFDVWGAGTTVTVSS
Jκ2        YTFGGGTKLEIK          JH2      --YFDYWGQGTTLTVSS
Jκ3 (Ψ)  [ITFSDGTRLEIK]          JH3      -AWFAYWGQGTLVTVSA
Jκ4        FTFGSGTKLEIK          JH4      YYAMDYWGQGTSVTVSS
Jκ5        LTFGAGTKLELK                   .. . ***.** .***
           .***.*****.*
```

```
DFL16.1      TTTATTACTACGGTAGTAGCTAC      DFL2.5       TCTACTATGGTAACTAC
         a   F  I  T  T  V  V  A              a    S  T  M  V  T
         b    L  L  L  R  *  *  L             b     L  L  W  *  L
         c     Y  Y  Y  G  S  S  Y            c      Y  Y  G  N  Y

DFL16.2      TTCATTACTACGGCTAC            DFL2.6       CCTACTATGGTTACGAC
         a   F  I  T  T  A                  a    P  T  M  V  T
         b    S  L  L  R  L                 b     L  L  W  L  R
         c     H  Y  Y  G  Y                c      Y  Y  G  Y  D

DFL2.2       TCTACTATGATTACGAC            DFL2.7       CCTACTATGGTAACTAC
         a   S  T  M  I  T                  a    P  T  M  V  T
         b    L  L  *  L  R                 b     L  L  W  *  L
         c     Y  Y  D  Y  D                c      Y  Y  G  N  Y

DFL2.3       TCTACTATGGTTACGAC            DFL2.8       CCTAGTATGGTAACTAC
         a   S  T  M  V  T                  a    P  S  M  V  T
         b    L  L  W  L  R                 b     L  V  W  *  L
         c     Y  Y  G  Y  D                c      *  Y  G  N  Y

DFL2.4       TCTACTATGGTTACGAC            DQ52         CAACTGGGAC
         a   S  T  M  V  T                  a    Q  L  G
         b    L  L  W  L  R                 b     N  W  D
         c     Y  Y  G  Y  D                c      T  G
```

Fig. 11. The J and D elements of the mouse. All four J elements of the heavy chain (Solin and Kaartinen, 1992) are functional. The light chain κ-locus has a total of six J elements, of which the third (shown) and the sixth (not shown) are pseudo-genes (Ponath et al., 1989). For the D-element, all three reading frames can in principle be used, provided stop codons are eliminated (Gu et al., 1991). The D_{FL2} group has two more members whose exact sequence is unclear (Kurosawa and Tonegawa, 1982; Ichihara et al., 1989). Thus, the mouse seems to have 12 D elements.

The response of the immune system to a particular antigen occurs in two distinct stages. The first stage, outlined above, already takes place in the development of B-cells before the antigen is ever seen: the random assortment of the genetic elements to prepare a library of binding molecules, by genetic rearrangement. The second stage appears to be antigen driven (Gearhart et al., 1988; Kocks and Rajewsky, 1989). A positive selection of responding B-cell clones leads to random somatic mutation (Tonegawa, 1983; Gearhart et al., 1988) of the rearranged antibody genes, and further positive selection of tight binders.

One of the crucial questions from the protein engineer's point of view is the structural logic of these processes, and the reasons why certain point mutations were selected (or at least tolerated as being equivalent). To answer these questions, we have to consider the three-dimensional structures and the locations of these variations. Fig. 6C illustrates the location of these mutations. On the sequence level, the deviations of both chains from their respective genomic sequence are shown in Fig. 10. All chains contain somatic mutations (with the exception of T15 itself). It remains unclear why M167-like antibodies and M603-like antibodies accumulate somatic mutations more frequently than T15-like antibodies (Feeney et al., 1988).

It can clearly be seen that the somatic mutations are scattered about the whole protein (Fig. 6C). However, the genetic variability, caused by the rearrangement process including the N-region diversity or the "nibbling" of the D-element, brings about the most dramatic changes in the binding site region. In Fig. 6, models of the three antibody prototypes M603, M167 and T15 are shown, emphasizing the location of all diversity generating features in space: the "genes-in-part" (the V, D, J region), the N-region diversity and the somatic mutations. While there is some uncertainty about the details of the models of M167 and T15, the general location of the affected residues with respect to the binding site can hardly be disputed (Padlan et al., 1976).

5.2 The Binding Site of M603

To understand the interactions between hapten and antibody more fully, it is necessary to consider the three-dimensional structure of the binding site of McPC603 (Segal et al., 1974; Satow et al., 1986) in detail (Fig. 12). The phosphorylcholine antigen is about 81% buried, and bound with a thermodynamic affinity constant of about $1.6 \times 10^5\ M^{-1}$ (Metzger et al., 1971). The quaternary ammonium ion is buried, while the phosphate is more at the protein surface. This is a direct consequence of the immunogen structure, which is always a diester of phosphate, and thus the phosphate is tethered to the immunogenic carrier, be it a protein or the teichoic acid (polysaccharide) of the microbial cell surface (Fig. 8, Fig. 9).

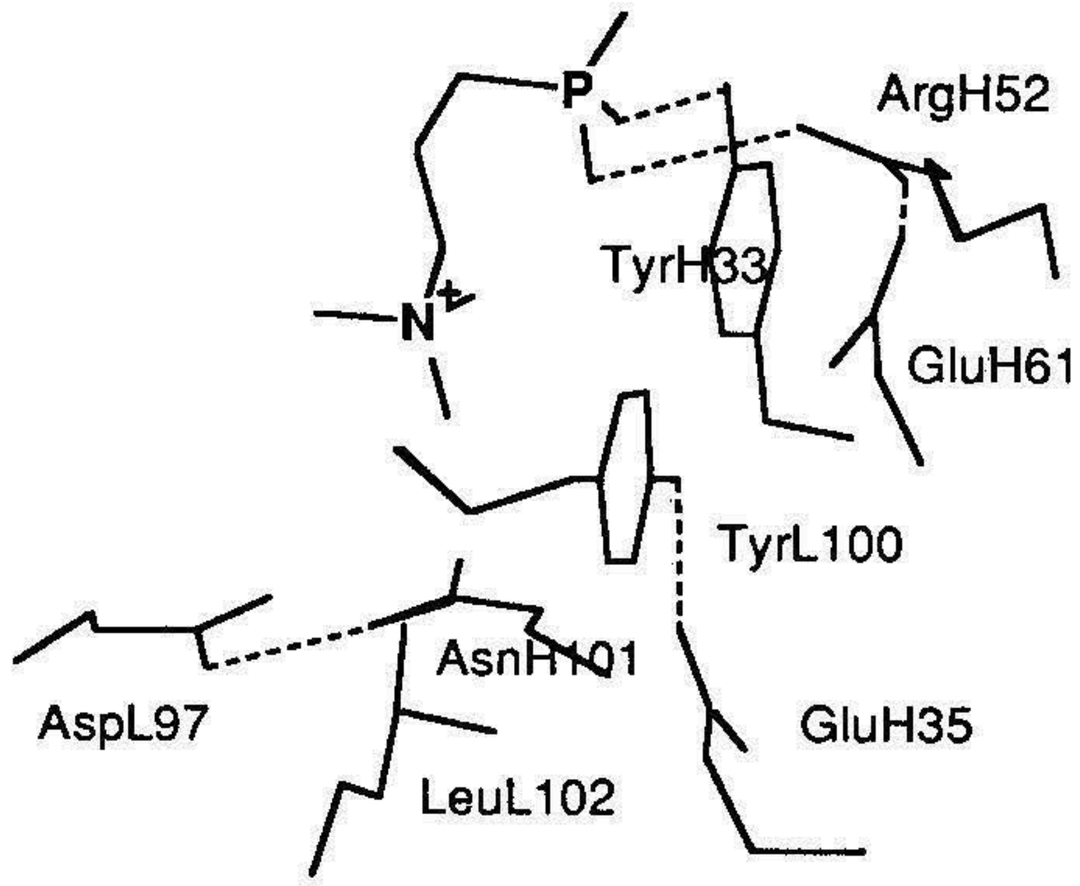


Fig. 12. Binding site of the antibody McPC603

The quaternary ammonium ion interacts with negative charges in the protein. These are contributed[2] in McPC603 by AspL97 and GluH35. Their closest carboxyl oxygen atoms are at distances of 4.7 and 5.8 Å away from the ammonium nitrogen. Both carboxyl groups are kept at this position by interactions of these moieties with other hydrogen bond donors of the protein. Interestingly, these very critical interactions are both contributed by residues from the other chain: AsnH101 (to interact with AspL97) and TyrL100 (to interact with GluH35). This means that for the generation of this particular binding site, the exact light and heavy chains found in McPC603 must be selected. It also shows very clearly that in the immune response to this antigen, the question of which domain is "more important" for antigen recognition is ill-posed, as both are equally needed. Unsurprisingly, single domains (either V_H or V_L alone) display no measurable antigen binding.

The phosphate group of the antigen makes two important interactions with the antibody: TyrH33 forms a hydrogen bond to one of the phosphate oxygens. ArgH52 forms a salt bridge with the phosphate. Several ^{31}P-NMR experiments show (Gettins et al., 1977; Goetze and Richards, 1978) that the phosphorylcholine is bound as the dianion at pH 8, at which most experiments have been done. There is no indication of a bidentate binding of the guanidinium group to the phosphate, as this seems incompatible with the observed electron density, although this might seem appealing from an organic chemist's point of view. Instead, the guanidinium group appears to be kept in position by an interaction with GluH61. One should not, however, overinterpret the structure as if the distances were very accurate, as it could only be solved to 2.7 Å resolution.

In direct van der Waals contact with the phosphorylcholine antigen are a number of hydrophobic groups, most notably TyrL100, LeuL102 and TrpH107.

---

[2] In the numbering of antibodies, 2 different systems have come into use. The first is the consensus numbering scheme proposed by Kabat et al. (see e.g. Kabat et al., 1991). The second is a sequential numbering preferred by crystallographers. In this article, the sequential numbering as in the crystal structure of McPC603, (entry 2MCP in the Brookhaven database) will be used for referencing equivalent positions of McPC603, M167 and T15, unless indicated otherwise.

An interesting and not fully resolved question are the contributions of additional amino acids to the binding of the antigen, which are not in direct contact with the antigen. Apart from the very difficult (and perhaps infertile) discussion about which residues contribute directly or indirectly to the shape of the binding site (in a sense, it is most of the amino acids of the antibody variable domains), electrostatic long range forces may be implicated (Novoutny and Sharp, 1992; Lee et al., 1992). Electrostatic calculations were carried out for the antibody M167 (Aggarwal and Plückthun, unpublished). They suggest a rapid decrease of strength with distance such that essentially no effect is left at 20 Å, in line with measurements in other proteins (Russel and Fersht, 1987; Russel et al., 1987). This is born out by experiment in some cases (see below), but the mutation AsnH106 → Asp in M167 decreases PC binding by a factor of 2. While this is not much, it is found in two different proteins (Schweder and Plückthun, unpublished), and is far larger than predicted by electrostatic calculations. It appears that this residue, also 20 Å away from PC, may give rise to different loop conformations, even though situated at the tip of a loop, perhaps because of a different type of turn preferred by Asn or Asp. This way, the interaction of residues at the bottom of the loop may again be better placed.

Another type of an at first seemingly electrostatic effect was a spontaneous Asp → Ala mutation (position H113 of T15 in Fig. 10a), which was suggested to interrupt a crucial salt bridge to ArgH100 (Fig. 10a), which would then be able to interfere with the PC binding pocket (Chien et al., 1989).

A comparison of the sequences of M603, T15 and M167 shows immediately that important interactions in the binding site must be different, although all are derived from the same V_H gene (Fig. 6, 10, 12). The reason comes almost exclusively from the N-region diversity, since the nature of the D segment and the J segment are identical between the three antibodies as well. The severe truncation (and somatic mutations) can sometimes mutilate the sequence of the D segment beyond recognition, but a comparison with all known D segments shows that $D_{FL16.1}$ is predominantly used. In other mouse strains, different D segments have been found in the anti-phosphorylcholine response (Clarke et al., 1983). Interestingly, heterologous recombinations of the light and heavy chains of the three myeloma protein chains did not lead to antigen binding (Hamel et al., 1986), but the structural reasons are not fully understood. That there is a correlation between the choice of the light chain and CDR-H3 is also shown in sequences of anti-PC antibodies using the $V_\kappa 24$ (M167) light chain which all appear to have the unusually long CDR-H3 of M167 (Gearhart and Bogenhagen, 1983; Claflin et al., 1987; Malipiero et al., 1987).

In McPC603, TrpH107 is in direct contact with the antigen. The antigen must be positioned at least somewhat differently in T15, as this CDR3 loop contains one additional amino acid. While sequence alignment would put the insertion right next to this tryptophan, functional arguments can be used to accommodate the additional amino acid more at the tip of the loop, consistent with the results from database searches of loops connecting β-strands in other proteins. A radically different situation is found in M167. Not only is the loop three amino acids longer than in M603, but the tryptophan residue is missing

altogether. Instead, there is another aspartate (H103 in the consensus numbering of Fig. 10), which may be involved in electrostatic interactions with the quaternary ammonium group.

This is demonstrated by converting AspH101 or AspH103 to Asn in M167. Both lead to a severe reduction in binding. This result demands that the quaternary ammonium group is in almost van der Waals contact with these carboxylates, since the attractive energy (inversely proportional to distance) of one positive charge distributed on the quaternary ammonium ion must be larger than the repulsive force of the two negative charges of the phosphate group. Therefore, the position of the quaternary ammonium ion appears to be different in different anti-PC antibodies (Schweder, Aggarwal, Guth and Plückthun, unpublished).

The D segment can in principle be read in all three reading frames, depending on the frame in which it is joined to V_H, which in turn is governed by nuclease digestion or N-region diversity. However, in almost all antibodies, the reading frame with the most tyrosines is found. Tyrosine seems to be very useful for protein-protein interactions (with the antigen), but also for antibody-DNA interactions or even antibody-hapten interactions (Padlan, 1990a; Mian et al., 1991). Most likely, this is because tyrosines are multifunctional: they can act as hydrogen-bond donors, interact by hydrophobic interactions and be involved in aromatic-aromatic interactions (Burley and Petsko, 1985). The tyrosines of the binding site of antibodies are exposed, which is very unusual, whereas normally tyrosines are buried deeper in the interior of proteins. There is no mechanistic reason on the DNA level, why one reading frame should be favored over all the others. The most economical hypothesis is therefore that there is a selection for protein function favoring the reading frame with the tyrosines.

The light chains of M603, T15 and M167 are derived from three different V_L genes and show therefore much more diversity than the heavy chains. There is only one length variation, namely in CDR1 of M167 (Fig. 10). This loop is one amino acid shorter. Recently, the X-ray structure of a mutant of the V_L domain of M603 was solved at high resolution, in which the five amino acids L31-L35 (NSGNQ) were replaced by the four amino acids L31-L34 (YKDG), derived from CDR1 of M167, on the M603 framework (Steipe et al., 1992). The effect was found to be entirely local involving only this loop, with the rest of the structure almost identical.

There is no N-region diversity in the light chains, and the J region selected is always the same (Fig. 10). J regions of the kappa light chains (Fig. 11) are most conspicuously different in the first amino acid. This amino acid, LeuL102, is in direct contact with the antigen. Its mutation leads to dramatic loss in binding (see below), making the choice of this particular J element structurally understandable.

Another most interesting difference between the three light chains is at position L97 (see Footnote 2), where M603 contains Asp (in direct contact with the quaternary ammonium ion), whereas the other two chains contain a hydrophobic residue (Phe or Leu) (Fig. 10). The presence of a hydrophobic residue correlates with the presence of Asp at position H101 (in T15 and M167)

whereas M603 contains an Asn at this position. Position H101 is also in direct contact with the antigen (Fig. 12). This observation is very noteworthy, as the genome of the mouse encodes Asp at position H101 (as retained in the sequence of T15 and M167, which is the germline version (Crews et al., 1981; Perlmutter et al., 1984; Malipiero et al., 1987; Claflin et al., 1987)). Thus there appears to be a forced mutation Asp → Asn, when the light chain of M603 is chosen, probably to avoid the juxtaposition of two negative charges (Claflin et al., 1987; 1989). Indeed, this mutation (AsnH101 → Asp) leads to a drastic decrease of antigen binding (see below). Sequencing the heavy chains of a variety of M603-related monoclonal antibodies showed that the selection of this particular light chain seems to be coupled to this somatic mutation (Claflin et al., 1987).

Figures 6 and 10 show the location of other somatic mutations in the sequence and the 3D-structure. They are scattered all over the protein. Most could be called conservative, but others clearly are not: In M167, GlyL74 → Arg would be expected to locally alter the conformation of the chain slightly, but the mutation probably can be accommodated. This mutation has no effect on binding. This result clearly shows that electrostatic effects play only a minor role at large distances (the distance to the hapten is about 20 Å), consistent with electrostatic calculations using the program DELPHI (Sharp and Honig, 1990). In contrast, the somatic mutation changing the genomically encoded TyrL55 into Ser present in M167 improves binding 24-fold. This framework residue, also about 20 Å away from the hapten, makes contact to CDR-H3, and thereby probably improves the geometry of interaction of the directly contacting amino acids.

The observed somatic mutations have been found in secreted functional antibodies, and the most drastic ones leading to no functional protein or perhaps even to no protein at all, have probably been selected against in the immune response of the mouse. Thus, the observed somatic mutations (and there are many more documented in other antibodies of the anti-PC response (Perlmutter et al., 1984; Malipiero et al., 1987, and references therein)) give an indication of the tolerance of the antibody structure. It is, of course, not a complete sampling by any means. There is no direct evidence that somatic mutations are totally random at the genetic level, and there may well be mutational hotspots, imposed by the DNA structure or sequence. Now, however, with the techniques of site-directed mutagenesis and facile expression in *E. coli*, we may take a new look at these questions with a very powerful set of tools.

6 Mutational Analysis of the Anti-PC Antibodies

To study the effect of single amino acid contributions to antibody-antigen interactions quantitatively, a variety of mutant antibodies of the three prototypes M603, T15 and M167 was produced and their binding constants were

measured. At the same time, structural studies of the recombinant proteins were initiated, both by NMR and crystallography. Finally, theoretical investigations were carried out by using molecular dynamics (to obtain structural information), and docking calculations (to obtain information about possible binding interactions) to correlate these theoretical results with experimental data. It was the intention of this combined approach to understand the structural logic leading to antigen binding in these particular antibodies.

The methodology used was to synthesize the genes of the antibodies McPC603 (Plückthun et al., 1987), T15 (Plückthun and Pfitzinger, 1991) and M167 (Plückthun and Pfitzinger, unpublished) and express them as F_v fragments, single-chain F_v fragments or F_{ab} fragments in *Escherichia coli*, as described above. With this strategy, mutations can be constructed efficiently, and the proteins can be produced quite rapidly.

Several methods have previously been used to measure the free energies of antigen binding of naturally occuring anti-phosphorylcholine antibodies. Most frequently, a semiquantitative method has been employed, in which the binding of the anti-PC antibody to PC bound to a solid phase is inhibited by the compound to be tested. This has been carried out using large amounts of antibodies with the precipitin reaction (Leon and Young, 1971; Young and Leon, 1977) or with much smaller amounts in an ELISA or RIA format (see e.g. Cook et al., 1982; Andres et al., 1981; Claflin et al., 1985), and when performed with great care, it will allow fairly rapid comparisons between different soluble analogs of PC. This method requires bivalent antibodies (see below). A more direct and unequivocal method is equilibrium dialysis (Metzger et al., 1971; Skerra and Plückthun, 1988). Its main disadvantage is the comparatively high protein concentration necessary, if binding constants are weak. Perhaps the most convenient technique, and that used in most of the mutant studies, is the titration of the intrinsic protein fluorescence (Pollet and Edelhoch, 1973; Pollet et al., 1974; Glaudemans et al., 1977; Manjula and Glaudemans, 1978; Bennett and Glaudemans, 1979; Glockshuber et al., 1990a). Apparently, the response of the tyrosine fluorescence to antigen binding must be especially important, as significant differences upon antigen binding are seen only if the excitation is at 280 nm. It remains unknown, however, which tyrosine residues may be the most crucial ones in sensing the antigen. Other techniques used to observe antigen binding have included ^{3}H-, ^{13}C-, and ^{31}P-NMR (Goetze and Richards, 1977a, b; 1978; Gettins et al., 1977; 1982), difference absorption, fluorescence polarization and circular dichroism (Pollet et al., 1974; Young et al., 1985).

As a semiquantitative first test, the binding of the mutant protein to a phosphorylcholine affinity column was investigated. Independent measurements showed that a mutant with a 60-fold reduction in affinity to PC could be purified (TyrH33 → Phe) (Glockshuber et al., 1991), whereas a 300-fold reduction in binding led to a loss of binding to the column (TyrH33 → His) (Stadlmüller and Plückthun, unpublished). The purification of the mutant proteins by affinity chromatography is also a selection for correctly folded protein, because the purification of poorly binding mutants is not possible with this method. A

new procedure, however, based on the purification with a tail of 5 histidine residues by immobilized metal affinity chromatography (Arnold, 1991) could be used to overcome this problem and to purify "inactive" mutants as single-chain F_v fragments (Skerra et al., 1991; Lindner et al., 1993).

The affinity ligand in column chromatography (and ELISA) is a phosphodiester of choline (Fig. 9) and thus carries only one negative charge at pH 8.0. It is possible that binding of the two-fold negatively charged PC molecule (at pH 8.0) might be differently affected by some mutations than binding to the column.

Changing most of the amino acids in direct contact with the trimethylammonium group abolishes binding by the affinity column below the limit of detection. It was found that the pair of residues AspL97/AsnH101 must contain exactly one negative charge (Table 1). Neither the combination AspL97/AspH101, nor the combination LeuL97/AsnH101 leads to detectable binding on a hapten affinity column. The rationale to make these mutations was given by the fact that in the V_{T15} germline gene of the mouse, residue H101 is an Asp. While the light chain of M603 contains an Asp at L97, the other two light chains of M167 and T15 contain Leu and Phe at this position, respectively. This may not be too surprising, as the quaternary ammonium group is somewhat hydrophobic, in spite of its charge. If these light chains are used, the negative charge at H101 must be retained. Unsurprisingly, a spontaneous Asp H101 → Asn mutant in T15 lost the ability to bind PC (Kobrin et al., 1991).

Apparently, the pairing of the V_{T15} heavy chain with the light chain of M603 (containing an Asp) leads to a forced somatic mutation in the heavy chain (Claflin et al., 1987; 1989). The juxtaposition of AspL97/AspH101 may result in a mutual local repulsion, leading to a maximal avoidance of these groups. This appears to be sterically possible in the structure. Probably, the binding interactions then become insufficient. Alternatively, more severe disturbances of V_H/V_L binding by this local repulsion cannot be excluded. These observations raise the question of how the M603 antibody was initially selected, before the AspH101 → Asn mutation improved the binding. Presumably, the (weak) binding of the genomic version might still be better than of most other (nonspecific) antibodies, *in the mutivalent B-cell surface or lgM state*. This single somatic mutation would then immediately improve this antibody. The opposite somatic mutation, AspL97 → Asn or Leu, should also lead to the desired increase in binding, but has not yet been observed to occur. It is not known, whether

Table 1. Pairs of residues in contact with the quaternary ammonium ion in M603

Position L97	Position H101	Binding constant for PC (M^{-1})
Asp (w.t.)	Asn (w.t.)	1.6×10^5
Asp	Asp	n.d.
Leu	Asp	n.d.
Leu	Asn	7.8×10^4

Mutant proteins carrying the pair of residues indicated in each line were examined.
n.d. = binding not detectable

particular mutational hotspots or just a bias in the known sequences is the reason why this change has not been found in the family of M603-like antibodies. Recently, the AspL97/AsnH101 arrangement was also introduced into M167 (Schweder and Plückthun, unpublished), and it leads to functional antigen binding but decreased stability of the protein.

Another pair of interactions between the light and the heavy chain may be crucial for the selection of the particular chains in the anti-PC response. TyrL100 makes a hydrogen bond to GluH35. The mutations GluH35 → Gln as well as TyrL100 → Phe or TyrL100 → Trp all abolish detectable antigen binding. Probably both the positioning of the phenyl ring of TyrL100 as well as the positioning of GluH35 is required, one for hydrophobic interactions of the ring edge with the antigen, the other for electrostatic interactions with the quaternary ammonium ion. Neither Phe nor the much larger Trp at this position leads to binding, although the latter might provide an acceptable lining of the pocket. It is possible that neither of these residues fixes GluH35 in the required position. Gln might still provide the H-bond acceptor function for TyrL100, provided it is oriented with its carbonyl group proximal to the tyrosine hydroxyl group. In conclusion, the particular interaction between TyrL100 and GluH35 seems to be required, and it cannot be pinpointed to only one of the partners. A mutation of GluH35 to Ala was also found to be the cause of binding loss in a spontaneous mutant of a T15-related antibody (Rudikoff et al., 1982).

The residues in contact with the phosphate group were also scrutinized (Glockshuber et al., 1991). TyrH33 makes a hydrogen bond to one of the oxygens of the phosphate group of PC. Its replacement by Phe was tested with a variety of PC-analogs (Table 2). Significant loss of binding for M603 was seen with PC (two negative charges at pH 8.0) and a butyric acid derivative (one negative charge), but not with a sulfate or sulfonate derivative (one negative charge), nor with a butyric acid methyl ester (no negative charge). This immediately shows that the importance of the H-bond donor function of TyrH33 is not correlated with the antigen charge. Charge, relative fluorescence increase and relative change in binding affinity by the TyrH33 → Phe exchange are not correlated, either. Interestingly, the absolute change in fluorescence upon binding of various haptens is hardly changed by the TyrH33 → Phe mutation, and this residue thus does not seem to be contributing significantly to the fluorescence change seen in the titration. It appears therefore, that the hydrogen

Table 2. Binding of haptens to M603 w.t. and Tyr → PheH33

Hapten	K_{ass} ($10^3\ M^{-1}$) w.t.	K_{ass} ($10^3\ M^{-1}$) Tyr → PheH33
$(CH_3)_3N^+CH_2CH_2OPO_3^{2-}$	160	2.5
$(CH_3)_3N^+CH_2CH_2OSO_3^-$	4.0	6.2
$(CH_3)_3N^+CH_2CH_2CH_2SO_3^-$	7.7	6.4
$(CH_3)_3N^+CH_2CH_2CH_2CO_2^-$	93	19
$(CH_3)_3N^+CH_2CH_2CH_2CO_2CH_3$	1.0	0.85

Table 3. Single point mutants of M603 F_v fragments not binding to the affinity column

Light chain	Heavy chain
TyrL100 → Phe	TyrH33 → Arg
TyrL100 → Trp	TyrH33 → His
LeuL102 → Phe	TyrH33 → Lys
LeuL102 → Tyr	GluH35 → Gln
	ArgH52 → Lys

bond between the sulfate, sulfonate or ester and TyrH33 is not formed at all or has a very non-ideal geometry, such that its loss is not detected energetically.

Introducing another positive charge at position H33 (Arg, Lys, His) did not lead to detectable binding (Table 3). This may be due to either very strict geometric requirements of antigen binding, incompatible with any of these residues or alternatively a local avoidance of the positive charge in H33 and ArgH52. At position H52, lysine could not substitute for the original arginine. Thus, the presence of TyrH33 and ArgH52 might explain the unexceptional selection of the V_{T15} gene in anti-PC antibodies.

However, there seem to be differences in the family of anti-PC antibodies. The mutation TyrH33 → Phe is not purifyable by affinity chromatography in M167, nor is ArgH52 → Lys. However, the latter one does become purifyable in the context of the additional mutation AsnH106 → Asp, perhaps because CDR3 has become slightly altered in conformation. It has been reported that the substitution of TyrH33 by His or Phe had hardly any effect in the antibody S107, a very close relative of T15 (Jackson et al., 1991). However, the latter data were not measured by direct antigen binding, but by the inhibition of the catalytic activity of these antibodies (see below). Antigen affinity chromatography of the mutant TyrH33 → His was unsuccessful in T15 (Pfitzinger and Plückthun, unpublished).

Before a complete understanding of these effects can be reached, more mutations must be compared in M603, T15 and M167. Perhaps even more important, however, is the solution of a high-resolution structure of some of these recombinant proteins and their mutants.

As a first step in this direction, the recombinant V_L domain of M603 was crystallized and its structure solved at 2.0 Å resolution (Glockshuber et al., 1990b; Steipe et al., 1992). In the crystal, V_L dimers are seen. Several observations from this structure determination are noteworthy:

(1) The structure is essentially identical to the same V_L domain in the context of the F_{ab} fragment. This shows that the sequence of each domain contains the complete structural information. Some changes are seen, however, such as a rigid body movement of CDR3, possibly brought about by contacts between heavy and light chain in the F_{ab} fragment.

(2) The structure of the interface is particularly well conserved, although the interface is making rather different contacts in V_L-V_H and V_L-V_L. The overall topology of the V_L dimer resembles the V_L-V_H pairing (as would be expected from the fact that both had a common ancestor), but the sequences of V_L and V_H have sufficiently diverged to change a number of contacts.
(3) The crystal packing in the V_L structure is such that the CDRs point into a very large solvent channel and no significant crystal contacts seem to influence their conformation. Thus the replacement of a loop of CDR1 in M603 by a piece of sequence of M167 (L31-L35, NSGNQ → YKDGΔ) could be crystallized isomorphously. This replacement was an entirely local change, not influencing the rest of the structure.

Molecular dynamics trajectories were calculated for the V_L domain, starting from either the structure of the F_{ab} fragment or the V_L dimer. The differences seen (Steipe et al., 1991) did not interconvert (Aggarwal et al., in preparation), but each of the structures appeared to relax locally. Most interestingly, the structure of CDR3 seen in the crystal structure, appeared to be metastable.

This same loop mutation has also been analyzed by antigen binding in the F_v fragment. Unexpectedly, a 15-fold increase in binding of *N*,*N*,*N*-trimethyl butyric methyl ester (Stadlmüller and Plückthun, unpublished) was observed. Under the assumption that this mutation only leads to a local change in the F_v structure (as in the V_L crystal structure), and taking into account the observation that all these amino acids are 10–15 Å away from the antigen, we must postulate alternate binding modes of this substrate bringing it in closer contact to these residues. This question can ultimately only be solved by obtaining the crystal structure of the mutant protein in complex with the antigen. It is very likely, however, that alternate binding modes are a frequent occurrence in studying a series of ligands with series of similar mutant proteins.

It was also found that the two mutations showing a significant improvement in the binding of *N*,*N*,*N*-trimethyl ammonium butyric acid, GluH61 → Gln (5-fold) and the loop mutation (L31-L35, NSGNQ → YKDGΔ) (15-fold), were not additive, but their combination resulted only in a 7-fold increase. Again, this implies alternative binding modes of this hapten in the different mutants.

These questions have also been addressed with docking calculations (Goodsell and Olson, 1990; Aggarwal and Plückthun, unpublished). This approach relies on a grid calculation of the potential field, using a static protein model. The hapten is given torsional degrees of freedom and performs a random walk on the grid, with steps selected by Monte-Carlo criteria. The hapten PC can be docked remarkably well in a position corresponding to the one found in the crystal structure. Furthermore, *N*,*N*,*N*-trimethyl ammonium butyric acid docks in a similar conformation, consistent with its sensitivity to the TyrH33 → Phe mutation. Other esters of PC, however, gave rise to other binding conformations, bringing them in closer contact to CDR1 of the light chain. The quaternary ammonium group seems to act as a ball in a socket, formed by the binding pocket.

The amino acid GluH61 was changed to Gln in M603, but the effect on PC binding was only marginal. This shows that the positioning of ArgH52, making a hydrogen bond to GluH61, cannot be very critical. While one may argue that Gln could still form a hydrogen bond between its side-chain carbonyl group and arginine, the mutation GluH61 → Ala was introduced in M167 (Schweder and Plückthun, unpublished). Again, only a slight change of PC binding was observed, improving PC binding. This small change argues for the positioning of ArgH52 being immaterial and a relatively low importance of long-range electrostatic interactions. Alternatively, both effects may cancel fortuitously. Electrostatics calculations (Aggarwal and Plückthun, unpublished) do predict a much larger repulsive effect of GluH61.

Interestingly, the improvement of binding of *N*,*N*,*N*-trimethyl ammonium butyric acid by the GluH61 → Gln mutation in M603 is much more pronounced than the binding of PC (Stadlmüller and Plückthun, unpublished). As this hapten has no negative charge, the effect cannot be due to electrostatic repulsion. Rather, a more direct interaction cannot be excluded, again suggesting an alternative binding mode of this ester.

7 Catalysis

Enzymes and antibodies have in common that they specifically bind only one ligand or a closely related set of similar molecules. Yet, enzymes cause a chemical transformation, and antibodies and other binding proteins do not. To understand the principal differences, one must first discuss the mechanistic basis of rate accelerations in enzymes.

The active site of the enzyme is a device to place catalytic groups optimally in space. The substrate may be stripped of its solvent shell, may be bent into a productive conformation (at the expense of "intrinsic" binding energy, i.e. the binding energy actually seen is already diminished by this amount), and may be surrounded by nucleophiles, acids, bases, metal ions and a second reactant at the optimal relative geometry. All of this can be summarized by the pithy statement of J. R. Knowles that enzyme catalysis compared to chemical catalysis is "not different, just better" (Knowles, 1991). This refers to the optimization of geometries, pKa values, and intrinsic binding constants. Clearly, depending on the particular reaction, enzymes use a large variety of catalytic devices from general acid/base catalysis to nucleophilic catalysis to Lewis acids (in the form of metal ions) and bases.

One of the crucial intellectual breakthroughs in understanding the structure of the enzyme active site is due to Haldane (1930) and Pauling (1946). It is the realization that the structural complementarity of the active site is directed to the transition state of the reaction in many cases.

This concept has been discussed in more detail by Jencks (1975) and Kraut (1988). There are several lines of experimental evidence for the correctness of this concept.

(1) The direct structural evidence from the crystallography of several enzyme-inhibitor complexes has demonstrated that the geometry of the active site is more complementary to the transition state than the ground state (the substrate) (Kraut, 1988; Lolis and Petsko, 1990). Unfortunately, not very many structures of complexes have been solved yet.
(2) Stable analogues to the transition state have been synthesized for many enzymes. In a large number of cases, these compounds were shown to be inhibitors with tighter binding than the substrate itself (Wolfenden, 1976).
(3) There are enzymes such as tyrosyl-tRNA synthetase that do not seem to use any other catalytic device than the transition state complementarity of the active site. In the partial reaction tyrosine + ATP → tyrosyl-AMP + PPi, there is no known acid/base catalysis (Borgford et al., 1987).
(4) Even a reaction which normally heavily relies on chemical catalysis such as peptide hydrolysis catalyzed by serine proteases uses transition state complementarity. This has been postulated from crystallography (Kraut, 1988) and demonstrated by "stripping" the catalytic triad Ser-221, His-64 and Asp-32 and converting these residues to alanines. This modified enzyme is now about 10^6 times slower than the wild-type, but catalyzes peptide hydrolysis still about 2700-fold over the reaction in solution (Carter and Wells, 1988; 1990).

It is these observations that led to the development of catalytic antibodies. The concept was originally proposed by Jencks (1969). It was to elicit antibodies against a stable analog of the transition state. Early attempts to realize this concept were hampered by the fact that in many cases the rate accelerations were far too small to be seen in a polyclonal serum, as the newly elicited antibodies make up only a small fraction of all antibodies. However, after the development of monoclonal antibodies (Köhler and Milstein, 1975), this concept was realized very successfully (for reviews see e.g. Shokat and Schultz, 1990; Lerner et al., 1991). It is not intended to give an overview about this rapidly developing field in this chapter. Rather, two aspects are to be emphasized: (1) the investigation of the phosphorylcholine binding antibodies as an example and (2) the role of protein engineering and structural research in an improved understanding and the development of new catalytic antibodies.

The family of phosphorylcholine binding antibodies was one of the first examples for which catalytic activity was demonstrated. The reaction catalyzed is the hydrolysis of a carbonate ester (Fig. 13), whose tetrahedral intermediate after water attack is structurally similar to phosphorylcholine or a phosphorylcholine-ester. The transition states leading to and away from this tetrahedral intermediate may, by Hammond's postulate, reasonably be expected to be similar in structure and energy to the tetrahedral intermediate.

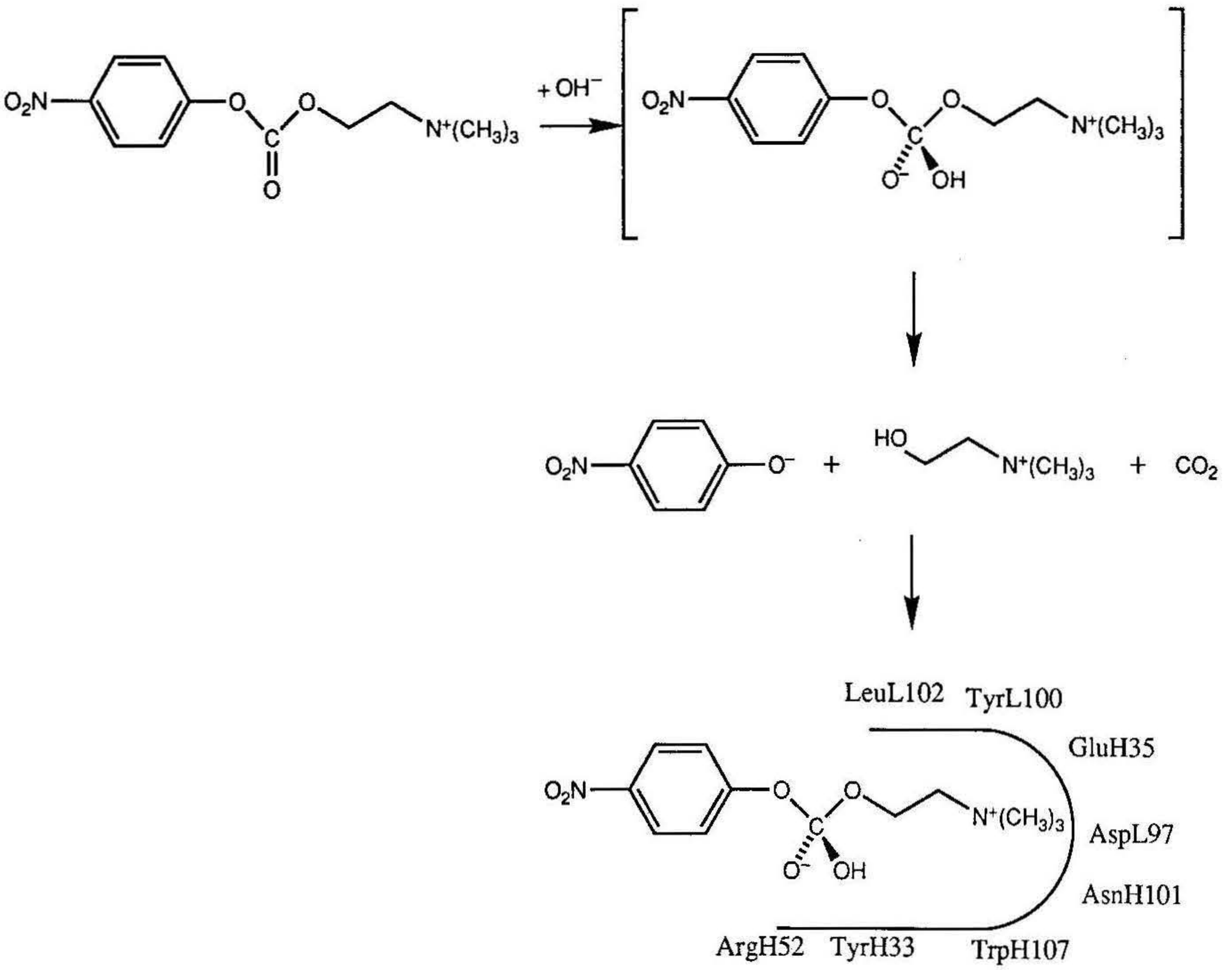


Fig. 13. Reaction catalyzed by the antibody McPC603. Below the tetrahedral intermediate, the binding pocket of the antibody is drawn schematically (see also Fig. 12)

This reaction was studied with the whole antibody M167 and T15 (Pollack et al., 1986; Pollack and Schultz, 1987), and their F_{ab} fragments as well as with the recombinant F_v fragment of M603 (Plückthun et al., 1990; Plückthun and Stadlmüller, 1991) (Table 4). The PC affinity is very similar in all cases, and the identity of the PC affinity between the F_v fragment and the whole antibody has been shown (Skerra et al., 1988; Glockshuber et al., 1990a). Therefore, the observed differences on K_M and k_{cat} are likely due to differences in the structure of the three binding sites.

For S107, an antibody very closely related in sequence to T15, several mutants were constructed (Jackson et al., 1991) and the kinetic parameters investigated. The mutation TyrH33 → His improved the catalytic efficiency, presumably by acting as a general base, and did not alter the inhibition constant by PC. It remains to be resolved, why direct measurements of PC binding to the TyrH33 → His mutant in M603, T15 and M167 always indicated much poorer binding (Stadlmüller, Pfitzinger, Schweder and Plückthun, unpublished).

The catalytic reaction in the w.t. antibodies did not seem to involve any kind of acid/base catalysis, as indicated by the pH-dependence (Pollack et al., 1987). Rather, a direct attack by OH^- is consistent with the kinetic data. The antibody then merely acts as a "super-solvent", providing an environment more suitable

Table 4. Kinetic constants of anti-phosphorylcholine antibodies in the hydrolysis of *p*-NO_2-phenylcholine carbonate

Antibody	fragment	produced in	k_{cat} (min^{-1})	K_M (mM)	Ref.
T15	IgA	mouse	0.32	0.71	Pollack & Schultz, 1987
M167	IgA	mouse	0.4	0.21	Pollack et al., 1986
M603	F_v	E. coli	0.045	1.3	Plückthun et al., 1990

for the transition state than the solvent water does. The rather modest catalytic efficiency is very reminiscent of subtilisin having been stripped of its catalytic triad (see above).

In more general terms, the use of molecular biology, and particularly *E. coli* expression systems, is likely to expand the possibilities in the field of catalytic antibodies greatly. It is unlikely that merely by raising monoclonal antibodies, i.e. by a selection for binding of a transition state analog (or, for that matter, any other substrate or product analog), general methods for the generation of functionalities leading to very efficient chemical catalysis with enzyme-like rate accelerations can be found. While there have been examples to the contrary (reviewed in Lerner et al., 1991), the use of protein engineering seems a more general alternative to improve upon a given antibody. First examples of this kind have been reported, such as the introduction of a metal binding site (Iverson et al., 1990), but in the latter case, catalysis has not yet been demonstrated.

In the long run, the bacterial expression technology may lead to direct screening and selection methods for catalytic activity. While the levels of catalytic activity of most catalytic antibodies are currently still too low to be detectable, this may be overcome by continued improvements in expression technology.

8 Avidity

It is occasionally surmised that the murine anti-phosphorylcholine response leads to antibodies with "poor" affinity, similar to the "low" affinity seen for most anti-polysaccharide antibodies. Yet, they are reproducibly elicited, when the animal is immunized with inactivated PC-carrying microorganisms, and they are generally believed to be of protective value. The solution of this apparent dilemma lies in the multivalence of antibodies. Briefly, after binding of one "arm" of the antibody to the surface polysaccharide at a phosphorylcholine moiety, the other "arm" now senses a higher local concentration of phosphorylcholine residues, since they are regularly spaced on the surface polysaccharide.

There have been attempts to quantify this phenomenon (Crothers and Metzger, 1972; Karush, 1976).

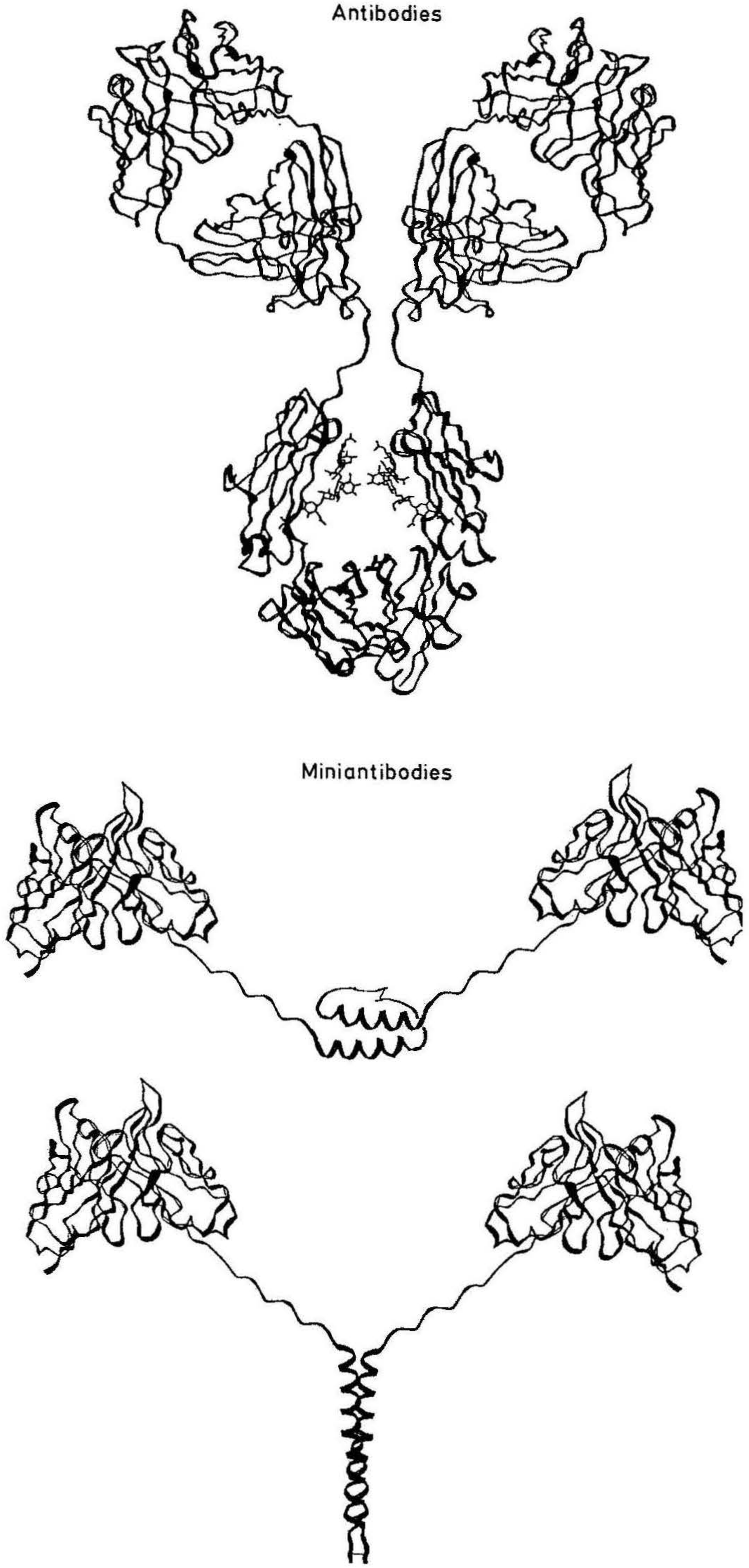
The quantitation problem essentially consists of the calculation of the increase in local concentration after the first "arm" has been bound. Although crystal structures of antigen binding sites are available (Reviews: Alzari et al., 1988, Davies et al., 1990), the structural knowledge of the constant F_c part and the geometry and flexibility of the F_{ab} fragments (the "arms") is far less advanced (for a review see e.g. Burton, 1990). For instance, the energy required to "bend" the antigen or the antibody is not easily computed. Therefore, a fair number of adjustable parameters exist, and it is not easy to make meaningful comparisons between theory and experiment. Furthermore, the spacing of the surface bound antigen in two dimensions is not accurately known, either.

It has been observed that anti-phosphorylcholine antibodies show rather different affinities to different phosphorylcholine containing polysaccharides, although the *thermodynamic* binding affinity to phosphorylcholine itself is much more similar (see e.g. Glaudemans et al., 1977; Manjula and Glaudemans, 1978; Andres et al., 1981; Claflin et al., 1985). It is not quite clear from these data, however, whether this is due to the spacing or the orientation of the phosphorylcholine moieties on the different polysaccharide surfaces, or whether some of the differences might be due to additional direct interactions with antigenic determinants in the polysaccharide, other than with the phosphorylcholine moiety itself.

The importance of multivalence, however, is directly seen experimentally, since no specific surface binding in assays like ELISA is detected for monomeric fragments. Apparently, they would be only poorly resistant to the required washing steps. This is not only true for the interaction with phosphorylcholine-containing surface polysaccharides, but also for phosphorylcholine haptens bound to a carrier-protein, used for coating the surface of the well. Nature's antibody design has solved this problem of divalence by using the C_H2 and C_H3 domains (Fig. 1, Fig. 4) which dimerize to form homodimers, and thereby dimerize two F_{ab} fragments to give the familiar Y-shape of the antibody.

For a variety of reasons, however, alternative means of dimerization have now been investigated. (1) The C_H2 domain of most antibody classes is glycosylated and the domains only interact via their sugar residues (reviewed in: Padlan, 1990b). This means that an analogous situation cannot be achieved in prokaryotic expression hosts, which are incapable of glycosylation. (2) The F_c part is the place of many biological effector functions of an antibody, including binding to the class-specific F_c receptors and complement (Morgan and Weigle, 1987). There may be instances when it is desirable to uncouple bivalence and effector functions. (3) The small size of functional F_v fragments, when combined

Fig. 14. *Top.* Molecular model of the whole antibody KOL. *Middle*: Molecular model of the dimeric miniantibody construct of M603 with antiparallel bundle helices. *Bottom*: Molecular model of the dimeric miniantibody construct carrying leucince zipper extensions. For details, see Pack and Plückthun, 1992


Antibodies
Miniantibodies

with a small dimerization domain, might take advantage of the multivalence benefits without sacrificing the small size. The small size is important in medical applications in tumor diagnostics and therapy, as it leads to altered pharmakokinetics, better tumor penetration and possibly low immunogenicity (Colcher et al., 1990).

The use of amphiphilic helices as dimerization devices was found advantageous and compatible with the above requirements (Pack and Plückthun, 1992) (Fig. 14). Both the helix from a previously designed four-helix-bundle (Eisenberg et al., 1986) and the coiled-coil helix from a leucine zipper, the natural dimerization motif of eukaryotic transcription factors (O'Shea et al., 1991). Although there is no direct structural evidence yet, it is expected from the design that the helices from the helix-bundle arrange in an antiparallel fashion (by analogy with the four-helix-bundle), while the helices from the leucine zipper are a parallel coiled-coil (O'Shea et al., 1991). Both constructs have also been stabilized covalently with an extension carrying a cysteine residue that forms a disulfide bond in vivo. These designed divalent fragments were found to spontaneously form in vivo in *E. coli* by ultracentrifugation and gel filtration measurements. They are largely dimers, and show the expected increase in avidity, almost equal to a natural antibody at a molecular weight of about a single F_{ab} fragment.

9 Conclusions and Future Approaches

The advances in molecular biology and especially production in bacteria, summarized in Sect. 4, have generated a very powerful set of tools for manipulating antibodies with far greater ease than previously possible. This has also lead to a more facile access to structural studies.

The anti-PC response of the mouse can be taken as a paradigm for an immune response. By the continued use of site-directed and random mutagenesis, it may be possible to generate a complete structural picture of an immune response including the structural and energetic changes brought about by somatic mutations. It is likely that many structural conclusions derived from studying an anti-hapten response in detail can be applied to anti-protein responses as well. There are many technical reasons why the use of these anti-PC antibodies turned out to be beneficial, mostly the availability of a three-dimensional structure from the outset of these studies (Segal et al., 1974).

At the time of writing, the antibody McPC603 was still the only catalytic antibody with a known structure. While this will undoubtedly change, engineering experiments in this system could be used to clarify the importance of defined interactions in stabilizing the ground state and the transition state complexes.

A careful analysis of this selection-based system of antibody binding to foreign molecules will certainly be an inspiration to the bioorganic chemist for some time to come.

References

Alzari PM, Lascombe M-B, Poljak RJ (1988) Ann Rev Immunol 6:555–580.
Andres CM, Maddalena A, Hudak S, Young NM, Claflin JL (1981) J Exp Med 154:1584–1598.
Arnold, FH (1991) Biotechnology 9:151–156.
Bennett LG, and Glaudemans CJ (1979) Biochemistry 18:3337–3342.
Behr T, Fischer W, Peter-Katalinic J, and Egge H (1992) Eur J Biochem 207, 1063–1075.
Better M, Chang CP, Robinson RR, and Horwitz AH (1988) Science 240:1041–1043.
Berry MJ, and Davies J (1992) J Chromatography 597:239–245.
Bhat TN, Bentley GA, Fischmann TO, Boulot G, and Poljak RJ (1990) Nature 347:483–485.
Borgford TJ, Gray TE, Brand NJ, and Fersht AR (1987) Biochemistry 26:7246–7250.
Branden C, Tooze J (1991) Introduction to protein structure, Garland, New York.
Bruccoleri RE, Haber E, and Novotny J (1988) Nature 335:564–568.
Bruccoleri RE, and Karplus M (1987) Biopolymers 26:137–168.
Buchner J, and Rudolph R (1991) Biotechnology 9:157–162.
Burley SK, and Petsko GA (1985) Science 229:23–28.
Burton DR (1990) in "F_c receptors and the action of antibodies", Am Soc Microbiol, Washington, DC, 31–54.
Calabi F, Neuberger MS (eds) (1987) Molecular genetics of immunoglobulin, Elsevier, Amsterdam (New Comprehensive Biochemistry 17).
Carter P, and Wells JA (1988) Nature 332:564–568.
Carter P, and Wells JA (1990) Proteins 7:335–342.
Chien NC, Roberts VA, Giusti AM, Scharff MD, and Getzoff ED (1989) Proc Natl Acad Sci USA 86:5532–5536.
Chothia C, and Lesk AM (1987) J Mol Biol 196:901–917.
Chothia C, Lesk, AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, Colman PM, Spinelli S, Alzari PM, and Poljak RJ (1989) Nature 342:877–883.
Claflin JL, and Berry J (1988) J Immunol 141:4012–4019.
Claflin JL, Berry J, Flaherty D, and Dunnick W (1987) J Immunol 138:3060–3068.
Claflin JL, George J, Dell C, and Berry J (1989) J Immunol 143:3054–3063.
Claflin JL, Hudak S, and Maddalena A (1981) J Exp Med 153:352–364.
Claflin JL, Hudak S, Maddalena A, and Bender T (1985) J Immunol 134:2536–2543.
Clarke SH, Claflin JL, Potter M, and Rudikoff S (1983) J Exp Med 157:98–113.
Clarke SH, and Rudikoff S (1984) J Exp Med 159:773–782.
Co MS, Deschamps M, Whitley RJ, and Queen C (1991) Proc Natl Acad Sci USA 88:2869–2873.
Colcher D, Bird R, Roselli M, Hardman KD, Johnson S, Pope S, Dodd SW, Pantoliano MW, Milenic DE, and Schlom J (1990) J Natl Cancer Inst 82:1191–1197.
Cook WD, Rudikoff S, Giusti AM, and Scharff MD (1982) Proc Natl Acad Sci USA 79:1240–1244.
Cosenza H, and Köhler H (1972) Proc Natl Acad Sci USA 69:2701–2705.
Crews S, Griffin J, Huang H, Calame K, and Hood L (1981) Cell 25:59–66.
Crothers DM, and Metzger H (1972) Immunochemistry 9:341–357.
Davies DR, Sheriff S, and Padlan EA (1988) J Biol Chem 263:10541–10544.
Davies DR, Padlan EA, and Sheriff S (1990) Ann Rev Biochem 59:439–473.
Eisenberg D, Wilcox W, Eshita SM, Pryciak PM, Ho SP, and DeGrado WF (1986) Proteins 1:16–22.
Esser C, and Radbruch A (1990) Ann Rev Immunol 8:717–735.
Fine RM, Wang H, Shenkin PS, Yarmush DL, and Levinthal C (1986) Proteins 1:342–362.
Fischer W (1988) Adv Micr Physiol 29:233–302.
Feeney AJ, Clarke SH, and Mosier DE (1988) J Immunol 141:1267–1272.
Freund C, Ross A, Guth B, Plückthun A, and Holak T (1993) FEBS Lett 320:97–100.
Gearhart PJ, and Bogenhagen DF (1983) Proc Natl Acad Sci USA 80:3439–3443.
Gearhart PJ, Johnson ND, Douglas R, and Hood L (1981) Nature 291:29–34.
Gearhart PJ, Lawler AM, Levy NS, Lebecque SG, and Malipiero UV (1988) Ann NY Acad Sci 546:57–62.
Gettins P, Potter M, Leatherbarrow RJ, and Dwek RA (1982) Biochemistry 21:4927–4931.
Gettins P, Potter M, Rudikoff S, and Dwek RA (1977) FEBS Lett 84:87–91.
Glaudemans CPJ, Manjula BN, Bennett LG, and Bishop CT (1977) Immunochemistry 14:675–679.
Glockshuber R, Malia M, Pfitzinger I, and Plückthun A (1990a) Biochemistry 29:1362–1367.

Glockshuber R, Stadlmüller J, and Plückthun A (1991) Biochemistry 30:3049–3054.
Glockshuber R, Steipe B, Huber R, and Plückthun A (1990b) J Mol Biol 213:613–615.
Go N, and Sheraga HA (1970) Macromolecules 3:178–187.
Goetze AM, and Richards JH (1977b) Biochemistry 16:228–232.
Goetze AM, and Richards JH (1978) Biochemistry 17:1733–1739.
Goetze AM, and Richards JH (1977a) Proc Natl Acad Sci USA 74:2109–2112.
Goodsell DS, and Olson AJ (1990) Proteins 8:195–202.
Gorman SD, Clark MR, Routledge EG, Cobbold SP, and Waldmann H (1991) Proc Natl Acad Sci USA 88:4181–4185.
Gu H, Kitamura D, and Rajewsky K (1991) Cell 65:47–54.
Haldane JBS (1930) Enzymes, Longmans, Green, and Co. London; (1965) MIT Press, Cambridge, Mass.
Hamel PA, Klein MH, and Dorrington J (1986) Mol Immunol 23:503–510.
Hasemann CA, and Capra JD (1990) Proc Natl Acad Sci USA 87:3942–3946.
Hiatt A, Cafferkey R, and Bowdish K (1989) Nature 342:76–78.
Hilbert DM, and Cancro MP (1988) J Immunol 140:4364–4371.
Hochman J, Gavish M, Inbar D, and Givol D (1976) Biochemistry 15:2706–2710.
Holm L, and Sander C (1992) J Mol Biol 225:93–105.
Holmberg D, Freitas AA, Portnoi D, Jacquemart F, Avrameas S, and Coutinho A (1986) Immunol Rev 93:147–169.
Hoogenboom HR, Marks JD, Griffiths AD, and Winter G (1992) Immunol Rev 130:41–68.
Horne C, Klein M, Polidoulis I, and Dorrington KJ (1982) J Immunol 129:660–664.
Horwitz AH, Chang CP, Better M, Hellstrom KE, and Robinson RR (1988) Proc Natl Acad Sci USA 85:8678–8682.
Huston JS, McCartney J, Tai MS, Mottola-Hartshorn C, Jin D, Warren F, Keck P and Oppermann H (1993) Int Rev Immunol, in press.
Ichihara Y, Hayashida H, Miyazawa S, and Kurosawa Y (1989) Eur J Immunol 19:1849–1854.
Inman JK, Merchant B, and Tacey SE (1973a) Immunochemistry 10:153–163.
Inman JK, Merchant B, Claflin L and Tacey SE (1973b) Immunochemistry 10:165–174.
Iverson BL, Iverson SA, Roberts VA, Getzoff ED, Tainer JA, Benkovic SJ, and Lerner RA (1990) Science 249:659–662.
Jackson DY, Prudent JR, Baldwin EP, and Schultz PG (1991) Proc Natl Acad Sci USA 88:58–62.
James MNG, and Sielecki AR (1983) J Mol Biol 163:299–361.
Janin J, Wodak S, Levitt M, and Maigret B (1978) J Mol Biol 125:357–386.
Jencks WP (1969) in Catalysis in Chemistry and Enzymology. McGraw-Hill, New York, pp.282.
Jencks WP (1975) Adv Enzymol 43:219–410.
Jennings HJ, Lugowski C, and Young NM (1980) Biochemistry 19:4712–4719.
Jones PT, Dear PH, Foote J, Neuberger MS, and Winter G (1986) Nature 321:522–525.
Jones TA, and Thirup S (1986) EMBO J 5:819–822.
Kabat EA, Wu TT, Perry HM, Gottesmann KS and Foeller C (1991) in Sequences of Proteins of Immunological Interest, 5th ed. US Dept of Health and Human Services, Public Health Service, Natl Inst of Health, Bethesda, MD.
Karush F (1976) Contemp Top Mol Immunol 5:217–228.
Kettleborough CA, Saldanha J, Heath VJ, Morrison CJ, and Bendig MM (1991) Protein Eng 4:773–783.
Kirkpatrick S, Gelat CD, Jr and Vecchi MP (1983) Science 220:671–680.
Klein M, Kortan C, Kells DIC, and Dorrington KJ (1979) Biochemistry 18:1473–1481.
Klinman NR, and Press JL (1975) Transplant Rev 24:41–83.
Knowles JR (1991) Nature 350:121–124.
Knappik A. Krebber C, and Plückthun A (1993) Biotechnology 11:77–83.
Kobrin BJ, Buhl S, Shulman MJ, and Scharff MD (1991) J Immunol 146:2017–2020.
Kocks C, and Rajewsky K (1989) Ann Rev Immunol 7:537–559.
Köhler G, and Milstein C (1975) Nature 256:495–497.
Köhler J, Kikuno R, and Plückthun A (1991) In:Maksic ZB, Eckert-Macsic M (eds) Molecules in natural science and medicine, an encomium for Linus Pauling. Ellis Horwood, New York, p441–465.
Kraut J (1988) Science 242:533–540.
Kurosawa Y, and Tonegawa S (1982) J Exp Med 155:201–218.
Lee FS, Chu ZT, Bolger MB, and Washel A (1992) Protein Eng 5:215–228.
Leon MA, and Young NM (1971) Biochemistry 10:1424–1429.

Lerner RA, Benkovic SJ, and Schultz PG (1991) Science 252:659–667.
Lindner P, Guth B, Wülfing C, Krebber C, Steipe B, Müller F, and Plückthun A (1992) Methods: A companion to Meth Enzymol 4:41–56.
Lolis E, and Petsko GA (1990) Ann Rev Biochem 59:597–630.
Maeda H, Matsushita S, Eda Y, Kimachi K, Tokiyoshi S, and Bendig MM (1991) Hum Antib Hybridomas 2:124–134.
Malipiero UV, Levy NS, and Gearhart PJ (1987) Immunological Rev 96:59–74.
Manjula BN, and Glaudemans CPJ (1978) Immunochemistry 15:269–271.
Martin ACR, Cheetham JC, and Rees AR (1989) Proc Natl Acad Sci USA 86:9268–9272.
McCafferty J, Griffiths AD, Winter G, and Chiswell DJ (1990) Nature 348:552–554.
McNeil HP, Chesterman CN, and Krilis SA (1991) Adv Immunol 49:193–280.
Mercolino TJ, Arnold LW, and Haughton G (1986) J Exp Med 163:155–165.
Metzger H, Chesebro B, Hadler NM, Lee J, Otchin N (1971) In:Amos B (ed) Progress in immunology: Proceedings of the 1st International Congress of Immunology, Academic Press, New York, p253–276.
Mian IS, Bradwell AR, and Olson AJ (1991) J Mol Biol 217:133–151.
Morgan EL, and Weigle WO (1987) Adv Immunol 40:61–134.
Morrison SL, and Oi VT (1989) Adv Immunol 44:65–92.
Novotny J, Rashin AA, and Bruccoleri RE (1988) Proteins 4:19–30.
Novotny J, and Sharp K (1992) Prog Biophys Molec Biol 58:203–224.
O'Shea EK, Klemm JD, Kim PS, and Alber T (1991) Science 254:539–544.
Pabo CO, and Suchanek EG (1986) Biochemistry 25:5987–5991.
Pack P, and Plückthun A (1992) Biochemistry 31:1579–1584.
Padlan EA (1990a) Proteins 7:112–124.
Padlan EA (1990b) In: "F_c receptors and the action of antibodies", Am Soc Microbiol, Washington, DC, 12–30.
Padlan EA, Davies DR, Rudikoff S, and Potter M (1976) Immunochemistry 13:945–949.
Page MJ, and Sydenham MA (1991) Biotechnology 9:64–68.
Pauling L (1946) Chem Eng News 24:1375–1377.
Perlmutter RM, Crews ST, Douglas R, Sorensen G, Johnson N, Nivera N, Gearhart PJ, and Hood L (1984) Adv Immunol 35:1–37.
Péry P, Petit A, Poulain J, and Luffau G (1974) Eur J Immunol 4:637–639.
Plückthun A (1991a) Biotechnology 9:545–551.
Plückthun A (1991b) Methods 2:88–96.
Plückthun A (1992) Immunol Rev 130:151–188.
Plückthun A, Glockshuber R, Pfitzinger I, Skerra A, and Stadlmüller J (1987) Cold Spring Harbor Symp Quant Biol 52:105–112.
Plückthun A, Glockshuber R, Skerra A, and Stadlmüller J (1990) Behring Inst Mitt 87:48–55.
Plückthun A, and Pfitzinger I (1991) Annals New York Acad Sci 646:115–124.
Plückthun A, and Stadlmüller J (1991) Catalytic antibodies, CIBA Symposium series 159:103–117.
Pollack SJ, Jacobs JW, and Schultz PG (1986) Science 234:1570–1573.
Pollack SJ, and Schultz PG (1987) Cold Spring Harbor Symp Quant Biol 52:97–104.
Pollet R, and Edelhoch H (1973) J Biol Chem 248:5443–5447.
Pollet R, Edelhoch H, Rudikoff S, and Potter M (1974) J Biol Cem 249:5188–5194.
Ponath PD, Boyd RT, Hillis DM, and Gottlieb PD (1989) Immunogenetics 29:389–396.
Potter M (1970) Fed Proc 29:85–91.
Potter M (1971) Ann NY Acad Sci 190:306–321.
Potter M (1972) Physiol Rev 52:631–719.
Potter M, and Leon MA (1968) Science 162:369–371.
Potter M, and Lieberman R (1970) J Exp Med 132:737–751.
Putlitz J, Kubasek WL, Duchene M, Marget M, von Specht B-U, and Domdey H (1990) Biotechnology 8:651–654.
Queen C, Schneider WP, Selick HE, Payne PW, Landolfi NF, Duncan JF, Avdalovic NM, Levitt M, Junghans RP, and Waldmann TA (1989) Proc Natl Acad Sci USA 86:10029–10033.
Reth M, and Leclercq L (1987) In:Mol genetics of immunoglobulin. Elsevier, Amsterdam, 111–134 (New Comprehensive Biochemistry 17)
Riechmann L, Clark M, Waldmann H, and Winter G (1988) Nature 332:323–327.
Rodwell JD, Gearhart PJ, and Karush F (1983) J Immunol 130:313–316.
Rudikoff S, Giusti AM, Cook WD, and Scharff MD (1982) Proc Natl Acad Sci USA 79:1979–1983.
Russell AJ, and Fersht AR (1987) Nature 328:496–500.

Russell AJ, Thomas PG, and Fersht AR (1987) J Mol Biol 193:803–813.
Satow Y, Cohen GH, Padlan EA, and Davies DR (1986) J Mol Biol 190:593–604.
Segal DM, Padlan EA, Cohen GH, Rudikoff S, Potter M, and Davies DR (1974) Proc Natl Acad Sci USA 71:4298–4302.
Sharp KA, and Honig B (1990) Annu Rev Biophys Biophys Chem 19:301–332.
Shenkin PS, Yarmush DL, Fine RM, Wang H, and Levinthal C (1987) Biopolymers 26:2053–2085.
Shokat KM, and Schultz PG (1990) Ann Rev Immunol 8:335–363.
Skerra A, Pfitzinger I, and Plückthun, A (1991) Biotechnology 9:273–278.
Skerra A, and Plückthun A. Protein Engineering 4:971–979.
Skerra A, Plückthun A (1988) Science 240:1038–1041.
Solin ML, and Kaartinen M (1992) Immunogenetics 36:306–313.
Steipe B, Plückthun A, and Huber R (1992) J Mol Biol 225:739–753.
Stenzel-Poore MP, Hall TJ, Heusser CH, Faust CH, and Rittenberg MB (1987) J Immunol 139:1698–1703.
Summers NL, Carlson WD, and Karplus M (1987) J Mol Biol 196:175–198.
Summers NL, and Karplus M (1989) J Mol Biol 210:785–811.
Tempest PR, Bremner P, Lambert M, Taylor G, Furze JM, Carr FJ, and Harris WJ (1991) Biotechnology 9:266–271.
Tonegawa S (1983) Nature 302:575–581.
Tramontano A, Chothia C, and Lesk AM (1990) J Mol Biol 215:175–182.
van Gunsteren WF, and Berendsen HJC (1990) Angew Chemie (Int Ed) 29:992–1025.
Verhoeyen M, Milstein C, and Winter G (1988) Science 239:1534–1536.
Ward ES, Güssow D, Griffiths AD, Jones PT, and Winter G (1989) Nature 341:544–546.
Williams KR, and Claflin JL (1982) J Immunol 128:600–607.
Wilson IA, Stanfield RL, Rini JM, Arevalo JH, Schulze-Gahmen U, Fremont DH, and Stura EA (1991) in Catalytic Antibodies, Ciba Foundation Symposium 159:13–39.
Wolfenden R (1976) Ann Rev Biophys Bioeng 5:271–306.
Wolfe J, and Claflin JL (1980) J Immunol 125:2397–2401.
Wu TT, and Kabat EA (1970) J Exp Med 132:211–250.
Yancopoulos GD, and Alt FW (1986) Ann Rev Immunol 4:339–368.
Young NM, and Leon MA (1977) Immunochemistry 14:757–761.
Young NM, Williams RE, and Claflin JL (1985) Mol Immunol 22:305–311.