Structural Studies of Immunoglobulins

R. R. Porter

In 1946, when I was starting work as a research student under the supervision of F. Sanger, the second edition of Landsteiner's book *The Specificity of Serological Reactions* (*1*) reached England. In it was summarized the considerable body of information available on the range of antibody specificity, and much of it was Landsteiner's own work or that by others using his basic technique of preparing antibodies against haptens and testing the ability of the hapten to inhibit the precipitation of the antiserum and the conjugated protein. Also described in this book was the work in Uppsala of Tiselius and Pederson in collaboration with Heidelberger and Kabat, in which they showed that all rabbit antibodies were in the γ globulin fraction of serum proteins and that they had a molecular weight of 150,000. This combination of an apparently infinite range of antibody combining specificity associated with what appeared to be a nearly homogeneous group of proteins astonished me and indeed still does.

Active Fragments of Antibodies

The preparation of antibodies by dissociation of specific precipitates with strong salt solutions or in acidic conditions had been described, as had the preparation in fair yield of γ globulin fractions by salting-out techniques from whole serum, so an experimental approach to the structural basis of antibody combining specificity was possible. A start had indeed been made by showing that the whole molecule was not required for the combining specificity. Parventjev (*2*) had introduced pepsin treatment of serum as a commercial method of purification of horse antitoxins, and Petermann and Pappenheimer (*3*) studied the reaction but used purified horse antibody against diphtheria toxin rather than whole serum for the peptic digestion. They showed that a product able to flocculate with the toxoid or neutralize toxin could be obtained and that it had a molecular weight of 113,000, that is, substantially less than that of the original molecule. Petermann (*4*) showed later than human γ globulin could be split by papain to give what she estimated, by using the ultracentrifuge, to be quarter molecules. No antibody activities were, however, investigated in this study.

At about the same time, Landsteiner (*5*) was extending his investigation of the antigenic specificity of protein antigens and had found that crude but apparently low-molecular-weight peptides from one acid digest of silk fibroin could inhibit the precipitation of soluble fibroin with its rabbit antiserum. This finding, in conjunction with the hapten studies and others, suggested that antigenic sites—and presumably, therefore, antibody-combining sites—were small, certainly very much smaller than the antibody molecules; further attempts to obtain antibody fragments that retained the power to combine with the antigen seemed worthwhile. Testing for such active fragments was by their ability to inhibit the combination of the antigen and whole antibody. However, although a variety of conditions of hydrolysis by acids or enzymes were investigated (*6*), only papain gave an active product; this appeared, from NH_2-terminal amino acid assay, to be the quarter molecules previously described by Petermann (*4*). There was no doubt that the combining site was in these smaller fragments, and hence a substantial reduction in the magnitude of the structural problem had been achieved, but protein molecules of molecular weight 40,000 were still a formidable prospect. This work was carried out in Sanger's laboratory in Cambridge. With his guidance, NH_2-terminal amino acid assay and determination of the terminal sequences were attempted but proved unhelpful in that they suggested that the rabbit γ globulins and antibodies had a single open polypeptide chain and that the biologically active quarter molecules also had the same alanine NH_2-terminal acid. The possibility that there might be blocked NH_2-terminal residues was not considered.

A return to papain digestion of rabbit γ globulin was made 7 years later, but in place of the crude enzyme preparation used earlier, a crystalline enzyme (*7*) was used in much lower concentrations, at one-hundredth the weight of substrate (*8*). Under these conditions a number of points missed previously became apparent. First, there was a very high recovery of total protein after dialysis of the digest—very few small peptides were formed in spite of the wide specificity of the enzyme. The products, all of very similar size (sedimentation coefficient 3.5*S*), were one-third the size of the original rather than one-quarter. Most surprising, one of the products of digestion crystallized very easily in diamond-shaped plates (Fig. 1) during the dialysis at neutrality in the cold room. This last observation, suggesting that a protein that itself would never crystallize could give a fragment that presumably was more homogeneous and hence able to crystallize, was quite unexpected and indeed was unacceptable. The crystals were dismissed as coming from the less soluble amino acids and were discarded without further consideration for several months. Fortunately, my neighbor in the adjoining laboratory at the National Institute of Medical Research in London was the x-ray crystallographer Olga Kennard, and when I eventually asked her opinion she immediately gave the view that they were protein crystals. They were then identified as the material in the third peak obtained by fractionation of the digest products on carboxymethyl cellulose (Fig. 2). They were named fraction III and are now known as the Fc fragment. Fractions I and II were the components of the digest which retained the combining specificity of the original antibody. Fraction

Copyright © 1973 by the Nobel Foundation. The author is professor of biochemistry, University of Oxford, Oxford, England. This article is the lecture he delivered in Stockholm, Sweden, on 11 December 1972, when he received the Nobel Prize in Physiology or Medicine, a prize he shared with Professor Gerald M. Edelman. Minor corrections have been made by the author. The article is published here with the permission of the Nobel Foundation and will also be included in the complete volume of *Les Prix Nobel in 1972* as well as in the series Nobel Lectures (in English) published by the Elsevier Publishing Company, Amsterdam and New York. Professor Edelman's lecture will be published in a subsequent issue.

III had no such activity, but did carry most of the antigenic specificity of the rabbit γ globulins when tested with antiserum from goats, rats, and guinea pigs. More detailed studies (*9*) showed that fractions I and II were very similar chemically and antigenically, and later Nisonoff and colleagues (*10*) showed that the distinction between these two fractions was artificial. If a basic fraction of γ globulin was treated with papain, two molecules of II and one of III were obtained, and an acidic fraction of γ globulin gave two molecules of I and one of III. The slight differences in change between fractions I and II reflected the charge heterogeneity of the starting material.

Nisonoff *et al.* (*11*) had also returned to the peptic digestion of γ globulins. Using the rabbit protein rather than horse antitoxin, they showed that on reduction the product with molecular weight of 100,000, comparable to that reported earlier, would give half molecules very similar to fractions I and II. The latter are now named Fab, and the peptic digest product is termed $(Fab')_2$.

The papain digest studies established that γ globulin, now named immunoglobulin gamma or IgG, was formed from three globular sections that were probably rather tightly folded, inasmuch as they were exceptionally resistant to further degradation by papain. The Fc fragment was apparently common to all molecules, whereas the two identical Fab fragments each carried a combining site and with it the inherent variability associated with the whole antibody. An attempt was made (*9*) to relate this tripartite structure with the supposedly single open polypeptide structure deduced from end group analysis of rabbit IgG. Of course, it made no sense, and progress depended upon the demonstration by Edelman (*12*) that human IgG and therefore presumably IgG's of all species were multichain proteins. It followed that there must be blocked NH_2-terminal amino acids and that estimation of the free NH_2-terminal amino acid was of only limited significance.

The Four-Peptide-Chain Structure

The solution of the gross structure of immunoglobulins depended upon establishing the relationship between the peptide chains identified by Edelman and the products of papain digestion. This was achieved easily when

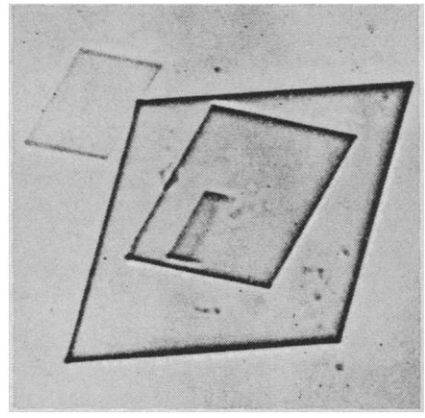


Fig. 1. Crystals formed from a papain digest of rabbit IgG, the Fc fragment.

the conditions of isolation of the chains were modified by using reduction in the absence of denaturing agents—conditions in which predominantly interchain disulfide bonds are broken. No fall in molecular weight followed such reduction, but the chains were dissociated and separated when chromatographed on Sephadex columns in acetic or propionic acid to give heavy and light chains with molecular weights of approximately 50,000 and 20,000, respectively. These chains now remained soluble at neutrality and retained antigenic specificities. A double diffusion plate using antiserum specific to Fab or Fc showed that Fab contained antigenic sites common to both heavy and light chains, but Fc had sites common to heavy chains only (Fig. 3). This led to the postulated four-chain structure (*13*) (Fig. 4). More detailed studies were in agreement with this structure (*14*) and confirmed the position of

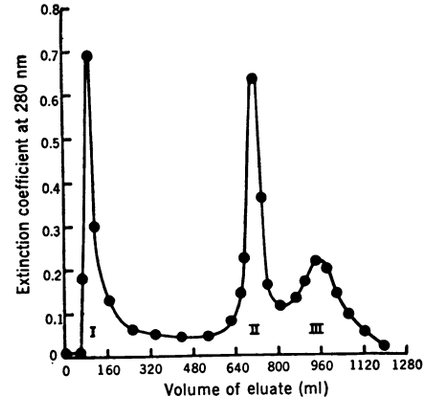


Fig. 2. Fractionation of a papain digest of rabbit IgG on carboxymethyl cellulose in sodium acetate buffer at *p*H 5.5, with a gradient from 0.01 to 0.9*M*. Fractions I and II (Fab) carry the antibody combining sites, and fraction III (Fc) will crystallize easily.

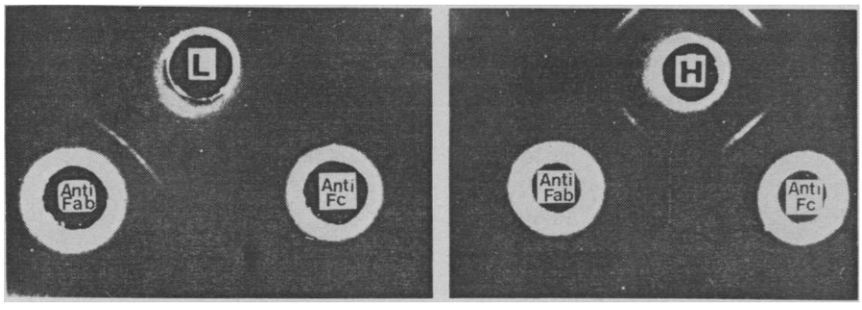
papain hydrolysis to be about the midpoint of the heavy chain.

It was some years before the rather complex arrangement of the inter- and intrachain disulfide bonds of the rabbit heavy chains was resolved (*15*) (Fig. 5). The mechanism of blocking of the alpha-amino groups of the NH_2-terminal amino acids also proved more difficult to establish than expected, as it was found to be due to the ringed residue pyrrolidonecarboxylic acid (PCA) (*16*), which was well known as an artifact arising from NH_2-terminal glutamine residues. All attempts to find NH_2-terminal glutamine in IgG were unsuccessful, even when it was isolated under conditions in which conversion to PCA during handling appeared to be excluded. There is evidence that glutamine is the residue incorporated into the peptide chain during synthesis (*17*), but PCA appears to be the terminal residue present in the immunoglobulins in the blood; it has been suggested that enzymatically catalyzed cyclization occurs intracellularly. It has been assumed that PCA is the only blocked NH_2-terminal residue in immunoglobulin molecules from all species, but there appears to have been little careful study.

Antibody Combining Sites

While the four-chain model clarified many aspects of the structure of antibodies, it made no contribution to our understanding of what features gave the possibility of forming antibodies of innumerable different specificities. It seemed at the time that the difficulties increased as the possibilities of variation were reduced. I made repeated unsuccessful attempts (*18*) to obtain digest products of Fab which still bound the antigen; in fact, progress has just been reported in isolating a peptic digest product of a mouse myeloma protein, MOPC 315. This fragment, named Fv, appears to be formed from the NH_2-terminal half of the Fab molecule—that is, of the light and heavy chains, and retains its full affinity for a dinitrophenyl hapten (*19*).

Understanding of the origin of the multiple binding sites came of course with the discovery of the phenomenon of the variable and constant parts in Bence Jones proteins (*20*). Earlier it had been shown that the urinary Bence Jones proteins were the equivalent of the light chains of the myeloma protein in the blood of the same patient



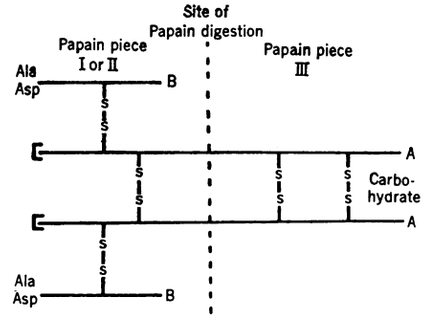


Fig. 3 (left). Double diffusion of the heavy and light chains of rabbit IgG against goat antiserum to rabbit Fab and goat antiserum to rabbit Fc. The light chain reacts only with antiserum to Fab while the heavy chain reacts with both antiserums; that is, Fab contains parts of heavy and light chains while Fc contains only parts of the heavy chain. Fig. 4 (right). The four-chain structure of rabbit IgG postulated on the basis of the double diffusion experiment of Fig. 3 and supporting chemical evidence.

(*21*). This observation of the variability in amino acid sequence in the NH_2-terminal 107 residues of the human kappa Bence Jones proteins while the remainder were constant immediately made possible an understanding as to how millions of different combining sites could be formed within the same structural framework. That the phenomenon was common to the NH_2-terminal 110 residues or so of the heavy chains was shown (*22*), and it was clear that the combining site was likely to be formed from the variable sections of both these chains.

Many lines of evidence have been brought to bear in an attempt to define more precisely in chemical terms just which residues in the variable regions of the heavy and light chains are likely to be directly concerned in determining the specificity of the binding site. They have been reviewed recently (*23*), and so the main conclusions could perhaps be just listed here:

1) The size of the antigenic site appears to be of the order of a hexapeptide or hexasaccharide.

2) This is comparable to the size of the substrate of a hydrolytic enzyme such as lysozyme. In this enzyme 15 to 20 amino acid residues have been identified as probable "contact amino acids," that is, residues forming a bond with the substrate. Hence, this appears to be the likely number of residues to line the antibody combining site and to play a direct role in determining specificity. If any residue could occupy any of these 15 to 20 positions, the possible number of variants is indeed high.

3) There are at least three hypervariable sections in each of the variable regions in heavy and light chains. They have been demonstrated rather clearly in the plots of Kabat and Wu (*24*) of the frequency of occurrence of different residues in myeloma proteins along the 110 or so positions of the variable regions. This hypervariability is also apparent in sequence studies of the heavy chains of rabbit IgG, and both lines of evidence agree in suggesting that in most cases the hypervariability is confined to one or two positions; but in the region of position 96 to 110 of the heavy chains, four or five positions may be exceptionally variable. In the rabbit γ chain (heavy chain of IgG) there is a section in this position across which no satisfactory sequence has been obtained, presumably because of the complexity of the sequences (*25*).

4) Several pieces of evidence suggest that these six hypervariable sections in the two variable regions may be brought together in the intact molecule to contribute to the structure of the combining site. The most direct evidence comes from affinity labeling studies in which an antibody is allowed to bind a hapten to which a reactive group has been attached. Covalent reaction will follow, and after subsequent hydrolysis the labeled peptides can be identified and placed by comparison with the known amino acid sequence. A variety of affinity labeling techniques have been introduced and used both with natural antibodies and with a mouse myeloma protein showing high affinity for the dinitrophenyl group. Though the work is in some cases incomplete, all studies agree in that the labeled reagents are found attached to residues in or near one of the six hypervariable sections.

5) It is likely that there is a hydrophobic region adjacent to the combining site which may not contribute to specificity but could increase the affinity of binding for an appropriate antigenic site.

The precise details of the combining site must await the completion of the crystallographic studies on immunoglobulins and their fragments, now being undertaken in several laboratories. There will be considerable interest in seeing how far the above predictions from chemical studies prove to be correct when the full structures become available.

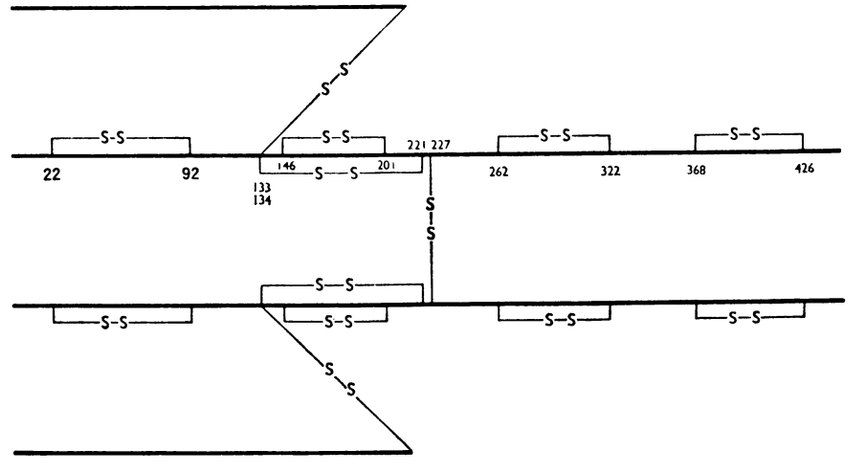


Fig. 5. The structure of rabbit IgG. The rather complex arrangement of inter- and intrachain disulfide bonds in the heavy chain is shown.

The Genetic Origin of the Multiple Forms of Antibodies

While the discovery of the variable region—and particularly of the hypervariable sections in it—seemed likely to answer the basic question about the structural origin of multiple binding specificities, it raised very difficult problems as to the genetic origin of these many different amino acid sequences. This topic has formed the basis of many discussions and reviews, and decisive evidence in favor of any of the many theories advanced is lacking. I only wish to make a brief comment here about the sequence work on the variable region of rabbit IgG heavy chains which developed as part of the other structural work discussed above.

There are perhaps two main points under immediate discussion. First, are heavy and light chains each the product of two genes, one coding for the variable region and one for the constant? Second, what are the multiple genes that code for the variable regions: germ line genes, or the product of somatic mutation of a much smaller number of germ line genes?

In each case it would be of obvious value if allelic variants able to act as genetic markers could be found in both the variable and constant regions. Markers in the constant regions have, in fact, been found in many of the different classes and subclasses of immunoglobulins of different species. The phenotypic character followed is the antigenic specificity of the immunoglobulin. This specificity has been correlated with amino acid changes, but no direct identification of the specificity and the sequence change—by demonstration of, say, the inhibitory power of a small peptide—has proved to be possible. Presumably a much larger section of molecule is necessary for the integrity of the antigenic site. However, in many cases there is no doubt that the genetic markers of the constant region have been found, inasmuch as they can be shown to be present in the Fc fragment. Allocation of genetic markers to the variable region is less clear, although it is likely that this is where the rabbit "a" locus allotypes originate.

In 1963, Todd (*26*) made the surprising observation that the "a" locus allelic specificities of rabbit immunoglobulins were common to IgG and IgM (an immunoglobulin with μ instead of γ as the heavy chain). As the structural work progressed it became clear that these antigenic specificities were carried by the γ and μ chains, although these chains differed obviously in chemical structure and hence in their structural genes. It then seemed possible that the "a" locus specificities could be determined by the variable regions that might be common to both chains. This observation of Todd's was the first to raise the possibility that two genes might be concerned in the structure of the heavy chain, and now two examples of crossovers out of about 400 offspring have been reported between "a" locus allelic specificities and specificities undoubtedly determined by structure of the constant parts of the γ chains. If extended, these studies will strengthen the evidence that two genes code for the γ chains. However, the establishment of the structural basis of the "a" locus specificities depends at present on correlation of amino acid sequences in the variable region with the specificity and lack of any similar correlation in the constant region sequences. This is rather indirect, but as far as it is acceptable, such correlation extending to about 16 positions in the γ chains has been found for pooled rabbit IgG (*27*). Work with homogeneous rabbit antibodies against polysaccharide has confirmed this correlation for some but not all these positions (*28*).

These "a" locus markers could clearly be of decisive importance in genetic studies, and indeed their existence in the variable region would be taken by some as strong evidence against the likelihood of many million copies of the variable regions being present in the germ line. It is, therefore, worthwhile to attempt to obtain unequivocal evidence that these apparent genetic markers of the variable region are indeed such. Alternatively, however, it should also be possible, using the chemical evidence now obtained, to follow directly the inheritance of a given allelic peptide rather than of an antigenic specificity. One such peptide that occurs in two forms among different rabbits can be identified rather easily by autoradiography after reaction of the half cystine residue at position 92 with [^{14}C]iodoacetic acid (*29*). Early evidence suggests that this peptide is indeed behaving as an heritable character; if confirmed, this will be direct evidence for a structural gene marker in the variable region. Further work should contribute to knowledge of the genetic origin of the variable region.

Some aspects of the structural studies of immunoglobulins have reached completion in that full chemical structures are now available for several human myeloma proteins and almost complete structures for rabbit immunoglobulins. The solution of the structural basis of the combining specificity of antibodies, which seemed to me the central problem, also appears to be nearing completion. There is still a role for structural work in the solution of the genetic origins of antibodies, and obviously there are many other applications not discussed here. Interaction of immunoglobulins with complement components and cell surfaces are two that are already arousing rapidly increasing interest.

References

1. K. Landsteiner, *The Specificity of Serological Reactions* (Harvard Univ. Press, Cambridge, Mass., 1946).
2. I. A. Parventjev, U.S. Patent No. 2,065,196 (1936).
3. M. L. Petermann and A. M. Pappenheimer, *J. Phys. Chem.* **45**, 1 (1941).
4. M. L. Petermann, *J. Amer. Chem. Soc.* **68**, 106 (1946).
5. K. Landsteiner, *J. Exp. Med.* **75**, 269 (1942).
6. R. R. Porter, *Biochem. J.* **46**, 479 (1950).
7. J. R. Kimmel and E. L. Smith, *J. Biol. Chem.* **207**, 515 (1954).
8. R. R. Porter, *Nature* **182**, 670 (1958).
9. ———, *Biochem. J.* **73**, 119 (1959).
10. J. L. Palmer, W. J. Mandy, A. Nisonoff, *Proc. Nat. Acad. Sci. U.S.A.* **48**, 49 (1962).
11. A. Nisonoff, F. C. Wissler, L. N. Lipman, D. L. Woernley, *Arch. Biochem. Biophys.* **89**, 230 (1960).
12. G. M. Edelman, *J. Amer. Chem. Soc.* **81**, 3155 (1959).
13. R. R. Porter, in *Basic Problems of Neoplastic Disease*, A. Gellhorn and E. Hirschberg, Eds. (Columbia Univ. Press, New York, 1962), pp. 177–194.
14. J. B. Fleischman, R. R. Porter, E. M. Press, *Biochem. J.* **88**, 220 (1963); M. J. Crumpton and J. M. Wilkinson, *ibid.*, p. 228; R. H. Pain, *ibid.*, p. 234.
15. I. J. O'Donnell, B. Frangione, R. R. Porter, *ibid.* **116**, 261 (1970).
16. E. M. Press, P. J. Piggot, R. R. Porter, *ibid.* **99**, 356 (1966); J. M. Wilkinson, E. M. Press, R. R. Porter, *ibid.* **100**, 303 (1966).
17. D. I. Stott and A. J. Munro, *ibid.* **128**, 1221 (1972).
18. R. R. Porter, *Brookhaven Symp. Biol.* **13**, 203 (1960).
19. D. Inbar, J. Hochman, D. Givol, *Proc. Nat. Acad. Sci. U.S.A.* **69**, 2659 (1972).
20. N. Hilschmann and L. C. Craig, *ibid.* **53**, 1403 (1965); K. Titani, E. Whitley, L. Avogardo, F. W. Putnam, *Science* **149**, 1090 (1965).
21. G. M. Edelman and J. A. Gally, *J. Exp. Med.* **116**, 207 (1962).
22. E. M. Press and N. M. Hogg, *Nature* **223**, 807 (1969); G. M. Edelman, B. A. Cunningham, W. E. Gall, P. D. Gottlieb, U. Rutishauser, M. J. Waxdal, *Proc. Nat. Acad. Sci. U.S.A.* **63**, 78 (1969).
23. R. R. Porter, in *Contemporary Topics in Immunochemistry*, F. P. Inman, Ed. (Plenum, New York, 1972), vol. 1, pp. 145–164.
24. T. T. Wu and E. A. Kabat, *J. Exp. Med.* **132**, 211 (1970); E. A. Kabat and T. T. Wu, *Ann. N.Y. Acad. Sci.* **190**, 382 (1971).
25. R. G. Fruchter, S. A. Jackson, L. E. Mole, R. R. Porter, *Biochem. J.* **116**, 249 (1970).
26. C. W. Todd, *Biochem. Biophys. Res. Commun.* **11**, 170 (1963).
27. L. E. Mole, S. A. Jackson, R. R. Porter, J. M. Wilkinson, *Biochem. J.* **124**, 301 (1971).
28. J. B. Fleischman, *Biochemistry* **10**, 2753 (1971); J. C. Jaton and D. G. Braun, *Biochem. J.* **130**, 539 (1972).
29. L. E. Mole, unpublished data.