PROFESSOR R. R. PORTER

Biochem. J. (1967) **105**, 417
Printed in Great Britain

The Structure of the Heavy Chain of Immunoglobulin and its Relevance to the Nature of the Antibody-Combining Site

THE SECOND CIBA MEDAL LECTURE

By R. R. PORTER*

Delivered at a meeting of The Biochemical Society on 20 *May* 1967 *at CIBA Laboratories Ltd., Horsham, Sussex*

I wish to summarize the recent work, carried out at St Mary's Hospital Medical School, on the chemical structure of the heavy chains of a human pathological IgG† (Daw) and of rabbit IgG prepared from pooled normal sera. I would then like to discuss what deductions can be drawn from these results about the structural basis of the specific combining power of antibodies.

In order to make this work comprehensible, however, I must give a brief summary of the general information on immunoglobulin structure that has accumulated so rapidly in the last five or six years. There are several excellent reviews (Fleischman, 1966; Cohen & Milstein, 1967*a*,*b*; Cohn & Lennox, 1967), so I will give only the very minimum amount of information necessary for the present purpose and, for brevity, will give it without acknowledgement to the many papers on which it is based.

IgG from all species examined has a basic structure of two heavy and two light chains joined together by disulphide bonds, as shown in Fig. 1. In every individual the molecules may have either of two types of light chain, κ- and λ-chains. Similarly, the heavy chains (γ-chains) exist in some and probably all species in several forms in each individual, giving rise to molecules of different subclasses. In human serum, for example, four subclasses have been identified and are now known as γ_1, γ_2, γ_3 and γ_4.

In addition, allelic variants are known of both heavy and light chains, there being only one variant, of course, in an individual homozygous for this characteristic.

With one apparent exception, any antibody specificity may be found in molecules of either K- or L-type, of any subclass or of any of the allelic variants. That is, the antibody-combining site is presumably in a section of the molecule structurally independent of those sections determining the other characteristics.

In contrast with the obvious complexity of the IgG of any individual or in any preparation of purified antibody, myeloma proteins appear to be entirely homogeneous examples of IgG. They are found in the serum of individuals from two species (man and mouse) that suffer from myelomatosis—a neoplasm of the cells synthesizing immunoglobulins, often present in the bone marrow. The myeloma proteins are believed to be the products of a single clone of cells. Each protein belongs to only one type, one subclass and one allelic variant, and, as far as is known at present, the heavy and light chains of each protein each have a single amino acid sequence. It is postulated that myeloma proteins represent a single homogeneous example of the many different kinds of IgG present in normal serum. As it is believed that all IgG is an antibody against something, it would follow that each myeloma will be a highly specific antibody. Several reports of myeloma proteins with antibody-like activity have indeed appeared recently (Zetterval, Sjöquist, Waldenström & Winblad, 1966; Metzger, 1967).

Of great experimental importance are the Bence-Jones proteins. These have been known for more

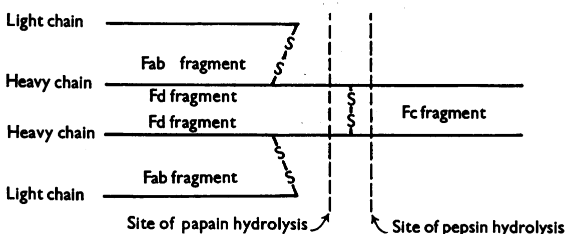

Fig. 1. Polypeptide structure of IgG. The nomenclature of the fragments is that recommended by the World Health Organisation (1964).

* Address: Department of Immunology, St Mary's Hospital Medical School, London, W. 2. Present address: Department of Biochemistry, University of Oxford.

† Abbreviations: IgG, immunoglobulin G; IgA(T), equine T- or B_2-globulin; [Glu (in sequences), blocked *N*-terminal glutamyl residue; PCA (in sequences), *N*-terminal pyrrolid-2-one-5-carboxyl residue; Glx (in sequences), glutamyl or glutaminyl residue; Asx (in sequences), aspartyl or asparaginyl residue.

than 100 years, and are found in large amounts in the urine of about half of all myeloma patients. It has now been shown that they are very similar to and probably identical with the light chains of the myeloma protein found in the same individual. Hence they offer an easily available source of homogeneous light chains from mouse or man, and they may be either K- or L-type and one or other of the known allelic variants. The Bence-Jones proteins are the material on which all the earliest amino acid sequence studies have been made.

All IgG can be split by papain to give two identical fragments (Fab fragments), which, if prepared from a known antibody, will each carry an antibody-combining site. The third fraction (Fc fragment), which crystallizes easily from the IgG of some species, is a dimer of the *C*-terminal halves of the two heavy chains. Fab fragments each contain one light chain and the *N*-terminal half of the heavy chain, known as the Fd fragment. Hence the antibody-combining site must be contained in either the light chain or the Fd fragment, or be formed jointly by both. Though the evidence is not conclusive (see Weir & Porter, 1966), we favour the view that the site is in the Fd fragment, with the light chain playing only a semi-specific role by stabilizing the steric structure of Fd fragment. As this is rather an important point, it is perhaps worth while showing some of our results (Table 1) that lead us to this view and that also illustrate the difficulty in arriving at a decisive answer. The heavy and light chains can only be separated in dissociating media that cause some loss of affinity for the antigen. In this case, 60% of the original activity is recovered if, after dissociation in acetic acid, the chains are dialysed back to neutrality without separation. If separated, the light chains have no activity and the heavy chains 20% activity. Mixing of the two restores 55% activity, but 35% activity is recovered if non-specific, rather than specific, light chains are added to specific heavy chains. Specific light and non-specific heavy chains have no activity. Thus the answer is not clear, but there seems no doubt that heavy chains alone can carry an active site but that specific light chains can also make a contribution.

Table 1. *Antigen-binding activity of mixtures of specific and non-specific peptide chains of IgA (T)*

Results are expressed as percentages of the amount of antigen bound [mol. of antigen/mol. of IgA (T), or mol. of antigen/2 mol. of each peptide chain]. The results are taken from Weir & Porter (1966).

Preparation	Mean antigen-binding activity (%)
Reduced alkylated anti-(diphtheria toxin) IgA (T)	100
Anti-(diphtheria toxin) heavy chain	21
Anti-(diphtheria toxin) light chain	2
Anti-(diphtheria toxin) heavy chain + anti-(diphtheria toxin) light chain	55
Anti-(diphtheria toxin) heavy chain + anti-lecithinase light chain	34
Anti-(diphtheria toxin) heavy chain + anti-(tetanus toxin) light chain	35
Anti-lecithinase heavy chain + anti-(diphtheria toxin) light chain	4
Anti-(tetanus toxin) heavy chain + anti-(diphtheria toxin) light chain	2

According to present ideas on protein synthesis, the amino acid sequence of a protein determines its steric structure and specific biological activity—no other information is required. Anfinsen and his colleagues have provided evidence to show that this is true for ribonuclease and several other proteins. Workers in two Laboratories have suggested that this is true also of antibodies (Haber, 1964; Whitney & Tanford, 1965). Fab fragments whose steric structures were totally disrupted by reduction in 6 M-guanidine recovered some specific binding activity when allowed to refold and re-form disulphide bonds in very dilute solution. Only 10–15% activity was recovered and it is difficult to prove that all the steric structure was destroyed in all molecules, but, with that reservation, this result suggests that in Fab fragments, as in enzymes, the steric structure is determined entirely by amino acid sequence, and hence the configuration and specificity of the antibody-combining site is determined by sequence. This basic assumption is made in all present theories of antibody formation. If it is correct, then certain amino acid sequences in the Fd fragment and also perhaps in the light chains would be expected to be characteristic of each antibody specificity.

Sequence studies on immunoglobulins began with the Bence-Jones proteins, first from humans and then from mice, chiefly those of K-type, although much information on the L-type has been published very recently (Titani, Wikler & Putnam, 1967; Milstein, Clegg & Jarvis, 1967). The outstanding finding has been the recognition of a stable section and a variable section when the amino acid sequences of different Bence-Jones proteins of the same type are compared. Fig. 2 shows a recent summary of the data from several Laboratories based on the almost complete sequence of one human κ-chain and partial sequences of 20 other κ-chains. The series of squares represents the linear sequence of 214 residues; the blank squares represent residues constant in all proteins examined and the shaded squares those that have been found to vary between proteins. The numbers inside the squares give the number of different variants found

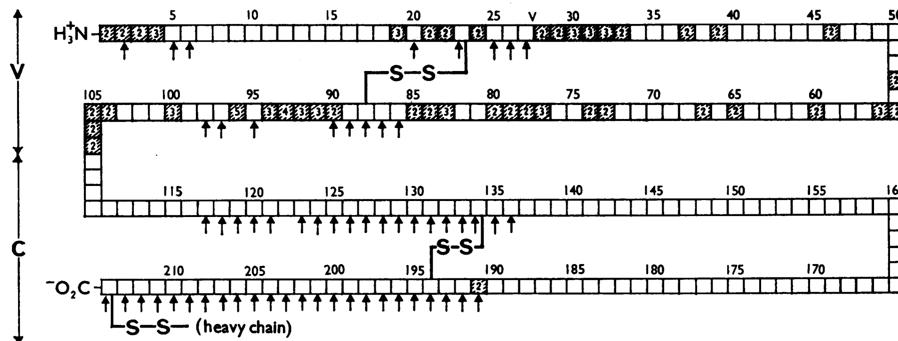

Fig. 2. Residue positions of human Bence-Jones protein, K-type, showing variable and constant residues. Blank squares indicate those positions where no changes have been found. Hatched squares indicate positions where the residues vary. The number in each square gives the number of different residues found so far in that position. Numbers above the squares give the numbering of the residues from the N-terminal end. Arrows show positions that have been found to be constant in five or more different Bence-Jones proteins. V, Variable region; C, constant region. The results are from Cohen & Milstein (1967*a*).

in those positions and the arrows show the residues that are constant in five or more proteins. Forty-three variable positions have been found in residues 1–107 and one in residues 108–214. The latter, at position 191, correlates with an allelic change and hence is constant in all Bence-Jones proteins of the same allelic character. Presumably the number of variable positions in residues 1–107 will rise further as more complete data become available. Not surprisingly, this finding has aroused intense interest and much speculation as to its biosynthetic origin, and many theories have been offered to explain it. Further, it has been assumed that this great variability (no two identical proteins have been found so far) must occur also in the N-terminal half of the light chains from normal IgG, and that a similar phenomenon of a variable N-terminal section occurs also in heavy chains. Two such regions could give rise to a very wide range of changes of configuration in this part of the molecule and would offer a satisfactory explanation of antibody specificity in terms of amino acid sequence.

This is all most plausible, but it will be difficult to prove unless several myeloma proteins with antibody activity against well characterized antigenic sites are found. As, in our view, the Fd fragment may be of greater importance than the light chain, it seemed essential to examine the Fd section of the heavy chain from a myeloma protein to decide if indeed there is a constant and a variable region and, if so, to fix the positions at which the transition occurs. It also seemed essential to examine the Fd section of the heavy chain of IgG from normal and immunized animals, as, until myeloma proteins with known antibody activity are isolated, only this will make possible a direct assessment of the relation of antibody specificity to amino acid sequence in several antibodies. We have tried to work with both problems together, hoping that work on the myeloma proteins, while of great interest in itself, would also help to guide the more difficult problem of the normal heavy chains, where complex amino acid sequences were expected.

Structures of the heavy chains from a pathological human IgG (Daw) and from normal rabbit IgG

N-*Terminal sequence.* The first problem was the N-terminal end of the heavy chains, as in both proteins the α-amino group is unreactive. The same observation has been made in a number of proteins, particularly in plant viral proteins, and in each case it has been found that the α-amino group was acetylated. Attempts to find an acetylated α-amino group in the heavy chain of IgG (Daw) were unsuccessful, so the N-terminal peptide was isolated (Porter & Press, 1965; Press, Piggot & Porter, 1966) by a technique similar to that used originally by Narita (1958) in his investigation of the blocked N-terminal sequence of tobacco-mosaic virus. The heavy chain was digested to small peptides by Pronase, a proteolytic enzyme of very wide specificity, and the complex mixture was run quickly through a small column of the acidic resin Dowex 50 at neutral pH. All peptides with a free α-amino group were adsorbed, so that there was a rapid concentration in the eluate of the N-terminal peptide with the blocked α-amino group. After further purification, the tripeptide [Glu-Val-Thr was isolated in a yield of 0·75 mole of peptide/mole of heavy chain, and its sequence was determined.

The presence of the blocked glutamic acid residue raised doubts as to its significance, as peptides with N-terminal glutamine will easily form a ring in mild acid conditions to give a pyrrolid-2-one-5-carboxyl peptide with no free α-amino group, and hence the peptide isolated might well have been derived from an internal sequence and be an artifact of the isolation procedure. This possibility was eliminated by allowing all the α-amino groups to react with 1-fluoro-2,4-dinitrobenzene immediately the enzyme digestion at pH 8·1 was complete. Glutamine is stable under these conditions, but the same peptide was isolated again in similar yield, and hence it was concluded that this was the true N-terminal sequence. Investigation of the properties such as lability to alkali, electrophoretic mobility and amide content of the isolated dipeptide, and comparison with synthetic PCA-Val proved that the N-terminal sequence was indeed PCA-Val-Thr.

Earlier studies of the N-terminal sequence of the heavy chain from normal rabbit IgG had led to inconclusive results, but, when they were repeated with the knowledge that pyrrolid-2-one-5-carboxylic acid might be the N-terminal residue, it was found that pyrrolid-2-one-5-carboxylic acid was, in fact, N-terminal in all molecules (Wilkinson, Press & Porter, 1966). There were, however, different residues in the adjacent positions, so that several peptides were obtained in low yield; this was the cause of the earlier difficulties. The N-terminal sequences of the heavy chain of rabbit IgG are PCA-Ser-Val-Glu, PCA-Ser-Leu-Glu and PCA-Gln, present in approximately 50%, 20% and 20% of the molecules respectively.

Digestion of heavy chains with cyanogen bromide. Though the N-terminal sequences can be and have been extended further, any attempt to determine a full sequence of the heavy chain, containing about 450 residues, depended on finding a technique to break the chain into relatively large pieces, which could be aligned easily but which were sufficiently small for detailed sequence studies. As mixed sequences were to be expected, some kind of reference point would be of great value as it might help to distinguish between sequences from distinct sections and variant sequences of the same section. Splitting at the methionine residues with cyanogen bromide (Gross & Witkop, 1961) seemed to be the best approach, as there are only a small number of methionine residues and the resulting peptides are characterized by a C-terminal homoserine residue, except the peptide from the C-terminal end of the chain.

With the heavy chain of IgG (Daw), this method proved successful (Piggot & Press, 1967) and gave the alignment shown in Fig. 3. The argument in support of this arrangement is, very briefly, as follows. Fragment 2*b* contains the N-terminal sequence, and fragment 5 is the only one that does not contain homoserine. Fragment 5 does have the same C-terminal residue, glycine, as does the whole heavy chain, and hence fragments 2*b* and 5 are the N-terminal and C-terminal fragments respectively. If the Fd and Fc fragments, rather than whole heavy chain, are treated with cyanogen bromide, then Fd fragment gives fragments 2*b*, 4 and 2*a*′; the last-named has a similar size and composition to fragment 2*a*, but contains no homoserine. Fc fragment gives fragments 5 and 3 and a small peptide 2*a*″ with C-terminal homoserine, which together with fragment 2*a*′ agrees well with the composition of fragment 2*a*. Hence no other arrangement is possible, and confirmation was obtained by the isolation in low yield from a cyanogen bromide digest of whole heavy chain of a large fragment, 1*a*, the composition of which agreed well with that of the sum of fragments 4 and 2*a*. It might have arisen from incomplete splitting at the methionine residues between these two sections, but analysis of fragment 1*a* suggested that this methionine residue might be missing from about 20% of the molecules.

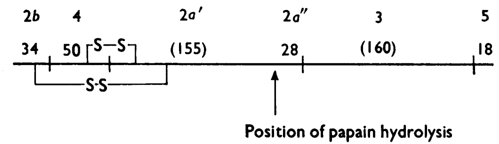

Fig. 3. Position of methionine residues in the heavy chain of a human pathological IgG (Daw). The upper numbers identify the fractions, and the lower numbers are the total number of residues between each pair of methionine residues indicated by the vertical bars. The results are from Piggot & Press (1967).

Investigation of the products of cyanogen bromide cleavage of Fab fragment and subsequent reduction of the disulphide bonds has led to the positioning of some of these, as shown in Fig. 3 (Piggot & Press, 1967). This then provides a satisfactory basis for detailed sequence studies of the Fd section of IgG (Daw) heavy chain.

The rabbit IgG heavy chain is a more complex problem, but progress has been made by a similar approach. There are about 5–6 methionine residues in the heavy chain, and the alignment of the cyanogen bromide fragments is shown in Fig. 4. The arrangement of the fragments in the Fc section is due in part to the work of Hill and colleagues (Hill, Delaney, Lebovitz & Fellows, 1966*b*) at Duke University, who have reported a partial sequence for some 240 residues in the Fc section. Some of the disulphide bonds have been placed by Prahl

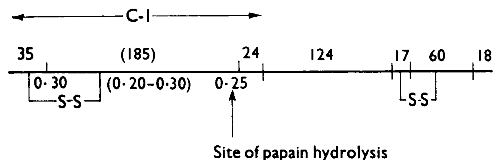

Fig. 4. Position of methionine residues in the heavy chain of normal rabbit IgG from pooled serum. Full methionine residues are indicated by a full bar and fractional methionine residues by a half bar. Numbers above the line give the number of residues between methionine residues. Numbers under the line give the content of fractional residues in the position indicated.

(1967*a*) and also by Givol & Porter (1966), as shown in Fig. 4.

It can be seen that although the methionine residues in the Fc section appear to be present in all molecules, as in the heavy chain of IgG (Daw), this is not true of the Fd section and no clean splitting is achieved in this part of the chain. There is a methionine at position 35, similar to that in IgG (Daw) heavy chain, but it is present in only a third of the molecules. Again, there is a methionine residue 24 residues from the *C*-terminal end of fragment C-1. In this case J. W. Prahl (unpublished work) has shown that its presence correlates with allotype, as, in a preparation of IgG from rabbits homozygous at the *Aa*3 locus, methionine is present in 80% or more of the molecules, and this methionine residue is replaced by threonine in the heavy chain of IgG determined by the *Aa*1 and *Aa*2 loci. In the IgG from pooled rabbit sera this methionine residue appeared to be present in 20–25% of the molecules. There is at least one other fractional methionine residue nearer the centre of fragment C-1 than either of the two that have been identified.

Sequence studies. Though sequence work is arduous to carry out, its results can be summarized quickly, and Fig. 5 gives the data obtained so far on the heavy chain of IgG (Daw). The sequence of the cyanogen bromide fragments 2*b* (Piggot & Press, 1967) and 4 (Press, 1967) is complete and the extension into fragment 2*a*′ (N. M. Hogg, unpublished work) has begun. Sequence data on fragment 2*a*″ are also almost completed (L. A. Steiner, unpublished work) and should lead to the establishment of the positions of the interchain disulphide bonds. One of the main objectives of the work with the heavy chain of myeloma proteins has been to find if the phenomenon of constant and variable sections, which is so striking a feature in the Bence-Jones proteins, also occurs here and, if so, to place the point at which the change occurs. A comparative study has been commenced therefore of the structure of a second myeloma protein

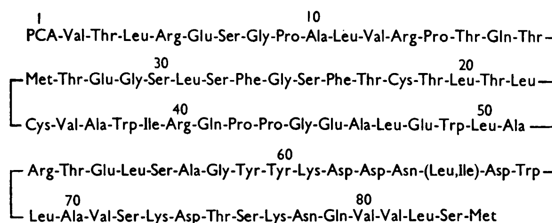

Fig. 5. *N*-Terminal sequence of the heavy chain of a human pathological IgG (Daw). The results are from Press (1967).

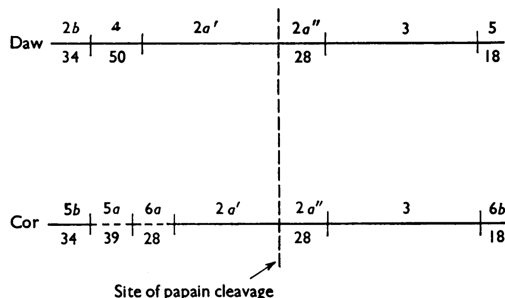

Fig. 6. Comparison of the alignments of the fragments obtained by CNBr splitting at the methionine residues in the heavy chains of two pathological human IgG that are of the same subclass, γ_1, and allelic character, *Gm*(*a*+)(*f*−). The upper numbers identify the fragments and the lower numbers give the number of residues between methionine residues. The results are from Press & Piggot (1967).

(Cor), which is of the same subclass and carries the same allelic characters as IgG (Daw). The *N*-terminal sequence is PCA-Val-Thr, as in IgG (Daw), but no other sequence data are available yet. The equivalent sections have been obtained from cyanogen bromide cleavage and aligned as shown in Fig. 6, and composition alone shows that there must be at least two differences in the section equivalent to fragment 2*b*, more in fragment 4, and some in fragment 2*a*′ (Press & Piggot, 1967). Though the data are very incomplete, it seems likely that variations will be found scattered over at least 100 positions, perhaps in a similar manner to that found in Bence-Jones proteins. Variations are found also in the *C*-terminal 18 residues of human myeloma chains (Prahl, 1967*b*) (Fig. 7), but here the differences correlate with subclass, and within one subclass the sequence appears to be constant in the small number of individual proteins that have been examined.

The sequences of K- and L-type Bence-Jones proteins of both men and mice have been compared and obvious homologies found and, indeed, homologies have also been found between these sequences

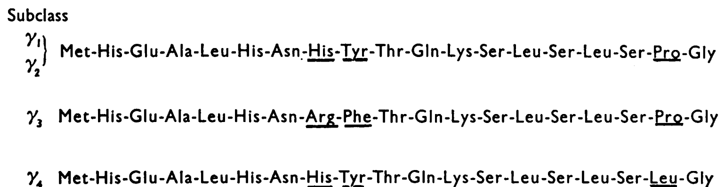

Fig. 7. *C*-Terminal sequence of the heavy chains from myeloma proteins of different subclasses. A bar shows where a replacement occurs. The results are from Prahl (1967*b*).

and that of the Fc portion of rabbit IgG heavy chain (Hill, Delaney, Fellows & Lebovitz, 1966*a*). Many deductions have been drawn from these data and they have been used to support numerous theories that explain both the evolutionary origin of the proteins and the genetic basis of the high variability of parts of the chain. These theories have been discussed critically (Cohn & Lennox, 1967), though no obvious answer has been arrived at yet. Clearly, this work is of great interest, and the sequence available from both the Fd section from IgG (Daw) and the rabbit Fd section (discussed below) can also be used for this purpose, but I wish to concentrate primarily on the relevance or otherwise to the structural basis of antibody specificity, so I will describe the sequence studies on the rabbit Fd section. This is the only section of immunoglobulins investigated so far where direct correlation of structure and specificity may perhaps be expected.

Rabbit Fd section. This work is the most difficult because of the variability of sequence. This has been shown at the *N*-terminal end and that it is widespread is suggested by the three or more fractional methionine residues discussed above. The variability should not, however, continue throughout the whole Fd section, as all molecules have certain constant features, for example the heavy–light interchain disulphide bond and also the three allelic variations that are known to be present in the Fd section (Kelus, 1963; Stemke, 1964), and hence will be constant in such a preparation from an animal homozygous in this respect. If the extent of the variable section of the Fd section is similar to that of the variable section of Bence-Jones protein, i.e. about 100 residues, then it might not be possible to establish a coherent sequence over this part. The first objective then was to try to obtain the sequence of the stable sections and to try to define their position in the molecule and, of course, to extend the information as far as possible.

First, the *N*-terminal sequence was extended by tryptic digestion to give a terminal arginine-containing nonapeptide present in about 75% of all molecules. Then investigation of the sequence of

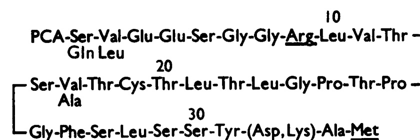

Fig. 8. *N*-Terminal sequence of the heavy chains of IgG from pooled rabbit serum. Double residues show that replacements have been found in these positions. A bar shows that a replacement occurs but it has not been identified. The results are from Wilkinson (1967).

the terminal section released by the cyanogen bromide splitting of the methionine present in about 30% of all the molecules has given the results shown in Fig. 8 (Wilkinson, 1967). These molecules will certainly be similar in sequence to, but may not be identical with, the other two-thirds of the molecules that do not have methionine residues in this position. Homology with the equivalent section of IgG (Daw) heavy chain is apparent (Wilkinson, 1967). It can be seen that there is variation at positions 2, 3, 23 and 35 and recent evidence suggests that it will also be found at position 9 (J. W. Prahl & J. M. Wilkinson, unpublished work), i.e. five positions out of 35, and with the experimental technique available it is unlikely that a variant present in less than 20% of the molecules would be detected. The variation at position 9 is of particular interest, as it appears to correlate with allotype. The arginine is missing in this section when isolated from heavy chain of IgG from a rabbit that is homozygous at the *Aa*3 locus. In agreement with the finding of the methionine–threonine replacement that also correlates with the same allotype (referred to above), the quantitative results suggest a 20–25% content of protein determined by the *Aa*3 locus in the pooled IgG. Thus the nature of the two residues at position 9 and at about position 220 is correlated in each case with the same allele, *Aa*3. Hence it is likely that both are coded for by the same gene. As position 220 is at the *N*-terminus of the Fc section (though sometimes split out during papain digestion), and position 9 is near the *N*-terminus of the Fd section,

it is likely that both the Fd section and the Fc section are coded for by the same gene, and that therefore one gene codes for the whole of the heavy chain. Contrary suggestions have been put forward to explain several phenomena that seem to be more easily explicable if the heavy chain is in fact two chains (see Cohen & Porter, 1964). For example, the presence of a variable and constant section, if true for myeloma heavy chain as in the Bence-Jones proteins, could be explained by a two-gene hypothesis (see Cohen & Milstein, 1967*b*).

For the main part of the molecule, splitting into further large sections, which might behave distinctively (even though not entirely homogeneous) and hence be capable of isolation and alignment, was necessary. The presence of only fractional methionine residues in fragment C-1 prevented the further use of cyanogen bromide, and so blocking of lysine residues with *S*-ethyl trifluorothioacetate and tryptic hydrolysis (Goldberger & Anfinsen, 1962) at the six arginine residues were used. This method has led to the isolation of six peptides whose composition is shown in Table 2 (Cebra, 1967). There were two large peptides, T_2 and T_3, two medium-sized peptides, T_1 and T_4, and two small peptides, T_5 and T_7, which by composition appear to account for 90% or more of fragment C-1. The yields of isolated peptides were 20–30%, which seems poor but is similar to the yields obtained when peptides of that size are isolated from proteins of unique sequence. Further, the content of certain peptides containing uncommon amino acids such as histidine or homoserine can be estimated in the crude fractions at an early stage, and on this basis the yield rises to 60–70%, suggesting that they are indeed derived from different sections of the chain and are not from varying overlapping sequences. Peptide T_2 is clearly *C*-terminal as it contains homoserine, and its sequence agrees well in the overlap with the partial sequence of the *N*-terminal end of fragment Fc given by Hill *et al.* (1966*b*). Peptide T_5 is the *N*-terminal end of fragment C-1 previously isolated, and the distinctive compositions of peptides T_1, T_3 and T_4, when compared with the known sequence of IgG (Daw) heavy chain, suggest the order of peptides T_5, T_1, T_4, T_3, T_2, with the hexapeptide T_7 not yet placed. As no overlaps have been established, this is only guesswork, but the conclusion that this series of peptides will account for most of the sequence of fragment C-1 seems probable. This implies that the position of five of the six arginine residues is relatively constant in most molecules. It is not possible yet to decide whether the failure to account for the sixth arginine residue, which should be present in quite a small peptide, is technical or whether it is in fact present as several fractional residues in different positions and hence hard to identify. The sequence of the *N*-terminal peptide T_5 was known (Wilkinson *et al.* 1966); that of peptide T_7 is Thr-Phe-Pro-Ser-Val-Arg, and it had been isolated previously from a tryptic digest of

Table 2. *Analysis of tryptic peptides of fragment C*-1

	Amino acid composition (moles/mole)							
	Peptide T_1	Peptide T_2	Peptide T_3	Peptide T_4	Peptide T_5	Peptide T_7	Sum of peptides	Fragment C-1 (average value)
Lys	2·0	5·2	3·0	0	0	0	10	12
His	0	1·0	0	0	0	0	1·0	1·6
Arg	1·0	0	1·0	1·0	1·0	1·0	5·0	6·3
Asp	1·7	4·0	6·2	5·5	0	0	17	17
Thr	5·3	7·8	9·3	3·3	0	1·0	26	30
Ser	4·2	9·8	7·2	1·2	1·8	1·0	25	28
Glu	2·0	2·9	4·4	6·4	3·0	0	18	15
Pro	2·5	12	7·7	2·3	0	0·9	24	22
Gly	4·0	4·2	9·0	0	2·2	0	19	21
Ala	3·0	3·1	3·9	0	0	0	10	14
Val	2·0	7·9	8·2	6·8	1·1	0·7	26	23
Ile	2·0	1·1	1·0	1·1	0	0	5·2	7·2
Leu	2·6	5·0	7·3	0	0	0	15	17
Tyr	2·1	0·9	2·8	1·1	0	0	6·9	9·4
Phe	1·4	2·0	2·1	1·1	0	0·8	7·4	7·4
$CyS \cdot CH_2 \cdot CO_2H$	0·9	2·3	3·1	1·0	0	0	7·3	7·4
Homoserine	0	1·0	0	0	0	0	1·0	1·8
Trp	0	0	0	2·0	0	0	2·0	4·0
Total residues	37	70	76	29	9·1	5·4	227	245

heavy chain by Hill *et al.* (1966*b*). The partial sequences of peptides T_2 and T_3 are given in Figs. 9 and 10 (Cebra, 1967).

The only comment that I wish to make on these sequences in this context is that, though tryptic peptides accounting for about ten *N*-terminal residues are missing from both peptide T_2 and peptide T_3 at present, the sequences of about 60 residues in each show clearly that they are from distinct sections of Fd fragment.

Though this work is obviously very incomplete, it is clear that it has gone further than expected and it now seems likely that a basic sequence can be established for 90%, and perhaps for all, of the Fd section of heavy chain of normal rabbit IgG. The work has been repeated on a purified anti-DNP-antibody and the same peptides have been obtained with similar compositions and in similar yields. No sequence work has been attempted on this material yet. Of the variations that have been found in the sequence of pooled IgG, one and probably another are related to allotype, and at least four others are not. There is no evidence that the latter four variants are related to antibody specificity. They might be related to the presence of different subclasses in the pooled IgG, though no subclasses have been identified as yet in the rabbit, and they may be related to allelic variants other than the three main allotypes that have been identified. Many other fractional residues probably exist—any present in less than 20% of the molecules are unlikely to be detected by present techniques. The methionine–threonine replacement, although present in 20–25% of the pooled IgG, was missed until IgG from a rabbit homozygous at the *Aa*3 locus was examined. Again, the variant sequences in the *C*-terminal 18 residues of human IgG heavy chain of different subclasses (Fig. 7) were not detected in pooled IgG where 85% of all the molecules (γ_1 and γ_2) will have a common sequence.

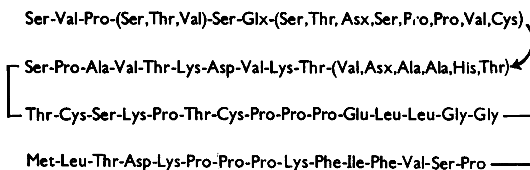

Fig. 9. Partial amino acid sequence of peptide T_2, a tryptic peptide of the CNBr fraction C-1 from normal rabbit heavy chain. The results are from Cebra (1967).

If we anticipate the data, which will take some time to obtain, and assume that it will confirm the pattern of the very incomplete evidence given here, then we would expect to find: (1) that the heavy chains of myeloma proteins will show a variable section of at least 100 residues at the *N*-terminal end, and a constant section about three times as long; (2) that a coherent sequence will be obtainable from the heavy chain of pooled IgG throughout most and perhaps all its length and that a specific antibody will have a very similar sequence; (3) that all the variants to be detected in pooled IgG heavy chain will be unrelated to antibody specificity but will be due to mixtures of subclasses or to the three known allelic variants or to other, as yet unrecognized, allelic variants. Can any conclusions be drawn as to the structural basis of antibody specificity? I am afraid that the answer is not many, even when all these assumptions are made. First, if there is indeed a variable section of both the myeloma-protein heavy and light chains comparable with that in the Bence-Jones proteins, it seems obvious that this must be the structural basis of antibody specificity and indeed perverse to look for any other explanation. Such a phenomenon is unknown among other proteins and it seems most probable that it is related to the special function of immunoglobulins, i.e. a specific affinity for a very wide range of structural features on the molecules that initiate or stimulate their synthesis. Until a number of myeloma proteins with indisputable antibody function have been found, however, final proof of this hypothesis is impossible. Further, there is yet another form of complexity found in immunoglobulins—idiotypic specificity (see Gell & Kelus, 1967), in which an antibody in one individual may have a unique antigenic specificity not common to another antibody of the same specificity in other individuals of the same species. Idiotypic specificity is always located in the Fab section of the molecule. Hence there appear to be structural variations other than those due to subclass or allelic differences that cannot be directly

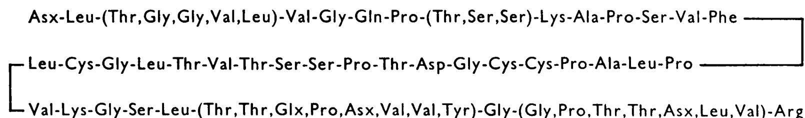

Fig. 10. Partial amino acid sequence of peptide T_3, a tryptic peptide of the CNBr fraction C-1 from normal rabbit heavy chain. The results are from Cebra (1967).

related to the configuration of the antibody-combining site. Myeloma proteins also show a similar individual antigenic specificity and the possibility must exist that in both cases it is dependent on structural features of the variable sections of both chains, or either chain, that do not control the structure of the combining site; i.e. part at least of the changes in the variable section can occur independently of antibody specificity.

All evidence (e.g. from the work on animals born and raised in a germ-free environment) suggests that the IgG of a normal animal is a mixture of very many antibodies against substances encountered in its environment. Unless the animal is recovering from a specific infection it is most unlikely that the content of any one antibody in the serum will exceed more than 1% of the total IgG, and it will probably be very much less. If, in fact, each antibody has a specific sequence in a part of the Fd section, then it may seem surprising that it was possible to arrive at a coherent sequence from such a complex mixture as pooled normal IgG and that if, in fact, such a sequence could be established throughout (with allowance for the variability related to the other factors) this might be taken as evidence against the belief that amino acid sequence is the basis of antibody specificity. It is clear that the results presented here do not allow any definite conclusion on this point. In the first place there could well be a highly variable sequence of, say, 20 residues that control antibody specificity. Such a section would be present in so many different forms that it would be undetectable in pooled IgG, and, until overlapping sequences have given a complete alignment of the peptides T_1–T_7, the existence of such a sequence cannot be excluded. If the variable section was, say, 50 residues, most of which varied from one antibody specificity to another, then the whole of such a section would remain unaccounted for. This is unlikely, as the results suggest that the fragments reported here make up about 90% of the Fd section. However, as F. H. C. Crick pointed out to me, it may be that there is a variable section of, say, 50 residues, perhaps scattered through the first 100 residues of the Fd section, but for each antibody specificity only 10 need be changed. That is, 80% of all the positions of the variable section and 100% of positions in the stable section will be identical in a mixture of many antibodies. Such a system could easily give enough variability to account for all known antibody specificities and yet would remain undetectable by present techniques of protein chemistry. Indeed, recoveries better than 90% of each peptide and proof that each position in each peptide was identical in more than 90% of the molecules would be necessary to disprove such an arrangement. Such technical perfection has not been achieved in the sequence studies of any protein. It is doubtful whether such a system exists, however, as, if the Bence-Jones proteins are taken as model, the only example of two nearly complete sequences (Gray, Dreyer & Hood, 1967) shows differences in 44 positions, together with an insertion of four other residues found in only one of the proteins. This suggests that changes of much more than 20% of the variable positions can occur. However, even if only 20% of the variable positions change in any one antibody specificity, comparison of the complete sequence of one or two purified antibodies with the apparent sequence of pooled IgG should show up the postulated ten variant positions. We can therefore claim that, if the work discussed here on the sequence of the heavy chains of pooled IgG and on anti-DNP-antibody can be taken to completion, it appears to offer a feasible experimental approach to obtaining an answer to the question: does amino acid sequence alone control antibody specificity and, if so, how is it achieved?

I thank members of the Department of Immunology, St Mary's Hospital Medical School, for permission to quote their unpublished results, and acknowledge the generous gifts of rabbit serum of known allotype from Dr A. S. Kelus and of characterized myeloma sera from Dr H. G. Kunkel and Dr E. F. Osserman and Daw serum from Dr V. Wynn. I thank the Medical Research Council for their essential financial support.

REFERENCES

Cebra, J. J. (1967). *Cold Spr. Harb. Symp. quant. Biol.* **32** (in the Press).
Cohen, S. & Porter, R. R. (1964). *Advanc. Immunol.* **4**, 287.
Cohen, S. & Milstein, C. (1967*a*). *Nature, Lond.*, **214**, 449.
Cohen, S. & Milstein, C. (1967*b*). *Advanc. Immunol.* **7** (in the Press).
Cohn, M. & Lennox, E. S. (1967). *Annu. Rev. Biochem.* **36** (in the Press).
Fleischman, J. B. (1966). *Annu. Rev. Biochem.* **35**, 835.
Gell, P. G. H. & Kelus, A. S. (1967). *Advanc. Immunol.* **7** (in the Press).
Givol, D. & Porter, R. R. (1966). *Israel J. Chem* **4**, 68P.
Goldberger, R. F. & Anfinsen, C. B. (1962). *Biochemistry*, **1**, 401.
Gray, W., Dreyer, W. & Hood, L. (1967). *Science*, **155**, 465.
Gross, E. & Witkop, B. (1961). *J. Amer. chem. Soc.* **83**, 1510.
Haber, E. (1964). *Proc. nat. Acad. Sci., Wash.*, **52**, 1099.
Hill, R. L., Delaney, R., Fellows, R. E. & Lebovitz, H. E. (1966*a*). *Proc. nat. Acad. Sci., Wash.*, **56**, 1762.
Hill, R. L., Delaney, R., Lebovitz, H. E. & Fellows, R. E. (1966*b*). *Proc. Roy. Soc.* B, **166**, 159.
Kelus, A. S. (1963). *Biochem. J.* **88**, 4P.
Metzger, H. (1967). *Fed. Proc.* **26**, 529.
Milstein, C., Clegg, J. B. & Jarvis, J. M. (1967). *Nature, Lond.*, **214**, 270.
Narita, K. (1958). *Biochim. biophys. Acta*, **28**, 184.
Piggot, P. J. & Press, E. M. (1967). *Biochem. J.* **104**, 616.
Porter, R. R. & Press, E. M. (1965). *Biochem. J.* **97**, 32P.
Prahl, J. W. (1967*a*). *Biochem. J.* **104**, 647.
Prahl, J. W. (1967*b*). *Biochem. J.* (in the Press).

Press, E. M. (1967). *Biochem. J.* **104**, 30c.
Press, E. M. & Piggot, P. J. (1967). *Cold Spr. Harb. Symp. quant. Biol.* **32** (in the Press).
Press, E. M., Piggot, P. J. & Porter, R. R. (1966). *Biochem. J.* **99**, 356.
Stemke, G. W. (1964). *Science*, **145**, 403.
Titani, K., Wikler, M. & Putnam, F. W. (1967). *Science*, **155**, 828.
Weir, R. C. & Porter, R. R. (1966). *Biochem. J.* **100**, 69.
Whitney, P. L. & Tanford, C. (1965). *Proc. nat. Acad. Sci., Wash.*, **53**, 524.
Wilkinson, J. M. (1967). Ph.D. Thesis: University of London.
Wilkinson, J. M., Press, E. M. & Porter, R. R. (1966). *Biochem. J.* **100**, 303.
World Health Organisation (1964). *Bull. World Hlth Org.* **30**, 447.
Zetterval, O., Sjöquist, J., Waldenström, J. & Winblad, S. (1966). *Clin. exp. Immunol.* **1**, 213.