The Covalent Structure of a Human γG-Immunoglobulin. XI. Functional Implications*

Gerald M. Edelman

ABSTRACT: Examination of the complete amino acid sequence of the γG1-immunoglobulin Eu and comparison with other immunoglobulins has revealed a number of relationships between the variable (V_H and V_L) and constant (C_H and C_L) regions of heavy and light polypeptide chains: (1) V_H and V_L are homologous to each other but are not obviously homologous to C_H or C_L. (2) C_H consists of three linearly arranged adjacent homology regions, C_H1, C_H2, and C_H3, each of which is similar in size and is closely homologous to the others and to C_L. (3) Each variable region and each constant homology region contains one intrachain disulfide bond; the order of the regions leads to a linear and periodic distribution of these bonds in the molecule. (4) The region in the middle of the heavy chain containing all the interchain disulfide bonds has no homologous counterpart in other portions of the molecule. The data prompt the hypothesis that in the whole molecule, homology regions have similar three-dimensional structures, each consisting of a compact domain which is stabilized by a single disulfide bond and which contributes to at least one active site. The existence of V and C homology regions supports the hypothesis that immunoglobulins evolved by duplication of two genes: V genes, which specified antigen binding functions and C genes, which specified effector functions. The present other studies and suggest that V genes may be translocated to C genes to form a single VC gene in the lymphoid cell.

Classical studies have indicated that antibodies have two kinds of functions in the immune response: (1) they interact with a wide range of chemically different antigenic determinants in a more or less specific fashion (antigen binding functions), and (2) they mediate physiological reactions of immunity such as complement and cell fixation (effector functions). Specific binding of antigens ultimately depends upon amino acid sequences in the variable (V) regions of light and heavy chains of antibody molecules; the diversity of these sequences in different molecules leads to the range of specificities required for a selective immune response (for a review, see Edelman and Gall, 1969). In contrast, effector functions are carried out by regions of relatively constant sequence (C regions) which, like enzymes, appear to have evolved for a relatively restricted set of interactions.

It is the purpose of this paper to discuss some functional and evolutionary implications of the studies on the covalent structure of the γG1-immunoglobulin Eu which have been reported in the previous papers of this series (Edelman *et al.*, 1968; Waxdal *et al.*, 1968a,b; Gall *et al.*, 1968; Cunningham *et al.*, 1968, 1970; Gottlieb *et al.*, 1970; Rutishauser *et al.*, 1970; Bennett *et al.*, 1970; Gall and Edelman, 1970). Amino acid sequence analyses have shown that the V and C regions of light chains are adjacent and do not overlap (Hilschmann and Craig, 1965; Titani *et al.*, 1967; Cunningham *et al.*, 1968). One of the most striking features that emerges from the completed sequence of a whole immunoglobulin molecule (Edelman *et al.*, 1969) is the arrangement of *both* its heavy and light polypeptide chains into linearly connected V and C regions within which there are strong homologies. Variation in the sequences of V_H and V_L regions for antigen binding functions and at the same time conservation of the sequences of C_H and C_L regions for effector functions appear to require special genetic and evolutionary mechanisms.

Materials and Methods

Homologies. Protein sequences were examined for homology by direct inspection for identical and chemically similar residues and also by the procedure of Fitch (1966). In this procedure, polynucleotide triplets corresponding to the amino acids in the sequences are compared by means of a computer program. A length of sequence (LSE) is chosen and compared with successive stretches of another region until all possible comparisons have been made. The minimal number of mutations required (MR) to convert one sequence into another is tabulated. The frequency of each MR value is plotted on cumulative probability paper; deviations from a straight line suggest that a resemblance is not fortuitous.

Distribution of Residues. The distribution of hydrophobic and hydrophilic amino acids in both light and heavy chains was determined by inspection as well as by comparisons using a computer. Similar comparisons were made for each individual amino acid.

Results and Discussion

The amino acid sequence of each polypeptide chain was examined for similar characteristics, for unusual distributions of amino acids, and for homologies with other immunoglobulin chains. The characteristic features of each chain of protein Eu are summarized in Table I.

Distribution of Amino Acids in the Sequence. In Figure 1 is shown the complete amino acid sequence of the light chain, which consists of 214 residues. So far, this is the only

* From The Rockefeller University, New York, New York 10021. *Received December 8, 1969.* This work was supported in part by grants from the National Science Foundation (GB 8371) and the U. S. Public Health Service (AM 04256 and AI 09273).

TABLE I: Characteristics of the Polypeptide Chains of the γG-Immunoglobulin Eu.

	Light Chain		Heavy Chain	
	V_L Region (1–108)	C_L Region (109–214)	V_H Region (1–114)	C_H Region (115–446)
Subgroup	κI		γI	
Subtype or subclass		K		γG1
Genetic grouping		Inv 3+ (Val_{191})		Gm 1−(Glu_{356}, Met_{358}) Gm 4+ (?Arg_{214})

complete sequence of a κ chain from a serum immunoglobulin, although partial sequences have been determined in several other cases (Niall and Edman, 1967). This sequence is fundamentally similar to those previously determined for type-K Bence–Jones proteins of subgroup I (Milstein, 1967). Fourteen new variable positions were observed in Eu; several of these positions have since been found in Bence–Jones proteins (Milstein, 1969). Unlike any other κ chain so far examined, the Eu light chain has a glycyl residue at position 108. In the absence of further data, it is not possible to tell whether this is a variable position or whether it represents a rare allotypic difference. Thus the variable region for this subgroup is either 107 or 108 residues long. The valyl residue at position 191 is correlated with the fact that protein Eu is Inv 3+ (Baglioni *et al.*, 1966; Edelman *et al.*, 1968).

The amino acid sequence of the entire heavy chain (446 residues) is presented in Figure 2. The NH_2 terminus is pyrrolidonecarboxylic acid as in the human immunoglobulins Daw (Press *et al.*, 1966) and He (Cunningham *et al.*, 1969). Cleavage by trypsin to form Fd(t) and Fc(t) fragments occurs at Lys-222 (Gall *et al.*, 1968). The polysaccharide is attached at Asx-297. This residue begins the sequence Asx-Ser-Thr; the sequence Asx-X-Thr has been found at the point of attachment of carbohydrate in several other glycoproteins (Eylar, 1966; Catley *et al.*, 1969). It is likely that similar transferases (Gottschalk, 1969) are used to attach the first sugar residue to the completed polypeptide chain in these proteins as well as in γG-immunoglobulins. Glu-356 and Met-358 may be associated with the Gm 1 specificity (Rutishauser *et al.*, 1968; Thorpe and Deutsch, 1966), and Arg-214 may be associated with the Gm 4 specificity (Edelman *et al.*, 1969) (Table I).

Comparison of the sequence of the Eu γ chain with that of another protein (He) having the same genetic type has shown that the variable regions belong to at least two subgroups (Cunningham *et al.*, 1969). Eu has been assigned tentatively to subgroup I and He to subgroup II. The two proteins differ in 83 of the first 114 positions, but the sequence of He becomes identical with that of Eu after position 114. Comparison of He with other subgroup II proteins indicates

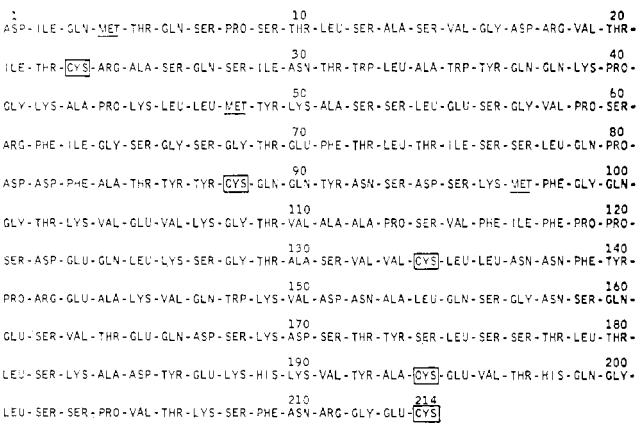

FIGURE 1: Complete amino acid sequence of the light chain. Methionyl residues are underlined and half-cystinyl residues are in boxes.

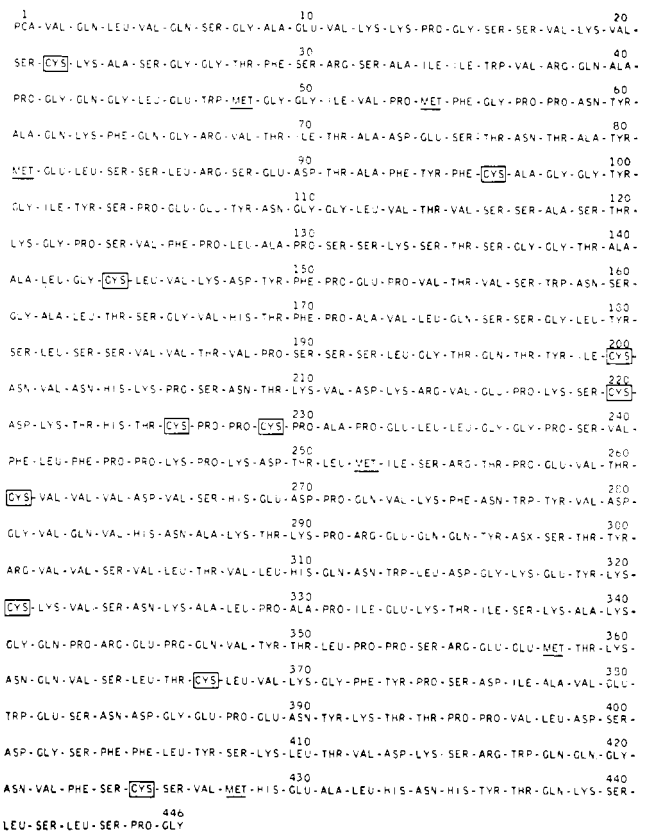

FIGURE 2: Complete amino acid sequence of the heavy chain. Methionyl residues are underlined and half-cystinyl residues are in boxes.

TABLE II: Notable Distributions of Amino Acids in Heavy and Light Chains of the γG-Immunoglobulin Eu.

Hydrophobic and hydrophilic residues	Relatively hydrophobic from residues 149 to 203 in region C_H1; only one Glu and one His and no Lys, Arg, or Asp in this stretch. Relatively hydrophilic from residue 210 to residue 225 of the hinge region of the heavy chain.
Residues with ionizable side chains	Relatively high density of charged residues in regions C_H2 and C_H3.
Histidine	None in V regions.
Proline	High density in stretch from residues 227 to 247.
Glycine	Higher density in regions V_L, V_H, and C_H1; lower density in regions C_H2 and C_H3.
Alanine	Lower density in regions C_H2 and C_H3.
Half-cystine	Periodic distribution in homology regions. All half-cystines contributing to interchain bonds within ten residues of each other in middle of heavy chain.
Isoleucine	Higher density in V regions.
Tyrosine	Clusters between residues 99 to 108 of heavy chain and 86 to 91 of light chain.
Tryptophan	Conserved in homology regions (see Figure 3).

that the $V_{\gamma II}$ region is 118 residues long. Although definitive assignments must await determination of the sequences of several other proteins in both subgroups, these data suggest that V_H regions vary in length from 114 to 118 residues as compared with 107 to 112 residues for V_L regions.

The sequence variations of heavy-chains resemble those of light chains and most of the substitutions may be accounted for by single base changes in the genetic code. There is some suggestion that the heavy-chain variable regions may contain more variable positions (Cunningham *et al.*, 1969). Nevertheless, the distributions and types of changes indicate that the mechanism of variation is similar in both kinds of chains.

Both the light and the heavy chains have been examined for unusual distributions of amino acids (Table II). Although other chains may show different distributions as a result of variability, certain features deserve comment.

The V and C regions of light and heavy chains have similar amino acid compositions. There is no unusual or asymmetric distribution of hydrophilic and hydrophobic residues, with the exception of one portion (residues 149–203) of region C_H1. This portion of C_H1 has a preponderance of hydrophobic residues and very few amino acids with ionizable side chains (Table II and Figure 2). In contrast, the neighboring portion of the heavy chain (residues 210–255) in the so-called "hinge region" contains mainly hydrophilic residues. There is a higher density of amino acids with ionizable side chains in the Fc region (residues 223–446) than in the rest of the molecule.

Histidine is absent in both V_H and V_L regions of protein Eu. Proline shows a particularly high density in the "hinge region." Both glycine and alanine are represented in lower density in the Fc portion of the molecule. An unequal distribution of glycyl residues in V and C regions of κ chains has been previously noted (Kabat, 1967); this is also true

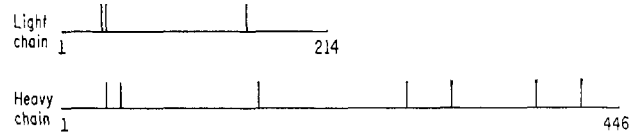

FIGURE 3: Location of tryptophanyl residues (vertical bars) in the light and heavy chains.

of the Eu κ chain. Isoleucine is present in higher density in the V regions. Tyrosyl residues show clusters near the ends of the variable regions of both heavy and light chains (Table II and Figures 1 and 2). Tryptophan has a periodic distribution (Figure 3) which appears to be related to a series of homologies in the amino acid sequence (Edelman *et al.*, 1969) and also possibly to the role of this hydrophobic amino acid in stabilizing the structure.

A periodic distribution is even more strikingly reflected in the half-cystine residues. Eu contains no free SH groups and each half-cystinyl residue (boxed in Figures 1 and 2) participates in a disulfide bond (Gall *et al.*, 1968; Gall and Edelman, 1970). Of the 16 disulfide bonds, 4 are interchain and 12 are intrachain. Each half-cystine participating in an intrachain bond is linked to the succeeding cystine in the chain. All of the half-cystinyl residues which contribute to the interchain bonds are located between residues 220 and 229 in what may be the "hinge region" of the molecule (Feinstein and Rowe, 1965).

Homologies and the Evidence for Gene Duplication and Translocation. The determination of the sequence and precise location of the disulfide bonds provides an unequivocal confirmation of the arrangement of light and heavy chains as well as the disposition of the proteolytic fragments. The primary structure shows a periodic arrangement (Figure 4). From the NH_2 termini to half-cystines V, the light and heavy chains can be aligned with corresponding intrachain disulfide bonds in similar positions. The disulfide loops are linearly and periodically disposed and are approximately of the same size. Each light chain is linked to its corresponding heavy chain by a disulfide bond between half-cystines V. Half-

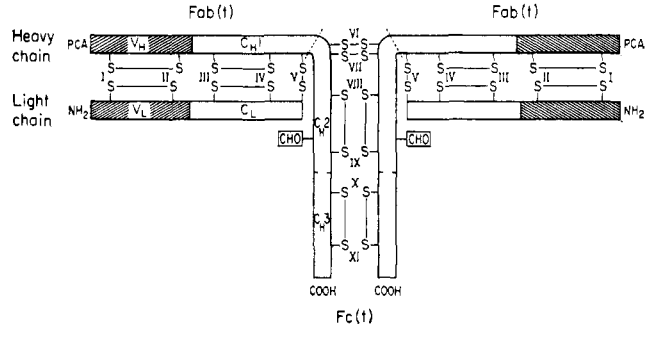

FIGURE 4: Overall arrangement of chains and disulfide bonds of γG1-immunoglobulin Eu. Half-cystinyl residues are I–XI; I–V designate corresponding residues in light and heavy chains. PCA, pyrrolidonecarboxylic acid; CHO, carbohydrate. Fab(t) and Fc(t) refer to fragments produced by trypsin, which cleaves the heavy chain as indicated by dashed lines above half-cystinyl residues VI. V_H, V_L: variable regions of heavy and light chains. C_L, constant region of light chain. C_H1, C_H2, C_H3: homology regions comprising C_H or constant region of heavy chain.

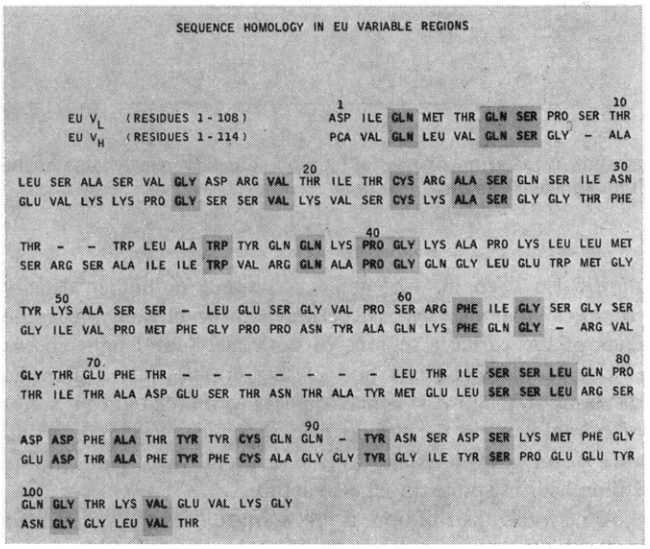

FIGURE 5: Comparison of V_H and V_L regions of protein Eu. Identical residues are shaded. Deletions indicated by dashes are introduced to maximize the homology.

cystines VI and VII link the heavy chains of each half-molecule together, and the cleavage by trypsin at Lys-222 to form Fab(t) and Fc(t) fragments occurs just four residues from half-cystine VI (residue 226). Although the structure must be special in this region and the distribution of portions of the three-dimensional structure must be affected by it (Valentine and Green, 1967; Pilz *et al.*, 1970), it is obvious that the Fab and Fe fragments are not biosynthetic subunits of the molecule.

Comparisons of the sequences of heavy and light chains for the presence of homologies indicate that the linear and periodic relationships described above are not fortuitous.

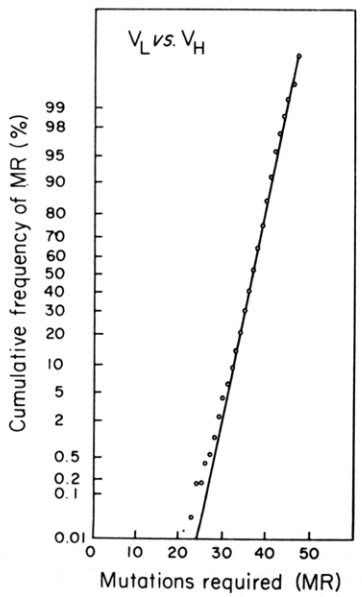

FIGURE 6: Cumulative probability plot resulting from comparison of V_L and V_H regions according to the program of Fitch using same deletions as in Figure 5.

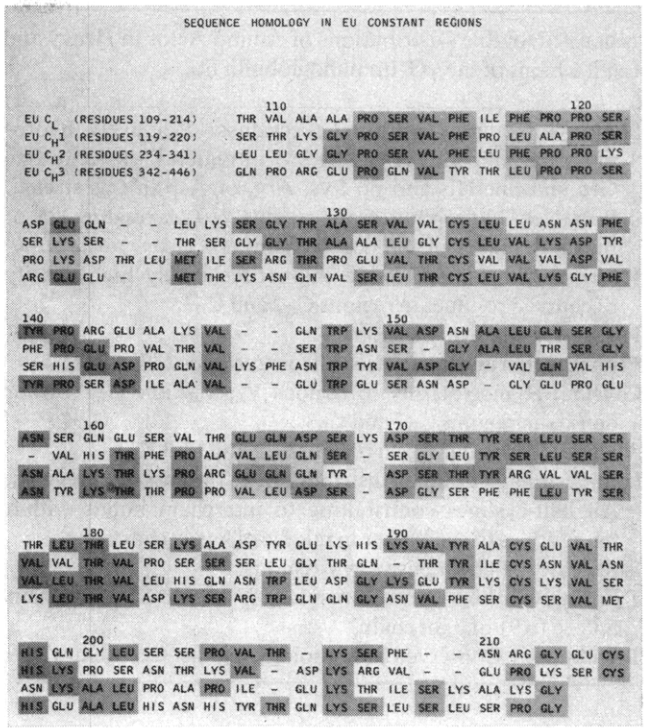

FIGURE 7: Comparison of the amino acid sequences of C_L, C_H1, C_H2, and C_H3 regions. Deletions, indicated by dashes, have been introduced to maximize homologies. Identical residues are darkly shaded; both dark and light shading are used to indicate identities which occur in pairs in the same positions.

The V_H and V_L regions are compared in Figure 5. Of 104 residues there are 25 identical residues in similar positions. Homology of these regions was also detected by the method of Fitch (Figure 6) although the deviation from the straight line in the cumulative probability plot was not striking. It should be pointed out that these comparisons (Figure 5 and Figure 6) were made with similar deletions, and the bulk of the evidence suggests that V region sequences from both kinds of chains evolved from a common precursor.

A similar comparison of C_H and C_L regions is shown in Figure 7; the indications of homology between these regions are very striking and there is little doubt that the constant regions are homologous. In a stretch of 100 residues any two of the four regions are identical in 29–34 positions. These direct comparisons are confirmed by computer analyses (Figure 8), which further substantiate the conclusion that C_L is homologous to three successive adjacent regions of the heavy-chain constant region, designated C_H1, C_H2, and C_H3. The stretch in the heavy chain from residue 221 to 233 in the hinge region (see Figures 2 and 4) is not homologous to any other portion of the light or heavy chains. This region, which has a high proline content, contains the two closely neighboring interchain disulfide bonds. As pointed out previously (Gall *et al.*, 1968), the arrangement of these disulfide bonds may have arisen in evolution by misalignment and unequal crossing-over.

In contrast to the homologies discussed above, comparisons of the variable with the constant regions show little if any deviation in the cumulative probability plots (Figure 9) and direct visual inspection also fails to indicate

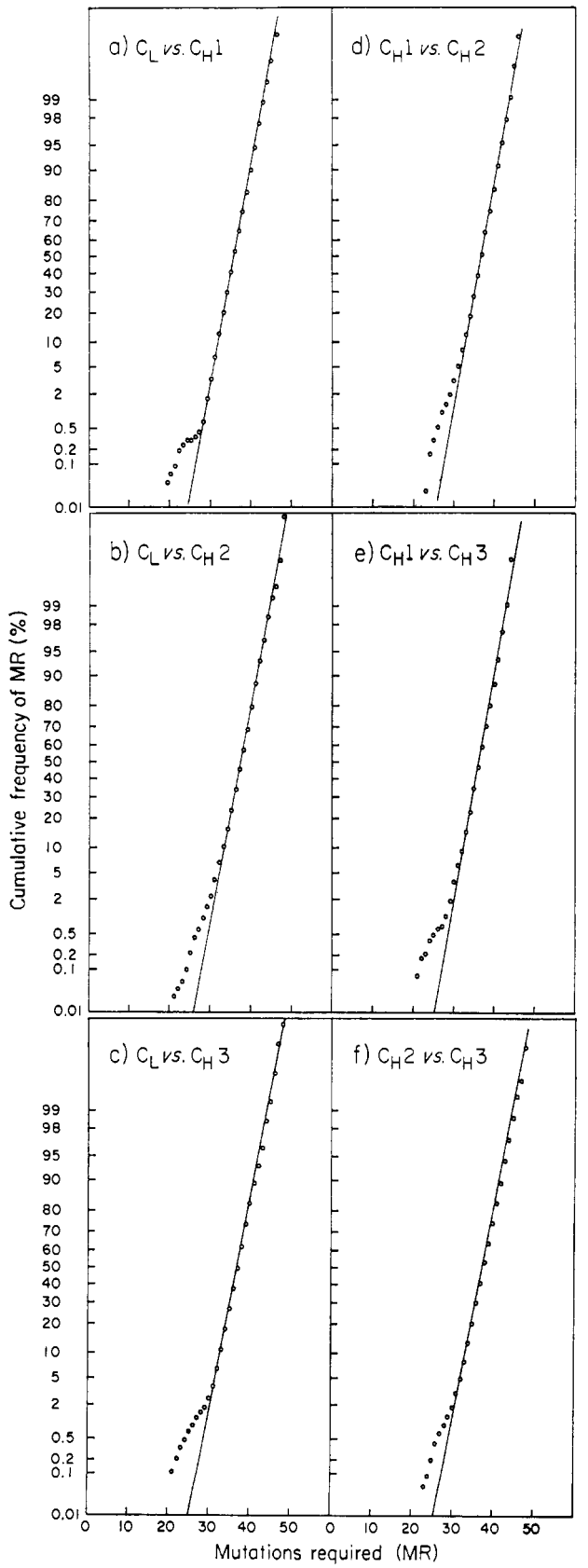

FIGURE 8: Cumulative probability plots comparing C_L, C_H1, C_H2, and C_H3 regions.

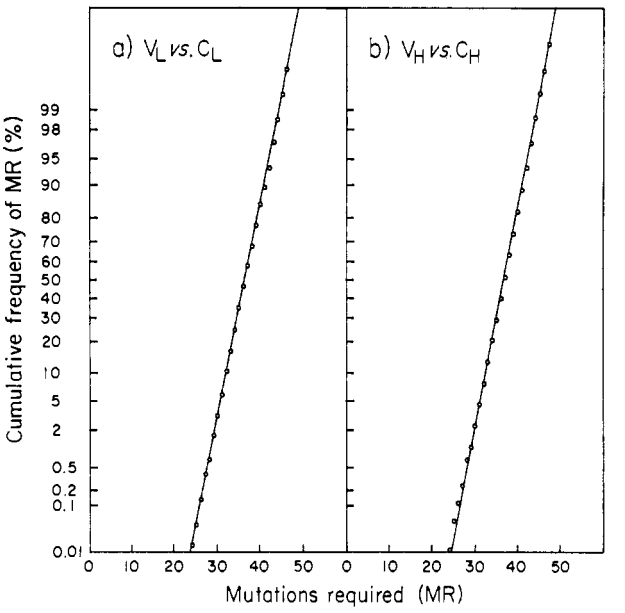

FIGURE 9: Cumulative probability plots comparing V_L and C_L and V_H and C_H regions.

homologies. Nevertheless, the similar amino acid compositions, lengths, and disulfide-bond distributions in these regions suggest the possibility that they may have been derived from a common ancestor. Moreover, it is possible that homologies between these regions may have been obscured by the diversification mechanism.

All of the homologies among the portions of the polypeptide chains of Eu are summarized in Figure 10. The relationships derived from the amino acid sequence provide convincing evidence that the immunoglobulin molecule evolved by successive duplication of precursor genes as suggested by Hill *et al.* (1966) and Singer and Doolittle (1966). The precursor gene apparently was sufficient in size to specify a polypeptide of 107–110 residues and duplication occurred as a tandem event for the C_H region. The order of emergence of C_L or C_H genes is not apparent from the data. A comparison of the sequences of constant homology regions with sequences in the constant regions of μ, α, δ, and ϵ chains may reveal whether any of these regions are conserved or to what extent they have been modified in the different classes. At present, however, the data are too scanty to specify in detail the order of evolutionary emergence of either heavy chains and their homology regions or of the immunoglobulin classes (Marchalonis and Edelman, 1968).

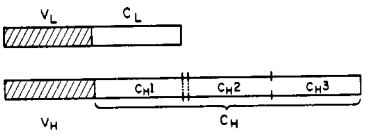

FIGURE 10: Diagram summarizing internal homologies in the structure of γG1-immunoglobulin Eu. Variable regions V_H and V_L are homologous. The constant region of the heavy chain (C_H) is divided into three regions, C_H1, C_H2, and C_H3, which are homologous to each other and to the C region of the light chain. The hinge region is located between the dotted lines.

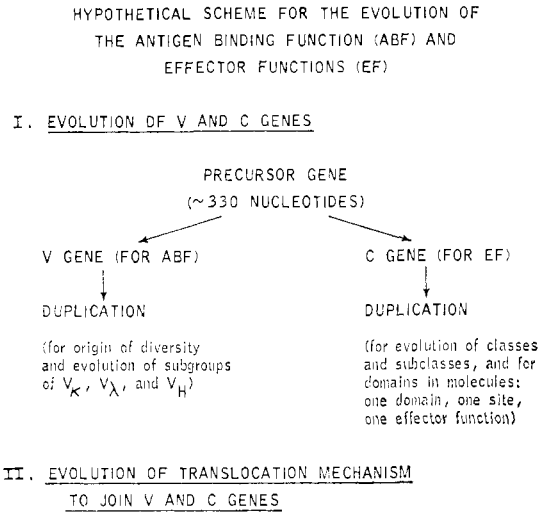

FIGURE 11: Hypothetical scheme for evolution of antigen binding functions (ABF) and effector functions (EF) of immunoglobulins.

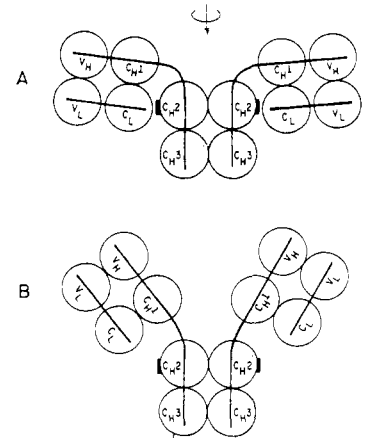

FIGURE 12: The domain hypothesis. Homology regions (see Figure 10) which constitute each domain are indicated. V_L, V_H—domains made up of variable homology regions. C_L, C_H1, C_H2, C_H3—domains made up of constant homology regions. Within each of these groups, domains are assumed to have similar three-dimensional structures and each is assumed to contribute to an active site. The V domain sites contribute to antigen recognition functions and the C domain sites to effector functions. (A) Hypothetical arrangement of domains in the free immunoglobulin molecule. The arrow refers to a dyad axis of symmetry. (B) Suggested rearrangement after antigen binding. (▮) Complement binding site.

The failure to find homologies between V and C regions suggests that V and C genes diverged early in evolution, the V genes for antigen binding functions and the C genes for effector functions (Figure 11). This hypothesis would account for the regional differentiation of these two fundamentally different functions within the molecule, but it requires the evolution of a mechanism to link the genes or their products. The evidence on nonallelic subgroups and genetic markers (Milstein *et al.*, 1969; Hood *et al.*, 1967) and the comparison of the heavy chains of Eu and He (Cunningham *et al.*, 1969) support the idea that the heavy and light polypeptide chains of the immunoglobulin molecule are each specified by two genes, V and C. It has been proposed that a V gene for a particular chain may be translocated to one of the two parental C genes to form a single VC gene in lymphoid cell precursors. This translocation hypothesis (Edelman and Gally, 1968; Edelman and Gall, 1969) may account in part for clonal expression and allelic exclusion in antibody production. Translocation may therefore be an important means for irreversible differentiation and commitment of a lymphoid precursor cell.

Comparison of the V regions of γ and μ chains (Wikler *et al.*, 1969) further substantiates this hypothesis and suggests that the same set of V genes may be used to specify all heavy-chain classes. If this is verified, the other main sets of V genes would be those specifying variable regions of κ and λ chains which are found in all classes of immunoglobulins. An arrangement of V genes into heavy and light chain sets might simplify complementation by V_H and V_L regions to form antibody sites. We have previously suggested (Edelman and Gally, 1964) that pairing of light and heavy chains is unrestricted, *i.e.*, if there are p V_L regions and q V_H regions then there are $p \times q$ possible interactions. Such a mechanism could serve powerfully to increase the variability of immunoglobulins. The alternative to relatively unrestricted chain interactions is that each V_L region must be paired with a unique V_H region in the cell (Tanford, 1968). Unique pairing would require an unprecedented mechanism for specific complementation of V_L and V_H genes in the lymphoid cell (Edelman and Gall, 1969) and it therefore seems unlikely. The apparent light-chain–heavy-chain specificity is more likely a result of clonal selection. This in turn implies that an immunoglobulin which can serve effectively as an antibody has particular light and heavy chains with high affinities for each other.

Domain Hypothesis. Homologies within the sequences of the polypeptide chains suggest that each homology region of immunoglobulin molecules may be folded in a compact domain (Figure 12A) (Edelman *et al.*, 1969; Edelman and Gall, 1969). This hypothesis is supported by the alignment of disulfide bonds and symmetry axes, as well as by the location of the sites of cleavage of the molecule by proteolytic enzymes (Solomon and McLaughlin, 1969; Karlsson, *et al.*, 1969; Turner and Bennich, 1968). Each domain is stabilized by a single disulfide bond and is linked to neighboring domains by less compactly folded stretches of the polypeptide chain. Within each group of V or C homology regions the domains would have similar but not identical tertiary structures, and each domain would contribute to at least one active site mediating a function of the class of immunoglobulins to which it belongs.

The domain hypothesis is consistent with the hypotheses of gene duplication and translocation. It is also consistent with the suggestion (Singer and Thorpe, 1968) that a pseudo-symmetry axis exists between the light and heavy chains similar to the rotation axis which passes through the two disulfide bonds connecting the heavy chains (Terry *et al.*, 1968). Proof for the existence of domains may come from electron microscopic (Valentine and Green, 1967) and X-ray crystallographic studies of immunoglobulins (Poljak *et al.*, 1967; Goldstein *et al.*, 1968; Terry *et al.*, 1968). The hypothesis predicts that limited proteolysis of Fab fragments might yield halves of Fd fragments; similar treatment of Bence-Jones proteins has already been shown to produce V_L and

C_L fragments (Karlsson *et al.*, 1969). Additional chemical evidence may come from determination of the location of sites for complement fixation and skin fixation in the C_H regions. Recent studies of Kehoe and Fougereau (1970) suggest that a complement binding site is located in the C_H2 domain.

The finding of additional intrachain disulfide bonds between homology regions may weaken the domain hypothesis. Two closely staggered disulfide bonds have been found in protein Daw (Press and Hogg, 1969), but because of their arrangement they could be located within the same domain. It is reasonable to suggest that the nonhomologous hinge region and the interchain disulfide bonds (see Figure 10) are not in a domain. The arrangement of the relatively hydrophilic portions of this region in an extended conformation near the surface of the molecule may result in the difference in susceptibility to reduction of interchain and intrachain bonds (Cecil and Wake, 1962).

At present, the only hypothetical domains to which a function can be definitely assigned are the C_H2 domain and the V_H and V_L domains. Affinity-labeling experiments (Singer and Thorpe, 1968) indicate that tyrosyl residues in the V regions are directly concerned with antigen binding. Examination of the primary structure of V_L and V_H regions indicates that the disulfide bridges, which appear to be constant features of these regions, must lie close to the tyrosyl residues implicated in antigen binding (residues 86 of the light chain and 94 of the heavy chain). This leads to the suggestion that the antigen combining site may be fixed and stabilized by these bonds. Amino acid variations in branches of each chain connected by the intrachain bond in the V region may be sterically arranged around the bond as a center. Neighboring domains, consisting of the closely homologous C_L and C_H1 regions, may provide additional stabilization of the antigen combining site in the face of the amino acid sequence variation. Because light chains are common to all immunoglobulin classes, it would be expected that C_H1 domains would tend to be conserved, *i.e.*, they would be similar insequence in the various immunoglobulin classes.

The polar distribution of variable domains with respect to the constant domains (see Figure 12A) is in accord with the idea that the function of antigen binding does not depend upon effector functions. It is nevertheless likely that the antigen binding and effector functions must work in concert with high efficiency, for example, to trigger cell maturation and division in the selective immune response (Edelman and Gall, 1969). It has been suggested (Cohn, 1970) that effector functions such as the triggering of cell maturation and complement binding depend upon a conformational change in the molecule following binding of the antigen (Feinstein and Rowe, 1965). There is a difference in the dimensions of the Eu immunoglobulin molecule as measured by low-angle X-ray scattering (Pilz *et al.*, 1970) and the dimensions of rabbit γG antibodies measured by electron microscopy after interaction with the dinitrophenyl hapten (Valentine and Green, 1967). The calculated radius of gyration based on the electron microscopic data for hapten-bound antibody is smaller than the measured value of free immunoglobulin in solution (Pilz *et al.*, 1970). This discrepancy has been explained in terms of collapse of the somewhat flexible structure following antigen binding. In addition, an observed decrease in frictional coefficient that results after hapten binding has been taken to mean that the molecule assumes a more compact conformation upon interacting with the hapten (Warner *et al.*, 1970).

All of these observations may be interpreted in terms of the domain hypothesis as shown in Figure 12B. Pivoting within the molecule could unmask the complement binding site (Kehoe and Fougereau, 1969) in the C_H2 domain. A similar rearrangement may occur upon heating immunoglobulins, thus allowing them to bind complement components more efficiently.

Although there have been several studies on the intracellular assembly of immunoglobulins from their polypeptide chains, no single mechanism has been defined (Askonas and Williamson, 1967; Scharff *et al.*, 1967). The domain hypothesis suggests that in this assembly each domain is folded before interaction with its corresponding domain across a symmetry or pseudosymmetry axis. This is in accord with evidence (Jaton *et al.*, 1968) that the information for folding of each domain resides mainly in the primary structure of that domain. Experiments on reconstitution of immunoglobulins *in vitro* (Olins and Edelman, 1964) have shown that spontaneous assembly can occur from mixtures of light and heavy chains, indicating that pairing of the domains comprising V_LC_L and V_HC_H1 regions is a key step in assembly of the molecule. The insolubility of isolated heavy chains which is frequently observed *in vitro* (Olins and Edelman, 1964) may be the result of incorrect pairing of V_HC_H1 regions between more than two chains. Because they are composed of two light chains, dimeric forms of Bence-Jones proteins (Gally and Edelman, 1964, 1965) may have tertiary structures very similar to that of the immunoglobulin domains in which V_HC_H1 is paired with V_LC_L. If so, one would predict that the interchain disulfide bond in the Bence-Jones dimer is at one end of the molecule in the three-dimensional structure. Such dimers may have functioned as monovalent antibodies during evolution and the evolution of divalency could have occurred by addition of the homology regions C_H2 and C_H3.

The presence of homologous regions in immunoglobulins and the arguments for their arrangement into domains raise the possibility that other proteins that contain polypeptide chains of great length are similarly organized. It would not be surprising to find that tandem homology regions occur more generally in large proteins. Gene duplication has been implicated in the evolution of proteins such as ferredoxin (Eck and Dayhoff, 1966) and haptoglobin (Beach and Dixon, 1968). The periodic pattern found in immunoglobulins is more extensive than that found in the primary structure of other proteins, however. The variety of shapes that could result from the large number of ways of arranging repeated homologous regions across symmetry and pseudosymmetry axes in multichain structures may have evolutionary advantages.

There is an additional general implication for protein chemistry of the structural findings on Bence-Jones proteins and immunoglobulins. The determination of tertiary by primary structure (Perutz *et al.*, 1965) remains a fundamental and formidable problem. Bence-Jones proteins may provide powerful experimental models for the analysis of this "folding" problem, for they have several desirable properties: (1) They represent an enormously diverse set of protein sequences. (2) The sequences are disposed around several

basic patterns (*i.e.*, κ and λ chains and their subgroups). Each pattern has presumably evolved to serve several functions, the V region for variation in shape, the C region for conservation of shape and interchain binding. (3) There is evidence (Edelman and Gally, 1962) to suggest that each sequence has a different tertiary structure. (4) They may be crystallized, and V regions can be cleaved from C regions and separately crystallized (Solomon and McLaughlin, 1969). Furthermore, the molecular weights of the chains are relatively low.

From a detailed analysis of sequence and crystallographic data on a number of Bence–Jones proteins, rules may emerge which shed light on the folding problem. In any event, correlations of such studies with those on crystallography of Fab and Fc fragments and whole γG-immunoglobulins should be particularly valuable in determining the conformations of antigen-combining sites and other functionally important regions of antibody molecules.

Acknowledgment

I am grateful to Dr. W. Einar Gall for his help in analyzing the data.

References

Askonas, B. A., and Williamson, A. R. (1967), *Cold Spring Harbor Symp. Quant. Biol. 32*, 223.

Baglioni, C., Zonta, L. A., Cioli, D., and Carbonara, A. (1966), *Science 152*, 1517.

Beach, J. A., and Dixon, G. H. (1968), *Nature (London) 218*, 736.

Bennett, C., Konigsberg, W. H., and Edelman, G. M. (1970), *Biochemistry 9*, 3181.

Catley, B. J., Moore, S., and Stein, W. H. (1969), *J. Biol. Chem. 244*, 933.

Cecil, R., and Wake, A. G. (1962), *Biochem. J. 82*, 401.

Cohn, M. (1970), *in* Control Processes in Multicellular Organisms, Wolstenholme, G. E. W., and Knight, J., Ed, Ciba Foundation Symposium, London, J. and A. Churchill, p 255.

Cunningham, B. A., Gottlieb, P. D., Konigsberg, W. H., and Edelman, G. M. (1968), *Biochemistry 7*, 1983.

Cunningham, B. A., Pflumm, M. N., Rutishauser, U., and Edelman, G. M. (1969), *Proc. Nat. Acad. Sci. U. S. 64*, 997.

Cunningham, B. A., Rutishauser, U., Gall, W. E., Gottlieb, P. D., Waxdal, M. J., and Edelman, G. M. (1970), *Biochemistry 9*, 3161.

Eck, R. U., and Dayhoff, M. O. (1966), *Science 152*, 363.

Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., and Waxdal, M. J. (1969), *Proc. Nat. Acad. Sci. U. S. 63*, 78.

Edelman, G. M., and Gall, W. E. (1969), *Annu. Rev. Biochem. 38*, 415.

Edelman, G. M., Gall, W. E., Waxdal, M. J., and Konigsberg, W. H. (1968), *Biochemistry 7*, 1950.

Edelman, G. M., and Gally, J. A. (1962), *J. Exp. Med. 116*, 207.

Edelman, G. M., and Gally, J. A. (1964), *Proc. Nat. Acad. Sci. U. S. 51*, 846.

Edelman, G. M., and Gally, J. A. (1968), *Brookhaven Symp. Biol. 21*, 328.

Eylar, E. H. (1966), *J. Theor. Biol. 10*, 89.

Feinstein, A., and Rowe, A. J. (1965), *Nature (London) 205*, 147.

Fitch, W. M. (1966), *J. Mol. Biol. 16*, 9.

Gall, W. E., Cunningham, B. A., Waxdal, M. J., Konigsberg, W. H., and Edelman, G. M. (1968), *Biochemistry 7*, 1973.

Gall, W. E., and Edelman, G. M. (1970), *Biochemistry 9*, 3188.

Gally, J. A., and Edelman, G. M. (1964), *J. Exp. Med. 119*, 817.

Gally, J. A., and Edelman, G. M. (1965), *Biochim. Biophys. Acta 94*, 175.

Goldstein, D. J., Humphrey, R. L., and Poljak, R. J. (1968), *J. Mol. Biol. 35*, 247.

Gottlieb, P. D., Cunningham, B. A., Rutishauser, U., and Edelman, G. M. (1970), *Biochemistry 9*, 3155.

Gottschalk, A. (1969), *Nature (London) 222*, 452.

Hill, R. L., Delaney, R., Fellows, R. E., Jr., and Lebovitz, H. E. (1966), *Proc. Nat. Acad. Sci. U. S. 56*, 1762.

Hilschmann, N., and Craig, L. C. (1965), *Proc. Nat. Acad. Sci. U. S. 53*, 1403.

Hood, L., Gray, W. R., Sanders, G. G., and Dreyer, W. J. (1967), *Cold Spring Harbor Symp. Quant. Biol. 32*, 133.

Jaton, J. C., Klinman, N. R., Givol, D., and Sela, M. (1968), *Biochemistry 7*, 4185.

Kabat, E. (1967), *Proc. Nat. Acad. Sci. U. S. 57*, 1345.

Karlsson, F. A., Peterson, P. A., and Berggard, I. (1969), *Proc. Nat. Acad. Sci. U. S. 64*, 1257.

Kehoe, J. M., and Fougereau, M. (1970), *Nature (London) 224*, 1212.

Marchalonis, J. J., and Edelman, G. M. (1968), *J. Exp. Med. 127*, 8911

Milstein, C. (1967), *Nature (London) 216*, 330.

Milstein, C. (1969), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 2*, 301.

Milstein, C., Milstein, C. P., and Feinstein, A. (1969), *Nature (London) 221*, 151.

Niall, H. D., and Edman, P. (1967), *Nature (London) 216*, 262.

Olins, D. E., and Edelman, G. M. (1964), *J. Exp. Med. 119*, 789.

Perutz, M. F., Kendrew, J. C., and Watson, H. C. (1965), *J. Mol. Biol. 13*, 669.

Pilz, I., Puchwein, G., Kratky, O., Herbst, M., Haager, O., Gall, W. E., and Edelman, G. M. (1970), *Biochemistry 9*, 211.

Poljak, R. J., Dintzis, H. M., and Goldstein, D. J. (1967), *J. Mol. Biol. 24*, 351.

Press, E. M., and Hogg, N. M. (1969), *Nature (London) 223*, 807.

Press, E. M., Piggot, P. J., and Porter, R. R. (1966), *Biochem. J. 99*, 356.

Rutishauser, U., Cunningham, B. A., Bennett, C., Konigsberg, W. H., and Edelman, G. M. (1968), *Proc. Nat. Acad. Sci. U. S. 61*, 1414.

Rutishauser, U., Cunningham, B. A., Bennett, C., Konigsberg, W. H., and Edelman, G. M. (1970), *Biochemistry 9*, 3171.

Scharff, M. D., Shapiro, A. L., and Ginsberg, B. (1967), *Cold Spring Harbor Symp. Quant. Biol. 32*, 235.

Singer, S. J., and Doolittle, R. F. (1966), *Science 153*, 13.

Singer, S. J., and Thorpe, N. O. (1968), *Proc. Nat. Acad. Sci. U. S. 60*, 1371.

Solomon, A., and McLaughlin, C. L. (1969), *J. Biol. Chem. 244*, 3393.

Tanford, C. (1968), *Accounts Chem. Res. 1*, 161.
Terry, W. D., Matthews, B. W., and Davies, D. R. (1968), *Nature (London) 220*, 239.
Thorpe, N. O., and Deutsch, H. F. (1966), *Immunochemistry 3*, 329.
Titani, K., Wikler, M., and Putnam, F. (1967), *Science 155*, 829.
Turner, M. W., and Bennich, H. (1968), *Biochem. J. 107*, 171.
Valentine, R. C., and Green, N. M. (1967), *J. Mol. Biol. 27*, 615.
Warner, C., Schumaker, V., and Karush, F. (1970), *Biochem. Biophys. Res. Commun. 38*, 125.
Waxdal, M. J., Konigsberg, W. H., and Edelman, G. M. (1968a), *Biochemistry 7*, 1967.
Waxdal, M. J., Konigsberg, W. H., Henley, W. L., and Edelman, G. M. (1968b), *Biochemistry 7*, 1959.
Wikler, M., Kohler, H., Shinoda, T., and Putnam, F. W. (1969), *Science 163*, 7.

Nitration of Human Serum Albumin and Bovine and Human Goiter Thyroglobulins with Tetranitromethane*

P. Gordon Malan† and Harold Edelhoch

ABSTRACT: Tetranitromethane was used as a chemical probe to examine the physical and chemical behavior of tyrosyl residues of human serum albumin and of bovine and human goiter thyroglobulins. Only 50% of the tyrosyl residues of human serum albumin could be nitrated using excess reagent at pH 8.0. About 70% of the 125 tyrosyl residues of both thyroglobulins were nitrated at high levels of tetranitromethane. Sedimentation velocity ultracentrifugation indicated little structural change in the proteins until high levels of nitration were reached. Human goiter thyroglobulin (containing 0.06% iodine) was chemically iodinated with tracer ^{131}I. Subsequent nitration with tetranitromethane released about 20% of the iodine. Spectrophotometric titration data of nitrophenol and phenol groups in nitrated serum albumin were analyzed by the Linderstrøm-Lang theory. The plot showed two regions where the electrostatic interaction factor was about four and two times the value obtained by other workers for the ionization behavior of native serum albumin over the same pH range. The nitrotyrosyl groups in human serum albumin appear to be represented by a range of pK values.

Ultraviolet difference spectroscopy, using 8 M urea as perturbant, showed red-shifted nitrotyrosyl difference peaks in nitrated thyroglobulin and serum albumin. In contrast to iodotyrosyl residues which show blue-shifted spectra, nitrotyrosyl residues are exposed to the solvent. This difference in behavior presumably reflects differences in polarity between nitro- and iodophenol groups. Nitration of free tyrosine was also examined and the results indicated that nitrotyrosine and other by-products were formed.

This report is part of a continuing study to characterize the chemical and physical behavior of the tyrosyl residues of thyroglobulin (Edelhoch and Perlman, 1968). The function of thyroglobulin is to synthesize (and store) the thyroid hormones, triiodothyronine and thyroxine. Since the hormones are formed in the protein by iodination and coupling of tyrosyl residues, the latter constitute what could be referred to as the active sites of this protein. In bovine thyroglobulin containing about 1% iodine (distributed in about 10 monoiodotyrosyl, 10 diiodotyrosyl, 5 thyroxyl, and less than 1 triiodothyronyl residues) about 35% of the iodinated tyrosyl residues have been coupled to form iodothyronyl residues. This coupling of iodotyrosyl residues, which occurs when only a relatively small fraction (~20%) of the tyrosyl groups have been iodinated, results in a high efficiency of conversion of iodine from precursor to hormonal forms, *i.e.*, about 45% (de Crombrugghe *et al.*, 1967).

The present study is an evaluation of the reactivity of the tyrosyl residues in thyroglobulin with tetranitromethane,[1] a reagent which attacks the same sites as iodine. Since the reaction mechanism of nitration of TNM is different from iodination (Bruice *et al.*, 1968), the influence of the environment of the tyrosyl residues in controlling their reactivity is evaluated. This reagent should also be useful in future studies designed to distinguish between those tyrosyl residues which can couple to form hormone and those which cannot. If the sequence of thyroglobulin becomes known it should then be possible to identify the residues which form hormones on iodination but are prevented by nitration.

It has been shown in several enzymes that tyrosyl residues which constitute part of the active site are preferentially nitrated with TNM (Riordan *et al.*, 1967; Cuatrecasas *et al.*, 1968; Skov *et al.*, 1969). Similar studies with iodination have revealed greater reactivity of tyrosyl residues in active site regions than those that are not involved in the function of the

* From the Clinical Endocrinology Branch, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Maryland 20014. *Received April 27, 1970.*
† Postdoctoral Fellow of the Arthritis Foundation, New York, N. Y.

[1] Abbreviations used are: TNM, tetranitromethane: HSA, human serum albumin; Tg, thyroglobulin; NO_2Tyr, 3-nitrotyrosine.