

*THE COVALENT STRUCTURE OF  
AN ENTIRE  $\gamma$ G IMMUNOGLOBULIN MOLECULE\**

BY GERALD M. EDELMAN, BRUCE A. CUNNINGHAM, W. EINAR GALL,  
PAUL D. GOTTLIEB, URS RUTISHAUSER, AND MYRON J. WAXDAL

THE ROCKEFELLER UNIVERSITY

*Communicated by Theodore Shedlovsky, March 21, 1969*

*Abstract.*—The complete amino acid sequence of a human  $\gamma$ G1 immunoglobulin (Eu) has been determined and the arrangement of all of the disulfide bonds has been established. Comparison of the sequence with that of another myeloma protein (He) suggests that the variable regions of heavy and light chains are homologous and similar in length. The constant portion of the heavy chain contains three homology regions each of which is similar in size and homologous to the constant region of the light chain. Each variable region and each constant homology region contains one intrachain disulfide bond. The half-cystines participating in the interchain bonds are all clustered within a stretch of ten residues at the middle of the heavy chains.

These data support the hypothesis that immunoglobulins evolved by gene duplication after early divergence of V genes, which specified antigen-binding functions, and C genes, which specified other functions of antibody molecules. Each polypeptide chain may therefore be specified by two genes, V and C, which are fused to form a single gene (translocation hypothesis). The internal homologies and symmetry of the molecule suggest that homology regions may have similar three-dimensional structures each consisting of a compact domain which contributes to at least one active site (domain hypothesis). Both hypotheses are in accord with the linear regional differentiation of function in antibody molecules.

---

Antibodies or immunoglobulins can interact with a wide range of different antigenic determinants and, after specific binding to an antigen, they play a fundamental part in physiological functions of the immune response. The specificity of antigen binding depends ultimately upon amino acid sequences of the variable or V regions of antibody molecules. It is the diversity of these sequences which results in the range of specificities required for a selective immune response. In contrast, other regions of the antibody molecule have relatively constant sequences and are responsible for physiological functions. Like enzymes, these C regions appear to have evolved for a restricted set of interactions. This unusual picture of intramolecular differentiation has emerged from studies of the structure of immunoglobulins from different animal species.<sup>1</sup> To date, only portions of immunoglobulin molecules have been subjected to amino acid sequence determination.

We now report the amino acid sequence of an entire human  $\gamma$ G1 immunoglobulin (molecular weight 150,000), the location of all disulfide bonds, the arrangement of light and heavy chains, and the length of the heavy chain V region.

*Materials and Methods.*—The isolation of the myeloma protein Eu<sup>2</sup> and the preparation of its CNBr fragments<sup>3, 4</sup> have been described. Similar methods were used for the isolation of the  $\gamma$ G1 myeloma protein He and for the preparation of its CNBr fragments.

We have previously described the methods used for enzymatic digestion with trypsin, chymotrypsin, and pepsin, gel filtration, ion exchange chromatography, high voltage paper electrophoresis, determination of NH<sub>2</sub>-terminal and COOH-terminal residues, amino acid analysis, and determination of amino acid sequences by the dansyl-Edman procedure.<sup>4-6</sup>

The positions of glutamine and asparagine were assigned<sup>9</sup> by determining the electrophoretic mobility of peptides and by amino acid analysis of the peptides after enzymatic hydrolysis. The half-cystinyl residues contributing to each intrachain disulfide bond were determined<sup>10</sup> using the diagonal electrophoresis method.<sup>11</sup>

*Results.*—The organization of the whole molecule is shown in Figure 1; an unequivocal proof of the arrangement of the two identical light chains and two identical heavy chains has already been given.<sup>4</sup> Each light chain is linked to its neighboring heavy chain by a disulfide bond between corresponding half-

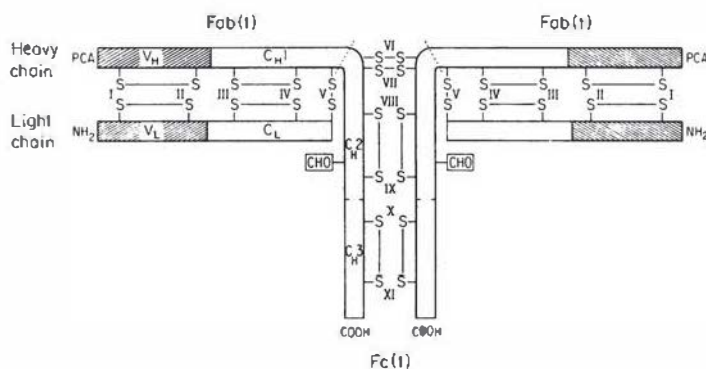


FIG. 1.—Over-all arrangement of chains and disulfide bonds of  $\gamma$ G1 immunoglobulin Eu. Half-cystinyl residues are numbered I–XI; numbers I–V designate corresponding residues in light and heavy chains. PCA: pyrrolidonecarboxylic acid. CHO: carbohydrate. “Fab(t)” and “Fc(t)” refer to fragments produced by trypsin, which cleaves the heavy chain as indicated by dashed lines above half-cystinyl residues VI. V<sub>H</sub>, V<sub>L</sub>: variable regions of heavy and light chains, C<sub>L</sub>: constant region of light chain. C<sub>H</sub>1, C<sub>H</sub>2, C<sub>H</sub>3: homology regions comprising C<sub>H</sub> or constant region of heavy chain.

cystines V. Half-cystines VI and VII form bonds linking the half-molecules via the heavy chains. Trypsin cleaves the molecule at lysyl residue 222 to form two Fab(t) and one Fc(t) fragments.<sup>2, 5</sup>

There are several strikingly linear arrangements in the primary structure. From their amino termini to half-cystines V, the light and heavy chains can be aligned or put in register. The intrachain disulfide bonds are linearly and periodically disposed.<sup>12, 13</sup> In accord with the alignment of light and heavy chains, corresponding intrachain disulfide bonds are in similar positions and the disulfide loops are of approximately the same size.

Previous studies<sup>7</sup> have suggested that V regions of light and heavy chains have similar lengths and begin at the NH<sub>2</sub>-termini; this will be confirmed below. The C<sub>L</sub> region of the light chain has the same length as V<sub>L</sub>, but the C<sub>H</sub> region of the heavy chain is about three times as long. C<sub>H</sub> may be divided into three

homologous regions of approximately equal length:  $C_H1$ ,  $C_H2$ , and  $C_H3$  (Fig. 1).

We have already reported the amino acid sequence of the first 87<sup>7</sup> and the last 224 residues<sup>8</sup> of the heavy chain as well as the partial sequence of the entire light chain.<sup>6</sup> The complete amino acid sequence of the light chain (214 residues<sup>8</sup>) is shown in Figure 2. Positions of the half-cystinyl residues may be compared with Figure 1 and the methionyl residues may be correlated with previous studies on the CNBr fragments of Eu.<sup>3, 4</sup> The variable region extends through residue 108. In accord with other studies,<sup>1</sup> valine 191 is related to the Inv specificity.<sup>2</sup>

The complete sequence of the heavy chain (446 residues) is presented in Figure 3 which may be compared with Figure 2 for alignment with the light chain sequence. Isolation of a single glycopeptide<sup>8</sup> indicated that the polysaccharide

```

1          10          20
ASP-ILE-GLN-MEI-THR-GLN-SER-PRO-SER-THR-LEU-SER-ALA-SER-VAL-GLY-ASP-ARG-VAL-THR-
           30          40
ILE-THR-CYS-ARG-ALA-SER-GLN-SER-ILE-ASN-THR-TRP-LEU-ALA-TRP-TYR-GLN-GLN-LYS-PRO-
           50          60
GLY-LYS-ALA-PRO-LYS-LEU-LEU-MET-TYR-LYS-ALA-SER-SER-LEU-GLU-SER-GLY-VAL-PRO-SER-
           70          80
ARG-PHE-ILE-GLY-SER-GLY-SER-GLY-THR-GLU-PHE-THR-LEU-THR-ILE-SER-SER-LEU-GLN-PRO-
           90          100
ASP-ASP-PHE-ALA-THR-TYR-TYR-CYS-GLN-GLN-TYR-ASN-SER-ASP-SER-LYS-MET-PHE-GLY-GLN-
           110          120
GLY-THR-LYS-VAL-GLU-VAL-LYS-GLY-THR-VAL-ALA-ALA-PRO-SER-VAL-PHE-ILE-PHE-PRO-PRO-
           130          140
SER-ASP-GLU-GLN-LEU-LYS-SER-GLY-THR-ALA-SER-VAL-VAL-CYS-LEU-LEU-ASN-ASN-PHE-TYR-
           150          160
PRO-ARG-GLU-ALA-LYS-VAL-GLN-TRP-LYS-VAL-ASP-ASN-ALA-LEU-GLN-SER-GLY-ASN-SER-GLN-
           170          180
GLU-SER-VAL-THR-GLU-GLN-ASP-SER-LYS-ASP-SER-THR-TYR-SER-LEU-SER-SER-THR-LEU-THR-
           190          200
LEU-SER-LYS-ALA-ASP-TYR-GLU-LYS-HIS-LYS-VAL-TYR-ALA-CYS-GLU-VAL-THR-HIS-GLN-GLY-
           210          214
LEU-SER-SER-PRO-VAL-THR-LYS-SER-PHE-ASN-ARG-GLY-GLU-CYS

```

FIG. 2.—Complete amino acid sequence of the Eu light chain. Half-cystinyl residues are in boxes and methionyl residues are underlined.

portion of the molecule is attached at Asx residue 297.<sup>14</sup> In a previous study<sup>8</sup> we have suggested that glutamyl residue 356 and methionyl residue 358 may be associated with Gm 1 specificities. The sequence of Eu (Gm 4+) between residues 211–252 can be compared with the partial sequence of immunoglobulin Daw<sup>17</sup> (Gm 4–). The presence of arginine in position 214 of Eu and lysine in a comparable position of Daw may be associated with their Gm 4 specificities.<sup>18</sup>

Of particular significance is the determination of the point at which  $V_H$  ends and  $C_H$  begins. A CNBr fragment comparable to fragment  $H_4$  was isolated from myeloma protein He which has the same Gm specificity as protein Eu. The sequence of the amino terminal portion of the CNBr fragment from He differed from that of the  $H_4$  fragment from Eu.<sup>21</sup>

1  
 PCA - VAL - GLN - LEU - VAL - GLN - SER - GLY - ALA - GLU - VAL - LYS - LYS - PRO - GLY - SER - SER - VAL - LYS - VAL - 20  
 SER - CYS - LYS - ALA - SER - GLY - GLY - THR - PHE - SER - ARG - SER - ALA - ILE - ILE - TRP - VAL - ARG - GLN - ALA - 40  
 PRO - GLY - GLN - GLY - LEU - GLU - TRP - MET - GLY - GLY - ILE - VAL - PRO - MET - PHE - GLY - PRO - PRO - ASN - TYR - 60  
 ALA - GLN - LYS - PHE - GLN - GLY - ARG - VAL - THR - ILE - THR - ALA - ASP - GLU - SER - THR - ASN - THR - ALA - TYR - 80  
MET - GLU - LEU - SER - SER - LEU - ARG - SER - GLU - ASP - THR - ALA - PHE - TYR - PHE - CYS - ALA - GLY - GLY - 100  
 GLY - ILE - TYR - SER - PRO - GLU - GLU - TYR - ASN - GLY - GLY - LEU - VAL - THR - VAL - SER - SER - ALA - SER - THR - 120  
 LYS - GLY - PRO - SER - VAL - PHE - PRO - LEU - ALA - PRO - SER - SER - LYS - SER - THR - SER - GLY - GLY - THR - ALA - 140  
 ALA - LEU - GLY - CYS - LEU - VAL - LYS - ASP - TYR - PHE - PRO - GLU - PRO - VAL - THR - VAL - SER - TRP - ASN - SER - 160  
 GLY - ALA - LEU - THR - SER - GLY - VAL - HIS - THR - PHE - PRO - ALA - VAL - LEU - GLN - SER - SER - GLY - LEU - TYR - 180  
 SER - LEU - SER - SER - VAL - VAL - THR - VAL - PRO - SER - SER - SER - LEU - GLY - THR - GLN - THR - TYR - ILE - CYS - 200  
 ASN - VAL - ASN - HIS - LYS - PRO - SER - ASN - THR - LYS - VAL - ASP - LYS - ARG - VAL - GLU - PRO - LYS - SER - CYS - 220  
 ASP - LYS - THR - HIS - THR - CYS - PRO - PRO - CYS - PRO - ALA - PRO - GLU - LEU - LEU - GLY - GLY - PRO - SER - VAL - 240  
 PHE - LEU - PHE - PRO - PRO - LYS - PRO - LYS - ASP - THR - LEU - MET - ILE - SER - ARG - THR - PRO - GLU - VAL - THR - 260  
CYS - VAL - VAL - VAL - ASP - VAL - SER - HIS - GLU - ASP - PRO - GLN - VAL - LYS - PHE - ASN - TRP - TYR - VAL - ASP - 280  
 GLY - VAL - GLN - VAL - HIS - ASN - ALA - LYS - THR - LYS - PRO - ARG - GLU - GLN - GLN - TYR - ASP - SER - THR - TYR - 300  
 ARG - VAL - VAL - SER - VAL - LEU - THR - VAL - LEU - HIS - GLN - ASN - TRP - LEU - ASP - GLY - LYS - GLU - TYR - LYS - 320  
CYS - LYS - VAL - SER - ASN - LYS - ALA - LEU - PRO - ALA - PRO - ILE - GLU - LYS - THR - ILE - SER - LYS - ALA - LYS - 340  
 GLY - GLN - PRO - ARG - GLU - PRO - GLN - VAL - TYR - THR - LEU - PRO - PRO - SER - ARG - GLU - GLU - MET - THR - LYS - 360  
 ASN - GLN - VAL - SER - LEU - THR - CYS - LEU - VAL - LYS - GLY - PHE - TYR - PRO - SER - ASP - ILE - ALA - VAL - GLU - 380  
 TRP - GLU - SER - ASN - ASP - GLY - GLU - PRO - GLU - ASN - TYR - LYS - THR - THR - PRO - PRO - VAL - LEU - ASP - SER - 400  
 ASP - GLY - SER - PHE - PHE - LEU - TYR - SER - LYS - LEU - THR - VAL - ASP - LYS - SER - ARG - TRP - GLN - GLU - GLY - 420  
 ASN - VAL - PHE - SER - CYS - SER - VAL - MET - HIS - GLU - ALA - LEU - HIS - ASN - HIS - TYR - THR - GLN - LYS - SER - 440  
 LEU - SER - LEU - SER - PRO - GLY - 446

FIG. 3.—Complete amino acid sequence of the Eu heavy chain. Half-cystinyl residues are in boxes and methionyl residues are underlined.

A comparison of the sequence of the two fragments from residues 101 to 121 is given in Figure 4. The sequences become identical at residue 115 (Eu numbering). Further studies<sup>21</sup> confirmed that from residue 115 to residue 252 the sequence of the He fragment was identical to that of Eu H<sub>1</sub>. In addition, tryptic fingerprints of the Fe fragments from Eu and He were identical. These

	101	105	110	115	120
EU	-GLY-ILE-TYR-SER-PRO-GLU-GLU-TYR-ASN-GLY-GLY-LEU-VAL-THR-VAL-SER-SER-ALA-SER-THR-LYS-				
HE	-THR-LEU-ALA-PHE-ASN-VAL-TRP-GLY-GLX-GLY-THR-LYS-VAL-ALA-VAL-SER-SER-ALA-SER-THR-LYS-				

FIG. 4.—Comparison of the amino acid sequence of the Eu heavy chain from residue 101–121 with the corresponding sequence of the heavy chain of myeloma protein He.

data suggest that the transition between  $V_H$  and  $C_H$  is located in the vicinity of residue 114 (Eu numbering). Studies on a number of additional proteins and a search for  $V_H$  region subgroups<sup>7</sup> will be required to locate this point definitively.

*Discussion.*—The present studies provide proof of the covalent structure and arrangement of chains in  $\gamma G1$  immunoglobulin. The half-molecule of Eu is the largest protein unit (446 + 214 residues) for which a complete amino acid sequence has been determined. In a protein of this size, one cannot neglect the possibility of small errors in sequence assignment; for this reason we are carrying out a number of further checks using various methods of peptide cleavage and fractionation.

One of the most striking features of the immunoglobulin molecule that emerges from the completed sequence is the sharp demarcation of its polypeptide chains into linearly connected regions that are associated with different functions. Variations in the sequences of paired  $V_H$  and  $V_L$  regions for the function of antigen binding in the selective immune response, and at the same time, conservation of sequence in  $C_H$  and  $C_L$  regions for other immunological functions appear to require special genetic and evolutionary mechanisms.<sup>22</sup>

The amino acid sequence of Eu provides convincing evidence that the immunoglobulin molecule evolved by successive duplication of precursor genes.<sup>23, 24</sup> Our analysis of a complete heavy chain has revealed an additional homology region ( $C_{H1}$ ), the structure of which was not previously known. Earlier comparisons<sup>7, 8</sup> of polynucleotide sequences corresponding to both chains of Eu showed evidence of homology between  $V_H$  and  $V_L$  and homologies among  $C_L$ ,  $C_{H2}$ , and  $C_{H3}$ . A complete comparison of the amino acid sequences of  $C_L$ ,  $C_{H1}$ ,  $C_{H2}$ , and  $C_{H3}$  is given in Figure 5. In a stretch of 100 residues, any two regions are identical in 29 to 34 positions. It is noteworthy that the stretch in the heavy chain from residue 221 to 233 which contains the interchain disulfide bonds<sup>5</sup> has no homologous counterpart in other portions of heavy or light chains.

In the data accumulated so far, little or no homology has been found between V and C regions. This prompts the speculation that V genes and C genes diverged early in the evolution of antibodies to serve two major groups of functions: antigen recognition functions (ARF) and effector functions (EF) such as interaction with cells and complement. The order of emergence of  $C_L$  genes or  $C_H$  genes from a precursor gene is not apparent from the data. A comparison with sequences of  $\mu$ ,  $\alpha$ ,  $\delta$ , or  $\epsilon$  chains may indicate similarities in their  $V_H$  regions<sup>25</sup> and may show whether any of the homology regions are conserved in constant regions of heavy chains of these classes.<sup>26</sup>

Early evolutionary divergence of V and C genes is consistent with the evidence<sup>27, 28</sup> that each chain is specified by two genes, V and C, and the hypothesis<sup>7, 22, 29</sup> that V gene episomes are translocated to C genes to form a single VC

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

## LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

## FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

## E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.