

IMMUNOGLOBULIN GENES

Paul D. Gottlieb

From the Department of Biology and Center for Cancer Research,
Massachusetts Institute of Technology, Cambridge, Massachusetts

For The New England Journal of Medicine

The production of antibodies in response to infection or vaccination is perhaps the most impressive adaptive response of higher organisms to their environment. The molecular mechanisms which underlie the great range and specificity of antibody responses have long been a subject of great interest, as has the molecular basis for inherited immunodeficiencies.

The molecular ^{structure} nature of the antibody ^{repertoire} library, the number and arrangement of the genes which encode it, and the genetic and molecular mechanisms which govern its expression are the subject matter of this review. With the advent of recombinant DNA technology and its application to the immune system, it has been possible for the first time to examine antibody genes at the DNA level, both in the germ line and in somatic cells. From comparisons of such analyses with results of genetic and structural studies of antibody molecules has emerged a consistent picture which is by no means complete, but which can account for much of the enormous range and diversity of the antibody response, and for many aspects of its regulation.

Among the issues clarified by work on immunoglobulin (Ig) genes at the nucleic acid level are:

- 1) As suggested by the "2 gene - 1 polypeptide" model¹, V and C regions for each Ig chain are encoded separately in the germ line and must be rearranged at the DNA level in somatic cells to form a functional antibody gene. However, at least two segments of DNA must be assembled to form the V-region, and even in a functional Ig gene, V and C coding sequences are not contiguous.

- 2) The germ line contains extensive V-region sequence diversity, but at least some diversity is generated by somatic diversification mechanisms;
- 3) The non-random deletions ^{of aag} in ~~so-called~~ heavy chain disease proteins and in abnormal myeloma L-chains can be explained on the basis of the structure of Ig genes;
- 4) The nucleotide sequence and arrangement of Ig genes suggest plausible mechanisms by which Ig genes are assembled and operated upon by the cell to lead to synthesis and secretion of antibody molecules.
- 5) A sequence abnormality associated with at least one apparently non-functional antibody gene segment provides an example of how mutation or deletion of Ig genetic material might lead to heritable immune deficiency.

The Antibody Molecule

Studies of immunoglobulin (Ig) structure provided the first clues concerning the nature and arrangement of antibody genes². An Ig molecule is a multichain structure consisting of two heavy (H) and two light (L) polypeptide chains held together by non-covalent forces and disulfide bonds. There exist a number of ^{The} classes and subclasses of H-chains (e.g. ^{main IgM, A, D, E} for mouse, μ , $\gamma 1$, $\gamma 2a$, $\gamma 2b$, α , ϵ , δ), and the physical and physiological activities (or effector functions) of an Ig molecule reflect the class or subclass of H chain it contains. Comparison of a number of H chains of the same class from one or several individuals indicates that they are essentially identical in sequence from approximately residue 118 through the carboxy terminus (the constant or C_H region), but that they differ extensively in sequence in their amino terminal ^(see fig) 117 residues (the variable or V_H region). Similarly, L chains of a given type have essentially identical sequences from residue 108 through the C-terminus (C_L region), but generally differ from each other in the amino terminal 107 residues (V_L region). V_H and V_L regions of an Ig molecule associate in pairs to form the antigen-binding sites. C_H regions of the H chains associate to render the Ig molecule bivalent with respect to antigen-binding sites, and to form structures which mediate the effector functions of the molecule. ^{relate to Fab & Fc diagram} Close examination of the amino acid sequence of C_H regions for internal sequence homology indicated that they are composed of tandemly-arranged units of approximately 110 amino acids which are structurally homologous to each other and to C_L³. Pairing of L with H chains and H chains with each other generates a 4-chain molecule consisting of linearly arranged domains: e.g. V_H-V_L, C_H¹ - C_L, C_H² - C_H², C_H³ - C_H³ for an Ig molecule (Figure 1). The existence of such domains

has been verified by X-ray crystallographic analysis⁴, and it is likely that each domain performs a particular function. The $C_H^1 - C_L$ and $C_H^2 - C_H^2$ domains are joined by a segment of H chain sequence known as the hinge which is extended in space and contains the cysteinyl residues involved in H-H and H-L disulfide bonds. It is likely that this region contributes to the flexibility necessary for effective bivalent binding functions.

V and C Genes are Separate But Linked

If, ^{for example,} we examine any one of the several kinds of Ig chains - e.g. the kappa (κ) light chain - in a single individual, we find a very large number of different κ chains which have the same C_κ but different V_κ sequences. If each κ chain were encoded in the germ line like a conventional protein, then there would be one gene for each κ chain that an individual can produce. That this is probably not the case was suggested by the finding in man of genetic polymorphisms involving amino acid substitutions in the C_κ region (the Inv markers)⁵. It is impossible to account for the origin and maintenance of such polymorphisms on the basis of a "one gene - one κ chain" germ line model. This and other considerations⁶ led to the suggestion that the C_κ region was encoded in the germ line by a single gene, and in order for a cell to synthesize a κ chain, one of an unknown number of V_κ genes must be joined with the C_κ gene to produce a $V_\kappa C_\kappa$ gene which can be stably inherited by daughter cells and which codes for an entire κ chain. As shown below, ^{that} direct inspection of Ig genes ^{are separated} at the DNA level has shown the ~~separateness~~ of V and C genes ~~embodied in the~~ ^{IN} "two genes - one polypeptide" model to be substantially correct for all of the types of mouse Ig chains studied to date, ^{exemplifying the 2g 1p model}

The genes encoding κ and lambda (λ) L chains are unlinked to each other and to H chain genes in rabbit and probably in man and mouse as well. However, the C_H genes for the several varieties of H chains produced by an individual, all appear to be very closely linked to each other on a single chromosome in all species studied⁶.

Evidence for genetic linkage of the V and C genes exists for several kinds of Ig chain. For example, serologically detectable amino acid sequence polymorphisms (called allotypes) have been well characterized in both the V_H and C_H regions of rabbit H chains. In breeding studies involving rabbits heterozygous for both V_H and C_H allotypes, 99% of the progeny inherited parental combinations of V_H and C_H allotypes⁷. This suggested close genetic linkage of V_H and C_H genes on a single chromosome. Moreover, in studies of antibody molecules produced by individual heterozygous rabbits, 99% of IgG molecules and >90% of IgA molecules contained V_H and C_H allotypes coded for by a single chromosome⁸. Thus the joining of V_H and C_H genes to make an intact $V_H C_H$ gene coding for a rabbit H chain appears to most often involve an intrachromosomal (genes from a single chromosome) event.

Patterns of V-Region Amino Acid Sequence Diversity

The extent and origin of V region sequence variability has been a major question in immunology. A brief examination of the amino acid sequence variability of mouse λ and κ light chains will serve to describe the nature of the variability observed. It will be apparent from comparison of this information with results of Ig gene analysis at the DNA level that a detailed understanding of the origin of antibody variability is gradually emerging.

Amino acid sequence analysis of myeloma H and L chains has provided much of our knowledge of the details of Ig V region sequence diversity. There is evidence to suggest that the picture of variability obtained from such analyses is a faithful representation of the repertoire of V regions employed in humoral immune responses⁹. In the mouse, less than 5% of serum Ig contain λ_1 chains, and this L chain type is of relatively minor importance for most mouse antibody responses. In analyzing the variability of Balb/c mouse myeloma λ_1 light chains, Cohn, Weigert and coworkers observed a very simple pattern of variability which suggested that the mouse λ_1 V region is encoded in the germ line by a single gene. Twelve of 18 V_{λ_1} regions studied ^{of the myelomas} had identical amino acid sequences (called the λ_0 sequence) which corresponds to the sequence of the germ line gene (see below). The remaining V_{λ_1} sequences were substantially the same, but differed from λ_0 and from each other by one or two amino acids in the so-called hypervariable or complementary determining regions (CDR)¹⁰.

^{6 mouse L cell} V_{λ_1} , V_K and V_H regions ^{hypervariable or complementary (CDRs)} each have three CDRs (Figure 1) in which comparison with other chains of the same type reveals the largest amount of variability¹¹. X-ray crystallographic analysis ^{on it} (of intact Ig and Fab fragments) has shown that the CDRs of both L and H chains generally lie in the antigen-binding site of the molecule¹². Thus the great variability in the CDRs of both H and L chains directly contributes to the enormous library of antigen-binding sites so important to a selective immune response. The non-CDR portions of the V region are called the framework, and in the case of V_{λ_1} regions there appears to be only one framework sequence shared by all λ_1 chains studied to date.^{9, 10}

Understanding the variability of $V_{\lambda 1}$ regions thus amounts to understanding the origin of the CDR amino acid sequence diversity. Many of the amino acid sequence substitutions observed in the $V_{\lambda 1}$ CDRs could be accounted for by single base substitutions in the DNA sequence for the λ_0 V gene. However, several substitutions required two or three base substitutions. Somatic mutation, either spontaneous or directed, is viewed by many as the most likely origin of the CDR amino acid sequence variability of mouse λ_1 chains, but the question of whether the $V_{\lambda 1}$ diversity observed reflects physiological diversity or is peculiar to the repeatedly passaged myelomas is not yet settled.

The amino acid sequence variability of mouse κ L chains is more extensive, and more nearly approximates that observed in human H-chains and κ and λ L-chains. Comparison of approximately two hundred whole or partial Balb/c mouse myeloma V_{κ} sequences indicates sequence variability throughout most of the V region. However, when the number of different amino acids observed at each position of the V region are enumerated, the most variable positions again correspond to the CDRs (Figure 1). Moreover, the V_{κ} sequences can be arranged into groups on the basis of similarity and in some cases, identity, of framework sequences.^{11, 13, 14} While somewhat arbitrary criteria for framework sequence similarity have often been employed (e.g. V_{κ} regions differing at more than 3 of the first 23 amino acids belong to different V_{κ} groups), it nonetheless appears that the mouse V_{κ} repertoire can be accounted for by approximately fifty such groups. Proteins from each group can be further divided into subgroups on the basis of more extensive sequence analysis of both framework and CDR regions,^{11, 13, 14} and the present view, supported by analysis at the DNA level (see below), is that each V_{κ} subgroup is determined

by a single germ line gene, similar to the situation seen for mouse $V_{\lambda 1}$ regions, and that a group is encoded by ~6-10 germ line genes. The possible origin of additional non-germ line CDR variability will be discussed below.

While far less data is available for V_H region diversity, estimates of at least 20 V_H groups have been obtained in the mouse.¹⁵

Analysis of Ig Genes at the DNA Level

With the above discussion of Ig structure, sequence diversity and gene organization based on genetic and protein structure analysis as an incomplete but adequate background, we shall now turn to the analysis of Ig genes at the DNA level. As in the analysis of amino acid sequence diversity, myelomas have provided the key to the study of Ig genes at the DNA level. Messenger RNA molecules (mRNA) which direct the synthesis of the H and L chains characteristic of each myeloma are present in sufficient quantity intracellularly to allow their isolation.¹⁶ By virtue of their sequence complementarity to DNA coding for the Ig chain, such purified mRNAs have served as probes to identify and isolate the Ig genes themselves. Hozumi and coworkers¹⁷ employed radiolabelled mouse κ mRNA to examine the arrangement of Ig genes in embryonic and myeloma DNA to test the "2 gene - 1 polypeptide" hypothesis - i.e. that V and C genes encoded separately in germ line DNA are joined at the DNA level in an antibody-producing cell to form a gene coding for an entire Ig chain. Their studies made use of a restriction endonuclease,¹⁸ one of a library of bacterial enzymes, each of which cleaves DNA specifically at a different characteristic short nucleotide sequence and thereby reduces high molecular weight cellular DNA to a complex mixture of fragments (called restriction fragments). Hozumi

et al.¹⁷ cleaved mouse embryonic DNA with a particular restriction endonuclease and showed that DNA complementary to the portions of a myeloma κ mRNA corresponding to the V and C regions appeared in separate DNA fragments. In contrast, when DNA from the κ -producing myeloma was cleaved with the same restriction endonuclease, a single restriction fragment contained DNA complementary to both V and C portions of the κ mRNA. Thus, at some stage of differentiation of a κ -producing cell, rearrangement at the DNA level brings separately encoded V and C genes into close proximity to create a functional VC gene for a κ chain. A similar process occurs for λ_1 ¹⁹ chains and probably for H chains.^{20, 21}

Gene Cloning

Analysis of the nucleotide sequence and fine structure of Ig genes has been made possible by application of recombinant DNA technology. Restriction fragments of DNA (see above) are incorporated into a specially modified bacteriophage^{22, 23} or plasmid²⁴ (a vector) and grown in bacteria so as to allow isolation (cloning) of a single pure vector containing the gene of interest. This can then be produced in large quantity for DNA sequence or other analysis. The cloning of Ig genetic material has been accomplished by two major routes. The enzyme reverse transcriptase (employed by RNA viruses to synthesize a DNA copy of themselves for insertion into a host chromosome) can be employed in vitro to synthesize DNA complementary to an isolated mRNA (called cDNA).²⁵ The cDNA copy can be incorporated into a suitable vector and produced in quantity for DNA sequence or other analysis,²⁶ or used as a probe for isolating genomic Ig DNA (the second major route). In one form of this latter approach,^{27, 28} sperm, embryonic, myeloma or somatic cell DNA is digested with a restriction

enzyme (see above), resulting fragments are subjected to electrophoresis and/or chromatography, and fractions are tested for DNA complementary to radiolabelled Ig cDNA or mRNA probes. Fractions containing complementary DNA are still highly impure, but can be incorporated into vectors and used for infection at sufficient dilution such that each colony of bacteria (if the vector is a plasmid) or each plaque (if the vector is a phage) harbors only one vector containing a single foreign DNA restriction fragment. Re-testing with radiolabelled probe allows identification of colonies or plaques containing an Ig gene segment, and that vector can then be recovered, recloned to be sure it is pure, and produced in quantity for detailed structural analysis.

Mouse Embryonic and Myeloma λ_1 Genes - A Simple System

Using the latter approach Tonegawa and coworkers cloned separate fragments containing $V_{\lambda 1}$ and $C_{\lambda 1}$ from mouse embryo DNA¹⁹ and a single fragment containing both $V_{\lambda 1}$ and $C_{\lambda 1}$ from a λ_1 -producing myeloma.²⁹ Embryonic mouse DNA appears to contain only one $V_{\lambda 1}$ gene,¹⁹ and this and the limited amino acid sequence diversity seen in myeloma λ_1 chains^{9, 10} has made the mouse λ_1 system the simplest for studying the rearrangement of antibody genes.

The structures of $V_{\lambda 1}$ - and $C_{\lambda 1}$ -containing embryonic fragments and of the myeloma λ_1 gene are diagrammed in Figure 2. Four findings are especially noteworthy:

1. DNA coding for most of the hydrophobic leader sequence (important in biosynthesis and secretion of Ig)³⁰ is separated from the rest of the leader and from the amino terminus of the V region by 93 bases of non-coding sequence;³¹

2. Although the length of the V region as determined from protein sequence analysis is 110 amino acids, the embryonic $V_{\lambda 1}$ gene only extends to residue 97;³¹

3. The embryonic DNA coding for the remainder of the V region (called the J segment) is present in the same restriction fragment as the $C_{\lambda 1}$ gene, but separated from it by approximately 1200 bases (1.2 kilobases);^{29, 31}

4. DNA rearrangement is apparent in myeloma DNA such that V and J segments are contiguous, but a complete VC gene with coding sequence running continuously from amino terminus to carboxy-terminus of the λ_1 chain does not exist at the DNA level. Even in the myeloma DNA, $V_{\lambda 1}$ and $C_{\lambda 1}$ regions lie 1.2 kb apart and the continuity of the hydrophobic leader sequence is interrupted by a 93 base intervening sequence.^{29, 31}

Thus in agreement with the "2 gene - 1 polypeptide" hypothesis,¹ V and C sequences are encoded separately in the germ line and are rearranged to form an active VC gene in antibody producing cells. However, the coding regions (called exons) of the resulting mature λ_1 gene are interrupted by non-coding intervening sequences (called introns),³² and the gene is referred to as a "split gene".

"Split genes" of the type described for the mouse λ_1 myeloma gene and for other Ig genes to be discussed below are not unique to the immunoglobulins, but appear to be common in the genomes of eukaryotes and their viruses.^{33, 34} The functional advantages of introns within genes are not known, although several theoretical suggestions have been offered.^{19, 32, 34, 45} However, their existence places constraints upon the expression of such genes.

In order for an intact λ_1 chain to be synthesized by the λ_1 myeloma tumor described above, a mRNA in which the coding regions for hydrophobic leader, V_L and C_L are contiguous must first be made. This appears to involve transcription of the entire gene, including the non-coding sequences, and processing or splicing this primary transcript while still in the nucleus to remove the introns and generate an active mRNA with all coding regions contiguous. We shall discuss splicing in more detail when we come to κ and H-chain genes.

Mouse κ L-Chains - A Complex System

The mouse λ_1 system has served well to describe some of the basic structural features of Ig genes and the rearrangements required for their expression. Unlike the λ_1 system with its limited amino acid sequence diversity generated from what appear to be single germ line V and J genes, the enormous amino acid sequence diversity of mouse κ chains is more representative of that found in human Ig chains. Multiple germ line genes had been suspected because the amino acid sequence and length differences which distinguish V regions of different groups of κ myeloma proteins were too extensive for any somatic diversifying mechanism to generate the different groups from a single germ line gene.³⁶ With the advent of gene cloning, Seidman and coworkers³⁷ demonstrated that each of two radiolabelled probes representing two different V_κ groups detected a different set of 5 - 10 DNA fragments when hybridized with restriction fragments of myeloma DNA which had been fractionated by chromatography and gel electrophoresis (see above). Thus different groups were indeed determined by different genes, and several genes appeared to specify each group.

Seidman and coworkers cloned complementary DNA from two fractions which hybridized with one of their V_{κ} probes.²⁷ Nucleotide sequence analysis revealed that the two V_{κ} regions were very similar in sequence to each other and to the V_{κ} chain secreted by the myeloma, but neither was identical to the secreted L chain. Most of the sequence differences were in the CDR portions of the V_{κ} region. Moreover, when the two clones were mixed, rendered single-stranded by denaturation, and then renatured and examined in the electron microscope for ability to renature with each other (a technique called heteroduplex analysis),³⁸ they were found to do so, indicating extensive (but not complete) sequence homology over a length of at least 3000 nucleotides which included flanking as well as coding sequences.³⁷

Analysis of mouse V_{κ} sequences at the protein level has suggested a minimum of 50 groups necessary to account for the framework diversity observed in myeloma κ chains,¹⁰ and a model of the mouse V_{κ} locus which depicts features described above and below is presented in Figure 3. If each V_{κ} group is determined in the germ line by approximately 6 genes as the results described above suggest,³⁷ then a minimum of about 300 V_{κ} genes would be required. However at the protein level, even within a group, the extent of amino acid sequence diversity in the CDRs is too great to be explained simply on the basis of six germ line genes per group.^{13, 14} Such extensive CDR sequence diversity can be accounted for if the member genes within a subgroup were permitted to pair with each other in the somatic cells. Such pairing could result in genetic recombination or in gene conversion, both of which could generate new sequences not present in the germ line.³⁹ Pairing among members of a group would be favored by the extensive sequence homology flanking the coding regions³⁷ (Figure 3), and since differences among the member genes are most concentrated in the CDRs, most of the new sequence diversity introduced would also lie in the CDRs.

Amino acid sequence analysis of myelomas suggests that recombination within framework segments of members of different V_K groups is rare or non-existent. Seidman and coworkers³⁷ have suggested that the sequences flanking the coding regions of different groups are non-homologous, and this might prevent mispairing and therefore improper recombination between different groups. This would serve as a safety device. Non-homologous crossing over which can occur following mispairing in the germ line can lead to gene loss or gene amplification. The constraint that such mispairing can only occur between member genes within a group would allow limited sequence diversification by this mechanism, but would protect the species from catastrophic gene loss of the entire κ gene system.³⁷ Only a single V_K group would be at jeopardy in any mispairing event. It has been suggested that the limited sequence diversity of the mouse λ_1 system reflects such a catastrophic gene loss.³⁷

In contrast to the above mechanisms, Kabat and coworkers have suggested that the V-region sequence diversity of an Ig chain is generated by assembly during embryogenesis of "minigenes" which are separate in the germ line and which code for each of the three CDRs and each of the four framework segments which surround the CDRs⁴⁰ (Figure 1). This hypothesis was prompted by results of computer analyses of myeloma protein amino acid sequences. These revealed, for example, that some proteins had identical frameworks preceding the first CDR but different frameworks between the first and second CDR, while in others the positions of identical and different frameworks were reversed. This implied assembly of V genes from smaller independent units. Recently, computer studies of DNA sequences of several cloned L-chain genes have revealed

unusual shared sequences in the framework segments flanking the CDRs.⁴¹ Wu and coworkers suggest that these may serve as recognition sites in the assembly of minigenes or in a somatic mutation mechanism which generates CDR sequence diversity.⁴¹ Testing the validity of the minigene hypothesis will require detailed comparison of Ig gene organization in germ line and somatic DNA.

As with the $V_{\lambda 1}$ genes, embryonic V_{κ} genes only code through approximately residue 97 of the protein V_{κ} region, and joining with a separately-encoded J_{κ} DNA segment is required to form a complete κ V-region.^{27, 35} It had been noted from studies of human and mouse V_{κ} amino acid sequences that group distinctions were no longer possible towards the end of the V region,⁴² and the existence of several J segments with which members of any V_{κ} group could join to generate an intact V coding region provides a clear explanation for this observation. Studies by Sakano et al.³⁵ and by Max et al.⁴³ have demonstrated the presence of five J_{κ} segments, separated from each other by approximately 300 bases, in the DNA 2.5 - 4.0 kilobases preceding the C_{κ} gene (Figure 3). DNA sequence analysis has shown that 4 of the 5 J_{κ} segments code for amino acid sequences that are observed in myeloma κ chains and are not restricted to any one V_{κ} group. The sequence of the fifth J-like segment (J3) was unlike any ever observed in myeloma proteins, and we shall return to this later.

Although no additional J segments could be found in the vicinity of C_{κ} ,^{35, 43} inspection of myeloma κ sequences indicated more variability in the region corresponding to the V-J junction (which lies within the third CDR)¹³ than could be accounted for simply by faithful V-J joining. The

possibility was suggested that the act of V-J joining at the DNA level might generate additional sequence diversity.^{35, 43} Inspection of nucleotide sequences following an embryonic V_K gene and preceding each of the five J_K segments has suggested a plausible mechanism by which this might occur. (Figure 4). Closely following the last codon (95) of all three V_K genes studied thus far is a DNA sequence which represents a nearly perfect palindrome - i.e. reading one DNA strand from left to right for 7 residues is nearly identical to reading the complementary strand from right to left (Figure 4a). The only imperfection is caused by an AT base pair at the center of symmetry. Closely preceding each of the five J_K is the identical near-perfect palindrome but the orientation of the entire double stranded segment is inverted 180° (Figure 4a). In many cases palindromes have been shown to act as recognition sites for DNA modifying enzymes (e.g. restriction enzymes)¹⁸ or other DNA-binding proteins (e.g. λ repressor).⁴⁴ The orientations of the palindromes following V_K and preceding J_K are such that a structure of the type shown in Figure 4b might be formed by pairing of complementary palindromic sequences within each DNA strand rather than between strands.^{35, 43} In the structure formed (called an inverted repeat stem structure) V_K and J_K are brought in close proximity, and enzymes which catalyze V-J joining might then operate, resulting in deletion from the chromosome of the stem structure and all of the DNA lying between the V and J segments joined.^{35, 37, 43} In order for the proposed stem structure to be formed, some DNA duplex would have to melt, at least locally, and V and J segments separated by thousands of nucleotides must come close together so that the intra-chain duplex can form. The forces which might bring this about and the time in development at which it might occur are unknown. However, similar inverted repeats are known to exist

at the ends of bacterial transposable elements⁴⁵ which can form such stem structures and which appear to undergo deletion from the bacterial chromosome by the above mechanism. Sakano et al.³⁵ have made the provocative suggestion that the ancestral event which permitted evolution of Ig V regions encoded by separate V and J genes involved a DNA segment analogous to bacterial transposable elements. Thus far in the analysis of the genome of higher organisms, the Ig genes have been the only ones shown to require DNA rearrangement for expression.

The joining of V and J by the above mechanism can also explain the exceptional variability at position 96 of mouse κ chains.^{15, 35, 43} As shown in Figure 4c, cutting and rejoining of DNA strands at slightly different positions to form contiguous VJ segment could result in several different amino acids at position 96. Several κ chains whose J peptide sequences cannot be accounted for by embryonic J_{κ} segments can be explained by recombination events 3 and 4 shown in Figure 4c. Nucleotide sequence analysis in the region of mouse $V_{\lambda 1}$ and $J_{\lambda 1}$ genes suggests a similar mechanism for V-J joining.³⁵ Comparable analyses have not been performed with H chains, but protein sequence results of Schilling et al.²⁰ suggest the existence of several J_H segments, and a large amount of sequence and length variability at the V-J junction. The origin of this variability is not known, but Hood has suggested that the mechanism described in Figures 4b and 4c may not account completely for junctional sequence diversity in the H-chain system, and that some diversity at the V_H - J_H junction may be provided by an as yet unidentified DNA segment which he calls the D segment.²¹

Splicing and Possible Sources of Immunodeficiency

Once a V_K-J_K joining event has occurred in one of a pair of homologous chromosomes, that cell and its descendants are thought to be committed to produce only that V_K region. Transcription of the resulting κ gene does not, however, directly yield a messenger RNA (mRNA) which can be translated to produce that κ chain. As shown in Figure 2, introns must first be spliced out of the primary transcript, just as introns in RNA transcripts in globin, ovalbumin, SV40, human adenovirus, and many other genes must be removed to make all of the coding regions contiguous.^{33, 34} While the mechanisms by which introns are spliced out are not known, it is likely that the nucleotide sequences at the ends of each intron adjacent to the coding regions provide at least part of the recognition site for the enzyme(s) which accomplish(es) splicing. In nearly all of the introns studied to date, it is possible to choose cutting points (i.e. the boundaries of coding region and intron) such that an intron begins with the sequence GT and ends with AG.⁴⁵ The nucleotide sequences following the four J segments whose sequences appear in myeloma proteins also begin with GT.^{35, 43} However, the intron following the fifth J_K (J_3 , Figure 3) does not begin with GT, and it may be significant that this is the only one of the five J_K segments for which a protein sequence has not been observed. If the initial GT sequence is required for splicing to occur, then the failure to observe the sequence of J_3 in a protein may reflect the failure of primary RNA transcripts of V-J genes involving that J to mature to functional mRNA's. If this were so, then it would constitute an inherited immunodeficiency in which the Balb/c mouse would be unable to produce certain antibodies. Whether this places the Balb mouse at some disadvantage is not known, but it represents an inherited defect at the level of antibody genes,

and serves as an example of the type of inherited lesion which might affect Ig genes of man as well. Similarly, absence of the palindrome sequences following certain V genes or preceding J genes due to mutation or deletion might lead to failure of that V (or J) to contribute to an active Ig gene, and possibly leave the organism at a disadvantage. Considering the large number of V genes observed it would be surprising if loss of expression of one V gene or V gene group could seriously affect immune responsiveness, but the existence of immune response genes which appear to map to Ig V loci suggests that this may be the case.⁹

H-Chain Genes and a Molecular Explanation for Heavy Chain Disease

Whereas to date most of the analysis of Ig genes at the DNA level has focussed on L-chains, analysis of the heavy chain gene complex promises to answer many of the outstanding questions concerning regulation of the antibody response. Sakano and coworkers⁴⁶ have cloned a DNA segment from a IgG1-producing mouse plasmacytoma which codes for the C_H1 region. Results indicate that the mouse C_H1 region is encoded by four non-contiguous DNA segments separated by three introns (Figure 5). The four coding segments correspond to the sequences of the C_H1, hinge, C_H2, and C_H3 regions as designated on the basis of amino acid sequence³ and of x-ray crystallographic analysis.⁴ The nucleotide sequences at the ends of the introns separating the hinge and C_H2 and C_H2 and C_H3 coding segments are similar to the putative splice sites of L-chain introns and those of other genes.^{35, 43, 46} Thus the contiguity of these coding segments in the mRNA undoubtedly results from splicing of a primary RNA transcript. Surprisingly, these generally conserved splice-related nucleotide sequences are not present in the intron separating

C_H1 from the hinge, and it is not known whether this reflects another mechanism for achieving contiguity of these segments, or whether there is some other explanation.

Studies of a cloned myeloma α chain gene by Early et al.⁴⁷ have suggested that the exons which encode C_H1 , C_H2 and C_H3 segments of the α chain are also separated by introns. These studies have not yet been carried to the sequence level, and it is not known whether the α -chain hinge constitutes an separate exon.

The findings that H-chains are encoded by "split genes" and the separate coding regions of the $C_{\gamma1}$ gene correspond to the several C_H domains and hinge provide a likely explanation for the non-random deletions observed in so-called heavy chain disease proteins.⁴⁸ These are myeloma H-chains of several different classes and subclasses which have been shown to contain internal deletions by amino acid sequence analysis. Two common deletions involve absence of part of the V_H region and the entire C_H1 region, and deletion of the hinge region only. The non-random end points of these deletions (see Figure 5) might be explained by sequence alterations at the DNA level which interfere with the generation of a normal mRNA by splicing. Mutation might cause RNA coding for one or more domains to be spliced out of the primary RNA transcript, resulting in a short mRNA from which the aberrant protein is synthesized. In no case has the gene coding for a heavy chain disease protein been cloned and sequenced, and it is possible that aberrant DNA rearrangements during formation of an active gene might give rise to a heavy chain disease protein.

Similar mechanisms could account for aberrant L-chains which lack a portion of V but resume at J (SAC), or lack a portion of V and the entire J, resulting at the beginning of C (SM).⁴⁸

A Possible Mechanism for the H-Chain Switch

Whereas different sets of V and J genes are employed for κ and γ chains, results of amino acid sequence, serological and genetic analysis have suggested that all H-chain classes and subclasses share the same set of V-regions.³⁶ The various H-chain constant regions are encoded by closely-linked genes, and as stated above, serological and other data in rabbits and mice suggests that V_H and C_H gene loci are linked to each other.^{8, 9}

Preliminary studies of the H-chain V and C region loci at the DNA level have suggested a possible explanation for an important aspect of immunoglobulin expression known as the switch.⁴⁹ Developmentally, the IgM class of antibody appears to be the first one expressed by B lymphocytes at the cell surface. Somewhat later B cells appear which express both IgM and IgD at the cell surface and descendants of these cells are generated which produce yet other H-chain classes or subclasses.⁴⁹ It is widely held that the V_H expressed in the IgG or IgA produced by such descendants is the same as that present on the surface IgM of the progenitor B-cells.⁵⁰

Preliminary analysis of the introns preceding C_μ and C_α genes in μ - and α -producing myelomas, respectively, has suggested a model by which the switch might be accomplished. Evidence from Hood and coworkers²¹ suggests that a 5 kilobase segment of the intron which precedes C_μ in the germ line and in a μ -producing myeloma is found preceding the C_α gene in an α -producing myeloma.

It is possible that the intron which separates the J segments from C μ in both the germ line and in cells committed to μ -chain production contains sites involved in switching the VJ segment of an active μ -chain gene from C μ to another C $_H$ gene in response to an appropriate signal (Figure 6).

For example, if a switch sequence (S) preceding C μ pairs with an inverted switch sequence (S') preceding another C $_H$ gene to form an inverted stem structure of the type shown in Figure 4, action of appropriate recombinational enzymes could result in deletion of all DNA between the sites. The VJ segment and portions of the intron preceding the S site in the active μ gene would then be contiguous with intron DNA which preceded the newly active C $_H$ gene in the germ line (Figure 6). The deletion of C μ and other C $_H$ genes upon switching is consistent with data of Honjo and Kataoka⁵¹ who have suggested an order for C $_H$ genes on the basis of the C $_H$ genes detectable in the genomes of mouse myelomas expressing different classes and subclasses of immunoglobulins. In the model presented in Figure 6, the switch sites which precede the S $_{\gamma 1}$ site might still be functional, and it is possible that with an appropriate signal, another switch might occur with deletion of the previous active C $_H$. However, if switch (S) sites were distributed as shown in the inset in Figure 6, then only a single switching event from μ to another H-chain class would be possible. Signals which might determine whether the switch will be to one C $_H$ gene or another may involve the nature of the antigen, the anatomical site of the stimulus, or the mode of presentation of the antigen. It should be emphasized that the model shown in Figure 6 has not been proved, but it is consistent with current knowledge of the structure and function of the H-chain locus and provides a useful working hypothesis.

Conclusion

Our present view of the structure and function of antibody genes represents the synthesis of results from three major experimental approaches:

- 1) the serology and genetics of antibodies;
- 2) protein chemical and x-ray crystallographic analysis of antibodies;
- 3) gene cloning followed by heteroduplex and nucleic acid sequence analysis of antibody genes.

Analysis of human antibody genes at the DNA level has not yet been performed, but the great similarity between the human system and experimental animal models seen in approaches 1) and 2) inspire the confidence that gene structures and mechanisms at the DNA level will be analogous.

It is apparent that the extensive amino acid sequence diversity of an individual's antibody repertoire results from the interplay of germ line content and somatic mechanisms. The collection of separate V and J segments inherited from both parents become rearranged in the somatic cells to form active (though split) antibody genes. In general, this rearrangement involves V and J from the same chromosome, and it is thought to commit a cell to production of one H and one L chain. V region sequence diversity comes from 1) initial germ line content; 2) possibly generation of additional intra-V diversity by recombination or gene conversion among members of a V-gene group due to mispairing which is favored by extensive homology of both coding and flanking sequences. However, the possibilities that CDR diversity reflects insertion of "minigenes" into the V-region framework or is the result of

somatic mutation^{40, 41} (either mechanism could explain results in the mouse λ_1 system) must be considered; 3) DNA joining of (presumably) any V to any linked J segment, with or without generating additional V-J junctional diversity in the third CDR (Figure 4). Finally, if the commitment of a cell to synthesis of one H-and one L chain is random as is supposed, an individual's antibody repertoire should be the product of the number of different H and different L chains it can make (H times L).³⁶

Knowledge of the detailed structure and arrangement of Ig genes suggests points at which heritable mutations might bring about antibody gene defects which could lead to immunodeficiencies. Germ line libraries of V and/or J sequences undoubtedly differ in the human population, and adventitious catastrophic gene loss of the kind suggested for the mouse λ_1 system³⁷ could profoundly reduce antibody variability. Moreover, the apparently inoperative J3 segment in the Balb/c mouse, possibly reflecting inability to undergo effective splicing due to what may be a faulty recognition site,^{35, 43} is an example of a molecular lesion which could prevent expression of any V gene joined to J3. Since there appear to be a total of five J_κ segments, this could result in a 20% depression of the κ chain diversity of that mouse. If V-J3 joining occurs, 20% of the B cell population may also be non-functional. Similar defects within a C_H gene donated by one or both parents could lead to reduced expression or total absence of that H-chain class. Similarly, defective intron sequences involved in V-J joining (e.g. the palindrome in Figure 4), could preclude effective joining of some V's and/or J's and thereby reduce variability. Finally, defects at any level of the enzymatic machinery needed to rearrange antibody gene DNA or splice RNA transcripts to produce functional

mRNAs could lead to partial or total absence of functional antibodies. It is not known whether any significant immunodeficiency results from mechanisms of the kind described here, but further elucidation of the genetics, structure and function of antibody genes should provide answers to this and other questions concerning immune function.

References

1. Hood LE: Two genes, one polypeptide chain - fact or fiction? Fed Proc 31:177-187, 1972
2. Edelman GM, Gall, WE: The antibody problem. Annu Rev Biochem 38:415-466, 1969
3. Edelman GM: The covalent structure of a human γ G-immunoglobulin. XI. Functional implications. Biochemistry 9:3197-3205, 1970
4. Davies DR, Padlan EA, Segal DM: Three-dimensional structure of immunoglobulins. Annu Rev Biochem 44:639-667, 1975
5. Terry WD, Hood LE, Steinberg AG: Genetics of Ig κ chains: chemical analysis of normal human light chains of differing InV types. Proc Natl Acad Sci USA 63:71-77, 1969
6. Edelman GM, Gally JA: The genetic control of immunoglobulin synthesis. Annu Rev Genet 6:1-46, 1972
7. Mage R, Lieberman R, Potter M, Terry WD: Immunoglobulin allotypes. in The Antigens Vol. I (Sela M editor) Academic Press, NY pp. 299-376, 1973
8. Knight KL, Malek TR, Hanly WC: Recombinant rabbit secretory immunoglobulin molecules: alpha chains with maternal (paternal) variable-region allotypes and paternal (maternal) constant-region allotypes. Proc Natl Acad Sci USA 71:1169-1173, 1974
9. Weigert M, Riblet, R: The genetic control of antibody variable regions in the mouse. Springer Seminars in Immunopathology 1:133-169, 1978

10. Weigert M, Riblet R: Genetic control of antibody variable regions.
Cold Spr Hrb Symp Quant Biol 41:837-846, 1976
11. Wu TT, Kabat EA: An analysis of the sequences of the variable regions
of Bence-Jones proteins and myeloma light chains and their implications
for antibody complementarity. J Exp Med 132:211-250, 1970
12. Amzel LM, Poljak RJ, Saul F, Varga JM, Richards FF: The three dimensional
structure of a combining region-ligand complex of immunoglobulin NEW at
3.5Å resolution. Proc Natl Acad Sci USA 71:1427-1430, 1974
13. Weigert M, Gatmaitan L, Loh E, Schilling J, Hood L: Rearrangement of genetic
information may produce immunoglobulin diversity. Nature 276:785-790, 1978
14. McKean DJ, Bell M, Potter M: Mechanisms of antibody diversity: multiple
genes encode structurally related mouse κ variable regions. Proc Natl Acad
Sci USA 75:3913-3917, 1978
15. Barstad P, Hubert J, Hunkapiller M, Goetze A, Schilling J, Black B, Eaton B,
Richards J, Weigert M, Hood L: Immunoglobulins with hapten-binding activity:
structure-function correlations and genetic implications. Eur J Immunol
8:497-503, 1978
16. Honjo T, Swan D, Nau M, Norman B, Packman S, Polsky F, Leder P: Purification
and translation of an immunoglobulin λ chain messenger RNA from mouse myeloma.
Biochem 15:2775-2779, 1976
17. Hozumi N, Tonegawa S: Evidence for somatic rearrangement of immunoglobulin
genes coding for variable and constant regions. Proc Natl Acad Sci USA 73:
3629-3632, 1976

18. Roberts RJ: Restriction endonucleases. Crit Rev Biochem 4:123-164, 1976
19. Brack C, Hiram M, Lenhard-Schuller R, Tonegawa S: A complete immunoglobulin gene is created by somatic recombination. Cell 15:1-14, 1978
20. Schilling J, Clevinger B, Davie J, Hood L: Amino-acid sequence analyses of homogeneous antibodies to dextran: diversity patterns suggest that DNA rearrangements between V and J segments occur in heavy chains. Nature (submitted)
21. Hood LE: Personal communication
22. Blattner FR, Williams B, Blechl AE, Denniston-Thompson K, Faber HE, Furlong LA, Grunwald DJ, Kiefer DO, Moore DD, Schumm JW, Sheldon EL, Smithies O: Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science 196:161-169, 1977
23. Leder P, Tiemeier D, Enquist L: EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λ gt WES system. Science 196:175-177, 1977
24. Collins J, Hohn B: Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage λ heads. Proc Natl Acad Sci USA 75:4242-4246, 1978
25. Efstratiadis A, Maniatis T, Kafatos FC, Jeffrey A, Vournakis JN: Full length and discrete partial reverse transcripts of globin and chorion mRNAs. Cell 4:367-378, 1975

26. Maniatis T, Kee SG, Efstratiadis A, Kafatos FC: Amplification and characterization of a β -globin gene synthesized in vitro. Cell 8: 163-182, 1976
27. Seidman JG, Leder A, Edgell MH, Polsky F, Tilghman SM, Tiemeier D, Leder P: Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. Proc Natl Acad Sci USA 75:3881-3885, 1978
28. Lenhard-Schuller R, Hohn B, Brack C, Hiram M, Tonegawa S: DNA clones containing mouse immunoglobulin κ chain genes isolated by in vitro packaging into phage λ coats. Proc Natl Acad Sci USA 75:4709-4713, 1978
29. Brack C, Tonegawa S: Variable and constant parts of the immunoglobulin light chain of a mouse myeloma cell are 1250 nontranslated bases apart. Proc Natl Acad Sci USA 74:5652-5656, 1977
30. Burstein Y, Schechter I: Amino acid sequence of the NH_2 -terminal extra piece segments of the precursors of mouse immunoglobulin λ_1 -type and κ -type light chains. Proc Natl Acad Sci USA 74:716-720, 1977
31. Bernard O, Hozumi N, Tonegawa S: Sequences of mouse immunoglobulin light chains before and after somatic changes. Cell 15:1133-1144, 1978
32. Gilbert W: Why genes in pieces? Nature 271:501, 1978
33. Berget SM, Moore C, Sharp PA: Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA 74:3171-3175, 1977
34. Darnell JE Jr: Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science 202:1257-1260, 1978

35. Sakano H, Hüppi K, Heinrich G, Tonegawa S: Sequences at somatic recombination sites of immunoglobulin light chain genes: Their implication for the mode of recombination, antibody diversity and evolution. *Nature* 280:288-294, 1979
36. Gally JA, Edelman GM: Somatic translocation of antibody genes. *Nature* 227:341-350, 1970
37. Seidman JG, Leder A, Mau M, Norman B, Leder P: Antibody diversity. *Science* 202:11-17, 1978
38. Ferguson J, Davis RW: Quantitative electron microscopy of nucleic acids. in *Advanced Techniques in Biological Electron Microscopy II*. (Koehler JK, editor) Springer-Verlag, Berlin, 1978 pp. 123-171
39. Edelman GM, Gally JA: Antibody structure, diversity and specificity. *Brookhaven Symposia in Biology* 21:328-343, 1969
40. Kabat EA, Wu TT, Bilofsky H: Variable region genes for the immunoglobulin framework are assembled from small segments of DNA - a hypothesis. *Proc Natl Acad Sci USA* 75:2429-2433, 1978
41. Wu TT, Kabat EA, Bilofsky H: Some sequence similarities among cloned mouse DNA segments that code for λ and κ light chains of immunoglobulins. *Proc Natl Acad Sci USA* 76, in press
42. Smith GP, Hood L, Fitch WM: Antibody diversity. *Annu Rev Biochem* 40: 969-1012, 1971
43. Max EE, Seidman JG, Leder P: Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. *Proc Natl Acad Sci USA* 76:3450-3454, 1979

44. Ptashne M, Backman K, Humayun MZ, Jeffrey A, Maurer R, Meyer B, Sauer RT: Autoregulation and function of a repressor in bacteriophage lambda. *Science* 194:156-161, 1976
45. Cohen SN: Transposable genetic elements and plasmid evolution. *Nature* 263:731-738, 1976
46. Sakano H, Rogers JH, Hüppi K, Brack C, Traunecker A, Maki R, Wall R, Tonegawa S: Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. *Nature* 277:627-633, 1979
47. Early PW, Davis MM, Kaback DB, Davidson N, Hood L: Immunoglobulin heavy chain organization in mice: Analysis of a myeloma genomic clone containing variable and alpha constant regions. *Proc Natl Acad Sci USA* 76:857-861, 1979
48. Franklin EC, Prelli F, Frangione B: Human heavy chain disease protein WIS: Implications for the organization of immunoglobulin genes. *Proc Natl Acad Sci USA* 76:452-456, 1978
49. Vitetta ES, Cambier J, Forman J, Kettman JR, Yuan D, Uhr JW: Immunoglobulin receptors on murine B lymphocytes. *Cold Spr Hrb Symp Quant Biol* 41:185-191, 1976
50. Sledge C, Fair DS, Black B, Krueger RG, Hood L: Antibody differentiation. Apparent sequence identity between variable regions shared by IgA and IgG immunoglobulins. *Proc Natl Acad Sci USA* 73:923-927, 1976
51. Honjo T, Kataoka T: Organization of immunoglobulin heavy chain genes and allelic deletion model. *Proc Natl Acad Sci USA* 75:2140-2144, 1978

ACKNOWLEDGMENT

The author is grateful to Drs. L. Hood, J.M. Davie and E.A. Kabat for discussions and/or access to data prior to publication, to Ms. Lois Hill for typing the manuscript, and to Ms. Francia Linden for graphic art work.

FIGURE LEGENDS

1. Diagram of an immunoglobulin molecule. CDR = complementarity determining region.

2. Rearrangement of embryonic DNA and splicing of RNA lead to production of a mature mouse λ_1 L chain messenger RNA. $L_{\lambda 1}$ = hydrophobic leader;
 — = DNA; ~~~ = RNA; ● = DNA rearrangement site; ○ = RNA splicing site.

3. Suggested model of the mouse κ L chain locus. Double stranded DNA is shown as a ribbon. Each group of embryonic V_κ genes appears as a stack of coils, and each gene corresponding to a subgroup is assigned a letter. Sequences flanking the coding regions are homologous among members of a group and are shown by cross-hatching patterns unique for each group. Member genes of a group are stacked to indicate homologous pairing which may permit generation of additional sequence diversity (see text). Subgroup E of group 1 is folded out of the stack to indicate that pairing need not only involve adjacent genes. Note that reciprocal recombination between subgroups 1D and 1F would lead to a deletion. ● = DNA rearrangement site; ○ = RNA splicing site; Note that no RNA splicing site follows J3, possibly explaining failure to observe the sequence of J3 in myeloma κ chains.^{35,43}

4. A possible mechanism for DNA rearrangement which permits V_κ - J_κ joining as well as somatic generation of sequence diversity in CDR3. (a) Double stranded DNA showing the palindrome following V_κ and the same palindrome inverted preceding J1. The first complete codon in J1 is numbered 96 since in some instances it becomes residue 96 of a completed V region. Bases numbered 1 through 9 are discussed below. (b) Only the top DNA strand of part (a) is

shown. Palindromic sequences following V_K and preceding $J1$ have annealed to form an inverted repeat stem structure and bring bases 1-5 and 6-9 into proximity to allow site-specific recombination. The complementary DNA strand (not shown) probably forms a similar structure. (c) Breaking and rejoining of the DNA strand shown in part (b) at different points between bases 1 and 9 can lead to codons determining three different amino acids at residue 96 of the newly formed V_K-J1 segment. The inverted repeat stem structure shown in part (b) has been deleted.

5. Structure of a mouse $C_{\gamma 1}$ gene⁴⁶ compared with that of a complete $V_H J_H C_{\gamma 1}$ chain. Separate $C_{\gamma 1}$ coding regions are made contiguous by splicing at the RNA level. Splicing sites (O) were observed preceding C_{H1} , between the hinge and C_{H2} , and between C_{H2} and C_{H3} , but not between C_{H1} and the hinge.
6. Hypothetical structure of the H chain locus which includes a possible mechanism for the "switch". The top line shows the H chain locus of a cell committed to producing $V_H J1 C_{\mu}$. Lying between J_n (the number of J segments for the H chain locus is not known) and C_{μ} are switch sites (S), and each C_H gene is preceded by a switch site (S'). Only the C_H genes studied by Honjo and Kataoka⁵¹ are shown, and the order of C_H genes is that suggested by them. The second line shows how $V_H J1$ may be switched from C_{μ} to $C_{\gamma 1}$ through formation (perhaps) of an inverted repeat stem structure involving the $S_{\gamma 1}$ and $S'_{\gamma 1}$ sites if these are appropriate palindromes (see Figure 4). The resulting $V_H J1 C_{\gamma 1}$ gene would contain a new DNA junction between $S_{\gamma 2b}$ and $C_{\gamma 1}$. In the model as shown, a further switch to $C_{\gamma 2b}$, $C_{\gamma 2a}$ or C_{α} could occur by the same mechanism. However, if the original order of S sites were that shown in the inset at the top of the figure, only a single switching

event would be possible.

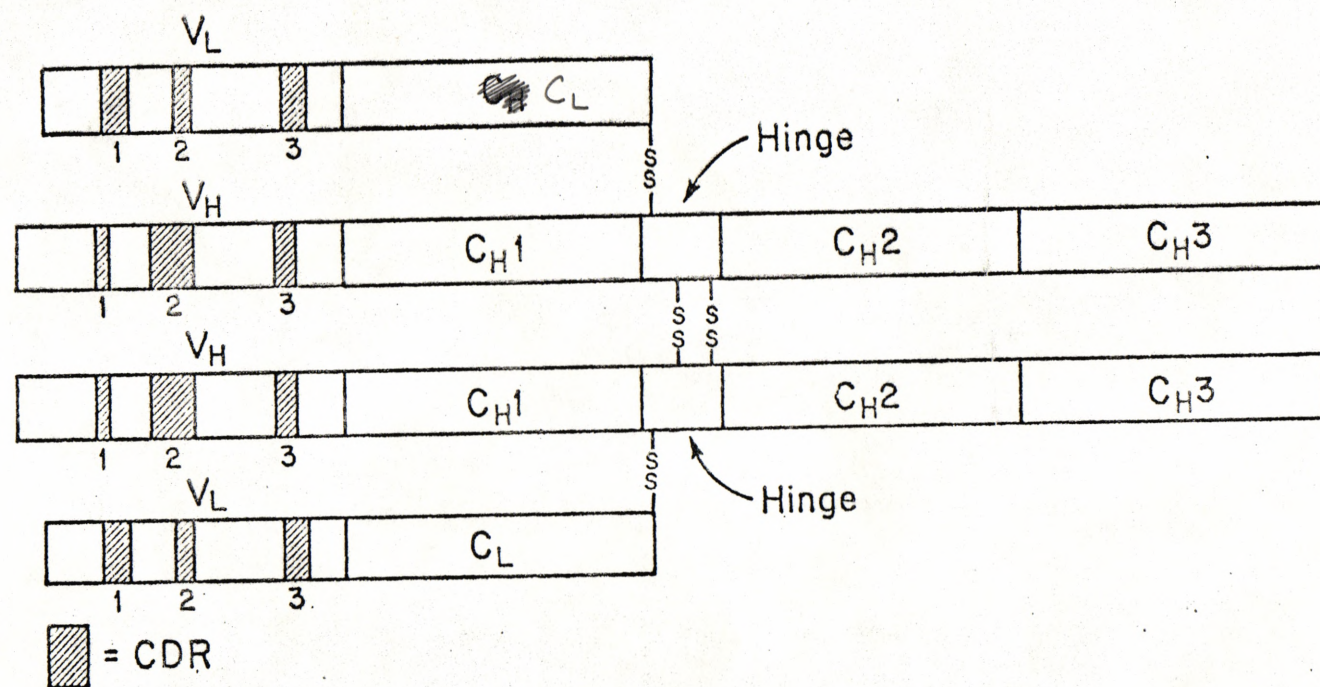


FIGURE 1

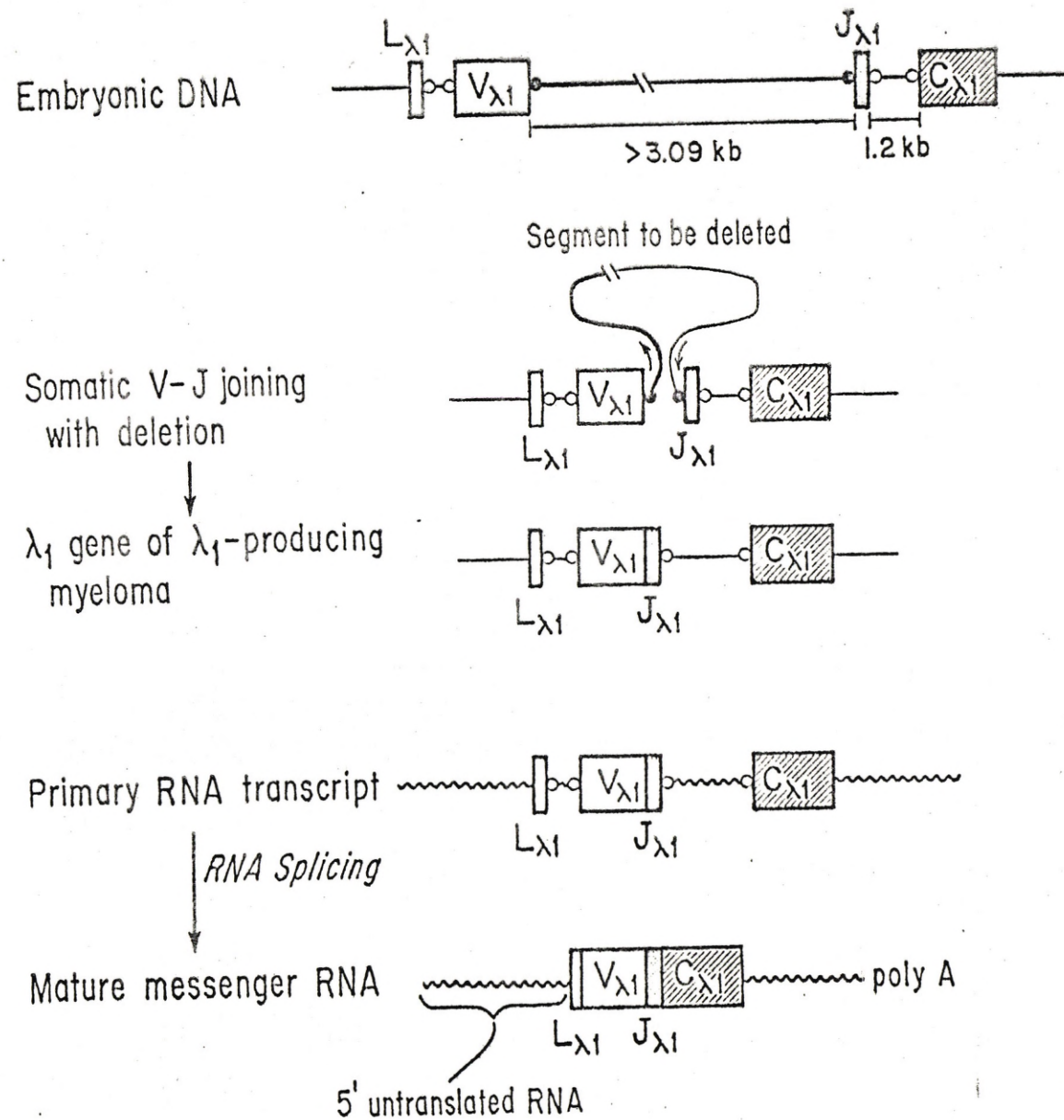
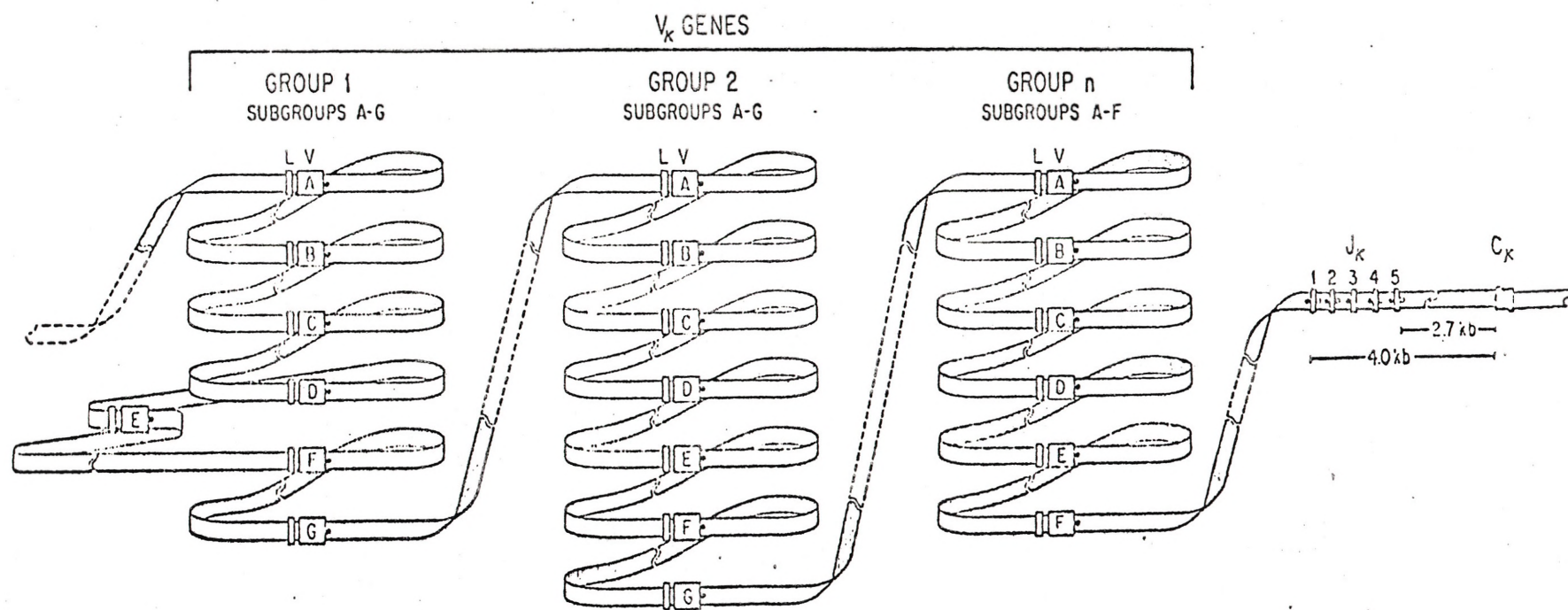


FIGURE 2

FIGURE 3



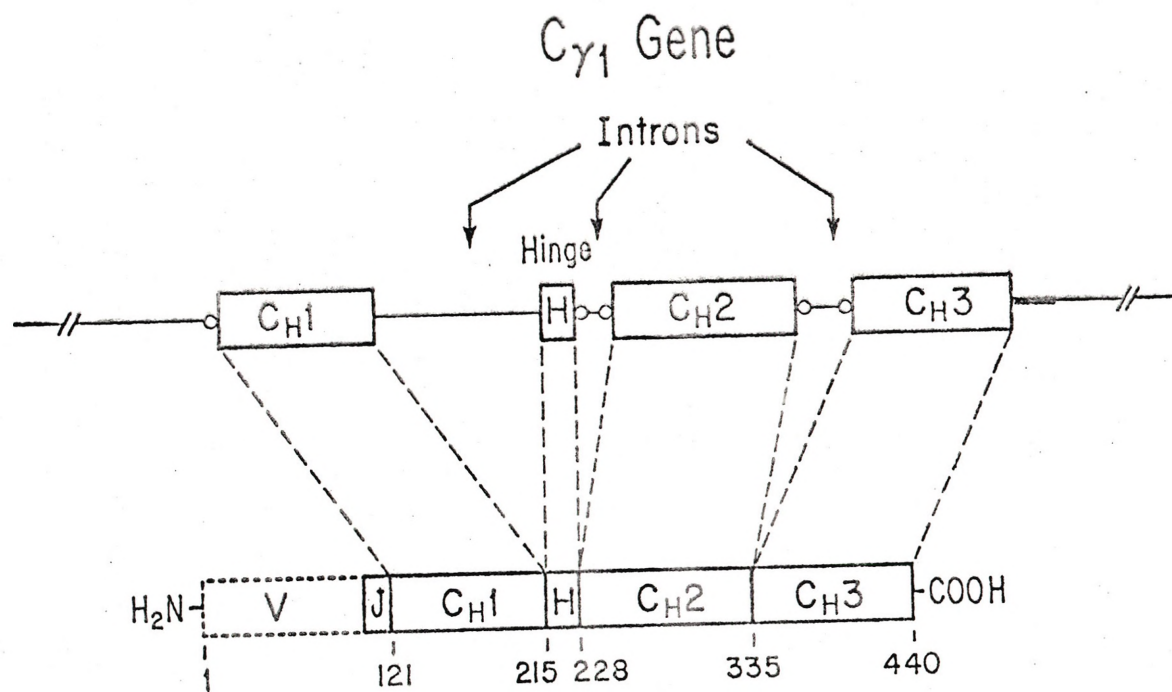


FIGURE 5

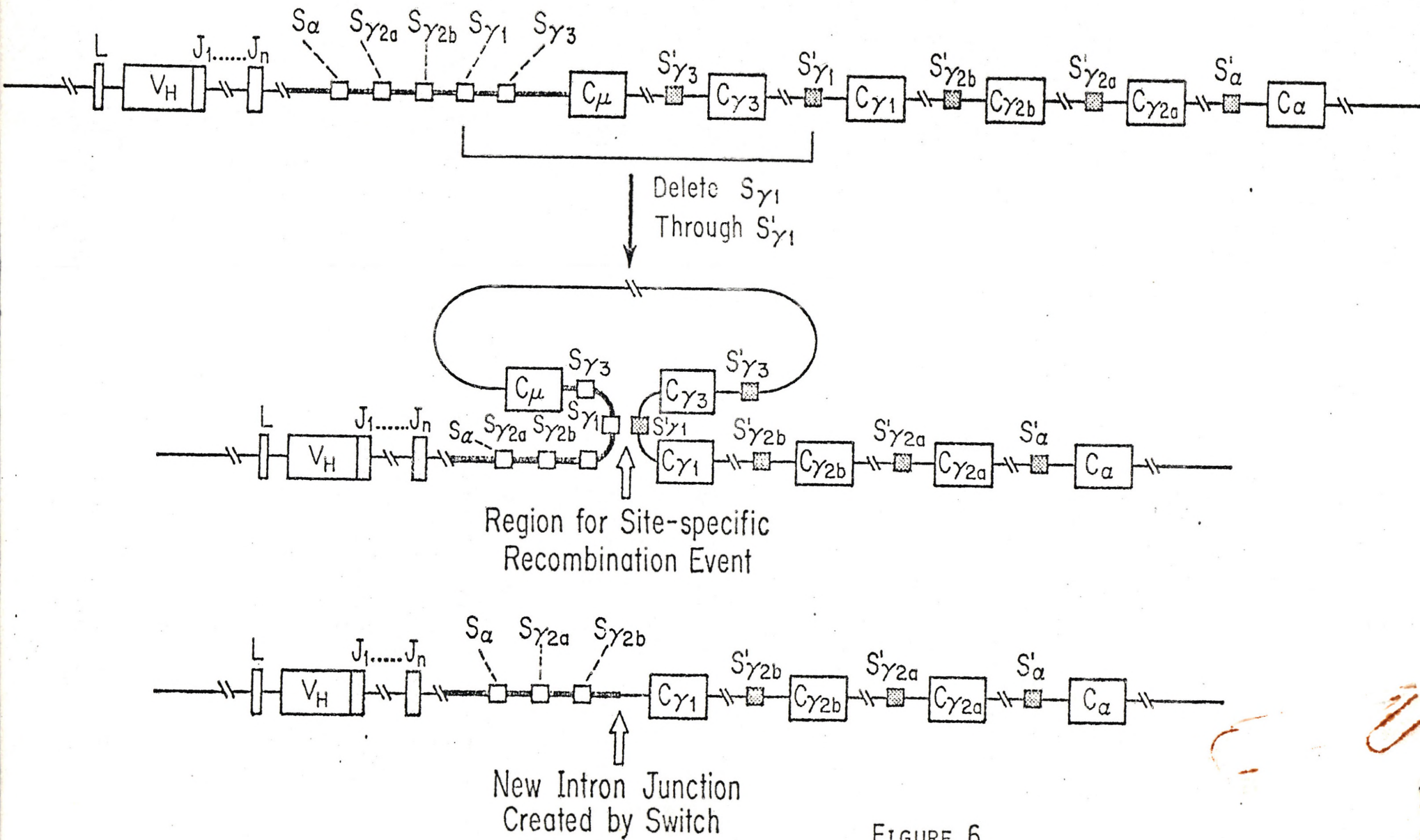
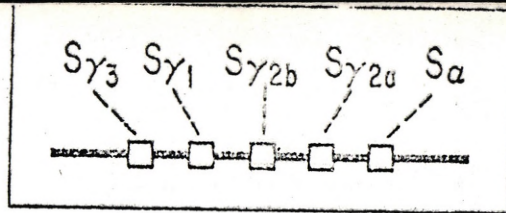


FIGURE 6

Ig molecules 2 ident light chains 2 ident heavy chains
each chain composed of a series of homologous domains;

V-region, begins at amino term, 1000's of different sequences

C-regions embody various effector functions of mol; few diff
classes of Ig (M,D,G,E,A) have diff bio funcs and are disting by
diff H chains (u,delta,gamm,epsilon,alpha)

L chains are of kappa or lambda class (ie containing V regions
and either C lambda or C kappa)

for most classes Ig secreted as 4-chain monomer IgM and IgA are
also secreted as pentamers and dimers; Ig monomers linked
covalently to a J or joining chain molecule

Ig genes are "split genes" Gene segments (exons) that code for V
and C domains separated from each other by noncoding DNA
sequences (introns)

exons are joined together by RNA-splicing events to generate
functional Ig messenger RNA's

chromosomal loc and organiz of mouse K, lambda and H chain genes
clusters are known

H chain cluster, a continuous stretch of DNA 200 kb (1kb=1000 BP)
on mouse chromosome 12 has been cloned and sequenced

This is how it looks in germline...SLIDE

Ig genes may comprise up to 1-2% of mouse genome

active, as opposed to germline, L and H chain genes generated by
somatic DNA rearrangements

a germline L chain V region is joined to a J region (not J chain)
or a H chain V region to a D and J region

this is called V region formation because combinatorial (or
mixing) events make complete V regions
contributes to the immense diversity of antibodies

Ig H chain genes undergo a second kind of DNA rearrangement called "class switching" in which the rearranged V is moved from the C-mu region and linked to another H chain C region accounts for the transition from M to G E or A expression

Ig messenger RNA's are generated from large nuclear precursors by RNA-processing events like RNA splicing

splicing generally produces a single RNA, however variants in splice result in differences between membrane-receptor and secreted forms from a single RNA transcript

more complex pathways allow for co-expression of surface IgM and IgD in B lymphocytes

When do these DNA and RNA rearrangements occur in B cell development?

Next SLIDE shows simplified scheme of B cell development based on changes in Ig gene expression and by existence of tumor cell lines corresponding to each of the developmental steps

1. pre-B cell contains a rearranged mu gene (ie the C-mu region gene has a rearranged V region in front of it)
expressed as cytoplasmic mu chain

L chain gene are not rearranged or detectably expressed

2. next cell is the immature B cell (contains surface IgM composed of 2 mu chains and 2 identical light chains (L or K). Thus there has been rearrangement of the L chain V region and activation of the L chain gene.

3. next cell is the B cell -steps involve RNA rather than DNA rearrangements. B cells coexpress mu and delta H chains bearing the same V region in surface IgM and IgD.

when a specific ag interacts with the Ig receptor B cells shift from expression of membrane M and D to secretion of IgM

4. plasma cell stages express and secrete the other classes of Ig

(G,A,E)

th transition from M secretion to production of other Ig classes involves a "class switching" DNA rearrangement.

The rearranged V region (1st expressed with C-mu) is joined to another H-chain C region. Mechanism of class switch appears different from that used in V region formation.

Altho best studied in the mouse, this knowledge can be extrapolated to man. Humoral immune response defects in man could be due to abnormalities in these mechanisms of DNA and RNA rearrangement.

apparent defects in Ig gene expression may turn out to be secondary to T help problems rather than structural alterations in H and L chain genes

-SCID-some lack B cells, others T cell factors which promote B cell development from stem cells

-Hypogamma associated with thymoma. Pts fail to exhibit pre-B cells with cyto mu. Lack mature circulating B cells and plasma cells.

seems to be clear failure to rearrange the V region and therefore failure to generate an active mu chain gene

-X-linked agamma. Young males with nl BM pre-B with large amts of cyto mu chains without light chains. B cell defect appears to be failure to generate L chain genes to generate a complete Ig molecule. Aberrant rearrangement of L chain genes or none at all.

-In 1 case of X linked hypogamm reported by Max Cooper the pre-B cells contained a rearranged and expressed mu gene + an aberrantly rearranged light vchain gene that does not make functional message

-in dysgamma type 1 (lo G,A, hi M) could be a block in class switching with hi M as a compensatory response

-cvid there are nl B # but make lo amts of Ig and cannot be activated by immunization in vivo or in vitro. Could be a generalized failure of RNA splicing.

UNIVERSITY OF CALIFORNIA—(Letterhead for Interdepartmental Use)

-Selective IgA defic-cells bear alpha but do not secrete. Could be structural gene defects affecting IgA secretion, eg structural alteration of carboxy-terminal region which make it unable to rx with J chain or chainges that eliminate glycosylation

-AT- IgA and E deficiency. Gene mapping studies from AT lymphoid lines indciate apparently nl alpha. Conclusion is that class switching is abn and does not activate IgE and A expressione

gene replacement therapies clearly far in future, however these natural expts can continue to help us understand Ig regulation.

VNL

