

Immunity and the Generation of Antibody Diversity by Jon Waggle

We are surrounded by pathogens. Bacteria, viruses, fungi, and other microorganisms, numbering in the thousands, are constantly seeking entry into the favorable growth medium our bodies provide. If these pathogens successfully gain entry into our bodies, they can interfere with the normal function of the infected tissue causing a number of ailments up to and including death. It is therefore imperative that our bodies are able to prevent these pathogens from gaining entry and, in the event that they do, eliminate them. This is the function of the immune system. The immune system is a complex network of interdependent systems which function to confer protection to the host from potentially harmful molecules, living or non-living, called antigens (Ag). The immune system accomplishes this task using a variety of methods. Antigens can be eliminated either by the inflammatory response or by the action of protein complement. They can also be engulfed by phagocytic cells of the immune system called macrophages. These functions however, are just as effective at removing host tissue as foreign tissue and therefore are not considered to be the major barrier to infection. The most important functions of the immune system include those which allow the host to quickly and efficiently remove only those specific antigens that are foreign and potentially pathogenic. This task requires that the immune system have a method of recognizing each antigen with tremendous specificity in order to avoid an immune response to host tissues. One way in which the immune system accomplishes this is through the production of antibodies by the B cell. Each antibody must be able to bind specifically to one part of an antigen, therefore a vast number of unique antibodies is required. The complex mechanism of antibody creation has been estimated to be able to produce *billions* of different structures which accounts for the widely effective function of these molecules. (1) This review is an attempt to introduce the reader to the genetic mechanisms responsible for the generation of this antibody diversity in the developing B cell that is one of the foundations of the human immune response.

Before we can effectively cover this topic however, we need a quick overview of human immunity. Immunity can be divided into two major categories; Innate (nonspecific) or Acquired (specific) immunity. Innate immunity is so called because it consists of cells and molecules that exist in the host prior to antigen exposure and are available immediately upon the inaugural contact with the antigen. It is comprised of the antigen nonspecific actions of phagocytic cells in the blood and tissues, circulating molecules such as protein complement, and a lymphocyte population known as natural killer (NK) cells. It comprises four types of barriers: anatomical, physiological, endocytic and phagocytic, and inflammatory. Anatomical barriers include the epithelium lining the surface of every orifice. Physiological barriers include temperature, pH, oxygen tension, and various soluble factors. Endocytic and phagocytic activity is carried out by monocytes, neutrophils and macrophages. The inflammatory response functions by altering the blood flow to the infected area to facilitate removal of pathogen and also to increase the temperature. (1)

Acquired immunity demonstrates the presence of a functional immune system that is capable of specifically recognizing and selectively eliminating antigens. Unlike innate immunity, acquired immunity only functions *after* exposure to antigen. It does not develop a measurable immune response upon the inaugural contact with antigen. Instead, efficient immune responsiveness develops over a period of time. This delay occurs because lymphocyte clones possessing specific receptors for the antigen are required for the expression of the acquired immune response. These clones must expand in number (clonal expansion) before strong

responses can be demonstrated. The initial delay in immune responsiveness is compensated for in subsequent exposures to the antigen, since numerous antigen-specific clones are in place to generate a rapid immune response. The rapid response following secondary exposure to previously encountered antigens is called immunological memory, and is the foundation for successful immunization campaigns.(2) The acquired immune system must also be able to distinguish between self and nonself to prevent the elimination of host tissues. Defects in this system can have serious crippling consequences including arthritis and death. Thus, the acquired immune system displays specificity, diversity, memory, and self/nonself recognition. The effects demonstrated by the specific immune system are generated by an interaction between two distinct classes of effector systems; cell-mediated and humoral immunity.

Humoral immunity refers specifically to those elements of the immune system which are extracellular in their activity. The primary example of this is the antibody (Ab). Antibodies can be secreted, from differentiated B cells called plasma cells, into the extracellular environment and remain effective. Antibodies bound to antigen can cause a variety of cell-mediated responses depending on what type of antibody is attached.(3)

Cell-mediated immunity is carried out by the functional cells of the immune system; T lymphocytes and antigen presenting cells(APC). T cells are produced in the bone marrow but reach maturity in the thymus. It is here that they come to express a unique antigen-binding receptor on their surface. This receptor is structurally different than the antibody receptor and therefore has a slightly different function in that they can only recognize antigen in association with cell-membrane proteins known as major histocompatibility complex (MHC) molecules. When a naive T cell encounters antigen associated with an MHC molecule on a cell, the T cell proliferates into a memory T cell or either a T helper (T_H) cell or a T cytotoxic (T_C) cell. (4)

As mentioned earlier, T cells can only recognize antigen in combination with MHC on the surface of cells. These presentation cells are called antigen presenting cells and they include macrophages and dendritic cells. These specialized cells internalize antigen, either by phagocytosis or endocytosis, and then re-express a part of that antigen, together with the MHC molecule, on their membrane. The helper T cell then recognizes the antigen/MHC complex and secretes soluble protein products called cytokines which regulate the activity of B cells and other cells of the immune system. In fact, antigen alone cannot stimulate the proliferation and differentiation of B cells. Antigen plus cytokines secreted by an activated helper T cell, which was in turn activated by the APC presenting the MHC/antigen complex, are necessary for inducing the B cell to activity. It is a seemingly redundant mechanism however, it is essential that the helper T cell be activated only when necessary, otherwise they would bind to host tissues and the individual would suffer the effects of an autoimmune disease. The MHC complex on the surface of the APCs is the cofactor which sets off the immune response.(4)

Innate immunity differs from acquired immunity in that it does not exhibit the tremendous diversity as antigens are eliminated wholly as opposed to a specific recognition event. Nor does the innate system exhibit memory, since heightened immune responses following reexposure are not observed. These differences, however, do not mean that both systems oppose each other in function. In fact, acquired immune responses depend on cells of the innate system such as macrophages and dendritic cells for antigen presentation. It is this codependence that explains the network example given earlier.

Now that a short overview of this system has been given we can move to a more thorough examination of that aspect of the immune system which is possibly the most important and is assuredly the most effective; the generation of antibody specificity.

B cells derive their name from the fact that they develop in the bone marrow. They leave the bone marrow expressing a unique antigen-binding glycoprotein receptor on their membrane

called an antibody molecule. The antibody molecule itself has two separable functions. One is to bind specifically to molecules from the pathogen that elicited the immune response; the other is to initiate the effector functions of the antibody such as the recruitment of other cells and molecules to destroy the pathogen once the antibody is bound to it. The antigen-binding region varies extensively between antibody molecules and is thus known as the variable region. The region of the antibody molecule that engages the effector functions of the immune system does not vary in the same way and is thus known as the constant region.(5)

All antibodies are constructed in the same way from four polypeptide chains. When antibodies are treated with agents that cleave disulfide bonds, two subunits can be distinguished.(5) One, a polypeptide chain of approximately 50kDa, is termed the heavy or H chain, and the other, of 25 kDa, is termed the light or L chain. The two chains are present in an equimolar ratio, and each intact antibody contains two heavy and two light chains. The two heavy chains are linked to each other by disulfide bonds and each heavy chain is linked to a light chain by a disulfide bond. In any one immunoglobulin molecule, the two heavy chains and the two light chains are identical, so the molecule has a two-fold axis of symmetry.(6)

There are only two types of light chains, which are termed lambda (λ) and kappa (κ) chains. No functional difference between the two has been found. By contrast, there are five main heavy-chain classes and these determine the functional activity of an antibody molecule. The five functional classes of immunoglobulin are immunoglobulin M (IgM), immunoglobulin D (IgD), immunoglobulin E (IgE), immunoglobulin A (IgA), and immunoglobulin G (IgG). Their heavy chains are denoted by the Greek letters μ , δ , ϵ , α , and γ respectively.(7)

Amino acid sequencing has shown that each immunoglobulin heavy and light chain contains distinct subregions, or domains, of about 110 amino acids in length. The light chain always has two regions, the heavy chains of IgA, IgD, and IgG have four regions and the heavy chains of IgM and IgE have five regions.(8) All of these subregions show similar patterns of amino acid sequences, suggesting that the immunoglobulin chains have evolved by repeated duplication of an ancestral gene corresponding to one subregion. It has been proposed that these domains originated in prokaryotes.(9)

X-ray crystallographic analysis and amino acid sequencing studies have revealed an intrachain disulfide bond forms a loop of about 60 amino acids which gives the domain a characteristic compact structure called an immunoglobulin fold. This structure consists of a 'sandwich' of two β pleated sheets, connected with the intrachain disulfide bond, each containing three or four antiparallel β strands of amino acids.(10) If the amino acid sequences of two immunoglobulins is compared it can be seen that the amino-terminal sequences of both heavy and light chains vary greatly between different antibodies and that this variability is limited to the first 110 amino acids, corresponding to the first subregion. The carboxy-terminal sequences are constant between immunoglobulin chains, either light or heavy, of the same type. The variable subregions are termed V regions and the constant subregions are called C regions.(11)

Sequence variability is not evenly distributed throughout the V regions. Many amino acids are conserved, particularly those that are important in determining the structure of the V domain. These relatively invariant regions are called framework regions, designated FW1, FW2, FW3, and FW4. By contrast, there are three regions, roughly from residues 28 to 35, from 49 to 59 and from 92 to 103, that show extensive variability. These have been denoted hypervariable regions and are designated HV1, HV2, and HV3. The framework regions actually form the β sheets that provide the structural framework of the domain, with the hypervariable region sequences corresponding to three loops at one edge of each sheet that are juxtaposed in the folded protein.(12) Thus, not only is sequence diversity focused on particular parts of the variable regions but it is localized to a particular part of the surface of the molecule. Additionally,

when the variable heavy chain region (V_H) and the variable light chain region (V_L) pair up in the antibody molecule, the hypervariable loops from each domain are brought together, creating a single hypervariable site at the tip of the antibody called the antigen-binding site. As these loops determine specificity by forming a surface complementary to the antigen, they are also called the complementarity determining regions, or CDRs.(13) Thus, the diversity contributed to the antibody by each chains variable region is compounded by the combination of each heavy and light chain to determine the final antigen specificity. This so called combinatorial diversity is a major contributor to antibody diversity.

Virtually any substance can elicit an antibody response. Functional groups on the surface of antigens, called epitopes or antigenic determinants, trigger this response when an antibody with the proper variable regions binds to them.(14) The number of chemical conformations that are possible is staggering and therefore, the number of unique antibodies must also be quite large. This vast number of antibodies required cannot be generated by the antibody production mechanisms that have been discussed. There is obviously additional diversity generating mechanisms, at the genetic level, that can account for the necessary diversity.

Before it was possible to examine immunoglobulin genes directly, there were two main hypotheses for the origin of this diversity. According to the germline theory, there is a separate gene for each antibody molecule and the antibody repertoire is largely inherited. In contrast, somatic mutation theories were based on the idea that a limited number of inherited antibody genes undergo mutation in B cells during the lifetime of an individual to generate the observed repertoire.(15) Recent research and the cloning of immunoglobulin genes has shown that the antibody repertoire is generated from a large but limited number of antibody genes that undergo DNA rearrangement.(16) Thus, both theories explain part of antibody diversity.

Detailed analysis of rearranged and germline genes for antibody molecules show that the κ and λ light chains and the heavy chains are encoded by separate multigene families situated on different chromosomes.

Table 1

Chromosomal Location of Gene Families

Gene	Chromosome Number
λ Light chain	22
κ Light chain	2
Heavy chain	14

Each of these multigene families contains a series of coding sequences called gene segments. The κ and λ light-chain families contain L(leader), V(Variable), J(Junctional), and C(Constant) gene segments; the heavy-chain family contains these same families and an additional group called the D(Diversity) region. A group of V gene segments is the first group, located 5' to the others, with each segment being preceded by a short L gene sequence that codes for a leader sequence that leads the heavy or light chain through the endoplasmic reticulum but is cleaved from the nascent polypeptide before assembly of the finished product. Each V gene segment is separated from the next by a noncoding sequence called an intron. This same arrangement is demonstrated in the other multigene families. A number of D, J, and C gene segments are separated from each other by an intron sequence. The estimated number of each gene

segment of each family is shown in the table below.

Table 2

Number of Coding Segments for Each Gene Family

Segment	kappa-light chain	lambda-light chain	Heavy chain
Variable (V)	100	100	75-250
Diversity (D)	0	0	30
Junctional (J)	5	6	6
Constant (C)	1	6	unknown

The sequence of rearrangement is slightly different between light and heavy chains in the developing B cell. In the light chains, the V and J gene segments are brought together at the DNA level by a recombination mechanism to be described later. The recombined VJ segment is preceded by a L segment and is followed by an intron sequence and a C segment. It isn't until this product is spliced after transcription that the C segment is joined to the Variable region. Heavy chain variable regions are constructed slightly differently. A DNA segment not present in light chains, called the D segment, is first joined to a J segment by the same recombination sequence utilized in light chains. This DJ segment is then recombined with the V segment to form the complete heavy chain variable domain. As in the light chain, the variable region coding segment, now called a VDJ segment, is joined to the C segment by RNA splicing after transcription not by recombination.(17) These recombination events are the major contributors to the generation of antibody diversity in the B cell. The multiple number of segments available for recombination in each multigene family represents the first level of diversity. Additionally, the process of recombination itself contributes to the level of diversity as is explained below.

When the non-coding DNA flanking the different heavy- and light-chain variable-domain gene segments is compared, conserved blocks of sequence can be identified at the 3' side of the V segments, the 5' side of the J segments, and on both sides of the D_H segments. The conserved sequences consist of a block of seven nucleotides (heptamer) and a block of nine nucleotides (nonamer) separated by 12 or 23 base pairs of 'spacer' DNA that are not conserved. These sequences are often referred to as recombination signal sequences (RSS). The heptamer sequence can be CACAGTG or its inverse compliment. The nonamer sequence can be ACAAAAACC or its inverse compliment. The spacing of 12 or 23 is required and is known as the 12/23 rule. This rule states that recombination only occurs between a heptamer and nonamer pair separated by a 12 base pair spacer and a pair separated by a 23 base pair spacer. Thus for heavy chains, the V segments cannot recombine directly with the J segments as both are flanked by recombination signals with 23 base pair spacers. It appears that the 12/23 rule arises from the fact that this number of base pairs corresponds to one or two complete turns of the DNA double helix, so that the heptamer and nonamer sequences are found on the same side of the DNA double helix, presumably where they can be recognized by the enzymes that mediate recombination.(7)

Although the details of the recombination process are not completely understood at the molecular level, there are several steps that are known to be required. The signals flanking the two gene segments to be recombined are aligned, presumably by the interaction of proteins that

bind to the signals. This is based on the 12/23 rule. The two ends of the variable-domain are not joined directly but undergo further processing that generates additional diversity at the junctions and is therefore known as junctional diversity. Junctional diversity is generated by two different processes that add nucleotides when the two double-stranded ends of the gene segments are joined during recombination. One gives rise to extra nucleotides known as P-nucleotides because they tend to be palindromic, and occurs in both heavy- and light-chain genes. The other generates extra residues, known as N-nucleotides because they are non-template encoded, and occurs only during heavy-chain recombination.

P-nucleotides arise when hairpin bends in the DNA at the ends of the immunoglobulin gene segments are cleaved asymmetrically. We will use the joining of the D segment to the J segment in heavy-chain rearrangement as an example. A single-strand break is created in each heptamer after the two gene segments to be joined are brought together. This break occurs randomly, at any base in the area and is therefore a source of diversity in itself. A hairpin end is formed when the phosphodiester bond in the complementary strand of the DNA molecule reacts with its counterpart in the opposite strand. This sequence creates a double stranded break that removes the signal sequences and joins the two ends of the DNA encoding the immunoglobulin gene segment to form the hairpin.(18) Before the two gene segments can be joined however, this bend must be cut. The cut is not usually at the very end; instead, the DNA is cleaved asymmetrically, leaving an overhanging stretch of single-stranded DNA containing one or more nucleotides from the opposite strand. Since this stretch of DNA contains a sequence of bases that were originally complementary to each other in the double-stranded DNA it is therefore a palindrome. Each single-stranded stretch of DNA on the two gene segments now pair with each other, with any unpaired bases at the ends of the segments being removed by exonuclease, leaving single-stranded gaps that can be filled in with complementary nucleotides.(18) The removal of unpaired bases from the end single-stranded segments means that the palindromic nature is sometimes lost. Nevertheless, this process alters the sequence of the joined gene segments at the point of joining, creating new diversity that is based on the precise site of cleavage of the hairpin.

The addition of N-nucleotides also involves the asymmetrical cleavage of the hairpin ends generated in recombination. However, this process only occurs at the V_H to D_H and D_H to J_H joints of the heavy-chain variable gene. The bases are added by the enzyme terminal deoxynucleotidyl transferase (TdT), which does not require a template strand and thus adds a random sequence of N-nucleotides at the joint, generating tremendous diversity at the site of gene segment joining. Following the addition of these nucleotides, the strands pair up and just as in P-nucleotide addition, those bases that do not pair up are cleaved by exonuclease and the single-strand stretches are then filled in with complementary bases.(19)

Little is known about the enzymes that catalyze recombination of antibody gene segments. Two genes have been identified that allow recombination of artificial immunoglobulin gene constructs to occur in non-antibody producing cells; they have been called RAG-1 and RAG-2, for recombination activator genes. Fibroblasts cotransfected with both proteins showed a 1000-fold increase in the frequency of V(D)J recombination. (20) However, these proteins are not all that is required as recombination of antibodies has been defective in mice when RAG-1 and RAG-2 are both expressed.(20) This is currently a hot topic of research.

When nucleotides are added or removed from a gene sequence as occurs during gene segment joining, it often results in a change in the reading frame of the gene. This can have several outcomes. The new reading frame may generate a premature stop codon which will fail to produce a functional product. More importantly it can also generate an alternate series of amino acids thereby further increasing the level of diversity.(21)

Overall, gene rearrangement in B cell development produces a single variable region for each antibody it will ever secrete. This occurs because although all cells have two sets of genes for each immunoglobulin chain, only one set is expressed in B cells; this is known as allelic exclusion.

Diversity in antibody molecules arises from several sources. Inherited variability arises from the presence of many variable-region gene segments in the genome. Additional diversity results from the formation of a complete variable-region gene by the random recombination of separate V, D, and J gene segments. Furthermore, the joining process itself is a source of diversity as a consequence of imprecise joining of the gene segments and the introduction of P- and N-nucleotides. A third source of diversity is the association of variable regions of the heavy and light chains to form the antigen-binding site. Finally, reading frame choices further diversify the antibody repertoire. The combination of all these sources of diversity creates a vast number of antibody specificities from a relatively limited number of genes.

References

1. Kuby, J, **Immunology 2nd edition**, W.H. Freeman and Company, 1994.
2. Lieber, M., R., **The Mechanism of V(D)J Recombination: A balance of Diversity, Specificity, and Stability**, *Cell*, Cell Press, Vol. 70, pp. 873-76, Sept. 18, 1992.
3. Oettinger, M., **Activation of V(D)J recombination by RAG1 and RAG2**, *TIG*, Elsevier Science publishers, Vol. 8 No. 12, Dec.1992.
4. Kuby, J, **Immunology 2nd edition**, W.H. Freeman and Company, 1994, p.59.
5. Kuby, J, **Immunology 2nd edition**, W.H. Freeman and Company, 1994, p.109
6. Sheriff, P, Sheriff, S, **Antibody-Antigen Complexes**, *Annual Review of Biochemistry*, Annual Reviews, Inc., Palo Alto, Vol. 59 1990, p. 493.
7. Stites, D. P., Terr, A. I., **Medical Immunology 9th edition**, Appleton and Lange, 1997, p. 356.
8. Janeway, C. A., Travers, P., **Immunobiology: The Immune System in Health and Disease**, Garland Publishing, 1994, p. 3:3.
9. Rolink, A., Melcher, F., **Molecular and Cellular Origins of B Lymphocyte Diversity**, *Cell*, Cell Press, Vol. 66, pp. 1081-94, Sept. 20 1991, p. 1081.
10. Sheriff, P, Sheriff, S, **Antibody-Antigen Complexes**, *Annual Review of Biochemistry*, Annual Reviews, Inc., Palo Alto, Vol. 59 1990, p. 501.
11. Janeway, C. A., Travers, P., **Immunobiology: The Immune System in Health and Disease**, Garland Publishing, 1994, p. 3:3.
12. Janeway, C. A., Travers, P., **Immunobiology: The Immune System in Health and Disease**, Garland Publishing, 1994, p. 3:9.
13. Stites, D. P., Terr, A. I., **Medical Immunology 9th edition**, Appleton and Lange, 1997, p. 509
14. Weaver, D. T., **V(D)J Recombination and Double-Strand Break Repair**, *Advances in immunology*, 58:29-86, 1995.
15. Kuby, J, **Immunology 2nd edition**, W.H. Freeman and Company, 1994, p.176.
16. Stites, D. P., Terr, A. I., **Medical Immunology 9th edition**, Appleton and Lange, 1997, p. 354
17. Lewis, S.M.,**The Mechanism of V(D)J Joining: Lessons from Molecular, Immunological, and comparative Analyses**, *Advances in immunology*, 56:27-150, 1994.
18. Lewis, S.M.,**The Mechanism of V(D)J Joining: Lessons from Molecular, Immunological, and comparative Analyses**, *Advances in immunology*, 56:27-150, 1994.
19. Stites, D. P., Terr, A. I., **Medical Immunology 9th edition**, Appleton and Lange, 1997, p. 358
20. Oettinger, M., **Activation of V(D)J recombination by RAG1 and RAG2**, *TIG*, Elsevier Science publishers, Vol. 8 No. 12, Dec.1992.

21. Raaphorst, F., M., Raman, C.,M., Nall, B.,T., Teale, J.,M., **Molecular mechanisms governing reading frame choice of immunoglobulin diversity genes**, Jan. 1997.

Copyright © 1997 Jon Waggle and Koni Stone