SPECIFICITY, DIVERSITY, AND IMMUNOGLOBULIN GENES

ANTIBODY SPECIFICITY. This can be thought of in terms of the 'goodness of fit' (affinity) between an antigenic determinant and a lymphocyte receptor or antibody. The better the fit, the more that cell or antibody seems to be specific for the determinant. It can be amazingly good: antibodies with association constants (K_a) in the range of 10^{15} liters/mole have been described[1].

Binding between Ag and Ab is non-covalent, depending like all protein interactions on charge-based ionic bonds, hydrogen bonding, van der Waals forces, and hydrophobic interactions between the amino acid side chains of the antibody's CDRs and corresponding residues on the antigenic determinant.

CROSS-REACTIVITY. This refers to the tendency of one antibody to react with more than one antigen. Again, it has to do with fit. We say colloquially that an antibody is 'against' mumps virus if it was obtained from an donor immunized with mumps, or if it reacts with mumps with high affinity); but we must remember that what the antibody really has is a combining site made up of CDRs whose position, charge, and hydrophobicity distribution is such that an antigenic determinant of the mumps virus binds it with observable affinity. Other antigenic determinants might also fit it; if they did so detectably, we would say that the antibody *cross-reacted* with those determinants. T cell mediated immunity has similar specificity and cross-reactivity.

> An example: some antigenic determinants on the lethal smallpox virus are similar to ones on the almost harmless cowpox virus. If you immunize a person with cowpox, the cross-reactivity with smallpox is such that he or she will be immune to that, too.
>
> Another example: formaldehyde will change incredibly poisonous tetanus toxin into a harmless **toxoid**, which retains antigenicity; if you're immunized with toxoid, your antibodies cross-react and neutralize toxin. Is that not *totally* cool?

To activate a B cell to produce antibody is a complex process. For purposes of our discussion now, it can be divided into two phases. First, binding of antigen to the B cell's receptors (membrane-bound versions of the antibody it will eventually release) occurs with a particular K_a. If this binding is strong enough, the second step, activation of the B cell, can take place. So an antigen which binds with low affinity may never activate the cell; but if another antigen comes along which not only binds but activates, the *product* of the cell (the secreted antibody) may combine with the low affinity antigen **well enough to be inconvenient.**

> A non-trivial example: Human heart valves contain an antigen, laminin, which cross-reacts with Group A streptococci. Obviously, the antigen in the valves does not normally activate the corresponding B cells, or we'd all have an autoimmune disease. When people get a streptococcal infection, the streptococcal antigens do activate these B cells because they bind to them with sufficient affinity. Then the released antibody can react with heart valves; with low affinity, it is true, but occasionally, in some people, with enough affinity to lead to a destructive inflammatory process: rheumatic heart disease.

CLONAL SELECTION THEORY. Each B cell makes only one specificity of antibody. How is it that one lymphocyte is anti-X, and another is anti-Y, etc.? The old theories were of the instructive type, that is, they said that the antigen told the immune system in some way to make an antibody of appropriate conformation, the way a potter's mold informs the pot. The trouble

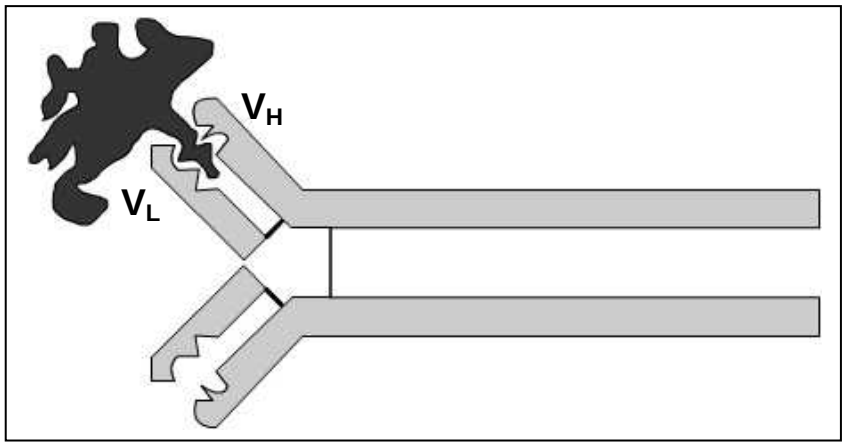
[1] This means that even when the antigen is diluted to 10^{-15} M half of the antibody combining sites will still be occupied by antigen.

with these theories was that they were Lamarckian; they implied that the outside world could instruct a cell to change its genetic information in some specific way so that a new protein was made. Lamarckian mechanisms just don't exist. In the 1950s, David Talmage here at CU and MacFarlane Burnet in Australia formulated the Clonal Selection Theory. It said that each cell of the immune system is programmed to make ***only one*** antibody (T cells weren't known yet, but the theory is true for them, too); that the choice of which antibody the cell will make is ***random***, not dependent on outside information; and that the entire population ***preexists*** in a normal individual, even before any contact with antigens. When a new antigen is introduced into the body, it comes into contact with a huge number of lymphocytes, and when it encounters one to whose receptors it binds with sufficient affinity, it activates it, resulting in selection of that clone to proliferate and secrete that antibody. **The best-fitting clones are selected by antigen.** This theory was initially hard to accept, but was proven to be true; and it is the central concept in immunology. Clonal selection is Darwinian; survival of the fittingest, as it were.

> How can we tell the theory is right? Here's the key experiment. An antigen, X, was coupled with an intensely radioactive label, such that any cell it bound to would quickly die from radiation. If instruction was the mechanism, enough Ag given would bind all the animal's B cells, because they would be 'non-specific' until instructed; so no antibody could be made. If selection was correct, only the pre-existing B cells with randomly expressed receptors for X would bind and die; all others would survive. A week after getting radioactive X, the animal was immunized with nonradioactive X and with Y, an unrelated antigen. It responded normally to Y, but not to X. Selection of pre-existing clones was proven.

ANTIBODY GENETICS. If B and T cell receptors are made even in the absence of antigen, there must be information in the genes for their structure. Most estimates are that our species can make a hundred trillion antibodies; does that require 10^{14} genes?

> **ASK YOURSELF:** How many genes does it take to assemble a person, anyway?

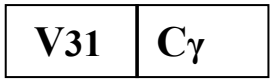


When the structure of antibodies was determined, it started to become apparent how we could get more diversity with fewer genes. Look again at an antibody molecule. The combining site (with which the epitope binds) is made up of the V (variable from one antibody to another) regions of H and L chains. If H and L chains are under separate genetic control and any two can associate randomly, then with 10^7 L chains and 10^7 H chains we can make $10^7 \times 10^7 = 10^{14}$ antibody combining sites; we could make 10^{14} antibodies with only 2×10^7 genes. This is a step in the right direction, but the real system is even more amazingly efficient.

ALLOTYPIC EXCLUSION, PART 1. (You may want to look back at the definition of allotypes.) The lambda, kappa, and H chain gene families are all on different chromosomes. A potential problem arises because, since we're diploid, each cell has two copies of each gene, maternal and paternal. Shouldn't that one cell make two different H chains and four different L chains, and therefore, by random combination, many different antibodies? It doesn't happen; only **one** H chain (maternal or paternal in origin) and **one** L chain (either kappa or lambda, either maternal or paternal) are synthesized in any **one** B cell. All the other genes are silenced. Though the *person* can make two allotypes, each *individual B cell* makes only one[2]. It is reminiscent of the way each female cell turns off one or the other of its X chromosomes; and this fact makes our job of understanding how we generate antibody diversity a little simpler.

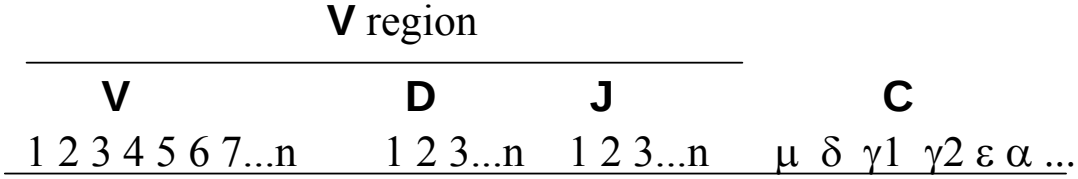
RECOMBINATION. If one probes germ line (sperm or egg) DNA, one finds many genes encoding Variable domains, but only one of each Constant region gene (i.e., one gene for to constant domains of delta, one gene for the constant domains of mu, etc.) per haploid genome. What exists in germ line must be something like this:

Whereas what is expressed by a B cell might be:



So it was found that the DNA *rearranges* in developing B cells, to bring one of many V's together with the correct C so that the unit could be copied into messenger RNA. Changing the relative positions of two pieces of DNA is called **recombination**. But soon, the 'V domain' genes were found to be broken up into smaller sets of minigenes.

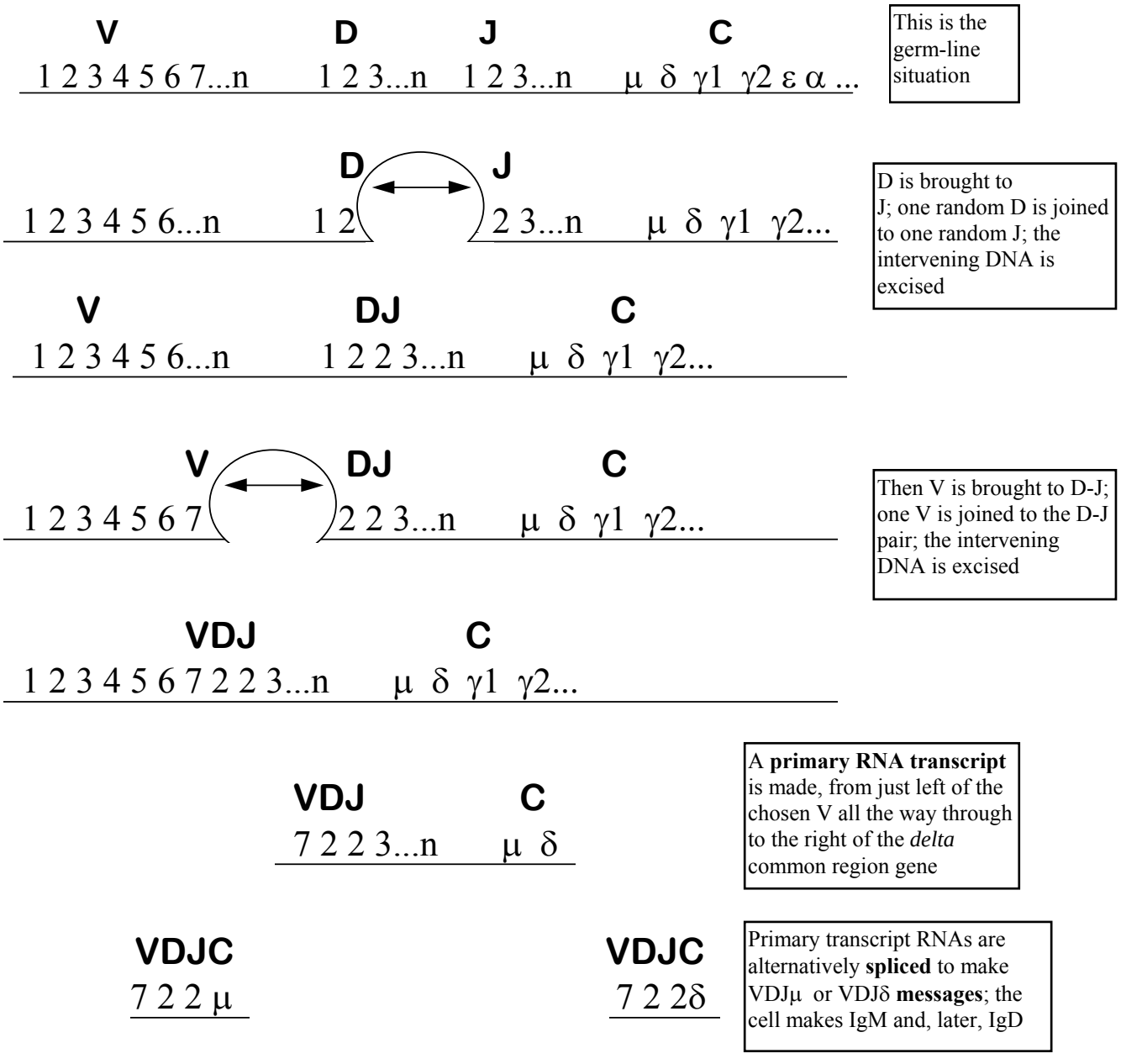
FINE STRUCTURE OF VARIABLE DOMAIN GENES. We need to agree on some terminology here: let's call the DNA which codes for the variable domain of an L or H chain the 'V domain gene region' rather than the V locus. This is because it turns out that at the DNA level, the information to code for a variable domain is actually **broken up** into segments or 'minigenes.' The variable domain region of heavy chain genes is broken up into multiple **V, D, and J gene segments**; the V region of light chains into V and J segments; so generically we say 'V(D)J.' The cell will choose one of its Vs, one D, and one J to make a V_H domain gene.



ASK YOURSELF: Let's say a complete V domain of the H chain is dinner. If the restaurant offered 10 soups, 15 main courses and 20 desserts, how many items did the chef have to prepare? How many different soup-main course-dessert dinners could they serve?

[2] This is good, because you wouldn't want a B cell that responds to mumps to also release antibody against, say, your kidneys. Specificity is important.

Heavy Chain: As we just said, its variable domain is coded by V, D, and J segments. The developing B cell first brings one random D segment close to one J; the DNA is cut, the intervening DNA is discarded and the ends joined. It then brings a V segment up to the recombined DJ, and repeats the cutting and joining process (there are splice acceptor and donor sites adjacent to each segment). Then the entire region from the assembled VDJ unit through to the end of the delta (of IgD) constant region gene, is transcribed into nuclear RNA. These **primary RNA transcripts** are alternatively processed by splicing, first to make only VDJ-mu, and later to make both VDJ-mu and VDJ-delta messages. Read the sequence down:



Light Chain: L chain gene rearrangement is similar, but they have only V and J segments, no D; and only one C domain gene.

There seem to be about 65 V's in the H family, 27 D's (in the H family only,) and 6 J's. At first glance this seems unnecessarily complicated, but look at how much diversity you get out of this arrangement. If you can recombine any V with any D and any J, 65 x 27 x 6 =10,500 different H chains made out of 98 gene segments! This gives us a lot more diversity, and we are not done yet.

> **RAG recombinases.** The enzymes that do the recombination of antibody and T cell receptor DNA are called RAG-1 and RAG-2 recombinases. The recombinases first bind splice signals to the right of a D segment and the left of a J segment, pull them together, and then cut and splice. Then they look for a splice sequence to the right of a V segment and do it again. If RAGs are knocked out, mice make neither B nor T cells. It happens in humans, too—very rarely (Omenn Syndrome). The RAG gene system appears in evolution with the jawed vertebrates; lampreys don't have it.

DIVERSITY GENERATION THROUGH SOMATIC VARIATION. The V-D and D-J joins are 'sloppy.' The cell has randomizing mechanisms: First, exonucleases for **chewing away** a few nucleotides after the DNA is cut but before two gene segments (D to J, V to DJ) are joined. Second, for **adding** a few nucleotides as well, an enzyme called terminal deoxynucleotidyl transferase, TdT, which doesn't use a template so its additions are random. Thus you can't predict the sequence at the joining area (which is called an 'N' region); it might be deduced that V7 has joined to D2, let's say, but in this cell there's an extra alanine and tyrosine there, and in that one there is a leucine missing. This produces a **lot** more **completely random** diversity. There is a price for it: two times out of three the N region, being of random length, will create a frame-shift mutation, that is, a nonsense codon which terminates transcription.

ALLOTYPIC EXCLUSION, PART 2. Sloppy recombination often ends up with a frame-shift mutation; when one examines a particular B cell, one often finds H and L genes that have been abortively rearranged, that is, in such a way as to produce nonsense codons. When this happens the cell tries again with the other allele; if things work, it goes on to become a B cell, if not, complete antibody cannot be made and the cell dies. If it gets lucky on the first try, though, it doesn't try the other allele. So although any one cell is theoretically capable of making 2 H chains (by rearranging both maternal and paternal loci,) and 4 light chains (maternal and paternal, kappa and lambda,) that never happens; it makes only one of each—all other alleles are *excluded*.

> **Receptor editing.** Although we just said that a B cell tries to rearrange each allele just once, that isn't strictly true. In some cells, when a rearrangement is detected as faulty (say a stop codon is generated), or when an anti-self receptor has been displayed, if the recombinases (RAG genes) are still active it can 'try again.' Sometimes this results in a successful cell. The process is called receptor editing.

GERM LINE VERSUS SOMATIC MUTATION. There used to be two schools of thought about antibody diversity: one said that all diversity (all the V genes) were in the germ line; if you looked at a fertilized ovum you could predict all potential antibodies that potential individual would be able to make. The other said that only a few were there. It postulated that during embryonic lymphoid development these genes underwent repeated (somatic) mutation until a full complement of antibodies was produced. Both theories, it turns out, were partly right. As we've just discussed, a lot of our diversity is in the germ line (that is, in the individual V, D, and J segments you're born with). Much more diversity is also generated by variable ('sloppy') V/J and V/D joining that generates N regions. And, finally and surprisingly, a new idea: the recombined V(D)J unit is **'hypermutable';** each time a B cell divides *after antigenic stimulation* there is a good chance that one of the daughters will make a slightly different antibody.

Selection of the best-fitting mutants after antigenic stimulation allows a gradual increase of affinity during an immune response—an exceptionally nice design feature called **affinity maturation**; it's why, for example, the longer you wait before making a monoclonal antibody, the better it is likely to be. (For T cells, there's *no* somatic mutation after contact with antigen.)

> **How it works.** Activation-Induced (Cytidine) Deaminase (AID) converts random cytosines in the CDR gene regions to uracil. So a C:G pair becomes a uracil: guanine mismatch. The uracil bases are excised by the repair enzyme uracil-DNA glycosylase. Error-prone DNA polymerases then fill in the gap, creating mostly single-base substitution mutations, so at the end of cell division one daughter may be making a different (worse? better?) antibody.

ASK YOURSELF: The mutation is random; can you think why the response gets better (higher average affinity), not worse?

CLASS SWITCHING. A single mature B cell starts by making both IgM and IgD, which it puts into its membrane as receptors, and then later it may switch to making IgG, IgE, or IgA. In all cases, **the V domain stays the same but the C region of the H chain changes**. As may be anticipated by now, what happens is that the cell which has put its particular H-chain VDJ combination together with its mu and delta genes (as shown in the diagrams) goes back to its DNA, does a loop-out of mu and delta, and puts VDJ next to the C-region gene of gamma or epsilon or alpha, while excising and discarding intervening DNA. The new mRNA, then, may be VDJα or VDJγ or VDJε. Thus a cell which is making IgM can go on to make IgG, but a cell making IgG cannot go back to making IgM; the mu information is *physically gone*. 'M to G' or 'M to A' or 'M to E' switches are common in antibody responses, and require T cell help; without it, only IgM responses are possible.

Remember that the cell switches heavy chain class, but doesn't switch light chains; they remain the same throughout the B cell's life.

SUMMARY OF RECEPTOR DIVERSITY MECHANISMS

Mechanism	B cell details		T cell details	
2-chain receptors (combinatorial diversity)	Each chain provides half the receptor's CDRs	Heavys times Lights	Each chain provides half the receptor's CDRs	Alphas times Betas
Recombination of germ-line segments (combinatorial diversity) RAG-1 & RAG-2	H chains: 65 V, 27 D, 6 J = 10,500 combos L chains: 35 V, no D, 5 J = 175 combos	10,500 times 175 = about 2 million antibodies	β chains: 50 V, 2 D, 13 J = 1,300 combos α chains: 70 V, no D, 60 J = 4,200 combos	4,200 times 1,300 = about 5.5 million T cells
'Optional diversity'	B cell can choose κ or λ L chains	Roughly doubles number of antibodies	There are also T cells with γ/δ receptors	Perhaps 5% of T cells are γ/δ
N region diversity (somatic)	Random nucleotides added or subtracted at VD and DJ joins	Estimated to produce 100 times more diversity than the germ line	Random nucleotides added or subtracted at VD and DJ joins	Estimated to produce 10,000 times more diversity than the germ line
Somatic hypermutation	After exposure to antigen	Mutation rate is about 1 in 10^4 cell divisions	Does not occur	
Total diversity including somatic hypermutation	~10^{14} antibodies possible; many fewer found in blood		~10^{11} TCR (T cell receptors) possible; about 10^8 found in blood	

Learning Objectives for Antibody Genes

1. Define:
   toxoid
   DNA recombination
   RNA splicing
   somatic mutation

2. Define cross-reactivity. Give an example of a non-self antigen which cross-reacts with a self antigen. Explain, in terms of lymphocyte activation, how a self antigen might not itself elicit antibody, but might react with antibody elicited by a cross- reacting antigen.

3. Discuss the Clonal Selection Theory in term of: the number of different receptor specificities it postulates per cell; the role antigen plays in the initial expression of receptors; the role of antigen in clonal selection; an experiment which provides strong evidence for the theory; how it differs from an instructional theory; whether it is Darwinian or Lamarckian.

4. Define allotypic exclusion. Demonstrate your knowledge of the concept by first stating the number of chromosomes in a cell which bear H or L genes, and then the number that actually contribute to a single B cell's antibody product.

5. Draw a diagram of the heavy and light chain gene regions of human DNA. Indicate V, (D,) J, and C subregions. Show how a heavy or light chain gene is assembled out of these subregions during the differentiation of a B cell.

6. Describe the somatic recombination model which explains how antibodies of the same specificity (that is, with the same CDRs and idiotype) can be found in two or more different classes ('class switching').

7. Calculate the minimal number of genes required to code for a million different antibody molecules, based on the (outdated) concept of 'one gene-one H or L chain'. Show how breaking the variable region gene up into V, D and J subregions requires fewer genes.

8. Define somatic mutation, and describe the essential difference between the somatic mutation and germ line hypotheses of immunological diversity.

9. Define somatic *hyper*mutation and distinguish it from the somatic mutation mechanism that produces N-region diversity.