Modelling the Diversity of the Immune System

Catherine Beauchemin, Department of Physics, University of Alberta

April 20, 2002

Contents

1 Immunoglobin structure (Kuby, Chapter 4)

- The first 100-110 or so amino acids of the amino-terminal region of a light or heavy chain varies greatly among antibodies of different specificity. These segments of highly variable sequence are called *V regions*: V_L in light chains and V_H in heavy. In fact, most of the differences among antibodies fall within areas called complementarity-determining regions (CDRs), and it is these CDRs, on both light and heavy chains, that constitute the antigen-binding site of the immunoglobulin molecule. (p.85)
- Myeloma cells (a plasma cell that can divide) yields antibodies called "myeloma protein". They are a result of a disease called "multiple myeloma". (p.86)

Class	Heavy chain	Subclasses	Light chain	Molecular formula
IgG	γ	$\gamma 1$, $\gamma 2$, $\gamma 3$, $\gamma 4$	κ or λ	$\gamma_2\kappa_2$ or $\gamma_2\lambda_2$
IgM	μ	—	κ or λ	$(\mu_2\kappa_2)_n$ or $(\mu_2\lambda_2)_n$ for $n = 1$ or 5
IgA	α	$\alpha 1$, $\alpha 2$	κ or λ	$(\alpha_2\kappa_2)_n$ or $(\alpha_2\lambda_2)_n$ for $n = 1$, 2, 3, or 4
IgE	ϵ	—	κ or λ	$\epsilon_2\kappa_2$ or $\epsilon_2\lambda_2$
IgD	δ	—	κ or λ	$\delta_2\kappa_2$ or $\delta_2\lambda_2$

Table 1: Chain composition of the 5 immunoglobulin classes in humans. (p.87, Table 4-1)

- Constant light chains come in 2 types: κ and λ. There are 4 λ subtypes in humans and are characterized by different single amino acids at 2 or 3 positions. Each antibody can only contain 1 light chain type. See Table 1. (p.86)
- Constant heavy chains come in 5 types: μ, δ, γ, ϵ and α called isotypes. The lenght fo the constant region is approximately 330 amino acids for isotypes δ, γ, and α, and 440 for isotypes μ and ϵ. (p.86-87)
- An antibody molecule has 2 identical heavy chains and 2 identical light chains or a multiple of this basic 4-chain structure. See Table 1. (p.87)
- Variable regions are subdivided into hypervariable regions (HV) and framework regions (FR). The FR is much less variable than the HV. The HV is also called the complementary-determining region (CDR). There are 6 CDRs on an antibody (3 in V_L and 3 in V_H). (p.89)
- Antigen-antibody contact: 15-22 amino acids in the antibody contact the same number of residues in the protein antigen: an area of approx. 650 Å^2 to more than 900 Å^2. (p.90)
- In general, more residues in the heavy-chain CDRs appear to contact antigen than in the light-chain CDRs. Thus the V_H domain often contrubutes more to antigen binding than the V_L domain. (p.91)
- All of the hybrid antibodies bound to the HIV glycoprotein antigen, indicating that the heavy chain alone was sufficient to confer specificity. However, one should not conclude that the light chain is largely irrelevant; in some antibody-antigen reactions, the light chain makes the most important contribution. (p.92)
- In fact, upon antigen binding, the CDR1 region of the light chain moves as much as 1 Å and the heavy chain CDR3 moves 2.7 Å. Thus, in addition to variability in the length and amino acid composition of the CDR loops, the ability of this loops to significantly change conformation upon antigen binding enables antibodies to assume a shape more effectively complementary to that of their epitopes. (p.92) Antigen can also undergo conformational changes upon binding with antibody. (p.92)
- The C_H1 and C_L domains also may contribute to antibody diversity by allowing more random associations between V_H and V_L domains than would occur id this association were driven by the V_H/V_L interaction alone. (p.93)
- The presence of C_H1 and C_L domains appears to increase the number of stable V_H and V_L interactions that are possible, thus contributing to the overall diversity of antibody molecules that can be expressed by an animal. (p.93)
- Skipped p.95–102.
- Antigenic determinants on Immunoglobins (p.102):
  - Isotype: isotypic determinants are constant-region determinants that distinguish each Ig class and subclass within a species.

  - Allotype: allotypic determinants are subtle amino acid differences encoded by different alleles of isotype genes. Allotypic differences can be detected by comparing the same antibody class among different inbred strains.
  - Idiotype: idiotypic determinants are generated by the conformation of the amino acid sequences of the heavy- and light-chain variable regions specific for each antigen. Each individual determinants is called an idiotope, and the sum of the individual idiotopes is the idiotype.

- Ig-α and Ig-β heterodimer (part of B-cell receptor), T-cell receptor and accessory proteins, class I and II MHC molecules are all part of the Ig superfamily (p.104) i.e. proteins whose corresponding genes derived from a common primordial gene encoding the basic domain structure. The primary amino acid sequence of these proteins suggests that they all contain typical immunoglobulin-fold domains consisting of about 110 amino acids, arranged in pleated sheets of antiparallel β strands, usually with an invariant intrachain disulfide bond spanning 50-70 residues.

- Skipped p.104–110.

2 Biophysics basics (Jack, Jan. 30, 2003)

- An amino acid string has to extremities called the N-terminal and the C-terminal.

- The N-terminal is the beginning of the chain and is usually burried inside the structure when the amino acid string is folded.

- The C-terminal usually sticks out of the amino acid structure when the string is in its folded state. This terminal is so called because it terminates with a carboxyl (COOH).

- See book *Biochemistry* by L. Stryer for more info.

- There are 20 different amino acids. Amino acids can be divided across a spectrum of charge groups. They can be charged, neutral and polar (permanent dipole) or neutral and non-polar. Charged and neutral polar amino acids are called hydrophylic and the neutral non-polar are hydrophobic. Those classes are not clear cut as the different amino acids spread out over the spectrum ranging from hydrophilic to hydrophobic based on their charge and polarity.

- Since all amino acids have various charges or dipole moment, they will tend to group together in such a way as to diminish the energy. A lot of work is done on trying to understand and predict the ways in which a particular amino acid string will fold depending on the available energy and the environment it is in.

- Amino acid strings form highly organized structures. The levels of organization are:
  - Primary structure: the sequence in which the amino acids are organized within the string.
  - Secondary structure: the motifs formed by the string (e.g. 3 β-sheets and 2 α-helix).
  - Tertiary structure: The 3-D organisation of the structures (usually represented by ribbon diagrams)
  - Quaternary structure: The functional domains (e.g. in the case of a T-cell this could be the recognition domain that make up the T-cell receptor)

- There are three type of representation for folded amino acid strings like proteins, they are
  - ribbon
  - ball-and-stick
  - space filling

- The ribbon diagram is used to represent the various conformations amino acid strings adopt, principally their tertiary structure. The various structures are:
  - α-helix.
  - β-sheet.
  - pleated-sheet.
  - random conformation between organized secondary structures.
- The ball-and-stick diagram represents atoms by balls and ionic links by sticks.
- The space filling diagram represents entire molecules by balls. This type of diagram is usually used to illustrate the charge distribution of a structure to visualize its binding or repulsive potential.
- The following are the various Coulomb forces by strength
  - Strong: ionic bonds, covalent bonds.
  - Weak: Van Der Walls, etc.
  - Weakest: hydrogen bonds.

3 Immunoglobin diversity (Kuby, Chapter 5)

- In germ-line DNA, multiple gene segments encode portions of a single immunoglobulin. During B-cell maturation in the bone marrow, certain of these genes segments are randomly shuffled by a dynamic genetic system capable of generating more than 10^{10} combinations. (p.115)
- After antigenic stimulation of a mature B cell in peripheral lymphoid organs, further rearrangement of constant-region gene segments can generate changes in the isotype expressed, producing changes in the biological effector functions without changing the specificity of the immunoglobulin molecule. Thus, mature B cells contain chromosomal DNA that is no longer identical to germ-line DNA. Genomic rearrangement is an essential feature of lymphocyte differentiation, an no other vertebrate cell type has been shown to undergo this process. (p.115)
- In 1976, S. Tonegawa and N. Hozumi found the first direct evidence that separate genes encode the V and C regions of immunoglobulins and that the genes are rearranged in the course of B-cell differentiation. (p.117)
- The Dreyer and Benett two-gene model — one gene encoding the vairable region and one gene encoding the constant region — applied to both heavy- and light-chain genes. (p.118)
- The κ and λ light-chain families contain V, J, and C gene segments; the rearranged VJ segments encode the variable region of the light chains. The heavy-chain family contains V, D, J, and C gene segments; the rearranged VDJ gene segments encode the variable region of the heavy chain. In each gene family, C gene segments encode the constant regions. (p.118–119)
- Note: It takes 3 nucleotides to encode 1 amino acid. (p.119)
- Light Chains:
  - The λ multigene family in human germ line contains 30 V_λ gene segments, 4 J_λ gene segments, and 4 C_λ gene segments. The V_λ and the 3 functional J_λ gene segments encode the variable region of the light chain, and each of the 3 functional C_λ gene segments encodes the constant region of one of the three λ-chain subtypes (λ1, λ2, and λ3). FIXME: are the 4^{th} J and C gene segments pseudogenes in humans too? (p.120) In human, any of the functional V_λ genes can combine with any of the 4 functional J_λ-C_λ combinations. (p.120–121)

  - The κ-chain multigene family in the human contains approximately 40 V_κ gene segments, 5 J_κ segments (FIXME: one of which is a pseudogene), and a single C_κ segment. The V_κ and J_κ gene segments encode the variable region of the κ light chaing, and the C_κ gene segment encodes the constant region. Since there is only 1 C_κ gene segment, there are no subtypes of κ light chains. (p.120) In human, any one of the V_κ gene segments can be joined with any one of the functional J_κ gene segment. (p.121)

- Heavy Chains:

  - The V_H gene segment was found to encode amino acids 1 to 94 and the J_H gene segment was found to encode amino acids 98 to 113. An additional nucleotide sequence, D_H, observed between the V_H and J_H gene segments corresponds to amino acids 95 to 97 of the heavy chain. The D_H gene segment encodes those amino acids located within the third complementarity-determining region (CDR3) and was designated D for diversity, because of its contribution to the generation of antibody diversity. (p.120)
  - The heavy-chain multigene family in humans has been shown to contain 51 V_H gene segements upstream from a cluster of 27 functional D_H gene segments. Downstream from the D_H gene segments are 6 functional J_H gene segments, followed by a series of 5 C_H gene segments. (p.120)
  - In humans and mice, the C_H gene segments are arranged sequentially in the order C_μ, C_δ, C_γ, C_ϵ, C_α. This sequential arrangement is generally related ot the sequential expression of the immunoglobulin classes in the course of B-cell development and the initial IgM response of a B-cell to its first encouter with an antigen. (p.120)

- The heavy-chain variable region genes rearrange first, then the light-chain variable region genes. This process of variable-region gene rearrangement produces mature, immunocompetent B cells; each such cell is committed to produce antibody with a binding site encoded by the particular sequence of its rearranged V genes. (p.120)

- Recombination signal sequences (RSS) flank certain germ-line gene segments. The sequences function as signals for the recombination process that rearranges the genes. They contain sequences of 12 to 23 base pairs corresponding to one and two turns of the DNA helix; consequently they were named one-turn recombination signal sequences and two-turn signal sequence. Signal sequences having a one-turn spacer can join only with sequences having a two-turn spacer (the so-called one-turn/two-turn joining rule). (p.122)

- The V_κ signal sequence has a one-turn spacer, and the J_κ signal sequence has a two-turn spacer. In the λ light-chain, the V_λ signal has a two-turn spacer and the J_λ signal has a one-turn spacer. In heavy chain, the signal sequences of the V_H and J_H gene segments have two-turn spacers, the signals on either side of the D_H gene segment have one-turn spacers. (p.122)

- In a nonproductive rearrangement, the resulting VJ or VDJ unit will contain numerous stop codons, which interrupt translation. But when gene segments undergo a productive rearrangement, the reading frame is maintained such that the resulting VJ or VDJ unit can be translated in its entirety, yielding a complete antibody. (p.126–127)

- B cells, like all somatic cells, are diploid and contain both maternal and paternal chromosomes. However, it expresses the rearranged heavy-chain genes from only one chromosome and the rearranged light-chain genes from only one chromosome. This process called allelic exclusion insures that functional B cells never contain more than one $V_H D_H J_H$ and one $V_L J_L$ unit because the expression of both alleles would render the B cell multispecific. (p.127)

- Studies suggest that expression of the heavy- and light-chain proteins may indeed prevent gene rearrangement of the remaining alleles and thus account for allelic exclusion. (p.128)

- To date, 7 means of antibody diversification have been identified in humans: (p.128)
  - **Multiple germ-line gene segments**
    e.g. human germ-line DNA contains 51 V_H, 27 D_H, 6 J_H, 40 V_κ, 5 J_κ, 30 V_λ, and 4 J_λ gene segments. (p.128)
  - **Combinatorial V-(D)-J joining**
    Possible combinations for V-(D)-J joining: for heavy-chain $51 \times 27 \times 6 = 8262$, κ light-chain $40 \times 5 = 200$, and λ light-chain $= 30 \times 4 = 120$. (p.128–129)
  - **Junctional flexibility**
    Recombination involves both the joining of recombination signal sequences to form a signal joint and the joining of coding sequences to form a coding joint. Although the signal sequences are always joined precisely, joining of the coding sequences often is imprecise. (p.129) Junctional flexibility leads to many nonproductive rearrangements, but it also generates several productive combinations that encode alternative amino acids at each coding joint. The amino acid sequence variation generated by junctional flexibility fall within CDR3 in immunoglobulin heavy- and light-chain DNA. Since CDR3 makes a major contribution to antigen binding by the antibody molecule, amino acid changes generated by junctional flexibility can make a major contribution to antibody diversity. (p.130)
  - **P-region nucleotide addition (P-addition)**
    After the initial single-strand DNA cleavage at the junction of a variable-region gene segment and attached signal sequence, the nucleotides at the end of the coding sequence turn back to form a hairpin structure. This hairpin is later cleaved by an endonuclease. This second cleavage sometimes occurs at a position that leaves a short single strand at the end of the coding sequence. Subsequent addition (P-addition) of complementary nucleotides to this strand by repair enzymes generates a palindromic sequence in the coding joint, and so these nucleotides are cllaed P-nucleotides. (p.130)
  - **N-region nucleotide addition (N-addition)**
    N-nucleotides are added in heavy-chains only during the D-J and V to D-J joining process by terminal deoxynucleotidyl transferase (TdT) catalyzed reaction. Up to 15 N-nucleotides can be added to both the $D_H - J_H$ and $V_H - D_H J_H$ joints. Thus a complete heavy-chain variable region is encoded by a $V_H N D_H N J_H$ unit. The additional diversity is large because N regions appear to consist of wholly random sequences. Since this diversity occurs at V-D-J coding joints, it is localized in the CDR3 of the heavy-chain genes. (p.130–131)
  - **Somatic hypermutation**
    Normally, somatic hypermutation occurs only within germinal centers. It is targeted to rearranged V-regions located within a DNA sequence containing about 1500 nucleotides, which includes the whole of the VJ or VDJ segment. Somatic hypermutation occurs at a frequency approaching 10^{-3} per base pair per generation. This is at least a hundred thousand-fold gither that the spontaneous mutation rate, about 10^{-8}/bp/generation, in other genes. Since the combined length of the H-chain and L-chain variable-region genes is about 600 bp, then an average of one mutation will be introduced by somatic hypermutation in every one or two cell divisions. Most of the mutations are nucleotide substitutions rather than deletions or insertions. SH introduces these substitutions in a largely, but not completely, random fashion. Certain nucleotid motifs within V_H and V_L my be especially susceptible. Although SH occurs throughout the VJ and VDJ segment, in mature B cells they are clustered within the CDRs of the V_H and V_L sequences. (p.130)
  - **Combinatorial association of light and heavy chains**
    Possible combinations for all possible heavy and light chains: $8262 \times (200 + 120) = 2.64 \times 10^6$. (p.129)
- Since CDR1 and CDR2 sequences are encoded by V segments only while CDR3 is encoded by $V_L - J_L$

Complex	HLA							
MHC class	II			III		I		
Region	DP	DQ	DR	C4,C2,BF		B	C	A
Gene products	DP$\alpha\beta$	DP$\alpha\beta$	DP$\alpha\beta$	C' proteins	TNF-αTNF-β	HLA-B	HLA-C	HLA-A

Table 2: Human HLA complex. (p.174, Figure 7-1)

and $V_H - D_H - J_H$ junctions, junctional flexibility as well as P- and N-additions occur only in CDR3 while somatic hypermutations can occur in CDR1, CDR2 and CDR3. (p.130)

- Affinity maturation normally takes place within germinal centers. (p.130)
- A particular immunoglobin can exist in a membrane-bound form or in a secreted form. The two forms differ in the amino acid sequence of the heavy-chain carboxyl-terminal domains (C_H3/C_H3 in IgA, IgD, and IgG and C_H4/C_H4 in IgE and IgM). The secreted form has a hydrophilic sequence of about 20 amino acids in the carboxyl-terminal domain; this is replaced in the membrane-bound form with a sequence of about 40 amino acids containing a hydrophilic segment that is outside the cell, a hydrophobic transmembrane segment, and a short hydrophilic cytoplasmic segment at the carboxyl-terminus. (p.135)
- Mature naive B cells produce only membrane-bound antibody, whereas differentiated plasma cells produce secreted antibodies. Presumably, some mechanism exists in naive B cells and in plasma cells that directs RNA processing preferentially toward the production of mRNA encoding either the membrane form or the secreted form of an immunoglobulin. (p.135)
- Skipped p.139–144.

4 MHC (Kuby, Chapter 7)

- Something accepted as self is *histocompatible*, but rejected as self is called *histoincompatible*. (p.173)
- QQ: "most T cells recognize antigen only when it is combined with an MHC molecule". Isn't this always true? (p.173)
- Because MHC molecules act as antigen-presenting structures, the partucular set of MHC molecules expressed by an individual influences the repertoire of antigens to which that individual's T_H and T_C cells can respond. (p.173)
- The MHC is referred to as the HLA complex in humans and as the H-2 complex in mice. See Table 2 for the organization of the HLA complex. (p.173)
- QQ: Why is the human and mice complex differently named? Is there a different name for each species? Is saying human HLA complex redundant? (p.173)
- MHC genes are organized into regions encoding 3 classes of molecules: (p.174)
  - **Class I MHC genes** encode glycoproteins expressed on the surface of nearly all nucleated cells and their major function of the gene products is presentation of peptide antigens to T_C cells.
  - **Class II MHC genes** encode glycoproteins expressed primarily on antigen-presenting cells (macrophages, dendritic cells, and B cells), where they present processed antigenic peptides to T_H cells.
  - **Class III MHC genes** generally encode various secreted proteins that have immune functions, including components of the complement system and molecules involved in inflammation.

FIXME ADD FIGURE

Figure 1: (LEFT) Inheritance of HLA haplotypes in a typical human family. (RIGHT) A new haplotype (R) arises from recombination of maternal haplotypes. (p.176, Figure 7-2c and d)

- QQ: What cells are nucleated, not-nucleated? (p.174)
- QQ: So Class III MHC products are only for signalling, not recognition? If it is so different, why does it not carry a different name? Is it just 'cause it's in the same spot? (p.174)
- The loci constituting the MHC are highly polymorphic, i.e. many alternative forms of the gene, or alleles, exist in each locus. (p.175)
- QQ: Clarify the meaning of locus. (p.175)
- Each set of alleles is referred to as a haplotype. An individual inherits one haplotype from the mother and one from the father. The alleles are codominantly expressed, i.e. both maternal and paternal genes products are expressed in the same cells. (p.175)
- QQ: What does homo/heterozygous means? (p.175)
- QQ: Can both parents' MHC be expressed at any time? How does it work? (p.175)
- If one carries different haplotype from each parent, one is histocompatible with both strains and able to accept grafts from either strains. (p.175)
- QQ: Explain all this gene stuff and figure 7-1 and table 7-1. (p.174–175)
- QQ: can you be heterozygous in a loci and homozygous in another? Is that a symptom of inbreding? (p.175)
- When the father and mother have different haplotypes, there is 1 in 4 chance that siblings will be histocompatible with each other, but no chance any can ever be compatible with the parents. See Figure 1. (p.175)
- Genetic recombinations or crossovers between parental haplotypes occurs at a frequency of around 0.5% and contributes significantly significantly to diversity of HLA loci in humans. See Figure 1. It is rare for any 2 unrelated individuals to have completely matched sets of identical HLA genes. (p.175–177)
- QQ: What does the 0.5% designates? How rare is it for any 2 unrelated individuals to have completely matched sets of identical HLA genes? (p.175–177)
- QQ: Remind me what a phenotype is. (p.177)
- Class I MHC molecules contain a large α chain associated noncovalently with the much smaller β_2-microglobulin. The α chain is a polymorphic transmembrane glycoprotein encoded by genes within the A, B, and C regions of the HLA complex while the β_2-microglobulin is a protein encoded by a gene located on a different chromosome. Association of the α chain with β_2-microglobulin is required for expression of a class I molecules on cell membranes. (p.178)
- The α chain is organized into: (p.178–179)
  - 3 external domains (α1, α2, and α3), each containing approximately 90 amino acids;
  - a transmembrane domain of about 25 hydrophobic amino acids followed by;
  - a shorter stretch of charged (hydrophilic) animo acids;

  - and a cytoplasmic anchor segment of 30 amino acids.

- The $\alpha 1$ and $\alpha 2$ domains organize as a deep groove, or cleft, approx. 25 Å × 10 Å × 11 Å, with 2 long α helices as sides and a β sheet of 8 antiparallel β strands as the bottom. This peptide-binding cleft is large enough to bind a peptide of 8–10 amino acids. (p.179)

- The $\alpha 3$ and β_2-microglobulin's are each organized inot 2 β pleated sheets each formed by antiparallel β strands of amino acids. This structure, known as the immunoglobulin fold is the reason why $\alpha 3$ and β_2-microglobulin are said to be members of the immunoglobulin superfamily. (p.179)

- The $\alpha 3$ domain appears to be highly conserved among class I MHC molecules and contains a sequence that interacts with the CD8 membrane molecule present on T_C cells. (p.179)

- The assembly of class I molecules is believed to occur by the initial interaction of the β_2-microglobulin with the folding class I α chain. This metastable "empty" dimer is then stabilized by the binding of an appropriate peptide to form the native trimeric class I structure consisting of the α chain, β_2-microglobulin, and a peptide. This complete molecular complex is ultimately transported to the cell surface. In the absence of β_2-microglobulin, the class I MHC α chain is not expressed on the cell membrane. (p.179–180)

- QQ: what happens to a MHC class I missing β_2-microglobulin? (p.180)

- QQ: So class I MHC is never displayed empty, i.e. without a peptide on it? (p.180)

- Class II MHC molecules contain two different polypeptide chains, an α chain and a β chain, which associate by noncovalent interaction. They also contain external domains, a transmembrane segment, and a cytoplasmic anchor segment. Each chain contains two domains, the $\alpha 1$ and $\beta 1$ domains are membrane-distal and are analogous in shape to the $\alpha 1$ and $\alpha 2$ of the MHC I and also form a similar cleft, while the $\alpha 2$ and $\beta 2$ domains are membrane-proximal and have structures analogous in shape to the $\alpha 3$ and β_2-microglobulin of the MHC I and also form immunoglobulin folds, making them members of the immunoglobulin superfamily. (p.180)

- QQ: explain the dimer of dimers story (p.180)

- Any one individual expresses only up to 6 different class I molecules and up to 12 different class II molecules, and yet this limited number of MHC molecules must be able to present an enormous array of different antigenic peptides. Thus a given MHC molecule can bind numerous different peptides, and some peptides can bind to several different MHC molecules. (p.181)

- Details of peptide binding by class I and class II MHC molecules is presented in Table 3. (p.181–182)

- The dissociation constant K_D of the peptide-MHC molecule complex is approximately 10^{-6}; the rate of association is slow but the rate of dissociation is even slower. In other words, the peptide-MHC molecule association is very stable under physiologic conditions such that most of the MHC molecules expressed on the membrane will be associated with a peptide of self or nonself origin. (p.183)

- Class I MHC molecules bind peptides derived from endogenous intracellular proteins, which are digested into peptides within the cytosol and present to $CD8^+$ T cells. Each type of class I MHC molecule and additionally each allelic variant of a class I MHC molecule binds a distinct set of peptides. Because a single nucleated cell expresses about 10^5 copies of each classs I molecule, then many different peptides will be expressed simultaneously on the surface of the cell by class I MHC molecules. (p.183)

- QQ: What is meant by the above? How does the 10^5 comes about? (p.183)

- The presented self-peptides are derived from common intracellular proteins such as cytochrome c, histones, and ribosomal proteins. In an altered self-cell, some of the self-peptides are replaced with peptides from viral proteins. (p.183)

	Class I molecules	Class II molecules
Peptide binding domain	$\alpha 1/\alpha 2$	$\alpha 1/\beta 1$
Nature of the peptide cleft	Closed at both ends	Open at both ends
General size of bound peptide	8–10 amino acids	13–18 amino acids
Peptide motifs involved in binding to MHC molecule	Anchor residues at both ends of peptide; generally hydrophobic carboxyl-terminal anchor	Anchor residues distributed along the length of the peptide
Nature of bound peptide	Extended structure in which both ends interact with MHC cleft but middle arches up away from MHC molecule	Extended structure that is held at a constant elevation above the floor of the MHC cleft

Table 3: Peptide binding by class I and class II MHC molecules. (p.182, Table 7-2)

- QQ: What is cytochrome, histone, rhibosomal proteins? Are they expandable? What other than a virus would get expressed as a non-self peptide? (p.183)
- QQ: I do not understand their calculation with 2000 peptides and 100-4000 copies. (p.183)
- The nonamers peptides (3 amino acids long) are the most frequently isolated peptide size from class I MHC molecules since they have a 100- to 1000-fold higher affinity than peptides of size 8 or 10. This suggests that 9 it is the cleft which is closed at both ends that places restrictions on the peptide size. (p.183)
- The majority of contacts between class I MHC molecules and peptides involve residue 2 at the N-terminal and residue 9 at the C-terminal of the nonameric peptide, such that the bound peptide assumes an extended structure, interacting with the MHC cleft at both ends but arching away from the floor of the cleft in the middle, permitting peptide diversity in between without compromising binding. It is thought that the amino acids that arch away from the MHC are more exposed and therefore interact directly with the TCR. (p.184)
- Class II molecules generally bind self or non-self exogenous membrane-bound proteins presumably internalized by phagocytosis or proteins associated with the vesicles of the endocytic processing pathway and presented to $CD4^+$ T cells. Almost all class II MHC molecules investigated so far have been shown to bind the same invariant-chain peptide. (p.185)
- QQ: Almost all class II MHC molecules investigated so far have been shown to bind the same invariant-chain peptide. What is meant by this? (p.185)
- Because the peptide-binding cleft in class II is open at both ends, it allows a bound peptide to extend beyond the ends (typicall 13–18 amino acids). Despite this, binding characteristics are determined by a central core of 13 amino acids residues. The peptides that bind to a particular class II molecule often have internal conserved "motifs", typically a sequence of 7–10 amino acids that provide the major contact points, instead of conserved anchor residues. Generally, this sequence has an aromatic or hydrophobic residue at the N-terminal and 3 additional hydrophobic residues in the middle portion and C-terminal end of the peptide. Aditionally, over 30% of peptides contain a proline residue at position 2 and another cluster of proline at the C-terminal end. (p.185–186)
- While diversity of antibodies is a dynamical process that changing overtime within an individual, MHC diversity does not change in time in a given individual but it may differ significantly from that of another individual. The diversity of the MHC within a species stems from polymorphism (the presence of multiple alleles at a given genetic locus) and the presence of duplicated genes with similar or overlapping functions, i.e. polygenic. (p.186)

- QQ: What is the difference between different alleles of a same genes and duplicated genes with similar functions? (p.186)
- Analysis of human HLA of European descent has revealed that class I molecules have approximately 60 A alleles, 110 B alleles, and 40 C alleles while the class II HLA-DR beta-chain genes may vary from 2 to 5 in different haplotypes and have approximatels 122 alleles of DRB and only 1 DRA chain.
- HLA haplotypes in humans should contain completely random combinations of the multiple alleles, but it is not the case, i.e. certain allelic combinations occur more frequently: this is called *linkage disequilibrium* and could be caused by:
  - not sufficient time or number of generations have elapsed to have reached equilibrium (p.186)
  - selective effects would prevale when combinations yield resistance to certain diseases or generate harmful effects (p.187)
  - crossover may be more frequent in certain DNA sequence regions (hotspots) then in others. (p.187)
- The polymorphic amino acids of class I and II MHC molecules are clustered in the membrane-distal $\alpha 1/\alpha 2$ and $\alpha 1/\beta 1$ domain respectively. This means that allelic differences contribute to differences in the ability of MHC molecules to interact with a given antigenic peptide. (p.188)
- Classical genes in class I and II form the class I and II MHC molecule that we know while the nonclassical genes may sometimes form MHC-like molecules that can present very specific things to T cells (e.g. H-2M-encoded class I molecule is uniquely suited to present peptides from prokaryotic organisms that are able to grow intracellularly). (p.189–190)
- Generally, the classical class I MHC molecules are expressed on most somatic cells, but the level of expression may vary. For example, the highest levels of class I molecules are expressed by lymphocytes, where they constitute approximately 1% of the total plasma-membrane proteins, or some 5×10^5 molecules. (p.190)
- Since the MHC alleles are codominantly expressed, a heterozygous individual expresses on its cells the gene products encoded by both alleles at each MHC locus. In other words, a heterozygous human individual expresses the A, B, and C alleles from each parent, i.e. 6 different class I MHC molecules on the membrane of each nucleated cell. (p.190)
- QQ: difference between nucleated and somatic cell? (p.190)
- A single virus-infected cell should be envisioned as having various class I molecules on its membrane, each displaying different sets of viral peptides. Consequently, different individuals within a species will have the ability to bind different sets of viral peptides. (p.190)
- Class II molecules are expressed only by antigen-presenting cells which are macrophages, dendritic cells and B cells; thymic epithelial cells and some other cells can be induced to express it. Further, expression in some cases depends on the differentiation stage (e.g. not expressed on pre-B cells but present on mature B cells, expressed only by monocytes and macrophages that have interacted with an antigen) (p.191)
- Because each of the classical class II MHC molecules is composed of two different polypeptide chains, which are encoded by different loci, a heterozygous individual expresses not only the parental class II molecules but also molecules containing the α and β chains from different chromosomes. Since the human MHC contains 3 classical class II genes (DP, DQ, and DR), a heterozygous individual expresses 6 parental class II molecules and 6 molecules containing α and β chain combinations from either parent. The number of different class II molecules expressed by an individual is further increased by the presence of multiple α- and β-chain genes in humans. (p.191)

Gene	Chain type	V	D	J	C
α	heavy	50	–	70	1
δ	heavy	3	3	3	1
β	light	57	2	13	2
γ	light	14	–	5	2

Table 4: TCR multigene families in humans. (p.220, Table 9-1)

- QQ: I don't understand their calculations! (p.191)
- Regulation mechanisms of MHC expression by cytokines:
  - interferon gamma (IFN-γ) up-regulates class I transcription and II expression on a variety of cells including non-antigen-presenting cells. (p.191)
  - IL-4 increases increases expression of class II molecules by resting B cells. (p.191)
  - IFN-γ down-regulates expression of class II molecules by B cells. (p.191)
  - corticosteroids and prostaglandines also decrease expression of class II molecules. (p.191)
- QQ: confusion about the effect of IFN-γ? (p.191)

5 TCR (Kuby, Chapter 9)

- The molecule responsible for T-cell specificity was found to be a heterodimer composed of either α and β or γ and δ chains. The majority of T cells express the $\alpha\beta$ heterodimer; the remaining express the $\gamma\delta$ heterodimer; the exact proportion vary among organs and species. (p.217)
- A cell that expresses TCR has approximately 10^5 TCR molecules on its surface. (p.217)
- Each chain in a TCR has two domains containing an intrachain disulfide bond that spans 60-75 amino acids. The amino-terminal domain in both chains exhibits marked sequence variation, but the sequences of the remainder of each chain are conserved. Thus the TCR domains — one variable (V) and one constant (C) — are structurally homologous to the V and C domains of immunoglobulins, and the TCR molecule resembles an Fab fragment. (p.217)
- The TCR variable domains have three hypervariable regions, which appear to be equivalent to the complementarity determining regions (CDRs) in immunoglobulin light and heavy chains. There is an additional area of hypervariability (HV4) in the β chain that does not normally contact antigen and therefore is not considered a CDR. (p.217–218)
- Functional TCR genes are produced by rearrangements of V and J segments in the α-chain and γ-chain families and V, D, and J segments in the β-chain and δ-chain families. See Table 4. (p.219)
- The α chain, like the immunoglobulin L chain, is encoded by V, J, and C gene segments. The β chain, like the immunoglobulin H chain, is encoded by V, D, J, and C gene segments. Rearrangement of the TCR α- and β-chain gene segments results in VJ joining for the α chain and VDJ joining for the β chain. See Table 4. (p.220)
- TCR α-chain DNA has two C gene segments, but their protein products differ by only a few amino acids and have no known fuctional difference. (p.220)
- Rearrangement of the TCR β-chain genes exhibits allelic exclusion. Allelic exclusion appears to be less stringent for the TCR α-chain genes.

- Since allelic exclusion is not complete for the TCR α chain, there are rare occasions when more than one α chain, there are rare occasions when more than one α chain is expressed on the membrane of a given T cell. (p.224)
- In TCR genes, combinatorial joining of V gene segments appears to generate CDR1 and CDR2, whereas junctional flexibility and N-region nucleotide addition generate CDR3. (p.224)
- It is possible for a V_β gene segment to join directly with a J_β or a D_β gene segment, generating a $(VJ)_\beta$ or a $(VDJ)_\beta$ unit. Alternative joining of δ-chain gene segments generates similar units; in addition, one D_δ can join with another, yielding $(VDDJ)_\delta$ and, in humans, $(VDDDJ)_\delta$. (p.226)
- Whereas the addition of N-region nucleotides occurs only in the Ig heavy-chain genes, it occurs in the genes encoding all the TCR chains. As many as 6 nucleotides can be added by this mechanism at each junction, generating up to 5461 possible combinations, assuming random selection of nucleotides. (p.226)
- Estimates suggest that the combined effects of P- and N-region nucleotide addition and joining flexibility can generate as many as 10^{13} possible amino acid sequences in the TCR junctional regions alone. (p.226)
- In contrast to the limited diversity of CDR1 and CDR2, the CDR3 of the TCR has even greater diversity that that seen in immunoglonulins. Diversity in the CDR3 region of the TCR is generated by junctional diversity in the joining of V, D, and J segments, joining of multiple D gene segments and the introduction of P and N nucleotides at the V-D-J and V-J junctions. (p.226)
- Unlike the Ig genes, the TCR genes do not appear to undergo extensive somatic mutation. The absence of somatic mutation in T cells ensures that T-cell specificity does not change after thymic selection and therefore reduces the possibility that random mutation might generate a self-reactive T cell. (p.226)
- Membrane-bound immunoglobulin on B cells associates with another membrane protein, the Ig-α/Ig-β heterodimer, to form the B-cell antigen receptor. Similarily, the T-cell receptor associates with CD3, forming the TCR-CD3 membrane complex. In both cases, the accessory molecule participates in signal transduction after interaction of a B or T cell with antigen it does not influence interaction with the antige. (p.226)
- $CD4^+$ T cells recognize antigen that is combined with class II MHC molecules and function largely as helper cells, whereas $CD8^+$ T cells recognize antigen that is combined with class I MHC molecules. They function largely as cytotoxic cells. (p.228)
- Both CD4 and CD8 have two key abilities that allow these membrane molecules to be classified as coreceptors: recognition of the peptide-MHC complex and signal transduction. (p.228)
- The affinity of T-cell receptors for peptide-MHC complex is low to moderate, with K_d values ranging from 10^{-4} to 10^{-7} M. Thus, this interaction is weak compared with antigen-antibody interactions, which generally have K_d values ranging from 10^{-6} to 10^{-10} M. (p.230)
- No other molecules than TCR is required for recognition, at least not for binding of TCR to a peptide-class I complex. While the accessory molecules such as CD3, CD4, CD8, etc. may enhance that binding, they are not necessary for it. (p.231)
- TCR by itself can recognize and bind the specific antigen presented by the MHC class I molecule. (p.232)
- CDR1 of the TCR β chain may contact the same part of the peptide that the MHC does. Therefore, the use of the terms epitope for the part of the peptide that reacts with the TCR and agretope for the part that reacts with MHC may need reexamination. (p.234)

- T cells respond not only to complexes of a foreign antigenic peptide plus self-MHC molecule but also to foreign MHC molecules (histocompatibility antigens) alone. This T-cell response leads to rejection of allogeneic grafts. Some evidence suggests that this alloreactivity results from the ability of T cells specific for an antigenic peptide plus self-MHC molecule to cross-react with various allogenic peptide-allogeneic MHC complexes. (p.236)

6 Cell biology basics (John, Feb. 13, 2003)

- Nucleotides are the building blocks of nucleic acid. They come in 4 types: G, C, A, and T. They are complementary in the following way: A-T and G-C.
- Each nucleic acid unit is made of a base (a nucleotide G, C, A or T), a sugar and a phosphate. Units link at the OH group of their sugar to form a hydrogen bound. (see [1] p.58) FIXME: add picture.
- A nucleic acid is the name given to a string of nucleotides.
- A codon is the name given to a group of 3 nucleotides that are to form an amino acid in RNA.
- It takes 3 nucleotides to make 1 amino acid. However, there are not $4^3 = 64$ different amino acids but rather only 20, see Table 5. This is due to repetition, i.e. a few groups of 3 nucleotides can yield the same amino acid. Further, some codons do not yield a protein, rather they indicate the end of a reading sequence, those are called stop codons.
- Amino acids with charged (basic or acidic) side chains are very polar and are nearely always found in the outside of protein molecules. The charge depends on the pH of the solution they are in. At pH=7 side-chains of all 5 basic and acidic amino acids are charged. See Table 6.
- Amino acids with uncharged polar side chains are relatively hydrophylic and are usually on the outside of proteins. See Table 6.
- Amino acids with non polar side chain are hydrophobic and tend to cluster together on the inside of proteins. See Table 6.
- The isoelectric point is the point where an amino acid has a net charge of zero.
- The 5' end and 3' end are the names by which one refers to the start and the end of the nucleic acid respectively. They are linked to the numbering of carbon joints in the sugar of a given nucleic acid unit. The 5' is the carbon join where the phosphate attaches while the 3' is the joint where the OH group is located i.e. where the phosphate of the next nucleic acid unit links and therefore is the growing end of the chain.
- Enzyme: typically a protein that facilitates cutting/pasting.
- Electrophoresis: the process through which one separates proteins, DNA or RNA by size or charge.
- Protein: a repetitive structure of amino acids. FIXME: add picture.
- Glycoprotein: a protein with carbohydrates attached to its side chains.
- Glycosylation: the process through which carbohydrates are attached to a protein's side chains.

First	Second				Third
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	*stop*	*stop*	A
	Leu	Ser	*stop*	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Table 5: Sequences of codon structure for all 20 amino acids and stop codon. Note that AUG can either be Methionine (Met) or a start codon.

7 Calculation for optimal epitope size [7]

In this article, the authors calculate the optimal epitope size, r, i.e. the number of complementary units (amino acids) on receptors and antigens that need to interact to generate the minimal affinity needed for B or T cells activation.

They start by calculating $Pr(N, N'; n)$, the probability that a receptor repertoire of size n has the property that all of N foreign antigens are recognized by at least one receptor in the repertoire but that none of N' self-antigens are recognized. Then r is such that $Pr(N, N'; n)$ is maximized. Let P_S be the probability that a random receptor recognizes a random antigen, so that the probability that a random receptor fails to recognize a random antigen is $P_F = 1 - P_S$, then

$$Pr(N, N'; n) = (1 - {P_F}^n)^N \cdot ({P_F}^n)^{N'}$$

and the value of P_F that maximizes $Pr(N, N'; n)$ is

$$\begin{aligned}
\frac{\mathrm{d}Pr}{\mathrm{d}P_F} &= 0 &&= N\left(1-{P_F}^n\right)^{N-1} \cdot \left(-n{P_F}^{n-1}\right) \cdot {P_F}^{nN'} + \left(1-{P_F}^n\right)^N nN'{P_F}^{nN'-1} \\
\frac{nN\left(1-{P_F}^n\right)^N {P_F}^n {P_F}^{nN'}}{P_F\left(1-{P_F}^n\right)} &= &&\frac{nN'\left(1-{P_F}^n\right)^N {P_F}^{nN'}}{P_F} \\
\frac{N{P_F}^n}{\left(1-{P_F}^n\right)} &= &&N' \\
\frac{N}{N'} &= &&\frac{1}{{P_F}^n} - 1 \\
P_F &= &&\left(1+\frac{N}{N'}\right)^{-1/n} \approx \; 1 - \frac{1}{n}\ln\left(1+\frac{N}{N'}\right),
\end{aligned}$$

where the approximation used, i.e. $x \approx 1 + \ln x$ holds for $x \approx 1$. In our case, $x = (1 - N/N')^{-1/n}$ and because $N >> N'$ and $n >> 1$, then $x \approx (N'/N)^{1/n} \approx 1$. This means that the probability that a random receptor

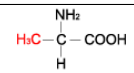
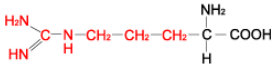
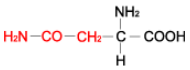
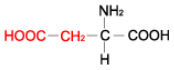
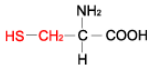
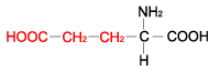
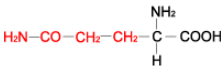
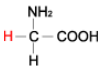
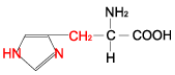
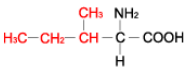
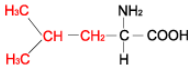
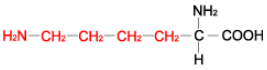
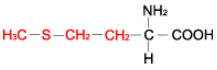
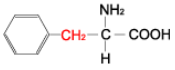
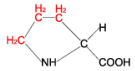
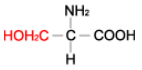
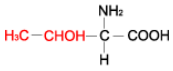
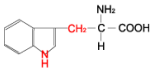
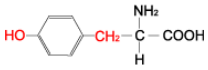
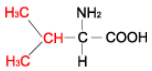
Name	Short	Type	Iso pt	Structure
Alanine	Ala (A)	nonpolar	6.00	
Arginine	Arg (R)	charged polar (basic)	11.15	
Asparagine	Asn (N)	uncharged polar	5.41	
Aspartic acid	Asp (D)	charged polar (acidic)	2.77	
Cysteine	Cys (C)	uncharged polar	5.02	
Glutamic acid	Glu (E)	charged polar (acidic)	3.22	
Glutamine	Gln (Q)	uncharged polar	5.65	
Glycine	Gly (G)	nonpolar	5.97	
Histidine	His (H)	charged polar (basic)	7.47	
Isoleucine	Ile (I)	nonpolar	5.94	
Leucine	Leu (L)	nonpolar	5.98	
Lysine	Lys (K)	charged polar (basic)	9.59	
Methionine	Met (M)	nonpolar	5.74	
Phenylalanine	Phe (F)	nonpolar	5.48	
Proline	Pro (P)	nonpolar	6.30	
Serine	Ser (S)	uncharged polar	5.68	
Threonine	Thr (T)	uncharged polar	5.64	
Tryptophan	Trp (W)	nonpolar	5.98	
Tyrosine	Tyr (Y)	uncharged polar	5.66	
Valine	Val (V)	nonpolar	5.96	

Table 6: Names, abreviation, charge type, and structure of all 20 amino acids.

recognizes a random antigen is

$$P_S = 1 - P_F \approx \frac{1}{n} \ln\left(1 + \frac{N}{N'}\right).$$

The authors go on to use values of $n = 10^7$, $N = 10^{16}$, and $N' = 10^6$ for mice lymphocytes and find that their resulting $P_S = 2.3 \times 10^{-6}$ is smaller than the observed 13×10^{-6}. While the logarithmic nature of P_S renders it insensitive N and N', this calculation should be rechecked for as precise as possible an n value.

Then using an m letter alphabet to represent amino acids, he goes on to estimate r the minimum length of contiguous complementary residues in a string of length l needed for interaction to generate minimal affinity for activation. Using algebraeic concerns, he determines that

$$P_S = m^{-r}\left[(l-r)\frac{(m-1)}{m} + 1\right],$$

and assuming that $l >> r > 1$, he finds that

$$\begin{aligned} r &= -\ln_m P_S + \ln_m\left[\frac{l(m-1)}{m}\right] \\ r &= \ln_m(nl) + \ln_m\left[\frac{(m-1)}{m}\ln\left(1 + \frac{N}{N'}\right)\right]. \end{aligned}$$

Those 2 different expressions for r are quite insensitive to changes in l and the ratio N/N' but very sensitive to changes in m. They estimate m using the following method. Taking $r = 15$ and since there are only 20 amino acids, then there are 20^{15} possible epitopes. If the repertoire is complete, then it contains 10^7 receptors which must be able to recognize 20^{15} epitopes or each receptor should be able to recognize $20^{15}/10^7 = 3.2 \times 10^{12}$ epitopes. Further, since each amino acid in a receptor can on average bind to $20/m$ amino acids on the antigen, if the epitope is of length r, then each receptor can on average bind to $(20/m)^{15}$ epitope which should be equal to 3.2×10^{12}, yielding $m = 2.9$ or 3 since m has to be an integer. This tells us that an alphabet of m letters best represent the various amino acid types and in the case of $m = 3$ those classes correspond to negative, positive and neutral amino acids, where negative complements positive and neutral complements neutral. Note that binding would be asymmetric if negative, positive and neutral amino acids occur in equal proportions.

8 Questions

8.1 Kuby, Chapter 4 [3]

- What is *antigenic determinent*, *monoclonal or polyclonal antibodies* and *serum*? (p.83)
- What are the α, β and γ fractions of the serum? (p.84)
- What implications do "glycoproteins, carbohydrates and glycosylation" have? (p.85)
- What does "immunized by fragments of rabbit IgG" mean? (p.85)
- What is the Δ light chain associated with IgD? (p.87)
- What would be a multiple of the 4-chain structure? (p.87)
- Explain how to read ribbon diagrams (p.88)
- Explain conformation of domains and intrachain loops of 60 amino acids (p.87 and Fig 4-2b p.84)
- Explain hydrophobic/phylic (p.89) structure within ribbon.
- Explain alleles of isotype genes. (p.103)
- Do residues refer to amino acids found in random conformation occuring between tertiary structures. (p.104)

8.2 Kuby, Chapter 5 [3]

- What is germ-line DNA ? a germ cell? vs functional genes? (p.115)
- Usually the N-terminal is hidden but they say that for Ig, variable region is in N-terminal. Is it then not hidden in Ig? (p.116)
- What is a germ cell vs somatic cell? (p.117)
- Please explain DNA, genes, RNA, mRNA, chromosomes, genes, restriction-endonuclease, 5', cloning, sequencing, recombination? (p.118)
- Explain the following process. Each V gene segment is preceded at its 5' end by a small exon that encodes a short signal or leader (L) peptide that guides the heavy or light chain through the endoplasmic reticulum. The signal peptide is cleaved from the nascent lignt and heavy chains before assembly of the finished immunoglobulin molecule. (p.119)
- What is in-between blocks (separators of 7 kilobases for example)? (p.119)
- What are noncoding genes for? (p.119)
- What is a nucleotide codon sequence and how can it be compared to an amino acid sequence? (p.119)
- What is a base pair? (p.119)
- How is it that 39-base pair codons correspond to 13 amino acids? (p.119)
- What is a pseudogene? What is its use? (p.119)
- NEW!!! Are the 4^{th} J_λ and C_λ gene segments pseudogenes in humans too? (p.120)
- NEW!!! Is there 1 pseudogene among the J_κ gene segments in humans too? (p.120)
- What is a coding exon and a noncoding intron? (p.120)
- Is the leader sequence what causes the polypeptide to be pulled into the lumen of the rough endoplasmic reticulum? (p.121)
- What is an enzyme? (p.123)
- Explain the whole process of enzymatic joining of gene segments (p.123)
- How are "in-phase" and "out-of-phase" joining defined? Is it just characterized by the random occurence of a stop signal before the end of the gene in the newly joined sequence? (p.126)
- What is an allele? (p.126)
- Wha is a chromosome? What does diploid mean? Where are your parents' chromosomes kept? yours? (p.127)
- NEW!!! Why does recombination after obtention of a μ heavy-chain tries κ before λ? Why is κ favored? Is this only in mice or in humans too? Reason? (p.127–128)
- Why can't they calculate the contribution of each of the 7 factors of diversity? (p.128)
- Why can any given individual's Ig loci contain a different number of a particular type of gene segment (e.g. 21 C and 13 V)? (p.128)
- What is junctional flexibility, P- and N-nucleotide addition? (p.128)
- Does affinity maturation only take place within germinal centers? What are germinal centers? (p.131)

- This seems to suggest that B-T cell interactions form germinal centers: germinal centers, structures that form in secondary lymphoid organs within a week or so of immunization with an antigen that activates a T-cell-dependent B-cell response. (p.131)
- Can an Ig not have a C_H/C_L? Does the joining between V-C same as class switching? Does it carry all Cs until switching? (p.132–133)
- Can memory cells and virgin cells produce secreated Ig? (p.135)
- Is there a use to being able to produce IgM and IgD? (p.135)
- What's in a B-cell? Draw diagram? Is there golgi? RER? (p.137)

8.3 Kuby, Chapter 9 [3]

- Are the $\alpha - \beta$ heterodimers the same as those that make up tubulin? (p.215)
- Unlike the antibody molecule, which is bivalent, the TCR is monovalent. What does this mean? (p.217)
- If α does not have allelic exclusion, is it the same with γ? (p.224)
- Is this statement true? How much do we know about the following statement? No other molecules than TCR is required for recognition, at least not for binding of TCR to a peptide-class I complex. While the accessory molecules such as CD3, CD4, CD8, etc. may enhance that binding, they are not necessary for it. (p.231)

8.4 Jerne 1974 [6]

- Is it true that one cell makes only 1 antibody? (p.376)
- About IgM and IgD, etc? the one cell law refering only to V-genes? (p.376)
- Are there non-committed stem cells? When do they commit? (p.376)
- What is allelic exclusion (p.376)
- Explain (D) (p.377)
- How different in size are the repertoire of receptors and epitopes? (p.381)
- Can an idiotope (epitope) recognize a paratope (receptor)? (p.382)
- "Identical paratopes on 2 antibody molecules does not necessarily imply that these molecules present identical idiotopes." What is meant? Is it true? (p.382)
- What accounts for the duality of response (negative vs positive)? Is their explaination accurate? Describe in details (p.382)
- Is the concentration of individual paratopes and idiotopes sufficient to establish an internal functional network? (p.383)
- Does the network theory makes sense in the light of autoimmunity? (p.384)
- Can an antigen remove antibodies? (p.384)
- Have all problems (1)-(8) been answered? (p.384)
- Can't immune memory be transmitted to our offsprings? (p.387)

8.5 Percus, Percus and Perelson 1993 [7]

- What is the size of the binding regions on immunoglobulin and T-cells? Is it really 15? What about the differences between T-cell and antibodies? (p.1691)

8.6 Hightower, Forrest, and Perelson 1995 [4]

- Remind me the difference between phenotype and genotype? (p.1)
- They are using 4 equal size libraries, this should not be correct! (p.2)
- Is the way they define fitness appropriate? (p.2)
- Do they say how individuals get selected? How is pressure applied?

8.7 Hightower, Forrest, and Perelson 1996 [5]

- What is Lamarckian inheritance? (p.1)

8.8 Smith, Forrest, Perelson et al. 2001 [8]

- No questions

8.9 Borghans, Noest, and De Boer 1999 [2]

- What is the second signal delivered by APCs? (p.569)
- Is it true that the innate immune system prevides signals about the context of antigenic epitopes depending on 1) the organ where the epitope is detected, 2) the presence of conserved pathogen-associated molecular patterns, and perhaps 3) tissue damage and thus signal whether the Ag should be attacked or tolerated? (p.569)
- Suggests that the memory retains type of immune response to be made as well as quantitative form in terms of precursor frequencies. (p.570)
- What is physiologically different in lymphocytes having a second encounter with Ag that allows them to no longer wait for instructions from the innate system? (p.570)
- Why can't LCMV-specific response be induced by LCMV infection in LCMV-transgenic mice that had been tolerized with LCMV peptides? (p.570)
- Here are the equations:

$$\begin{aligned} P_i &= 1-(1-p)^R \\ R &= R_0(1-p)^{fS} \end{aligned}$$

where p is prob to respond to a random epitope, R_0 is # of different lymphocytes clones, f is fraction of self epitopes S that induce self tolerence, R is functional repertoire after tolerence induction, i.e. number of lymphocytes that do not respond to any of the fS self tolerizing epitopes, P_i is the prob. of mounting an immune reponse. For complete self tolerance, i.e. $f = 1$, the optimal p is $p = 1/fS = 1/S$. For ignored self we have:

$$\begin{aligned} P_S &= P_t-(1-p)^R \\ P_t &= (1-p\alpha)^R \\ \alpha &= 1-(1-p)^{(1-f)S} \end{aligned}$$

where α is the fraction of potentially autoreactive clones in functional repertoire, $p\alpha$ is the fraction of truly autoaggressive clones in functional repertoire responding to pathogens, P_t is the probability of remaining self tolerant, and P_S is the probability to mount an immune response but to remain tolerant to ignored self.

9 Project ideas, testing models

- Build antigen, TCR and antigen and calculate values found in [7] through statistical analysis of several runs of simulation.
- Redo work presented in [4] using the right number of libraries and their right size.

References

[1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson. *Biologie moléculaire de la cellule, 3ième édition*. Médecine-Sciences / Flammarion, Paris, 1995.

[2] J. A. M. Borghans, A. J. Noest, and R. J. d. Boer. How specific should immunological memory be? *J. Immunol.*, 163(2):569–575, 15 July 1999.

[3] R. A. Goldsby, T. J. Kindt, and B. A. Osborne. *Kuby Immunology* 4^{th} *Edition*. W. H. Freeman and Company, New York, 2000.

[4] R. R. Hightower, S. Forrest, and A. S. Perelson. The evolution of emergent organization in immune system gene libraries. In L. J. Eshelman, editor, *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 344–350, San Francisco, CA, 1995. Morgan Kaufmann.

[5] R. R. Hightower, S. Forrest, and A. S. Perelson. The Baldwin effect in the immune system: Learning by somatic hypermutation. In R. K. Belew and M. Mitchell, editors, *Adaptive Individuals in Evolving Populations*, pages 159–167. Addison-Wesley, Reading, MA, 1996.

[6] N. K. Jerne. Towards a network theory of the immune system. *Ann. Inst. Pasteur Imm.*, 125 C:373–389, 1974.

[7] J. K. Percus, O. E. Percus, and A. S. Perelson. Predicting the size of the T-cell receptor and antibody combining region from consideration of efficient self-nonself discrimination. *Proc. Natl. Acad. Sci. USA*, 90:1691–1695, 1993.

[8] D. J. Smith, A. S. Lapedes, S. Forrest, J. C. deJong, A. D. M. E. Osterhaus, R. A. M. Fouchier, N. J. Cox, and A. S. Perelson. Modeling the effects of updating the influenza vaccine on the efficacy of repeated vaccination. In A. D. M. E. Osterhaus, N. J. Cox, and A. W. Hampson, editors, *Options for the Control of Influenza Virus IV*, volume 1219 of *International Congress Series*, pages 655–660, Amsterdam, 2001. Excerpta Medica (Elsevier). Proceedings of the World Congress on Options for the Control of Influenza Virus IV, Crete, Greece, 23–28 September 2000.