DNA METHODS FOR HLA TYPING

A WORKBOOK FOR BEGINNERS

Carolyn Katovich Hurley, Ph.D.
C. W. Bill Young Marrow Donor Recruitment and Research Program
Department of Oncology, Georgetown University School of Medicine,
Washington, DC 20057

Version 7

PURPOSE OF THE MANUAL

The purpose of this manual is to provide the reader with the basic background needed to understand the HLA system and the molecular biology methods used to identify HLA "types". It is assumed that the reader has a college education and has taken courses in basic biology and biochemistry. The best results will be obtained if the reader starts at the beginning of the manual and follows all of the instructions in the manual. The manual is meant to be a workbook and spaces for answers are provided in the text. Answers are provided at the end of the manual. Readers are encouraged to discuss their questions with their laboratory director.

TABLE OF CONTENTS

Reference material:

Current publications and web site listing all class I and II nucleotide sequences and nomenclature. The following were current when this manual was written:

Marsh, S.G.E., et al. Nomenclature for factors of the HLA system, 2004. Tissue Antigens 65:301- 368, 2005.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P., Marsh, S.G.E. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Research 31:311-314, 2003.

HLA nomenclature websites:

http://www.anthonynolan.org.uk/HIG/
http://www.ebi.ac.uk/imgt/hla/

OVERVIEW OF HLA TYPING

Hematopoietic stem cells found in the bone marrow differentiate to become blood cells. These cells play important roles in fighting infection (white blood cells or lymphocytes), in transporting oxygen to the tissues (red blood cells), and in causing blood clotting (platelets). Certain types of cancer affect these stem cells and can be fatal to the patient. One therapy for the treatment of these blood diseases like leukemia and aplastic anemia is bone marrow (or hematopoietic stem cell) transplantation. In this procedure, the abnormal bone marrow of the patient is destroyed by irradiation and chemotherapy. The patient (recipient) is then transfused with stem cells from another individual (donor). The transfused cells travel to the cavities in the bones and grow there forming new blood cells.

In order for hematopoietic stem cell transplantation to be successful, the patient and the donor of the stem cells must be matched for a group of proteins called HLA molecules (or HLA antigens). This manual describes HLA molecules and discusses the techniques that are used to determine the HLA "types" of the patient and potential donors. If the patient and a potential donor have the same HLA type, the transplant has a good chance of being successful. If a patient and a donor are HLA mismatched, the donor's stem cells may be destroyed by the patient's immune system (graft rejection) or the immune system cells in the donor's stem cell preparation may attempt to destroy the patient's own cells (graft vs. host disease). These events can result in major complications or death of the patient.

This manual is designed to be used by readers who will participate in using molecular biology techniques to identify HLA types. It begins with a description of some basic molecular biology concepts.

CHAPTER 1

GENERAL CONCEPTS IN MOLECULAR BIOLOGY

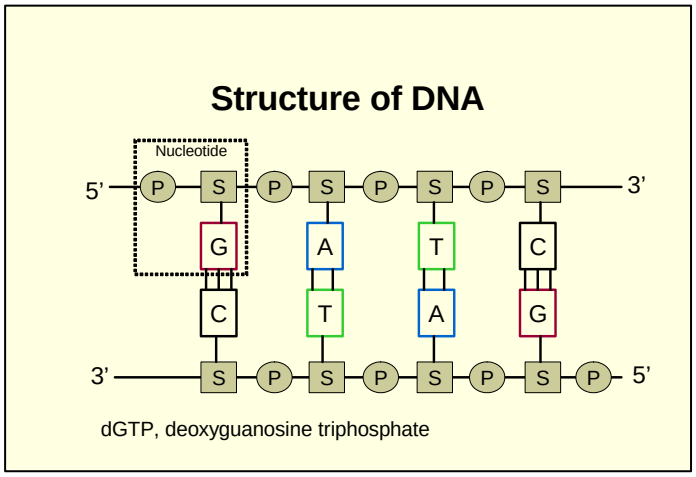
The purpose of this chapter is to describe basic principles regarding DNA and to describe the methods used to study DNA.

I. Structure of DNA and RNA

A. DNA (deoxyribonucleic acid) is composed of a phosphate and sugar (deoxyribose) backbone attached to bases (Adenine, Thymine, Cytosine, Guanine) [Figure 1-1]. The combination of a base, a sugar, and a phosphate group is called a nucleotide (e.g., dATP is an abbreviation of deoxyadenosine triphosphate; dNTP is an abbreviation which means any nucleotide). Nucleotides are the basic building blocks of DNA.

B. Nucleotides are linked together to form a linear chain of nucleotides. One end of this single DNA strand is called 5'; the other end is called 3'. The two ends have different structures. It is common practice to put the 5' nucleotide on the left side of the page when writing down a sequence of nucleotides (e.g., 5' TAAGGCT 3').

Figure 1-1



C. Two strands of DNA form a ladder (or double helix) by base pairing (G-C and A-T). The bases are paired through hydrogen bonds. G-C pairs form 3 hydrogen bonds while A-T pairs form 2 hydrogen bonds. This difference means that it is harder to break the bonds that hold together a G-C pair than the bonds holding together an A-T pair.

D. The two strands of DNA that form a double helix run in opposite directions.

E. The length of double-stranded DNA is measured in base pairs. 1000 base pairs is called a kilobase (1 kb).

F. The two strands of DNA are complementary to one another; if you are given the sequence of one strand, you automatically know the sequence of the second strand. To save space, the sequence of only one strand of the DNA, the coding strand, is reported in the literature. The sequence of the coding strand is the same sequence as found in the mRNA that is transcribed from that gene and is always written with the 5' end on the left (like a sentence). [More on mRNA in Chapter 2.]

G. RNA (ribonucleic acid) is composed of phosphate, ribose, and bases (A,C,G, Uracil). RNA containing the coding information for a protein (messenger RNA or mRNA) is single stranded.

II. Restriction endonucleases

A. Restriction endonucleases (RE) are enzymes that cut double-stranded DNA at a specific sequence of base pairs. These sequences are palindromes, that is, they read the same forwards (coding strand) as backwards (noncoding strand). For example, the restriction enzyme EcoRI cleaves DNA at the sequence:

---GAATTC--- (coding strand)
---CTTAAG--- (noncoding strand)

B. Depending on the restriction enzyme, the ends generated after cleavage are blunt (or flat) or have a 5' or 3' single strand protrusion (cohesive or sticky end).

Blunt ends are generated by the RE SmaI:

```
5' ----CCCGGG---- 3'  ===>        5' ---CCC 3'      5' GGG--- 3'
3' ----GGGCCC---- 5'              3' ---GGG 5'       3' CCC--- 5'
```

A 5' single strand protrusion is formed by the RE EcoRI:

```
5' ----GAATTC---- 3'  --->        5' ----G 3'       5' AATTC---- 3'
3' ----CTTAAG---- 5'              3' ----CTTAA 5'       3' G---- 5'
```

QUESTION 1: Borrow a molecular biology protocol book or a molecular biology catalog and look up the recognition sequence for the restriction enzyme PstI. Write out the sequence of a double stranded DNA with the restriction enzyme site in it. Label the 5' and 3' ends. Draw the fragments generated after cleavage with the 5' and 3' ends labeled. What kind of protrusion is generated?

C. Cohesive-ended fragments can be ligated (or linked) to one another only if their protrusions are compatible (have complementary sequences). Any DNA fragments which have blunt ends can be ligated together.

D. The number of nucleotides defining the cleavage site of a restriction enzyme can vary. RE with shorter recognition sequences cut DNA more frequently than those with longer recognition sequences. Assuming that a piece of DNA has an equal content of each base (A, C, G, and T), a RE with a 4-base recognition sequence will cleave the DNA, on average, every 4^4 (256) bases compared to every 4^6 (4096) bases for a RE with a 6-base recognition sequence.

Note: The probability of cleavage for a RE that cuts GGCT is calculated as the probability of finding that sequence in the DNA. Probability of a specific base, e.g., G, occurring is 1/4 (based on 4 bases: A,C,G,T) so probability of the specific sequence GGCT is 1/4x1/4x1/4x1/4 = 1/256.

E. RE are endonucleases. If a restriction site appears at the end of a DNA fragment, the restriction enzyme will not cut the DNA.

5' GAATTCGACTGCCATA 3'
3' CTTAAGCTGACGGTAT 5' will not be cut by EcoRI

QUESTION 2: Write down the DNA sequence cleaved by the RE NotI. Will NotI cleave genomic DNA more or less frequently than EcoRI? [Hint: You will need to look in a book to find the recognition sequence of this enzyme before you can answer the question.]

III. Denaturation/Hybridization

A. Denaturation disrupts the hydrogen bonds which hold the bases and, hence, the double stranded DNA together. Denaturation can be accomplished by heating the DNA or by treating the DNA with alkali (NaOH) or polar solvents (e.g., dimethyl sulfoxide (DMSO), formamide) which break hydrogen bonds.

B. The melting temperature (Tm) is defined as the temperature at which 50% of the DNA is hybridized (i.e., found in a double stranded form) and 50% is denatured. The Tm for a short piece of DNA can be estimated by [4 x G+C pairs] + [2 x A+T pairs]. The Tm is influenced by the base composition and the length of the double stranded DNA. Heating DNA at 94^{o}C or higher will usually denature all double stranded DNA regardless of the length or base composition. For example, the Tm of the double stranded DNA sequence:

AATGCGGAT
TTACGCCTA

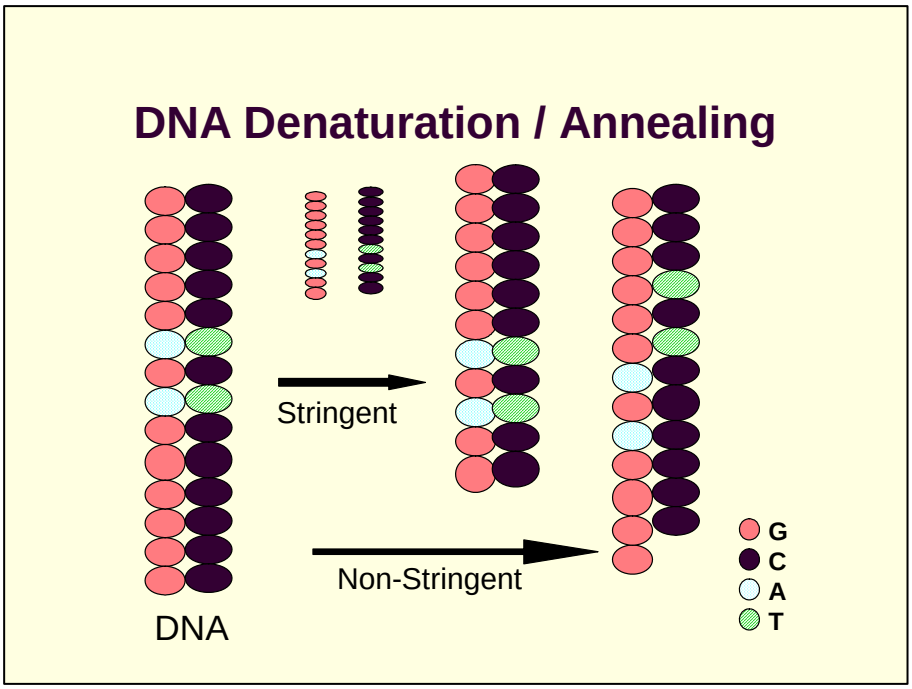
is (4x4)+(2x5)=26^{o}C.

QUESTION 3: Write out the sequence of any piece of DNA that is 18 base pairs in length and determine its approximate melting temperature. Is the melting temperature higher or lower than the example shown above? What would be the melting temperature of the 18 base pair sequence if it was made up of only G-C pairs? Only A-T pairs?

C. Hybridization or reannealing of DNA regenerates the base pairs and yields double stranded DNA. The efficiency of hybridization depends on:

1. Concentration of DNA: For example, if we denature the DNA found in a human cell, it will take a long time for one strand of an HLA gene to find its complement among all the other genes present. In contrast, the strand can rapidly find its complement in a solution that contains only many copies of that HLA gene.

Figure 1-2



2. Time that the hybridization is given to take place: The more complex the mixture of DNA, the more time a single stranded piece of DNA needs to find its match. For example, one strand of an HLA gene will have a hard time finding its complement in a mixture containing all the genes found in a human cell. The longer the hybridization time, the more likely the strands are to find one another and anneal.

3. Base composition of the DNA: Stretches of GC pairs tend to anneal more rapidly than stretches of AT pairs to form double stranded DNA. This probably happens because the formation of 3 hydrogen bonds between a single GC pair stabilizes the hybridized strands to a greater extent than the 2 hydrogen bonds formed by an AT pair. TMAC (tetramethylammonium chloride) is a chemical which causes DNA to reanneal at a rate related only to its length so that the time of

reannealing does not depend on the numbers of GC pairs.

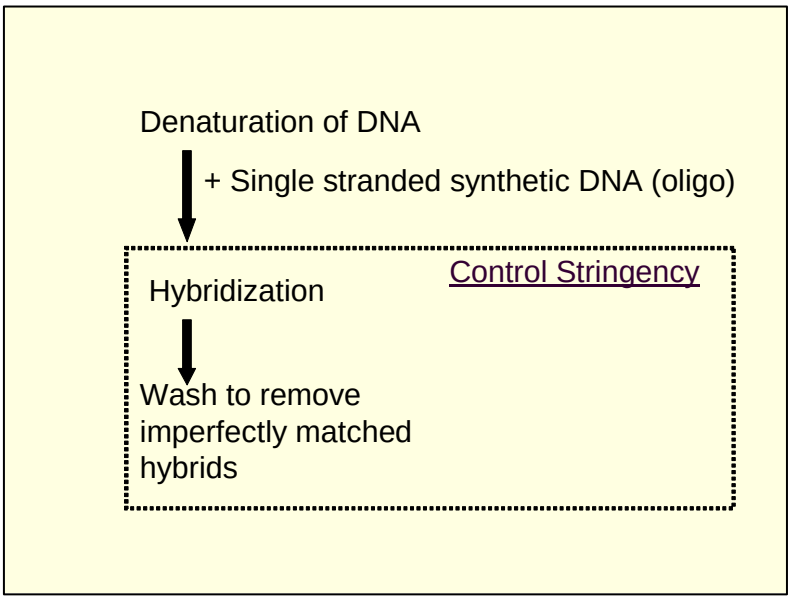
D. Single strands of DNA will bind to imperfectly matched single strands under conditions of low stringency [Figure 1-2]. Under high stringency conditions, only perfect matches are found. Usually, salt concentration and temperature are used to control the stringency.

1. Temperature of the reaction: High temperatures cause the DNA strands to move more rapidly. Imperfectly matched DNA hybrids will dissociate more rapidly at high temperatures. Therefore, higher temperatures favor the generation of perfectly matched hybrids. Obviously, if the temperature is too high, the two strands will not reanneal.

2. Salt concentration of the reaction mixture: High concentrations of salt allow imperfectly matched hybrids to be formed. At low salt concentrations, only perfectly matched hybrids will form.

3. Low salt and high temperature create high stringency conditions. The stringency of the match is controlled during the incubation of single stranded DNA (hybridization) or during the wash following the hybridization reaction. The temperature is the easiest parameter to adjust to obtain perfect matches. If using a short piece of single stranded DNA for hybridization, the final high stringency wash is often carried out 3-5^{0}C below the melting temperature to keep the perfectly matched strands hybridized but to eliminate imperfectly matched hybrids [Figure 1-3].

Figure 1-3



E. Probes are used to detect specific DNA sequences by hybridization. For HLA typing, these probes are synthetic single stranded DNA (oligonucleotides) which are 12-26 nucleotides in length. Probes are often called sequence specific oligonucleotide probes (SSOP).

1. Because many copies of these probes can be added to create a high concentration, hybridization takes place rapidly.

F. One strand of the DNA is often attached to a solid support (e.g., membrane or bead) to increase the speed of hybridization and to aid in the detection of a successful hybridization reaction.

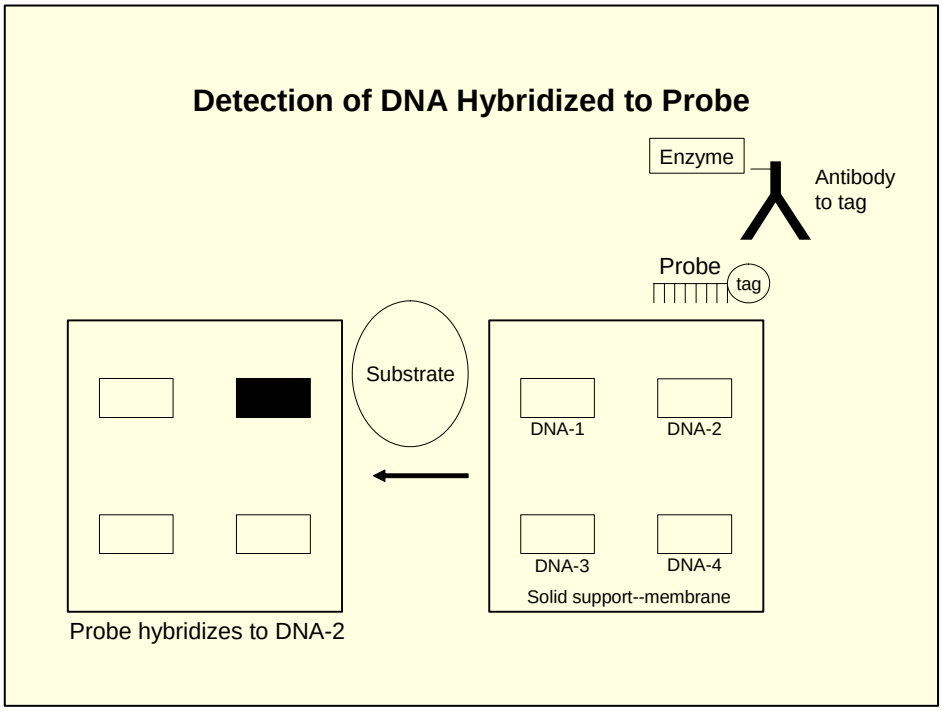
IV. Detection of hybridization

A. If the probe is labeled, hybridization can be detected.

1. Probes can be labeled using radioactive phosphate (P32) or by adding a modified or unusual nucleotide to a probe. The label is added to the probe either during or after synthesis of the probe. Methods to do this are described later.

2. Binding can be detected by autoradiography (to detect P32) or through detection of color, chemiluminescence (light emitted by a chemical reaction) or fluorescence. Detectors may be X-ray film, ELISA plate readers, fluorescence or chemiluminescence detectors, or the human eye. Figure 1-4 illustrates a detection system.

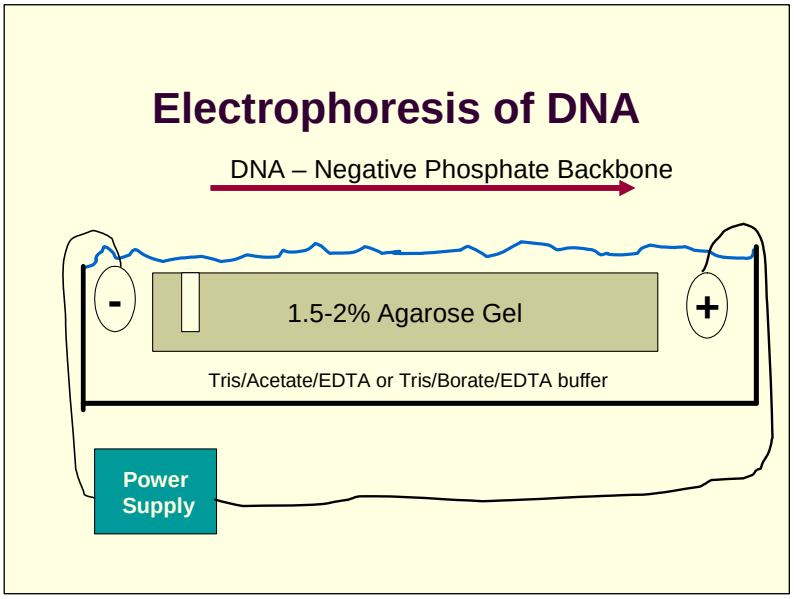
Figure 1-4



V. Electrophoresis

A. Electrophoresis through agarose or polyacrylamide gels is the standard method to separate, identify, and purify DNA fragments [Figure 1-5]. DNA is negatively charged (due to the phosphate backbone) and moves away from the negative pole (cathode) and toward the positive pole (anode) in an electric field. Pieces of DNA are identified by size. Small pieces of DNA move more rapidly than large pieces of DNA. DNA is visualized on the gel by UV light if the DNA is stained with ethidium bromide or by X-ray film if the DNA is labeled with a radioactive isotope.

Figure 1-5

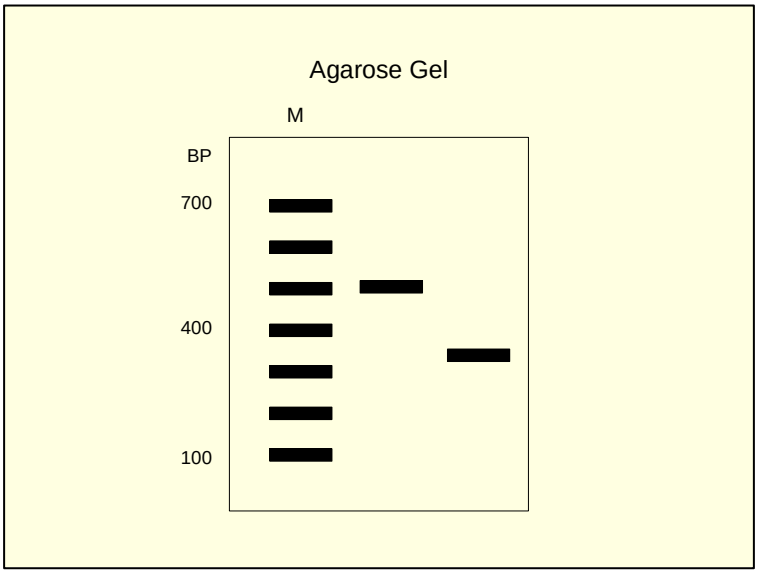


B. The gel itself is a matrix through which the pieces of DNA travel. Polyacrylamide gels are used to separate small fragments of DNA (5-500 base pairs) and are used for DNA sequencing or oligonucleotide purification. Agarose gels are used for DNA from 200 base pairs to 50,000 base pairs (50 kilobases). Two DNA fragments which are 50 bp and 55 bp will likely migrate at different rates and will be identified on an acrylamide gel but will migrate together (and not be resolved) on an agarose gel.

C. The percent composition of acrylamide or agarose determines how easily pieces of DNA of different sizes can travel through the matrix; therefore, the percent composition of the gel will determine the range over which DNA fragments are resolved (separated from one another).

QUESTION 4: Figure 1-6 is a picture of an ethidium bromide stained agarose gel. The lane labeled M contains commercially purchased marker DNA which contains DNA fragments that are multiples of 100 base pairs in length. The other two lanes contain pieces of DNA that have been electrophoresed in the gel. Label the positive and negative poles of the gel and indicate the direction of DNA migration. What are the approximate sizes of the two DNA fragments in the lanes?

Figure 1-6

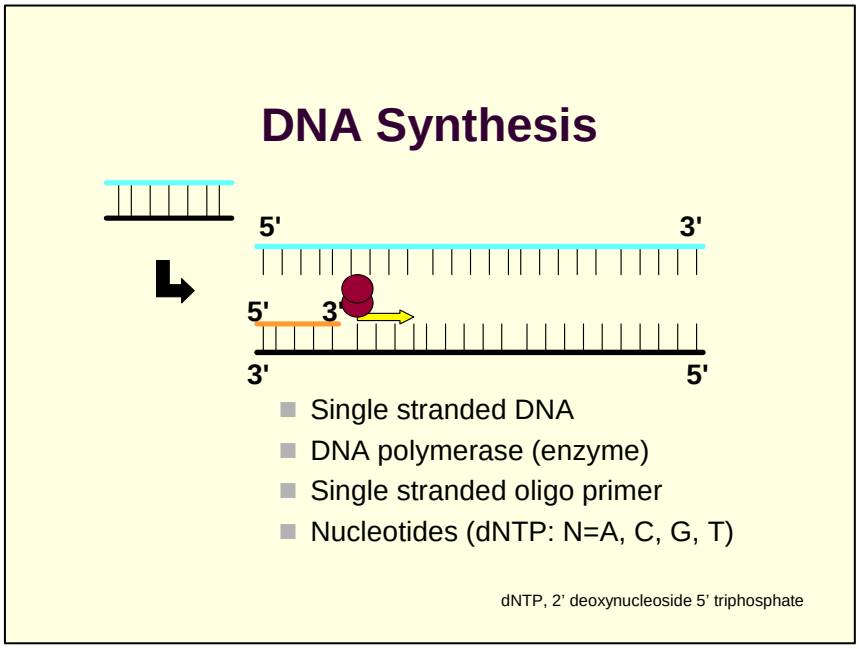


VI. Synthesis of DNA

A. Enzymes called DNA polymerases synthesize single strands of DNA [Figure 1-7]. The enzymes require:

1. A single stranded DNA template to copy. The polymerases synthesize a product whose sequence is complementary to the template. Thermostable DNA polymerase (Taq DNA polymerase), E. coli DNA polymerase I, the Klenow fragment of that polymerase,and T4 and T7 polymerases (including Sequenase) require a DNA template. When starting with a mRNA template, the enzyme reverse transcriptase copies mRNA to make complementary DNA (cDNA).

Figure 1-7



2. A single stranded DNA primer (a synthetic oligonucleotide) that is hybridized to the template. Polymerases usually add nucleotides on to the 3' end of a primer and extend the newly synthesized strand from 5' to 3'.

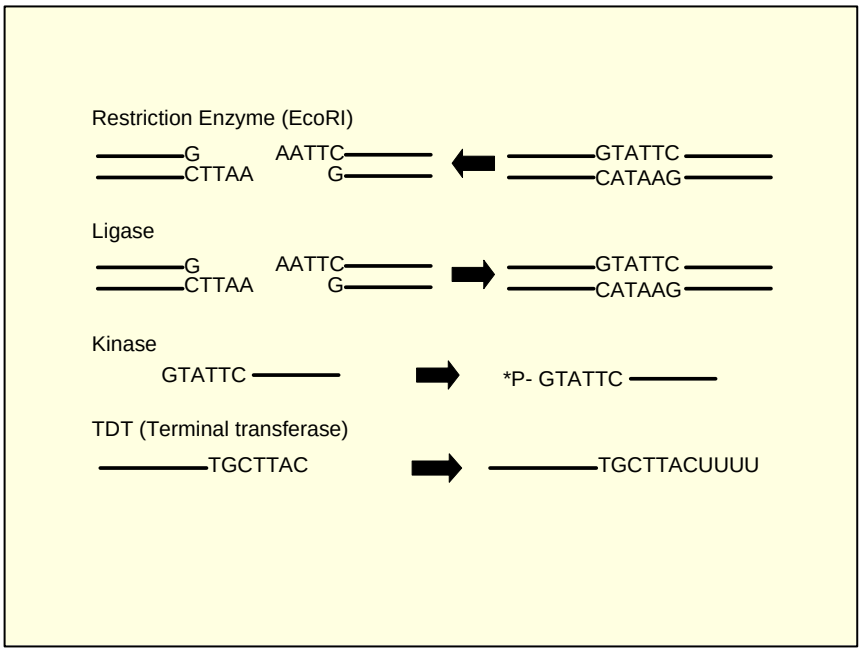
3. Nucleotides (dNTP) to form the new DNA strand.

B. If modified (e.g., radiolabeled) nucleotides are added during synthesis or if a modified (e.g., biotinylated) primer is used for synthesis, the newly synthesized DNA becomes labeled.

C. Single stranded DNA of a defined sequence (oligonucleotide) can be made in the laboratory using an automated DNA synthesizer and phosphoramidite chemistry. These pieces of DNA can be used as primers for DNA synthesis or as probes in hybridization reactions. Modified nucleotides might be used instead of A,C,G,T to provide more stability or to enhance hybridization to certain sequences.

D. DNA ligases (e.g., T4 DNA ligase) catalyze the formation of the phosphate backbone in double stranded DNA and can be used to join up DNA fragments [Figure 1-8]. For example, a piece of DNA cut with the RE EcoRI into two fragments can be put together again by incubating with a ligase.

Figure 1-8



E. Kinases (e.g., T4 polynucleotide kinase) add phosphates on to the 5' ends of DNA fragments [Figure 1-8]. This enables DNA to be labeled with radioactive phosphate (P32) so that hybridization can be detected.

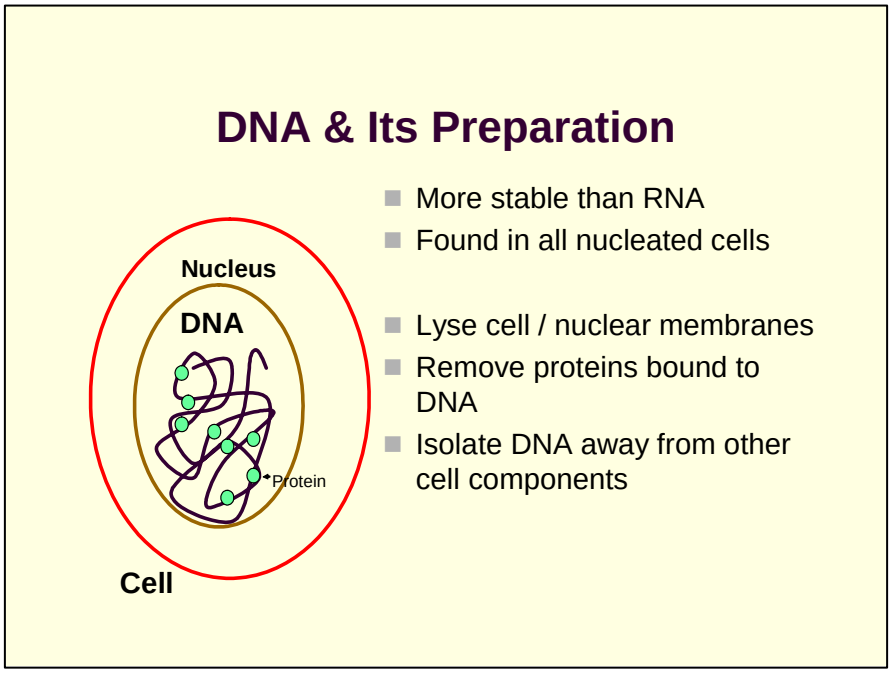
F. Terminal transferase adds nucleotides to the 3' ends of DNA molecules to make a single stranded tail [Figure 1-8]. Terminal transferase can be used to create homopolymer tails (tails of all one nucleotide) for cloning or for labeling the DNA.

VII. Extraction of DNA from cells

A. Any cell with a nucleus can be used as a source of DNA [Figure 1-9]. Red blood cells do not contain nuclei; however, other cells in the blood like white blood cells are a good source of DNA. Cell lines like transformed B cells are also a good source of DNA. Because transformed cells can be grown in culture in the laboratory, they provide an inexhaustible supply of DNA.

B. Many different protocols can be used to isolate DNA from cells. One protocol uses Triton X-100 (a detergent) to lyse the cell membrane releasing the nuclei. If the starting material was whole blood, these nuclei must then be washed extensively to wash away any hemoglobin released by the red blood cells. The heme portion of hemoglobin interfers with the gene amplification reaction used to determine HLA types.

Figure 1-9



C. The nuclei are lysed using another detergent, Tween-20, and the DNA freed from the proteins bound to it by treatment with Proteinase K, an enzyme which destroys proteins.

D. After the proteins are destroyed, the Proteinase K is also destroyed by incubation of the DNA at high temperatures (90^{o}C). It is important to destroy the Proteinase K because it can degrade the enzyme used in the HLA typing reaction.

E. Other methods may employ a solid phase to bind DNA for isolation.

F. A number of vendors sell kits to prepare DNA.

References:

Ausubel et al. Short Protocols in Molecular Biology. John Wiley and Sons, New York, N.Y. 2002.

Alberts et al. Molecular Biology of the Cell. Garland Publishing Co., New York, N. Y. 2002.

CHAPTER 2

HLA CLASS I PROTEINS AND GENES

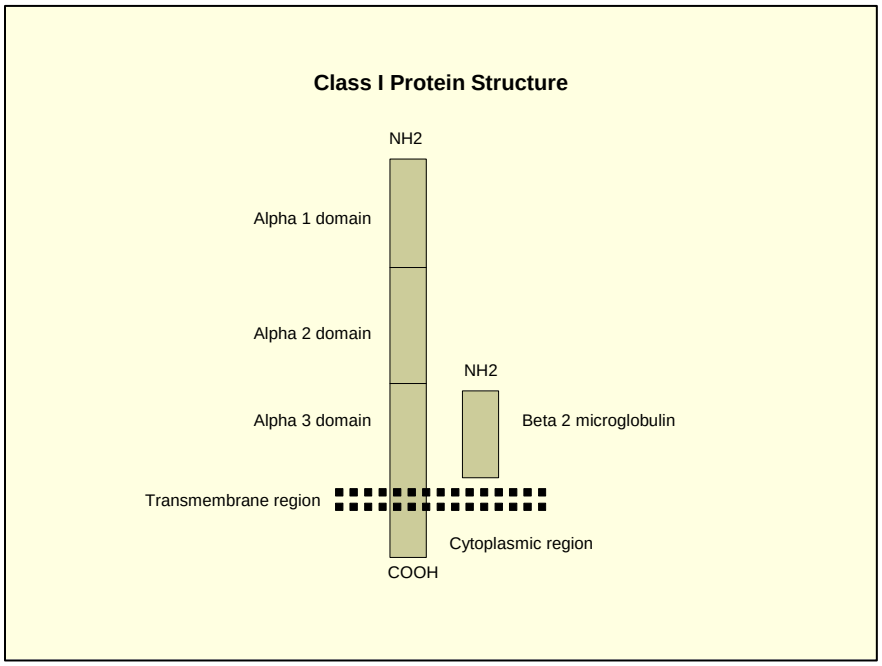
The purpose of this chapter is to describe an important group of HLA molecules, the class I molecules. Matching for bone marrow transplantation involves identification of these class I molecules, HLA-A, HLA-B and HLA-C. This chapter also discusses the general properties of proteins and genes.

I. HLA is an abbreviation for human leukocyte antigen. These protein molecules are expressed on the surfaces of our cells and play an important role in transplantation. Because of this, the HLA molecules of graft donor and recipient must be identified to find the best match. These molecules are often called histocompatibility molecules because of their role in compatibility of tissue.

II. Class I HLA molecules are histocompatibility molecules which are identified during HLA typing.

A. The class I molecules are found on the surface of essentially all nucleated cells in the body.

Figure 2-1



B. At least three different class I molecules, HLA-A, HLA-B, and HLA-C, are expressed on each cell of an individual. These molecules are very similar to one another. There are approximately 0.5-1 million class I molecules on the

surface of a cell.

C. Class I molecules are comprised of a cell membrane glycopolypeptide (44K MW, alpha or heavy chain) associated with a second polypeptide, beta-2 microglobulin (12K MW) [Figure 2-1]. Each chain is a linear string of amino acids. Each class I protein has a specific sequence of amino acids that differs from other proteins. The sequence of an HLA-A molecule is different from the sequence of an HLA-B molecule. [Note: Class II molecules also have alpha chains but they are different in amino acid sequence from class I alpha chains. Likewise, the HLA-A,B,C alpha chains differ from one another in protein sequence.]]

Table 2-1. One- and Three-Letter Codes for Amino Acids

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic Acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Trytophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

There are 20 amino acids (e.g., lysine, serine, leucine) [Table 2-1]. To save space, publications sometimes use a single letter to refer to an amino acid. For example, S is used for serine, L for leucine, and K for lysine.

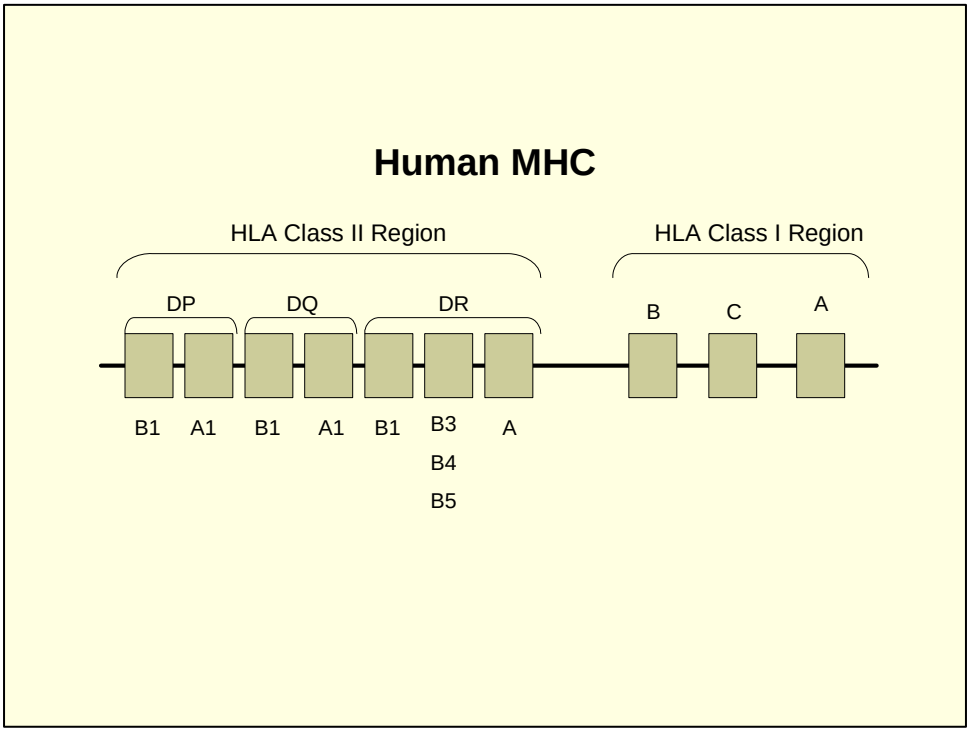
One end of the polypeptide chain is called the amino-terminus (or N-terminus) and the other end is called the carboxy-terminus (or C-terminus).

D. Each alpha (or heavy) chain can be divided into regions: three extracellular domains (each ~100 amino acids in length), transmembrane region and cytoplasmic tail. Each alpha chain is initially synthesized with a short leader peptide on its amino terminus. This peptide is removed from the polypeptide during transport to the cell surface and is not found in the mature protein.

III. Class I genes

A. Some of the information specifying the amino acid sequences of the HLA class I molecules is located on human chromosome 6 [Figure 2-2]. The sequence of base pairs containing the genetic information that specifies a protein sequence defines a gene. The genes that encode the three class I 44K MW alpha chains are located next to one another in the MHC. Beta-2 microglobulin is encoded on another chromosome.

Figure 2-2

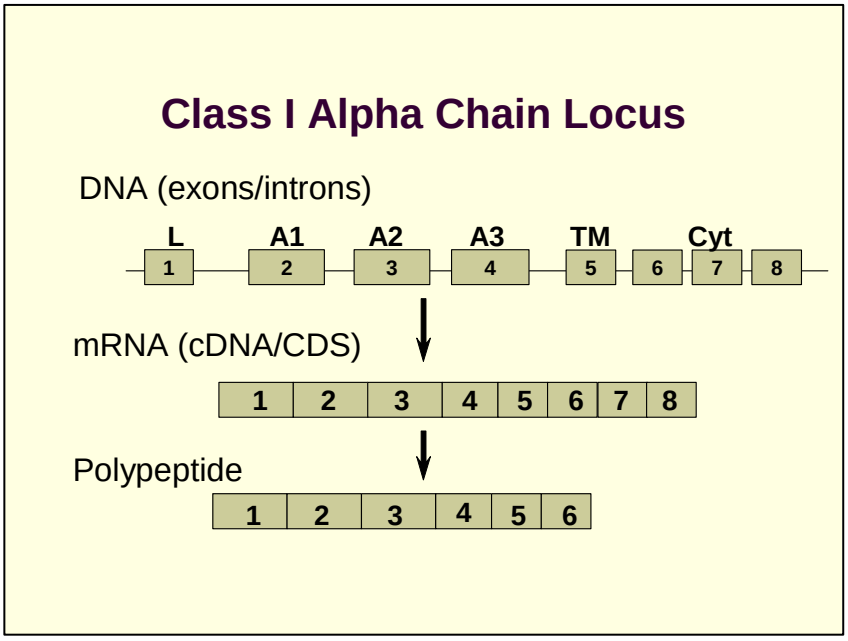


B. The genes that encode the alpha chains of the class I proteins are located next to one another. Beta2 microglobulin is encoded on another chromosome. This cluster of genes is part of the major histocompatibility gene complex (MHC). This complex encompasses approximately 3500 kb (3,500,000 bases) of DNA. When the location of a gene on a chromosome is known, that gene can also be called a locus.

C. The World Health Organization (W.H.O.) has a committee that assigns names to HLA loci. For example, HLA-A is the name of the locus that encodes the HLA-A alpha chain.

D. The information for a single gene is not found in a single stretch of base pairs but is found in multiple segments of double stranded DNA called exons

[Figure 2-3]. These segments are separated by intervening segments of DNA called introns. The entire set of exons containing the coding information for an entire polypeptide chain is called a "gene".Figure 2-3



IV. Transfer of information from DNA to RNA to protein.

A. The DNA encoding the HLA molecule is copied (transcribed) into mRNA. After processing, the mRNA contains only the exon sequences.

QUESTION 1: Figure 2-4 lists the genomic or DNA sequence of an HLA-A gene. Figure 2-5 lists the coding sequence (CDS) (=mRNA sequence) sequence from the same HLA-A gene. Use the information in Figure 2-5 to identify and box the exons which encode the HLA-A gene in the genomic DNA sequence in Figure 2-4. Remember that genomic DNA contains exons and introns. Hint: The vertical lines in Figure 2-4 and the underlining added to Figure 2-5 will help you find the exons.

Figure 2-4. Full length genomic sequence A*01010101

```
                    -291       -281       -271       -261       -251       -241       -231       -221       -211       -201
A*01010101  CAGGAGCAGA GGGGTCAGGG CGAAGTCCCA GGGCCCCAGG CGTGGCTCTC AGGGTCTCAG GCCCCGAAGG CGGTGTATGG ATTGGGGAGT CCCAGCCTTG

                    -191       -181       -171       -161       -151       -141       -131       -121       -111       -101
A*01010101  GGGATTCCCC AACTCCGCAG TTTCTTTTCT CCCTCTCCCA ACCTACGTAG GGTCCTTCAT CCTGGATACT CACGACGCGG ACCCAGTTCT CACTCCCATT

                     -91        -81        -71        -61        -51        -41        -31        -21        -11         -1
A*01010101  GGGTGTCGGG TTTCCAGAGA AGCCAATCAG TGTCGTCGCG GTCGCTGTTC TAAAGTCCGC ACGCACCCAC CGGGACTCAG ATTCTCCCCA GACGCCGAGG

                     10         20         30         40         50         60         70         80         90        100
A*01010101  |ATGGCCGTCA TGGCGCCCCG AACCCTCCTC CTGCTACTCT CGGGGGCCCT GGCCCTGACC CAGACCTGGG CGG|GTGAGTG CGGGGTCGGGAGGGAAACCG

                    110        120        130        140        150        160        170        180        190        200
A*01010101  CCTCTGCGGG GAGAAGCAAG GGGCCCTCCT GGCGGGGGCG CAGGACCGGG GGAGCCGCGC CGGGAGGAGG GTCGGGCAGG TCTCAGCCAC TGCTCGCCCC

                    210        220        230        240        250        260        270        280        290        300
A*01010101  CAG|GCTCCCA CTCCATGAGG TATTTCTTCA CATCCGTGTC CCGGCCCGGC CGCGGGGAGC CCCGCTTCAT CGCCGTGGGC TACGTGGACG ACACGCAGTT

                    310        320        330        340        350        360        370        380        390        400
A*01010101  CGTGCGGTTC GACAGCGACG CCGCGAGCCA GAAGATGGAG CCGCGGGCGC CGTGGATAGA GCAGGAGGGG CCGGAGTATT GGGACCAGGA GACACGGAAT

                    410        420        430        440        450        460        470        480        490        500
A*01010101  ATGAAGGCCC ACTCACAGAC TGACCGAGCG AACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG ACG|GTGAGTG ACCCCGGCCC GGGGCGCAGG

                    510        520        530        540        550        560        570        580        590        600
A*01010101  TCACGACCCC TCATCCCCCA CGGACGGGCC AGGTCGCCCA CAGTCTCCGG GTCCGAGATC CACCCCGAAG CCGCGGGACT CCGAGACCCT TGTCCCGGGA

                    610        620        630        640        650        660        670        680        690        700
A*01010101  GAGGCCCAGG CGCCTTTACC CGGTTTCATT TTCAGTTTAG GCCAAAAATC CCCCCGGGTT GGTCGGGGCG GGGCGGGGCT CGGGGGACTG GGCTGACCGC

                    710        720        730        740        750        760        770        780        790        800
A*01010101  GGGGTCGGGG CCAG|GTTCTC ACACCATCCA GATAATGTAT GGCTGCGACG TGGGGCCGGA CGGGCGCTTC CTCCGCGGGT ACCGGCAGGA CGCCTACGAC

                    810        820        830        840        850        860        870        880        890        900
A*01010101  GGCAAGGATT ACATCGCCCT GAACGAGGAC CTGCGCTCTT GGACCGCGGC GGACATGGCA GCTCAGATCA CCAAGCGCAA GTGGGAGGCG GTCCATGCGG

                    910        920        930        940        950        960        970        980        990       1000
A*01010101  CGGAGCAGCG GAGAGTCTAC CTGGAGGGCC GGTGCGTGGA CGGGCTCCGC AGATACCTGG AGAACGGGAA GGAGACGCTG CAGCGCACGG |GTACCAGGGG

                   1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
A*01010101  CCACGGGGCG CCTCCCTGAT CGCCTATAGA TCTCCCGGGC TGGCCTCCCA CAAGGAGGGG AGACAATTGG GACCAACACT AGAATATCAC CCTCCCTCTG

                   1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
A*01010101  GTCCTGAGGG AGAGGAATCC TCCTGGGTTT CCAGATCCTG TACCAGAGAG TGACTCTGAG GTTCCGCCCT GCTCTCTGAC ACAATTAAGG GATAAAATCT

                   1210       1220       1230       1240       1250       1260       1270       1280       1290       1300
A*01010101  CTGAAGGAGT GACGGGAAGA CGATCCCTCG AATACTGATG AGTGGTTCCC TTTGACACCG GCAGCAGCCT TGGGCCCGTG ACTTTTCCTC TCAGGCCTTG

                   1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
A*01010101  TTCTCTGCTT CACACTCAAT GTGTGTGGGG GTCTGAGTCC AGCACTTCTG AGTCTCTCAG CCTCCACTCA GGTCAGGACC AGAAGTCGCT GTTCCCTTCT

                   1410       1420       1430       1440       1450       1460       1470       1480       1490       1500
A*01010101  CAGGGAATAG AAGATTATCC CAGGTGCCTG TGTCCAGGCT GGTGTCTGGG TTCTGTGCTC TCTTCCCCAT CCCGGGTGTC CTGTCCATTC TCAAGATGGC

                   1510       1520       1530       1540       1550       1560       1570       1580       1590       1600
A*01010101  CACATGCGTG CTGGTGGAGT GTCCCATGAC AGATGCAAAA TGCCTGAATT TTCTGACTCT TCCCGTCAG|A CCCCCCCAAG ACACATATGA CCCACCACCC

                   1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
A*01010101  CATCTCTGAC CATGAGGCCA CCCTGAGGTG CTGGGCCCTG GGCTTCTACC CTGCGGAGAT CACACTGACC TGGCAGCGGG ATGGGGAGGA CCAGACCCAG

                   1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
A*01010101  GACACGGAGC TCGTGGAGAC CAGGCCTGCA GGGGATGGAA CCTTCCAGAA GTGGGCGGCT GTGGTGGTGC CTTCTGGAGA GGAGCAGAGA TACACCTGCC

                   1810       1820       1830       1840       1850       1860       1870       1880       1890       1900
A*01010101  ATGTGCAGCA TGAGGGTCTG CCCAAGCCCC TCACCCTGAG ATGGG|GTAAG GAGGGAGATG GGGGTGTCAT GTCTCTTAGG GAAAGCAGGA GCCTCTCTGG

                   1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
A*01010101  AGACCTTTAG CAGGGTCAGG GCCCCTCACC TTCCCCTCTT TTCCCAG|AGC TGTCTTCCCA GCCCACCATC CCCATCGTGG GCATCATTGC TGGCCTGGTT

                   2010       2020       2030       2040       2050       2060       2070       2080       2090       2100
A*01010101  CTCCTTGGAG CTGTGATCAC TGGAGCTGTG GTCGCTGCCG TGATGTGGAG GAGGAAGAGC TCAG|GTGGAG AAGGGGTGAA GGGTGGGGTC TGAGATTTCT

                   2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
A*01010101  TGTCTCACTG AGGGTTCCAA GCCCCAGCTA GAAATGTGCC CTGTCTCATT ACTGGGAAGC ACCTTCCACA ATCATGGGCC GACCCAGCCT GGGCCCTGTG

                   2210       2220       2230       2240       2250       2260       2270       2280       2290       2300
A*01010101  TGCCAGCACT TACTCTTTTG TAAAGCACCT GTTAAAATGA AGGACAGATT TATCACCTTG ATTACGGCGG TGATGGGACC TGATCCCAGC AGTCACAAGT

                   2310       2320       2330       2340       2350       2360       2370       2380       2390       2400
A*01010101  CACAGGGGAA GGTCCCTGAG GACAGACCTC AGGAGGGCTA TTGGTCCAGG ACCCACACCT GCTTTCTTCA TGTTTCCTGA TCCCGCCCTG GGTCTGCAGT

                   2410       2420       2430       2440       2450       2460       2470       2480       2490       2500
A*01010101  CACACATTTC TGGAAACTTC TCTGGGGTCC AAGACTAGGA GGTTCCTCTA GGACCTTAAG GCCCTGGCTC CTTTCTGGTA TCTCACAGGA CATTTTCTTC

                   2510       2520       2530       2540       2550       2560       2570       2580       2590       2600
A*01010101  CCACAG|ATAG AAAAGGAGGG AGTTACACTC AGGCTGCAA|G TAAGTATGAA GGAGGCTGAT GCCTGAGGTC CTTGGGATAT TGTGTTTGGGAGCCCATGGG

                   2610       2620       2630       2640       2650       2660       2670       2680       2690       2700
A*01010101  GGAGCTCACC CACCCCACAA TTCCTCCTCT AGCCACATCT TCTGTGGGAT CTGACCAGGT TCTGTTTTTG TTCTACCCCA G|GCAGTGACA GTGCCCAGGG

                   2710       2720       2730       2740       2750       2760       2770       2780       2790       2800
A*01010101  CTCTGATGTG TCTCTCACAG CTTGTAAAG|G TGAGAGCTTG GAGGGCCTGA TGTGTGTTGG GTGTTGGGTG GAACAGTGGA CACAGCTGTG CTATGGGGTT

                   2810       2820       2830       2840       2850       2860       2870       2880       2890       2900
A*01010101  TCTTTGCGTT GGATGTATTG AGCATGCGAT GGGCTGTTTA AGGTGTGACC CCTCACTGTG ATGGATATGA ATTTGTTCAT GAATATTTTT TTCTATAG|TG

                   2910       2920       2930       2940       2950       2960       2970       2980       2990       3000
A*01010101  TGA|GACAGCT GCCTTGTGTG GGACTGAGAG GCAAGAGTTG TTCCTGCCCT TCCCTTTGTG ACTTGAAGAA CCCTGACTTT GTTTCTGCAA AGGCACCTGC

                   3010       3020       3030       3040       3050       3060       3070       3080       3090       3100
```

```
A*01010101          ATGTGTCTGT GTTCGTGTAG GCATAATGTG AGGAGGTGGG GAGAGCACCC CACCCCCATG TCCACCATGA CCCTCTTCCC ACGCTGACCT GTGCTCCCTC

                         3110       3120       3130       3140       3150       3160       3170       3180       3190       3200
A*01010101          CCCAATCATC TTTCCTGTTC CAGAGAGGTG GGGCTGAGGT GTCTCCATCT CTGTCTCAAC TTCATGGTGC ACTGAGCTGT AACTTCTTCC TTCCCTATTA
```

Figure 2-5. Nucleotide CDS (coding sequence) A*01010101

```
                              -20                -15                -10                 -5                  1
A*01010101          ATG GCC GTC ATG GCG CCC CGA ACC CTC CTC CTG CTA CTC TCG GGG GCC CTG GCC CTG ACC CAG ACC TGG GCG G|GC

                             5                  10                 15                 20                  25
A*01010101          TCC CAC TCC ATG AGG TAT TTC TTC ACA TCC GTG TCC CGG CCC GGC CGC GGG GAG CCC CGC TTC ATC GCC GTG GGC

                            30                 35                 40                 45                  50
A*01010101          TAC GTG GAC GAC ACG CAG TTC GTG CGG TTC GAC AGC GAC GCC GCG AGC CAG AAG ATG GAG CCG CGG GCG CCG TGG

                            55                 60                 65                 70                  75
A*01010101          ATA GAG CAG GAG GGG CCG GAG TAT TGG GAC CAG GAG ACA CGG AAT ATG AAG GCC CAC TCA CAG ACT GAC CGA GCG

                            80                 85                 90                 95                 100
A*01010101          AAC CTG GGG ACC CTG CGC GGC TAC TAC AAC CAG AGC GAG GAC G|GT TCT CAC ACC ATC CAG ATA ATG TAT GGC TGC

                           105                110                115                120                 125
A*01010101          GAC GTG GGG CCG GAC GGG CGC TTC CTC CGC GGG TAC CGG CAG GAC GCC TAC GAC GGC AAG GAT TAC ATC GCC CTG

                           130                135                140                145                 150
A*01010101          AAC GAG GAC CTG CGC TCT TGG ACC GCG GCG GAC ATG GCA GCT CAG ATC ACC AAG CGC AAG TGG GAG GCG GTC CAT

                           155                160                165                170                 175
A*01010101          GCG GCG GAG CAG CGG AGA GTC TAC CTG GAG GGC CGG TGC GTG GAC GGG CTC CGC AGA TAC CTG GAG AAC GGG AAG

                           180                185                190                195                 200
A*01010101          GAG ACG CTG CAG CGC ACG G|AC CCC CCC AAG ACA CAT ATG ACC CAC CAC CCC ATC TCT GAC CAT GAG GCC ACC CTG

                           205                210                215                220                 225
A*01010101          AGG TGC TGG GCC CTG GGC TTC TAC CCT GCG GAG ATC ACA CTG ACC TGG CAG CGG GAT GGG GAG GAC CAG ACC CAG

                           230                235                240                245                 250
A*01010101          GAC ACG GAG CTC GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC TTC CAG AAG TGG GCG GCT GTG GTG GTG CCT TCT

                           255                260                265                270                 275
A*01010101          GGA GAG GAG CAG AGA TAC ACC TGC CAT GTG CAG CAT GAG GGT CTG CCC AAG CCC CTC ACC CTG AGA TGG G|AG CTG

                           280                285                290                295                 300
A*01010101          TCT TCC CAG CCC ACC ATC CCC ATC GTG GGC ATC ATT GCT GGC CTG GTT CTC CTT GGA GCT GTG ATC ACT GGA GCT

                           305                310                315                320                 325
A*01010101          GTG GTC GCT GCC GTG ATG TGG AGG AGG AAG AGC TCA G|AT AGA AAA GGA GGG AGT TAC ACT CAG GCT GCA A|GC AGT

                           330                335                340
A*01010101          GAC AGT GCC CAG GGC TCT GAT GTG TCT CTC ACA GCT TGT AAA G|TG TGA
```

B. The mRNA is translated into protein by the ribosomes. The ribosome reads the genetic code to convert RNA sequences into a protein sequence.

The genetic code consists of all possible triplet combinations of RNA bases. Each triplet (codon) specifies one amino acid. For example, the codon UCU (TCT in DNA) specifies the amino acid serine. An amino acid can be specified by more than one triplet. For example, UCC, UCA, and UCG also specify serine.

The class I protein sequences start with a methionine at the amino-terminus of the leader peptide encoded by an AUG codon in the mRNA. Some triplets are called "stop codons" and identify the end of the protein sequence (carboxy-terminus). UGA is a stop codon.

C. The class I leader peptide and part of the first amino acid of the alpha-1 domain are encoded in exon 1 [Figure 2-3]. The rest of the first domain is encoded in exon 2. Exon 3 encodes the second domain and exon 4 encodes the third domain. The remainder of the exons encode the transmembrane and cytoplasmic regions.

D. Intron 1 follows exon 1, intron 2 follows exon 2 and so forth.

E. If the gene is characterized by a sequence analysis of a DNA copy of the mRNA, that sequence is called a cDNA (complementary DNA) sequence. The cDNA sequence reported in the literature has the same sequence as the mRNA and is always written 5' (on the left) to 3'.

QUESTION 2: Use Figure 2-5 and Table 2-2 to translate the cDNA (mRNA) sequence for HLA-A into a protein sequence. The codon encoding the first amino acid in the leader sequence is indicated. Circle the stop codon.

V. The polypeptide specified by an alpha gene associates with the polypeptide specified by the beta-2 microglobulin gene to form a class I protein [Figure 2-1]. Polypeptide is a term used to indicate that the alpha and beta chains are not usually found alone but are found in an complex (also called a heterodimer).

Table 2-2. The Alphabet of the Genetic Code[1]							
UUU[2]	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Term[3]	UGA	Term
UUG	Leu	UCG	Ser	UAG	Term	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

[1] mRNA codons which specify a particular amino acid.
[2] U is found in mRNA; T is found in DNA.
[3] Term indicates a termination codon which halts protein synthesis.

References:

Marsh, S.G.E., et al. Nomenclature for factors of the HLA system, 2004. Tissue Antigens 65:301- 368, 2005.

http://www.anthonynolan.org.uk/HIG; http://www.ebi.ac.uk/imgt/hla/

Nelson and Cox. Lehninger Principles of Biochemistry W.H. Freeman. 4th Edition Worth Publishing Co., New York, N.Y. 2004.

Alberts et al. Molecular Biology of the Cell. Garland Publishing Co., New York, N.Y. 2002.

Janeway et al. Immunobiology. Garland Publishing Inc., New York, N.Y. 2005.

CHAPTER 3

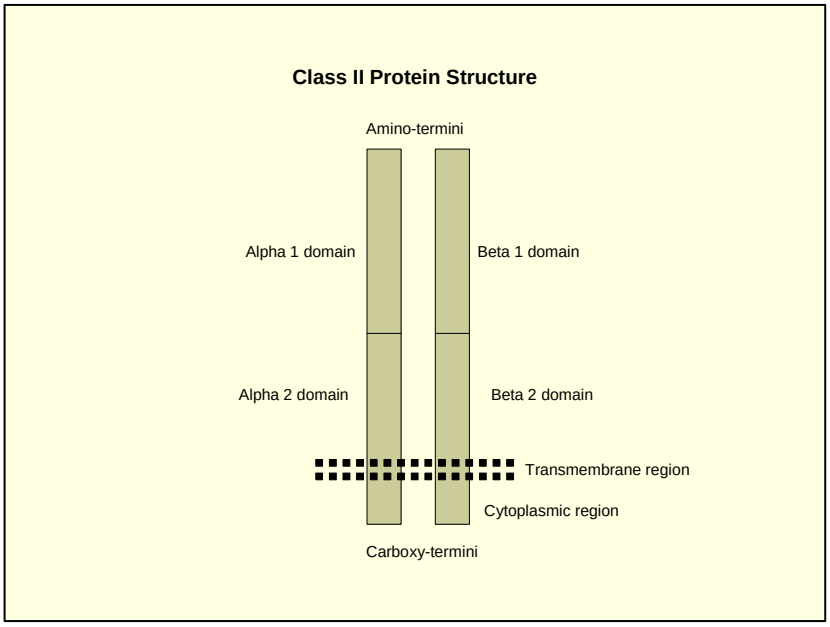
HLA CLASS II PROTEINS AND GENES

The purpose of this chapter is to describe a second important group of HLA molecules, the class II molecules. Matching of donor and recipient for bone marrow transplantation involves identification of at least one of the class II molecules, HLA-DR. Many of the characteristics of the class II molecules are similar to the class I molecules discussed in Chapter 2.

I. Class II HLA molecules

A. These molecules are found on the surface of cells of the immune system like B cells and dendritic cells although other cell types may express these molecules under certain conditions (e.g., under the influence of cytokines).

Figure 3-1



B. Three different class II molecules, HLA-DR, HLA-DQ, and HLA-DP, are expressed on immune system cells by an individual. These molecules are very similar to one another. There are approximately 0.5-1 million class II molecules on the cell surface.

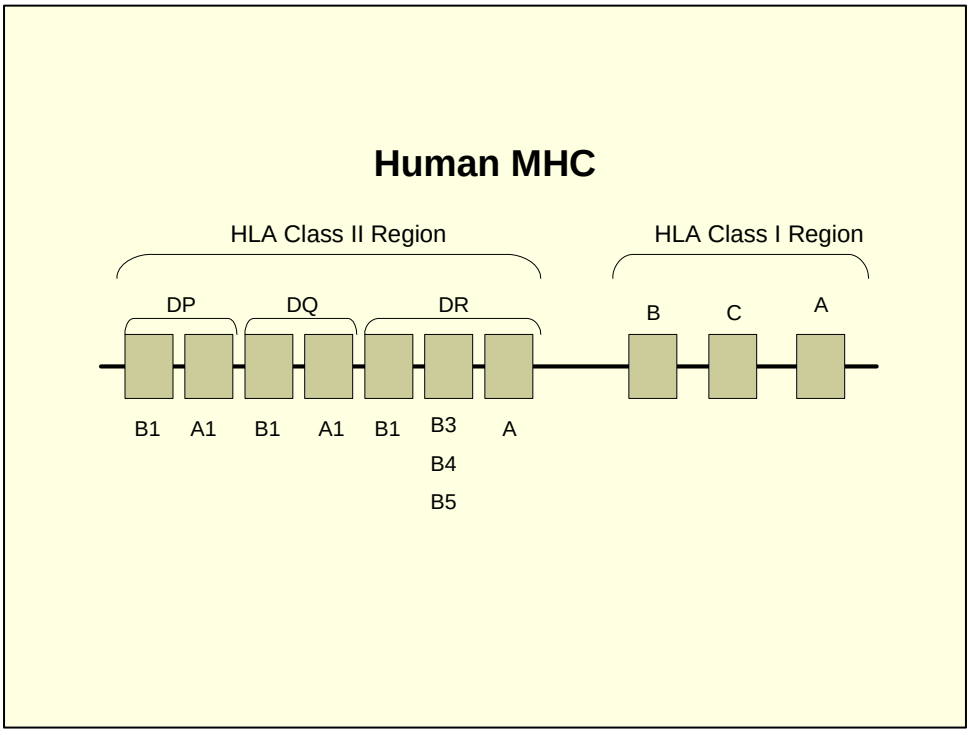
C. The class II molecules are cell membrane glycoproteins consisting of an alpha polypeptide chain (~34K MW) and a beta polypeptide chain (~28K MW) [Figure 3-1]. Each protein has a specific sequence of amino acids that differs from other proteins. The sequence of an HLA-DR molecule is different from the sequence of an HLA-DP molecule.

D. Each class II polypeptide chain can be divided into regions: two extracellular domains (each ~100 amino acids in length), a transmembrane region, and a cytoplasmic tail. Each polypeptide is initially synthesized with a short leader peptide on its amino terminus. This peptide is removed from the polypeptide during transport to the cell surface after synthesis and is not found in the mature protein.

II. Class II genes

A. The information specifying the amino acid sequences of the HLA class II molecules is located on human chromosome 6 [Figure 3-2]. The sequence of base pairs containing the genetic information that specifies a protein sequence defines a gene.

Figure 3-2



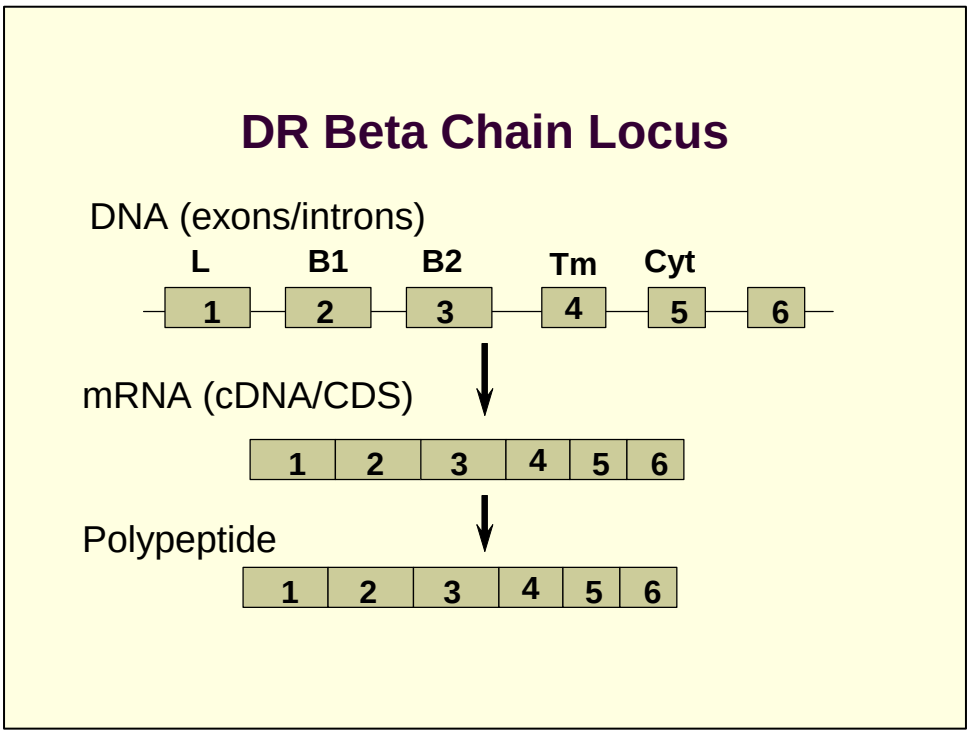
B. The genes that encode the alpha and beta chains of the class II proteins are located next to one another in the MHC.

C. DQB1 is the name of the locus that encodes the DQ beta chain and DQA1 is the name of the locus that encodes the DQ alpha chain. The numbers (e.g., DQA1) were added because there are other class II-like genes in the MHC.

D. The information for a single gene is found in multiple segments of double stranded DNA called exons [Figure 3-3].

QUESTION 1: Go to the website http://www.anthonlynolan.org.uk/HIG and find a copy of the current listing of the W.H.O. nomenclature for HLA alleles. What is the name of the locus that encodes the DR alpha chain?

Figure 3-3



CHAPTER 4

HLA ALLELES AND INHERITANCE

The purpose of this chapter is to describe the HLA diversity in the human population. It is this diversity that makes it is so difficult to find an unrelated donor with the same HLA alleles as a patient.

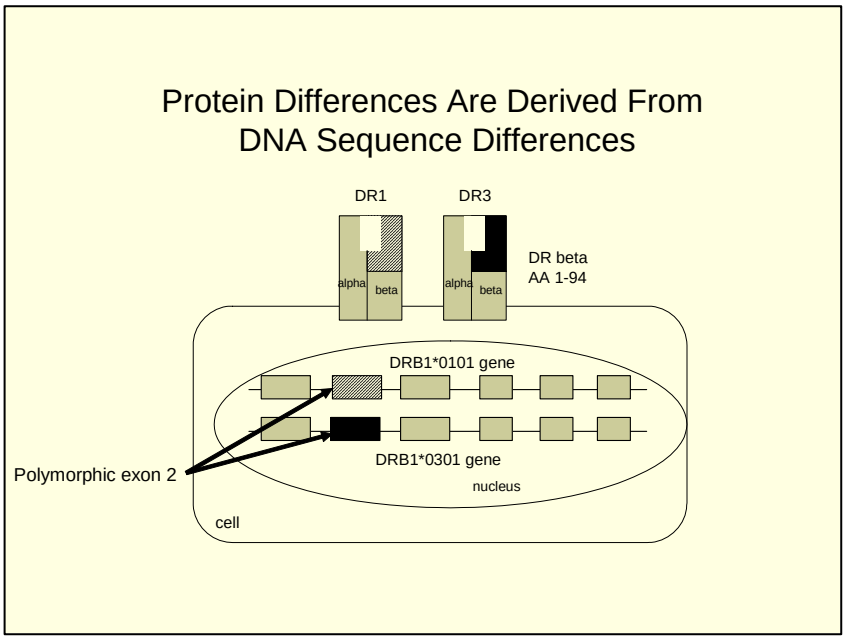
I. HLA genes are highly polymorphic (many forms). This means that, if we look at an HLA gene in unrelated individuals in a large population of humans, we will see that many people have a different DNA sequence for that gene. Each different DNA sequence is called an allele. These different gene sequences give rise to many different class I and class II allelic products (polypeptides).

A. Some loci have many alleles. For example, the DRB1 locus has over 560 alleles. Other loci have only a few alleles. For example, DRA has three alleles.

B. It is unlikely that two unrelated individuals carry the same class II alleles at DR, DQ and DP loci. A ballpark estimate for common HLA alleles is that only 1 person in 20,000 people from the same ethnic group will carry the same HLA alleles as another person in that group.

C. The two alleles carried by an individual is called their genotype.

Figure 4-1



II. In the class II alleles, the second exons which encode the first domain of each

polypeptide chain contain the majority of the allelic differences. Allelic differences can also be found in other exons. HLA class II typing usually focuses on defining the differences in the second exon [Figure 4-1].

QUESTION 1: Go to the web, http://www.anthonynolan.org.uk/HIG, and find a copy of the current list of the cDNA sequences of the DRB1 alleles. Practice using the alignment tool in the IMGT/HLA Sequence Database section.

III. The W.H.O. assigns names to HLA alleles. For example, alleles at the DRB1 locus are called: DRB1*010101, DRB1*010201, DRB1*0402.

A. Each HLA allele is designated by the name of the gene or locus followed by an asterisk and a 4-8 digit number indicating the allele [Figure 4-2]. For example, DPB1*0102 is an allele of the HLA-DP B gene; DQB1*030101 is an allele of the HLA-DQ B gene; DQA1*060101 is an allele of the HLA-DQ A gene. The first two numbers in the numerical designation of each allele are often based on the serologic type of the resultant protein (explained in Chapter 5) and/or the similarity to other alleles in that group. The next two numbers in the allele designation refer to the order in which the gene was discovered. The allele name should be viewed as simply a unique name for the allele and does not necessarily imply anything about its relationship to other alleles with similar names.

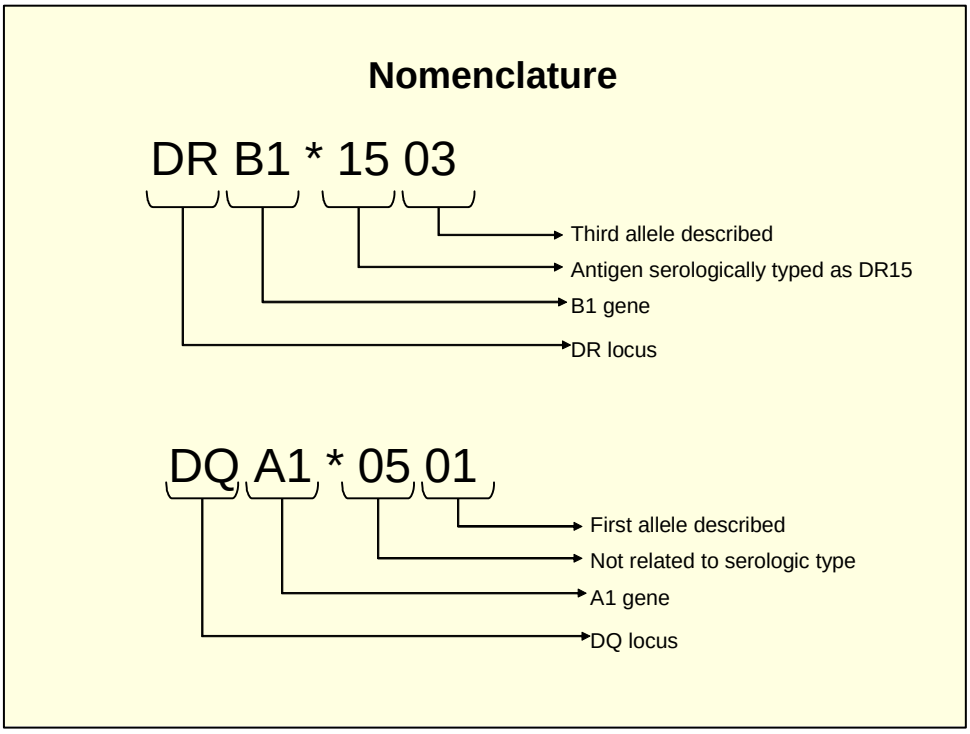
QUESTION 2: Use the W.H.O. nomenclature report to list a few alleles of the DQB1 locus.

B. Sometimes shorter names are used such as DRB1*0101. This designation means that the typing does not distinguish among alleles whose names all start with DRB1*0101. These alleles include DRB1*010101, DRB1*010102 and DRB1*010103.

C. Bone marrow donor registries may use letter codes to designate subsets of HLA alleles. For example, the letters AF in DRB1*14AF mean *1401 or *1409. For example, DRB1*04ABC means DRB1*0403 or *0404 or *0406 or *0407 or*0408 or *0410 or *0411 or *0417 or *0419 or *0420 or *0423. The letter codes can be found at http://bioinformatics.nmdp.org.

QUESTION 3: The DRB1*0101 allele is found in 1.9% of individuals in a population. The DRB1*0301 allele is found in 7%. What is the probability of finding an individual from that population who carries both alleles?

Figure 4-2



IV. DNA sequence differences between alleles can result in differences in the protein sequence or can be silent at the protein level.

A. An example of alleles that do not differ at the protein sequence level is DRB1*110101 and DRB1*110102. Both alleles have the same protein sequence but differ in the codons used to specify that sequence (silent or synonymous substitutions). The W.H.O committee indicates silent changes by adding an additional two digits to the allele name. Since the immune system detects differences in the HLA proteins, silent differences that do not change the protein sequence are not considered important in selecting an HLA matched donor.

B. N is added to indicated a null or nonexpressed allele (e.g., DRB4*01030102N). Since the N is optional, this allele is also correctly named DRB4*01030102. Null alleles are important to consider in the selection of an HLA matched donor. Digits 7 & 8 of this allele indicate that it differs from DRB4*0103010<u>1</u> in the intron or 5' or 3' regions of the gene (that is, the exon sequences are identical).

QUESTION 4: Using the class II sequence report, translate the first 20 codons of DRB1*010101 and DRB1*130201 into amino acids using Table 2-2.

QUESTION 5: Write down a different DNA sequence that would encode the same polypeptide.

ACT	GGT	TAC	TTC	GAG
Thr	Gly	Tyr	Phe	Glu
T	G	Y	F	E

V. Each person carries two copies of each HLA-DR, -DQ and -DP alpha and beta gene, one inherited from their mother and one from their father [Figures 4-3 and 4-4]. Furthermore, since all of the HLA genes are found on a single chromosome, each person has inherited one copy of chromosome 6 carrying one HLA gene complex from their mother and one copy carrying a second HLA gene complex from their father.

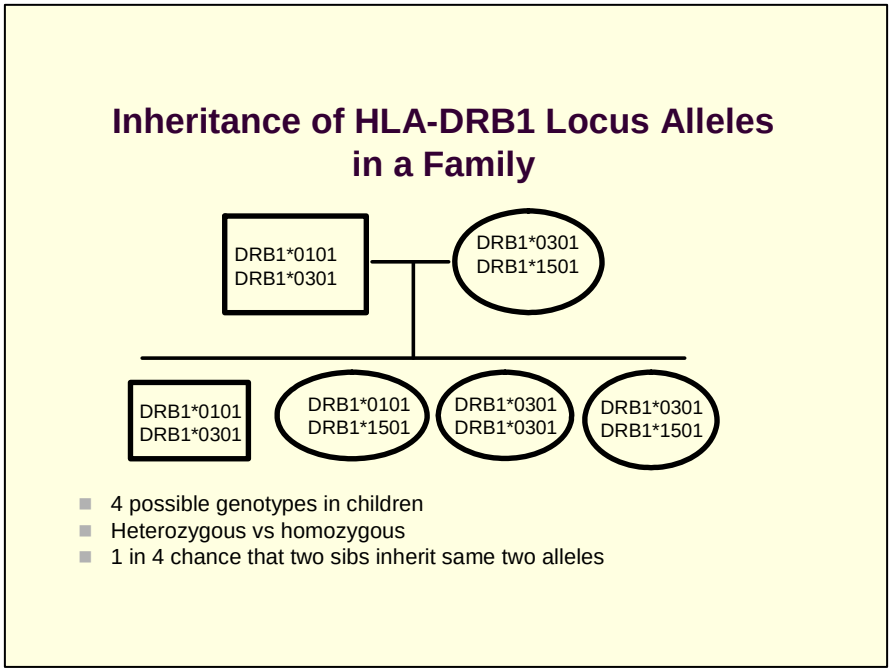
A. The HLA genes are codominantly expressed, that is, both copies encode proteins that are expressed on a single cell.

B. A person can have two identical alleles of a single gene (homozygous) or

may have two different alleles of a single gene (heterozygous).

C. A parent and child share one chromosome or haplotype (haploidentical). Two siblings (brothers and sisters) have a 1 in 4 chance of receiving the same two copies of chromosome 6 from their parents and becoming HLA identical. Thus, the class II genes, DR, DQ, DP, are inherited as a package. Traditionally, chromosomes from the father are labeled "a" and "b" and chromosomes from the mother are labeled "c" and "d".

Figure 4-3

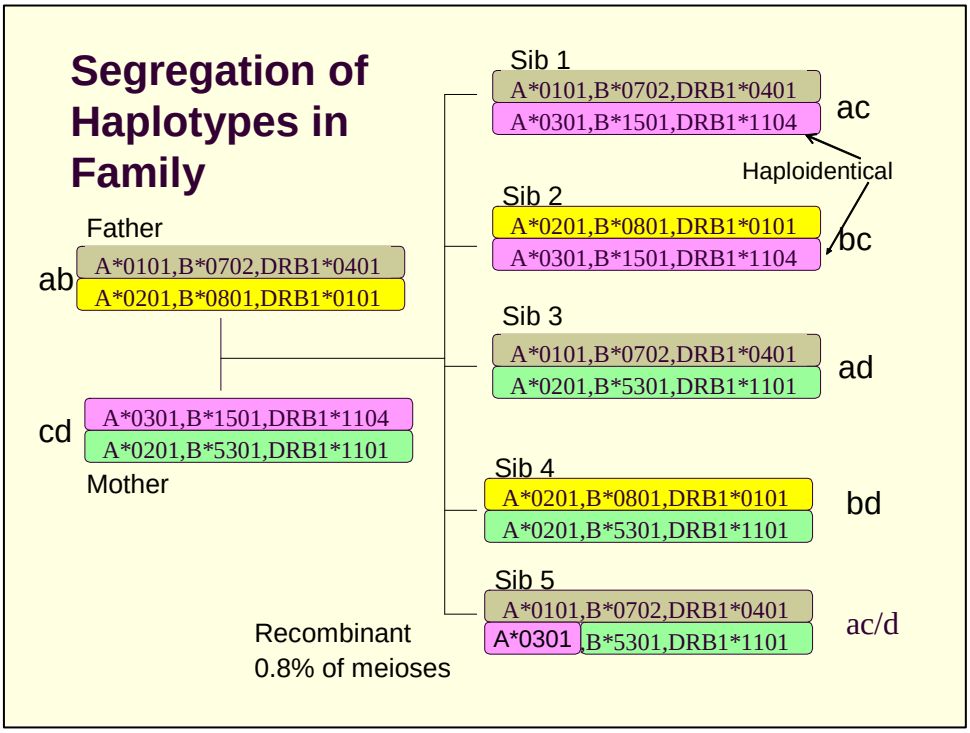


D. Sometimes the two copies of chromosome 6, which carry the class II genes, exchange gene segments, a process called reciprocal recombination [Figure 4-4]. This exchange reshuffles the DR, DQ, DP combinations. If this happens in the germ cells (egg and sperm), that person's offspring may inherit the new combination.

QUESTION 6: Father is DRB1*0402,DRB1*1103 and mother is DRB1*0101,DRB1*0301. What are the possible DRB1 allele combinations inherited by their children?

Can two children be DR identical (i.e., share DRB1 alleles)? Will any of the children be homozygous?

Figure 4-4



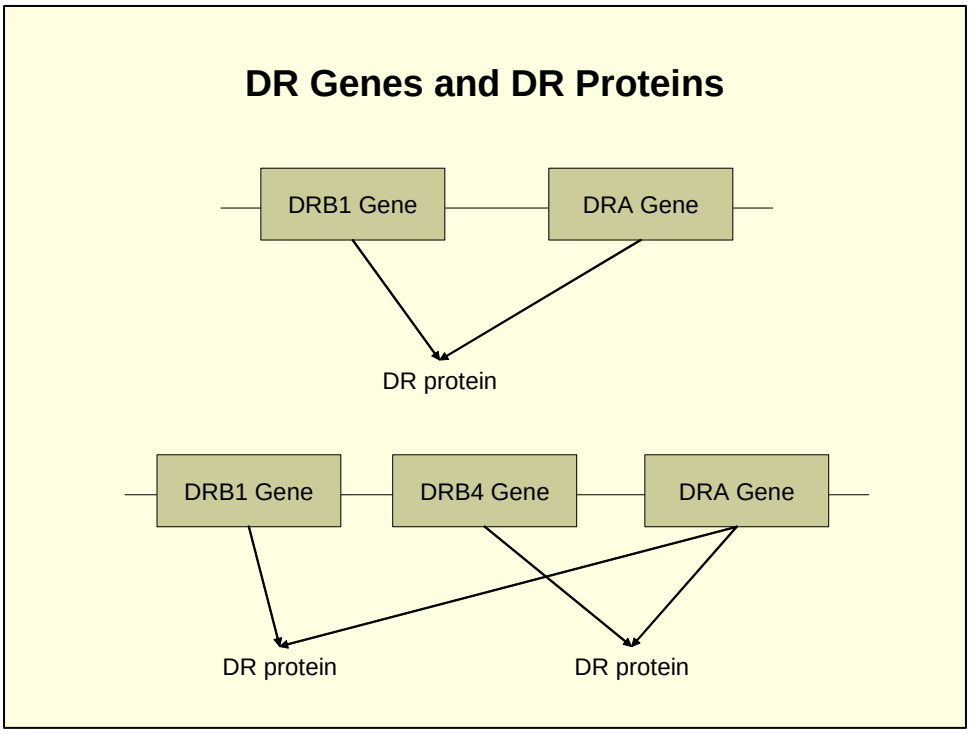
QUESTION 7: Father is a: DRB1*0403,DPB1*0201; b: DRB1*1101,DPB1*0201 and mother is c: DRB1*0101,DPB1*0201; d: DRB1*0302,DPB1*0101. What are the possible DR,DP allele combinations inherited by their children?

Could a child be DRB1*0403,DPB1*0201; DRB1*0101,DPB1*0101?

VI. The number of genes in the MHC may vary among different individuals carrying different haplotypes or different copies of chromosome 6.

A. Some chromosomes carry only one expressed DR beta gene [Figure 4-5]. An expressed gene is one that encodes a polypeptide. The DR beta locus is called DRB1.

Figure 4-5



B. Other copies of chromosome 6 carry two expressed DR beta genes. One locus is DRB1. The second beta chain locus is called DRB3 or DRB4 or DRB5. Individuals who carry haplotypes containing two DR beta loci express two different DR molecules encoded by that haplotype [Figure 4-5].

C. Each chromosome carries a single DQA1, DQB1, DPA1, and DPB1 gene and, thus, encodes a single DQ and a single DP molecule.

D. Some class II genes in the MHC are pseudogenes (e.g., DRB2) and are defective in some way. Others (e.g. DM,DO) specify proteins that are not considered important for transplantation matching.

VII. Particular HLA-DR alleles are often found together on chromosomes that carry two DR beta loci.

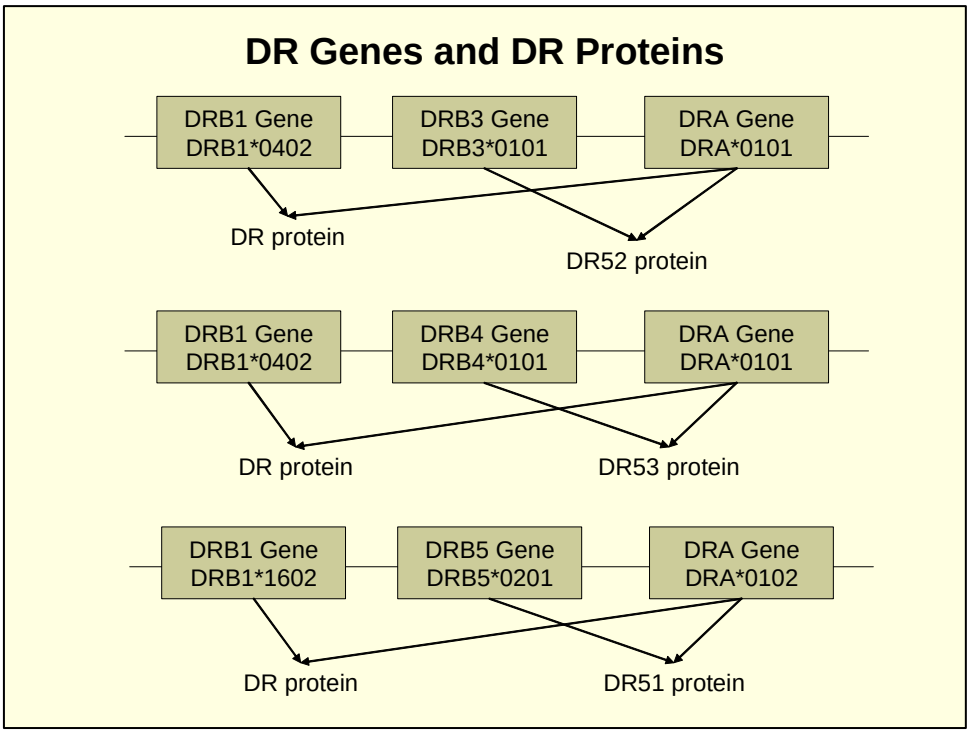
A. DRB1 and DRB4. Alleles whose names start with DRB1*04, DRB1*07, and DRB1*09 are usually found on the same chromosome as alleles whose names begin with DRB4 [Figure 4-6].

B. DRB1 and DRB3. Alleles whose names start with DRB1*03, DRB1*11, DRB1*12, DRB1*13, and DRB1*14, are usually found on the same chromosome as alleles whose names begin with DRB3 [Figure 4-6].

C. DRB1 and DRB5. Alleles whose names start with DRB1*15 or DRB1*16 are usually found on the same chromosome as alleles whose names begin with DRB5 [Figure 4-6].

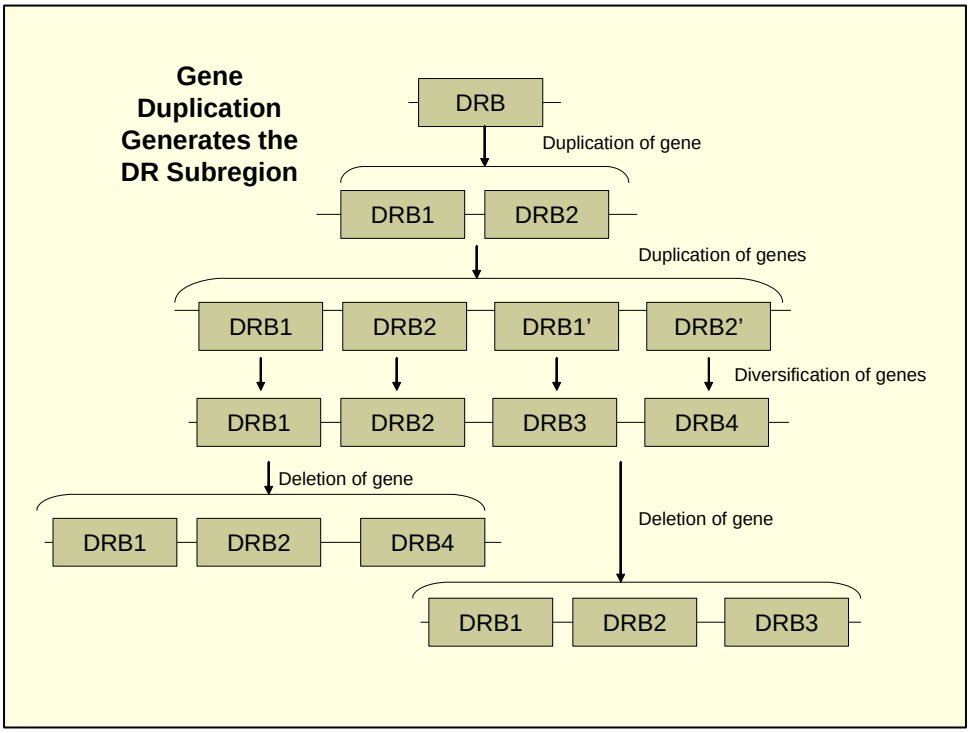
D. These associations are common but not always found. For example, DRB1*1501 can be found on a chromosome without a DRB5 allele and DRB5 can be found on a chromosome without a DRB1*1501 allele.

Figure 4-6



E. The names of the DRB loci (DRB1, DRB3, DRB4, DRB5) are derived from the hypothesized evolutionary origin of the DRB genes [Figure 4-7]. These genes are thought to have arisen from duplication, deletion, and diversification over evolutionary time.

Figure 4-7



Question 8: Draw the class II loci present on the two copies of chromosome 6 from a person who carries DRB1*1304, DRB1*0402, DRB3*0201, DRB4*0101, DQA1*0201, DQB1*0201, DPA1*0104, DPB1*0401 alleles.

Is the person homozygous or heterozygous for DQ and DP alleles?

VIII. The class II genes are closely related to one another and share many segments of sequence. This sharing may result in difficulties in distinguishing specific HLA alleles.

A. Some alleles fall into families based on similarities in their DNA sequences. These families are designated by their similar nomenclature. For example, alleles, DRB1*080101, DRB1*080203, DRB1*080302, DRB1*080404, are very similar in sequence to one another as denoted by the first two digits 08.

B. Some segments of each class II gene sequence are shared by all alleles.

C. Other segments of the gene sequences are polymorphic or vary among alleles. Alleles from different allele families may share these polymorphic segments of sequence.

QUESTION 9: Compare the DNA sequences of DRB1*010201 and DRB5*0202 and identify sequences that are shared. Some sequences are common to many DRB alleles; other sequence segments are found only in a few alleles. Identify the shared sequences in these two alleles. [Hint: sequences are usually compared to one sequence (e.g., DRB1*0101) and nucleotides that are identical are indicated by a dash (-).]

Compare the DNA sequences of DRB1*0103 and DRB1*130101 and identify sequences that are shared.

Observe the nucleotides that differ between alleles.

IX. Class I loci are also highly polymorphic.

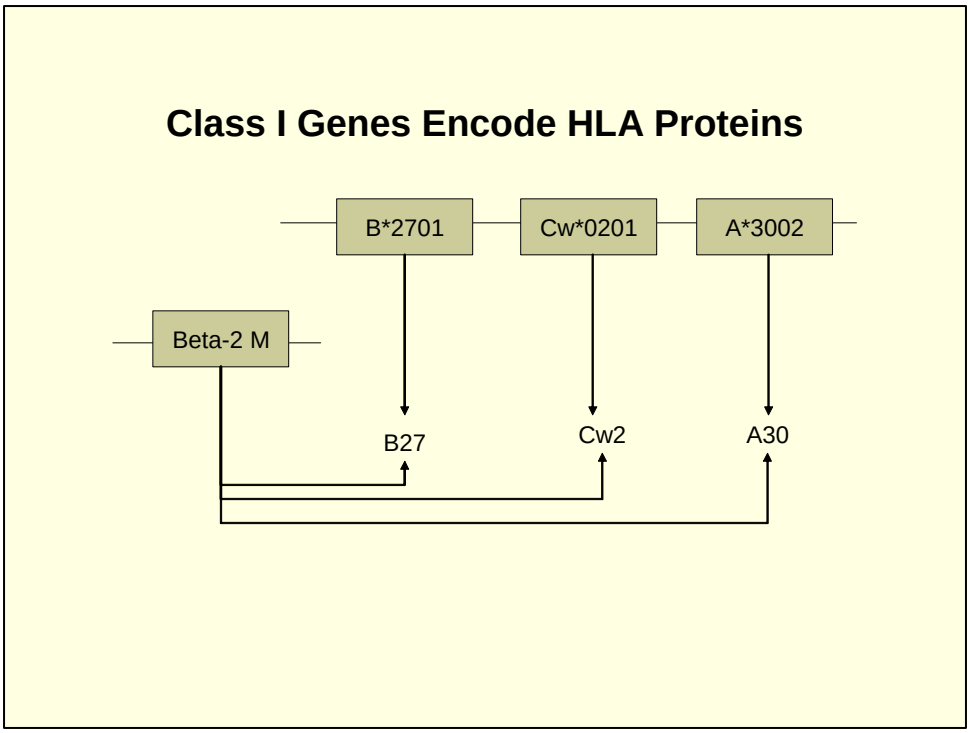
A. In the class I alleles, the second and third exons which encode the first and second domains of the alpha chain contain the majority of the allelic differences.

B. The class I loci are very similar to one another. The sequences that identify an A locus allele from a B or C locus allele are located in the first exon (encoding the leader) or in the 3' exons (encoding the transmembrane and cytoplasmic regions) or in the introns separating the exons.

C. Alleles are designated by numbers. For example, A*0202 and A*03010101 are alleles of the HLA-A locus. B*2701 is an allele of the HLA-B locus [Figure 4-8].

D. Null or nonexpressed alleles may be indicated by an “N” in the allele name e.g., A*2411N. An “L” indicates an allele which exhibits decreased protein expression. An “S” indicates that the HLA protein product is secreted. While null alleles are important to consider in matching of patient and donor, the importance of low vs. normal expression levels or secreted vs cell surface molecule is not yet known. “Q” means that the level of expression is not known but expected to be low or absent.

Figure 4-8



QUESTION 10: Use the W.H.O. nomenclature report to list a few alleles of the C locus. The locus is called Cw so that it is not confused with the complement loci.

E. The HLA genes are closely related to one another and share many segments of sequence. This sharing may result in difficulties in distinguishing specific HLA class I alleles or molecules.

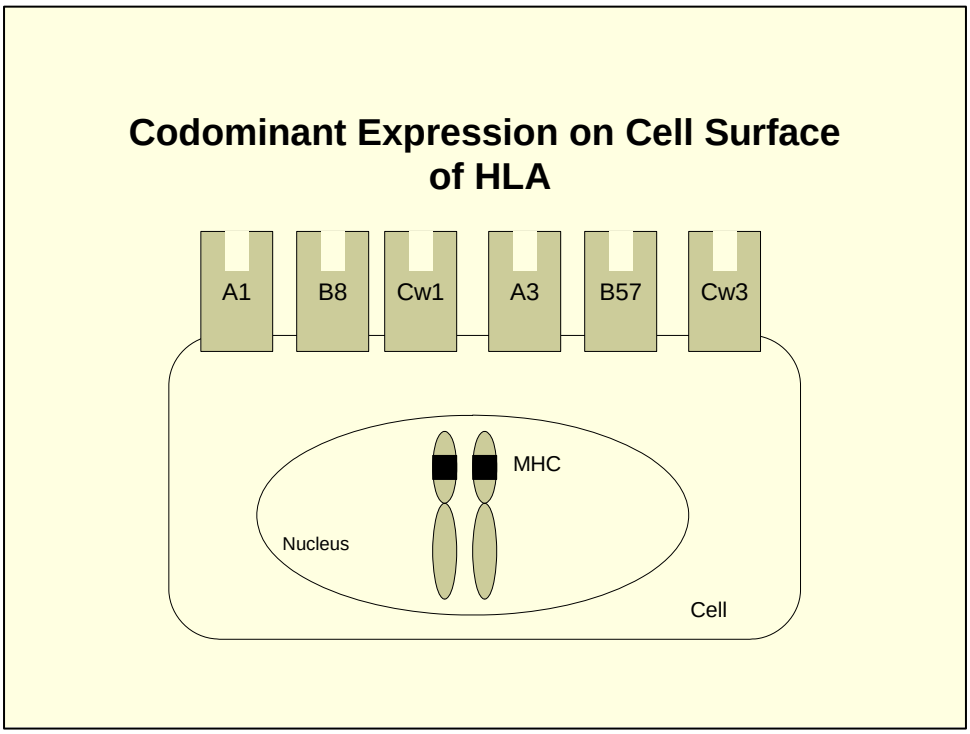
QUESTION 11: Go to the web page: http://www.anthonynolan.org.uk/HIG and find a copy of the current list of the cDNA sequences of the class I alleles. Compare the nucleotide

sequences of B*070201, B*0801, and B*4201 and identify sequences that are shared between alleles.

X. Each person carries two copies of each HLA-A, -B, and -C locus, one inherited from their mother and one from their father.

A. These genes are codominantly expressed, that is, both copies encode proteins that are expressed on a single cell [Figure 4-9]. If this is a cell of the immune system, it will also express all the class II alleles the cell carries.

Figure 4-9



B. A parent and child share one chromosome or haplotype (haploidentical). Two siblings have a 1 in 4 chance of receiving the same two chromosomes (HLA identical).

C. Sometimes reciprocal recombination alters these allele combinations.

XI. Other class I molecules (HLA-E, HLA-G) are expressed in specific tissues at specific times. For example, HLA-G is expressed at the maternal/fetal interface. These molecules are less polymorphic than HLA-A, -B, and -C and are not considered

important for HLA matching.

References:

Parham, P., Lomen, C.E., Lawlor, D.A., Ways, J.P., Holmes, N., Coppin, H.L., Salter, R.D., Wan, A.M., Ennis, P.D. 1988. Nature of polymorphism in HLA-A, -B, -C molecules. Proc. Natl. Acad. Sci. USA 85:4005-4009.

Parham, P. and Ohta, T. Population biology of antigen presentation by MHC Class I molecules. Science 272:67-74, 1996.

CHAPTER 5

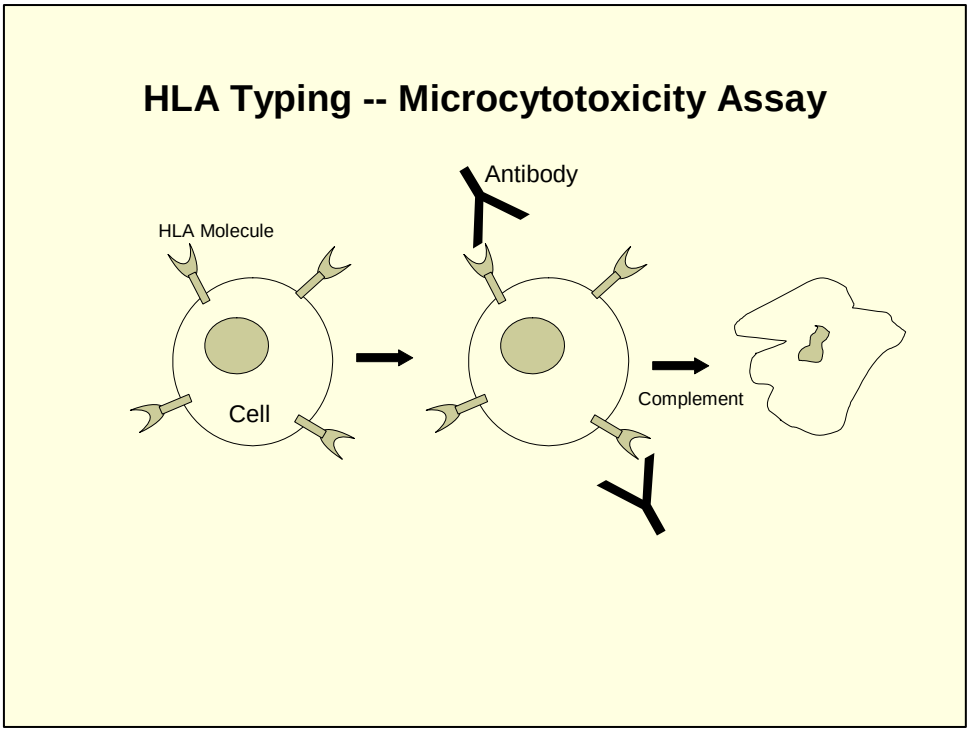
HLA TYPES DEFINED BY SEROLOGY

The purpose of this chapter is to describe the terms used by HLA serologists to describe the many forms of the HLA molecules found in the human population.

I. Serology is used to identify the HLA proteins on the surfaces of cells.

A. The different forms of the HLA proteins found in the human population may be detected serologically using antibodies (human alloantisera or monoclonal antibodies) in a test called a microcytotoxicity assay [Figure 5-1].

Figure 5-1



B. The alloantisera utilized in this assay are obtained from humans who have been sensitized to foreign HLA molecules by pregnancy or previous transplant. These antibodies are used as reagents to identify serologic determinants or specificities (or HLA types) by reacting with the HLA molecules present on the cell surfaces.

C. Because humans mount an immune response to foreign HLA molecules, the HLA molecules are often called antigens. An antigen is any substance that an antibody can bind.

II. Each serologic specificity (or HLA type) is designated by a letter indicating the kind of HLA antigen (A, B, C, DR, DQ, DP) and a number. The number indicates the order in which the type was originally discovered or its relationship to other defined antigens. For examples, DQ7 is a serologic specificity localized on an HLA-DQ antigen, DR4 is a serologic specificity localized on an HLA-DR antigen, and B8 is a serologic specificity localized on an HLA-B antigen.

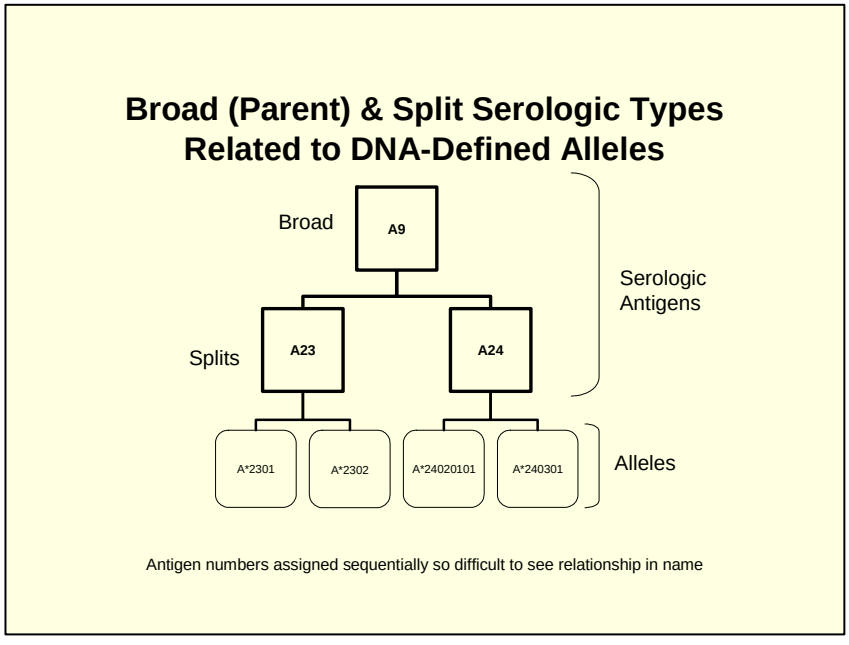
QUESTION 1: Go to the web http://www.anthonynolan.org.uk/HIG and find a list of the accepted serologic specificities.

III. Broad antigens and splits

A. Some antibodies define clusters of HLA proteins (broad antigens) that are similar and that are indistinguishable using a particular antibody preparation [Figure 5-2]. For example, the A2 specificity is likely found on at least 90 different HLA-A molecules.

B. Other more specific antibodies may allow the definition of subdivisions or "splits" of broad antigens. A nontechnical example of splits which may be clearer is that a stranger may recognize you as a member of the Jones family (broad specificity) but a friend may recognize you as Mary Jones (a split).

Figure 5-2



EXAMPLE: Antibodies define A9, a broad antigen [Figure 5-2]. A9 was the 9th HLA-A specificity described. The splits of A9 have been labeled A23 and A24. Thus, the specificity A23 is often called a split of the broad specificity A9 and can be also written as A23(9). Because these serologic specificities were named as they were discovered, the splits of A9 were designated as A23 and A24 because they were the 23rd and 24th HLA-A antigens discovered. An individual who types as A24 using A24-specific antibodies would also type as A9 using A9-specific antibodies but they would not type as A23.

C. Different HLA alleles defined by DNA typing can specify HLA proteins which are indistinguishable using serologic typing. For example, an individual carrying the DRB1*040101 allele would have the same serologic type (DR4) as an individual carrying the DRB1*0412 allele. Thus, DRB1*040101 and DRB1*0412 are splits of the broad specificity DR4. These splits are identified by DNA typing but, because we do not have serologic reagents specific enough to define this split, we can not serologically distinguish between the antigens specified by the two alleles, DRB1*040101 and DRB1*0412.

QUESTION 2: Go to the library and find a copy of G.M.T. Schreuder, et al. 2005. The HLA Dictionary 2004: a summary of HLA-A,-B,-C,-DRB1/3/4/5,-DQB1 alleles and their association with serologically defined HLA-A,B,-C,-DR and –DQ antigens. Tissue Antigens 65:1-55. Look up the tables that list the serologic specificities of HLA-A,-B,-DR and links them to the alleles.

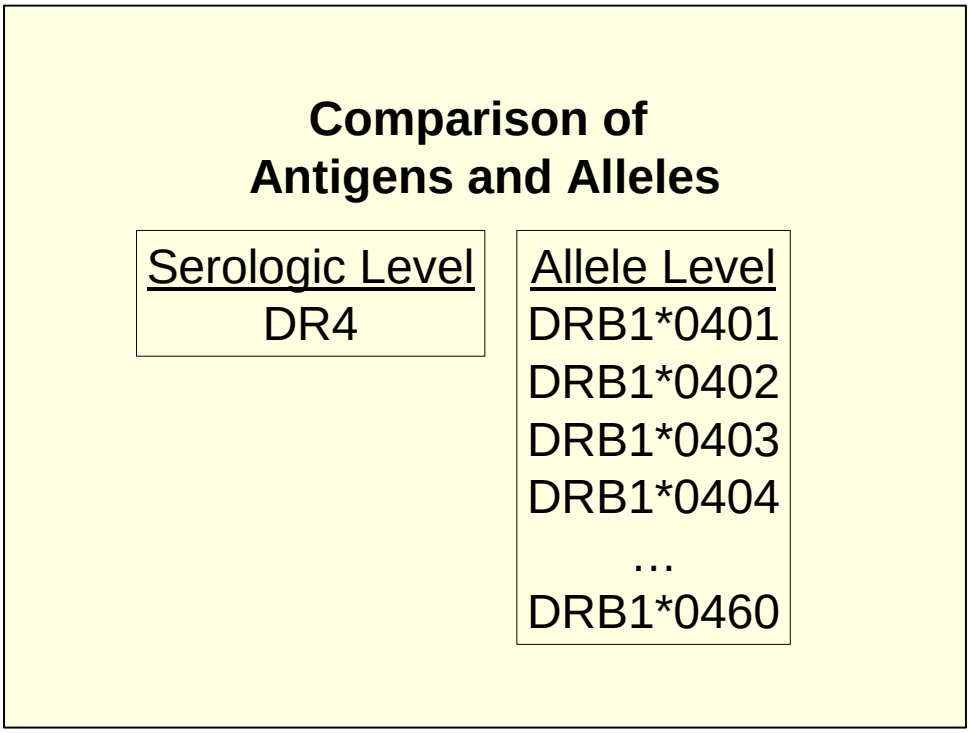
D. Although HLA types have been defined using serology for many years, the available antibodies lack the resolution required to identify all of the specific products of the HLA alleles [Figures 5-3]. For example, serology can not determine whether a donor and recipient who are typed as DR4 carry the same alleles of DR4. This is one reason that, in some situations such as typing for bone marrow transplantation, that serology has been replaced by DNA-based typing methods.

QUESTION 3: Do the donor and recipient below carry the same DR alleles?

Donor: DR8,DR3

Recipient: DR8,DR3

Figure 5-3



QUESTION 4: Father is DR4, DR11 and mother is DR1, DR3. What are the possible DR types inherited by their children?

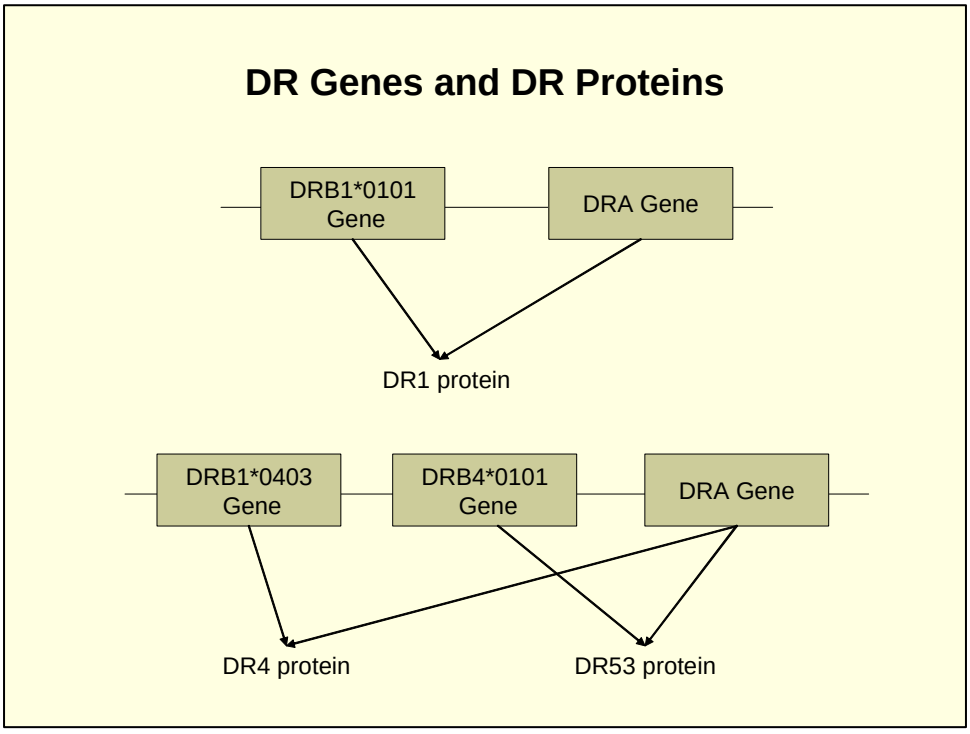
What are one possible set of DR alleles that might be found in the family?

E. Cellular assays such as the mixed lymphocyte culture (MLC) measure the differences in class II proteins between individuals. A cellular assay is more sensitive in detecting HLA differences than serologic typing since even a single amino acid differences can cause stimulation of white blood cells. Because of the difficulty in generating and maintaining cellular reagents and because of difficulties in interpreting experimental results, these types of assays have been replaced by DNA-based typing methods.

IV. The HLA genes encode specific HLA antigens.

A. The DR molecules encoded by the DRA and DRB1 loci carry the DR1, DR15, DR16, DR3, DR4, etc. serologic specificities [Figure 5-4].

Figure 5-4



B. Molecules encoded by the DRA and DRB3 loci carry the DR52 serological specificity Molecules encoded by the DRA and DRB4 loci carry the DR53 serological specificity. Molecules encoded by the DRA and DRB5 loci carry the DR51 serological specificity.

C. Molecules encoded by the DQA1 and DQB1 loci carry DQ serological specificities. Molecules encoded by the DPA1 and DPB1 loci carry DP serological specificities. [Note: There are very few alloantisera that detect DP types.]

D. Molecules encoded by the A, B, and C loci carry HLA-A, HLA-B, and HLA-C serologic specificities. HLA-C is poorly defined by serologic reagents.

E. Individuals who carry haplotypes (or chromosomes) with two DR beta loci express two different DR molecules specified by that haplotype and, since the allele associations are usually fixed, the combination of DR molecules can usually be predicted [Figure 5-4]. [Review Chapter 4.]

1. Individuals who express a DR4 molecule usually also express a DR53

molecule and carry a DR4 beta chain allele (e.g., DRB1*0402) and a DR53 beta chain allele (e.g., DRB4*01010101) [Figure 5-4]. This is also true for DR7 and DR9.

2. Individuals who express a DR3 molecule usually also express a DR52 molecule and carry a DR3 beta chain allele (e.g., DRB1*030101) and a DR52 beta chain allele (e.g., DRB3*01010201). This is also true for DR11, DR12, DR13, and DR14.

3. Individuals who express a DR15 molecule usually also express a DR51 molecule and carry a DR15 beta chain allele (e.g., DRB1*150101) and a DR51 beta chain allele (e.g., DRB5*010101). This is also true for DR16.

QUESTION 5: Draw the DR loci present in the MHC of a person who is serologically typed as DR11, DR52.

List one set of possible alleles. [Hint: There are more than one possible allele for each locus, just pick one.]

F. There are exceptions to these associations. For example, some DR7 haplotypes carry a DRB4 allele but it is not expressed as a DR53 molecule on the cell surface because of a defect in the gene (e.g., a termination codon halting protein synthesis). The allele specifying this null allele is sometimes labeled with an “N” (e.g., DRB4*01030102N).

QUESTION 6: Father expresses HLA-A2,A3,B27,B53 and mother expresses HLA-A2,A11,B51,B71. What are the possible HLA-A and -B types that their children might express?

One of the children expresses HLA-A2,B27,B71. What are the most likely haplotypes carried by the parents?

V. Complications.

A. Unfortunately, the serologic types associated with some HLA alleles are not yet known. For example, the serologic type specified by B*0808 is not known.

B. In some cases, the name of the allele does not reflect its serologic type. For example, the HLA molecule specified by B*5002 is serologically typed as B45.

QUESTION 7: What is the serologic type assigned to the following alleles: A*680101, B*1301, B*1804, DRB1*030101, DRB1*1122?

QUESTION 8: A patient carries A*02010101 and A*2409N alleles. What is the serologic type of this patient? Would a donor serologically typed as A2, A24 be a match for this patient?

References:

G.M.T. Schreuder, et al. 2005. The HLA Dictionary 2004: a summary of HLA-A,-B,-C,-DRB1/3/4/5,-DQB1 alleles and their association with serologically defined HLA-A,B,-C,-DR and –DQ antigens. Tissue Antigens 65:1-55.

HLA websites with serology information:

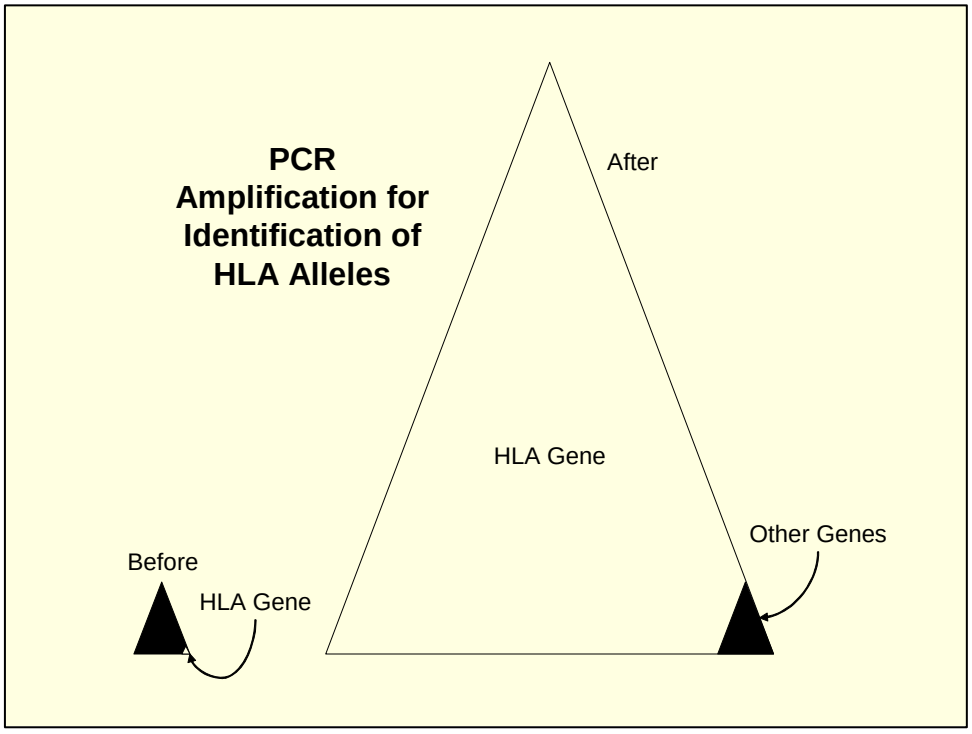
http://www.anthonynolan.org.uk/HIG/
http://www.ebi.ac.uk/imgt/hla/

CHAPTER 6

GENE AMPLIFICATION USING THE POLYMERASE CHAIN REACTION

The polymerase chain reaction (PCR) is a rapid way of isolating large quantities of specific HLA genes for HLA typing [Figure 6-1]. Using this method, we can generate millions of copies of a specific gene.

Figure 6-1



I. PCR is a method of DNA synthesis [Figure 1-7] using:

A. A DNA polymerase that is not destroyed when heated (thermostable): Taq polymerase. This enzyme comes from bacteria that live in hot springs.

B. DNA template: Crude cell lysates containing heat denatured DNA can be used although better results are obtained with DNA that has been more extensively purified.

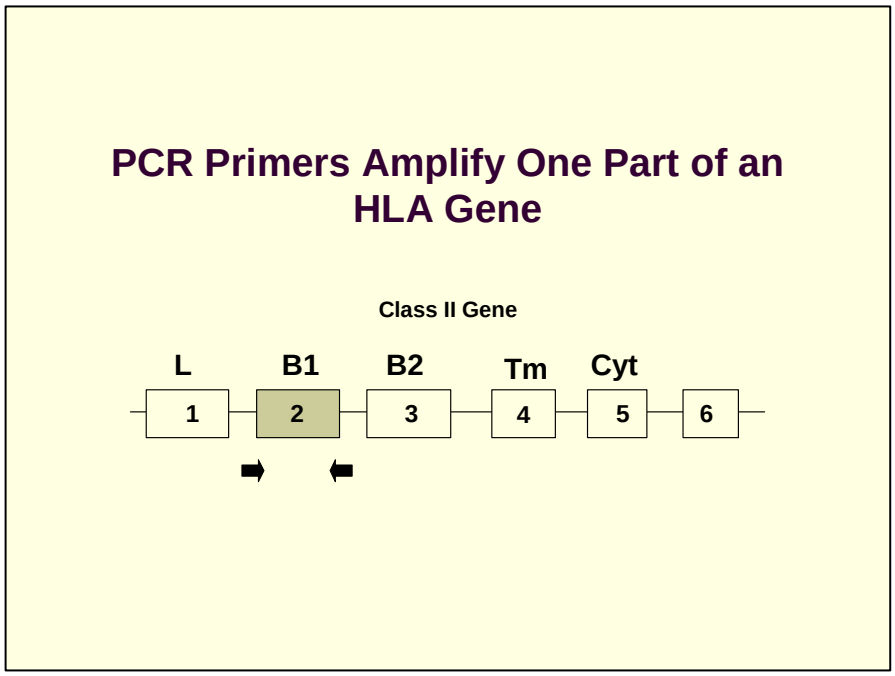
C. Nucleotides.

D. Primers.

1. One of the advantages of PCR is that it uses two primers so that both

DNA strands are copied. The two single stranded primers must flank the region to be amplified. Since most of the polymorphism of the class II genes is found in the second exon, primers usually are designed to flank this region [Figure 6-2]. Primers for class I alleles usually flank the polymorphism found in exons 2 and 3.

Figure 6-2



2. The primers hybridize to opposite strands of the DNA ladder. One primer is called a forward (or 5' or sense or coding) primer and the other primer is called a reverse (or 3' or antisense or noncoding) primer. Remember the discussion in Chapter 1 about DNA synthesis.

3. The primer sequences must complement (match) their target sequence and be sufficiently long (20-30 nucleotides) to bind to only the HLA gene that you want to amplify.

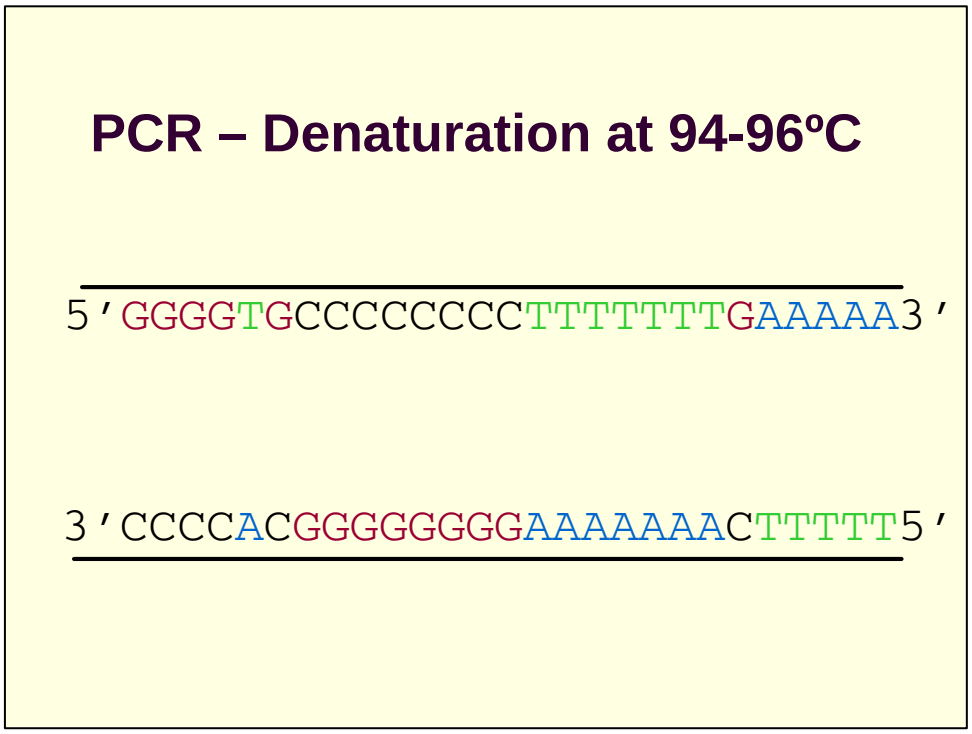
4. A primer should not contain any stretches of sequence that would anneal to the other primer (form primer dimers). For example, 5' AGCACTTTT and 5' TCCATAAAA would not be a good choice because the stretches of Ts and As would anneal. The polymerase would add complementary nucleotides to make a short double strand DNA called a primer dimer.

```
5′   AGCACTTTT 3′          ===> 5′AGCACTTTTATGGA
     3′   AAAATACCT 5′           TCGTGAAAATACCT
```

II. DNA is amplified by using a three step procedure:

A. DNA denaturation (94-96^{o}C) to generate a single stranded template [Figure 6-3].

Figure 6-3



.

Annealing of the primers (45-65^{o}C) using hybridization conditions that guarantee that the primers will bind to perfectly matched sequences (target sequence) and not to sequences that are not matched [Figure 6-4]. The temperature of annealing controls, in part, the specificity of the amplification [Hint: Remember Chapter 1]. The higher the temperature, the more specific the amplification until you get to the melting temperature of the primer. At that temperature, amplification does not work very well, if at all.

Figure 6-4

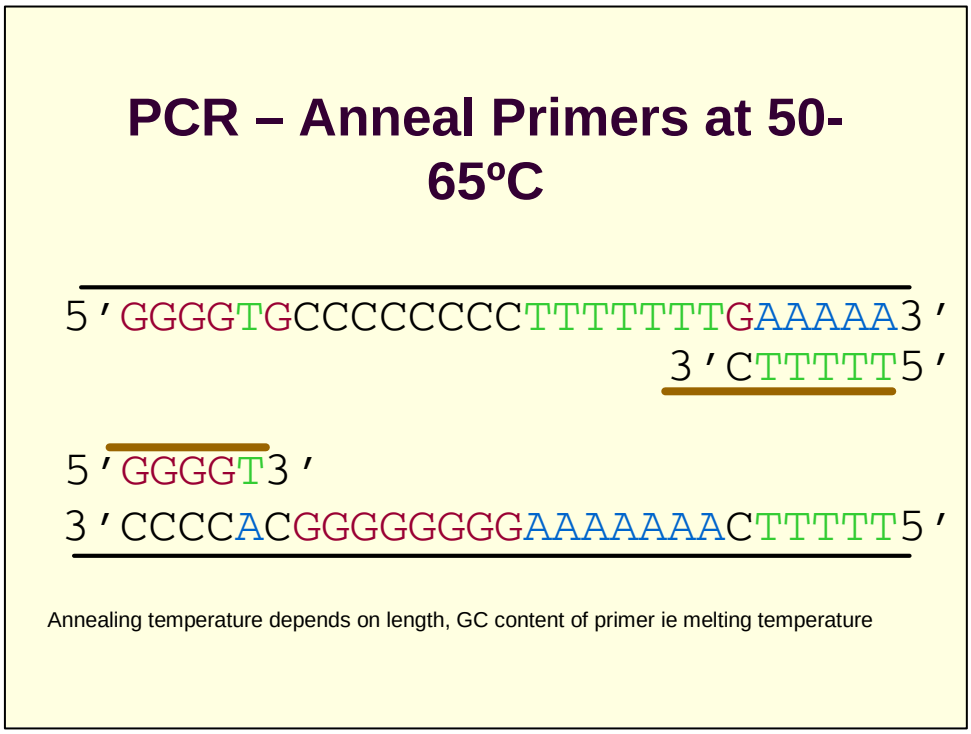
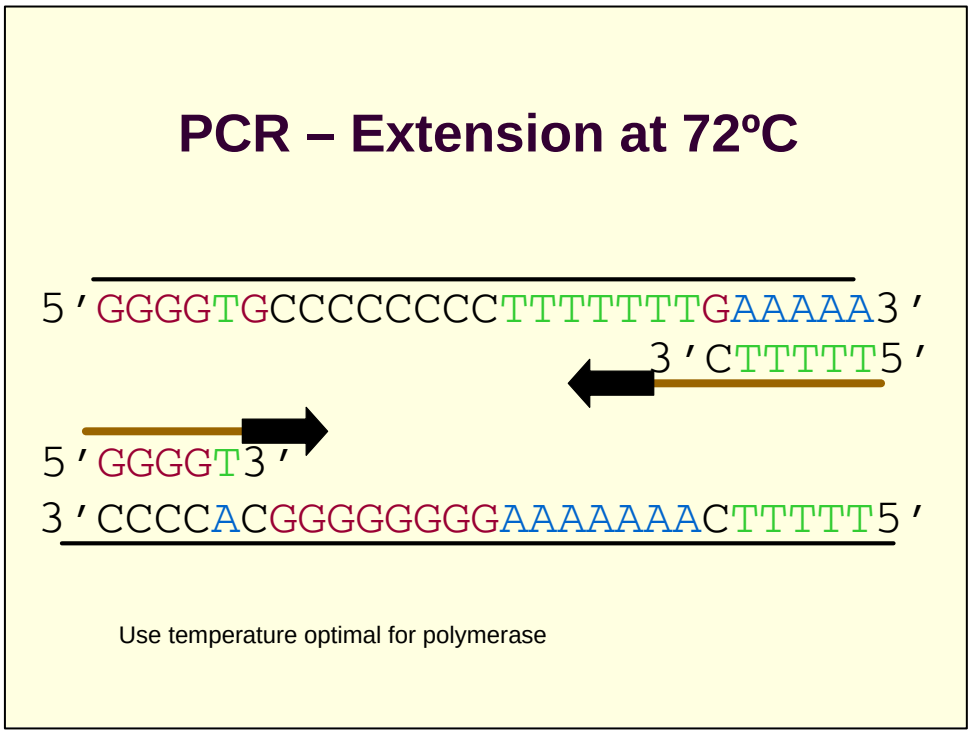


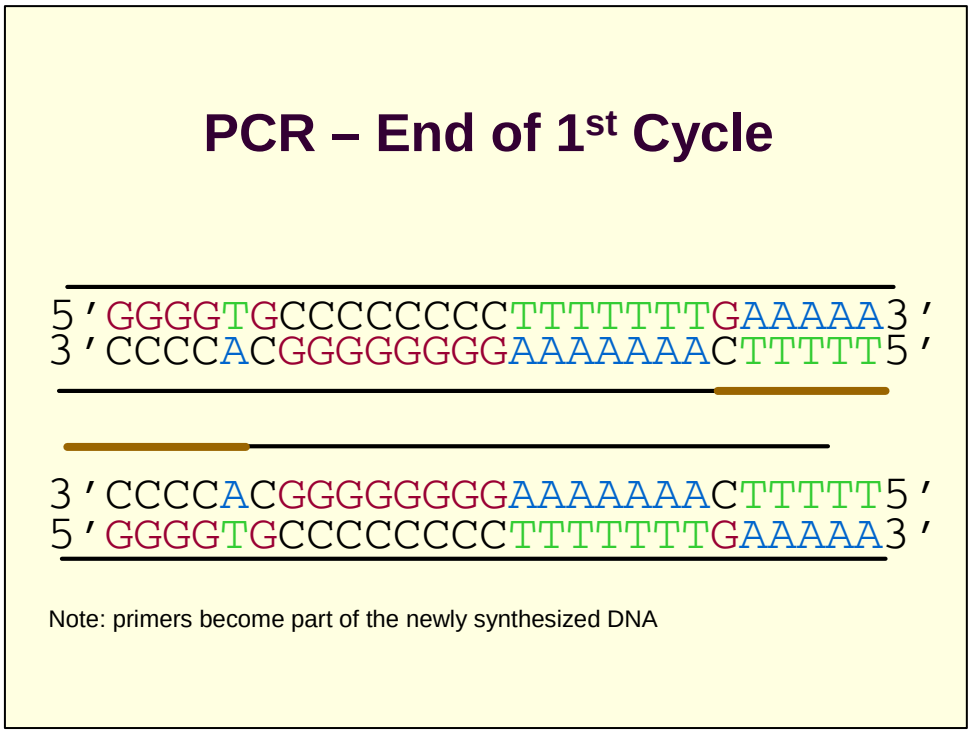
Figure 6-5



C. Extension (synthesis of DNA) (around 72°C) [Figure 6-5].

D. At the end of the first cycle, amount of target DNA is doubled [Figure 6-6].

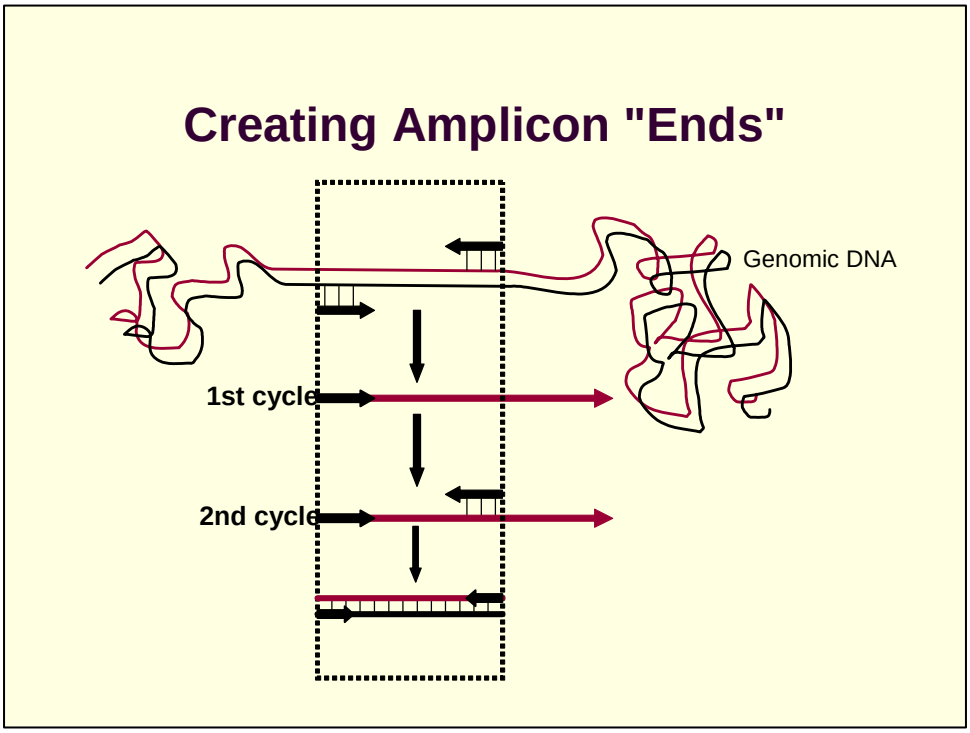
Figure 6-6



E. The three steps are repeated over and over by simply changing the temperature of the reaction mix using an instrument called a thermal cycler. The newly synthesized strands serve as templates for synthesis in the next cycle (another advantage of PCR). Usually 25-30 cycles of amplification are carried out to yield millions of copies of the gene of interest.

F. Figure 6-7 illustrates how the reaction yields an amplicon of defined length after many cycles. Question 2 gives you a chance to demonstrate this to yourself.

Figure 6-7.



QUESTION 1: What would happen if you used a primer set that had the following sequences to amplify a gene from human DNA:

5' CC
5' TA

Would primers that were 20 bases long be more or less specific in priming the amplification reaction? Why? [Hint: Remember the discussion about the specificity of restriction enzymes. What is the probability that you will find a two nucleotide-long sequence in the DNA compared to the probability of finding a 20 nucleotide-long sequence?]

QUESTION 2: Draw out a PCR reaction amplifying the DNA listed below using primers listed. What happens at each step of the amplification (denaturation, annealing, extension)? What do the amplified products look like after the second cycle of amplification? After the third cycle?

5' AATAATAATAATAATAATTATGGCGGCTATCGGCGGCGGCTTTATTATTATTATTCCCC 3'
3' TTATTATTATTATTATTAATACCGCCGATAGCCGCCGCCGAAATAATAATAATAAGGGG 5'

Primers: 5'TTATG and 5'AAAGC

How many base pairs in length is the amplified fragment?

If you started with one piece of DNA, after 3 cycles, how many copies do you have? After 10 cycles?

QUESTION 3: The sequences of the two primers that are used to amplify HLA-A alleles are:

Forward: 5’ CCC AGA CGC CGA GGA TGG CCG 3’
Hint: 5’UTR-exon 1 boundary
Reverse: 5’ GCA GGG CGG AAC CTC AGA GTC ACT CTC T 3’
Hint: Intron 3; remember 5’ to 3’ and complementary DNA strands running in opposite directions

Use the HLA-A sequence in Figure 2-4 to locate these sequences in the DNA. Remember that DNA synthesis proceeds 5' to 3' and that we are only interested in the polymorphic regions (second and third exon).

Why can't you find the primer sequences in the cDNA sequence of HLA-A in Figure 2-5?

QUESTION 4: Locate the C locus primers used to amplify exons 2 and 3: AGCGAGG(GT)GCCCGCCCGGCGA and GGAGATGGGGAAGGCTCCCCACT. [Hint: (GT) means that two primers are synthesized, one with G at this position and one with T in this position. Hint: Look at the genomic sequences because the primers anneal in the

introns.

Note: Primers described in Cereb, N. et al. 1996. Nucleotide sequences of MHC class I introns 1,2, and 3 in humans and intron 2 in nonhuman primates. Tissue Antigens 47:498-511. Correction in Tissue Antigens 48:235-236, 1996.

QUESTION 5: Design primers to amplify the second exon of the DQA1 alleles. You will need to find a publication listing all the DQA1 alleles to do this [Hint: WHO HLA Nomenclature web site.].

III. To determine if DNA has been amplified following the PCR reaction, the PCR product is often analyzed by gel electrophoresis to identify a fragment of a specific size. For example, amplification of exon 2 of the DRB genes will yield a fragment of DNA approximately 270 base pairs in length and amplification of exons 2 and 3 of an HLA-B gene will yield a fragment of DNA approximately 1000 base pairs in length.

   A. Note that a primer dimer may also detected on the gel during electrophoresis. It will be smaller that the HLA amplicon so it will move farther down the gel.

IV. Advantages of PCR

   A. Large quantities of the HLA gene of interest will allow the rapid detection of the HLA type.

   B. Amplification of a specific HLA gene means that detection of an HLA type will not be influenced by the presence of other HLA genes or non-HLA genes during probe hybridization or sequencing.

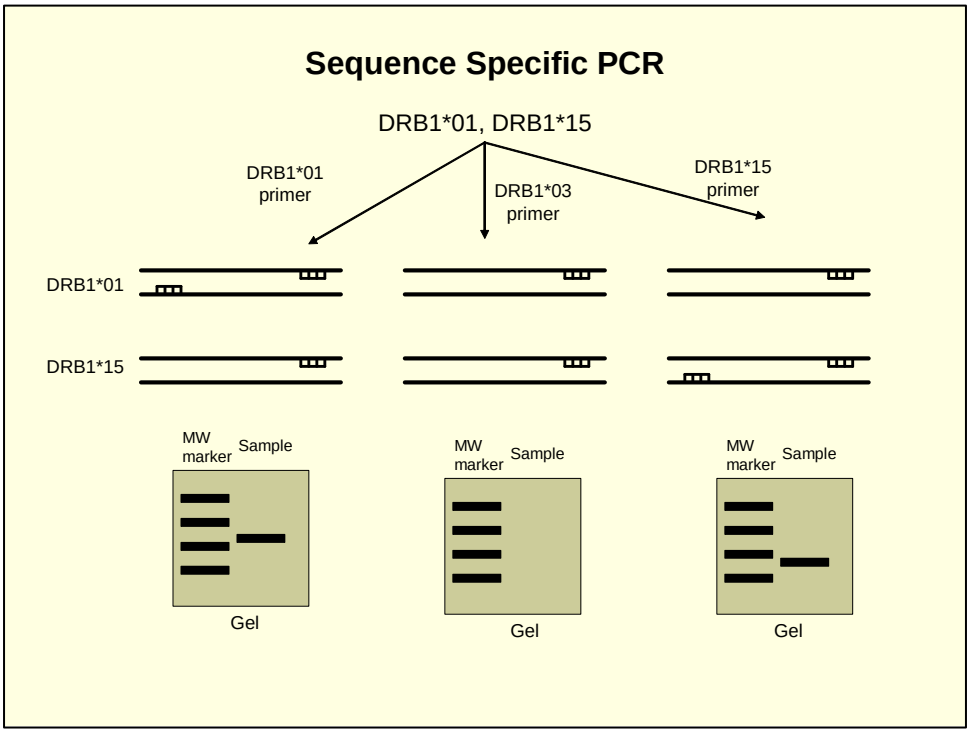
V. Contamination with previously amplified DNA is the biggest potential problem. It is relatively easy to contaminate the work area with amplified DNA because there are so many copies of an HLA gene following amplification.

QUESTION 6: Assume that a PCR amplification generates 27 million (27×10^6) copies of a piece of DNA in a 100 microliter reaction volume. How many copies will be in a 1 microliter drop that was sucked into the barrel of a pipetter?

VI. Allele or group-specific PCR or sequence-specific primer (SSP) typing. A modification of this approach is called ARMS (amplification refractory mutation system).

A. In this technique, PCR primers are designed to anneal only to a specific set of alleles or to a single allele [Figure 6-8]. One or both primers include sequences unique to the allele(s). These unique sequences should be located at the 3' end of the primer for maximum specificity in the annealing step. Remember, that to get efficient amplification, both PCR primers must anneal to the DNA.

Figure 6-8



B. The primer set can be designed to give an amplified fragment of a specific

size which can be detected by gel electrophoresis.

C. Failure to use the appropriate amplification conditions can cause amplification that will give either false positives (wrong alleles amplified) or false negatives (correct allele not amplified).

D. Primers used in the ARMS system may have an additional mismatch incorporated in their sequence. This makes the primers mismatched to all alleles; however, the allele(s) which amplify are less mismatched than the alleles which should not amplify.

E. The 3' end of the primer is most important in controlling the specificity of annealing and amplification. Differences between the primer and template near the 5' end of the primer are likely <u>not</u> to affect annealing of the primer and will give rise to false positives (i.e. amplification occurs even though mismatched).

QUESTION 7: Design a forward PCR primer that will amplify all the DRB1*04 alleles and not most of the other DR alleles. You can use 5' CCG CTG CAC TGT GAA GCT CT as a reverse primer (end of exon 2) but you need to design a forward primer. You will have to look at the DRB1 allele sequences that you looked at in Chapter 4.

Can you design a forward PCR primer that will amplify only the DRB1*11 alleles?

Can you design a primer set that will amplify only DRB1*1101 alleles?

VII. The amplified DNA is used to identify HLA types as described in the next chapters.

References:

Ausubel et al. Short Protocols in Molecular Biology. John Wiley and Sons, New York, N.Y. 2002.

Bunce, M. O'Neill, C.M., Barnardo, M.C., Krausa, P., Browning, M.J., Morris, P.J. & Welsh,

K.I. 1995. Phototyping: comprehensive DNA typing for HLA-A,B,C DRB1, DRB3, DRB4,DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens 46:355-367.

Cereb, N., Kong, Y., Lee, S., Maye, P., Yang, S.Y. 1996. Nucleotide sequences of MHC class I introns 1,2, and 3 in humans and intron 2 in nonhuman primates. Tissue Antigens 47:498-511. Correction in Tissue Antigens 48:235-236, 1996.

Krausa, P., Bodmer, J.G., and Browning, M.J. 1993. Defining the common subtypes of HLA A9, A10, A28 and A19 by use of ARMS/PCR. Tissue Antigens 42:91-99.

Mullis, K. The unusual origin of the polymerase chain reaction. Scientific American, April 1990.

Olerup, O. and Zetterquist, H. 1992. HLA-DR typing by PCR amplification with sequence specific primers (PCR-SSP) in 2 hours: An alternative to serological DR typing in clincial practice including donor-recipient matching in cadaveric transplantations. Tissue Antigens 39:225-235.

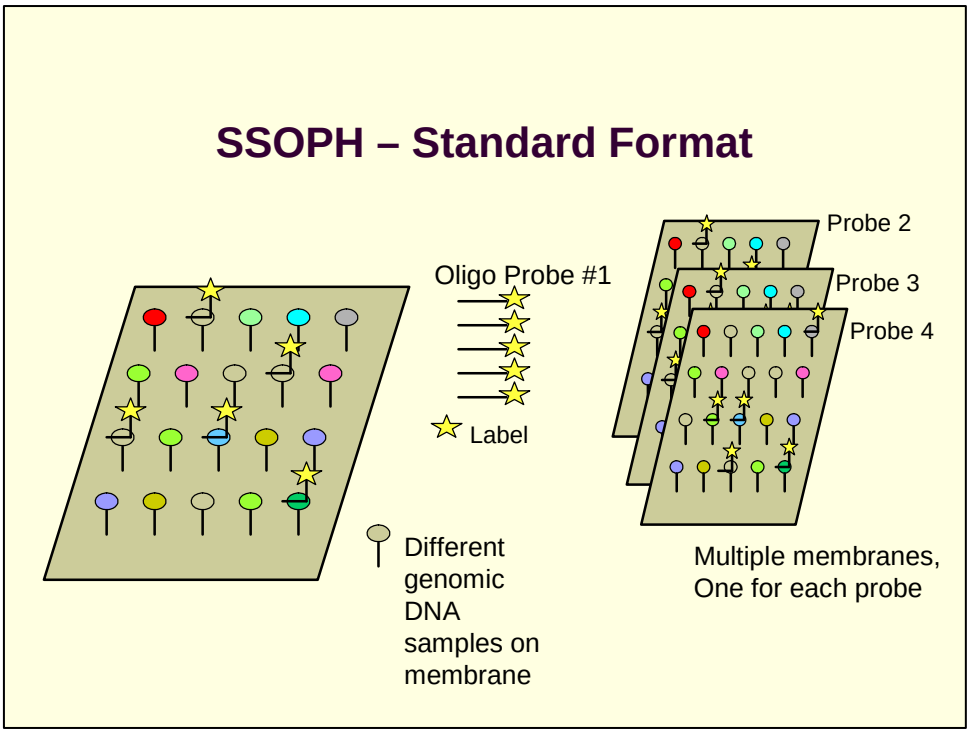
Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487-91.

CHAPTER 7

USE OF OLIGONUCLEOTIDE PROBES TO DETECT SPECIFIC DNA SEQUENCES

The amplified DNA generated using the PCR is used to identify HLA types. This chapter discusses the use of oligonucleotide probe hybridization (Figure 7-1) to detect specific HLA alleles.

Figure 7-1

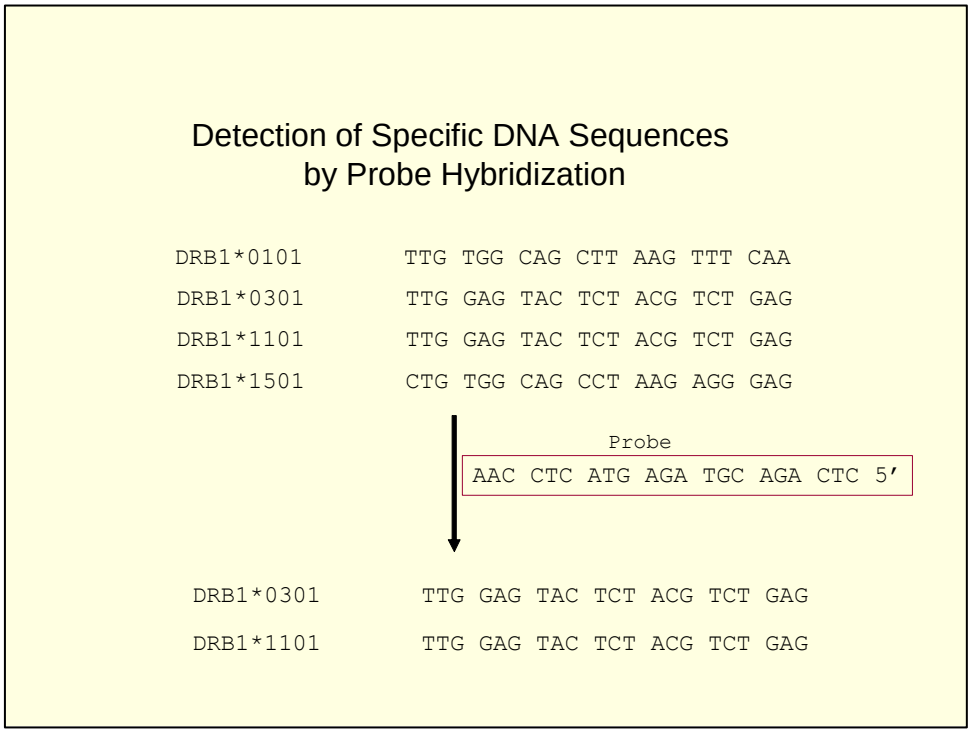


I. Selection and synthesis of probes (oligonucleotides) to detect specific alleles (SSOP=sequence specific oligonucleotide probes) (also called SSO) by hybridization (SSOPH).

A. Oligonucleotides ("oligos") (single stranded DNAs) must complement their target sequence and be sufficiently long (~18 nucleotides) to allow the use of hybridization conditions that guarantee discrimination between the target sequence and other closely related sequences [Figure 7-2]. The more differences between the matched and mismatched sequences, the easier it is to establish specific hybridization conditions. If the sequences are mismatched for only one nucleotide, the mismatched base should be placed in the middle of the oligonucleotide probe. Stringent washes are more likely to remove single base mismatches in this position; it is more destabilizing

than a mismatch at one end of the oligonucleotide.

Figure 7-2



B. The oligo should be approximately 50% G+C (if possible) and should not contain any complementary sequences that might cause the oligo to anneal to itself (form a hairpin-like structure). For example, AAAAATGCCGCTATTTTT would not be a good choice.

C. The oligo can be complementary to either strand of the DNA. Most people synthesize oligos which are identical in sequence to the coding strand (the sequence published in the literature) so that they are easier to find in the HLA allele sequences.

II. Labeling of oligo probes to detect binding. Probes are usually end-labeled, tailed, or contain modified nucleotides. Labeling can be nonradioactive or radioactive (P32). If probes are attached to beads, the bead can be fluorescently tagged.

III. Hybridization of probes to DNA samples.

A. Amplified DNA is linked to a solid support, denatured, and then hybridized to a labeled SSOP [Figures 7-1, 7-3]. Support can be a membrane, plastic plate or a bead.

B. The conditions of the hybridization and/or wash to remove nonspecifically bound probe are very important in controlling specificity of hybridization [Figure 7-4]. Most people hybridize using nonstringent conditions and utilize stringent conditions for washing.

Figure 7-3

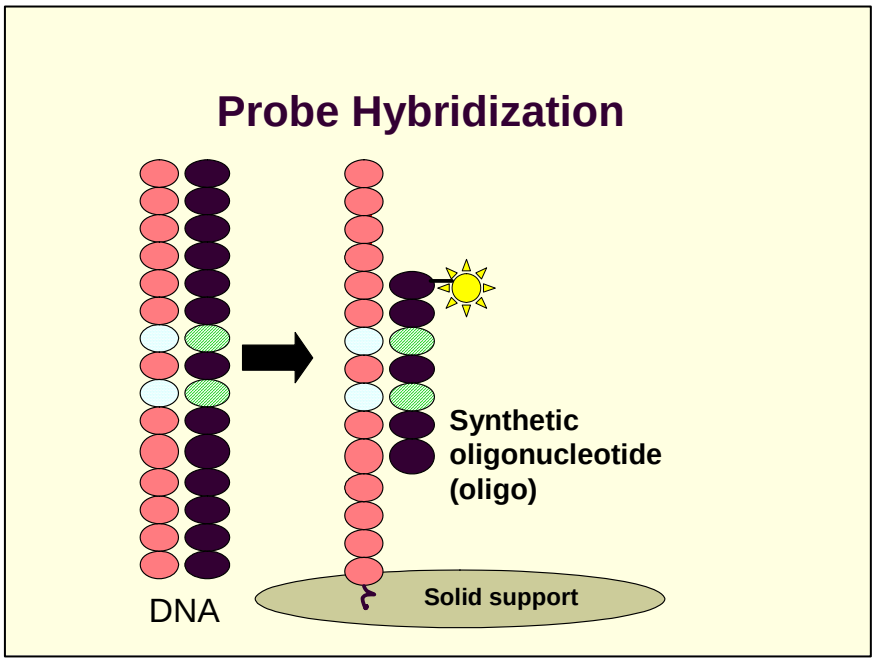
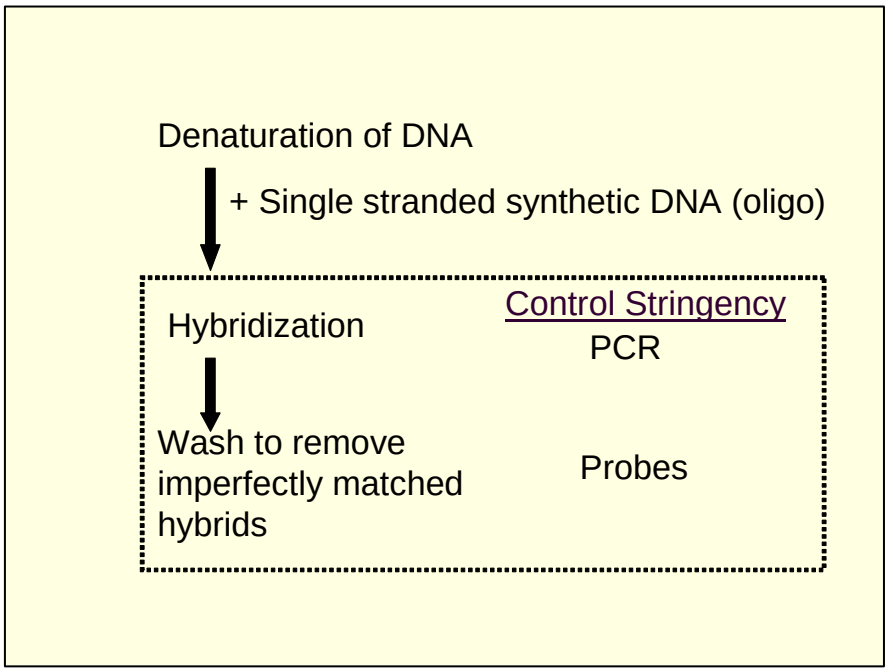


Figure 7-4

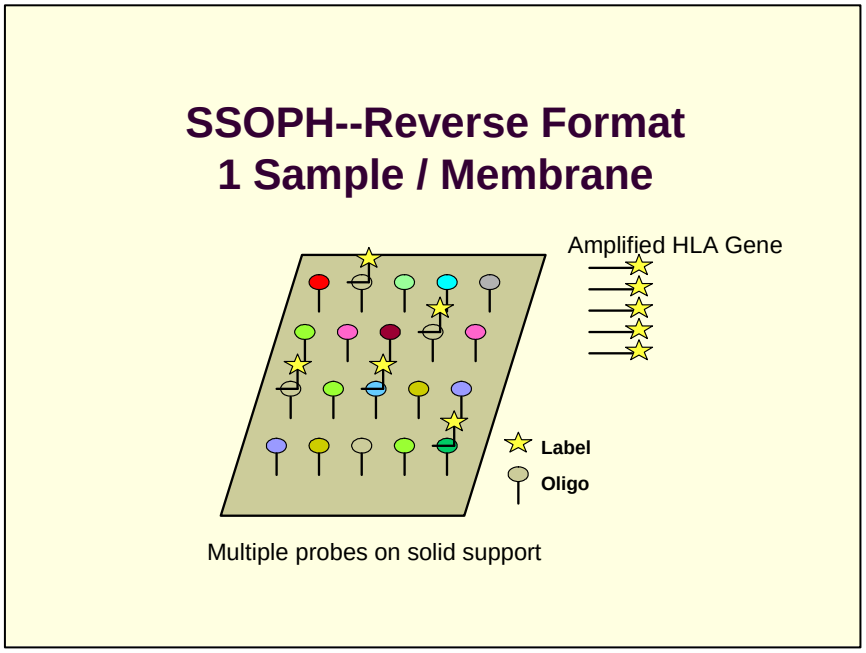


C. In one form of this technique, samples of DNA from many people are applied to a single membrane and hybridized with a single probe [Figure 7-1]. Each assay requires a separate membrane for each probe.

D. Alternatively, the probes may be linked to the solid support (like a 24 well plate) and then hybridized to labeled, denatured, amplified DNA (reverse dot blot) [Figure 7-5]. Support can be a membrane, plastic plate or a bead.

E. Another variation on the SSOP technique is the "chip" technology. In this methodology, short overlapping oligonucleotides representing the entire sequences of alleles are attached to a solid support about the size of a dime (a "chip")]. Labeled denatured amplified DNA is incubated with this chip and the binding of the DNA to individual oligos is detected by a excitation of the label with a laser].

Figure 7-5



COMMENT: Both probes and primers are single stranded pieces of synthetic DNA. When the oligonucleotide is used to prime DNA synthesis, it is called a primer. When the oligonucleotide is used for hybridization and that hybridization event is used to define an HLA type, the oligonucleotide is called a probe.

IV. Difficulty with this system:

A. Alleles often share nucleotide sequences so probes usually detect more than one allele. One may need to use multiple probes for typing a single allele. One may also need to carry out group specific amplification [Figures 7-6].

Identification of individual alleles is called high resolution or allele level typing. Typing which narrows down the allele possibilities but still includes more than one possible allele is termed intermediate resolution testing. An example is a sample typed as (DRB1*1101 or DRB1*1104) and (DRB1*0302 or DRB1*0303).

B. Contamination with other DNAs may give false positives. For example, if DNA amplified in a previous assay contaminates a pipettor, that DNA may be transferred into a DNA sample in another assay.

QUESTION 1: The two sequences below differ by a single nucleotide. Design a 10 base long oligo probe to detect sequence #1.

#1 5' ATACAGAGGTACTACGCCTAATATGGCGCTA
#2 5' ATACAGAGGTACTACACCTAATATGGCGCTA

What is the G+C content of the oligo that you have designed? What is its approximate melting temperature?

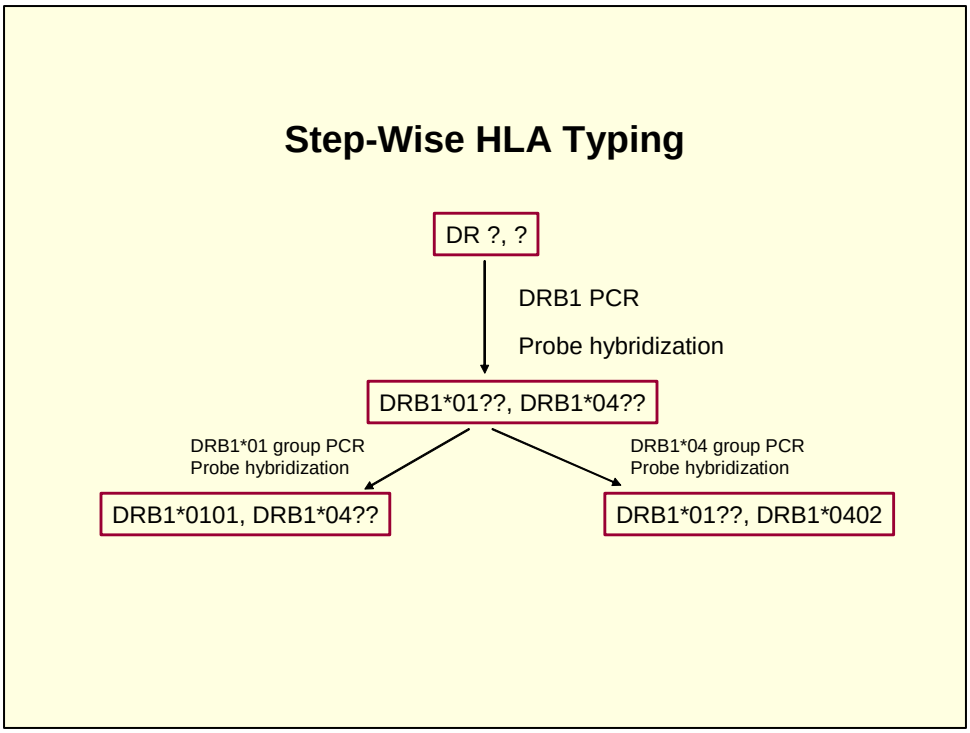
If you carry out the hybridization wash at 50^{o}C, what will happen?
What would happen if you put the discriminating nucleotide at the 3' or 5' end of the probe?

QUESTION 2: Using the nomenclature website that lists all of the DRB allele sequences, design an oligonucleotide to specifically detect DRB1*11 alleles. This oligo will be equivalent to a DR11-specific antibody used in serologic HLA typing. If we use the oligonucleotide probes to define "serologic types", this is called antigen level or low resolution typing.
[Note that the probe will also hybridize to a few non-DRB1*11 alleles like DRB1*0415 and DRB1*0308.]

Design an oligonucleotide to detect all DRB alleles.

Design an oligonucleotide to detect only the DRB1*1001 alleles. If we use the oligonucleotide probes to define alleles, this is called allele level or high resolution typing.

Figure 7-6



QUESTION 3: How would you design a protocol to identify a DRB1*0102 allele from a DRB1*0101 allele?

References (a few examples):

Bugawan, T. L., Apple, R., and Erlich, H. A. 1994. A method for typing polymorphism at the HLA-A locus using PCR amplification and immobilized oligonucleotide probes. Tissue Antigens 44:137-147.

Bugawan, T. and Erlich. 1991. Rapid typing of HLA-DQB1 DNA polymorphism using

nonradioactive oligonucleotide probes and amplified DNA. Immunogenetics 33:163.

Middleton, D., Williams, F., Hamill, M. A., et. al. 2000. Frequency of HLA-B alleles in a Caucasoid population determined by a two-stage PCR-SSOP typing stategy. Human Immunol 61:1285-1297.

Fernandez-Vina, M., Lazaro, A.M., Sun, Y., Miller, S., Forero, L., and Stastny, P. 1995. Population diversity of B-locus alleles observed by high-resolution DNA typing. Tissue Antigens 45:153-168.

Gao, X., Fernandez-Vina, M., Shumway, W., and Stastny, P. 1990. DNA typing for class II HLA antigens with allele-specific or group-specific amplification: typing for subsets of HLA-DR4. Human Immunol. 27:40-50.

Hurley, C.K., T. Tang, J. Ng, and R.J. Hartzman. 1997. HLA typing by molecular methods. In Manual of Clinical Laboratory Immunology. Rose, N.R., Conway de Macario, E., Folds, J.D., Lane, H.C., and Nakamura, R.M. eds. ASM Press, Washington, DC, pp. 1098-1111.

Fulton, R.J., R.L. McDade, P.L. Smith, L.J. Kienker, and J.R. Kettman Jr. 1997. Advance multiplexed analysis with the FlowMetrix ™ system. Clinical Chemistry 43: 1749-1756.

Nevinny-Stickel, C., Bettinotti, M.d.l.P., Andreas, A., Hinzpeter, M., Muhlegger, K., Schmitz, G., and Albert, E.D. 1991. Nonradioactive HLA class II typing using polymerase chain reaction and digoxigenin-11-2'-3'-dideoxy-uridinetriphosphate-labeled oligonucleotide probes. Human Immunology 31:7-13.

Saiki, R.K., Walsh, P.S., Levenson, C.H., and Erlich, H.A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. USA 86:6230-6234.

CHAPTER 8

SEQUENCING OF HLA GENES

It is not yet know how many HLA alleles exist in the human population. It is possible that there will be so many HLA alleles that SSOP typing will require huge panels of oligonucleotide probes and SSP will require huge primer panels to identify alleles. It may be more practical and informative to use DNA sequencing to determine if a specific donor and recipient carry the same HLA alleles.

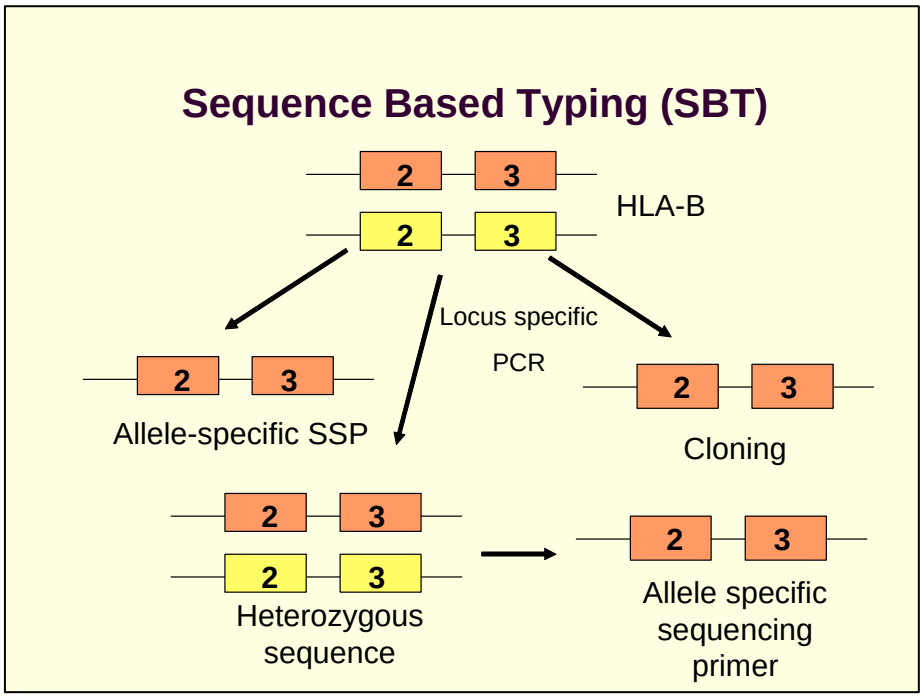
I. DNA sequencing can be used to determine the exact base sequence of the DNA or mRNA (cDNA) encoding an HLA molecule.

II. Method.

A. The DNA of the gene must be amplifed by PCR [Figure 8-1].

1. PCR is more commonly used today to obtain enough copies of a gene to sequence (Chapter 6). The primers used for PCR amplification determine how many alleles are coamplified. For example, if primers that amplify all HLA-B alleles are used, then the amplified product will contain both HLA-B alleles expressed by an individual.

Figure 8-1



2. If a sequence specific primer set is used for PCR, each of the two alleles may be isolated for sequencing

   Cloning is more work and would likely not be used for HLA typing. To make a DNA library or to "clone" a piece of DNA, pieces of DNA or DNA copies of the mRNA (cDNA) are inserted into a vector. Vectors are pieces of DNA that can replicate like a plasmid or virus. Each vector containing an inserted piece of DNA is propagated in bacteria and, as the bacteria replicate, millions of copies of the vector and inserted DNA are made.

   Cloning can be used to isolate one HLA allele for characterization from a heterozygous individual. PCR amplified DNA (e.g., HLA-B PCR product) is cloned into a vector and individual bacterial colonies isolated. Each colony carries one vector and, hence, one of the two HLA alleles.

   DNA/cDNA libraries can be made that contain most of the pieces of DNA (genomic library) carried by a cell or most of the mRNAs (cDNA library) expressed by a cell. From these libraries, the genes or mRNAs (cDNAs) that encode the HLA molecules can be isolated.

QUESTION 1: If a person was typed as DRB1*0101 and DRB1*0701, how would you obtain only the DRB1*0701 allele for sequence analysis?

B. DNA sequencing is a method of DNA synthesis (Chapter 1) requiring:

   1. A single stranded template containing the PCR-amplified copies of the HLA gene being sequenced.

   2. One primer that anneals to the template (called the sequencing primer).

   3. A DNA polymerase like Taq polymerase.

   4. Nucleotides.

      a. dATP, dTTP, dGTP, dCTP.

      b. Dideoxynucleotides (ddATP,ddTTP,ddCTP,ddGTP). When the polymerase incorporates a dideoxynucleotide into the growing

DNA strand, synthesis stops at that point.

5. Some method of labeling is used to identify the newly synthesized DNA strand.

   a. The dideoxynucleotide could be labeled with a fluorescent dye

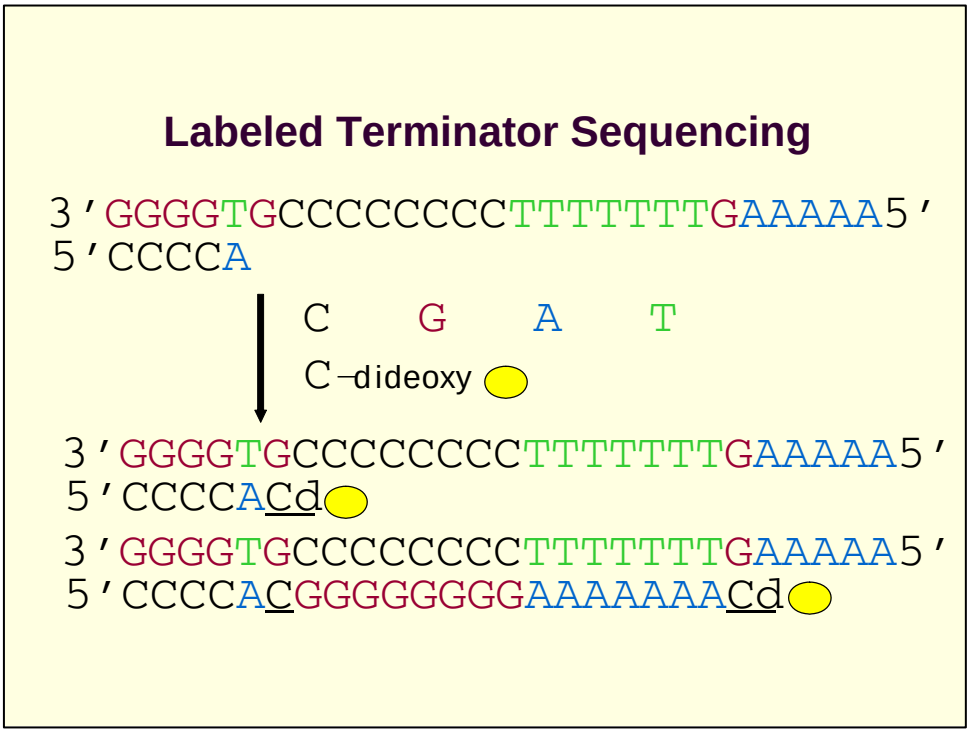
   b. Or the primer could be labeled with a fluorescent dye

C. The method used for sequencing is called the Sanger chain termination sequencing method [Figure 8-2]:

1. The amplified DNA is denatured and the primer is annealed.

2. The DNA is divided into 4 aliquots. A different chain-terminating nucleotide is added to each aliquot in addition to all four of the normal nucleotides. For example, ddCTP + dATP, dTTP, dGTP, dCTP are added to one tube [Figure 8-2].

Figure 8-2

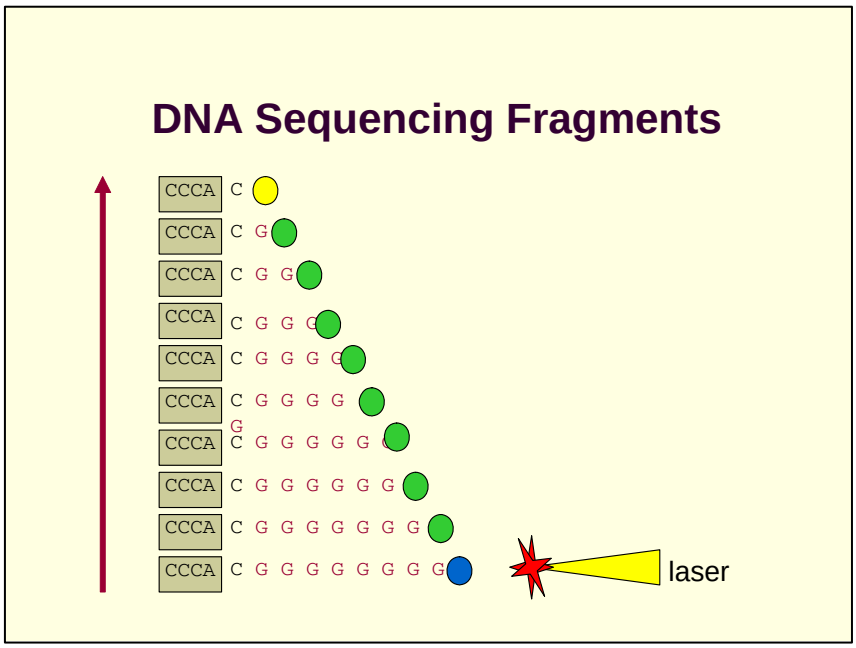


. As the polymerase is synthesizing a complementary DNA strand (the sequence shown in Figure 8-2), it has a choice of nucleotides for the incorporation. If it uses a normal nucleotide, synthesis proceeds. If the polymerase incorporates a dideoxynucleotide, synthesis halts.

Remember that many identical strands of DNA are being synthesized at the same time.

4. Each reaction generates populations of labeled oligonucleotides of different lengths that begin from a fixed point (the primer) and terminate randomly at the residue represented by the ddNTP in that aliquot [Figure 8-3].

Figure 8-3



5. The populations of oligonucleotides of different lengths are resolved by electrophoresis on a polyacrylamide like gel. If different colored fluorescent dyes are used to label each dideoxy aliquot, the four aliquots can be run in the same lane. A laser reads each color as the fragments pass by a detector (automated sequencer).

6. The “read” from a single sequencing primer is usually more than the length of an HLA exon but doesn’t usually include all of the exons in the amplicon. Usually multiple sequencing primers are used in different reactions to produce fragments covering all of the exons included in the amplicon. Sense and antisense primers are used to obtain the sequence of both strands of the DNA being sequenced.

QUESTION 2: Draw out the sequencing reaction for the following piece of single stranded DNA using the primer listed below. What would the sequencing gel look like?

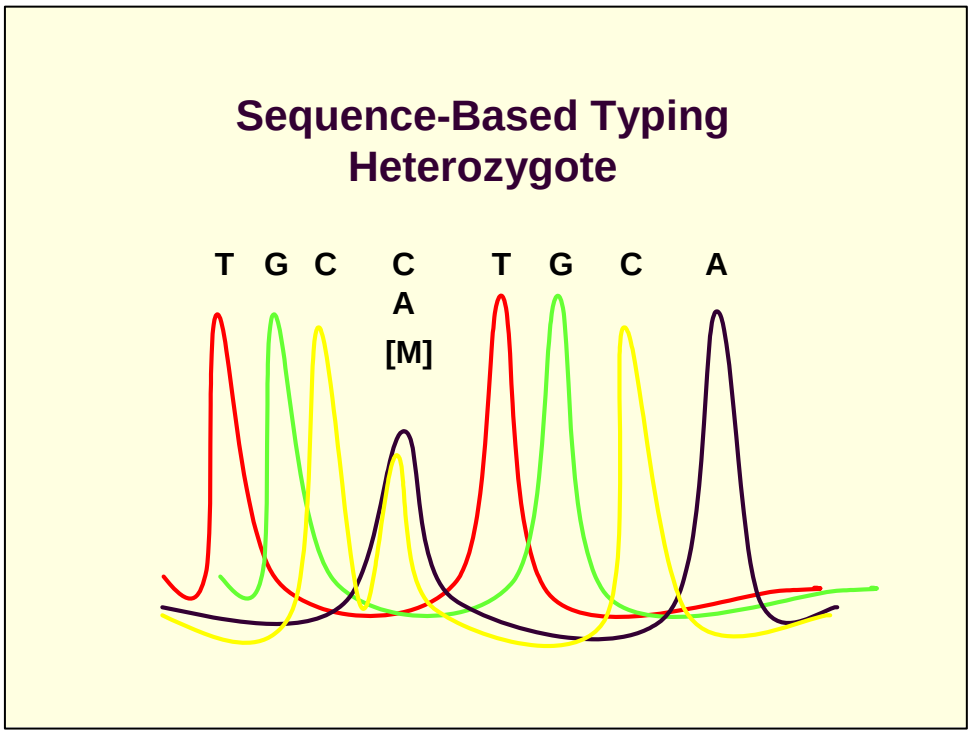
3' AAAAAATGCCGAATCCGATACGTCGGGCATT 5'

Primer: 5' TTTTTTA 3'

D. In sequence based typing (SBT), amplified HLA alleles are sequenced directly (i.e., without cloning) to identify the alleles carried by the individual.

1. Depending on the PCR primers used, the sequence may contain either a single allele or two alleles mixed together [Figure 8-4].

Figure 8-4



hen two alleles are sequenced simultaneously, positions where the two alleles differ in sequence will show two nucleotides. For example, in Figure 8-4, the fourth position shows both a C and an A. This is labeled as M meaning both C and A. Table 8-1 lists the codes used for multiple nucleotides.

b. A sequence-specific sequencing primer may be used to produce the sequence of a single allele from a PCR reaction containing both alleles at a locus.

2. A software program is used to identify the alleles based on their sequence.

Table 8-1

IUB Codes for Multiple Nucleotides

R	A/G (puRine)
Y	C/T (pYrimidine)
K	G/T (Keto)
M	A/C (aMino)
S	G/C (Strong 3H)
W	A/T (Weak 2H
N	A/C/T/G aNy base

IUB, International Union of Biochemistry Nomenclature Committee

III. This is the ideal method for identifying the HLA alleles carried by an individual but is slower, more labor intensive, and more expensive than other typing methods.

References:

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Kotsch, K., J. Wehling, S. Kohler, and R. Blasczyk. 1997. Sequencing of HLA class I genes based on the conserved diversity of the noncoding regions: sequencing-based typing of the HLA-A gene. Tissue Antigens 50:178-191.

Versluis, L.F., Rozemuller, E., Tonks, S., Marsh, S. G. E., Bouwens, A. G. M., Bodmer, J. G., and Tilanus, M. G. J. 1993. High-resolution HLA-DPB typing based upon computerized

analysis of data obtained by fluorescent sequencing of the amplified polymorphic exon 2. Human Immunology 38:277-283.

McGinnis, M.D., Conrad, M.P., Bouwens, A.G.M., Tilanus, M.G.J., and Kronick, M.N. Automated, solid-phase sequencing of DRB region genes using T7 sequencing chemistry and dye-labeled primers. Tissue Antigens 46:173-179, 1995.

Petersdorf, E.W. and Hansen, J.A. A comprehensive approach for typing alleles of the HLA-B locus by automated sequencing. Tissue Antigens 46:73-85, 1995.

CHAPTER 9

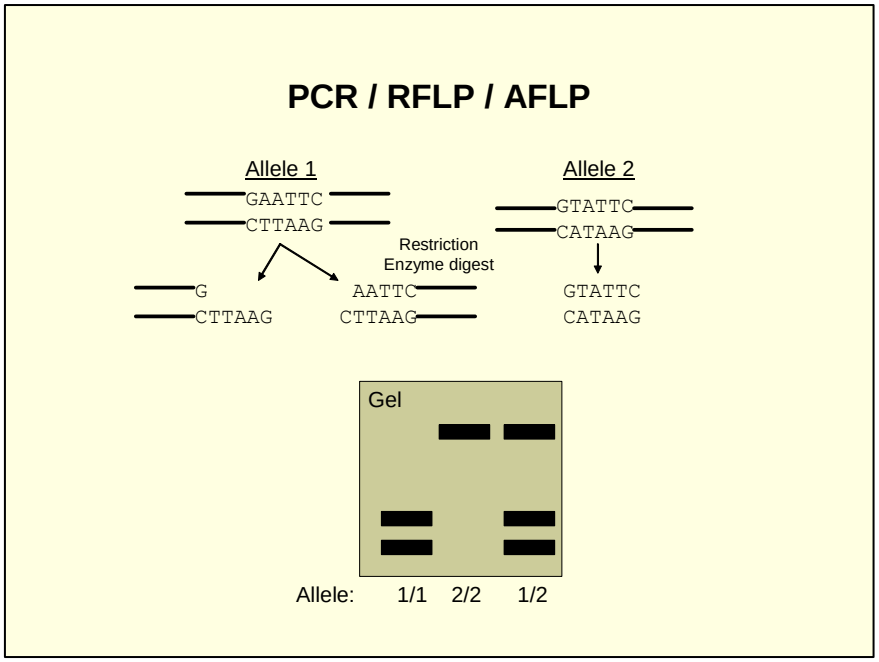
OTHER MOLECULAR BIOLOGY TECHNIQUES FOR HLA TYPING

While PCR/SSOP and SSP typing are the most common techniques used for HLA typing, other techniques have been described.

I. PCR/Restriction Fragment Length Polymorphism (RFLP)/AFLP [Figure 9-1].

A. This method can be used if two alleles differ by the presence of a restriction enzyme site. Following PCR amplification, the amplified DNA is cleaved with the restriction enzyme and the alleles are identified by the fragmentation pattern upon gel electrophoresis.

Figure 9-1



B. Two problems with the technique are a failure to get complete cleavage which may make a homozygote look like a heterozygote and difficulty in finding appropriate restriction sites in all of the alleles at a locus.

II. Denaturing gradient gel electrophoresis.

A. As amplified double stranded DNA is electrophoresed through a denaturing gradient gel, it reaches a point in the gradient where it denatures. This point is determined by the sequence of the DNA.

B. Not widely used because the results may be difficult to interpret.

III. Single stranded conformational polymorphism (SSCP).

A. PCR-amplified DNA is denatured and electrophoresed on a polyacrylamide gel. Each single strand moves at a position related to its conformation as determined by its sequence.

B. Not widely used because the results may be difficult to interpret, this method could be useful in comparing the alleles of two individuals.

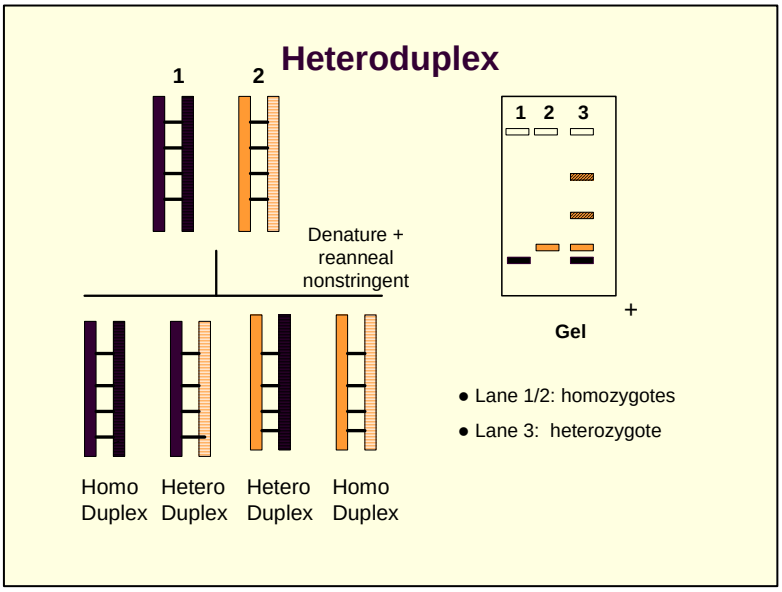
IV. Heteroduplex formation [Figure 9-2].

A. Amplified DNA is denatured and allowed to reanneal under nonstringent conditions. If DNA strands are present which do not perfectly match (e.g., in a person heterozygous for the gene amplified), these will form heteroduplexes in addition to homoduplexes. These heteroduplexes will have an altered conformation compared to the homoduplexes and will migrate differently in an electric field.

B. Sometimes a labeled reference DNA is added. This DNA is designed to anneal to the gene of interest creating additional heteroduplexes. In this case, only the labeled heteroduplexes are detected. This variation is called reference strand mediated conformational analysis (RSCA).

C. This method could be used to compare the alleles of two individuals or to determine an HLA type if compared to known alleles.

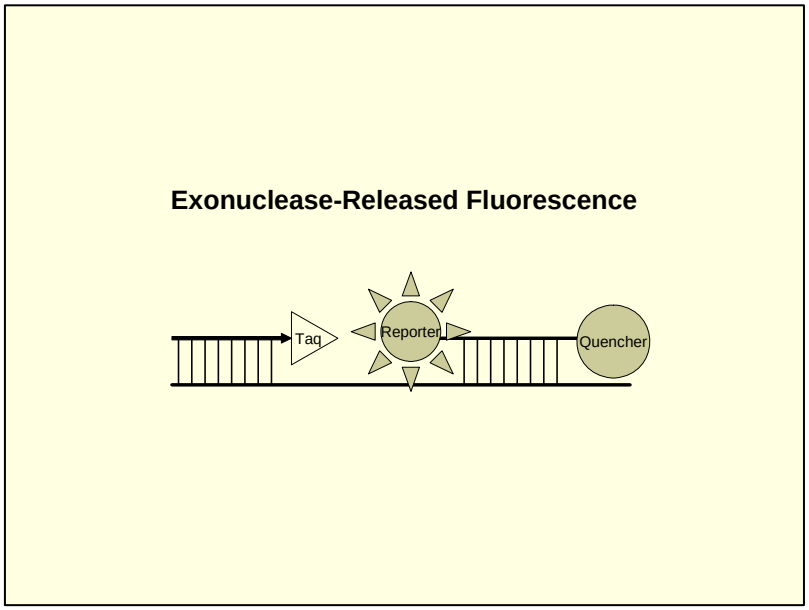
Figure 9-2



V. Exonuclease-released fluorescence [Figure 9-3].

A. An sequence specific oligonucleotide labeled with reporter and quencher dyes is hybridized to target DNA. Addition of Taq polymerase and a locus-specific primer located 5' of the probe causes the reporter dye to be released during PCR. Once the reporter dye is separated from the quencher dye, fluorescence is produced indicating that the probe had hybridized to the DNA.

Figure 9-3



VI. Pyrosequencing. A modification of the sequencing reaction, this method adds only one nucleotide at a time. Addition to the growing chain is detected. This method is used for sequencing short segments of DNA and can potentially be used to determine which polymorphisms are linked together in a sequence.

References:

RFLP:

Maeda, M., Uryu, N., Murayama, N., Ishii, H., Ota, M. Tsuji, K., and Inoko, H. 1990. A simple and rapid method for HLA-DP genotyping by digestion of PCR-amplified DNA with allele-specific restriction endonucleases. Human Immunology 27:111-121.

Olerup, O. 1990. HLA class II typing by digestion of PCR-amplified DNA with allele-specific restriction endonucleases will fail to unequivocally identify the genotypes of many homozygous and heterozygous individuals. Tissue Antigens 36:83-87.

SSCP:

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA 86:2766-2770.

Lo, Y.M.D., Patel, P., Mehal, W.Z., Fleming, K.A., Bell, J.I., and Wainscoat, J.S. 1992. Analysis of complex genetic systems by ARMS-SSCP: application to HLA genotyping. Nucleic Acids Res. 20-1005-1009.

Heteroduplex/RSCA:
Savage et. al. 1996. A rapid HLA-DRB1*04 subtyping method using PCR and DNA heteroduplex generators. Tissue Antigens 47:284-292.

Summers, C., Morling, F., Taylor, M., Yin, J. L., and Stevens, R. 1994. Donor-recipient HLA class I bone marrow transplant matching by multilocus heteroduplex analysis. Transplantation 58: 628-629.

Arguello, J.R., Little, A-M, Bohan, E. et. al. 1998. High resolution HLA class typing by reference strand mediated conformation analysis (RSCA). Tissue Antigens 52: 57-66.

Exonuclease:
Faas et. al. 1996. Sequence specific priming and exonuclease-released fluorescence detection of HLA-DQB1 alleles. Tissue Antigens 48:97-112.

Slateva, K., Albis-Camps, M., Blasczyk, et.al. 1998. Fluorotyping of HLA-A by sequence specific priming and fluorogenic probing. Tissue Antigens 52: 462-72.

Pyrosequencing
Wang, L. and F. Marincola. 2003. Applying pyrosequencing to HLA typing. ASHI Quarterly, first quarter.

Ramon, D., M. Braden, S. Adams, F. Marincola, and L. Wang. 2003. Pyrosequencing: A one-step method for high resolution HLA typing. J.Transl. Med. 1:9.

CHAPTER 10

INTERPRETATION OF DNA TYPING RESULTS

The hybridization results obtained with sequence specific oligonucleotide probes or the amplifications obtained with sequence specific primers are used to identify HLA types.

I. Most HLA alleles do not have a unique DNA sequence that characterizes that allele and that is found in no other allele (Chapters 6 and 7).

II. Detection of HLA types will often use a panel of SSOP (30-40 probes) to obtain low resolution or serologic typing level resolution results for a single locus e.g., DRB1 or HLA-A. The use of more probes will produce an intermediate level of typing resolution.

III. HLA types are obtained by comparing the positive and negative probe hybridizations (or the positive and negative sequence specific amplifications) to the known list of alleles (Figure 10-1).

Figure 10-1

DR allele	Probe 1	Probe 2	Probe 3	Probe 4	Probe 5	Probe 6
0101	■					
0301		■				■
0401			■			
0701				■		
1101		■			■	
1301		■			■	■

QUESTION 1:Based on Figure 10-1, what would be the typing if probe 1 and probe 3 were positive and the remainder negative?

What would the typing be if probe 2, probe 4, probe 5 were positive and the rest negative?

QUESTION 2: Align the sequences of the DRB1 alleles. If probe DR1001 (5'TGGCAGCTTAAGTTTGAA (codons 9-13)) is positive, one of the DRB1*01 alleles is present. If probe DR7007 (5'ACATCCTGGAAGACGAGC (codons 66-72)) is also positive,

the DRB1*0103 allele may be present. What would be your interpretation of the DR type for this sample if the probe DR1004 (5' GAGCAGGTTAAACATGAG (codons 9-14)) is also positive in addition to the probes listed above?.

How would your interpretation of DRB1*0103 change?

IV. Interpretation of typing results provides us with a genotype for an individual. Often the typing result yields more than one possible genotype and more testing will be required to determine which genotype is the correct one for that individual. Figure 10-2 is an example of a sequencing result which identified two possible genotypes: DRB1*010101, 030102 or DRB1*0104,0314.

QUESTION 3: Why can these two alternative genotypes be distinguished by sequencing of both alleles of the DRB1 locus together?

How would you determine which genotype is the correct one?

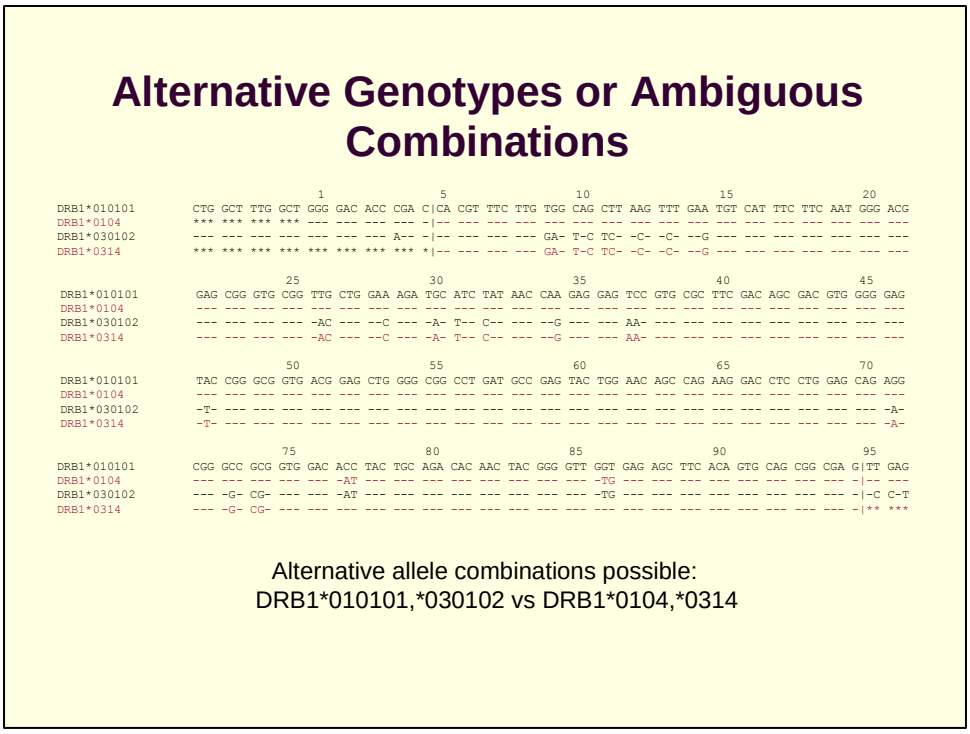
V. New HLA alleles continue to be described. If the same sample is typed with the same set of probes each year, the interpretation of the hybridization results might change over time. For example, if DNA from a sample was positive with probe DR1001AS in 1990, the result would be interpreted as DRB1*0101 or DRB1*0102 or DRB1*0103. If DNA from the same sample was typed again a few years later, a positive hybridization result with probe DR1001AS would be interpreted as DRB1*0101 or DRB1*0102 or DRB1*0103 or DRB1*0104. DRB1*0104 is a more recently described allele which carries the sequence detected by DR1001AS.

Note that the hybridization result does not change; only the interpretation of that result. This means that one must always review the interpretation of a typing result if the typing occurred some time in the past (e.g., over a year ago). This requires the knowledge of the sequences of the probes and primers used in the typing and the positive and negative hybridization results obtained with those probes and primers.

QUESTION 4: A donor on the unrelated bone marrow registry was typed as DRB1*11 using the probe 5703 (5' GCCTGATGAGGAGTACTG (codons 55-61)). What would be this person's DRB1 type based on all of the DRB1 alleles that have been identified to date? (Hint: Look at the list of current DRB1 alleles.)

A. The use of sequencing for typing is also susceptible to the same complication caused by the addition of new HLA alleles.

Figure 10-2



References:

Hurley, C.K. 1997. Acquisition and use of DNA-based HLA typing data in bone marrow registries. 1997. Tissue Antigens 49:323-328.

THE FUTURE

DNA based HLA typing methods have many advantages:

(1) The oligonucleotide reagents used are synthetic DNAs and can easily be resynthesized providing an unlimited source of typing reagents. The same DNA reagents can be utilized by laboratories all over the world to produce a consistent definition of HLA types.

(2) When new alleles are identified, new probes and primers can be designed to identify the new allele.

(3) The starting material for DNA based typing is easy to obtain from blood or buccal swabs.

(4) Approaches need to be developed to store and use DNA-based HLA data more effectively to avoid the complications caused by the addition of new HLA alleles altering the interpretation of typing results.

The same methods will be used to type for other polymorphic genes that may be important in matching donor and recipient for transplantation of hematopoietic stem cells.

Answers to the questions can be found in a separate file.