

IMMUNOLOGY

Antibody Diversity, Organization and Expression of Immunoglobulin Genes

Ashutosh Tiwari and Subrata Sinha
Department of Biochemistry
All India Institute of Medical Sciences
New Delhi - 110029

CONTENTS

[Introduction](#)

[Germ-Line Theory for Antibody Diversity](#)

[Somatic Mutation Theory](#)

[Multigene Organization of Immunoglobulin Genes](#)

[Variable Region Gene Rearrangements](#)

[Mechanism of DNA Rearrangements](#)

[Productive and Nonproductive Gene Rearrangement](#)

[Allelic Exclusion and Antigenic Specificity](#)

[Antibody Diversity and its Mechanism](#)

[Antibody Class Switching](#)

[Expression of Immunoglobulin Genes](#)

[Transcriptional Regulation of Immunoglobulin Genes](#)

[Translational Regulation](#)

Keywords

Antibody diversity, VDJ recombination, Clonal selection, Somatic Hypermutation, Class switching, Recombination Signal sequence (RSS), Allelic Exclusion

Introduction

The cells responsible for production of antibodies are the B-lymphocytes. They produce antibodies in response to foreign substances (antigens) taken into the body. Virtually any substance can elicit an antibody response. Functional groups on the surface of antigens, called epitopes or antigenic determinants, trigger this response when an antibody with the proper variable regions binds to them. The number of chemical conformations that are possible is staggering and therefore, the number of unique antibodies must also be quite large. A million different synthetic chemical compounds may serve as antigens and cause the production of antibodies. Since the genetic information for protein production resides in the genes, it might appear that a corresponding number of genes would be required to produce a million different antibodies. B-lymphocytes are capable of making a huge repertoire of antibodies - approximately 18 billion different molecules. The size of this repertoire exceeds the coding capacity of the human genome. This vast number of antibodies required cannot be generated by the antibody production mechanisms that have been discussed. There is obviously additional diversity generating mechanisms, at the genetic level, that can account for the necessary diversity. The genetic basis for this diversity lies in the organization of the immunoglobulin genes. In fact, the diversity is generated from a relatively small repertoire of parts by combining them in different ways.

Germ-Line Theory for Antibody Diversity

To explain the tremendous diversity of antibody structure, germ-line theory emerged. According to the germ-line theory, there is a separate gene for each antibody molecule and the antibody repertoire is largely inherited. The germ-line theory maintained that the genome contains a large repertoire of immunoglobulin genes, thus this theory invoked no special genetic mechanism to account for antibody diversity.

Somatic Mutation Theory

In contrast, somatic mutation theories were based on the idea that a limited number of inherited antibody genes undergo mutation in B cells during the lifetime of an individual to generate the observed repertoire. Recent research and the cloning of immunoglobulin genes has shown that the antibody repertoire is generated from a large but limited number of antibody genes that undergo DNA rearrangement, thus, both theories explain part of antibody diversity.

As the amino acid sequences of more and more antibodies were determined, it became clear that the sequence stability maintained in the constant region while some kind of diversifying mechanism generated the variable region. Neither the germ-line nor somatic-mutation theory gives a reasonable explanation for this central feature of immunoglobulin structure. Germ-line proponents found it difficult to explain diversity in the variable part of many heavy and light chain genes while preserving the constant region of each changed. Somatic mutation proponents found it difficult to explain the mechanism that could generate diversity in variable region of single heavy or light chain gene in the somatic cells without alteration in the constant region.

Dreyer and Bennett's Two-Gene Model

Dreyer and J. Bennett (1965), proposed a recombinational model which combined elements of the germ-line and somatic-mutation theories. They suggested that two separate genes encode a

single immunoglobulin heavy or light chain, one gene for the variable region and the other for the constant region. They speculated that these two genes must come together at the DNA level by some means to form a continuous message that can be transcribed and translated into single heavy or light chain. They also proposed that hundreds or thousands of variable region genes were carried in the germ line, whereas only single copies of constant region class and subclass genes exist.

In 1976, S. Tonegawa and N. Hozumi found the first direct evidence that separate genes encode the variable and constant regions of immunoglobulins and that the genes are rearranged in the course of B cell differentiation. Tonegawa was awarded the Nobel Prize for this work in 1987. They analyzed the DNA from embryonic cells and myeloma cells which represent the different stages of the cell development. The embryonic cell is an earliest stage of cell development and myeloma cells represent the mature developed antibody producing cells. DNA from embryonic cells and myeloma cells was treated with the same restriction endonuclease and the fragments were separated by size. They were analyzed for their ability to hybridize with the radio labelled mRNA probe. Two separate restriction fragments from the embryonic DNA hybridized with the mRNA, whereas only a single restriction fragment of the adult myeloma DNA hybridized with the same probe. Tonegawa and Hozumi suggested that during differentiation of lymphocytes from embryonic state to the fully differentiated plasma cell state, the variable (V) and constant (C) genes undergo rearrangement. In the embryonic state the V and C genes are separated by a large DNA segment that contains a restriction endonuclease site. In adult state during differentiation, the V and C genes are brought closer together by DNA rearrangement and the intervening DNA sequence is eliminated. On the treatment with restriction endonuclease, the two genes remain close to each other due to lack of restriction site between them. (Fig.1)

Multigene Organization of Immunoglobulin Genes

Detailed analysis of rearranged and germ-line genes for antibody molecules show that the κ and λ light chains and the heavy chains are encoded by separate multigene families situated on different chromosomes (Table1). In germ-line DNA, each of these multigene families contains several coding sequences, called gene segments, separated by noncoding regions. During B cell maturation, the gene segments are rearranged by splicing of noncoding segments and the coding gene segments brought together to form functional immunoglobulin genes. This is at present the most accepted view of immunoglobulin genes and their organization.

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

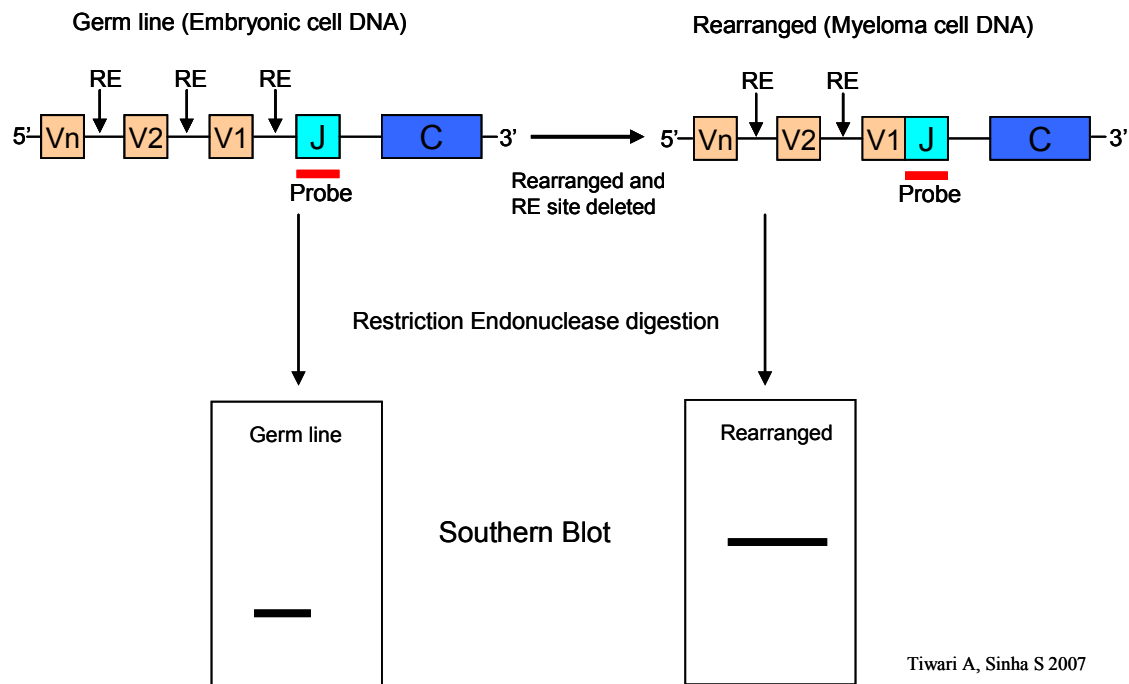


Fig.1: Diagrammatic representation of experiment showing that immunoglobulin DNA undergoes rearrangement during cell development. Note that only one of the several J gene segments present is shown. In this example, a probe that includes a J gene segment is used to identify RE digestion fragments that include all or portions of this segment.

Table1: Chromosomal Location of Gene Families (Human)

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

Lambda Chain Multigene Family

The gene organization of lambda chain locus is bit complex. When the nucleotide sequence was compared with the known amino acid sequence of the lambda chain variable region, an unusual discrepancy was observed. Only the first 97 amino acids of the lambda chain variable region corresponded to the nucleotide sequence, the remaining 13 C-terminal amino acids of the variable region were not matching with the corresponding nucleotide sequence. It has been found that many base pair away a separate; 39-base pair gene segment encoded the remaining 13 amino acid of the lambda chain variable region. This separate 39-bp gene segment is known as J

(joining) segment. Thus, a functional lambda variable region gene is made up of two coding segments; 5' V segment and 3' J segment which are separated by non coding DNA sequence. The lambda multigene locus is more complex in humans. There are 31 functional V gene segments, 4 J segments and 7 C segments. In humans lambda locus contains many V, J and C pseudogenes, a defective gene that is incapable of encoding protein; such genes are indicated with the psi symbol (ψ). In mouse germ-line lambda locus contains three V gene segments, four J gene segments and four C gene segments. (Fig.2a). The $J_{\lambda 4}$ gene segment is a pseudogene. The V_{λ} gene segment and the three functional J_{λ} gene segments encode the variable region of light chain, and each C_{λ} gene segment encodes the constant region of one of the three lambda chain subtypes ($\lambda 1$, $\lambda 2$ and $\lambda 3$).

Kappa Chain Multigene Family

In mouse kappa chain multigene family contains approximately 85 V gene segments each with a 5' upstream leader sequence. There are five J gene segments and a single C gene segments (Fig.2b). Since there is only one C gene segment, there are no subtypes of kappa light chain. In humans the kappa chain multigene family has similar to that of the mouse, contains forty V gene segments, five J gene segments and single C gene segment.

Heavy Chain Multigene Family

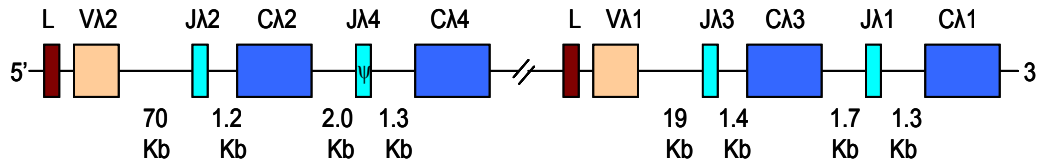
The gene organization of heavy chain is similar but more complex than that of the kappa and lambda light chain genes (Fig.2c). Unlike light chain, each heavy chain is encoded by four gene segments. An additional gene segment encodes part of the heavy chain variable region. Leroy Hood and his colleagues proposed that a third germ-line gene segment must join the V_H and J_H gene segments to encode the entire variable region of the heavy chain. This gene segment encodes the amino acids with in the third complementarity determining region (CDR3), was nominated as D for diversity, because of its contribution in the antibody diversity. In humans 51 V_H gene segments have been identified which are located upstream from a cluster of 27 functional D_H gene segments. Each V_H gene segment is preceded by a leader sequence. There are six functional J_H gene segment followed by a series of C_H gene segment present in the heavy chain locus. Each C_H gene segment encodes the constant region of immunoglobulin heavy chain isotype. The C_H gene segments consist of coding exons and noncoding introns. Each exon encodes a separate domain of heavy chain constant region. A similar heavy chain gene organization is present in mouse. In humans and mice, the C_H gene segments are arranged in a specific sequence in the order C_{μ} , C_{δ} , C_{γ} , C_{ϵ} , C_{α} . This sequential arrangement is related to the sequential expression of the immunoglobulin classes in the course of B cell development. Thus, when the B cell encounter first with an antigen the initial IgM response of B cell occurs.

Variable Region Gene Rearrangements

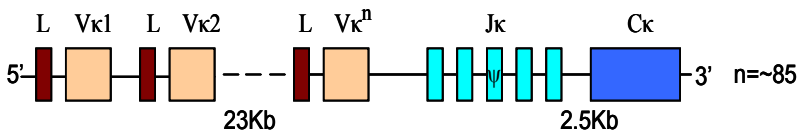
Mature B lymphocytes are the only cells which express the properly rearranged immunoglobulin genes of all classes and subclasses. During B cell maturation, variable-region gene rearrangement occurs in the bone marrow. This rearrangement occurs in a definite sequence and is related to B cell maturation process. The heavy chain variable region genes rearrange first, then the light chain variable region genes. At the end of this rearrangement process, each B cell has a single variable region DNA sequence for it's light chain and another for it's heavy chain. Finally this variable region gene rearrangement leads to mature B cells which are committed to

produce antibody with binding site encoded by the particular rearranged variable region gene sequence.

(a) λ -chain DNA



(b) κ -chain DNA



Tiwari A, Sinha S 2007

(c) Heavy-chain DNA

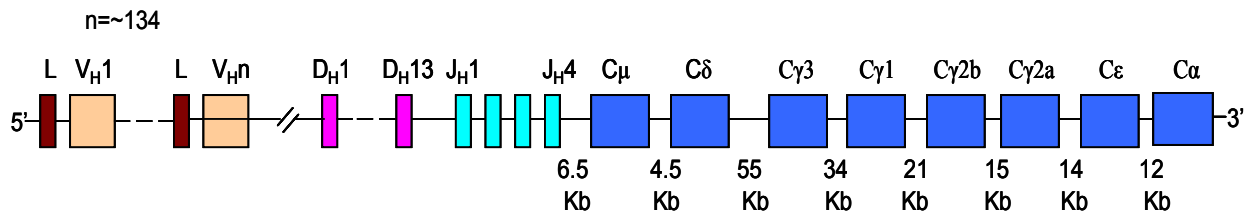
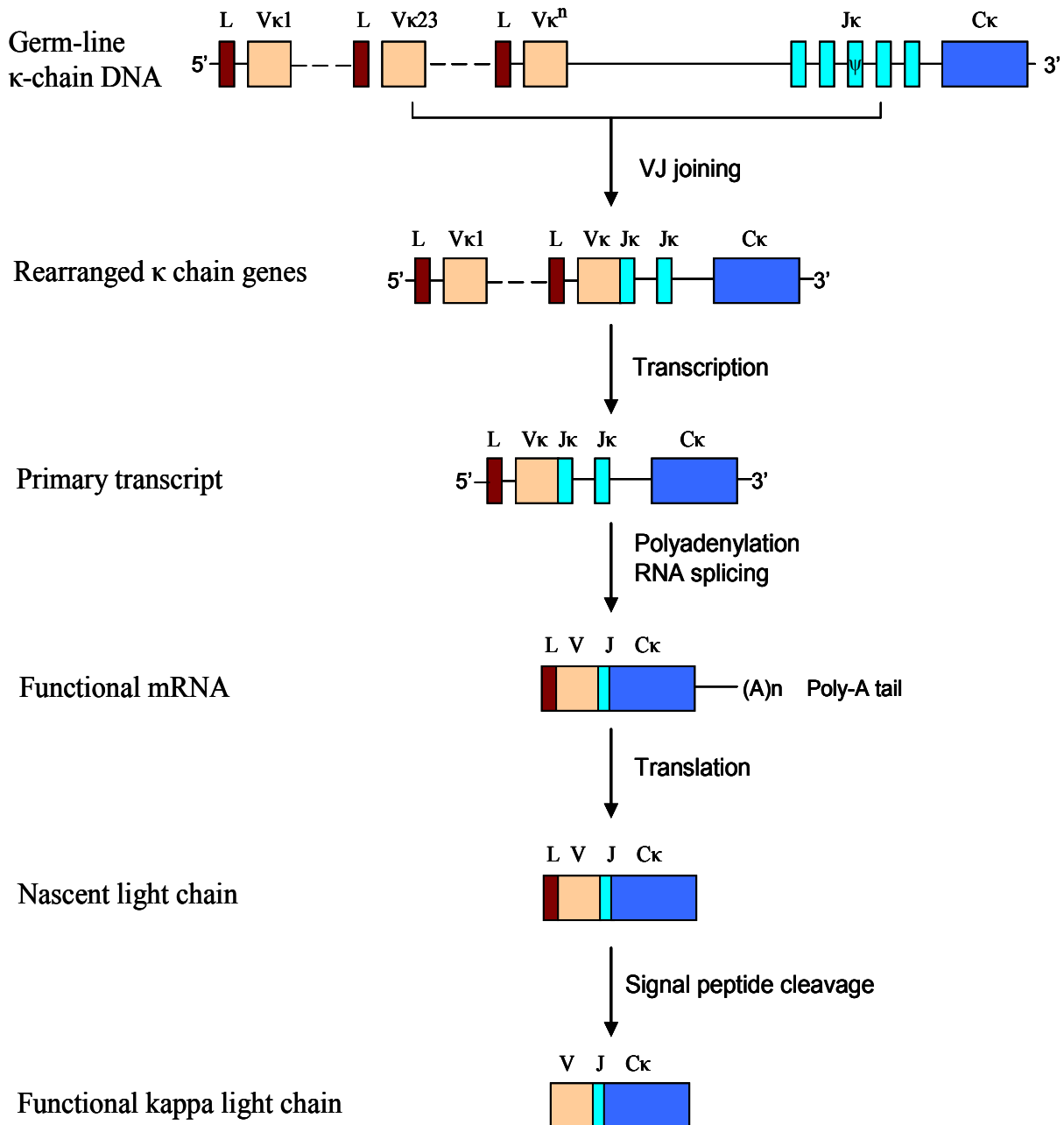


Fig.2: Immunoglobulin germ-line gene arrangement in the mouse: (a) lambda light chain, (b) kappa light chain and (c) heavy chain. The distance between the various gene segments is represented in kilobases (Kb)

Light Chain Gene Rearrangement

Light chain gene rearrangement start with the combination of one V gene segment with J gene segment and form VJ rearranged gene. During this process the intervening DNA sequence is removed. In humans or mouse kappa light chain, any one of the V_{κ} gene segments can be joined with any one of the functional J_{κ} gene segment. In humans, any of the functional V_{λ} genes can combine with any of the four functional J_{λ} - C_{λ} combinations. In mouse this rearrangement is more complicated. DNA rearrangement can join the $V_{\lambda 1}$ gene segment with either the $J_{\lambda 1}$ or the $J_{\lambda 3}$ gene segments. The $V_{\lambda 2}$ gene segment can be joined with the $J_{\lambda 2}$ gene segment.

Kappa and lambda light chain rearranged genes contain a short leader (L) exon, a noncoding sequence (introns), a joined VJ gene segment, a second introns and the constant region (Fig.3). The rearranged light chain gene sequence is transcribed by RNA polymerase from the leader exon through the C gene segment to the stop signal, generating a light chain primary RNA transcript. The processing of primary transcript is done by splicing and then the resulting light chain mRNA undergo translation.



Tiwari A, Sinha S. 2007

Fig.3: Kappa light chain gene rearrangement and processing of RNA to produce functional kappa light chain protein.

Heavy Chain Gene Rearrangement

Functional heavy chain gene requires two separate rearrangement events within the variable region. Variable region gene rearrangement starts with the combination of one of the D_H to one of the J_H gene segments. Then one of the many V_H gene segments joins the resulting D_HJ_H segment and generate a $V_HD_HJ_H$ unit that encodes the entire variable region. This rearranged variable region gene contains the following sequences from the 5' end to 3' end: a short leader exon, an intron, a joined VDJ segment, another intron and a series of C gene segments (Fig. 4).

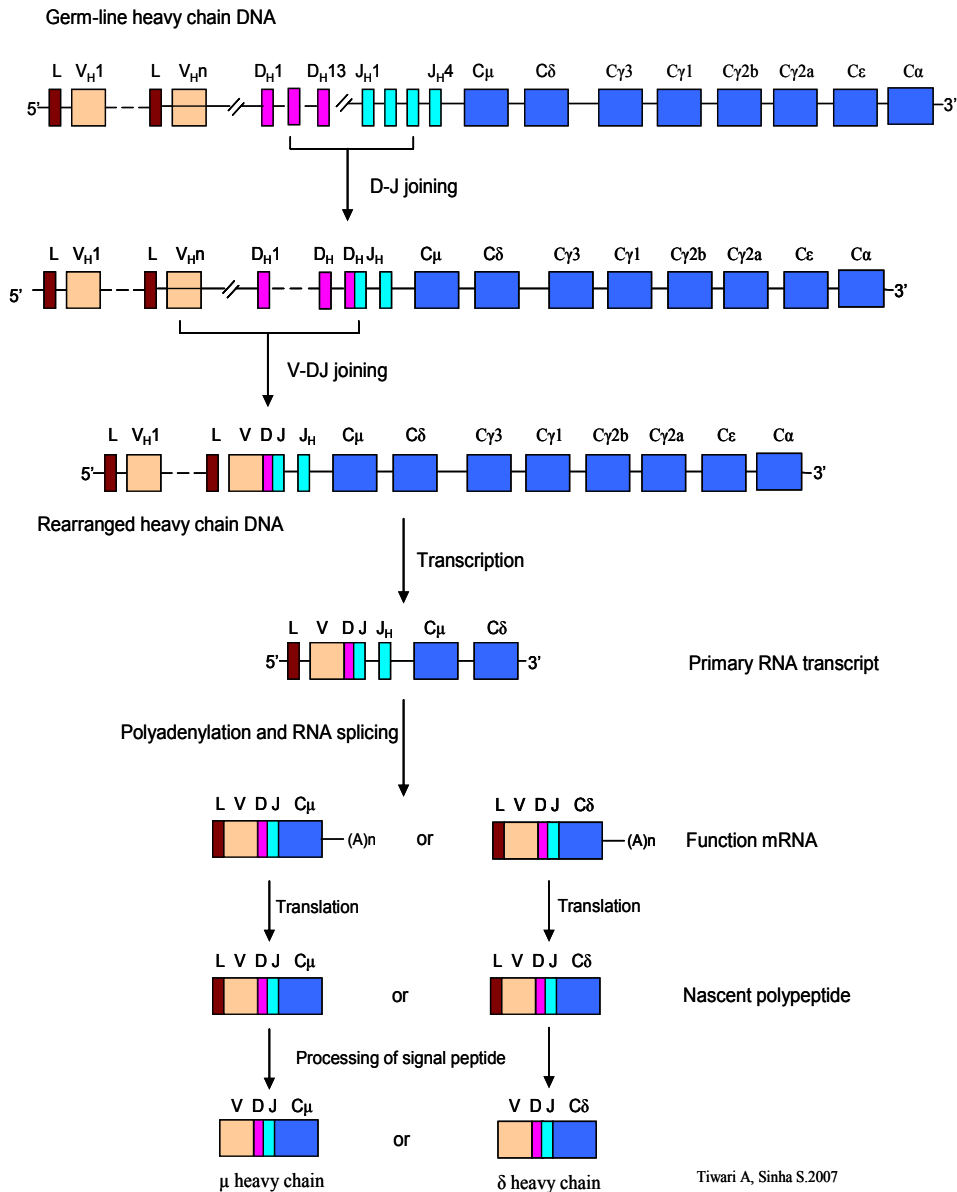


Fig. 4: Heavy chain gene rearrangement and processing of RNA transcript to generate function μ or δ heavy chain protein. Two DNA joinings occur to generate functional heavy chain gene. Expression of functional heavy chain gene involves differential RNA processing which generate several end products. In this example each C gene segment is drawn as single coding sequence; in actual each C gene is made of series of exons and introns

After the proper rearrangement, the RNA polymerase can bind to the promoter sequence and transcribe the entire heavy chain gene including the introns. Initially, both $C\mu$ and $C\delta$ gene segments are transcribed. After this RNS splicing remove the introns and process the primary transcript to generate mRNA including either $C\mu$ or the $C\delta$ transcript. These mRNA are then translated and leader peptide is cleaved to produce a fully functional μ and δ chains. So the mature B-cell is able to express IgM and IgD with identical antigenic specificity on its surface. It is only after antigenic stimulation that B-cell undergoes class switching to produce other classes and subclasses of antibodies.

Mechanism of DNA Rearrangements

Rearrangement of variable region DNA segments leads to the production of a functional gene in both heavy and light chain. This rearrangement leads to antigenic specificity of the antibody. The joining of V gene segment to C gene segment is a posttranscriptional process and takes place by RNA splicing.

DNA Rearrangement by Recombination Signal Sequence (RSS)

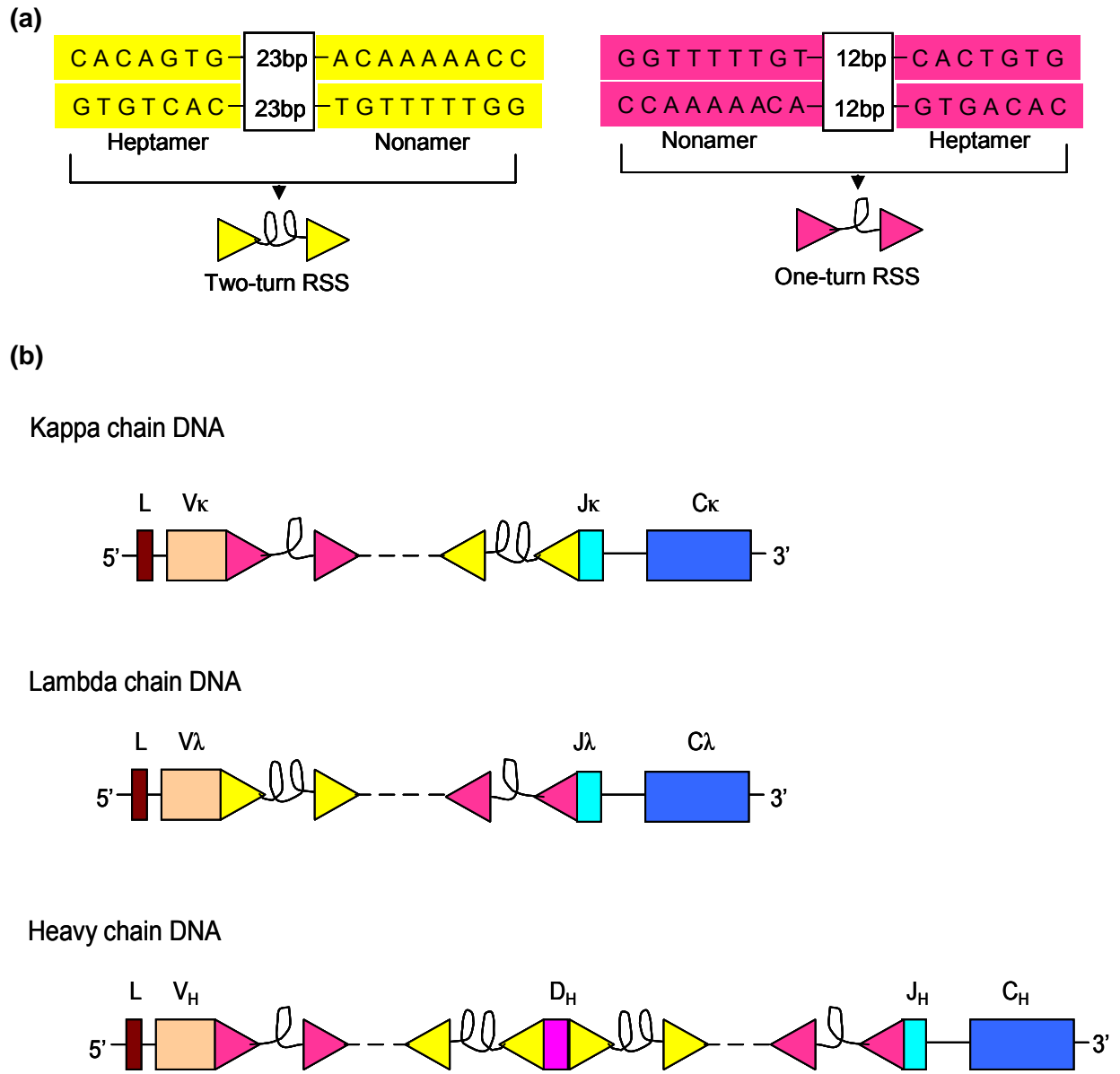
It has been found by the DNA sequencing studies that unique recombination signal sequences (RSSs) present in the flanking region of each germ-line V, D, and J gene segment. One RSS is located 3' end to each V gene segment and 5' end to each J gene segment. One RSS is present on the both sides of each D gene segment. The sequence of RSS provides signals for the recombination process that rearranges the genes. RSS are a combination of seven base pair long palindromic sequences (heptamer) and a nine base pair long AT rich conserved sequence (nanomer) with an intervening spacer DNA of either 12 base pairs or 23 base pairs length (Fig.5a). Since the 12 base pair length and 23 base pair length correspond to approximately one turn and two turns of the DNA helix; for this reason the sequences are called one-turn recombination signal sequence and two-urn recombination signal sequences. These recombination sequences are recognized by specific recombinases.

In lambda chain, $V\lambda$ signal sequence has two-turn spacer and the $J\lambda$ signal sequence has one-turn spacer. This order is reversed in kappa chain; the $V\kappa$ signal sequence has a one-turn spacer and $J\kappa$ signal sequence has two-turn spacer. In the heavy chain, the V_H and J_H signal sequences have two turn spacer, the signal on the either side of the D_H gene have one-turn spacer (Fig.5b). Signal sequences having one turn spacer can join only with sequences having a two turn spacer (one-turn/two-turn joining rule). This joining rule ensures the correct rearrangement of various gene segments and the segments of the same type do not join each other.

Joining of Various Gene Segment by Recombinases

The V-D-J or V-J joining is a precise mechanism which takes place at the junctions between RSS and coding sequences. This joining is catalyzed by specific enzymes collectively called V(D)J recombinases. These recombinases are encoded by two genes (recombination activating genes) designated as RAG-1 and RAG-2. These enzymes act synergistically and required to mediate V-D-J joining. The RAG-1 and RAG-2 proteins and the enzyme terminal deoxynucleotidyl

transferase (TdT) are the only lymphoid-specific gene products that have been involved in V(D)J rearrangement.



Tiwari A, Sinha S. 2007

Fig. 5: (a) Nucleotide sequence and diagrammatic representation of recombination signal sequences (RSSs). One-turn RSS and two-turn RSS are represented by red and yellow color respectively. (b) Location of RSSs in germ line immunoglobulin DNA. Both the RSS have characteristic locations within λ chain, κ chain and Heavy chain germ-line DNA

The V(D)J joining mediated by RAG-1 and RAG-2 recombinases occurs in two ways:

1. Deletional Joining

In this VDJ joining, intervening DNA between the two coding gene segments is deleted and removed. This joining occurs when the gene segments to be joined have the same transcriptional orientation. This process yields two products: a rearranged VJ unit that includes the coding joint, and a circular excision product consisting of the recombination signal sequences (RSSs), signal joint and intervening DNA (Fig.6).

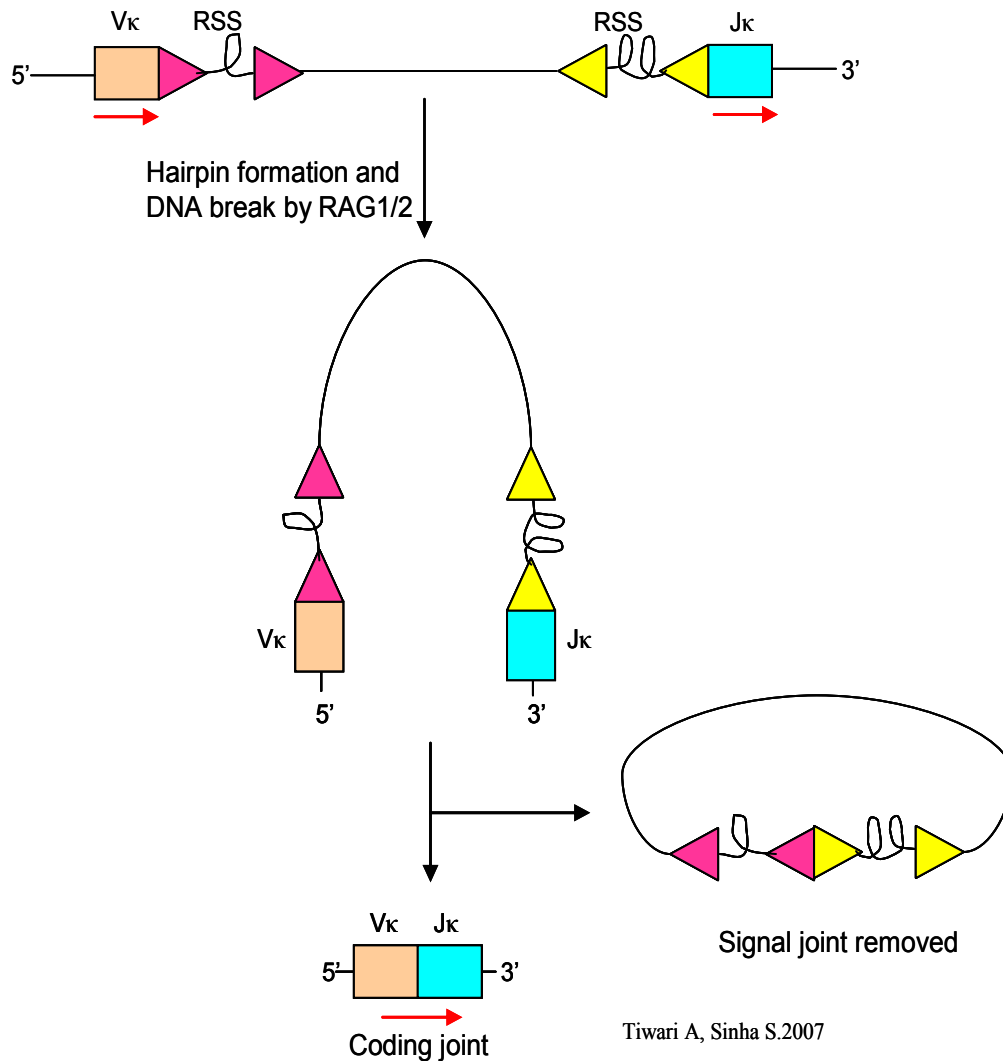


Fig. 6: Schematic representation of deletional joining of V and J gene segments of kappa light chain. Deletional joining occurs when the gene segments have the same transcriptional orientation

2. Inversional Joining

Inversional joining occurs when the gene segments have opposite transcriptional orientations. The frequency of two gene segments having opposite transcriptional orientation is less. In this case, the recombination signal sequences (RSSs), signal joint and intervening DNA are retained and the orientation of one of the joint segments is inverted (Fig.7). In the human kappa light chain locus, about half of the V_{κ} gene segments are inverted with respect to J_{κ} and their joining is mediated by inversion.

In both types of recombination, a few nucleotides may be added to or deleted from the cut ends of the coding sequences before they are rejoined. This is one of the reasons of the diversity of the immunoglobulin molecules.

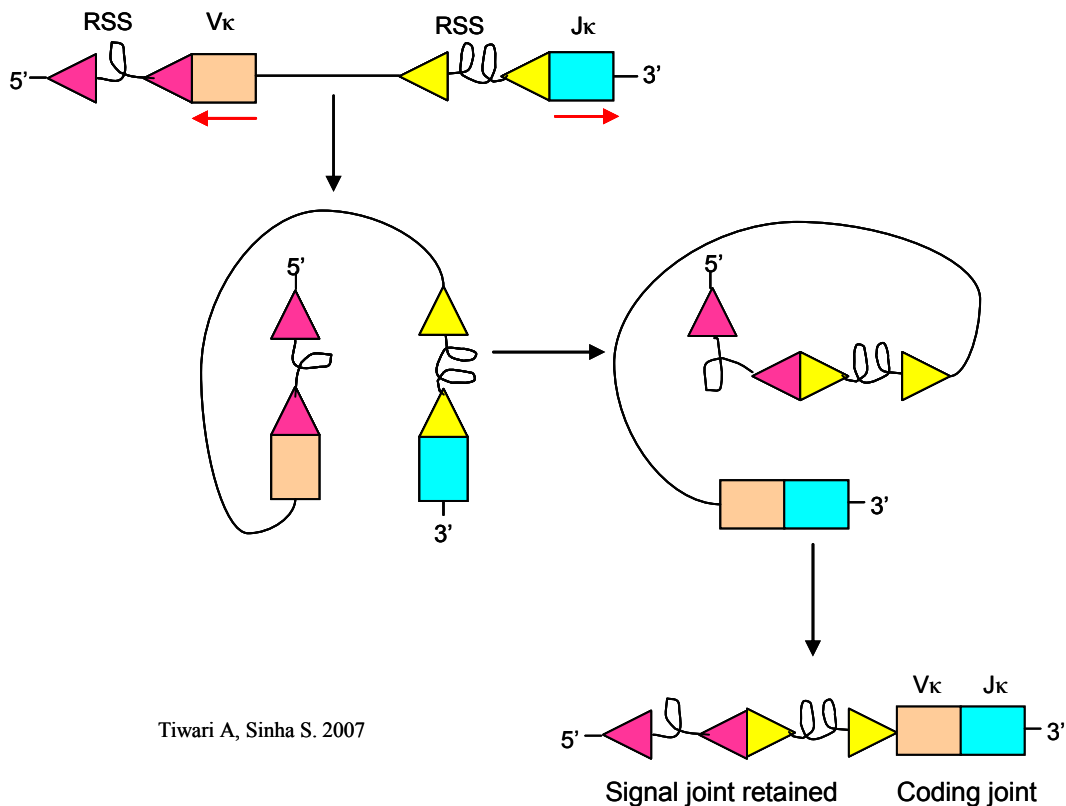


Fig. 7: Schematic representation of inversional joining of V and J gene segments of the kappa light chain. Inversional joining occurs when the gene segments have opposite transcriptional orientation

Productive and Nonproductive Gene Rearrangement

Gene rearrangement is an important event in the ultimate production of mature and fully functional immunoglobulin molecules. The VDJ joining involves the cutting of double stranded DNA at the junction of RSS and the coding gene segment followed by joining of the two different gene segments. The double stranded DNA breaks that initiate VDJ rearrangement are

introduced precisely at the junction of RSS and coding sequences, but the subsequent joining of coding sequences is imprecise. The consequences of imprecise joining may lead to change in sequence leading to frame shift mutations. The resulting VJ or VDJ gene segment is likely to contain numerous stop codons which interrupt translation. When the rearranged gene segments are joined in phase, the reading frame is maintained. In this productive rearrangement, VJ or VDJ gene segment can be translated fully and yielding a complete immunoglobulin. It is estimated that only one in three attempts at V_L - J_L joining, and one in three subsequent attempts at V_H - D_H - J_H joining are productive. Hence, less than 1/9 (11%) of pre B-cells in bone marrow progress to maturity and leave the bone marrow as mature immunocompetent B cells.

Allelic Exclusion and Antigenic Specificity

All B cells contain both parental and maternal chromosomes like all somatic cells. Even though a B cell is diploid, but in one cell rearranged heavy and light chain genes are expressed from one of the two alleles. The productive rearrangement of one allele on one chromosome leads to the inhibition of rearrangement and expression of the other allele. This process is called allelic exclusion. It is essential to ensure the antigenic specificity of B cell. Expression of both the alleles would render the cell multispecific.

A model was proposed by G.D.Yancopoulos and F.W. Alt to explain the mechanism of allelic exclusion (Fig.8). They suggest that once a productive gene arrangement is achieved, the functional protein is expressed and the presence of functional protein acts as signal to prevent the further gene rearrangement. Since first of all heavy chain gene rearrangement takes place, so according to this model the presence of μ heavy chains give signals to the B cell to turn off further gene rearrangement of the other heavy chain allele and to turn on rearrangement of the κ light chain genes. Once the productive gene rearrangement takes place in κ light chain, functional κ light chains are produced and then pair with μ heavy chains to form a complete immunoglobulin molecule. The presence of complete antibody then turns off further light chain gene rearrangement. If κ light chain gene rearrangement is nonproductive for both κ alleles, the rearrangement in the λ light chain genes begins. If both the λ alleles gene rearrangements are nonproductive, the B cell apparently stop the maturation process and soon dies by apoptosis. This hypothesis has been proved by various studies in the transgenic mice. These studies suggest that expression of heavy and light chain may in fact prevent gene rearrangement of the remaining alleles and thus account for allelic exclusion.

Antibody Diversity and its Mechanism

The vertebrate immune system can produce a vast variety of immunoglobulin molecules specific for correspondingly vast number of antigens. With the understanding of genetic rearrangements of various immunoglobulin gene segments, many possible mechanisms have been put forward to explain this antibody diversity. There are several means of antibody diversification known in mice and humans.

1. Multiple germ-line gene segments
2. Combinatorial V(D)J joining
3. Junctional flexibility
4. P region and N region nucleotide addition
5. Somatic hypermutation

6. Combinatorial association of light and heavy chains

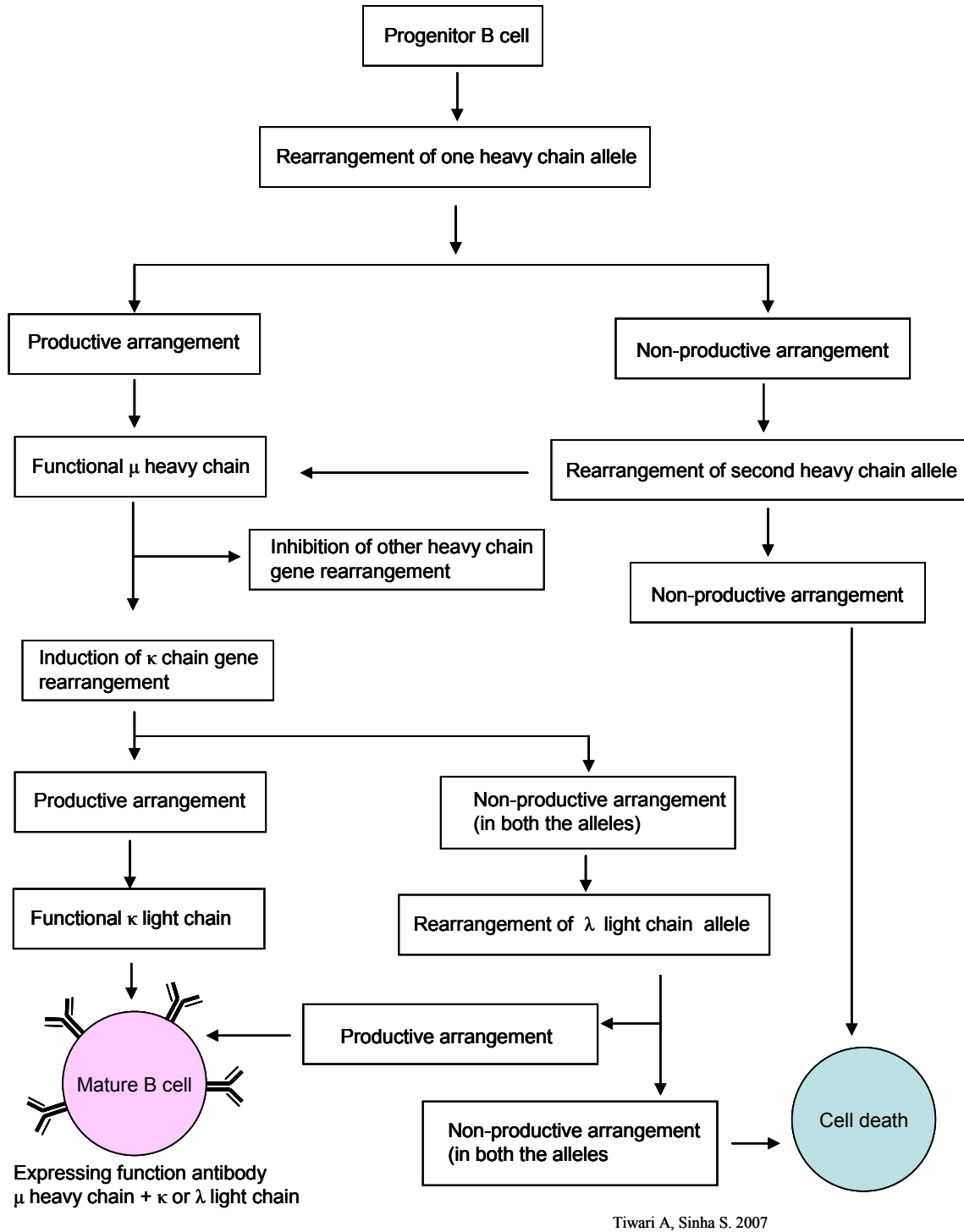


Fig.8: Schematic representation of allelic exclusion model proposed by Yancopoulos and Alt

The above mentioned avenues of diversification contribute considerably to the immense number of distinct antibodies generated in the mammalian immune system. Although the exact contribution of each of these events of diversification to total antibody diversity is not identified, but each of them contributes significantly to antibody diversity.

1. Role of Multiple Germ-line Gene in Antibody Diversity

The large number of variable germ-line genes makes it possible to produce heavy and light chains of different antigenic specificities that if only one or few variable region genes were present. The numbers of genes in various germ-line gene segments as listed below are based on the study of immunoglobulin loci of the human genome. The figures are likely to vary in large population-based results. The estimated number of each gene segment of each family is shown in Table 2.

Table: 2 No. of Coding Segments for Each Gene Family

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

Although the number of germ-line genes present in the mammalian genome is very few than predicted by early followers of the germ-line model, but multiple germ-line V, D and J gene segments clearly provide a great contribution to the diversity of the antigen-binding sites in antibodies.

2. Combinatorial V(D)J Joining

The contribution of multiple germ-line gene segments to antibody diversity is further enlarged by random rearrangements of these gene segments in somatic cells. Due to the large numbers of various gene segments and their random joining, a vast variety of combinations are possible (Table.3).

3. Junctional Flexibility Contributes to Diversity

It has been found very often that the joining of recombination signal sequences is precise while the joining of the coding sequence is imprecise. It is possible that V-J, D-J and V-(D)-J joining can take place at any of the several nucleotides present at the ends of joining segments and may produce various codon combinations leading to a change in amino acid sequence which is actually coded by the original DNA (Fig.9).

Table 3 : Combinatorial antibody diversity in human. The slight differences in numbers may be seen among different individuals. Here only the functional gene segments have been listed. The genome contains additional segments that are incapable of rearrangement or contain stop codons or both

Possible V-D-J joining combinations in heavy chain = 250 V X 30 D X 6 J genes = 45000

Possible V-J joining combination in kappa chain = 100 V X 5 J = 500

Possible V-J joining combination in lambda chain = 100 V X 6 J = 600

Possible combination of heavy and light chain = 45000 X (500+600) = 4.95 X 10⁷

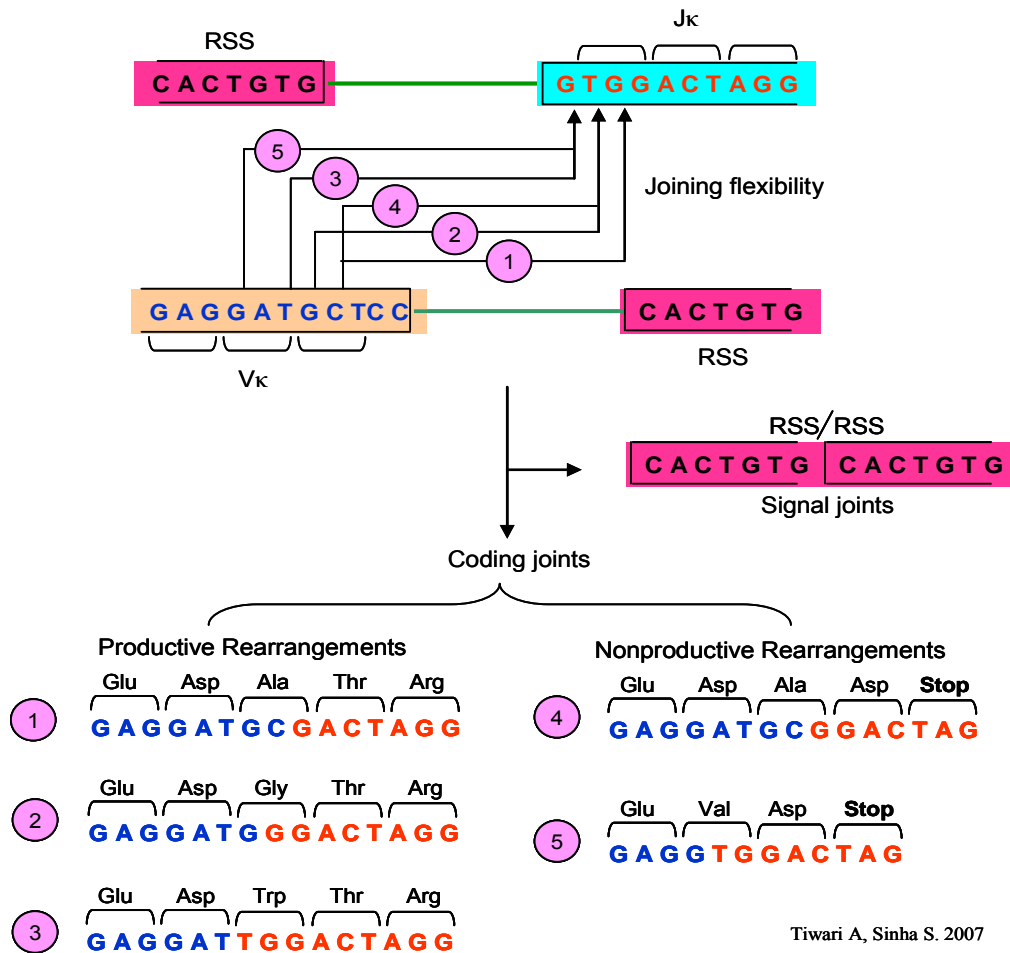


Fig. 9: Schematic representation of junctional flexibility in the joining of variable region gene segments. Joining of gene segments (represented by arrows) generates both productive (1, 2 and 3) and nonproductive rearrangements (4 and 5). Productive rearrangements lead to functional protein while nonproductive rearrangements contain stop codons and are not translated into protein

Junctional flexibility leads to many nonproductive rearrangements, but it also generates productive combinations that encode alternative amino acid at each coding joint, so increasing the antibody diversity. The amino acid sequence variation generated by junctional flexibility in the coding joint comes under the third complementarity determining region (CDR3) in immunoglobulin heavy chain and light chain DNA. Since CDR3 often makes a major contribution to antigen binding site, amino acid changes by junctional flexibility are important in the antibody diversity.

4. P and N Nucleotide Addition

During the V-J or V-DJ joining, the cleavage of the single strand of DNA at the junction of a variable region gene segment and signal sequence produces short sequence nucleotides at the end of the coding sequence. The end of coding sequence which is made up off short sequence turn back and form a hairpin loop structure (Fig.10 a). This hairpin is later cleaved by an endonuclease. This second cleavage sometimes occurs at a position that leaves a short single strand at the end of the coding sequence. The subsequent addition of complementary bases to this strand by repair enzymes generates a palindromic sequence in the coding joint. This is called P-nucleotide addition and the nucleotides are called P nucleotides. Variation at the site of second cleavage by repair enzymes leads to variation in the coding sequence.

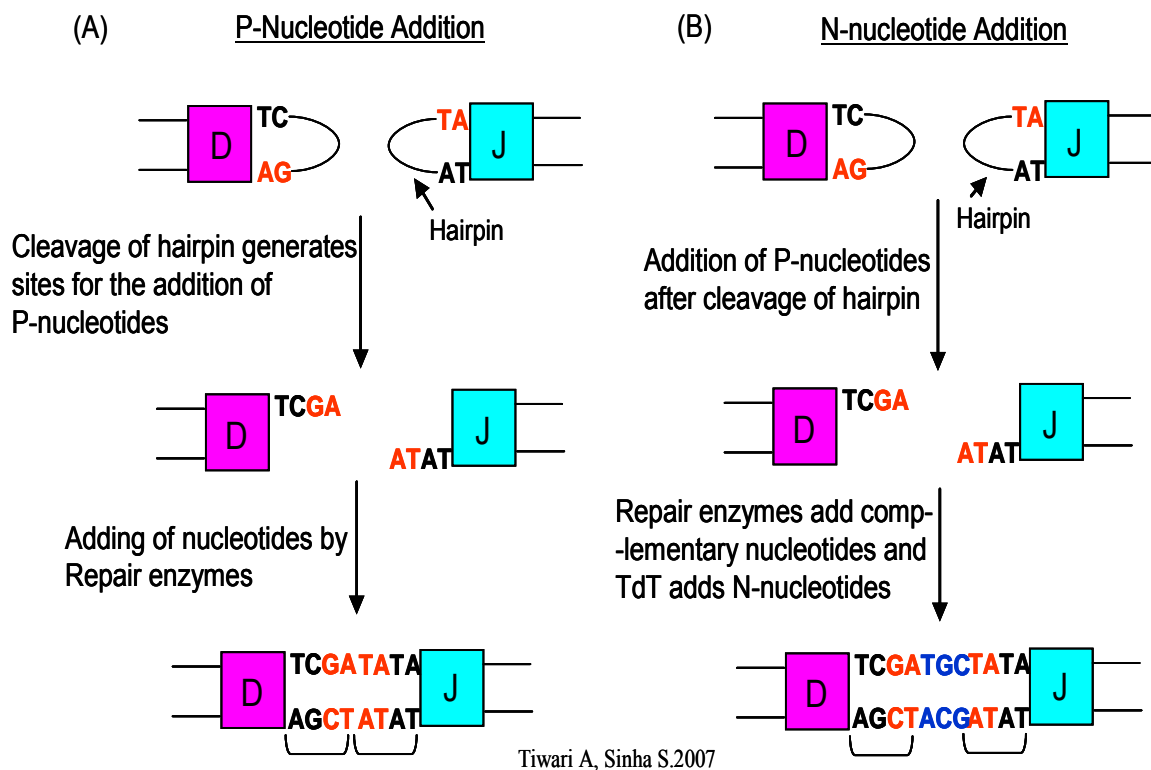


Fig.10: P-nucleotide and N-nucleotide addition during V region gene rearrangement. (a) If cleavage of hairpin generates single stranded end, repair enzymes add P nucleotides (in red) to produce palindromic sequence. (b) Besides P-nucleotides addition, addition of random nucleotides (in blue) by TdT can occur at coding joints of heavy chain

After rearrangement, the coding joints of heavy chain variable region have been shown to contain new short amino acid sequence that are not encoded by the germ-line V,D or J gene segments. These nucleotides are added by an enzyme named as terminal deoxynucleotidyl transferase (TdT). These nucleotides are called N-nucleotides. Maximum 15 nucleotides can be added to both the D_H - J_H and V_H - D_H J_H coding joints (Fig.10b). Thus, a complete functional heavy chain variable region is represented as $V_HND_HNJ_H$ unit. The diversity generated by N nucleotides is quite large because N region consist of entirely random sequences.

5. Somatic Hypermutation

It is apparent so far that antibody diversity is generated during formation of specific variable region gene rearrangement in developing B-cells. It is believed that once functional V-J and V-D-J gene rearrangement completed, the specificity and diversity is defined for ever and should not change anymore. But this is not true; in fact, the variable region genes undergo point mutations on antigenic stimulation that further increases the antibody diversity. The process by which additional antibody diversity generated in the rearranged variable region genes by point mutations is called somatic hypermutation. Somatic hypermutation is targeted to rearranged variable region genes located within a DNA sequence containing about 1500 nucleotides, which includes the whole of VJ or VDJ gene segment. The frequency of somatic hypermutation (10^{-3} per base pair per generation) is around hundred thousand fold higher than the spontaneous mutation rate (10^{-8} per base pair per generation) in other genes. In somatic hypermutation, most of the mutations are nucleotide substitutions rather than deletion or insertion. Somatic hypermutations occur throughout the VJ or VDJ segment and fall within the CDRs. Somatic hypermutations influence the overall affinity for antigens. Since CDRs are involved in antigen binding and somatic hypermutations are clustered within the CDRs, changes in CDR affinity to antigen may lead to change in selection pattern and affect the maturity. Normally the B-cells with high affinity receptors for antigens are preferentially selected for survival. The exact mechanism of somatic hypermutation is not known. Since these hypermutations are seen in B-cells in response to T-cell dependent protein antigens, it is possible that T-cells or their products are involved in somatic hypermutation.

6. Combinatorial Association of Heavy and Light Chains

Since the specificity of antibody is determined by the rearranged functional variable region genes of both light and heavy chains. The possible combinations of heavy and light chains are therefore also contributed in the antibody diversity. Human genome has the potential to generate 45000 heavy chain genes and 1100 light chains genes as a result of variable region gene rearrangements. If any of the heavy chain combines with any of the light chain, the potential combinations of heavy and light chain variable genes will be 4.95×10^7 . This number is quite higher than the amount of combinatorial diversity actually generated in an individual. It is not likely that all variable genes of heavy chain and light chain will pair with each other. The combination process is not completely random, because not all the variable gene segments are used at the same frequency. Some of them are used very frequently, others only occasionally and still some others almost never used.

It is difficult to calculate the precise number of antibody with different combining sites, but we know that it is quite high. In the variable region, very large numbers of new sequences are

generated by junctional flexibility, P and N nucleotide addition. In addition to these above mentioned events, the somatic hypermutation contributes extremely to the antibody diversity.

Antibody Class Switching

There are five classes of antibodies. An antibody of the same specificity can be expressed in each class and subclass. This is important because different classes and subclasses of antibodies have different effector function in spite of the same specificity. A naïve B-cell carries IgM and IgD on its surface, but after antigenic stimulation the heavy chain DNA can go further rearrangement in which the rearranged VDJ gene unit can combine with any of the C_H gene segment as a result B cell can generate other antibody classes. This process of changing antibody class is called as class switching or isotype switching. The mechanism of class switching is unclear, but it involves DNA flanking sequences (switch regions) located 2-3 Kb upstream from each C_H segment. These switch regions are made up off multiple copies of short repeats (GAGCT and TGGGG). It is proposed that there are some class specific recombinases which are capable of binding selectively to these sites. On binding to these sites, these recombinases bring out the splicing of the corresponding C_H region gene to the rearranged VDJ unit and leaving out the intervening DNA. Heavy chain class switching is not a random phenomenon but a regulated one. It is regulated by the various cytokines secreted by T_H cells. Class switching is mainly seen in response to B-cell activation by thymus dependent antigen. This type of activation requires the direct contact of the T_H cells and B-cells. The antibody class switching is minimal in B-cell activation by thymus independent antigens. These cells mainly produce IgM antibody. Interleukin 4 (IL-4) for example, induces class switching from C_μ to $C_\gamma 1$ or C_ϵ . In some cases, IL-4 has been observed to induce class switching in a successive manner (Fig.11).

Class switching depends upon the basically three elements: switch region, switch recombinases and cytokine signals that direct the isotype to which the B-cell switches.

Expression of Immunoglobulin Genes

The gene rearrangements leading to antibody diversity and class switching occurs at the DNA level, but post transcriptional processing of immunoglobulin primary transcripts is required to produce functional mRNAs. After the variable region gene rearrangements, primary transcripts contain intervening DNA sequences (non-coding introns) and J gene segments not lost during VDJ rearrangement. In heavy chain C gene segments are also organized as a series of coding exons and non-coding introns. The primary transcript must be processed to remove the intervening DNA sequences. Intervening DNA sequences are removed and the remaining coding regions are connected by a process called RNA splicing. There are some splice sites which are made up off small conserved sequence, present at the intron exon boundaries within a primary transcript. These splice site provide the signals at the position where splicing occurs. Processing of primary transcript occurs within the nucleus, afterward the functional mRNA is exported from the nucleus to cytoplasm where it undergoes translation by ribosomes and form functional heavy or light chains.

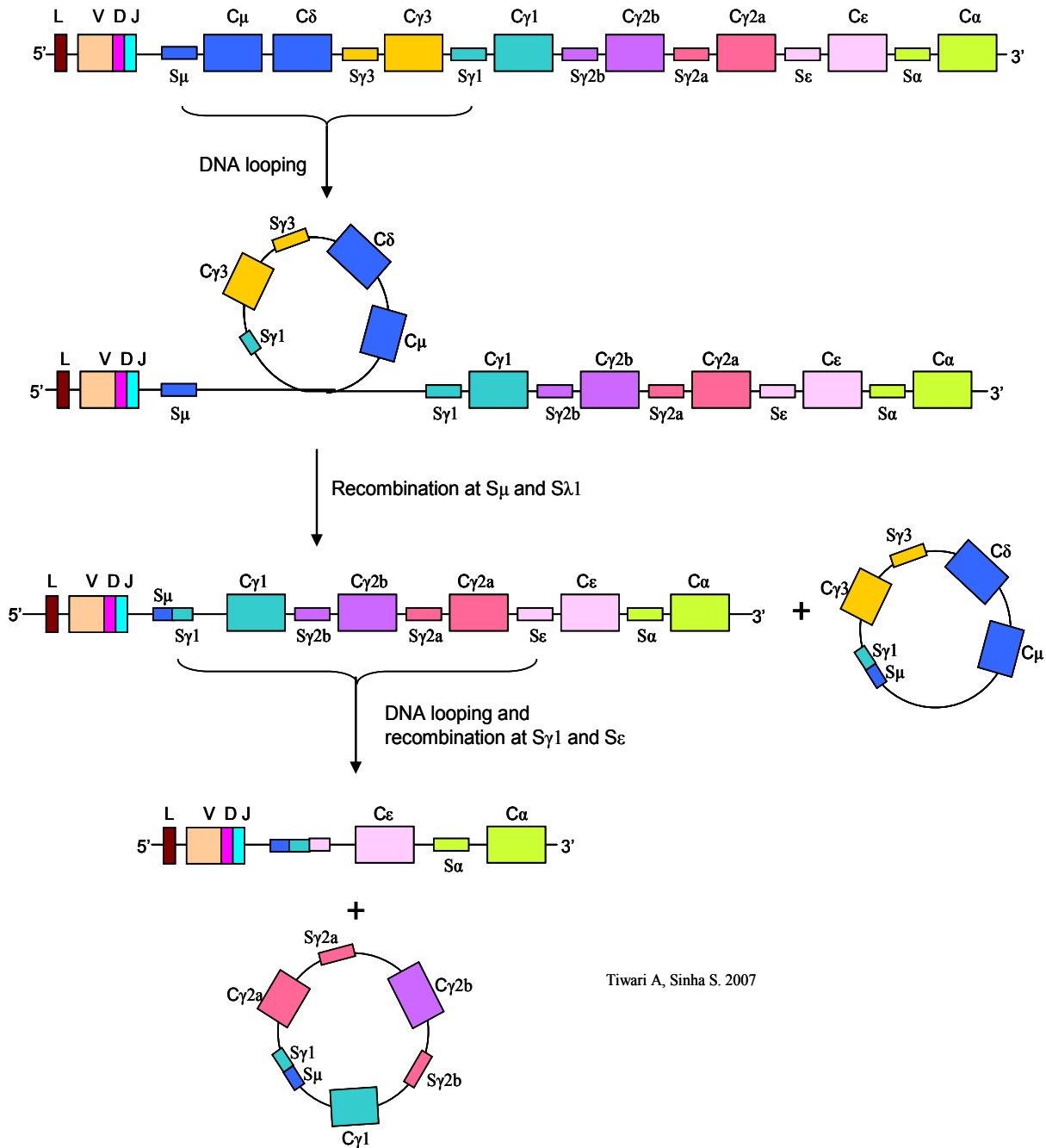


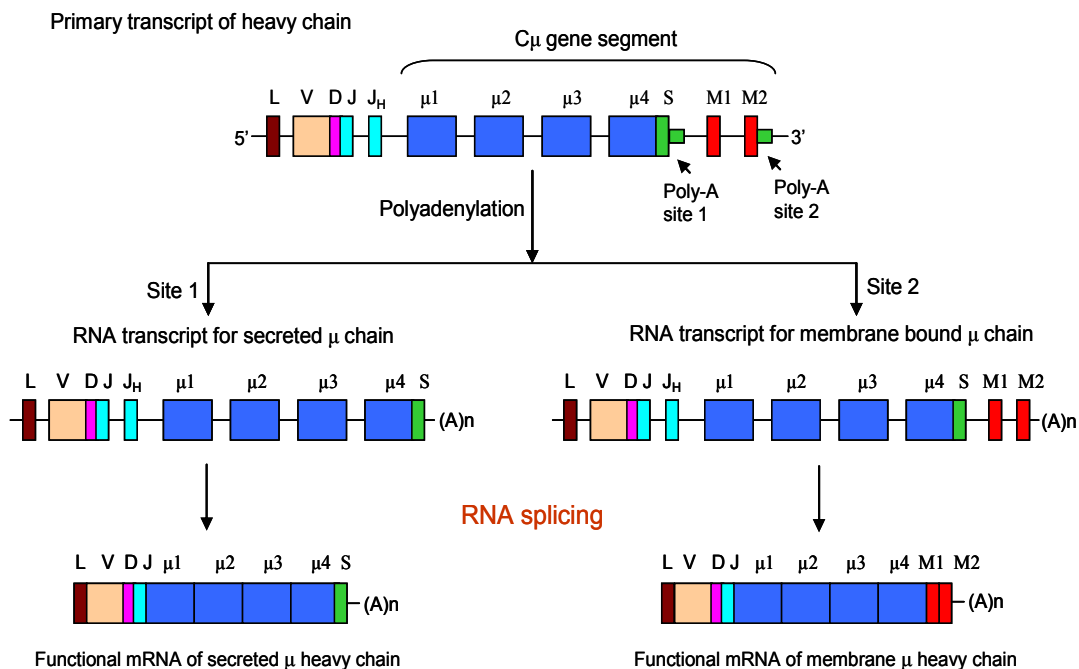
Fig.11: Schematic representation of proposed mechanism for class switching induced by interleukin 4 in heavy chain. Switch sites are present upstream from each C_H segment except C δ . Interleukin 4 induces sequential class switching from C μ to C λ 1 to C ϵ

Expression of Membrane or Secreted Immunoglobulin

A particular immunoglobulin can exist in either membrane bound or secreted form. The two forms of the immunoglobulin differ in the heavy chain carboxy terminal domains. The secreted form has a hydrophilic sequence of about 20 amino acids in the carboxyl-terminal domain, while

the membrane bound form contains a hydrophilic segment within the sequence of about 40 amino acids at the carboxyl-terminal end. Since heavy chain DNA contain a single C_H gene segment corresponding to each class and subclass, it is quite surprising the existence of two different forms of the particular immunoglobulin. The puzzle of existence of these two forms of immunoglobulin resolves by the sequencing studies of the C_μ gene segment. This gene has four exons $C_{\mu 1}$, $C_{\mu 2}$, $C_{\mu 3}$ and $C_{\mu 4}$ encoding the four domains of the IgM molecule. The $C_{\mu 4}$ exon has a specific sequence called as ‘S’ sequence at its 3’ end. The S sequence encodes 13 amino acid long hydrophilic region of the secreted IgM. There are two additional exons located 1.8 Kb downstream from the 3’ end of $C_{\mu 4}$ – called M1 and M2. The M1 exon encodes 26 amino acid long transmembrane domain and M2 encodes the cytoplasmic segment of the C_{H4} domain in membrane bound IgM. DNA sequencing revealed that all the CH gene segments have two additional downstream M1 and M2 exons.

The primary transcript of rearranged μ heavy chain gene contain two polyadenylation sites – Poly A1 and Poly A2 located at the 3’ end of $C_{\mu 4}$ exon (site1) and M2 exon (site2) respectively (Fig.12). These two polyadenylation sites act as signal sequence for cleavage of primary RNA transcript. If cleavage of primary transcript and polyadenylation occurs at site1, the M1 and M2 exons are lost. Excision of M1 and M2 exons produces mRNA which encodes secreted form of the heavy chain. If cleavage and polyadenylation of the primary transcript occurs at site2, then splicing removes the S sequence from 3’ end of the $C_{\mu 4}$ exon and joins the remainder of the $C_{\mu 4}$ exon with M1 and M2 exons. This final mRNA encodes the membrane form of the heavy chain.



Tiwari A Sinha S.2007

Fig.12: Schematic diagram of the primary transcript of a rearranged heavy chain gene showing the C_μ exons and polyadenylation site. Polyadenylation of the primary transcript at either site 1 or site 2 followed by RNA splicing generates functional mRNAs either for secreted or membrane bound μ chains

The differential processing of common primary transcript determines the fate of immunoglobulin whether the membrane form or secreted form will be produced. Mature plasma cells produce secreted antibodies, whereas naïve B-cells produce only membrane bound antibodies. It is not clear, why the B-cells have RNA processing preferentially towards the membrane bound immunoglobulins and plasma cells towards the secreted form.

Simultaneous Expression of IgM and IgD

Transcription of rearranged heavy chain genes in mature B-cells produces primary transcripts containing both the C μ and C δ gene segments. These two gene segments are situated relatively close to each other (only about 5 Kb apart) and there is no switch site located between them so the entire VDJC μ C δ region is transcribed into a single primary transcript about 15 Kb long containing four polyadenylation sites (Fig.13). Sites 1 and 2 are associated with C μ gene segment and sites 3 and 4 are located at similar places in the C δ gene segment. Cleavage at site 2 after polyadenylation leads to membrane μ heavy chain, while cleavage at site 4 leads to removal of C μ gene segment and production of membrane bound δ heavy chains. Cleavage at sites 1 or 3 leads to secretory μ heavy chain and secretory δ heavy chain mRNA transcripts respectively.

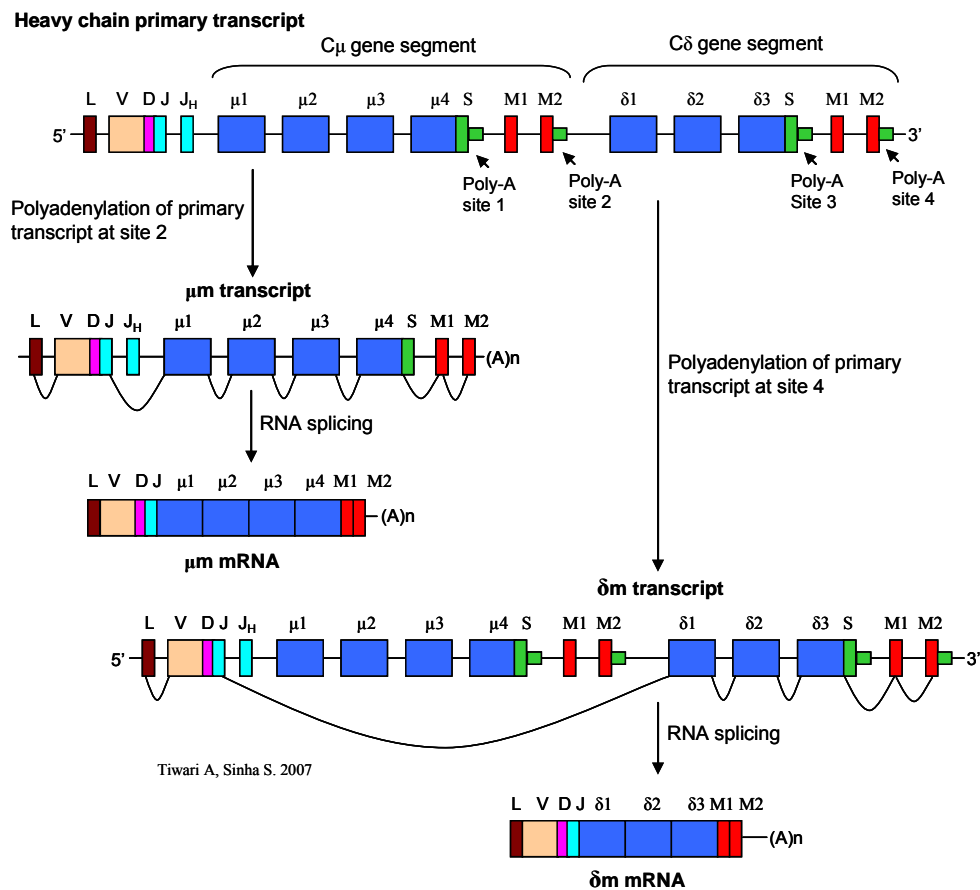


Fig. 13: Structure of rearranged heavy chain primary transcript showing C μ and C δ exons and Polyadenylation sites. Schematic representation of expression of membrane forms of μ and δ heavy chains by alternative processing

Transcriptional Regulation of Immunoglobulin Genes

Immunoglobulin genes are expressed specifically in B-cells and they express at different rates during various developmental stages. Like other eukaryotic genes, there are mainly three types of genetic regulatory elements found in immunoglobulin genes.

1. **Promoters:** contains relatively short sequences (highly conserved) rich in AT bases called TATA box. They are located about 200 bp upstream from the transcription initiation site that promotes initiation of RNA transcription in a specific direction.
2. **Enhancers:** Nucleotide sequences which enhance the transcription frequency. They could be located upstream or downstream from a gene whose transcription is to be regulated.
3. **Silencers:** Nucleotide sequences that down-regulate transcription. They are also located upstream or downstream from a gene like enhancers.

The locations of these regulatory elements in germ-line immunoglobulin DNA are shown in Fig.14. The actual process of transcription is performed by RNA polymerase II. Promoters are located quite far apart from the enhancers (250-300Kb) in the germ-line variable genes. Gene rearrangement in germ-line DNA brings the promoters and enhancer quite close (with in 2Kb distance from each other). Consequently, the transcription frequency of rearranged VJ or VDJ gene segments is enormously enhanced.

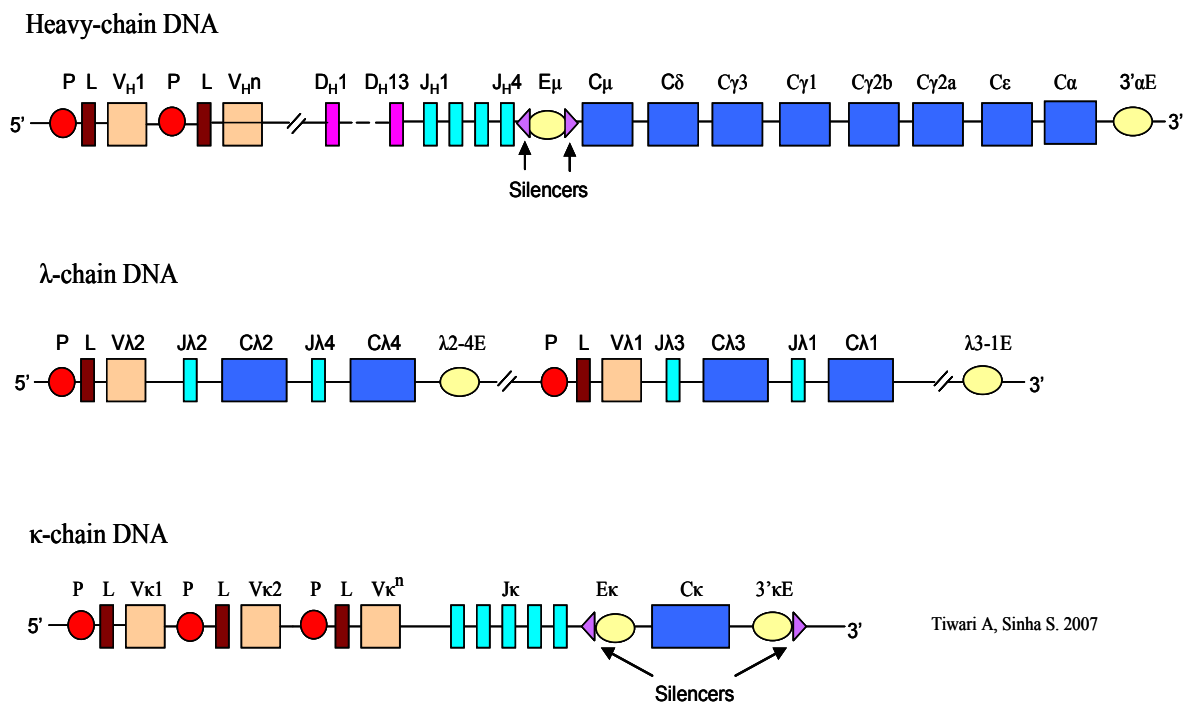


Fig. 14: Location of promoters (red circle), enhancers (yellow oval) and silencers (purple triangle) in mouse heavy chain, lambda light chain and kappa light chain germ line DNA. Variable region DNA rearrangements brings the enhancers close to promoter thus, enhances the transcription frequency

Translational Regulation

An mRNA with long half life is likely to be translated more than that the mRNA with short half life. For example, B cells contain 10 times more mRNA than mRNA, but they express more mIgD than mIgM.

Suggested Reading

1. Hozumi N, Tonegawa S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. (1976). *Proc. Natl. Acad. Sci. U.S.A.* 73 (10): 3628-32.
2. Janeway CA, Jr et al. *Immunobiology*. 2001, 5th ed. Garland Publishing.
3. Richard A Goldsby, Thomas J. Kindt, Barbara A. Osborne, Janis Kuby. *Immunology*, 2003, 5th edition. W.H. Freeman and company, New York.
4. Abul K. Abbas, Andrew H. Lichtman, Jordan S Pober. *Cellular and Molecular Immunology*. 4th Edition. W.B. Saunders Company.
5. Nossal G. J. V. and Lederberg J. Antibody production by single cells. 1958 *Nature* 181:1419-1420
6. Forsdyke, D.R. The Origins of the Clonal Selection Theory of Immunity. 1995 *FASEB. Journal* 9:164-66.