ORGANISATION AND EXPRESSION OF IMMUNOGLOBULIN GENES

GENERATION OF ANTIBODY DIVERSITY

One of the most important features of the vertebrate immune system is its ability to respond to apparently limitless array of foreign Ags.

Virtually each and every Ab molecule contains a unique AA sequence in its variable region.

The constancy in constant region and tremendous variation in variable region in a single protein molecule is a characteristic of Ab.

In germ-line DNA, multiple gene segments encode portions of a single Ig H or L chain…..

- ….These gene segments can not be transcribed or translated into complete chains until they are rearranged into functional genes.
- During B cell maturation in bm, certain of these gene segments are randomly shuffled to generate more than 10^6 combinations.
- A mature and immunocompetent B cell contains coding sequences for one functional H chain variable region and one light chain variable region.
- The individual B cell is thus antigenically committed to a specific epitope

After antigenic stimulation of a mature B cell in peripheral lymphoid organs, further rearrangement of constant region gene segment can generate changes in the isotype expressed.

Mature B cells contain chromosomal DNA that is no longer identical to germ-line DNA.

Genomic rearrangement is an essential feature of lymphocyte differentiation and no other vertebrate cell type has been shown to undergo this process.

For several decades, immunologists sought to imagine a genetic mechanism that could explain the tremedous diversity of Ab structure.

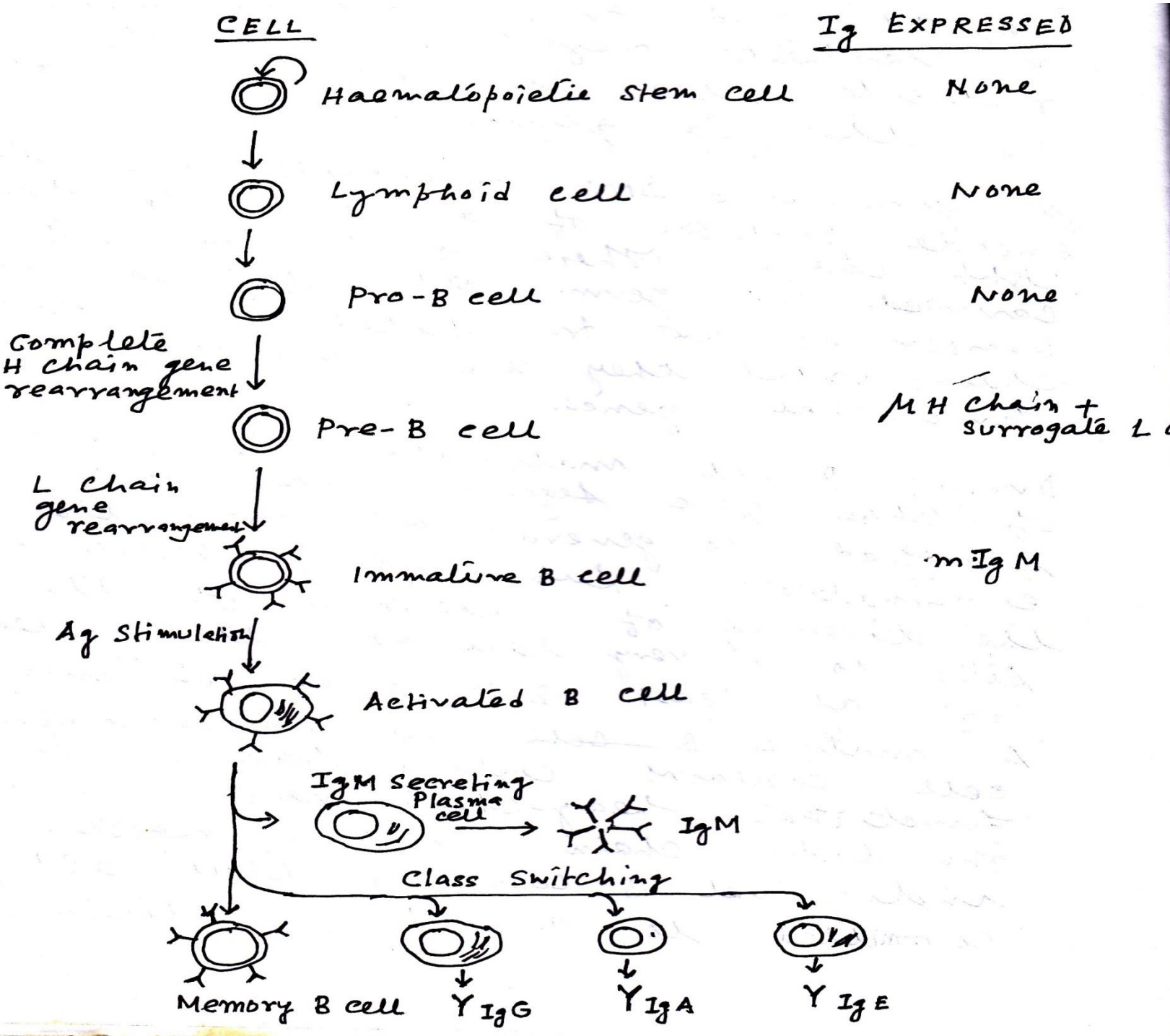
Two different sets of theories have emerged-

1. The Germ-line Theories
2. The Somatic Variation Theories

The GL theories maintained that the genome contributed by the germ cells (egg and sperm), contain a large repertoir of Ig genes. So, no special mechanism to account for Ab diversity.

In contrast, the somatic variation theories maintained that the genome contains a relatively small number of Ig genes, from which a large number of Ab specificities are generated in somatic cells by mutation and recombination.

Whether the diversity was generated by germ-line or somatic variation


CELL
Ig EXPRESSED
Haemalopoietic stem cell
None
Lymphoid cell
None
Pro-B cell
None
Complete H chain gene rearrangement
Pre-B cell
M H Chain + surrogate L c
L Chain gene rearrangement
Immature B cell
mIg M
Ag Stimulation
Activated B cell
IgM Secreting Plasma cell
IgM
Class Switching
Memory B cell
Y IgG
Y IgA
Y IgE

Tonegawa's Bombshell: Ig Gene Rearrangement

Tonegawa & Hozumi (1976) proved the first direct evidence that separate genes encode the V & C regions of Igs & that genes are rearranged in the course of B cell differentiation.

They used various REs to generate DNA fragments from embryonic cells and adult myeloma cells.

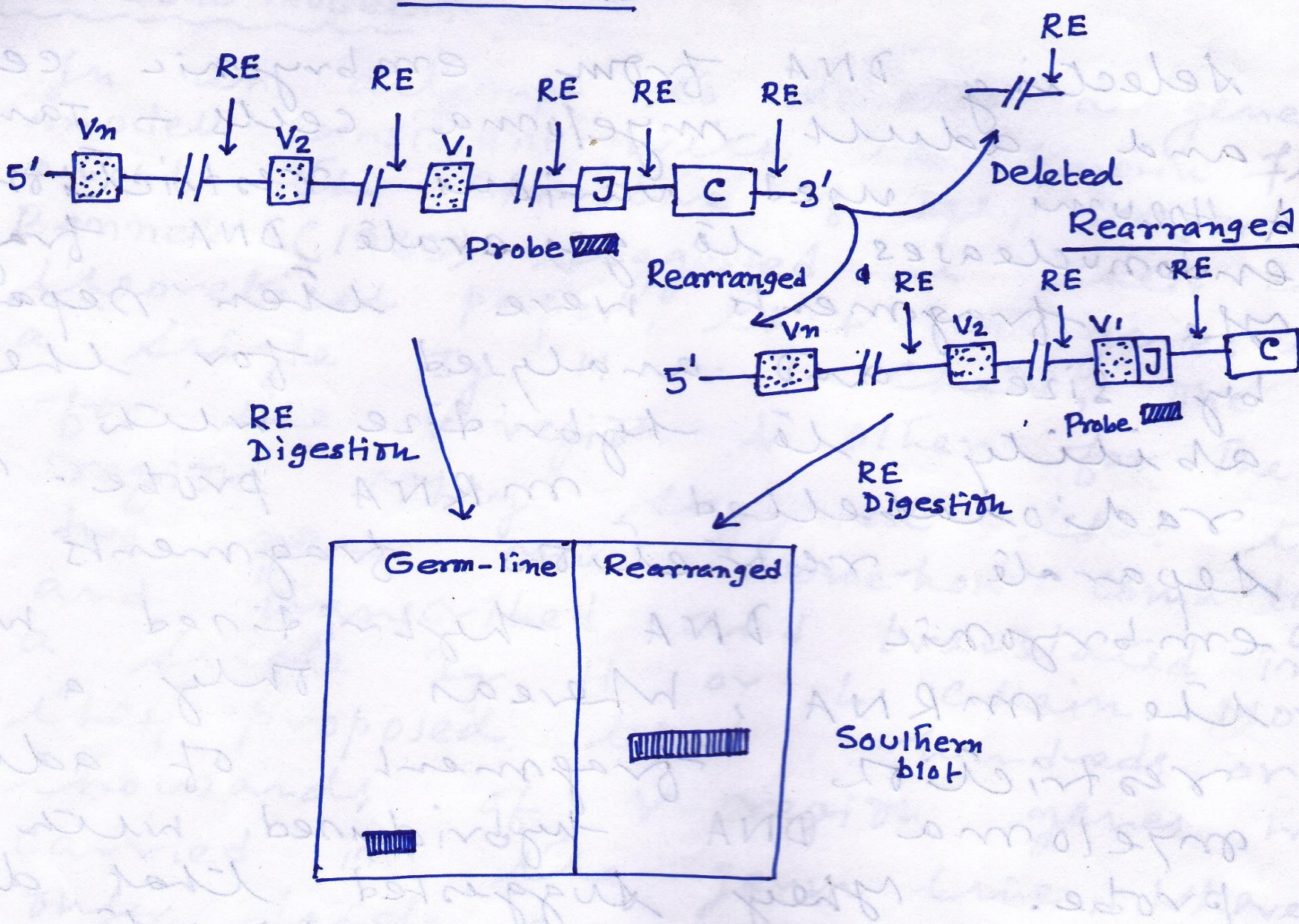
The fragments were then separated by size and analysed for their ability to hybridise with radiolabelled mRNA probe.

Two separate restriction fragments from embryonic DNA hybridised with mRNA, whereas, only a restriction fragment of adult myeloma DNA hybridised with same probe.

They suggest that during differentiation of lymphocytes from embryonic state to the fully differentiated plasma cell stage, the V and C genes undergo rearrangement.

In the embryo, the V and C genes are separated by a large DNA segment that contains a RE site; during differentiation, the V and C genes are brought closer together and the interveni9ng DNA sequence is eliminated.

Germ-Line


RE
RE
RE
RE
RE
RE
Vn
V2
V1
J
C
5'
3'
Probe
Deleted
Rearranged
Rearranged
RE
RE
RE
RE
Vn
V2
V1
J
C
5'
Probe
RE Digestion
RE Digestion
Germ-line
Rearranged
Southern blot

Multigene Organisation of Ig Gene

Cloning and sequencing of the L and H chain DNA revealed that the κ & λ chain and H chains are encoded by separate multigene families situated on different chromosomes.

Gene	Human	Mouse
λ	22	16
K	02	06
Heavy	14	12

The germ-line DNA each of these multigene families contains several coding sequences, called *'gene segments'* separated by non-coding regions.

During B cell maturation, these gene segments are rearranged and brought together to form functional Ig genes.

The κ & λ chain families contain V, J and C gene segments; the rearranged VJ segments encode the variable region of the L chains.

The H chain family contains V, D, J and C gene segments; the rearranged VDJ gene segments encode the variable region of H chain.

In each gene family, the C gene segment encodes the constant regions.

Each V gene segment is preceded at its 5'end by a small exon that encodes a short *'signal' or 'leader peptide'* that guides the H or L chain through ER.

The *'signal' or 'leader peptide'* is cleaved from the nascent L and H chains before assembly of the finished Ig molecule.

Thus, the amino acids encoded by this leader sequence do not appear in the Ig molecule.

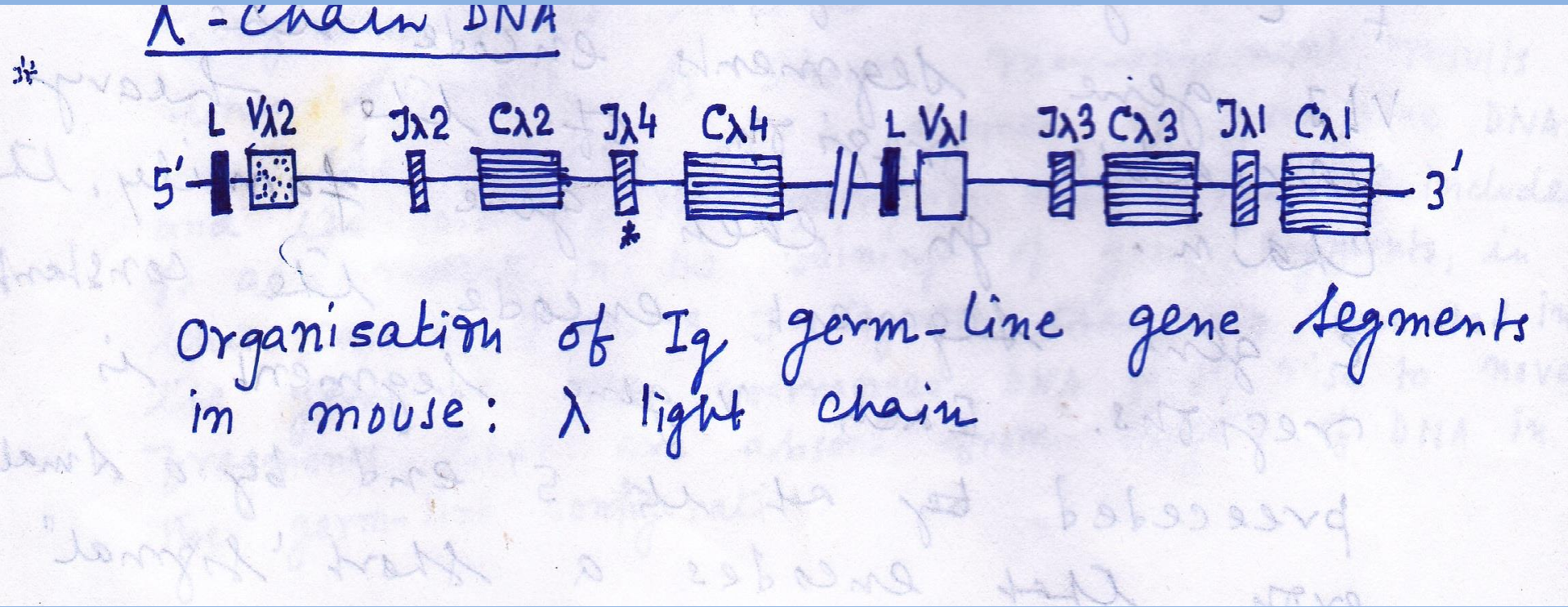
λ chain Multigene Family

The L chain variable region is encoded by two gene segments.

A functional λ variable region gene contains two coding segments – a 5’ V segment and a 3’J segment- which are separated by a non-coding DNA sequencein unaranged germ-line DNA.

The λ multigene family in the mouse germ-line contains two Vλ gene segments, four Jλ gene segments and four Cλ gene segments.

Organisation of Ig germ-line gene segments of mouse



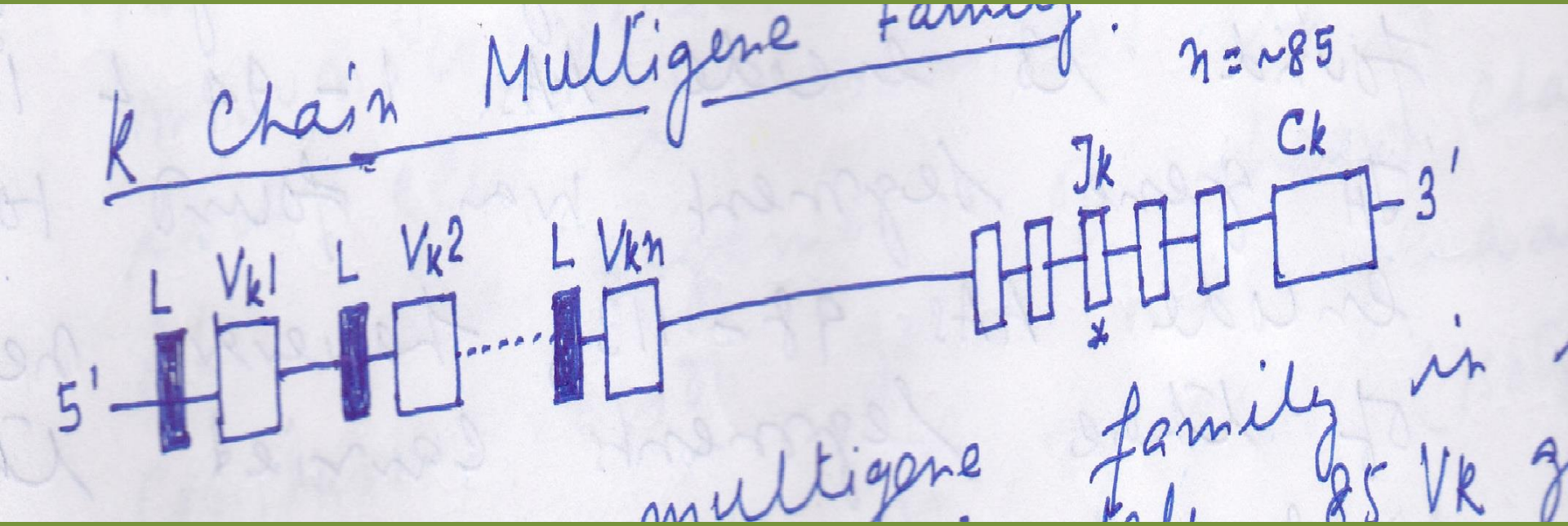
The Jλ4 is a ‘pseudogene’ a defective gene that is incapable of coding protein; such gene are indicated with psi (ψ) symbol.

Interestingly, the Jλ4’s constant region partner, Cλ4 is perfectly functional gene.

In humans, the λ locus is more complex. There are 31 functional Vλ gene segments, 4 Jλ segments, and 7 Cλ segments.

In addition to the functional gene segments, the human λ complex contains many Vλ, Jλ and Cλ pseudogenes.

Kappa Chain Multigene Family



The κ chain multigene family in the mouse contains approximately 85 Vκ gene segments, each with an adjacent leader sequence a short distance upstream (on the 5'side).

There are 5 Jκ gene segments (one is non-functional pseudogene) and a single Cκ gene segment.

As in the λ multigene family, the Vκ and Jκ gene segments encode the variable region of κ L chain, and the Cκ gene segment encodes the constant region.

Since, there is only one Cκ gene segment, there are no subtypes of κ light chains.

The κ chain multigene family in humans, which has an organisation similar to that of the mouse, contains approximately 40 Vκ gene segments, 5 Cκ gene segments.

Heavy Chain Multigene Family

The organisation of Ig H chain genes is similar to that of the κ and λ L chain genes.

An additional gene segment encodes the part of the H chain variable region. Hood et al. compared the H chain variable region AA sequence with the VH and JH nucleotide sequences.

The VH gene segments was found to encode AAs 1-94 and the JH gene segment was found to encode AAs 98-113. However, neither of the segment carried the information to encode AAs 95-97.

From these observations, Hood et al. proposed that a third germ-line segment must join with the VH and JH gene segments to encode the entire variable region of the H chain.

Variable Region Gene Rearrangements

VRGR occurs in an ordered sequence during B cell maturation in bone marrow.

The H chain VRGs rearrange first, then the L chain VRGs.

At the end of the process, each B cell contains a single functional variable region DNA sequence for its H chain and another for its L chain.

- The process of VRGR produces mature, immunocompetent B cells.
- Each such cell is committed to produce Ab.
- Rearrangement of H chain constant region genes will generate further changes in Ig Class (isotypes) expressed by a cell.

L Chain DNA Rearrangement

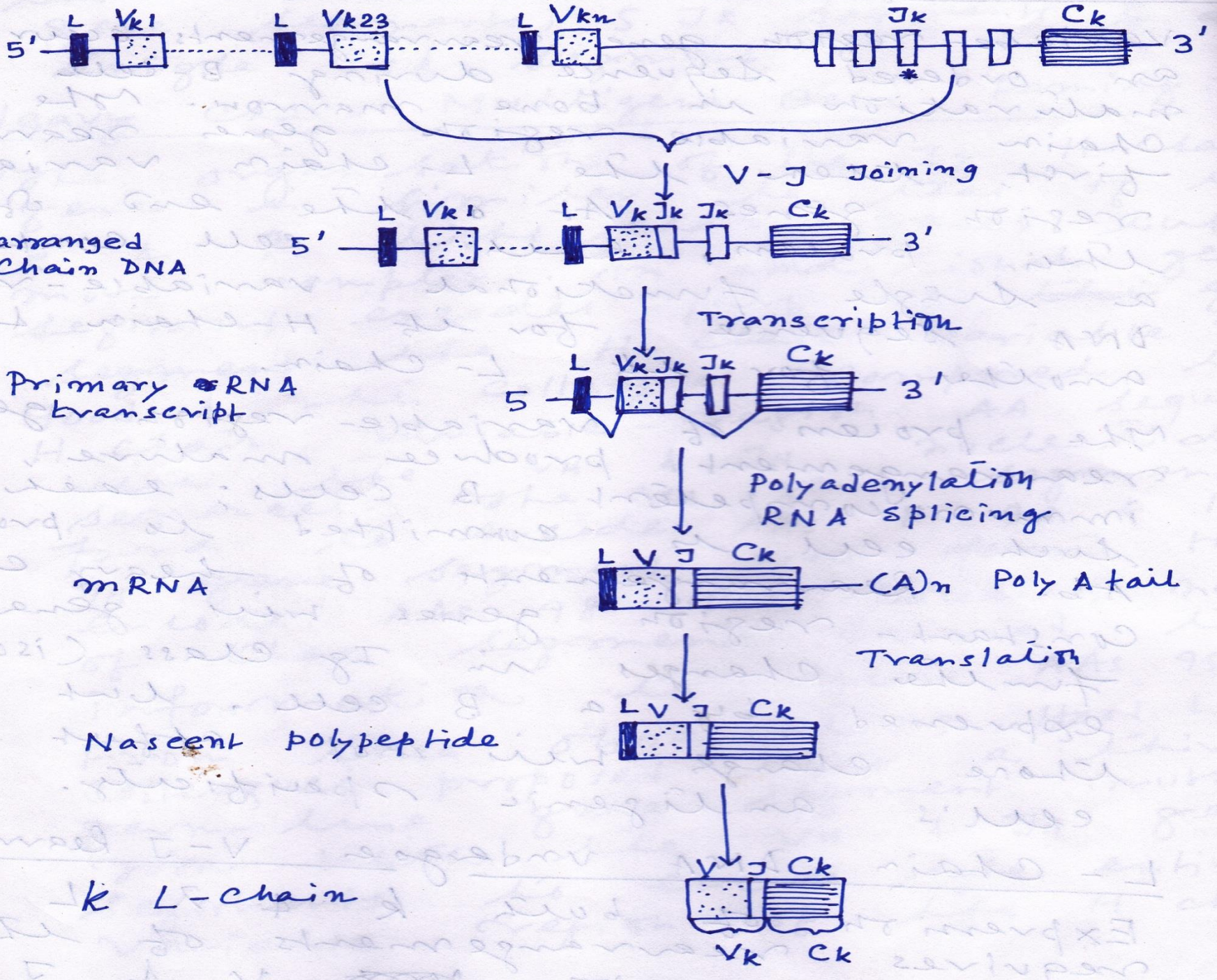
Expression of both κ and λ L chain requires rearrangements of the variable-region V and J gene segments.

Rearranged κ and λ genes contain the following regions in order from the 5'to 3'end- a short leader sequence (L) exon, a non-coding sequence (intron), a joined (VJ) gene segment, a second intron, and a constant region.

The introns in the primary RNA transcripts are removed by RNA processing enzymes, and the resulting L chain mRNA then exits from the nucleus.

The L chain mRNA binds to ribosomes and translated into the L chain protein.

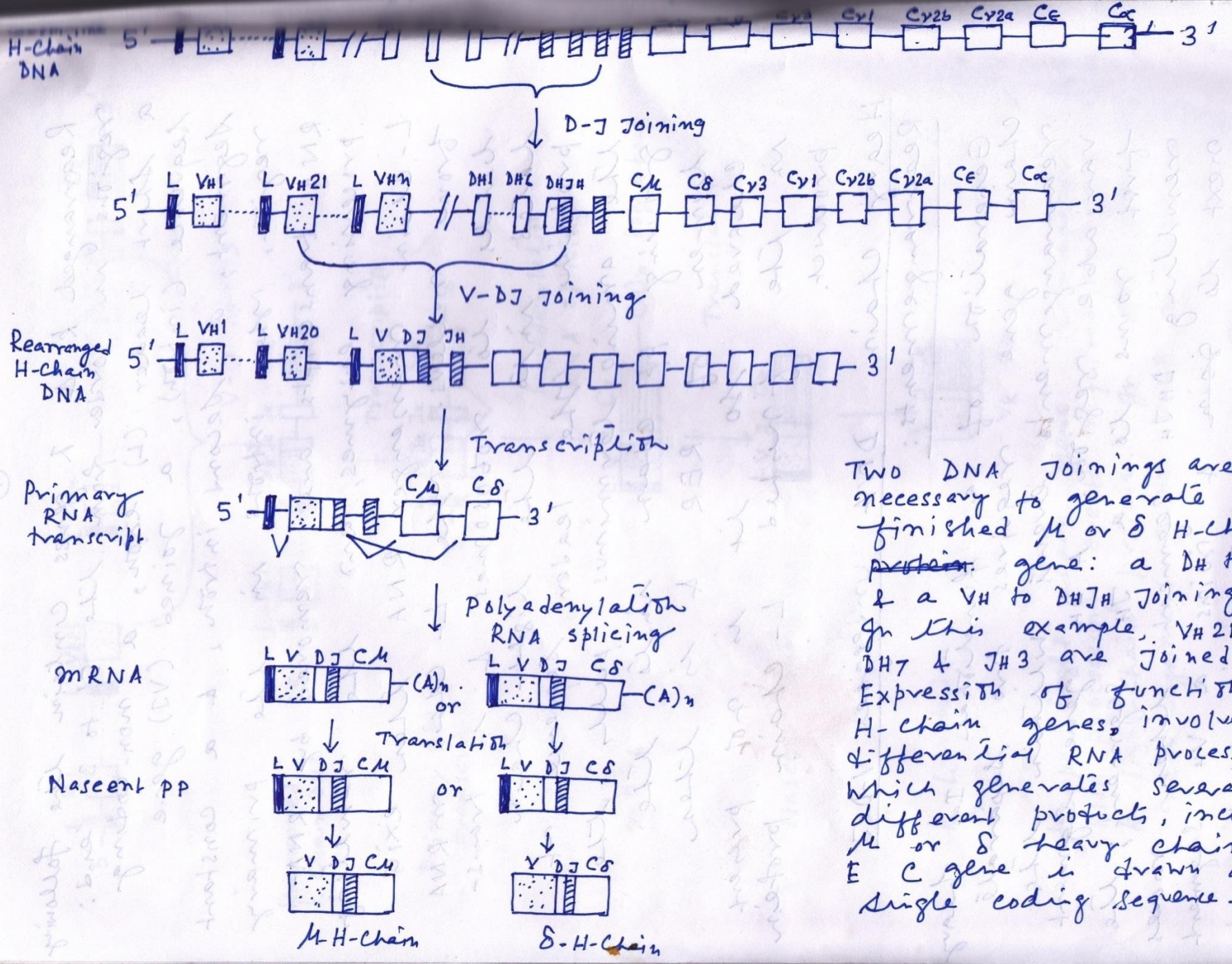
The L sequence at the amino terminus pulls the growing pp chain into the lumen of RER and is then cleaved. So it is not present in the finished L chain protein products.


L Vk1
L Vk23
L Vkn
Jk
Ck
5'
3'
V-J Joining
Rearranged k Chain DNA
L Vk1
L Vk Jk Jk
Ck
5'
3'
Transcription
Primary RNA transcript
L Vk Jk Jk
Ck
5
3'
Polyadenylation
RNA splicing
mRNA
L V J Ck
(A)n Poly A tail
Translation
Nascent polypeptide
L V J Ck
k L-chain
V J Ck
Vk
Ck

H Chain DNA V-D-J Rearrangements

Generation of functional Ig H chain gene requires two separate rearrangement events within the variable region.

A D_H gene segment first joins to J_H segment; the resulting D_HJ_H segment then moves next to and joins a V_H segment to generate $V_HD_HJ_H$ unit that encodes the entire variable region.



Two DNA joinings are necessary to generate finished μ or δ H-chain ~~protein~~ gene: a DH to JH & a VH to DHJH Joining. In this example, VH21, DH7 & JH3 are joined. Expression of functional H-chain genes, involves differential RNA processing which generates several different products, including μ or δ heavy chains. E C gene is drawn as a single coding sequence.

In H chain DNA, variable region rearrangement produces a rearranged gene consisting of following sequences, starting from 5’end: a short L exon, an intron, a joined VDJ segment, another intron and a series of C gene segments. A promoter sequence is located a short distance upstream from each H chain leader sequence

Once H chain gene rearrangement is accomplished, RNA Pol. can bind to the promoter sequence and transcribe the entire H chain gene, including the introns.

Initially, Cμ and Cδ gene segments are transcribed.

Differential polyadenylation and RNA splicing remove the introns and process the primary transcript to generate mRNA including either the Cμ or the Cδ transcript.

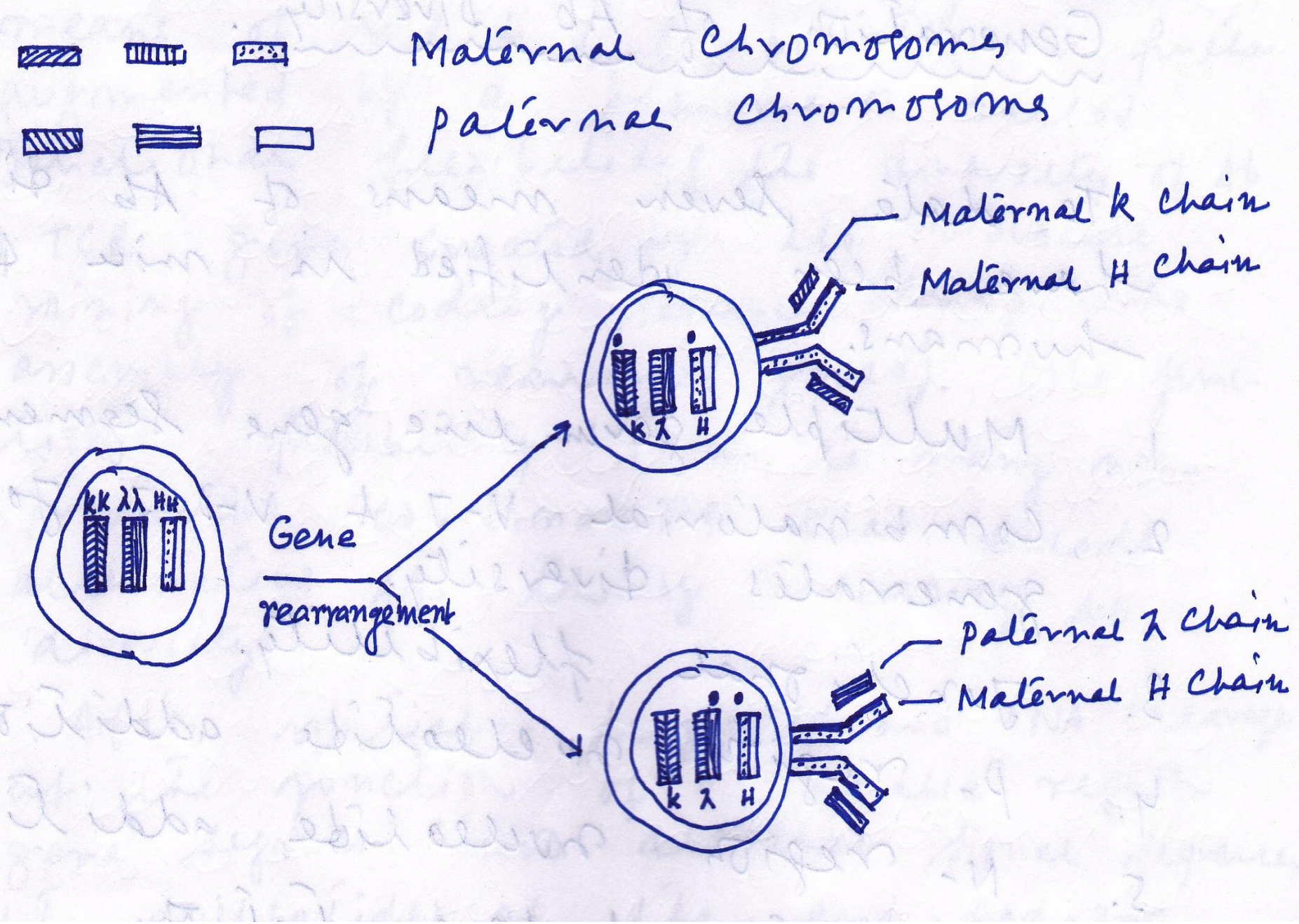
These two mRNAs are then translated and leader peptide is cleaved, generating finished μ and δ chains and allow the production of immunocompetent B cell to express both IgM and IgD.

Allele Exclusion

B cells, like all somatic cells are diploid and contain both maternal and paternal chromosomes. Even though a B cell is diploid, it expresses the rearranged H chain genes from only one chromosome and the rearranged L chain genes from only one chromosome. The process is called “allelic exclusion”. It ensures that functional B cell never contains more than one $V_HD_HJ_H$ and one V_LJ_L unit……

….This is, of course, essential for the antigenic specificity of the B cell, because the expression of both alleles would render B cell multispecific.

The phenomenon of allelic exclusion suggests that once a productive V_H-D_H-J_H rearrangement and a productive V_L-D_L rearrangement have occurred, the recombination machinery is turned off, so that the H and L chain genes on homologous chromosomes are not expressed…..


Maternal Chromosome
paternal Chromosome
Maternal k Chain
Maternal H Chain
kk λλ HH
Gene
rearrangement
k λ H
Paternal λ Chain
Maternal H Chain
k λ H

Generation of Antibody Diversity

To date seven means of Ab diversity have been identified in mice and humans.

1. Multiple germ-line gene segment.
2. Combinatorial V-J and V-D-J joining generates diversity.
3. Junctional flexibility.
4. P-region nucleotide addition.
5. N-region nucleotide addition.
6. Somatic hypermutation.
7. Combinatorial association of L and H chains.

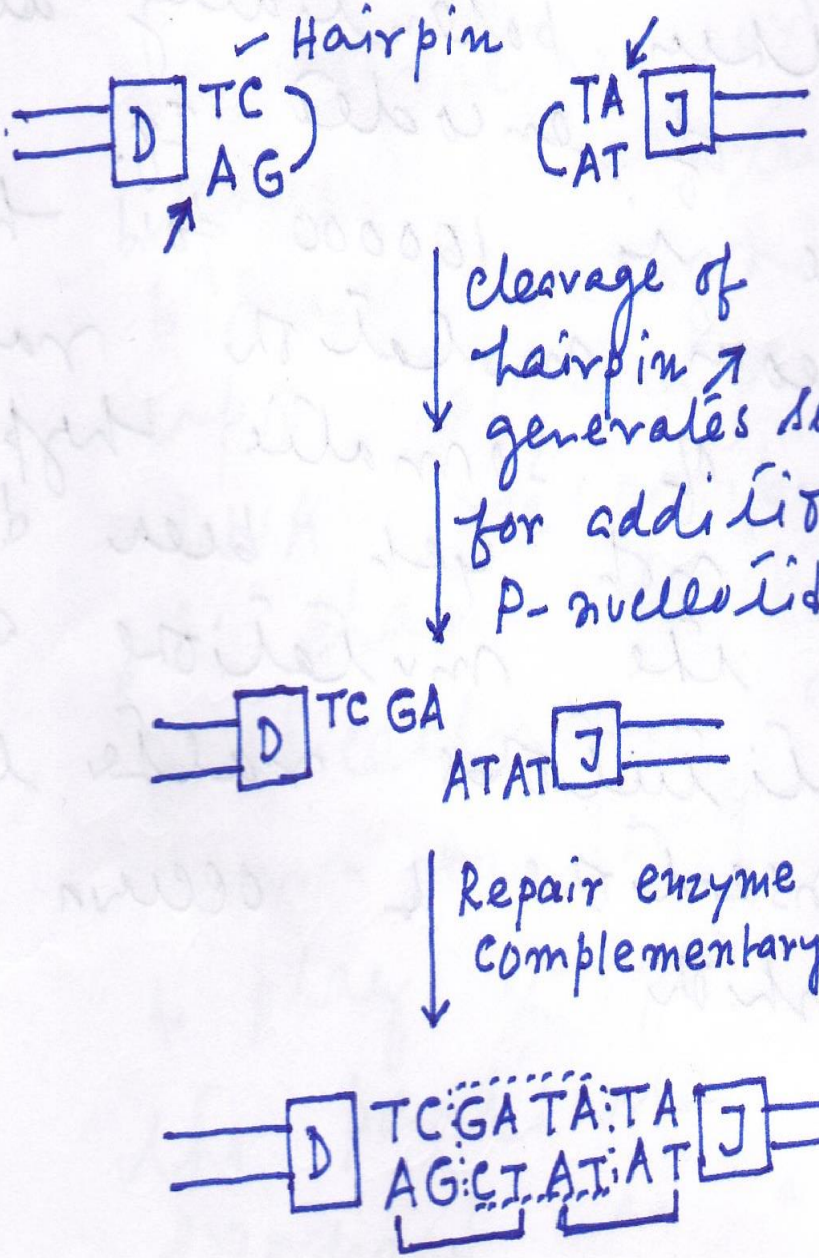
- The germ-line DNA of one human reveals 51V_H, 25D, 6J_H, 40V_K, 5J_K, 31$V\lambda$ and 4$J\lambda$ gene segmemnts. In addition to these functional gene segments, there are some pseudogenes.
- The contribution of multiple germ-line gene segments to Ab diversity is manifested by random rearrangement of these segments in somatic cells. It is possible to calculate how much diversity can be achieved by gene rearrangements although it is not possible to make an exact calculation of their contribution.

- The enormous diversity generated by means of V,D and J combination is further augmented by a phenomenon called *junctional flexibility* (the diversity of Ab and TCR genes created by the imprecise joining of coding sequence during the assembly of rearranged genes). The JF leads to many non-functional combinations that encode alternative AAs, thereby increasing Ab diversity.

- After initial single-strand DNA cleavage at the junction of a variable region gene segment and attached signal sequence turn back to form ‘hairpin’ structure. This hairpin structure is later cleaved by an endonuclease. This second cleavage sometimes occurs at a position that leaves a short single strand at the end of coding sequence. The subsequent addition of complementary nucleotides to this strand (“p-addition”) by repair enzyme generates a palindromic, and so these nucleotides are called “p-nucleotides”.

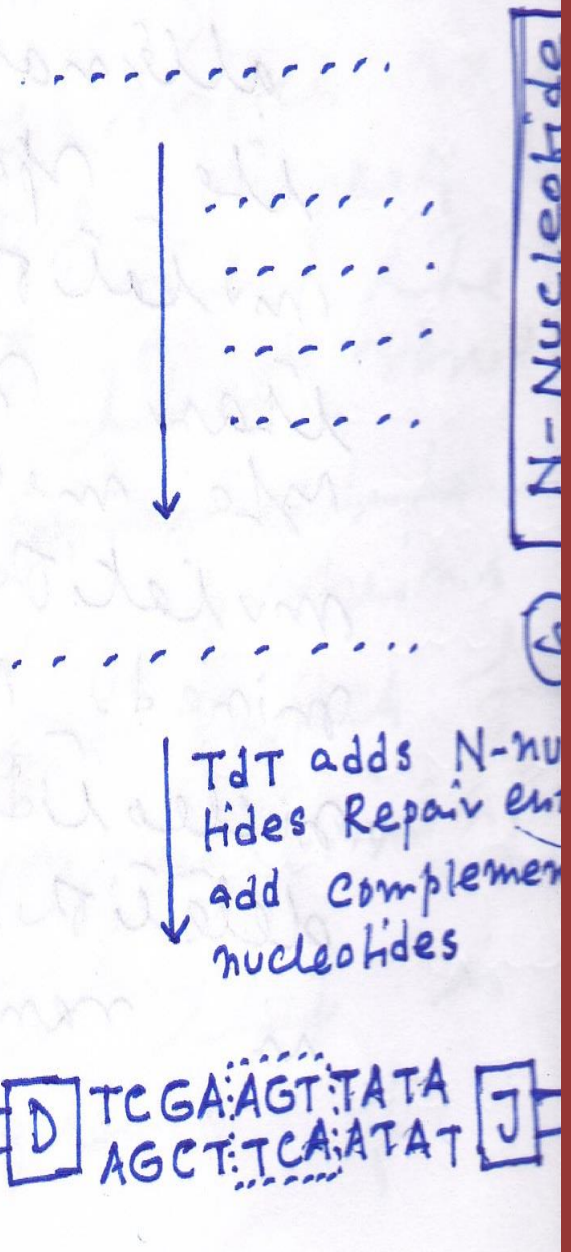
P-Nucleotide Addition

(a)



N-Nucleotide

(b)



Legend to the Previous Diagramm

(a)

If the cleavage of the hairpin intermediate yields a ds end on the coding sequence, p-nucleotide addition does not occur. In many cases, however, cleavage yields a ss end. During subsequent repair, complementary nucleotides are added, called p-nucleotides, to produce palindromic sequence.

(b)

Beside p-nucleotide addition, addition of random N-nucleotides by a terminal deoxytidyl transferase (TdT) can occur during joining of H chain coding sequence.

All the Ab diversity described so far operates during the formation of specific variable regions by gene rearrangements. The additional Ab diversity generates by a process called “somatic hypermutation” and as a result of which, individual nucleotides in VJ and VDJ units are replaced with alternatives, thus potentially altering the specificity of encoded Ig.

- In humans, there is the potential to generate 8262 H chain genes and 320 L chain genes as a result of variable-region gene rearrangements. Although, the number of different Ab combining sites, the immune system can generate, is difficult to calculate with precision. We know that it is quite high.

Antibody Production

- Antibodies can come in different varieties known as isotypes or classes. In placental mammals there are five antibody isotypes known as IgA, IgD, IgE, IgG and IgM. They are each named with an "Ig" prefix that stands for immunoglobulin, another name for antibody, and differ in their biological properties, functional locations and ability to deal with different Ags.

- The antibody isotype of a B cell changes during cell development and activation.
- Immature naïve B cells, which have never been exposed to an Ag, express only the IgM isotype in a cell surface bound form. B cells begin to express both IgM and IgD when they reach maturity—the co-expression of both these Ig isotypes renders the B cell 'mature' and ready to respond to Ag.
- B cell activation follows engagement of the cell bound Ab molecule with an Ag, causing the cell to divide and differentiate into an Ab producing cell called plasma cell.

- In this activated form, the B cell starts to produce Ab in a secreted form rather than a membrane-bound form. Some daughter cells of the activated B cells undergo class switching, a mechanism that causes the production of Abs to change from IgM or IgD to the other Ab isotypes, IgE, IgA or IgG, that have defined roles in the immune system.

Ig class switching (or **isotype switching** or **isotypic commutation** or **class switch recombination**(CSR)) is a biological mechanism that changes a B cell's production of Ab from one class to another, for example, from an isotype IgM to an isotype IgG. During this process, the constant region portion of the Ab H chain is changed, but the variable region of the H chain stays the same. Since the variable region does not change, class switching does not affect Ag specificity. Instead, the antibody retains affinity for the same Ag, but can interact with different effector molecules.

Mechanism

Class switching occurs after activation of a mature B cells via its membrane-bound Ab molecule (or B Cell Receptor) to generate the different classes of Ab, all with the same variable domains as the original Ab generated in the immature B cell during the process of V(D)J recombination but possessing distinct constant domains in their H chains.

- Naïve mature B cells produce both IgM and IgD, which are the first two heavy chain segments in the Ig gene locus. After activation by Ag, these B cells proliferate.
- If these activated B cells encounter specific signaling molecules via their CD40 and cytokine receptors (both modulated by T helper cells), they undergo Ab class switching to produce IgG, IgA or IgE Abs. During class switching, the constant region of the Ig H chain changes but the variable regions, and therefore antigenic specificity, stay the same. This allows different daughter cells from the same activated B cell to produce Ab of different isotypes or subtypes (e.g. IgG1, IgG2 etc.).

Order of exons in H chain

- μ - IgM
- δ - IgD
- γ3 - IgG3
- γ1 - IgG1
- Pseudogene similar to ε gene that is not used
- α1 - IgA1
- γ2 - IgG2
- γ4 - IgG4
- ε - IgE
- α2 - IgA2

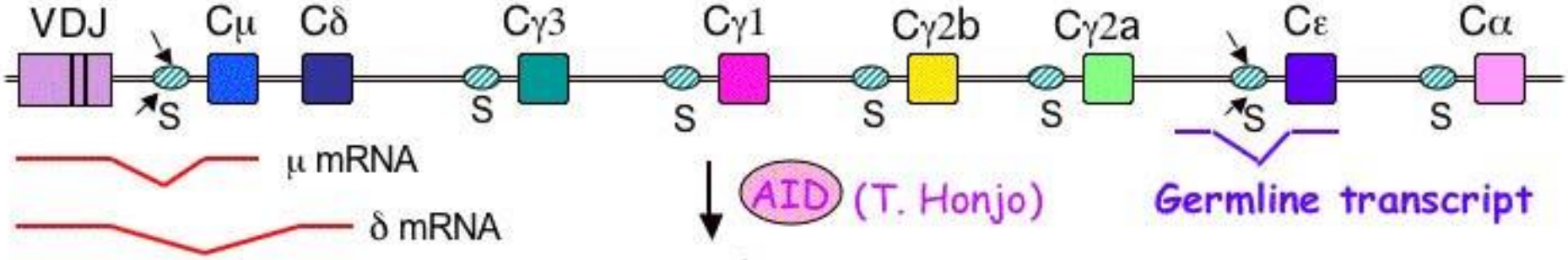
- Class switching occurs by a mechanism called class switch recombination (CSR) binding.
- Class switch recombination is a biological mechanism that allows the class of Ab produced by an activated B cell to change during a process known as class switching.
- During CSR, portions of the Ab H chain locus are removed, and the gene segments surrounding the deleted portion are rejoined to retain a functional Ab gene that produces Ab of a different isotype.

Double-stranded breaks are generated in DNA at conserved nucleotide motifs, called switch (S) regions, which are upstream from gene segments that encode the constant regions of Ab H chains; these occur adjacent to all H chain constant region genes with the exception of the δ-chain.

- DNA is nicked and broken at two selected S-regions by the activity of a series of enzymes, including Activation-Induced (Cytidine) Deaminase (AID), uracil DNA glycosylase and apyrimidic/apurinic (AP)-endonucleases. The intervening DNA between the S-regions is subsequently deleted from the chromosome, removing unwanted μ or δ H chain constant region exons and allowing substitution of a γ, α or ε constant region gene segment. The free ends of the DNA are rejoined by a process called non-homologous end joining (NHEJ).

Antibody Class Switch Recombination (CSR)

Heavy chain genes in IgM expressing **cell**

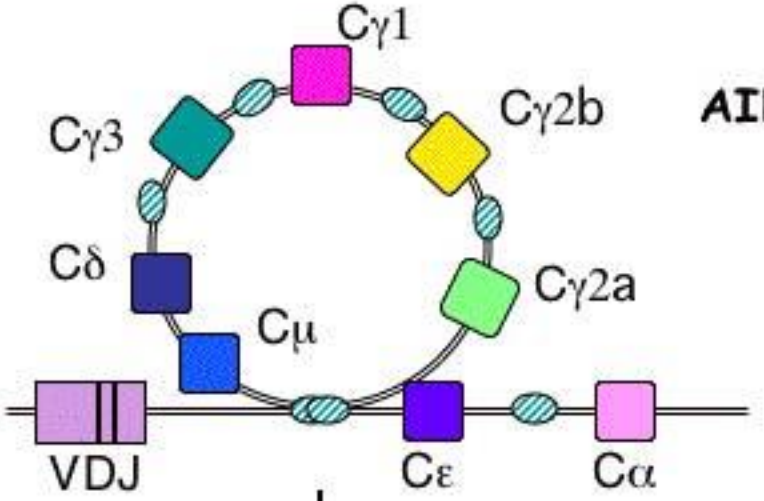


AID (T. Honjo)

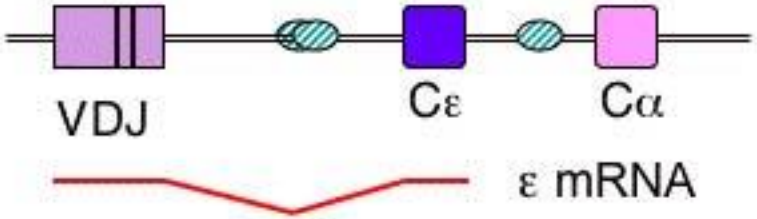
Germline transcript

AID=Activation Induced cytidine Deaminase

Switch recombination

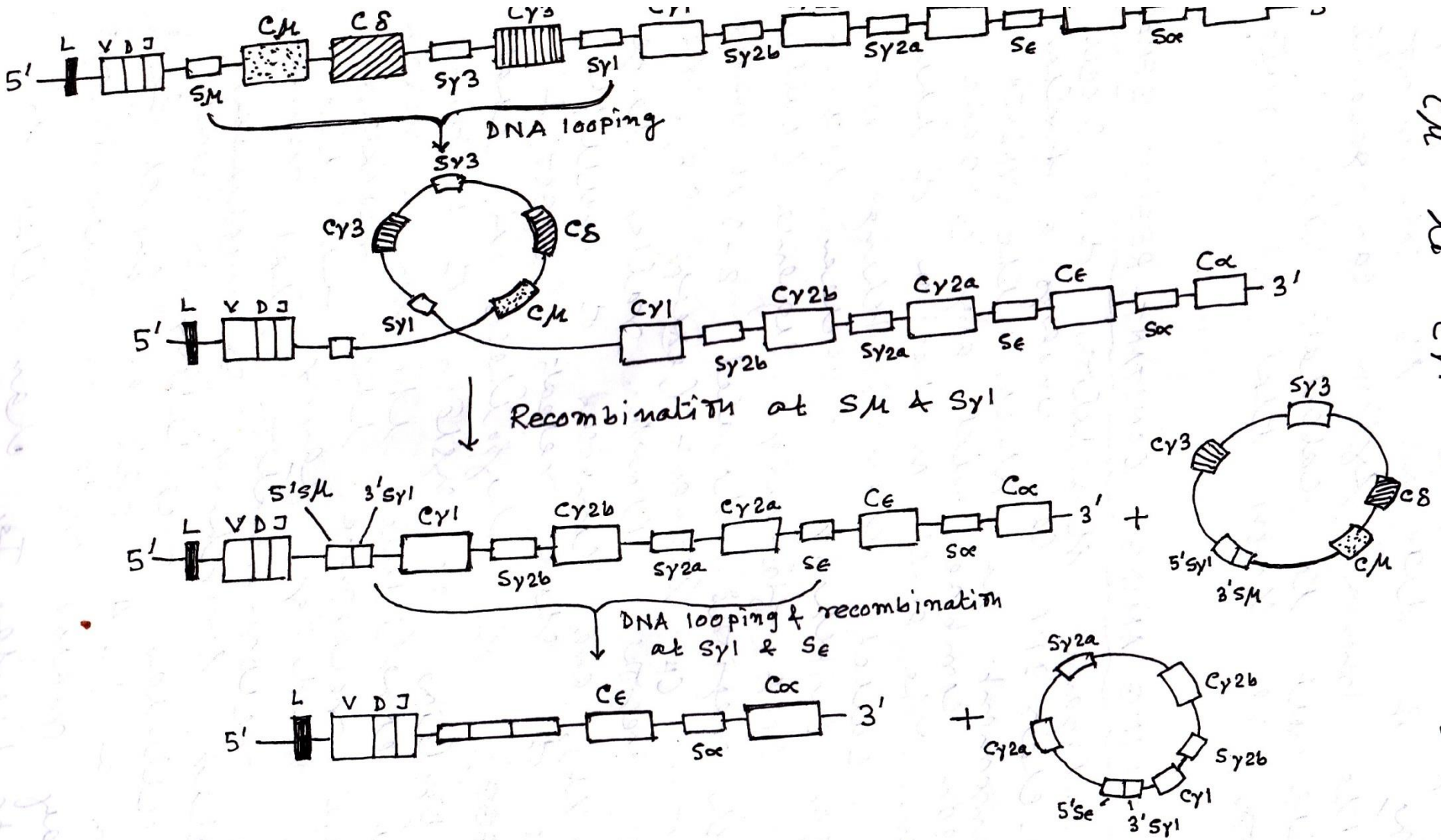


Heavy chain genes in IgE-expressing cell



CLASS SWITCHING AMONG CONSTANT REGION GENES

After antigenic stimulation of a B cell, the H chain DNA can undergo a further rearrangement in which the $V_HD_HJ_H$ unit can combine with any C_H gene segment. The exact mechanism of this process is called “Class Switching” or “Isotype Switching”, is unclear, but it involves DNA flanking sequence (called ‘switch regions’) located 2-3 kb upstream from CH segment (except Cδ). Intercellular regulatory proteins known as IL-4 acts as ‘switch factor’ and plays a major role in determining the particular Ig class that is expressed as a consequence of switching.



Proposed mechanism for class switching induced by IL-4 in rearranged Ig H-chain genes. A switch site is located upstream from each C_H segment except Cδ. Identification of the indicated circular excision products containing portions of the switch sites suggested that IL-4 induces sequential class switching from Cμ to Cγ1 to Cε.

Examinations of the DNA excision products produced during class switching from Cμ to $C\gamma^1$ showed that a circular excision product containing Cμ together with 5'end of the γ^1 switch region ($S\gamma^1$) and the 3'end of the μ switch region (Sμ) was generated.

Furthermore, the switch from C γ^1 to Cϵ produced circular excision products containing $C\gamma^1$ together with portions of the μ, γ and ϵ switch regions.

Thus class switching depends upon the interplay of three elements:

1. Switch regions
2. A switch recombinase and
3. The cytokine signals

EXPRESSION OF Ig GENES

The primary transcripts produced from rearranged H chain and L chain genes contain intervening DNA sequences that include non-coding introns and J gene segments not lost during V-(D)-J rearrangement.

Processing of primary transcript in the nucleus removes each of these intervening sequences to yield the final mRNA product.

The mRNA is then exported from nucleus to be translated by ribosomes into complete H or L chain.

Processing of an Ig H chain primary transcript can yield several different mRNAs, which explains how a single B cell can produce secreted/ membrane bound forms of a particular Ig and simultaneously express IgM and IgD.

Thus, a differential processing of a common primary transcript determines whether the secreted or membrane form of an Ig will be produced.

The mature naïve B cells produce only membrane-bound Ab, whereas differentiated plasma cells produce secretory Abs.

Differential RNA processing also underlies the simultaneous expression of membrane-bound IgM and IgD by mature B cells.

The transcription of rearranged H chain genes in mature B cells produce primary transcript containing both Cμ and Cδ gene segments.

The Cμ and Cδ gene segments are close together in the rearranged gene (only ~5 kb apart).

Since the mature B cell expresses both IgM and IgD on its membrane, both processing pathways must occur simultaneously.

Synthesis, Assembly and Secretion of Igs

Ig H and L chain mRNAs are translated on separate polyribosomes of RER.

Newly synthesised chains contain an amino-terminal leader sequence, which serves to guide the chains into lumen of the RER, where the signal sequence is then cleaved.

The assembly of L and H chains into the S-S linked and glycosylated Ig molecules occurs as the chains pass through the cisternae of RER.

The complete molecules are transported to the G. apparatus and then into secretory vesicles, which fuse with the plasma membrane.

THE END