VPM CLASSES

# GATE SCIENCE

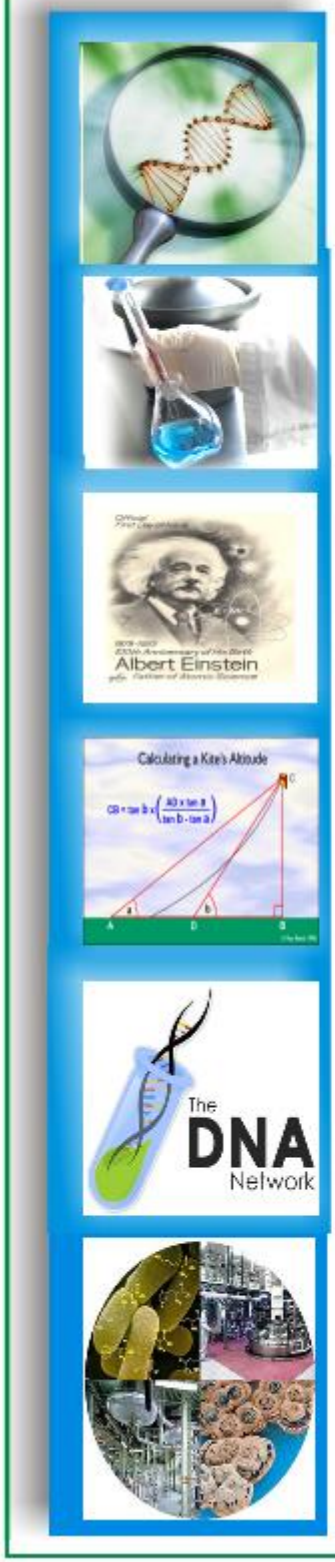
## BIOTECHNOLOGY

## SAMPLE THEORY

* GENOME ORGANISATION
* MOLECULAR STRUCTURE OF GENE
* CHROMOSOMES & CHROMATIN

9001894070

www.vpmclasses.com




# GATE SCIENCE - BIOTECHNOLOGY

## SAMPLE THEORY

- GENOME ORGANIZATION
- MOLECULAR STRUCTURE OF GENE
- CHROMOSOMES AND CHROMATIN

VPM CLASSES

For IIT-JAM, JNU, GATE, NET, NIMCET and Other Entrance Exams

Web Site www.vpmclasses.com E-mail-vpmclasses@yahoo.com

## 1. Genome Organization

### Genome

- The genome is defined as the totality of all the DNA of an organism. Most genomes (all eukaryotes and prokaryotes and some virus), are made up of DNA but a few viruses have RNA genome. The eukaryotic genome consists of two distinct parts: **Nuclear genome** and **Organelles** (mitochondrial and chloroplast) **genome**.
- The amount of DNA present in haploid genome of species is called **C-value**. C- value is characteristic of each species.
- The value ranges from $<10^6$ bp as in smallest prokaryote, mycoplasma to more than $10^{11}$ bp for amphibians.
- In lower eukaryotic organisms like yeast, amount of DNA increases with increasing complexity of organisms.
- In higher eukaryotes there is no correlation between increased genome size and complexity.
- This lack of correlation between genome size and genetic complexity is refers to **c-value paradox**.

### Genome size in some eukaryotes

| Organism | Genome size (Mb) |
|---|---|
| S. Cerevisiae (yeast) | 12 |
| A. Thaliana (mustard plant) | 120 |
| D. melanogaster (fruit fly) | 170 |
| H. sapiens (human) | 3,300 |
| H.vulgare (barley) | 5,300 |

### Genome Complexity

- Genome complexity is the total length of different sequences of DNA.Genome complexity can be measured through the renaturation kinetics of denatured DNA.
- Renaturation of DNA occurs through complementary base pairing.

- A DNA reassociation reaction is described by the **$Cot_{½}$**. If large DNA is sheared into uniform fragments and allowed to renature, then the rate of renaturation of denatured DNA is expressed as:

$$\frac{dc}{dt} = -KC^2$$

where k is the second-order rate constant. C is the concentration of single-stranded DNA at time t and the second order rate equation for two complementary strands coming together is given by the rate of decrease in C.

- The time for half of the DNA to renature (when $C/C_0 = 0.5$) is defined as $t = t_{½}$. Then,

$$\mathbf{C_0 t_{1/2} = \frac{1}{k}}$$

- The product of $C_0 \times t_{½}$ is called the $Cot_{½}$.
- A graph of the fraction of single-stranded DNA reannealed ($C/C_0$) as a function of $C_0t$ on a semilogarithmic plot is referred to as a **Cot** (pronounced "Cot")**curve**.
- $Cot_{½}$ of a reaction indicates the total length of different sequences that are present.
- The $Cot_{½}$ for the renaturation of the DNA of any genome is proportional to its complexity.

**Repetitive sequences**

- Two general types of genomic sequences: Nonrepetitive DNA (consists of sequences that are unique i.e. there is only one copy in a haploid genome) and **Repetitive DNA** (sequences that are present in more than one copy is a haploid genome).
- Prokaryotes contain only nonrepetitive DNA in lower eukaryotes, most of the DNA is nonrepetitive.
- In higher eukaryotes, upto half of the DNA often is occupied by repetitive components.
- The length of the repetitive DNA component tends to increase with overall genome size.
- The repetitive DNA sequences divided into two general classes: **Moderately repetitive DNA** sequences and **Highly repetitive DNA** sequences.

**Highly Repetitives Sequences**

- Highly repetitive DNA consists of very short sequences repeated (typically <100bp) many times in tandem in large clusters. It is sometime also known as **simple sequence DNA**.
- Distinctive base compositions of simple sequence DNA and thus differences in buoyant density result in the formation of satellite to the main DNA band. The sequence of these DNA is known as **satellite DNA**.
- One type of human satellite DNA, the haploid DNA repeats found in the centromere regions of chromosomes. Satellite DNAs are species specific.
- Satellite DNA sequences generally are not transcribed and are located most often in the heterochromatin associated with the centromeric regions of chromosomes.

**Minisatellites and Microsatellites**

- Minisatellites form clusters up to 20 kb in length, with 20-50 repeat units, each containing about 15 to 100 base pairs.
- **Telomeric DNA**, which in humans comprises hundreds copies of the motif 5'- TTAGGG-3' is an example of a minisatallite.
- Microsatellite clusters are shorter, usually <150bp, and the repeat unity is usually 13bp or less. The typical microsatellite consists of a 1-, 2-, 3-, or 4-bp unit repeated 10 to 20 times.

**Table: 1 Major classes of tandemly repeated human DNA**

| Class | Size of repeat unit (bp) | Location |
|---|---|---|
| Satellite DNA (100 kb to several Mb) | 5-171 | Mainly at centromere |
| alpha satellite | 171 | Centromeric hetrochromatin |
| beta satellite | 68 | Centromerric hetrochromatin of some chromosome (1, 9, 13, 14, 15, 21, 22, and y) |
| Minisatallite DNA (upto 20kb) | 9-64 | At or close to telomere |

| | | |
|---|---|---|
| Microsatellite DNA (less then 150 bp) | 2-13 | Dispersed through out all chromosome |

**Moderately repetitive sequences**

- Moderately repetitive DNA consists of relatively short sequences that are repeated typical 10-1000x in the genome. These sequences are dispersed throughout the genome.
- A fraction of the moderately repetitive DNA consists of mobile DNA elements (transposable), short sequences of DNA that have the ability to move to new locations in the genome and/or to make additional copies of themselves.

**2. Molecular Structure of Gene**

Chemically, a gene is formed of DNA, but what of DNA constitutes a gene, has been explained by the relationship between different genetic phenomenon and DNA molecule. **BENZER** has coined new terms to denote the relationship between DNA molecule and genetic phenomena.

1. **Recon–** It is the smallest unit of DNA, capable of undergoing crossing over and recombination. A recon may be as small as one nucleotide pair in DNA.
2. **Muton–** It is smallest unit of DNA which could undergo mutation. In its smallest expression, it represents a change in pair of nucleotides. Any change in base of triplet will modify the message carried by the codon.
3. **Cistron–** It is the unit of function. It is gene in the real sense which consists of number of nucleotides and which capable of synthesizing a polypeptide chain of enzyme. It means a cistron will consist of nucleotides three times the number of total amino acids (a sequence of three nucleotides codes for one amino acid molecule) present in the polypeptide chain for which it has the coded information.
4. **Complon–** It is the unit of complementation. It has been used to replace cistron. Certain enzymes are formed of two or more polypeptide chains; whose active groups are complementary to each other.
5. **Operon–** Operon is the combination of operator gene and sequence of structure genes which act together as a unit. Therefore, it is composed of several genes. The effect of operator gene may be additive or suppressive.

**6.** **Replicon–** It is the unit of replication. Several replicons construct a chromosome.

**Discontinuous Genes (Exons and Introns)**

In 1977, it was discovered that the biological information's carried by some genes is not continuous. It splits into several distinct units separated by regions of noncoding DNA. Such genes are called discontinuous genes or **split genes or mosaic genes**. The sections containing biological information are called **exons** and the intervening noncoding sequences are referred to as **introns**.

**Introns**

Introns are present in the genes of eukaryotes and their viruses. But these are absent in genes of prokaryotes and their viruses. These are present in archaebacteria but are absent in mitochondrial genome.

A single gene may contain no introns or may have as many as 52 as found in mammalian $\alpha$– collagen gene.

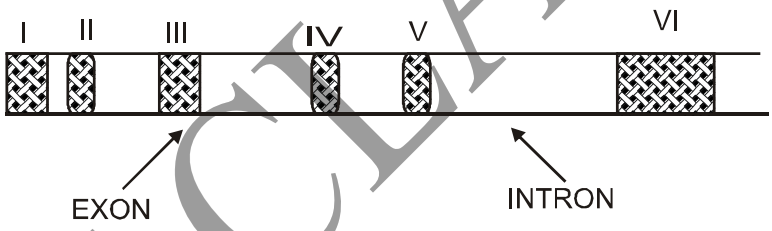



**Fig : 1** Diagram of an hypothetical discontinuous gene showing exons and introns.

**Gilbert hypothesis–** It says that during evolution, exons from different discontinuous genes can be shuffled forming new combinations of biological information. The shuffling may produce new functional proteins.

**Removal of Introns from Primary Transcript**

Transcription produces faithful copy of the coding strand of the gene. If the gene contains introns, its primary transcript will include copies of these. Therefore, interfere must be removed and the exon regions of the transcript be reattached to one another before translation. This process is called **splicing**.

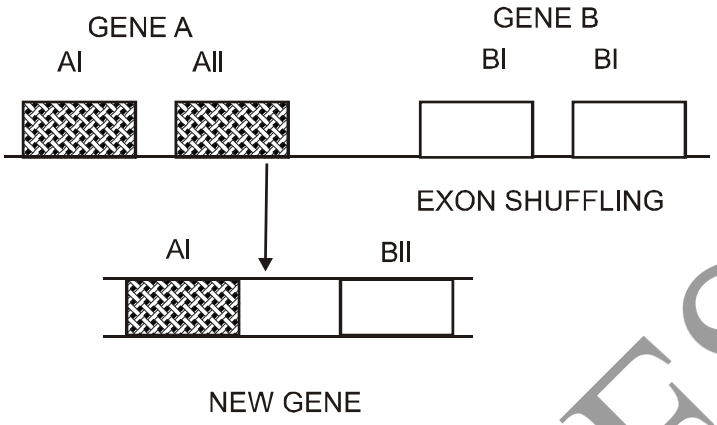



**Fig.: 2** Gilbert hypothesis about the origin of new genes.

**Gene**

- The primary function of the genome is to specify RNA molecules. Each region of the DNA helix that produces a functional RNA molecule constitutes a **gene**.
- A gene also includes all the DNA sequences required for synthesis of a particular RNA transcript.
- The number of genes in bacterial genomes is proportional to genome size.
- The bacterium with the smallest known genome, M genitalium has ~470 genes.
- The number of genes in a eukaryote varies from 6000 to 40,000 but does not correlate with the genome size or the complexity of the organism.
- Size of genes in higher eukaryotes varies greatly.
- Genes that are more than 100,000 nucleotides pairs in length are common.
- Portion of a gene in higher eukaryotes consists of noncoding DNA that interrupt the relatively short segments of coding DNA. The coding sequences are called exons; the intervening (noncoding) sequences are called **introns**.

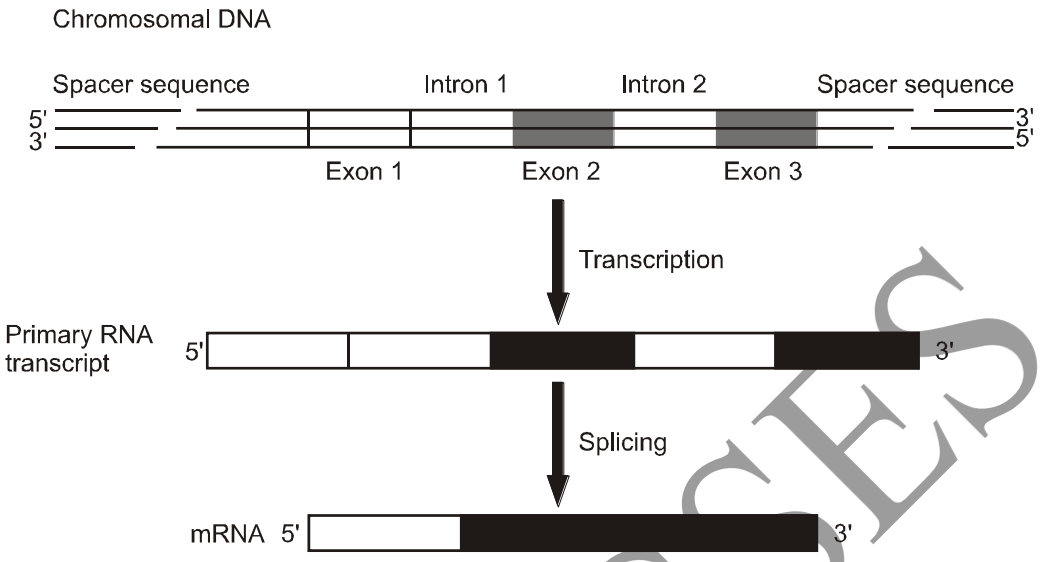



**Figure : 3 Structure of eukaryotes genes**
Most eukaryotic genes contain segments of coding sequences (exons) interrupted by noncoding sequences (introns). Both exons and introns are transcribed to yield a long primary RNA transcript. The introns are then removed by splicing to form the mature mRNA.

**Introns**

- Introns were first discovered in 1977, independently by **Phillip Sharp and Richard Roberts**.
- Introns are present in most genes of complex eukaryotes, although they are not universal.
- Many genes like interferons,histone genes, Ribonuclease genes (many), heat shock protein genes, G protein coupled receptors (many) lack introns.
- Introns are rare in genes of prokaryotes.
- Large genes consist of a long string of alternating exons and introns, with most of the gene consisting of introns.
- Introns range in size from about 50 nucleotides to 800,000 nucleotides. Exons are usually short, typically on the order of 150 nucleotides.
- All classes of genes may contain introns: nuclear genes coding for proteins, nucleolar genes coding for rRNA, and genes coding for tRNA. Introns also are found in mitochondrial genes and in chloroplast genes.

**Table: 2 Types of most common introns**

| Intron type | Where found |
|---|---|
| GU-AG introns | Eukaryotic nuclear pre-mRNA |
| AU-AC introns | Eukaryotic nuclear pre-mRNA |
| Group I | Eukaryotic nuclear pre-rRNA, organelle RNAs, some prokaryotic RNAs |
| Group II | Organelle RNAs, some prokaryotic RNAs |
| Pre-tRNA introns | Eukaryotic nuclear pre-tRNA |
| Archaeal introns | Various RNAs |

**Gene Families**

- In prokaryotes, closely related genes are often organized in operons & transcribed as polycistronic RNA. This doesn't occur in Eukaryotes.
- Many related eukaryotic genes can be functionally grouped as a set of genes called a **Gene Family**.

**1. Simple/classical Multigene Family**

- Families in which all the members have identical or nearly identical sequences belongs to simple multigene families.
- The rRNA genes are examples of simple multigene families.
- E-coli genome contains seven copies of its rRNA genes. Human cells contain about 200 rRNA gene copies per haploid genome, clustered on one chromosome.

**2. Complex Multigene Family**

- One of the best examples of this type of multigene family is the mammalian globin genes,
- The globins are the blood proteins that combine to make hemoglobin, each molecule of hemoglobin being made up of two $\alpha$-type and two $\beta$-type globins.

- The $\alpha$ (alpha)-subfamily in human contains three genes: zeta ($\zeta$) gene (expressed only in the early embryonic stage) and two copies of the alpha genes (expressed during the fetal and adult stages).
- In human $\beta$(beta)-globin gene cluster is longer than the $\alpha$-globin gene cluster and contains five genes.

**Pseudogenes**

- Pseudogenes are functionless gene variants that are present as a result of ancient historical accident.
- Pseudogenes are a type of evolutionary relic that indicates about the changing nature of genome. There are two main types of Pseudogene:

1. **Conventional pseudogene**

   A gene that has become inactive (non-functional) because of the accumulation of mutations. The globin pseudogenes are most common example of conventional pseudogenes. The human globin gene clusters contain five pseudogenes.

2. **Processed pseudogene**

   A gene that results from integration into the genome of a reverse-transcribed copy of an mRNA. Processed pseudogenes might integrate into the same chromosome or into a different chromosome. It does not contain any introns that were present in its parent gene. It also lacks promoter sequences present in the upstream of parent gene.

**3. Chromosomes and Chromatin**

- The complexes between eukaryotic DNA and proteins are called **chromatin**, which typically contains about twice as much protein as DNA. Major proteins of chromatin are the histones.
- Histones-small proteins containing a high proportion of basic amino acids (arginine and lysine) that facilitate binding to the negatively charged DNA molecule.
- There are five major types of histones-called **H1, H2A, H2B, H3, and H4**- which are very similar among different species of eukaryotes.
- In addition, chromatin contains an approximately equal mass of a wide variety of nonhistone chromosomal proteins.

- The basic structural unit of chromatin, the nucleosome, was described by **Roger Kornberg in 1974**.
- Nuclease digestion and the electron microscopic studies suggested that chromatin is composed of repeating 200-base-pair units, which were called nucleosomes.

**Table: 3 The Major Histone Proteins**

| Histone[a] | Molecular weight | Number of amino acids | Percentage lysine +arginine |
|---|---|---|---|
| H1 | 22,500 | 244 | 30.8 |
| H2A | 13,960 | 129 | 20.2 |
| H2B | 13,774 | 125 | 22.4 |
| H3 | 15,273 | 135 | 22.9 |
| H4 | 11,236 | 102 | 24.5 |

- More extensive digestion of chromatin with micrococcal nuclease was found to yield particles (called **nucleosome- core particles**) that correspond to the beads.
- Detailed analysis of these particles has shown that they contain 147 base pairs of DNA wrapped 1.67 times around a histone core consisting of two molecules each of H2A, H2B, H3, and H4 (the **core histones**).
- One molecule of the fifth histone, H1, is bound to the DNA as it enters each nucleosome core particle.
- This forms a chromatin subunit known as a **chromatosome**, which consists of 166 base pairs of DNA wrapped around the histone core and held in place by H1 (a linker histone).
- The packaging of DNA with histones yields a chromatin fiber approximately 10 nm in diameter that is composed of chromatosomes separated by linker DNA segments averaging about 50 base pairs in length.

**Euchromatin**

- It is transcriptionally active & more decondensed so that transcription machinery can work on it.

- In interphase nucleic, euchromatin appears in 30 nm form or more condensed form of 60-130 nm chromatin fibers.

**Heterochromatin**

- It is transcriptionally inactive & consist of highly repetitive DNA sequences, such as those present in centromere or telomeres.
- It is a highly condensed structure.
- As cell enter mitosis, their chromosomes become highly condensed. Thus, transcription ceases during mitosis.

**1.** One of the post replication modification in DNA is methylation of bases,in this context which of the following statements is incorrect

(A) Methylation protect cell's own DNA from digestion by its own restriction endonuclease

(B) Methylation result in localized conversion of B DNA to Z DNA.

(C) Methyl group are added predominantly to cytosine in bacterial cell and to adenine eukaryotes

(D) Methylation of bases is important in some aspects of DNA repair in bacteria.

**1(B)** Methylation is the process where we observe the phenomenon of addition of methyl group to adenine or cytosine base pair. It is done to discriminate between our own body DNA, so as to protect it from the restriction digestion that is to be carried out by our own restriction enzymes. Methylation process also plays a predominant role is host restriction mechanisms & DNA repair system. So the statement which is incorrect is that the methylation plays an important role in a localized conversion of B - DNA to Z - DNA as Z DNA may be a transient configuration of B - DNA.

**2.** The product of gene X is a cytosolic protein. A repressor of gene X expression is over expressed in a cell line using recombinant DNA technology. After the over expression, the cell line is observed to be susceptible to UV lights and higher mortality rates are observed upon even short exposures to the UV. What could be the function of the gene X related to?

(A) Photosynthesis

(B) DNA repair

(C) DNA replication

(D) X-chromosome assembly

**2.(B)** When a repressor stops the gene X function cells become susceptible UV lights and higher mortality rate. This proves that gene X must be playing part in DNA repair as UV induced pyrimidine dimers are corrected by DNA repair genes.