RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF ENGINEERING AND TECHNOLOGY

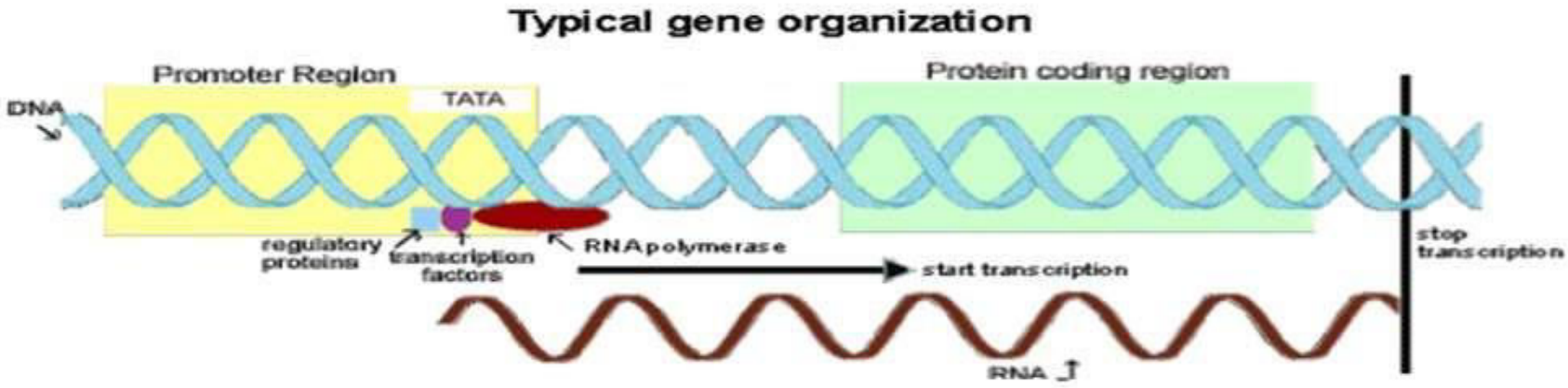
Department of Biotechnology

Gene prediction in prokaryotes

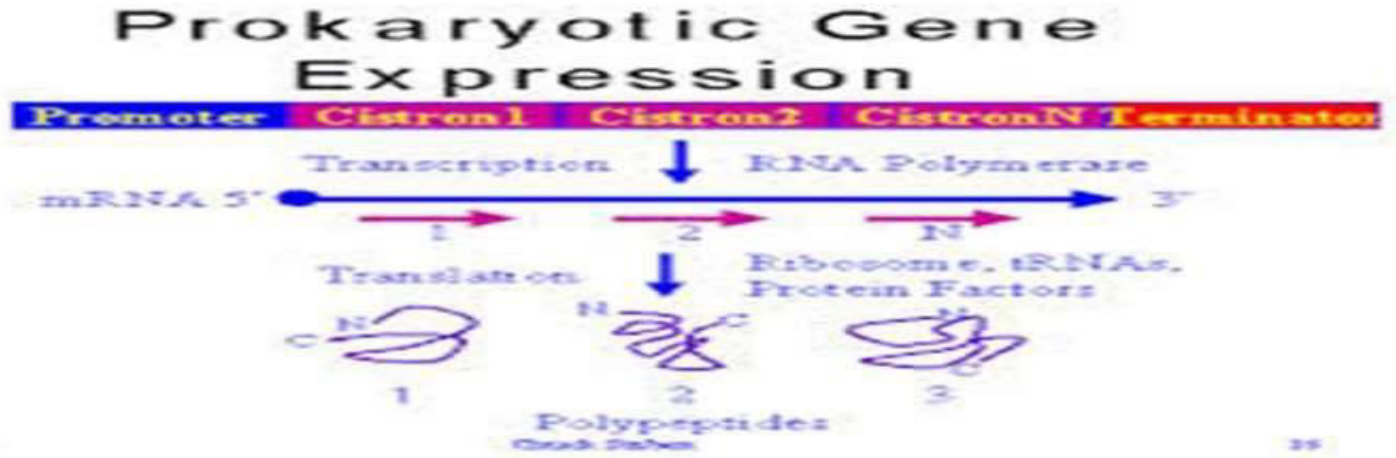
Genes are subunits of DNA, the information database of a cell that is contained inside the cell nucleus.

This DNA carries the genetic blueprint that is used to make all the proteins the cell needs.

Every gene contains a particular set of instructions that code for a specific protein



- In prokaryotes the primary control point is the process of transcription initiation .
- Different ways for regulation of gene expression in bacteria:
  - 1- Promoter recognition.
  - 2-Transcription elongation(Attenuation).
- Regulation of gene expression can be done by some operon pathways such as
  1.lac operon.
  2.tryptophan operon.



Eukaryotes vs Prokaryotes

✓Eukaryotic DNA wrapped around histones that might result in repeated patterns (periodicity of 10) for histone binding. The promotor regions might be near these sites so that they remain hidden.

✓Prokaryotes have no introns.

✓Promotor regions and start sites more highly conserved in Prokaryotes

✓Different codon use frequencies

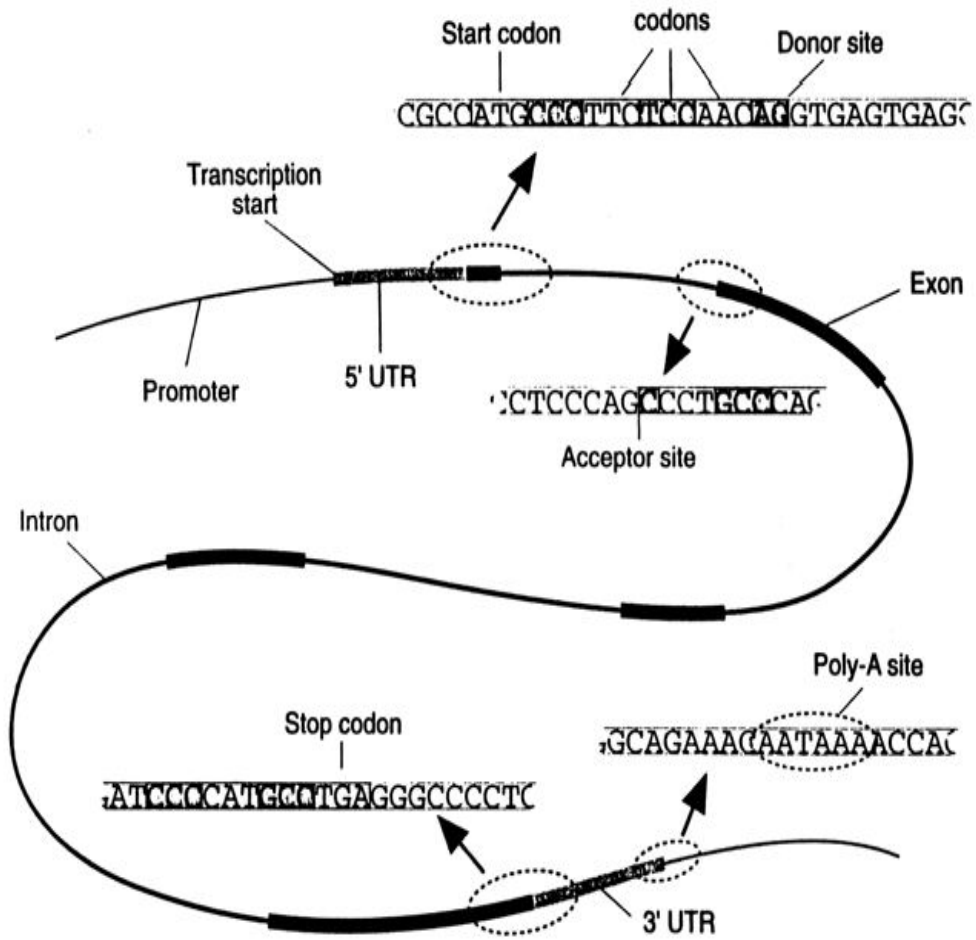


Fig. 8. The structure of a gene with some of the important signals shown.

Gene finding is species-specific

- Codon usage patterns vary by species
- Functional regions (promoters, splice sites, translation initiation sites, termination signals) vary by species
- Common repeat sequences are species- specific
- Gene finding programs rely on this information to identify coding regions

Pattern-based gene finding

- ORF finding based on start and stop codon frequency is a pattern-based procedure
- Other pattern-based procedures recognize characteristic sequences associated with known features and genes, such as ribosome binding sites, promoter sites, histone binding sites, etc.
- Statistically based.

Content-based gene finding

- Content-based gene finding methods rely on statistical information derived from known sequences to predict unknown genes
- Some evaluative measures include: "coding potential" (based on codon bias), periodicity in the sequence, sequence homogeneity, etc.

A standard content-based alignment procedure

- Select a window of DNA sequence from the unknown. The window is usually around 100 base pairs long
- Evaluate the window's potential as a gene, based on a variety of factors
- Move the window over by one base
- Repeat procedure until end of sequence is reached; report continuous high-scoring regions as putative genes

XYZ

abc

XYZ

abc

XYZ

abc

XYZ

abc

MCQs

1. A
2. A
3. A
4. A
5. A
6. A
7. A
8. A
9. A
10. A