

Basic Principles of Molecular Biology

by

DR. ERNELEA P. CAO

Institute of Biology and Natural Sciences Research Institute,
College of Science, University of the Philippines
Diliman, Quezon City

The following contains notes on the basic concepts involved in molecular biology, particularly on the nature of the genetic material (DNA), processes where it is involved, their controls and the overall significance of all of these in terms of the Central Dogma of Molecular Biology. In addition, using information derived from the Central Dogma, an overview of the concepts involved in gene manipulation will be presented. The notes follow an outline type of format. Key words and phrases are highlighted for easy reference and recall.

DNA: The genetic material

DNA Structure

- Double helix
- Made up of anti-parallel strands (one oriented 5' - 3', the other 3' - 5')
- Made up of polynucleotides
- **Nucleotide** = phosphate-(deoxyribose)sugar-nitrogen base
- Nucleoside = phosphate-sugar
- Nitrogen base = Adenine, Guanine (purines), Thymine, Cytosine (pyrimidines)
- Nitrogen base pairing occurs = between A and T; between C and G

Features that make the DNA a good genetic material

- Chemically **stable** (chemical forces stabilize the molecule and prevent vulnerability to denaturation or damage/changes)
- **Replicable** (built-in mechanism that allows faithful copying of the genetic information for transmission to the next generation of cells or organisms)
- **Translatable** (genetic information can be expressed in the form of proteins that represent the phenotype at the organismal level)
- **Mutable** (at the same time that the molecule is stable and that replication allows for faithful copying of the genetic information, spontaneous and induced mutations/changes are possible to allow for the formation of new forms or variations for increased adaptation and survival of the organism)

Chemical forces in the DNA molecule

- **Hydrogen bonds** between base pairs (2 H-bonds between A and T; 3 H-bonds between C and G)
- **Pi-pi electron interactions**/stacking interactions between base pairs
- **Covalent bonds** between phosphate and sugar (deoxyribose)
- **Ionic bonds** between DNA and histones

Central Dogma of Molecular Biology



Replication

- By **complementary base pairing** (each strand serves as a template for the synthesis of the new strand)
- **Bidirectional** (occurs in a 5' - 3' direction due to anti-parallel nature of the strands)
- Utilizes DNA polymerase III to add the new nucleotides (one strand is synthesized continuously, the other discontinuously producing the **Okazaki fragments**)
- **Semi-conservative** (daughter DNA molecules = 1 old and 1 new strand)
- Steps
 - **Initiation** (nicking and unwinding of the double helix, priming event - utilizes **primers**)
 - **Elongation** (addition of new nucleotides by DNA polymerase III, editing of wrong bases by DNA polymerase I -copy editing)
 - **Termination** (sealing of gaps by DNA ligase)

Transcription

- Selectively utilizes a portion of one DNA strand as template for the synthesis of RNAs
- Temporary unwinding and opening of DNA template (utilizes complementary base pairing phenomenon = A pairs with uracil (U) this time; C with G)
- Utilizes **RNA polymerase** for addition of new nucleotides (5'-3' direction)
- RNAs produced = **mRNA, tRNA, rRNA**

- Steps
 - **Initiation** (**promoter** binding, DNA unwinding)
 - **Elongation** (RNA polymerase moves along DNA template and adds complementary nucleotides)
 - **Termination** (may or may not be **rho**-dependent)

Translation (protein synthesis)

- Polypeptide chain is formed by the addition of amino acids based on the triplet sequence of bases in the mRNA molecule (produced by transcription)
- **Genetic code** = nucleotide sequence in nucleic acids specifies the amino acid sequence in proteins (**triplet code** = codons or three nucleotides are **non-overlapping** and **comma-less**, **degenerate** or has redundancy)
- Steps
 - **Initiation** (assembly of ribosome on mRNA)
 - **Elongation** (repeated cycles of amino acid delivery, peptide bond formation, and movement along the mRNA or translocation)
 - **Termination** (release of the polypeptide chain)

Control of Gene Expression

Kinds of genes:

1. **Structural** gene (codes for a protein)
2. **Regulatory** gene (regulates expression of structural gene)

In prokaryotes

- **Operon** concept (group of structural genes which are under the control of the same regulatory locus)
- **Polycistronic**
- *Lac* operon (an example of negative control by induction)
- *Trp* operon (an example of negative control by repression)

In eukaryotes

- **Monocistronic**
- Utilizes RNA (called **activator RNA**) as regulator of gene expression instead of repressor protein (as in prokaryotes)

Controls

- Post-transcriptional processing of mRNA or primary transcript
 - **Capping or methylation** of 5'-end
 - **Polyadenylation** or addition of poly-A tail at 3'-end
 - **Splicing**
 - **RNA editing**
- Post-translational processing
 - **Cleavage** (to remove signal peptides, to release mature fragments, to remove internal peptides, trimming of both N- and C-termini)
 - **Chemical modifications** (phosphorylation, ribosylation, glycosylation, etc.)

Gene manipulation

DNA/Gene cloning

- Facilitates the isolation and manipulation of fragments of an organism's genome by replicating them independently as part of an autonomous vector
- Involves the isolation of the gene(s) responsible for the expression of a protein or the formation of a product
- Gene or DNA fragment is placed in an autonomously replicating piece of DNA called a **vector** forming a **recombinant DNA**
- Propagation of the host organism containing the recombinant DNA forms a set of genetically identical organisms or a **clone**, hence the process is known as DNA/gene cloning

Applications of DNA cloning (often collectively known as genetic engineering)

- DNA sequencing (or the derivation of protein sequence)
- Isolation and analysis of gene promoters and other control sequences
- Investigation of protein/enzyme/ RNA function by large-scale production of normal and altered forms
- Identification of mutations, for example, gene defects leading to disease
- Biotechnology, the large-scale commercial production of proteins and other molecules of biological importance, for example, human insulin and growth hormone
- Engineering of plants and animals, and gene therapy
- Engineering proteins to alter their properties

Hosts and vectors

- Must be normally capable of being replicated and isolated independently of the host's genome (although some are designed to incorporate DNA into the host genome for longer term expression of cloned genes)
- Also incorporate a **selectable marker**, a gene which allows host cells containing the vector to be **selected** from amongst those which do not, usually by conferring resistance to a toxin, or enabling their survival under certain growth conditions
- Utilizes **plasmids, bacteriophages, cosmids, BACs, YACs**

Subcloning

- Refers to the transfer of a fragment of cloned DNA from one vector to another
- May be used to investigate a short region of a large cloned fragment in more detail, or to transfer a gene to a vector designed to express it in a particular species
- Steps
 - **Isolation** of plasmid DNA
 - **Digestion** (cutting) of the plasmid into discrete fragments with **restriction endonucleases**
 - **Separation** of the fragments by **agarose gel electrophoresis**
 - **Purification** of the desired **target** fragment
 - **Ligation** (joining) of the fragment into a new plasmid vector, to form a new recombinant molecule
 - Transfer of the ligated plasmid into an *E. coli* strain (**transformation**)
 - **Selection** of transformants
 - **Analysis** of recombinant plasmids

DNA libraries

- Sets of DNA clones, each of which has been derived from the insertion of a different fragment into a vector followed by propagation in the host
- **Genomic libraries** (prepared from random fragments of genomic DNA; but may be inefficient in finding a gene, particularly in large eukaryotic genomes where much of the DNA is non-coding)
- **cDNA libraries** (source is the mRNA which is copied to produce cDNA by **reverse transcription**; yield only the coding region)

Screening libraries

- Method for screening which clone in a library contains the gene of interest
- Based on the use of radioactively or fluorescently-labeled **DNA probe** (complementary or partially complementary to a region of the gene sequence and which can be detected by **hybridization**)
- **Probe** may be **oligonucleotide** sequence derived from the sequence of the protein product of the gene, or from a related gene from another species
- **Polymerase chain reaction (PCR)** is an important method for generation of probes
- Other screening methods involve the expression of the coding sequences of the clones in the library and identification of the protein product from its activity, or with a specific antibody

Analysis of a clone

- Structure of the cloned DNA fragment may be investigated further using **restriction mapping** or by **sequencing** of the entire fragment
- Sequence can be compared with other known sequences from databases and the complete sequence of the protein product determined
- The sequence is then available for manipulation in any of the applications of cloning described