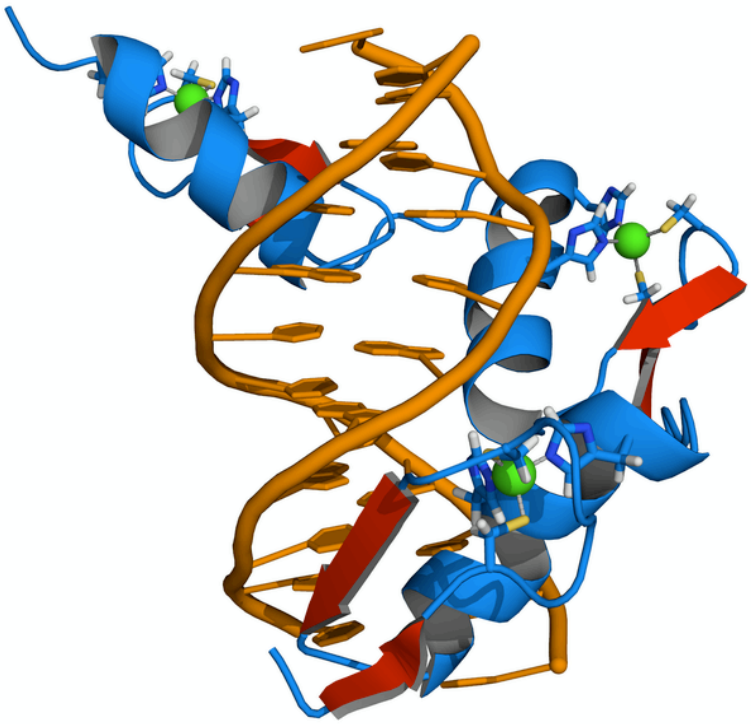
Imperial College London

Faculty of Medicine

Cellular and Molecular Science

Graduate Entry Programme

Year 1 - 2011/12



Course Leader: Dr Mick Jones

Tel: 020 8383 1643
Email: m.d.jones@imperial.ac.uk

https://education.med.imperial.ac.uk

Basic Molecular Biology

Proteins and Nucleic Acids

Structure & Function

The following notes should be read in conjunction with the 3 PowerPoint presentations by Dr Mick Jones; CMSLecture1.MDJ, CMSLecture2.MDJ & CMSLecture3.MDJ.

The material covered in these presentations should be material that you will have covered in your undergraduate BSc degree.

You should go through the material in your own time.

If you have any queries, then please contact Dr Mick Jones, who will answer any queries, and if required will schedule group tutorial sessions to cover any points.

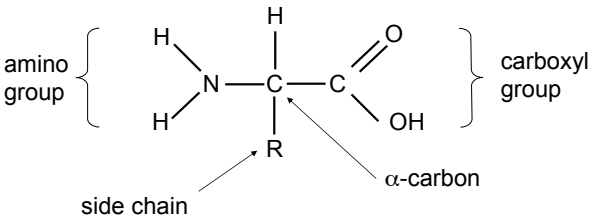
CMSLecture1.MDJ

Protein Structure

Learning Objectives:

- Outline the reaction by which amino acids are joined together.
- Appreciate the different types of bond that combine to stabilise a particular protein conformation.
- Sketch a trimeric peptide, illustrating the amino -terminus, carboxyl terminus and side chains.
- Distinguish between a α-helix and a ß-pleated sheet and appreciate the bonds that stabilise their formation.
- Understand the concepts of primary structure, secondary structure, tertiary structure & quaternary structure with respect to proteins.
- Outline how warfarin works with reference to the post translational modification of glutamate.

Anatomy of an Amino Acid

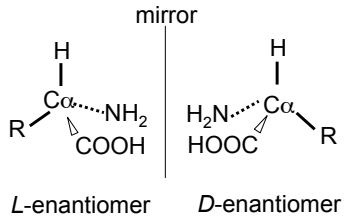


Substitutions at the R position or side chain, give rise to the 20 different amino acids e.g. R=CH_3 in alanine.

The whole of the amino acid minus the side chains is known as the backbone.

Chirality

The central Cα carbon atom is a chiral centre (from the Greek, meaning "handed") i.e. it has four different substituents bound to it.

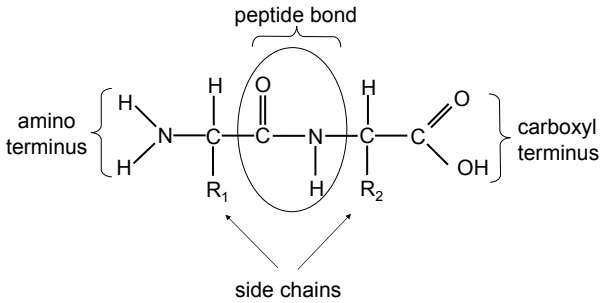


This gives rise to optical isomers (*enantiomers*) of each amino acids each of which is a mirror image of the other.

Glycine (Gly) has no side chain (only an H atom) and is therefore the only non-chiral amino acid.

Individual amino acids (R1 and R2) are joined in condensation reactions (i.e. water is lost) to form peptide chains.

Anatomy of a Peptide



The polypeptide chain of a protein rarely forms a disordered structure (random coil) as proteins generally have functions to fulfil, and these functions rely upon specificity. In turn, functionality requires a definite 3D structure or ***conformation*** of the polypeptide chain.

Proteins generally possess a degree of flexibility necessary for function e.g. muscle fibres

Proteins are held together by:

Covalent bonds (in which two atoms share electrons) are the strongest bonds within protein and exist in the primary structure itself. Covalent bonds can also exist as *disulphide bridges*. These occur when cysteine side chains within a protein are oxidised resulting in a covalent link between the two amino acids.

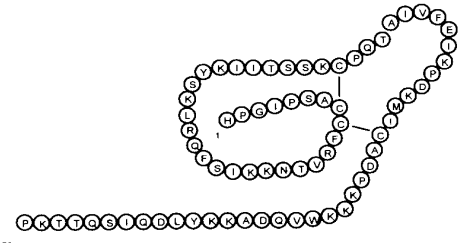
Hydrogen Bonds occur when two atoms bearing partial negative charges share a partially positively charged hydrogen, the atoms are engaged in a hydrogen bond (H-bond).

Ionic interactions arise from the electrostatic attraction between charged side chains e.g. Glu, Asp, Lys and Arg. They are relatively strong bonds, particularly when the ion pairs are within the protein interior and excluded from water.

Van der Waals Forces are transient, weak electrostatic attractions between two atoms, due to the fluctuating electron cloud surrounding each atom which has a temporary electric dipole. Although relatively weak and transient in nature, because of the sheer number of these interactions within a protein, they can still have a large part to say in the overall conformation of a protein.

Hydrophobic Interactions are a major force driving the folding of proteins into their correct conformation. They juxtapose hydrophobic side chains by packing them into the interior of the protein. This creates a hydrophobic core and a hydrophilic surface to the majority of proteins

Primary structure is the linear sequence of amino acids that make up the protein.

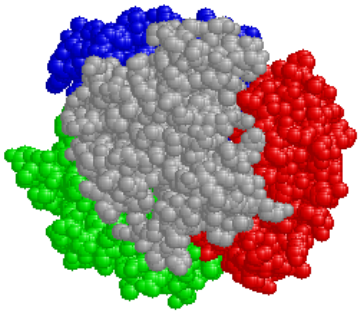


Secondary structure is defined as local structural motifs within a protein, e.g. α-helices and β-pleated sheets.

Tertiary structure is the arrangement of the secondary structure motifs into compact domains.

Quaternary structure is the three dimensional structure of a multimeric protein composed of several subunits

Tertiary structure



Quaternary structure

Even after synthesis, **(post translation)** the starting set of 20 amino acids can be modified to create novel amino acids, enhancing the capabilities of the protein e.g. hydroxylation, glycosylation, carboxylation.

Nucleic Acids and Chromosomes

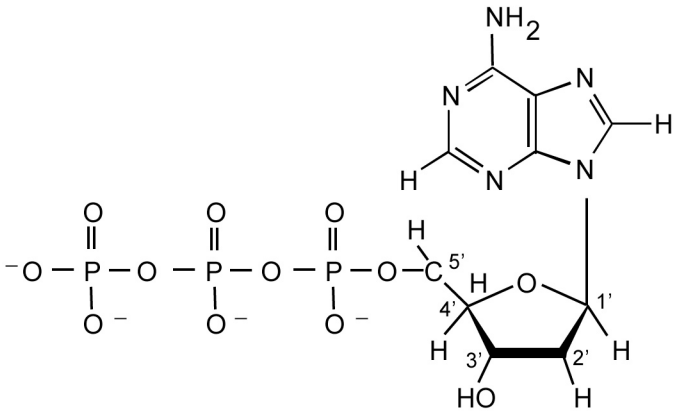
Learning Objectives:

- Draw the structure of a nucleotide labelling the sugar, base and phosphates and explain the difference between a nucleotide and a nucleoside.
- List the bases found in DNA and RNA and indicate which ones are purines and which ones are pyrimidines.
- Describe a single DNA chain and explain the difference between the 5' and 3' ends.
- Draw the structure of the double-stranded helix of DNA (not atomic structure) showing base-pairing, the major and minor grooves, and the directionality of the chains.
- Describe melting and re-annealing of complementary strands and what is meant by Watson-Crick base-pairing.
- Compare the genomes of *E.coli* and *Homo Sapiens.*
- Draw a diagram illustrating the packaging of DNA into nucleosomes and relate this to chromosome structure.
- Describe the human karyotype.

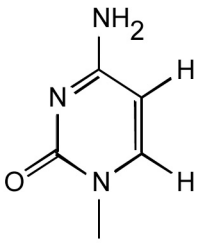
Nucleic Acids

- DNA and RNA are nucleic acids.
- They are macromolecules made up of a large number of nucleotides.
- A nucleotide is composed of a base, a sugar, and a phosphate group.
- A nucleoside is composed of a base and a sugar (no phosphate).
- The sugar in DNA is deoxyribose, the sugar in RNA is ribose.

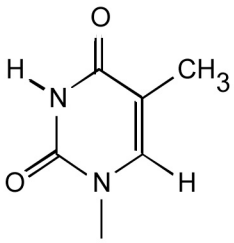
Nucleotide: deoxyadenosine 5'-triphosphate (dATP)



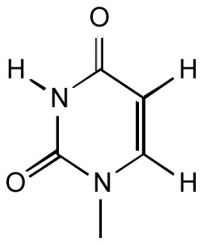
DNA and RNA bases:
Pyrimidines:



Cytosine
(C)

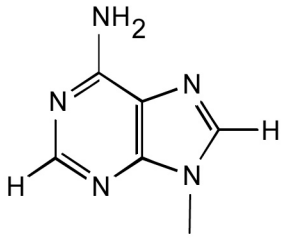


Thymine
(T)
DNA only

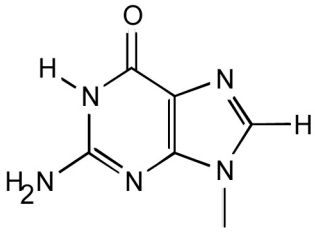


Uracil
(U)
RNA only

Purines:



Adenine
(A)

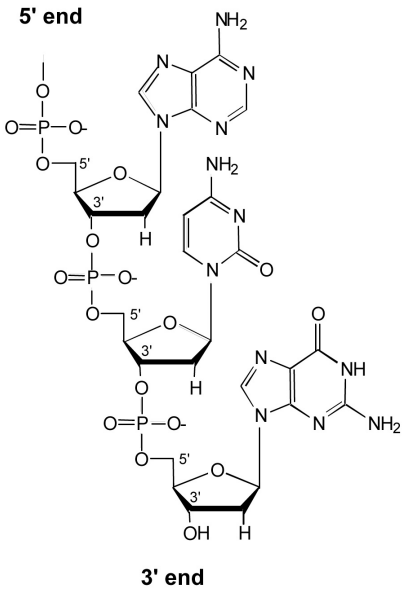


Guanine
(G)

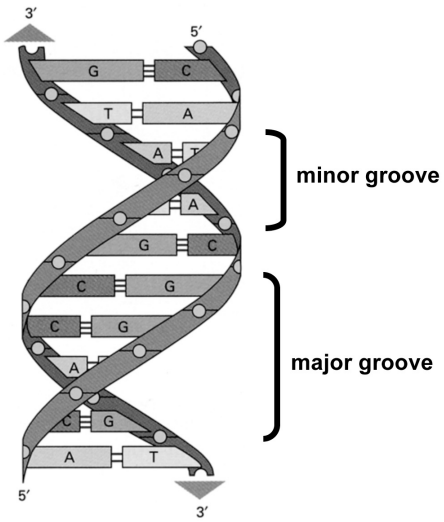
The nucleosides: (deoxy)cytidine, (deoxy)thymidine, (deoxy)uridine, (deoxy)adenosine, (deoxy)guanosine

The nucleotides: deoxyadenosine 5'-triphosphate (dATP); adenosine monophosphate (AMP) etc.

A single DNA chain

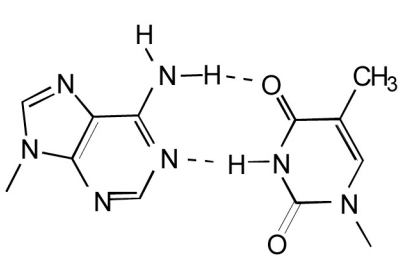


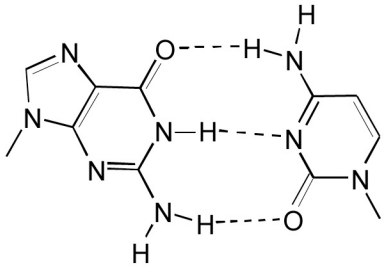
The DNA double helix



- **DNA** is a long chain of deoxyribose units linked by phosphodiester links.
- The phosphate on the 5' carbon is linked to the –OH on the 3' carbon along the chain.
- On each deoxyribose there is a base.
- The chain has two ends. The 5' end and the 3' end. It is not symmetrical.
- The primary sequence is the linear sequence of the bases. By convention, the nucleotide sequence is specified in the 5' to 3' direction.
- **The secondary structure** of DNA is a right-handed double helix. The two chains in the helix run in opposite directions.
- The deoxyribose and phosphate groups run along the outside of the helix, with the negative charges outside.
- The bases point inwards and the flat planes are perpendicular to the helix.
- The two chains are held together by hydrogen bonds between the bases.
- The two strands are complementary in their sequence due to the specificity of base-pairing. Adenine always pairs with Thymine; Guanine always pairs with Cytosine.

Watson-Crick base pairs





Adenine (A) Thymine (T)

Guanine (G) Cytosine (C)

2 hydrogen bonds → less stable

3 hydrogen bonds → more stable

Melting and re-annealing

- High temperature and/or low salt concentration causes the two strands to melt or disassociate.
- If you then lower the temperature or increase the salt concentration, the two melted strands will re-anneal into a double helix.
- Hybridisation: in a mixture of DNA with different sequences, the complementary strands will find each other in the mixture.

The *E.coli* genome:

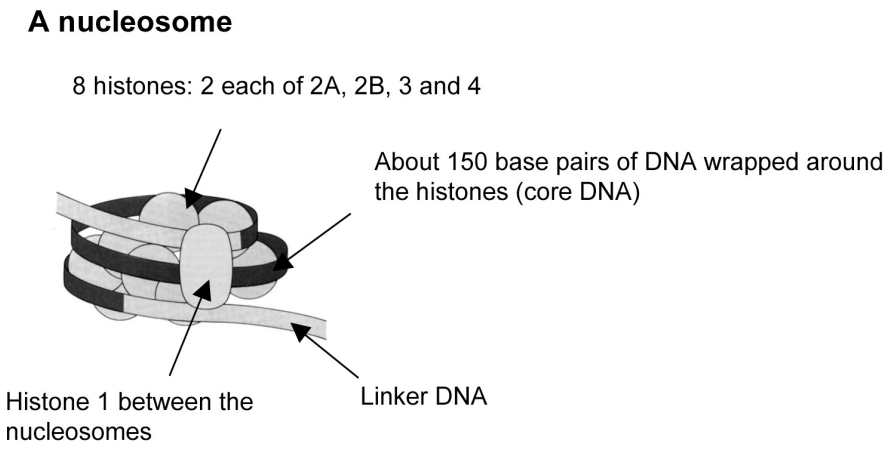
- *E.coli* has 4.7×10^6 base pairs in a single circular double-stranded molecule.
- The length of the *E.coli* DNA is 1.4 mm.
- The DNA in *E.coli* is tightly packaged – the bacterium is only 3 μm long.

The human genome:

- The human genome (haploid) consists of about 3 x 10^9 base pairs of DNA.
- The DNA is divided into chromosomes that each contain a linear double-helical DNA molecule of about 200 x 10^6 base pairs.
- Prior to cell division, the DNA condenses into discrete chromosomes, visible by microscopy.
- A diploid cell has 46 chromosomes; 22 pairs of 'normal' chromosomes and 2 sex chromosomes.

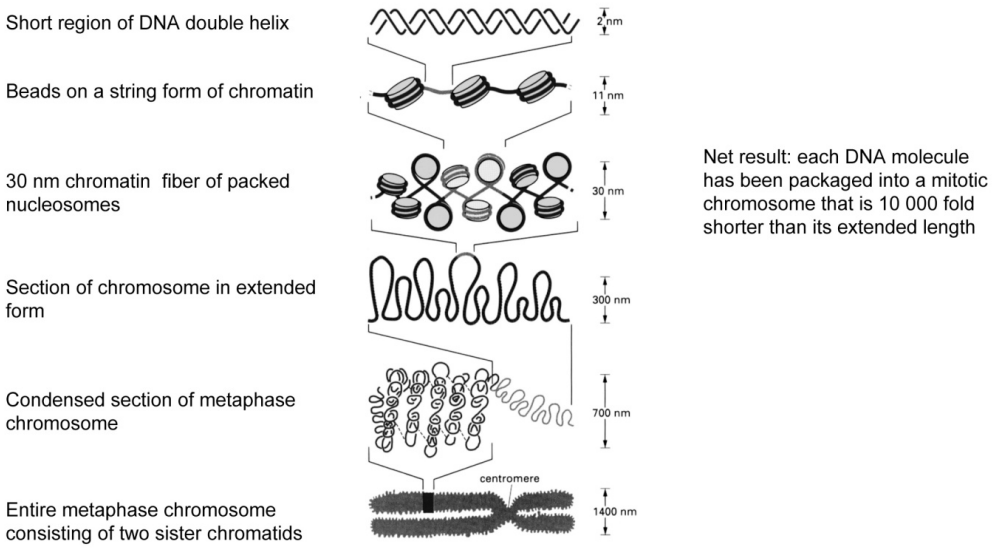
Packaging of eukaryotic DNA:

- The DNA in a diploid human cell is nearly 2 m long. To fit into cells, the DNA is tightly packaged into chromatin.
- Chromatin consists of DNA and proteins.
- The lowest level of packaging is the nucleosome, which consists of DNA wrapped around histone proteins.
- The nucleosomes form a chain, which pack into a helical array.



Packaging of DNA into a chromosome

As the DNA has already replicated, there are two identical copies and two identical chromatids (sister chromatids) for each chromosome at metaphase.



Nucleic Acids and Gene Expression

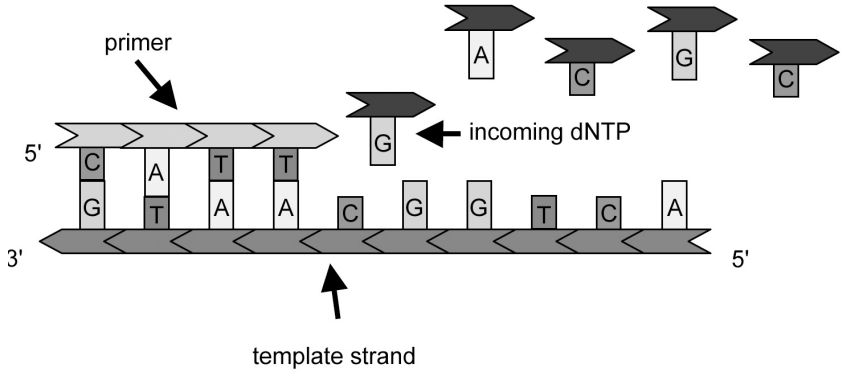
DNA replication, the cell cycle and mitosis

Learning Objectives:

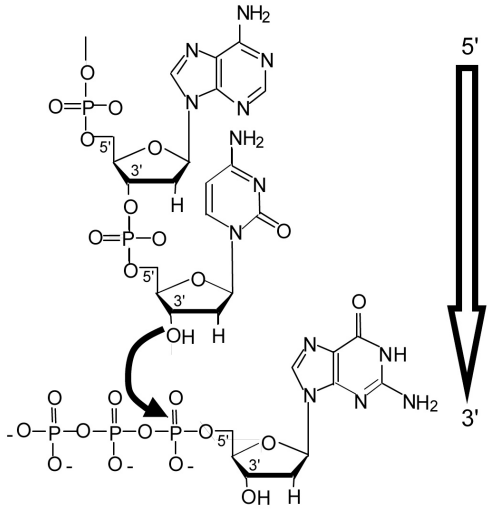
- Explain semi-conservative replication.
- Describe the reaction catalysed by DNA polymerases.
- Describe how nucleoside analogs can be used as drugs.
- Describe the functions of the components of the replication complex including the terms template, primer, leading strand, lagging strand, Okasaki fragment and replication fork.
- Describe how accuracy is maintained by proof-reading and the use of RNA primers.
- Draw a diagram showing replication of the *E.coli* chromosome.
- Describe the replication of mammalian chromosomes.
- Describe the different phases of the cell cycle.
- Draw a diagram showing how the chromosomes segregate at metaphase.

DNA replication

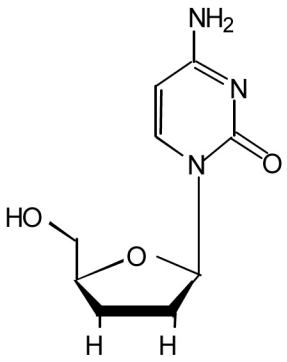
- DNA replication is semi-conservative. Each strand forms the template for a new strand of DNA.
- DNA polymerases add deoxynucleotide tri-phosphates to the 3' end of a DNA molecule.
- Synthesis is driven by the release of energy from the hydrolysis of the tri-phosphate.
- DNA polymerases need a template and a primer. They cannot start a new chain from scratch.



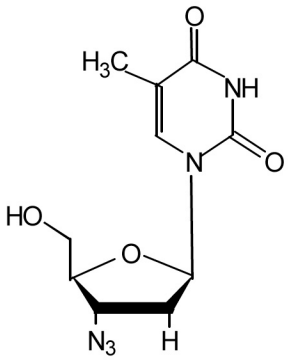
Enzyme reaction



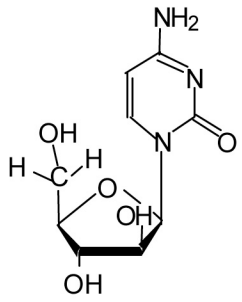
Some nucleoside analogs used as drugs



dideoxycytosine (ddC)
Drug for HIV (zalcitabine)

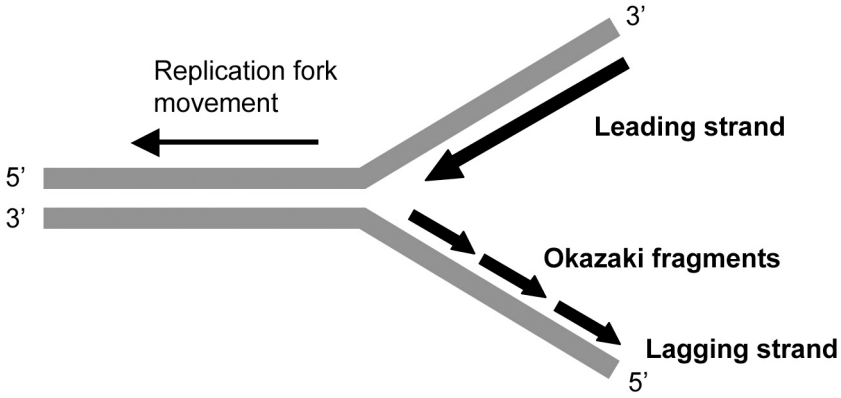


azidothymidine (AZT)
Drug for HIV (Zidovudine)



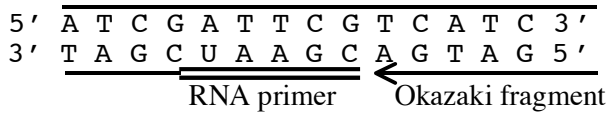
Cytosine arabinose
Used in Chemotherapy

DNA Replication



- Replication begins at discrete points on the DNA molecule and proceeds bi-directionally. The structure formed is called a replication fork.
- The replication fork is asymmetric. Both strands are synthesised in 5' to 3' direction.
- DNA synthesis is continuous on the leading strand
- On the lagging strand, synthesis is discontinuous.
- As the replication fork moves forward, single-stranded DNA is exposed on the lagging strand.
- When about 100-200 bases are exposed, a short RNA primer of about 5-10 bases is synthesised.
- The primer is synthesised by an RNA polymerase called primase.
- The primer is extended by a DNA Polymerase until the last RNA primer is reached.
- The stretches of DNA synthesised on the lagging strand are called Okazaki fragments.

Joining the Okazaki fragments
A special ribonuclease removes the RNA primer using a 5' to 3' exonuclease activity.

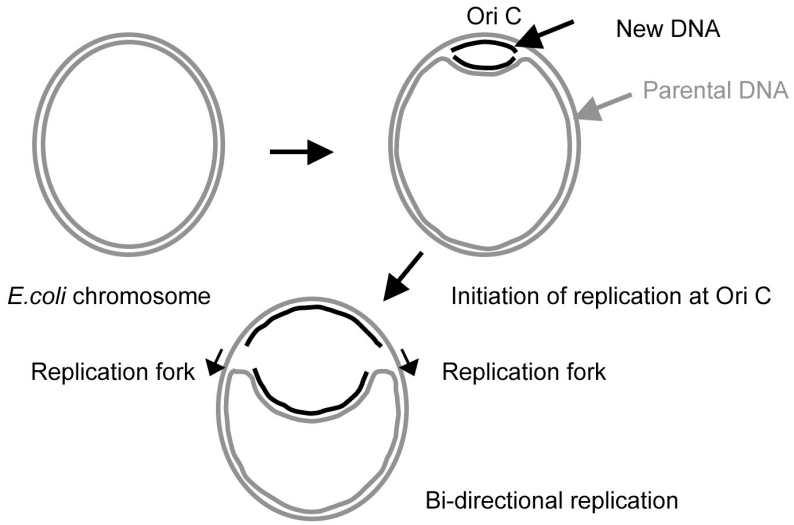


DNA polymerase then synthesises DNA through the RNA primer region.

DNA ligase joins the two adjacent strands of DNA together using ATP.

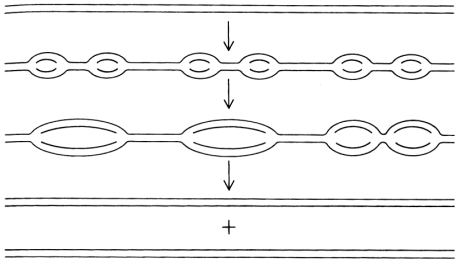
Accuracy of DNA replication
- DNA replication has an error frequency of about 1 change per 10^9 base pairs.
- Proof reading: DNA Polymerase has 3' to 5' exonuclease activities for proof reading.
- RNA Primers: Inaccurate RNA primers are replaced by accurate DNA.

Replication of the *E. coli* chromosome
In *E. coli* there is a single replication origin and the two forks meet on the other side of the circular chromosome.



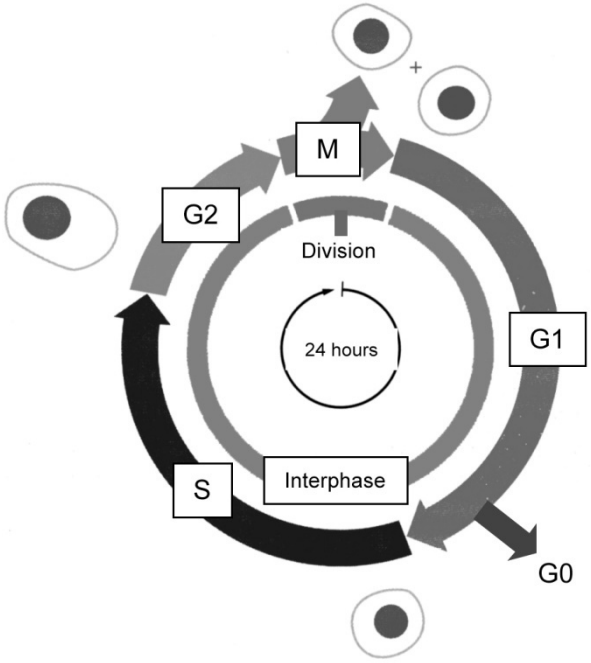
Replication of the eukaryotic genome

The enzymatic mechanism of replication in eukaryotes is very similar to that in *E. coli*. There are replication forks, leading and lagging strands and enzymes which carry out all the same functions as in *E. coli*, but the actual proteins used are different.



Multiple replication origins in eukaryotes

- Eukaryotic chromosomes are linear and very long.
- Multiple replication origins are distributed at intervals of about 100 kb.
- Each replication origin gives bidirectional replication forks.
- Replication is finished when all the forks have met.

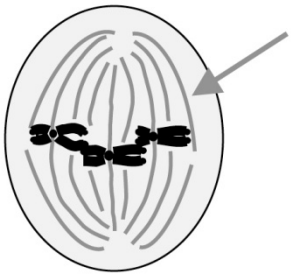


The mammalian cell cycle

A mammalian cell cycle takes about 24 hours in tissue culture.
M phase: Mitosis 1 hr; cell division, metaphase.
G1; gap phase 1; 10 hrs
S phase; DNA replication, 9 hrs
G2; gap phase 2; 4 hrs
G0; cells which have stopped dividing.
G1, G0, S and G2 are interphase.

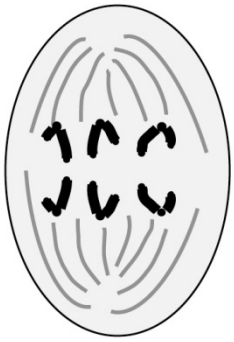
Metaphase and Anaphase of Mitosis

Metaphase



Condensed chromosomes, aligned on central plane of spindle

Anaphase



Sister chromatids move to opposite poles of spindle

CMSLecture2.MDJ

Nucleic Acids and Gene Expression

Gene Organisation & Transcription Part 1 - Gene Transcription

Learning Objectives:

- Describe the basic differences between DNA and RNA
- Describe what is meant by "transcription"
- List the major functional classes of RNA and the classes of RNA polymerases involved in synthesising each of these.
- Describe what is meant by a "gene promoter"
- Describe what is meant by a "transcription factor"
- Describe, with the aid of diagrams, the processes involved in transcribing a eukaryotic gene.

What Do We Mean By "Gene Expression"?

DNA → RNA → (Protein)

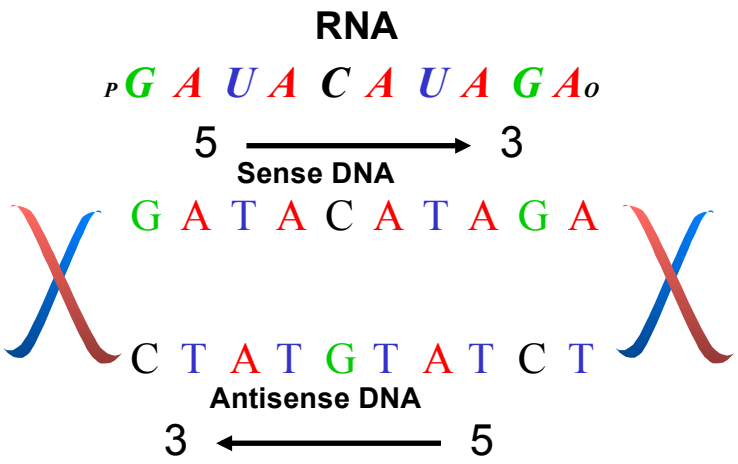
- The initial product of gene expression is always RNA
- In some cases this RNA is functional (e.g. tRNA, rRNA, snRNA)
- In many cases this RNA is the template for protein translation (eg mRNA)
- Gene expression occurs in the cell nucleus. RNA is exported to the cell cytoplasm, to be used in "Protein Translation"

Eukaryotic RNA Polymerases

Eukaryotic cells contain three types of RNA polymerases:

- RNA Polymerase I -Transcribes rRNA genes
- RNA Polymerase III- Transcribes tRNA and 5S RNA genes
- RNA Polymerase II - Transcribes genes encoding proteins into mRNA

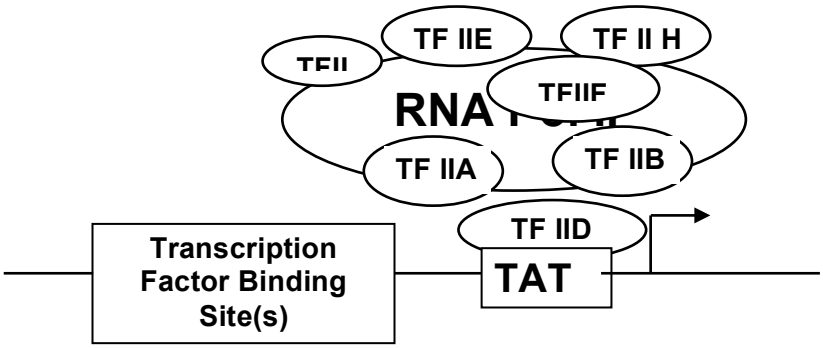
Transcription: Making an RNA copy of a DNA strand



3. Ribonucleotide bases are joined by phosphodiester bonds.

The RNA chain grows one base at a time in a 5' ->3' direction

Components of the Basal Transcription Complex



- The **Basal Transcription Complex** allows **RNA polymerase II** to be phosphorylated and then engage in transcription.
- In the absence of binding of other **Transcription Factors** this produces a **Basal** (low) level of transcription.
- Transcription factors "bend DNA" on binding. They can interact with each other and the Basal Transcription Complex to modulate transcription.

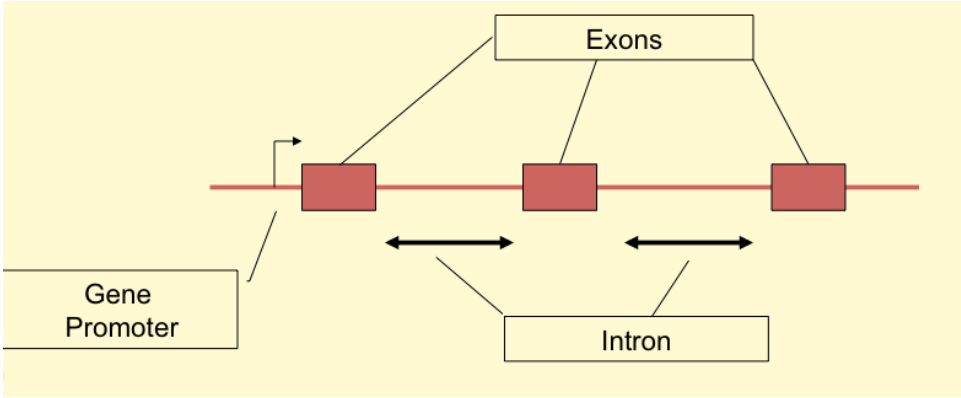
Nucleic Acids and Gene Expression

Gene Organisation & Transcription Part 2 - mRNA Processing

Learning Objectives

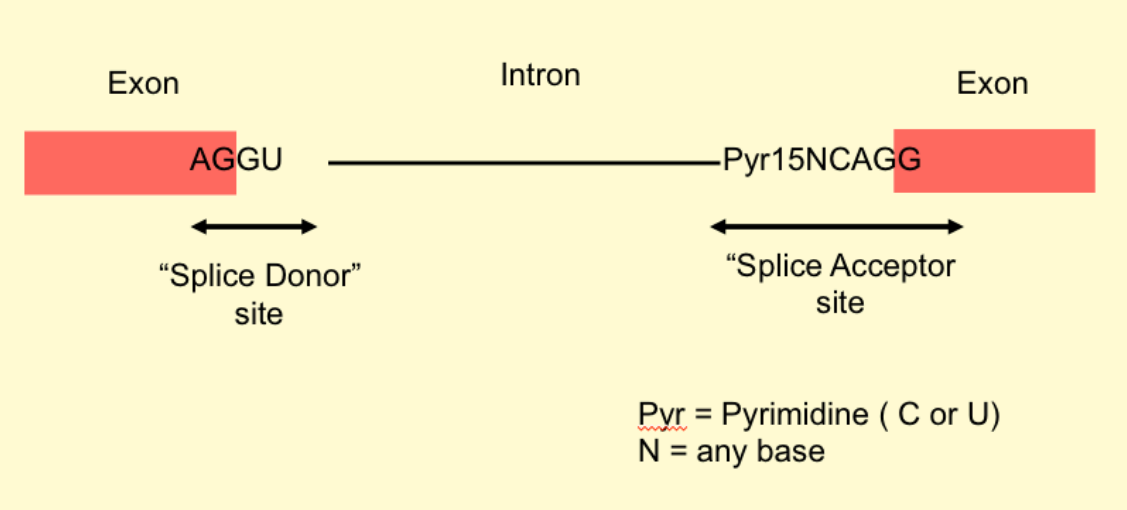
- Describe, with the aid of diagrams, the events that take place in pre- mRNA processing
- Define what is meant by a "splice donor site"
- Define what is meant by a "splice acceptor site"
- Describe the "lariat" intermediate in mRNA splicing
- Define the function of the "Spliceosome"
- Describe the addition of a "cap" and "poly A tail" to pre-messenger (hn-) RNA.
- With examples, describe how mutations in splice sites feature in human disease.

The Anatomy of a Gene



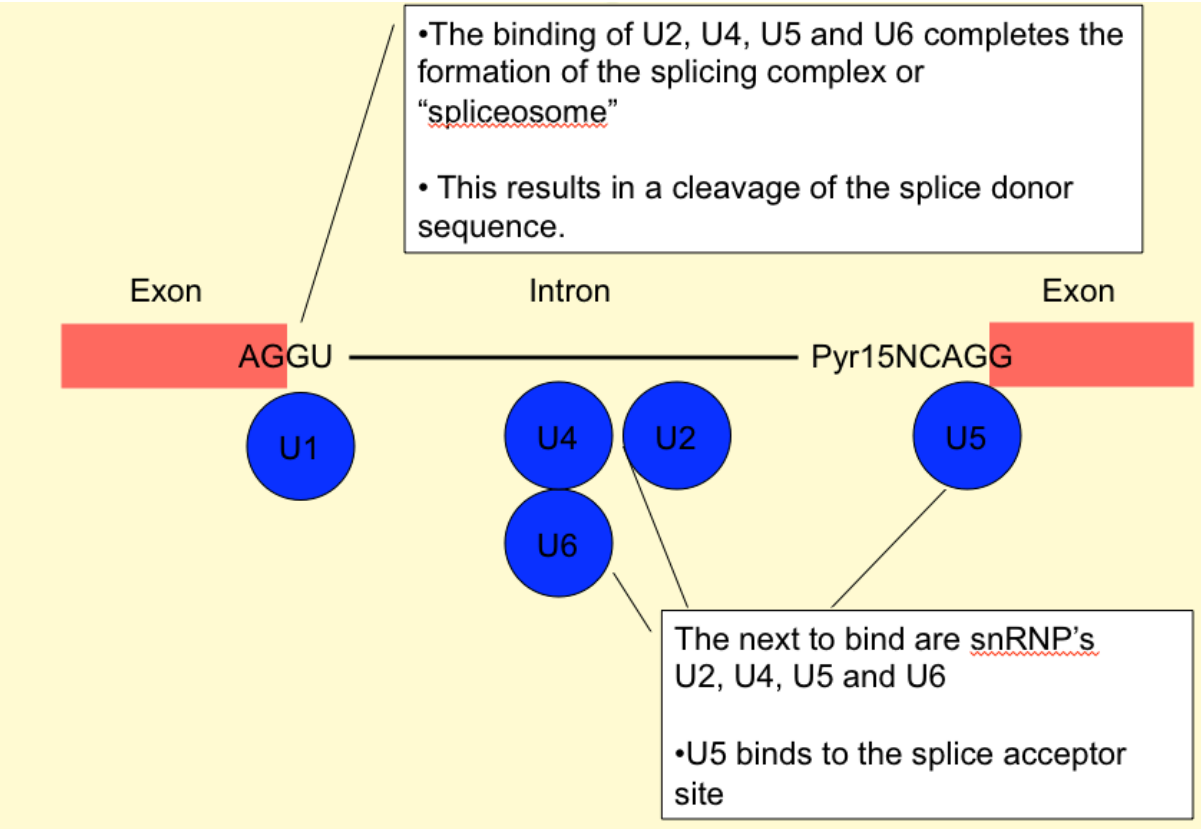
- The promoter lies at the start (5' end) of the gene
- The sequence information contained in the final mRNA is encoded "discontinuously" in the DNA of the gene
- Segments of the gene which contain sequences that form part of the final RNA are called "**exons**"
- "**Introns**" are sequences in the gene which are transcribed but are edited out of the final mRNA

The sequence of events in mRNA splicing



Introns start with the sequence "GU" and end with the sequence "AG"

The sequence of events in mRNA splicing



CMSLecture3.MDJ

Nucleic Acids and Gene Expression:
Protein translation & post-translational modification

Learning Objectives
- Outline the mechanisms by which ribosomes can translate a mRNA sequence into a protein sequence.
- Describe the role of aminoacyl tRNAs in ensuring the fidelity of the genetic code.
- State how a ribosome recognises the start and end of a sequence to be translated.
- Explain why some antibiotics inhibit protein synthesis in prokaryotes but not eukaryotes.
- Identify the features of a newly-synthesised protein that are required for it to enter the secretory pathway.
- Give examples of the ways in which newly-synthesised proteins can be post-translationally modified.

- There is a linear relationship between the information encoded within DNA (the genetic code) and the proteins that are synthesised using that information

- Three nucleotides encode one amino acid, hence the name “triplet code”; a group of three nucleotides is called a codon

- Protein synthesis always starts with Met =AUG and finishes with a stop codon =UAA, UAG or UGA

- transfer RNAs are the transporters of amino acids (bound to the 3’ terminus) to the ribosome; they contain an anticodon loop that binds to the corresponding codon on the mRNA

- Translation initiates with:

Step 1: dissociation of ribosome subunits (40S + 60S)
Step 2: assembly of preinitiation complex containing Met-tRNA + Initiation Factors + 40S subunit
Step 3: binding of mRNA to preinitiation complex
Step 4: binding of 60S subunit

- Translation elongation proceeds with:

Step 1: binding of new tRNA to A site
Step 2: catalysis of peptide bond between two amino acids by peptidyl transferase
Step 3: translocation of tRNA to P site and dissociation of first tRNA

- Translation termination occurs by:

Step 1: recognition of stop codon
Step 2: release of peptide chain
Step 3: dissociation of release factors and ribosomes

- Antibiotics are natural products of bacteria or fungi that can selectively inhibit prokaryotic protein synthesis because the translational machinery is complex and easily disrupted

- Synthesis of proteins destined for the secretory pathway to the cell surface occurs on Rough Endoplasmic Reticulum (RER) in the following stages:

Step 1: recognition of hydrophobic N-terminal signal sequence by signal recognition particle (SRP)
Step 2: binding of SRP to a receptor at the RER surface

Step 3: translocation of the growing protein into the lumen of RER
Step 4: cleavage of signal sequence and protein folding
Transmembrane proteins have additional hydrophobic sequences that stick in the membrane of the RER

• Post-translational modification - after synthesis many proteins, e.g. insulin, are modified further before they are fully functional. Modifications include:
- Proteolytic cleavage
- Disulphide bond formation
- Addition of carbohydrate (Glycosylation)
- Addition of phosphate (Phosphorylation)
- Addition of lipid groups (Acylation, Prenylation)

Nucleic Acids and Gene Expression

Analysis of Nucleic acids

Learning Objectives:
- Explain the term hybridisation, used for binding of a probe to a nucleic acid.
- Explain the concept of stringency of hybridisation, and the factors that contribute to stringency.
- Explain how the polymerase chain reaction (PCR) is used to amplify small amounts of DNA for subsequent analysis.
- Describe in general terms the way in which PCR primers would be selected to amplify a given DNA sequence.
- Describe the reactions carried out by restriction enzymes (restriction endonucleases) and explain their usefulness in analysis of DNA.

Cell-based DNA cloning

- Construction of recombinant DNA molecules *in vitro:* cutting a target DNA and a replicon with restriction endonucleases, so that the ends of the two DNA sequences are compatible. Joining the DNA fragments by using the enzyme DNA ligase.
- Transformation of the recombinant DNA molecules into host cells (bacteria, yeast)
- Selective propagation of individual cell colonies (selectable antibiotic resistance markers).
- Expansion of the cell culture and isolation of recombinant DNA

Restriction Endonucleases

- Enzymes that cleave DNA at specific recognition sites, usually 4-8bp palindromic sequences - produce "blunt" or "sticky" DNA ends.
- The longer the recognition site, the less frequently it occurs in DNA.

Separation of DNA fragments

- Electrophoresis - DNA is negatively charged due to its phosphate backbone and moves towards the anode (+ve electrode) when an electrical force is applied to it.
- When DNA is forced to travel through a porous gel matrix (agarose / polyacrylamide gel) small fragments are retarded less than large fragments and hence travel faster.
- After resolution, DNA can be isolated from the gel or transferred to a membrane to form a replica for hybridisation.

Nucleic Acid Hybridisation

- A *key method* for detecting specific nucleic acid sequences in which homologous single-stranded DNA or RNA molecules combine via homologous base-pairing to form double-stranded molecules.

- Standard assay involves a labeled nucleic acid probe (DNA, RNA or oligonucleotide) to identify homologous related molecules in a mixture of target unlabeled nucleic acids.

Hybridisation assays

- Target DNA is immobilised on a solid support - nylon or nitrocellulose membrane - which readily binds single-stranded nucleic acid (e.g. DNA) and then hybridised with a solution of labeled probe (radioactive or fluorescent).
- Examples:
  - Southern blot hybridisation (DNA target and DNA probe)
  - Northern blot hybridisation (RNA target and DNA probe)
  - Colony blot hybridisation (bacterial DNA target, DNA probe)
  - Chromosome *in situ* hybridisation (Chromosome target and DNA probe)
  - Tissue *in situ* hybridisation (RNA target and RNA probe)
  - Reverse hybridisation – Microarrays (immobilised DNA or oligonucleotide probe, target DNA solution)

Melting temperature and hybridisation stringency

- Denaturation of a probe DNA is achieved by heating until the hydrogen bonds between the bases holding the two strands together are disrupted.
- The energy needed to do this depends on strand length (longer strand = more hydrogen bonds to break), base composition (G-C pair has one more hydrogen bond than A-T) and chemical environment (monovalent cations stabilise the DNA duplex by neutralising charge on phosphate backbone; denaturants (formamide/urea) destabilise the DNA duplex).
- Melting temperature (Tm) - measure of nucleic acid duplex stability (Hybridisation is carried out at temperatures < 25˚C below Tm).
- Hybridisation stringency (i.e. the power to distinguish between related sequences) increases with increase in temperature and decrease in salt concentration.

Cell-free DNA cloning - Polymerase Chain Reaction (PCR)

- *In vitro* method to allow selective amplification of a specific target DNA within a heterogeneous collection of DNA sequences (e.g. total genomic DNA or complex cDNA population).
- Some sequence information is needed to design 2 primers (15 - 25 nucleotides in length), one complimentary to each strand of the DNA to be copied.
- Primers are specifically annealed to heat-denatured DNA.
- Thermostable DNA polymerase + dNTPs extend from the primers and generate new strands.
- Denature the DNA and repeat the cycle many times -> geometric increase**.**
- Applications: typing genetic markers; Detecting or introducing point mutations; cDNA cloning; genome walking; detecting gene expression; DNA sequencing.

DNA Sequence Analysis

- Understand the principle of Sanger ‘dideoxy’ chain termination DNA sequencing
- Resolution of DNA fragments

References

- “Human Molecular Genetics” - Strachan and Read
- “Genes” - Lewin
- “Molecular Biology of the Cell” - Alberts *et al*.