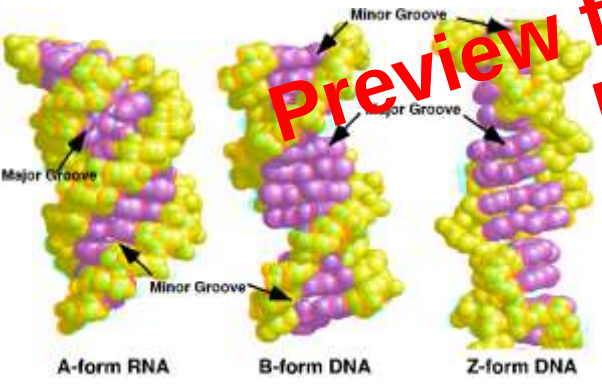
GENETICS 2018

DNA structure and organisation ✓✓

At the end of this lecture you should be able to:

- Describe the essential structural parameters of the DNA double helix
  - DNA is right handed – the right zig is over the left zag
  - The strands have opposite polarity and run 5' to 3'

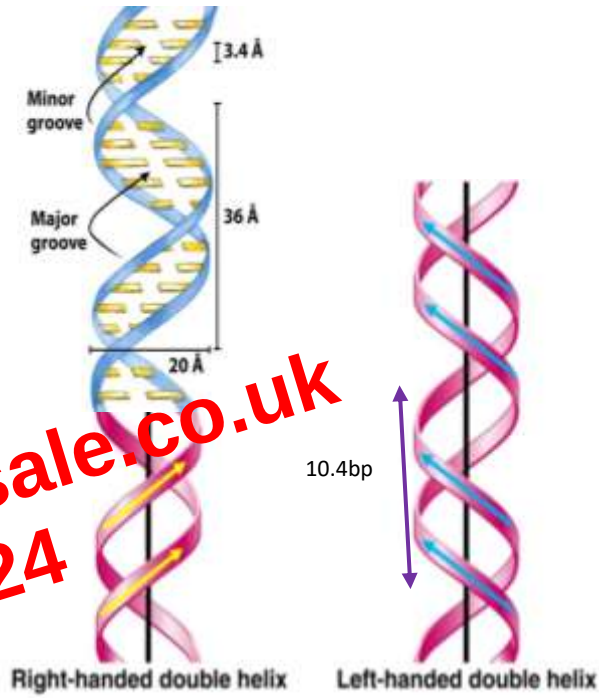
  - It has minor and major grooves – different atoms of the bases can be accessed in each groove, and most of the sequence specific binding information is in the major groove. Proteins bind in the grooves, and there is huge binding energy.
  - The helix has approximately 10.4 base pairs per turn and a 34.3° rotation per base (axial twist). One angstrom (Å) is about the size of a covalent bond.
  - Remember when drawing DNA that it is right handed, and if drawing a replication fork then all three helices are right handed
  - There are three forms of DNA:



A – found in crystalline structures with dehydration, the structure is short and distorted, and the bases not co-planar

B – standard r-handed DNA form found in most systems (co-planar)

Z – left-handed form which is longer and thinner, and can occur in the middle of B-form DNA (co-planar). Can form due to alternating purine-pyrimidine sequences such as GCGCGC.



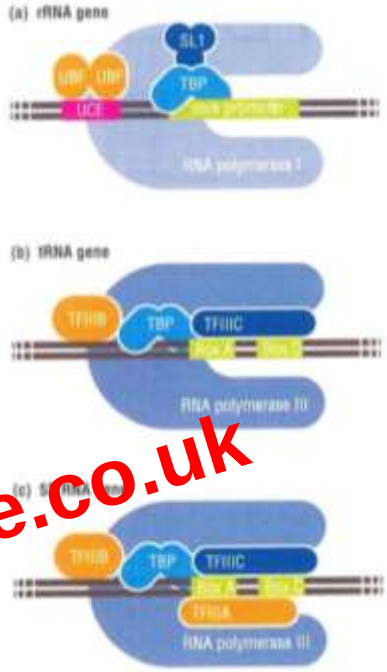
  - The structure of DNA was discovered by Watson and Crick, who used X-ray diffraction to show that the DNA was in a well ordered lattice/helix.
  - There are various forms of displacement of bases pairs, with the Z axis being the axis of the helix. These have different values in the different forms of DNA.
  - Do not confuse propeller twist (between bases) with the axial twist of the helix (34.3°)
  - Ambiguities:
    - Y = C or T
    - R = A or G
    - N = any nucleotide

- Polar mutations – indels which cause a premature stop codon, leaving a large section untranslated. Only happens in polycistronic (multi transcript) mRNA, and affects all later genes.

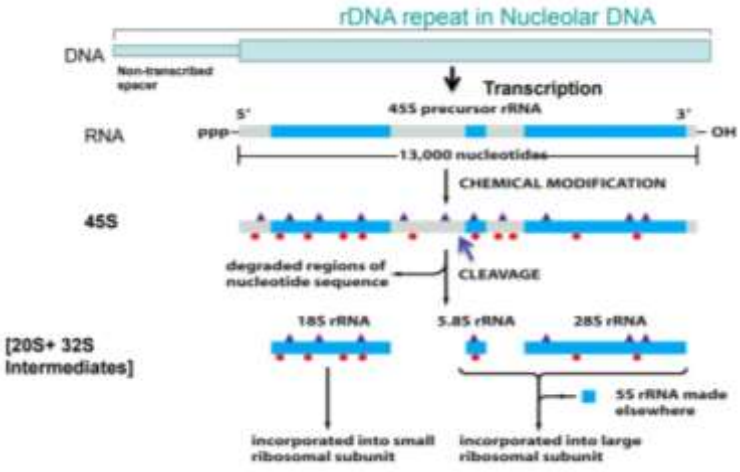
Eukaryotic transcription and RNA processing ✓✓

At the end of this lecture you should be able to:

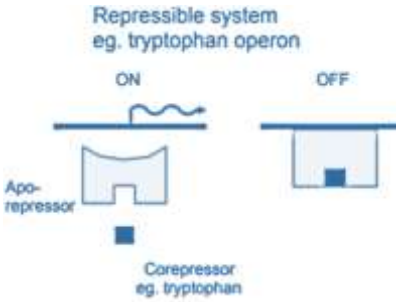
- Describe different RNA polymerases and their subunits
  - All three polymerases share common subunits
    RNA polymerase I – makes rRNA (stable)
    RNA polymerase II – makes pre-mRNA (high-turnover)
    RNA polymerase III – makes tRNA/5S rRNA (stable)
  - Eukaryotic RNA polymerases I and III rely on a distinct set of proteins to initiate transcription. Both enzymes are very similar to RNA polymerase II, with several of the core enzyme subunits identical in all three eukaryotic RNA polymerases. Pol I and III recognise different promoter sequences and have their own unique transcription factors.
  - TBP is a common feature across all three, to initiate transcription. TBP is a subunit of a larger complex: SLI, TFIID, and TFIIIB (for I, II and III respectively).
  - RNA polymerase I binds to a promoter containing a core promoter element and an upstream control element (UCE). TBP, which is part of a larger complex called SL1, helps RNA polymerase I to recognize the core promoter.
  - Genes transcribed by RNA polymerase III have an unusual promoter structure: some of the key promoter elements are located downstream of the transcription start site. There are different pre-initiation complexes formed by RNA polymerase III upstream of tRNA genes and the 5S RNA gene. Promoter recognition by RNA polymerase III is mediated by TBP – in TFIIIB - mirroring promoter recognition by polymerases I and II.

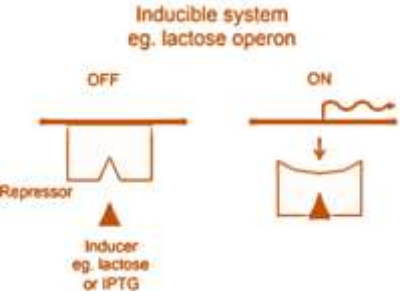


- Understand rRNA and how it is produced
  - rRNA makes up the majority of the RNA in a call, as cells *always* need ribosomes (but gene requirements vary)
  - 45S etc - the S is for svedburg units, which measure how fast it will sediment in a centrifuge
  - Different S values are reported do to imprecise measurement.
  - Chemical modifications are comprised mostly of base and sugar methylation, and pseudouridine addition (fake uracil, no one sure why)

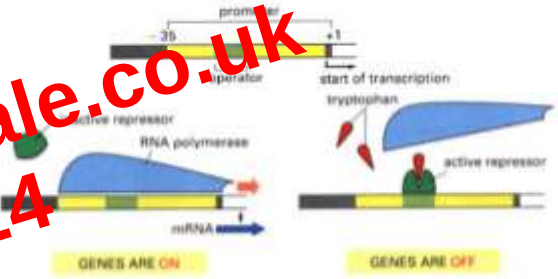


- Alternative sigma factors control different gene sets in bacteria. Genes are regulated by the availability of alternative sigma factors which recognise different promoters (giving different transcripts)
- These alternatives may be encoded by an invading virus
- In the host cell a change in conditions (eg low resources) triggers sigma factors to create the relevant suite of genes
- Phage T7 encodes its own RNA polymerase which recognises it's own promoters (an extreme version of having your own sigma factors)
- Repressible system – anabolic (synthesis), eg. Tryptophan operon
  - Are always on as standard, with the apo-repressor unbound
  - Corepressor (eg. Tryptophan) has to bind to the apo-repressor in order to change the repressor shape so that it can bind and repress the gene (stopping synthesis)



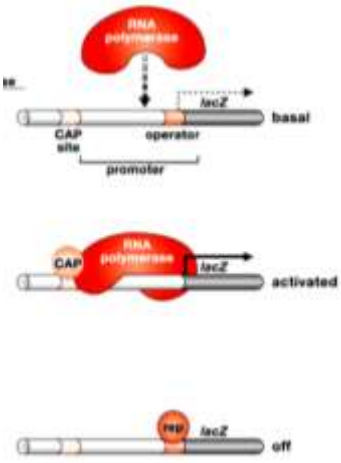


  - Inducible system – catabolic (breakdown), eg. Lactose operon
    - Are off as standard, with the repressor already bound
    - The inducer (eg. Lactose or IPTG) has to bind the repressor so that it changes shape and releases the DNA
  - Both ligands which bind the repressors are called effectors
  - Both above examples use sigma70
  - The tryptophan operon (repressible) is a common example. The transcription factors bind to the 'operator' – a location between the -35 box and the +1 start site. The bound transcription factor (and corepressor) prevents the RNA polymerase from binding. Repressors and activators often have helix-turn-helix motifs.



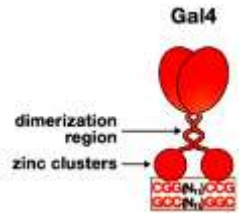
- The lac repressor only represses the lactose operon in the absence of its effector. The repressor still binds to an operator sequence, which is downstream form the promoter sequence when the RNA polymerase binds. This means that if the repressor dissociates it is ready to translate straight away.

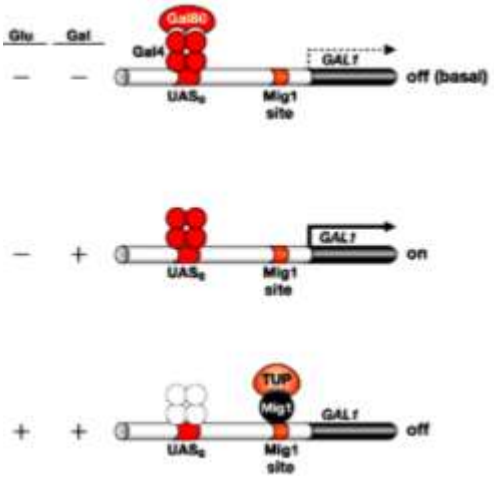
  - Lac can be activated by the CAP protein (catabolite activator protein/ CRP/cAMP receptor protein). It aids the recruitment of RNA polymerase in the absence of glucose (cell prefers glucose to lactose, so only uses a low level of lactose in the presence of glucose). The CAP protein binds upstream of the operator to promote lactose catabolism.



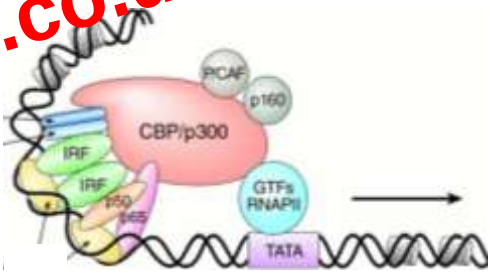
- In reality it's a little more complex, because there are three operator sites. Repressor dimers bound at the three sites interact with each other to form a higher-order complex. These dimers can block the CAP protein from binding

- Understand that transcription factors are modular (control of Gal1 in yeast)
  - GAL1 is positively regulated by galactose, and negatively regulated by glucose.
  - It has an upstream activating sequence for galactose (UAS_G) , which binds Gal4. This in turn binds Gal80 in the absence of galactose, which represses transcription.
  - In the presence of galactose Gal4 cannot bind Gal80, and GAL1 is expressed
  - In the presence of glucose and galactose, the Mig1 protein (and TUP) bind to the Mig1 site, repressing the GAL1 protein (as glucose is preferred over galactose
  - The zinc cluster binding domain only act to tether the activating domain near to the promoter (increasing its relative concentration to the transcription machinery) – any DNA binding domain works





  - The activation domain has two activating regions and a Gal80 binding region
  - Gal80 is not intrinsically inhibitory, it just hides the activating region on Gal4 (if activating region is stuck to Gal80 instead, then Gal80 acts as an activator) – anything which tethers the activating region close to the promoter activates
  - RNA polymerase II is not sat on the promoter waiting for activation because the transcription factors are not at the promoter before activation. This makes this system more similar to recruitment and tethering by CAP than activation by NtrC and MerR
  - Other proteins are recruited at induction and repression, mostly histone modifiers, and nucleosome remodelling machines
  - Higher-order complexes may contain even more additional components like the example on the right (don't learn)



- Describe the role of histone modifications involved in gene expression control
  - Chemical modifications of histone tails control many properties, like condensation, accessibility, and the processivity of transcription
  - Transcription alters the pattern of histone modification, and the modifications alter the transcribability of the chromatin. There is a high number of combinations. This forms the concept of the histone code.
  - The pattern of histone modifications definitely gives an indication of the recent rate of transcription, but they do not directly dictate the rate.
  - The primary signals must come from protein-DNA interactions, which are then related into DNA by signalling cascades.
  - This epigenetic code is not deterministic, more like a consensus of modifications caused by transcriptions.