Transcription in Eukaryotes

Introduction - RNA Pols

- Eukaryotic genes differ from prokaryotic ones in many ways
  - They're HUGE, because they contain "introns" that need to be removed before translation
- There are three RNA Pols in euks
  - Identification
    - Chromatographic purification and elution at different salt concentrations
    - Each has a different sensitivity to the FUNGAL TOXIN ALPHA-AMANITIN, which can be used to distinguish their activites
  - RNA Pol I
    - In nucleolus
    - Transcribes most rRNA
    - *Insensitive to alpha-amanitin*
  - RNA Pol II
    - In nucleoplasm
    - Trascribes all protein-coding genes (mRNA) and some small nuclear RNA (snRNA) genes
    - *Very sensitive to alpha-amanitin*
  - RNA Pol III
    - In nucleoplasm
    - Transcribes tRNA, 5S rRNA, U6 snRNA and certain other small RNAs
    - *Moderately sensitive to alpha-amanitin*
  - These, naturally, have to be IMMENSELY processive

Pol I Genes

- Responsible for the continuous synthesis of rRNA in interphase
  - Human cells contain five clusters of ~40 copies of the rRNA gene on different chromosomes
  - Each produces a 45S transcript, which is then cleaved to give 28S 18S and 5.8S RNA
  - Active transcription from these extensive gene clusters can be vividly seen under an electron micrograph as a "christmas tree"
- Promoter
  - Mammalian pre-rRNA gene promoters have a bipartite transcription control region
    - CORE ELEMENT - transcription start site (-36 to +6)
    - The UPSTREAM CONTROL ELEMENT (UCE) - increases transcription greatly (about -100), G-C rich
  - A specific protein called UPSTREAM BINDING FACTOR (UBF) binds to the UCE AND to a sequence in the upstream part of the core element
    - The two molecules of UBF then stick to each other, forming a LOOP of DNA between the two binding sites
  - An additional factor, SELECTIVITY FACTOR 1 (SL1) is involved
    - It binds and stabilises the UBF-DNA complex, and interacts with the other part of the core element
    - SL1 binding allows RNA Pol I to bind to the complex and initiate transcription
    - A protein called TBP (TATA-binding protein), a critical factor in eukaryotic transcription required by ALL RNA Pol genes in eukaryotes
    - SL1 contains several subunits
      - Other subunits - TBP-associated factors - TAFs
- Termination
  - Terminate at a specific 18bp sequence that is recognised by a specific protein

Pol II Genes

- Pol II is required for the transcription of most genes - it is responsible for mRNA synthesis is eukaryotes
- The mRNA produced by Pol II is extensively modified after synthesis (see later)
- Promoters
  - Most Eukaryotic promoters contain a sequence called the TATA BOX about -25 to -35 bp
    - This acts in a similar way to the -10 promoter sequence in prokaryotic transcription
    - But much more complicated, involving a whole range of Transcription Factors binding in a specific order because transcription is initiated
  - Some contain an INITIATOR ELEMENT instead of the TATA box
    - Located around the transcription start site
    - It seems that a further DNA-binding protein binds to this element
    - This then recruits TBP (TATA binding protein) which in turn leads to the cascade of events above
  - Some have neither, and are transcribed very slowly
- Upstream Regulatory Elements
  - The low activity of basal promoters is greatly increased by the presence of other elements located UPSTREAM of the promoter
  - Two common examples
    - The SP1 (or GC box) [~-90]
    - The CCAAT box [~ -50 to -70]
- Enhancers
  - Many genes have ENHANCERS
  - These are regulatory elements located MANY THOUSANDS of bp away from the promoter
  - Probably act by recruiting an activator protein and "bending over" to come into direct contact with the promoter (SEE BEFORE)
  - Some genes have associated SILENCERS
  - First observed in the genome of SV40 - a sequence of ~100bp can significantly increase transcription from a basal promoter, even far upstream

Pol III genes

- Promoter
  - In tRNA genes
    - The promoters for Pol III are WITHIN the GENE being transcribed
    - There are two highly conserved promoter sequences ("A box" and "B box"), which also code for highly conserved areas of the tRNA
    - First, TFIIIC binds to BOTH these boxes
      - Contains TBP (TATA BINDING PROTEIN)
    - TFIIB is then able to bind 50bp upstream from the A box, at the start of transcripiton
      - TFIIB seems to have no sequence specificity
      - It needs TFIIIC to "guide" it to be able to do that
    - TFIIB then allows RNA Pol III to BIND and INITIATE TRANSCRIPTION
  - In 5S RNA genes
    - Contains two sequences - the "A box" and "C box"
    - TFIIIA binds to the "C box"
    - TFIIIC is then able to bind over the whole thing
    - RNA Pol can now be recruited
  - Some also use upstream regulatory sequences
- Termination
  - Terminate with a run of Us associated with a GC-rich region

(C) Daniel Guetta, 2007