Computational Biology I
LSM5191
(2003/4)

Aylwin Ng, D.Phil

Lecture Notes:
Transcriptome:
Molecular Biology of Gene Expression I
Flow of information: DNA to polypeptide

DNA → Transcription → Addition of 5’ cap → Cleavage & addition of polyA tail at 3’ end → RNA splicing → Transport to cytoplasm → Translation → Polypeptide
TRANSCRIPTION – An Overview

1. Binding of RNA polymerase

2. Separation of DNA

3. Base pairing of first nucleoside triphosphate to starting base in DNA

4. Binding of second nucleoside triphosphate and formation of phosphodiester bond between its 5' phosphate and the 3' hydroxyl on previous nucleotide

5. Binding and addition of third nucleoside triphosphate; continuation of process with each successive nucleoside triphosphate, as RNA polymerase moves along template DNA
GENE EXPRESSION IN BACTERIA

• Genes with related functions often located next to each other.

• This cluster of genes comprise a single transcription unit called an OPERON, i.e., a single mRNA molecule contains the full set of genes of the operon.

Example: genes involved in lactose metabolism constitute the Lac Operon

I P O Z Y A

• Hence each prokaryotic mRNA encodes several polypeptides.

• mRNA is poly-cistronic (a cistron is defined as a genetic unit that encodes a single polypeptide).

• Poly-cistronic mRNA contains multiple ribosome-binding sites near start sites for all protein coding regions in the mRNA.
Some Conventions:

- **+1** is designated the transcription-initiation site in the DNA sequence.
- **+ve numbers** are assigned to base-pairs extending **DOWNSTREAM** in the direction of transcription.
- **-ve numbers** are assigned to nucleotide sequences **UPSTREAM** of the transcription start site.
**E. coli** PROMOTER SITES

- PROMOTER is where RNA polymerase binds.
- Various other proteins (activators, repressors) interact with DNA at or near the promoter to regulate transcription initiation.
- 2 regions ( -10 and -35 regions) in most *E. coli* promoters are critical for binding RNA polymerase holoenzyme (β', β, α, α, σ^70) via its σ^70 subunit (or initiation factor).
- After transcribing ~10 bp, σ^70 subunit is released.
- The core RNA polymerase (β', β, α, α) continues transcribing (chain elongation).
- ‘Strong’ promoters = promoters at which RNA pol initiates transcription at high frequency (dependent on enzyme’s affinity for promoter).

**Strong E. coli promoters**

```plaintext
ty r tRNA
rrn D1
rrn X1
rrn (DXE)_2
rrn E1
rrn A1
rrn A2
λ Pr
λ Pγ
T7 A3
T7 A1
T7 A2
fd VIII
```

```plaintext
TCTCAACGTAACACTTTACAGCGCG•••CGCTATTGGATATGC•GCCCGCGCTTCCCGATAAGGG
GATCAAAAAATACCTTGGTGCAAAAA•TTGGGATCCCTATAATGCCTCCGGTGAACAGACACACG
ATGCAATTTCCTGCCTTGCCTTTTCTCTGA••GCCGACTCCCTATAATGCCTCCGGTGAACAGACACG
CCCTGAAATTGAGGCTTGACTCTGAAA•GAGGAAGAGCTATAATGCCTCCGGTGAACAGACACG
CTGCAATTTCTTCTTGTGCCTCATCGCCGG•GAGGAAGAGCTATAATGCCTCCGGTGAACAGACACG
TTTCTAATTTCTTCTTGTGCCTCATCGCCGG•AATAACTCTCTTATAATGCCTCCGGTGAACAGACACG
GCAAATATATATACCTTGCCTCTAGTTACG•CCGGAAGGCGTTCCGGTGAACAGACACG
AACCTCGGCCGTGTTGACTATTTTA•CCTTCTGCGGGTGATATGG•TTGGCATCTACTAGAGGG
TATCTTGCGCCGTTGACTATTTTA•CCTTCTGCGGGTGATATGG•TTGGCATCTACTAGAGGG
TTTCTAATTTCTTCTTGTGCCTCATCGCCGG•GAGGAAGAGCTATAATGCCTCCGGTGAACAGACACG
GCAAATATATATACCTTGCCTCTAGTTACG•CCGGAAGGCGTTCCGGTGAACAGACACG
```


**Image**: Diagram showing the specific DNA sequences for various E. coli promoters.
Identification of PROMOTER SITES

- **DNase I footprinting assays** identify protein-DNA interactions.
- DNase I randomly hydrolyses phosphodiester bond.
- Low concentration of DNase I used → on average each DNA molecule is cleaved just once.
EUKARYOTIC TRANSCRIPTION

- The basic principles, in general, also apply to eukaryotic organisms.
- Transcription is initiated at a specific base pair and is controlled by the binding of trans-acting proteins (transcription factors) to cis-acting regulatory DNA sequences.
- However, eukaryotic cis-acting elements are often much further from the promoter they regulate, and transcription from a single promoter may be regulated by binding of multiple transcription factors to alternative control elements.
- Transcription control sequences can be identified by analysis of a 5′-deletion series.
- Eukaryotes have THREE (instead of just one) RNA polymerases (all large multi-subunit enzymes).
- Eukaryotic mRNAs are generally monocistronic. [cf Prokaryotes]
EUKARYOTIC RNA POLYMERASES

RNA Polymerase I:
- Transcribes gene encoding ribosomal RNA (45S precursor yielding 28S, 18S, 5.8S rRNAs)

RNA Polymerase II:
- Transcribes all protein-coding genes,
- Transcribes genes encoding small nuclear RNAs (U1, U2, U3 etc.)

RNA Polymerase III:
- Transcribes genes encoding transfer RNA,
- Transcribes gene encoding 5S rRNA,
- Transcribes gene encoding snRNA U6.
RNA POLYMERASE I

- Essential protein factors (rather than the polymerase) recognise DNA sequences around transcription start site.
- Key sequences recognised by these factors are located within 50 bases upstream of start site.
- SL1 factor recruits RNA polymerase I.
Brief Notes:
- TFIIB recruits RNA pol II.
- TFIIH phosphorylates C-term domain of RNA pol II.
- Phosphorylated form is able to initiate transcription.

RNA POLYMERASE II
RNA

Total RNA

Coding RNA
4% of total

Noncoding RNA
96% of total

Pre-mRNA (hnRNA)

mRNA

Pre-rRNA

rRNA

Pre-tRNA

tRNA

snRNA

snoRNA

scRNA
tmRNA and various other types

KEY

- **All organisms**
- **Eukaryotes only**
- **Bacteria only**
POST-TRANSCRIPTIOANL EVENTS in EUKARYOTES

- (1) Capping
- (2) Polyadenylation
- (3) RNA splicing
- (4) RNA transport
- (5) Translation
(1) CAPPING

- 5’ end of nascent mRNA is modified (capped).
- **Cap protects RNA from degradation by 5’→3’ exonuclease activity.**
- Capping only occurs in Eukaryotes!
- Addition of a Methylated Guanylate residue (NOT encoded by DNA).
- Rxn catalysed by guanylyl transferase.
- 3 phosphate molecules separate the G residue from the first nucleotide in the chain (whereas only 1 P separates the other nucleotides).
- Guanylate is joined via a 5’-5’ linkage rather than the std. 3’-5’ linkage which links nucleotides in a growing chain.
(2) POLYADENYLATION

- Cleavage at 3’ end of mRNA
- Addition of poly(A) tail at 3’ end of cleaved mRNA

**CPSF:**
Cleavage & Polyadenylation Specificity Factor

**CstF:**
Cleavage stimulation factor

Poly(A) site

5’

AAUAAAAA

G/U

3’

CPSF

5’

AAUAAAAA

G/U

3’

CstF

5’

AAUAAAAA

3’

CPSF

Endonucleolytic cleavage

5’

AAUAAAAA

3’

CstF

5’

(A)_{200}

3’

Poly(A) polymerase

Degradation
Role of polyadenylation

• To protect mRNA from degradation by exonucleases.

• Exonucleases ‘attack’ its free 3’ end and rapidly degrades mRNA.

• Appears to increase the efficiency by which an mRNA is translated.

Not all mRNAs (encoding proteins) are polyadenylated, e.g.mRNAs encoding Histones.

• mRNA fold itself into a double-stranded stem-loop structure which protects it from degradation.
EXONS & INTRONS

- Protein-coding regions of a gene are known as **EXONS**.

- Intervening regions that do not encode parts of protein are known as **INTRONS**.

- Introns are transcribed into mRNA, but remains in nucleus.

- Hence, primary RNA transcript must have its introns removed before being transported into the cytoplasm and translated.

- **RNA SPLICING** is the process whereby introns are removed.
(3) RNA SPLICING – what’s the mechanism?

- Clue: Short, conserved sequences at splice junctions.
RNA SPLICING

5' splice site

GU

Exon 1

Intron

AG

Exon 2

3' splice site

Cleavage at 5' splice site

5' GU

Exon 1

Intron

AG

Exon 2

3' splice site

Lariat formation

U

G

5'

2'A

AG

3'

Cleavage at 3' spliced site

U

G

5'

2'A

AG

3' 5'

Exon 1

Exon 2

3'
In vitro analysis (Ruskin et al., 1984)

Nuclear extract (from cells) incubated with radio-labelled RNA:

- Starting RNA
- Final spliced product
- Excised intron
Small nuclear RNAs (snRNAs)

- These are splicing ‘factors’, i.e. assist in the splicing process.

- NOTE: they are NOT proteins, but RNA molecules !!!

- But snRNAs associate with small nuclear ribonucleoproteins (snRNPs) to form a large ribonucleoprotein complex called a Spliceosome.
SPLICEOSOMES assembly
ALTERNATIVE SPLICING

- A mechanism for tissue-specific expression
- E.g. Hepatocytes generate fibronectin proteins that are different from those produced by Fibroblasts.
(4) RNA TRANSPORT

• Spliced mRNA must be transported out from the nucleus (across the nuclear membrane) into the cytoplasm for translation into protein.

• Heterogeneous nuclear RNPs (hnRNPs) is likely to mediate this transport by associating with mRNA in nucleus.

• In yeast, the Gle1 protein mediates this transport.

• The Gle1 protein contains a short nuclear export signal (NES) sequence.

• NES sequence is also present in HIV’s Rev protein, which is involved in regulating the nuclear-cytoplasmic transport of different HIV mRNAs.
(5) Translation (next lecture)