Computational Biology I

LSM5191

Aylwin Ng, D.Phil

Lecture 5 Notes:

Control of Gene Expression
A liver cell is obviously different from a nerve cell.

Do all cells in an individual have the same DNA content?
CONTROL OF GENE EXPRESSION
REGULATION OF GENE EXPRESSION

• **Any of these stages** could be used to regulate expression of specific genes in particular tissues.
• But in general, the **primary control of gene expression is at the level of transcription.**
TRANSCRIPTIONAL CONTROL

Control beyond the interplay of just the core factors & machinery essential for the transcriptional process.

I. Regulatory Sequence Elements

(A) Short regulatory elements

(B) Enhancers or Enhancer Elements

(C) Locus control regions

II. Transcriptional activators
I. Regulatory Sequence Elements
(A) Short regulatory elements

- Some elements are **commonly found in many** genes:
  - e.g. TATA, CCAAT and Sp1 boxes.

- Others are found **only in certain** genes:
  - e.g. heat-shock element
    - found only in genes whose transcription is increased in response to elevated temperature.

Pelham, 1982, Cell 30:517-28
<table>
<thead>
<tr>
<th>Element Name</th>
<th>Consensus sequence</th>
<th>Other genes containing sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>TATA box</td>
<td>TATA A/T A A/T</td>
<td>Very many genes.</td>
</tr>
<tr>
<td>CCAAT box</td>
<td>TGTGGCTNNNAGC CAA</td>
<td>α- and β-globin, albumin, HSV tk, cellular oncogenes: c-ras, c-myc, etc.</td>
</tr>
<tr>
<td>Sp1 box</td>
<td>GGGCGG</td>
<td>Metallothionein IIA, type II procollagen, dihydrofolate reductase, etc.</td>
</tr>
<tr>
<td>CRE</td>
<td>T/G T/A CGTCA</td>
<td>Somatostatin, fibronectin, α-gonadotrophin, c-fos, etc.</td>
</tr>
<tr>
<td>AP2 box</td>
<td>CCCCCAGGC</td>
<td>Collagenase, MHC class 1 antigen H-2K(^b), metallothionein IIA.</td>
</tr>
<tr>
<td>Heat-shock consensus</td>
<td>CTNGAATNTTCTAG A</td>
<td>Heat-inducible genes hsp83, hsp27, etc.</td>
</tr>
</tbody>
</table>

Adapted from Latchman, D., 1998, Gene regulation, Stanley Thornes Publ.
(B) Enhancer or Enhancer Elements

• These elements can **activate a promoter** when placed:
  • up to several kb from promoter,
  • in either orientation relative to promoter,
  • upstream or downstream of the transcribed region, or within introns.

• Genes exhibiting tissue-specific expression found to contain enhancers.

• A **tissue-specific enhancer** can activate the promoter of its own or another gene **only in one particular tissue** and not others:
How would you design an experiment to demonstrate this?

High-level transcription of Gene X in cell type A,

Low-level transcription of Gene X in cell type B,

Enhancer is transferred to unrelated gene

High-level transcription of Gene Y in cell type A, (Enhancer is active; Activates promoter)

Low-level transcription of Gene Y in cell type B, (Enhancer is inactive)
In vivo demonstration

Case example:
Tissue-specific expression of Insulin gene in vivo.

• Construct: Insulin gene enhancer element linked to gene encoding large T antigen (Ag) of SV40 virus.

• This construct introduced into a fertilized mouse egg.

• Egg returned to oviduct of mouse.

• Expression of large T Ag analyzed in all tissues of the transgenic mouse (using specific antibody).

• Expression of large T was detectable only in the pancreas (specifically in the ? cells of the pancreatic islets which produce insulin).

▷ Enhancer is therefore capable of conferring the specific pattern of insulin gene expression on an unrelated gene in vivo.

Possible mechanisms of action:
- By changing chromatin structure leading to nucleosome displacement,
- By direct interaction with the proteins of the transcriptional apparatus:
Models (a) and (b) cannot explain the following finding:

• Immunoglobulin enhancer activates equally well 2 promoters located 1.7kb and 7.7kb away on the same DNA molecule.

• Models (a) & (b) would postulate that the sliding of factors or the assembly of connecting molecules would stop at the 1st promoter.

Model (c) is able to explain the observation showing the critical importance of DNA structure on the action of enhancers.

Region between SV40 enhancer & promoter:
• Removal of multiples of 10 bases (1 helical turn) Activity.
• Removal of bases corresponding to half a helical turn Activity disrupted.


Takahashi et al., 1986, Nature 319:121-126
Enhancer-binding proteins actually bend the DNA so that interactions can occur between regulatory proteins bound at distant sites on DNA.

e.g. T-cell receptor α chain gene enhancer.

LEF-1 factor binds to a site at the centre of this enhancer, bends DNA, brings other regulatory factors into close proximity:
Werner and Burley, 1997, Cell 88:733-736
LOCUS CONTROL REGIONS (LCRs)

- LCR elements are sequences (additional to promoters & enhancers) that are necessary for high-level gene expression.

- Influence expression of adjacent genes in a position-independent manner, i.e. regardless of the position of the LCR in the genome.

- Act in a tissue-specific manner.

- LCRs function by affecting chromatin structure:
  - When gene is introduced transiently into cells (i.e. exogenous DNA is not packaged into chromatin) LCR has no effect on gene activity.
  - When gene is an integral part of the chromosome LCR affects gene activity.

- LCR induces DNase I hypersensitivity in adjacent regions (e.g. β-globin cluster).

- DNase I hypersensitivity is characteristic of active or potentially active genes.

- Gene lacking LCR will be subject to the influence of adjacent regulatory elements which might repress its expression by directing its organization into a closed chromatin conformation.
LCRs (Cont.d)

- LCR elements have been identified in α- and β- globin gene clusters, the major histocompatibility (MHC) locus, CD2 and lysozyme genes.
- In the β-globin gene cluster, LCR is located 10-20kb upstream of the β-globin genes.
- Its deletion leads to a lack of expression of any genes in the cluster.
- In Humans, this leads to a lethal disease (Hispanic thalassaemia), in which no functional haemoglobin is produced.

II. Transcriptional activators
Transcriptional Activators (Transcription Factors)

Proteins that bind to DNA in a sequence-specific manner and regulate the level of transcription.

Some characteristic structural features used to classify members into certain classes of Trans-activators:

- Helix-turn-helix motif
- Zinc finger motif
- The Leucine zipper

Transcriptional Activators

Genome-wide comparison of transcriptional activator families across Eukaryotes

Adapted from Tupler et al., 2001, Nature 409:832
Helix-turn-Helix motif

- Many transcriptional activators with this type of DNA-binding domain are called **homeodomain proteins**.
- Name derived from a group of *Drosophila* genes (homeotic genes) in which the conserved sequence encoding this structural motif was 1st observed.
- Mutations in these homeotic genes result in transformation of one body part into another during fly’s development.
- These genes encode regulatory proteins that **activate** or **repress** activity of other genes encoding proteins required for development of certain structures.

E.g. The Engrailed (Eng) protein binds the identical sequence recognized by Ftz (another homeodomain protein) and blocks gene induction by Ftz.

Adapted from Harrison, 1991, Nature 353:715
Zinc finger motif

- This motif is common in eukaryotic proteins.
- Estimated 1% of all mammalian genes code for zinc finger proteins.
- At least 6 different versions of this motif.
- The first identified was the **Cys$_2$His$_2$ finger**.
- Consensus sequence: Tyr/Phe-X-Cys-X$_{2-4}$-Cys-X$_3$-Phe/Tyr-X$_5$-Leu-X$_2$-His-X$_{3.4}$-His
- This structure binds one Zn$^{2+}$ ion through the 2 Cys and 2 His side chains.
- The transcriptional activators, TFIIIA (for the gene encoding 5S RNA of the ribosome) was the 1$^{st}$ to be identified bearing this motif.
Cys$_4$ Zinc finger motif

- The second type is the Cys$_4$ zinc finger.
- Found in more than 100 transcriptional activators.
- Steroid receptor superfamily or now known as nuclear receptors.
- 2 groups of 4 critical Cys bind a Zn$^{2+}$ ion.

- Cys$_2$His$_2$ proteins generally contain 3 or 4 repeating finger units and bind to DNA as monomers.

- Cys$_4$ proteins generally contain only 2-finger units and bind to DNA as dimers (either homodimers or heterodimers).

- The yeast Gal4 protein exhibits the Cys$_6$ zinc finger motif.
Leucine zipper

- Motif present in many transcriptional activators.
- Contains **hydrophobic leucine** at every 7th position in the C-terminal portion of their DNA-binding domains.
- These proteins **bind to DNA as dimers**.
- Dimers form via hydrophobic interactions between the C-terminal regions of the \( \alpha \)-helices, forming a coiled-coil structure.
- Hydrophobic side chains form a stripe down one side of the \( \alpha \)-helix.
- Hydrophobic stripes make up the interacting surfaces between the helices in the coiled-coil dimer.

![Basic DNA-binding domain](image)
Fos and Jun proteins:

- Examples of transcriptional activators bearing the leucine zipper motif:

- Jun can bind as a homodimer to the AP1 recognition sequence, TGAGTCAG, transcriptional induction of phorbol esters.

- Fos cannot bind to DNA alone, but can form a heterodimer with Jun.

- Jun-Fos heterodimer binds AP1 with 30-fold greater affinity than Jun homodimer.
Modular nature of Transcriptional activators

- transcriptional activators have modular structures (e.g. with a DNA-binding and activation domains).

- Classic domain-type structure seen in yeast transcriptional activators GCN4.

- GCN4 induces genes encoding enzymes of amino-acid (a.a.) biosynthesis in response to a.a. starvation.

- Expt: 60a.a. region (containing DNA-binding site) introduced into cells binds GCN4-responsive genes but fails to activate transcription.

- This only confirms the DNA-binding domain.

- A functional test is needed to identify the activation domain.
Identify activation domain:

Perform ‘**Domain-swap**’ experiment to locate activation domain of FactorA:

- Link various regions of FactorA to DNA-binding domain of FactorB

![Diagram]

Where is activation domain?

DNA-binding domain

**FactorA**

**FactorB**

DNA-binding domain

Reporter gene activation

Response element & binding site for FactorB

Reporter gene
An extreme example of (‘parasitic’) modularity:

- Herpes simplex virus (HSV) VP16 protein activates the transcription of viral immediate-early genes during lytic infection.
- VP16 contains a potent activation region.
- VP16 contains no DNA-binding domain & therefore cannot bind DNA itself.
- But following infection, VP16 complexes with cellular Oct-1 protein, which binds the sequence, TAATGAAT, in the viral promoters, and
- Activation is therefore achieved.
General features of Activating Domains:

- Do not show strong a.a. sequence similarity amongst transcriptional activators.
- But in many cases, activating domains contain a very high proportion of acidic a.a., a region of strong negative charge.
- E.g. 17 acidic a.a. residues were found in the 82-a.a. N-terminal activating domain of the glucocorticoid receptor.
- E.g. 17 acidic a.a. residues were found in the 60-a.a. activating domain of GCN4.
- Hence activating domains also known as ‘Acidic Blobs’ or ‘Negative Noodles’.
- It has been suggested that activating domains adopt an $\alpha$-helical or an anti-parallel $\beta$-sheet conformation.
Translational Control
TRANSLATIONAL CONTROL

Regulation at the level of translation.

Significance of Translational Control:

• Translational control tends to occur in situations where very rapid responses are required.

• Translational control viewed as supplementing the regulation of transcription, to meet the requirements of particular specialized cases.

• E.g. following heat shock, it is necessary to:
  • Shut down rapidly enzyme and structural protein synthesis,
  • Rapidly synthesize heat-shock proteins.
TRANSLATIONAL CONTROL

• Some interesting examples of how control is mediated by untranslabeled region of mRNA:

Ferritin expression:
(Ferritin is an iron-storage protein)

• Control is mediated by sequences in the 5’ untranslabeled region of ferritin mRNA.
• Sequences in this region can fold into a stem-loop structure.
• Stem-loop structure is stabilized by the Iron-response-element binding protein (IRE-BP) interacting with this structure.

• Presence of Iron:
• IRE-BP binds iron and dissociates from the stem-loop in the process,
• stem-loop structure unfolds,
• enhanced translation of gene encoding ferritin.

IRE-BP

Ire-BP

Fe

5’ 3’

Start of translation

5’ 3’

Start of translation

Nascent polypeptide

ribosome

IRE-BP

Stem loop

+ Fe
**TRANSLATIONAL CONTROL**

**Transferin receptor expression:**
(Transferin receptor brings iron into cell)

- Control is mediated by sequences in the **3' untranslated region** (important for the stability) of the transferin receptor mRNA.
- Sequences in this region can also fold into a stem-loop structure.
- Stem-loop structure is stabilized by the Iron-response-element binding protein (IRE-BP) interacting with this structure.

- Presence of Iron:
  - IRE-BP binds iron and dissociates from the stem-loop in the process,
  - stem-loop structure unfolds,
  - RNA becomes *susceptible to nuclease degradation* at a rapid rate.