Control of Gene Expression

Different cell types of an organism contain the same DNA but the DNA is expressed differently. External signals can cause a cell to change the expression of its genes.
Gene Expression can be regulated in many steps:

1. Transcriptional control
2. RNA processing control
3. RNA transport and localization control
4. post-transcriptional control \((\text{RNA Silencing})\)
5. Translational control
6. protein activity control and protein degradation
7. Epigenetic elements
Studies of the number of different mRNAs suggest that, at any one time, a typical human cell expresses 30-60% of its approximately 25,000 genes.
Transcriptional control

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

The transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act by a single signal. Many others are complex and resemble tiny microprocessors, responding to a variety of signals that they interpret and integrate in order to switch their neighboring gene on or off. These switching devices are found in are composed of two types of fundamental components:

(1) **Short sequence of DNA**
(2) **gene regulatory proteins that recognize and bind to this DNA.**

<table>
<thead>
<tr>
<th>NAME</th>
<th>DNA SEQUENCE RECOGNIZED*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
</tr>
<tr>
<td>Lac repressor</td>
<td>5’ AAATCGTGACGGCTAAATGAG 3’ TTAACGTTTCCTTGTTAA</td>
</tr>
<tr>
<td>CAP</td>
<td>TGTGACGTTAGCTCAGT</td>
</tr>
<tr>
<td>Lambda repressor</td>
<td>TATCACCGCCAGAGGTCT</td>
</tr>
<tr>
<td><strong>Yeast</strong></td>
<td></td>
</tr>
<tr>
<td>Gal4</td>
<td>CGGAGAAGCTTTCCCTCCG</td>
</tr>
<tr>
<td>Mata2</td>
<td>GCTGTTGTATACA</td>
</tr>
<tr>
<td>Gcn4</td>
<td>ATGACTCATAGTA</td>
</tr>
<tr>
<td><strong>Drosophila</strong></td>
<td></td>
</tr>
<tr>
<td>Kruppel</td>
<td>AAGCGCTTTAA</td>
</tr>
<tr>
<td>Bicoid</td>
<td>GGGATTAAGA</td>
</tr>
<tr>
<td><strong>Mammals</strong></td>
<td></td>
</tr>
<tr>
<td>Sp1</td>
<td>GGCGGCG</td>
</tr>
<tr>
<td>Oct1 Pou domain</td>
<td>ATGCCAATGGTTTA</td>
</tr>
<tr>
<td>GATA1</td>
<td>TGATAGACTTAC</td>
</tr>
<tr>
<td>MyoD</td>
<td>CAATTGTTA</td>
</tr>
<tr>
<td>p53</td>
<td>GGCGAAGCTCTCCGTCCAGA</td>
</tr>
</tbody>
</table>
The **Helix-Turn-Helix Motif** is One of the Simplest and Most Common DNA-Binding Motifs

The helix-turn-helix motif is composed solely of amino acids. A second important group of DNA-binding motifs includes one or more **zinc atoms** as structural components. Zinc-coordinated DNA-binding motifs are called **zinc fingers**.

Some Proteins Use **Loops** That Enter the Major and Minor Grooves to Recognize DNA. **Homodimerization and Heterodimerization** help Gene Regulatory proteins to bind DNA sequences.
Predict the DNA sequences Recognized by All Gene Regulatory Proteins.
Gel-Mobility Shift Assay,
ChIP,
Digestion insensitive sequence

line. The effect of the proteins in the extract on the mobility of the DNA fragment is analyzed by polyacrylamide-gel electrophoresis followed by autoradiography. The free DNA fragments migrate rapidly to the bottom of the gel, while those fragments bound to proteins are retarded; the finding of six retarded bands suggests that the extract contains six different sequence-specific DNA-binding proteins (indicated as C1-C6) that bind to this DNA
Switching the tryptophan genes on and off.

If the level of tryptophan inside the cell is low, RNA polymerase binds to the promoter and transcribes the five genes of the tryptophan (Irp) operon. If the level of tryptophan is high, however, tryptophan repressor (helix-turn-helix protein) is activated to bind to the operator, where it blocks the binding of RNA polymerase to the promoter. The repressor protein has to have two molecules of tryptophan bound to it. Whenever the level of intracellular tryptophan drops, the repressor releases its tryptophan and becomes inactive, allowing the polymerase to begin transcribing these genes. The promoter includes two key blocks of DNA sequence information, the -35 and -10 regions highlighted in yellow.
The mechanisms by which specific Aene regulatory proteins control gene transcription in procaryotes.

(A) NEGATIVE REGULATION
- Bound repressor protein prevents transcription
- Bound repressor protein
- ADDITION OF LIGAND switches GENE ON by REMOVING REPRESSOR PROTEIN
- GENE OFF
- Removal of ligand switches gene on by removing repressor protein
- Inactive repressor

(B) POSITIVE REGULATION
- Bound activator protein promotes transcription
- Bound activator protein
- RNA polymerase
- ADDITION OF LIGAND switches GENE OFF by REMOVING ACTIVATOR PROTEIN
- GENE ON
- mRNA
- Protein
- Removal of ligand switches gene off by removing activator protein
- 5’ -> 3’
LacZ, encodes the enzyme B-galactosidase which breaks down the disaccharide lactose to galactose and glucose. Lactose addition increases the concentration of allolactose, an isomer of lactose, which binds to the repressor protein and removes it from the DNA. Glucose addition decreases the concentration of cyclic AMP; because cyclic AMP no longer binds to CAP, this gene activator protein dissociates from the DNA, turning off the operon.
Gene regulation in eucaryotes by transcription

1. Eucaryotic RNA polymerase II requires five general transcription factors (27 subunits in toto), whereas bacterial RNA polymerase needs only a single general transcription factor, the σ subunit. The stepwise assembly of the general transcription factors at a eucaryotic promoter provides, in principle, multiple steps at which the cell can speed up or slow down the rate of transcription initiation in response to gene regulatory proteins.

2. Eucaryotic cells lack operons—sets of related genes transcribed as a unit and therefore must regulate each gene individually.

3. Each bacterial gene is typically controlled by one or only a few gene regulatory proteins, but it is common in eucaryotes for genes to be controlled by many (sometimes hundreds) of different regulatory proteins. This complexity is possible because many eucaryotic gene regulatory proteins can act over very large distances (tens of thousands of nucleotide pairs) along DNA, allowing an almost unlimited number of them to influence the expression of a single gene.

4. A central component of gene regulation in eucaryotes is Mediator, a 24-subunit complex, which serves as an intermediary between gene regulatory proteins and RNA polymerase. Mediator provides an extended contact area for gene regulatory proteins compared to that provided by RNA polymerase alone, as in bacteria.

5. The packaging of eucaryotic DNA into chromatin provides many opportunities for transcriptional regulation.
The presence of an independent DNA-binding and a transcription-activating domains in the yeast gene activator protein GAL4. (Y2HT)
Chromatine Organisation

- Short region of DNA double helix
- "Beads on a string" form of chromatin
- 30-nm chromatin fibre of packed nucleosomes
- Section of chromosome in an extended form
- Condensed section of chromosome
- Entire mitotic chromosome

Distances: 2 nm, 11 nm, 30 nm, 300 nm, 700 nm, 1,400 nm
**Epigenetic**: Heritable changes in gene expression without a change in DNA sequence

**Eu- and Hetero-Chromatin**
Structure of Chromatin is determinant
Heterochromatin - densely packed regions of DNA - no gene expression

Associated with methylation of cytosine and histone modification – acetylation
Methylation, as a tool for the epigenetic control of gene expression

**DNA methylation:** cytosine methylation exerts negative effects on gene regulation via the involvement of other proteins that bind specifically to DNA when it is methylated. In mammals, methylated cytosines occur exclusively in CG sequences, and CG methylation are maintained during DNA replication by DNA methyltransferase 1 (DNMT1).

In plants:

1. Symmetrically methylation:
   - **by MET1**
     - CG
     - GC
     - CH$_3$-CG
     - CH$_3$-GC
     - CH$_3$-CNG
     - CH$_3$-GNC
     - CH$_3$-CAG
     - CH$_3$-CTG
     - (Methyltransferase 1)
   - **by CMT3**

2. Asymmetrically methylation:
   - **by DRM2**
     - CNN
     - GNN
     - CH$_3$-CNN
     - CH$_3$-GNN
     - CCC, CCA, CCT
     - CAA, CAC, CAT
     - CTT, CTC, CTA
     - (Domains Rearranged Methyltransferase 2)
     - CGG
     - CCG
     - CAG
     - CTG

**DRM2** is the major de novo DNA methyltransferase establishing new cytosine methylation marks in a process directed by siRNAs termed RNA-directed DNA methylation (RdDM). RdDM can methylate cytosines in all sequence contexts. In Arabidopsis, homozygous partial loss-of-function *met1* mutants are very affected. **DRM2 and CMT3** have redundant functions, plants lacking both functions have morphological defects.
Histone methylation

Histones associate with DNA to accomplish the chromatin assembly, forming protein-DNA structures known as nucleosomes. A nucleosome core particle includes ~146 bp of DNA wrapped almost twice around a core histone octamer (2 molecules each of Histones H2A, H2B, H3 and H4).

Core histones can be reversibly modified by acetylation, methylation, phosphorylation, ubiquitination or ADP-ribosylation and these modifications have consequences for gene activation, gene repression and chromosome replication.

Lysines (Lys, K) at the N-terminal ends of the core histones are the predominant sites of known regulatory modifications (K4, K9 and K27)

Methylation of H3K9 results in binding of repressor proteins, such as HP1 (Heterochromatin Protein 1) that helps to establish highly compacted and transcriptionally inactive regions of chromatin known as heterochromatin

**Histone Methyletransferases:**

SUVH2: H3K9, H4K20

KRYPTONITE (SUVH4): H3K9

Clr4 (fission yeast): H3K9

SIL1: H3K9

...
Specific DNA methylation ↔ specific Histone methylation (dependance and independance)

CG methylation provides distinct and direct information for a specific histone methylation.

**Active genes**
- H3K4
- H3K36
- H3K27 me3
- H3K79
- H4K20 me2,3
- Loss of C methylation

**Silent genes**
- H3K9
- H3K27 me1
- H3K27 me2
- C methylation
- H3K27
- H4K20me1
- C methylation

**Active chromatin**
- H3K4

**Silent chromatin**
- DNA-independent

Mathieu et al., 2005

**CG methylation**
- CG methylation → increases in **H3K9me2**
- Loss of CG methylation → increases in **H3K4** methylation but no effect on H3K9 Methylation.

**H3K27**
- H3K27 mono- and dimethylation mark **silent** heterochromatin independently of DNA methylation.
- CG hypermethylation characteristic of heterochromatin specifically prevented **H3K27me3** trimethylation.
Specific Histone methylation ↔ specific DNA methylation

Either DNA methylation is triggered by histone methylation or vice versa histone methylation is initiated after DNA methylation???
Most important ways of locally chromatin alteration:

1. Histone modifications
2. Nucleosome remodeling
3. Nucleosome removal
4. Nucleosome replacement

Gene activator proteins use all four of these mechanisms by attracting histone modification enzymes, ATP-dependent chromatin remodeling complexes, and histone chaperones to alter the chromatin structure of promoters they control.

In general terms, these local alterations in chromatin structure are believed to make the underlying DNA more accessible, thereby facilitating the assembly of the general transcription factors, Mediator, and RNA polymerase at the promoter. These modifications provide favorable interactions for the binding of a large set of proteins that read a histone code for transcription initiation.

The alterations of chromatin structure that occur during transcription initiation can persist for variable lengths of time. In some cases, as soon as the gene regulatory protein dissociates from DNA, the chromatin modifications are rapidly reversed. However, this altered chromatin structure can persist, even after the gene regulatory protein that directed its establishment has dissociated from DNA. In principle, this memory can extend into the next cell generation because, chromatin structure can be self-renewing.
Six ways in which eucaryotic gene repressor proteins can operate

(A) competitive DNA binding

(B) masking the activation surface

(C) direct interaction with the general transcription factors

(D) recruitment of chromatin remodeling complexes

(E) recruitment of histone deacetylases

(F) recruitment of histone methyl transferase
The stable inheritance of DNA methylation patterns can be explained by maintenance DNA methyltransferases. DNA methylation patterns, however, are dynamic during vertebrate development. Shortly after fertilization there is a genome-wide wave of demethylation, when the vast majority of methyl groups are lost from the DNA. This demethylation may occur either by suppression of maintenance DNA methyltransferase activity, resulting in the passive loss of methyl groups during each round of DNA replication, or by a specific demethylating enzyme. Later in development, new methylation patterns are established by several de novo DNA methyltransferases that are directed to DNA by sequence-specific DNA-binding proteins where they modify adjacent unmethylated CG nucleotides. Once the new patterns of methylation are established, they can be propagated through rounds of DNA replication by the maintenance methyltransferases.

DNA methylation of the promoter region of a gene or of its regulatory sequences can interfere directly with the binding of proteins required for transcription initiation. In addition, the cell has a repertoire of proteins that specifically bind to methylated DNA thereby blocking access of other proteins. One reflection of the importance of DNA methylation to humans is the widespread involvement of errors in this mechanism in cancer progression.

**Genomic Imprinting Is Based on DNA Methylation**
Mammalian cells are diploid, containing one set of genes inherited from the father and one set from the mother. The expression of a small minority of genes depends on whether they have been inherited from the mother or the father: while the paternally inherited gene copy is active, the maternally inherited gene copy is silent, or vice-versa. This phenomenon is called genomic imprinting.

**Epigenetic Mechanisms Ensure That Stable Patterns of Gene Expression Can Be Transmitted to Daughter Cells**