6 Principles of Gene Regulation
Why expression of genes is regulated?
⇒ To synthesize just the amount of protein needed under specific conditions

How does the cell know to express which gene at what level?
⇒ Intra- and extracellular communication
General Scheme of the Signal Transduction Pathway

Signal

⇒

Sensor

⇒

Signal transduction (one or more steps)

⇒

Gene regulation

⇒

Feedback
6.1 Regulation at the level of DNA

6.2 Regulation at the level of transcription

6.3 Regulation at the level of translation

6.4 Regulation by proteolysis
6.1 Regulation at the Level of DNA
1. Structural changes of the DNA by
   - supercoiling
   - bending
   - DNA looping
2. GATC methylation
3. Programmed DNA rearrangements
4. Decryptification
Does supercoiling influence transcription?

Experimental systems:

*in vitro:*
- Plasmids containing a variable number of superhelical turns
- *In vitro* transcription: quantification of the mRNA
- **Result:** some promoters exhibit superhelicity-dependent transcription

*in vivo:*
- Addition of gyrase inhibitors
- Analyses of *gyr* or *topA* mutants
The following stimuli influence the linking number:

1. High osmolarity
2. Anaerobic growth
3. Growth phase
4. Temperature

Open question:

How do these stimuli influence gene expression?
How can bending influence gene expression?

There are two different kinds of bends:
- Intrinsic bends
- Protein-induced bends
Intrinsic bends are induced by AT-tracts occurring at the correct spacing:

AATAGCATCGAAATAGGCCACGTTTTAG
The DNA Sequence of the virF Promoter Region of *Shigella flexneri*

GGAGCCTCCAGTCTGAAGGGGCTTTATGCGTT
CCGTATAGGATATTATGATGCTGGAGGTTTTTGC
GACAGTCCACTCTCTTTTCCGAACGATTTACACAGA
TATTGCTAAGAAAAAGTGAACACCATAATATGGTT
ATAGTCCCTTTGAGTGCAAAATACTTAGCTTGTT
GCACAGAGAAATAGAAGCTGCATAAGCTCTTTT
CTTCYYYYYYYYYGTAAATAAAGTTAAATATAGGA
AAAATTACTTTAAATCTATCTTAAATAACGGAAAA
TTTTTGTTATAACAATCACACTTTACAGAAATTTTCTT
AGTTACTCTGTAAACACTAAATATAGTTTGGTT
ATTCTGTGTGAATTTATG

**AT-tracts**

**H-NS binding sites**
The Bend of the \textit{virF} Promoter is Temperature-Dependent

## Temperature-Dependent Expression of virF-lacZ Fusions

<table>
<thead>
<tr>
<th>Strain</th>
<th>β-Galactosidase Activity at 30°C</th>
<th>37°C</th>
<th>Ratio 37/30°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild-type</td>
<td>833</td>
<td>4180</td>
<td>5.02</td>
</tr>
<tr>
<td>hns</td>
<td>7537</td>
<td>5781</td>
<td>0.77</td>
</tr>
</tbody>
</table>
DNA-binding Proteins Introducing Bends

**CAP:** Homodimer
- CAP-cAMP-DNA complex
- Two discontinuous kinks of about 45°
- Stimulates transcription

**IHF:** Heterodimer
- Sharp bends: >140°
- May repress or stimulate transcription

**FIS:** Homodimer
- Bends DNA by 90°
DNA Looping

- One or more protein species bind to two different sites on the DNA separated by a certain distance
- The proteins interact with each other and loop out the DNA located between them

**Examples:**
1. The LacI repressor
2. The AraC repressor
The LacI Repressor

The AraC Repressor

R Schleif (2000) TIG 16: 559
Bacterial Methyltransferases

Catalyze the transfer of a methyl group from SAM to
- cytosines at the C-5 position
- adenines at the N-4 or N-6 position

Most bacterial methyltransferases are involved in restriction-modification systems to protect from invading DNA
The Orphan Methyltransferase EcoDam

Plays a critical role in many bacterial pathways

- gene regulation
- mismatch repair
- DNA replication
- Nucleoid structure determination

A Lobner-Olesen (2005) Curr. Opin. Mic. 8: 151
J Casadesus (2006) MMBR 70: 830
Monomeric EcoDam Binds and Methylates the N-6 Position of the Adenine in the Sequence

\[ \text{mGATCCTAG} \]
Promoters Exhibiting Increased Transcription when GATC is Hemimethylated

<table>
<thead>
<tr>
<th>Gene</th>
<th>-35</th>
<th>Promoter</th>
<th>-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>dnaA P2</td>
<td>AGAA\text{GA} TCTCTTTGCGGCAGTTTTAGGC TATGAT C</td>
<td>\text{GATCAT}</td>
<td>\text{GATC}</td>
</tr>
<tr>
<td>glnS</td>
<td>TTGTCA GCCTGTCCCGCTTTTATAA \text{GATCAT}</td>
<td>\text{ATCGT}</td>
<td>\text{GATC}</td>
</tr>
<tr>
<td>trpR</td>
<td>CT\text{GATC} CGCACGTTTTATGATATGC TATCGT</td>
<td>\text{GATC}</td>
<td>\text{GATC}</td>
</tr>
<tr>
<td>tnpP Tn5</td>
<td>GGAACC TTTCCCGTTTTTCCAGA TCTGAT C</td>
<td>\text{GATC}</td>
<td>\text{GATC}</td>
</tr>
</tbody>
</table>
Pap Phase Variation Model

$pap = \text{pyelonephritis-associated pili produced by UPEC = uropathogenic } E. \ coli$

$paplp$: coregulator

$papBA$: pilin A and B

J Casadeus (2006) MMBR 70: 830
The Pap OFF- to ON-Phase Transition Mechanism
Programmed DNA Rearrangements Within the Chromosome

1. Duplication, amplification: Gene dosage effect

2. Inversion:
   - Turn on and off of genes
   - Synthesis of different proteins

3. Deletion:
   - Fusion of the two parts of a gene
Silent and Cryptic Genes

Silent genes: Are not expressed at a detectable level

**Mechanism:**
Involves a *silencer* and a protein binding to the silencer
**Example:** *bgl* operon

Cryptic genes: code for inactive proteins

**Mechanism:**
Genes inactive because of a point mutation
**Examples:** *ebg* and *ilvGM* operons
The \textit{bgl} Operon

$\textit{bgl} = \text{beta-glucosides}$
Naturally Occurring $\beta$-Glucosides

- **Cellobiose**
- **Arbutin**
- **Salicin**
The two H-NS complexes interact and form a DNA loop thereby preventing binding of the RNAP.

Model of Silencing of the $bgl$ Operon

H-NS binding sites

$\text{bgl}$
The ebg Operon

$ebg = \text{evolved beta-galactosidase}$

$ebgR$: repressor; 327 aa
$ebgA$: $\beta$-galactosidase; 1030 aa; hexamer
$ebgC$: unknown function
The *ilvGM* Operon

- *ilvGM*: cryptic; valine-resistance
- Frameshift mutation: 300 amino acids (inactive protein)
- Removal of the frameshift: 521 amino acids (active protein)
Biological Significance of Decryptification?

Decryptification acts at the level of the population
What happens if two isogenic *E. coli* strains, one prototrophic and the other auxotrophic are co-cultivated in a glucose-limited chemostat?

- The auxotrophic strain will win!
Why auxotrophy can be an advantage for the cell?

1. Metabolic products of anabolic pathways can be slightly toxic or interfere with the regulation of other pathways.

2. A protein not used by the cell can interact with other cellular components, e.g., a transport protein can compete with another transport protein for the limited number of sites at the inner membrane.
**Silencer**

**Silencer** = DNA sequence which drastically reduces the expression of a downstream operon

**Example:** H-NS binding sites sandwiching the promoter of the *bgl* operon
H-NS

- A major component of the bacterial nucleoid
- Binds preferentially to curved DNA
- Is able to generate bends in noncurved DNA
- Affects the expression of a large number of genes
- Is considered a general **negative** transcriptional regulator

DW Ussery (1994) Biochimie 76: 968
RM Williams (1997) FEMS Microbiol. L. 156: 175
Models to Explain Gene Repression

1. Transcriptional silencing: H-NS binding to DNA occludes RNAP access through a generalized effect on DNA compaction

2. Direct competition: Occurs between H-NS and the RNAP for overlapping binding sites near the promoter

3. Modification of DNA supercoiling
Transcriptional Silencing

Observation:
Pap pili are present at 37°C, but their synthesis is shut-off at temperatures below 26°C

Direct Competition

**Observation:**

H-NS

-35   -10  proVWX

Binding of H-NS prevents binding of the RNAP

C Ueguchi (1993) EMBO J. **12**: 1039
6.2 Regulation at the Level of Transcription

Predominant level of regulation involving in all cases the RNA polymerase
6.2.1 Components of the DNA-Dependent RNA Polymerase
Two enzymes able to synthesize RNA:
- DNA primase (d*naG)
- RNA polymerase (RNAP)

RNAP is responsible for the synthesis of:
- mRNA
- tRNA
- rRNA
- small RNAs = ncRNAs (non-coding)

Two different forms of the RNAP:
Core enzyme: $\alpha^I\alpha^II\beta\omega$; interacts non-specifically with DNA
Holoenzyme: $\alpha^I\alpha^II\beta\omega\sigma\ E\sigma$
Function of the different subunits:

α¹ and α² : Identical in sequence but different in location

- α¹ interacts with β
- α² interacts with β'
\( \alpha \) consists of two domains:

**N-terminal domain** \((\alpha_{NTD})\):
- Dimerization and interaction with \( \beta \) and \( \beta' \), respectively

**C-terminal domain** \((\alpha_{CTD})\):
- Tethered to the \( \alpha_{NTD} \) by a flexible linker can bind to DNA (UP element) and is the interaction target of a number of transcription factors
**β subunit:**
- Binds rNTPs
- Catalytic domain of the polymerase
- Binds rifampycin (initiation) and streptolydigin (elongation)
- Binds ppGpp

**β' subunit:**
- Involved in unspecific binding to DNA
- Involved in termination and antitermination

**ω subunit:**
**Hypothesis:** Functions in RNAP assembly
Sigma Factors Contribute to a Range of Functions

1. Promoter recognition
2. Serving as contact point for transcription regulators
3. Promote strand separation
4. Influence early phases of elongation
5. Influence promoter-proximal pausing

Assembly of the RNAP

$$\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\omega$$

- The $\alpha$ dimer forms the scaffold on which $\beta$ and $\beta'$ assemble
- The $\omega$ subunit assists $\beta'$ binding to the $\alpha_2\beta$ sub-assembly
- The overall architecture resembles that of a crab claw:
  - $\beta$ forms one pincer
  - $\beta'$ forms the other pincer
Sigma Factors Are Categorized Into Two Classes

1. The major *sigma-70* class
   - are encoded in all bacterial genomes
   - are subdivided into four broad phylogenetic subfamilies
   - recognize –35 and –10 region

2. The minor *sigma-54* class
   - present in ~60% of bacterial genomes
   - recognize –24 and –12 region

MS Paget (2003) Genome Biol. 4: 203
Region 2 binds to the $\beta'$ subunit and -10 region
Region 4 binds to the $\beta$ subunit and -35 region

Recognizes the -24 region

By Function Two Groups of Sigma Factors Are Distinguished

- Housekeeping sigma factors
- Alternative sigma factors
Number of Sigma Factors

*Mycoplasma genitalium* 1

*E. coli* 7

*B. subtilis* 17

*Streptomyces coelicolor* 63
The Sigma Factors of *Escherichia coli*:

<table>
<thead>
<tr>
<th>σ factors</th>
<th>Upstream recognition sequence (−35 region)</th>
<th>Number of spacer nucleotides</th>
<th>Downstream recognition sequence (−10 region)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ70</td>
<td>TTGACA</td>
<td>17±1</td>
<td>TATAAT</td>
<td>Housekeeping genes at exponential growth</td>
</tr>
<tr>
<td>σ5 (σ38) *</td>
<td>CCGGCG</td>
<td>17 ± 1</td>
<td>CTATACT</td>
<td>Stationary phase expression</td>
</tr>
<tr>
<td>σH (σ32)</td>
<td>TNtCNCCCTTGAA</td>
<td>13–17</td>
<td>CCCCATtTA</td>
<td>Heat shock genes</td>
</tr>
<tr>
<td>σE (σ24)</td>
<td>GAACCTT</td>
<td>16 (ATAAA)</td>
<td>TCTGAT</td>
<td>Extreme heat shock, extracytoplasmic stress</td>
</tr>
<tr>
<td>σF (σ28)</td>
<td>TAAA</td>
<td>15</td>
<td>GCCGATAA</td>
<td>Flagella synthesis and chemotaxis</td>
</tr>
<tr>
<td>σN (σ54)</td>
<td>ttGGcaca</td>
<td>4</td>
<td>ttGCA</td>
<td>Nitrogen-regulated genes</td>
</tr>
<tr>
<td>Fecl (σ19)§</td>
<td>AAGGAAAAT</td>
<td>17</td>
<td>TCCTTT</td>
<td>Ferric citrate transport genes, extracytoplasmic stimuli</td>
</tr>
</tbody>
</table>
# Sigma Factors of *Bacillus subtilis*: 17

<table>
<thead>
<tr>
<th>$\sigma$ factors *</th>
<th>Gene</th>
<th>Upstream recognition sequence (−35 region)</th>
<th>Number of spacer nucleotides</th>
<th>Downstream recognition sequence (−10 region)</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vegetative factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^B$ ($\sigma^{37}$)</td>
<td>sigB, rpoD</td>
<td>TTGACA</td>
<td>17 ± 1</td>
<td>TATAAT</td>
<td>Housekeeping genes, early sporulation</td>
</tr>
<tr>
<td>$\sigma^C$ ($\sigma^{12}$)</td>
<td>sigH, spoOH</td>
<td>(T$_1$)GG(T$_1$)TT(T$_3$)A</td>
<td>14</td>
<td>GGGTAT</td>
<td>General stress response</td>
</tr>
<tr>
<td>$\sigma^D$ ($\sigma^{28}$)</td>
<td>sigD, flaB</td>
<td>AAATC</td>
<td>15</td>
<td>TA(T$_1$)TCG(T$_1$)TT(T$_3$)TAA</td>
<td>Postexponential gene expression</td>
</tr>
<tr>
<td>$\sigma^I$ ($\sigma^{30}$)</td>
<td>sigI</td>
<td>TAAA</td>
<td>15</td>
<td>GCCGATAT</td>
<td>Chemotaxis, flagellar gene expression</td>
</tr>
<tr>
<td>$\sigma^L$ ($\sigma^{54}$)</td>
<td>sigL</td>
<td>TGGCACA</td>
<td>5</td>
<td>TTGCANNN</td>
<td>Expression of degradative enzymes</td>
</tr>
<tr>
<td><strong>Sporulation factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^E$ ($\sigma^{29}$)</td>
<td>sigE, spoLGB</td>
<td>(T$_1$)ATA(T$_3$)A(T$_3$)</td>
<td>11</td>
<td>CATACA(T$_3$)</td>
<td>Early mother cell gene expression</td>
</tr>
<tr>
<td>$\sigma^F$ ($\sigma^{spolAC}$)</td>
<td>sigF, spoLAC</td>
<td>GCAT(T$_3$)</td>
<td>15</td>
<td>GG(T$_1$)A(T$_3$)A(T$_3$)T</td>
<td>Early forespore gene expression</td>
</tr>
<tr>
<td>$\sigma^G$</td>
<td>sigG, spoLGC</td>
<td>G(T$_3$)AT(T$_3$)A(T$_3$)</td>
<td>18</td>
<td>CAT(T$_1$)TA</td>
<td>Late forespore gene expression</td>
</tr>
<tr>
<td>$\sigma^H$ ($\sigma^{77}$)</td>
<td>sigK, spoLVB:spoLlC</td>
<td>AC</td>
<td>17</td>
<td>CATANNTA</td>
<td>Early mother cell gene expression</td>
</tr>
</tbody>
</table>

* Alternative designations are given in parentheses.
The availability and activity of $\sigma$ factors are controlled by at least three types of regulatory factors:

- **Anti-$\sigma$ factors**: bind and inhibit their cognate $\sigma$ factors
- **Appropriators**: alter the activity of a specific RNAP holoenzyme
- **Proteolysis**: $\text{Pro-}\sigma \rightarrow \text{active } \sigma$ factor
Regulation of the anti-\(\sigma\) factors can occur by:

- Antagonistically by anti-anti-\(\sigma\) factors: SpoIIAA and RsbW
- By co-anti-sigma factors: RseB
- By proteolysis: RseA, RsiW
Anti-Sigma Factors

Anti-sigma factor =
Protein forming a complex with a given sigma factor thereby preventing its interaction with the RNAP core enzyme
Anti-sigma factors can be located in two different compartments:

- In the cytoplasm: AsiA
- Anchored in the cytoplasmic membrane: RseA
AsiA Acts First as an Anti-\(\sigma\) Factor
And Then as an Appropriator

- Encoded by phage T4
- The 90-amino-acid protein binds specifically to the 4.2 region of \(\sigma^{70}\) and thereby blocks interaction with the –35 region = acts as anti-\(\sigma\) factor
- Acts as appropriator by deploying RNAP to T4 middle promoters
AsiA and Rsd Cause Conformational Changes of $\sigma_4$

Inactivation of the RseA Anti-Sigma Factor by the Sequential Action of Three Proteases

RseA belongs to the group of ECF σ factors.

RseB acts as a co-anti-σ factor and stabilizes the interaction between RseA and σ^E.

The Family of ECF Sigma Factors

ECF = extracytoplasmic function

- In most cases, the genes coding for an ECF sigma factor are cotranscribed with a gene coding for a negative regulator
- Negative regulator = anti-sigma factor
- Most anti-sigma factors are bitopic inner membrane proteins:
  - N-terminal domain binds to the sigma factor
  - C-terminal domain senses the stress factor
Occurrence of ECF Sigma Factors in Different Species

<table>
<thead>
<tr>
<th>Species</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em></td>
<td>2</td>
</tr>
<tr>
<td>2 (heat shock; iron uptake)</td>
<td></td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>7</td>
</tr>
<tr>
<td><em>Corynebacterium diphtheriae</em></td>
<td>7</td>
</tr>
<tr>
<td><em>Mycobacterium tuberculosis</em></td>
<td>10</td>
</tr>
<tr>
<td><em>Mesorhizobium loti</em></td>
<td>16</td>
</tr>
<tr>
<td><em>Streptomyces coelicolor</em></td>
<td>41</td>
</tr>
</tbody>
</table>
The Concept of Anti-Anti Sigma Factors

Anti-anti sigma factor =
Protein forming a complex with the anti-sigma factor when not phosphorylated

Two examples:
SpollAA
RsbW
The Anti-Anti Sigma Factor SpoIIAA

anti-sigma factor

anti-anti sigma factor
Modulation of $\sigma^B$ Activity

- **Anti-anti-sigma factor** + **Anti-sigma factor** + $\sigma^B$
- **Dephosphorylation**
  - ATP depletion: C-, PO$_4$-starvation, O$_2$-limitation
  - Phosphatase: heat, ethanol, NaCl, acid shock
- **Anti-anti-sigma factor** + **Anti-sigma factor** + E $\sigma^B$
The Five Promoter Elements

Always present:
1. The -35 element: TTGACA
2. The -10 element: TATAAT

May be present in addition alone or in combination:
3. The UP element
4. The extended -10 element
5. The discriminator region
DNA Elements and RNAP Modules That Contribute to Promoter Recognition

UP element:
- Centered at about – 42 and – 52 (A+T-rich DNA sequence)
- Stimulates transcription by recruiting $\alpha$CTD
- Consensus sequence:
  $$\text{AAA(A/T)(A/T)TA(A/T)TTTTnnAAAA}$$
Extended –10 motif: TGnTATAAT

- Makes contact with region 3.0
- -35 region dispensable
- Stimulates transcription by increasing the rate of association and by stabilizing the promoter complex
DNA Elements and RNAP Modules That Contribute to Promoter Recognition

**Discriminator region:**
- From – 6 to – 4
- Optimal sequence: GGG

SP Haugen (2008) PNAS 105: 3292
Anti-\(\alpha\) Factor

*B. subtilis* protein Spx:

- It sequesters activator and \(\sigma\)-region-4 binding surfaces on \(\alpha\)CTD
- Can activate or inhibit transcription at specific promoters during oxidative or disulfide stress
6.2.2 Initiation of Transcription
The different stages of initiation of transcription:

1. Free diffusion to a DNA molecule
2. Linear diffusion to the promoter
3. Specific binding to the promoter and formation of a closed complex:
   - RPc1: RNAP protects DNA from $-55$ to $-5$
   - RPc2: RNAP protects DNA from $-55$ to $+20$
4. Isomerization: formation of the open complex involves opening of bp $-9$ to $+2$
5. Formation of the ternary complex: binding of the first rNTP
6. Abortive initiation: short oligomeric products are synthesized and released
7. Promoter clearance
RNA Polymerase and its Interaction at Promoters with Canonical –35 and –10 Hexamers

Structural Transitions During the Steps of Transcription Initiation

β flap:
Rudder:

k
Structural Transitions During the Steps of Transcription Initiation

\[ \beta \text{ flap:} \]
\[ \text{Rudder:} \]
Structural Transitions During the Steps of Transcription Initiation

The Pathway of Transcription Initiation at Bacterial Promoters
Average transcription rate of *E. coli* RNAP:
50 -100 nucleotides per sec

The elongation cycle comprises three basic steps:

1. Binding of a template-complementary nucleoside triphosphate (NTP) into the active site

2. Chemical reaction of the RNA chain 3'-OH with the NTP $\alpha$-PO$_4$ catalyzed by a pair of bound Mg$^{2+}$ ions, resulting in one NMP addition to the RNA and liberation of pyrophosphate

3. Translocation of the nucleic acid assemblage
Stabilization of the elongation complex:
- Binding of dsDNA to the RNAP
- Formation of a 9 bp RNA/DNA hybrid
- ~ 5 nucleotides are bound in a protein channel
Modifications of the Processive RNAP:

1. **Pausing sites**: The transcribing complex pauses for a variable time before it resumes transcription.

2. **Termination sites**: The transcription complex is terminated.

3. **Arrest sites**: The transcribing complex is arrested and unable to resume elongation without additional factors.
**Pausing sites:**

When does pausing occur?  
What is its biological function?

**Parameters involved:**

1. Capacity of the nascent transcript to form a stem-loop structure
2. High GC content can inhibit strand opening
3. Transcription factors NusA and NusG  
   - **NusA**: plays a role in synchronizing the rates of transcription and translation  
   - **NusG**: enhances the rate of transcription
4. Effector molecule ppGpp: some pauses are enhanced, some reduced, some unchanged
Two proteins identified able to reactivate the arrested complex:

**GreA:** induces the cleavage of short fragments (2-3 nucleotides) from the 3' end of the mRNA

**GreB:** induces the cleavage of longer fragments (2-18 nucleotides) from the 3' end of the mRNA

Convert the active site of RNAP from „polymerizing“ to „ribonucleolytic“
The Transcription Elongation Factors GreA and GreB

- Prevent transcription elongation arrest
- Facilitate promoter escape
- Consist of a globular domain and a coiled-coil domain with an acidic tip
- Bind in the RNAP secondary channel
Arrested RNA Polymerase and GreB

Proofreading Steps During Elongation

Two mechanisms:

**Error prevention**: Template-dependent rNTPase activity; converts rNTPs to the corresponding rNDP

**Error correction**: Incorrectly incorporated rNMP is removed by hydrolytic cleavage (pyrophosphorolysis)

Each mRNA molecule is translated about 40 times