2 Organization of the Bacterial Chromosome

2.1 Principles to compact the bacterial chromosome

2.2 Organization of the bacterial chromosome into genes, repetitive sequences and regulator sequences

2.3 Rearrangement of the chromosome
**Bacterial Chromosomes: Size, Number and Topology**

<table>
<thead>
<tr>
<th>Bacterial species</th>
<th>Size [Mb]</th>
<th>Number</th>
<th>Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carsonella rudii</td>
<td>0.18</td>
<td>1</td>
<td>circular</td>
</tr>
<tr>
<td><em>M. genitalium</em></td>
<td>0.59</td>
<td>1</td>
<td>circular</td>
</tr>
<tr>
<td>Borrelia burgdorferi</td>
<td>0.97</td>
<td>1</td>
<td>linear</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>4.72</td>
<td>1</td>
<td>circular</td>
</tr>
<tr>
<td><em>Brucella melitensis</em></td>
<td>2.1+0.15</td>
<td>2</td>
<td>circular</td>
</tr>
<tr>
<td><em>Agrob. tumefaciens</em></td>
<td>2.8+2.1</td>
<td>2</td>
<td>circular, linear</td>
</tr>
<tr>
<td>Sorangium cellulosom</td>
<td>13.1</td>
<td>1</td>
<td>circular</td>
</tr>
</tbody>
</table>

**Multireplicon genome**
Characteristics of bacterial chromosomes:

1. Most chromosomes are circular and 0.5 to 10 Mbp in length
2. Most chromosomes contain prophages, IS elements and transposons = mobilome
3. Are fluid due to Horizontal Gene Transfer (HGT) and lateral gene transfer
4. Are condensed by supercoiling and histone-like proteins
Linear chromosomes:

- One central origin of replication called oriC
- Replication proceeds bidirectionally towards the two ends
- Ends of the linear chromosomes: Telomers
Distribution of Circular and Linear Chromosomes in Prokaryotes
Polyploidy

*E. coli, B. subtilis:* 2 – 4 chromosomes/cell
*Deinococcus sp.:* 4 – 10 copies
*Sinorhizobium meliloti* (bacteroids): ~24 copies
*Buchnera aphidicola:* ~120 copies
*Epulopiscium* sp.: 50,000 – 120,000 copies

**Factors Leading to Polyploidy**

Amplification provides resources to support
- Rapid growth and division
- Cell specialization and adaptation
- May enhance repair of genetic lesions
- Is allied with metabolic adaptation

JE Mendell (2008) *PNAS* **105:** 6730
Extreme Polyploidy in *Epulopiscium*

**Habitat:**
*Epulopiscium* sp. type B, which occurs in the intestinal tract of the unicornfish *Naso tonganus*

**Size:**
Lengths of 200–300 m and widths of 50–60 m

**Genome:**
*Epulopiscium* cells contain tens of thousands of copies of a fully replicated, 3.8 Mb genome
The *Epulopsium* Life Cycle

# Very Small Genes

## Examples for small genes:

<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>Number of aa</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. aureus</em></td>
<td>hld</td>
<td>26</td>
<td>cytolytic δ-lysin</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>spoVM</td>
<td>26</td>
<td>sporulation</td>
</tr>
<tr>
<td><em>E. faecalis</em></td>
<td>iad</td>
<td>22</td>
<td>inhibitor</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>mccA</td>
<td>7</td>
<td>inhibits protein synthesis</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>23S RNA</td>
<td>5</td>
<td>erythromycin resistance</td>
</tr>
</tbody>
</table>
How to identify small genes?

Theoretical considerations:

1. Canonical SD upstream

2. The N-terminal Met is formylated

3. The amino acid sequence of the peptide is identical with its gene sequence
How to identify small genes?

Experimental considerations:

1. Deletion of the coding sequence removes the peptide

2. Construction of a translational fusion with a reporter gene, e.g. \textit{lacZ}
2.1 Principles to Compact the Bacterial Chromosome

The *E. coli* chromosome is compacted nearly 1000-fold in a nucleoprotein complex = nucleotid.

Occupy about one fifth of the cell‘s volume and located in the center of the cell.
Two principles:

1. Arrangements of the DNA in superhelical domains; size of the superhelical loops: 10 – 100 kb
   Number of superhelical domains in \textit{E. coli}: 12 – 400

2. Histon-like proteins
2.1.1 Superhelicity
Description of Superhelicity

\[ L = T + W \]

**L** = linking number: is the number of times the two strands of a **closed DNA duplex** cross over each other

**T** = twisting number: represents the rotation of one strand about the other and represents the total number of turns of the duplex

**W** = Writhing number: represents the turning of the axis of the duplex in space

relaxed molecule: \( W = 0 \), and the linking number equals the twist

\[ L \text{ for DNA in the B conformation} = \frac{N}{10.5} \]
Topoisomerases

Definition:
Bind to DNA and make transient breaks in DNA, pass DNA strands through the breaks and then reseal the DNA.

They modulate three topological forms of DNA:
1. The linking number (Lk)
2. Catenanes
3. Knots

Type-1 topoisomerases: Alter Lk in steps of one
Type-2 topoisomerases: Alter Lk in steps of two
## DNA Topoisomerases of *Escherichia coli*

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Function (mechanism)</th>
<th>Subunit structure</th>
<th>Subunit molecular mass (kDa)</th>
<th>Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topoisomerase I</td>
<td>DNA relaxation (type-1)</td>
<td>α</td>
<td>105</td>
<td>topA</td>
</tr>
<tr>
<td>DNA gyrase (or topoisomerase II)</td>
<td>DNA supercoiling and relaxation (type-2)</td>
<td>α₂β₂</td>
<td>105 (α) 95 (β)</td>
<td>gyrA (α), gyrB (β)</td>
</tr>
<tr>
<td>Topoisomerase III</td>
<td>Decatenation (type-1)</td>
<td>α</td>
<td>73.2</td>
<td>topB</td>
</tr>
<tr>
<td>Topoisomerase IV</td>
<td>Decatenation; DNA relaxation (type-2)</td>
<td>α₂β₂</td>
<td>83.7 (α) 70.2 (β)</td>
<td>parC (α), parE (β)</td>
</tr>
</tbody>
</table>
Topoisomerases are Divided into Two Types

1. **Type-1 enzymes**: Alter the linking number **in steps of one** by passing DNA through single-stranded breaks; ATP-independent

2. **Type-2 enzymes**: Alter the linking number **in steps of two** by passing DNA through a double-strand break; ATP-dependent
Negative and Positive Superhelical Turns

Relaxed DNA
$Lk = 200$

$\Delta Lk = -2$

Negative supercoils
$Lk = 198$

$\Delta Lk = +2$

Positive supercoils
$Lk = 202$
Action of the Two Types of Topoisomerasases

Reverse Gyrase

- Introduces positive superhelical turns in DNA molecules in an ATP-dependent reaction
- Present only in hyperthermophilic microorganisms living above 70°C
- Is composed of a C-terminal topoisomerase fused to an N-terminal helicase-like domain

2.1.2 Histon-Like Proteins

Four major histone-like proteins:

1. HU
2. H-NS
3. IHF
4. FIS
5. Dps
Compaction of DNA Through Nucleosomes

Histone core of nucleosome

Linker DNA of nucleosome
Protein-Dependent Constraint of DNA Supercoiling

HU = **Heat Stable Nucleoid Protein**

- Heterodimer in *enterobacteria* (hupA and hupB)
- Binds to DNA with low or any specificity
- Binds with high affinity to bent DNA
- 30,000 dimers per cell
- Modulates the topology of the chromosome
- Participates in several cellular processes such as
  - the initiation of oriC-dependent DNA replication
  - transposition of phage Mu and Tn10
  - DNA inversion in *Salmonella typhimurium*
H-NS = Histone-Like Nucleic Structuring Protein

- 15.4 kDa protein encoded by *hns*
- Consists of two functional domains:
  - N-terminal dimerization (oligomerization) domain
  - C-terminal DNA-binding domain
- Binds specifically to *intrinsically curved* DNA
- Modulates the synthesis of a large number of gene Products; in most cases H-NS inhibits target gene expression
- Induced by cold shock
H-NS Compacts DNA Molecules

H-NS-Mediated Repression of the virF Promoter of *Shigella flexneri*

IHF = Integration Host Factor

- Heterodimeric protein encoded by *himA* and *himD*
- DNA-binding consensus: CAATNTATTGAATT
- Contacts the minor groove of DNA
- Bends DNA by more than 140°
- Involved in phage lambda site-specific recombination and packaging
- Affects gene expression serving either as activator or inhibitor
FIS = Factor for Inversion Stimulation

- Homodimeric protein encoded by gene *fis*
- DNA-binding to degenerate consensus: 
  \[(G/T)nYRnn(A/T)nYRnn(C/A)\]
- Induces DNA-bending (40° to 90° bend angles)
- Stimulates DNA inversion (Hin, Gin, Cin)
- Stimulates integration and excision of phage λ DNA
- Involved in *oriC*-directed DNA-replication
2.2 Organization of the Bacterial Chromosome into Genes, Repetitive Sequences and Regulator Sequences

*E. coli* chromosome:
- 90% of its nucleotide sequence could encode protein
- The non-coding genome is densely packed with regulatory signals for transcription initiation and termination
**Genes**

Average size of a bacterial gene: 1 kb

**Pseudogenes:**
- Code for dysfunctional proteins with homology to known proteins (unable to function properly)
- *M. leprae*: 27% of the genome occupied by pseudogenes

**Silent genes:**
- Not expressed at a detectable level
- Need a mutational event to become expressed
2.2.1 Repetitive Sequences

- Repetitive sequences without coding capacity
- Repetitive sequences with coding capacity
2.2.1.1 Repetitive Sequences Without Coding Capacity
Repetitive sequences may consist of

- Simple homopolymeric tracts of a single nucleotide
- Several multimeric classes of repeats in small or large number

Questions:

- Structure ?
- Function ?
The Three Families of Repetitive Sequences Identified in *Enterobacteriae*

1. REP = repetitive extragenic palindromic

2. ERIC = enterobacter repetitive intergenic consensus

3. Chi ($\chi$)
REP

- 38-bp imperfect palindrome
- Present always outside of coding regions, either at the end of a transcriptional region or within transcriptional units after genes

**Functions:**
- Termination of transcription (hair-pin structure)
- Increases mRNA stability; example: *malG* – 2x REP – *malF*; stabilizes *malG*
- Binding site for HU and DNA gyrase
- Binding site of DNAP I; involved in repair?

- Two derivatives: RIB and BIME
REP (Repetitive Extragenic Palindromic) = PU (Palindromic Unit)

- 40-bp long imperfect palindromes
- Part of mRNAs
- Located at the end of the transcript or between cistrons
BIME Elements

BIME = Bacterial Interspersed Mosaic Element

Distribution of short palindromic repeats on the *E. coli* K12 chromosome

Green lines: REP
Stars: BIME-1
Red dots: BIME-2
A Model for the Role of IHF Binding at BIME-1
ERIC

- 126 bp long with an inverted repeat in its center which may lead to the formation of a cruciform structure
- Stabilisation of the upstream DNA
- Might serve as a regions of homologous recombination
- Distribution as an RNA intermediate?
Chi ($\chi$)

- 5'-GCTGGTGG-3'
- Stimulates homologous recombination
- One chi site per 5.5 kb on the average
2.2.1.2 Repetitive Sequences with Coding Capacity
2.2.1.2.1  \textit{rrn}-Loci
Most bacterial species contain more than one *rrn* operon which code in most case for 16S – 23S – 5S rRNA:

*E. coli* and *S. typhimurium*: each seven copies

*B. subtilis*: 10 copies

*Mycobacterium*

Slow growing species have one
Fast growing species two copies
Location in Minutes of the Seven *rrn* Operons on the Chromosome of *Escherichia coli* K12
2.2.1.2.2 $rhs$-Loci

$rhs = \text{recombination hot spot}$
Organization of a Prototypical Rhs Element

Overall length: 8 - 10 kb

2.2.1.2.3 Mobile Elements

1. IS elements
2. Transposons
3. Transposing phages
Barbara McClintock

1902 - 1992

Nobel price: 1983
Maize from Middle America
Maize from Middle America
Bacteria can harbour three types of transposable elements:

- IS elements
- Transposons
- Transposing phages
Organization of IS Elements

Insertion sequence, IS1

IR Transposase gene IR

5′ GGTGATGCTGCCAACCTTACTGAT 3′
3′ CCACTACGACGGTTGAATGACTA 5′

5′ ATCAATAAGTTGGAGTCATTACC 3′
3′ TAGTTATCAACCTCAGTAATGG 5′
Characteristics of IS Elements

1. 0.7 – 2.5 kb in length
2. Most have terminal inverted repeats, 10 – 40 bp in length
3. Most generate short direct repeats during integration at target sites, 2 – 14 bp in length
4. More than 700 IS elements described, divided in about 20 families based on transposasase homology
5. Most of them code for just one protein, the transposase
Transposition of IS Elements

Insertion of IS element into chromosomal DNA

Target site

Cut

Chromosomal DNA

Cut

Inserted IS element

Gaps filled by DNA polymerase, DNA ligase

Host DNA

New DNA

Duplicated target site sequence
The *E. coli* K12 chromosome contains eight different IS elements:

<table>
<thead>
<tr>
<th>IS</th>
<th>Length (bp)</th>
<th>Copies/Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1</td>
<td>768</td>
<td>6-8</td>
</tr>
<tr>
<td>IS2</td>
<td>1,327</td>
<td>6</td>
</tr>
<tr>
<td>IS3</td>
<td>1,258</td>
<td>5</td>
</tr>
<tr>
<td>IS4</td>
<td>1,426</td>
<td>1</td>
</tr>
<tr>
<td>IS5</td>
<td>1,221</td>
<td></td>
</tr>
<tr>
<td>IS30</td>
<td>1,221</td>
<td></td>
</tr>
<tr>
<td>IS150</td>
<td>1,443</td>
<td></td>
</tr>
<tr>
<td>IS186</td>
<td>1,338</td>
<td></td>
</tr>
</tbody>
</table>
Function of IS elements:

1. Insertion mutations: knockouts, gene expression
2. Deletions, inversions
3. Transpositions (conservative, replicative)
4. Cotransposition
Principles of Transposition Frequency Regulation

1. Transcriptional repressor
2. Translational inhibitor (antisense RNA)
3. Translational frameshifting
4. Stability of the transposase
**Transposons:** Two classes

**Simple transposons:** Tn3 family

- IR
- *tnpA*  
- *tnpR*
- Antibiotic resistance gene(s)
- IR

**Composite transposons:** Tn5, Tn10

- IS element
- Antibiotic resistance gene(s)
- IS element

**Conjugative transposons**
The Simple Transposon Tn3

Transposon, Tn3

4,957 bp

- tnpA
- tnpB
- bla

Transposase
Resolvase
β-lactamase

Left inverted repeat (38 bp)

mRNAs

Right inverted repeat (38 bp)
The Composite Transposon Tn10

Transposon, Tn10

1,400 bp 9,300 bp 1,400 bp

IS10L
Inverted repeats of IS element

Tetracycline resistance gene ($Tc^R$)

IS10R
Inverted repeats of IS element

Inverted IS elements
The Transposing Phage Mu

a) Phage DNA present in virus particles

- Mu genome
- Invertible G segment
- Left-inverted repeat (IR-L)
- Right-inverted repeat (IR-R)

Host DNA

b) Prophage DNA

- Mu genome
- Invertible G segment

E. coli chromosome
- IR-L
- IR-R
- 5 bp direct repeat
2.2.1.2.4 Bacterial Retrons

Retron = A prokaryotic genetic element that produces multicopy DNA covalently linked to RNA (msDNA) by reverse transcriptase.

Discovered 1981 as satellite DNA in *Myxococcus xanthus*.

Later in clinical isolates of *E. coli*.
Generalized Version of a msDNA-RNA Compound

Retrons are composed of three genes:

1. msd: msDNA
2. msr: msRNA
3. ret: RT
Biosynthetic Pathway of msDNA Synthesis

msDNA exists in several hundreds copies per cell
Biological Function of msDNA

Unknown

Observation:
Cells producing msDNA exhibit an elevated mutation rate

Mechanism: unknown
2.2.1.2.5 CRISPR Loci

CRISPR = Clusters of Regularly Interspaced Short Palindromic Repeats

- Discovered 1987 in *E. coli*
- Regarded as defense systems against invading phage and plasmid DNA
- Widespread among eubacteria (~40%) and Archaea (~90%)
Typical Structure of a CRISPR Locus

Repeats: 21 to 48 bp
Spacers: 26 to 72 bp; show homology to phages and plasmids
Loci typically contain 50 units up to 300 units

CRISPRs Acquire Phage-Derived Spacers That Provide Immunity

The *Streptococcus thermophilus* CRISPR1 Locus

Newly Acquired Spacers in Phage-Resistant Mutants
Origin of the Spacer Sequences

1. Spacers are derived from preexisting sequences, either chromosomal or within transmissible genetic elements such as phages and conjugative plasmids
2. The extrachromosomal elements fail to infect the specific spacer carrier strain
3. About 65% of the spacers correspond to sequences in phages and conjugative plasmids, the remaining 35% to chromosomal sequences

Spacers Exhibit Homology to Genes Involved in

- Plasmid transfer
- DNA replication
- Phage assembly
- Replication partitioning
- Phage integration and excision
# CRISPRs from *E. coli* Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Gene</th>
<th>Element</th>
<th>Activity</th>
<th>Alignment&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECOR42</td>
<td><em>traI</em></td>
<td>Plasmid F</td>
<td>Helicase</td>
<td>gttcccgctgcgtgtatatgacgacaagagag</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gttcccgctgcgtgtatatgacgacaagagag</td>
</tr>
<tr>
<td>ECOR44</td>
<td>Unannotated</td>
<td>Phage P1</td>
<td>Unknown</td>
<td>cttttggcaagccaggattgtgaacacattaccgt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>cttttggcaagccaggattgtgaacacattaccgt</td>
</tr>
<tr>
<td>ECOR47</td>
<td><em>darB</em></td>
<td>Phage P1</td>
<td>Methylase</td>
<td>gctgtggcggaggcaacaggcaatcccgc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gctgtggcggaggcaacaggcaatcccgc</td>
</tr>
<tr>
<td>ECOR49</td>
<td><em>resD</em></td>
<td>Plasmid F</td>
<td>Resolvase</td>
<td>atcgactatatggcccatcaggctttgcagaaac</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>atcgactatatggcccatcaggctttgcagaaac</td>
</tr>
</tbody>
</table>

Two major questions:

1. By which mechanism the spacer regions inhibit gene expression?

2. By which mechanism spacer regions are generated?
Formation of New CRISPR Units

Random RNA recombination and reverse transcription OR Reverse transcription with random copy choice

Integrate (COG1518?)

Homologous recombination with genomic CRISPR region

dsDNA with CRISPRs and target-derived spacers

genomic DNA with new target-derived spacer
Putative Simplified Model of CRISPR Action
The Endoribonuclease Cas6 of *Pyrococcus furiosus*

2.3 Rearrangement of the Chromosome

Chromosomes can be rearranged by

- Homologous recombination between repetitive sequences
- By sequence-specific recombination
2.3.1 Amplification
**Tandem duplications**: Observed with many bacterial species

**Streptomyces**: Amplify certain chromosomal sequences

**S. fradiae**: 

- Amplifiable unit = 8.3 kb plus 2.2 kb direct repeat = 10.5 kb
- Can be amplified to up to 500 copies = 5.2 Mbp!
2.3.2 Deletions
Effect of small deletions:

can

1. Alter the gene dosage

2. Alter the gene product

3. Alter the pattern of gene expression
**Gene dosis:**

*B. subtilis:* Some strains harbor nine *rrn* operons

Homologous recombination between two adjacent loci results in the deletion of one *rrn* operon and of the DNA in between the two *rrn* operons
Gene product:

*Neisseria gonorrhoeae*: *opa* (opacity) genes

CTCTT repeat

**Mechanism**: Expression ON - OFF
Control Model for the Synthesis of Opa Protein
Slipped Mispairing During Replication

\[
\text{AATCTAGTATATA} \Rightarrow \text{TTAGGATCATATATGTGC}
\]

**AG**

**TT**

**C**

**GA**

\[
\text{AATC} \quad \text{TATATA}
\]

\[
\text{TTAGGATCATATAT}
\]

**INSERTION OF TA**

\[
\text{AATCCCTAGTATATA} \quad \text{TA}
\]

\[
\text{TTAG} \quad \text{ATATGTGC}
\]

**G**

**T**

**A**

**TA**

\[
\text{TC} \quad \text{TATACACG}
\]

\[
\text{AATATGTGC}
\]

**DELETION OF UNPAIRED TA**
Gene expression: Examples

*B. subtilis*: Synthesis of active sigma-K factor

*Anabaena*: Formation of heterocysts
DNA Rearrangement Creates an Active $\textit{sigK}$ Gene in the Mother Cells of \textit{Bacillus subtilis}

**Anabaena: Formation of heterocysts**


  - 55-kb deletion by recombination between two direct repeats: fusing $nifS$ to $rbcL$ and $rbcS$
Deletion of Two DNA Segments

VEGETATIVE CELL

HETEROCYST

11 kb

55 kb

11 kb circle

55 kb
2.3.3 Inversions

1. To regulate gene expression

2. To synthesize two different gene products
Examples:

1. The fimbrial operon of *E. coli* and phase variation of *Salmonella*
2. The G-segment of phage Mu and the C-segment of phage P1