central dogma + regulation

- tsx initiation control by transcription factors (TF)
- tsl initiation control by sRNA and RNA-binding proteins
- tsx termination control by sRNA and anti-terminators
- control of mRNA and protein degradation

Coupled to environmental signals

transcriptional initiation and termination

tsx init control by activators, repressors
Topic 2: Transcription Initiation Control

A. Mechanisms of tsx initiation in bacteria

1. Components:

- core enzymes of RNA polymerase:
  - sigma factor:

- B. subtilis has ~20 $\sigma$-factors (include sporulation, competence, …)
- generally, more complex the life style of organism, more sigma factors

• E. coli has 6 different $\sigma$-factors

- core promoter recognition sequences

- substitution of $\sigma$-factors recognize different set of promoters

<table>
<thead>
<tr>
<th>Factor</th>
<th>-35 Sequence</th>
<th>Separation</th>
<th>-10 Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^{70}$</td>
<td>TTGACA</td>
<td>16-18 bp</td>
<td>TATAAT</td>
</tr>
<tr>
<td>$\sigma^{32}$</td>
<td>CCCTTGAA</td>
<td>13-15 bp</td>
<td>CCCGATNT</td>
</tr>
<tr>
<td>$\sigma^{54}$</td>
<td>CTGAGA</td>
<td>6 bp</td>
<td>TTGCA</td>
</tr>
<tr>
<td>$\sigma^{28}$</td>
<td>CTAAA</td>
<td>15 bp</td>
<td>GCCGATTA</td>
</tr>
</tbody>
</table>
• E. coli’s σ-factors:
  
  – σ^{70} (\text{rpoD}): “house-keeping” genes (70-80% of genome) 
  ~1000 copies in growing cells, mostly occupied 
  
  – σ^{0} (\text{rpoS}): stress response (~70 genes) 
    post-txn control: tsl increased at low temp and high osmolarity 
    proteolysis inhibited upon carbon starvation and high temp 
    optimal recognition seq same as σ^{70}, but specificity different 
    uses anti- σ^{70} factor to block σ^{70}-RNAP association 
  
  – σ^{32} (\text{rpoH}): heat shock response (50-100 genes) 
    rapidly degraded; transiently stabilized by unfolded proteins 
    mRNA activated for translation at high temp 
    turns on chaperones and proteases 
    
  – σ^{54} (\text{rpoN}): nitrogen starvation response 
    activated when ammonia (preferred nitrogen source) is low 
    turns on genes to utilize alternative nitrogen sources 
  
  – σ^{F} (\text{fliA}): flagella biosynthesis 

• core promoter recognized by σ^{70}-factor: 
  
  canonical promoter has fuzzy motif 

• consensus sequence: 
  
  TTGACA_{17nt} \quad TATAAT 

occurrence of fuzzy promoter motifs in random sequences: 

• 3 out 6 matches in -35 region: \[
\frac{6!}{3! \cdot 3!} \cdot 0.25^3 \cdot 0.75^3 = 13\% 
\]

• degeneracy in spacing (16-18bp): 3 

• 4 out 6 matches in -10 region: \[
\frac{6!}{4! \cdot 2!} \cdot 0.25^4 \cdot 0.75^2 = 3.3\% 
\]

⇒ at given position in the genome, motif occurrence probability 

\approx 13\% \times 3 \times 3.3\% \approx 1.3\% 

or one occurrence every ~ 80bp, i.e., everywhere!
1. RNA polymerase recognition of promoters

- **core promoter** recognized by $\sigma^{70}$-factor:
  - Consensus sequence: TTGACA TATAAT
  - 17nt

- **strong promoters** (e.g., rRNA genes):
  - AT-rich (positions -40 to -60)

- **other**: extended -10
  - TGNTATAAT
  - Specificity determined mostly with the help of activators/repressors -- later

2. RNA polymerase-promoter interaction

- $\sigma^{70}$ bends DNA at -35; facilitates interaction with upstream activators
- Stabilizes open complex
- The $\beta$ and $\sigma$ domains of $\sigma^{70}$ are crucial for DNA binding
- Sigma N-terminus controls DNA binding
- N-terminus blocks DNA binding in holoenzyme
- DNA binding domains
- N-terminal region
- DNA displaces N-terminus in open complex
- N-terminal region
• pathway from initiation to elongation

recognition of -10/-35 regions
opening of -10 region (by twist-induced torque?)
stabilization of open complex

3. Quantitative aspects

• coverage size of RNAp

• tsx init rates: max ~100/min; typical 1/min

• abundance of RNAp

~ 5,000 core enzymes
~ 1/3 sigma factors
do the numbers make sense?
• core enzymes needed for tsx
  – rRNA genes: $6000\text{nt} \times 7 \text{ copies/\text{genome}} / 40\text{nt/RNAP} \sim 1000$ core enzymes
  – house-keeping genes: one firing per min, 1 min to cover 2500 bp (at 40nt/sec)
    ➞ one operon per enzyme ➞ 1000 core enzymes
  – multiplicity of 2.7 copies per cell (at 30min doubling) $\sim 5500$ core enzyme

• sigma factor usage
  – 1000 operons * 1 sec loading time/60 sec elongation time $\sim 20$ $\sigma$
  – multiplicity of 2.7: 50 $\sigma$
  – expt: 70% of RNAP do not release sigma right after initiation
    ➞ why so many $\sigma$s?

• available RNAP holoenzyme conc $\sim 30\text{ nM}$ [McClure, 1983]
  $\approx 0.5 \sim 1\mu\text{M}$ [Klumpp & Hwa, PNAS 2008]

• typical binding constants and rates

\[
\begin{align*}
\text{R} + \text{P} & \overset{K_1}{\underset{k_2}{\rightleftharpoons}} \text{I}_1 \overset{k_{-2}}{\underset{K_3}{\rightleftharpoons}} \text{I}_2 \overset{\text{RAP}}{\rightarrow} \\
\text{R} & \overset{k_2}{\underset{1}{\rightleftharpoons}} \text{P} \\

K_1 / K_{m} &= 10^{-3} \sim 1 \Rightarrow \bar{K}_1 = \sum_{n \neq j}^{N} e^{(C_j - G_n)/k_BT} = \frac{N \cdot K_1}{K_{m}} = 10^{4} \sim 10^{7} \text{ nM} \\
\text{– promoter binding typically very weak, i.e., } [\text{RNAP}] / \bar{K}_1 \ll 1 \\
\text{– opportunity for regulation, e.g., boost promoter binding probability}

k_2 &= 10^{-3} \sim 10^{-1} \text{ sec}^{-1} \\
\text{– fast end need not be faster} \\
\text{[cf: search kinetics]} \\
\text{– another opportunity for regulation}
\end{align*}
\]
4. TF-RNAP interaction

- Recruitment:

  e.g., CRP (activated by cAMP; aka CAP)

-- Class I CRP sites (-61.5, -71.5, -82.5, -92.5, -102.5)

-- Class II CRP sites (-41.5)
-- mechanism of activation? activator bypass experiments:

➔ glue-like attraction between CRP and α-CTD

(bacterial two-hybrid system)

-- repression via promoter exclusion
e.g., Lac repressor (inactivated by lactose, IPTG)

dimer of Lac repressor
• minimal unit for DNA binding
• large portion of molecule for ligand binding
  and allosteric control of the DNA-binding domain
• wt Lac tetramerizes
• Allosteric mechanisms of activation
  – NtrC (activated by phosphorylation under low nitrogen level): can activate $\sigma^{54}$ from 1-2 kbp away; has ATPase activity

• DNA looping
  – tight repression by lac tetramers
  – strong regulation by AraC