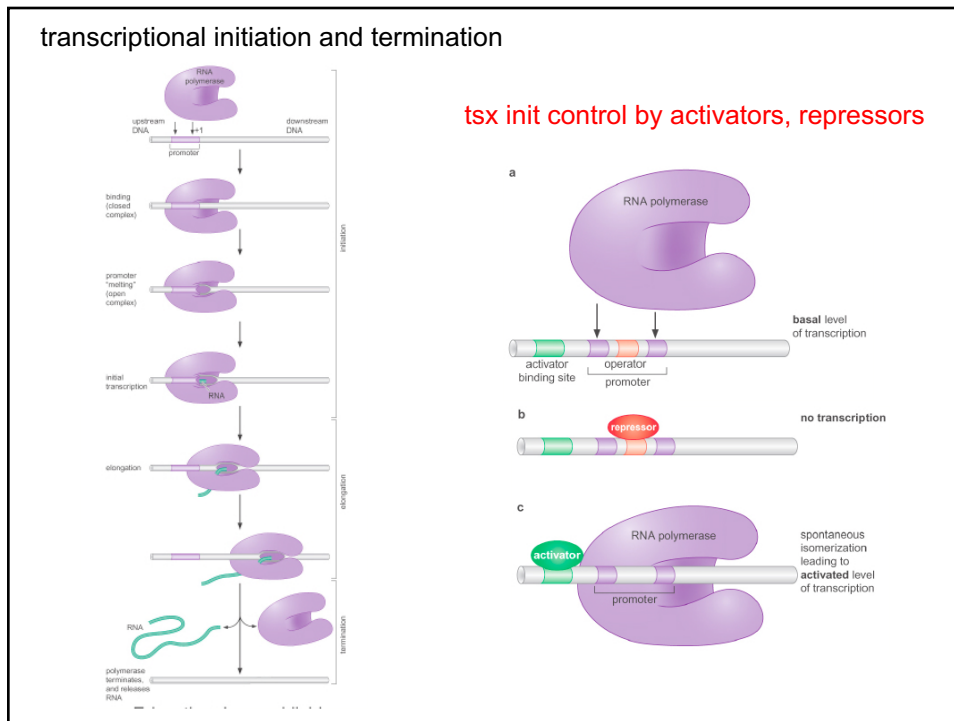


1



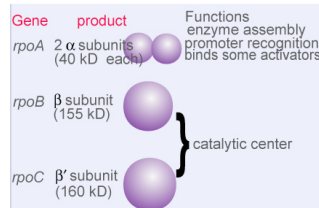
2

## Topic 2: Transcription Initiation Control

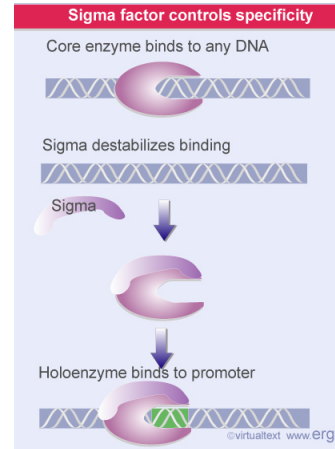
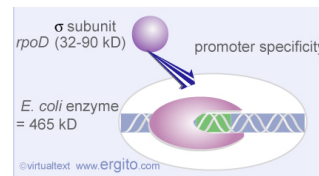
### A. Mechanisms of tsx initiation in bacteria

#### 1. Components:

- core enzymes of RNA polymerase:



- sigma factor:



3

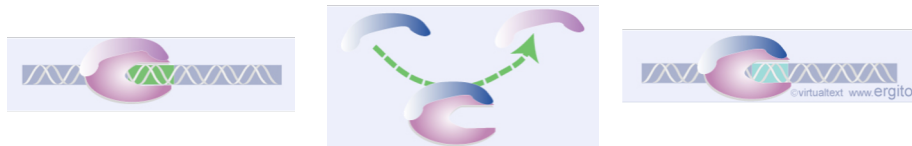
- *E. coli* has 6 different  $\sigma$ -factors

Gene	Factor	Use
<i>rpoD</i>	$\sigma^{70}$	general
<i>rpoS</i>	$\sigma^S$	stress
<i>rpoH</i>	$\sigma^{32}$	heat shock
<i>rpoE</i>	$\sigma^E$	heat shock
<i>rpoN</i>	$\sigma^{54}$	nitrogen
<i>fliA</i>	$\sigma^{28}$ ( $\sigma^F$ )	flagellar

- core promoter recognition sequences

Factor	-35 Sequence	Separation	-10 Sequence
$\sigma^{70}$	TTGACA	16-18 bp	TATAAT
$\sigma^{32}$	CCCTTGAA	13-15 bp	CCCGATNT
$\sigma^{54}$	CTGGNA	6 bp	TTGCA
$\sigma^{28}$ ( $\sigma^F$ )	CTAAA	15 bp	GCCGATAA

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



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

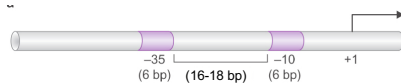
4

• *E. coli*'s  $\sigma$ -factors:

- $\sigma^{70}$  (*rpoD*): "house-keeping" genes (70-80% of genome)
  - ~1000 copies in growing cells; mostly occupied
- $\sigma^S$  (*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  $\sigma^{70}$ , but specificity different
  - uses anti-  $\sigma^{70}$  factor to block  $\sigma^{70}$ -RNAP association
- $\sigma^{32}$  (*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
- $\sigma^E$  (*rpoE*): heat shock response
  - induced by unfolded proteins in cell envelope
  - via cleavage of anti  $\sigma^E$ -factor
  - directs the expression of genes that restore envelope integrity
- $\sigma^{54}$  (*rpoN*): nitrogen starvation response
  - activated when ammonia (preferred nitrogen source) is low
  - turns on genes to utilize alternative nitrogen sources
- $\sigma^F$  (*fliA*): flagella biosynthesis

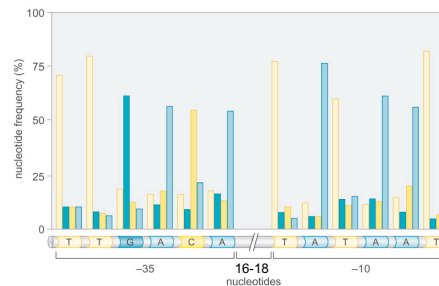
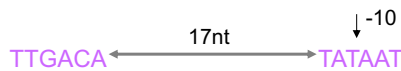
5

• core promoter recognized by  $\sigma^{70}$ -factor:



canonical promoter has fuzzy motif

• consensus sequence:



occurrence of fuzzy promoter motifs in random sequences:

• 3 out 6 matches in -35 region:  $\binom{6}{3} \cdot 0.25^3 \cdot 0.75^3 \approx 13\%$

• degeneracy in spacing (16-18bp): 3

• 4 out 6 matches in -10 region:  $\binom{6}{4} \cdot 0.25^4 \cdot 0.75^2 \approx 3.3\%$

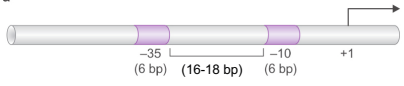
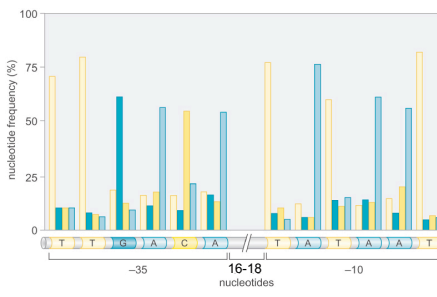
→ at given position in the genome, motif occurrence probability

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

or one occurrence every ~ 80bp, i.e., everywhere!


6


- core promoter recognized by  $\sigma^{70}$ -factor: canonical promoter has fuzzy motif

- consensus sequence:
 

$\xleftarrow{17\text{nt}}$ 
 $\downarrow$  -10

TTGACA
TATAAT
- strong promoters (e.g., rRNA genes):
 

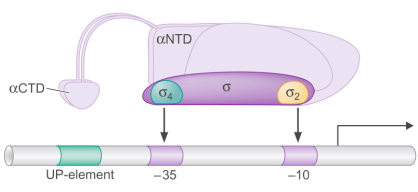
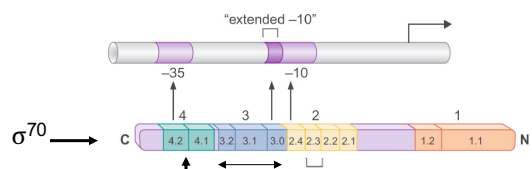
AT-rich (positions -40 to -60)
- other: extended -10
 

TGNTATAAT

specificity determined mostly with the help of activators/repressors -- later

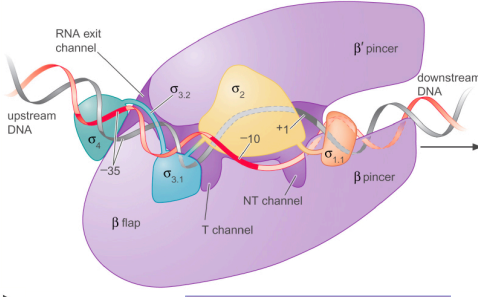
7

## 2. RNAP-promoter interaction

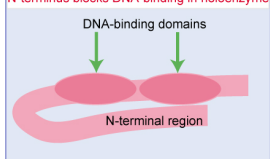
bends DNA at -35; facilitates interaction with upstream activators

stabilize open complex




Sigma N-terminus controls DNA-binding

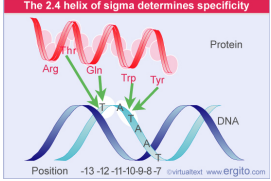
N-terminus blocks DNA-binding in holoenzyme



DNA displaces N-terminus in open complex

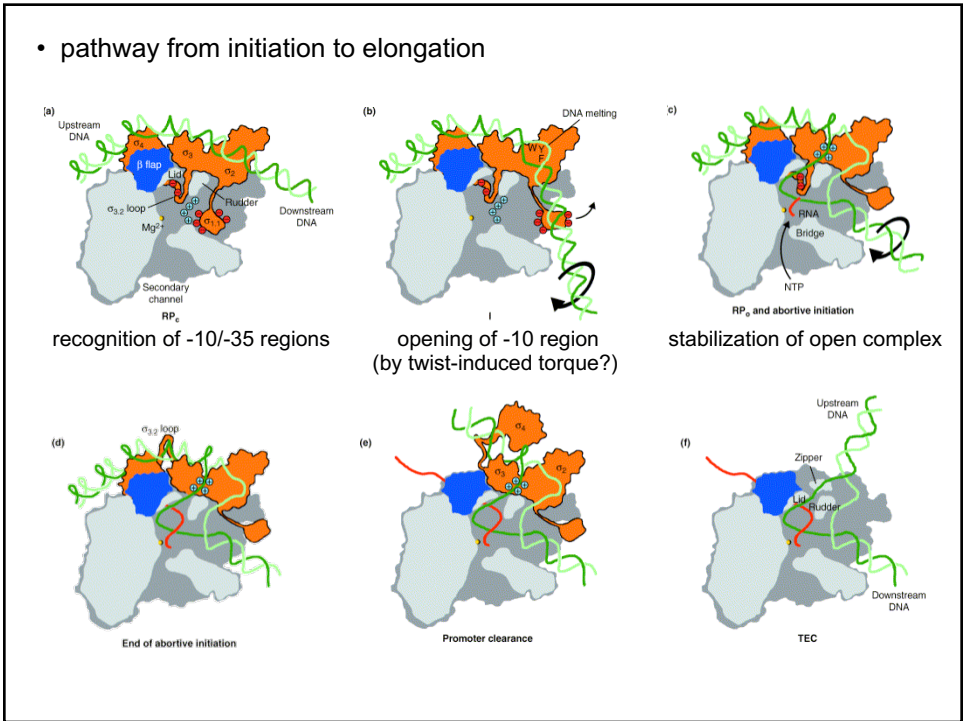


The 2.4 helix of sigma determines specificity



Position -13 -12 -11 -10 -9 -8 -7

8



9

### 3. Quantitative aspects

- coverage size of RNAP
- abundance of RNAP

**RNA polymerase changes size at initiation**

**Initiation complex** contains sigma and covers 75-80 bp

-50 -40 -30 -20 -10 1 +10 +20 +30

**Initial elongation complex** forms at 10 bases, may lose sigma, and loses contacts from -35 to -55

-50 -40 -30 -20 -10 1 +10 +20 +30

**General elongation complex** forms at 15-20 bases and covers 30-40 bp

-50 -40 -30 -20 -10 1 +10 +20 +30

©virtualltext www.ergifo.com

**All RNA polymerase is bound to DNA**

500-1000 core enzymes at loose complexes

500-1000 holoenzymes at loose complexes

500-1000 holoenzymes at promoter complexes

~2500 core enzymes engaged in transcription

©virtualltext www.ergifo.com

~ 5,000 core enzymes  
~ 1/3 sigma factors

**do the numbers make sense?**

10

- core enzymes needed for tsx
  - rRNA genes: 6000nt \* 7 copies/genome / 40nt/RNAP ~ 1000 core enzymes
  - house-keeping genes: one firing per min, 1 min to cover 2500 bp (@ 40nt/sec)
    - one operon per enzyme → 1000 core enzymes
  - multiplicity of 2.7 copies per cell (@ 30min doubling) ~ **5500 core enzyme**
- sigma factor usage
  - 1000 operons \* 1 sec loading time/60 sec elongation time ~ 20  $\sigma$
  - multiplicity of 2.7: **50  $\sigma$**
  - expt: 70% of RNAP do not release sigma right after initiation
    - why so many  $\sigma$  s?

11

- available RNAP holoenzyme conc ~ 30 nM [McClure, 1983]  
 $\approx 0.5 \sim 1 \mu\text{M}$  [Klumpp & Hwa, PNAS 2008]
- typical binding constants and rates
 

(rapid equilibria) (slow)

$$\text{R} + \text{P} \xrightleftharpoons{K_1} \text{I}_1 \xrightleftharpoons[k_{-2}]{k_2} \text{I}_2 \xrightleftharpoons{K_3} \text{RP}_O$$

(slow) (rapid equilibria)

$$K_1 / K_{ns} = 10^{-3} \sim 1 \Rightarrow \tilde{K}_1 = \sum_{n \neq j}^N e^{(G_1 - G_{ns})/kT} = \frac{N \cdot K_1}{K_{ns}} = 10^4 \sim 10^7 \text{ nM}$$
  - promoter binding typically very weak, i.e.,  $[\text{RNAP}] / \tilde{K}_1 \ll 1$
  - opportunity for regulation, e.g., boost promoter binding probability

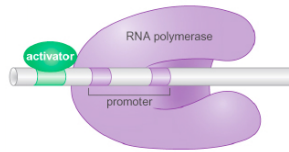
$k_2 = 10^{-3} \sim 10^{-1} \text{ sec}^{-1}$

  - fast end need not be faster [cf: search kinetics]
  - another opportunity for regulation

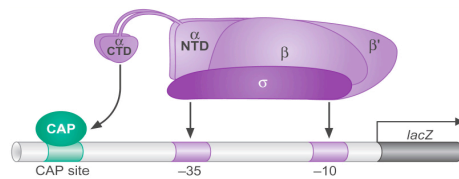
12

#### 4. TF-RNAP interaction

- Recruitment:

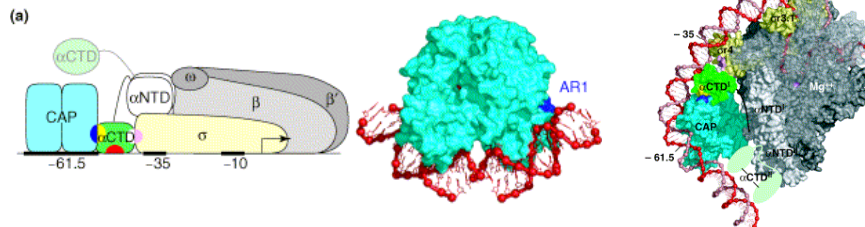


e.g., CRP (activated by cAMP; aka CAP)

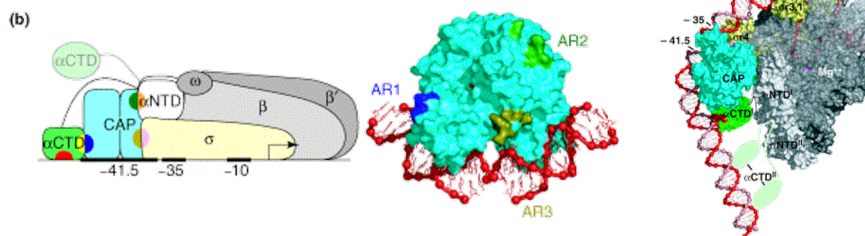


13

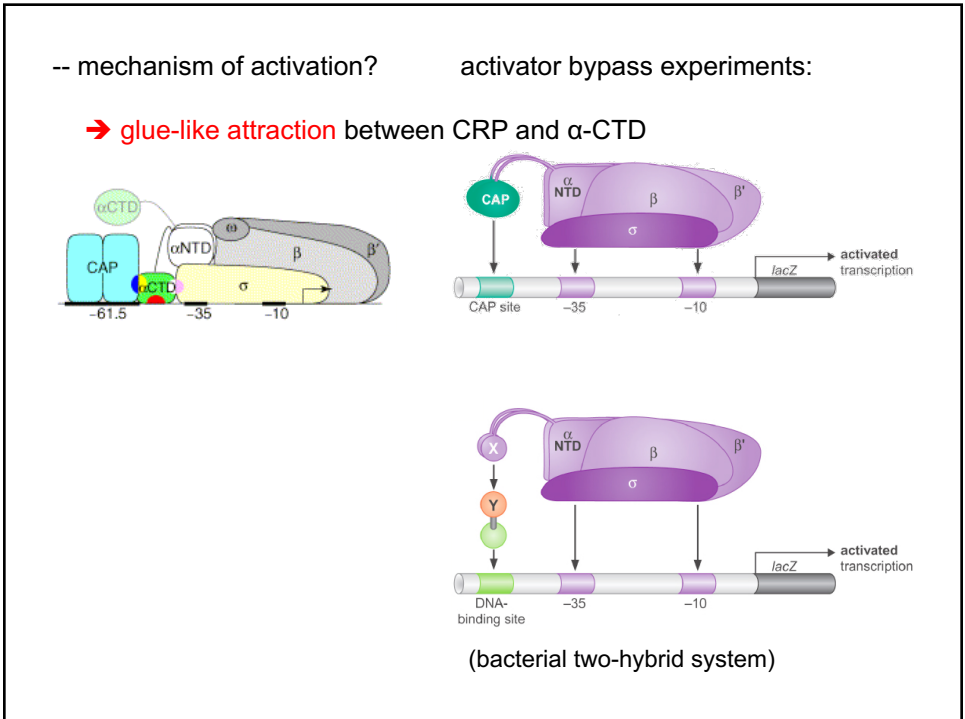
-- Class I CRP sites (-61.5, -71.5, -82.5, -92.5, -102.5)



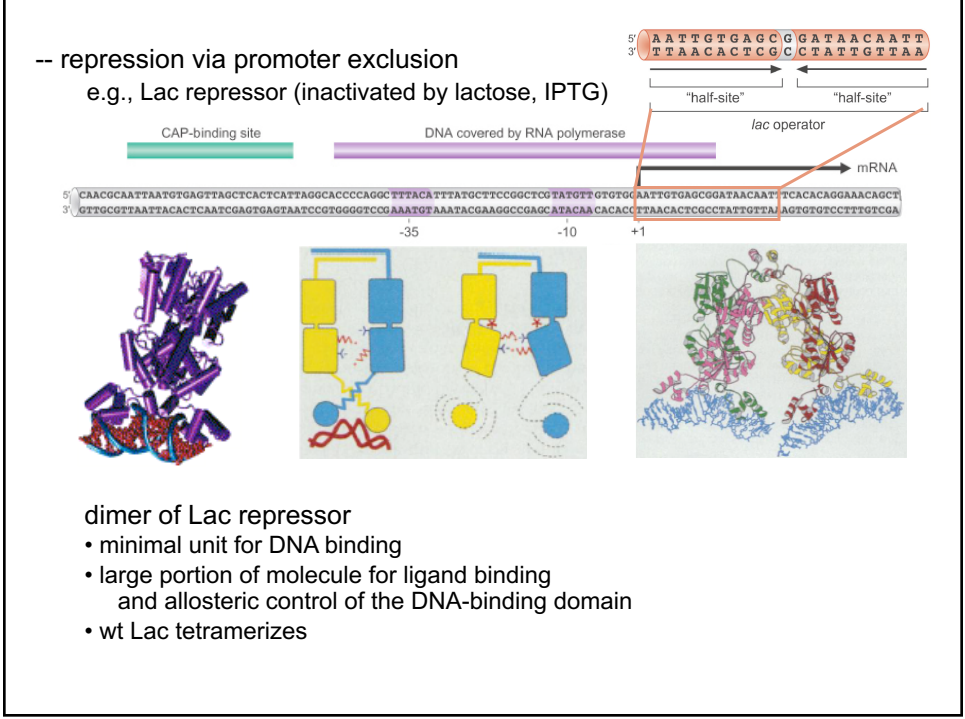
-- Class II CRP sites (-41.5)



14



15

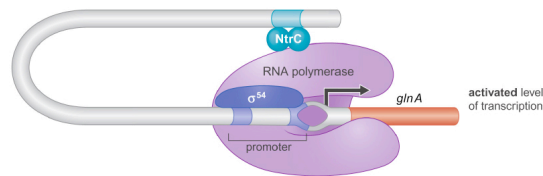


16

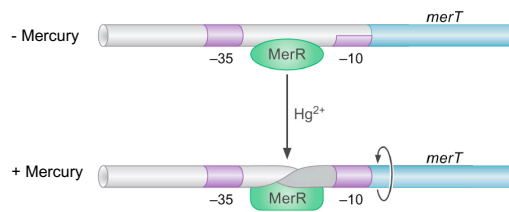


• Allosteric mechanisms of activation

- NtrC (activated by phosphorylation under low nitrogen level): can activate  $\sigma^{54}$  from 1-2 kbp away; has ATPase activity



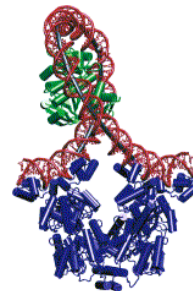
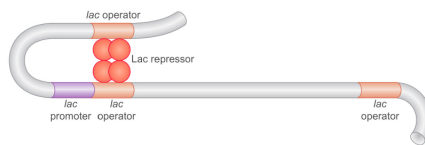
- MerR: activate  $\sigma^{70}$ -mediated *tsx* by twisting promoter DNA



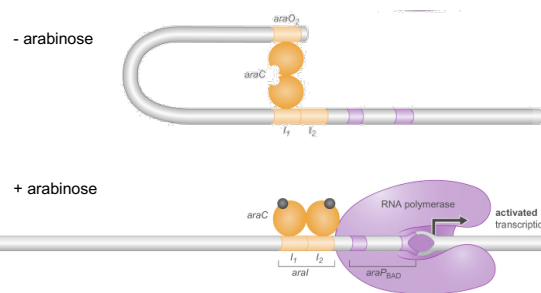
17

• DNA looping

- tight repression by lac tetramers



- strong regulation by AraC



18