

# MB&B 301b – Principles of Biochemistry II – Spring 2016

Based on lectures by Prof. Christian Schlieker and Prof. Karla Neugebauer  
James Diao

## Table of Contents

---

Lecture 1: Molecular Basis of Heredity.....	3
Lecture 2: Prokaryotic Genetics & Gene Structure .....	7
Lecture 3: Genetic Code.....	15
Lecture 4: DNA Replication I .....	23
Lecture 5: DNA Replication II .....	27
Lecture 6: DNA Modification & Repair.....	35
Lecture 7: Recombination & Transposition .....	43
Lecture 8: Transcription in Bacteria .....	53
Lecture 9: RNA Structure & Chemistry .....	59
Lecture 10: Catalytic RNA .....	67
Lecture 11: Protein Synthesis I.....	75
Lecture 12: Protein Synthesis II.....	81
Lecture 13: Prokaryotic Gene Expression I .....	87
Lecture 14: Prokaryotic Gene Expression II .....	91
Lecture 15: Regulation of Bacteriophage Lambda .....	97
Lecture 16: Cloning & Genomic Analysis .....	103
Lecture 17: Eukaryotic Genome I .....	113
Lecture 18: Eukaryotic Genome II .....	119
Lecture 19: Eukaryotic Transcription.....	125
Lecture 20: Eukaryotic RNA Processing .....	131
Lecture 21: Eukaryotic Translation.....	139
Lecture 22: Tumor Viruses & Oncogenes .....	146
Lecture 23: Development .....	146
Lecture 24: Immunology.....	146



# Lecture 1: Molecular Basis of Heredity

## A. The Scientific Process

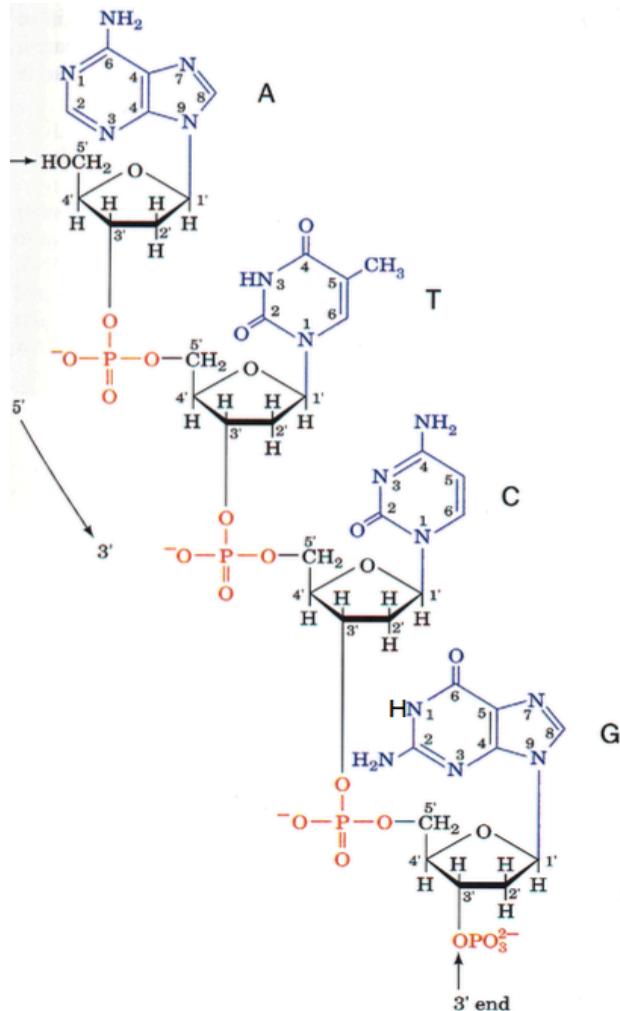
- a. Data is the direct output of an experiment; Interpretation is the conclusions you draw.
- b. Negative experimental outcomes can be controls – useful when properly conducted.

## B. Discovery of DNA as the genetic material

- a. Griffith first discovered transforming principle; that virulent smooth bacteria could transform rough benign bacteria.
- b. Avery's experiment demonstrate that DNA is the transforming principle.
  - i. Inactivated by DNases
  - ii. Insensitive to proteases
  - iii. Purified material had same molecular mass and elemental proportions of DNA

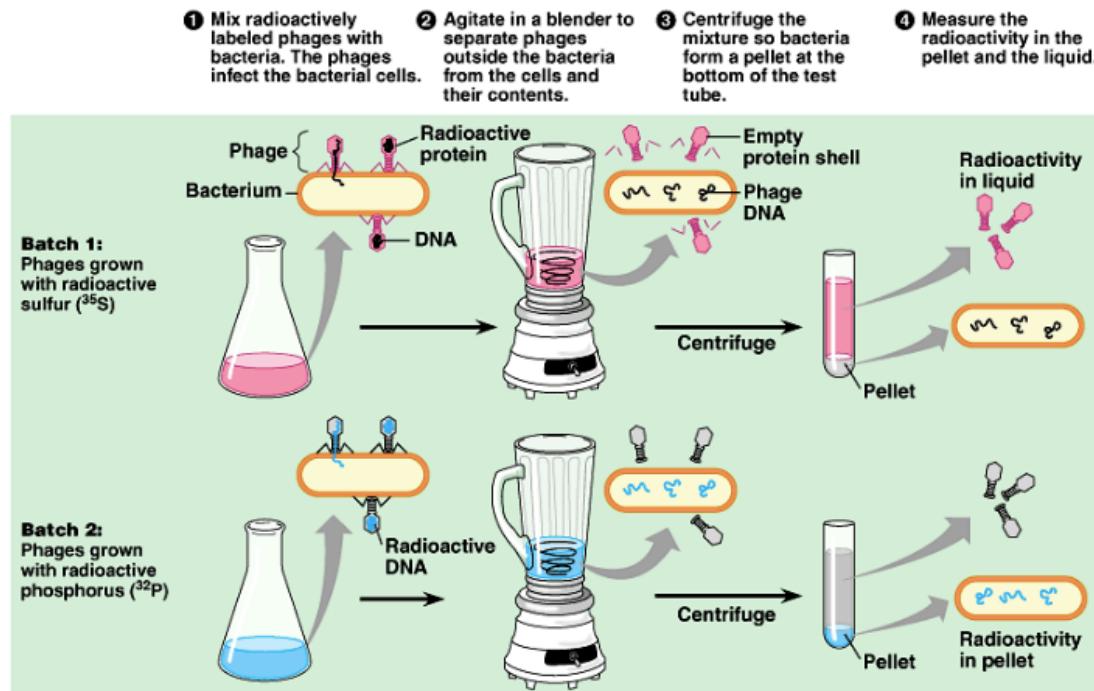
## C. Chemistry of DNA

- a. DNA: nitrogenous base (heterocyclic, planar) + ribose (5-sugar) + phosphate
- b. Nitrogenous base attached at 1' position via glycosidic linkage (note: C—N)
- c. Deoxyribose has 2' H, ribose has 2' OH.
- d.  $P_i$  connects sugars via 5'-3' phosphodiester bond.



#### D. Hershey-Chase “blender” experiment

- a. Proved that genes are made of DNA.
- b. E. coli good for molecular biology - large population, rapid generation, haploid genome.



(b) The experiment showed that T2 proteins remain outside the host cell during infection, while T2 DNA enters the cell.

Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

#### E. X-ray Diffraction of DNA Fibers (Wilkins & Franklin)

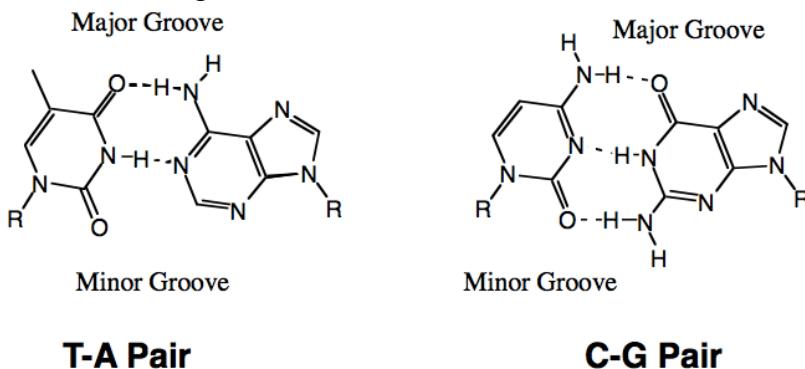
- a. Monoclinic – 2-fold axis of symmetry implies 2 antiparallel strands.
- b. Helical cross implies double helix.
- c. Strong reflection at 10<sup>th</sup> layer line (distance of 3.4 Å) suggests 10 units and 3.4 Å per turn.  
Actual helical repeat is 10-11 base pairs with helical diameter 20 Å.

#### F. Chargaff's Ratios for Genomic DNA

- a. Data shows that A/T and C/G are almost nearly 1; asymmetry between (A+T)/(C+G)
- b. Base composition is constant across different tissues, species, environment, etc.

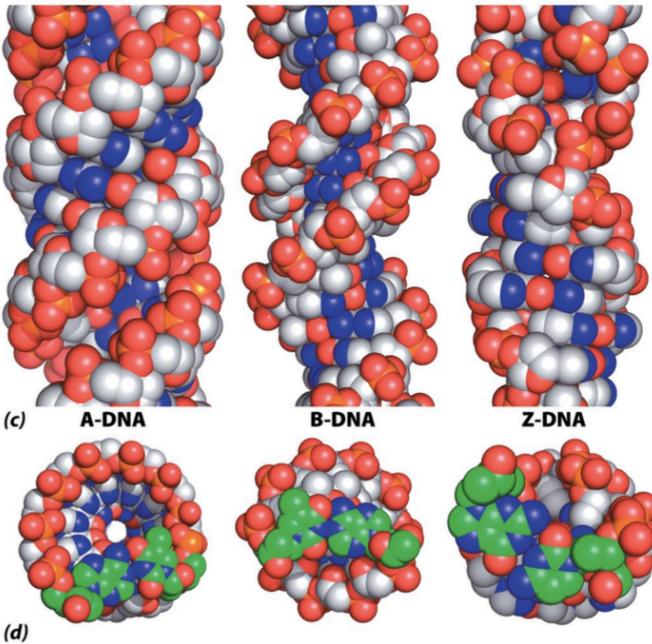
#### G. DNA Structure Base Pairing

- a. Watson & Crick started off with the wrong tautomeric forms of the bases (enol instead of keto). Corrected by Donague.
- b. Asymmetry of R groups creates a major groove opposite the R groups.  
The minor groove is on the same side.



## H. DNA Structure and Forms

- a. A and B forms are right-handed. Z form is left-handed.
- b. B form is relevant form of DNA. Wide/deep major groove, narrow深深 minor groove.
- c. A form is relevant form of RNA (some DNA). Deep major groove, shallow minor groove.
- d. Substantially different grooves; impacts function of DNA/RNA



## I. B-form DNA helix:

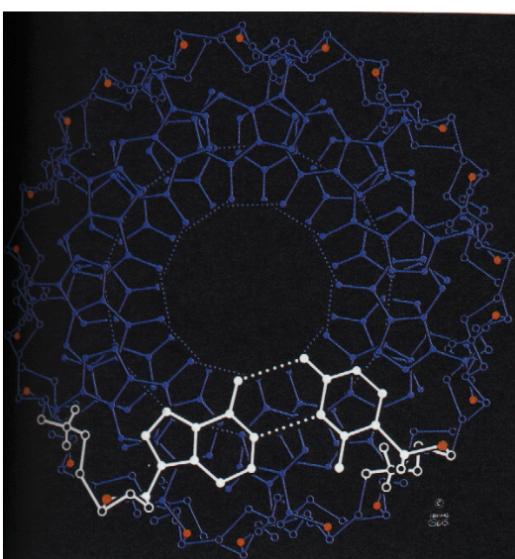
- a. Bases perpendicular to helical axis and ribose plane.
- b. Base pairs arranged at center of helix.

### A-form DNA/RNA helix:

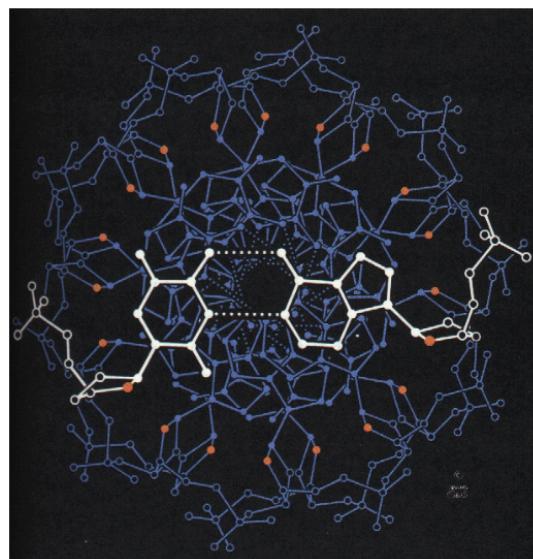
- c. Extremely deep major groove.
- d. Base pairs very displaced and tilted from perpendicular plane of helical axis.
- e. Broad vs. narrow groove? Helical RNA is recognized in the broad minor groove.

A-form  $\longleftrightarrow$  B-form driven by sugar puckering changes

### A-form DNA/RNA helix

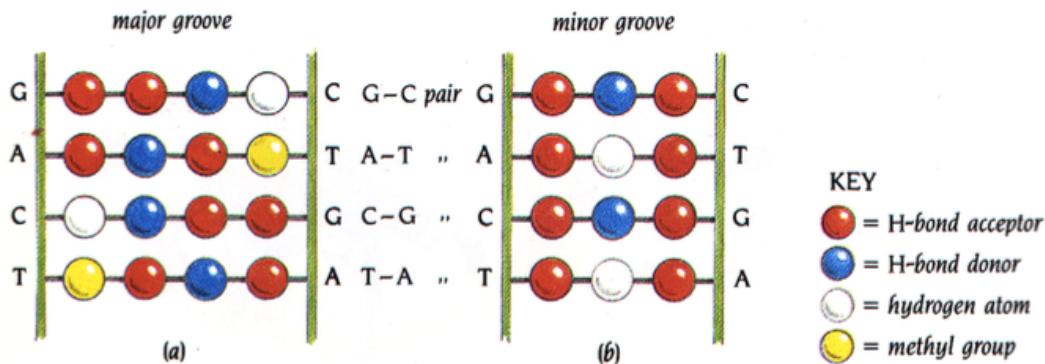


### B-form DNA helix



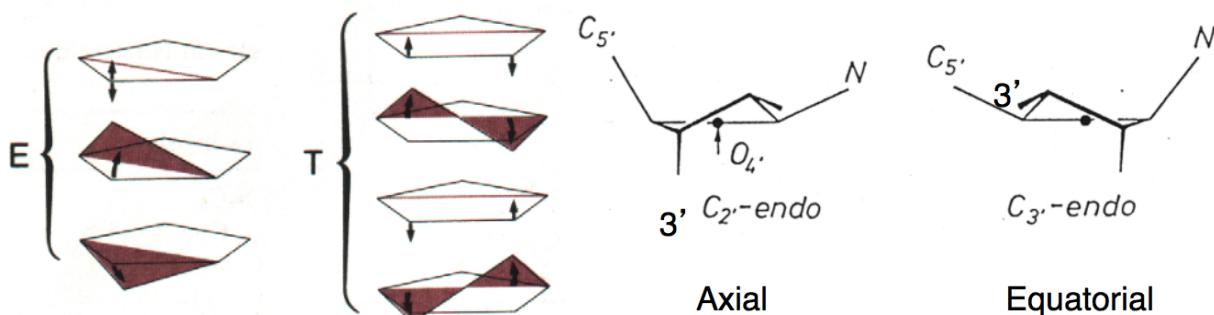
### J. Major and Minor Groove Chemistry

- a. Major groove is more chemically diverse: A-T  $\neq$  T-A, etc.
- b. Minor groove has equivalents: A-T = T-A, etc.



### K. Structural Constraints in DNA Conformation

- a. Nucleotide: 7 torsion angles
- b. Sugar pucker: E (envelope) has 4 coplanar; T (twist) has 3 coplanar, and 2 on diff sides.
  - i. B-form: C2'-endo sugars (DNA)
  - ii. A-form: C3'-endo sugars (RNA)
  - iii. Axial/equatorial difference leads to different helical geometries



- L. Syn: toward the base (rare in purines b/c small)  
 Anti: away from the base (vast majority).  
 Z-DNA: has syn guanines and a C3'-endo pucker.

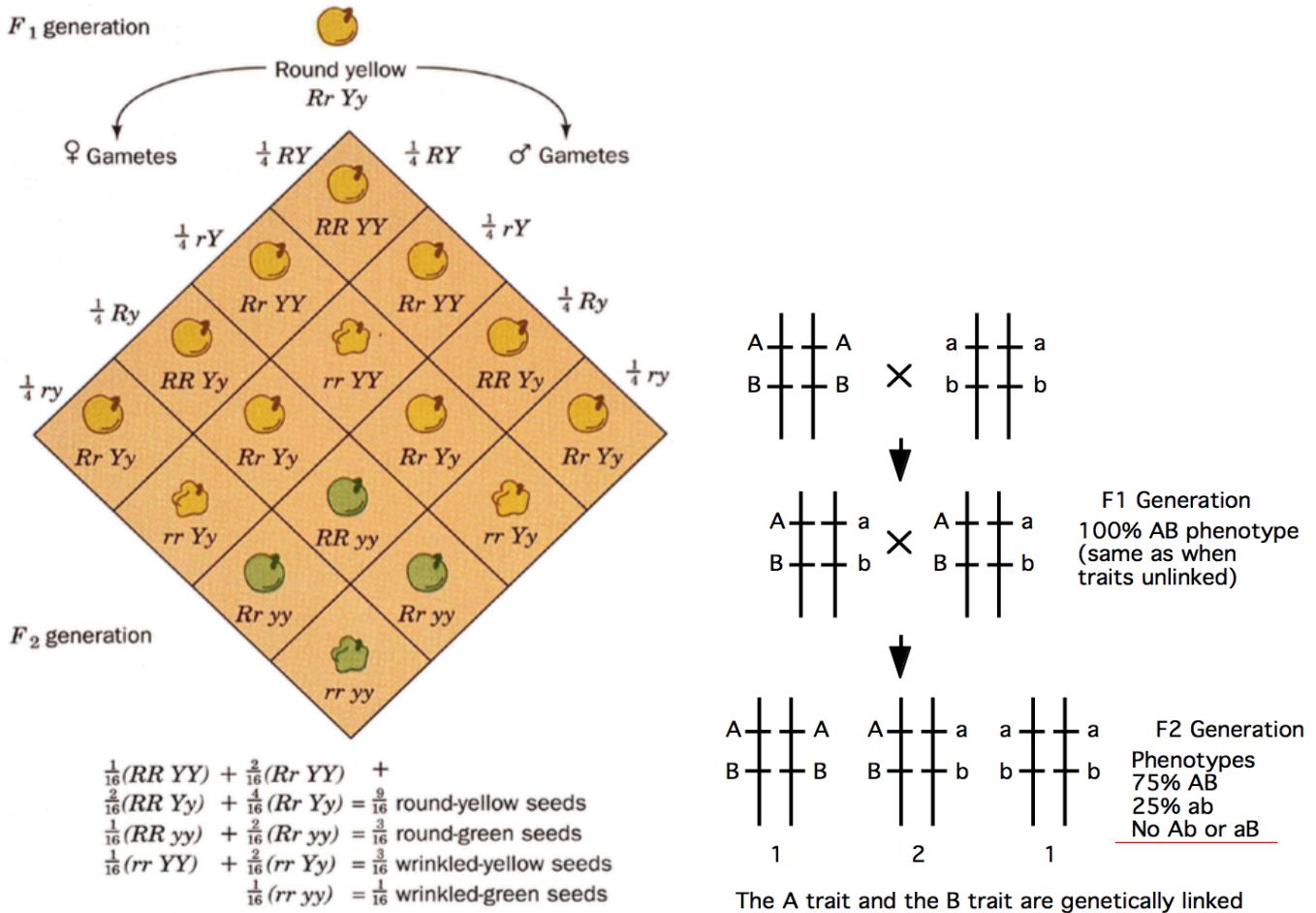
## Lecture 2: Phage/Bacterial Genetics & Gene Structure

### A. Mendel's Laws of Genetic Inheritance

- a. All organisms have 2 genes per trait, 1 from each parent.

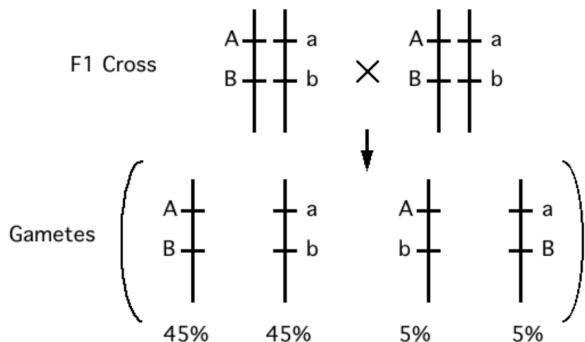
Different traits are inherited independently. Marginal distributions are the same (explore).

- b. Linked genes are an exception- prevents certain combinations, except by recombination.

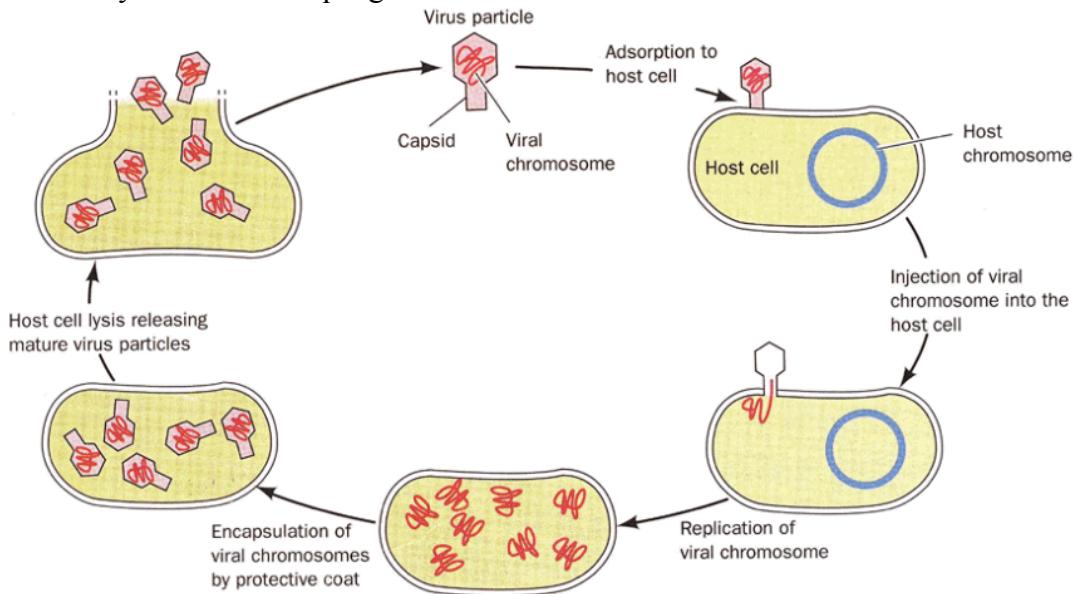


### B. Chromosomal recombination results in non-Mendelian ratios

- a. Recombination Frequency (RF) = #Recombinants / #Total Progeny \* 100
- b. Unlinked genes have RF = ~50% (Mendelian); Linked genes have RF < 50%  
Worries about multiple recombination.
- c. RF correlates with distances between genes: 1% RF = 1 centiMorgan (cM)  
Smaller distances require larger sample sizes.



### C. "Life" Cycle of Bacteriophages



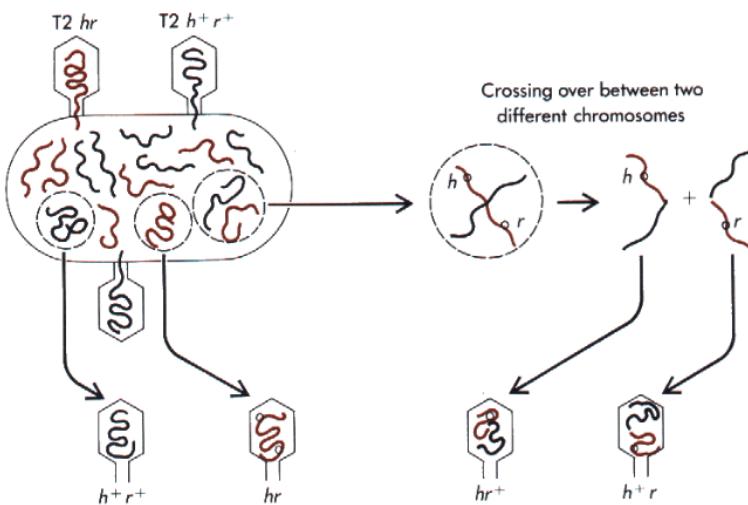
- a. Infect bacteria with phage at low titer [phage]/[medium] and spread on a plate.
- b. Plaques (holes in the bacterial layer) correspond to individual phage infections.

### D. Phage phenotypes

- a. Phenotypes include color, shape, size of plaque, different bacteria hosts, lysis rate.
- b. Seeing recombination is easier in haploid organisms (phages) b/c genotype = phenotype. There's no obscuring 2<sup>nd</sup> set of genes, no dominant/recessive, etc.
- c. Efficient:  $10^6 - 10^9$  individuals
- d. r<sup>+</sup>: wild-type, small. r: rapid-lysis, large. Coinfected: mottled plaque.

### E. Bacteriophage Recombination

- a. High titer of phage allows multiple phages to enter the same host. During infection/lysis, DNA recombination can produce mixed genotypes
- b. Spread an E. coli B and B/2 lawn. h<sup>+</sup> kills B cells, h kills both. h<sup>+</sup> results in turbid plaque.
- c. Ex: combining h/h<sup>+</sup> (can kill E. coli B/B2 cells) and r/r<sup>+</sup> (large, clear plaque). Adding only h/r and h<sup>+</sup>/r can produce h/r<sup>+</sup> and h<sup>+</sup>/r phenotypes.



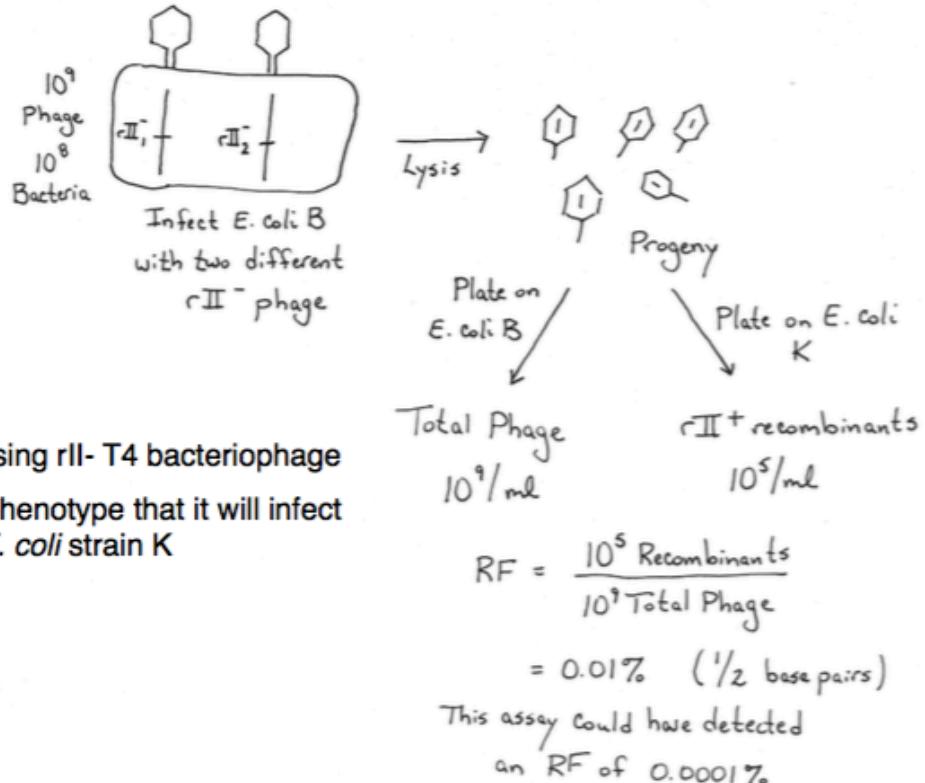
**Parental Genotype**

**Recombinant Genotype**

#### F. Ordering Genes by Successive Two Factor Crosses

- a. Count recombinants from AB x ab. 3 genes needs 3 experiments for absolute positions.  
Used to map the genome of the T4 map.
- b. Recombination analysis revealed circular viral genomes because distances from genes that should be far end up being close.
- c. People thought genes were indivisible because you can't detect intragenic recombination (the assay's resolution < length of gene), esp with experiments on pea/Drosophila.

#### G. Experiment: what is the smallest mutable unit in a gene?



- a. Normally, phages can infect B and K. The rII<sup>-</sup> mutated phage CANNOT infect K
- b. Recombination would combine wt rII<sub>1</sub> and rII<sub>2</sub> sites, rescuing the phenotype.
- c. Cross rII<sup>-</sup> x rII<sup>-</sup> and count rII<sup>+</sup> recombinants on K.
- d. Serial dilutions used for phage concentration.

#### H. Fine Mapping Structure of the Gene

- a. Huge sample sizes (scale of 10<sup>9</sup>) → nucleotide resolution (< 1 bp), RF of 0.0001%
- b. This was repeated with 1600 rII mutations; precisely mapped the rII gene.
- c. Mutation hotspot areas, maybe due to base rate of mutation (10<sup>-6</sup> – 10<sup>-10</sup>)
- d. Many big-picture implications
  - i. 1 gene can be mutated at multiple positions
  - ii. Genes are unambiguously linear (by mapping)
  - iii. Most mutations are at a single site (can be restored)
  - iv. Some mutations are deletions of 1+ mutable sites. Not resuable by back mutation.
  - v. Recognized that linkage map is a representation of Watson/Crick's DNA double helix molecule.

## I. Complementation vs. Recombination (Cis-Trans Assay)

- a. During coinfection, the two phages “complement” their deficiencies.
- b. Different genotypes may result in the same phenotype
  - i. Mutations in 1 gene vs. mutations in 1 of multiple NBI genes.
  - ii. Necessary but insufficient (NBI): subunits of protein; enzymes in pathway.
- c. rII phenotype is the ability to grow on strain K; 2 adjacent NBI genes (rIIA and rIIB).
- d. Two T4 phages with different rIIA/rIIB mutations can coinfect strain K.  
This results in a single hole (or patchy part) in a spot test.  
Same rIIA/rIIB mutations are unable to rescue the phenotype.
- e. Mutants that do NOT complement are in the same **cistron** (i.e. gene).

## J. Recombination vs. Complementation

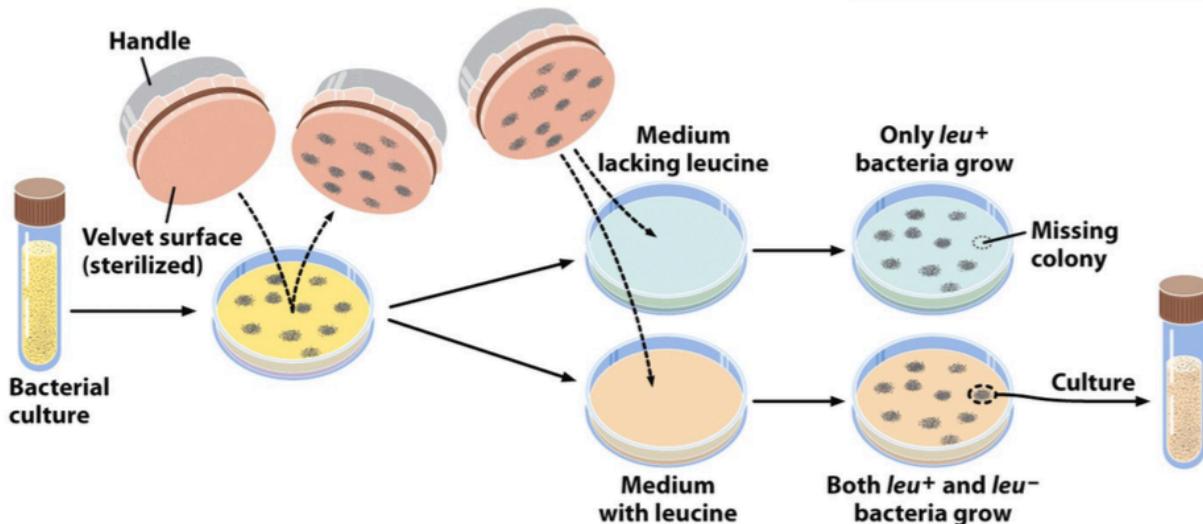
- a. Recombination is very rare, while complementation occurs in every infection.
- b. Recombination is assayed by many infections (spread on petri dish),  
while complementation is assayed by single infection (spot test).

## K. Bacterial Genetics

- a. Bacterial phenotypes can be assayed and genetically characterized: antibiotic resistance, ability to grow without some metabolic precursors, cellular mobility, phage infectivity, etc.
- b. Mutations can be spontaneous or induced (mutagens). They are stably inherited and can be selected upon.

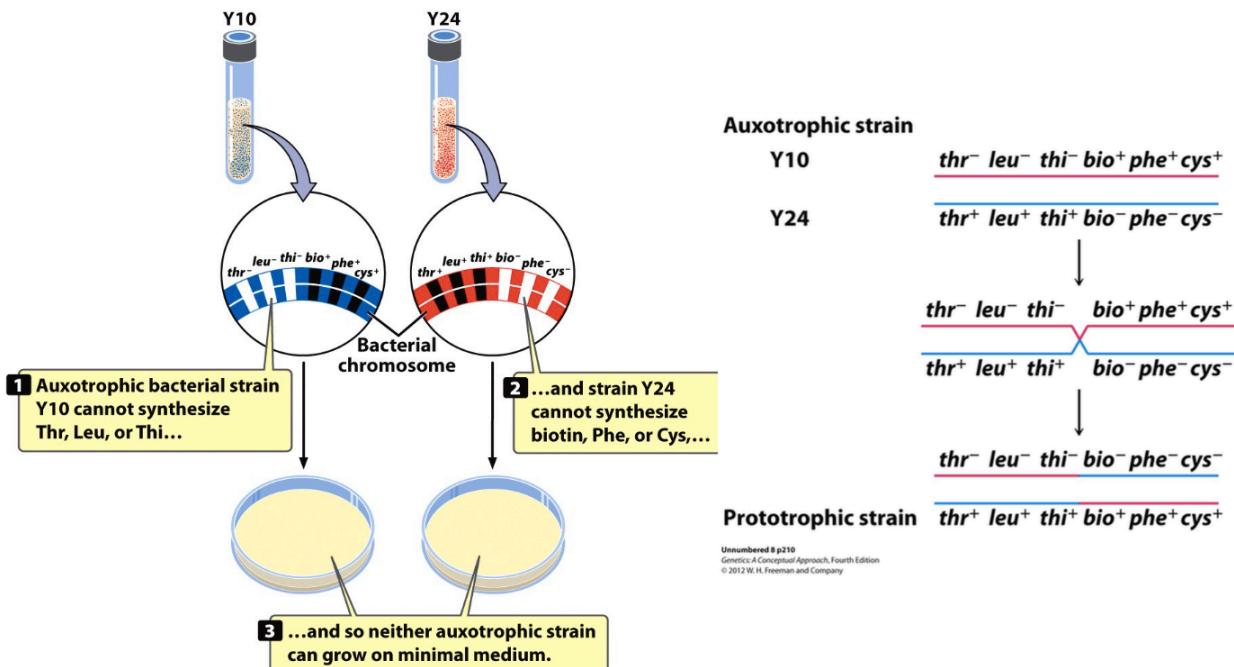
## L. Bacterial Mutant Selection via Replica Plating

- a. Procedure: Press sterile velvet cloth is on a master agar plate with bacterial colonies to pick up cells. Press against 1 or more new, sterile agar plate(s) to transfer these cells, which grow into mirror colonies.
- b. Selection for antibiotic-resistance: Add to 2<sup>nd</sup> plate to select for Strep<sup>R</sup> cells.
- c. Selection of auxotrophs: remove metabolic factor (leucine) on a 2<sup>nd</sup> plate, and replica master plate on the 3<sup>rd</sup> plate. Select for mutations in essential\_nutrient-coding genes.



## M. Bacterial Sex

- a. Lederberg-Tantum experiment: proved genetic exchange between bacteria.



- a. Strains able to “complement” each other are able to survive
- b. Challenge: diffusion of necessary metabolites.
- c. Solution: filter that excludes cells but allows medium exchange.

No colonies grew; physical contact is necessary.

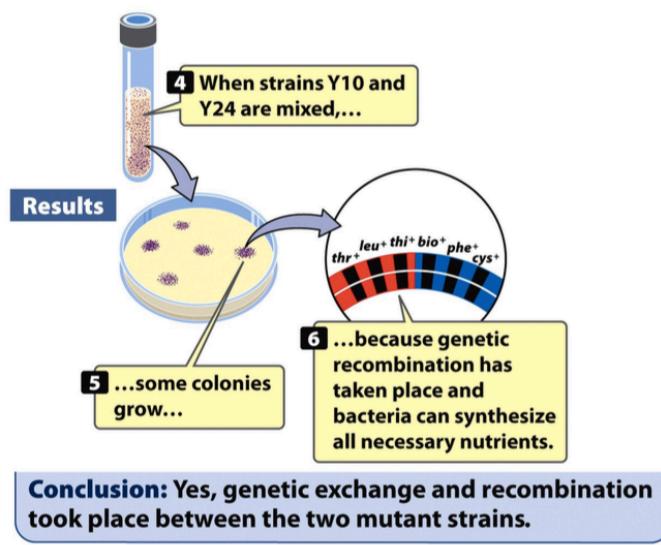
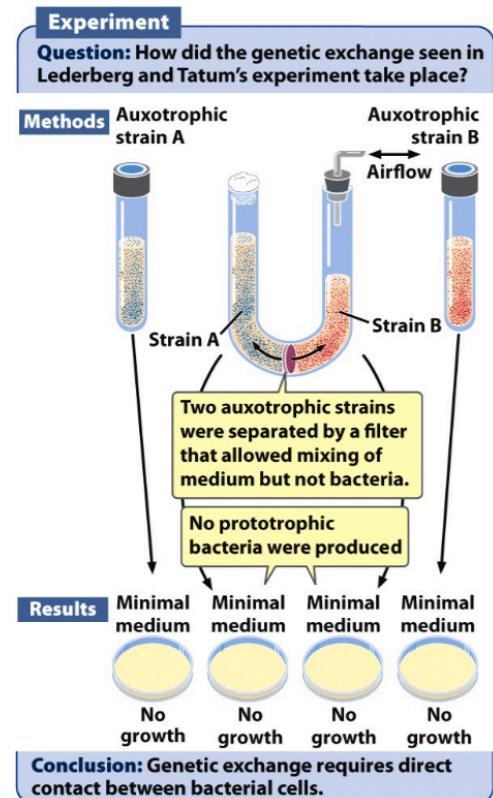


Figure 8.8 part 3  
Genetics: A Conceptual Approach, Fourth Edition  
© 2012 W.H. Freeman and Company



## N. F Factor: Mediates DNA Transfer between Bacteria

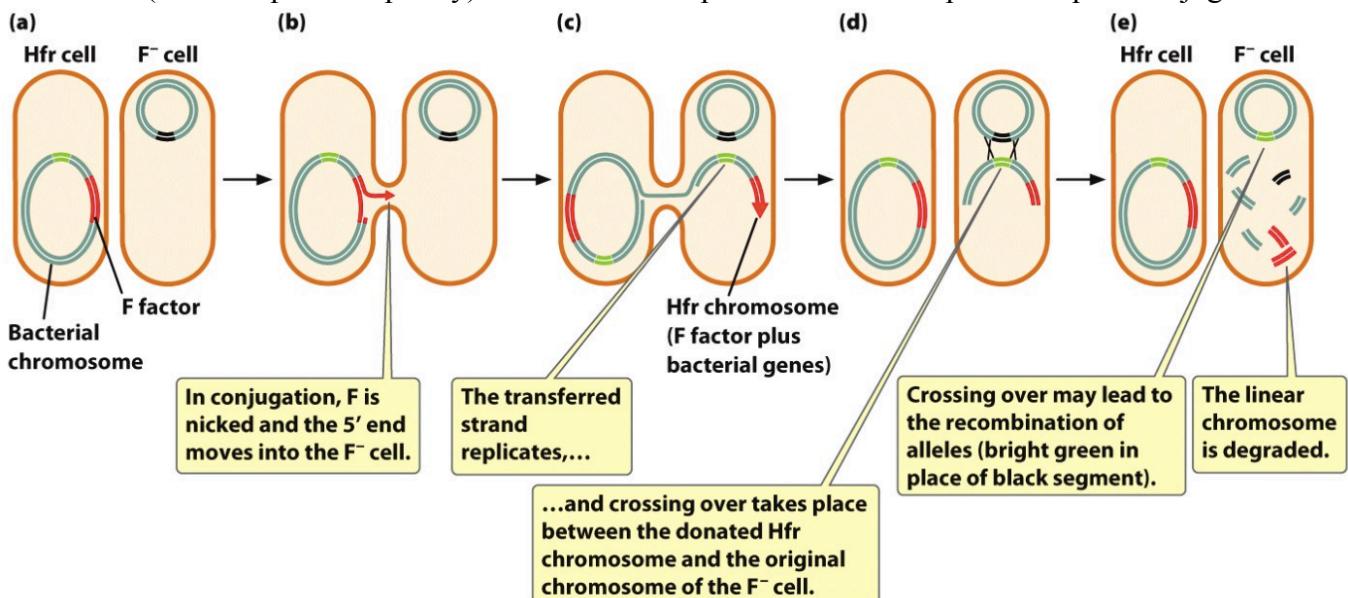
- F factor codes for a pilus to conjugate with an F- bacteria. After sex, both are F+.
- 2% the size of the 3000 kbp E. coli chromosome.

## O. High Frequency Recombinants

- High frequency recombinant (Hfr): integration of the F factor into the chromosome.
- When an Hfr transfers DNA via the F factor, it transfers the bacterial genome along with it. This starts at the site of integration and moves circularly around the chromosome.
- Transfer takes about 90 minutes; interruption results in partial transfer.

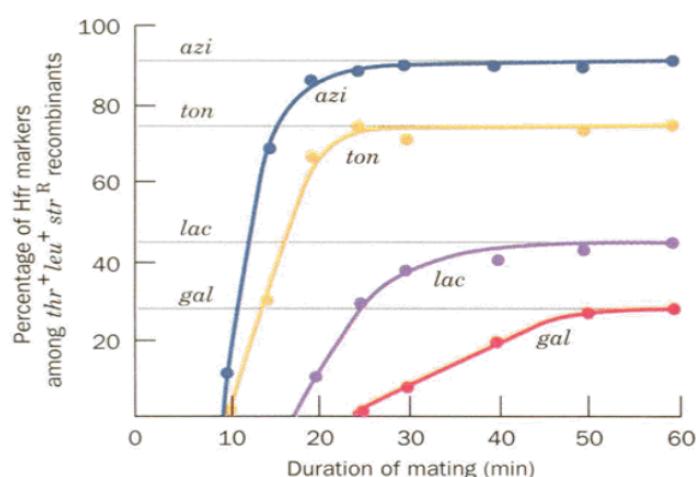
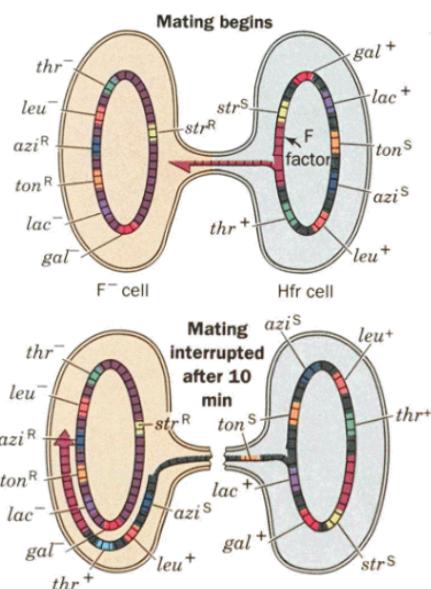
## P. Genomic Mapping using Bacterial Conjugation

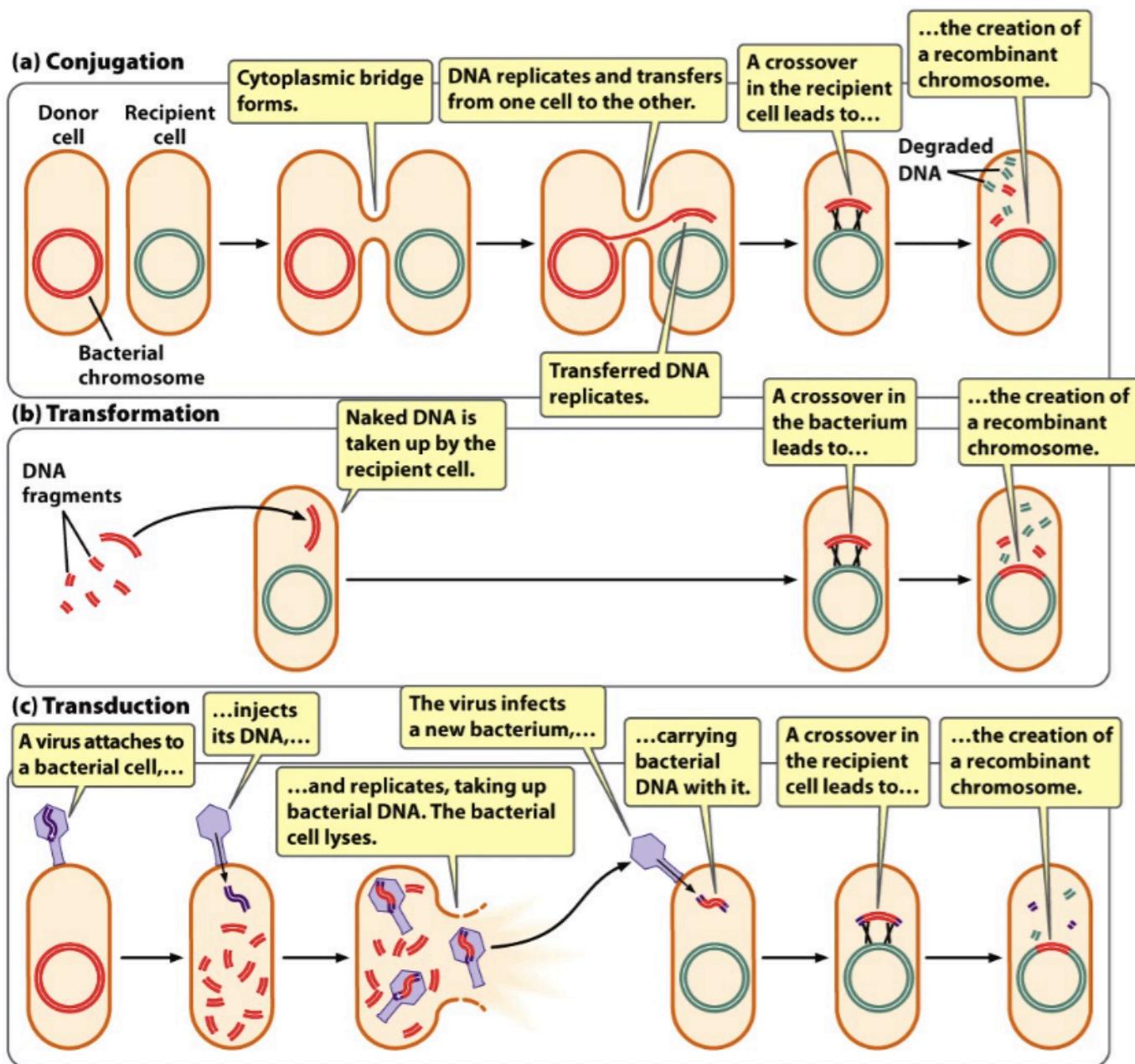
- Merodiploid: an essentially haploid organism carrying a 2<sup>nd</sup> copy of a part of its genome (unstable partial diploidy). Greek: meros = part. Occurs in recipient cell post-conjugation.



## Q. Gene Mapping by Interrupted Conjugation

- Wollman and Jacob: tracked recombinants over time to map bacterial genome (in minutes).
- Stop conjugation after different times (different transfer amount) to see how close genes are to the F-factor. The closest genes enter first, while more distal genes enter later.







## Lecture 3: Genetic Code

### A. The Coding Problem

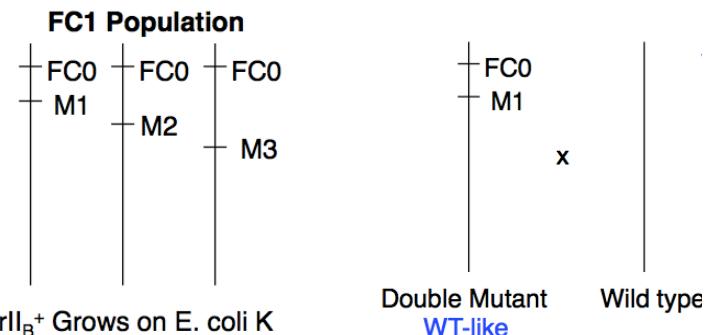
- a. How do the 4 nucleotides define codes for 20 amino acids?
- b. 2-base code is insufficient; only specifies  $4^2 = 16$  doublets.
- c. 3-base code specifies  $4^3 = 64$  triplets; implies degeneracy (redundancy)

### B. Gamow Code

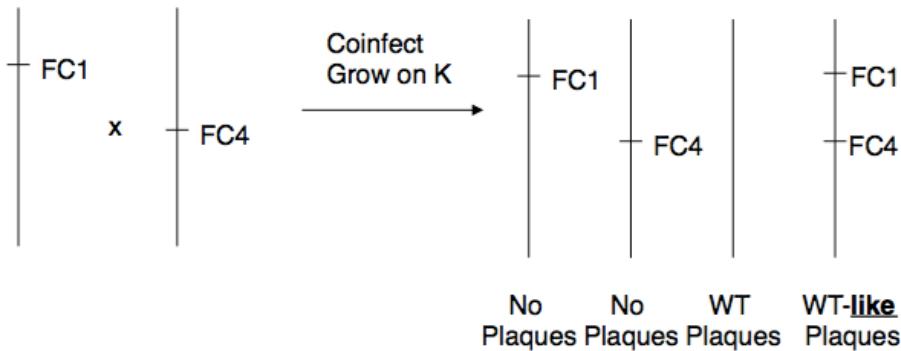
- a. Non-degenerate, overlapping (base pairs are shared in adjacent codes).
- b. DNA is itself the template for protein synthesis.
- c. The major groove forms distinctive cavities between bases that specific amino acids recognize and fit into. An enzyme then polymerizes them.
- d. Can be disproven by 1-1-2-2.

### C. Crick's experiments

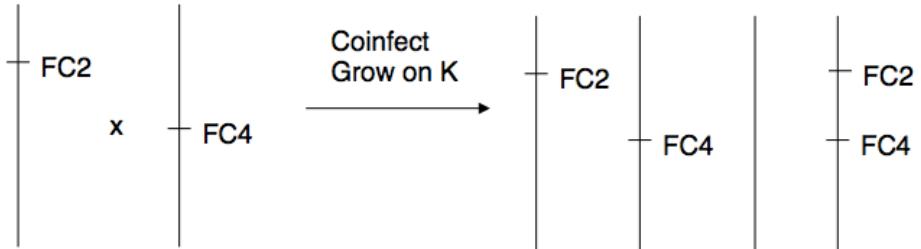
- a. T4 bacteriophage FC0 mutant: won't plaque E. coli K, gives an **r plaque** on E. coli B.
- b. Produced by proflavine: causes frameshift (+1 / -1).
- c. Step 1: use proflavine to produce M1, M2, M3 – frameshift mutations that partially rescue the phenotype ( $r\text{II}_B$  – ability to grow on K) when double-mutated (wt-like).



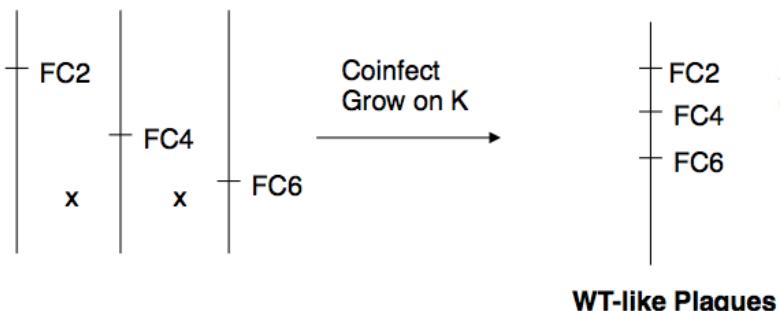
- d. Step 2: Cross FC0/FC1 with WT to isolate mutants
  - i. Replica plate to find B+/K- single-mutant strains.
- e. Step 3: Backcross mutant with FC0 to distinguish mutants
  - i. If mutant = FC1, some recombinants will be rescued (WT/WT-like)
  - ii. If mutant = FC0, recombinants still die.
- f. Repeat steps 1-3 to obtain generations of mutants (FC1 from FC0, FC2 from FC1, etc).
- g. Even x Odd: rescued double-mutant recombinants.



- h. Even x Even or Odd x Odd: double-mutant recombinants die.



- i. Triple cross with Even x Even x Even or Odd x Odd x Odd: triple recombinants survive.



- j. Conclusion: there is a reading frame of a triplet code.

Insertions and deletions would disrupt the reading frame → nonsense protein product.

#### D. Crick's Adaptor Hypothesis:

- Naïve idea: RNA can form 20 different cavities for amino acids to attach to.
- True idea: (m)RNA presents a sequence of H-bonding sites, and adaptor molecules (tRNA) fit onto the RNA template, carrying amino acids.
- tRNA contains an amino acid attachment site, and a template recognition site.

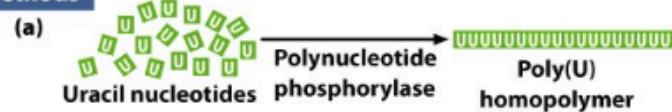
#### E. Nirenberg and Matthaei's filter binding experiment

- First codon identified by in vitro translation of homopolymer (polyA).
- Polynucleotide phosphorylase can polymerize dNTPs.
- High  $[Mg^{2+}]$  allows start-codon free translation.
- Radiolabeled each amino acid; the polymerized ones are precipitable by trichloroacetic acid (TCA).

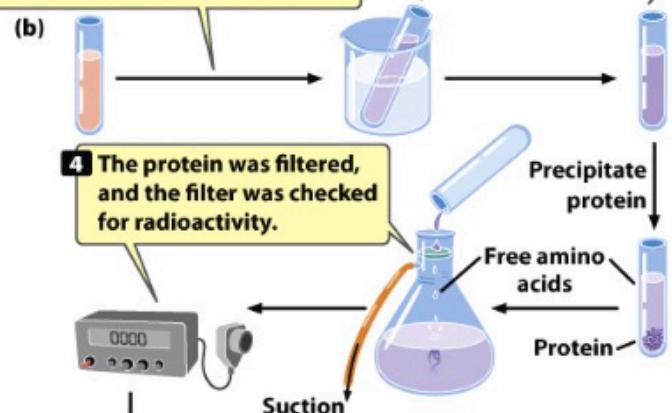
### Experiment

**Question:** What amino acids are specified by codons composed of only one type of base?

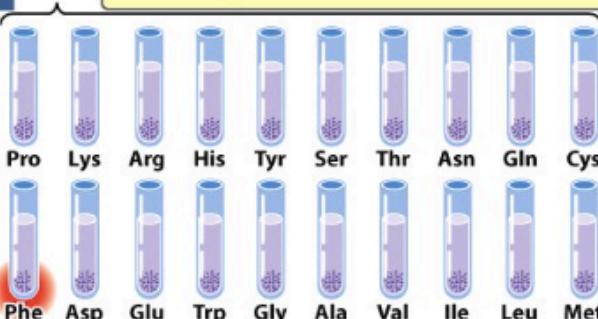
#### Methods



- 1 A homopolymer—in this case, poly(U) mRNA—was added to a test tube containing a cell-free translation system, 1 radioactively labeled amino acid, and 19 unlabeled amino acids.
- 2 The tube was incubated at 37°C.
- 3 Translation took place.



#### Results



- 5 The procedure was repeated in 20 tubes, with each tube containing a different labeled amino acid.
- 6 The tube in which the protein was radioactively labeled contained newly synthesized protein with the amino acid specified by the homopolymer. In this case, UUU specified the amino acid phenylalanine.

**Conclusion:** UUU encodes phenylalanine; in other experiments, AAA encoded lysine, and CCC encoded proline.

- a. Part 2: Fed dNTP in pairs into the translation system
- Poly A/C → CCC, CCA, AAC, CAC, ACA, ACC, and AAA
  - Relative frequency of each triplet depends on proportions of A and C
  - Conclusively proved degeneracy: Poly(UA/UC/UG) all encode Leu

Codon	Probability of Occurrence	Probability		Relative Amount of Amino Acid Incorporated
		Relative Incidence <sup>a</sup>	Amino Acid	
UUU	0.44	100	Phe	100
UUG	0.14	32	Leu	36
UGU	0.14	32	Cys	35
GUU	0.14	32	Val	37
UGG	0.04	9	Trp	14
GUG	0.04	9		
GGU	0.04	9	Gly	12
GGG	0.01	2		

- b. Part 3: Fed simple repeats

Table 15-3 Assignment of Codons Using Repeating Copolymers Built from Two or Three Nucleotides

Copolymer	Codons Recognized	Amino Acids Incorporated or Polypeptide Made	Codon Assignment
(CU) <sub>n</sub>	CUC UCU CUC . . .	Leucine Serine	5'-CUC-3' UCU
(UG) <sub>n</sub>	UGU GUG UGU . . .	Cysteine Valine	UGU GUG
(AC) <sub>n</sub>	ACA CAC ACA . . .	Threonine Histidine	ACA CAC
(AG) <sub>n</sub>	AGA GAG AGA . . .	Arginine Glutamine	AGA GAG
(AUC) <sub>n</sub>	AUC AUC AUC . . . UCA UCA UCA . . . CAU CAU CAU . . .	Polyisoleucine Polyserine Polyhistidine	5'-AUC-3' UCA CAU

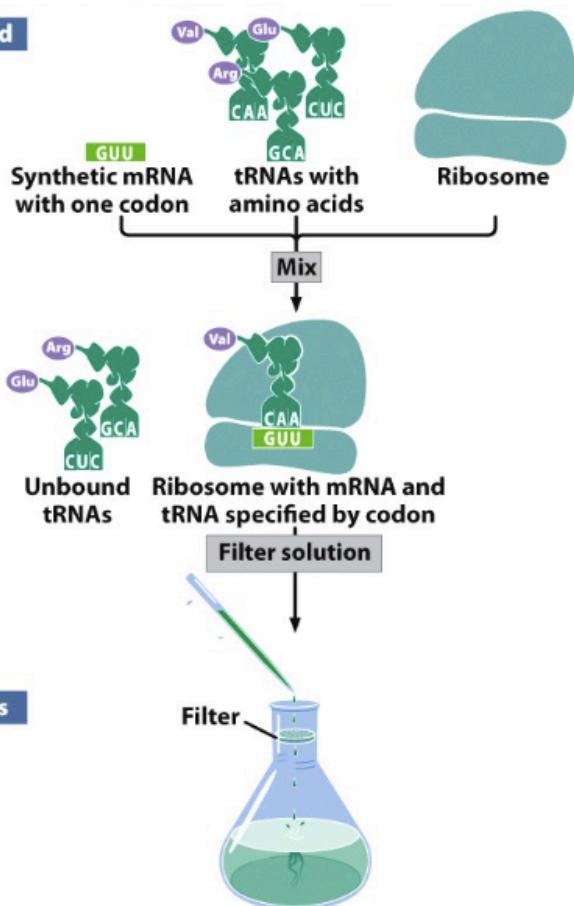
- Identified more degeneracy: some gave only 2 polymers (not 3)
  - Identified stop codons: some gave only di/tri-peptides (not polymers).
- UAG/UAA/UGA

## F. Triplet Binding Assay

### Experiment

**Question:** With the use of tRNAs, what other matches between codon and amino acid could be determined?

#### Method



#### Results

**Conclusion:** When an mRNA with GUU was added, the tRNAs on the filter were bound to valine; therefore the codon GUU specifies valine. Many other codons were determined by using this method.

Table 15-2 Binding of Aminoacyl tRNA Molecules to Trinucleotide-Ribosome Complexes

Trinucleotide						AA~tRNA Bound
5'-UUU-3'	UUC					Phenylalanine
UUA	UUG	CUU	CUC	CUA	CUG	Leucine
AAU	AUC	AUA				Isoleucine
AUG						Methionine
GUU	GUC	GUU	GUG	UCU*		Valine
UCU	UCC	UCA	UCG			Serine
CCU	CCC	CCA	CCG			Proline
AAA	AAG					Lysine
UGU	UGC					Cysteine
GAA	GAG					Glutamic acid

\*Note that this codon was misassigned by this method.

### G. The Standard Genetic Code

**Table 32-2** The “Standard” Genetic Code<sup>a</sup>

First position (5' end)	Second position				Third position (3' end)
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA Stop	UGA Stop	A
	UUG	UCG	UAG Stop	UGG Trp	G
C	CUU	CCU	CAU His	CGU	U
	CUC	CCC	CAC	CGC	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG	CCG	CAG	CGG	G
A	AUU	ACU	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met <sup>b</sup>	ACG	AAG	AGG Arg	G
G	GUU	GCU	GAU Asp	GGU	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA	GCA	GAA Glu	GGA Gly	A
	GUG	GCG	GAG Glu	GGG	G

<sup>a</sup>Nonpolar amino acid residues are tan, basic residues are blue, acidic residues are red, and nonpolar uncharged residues are purple.

<sup>b</sup>AUG forms part of the initiation signal as well as coding for internal Met residues.

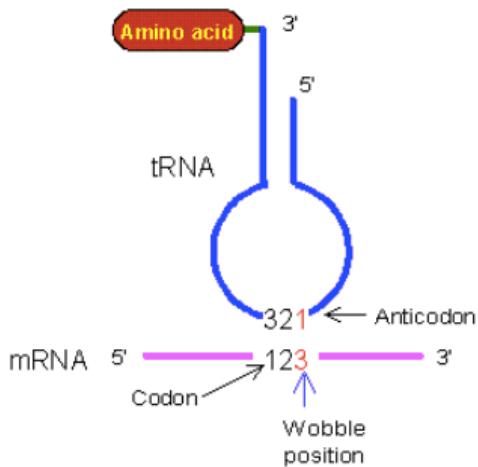
- a. Universal
  - i. Exceptions: mitochondria, archaeabacteria.
- b. NOT ambiguous
  - i. Exception: GUG sometimes read as Met/start; dependent on sequence context.
  - c. Highly degenerate (64 codons for 20 amino acids)
  - d. Three stop codons, not recognized by tRNA
  - e. Arrangement of codons is NOT random; evolved to minimize deleterious mutations.

H. Overlapping reading frames: present in some viral and mitochondrial genes, which have selection pressures for small genomes.

I. Codon usage varies significantly across organisms: this makes it difficult to overexpress human/mammalian genes in E. coli.

J. Wobble Hypothesis:

- a. Suggested after it was found that <61 tRNAs required to decode message.
- b. There is some degeneracy in the tRNAs.
- c. Some tRNAs recognize multiple codons with differences in the 3<sup>rd</sup> base (wobble position).  
tRNA(Ala) – GCU/GCC/GCA
- d. This is often due to inosine (post-transcriptional modification), which bonds with C/U/A.
- e. Similarly, U can bond with G in the 3<sup>rd</sup> position.



		Wobble bases				
tRNA		C	A	G	U	I
mRNA		G	U	C	A	C
					U	
mRNA		C	A	G	U	
tRNA		G	U	C	A	
		I	I	U	G	I

f. Codon is antiparallel to anticodon.



## Lecture 4: DNA Replication I

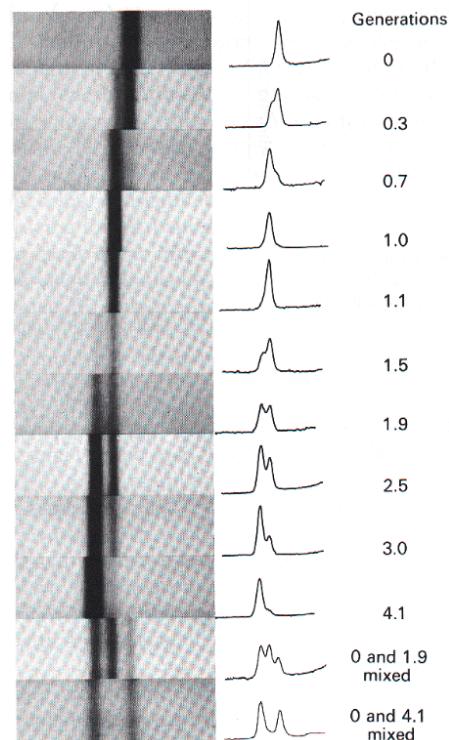
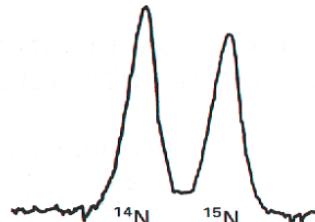
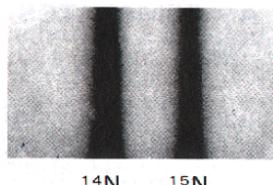
### A. Three models for DNA replication

- Conservative; copied without changing the original
- Semiconservative; separate and copy separately
- Dispersive; some mixture of the two

### B. Supercoiling is the main challenge to semiconservative model

### C. Isotopic Labeling of DNA using isotope N-15 in culture media.

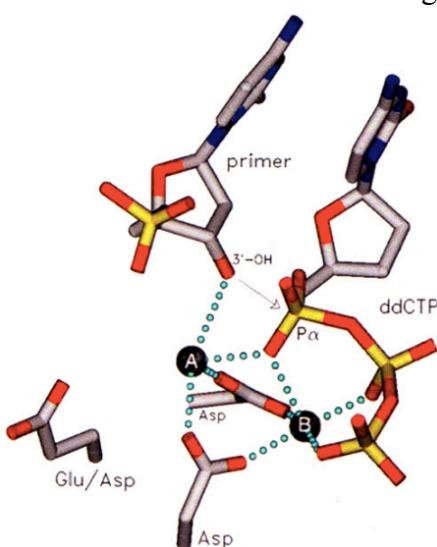
- Equilibrium density centrifugation



- Light and heavy DNA separated by ultracentrifugation
- Cells grown in N-15 media and then shifted to N-14.
- Generation 0: all heavy
- Generation 1: all intermediate
- Generation 2: half light, half intermediate
- Conservative model would always retain a heavy band and never have intermediate bands.
- Dispersive model would not have distinct bands; disproven.

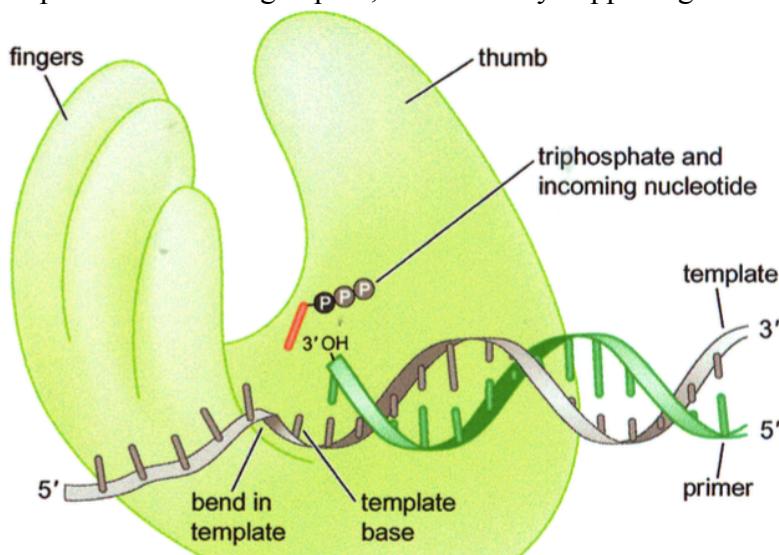
### D. DNA Polymerases

- All polymerases have a right-hand structure, with a catalytic site in the palm
- 2 key negatively charged residues coordinate metal ions – usu Mg<sup>2+</sup>
  - Positions the incoming nucleotide
  - Activates an –OH group for nucleophilic attack



**FIGURE 30-10 Schematic diagram for the nucleotidyl transferase mechanism of DNA polymerases.** A and B represent enzyme-bound metal ions that usually are Mg<sup>2+</sup>. Atoms are colored according to atom type (C gray, N blue, O red, and P yellow) and metal ion coordination is represented by green dotted lines. Metal ion A activates the primer's 3'-OH group for in-line nucleophilic attack on the incoming dNTP's α-phosphate group (arrow), whereas metal ion B acts to orient and electrostatically stabilize the negatively charged triphosphate group. [Courtesy of Tom Ellenberger, Harvard Medical School.]

- c. Primer strand: extended strand.
- d. Template strand: being copied; no chemistry happening to it.



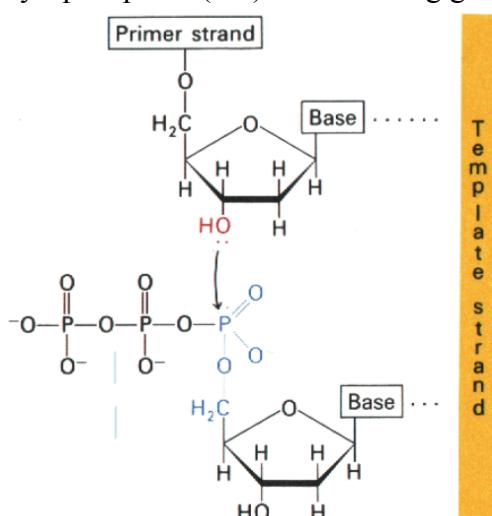
- e. Processivity: how long it continues adding nucleotides before it falls off the strand. Higher processivity makes a more processive pol.
- f. Pol proceeds 5' → 3' ALWAYS.

#### E. Types of DNA Polymerases in E. coli.

- a. Pol I: synthesizes short DNA segments in DNA repair and nick translation.  
Has BOTH 3'-5' proofreading and 5'-3' nick translation activity.
- b. Pol II: Deletion is not lethal, but participates in SOS response (DNA repair).
- c. Pol III: dominant; only 10-20 copies per cell. Most efficient, highly processive.

#### F. Polymerization reaction

- a. Polymerase uses the same active site for all 4 base pairs.
- b. The 3'-OH of the primer strand attacks the α-phosphate of the incoming nucleotide, which is aligned by base-pairing.
- c. Pyrophosphate (PP<sub>i</sub>) is the leaving group; high-energy phosphate drives the reaction.



### G. Polymerase Fidelity

- a. More accurate than expected from energetics of base-pairing.
- b. Induced fit: binding of correct dNTP “closes” the fingers to initiate catalysis.
- c. Error rate:  $10^4\text{-}10^5$  bp – lethal after just a few cell divisions.
- d. Replicative polymerases have 3'-5' exonuclease activity (proofreading).  
May be separate peptide domain (Pol I/II) or separate subunit (Pol III)
- e. DNA mismatch swings the ssDNA to the exonuclease site.

### H. Proofreading mechanism

- a. DNA mismatch leads to ssDNA, which favors binding into the exonuclease site, where it is degraded.
- b. The cost is high; 10% of the genome is degraded and resynthesized every cycle.

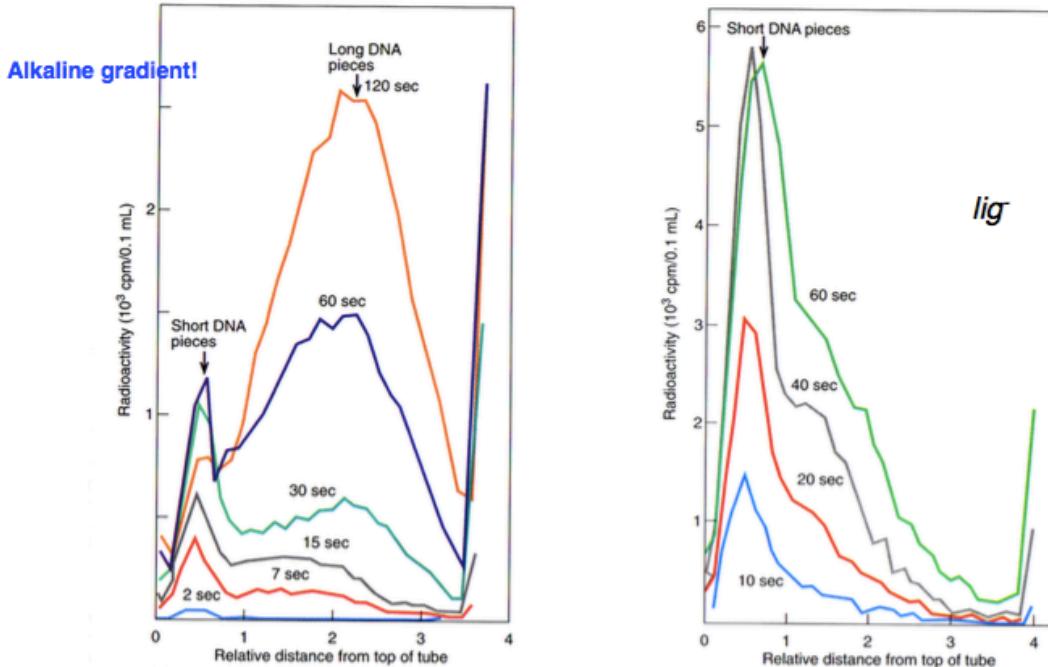
### I. 5'-3' Quiz

- a. dNTPs must be added to the 3' end because that's where the attacking  $-OH$  is. If it were added to the 5' end, the strand itself would have to supply the high-energy triphosphate.
- b. This is problematic because proofreading and spontaneous hydrolysis would completely shut down replication.

### J. The polarity problem

- a. Continuous 5'-3' replication is impossible in the lagging strand.
- b. Okazaki labeled replicating T4 DNA with short pulses of radioactive DNA precursor, and separated by size with centrifugation.
- c. At first there are many short pieces, but the longer pieces later accumulate.
- d. If DNA ligase is knocked out ( $lig^-$  assay), short DNA continues to predominate.

### The polarity problem answered?



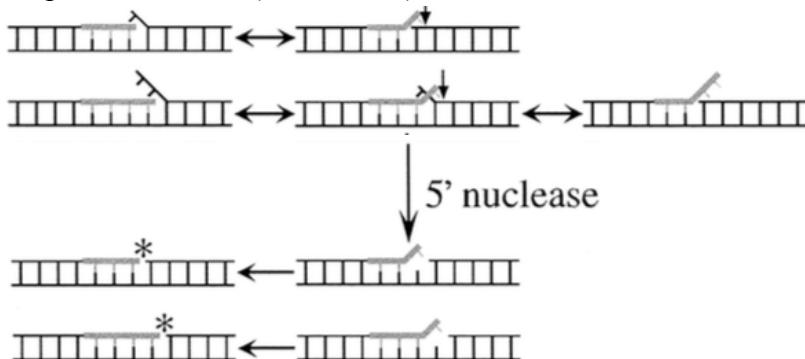
- e. U substitution for T necessitates repair using ligase, which results in more small fragments even in the leading strand.

K. Primase (DnaG) sets a 3'-primer (~5nt) every few hundred base-pairs.

This is extended by DNA polymerase until it reaches the 5' of the previous RNA primer.

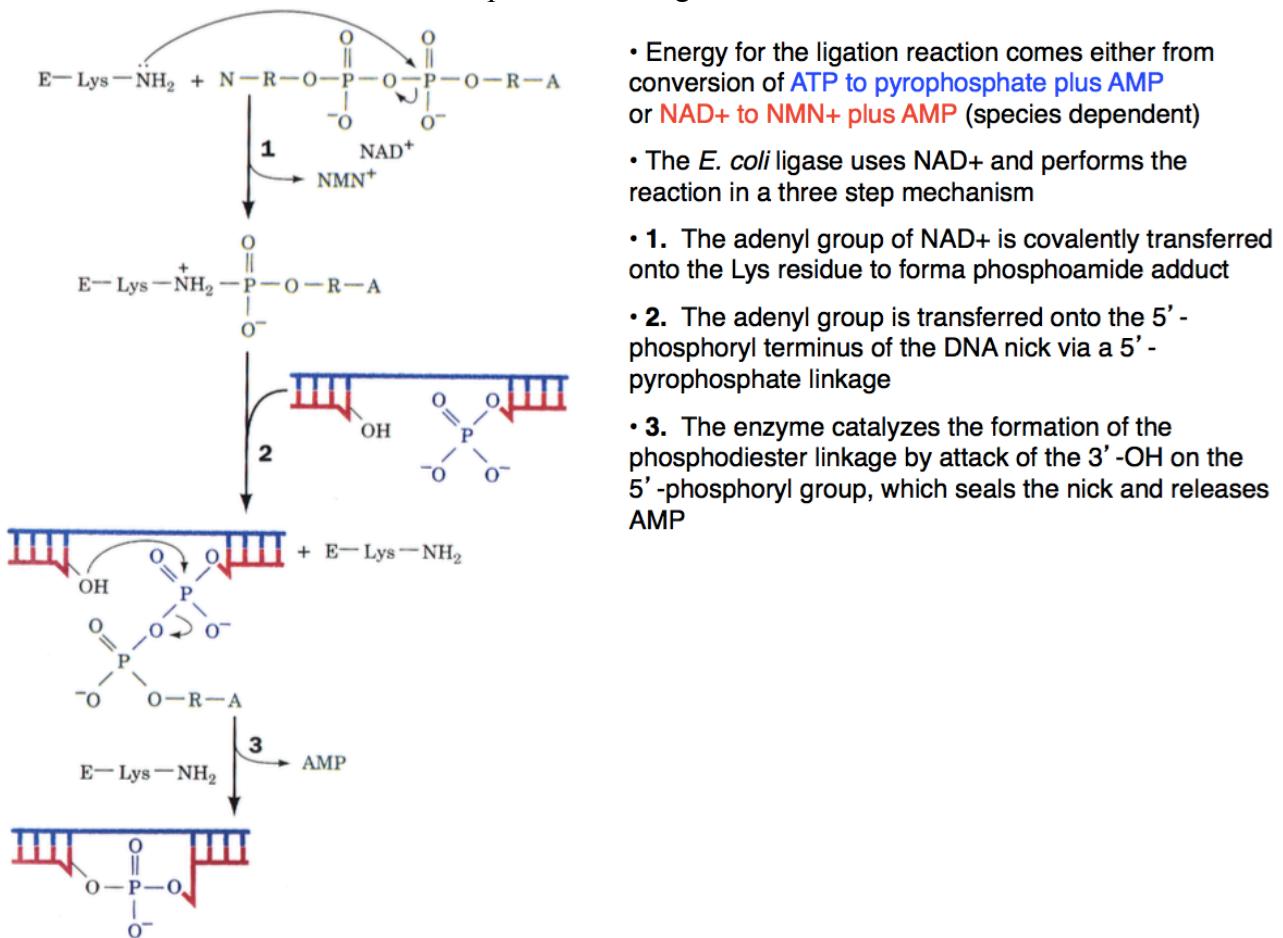
#### L. Nick Translation and Primer Removal

- Pol 1: 5' nuclease activity can chew through RNA primers to combine Okazaki fragments. This results in a flap, which is in equilibrium.
- Flap endonuclease (5' nuclease) makes a cut to create a flush-end nick, sealed by ligase.



#### M. Ligation

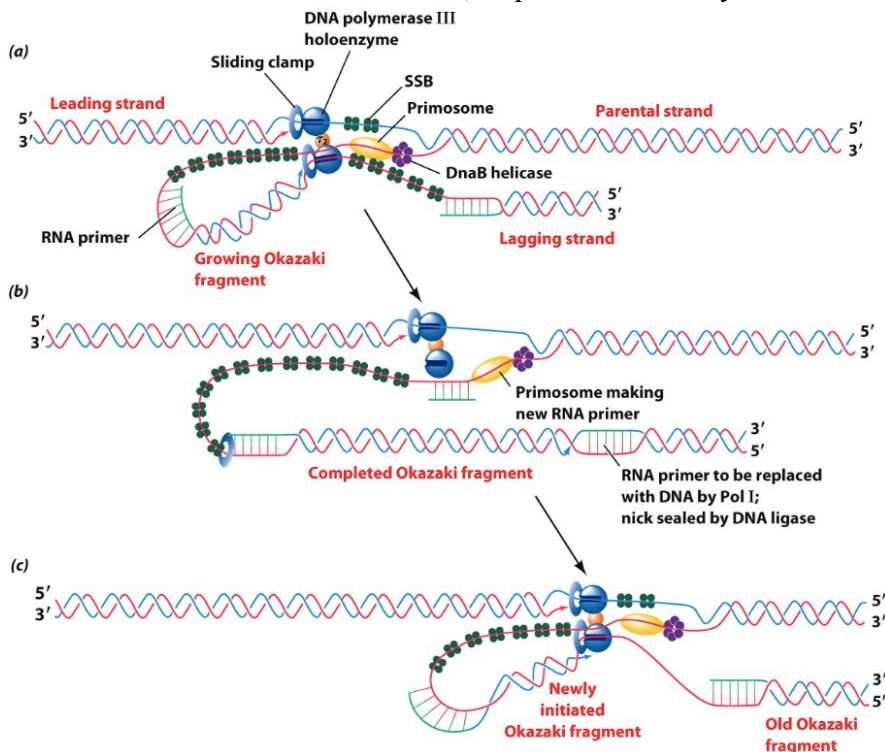
- Powered by ATP  $\rightarrow$  AMP and PP<sub>i</sub> // NAD<sup>+</sup>  $\rightarrow$  NMN<sup>+</sup> and AMP
- Can splice together 2 unrelated segments of DNA
- Nick translation: used to incorporate radiooligonucleotides.



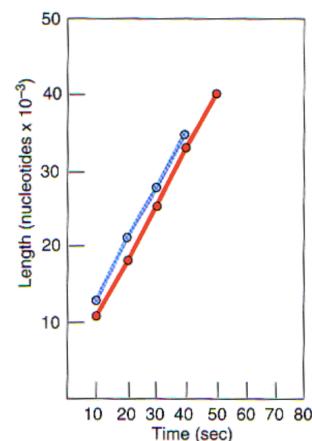
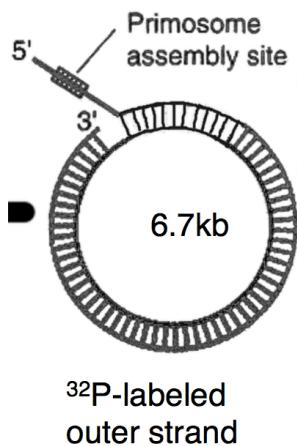
**FIGURE 31-16.** The reactions catalyzed by *E. coli* DNA ligase. In eukaryotic and T4 DNA ligases, NAD<sup>+</sup> is replaced by ATP so that PP<sub>i</sub> rather than NMN<sup>+</sup> is eliminated in the first reaction step. The numbered steps are described in the text.

## Lecture 5: DNA Replication II

- A. Learn Processivity experiment and processivity v. distributive
- B. DNA Pol III holoenzyme is a very complicated machine
  - a. Polymerization and proofreading (core)
  - b. Clamp loading complex (sliding clamp): prevents Pol III dissociation from the template DNA strand; allows more nt added to the growing strand per association (1,000-fold).
- C. Problems with DNA Replication
  - a. Primer must be removed; 5' nuclease of Pol I
  - b. Processivity; sliding clamp ( $\beta$ -subunit) and clamp loader (gamma complex) Pol II without  $\beta$ -subunit would only last 10-15 nts.
  - c. Exposed ssDNA; SSB
  - d. Supercoiling; topoisomerase (gyrase)
- D. Replisome model
  - a. Coordination of leading and lagging strands
  - b. Dimeric Pol III allows coordination; improves efficiency.



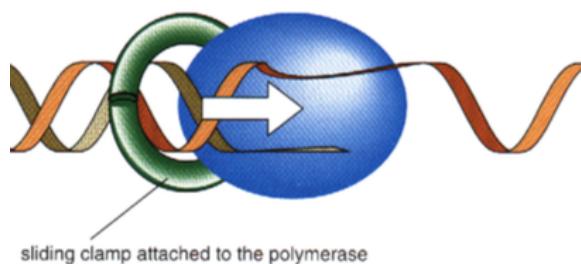
- E. Experiment: Measuring Pol III Leading Strand Processivity
  - a. Pre-primosome and DNA B helicase form replication forks that move at the same rate.



- b. Processive enzyme: binds substrate and fully extends before dissociating.
- c. Distributive enzyme: binds substrate, extends some nts, dissociates, and moves on.

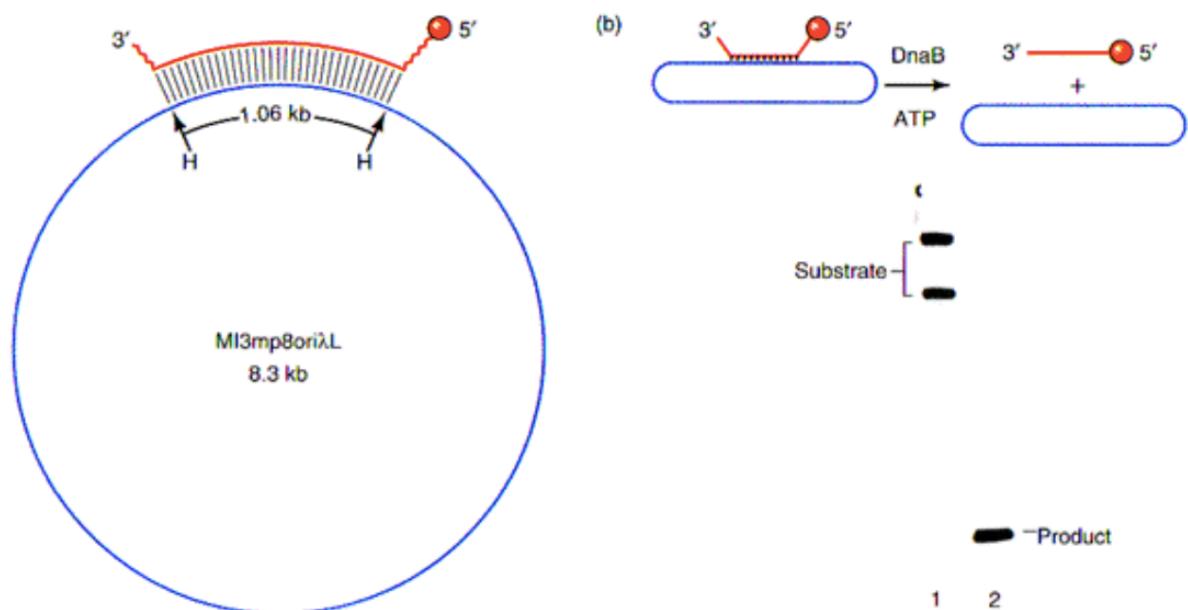
#### F. Processivity dependent on $\beta$ subunit

- a. Assembly is non-covalent but irreversible (ATP required for assembly).
- b.  $\beta$  subunit is a dimer that assembles as a circle around the helix.
- c. Is added to the helix by the clamp loader (gamma complex).



#### G. DNA helicase assay

- a.  $^{32}\text{P}$ -labeled ssDNA at the 5' end, and annealed to long ssDNA.
- b. DNA helicases (DnaB) can break the H-bonds and liberate the labeled strand.
- c. Two substrate bands are either linear/circular, or relaxed/supercoiled.

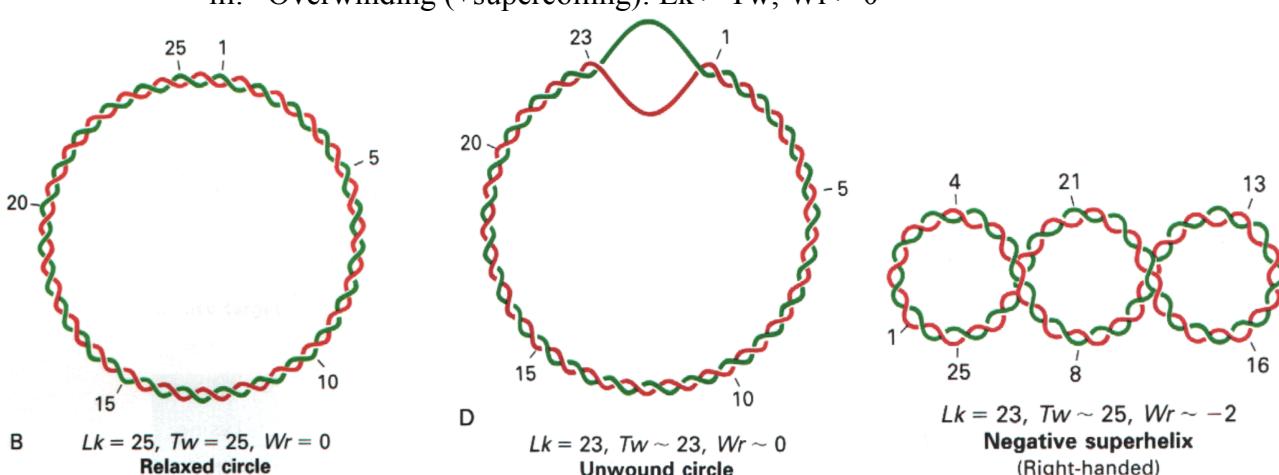


## H. Helicase and SSB

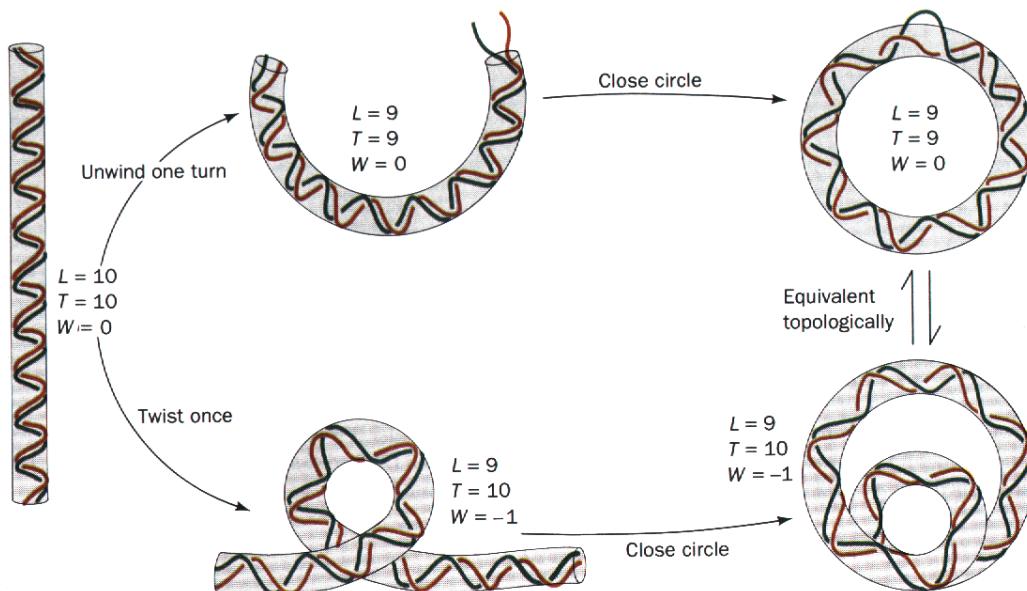
- a. Helicase uses ATP to break H-bonds.
- b. SSB prevents rehybridizing after unwinding. SSB are recycled when Pol reaches them.

## I. Superhelicity and Topoisomerases

- a. Normal biological function of DNA ONLY in proper topological state.
- b. Linking number ( $Lk$ ): number of times one strand wraps around the other
- c. Twist ( $Tw$ ): number of base-pairs divided by nt/turn (10.5 for B-DNA).
- d. Writhe ( $Wr$ ): turns of the duplex axis about the superhelical axis (???)
- e.  $Lk = Tw + Wr$ .
  - i. Relaxed:  $Lk = Tw$
  - ii. Underwinding (-supercoiling):  $Lk < Tw$ ;  $Wr < 0$
  - iii. Overwinding (+supercoiling):  $Lk > Tw$ ;  $Wr > 0$



## J. Topoisomerases affect supercoiling



**FIGURE 28-36.** Two ways of introducing one supercoil into a DNA with 10 duplex turns. The two closed circular forms shown (right) are topologically equivalent; that is, they are interconvertible without breaking any covalent bonds. The

linking number  $L$ , twist  $T$ , and writhing number  $W$  are indicated for each form. Strictly speaking, the linking number is only defined for a covalently closed circle.

### K. Introducing Negative Supercoiling

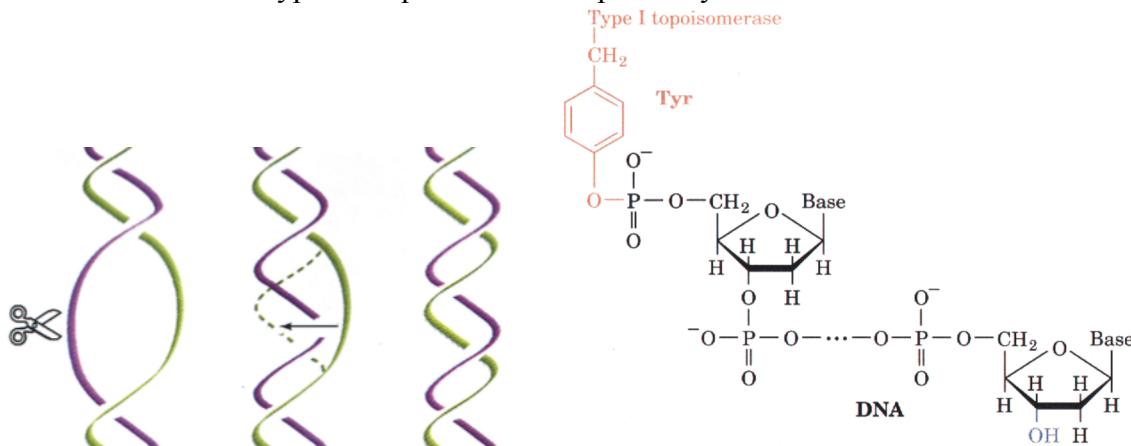
- Unwinding one loop and reattaching.
- Physically winding the duplex with 1 left-handed turn.

### L. Supercoiling and Electrophoresis

- The more supercoiled the DNA, the faster it moves through the gel.
- Incubating with topoisomerase leads to discrete bands for different Lk/Wr.

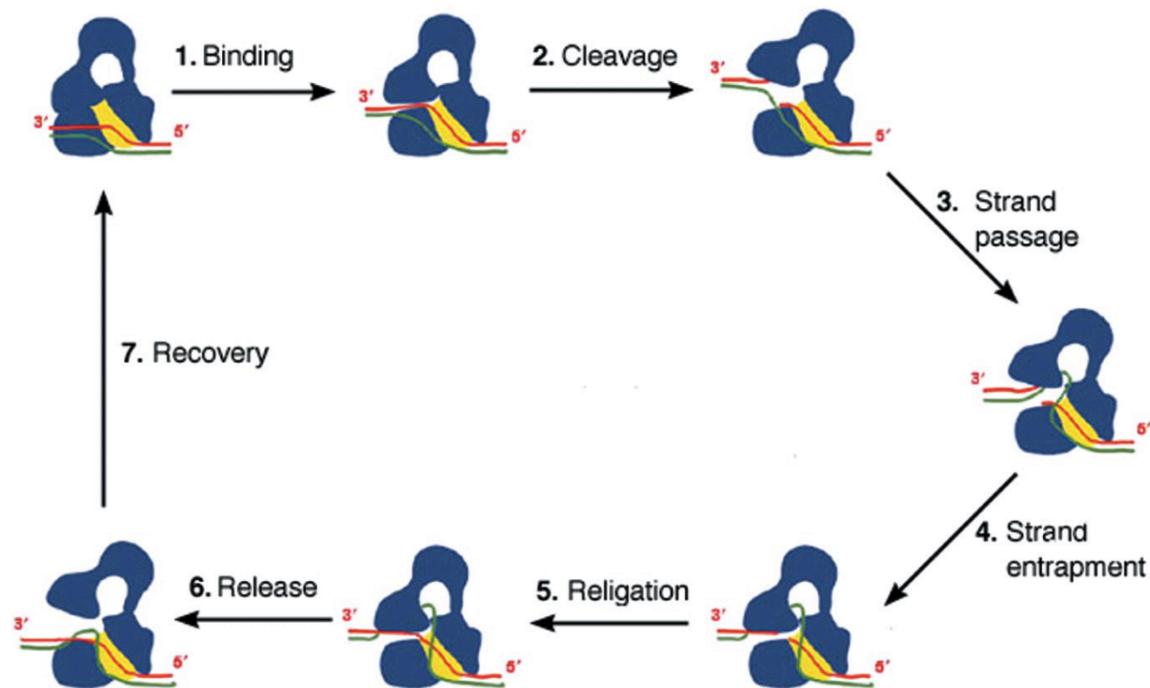
### M. E. coli topoisomerase I (type IA topo) is a nicking-closing enzyme

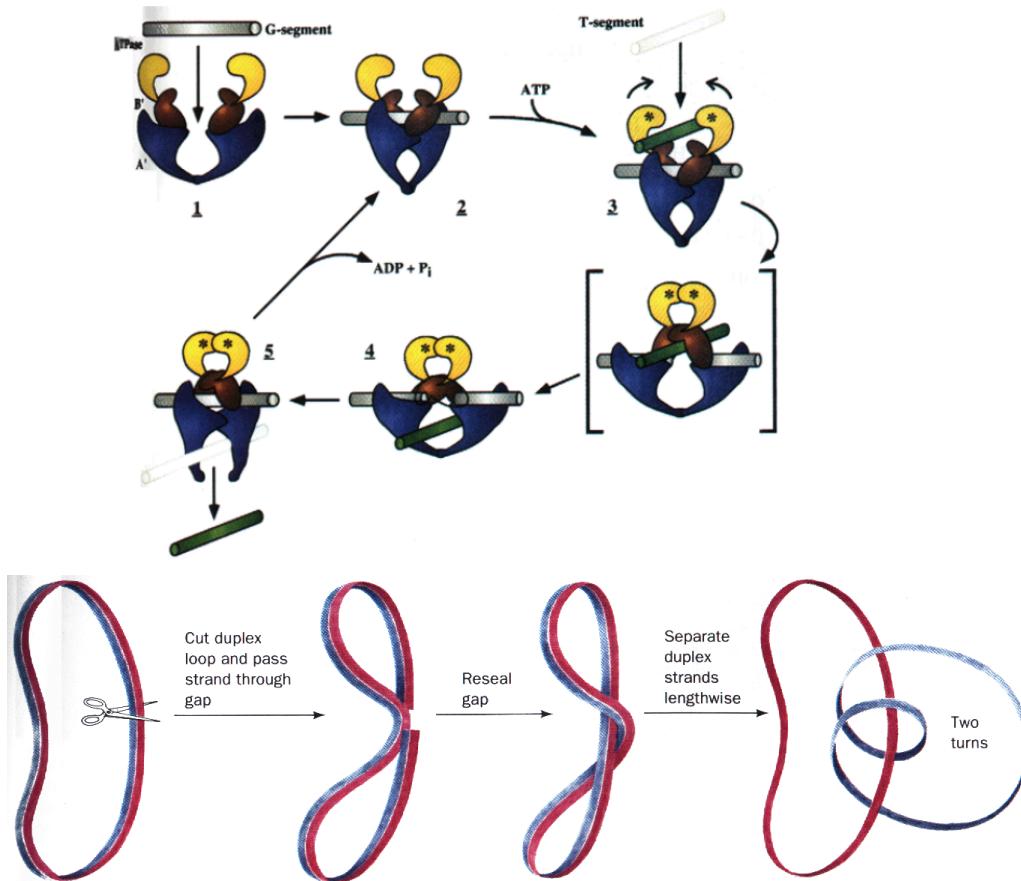
- Covalent topo-DNA (phosphotyrosine) intermediate
- Operates by a “strand passage” mechanism
- (1) Catenates single-stranded circles (2 strands with Lk=0 produces 2 strands with Lk=1)
- (2) Increases Lk
- Note: Type 1B topoisomerases operate by a “nick-rotate” mechanism.



### N. Reaction mechanism

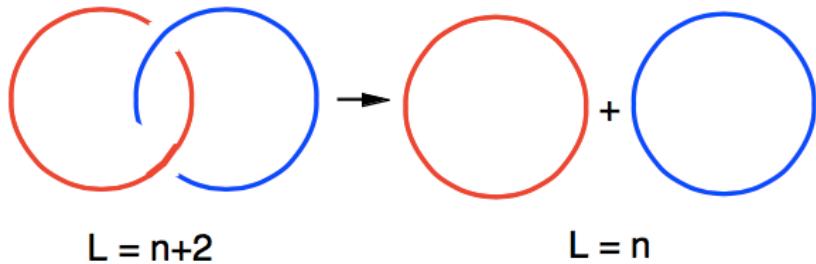
- Covalent phosphotyrosine intermediate between the DNA strand and a tyrosine residue.
- The enzyme acts as a hinge and clamp, which allows the complementary strand through the single-stranded nick → Lk increased by 1 until Wr = 0.





**FIGURE 28-43.** A demonstration, in which DNA is represented by a ribbon, that cutting a duplex circle, passing the strand through the resulting gap, and then resealing the break

changes the linking number by two. Separating the resulting strands (slitting the ribbon along its length; right), indicates that one strand makes two complete revolutions about the other.



O. Topoisomerase II (DNA gyrase)

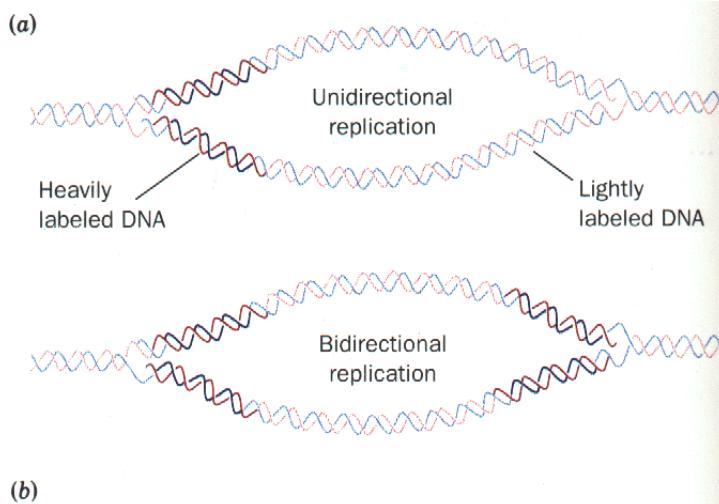
- DNA gyrase introduces negative supercoils into DNA ( $Lk \approx -2$ )
- Relaxes positive superhelical strain
- A phosphotyrosine intermediate conserves phosphodiester bond energy.
- Topo II involves a hinge and clamp mechanism

P. Quiz?

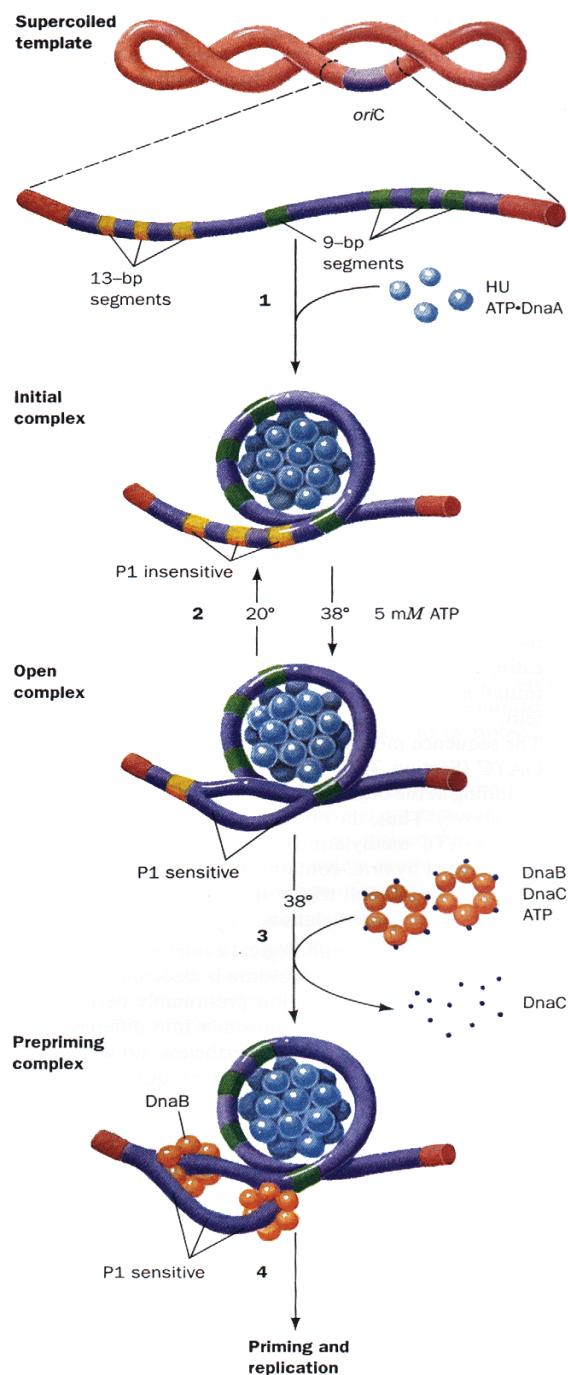
- Q. DNA Replication requires: Pol III, helicase, SSB, primosome, gyrase, processivity clamp and clamp loader, and ligase.

R. Directionality

- Initiation creates a bidirectional replication bubble
- Pulse chase with tritium-labeled nts can distinguish uni/bi-directionality.



**FIGURE 31-4.** The autoradiographic differentiation of unidirectional and bidirectional  $\theta$  replication of DNA. (a) An organism is grown for several generations in a medium that is lightly labeled with [ $^3$ H]thymine so that all of its DNA will be visible in an autoradiogram. A large amount of [ $^3$ H]thymine is then added to the medium for a few seconds before the DNA is isolated (pulse labeling) in order to label only those bases near the replication fork(s). Unidirectional DNA replication will exhibit only one heavily labeled branch point (above), whereas bidirectional DNA replication will exhibit two such branch points (below). (b) An autoradiogram of *E. coli* DNA so treated, demonstrating that it is bidirectionally replicated. [Courtesy of David M. Prescott, University of Colorado.]

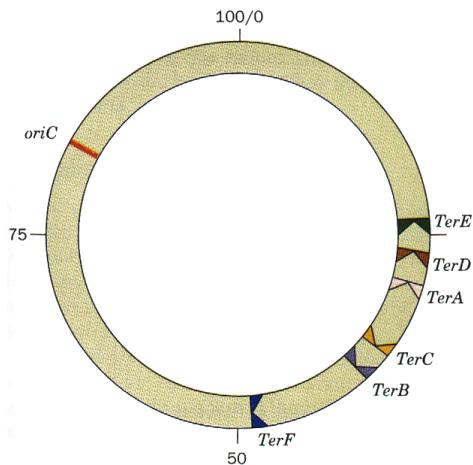


### S. Replication Initiation

- Replication initiated at single site in *E. coli*: oriC
- oriC contains 4 copies of a 9-bp sequence element, recognized by DNA-binding protein DnaA, which forms an initial complex (negatively supercoiled).
- This melts an AT rich stretch on the boundary of the oriC (becomes S1 nuclease sensitive; S1 is specific for ssDNA).
- This melted region is then recognized by the hexameric complex DnaB, DnaC to form a prepriming complex from which DnaC is released.
- Melted region expanded by gyrase/helicase/SSB activity; replication primed by primase.

### T. Replication Termination

- a. Termination signal is a large region flanked by 6 termination sites (TerA-TerF)
- b. The signals are directional; only signals termination in 1 direction.
- c. Arrest of replication requires Tus protein, which binds to the Ter sites and blocks helicase.



**FIGURE 31-26.** A map of the *E. coli* chromosome, in the same orientation as Fig. 27-18, showing the positions of the *Ter* sites (right) and the *oriC* site (left). The *TerC*, *TerB*, and *TerF* sites, in combination with *Tus* protein, allow a counter-clockwise-moving replisome to pass but not a clockwise-moving replisome. The opposite is true of the *TerA*, *TerD*, and *TerE* sites. Consequently, two replication forks that initiate bidirectional DNA replication at *oriC* will meet between the oppositely facing *Ter* sites.

### U. Replication Speed

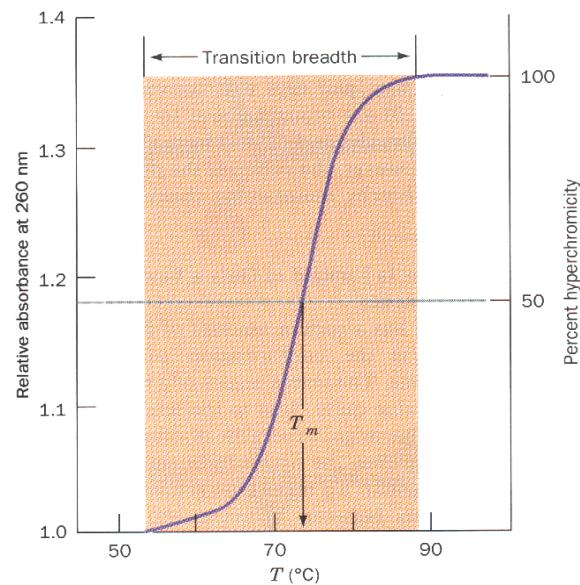
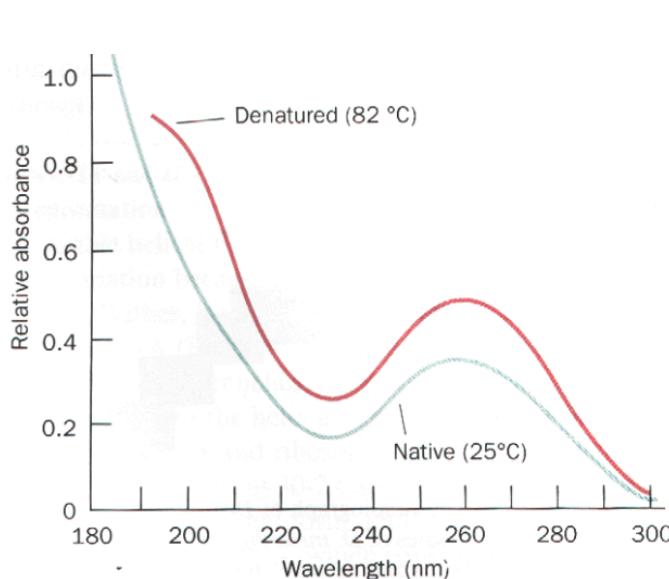
- a. 1000 nt/s in *E. coli*; very fast.



## Lecture 6: DNA Modification & Repair

### A. DNA Duplex Stability

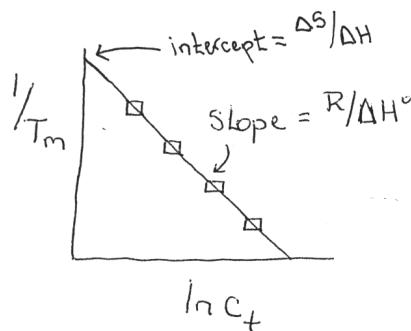
- a. DNA is denatured by heat, high pH, and urea/formamide
- b. Hyperchromicity: denatured DNA absorbs more UV; provides metric for stability
- c. DNA melting curve: plots relative absorbance (at 260 nm) against temperature.



- d.  $T_m$  (melt temperature): when  $\frac{1}{2}$  of the max absorbance has been reached; also when  $\frac{1}{2}$  of the transition from bound to free has occurred.

### B. $T_m$ Equation

- a.  $1/T_m$  is linearly correlated with  $\ln[\text{DNA}]$  with slope  $R/\Delta H^\circ$  and intercept  $\Delta S^\circ/\Delta H^\circ$ .
- b. Plotting the variables allows  $\Delta H^\circ$ ,  $\Delta S^\circ$ , and  $\Delta G^\circ$  to be calculated for the duplex.



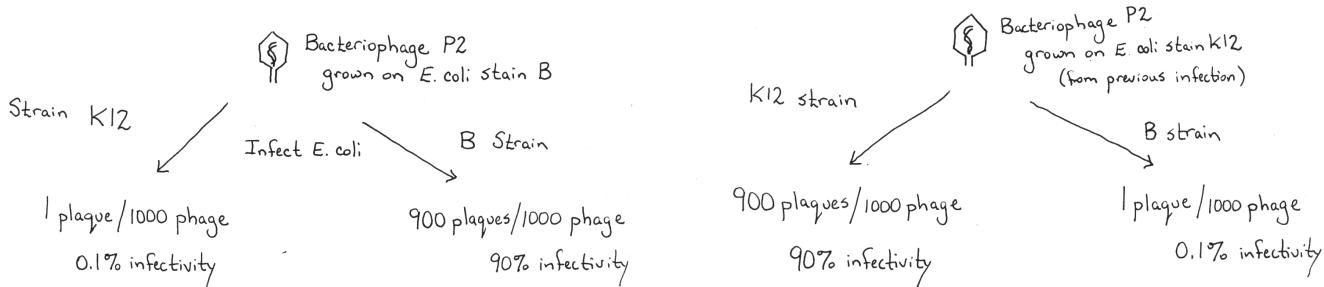
$$\frac{1}{T_m} = \frac{R}{\Delta H^\circ} \ln C_t + \frac{\Delta S^\circ}{\Delta H^\circ}$$

### C. Making incremental changes in DNA and measuring thermodynamics reveals the following:

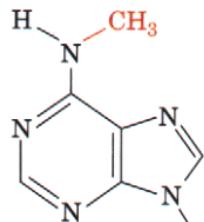
- a. G-C is 1.5-2.0 kcal/mol more stable than A-U
- b.  $T_m$  of duplex increases by  $0.4^\circ\text{C}$  for every % increase in GC content.
- c. Base stacking (more than H-bonding) contributes significantly to duplex stability.

#### D. Phage Restriction

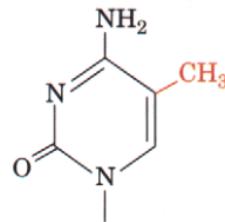
- a. Ability of phage to infect is dependent on the strain on which the phage was grown.
- b. This suggests that the environment of phage growth affects the genotype of the organism, independently of the sequence.



- c. DNA from resisted phages are degraded following infection
- d. The responsible enzymes (restriction modification enzymes) were identified:
- e. Methyl transferase: modifies DNA by transferring methyl from S-adenosyl methionine onto A (N6) or C (C5) at consensus sequences.



**N<sup>6</sup>-Methyladenine (m<sup>6</sup>A)**



**5-Methylcytosine (m<sup>5</sup>C)**

- f. Modification for type II restriction/modification enzymes generally occur within palindromic DNA sequences (ex: GAATTC)
- g. Restriction system consists of a restriction endonuclease (usu dimer) and a cognate methyl transferase (monomer)
- h. Methylation of cognate target sequence gives resistant to cleavage;

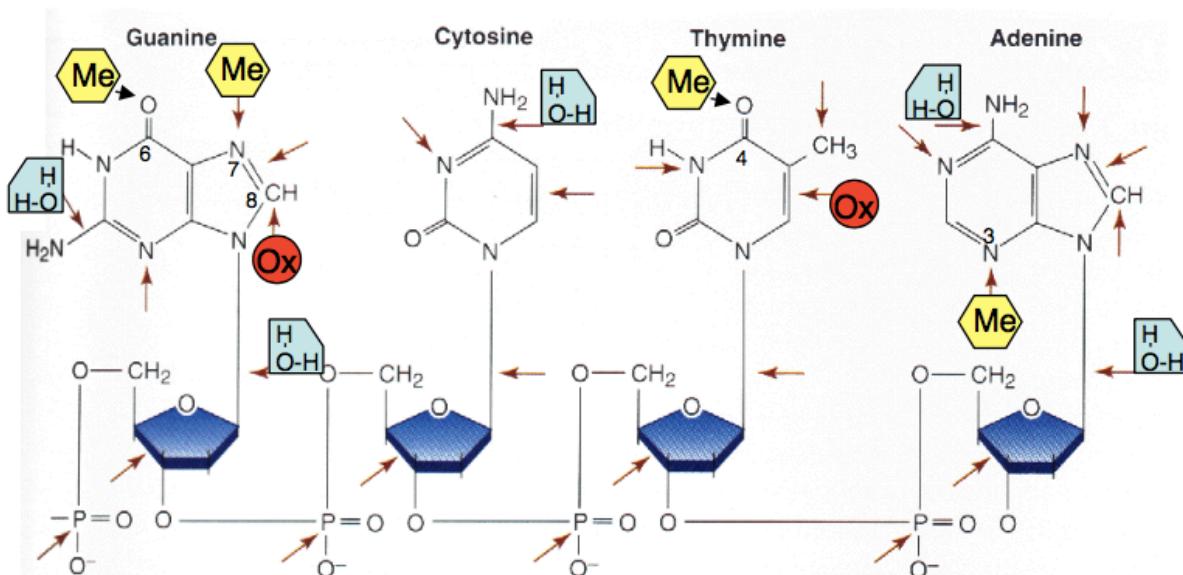
#### E. Epigenetic Modification – DNA Methyltransferase

- a. Methyltransferases recognize the consensus site and utilize S-adenosylmethionine as a cofactor to methylate a base.
- b. Mechanism: flips C out of the helix into the enzyme active site. Gln 237 fills in by H-bonding with G.

#### F. DNA Modification and Damage

- a. Endogenous/spontaneous: originates within the cell.
  - i. Replication errors
  - ii. Inherent instability of DNA
    - 1. Hydrolytic deamination of bases (esp C, also A and G)
    - 2. Hydrolysis of glycosidic bond (depurination), leads to apurinic [AP] sites.
- b. Environmental
  - i. Ionizing radiation: base damage and strand breaks
  - ii. UV produces cyclobutane pyrimidine dimers and other photoproducts
  - iii. Alkylation, acylation, oxidation.
- c. Deficiencies in repair cause up to 90% of human cancers

## G. List of Mutagenic Sites



**FIGURE 15.10**

### DNA sites subject to spontaneous chemical modifications.

Nucleotides are subject to various spontaneous chemical changes at sites indicated by arrows including (1) hydrolytic attack, (2) oxidative damage, and (3) methylation. The frequency and extent of chemical change vary from site to site.

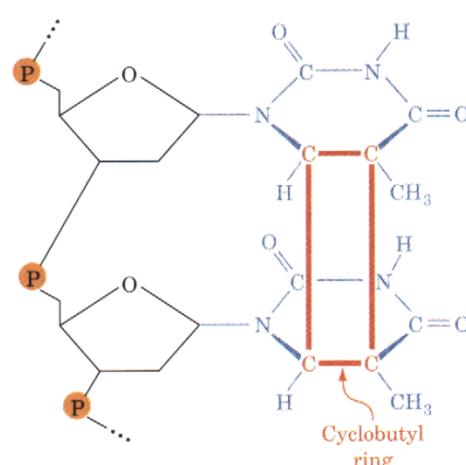
**Ox** Important sites of attack by reactive oxygen species (hydroxyl radicals, peroxide radicals). Significant products: 8-OxoG (mispairs with A), thymine glycol (blocks replication)

**H-O-H** Sites of hydrolytic deamination: **C >> A/G (~50-fold)** and glycosidic bond hydrolysis: **G, A >> C, T (~20-fold)**

**Me** Important sites of alkylation. G-N7 and A-N3 are most frequent, G-O6 and T-O4 are mutagenic

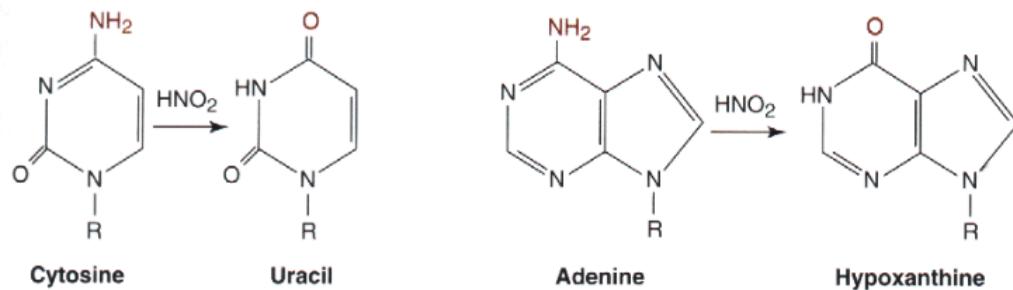
## H. UV Induced Crosslinking

- DNA absorbs at 260 nm (UV range)
- Photoexcitation by UV can connect T-T into thymidine dimers  
This kinks the DNA and blocks replication.
- If you're missing the repair enzymes, you get freckles and skin cancers.



### I. Hydrolytic Deamination (of C and A)

- a. Nitrous oxide
- b. Flips the site

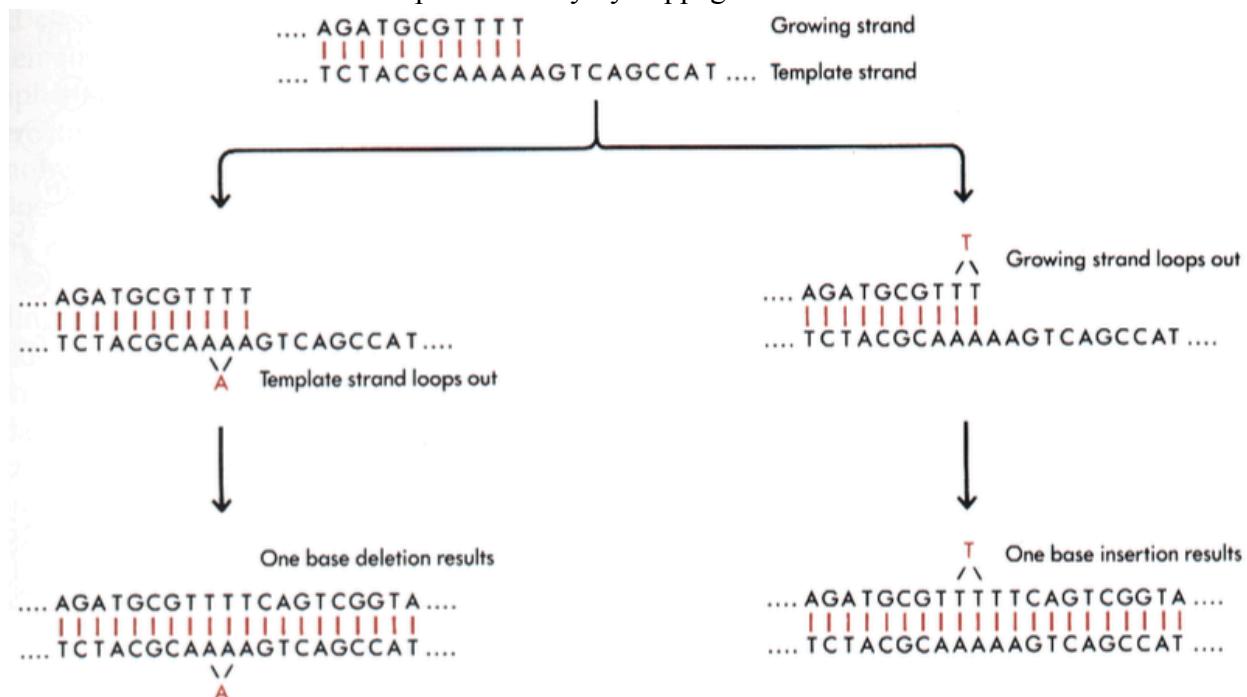


### J. Alkylation and Depurination

- a. DMS or Benz[a]anthracene (from cigarette smoke) alkylate DNA at G(7), A(N1), or C(N3)
- b. What would methylation of A and C do?
- c. G modification depurinates the base.

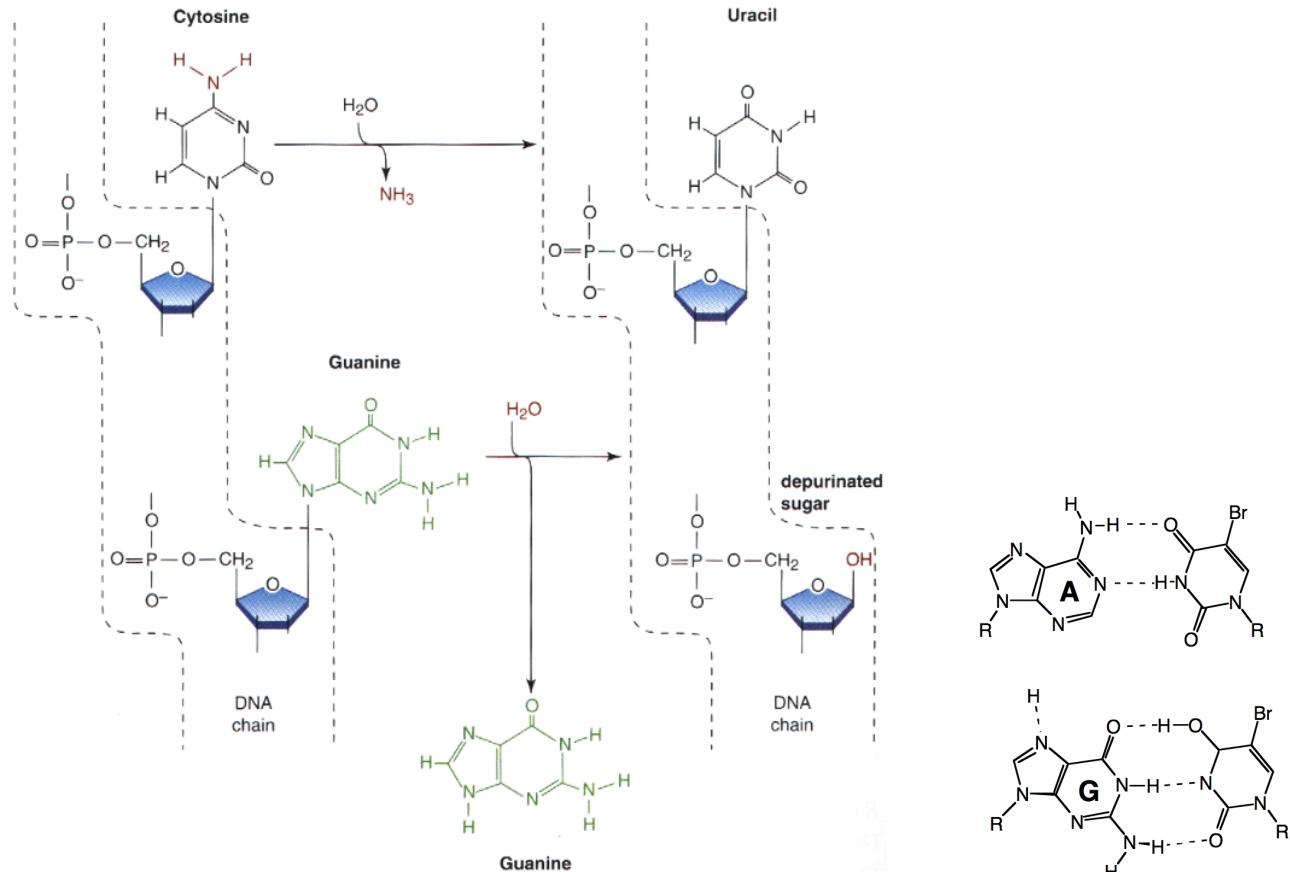
### K. Intercalators and DNA frameshifting

- a. EtBr intercalates between bases.
  - i. Extends the length of the helix
  - ii. Increases error rate; frameshift mutations
- b. Frameshift also occurs spontaneously by slippage of Pol.



## L. Spontaneous Mutations

- 5000 depurination events and 100 deamination events per genome per day.
- Tautomeric shifts in bases result in transition mutations ( $G \longleftrightarrow A$ ) or ( $C \longleftrightarrow T$ )
- Transversions are also observed with purine replaced by pyrimidine, or vice versa.
- Base analogue 5-bromouridine allows enol tautomer and drives transversions.



## M. DNA Repair: Overview

### How do cells deal with all this damage?

Although most of what we know comes from studies in *E.coli*, repair systems are very well conserved and defects in humans often dramatically increase susceptibility to cancers

#### A. Direct Reversal of the damage

Photoreactivation of cyclobutane dimers by photolyase

Removal of methyl and ethyl groups from G-O6, T-O4 and phosphate by Ada (and from A-N1 and C-N3 by AlkB)

#### B. Removal of Damaged Base – Base Excision Repair (BER)

Many specific DNA glycosylases (hydrolyze the glycosidic bond):

- Uracil N-glycosylase (Ung)
- 3-MeA DNA glycosylase II (AlkA: removes 3-MeA as well as 3-MeG and 7-MeG)
- MutM: removes A from A–8-oxoG and A–G mispairs
- MutY: removes 8-oxoG from 8-oxoG–C base pairs

#### C. Removal of a single-strand DNA segment that contains the damage

Nucleotide Excision Repair (NER) – UvrABCD, removes bulky, distorting damage (eg cyclobutane dimers)

Mismatch Repair for replication errors (MutS, L, H)

Very Short Patch Repair (VSP) for T–G mispairs

#### D. Replacement by Recombination

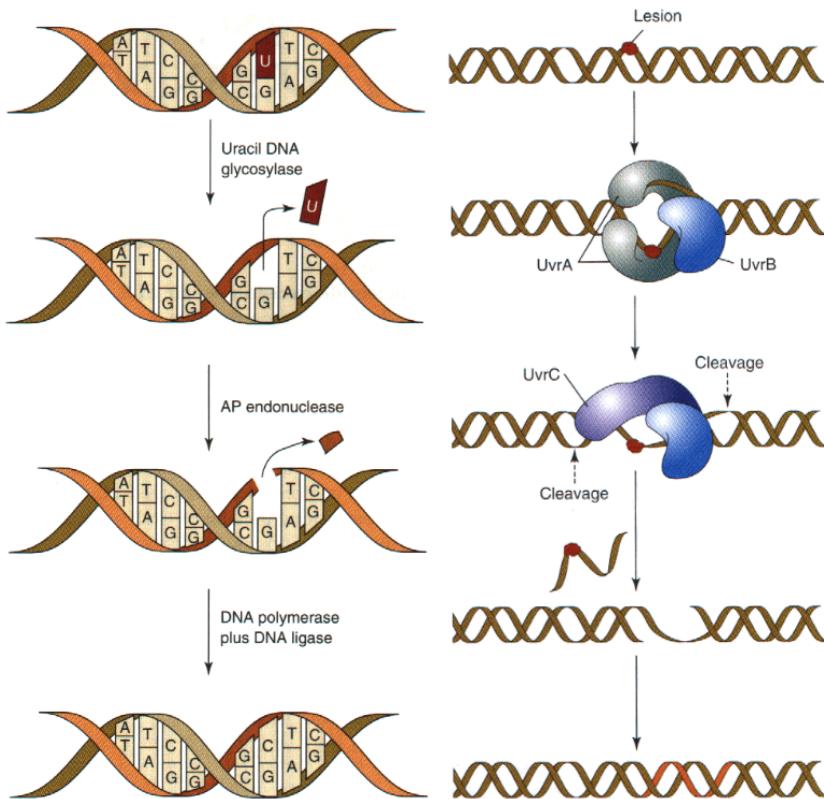
#### E. Damage Bypass – action of last resort!

## N. Glycosylases and Excision Repair

- a. Glycosylases find DNA lesions and break the glycosidic linkage between base & sugar.
  - i. Specific for uracil, hypoxanthine, methylated bases, ring-opened purines, etc.
  - ii. Cannot recognize thymidine (from deamination of 5-methyl C)- hotspots for C → T mutation.
- b. Endonucleases remove the abasic sugar and normal repair follows (Pol and lig).

## O. Nucleotide Excision Repair

- a. Removes thymidine dimers
- b. UvrABC enzymes recognize duplex perturbations, e.g. pyrimidine dimers, chemical adducts, and other bulky modifications.
- c. UvrABC makes 2 single-strand cuts, 1 turn apart (12-13 nt), and the gap is repaired normally by pol + lig.



## P. Photolyase

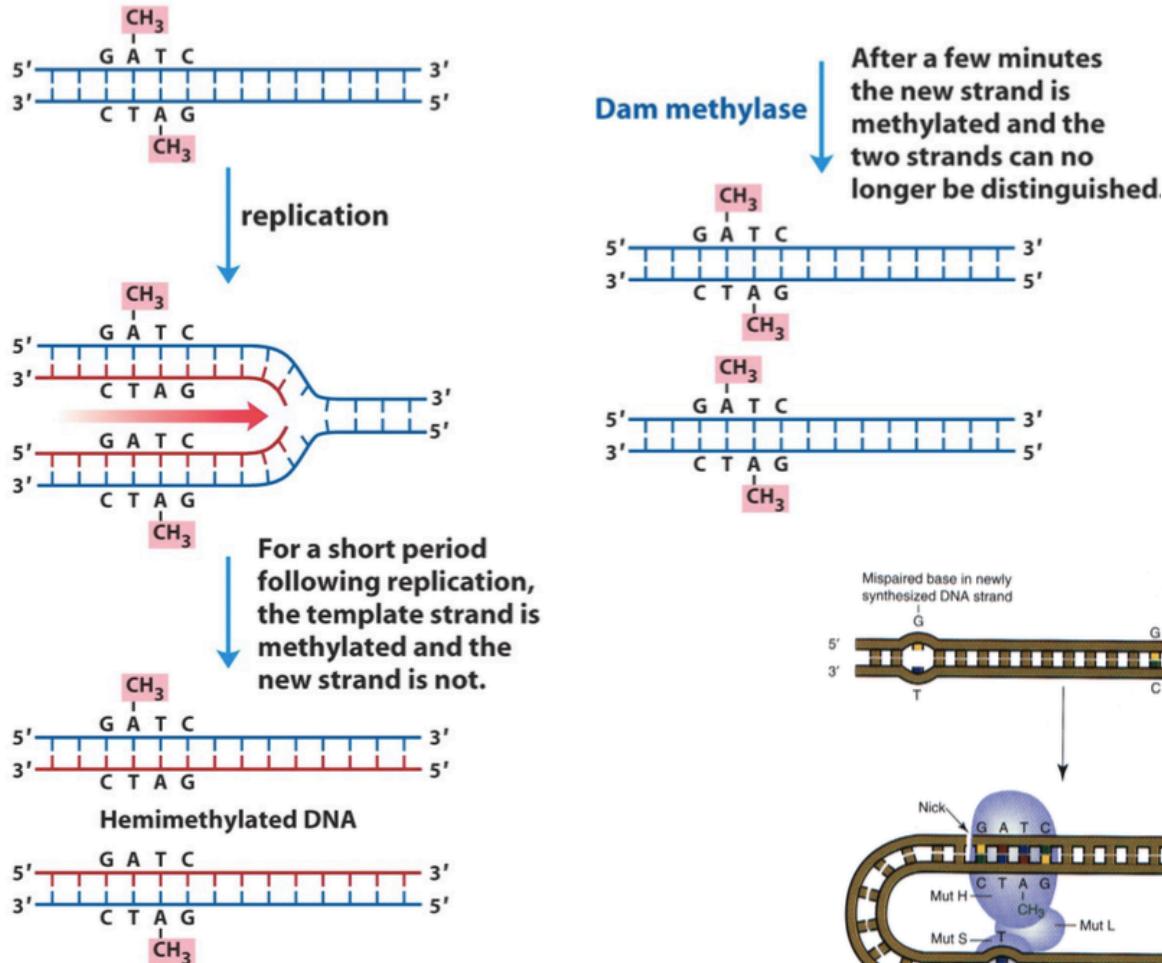
- a. Use visible light to reverse UV damage (pyrimidine dimers).
- b. Mechanism
  - i. 4-deazaflavin absorbs light (300-500 nm, visible) and transfers energy to FADH<sup>-</sup>, which passes it to the pyrimidine dimer.
  - ii. Pyrimidine anion reduces FADH<sup>•</sup> back to FADH<sup>-</sup>.

## Q. Alkyltransferases

- a. Problems
  - i. Methylation at G(O6) by agents like MNNG.  
(How is methylation of G mutagenic? What is the nature of the mutation?)
  - ii. Methylation of the backbone: slows/stalls replication
- b. Solution: Ada (O6-methylguanine-DNA methyltransferase)

- i. Transfers methyl-G to an active site Cys in the protein (coordinated by  $Zn^{2+}$ ).
- ii. Also removes methyl from DNA phosphotriesters
- iii. Sacrificial repair protein; not an enzyme (irreversibly modified by dealkylation).
- iv. Methylated Ada acts as a transcription factor to promote Ada<sup>++</sup>

#### R. Methylation distinguishes old and new DNA in bacteria

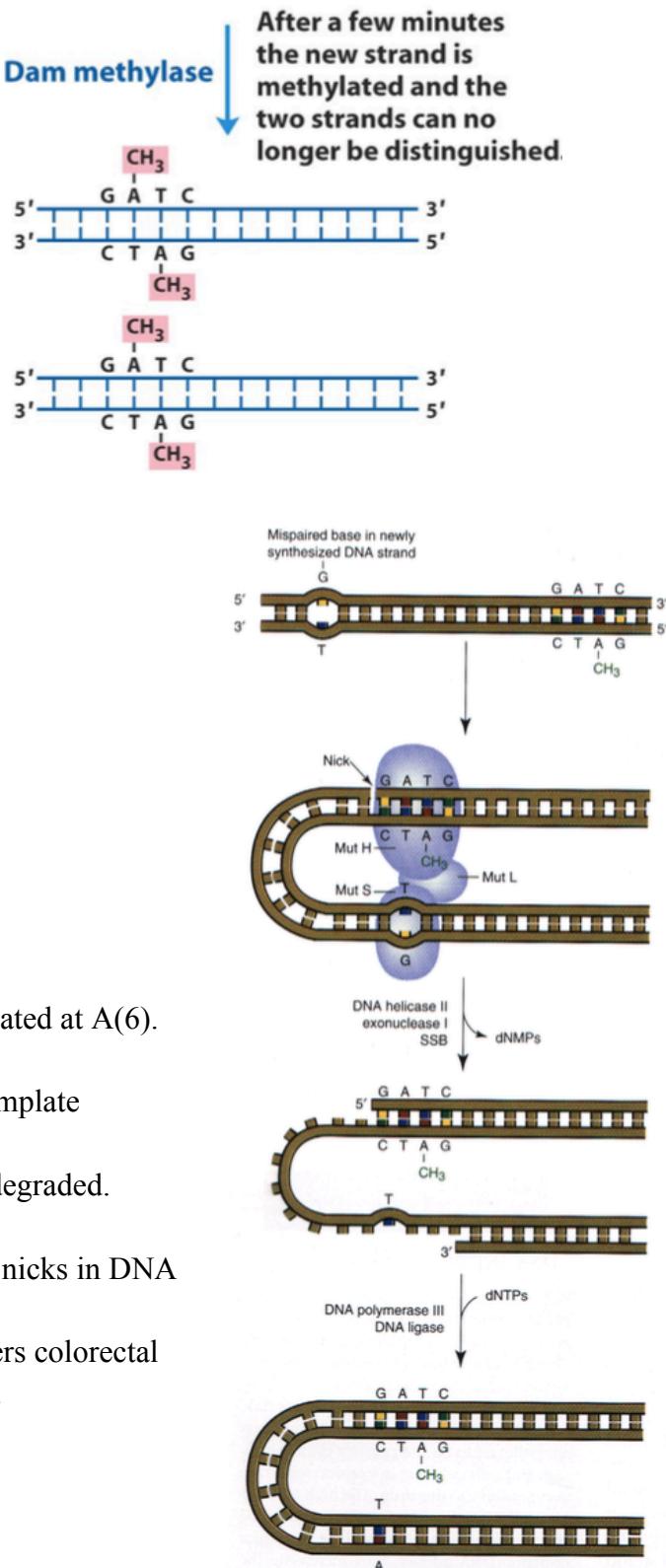


#### S. DNA Mismatch Repair in Bacteria

- a. Template strand (old strand) is methylated at A(6).
- b. MutS recognizes the mismatch
- c. MutH identifies which strand is the template
- d. MutL links the 2.
- e. DNA between CH<sub>3</sub> and mismatch is degraded.

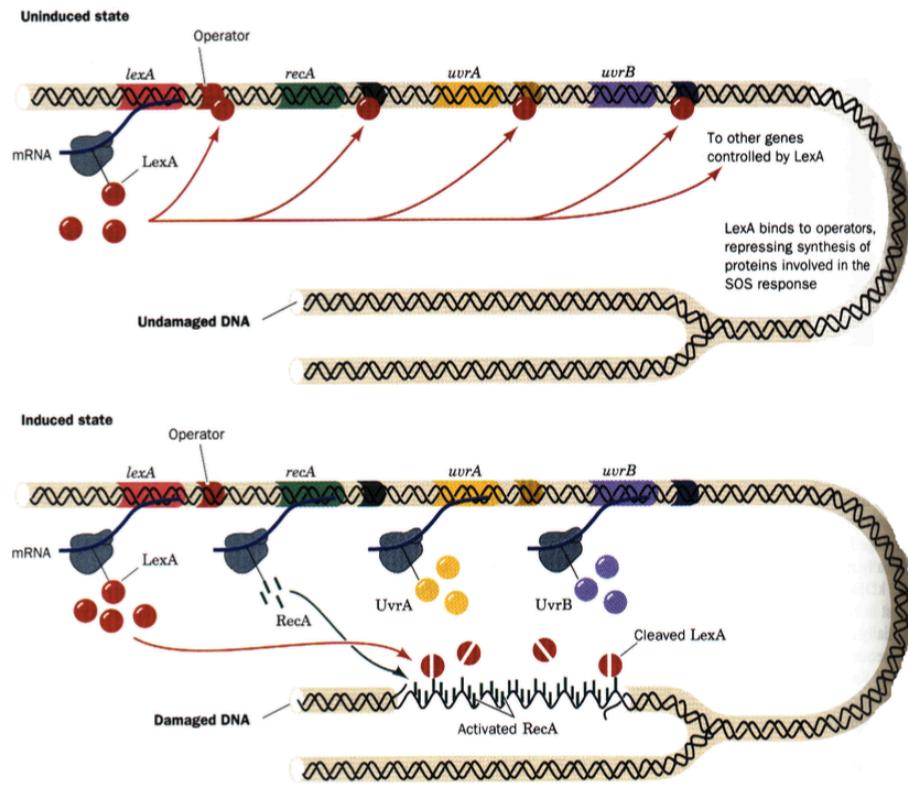
#### T. DNA Mismatch Repair in Eukaryotes

- a. MutS and MutL-like proteins identify nicks in DNA to distinguish new v. old.
- b. Loss of mismatch repair fidelity triggers colorectal cancer- defects in related repair genes.



## U. SOS Repair-Error Prone Repair

- SOS repair genes are induced by ssDNA++ from severe damage that stops DNA synthesis.
- RecA binds ssDNA → autocleavage of LexA → activation of SOS genes
- Effect: random incorporation of nucleotides around damaged areas.  
Able to repair AP sites 1/4 of the time.



## V. Summary

■ TABLE 15-2. Repair systems in *E. coli*

General mode of operation	Example	Type of lesion repaired	Mechanism
Direct removal of lesions	Alkyltransferases	O-6-alkylguanine	Transfer alkyl group from O-6-guanine to cysteine residue on transferase
	Photoreactivating enzyme (PRE)	UV photodimers	Splits dimers in the presence of white light
General excision	<i>uvrABC</i> -encoded exonuclease system	Lesions causing distortions in double helix, such as UV photoproducts, and bulky chemical adducts	Makes endonucleolytic cut on either side of lesion; resulting gap is repaired by DNA polymerase I and DNA ligase
Specific excision	AP endonuclease	AP sites	Makes endonucleolytic cut; exonuclease creates gap, which is repaired by polymerase I and ligase
	DNA glycosylases	Certain altered bases, such as deaminated bases (uracil, hypoxanthine), certain methylated bases, ring-opened purines, and other modified bases	Removes base, creating AP site, which is repaired by AP endonucleases
Postreplication	Mismatch-repair system	Replication errors resulting in base-pair mismatches	This system recognizes which strand is correct by detecting methylated A residues that are part of the 5'-GATC-3' sequence, and then excises bases from the newly synthesized strand when a mismatch is detected
	Recombinational repair	Lesions that block replication and result in single-stranded gaps	Recombinational exchange
	SOS system	Lesions that block replication	Allow replication bypass of blocking lesion, resulting in frequent mutations across from lesion

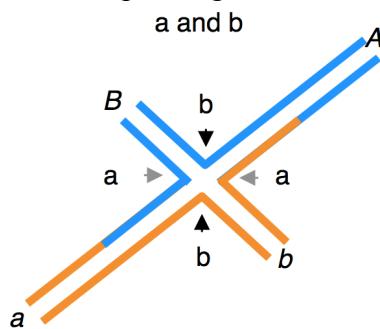
# Lecture 7: Recombination & Transposition

## A. Roles of Recombination

- a. DNA repair
- b. **DNA replication fork rescue**
- c. Meiotic chromosome segregation
- d. Telomere maintenance

## B. Holliday Model

- a. Strand exchange of 2 nicked homologous DNA segments.
- b. Ligation creates the Holliday Junction, which migrates to create the heteroduplex region.



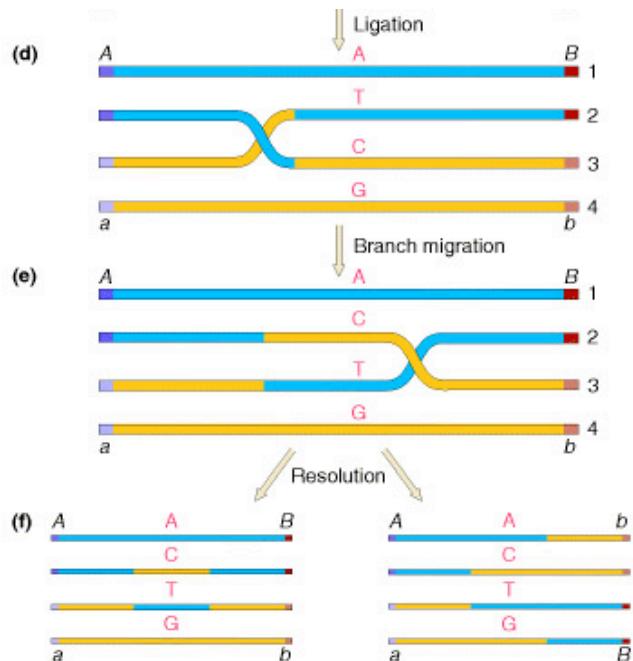
- c. Resolved by cutting 2 of the strands

- i. Horizontal cut (a, left) is non-crossover because only the middle part swapped.
- ii. Vertical cut (b, right) is crossover because the ends swapped.
- iii. Crossover comes from an odd number of vertical cuts.

## C. Isolation of First Recombination Mutant (recA) Based on Conjugational Recombination

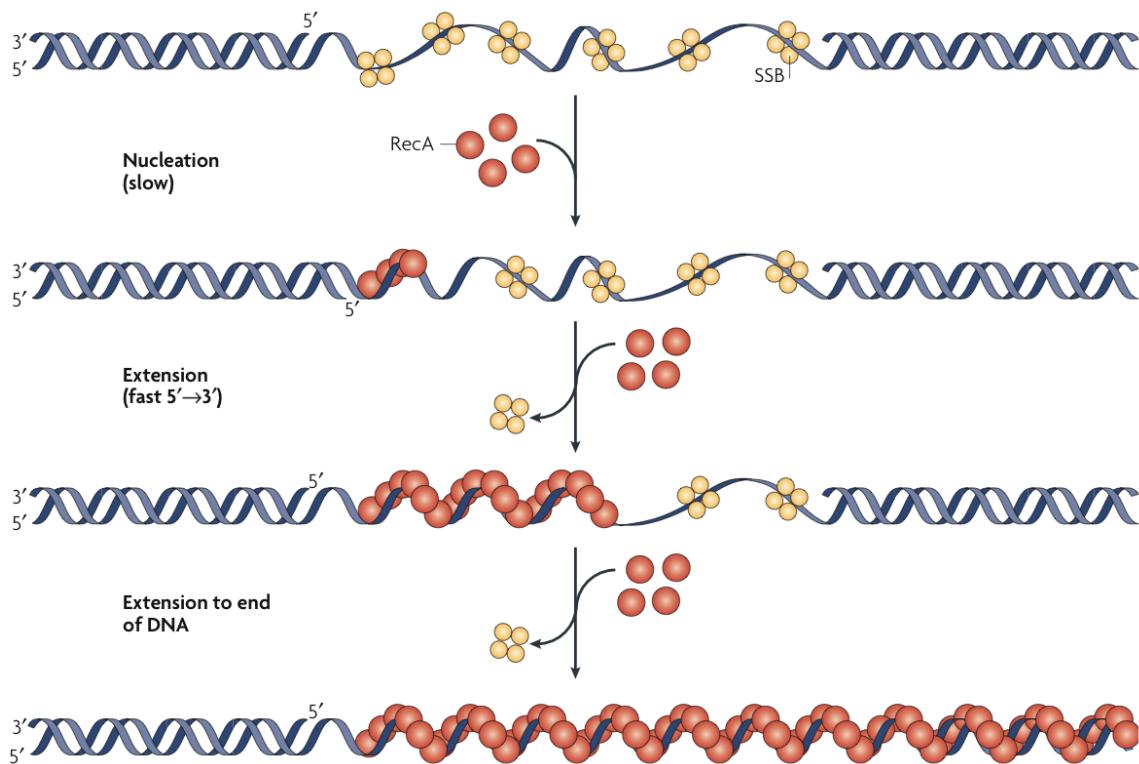
- a. A bunch of F- bacteria were randomly mutated and marked with Leu-, Ade+
- b. Each mutant was mated with oppositely marked F+ bacteria (Leu+, Ade-) and the progeny was scored for Leu+, Ade+ recombinants.
- c. 2 (recA) mutants were unable to undergo conjugational recombination and easily die upon UV (damaged DNA repair).

## D. Factors in the E. coli RecBC Pathway



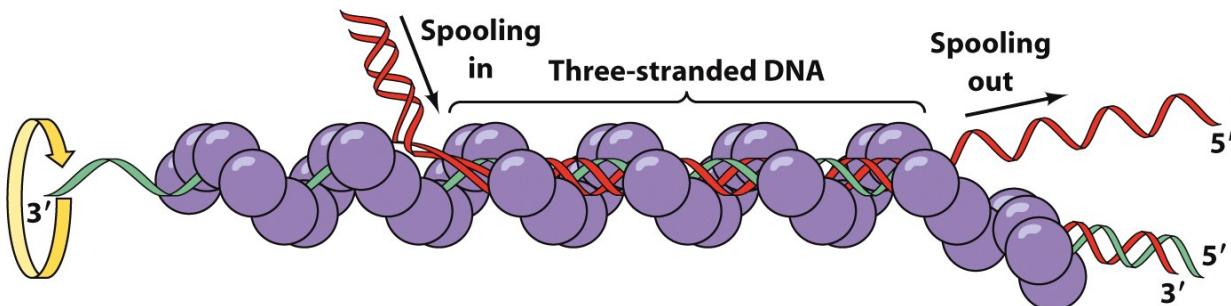
<u>Factors</u>	<u>Function</u>	<u>Mutant Phenotype</u>
recA:	Strand transfer; formation of HJ	No recombination
recBCD:	Generate ssDNA needed for HJ	1% recombination
rvuAB:	Branch migration	Mild recombination defect
rvuC:	Resolves HJ	Mild recombination defect

### E. Assembly of RecA filaments on ssDNA



- a. RecA decoration necessary for homologous recombination.
- b. Slow nucleation; aided by other proteins (e.g. RecFOR)
- c. Rapid growth of RecA-ATP (5'-3') filament (1 every 3 bases).
- d. RecA-DNA is extended ~50% to 18 bases/turn.
- e. Dissociation (5'-3') by ATP hydrolysis.

### F. Spooling Model of RecA-ssDNA + duplex DNA



- a. RecA holds ssDNA in an extended B-DNA conformation
- b. Spooling into 3-stranded helix.

### G. Recombination and the Chi (X) state

- a. Chi sequence: (crossover-hotspot-instigator): 8 nt recombination hotspot.
- b. Asymmetric and unidirectional (active when RecBCD approaches from 3' end)
- c. 13x more likely than random sequence; 1,000 occurrences in *E. coli* genome.
- d. In *E. coli*, 90% of Chi sites have 3' end pointing to terminus of replication.  
This is because you're always adding nts 5'-3', and if there's a problem (i.e. nick) that leads to a DSB, you want to be able to hit the Chi site while moving back, from 3'-5'.
- e. Creates highly invasive ssDNA to search for sequence complementarity.

## H. Two models of RecA-DNA pairing

- Base-triplet hypothesis: non-WC H-bonds form a triple-helix.
- Base-pair hypothesis: base pairs switch; the ssDNA sneaks in to sometimes WC-bp with 1 strand of the duplex.
- Preferential switched base-pairing at AT sites during homologous recognition.
- Just 6 nt can initiate pairing.

## I. Creation of ss-ends by RecBCD: Influence of Chi

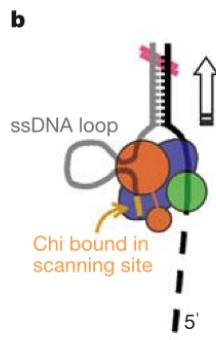
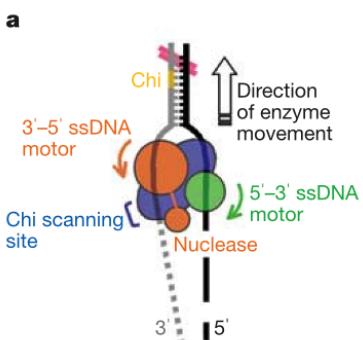
### a. Process

- 1) ds-break from DNA damage or **collapsed replication fork**.
- 2) RecBCD helicase binds to ds-break and begins unwinding dsDNA.
- 3) Digestion of 3' tail and sparse cuts to 5' tail
- 4) Chi site is bound to scanning site; triggers pause and flip: digestion of 5' tail and preservation of 3' tail.
- 5) RecA loaded on 3' tail
- 6) RecBCD dissociates

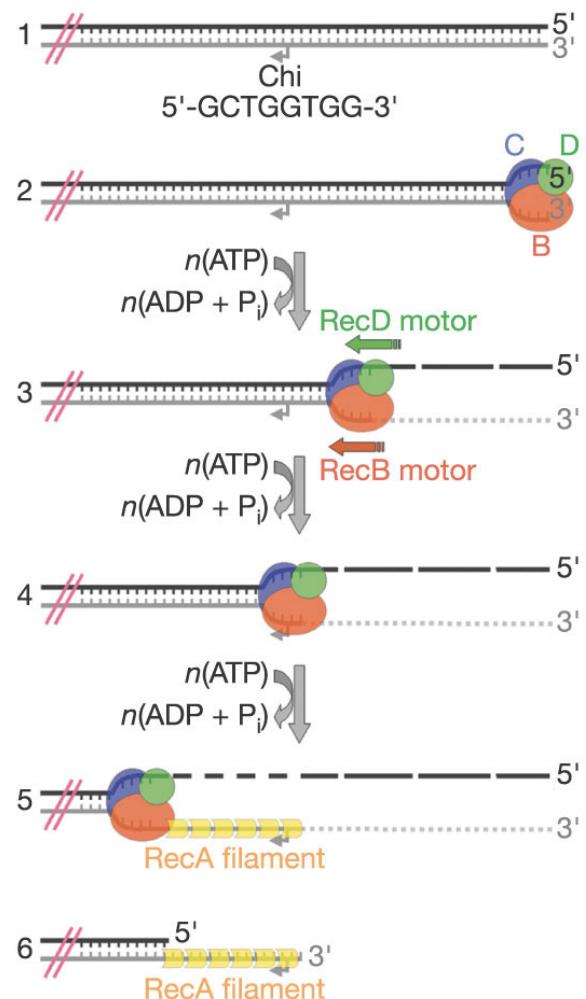
RecB: endonuclease and 3'-5' helicase

RecC: splits duplex, recognizes chi

RecD: 5'-3' helicase



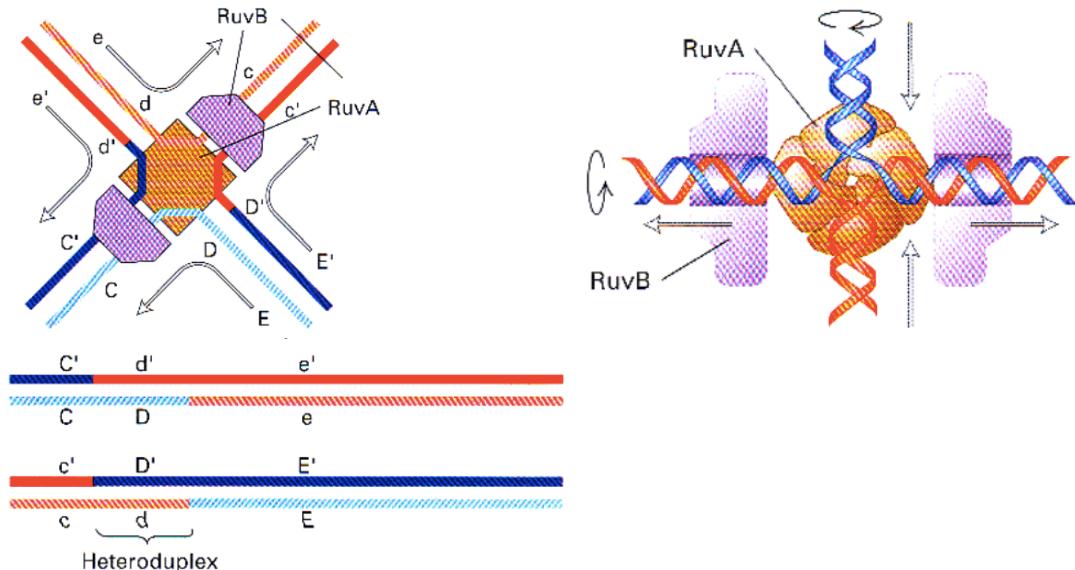
**Figure 5** Diagram outlining the changes in RecBCD that occur after encountering a Chi site. **a**, Before Chi, the enzyme progresses along duplex DNA using the bipolar motors of the RecB (orange) and RecD (green) subunits. The 3' tail is digested processively but the 5' tail is cut much less frequently. **b**, On encountering Chi, the RecC subunit (blue) binds tightly to the 3' tail, preventing further digestion of this strand. The 5' tail is now able to access the nuclease site more frequently and is degraded more fully. The enzyme continues to advance along the DNA resulting in a loop out from the RecB subunit that can be loaded with RecA protein.



- Preferential degradation is dependent on endonuclease active site accessibility (3' then 5').
- RecA binds to the ssDNA loop formed by chi-site halting (pauses and slows by  $\frac{1}{2}x$ )

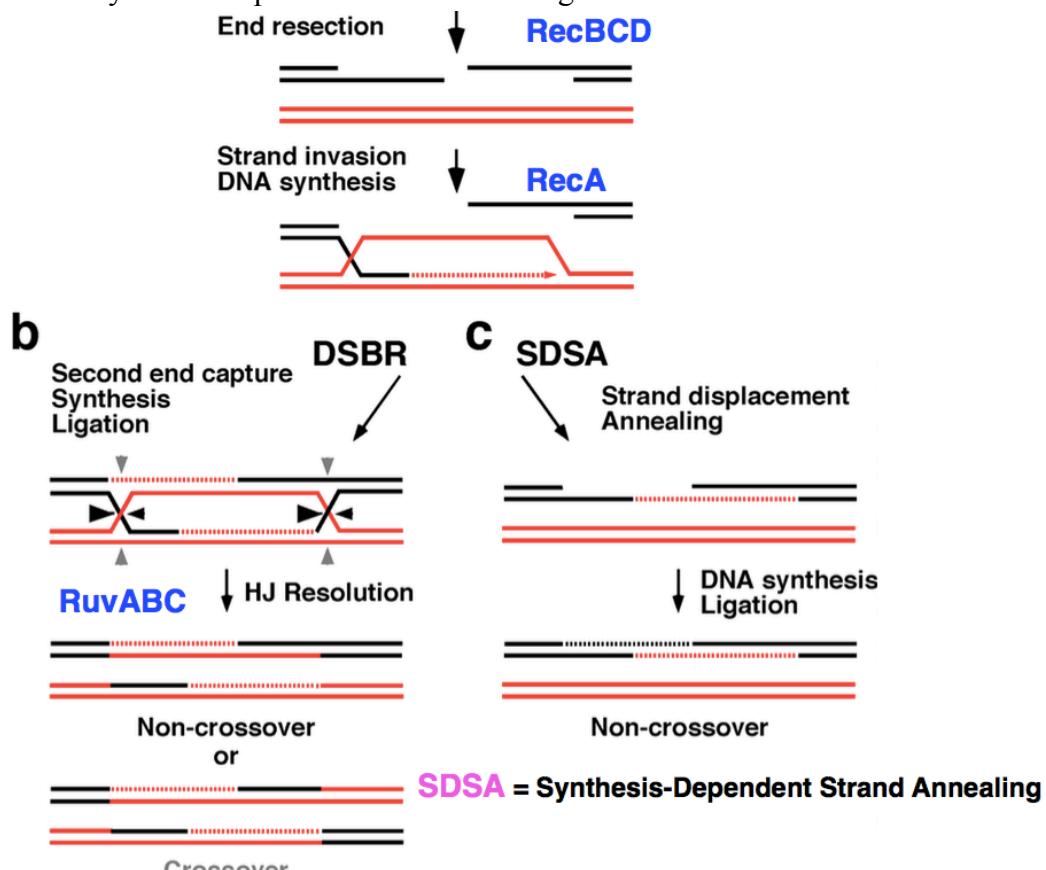
### J. General Recombination Factors: RuvABC

- RuvA binds the Holliday Junction and recruits RuvB.
- RuvAB is a “molecular screw” that promotes branch migration.
- RuvC is an endonuclease that cuts across the Holliday junction.
- The strands are ligated to generate recombinants.



### K. Recombinational Repair of ds-breaks

- DSBR: double-stranded break repair
- SDSA: synthesis-dependent strand annealing



**DSBR** = Double-Strand Break Repair

Sung & Klein, 2006, Nature Reviews MCB

## L. Recombinational DNA Repair of Stalled Replication Forks

- Replication forks stall at lesions/nicks on 1 strand.
- RecA is always involved, the others are often there too.
- Lesions include mismatches, modified bases, abasic sites, and dimerization (T-T).

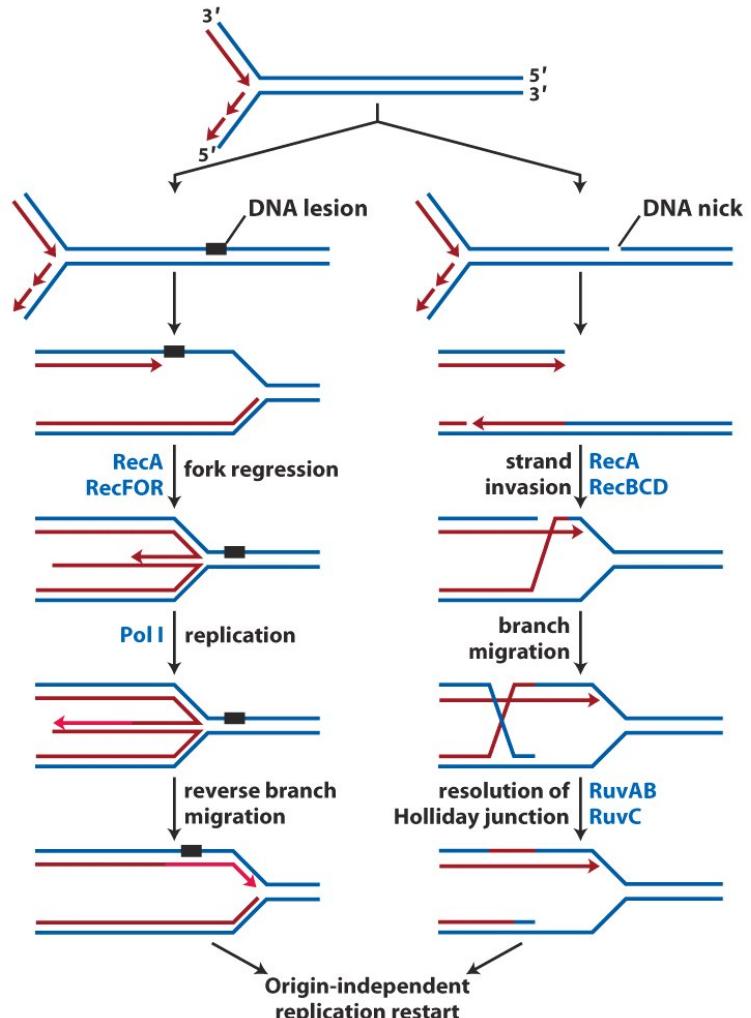
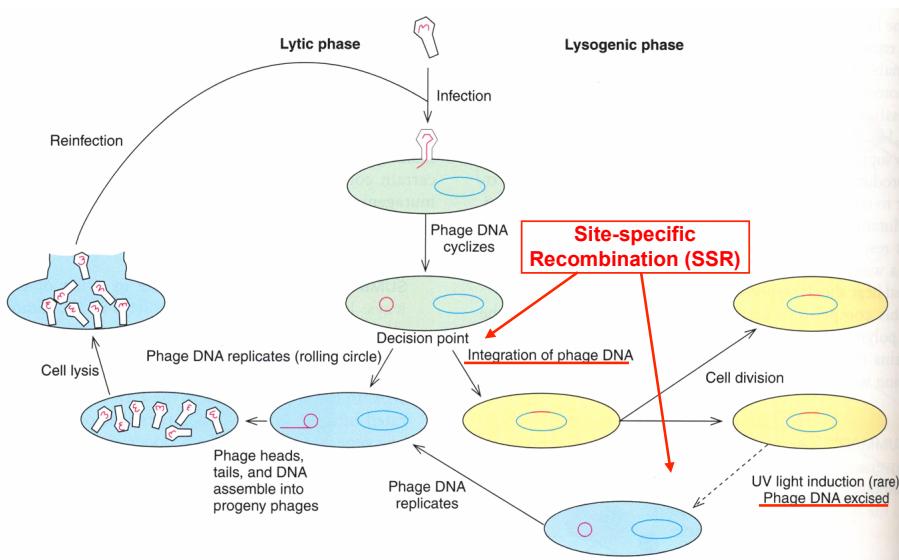
## M. DNA Lesions

- Strand 1 hits the lesion and strand 2 progresses past it.
- Fork regression:  
Strand 2 dissociates from the template, bends back, and is the template for strand 1.
- Reverse branch migration:  
Extended strand 1 base-pairs past the lesion and continues replication.
- Remaining lesion is removed via lecture 6 mechanisms.

## N. DNA nick

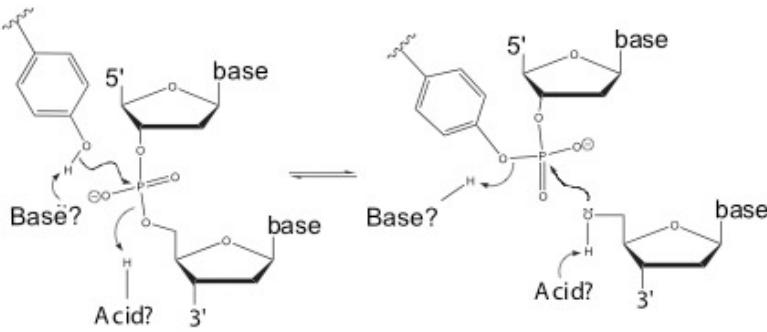
- Replication hits the nick.
- Strand 1 forms a ds-break.
- Strand 2 extension ligated to the strand 1 template.
- RecABCD-mediates strand 1 invading strand 2.
- Branch migration and 2<sup>nd</sup> invasion of strand 1 into strand 2.
- Resolution and continuation of replication.

## O. Site-specific Recombination in Lambda Phage Triggers Lysogeny (lambda prophage)



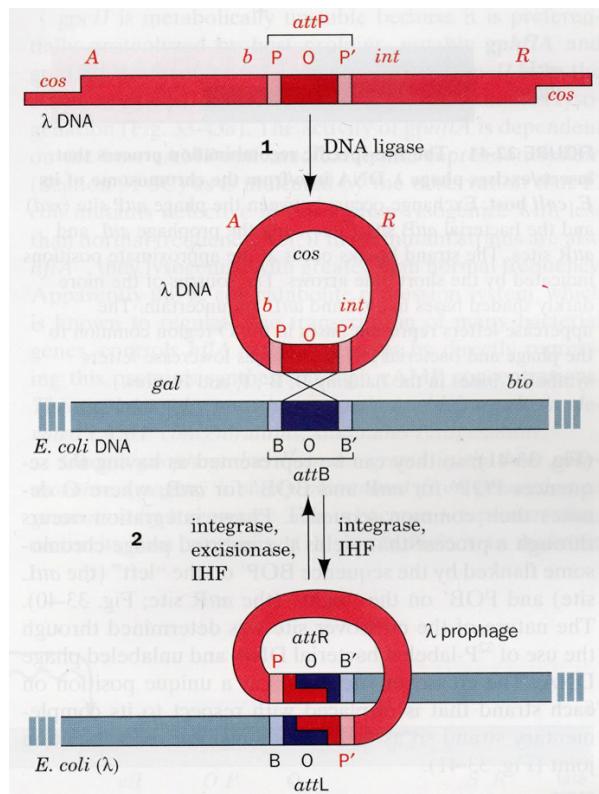
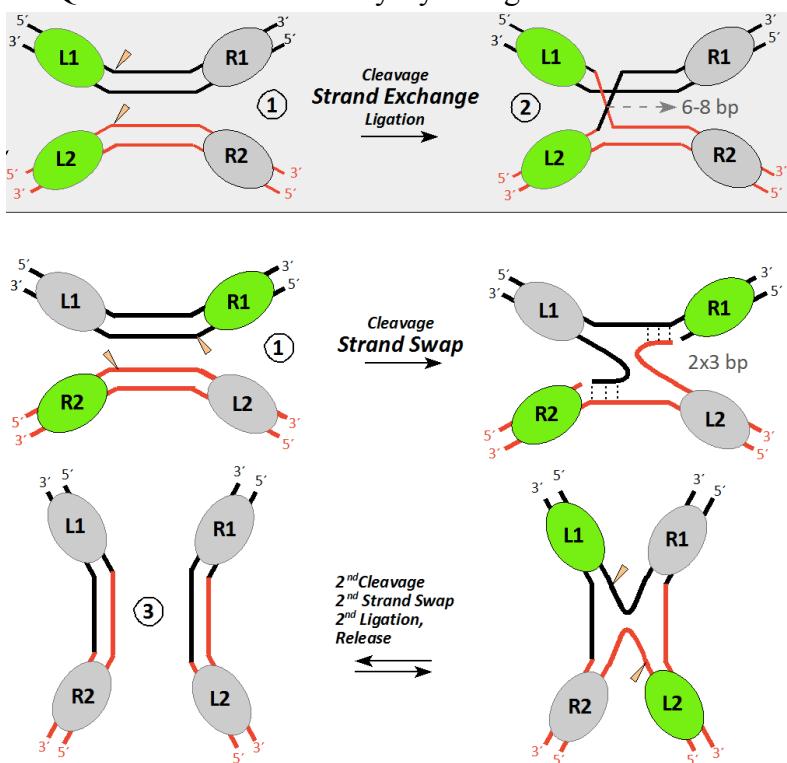
### P. Lambda Integration: Site-Specific Recombination (SSR) by a Tyrosine Recombinase

- Lambda integrase (Int) triggers SSR between phage attP and host attB sites.
- Cleavages initiated by conserved Tyr without ATP.
- Phosphodiester bond energy stored in transient phosphotyrosyl linkage to DNA 3' end.



- Product contains 7 bp heteroduplex

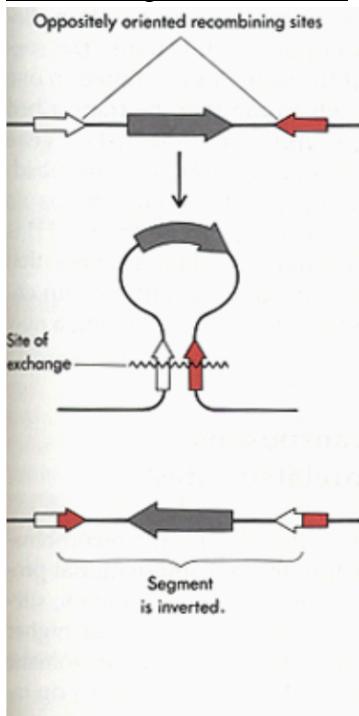
### Q. Mechanism of SSR by Tyr-integrase



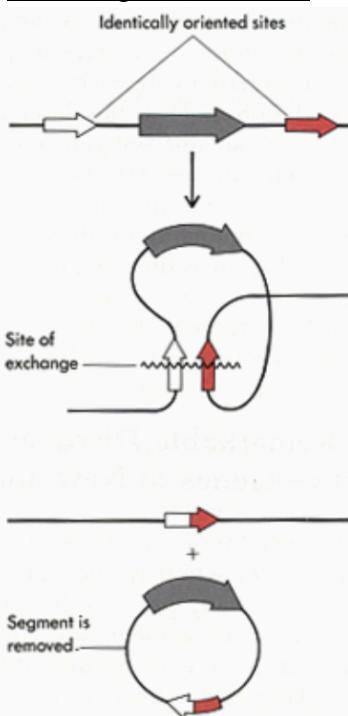
**FIGURE 33-40** Site-specific recombination in bacteriophage λ. This schematic diagram shows (1) the circularization of the linear phage λ DNA through base pairing between its complementary ends to form the *cos* site; and (2) the integration/excision of this DNA into/from the *E. coli* chromosome through site-specific recombination between the phage *attP* and host *attB* sites. The darker colored regions in the *att* sites represent the identical 15-bp crossover sequences (O), whereas the lighter colored regions symbolize the unique sequences of bacterial (B and B') and phage (P and P') origin. [After Landy, A. and Weisberg, R.A., in Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A. (Eds.), *Lambda II*, p. 212, Cold Spring Harbor Laboratory (1983).]

## R. Site-specific recombination can have distinct outcomes

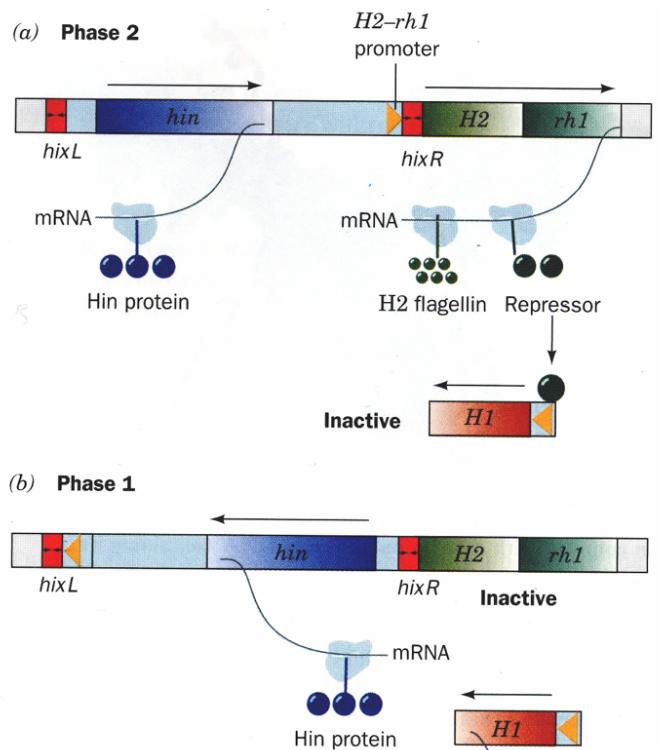
### Inverted repeats: inversion



### Direct repeats: excision



### (a) Phase 2



- a. Resolvase can catalyze the direct-repeat/excision reaction.

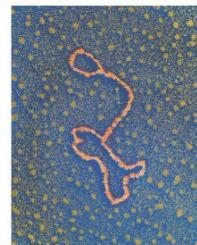
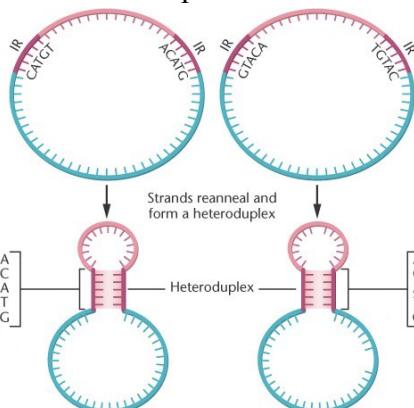
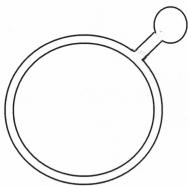
## S. SSR regulates gene expression in Salmonella

- a. Phase variation: switching antigenic proteins that make flagella
- b. Phase 2: H2-rh1 promoter synthesizes H2-flagellin and H1 repressor.
- c. Phase 1: Segment inversion removes promoter. H1-flagellin gene released from repression.
- d. Recombinase (Hin) protein mediates inversion and is expressed in either inversion. Closely related to resolvase.

## T. Transposons

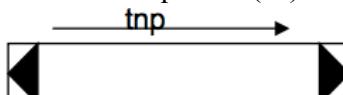
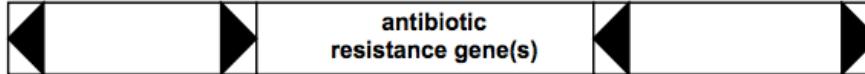
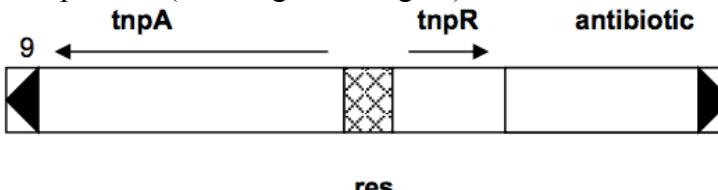
- a. R factors (resistance factors) move between bacteria via conjugation.
- b. Heteroduplex mapping found that R factor addition occurs with new sequences flanked by inverted repeats.
- c. Method: make heteroduplex of new vs old plasmid.

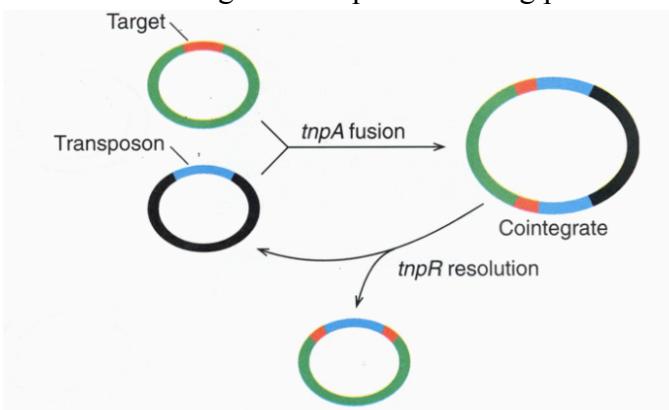
Make heteroduplex of new versus old plasmid:



- d. Transposition (Illegitimate Recombination)

- i. Transposition is immensely important for evolution of genomes

- ii. Transposons (Tn) are ubiquitous segments of DNA:
  - 1. Have terminal inverted (imperfect) repeats: 20-40 bp
  - 2. Encode transposase (tnp or tnpA) to recognize these ends
  - 3. Can move anywhere, creating precise duplication of the target sequence (3-13 bp) flanking their ends.
  - 4. Movement is regulated:  $10^{-5}$  to  $10^{-7}$  per generation
  - 5. Can catalyze other DNA rearrangements
- iii. HIV retrovirus in its DNA phase is a transposon
- iv. Rearrangements of immunoglobulin genes that result in antibody diversity occur by transpositional recombination.
- e. Classes of Bacterial Transposons
  - i. Insertion Sequence (IS)
 
  - ii. Composite Tn (includes R factors): two IS flanking some gene.
 
  - iii. Complex Tn (Tn3 or gamma-sigma): contains resolvase (res).
 
- f. Simple transposition (cut and paste)
  - i. Occurs in IS and composite Tn
  - ii. Both strands move and reconnect at the new site, leaving donor DNA ds-break.
- g. Replicative transposition (copy and paste)
  - i. Occurs in complex Tn
  - ii. Requires transposase and resolvase (site-specific recombinase, coded by tnpR)
  - iii. Step 1: tnpA gp catalyzes fusion with the target to form cointegrate  
Transposon replicates during cointegrate formation
  - iv. Step 2: tnpR gp resolves cointegrate into the target plasmid + inserted transposon, as well as the original transposon-bearing plasmid.



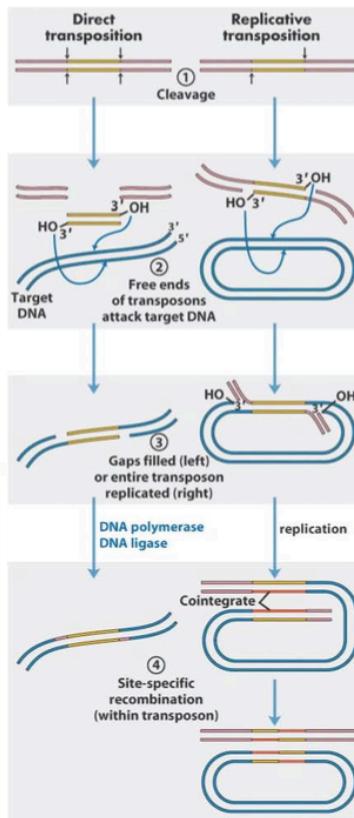
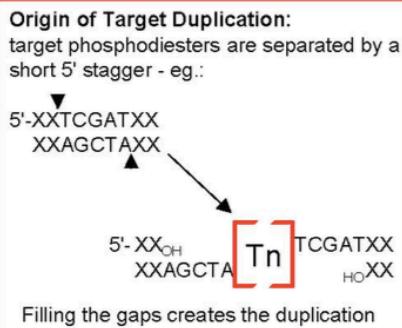
## U. How DNA Transposons Move

### Non-replicative (eg. Tn10)

Synaptic complex (transposase + 2 Tn ends formed. Transposase cuts both DNA strands at each Tn end, releasing Tn (note DSB in vector)

The 3'-OH ends attack the target phosphodiester backbone, resulting in a simultaneous cleavage of the target and joining of the Tn 3' end to the target 5'-P

Gaps at junctions repaired (Pol I + ligase)



### Replicative (eg Tn3, γδ)

Synaptic complex (transposase + 2 Tn ends formed. Transposase cuts single DNA strand at each Tn 3'-end

The 3'-OH ends attack the target phosphodiester backbone, resulting in a simultaneous cleavage of the target and joining of the Tn 3'-end to the target 5'-P

Since the 5'-Tn ends remain attached to the vector, the new junctions with target are equivalent to replication forks.

Replication of the Tn sequence (Pol III etc) followed by ligation to the 5'-end of the vector creates a cointegrate of donor and target

Some replicative Tns (incl. Tn3 & γδ) encode a Site-specific Recombinase (Resolvase) that excises the vector + one Tn copy from the cointegrate

## V. Review of Recombination Principles

### • HOMOLOGOUS RECOMBINATION

Critical for DNA repair

Requires recA

Requires long stretches of (nearly perfect) homology

Requires ATP hydrolysis for synapsis

Initiated by ssDNA, generated by RecBCD, which initially recognizes a duplex end

Produces a long heteroduplex (~1000 bp)

### • SITE-SPECIFIC RECOMBINATION

Requires a site-specific recombinase (e.g. Int, Hin or resolvase)

Two sites share a short stretch of sequence identity

Energy of phosphodiester bond captured in DNA-protein intermediate

Produces a short heteroduplex (2 bp for Ser-recombinases, 6-8 bp for Tyr-recombinases)

### • TRANSPOSITION

Requires transposase

No DNA homology required

Produces short direct repeats flanking the transposon

Initial strand transfer involves only the Tn 3' ends

DNA repair required (and sometimes complete Tn replication)

### IMPORTANCE OF TRANSPOSITION

- Genome remodeling
- Spread of antibiotic resistance
- Integration of DNA copy of HIV genome into human DNA
- Causes mutation in target gene



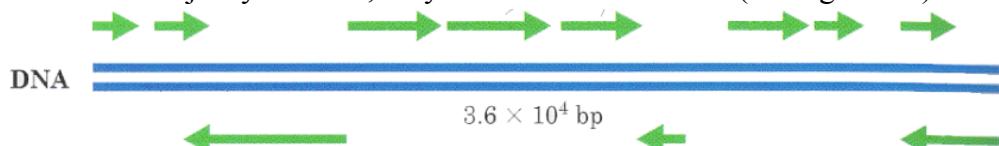
# Lecture 8: Transcription in Bacteria

## A. Overview of Transcription

- RNA polymerase (RNAP) catalyzes exclusively 5'-3' complementary addition of NTPs.
- A primary transcript starts with either pppA or pppG and ends with 3' OH.
- RNAP lacks editing function; accessory proteins carry out editing.
- DNA locally denatured at transcription start site forms an open complex before the first dNTPs are added
- RNAP extends RNA strand (elongation) until termination site.

## B. Orientation of the RNA Relative to the DNA

- In the vast majority of cases, only 1 strand is transcribed (coding strand).

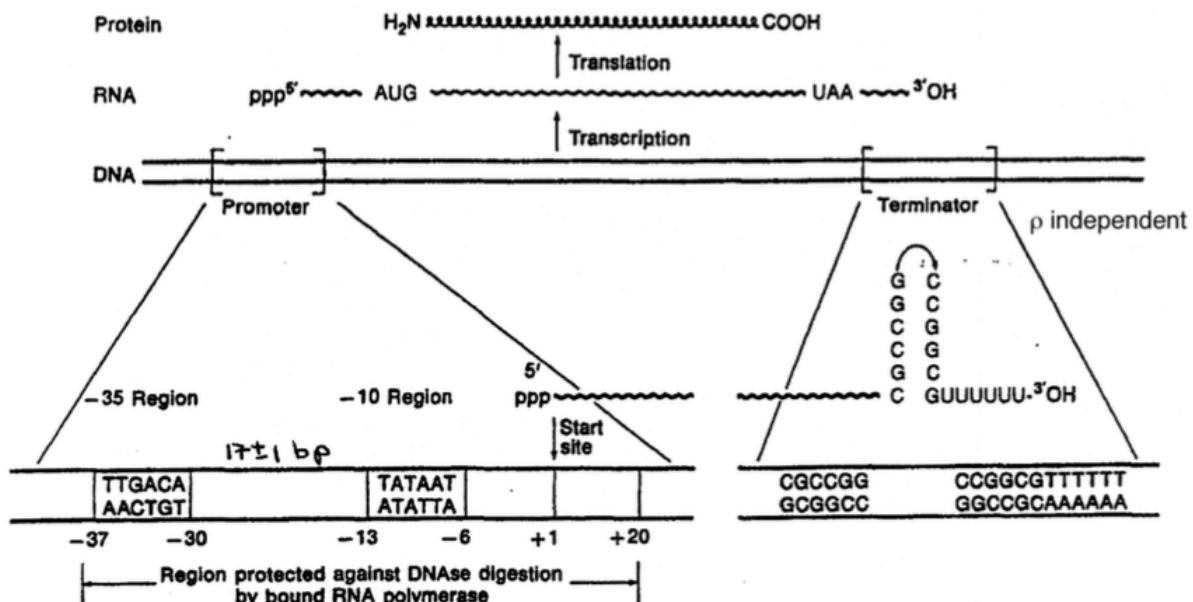


## C. Visualization of Transcription

- Simultaneous transcription on the same gene is okay.
- Genes that code rRNAs: very active; limited by sterics.
- Christmas tree architecture: imaging gives start, stop, and direction.
  - When mRNAs are decorated with ribosomes.
  - Widens from initiation to termination because RNA gets longer as it grows
- In prokaryotes: transcription coupled to translation. Ribosomes instantly engage mRNA.

## D. The Typical Bacterial Transcription Unit

- Transcription is a key upstream control point (flux)



**Figure 21-7**  
Important features of a typical transcription unit. DNA is shown with promoter and terminator regions expanded below. RNA is transcribed starting in the promoter region at +1 and ending after the stem and loop of the terminator. The protein resulting from translation of this RNA is shown above with its N and C termini indicated.

- b. RNA has untranslated regions (UTRs) before and after the start/stop codons; regulatory.  
Short in prokaryotes, very long in eukaryotes.
- c. Terminators can be Rho-dependent or Rho-independent.
- d. The signal for termination is found in the RNA itself.

#### E. Bacterial RNA Polymerase (RNAP)

- a.  $\beta'$ : DNA binding
- b.  $\beta$ : pol catalytic site (Zn)
- c. alpha: interacts with regulatory proteins
- d. sigma: recognizes promoter; initiates transcription; confers specificity.
- e. omega: promotes RNAP assembly
- f. GreA, GreB improve RNAP fidelity by cleaving stalled transcription elongation complexes about 2-10 nts from the 3' end. RNAP then restarts.

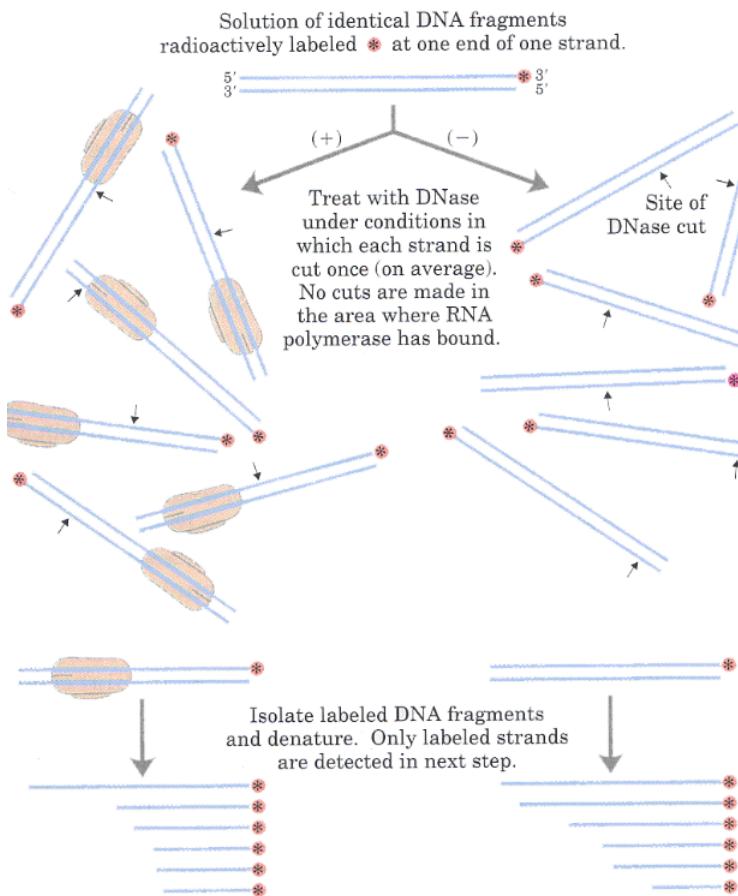
#### F. Identifying Promoter Elements

- a. +1 marks the RNA start site.
- b. There are -10 region and -35 promoter regions.
- c. Deviating from the consensus sequence downregulates gene activity.
- d. Closer to perfect consensus = stronger transcription, and vice versa.

This is critical for control. Weak promoters are more suitable for regulation

#### G. Identifying Promoter Elements: Enzymatic Footprinting of RNAP on DNA.

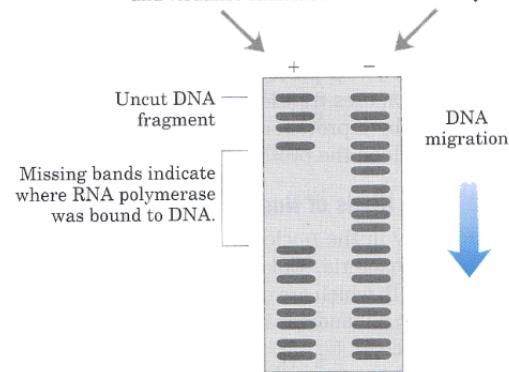
- a. How do I find regions in DNA of interest that bind to RNAP (or its sigma)?



**figure 1**

Footprint analysis of the binding site for RNA polymerase on a DNA fragment. Separate experiments are carried out in the presence (+) and absence (-) of RNA polymerase.

Separate fragments by polyacrylamide gel electrophoresis and visualize radiolabeled bands on x-ray film.



- b. DNase 1 is unable to cut regions that are bound to RNAP.
- c. Creates a gap in the isolated labeled DNA fragments

#### H. DMS Footprinting (Maxam and Gilbert)

- a. Much higher resolution because relies on small organic molecule (DMS) that can get closer to the Pol than the bulky DNase. Gives a more precise protection footprint.
- b. Mild DMS conditions: on average, 1 methylation event per DNA molecule.
- c. Reagents can remove methylated purines from DNA and then break the DNA there.
- d. Normally, 6' O in G's are protected from methylation by H-bonding.  
Local melting in protection footprints deprotects G's and amplifies methylation/cutting.

#### I. Transcription Initiation

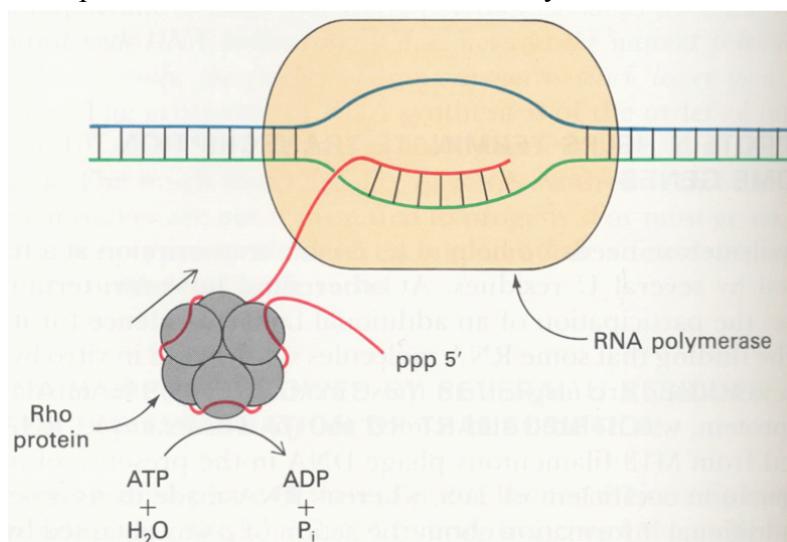
- a. RNAP binds at promoters: closed complex.
- b. Sigma factor binds and induces open complex at -10 promoter region; prevents procession.
- c. Promoter clearance: RNAP kicks off sigma after ~12 nts, adopts the ternary elongation complex (TEC), and processes along the promoter. Sigma is replaced by nusA, which senses termination signals.
- d. RNAP can transcribe 8-10 nts at a time; suggests inchworm/scrunching mechanisms.

#### J. Bacterial NRA Pol Structures Reveal Multiple Conformational Changes

- a. Aromatic side chains help stabilize nonpolar bases after initial melting, and positive side chains stabilize the negative phosphate backbone
- b. Abortive termination: half the time, the attempt to switch to elongation fails.  
Many conformational changes; many thermodynamic requirements.  
Forms a small transcript and is forced to restart.

#### K. Termination signals in nascent RNA

- a. Rho-independent
  - i. GC-rich stem with perfect Watson-Crick complementarity forms a hairpin followed by many Us. AU- regions are very weak, which eases dissociation.
  - ii. The hairpin's mechanism is uncertain. Could be a pulling out, a blocking action, or conformational changes.
- b. Rho-dependent
  - i. Rho binds 80 nt of C-rich/G-poor regions.
  - ii. Pol "stutters" on subtle differences in DNA conformation, and Rho catches up.
  - iii. ATP-powered RNA-DNA helicase activity releases the transcript.



## L. Inhibitors of Bacterial RNA Polymerase

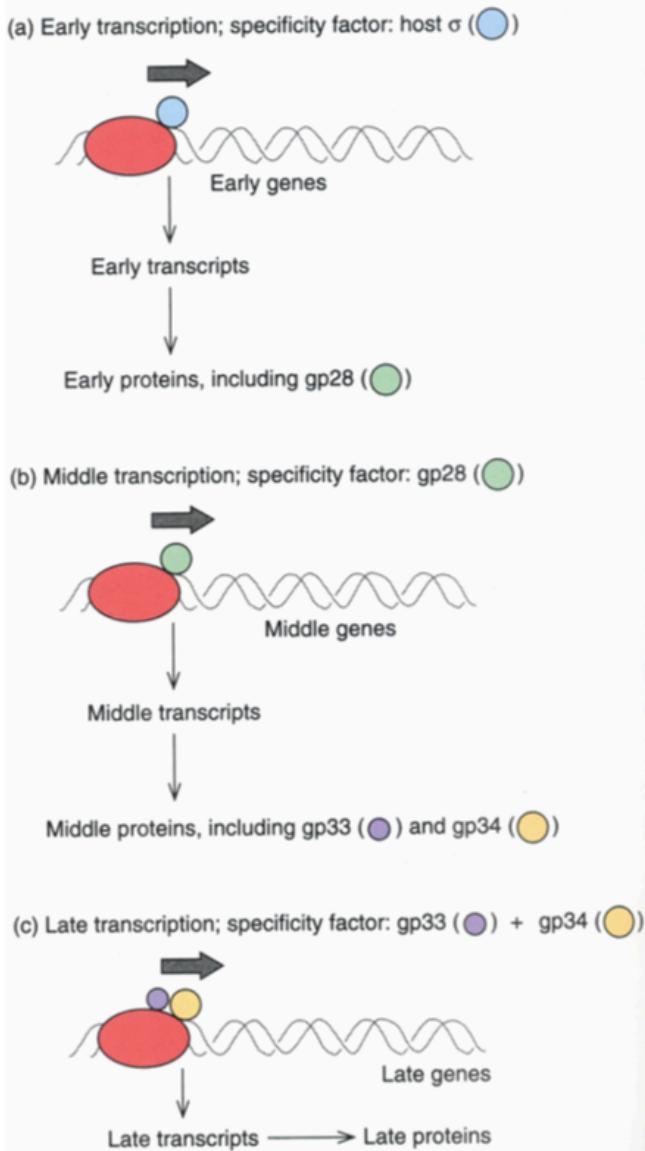
- a. Rifamycine binds  $\beta$ ; blocks initiation step via the exit channel (NOT binding/elongation)
- b. Actinomycin D: binds DNA template at CpG islands; blocks elongation

## M. Transcriptional Control by Sigma Factor Exchange

- a. Heat shock response: turns off most transcription, ramps up a few genes for hsp, which help refold or regrade partially unfolded proteins.
- b. In E. coli, normal sigma factor  $\sigma^{70}$  is replaced by heat shock sigma factor ( $\sigma^{32}$ ), which recognizes different promoter sequences.
- c. Both synthesis and stability of  $\sigma^{32}$  increase under heat shock conditions.
- d. Promoter for  $\sigma^{70}$  recognized by heat shock  $\sigma$  factor.

## N. Cascade of Sigma Factors Controls Gene Expression in Phages

- a. SPO1 DNA has “early” promoters recognizable by the host  $\sigma$  factor.
- b. These encode a phage  $\sigma$  factor recognizing the “middle” genes, which encode a phage  $\sigma$  factor that recognizes the “late” genes.
- c. Each phage  $\sigma$  factor binds more tightly to RNAP, thereby replacing the previous  $\sigma$  and achieving temporal regulation.

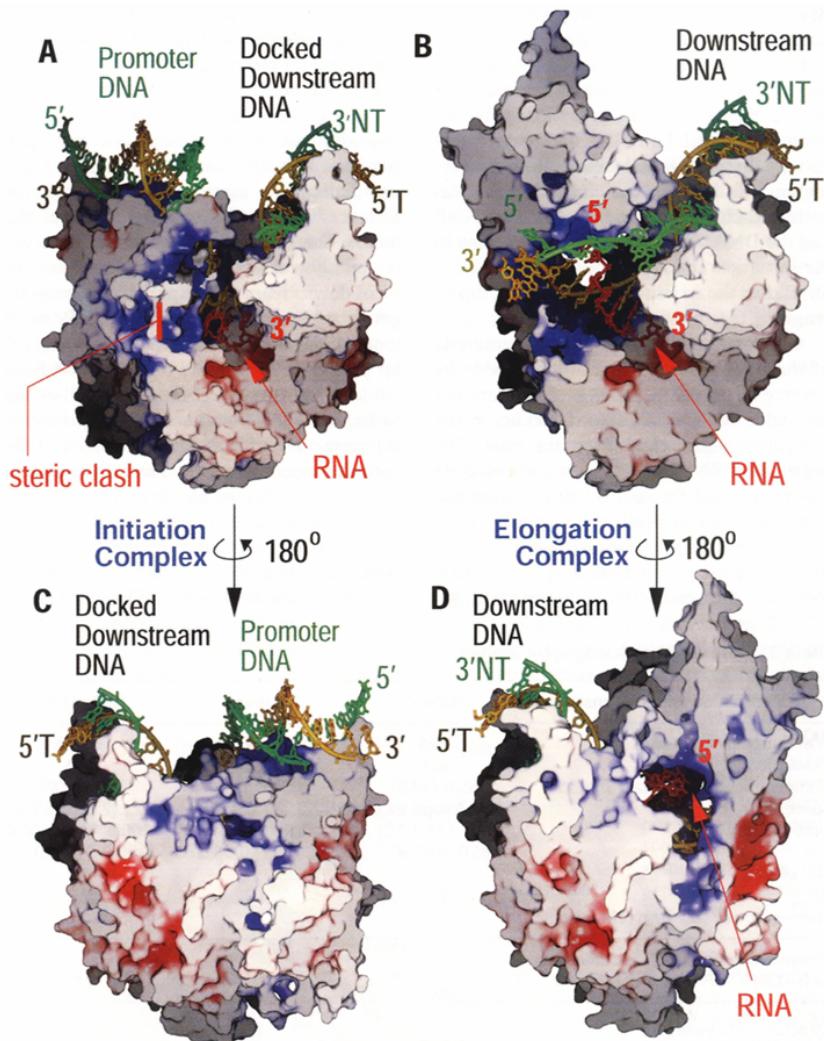


## O. Other Phages Encode their own RNA Polymerases

- a. T7 RNAP encoded by early gene; transcribes late genes.  
1 subunit; recognizes 23 bp promoter, 6x faster than E. Coli RNAP.
- b. SP6 RNAP: also 1 subunit.
- c. Both of these mimic the 3D structure of E. coli Pol 1

## P. Dramatic Conformational Change of T7 RNAP during Transition from Initiation to Elongation

- a. Crystal structure of initiation complex: unable to extend RNA because of sterics.
- b. Elongating complex changes dramatically when paired with 30 bp DNA and 17 nt RNA.
- c. Parts of the protein move as much as 70 Å.
- d. Promoter binding site is destroyed. A channel is created to accommodate the 7 bp heteroduplex in the active site. An exit tunnel for the RNA is created.
- e. The thumb domain separates the heteroduplex and directs the RNA into the exit tunnel.



**Fig. 4.** Formation of the RNA exit tunnel and upstream DNA binding site in the elongation complex. The solvent contact surface for the initiation complex conformation of T7 RNAP (A) with the observed upstream promoter DNA and heteroduplex along with the downstream DNA modeled from the elongation complex. The thumb domain has been removed from both (A) and (B) to allow a view into the heteroduplex binding site. The elongation complex (B) shows the disappearance of the promoter DNA binding site, the formation of a new channel that binds to heteroduplex and upstream DNA nonspecifically and a tunnel through which the transcript (red) exits. (C) and (D) are rotated by 180° about a vertical axis. The appearance of the tunnel that contains the 5' end of an RNA transcript (red) in the elongation complex (D). The positive electrostatic potential is blue and the negative is red. The two complexes have been oriented identically by superposition of their palm domains.



# Lecture 9: RNA Structure & Chemistry

## A. Types of Cellular RNA Molecules

- rRNA: stable, comes in defined sizes. Associate with proteins to form ribosomal subunits.
- tRNA: adaptor molecules (75-90 nt), more than the expected 20 types.  
Defined structure: anticodon loop to bind mRNA and AA-attachment site
- mRNA: short-lived, many sizes. Carry genetic info from DNA → Protein
- Small/noncoding RNAs (21-300 nt): roles in mRNA processing, modification, and translation, DNA replication, etc.
- Much more RNA than DNA in the cell: 6 pg of DNA, **30** pg RNA

## B. The Chemistry of RNA

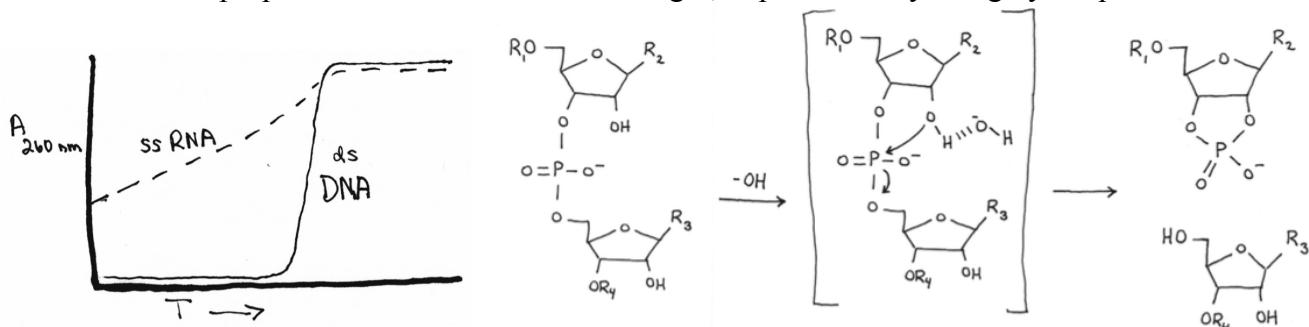
- RNA has 2'-OH (not 2'-H), which changes the sugar pucker/conformation.
- Thymine is replaced by Uracil (analog, lacks 5'-CH<sub>3</sub>)

## C. RNA Secondary Structure

- Random RNA structures will have 50-60% nt involved in base-pairing
- Structures include bulges, internal loops, and hairpins.
- Base-paired regions (e.g. hairpins) are still helical!
- A-form: found in RNA-RNA, RNA-DNA, and dehydrated DNA  
Tilted base pairs and ~11 bp/turn. Slightly fatter.
- RNA major groove is deep and narrow (hard to access).

## D. Melting curve of RNA v. DNA:

- Broad because RNAs have a range of secondary structures with very different stabilities.
- T<sub>m</sub> proportional to GC-content and length; duplex stability is highly cooperative.



## E. Base Hydrolysis of RNA

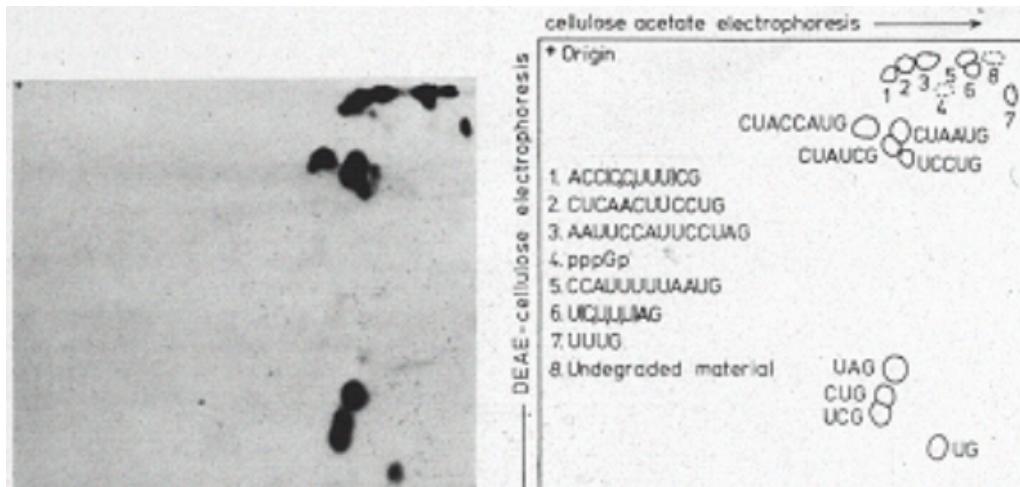
- OH<sup>-</sup> activates 2'-OH to attack and cyclize the phosphate
- More base detaches phosphate: yields mixture of 2'/3'-phosphate termini.
- Base hydrolysis provides the mechanism for enzyme-mediated RNA cleavage.

## F. Degradation of RNA

- Enzymes
  - i. RNase A (pancreatic) 3' to C/U (pyrimidines)
  - ii. RNase T1 3' to G
  - iii. RNase U2 3' to A/G (purines)

## G. RNA Sequencing

- RNA fingerprint: cleavage products of  $^{32}\text{P}$ -radiolabeled RNA separated by 2D paper chromatography, and visualized by autoradiography.
- Slow RNA sequencing: Make multiple cleavages and look at the fragment fingerprints to piece together the original.



- Rapid RNA sequencing: RNA is labeled at one end, partially cleaved at different nt sites (~1 cut per RNA) and run out on a gel (similar to DNA sequencing).

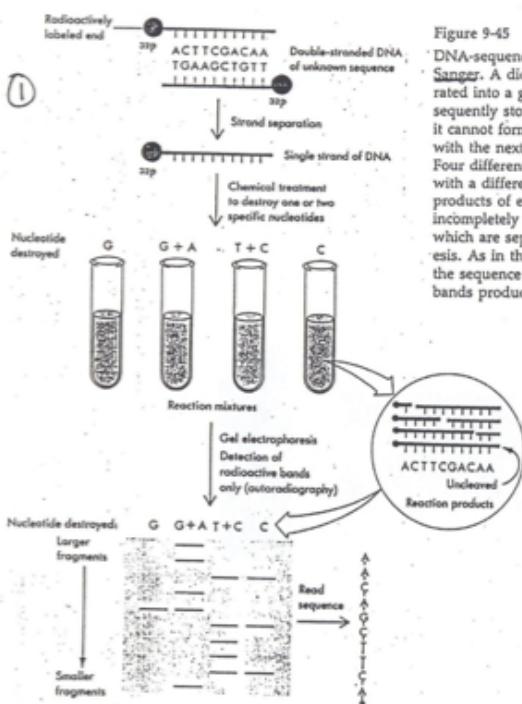


Figure 9-45  
DNA-sequencing method developed by Sanger. A dideoxynucleotide is incorporated into a growing DNA strand, subsequently stopping chain growth, since it cannot form a phosphodiester bond with the next incoming nucleotide. Four different reactions are run, each with a different dideoxynucleotide. The products of each reaction are a series of incompletely elongated segments, which are separated by gel electrophoresis. As in the Maxam-Gilbert method, the sequence can be read from the bands produced in the gel.

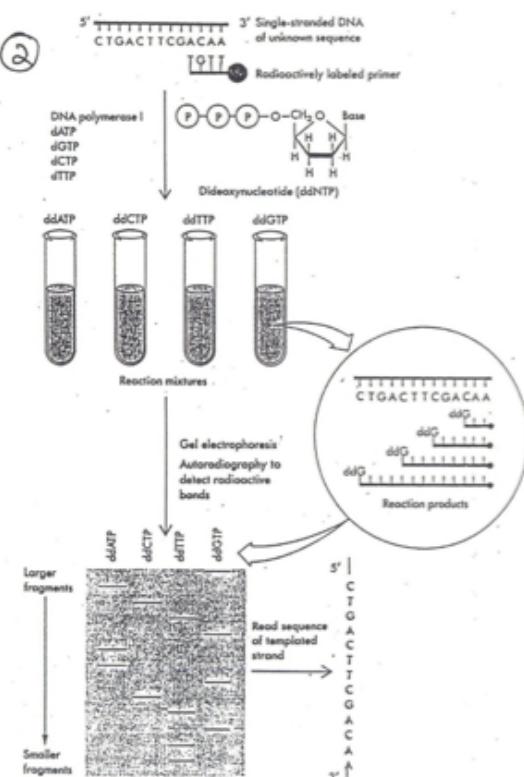


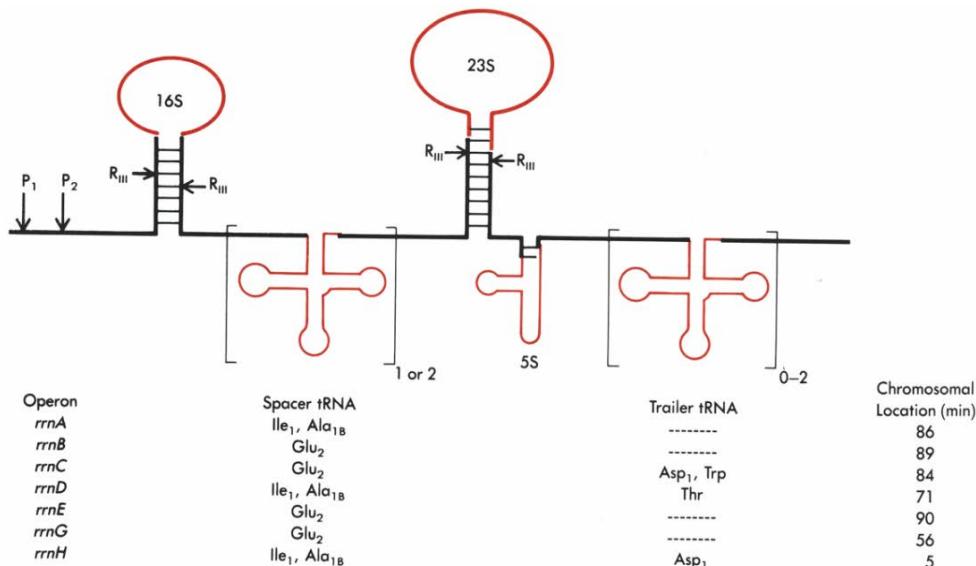
Figure 9-44  
DNA-sequencing method developed by Maxam and Gilbert. This method uses chemical reagents to destroy specific nucleotide bases and thus break the DNA molecule at specific sites. First the strands of the DNA molecule are labeled radioactively at one end (usually the 5' end), and the two strands are separated (only one will be sequenced). Then aliquots of the chosen strand are treated with four different chemical reagents that break the strand at one or two specific nucleotides; the treatment is limited so that at most a single residue of the susceptible base(s) in the molecule will react. Thus, in each reaction mixture, a nested set of radioactive fragments is generated, as shown here for only the reaction mixture that destroys C residues. Finally, gel electrophoresis is used to separate the products of each reaction by size. The pattern of radioactive bands seen on X-ray film immediately reveals the sequence.

## Rapid DNA Sequencing Methods

See Lehninger pp. 302-303 (6<sup>th</sup> ed)  
Voet & Voet pp. 176-180 (4<sup>th</sup> ed)

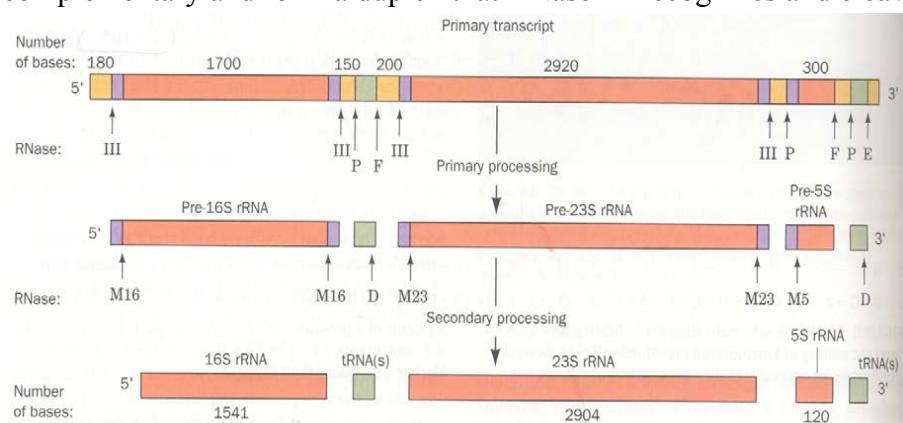
## H. Ribosomal RNA (rRNA)

- 80% of total RNA in cells, 50-65% of the mass of ribosomes
- Contain post-synthetically modified nts (psi and 2'-O-methyl).
- Svedberg (S) value: relative value for size; measures sedimentation rate  
i.e. how far the ribosome sinks into some kind of gradient upon centrifugation.
- Coded by tandem repeats of 7 different rrn (rRNA) operons
- Synthesized as precursor RNAs that are processed into mature rRNA.
- Bacteria: 70S ribosome made of 3 rRNAs
  - 30S subunit made of **16S rRNA** + 20 proteins
  - 50S subunit made of **23S rRNA, 5S rRNA** + 34 proteins
- Eukaryotes: 80S ribosome made of 4 rRNAs
  - 40S subunit has **18S rRNA**
  - 60S subunit has **28S rRNA, 5.8S rRNA, 5S rRNA** + proteins



## I. Processing of Pre-ribosomal RNA Transcripts

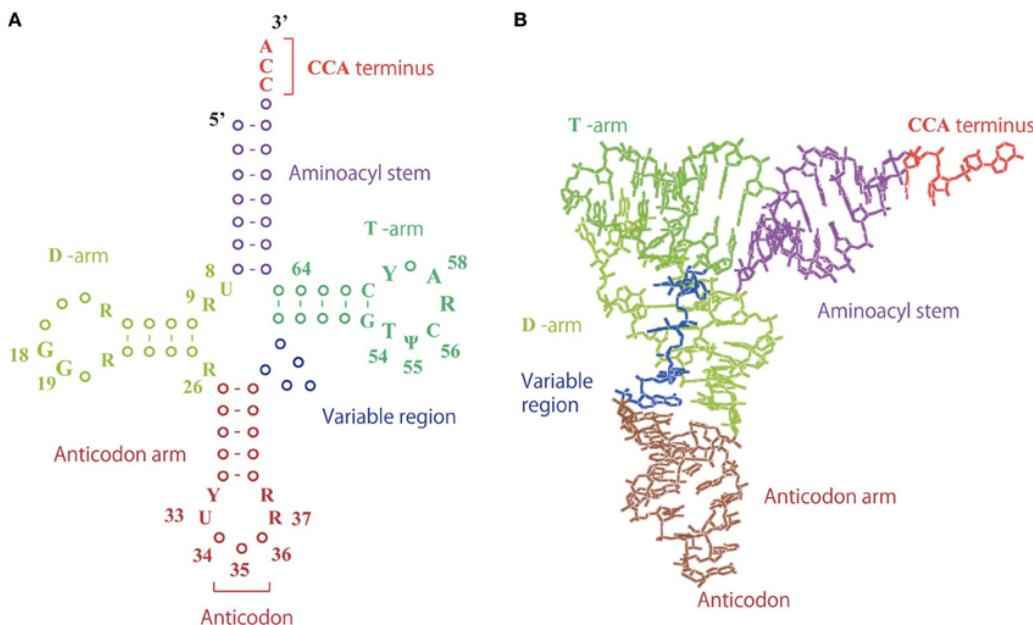
- Two important endonucleases: RNase P and RNase III.
  - RNase P cleaves at 5' end of tRNAs in the transcript; catalytic RNA component
  - RNase III cleave ds-RNA by accessing narrow major groove; homodimer.
- Stem-and-giant-loop secondary structure: sequences flanking 16S and flanking 23S are complementary and form a duplex that RNase III recognizes and cleaves.



- J. rRNA Secondary Structure: contains short duplexes (6-10 bp), internal bulges, hairpins, central hubs (from which hairpins radiate) and single stranded segments.
  - K. Compensatory Mutations to Determine 2° structure
    - a. If variation across species preserves WC pairing, then those nts are likely to be paired.
    - b. Non-canonical pairings exist in RNA structures.
  - L. rRNA 2° Structure Reveals 5.8S-23S (bacterial-eukaryotic) homology
    - a. 5.8S is similar to the 5'-end of E. coli 23S
  - M. tRNAs as Adaptors
    - a. 15% of total RNA
    - b. Stable in the cell
    - c. 75-90 nt
    - d. Made as precursors in rrn operons or multi-tRNA transcripts in bacteria.
    - e. 5'-end of all tRNAs is made by RNase P.

## N. tRNA Secondary Structure

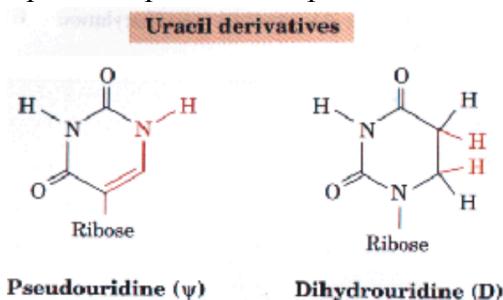
- a. Conserved cloverleaf shape (4 short helices)
  - b. Acceptor stem (top) and anticodon loop (bottom)
  - c. Anticodon loop has 7 nts, the middle 3 are the anticodon.
  - d. ss-CCA-OH added to 3'-end after transcription by CCA-adding enzyme.
  - e. Amino acid is attached to A on CCA-3'.
  - f. 15 invariant positions and 8 semi-variant positions in different tRNAs (impt for structure/function)
  - g. Variable arm can be 3-20 nts.



## O. Post-synthetic Modifications of tRNA

- a. 10/76 nts in tRNA<sup>Ala</sup> are modified; important but not essential.
  - b. Bulky R group added to purine near anticodon 3'-end to eliminate wobble involving 1<sup>st</sup> position of mRNA codon.

- c. Conserved: D arm named for dihydrouridine and t-psi-C loop named for pseudouridine.



P. L-Shaped 3° structure of tRNA

- 42/76 nt are duplexed, 71/76 are stacked.
- Structure dominated by base-stacking:** two pairs of helices stack to form a longer helix
- D-loop makes extensive non-WC interactions with the T-psi-C loop.

Q. Non-canonical Pairings in tRNA Structure

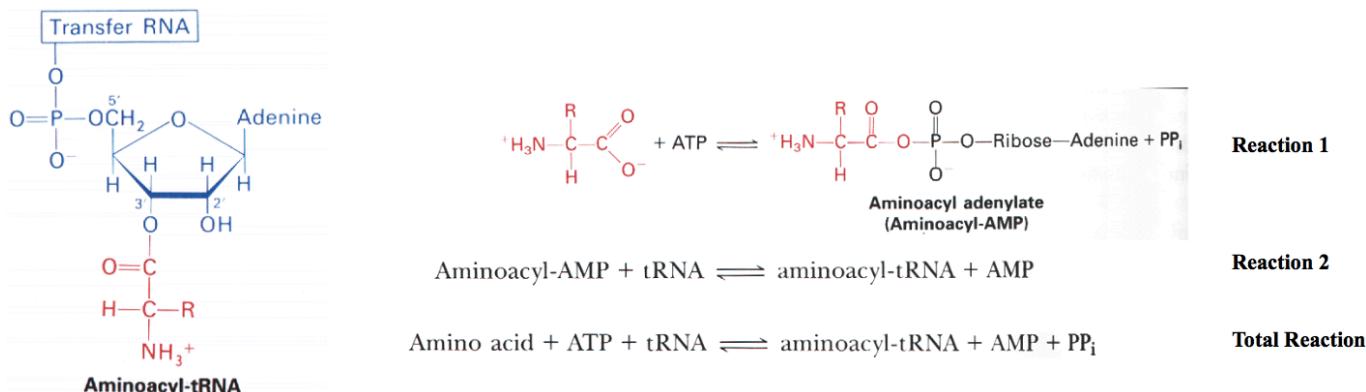
- 8/9 tertiary contacts between the 2 helical stacks are non-WC.
- Base triplets (GGC, AAU): an unpaired 3<sup>rd</sup> strand can recognize a helix by pairing.
- Strange pairings include GU, GA, and non-canonical GC/AT pairs.

R. tRNAs are Recognized by Multiple Cellular Components

- RNase P and processing enzymes
- tRNA modifying enzymes
- CCA transferases
- Aminoacyl-tRNA synthetases
- EF-Tu (the escort)**
- Ribosome

S. Isoacceptors: tRNAs that accept the same amino acid

T. tRNA Charging: Aminoacylation

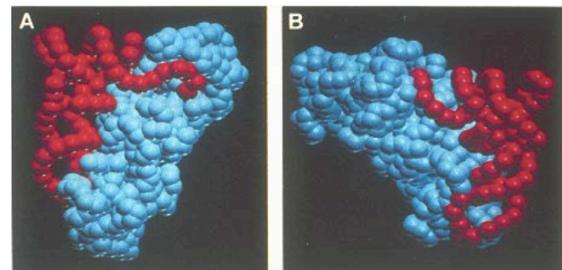


- Aminoacyl-tRNA synthetases use ATP to add AMP as a leaving group for aminoacylation.
  - Amino acid joined to 5'-alpha-phosphate of AMP, with release of PP<sub>i</sub>
  - Aminoacyl-AMP releases AMP while reacting with tRNA
- tRNA terminal 2'/3'-OH is coupled to AA-COO-, Rapid exchange between the two OH's
- Each synthetase is specific for a single amino acid, recognized by extensive H-bonding and hydrophobic interactions.
- Synthetases also make specific interactions to recognize multiple isoacceptor tRNAs.

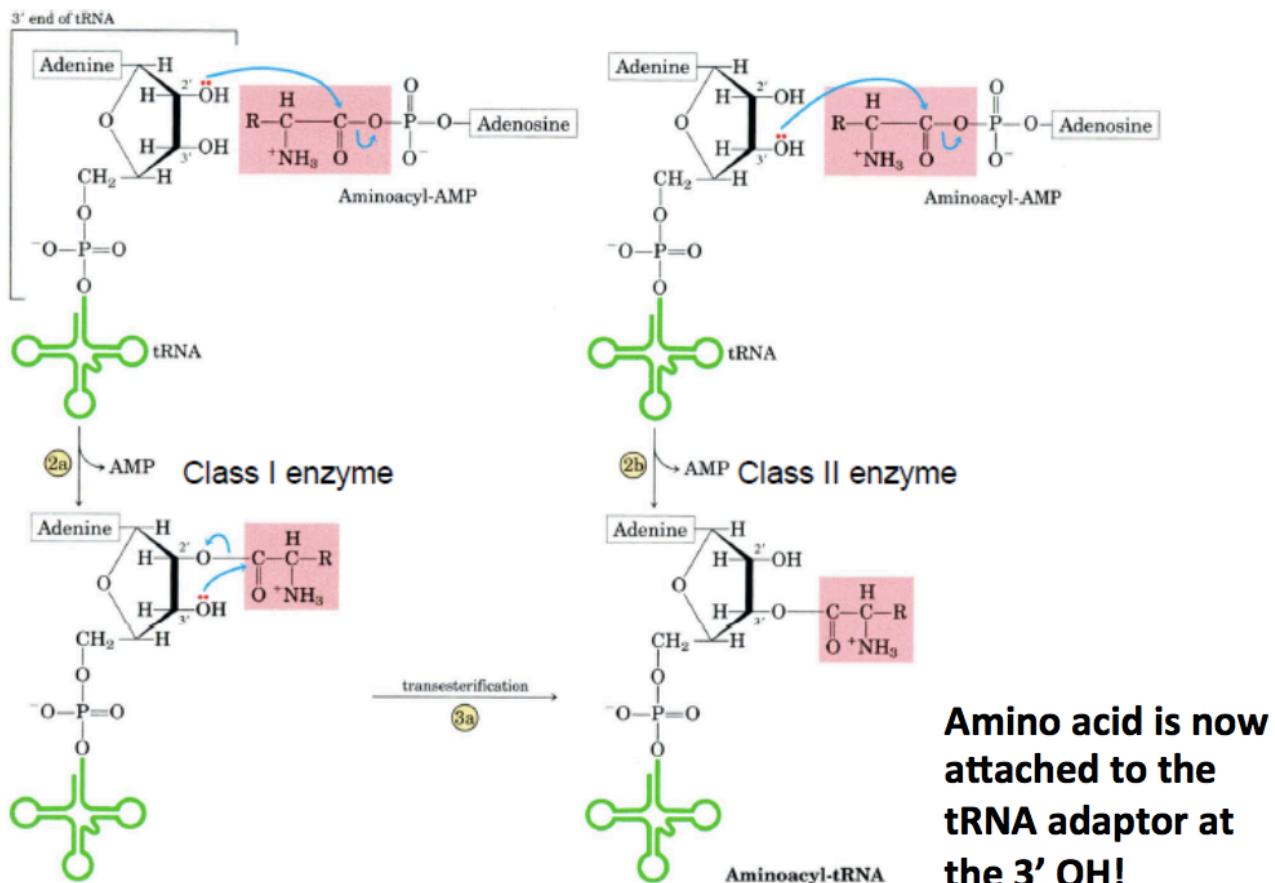
- e. “Identity elements” present in anticodon loop, acceptor stem, and variable loop.
  - i. Positive identity elements recruit the correct synthetase
  - ii. Negative identity elements resist the wrong synthetase
- f. Proofreading mechanism ensures that proper amino acid was added (for similar AA-synthetases, like Val v. Ile).

#### U. Class I and Class II Aminoacyl-tRNA Synthetases

- a. Two different structural classes based on:
  - (1) mechanism of reaction, and
  - (2) conservation of protein fold.
- b. Class I aminoacylates the 2'-OH, while Class II targets the 3'-OH. This means different angle of attack (opposite sides of approach).

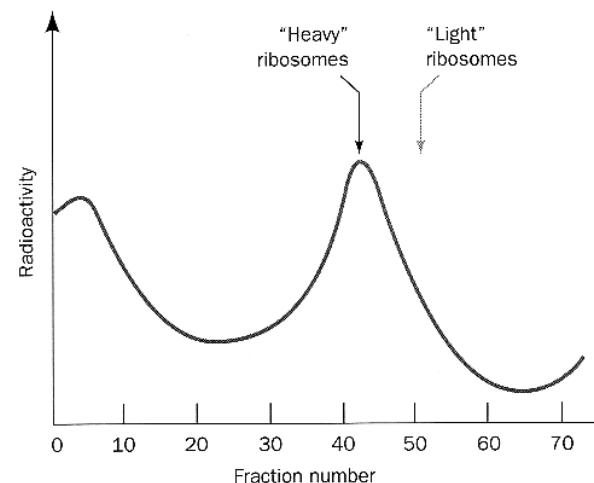
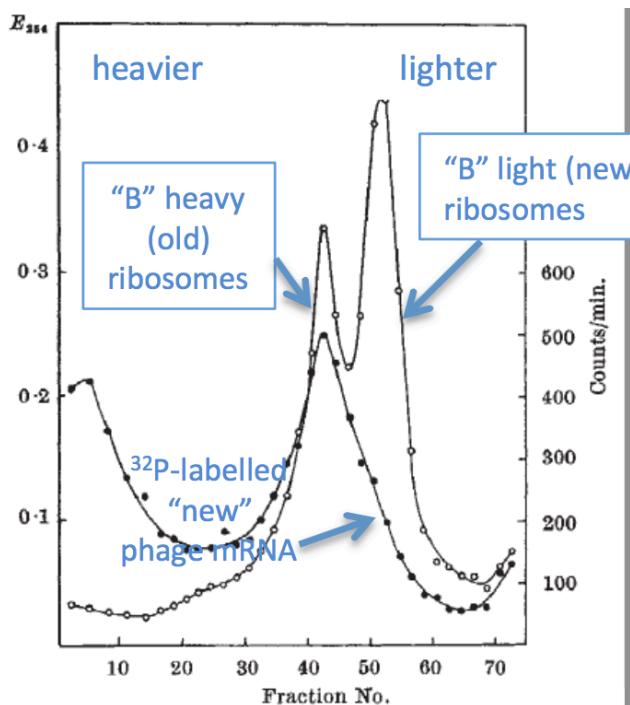


### Conversion of aminoacyl-AMP to aminoacyl-tRNA



## V. Brenner: Discovery of mRNA

- Initial thought: rRNA carries genetic information.
- E. coli were grown in heavy media ( $^{15}\text{N}$  and  $^{13}\text{C}$ ) to make heavy ribosomes.
- E. coli then shifted to light media containing  $^{32}\text{P}$  and infected with T4.
- T4 shuts down host expression and begins synthesizing ONLY phage proteins.
- After growth in  $^{32}\text{P}$ , cells were lysed and ribosomes isolated by centrifugation.
- If 1 ribosome-1 protein, then you'd expect the light ribosomes to be labeled with rRNA.  
But the only way you get a peak at heavy ribosomes is if new RNA was associating there.
- Density gradient centrifugation:



**FIGURE 29-6.** The distribution, in a CsCl density gradient, of  $^{32}\text{P}$ -labeled RNA that had been synthesized by *E. coli* after T4 phage infection. Free RNA, being relatively dense, bands at the bottom of the centrifugation cell (left). Much of the RNA, however, is associated with the  $^{15}\text{N}$ - and  $^{13}\text{C}$ -labeled “heavy” ribosomes that had been synthesized before the phage infection. The predicted position of unlabeled “light” ribosomes, which are not synthesized by phage-infected cells, is also indicated.

## W. mRNAs

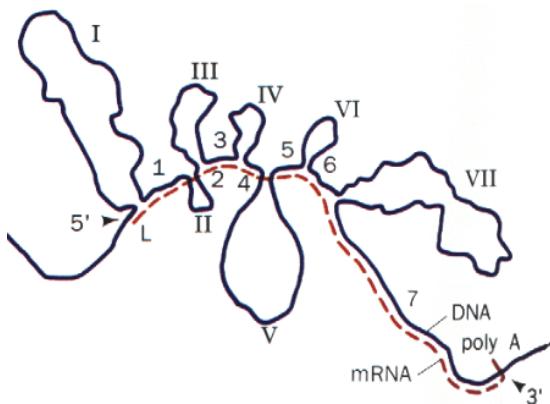
- mRNAs have different lengths for different protein sizes.
- mRNA is complementary to one strand of DNA; can hybridize to one strand.
- mRNA turnover is rapid; only comprises 5% of total cellular RNA
- mRNA associate transiently with ribosomes
- mRNA have leader and trailer regions flanking the coding region.  
Polycistronic mRNA usually have spaces between coding regions of gene products.
- mRNA secondary structure can control translation.



## Lecture 10: Catalytic RNA

### A. Pre-mRNA Splicing: Introns and Exons

- a. Surprise! Eukaryotic genes are discontinuous.
- b. Hybridization of a cDNA probe to a eukaryotic gene shows excised regions that get looped at and do not hybridize to mRNA (introns).



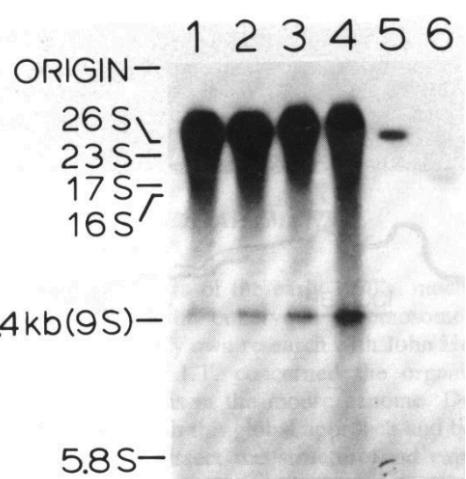
- c. Introns can be a few dozen/thousand nt; much larger than exons.
- d. Introns removed only at RNA level.

### B. Discovery of Catalytic RNA

- a. Tetrahymena thermophila: transcriptionally active macronucleus and inactive micronucleus; contains ~10,000 copies of rDNA in each macronucleus.
- b. Cech tried to purify lots of rDNA.
- c. Sequencing of rDNA shows that it contains a single intron (intervening sequence-IVS)
- d. rRNA is transcribed very quickly: 1 copy/s/gene.
- e. Powerful system for studying intron splicing

### C. Pre-rRNA Splicing Occurs in vitro

- a. Alpha-amanitin: inhibits transcription of all but rRNA in eukaryotes; used to make full-length (26S) body-labeled RNAs.
- b. Cech observed formation of 9S product from 26S full-length RNAs.
- c. Isolated pre-rRNA is incubated with:
  - i. Fractionated nuclear extracts; intron is efficiently spliced.
  - ii. Salt and NTP solutions, no extract; intron is STILL spliced.

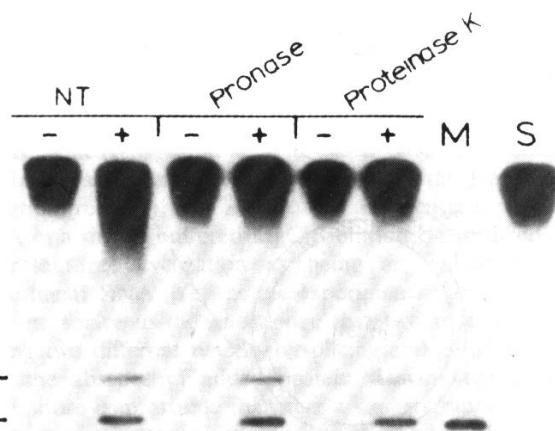


### D. Requirements for rRNA self-splicing are minimal

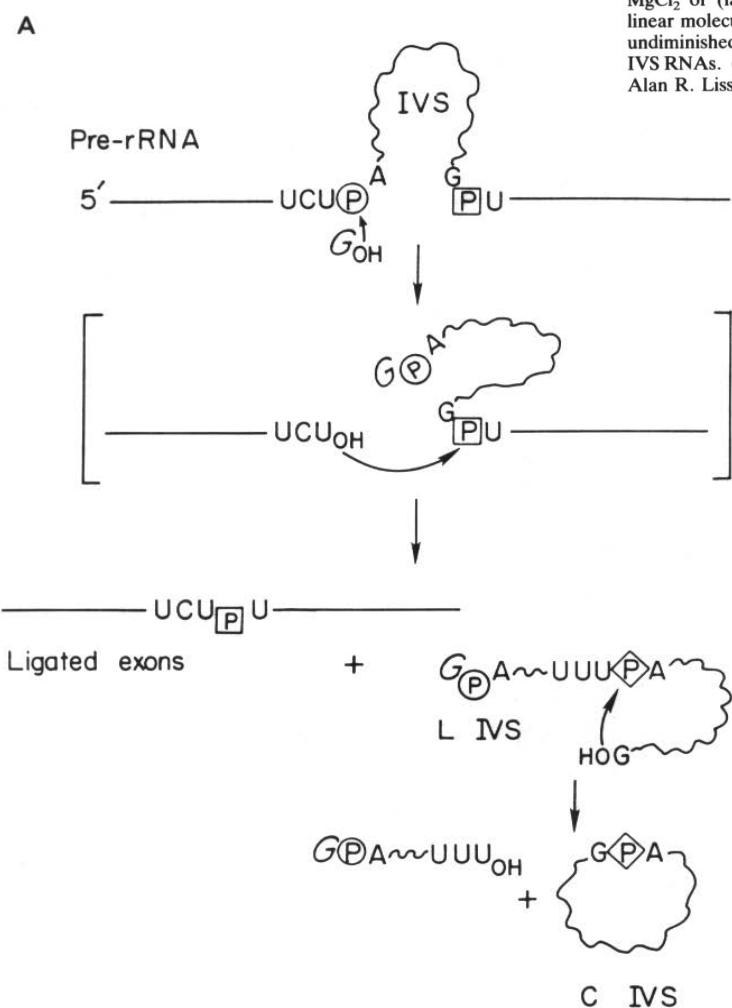
- a. Sequencing showed that intron had an extra G on the 5'-end.
- b. Intron release demands MgCl<sub>2</sub> and some kind of guanosine (G, GMP, GTP, etc).

### E. The Group I Intron is a True Catalytic RNA

- Splicing activity is resistant to phenol extraction, SDS, and proteases; Proteins uninvolved in catalysis.
- rRNA spliced in vivo and in vitro.
- Conclusion: rRNA Group I intron is a ribozyme



**Fig. 5.** IVS RNA excision and cyclization activities are resistant to protease treatment. (Lane S) Pre-rRNA, purified by SDS-phenol extraction. This RNA was then treated exhaustively with proteases or (NT) not treated. Incubation with GTP was performed, (lanes -) in the absence of  $MgCl_2$  or (lanes +) with 10 mM  $MgCl_2$ . Excision of the IVS RNA as a linear molecule (L) and subsequent conversion to a circular form (C) were undiminished by protease treatment. (Lane M) isolated linear and circular IVS RNAs. (Reproduced from Grabowski *et al.*, 1983, by permission of Alan R. Liss, Inc.)



**Fig. 4.** (A) Self-splicing of the *Tetrahymena* pre-rRNA by consecutive transesterification reactions. Straight lines, exons (mature rRNA sequences); wavy line, IVS; circle, 5' splice-site phosphate; square, 3' splice-site phosphate; diamond, cyclization site phosphate. (B) Model for the initial step involving nucleophilic attack by the 3' -hydroxyl group of guanosine on the phosphorus atom at the 5' splice site. The hypothesis of an in-line,  $S_N2(P)$  reaction with inversion of configuration around phosphorus was subsequently confirmed (McSwiggen & Cech, 1989). The hypotheses of acid-base catalysis and coordination of  $Mg^{2+}$  to the phosphate, enhancing the electrophilicity of the phosphorus atom and stabilizing the trigonal bipyramidal transition state, are untested. (Reproduced from Cech, 1987; copyright by the AAAS)

## F. Group I Intron Secondary Structure

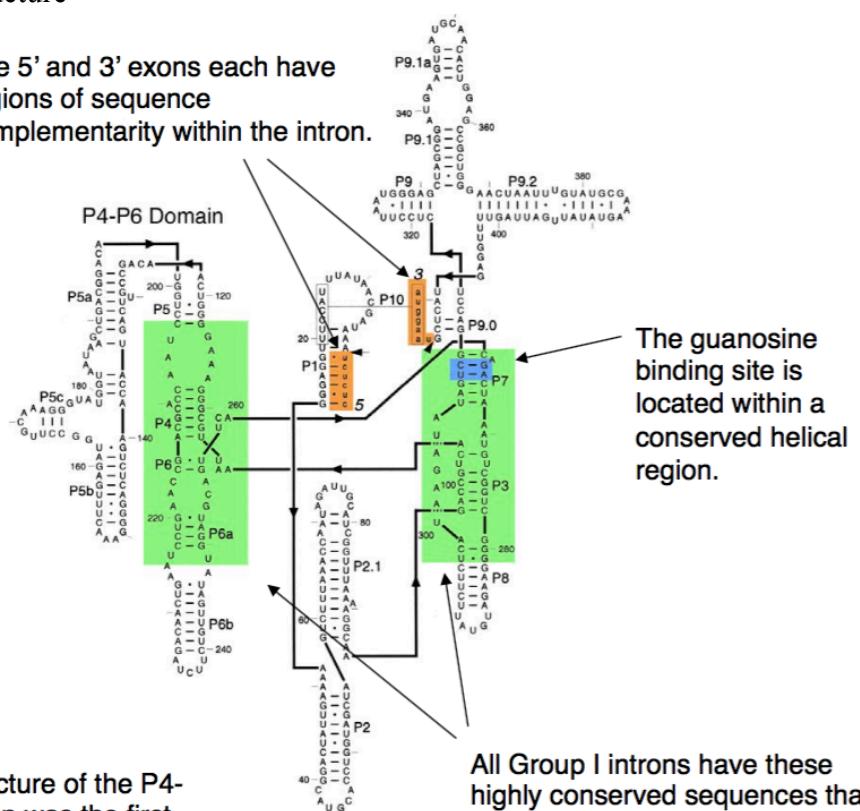
The 5' and 3' exons each have regions of sequence complementarity within the intron.

- There are now hundreds of examples of Group I introns.

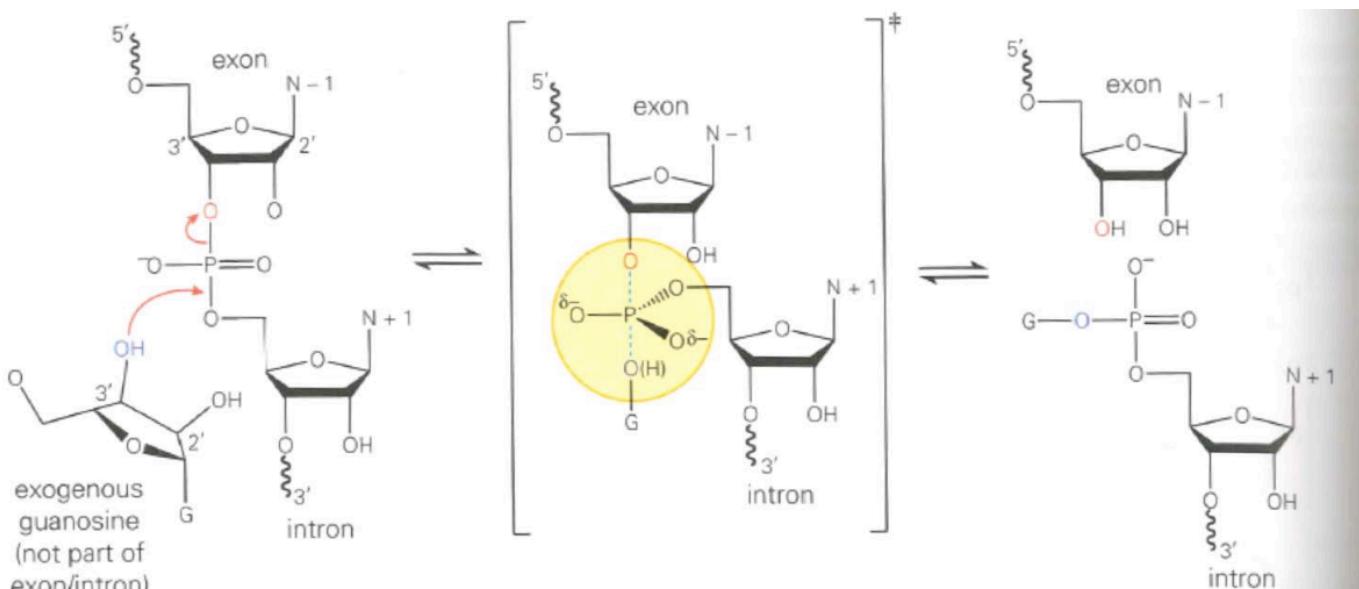
These are located in eukaryotic and prokaryotic genes - typically in rRNA or tRNAs.

- Sequence conservation and covariation were used to map the secondary structure of the RNA.

- Such a phylogenetic approach was used in combination with biochemistry to identify the G binding site.



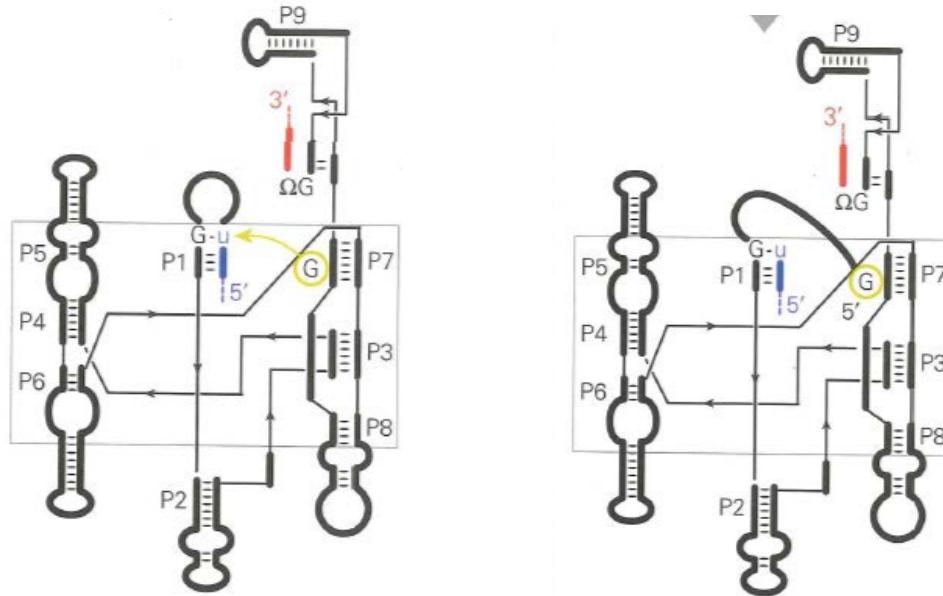
The high-resolution X-ray structure of the P4-P6 domain of the Group I intron was the first large RNA structure to be determined after tRNA. Doudna lab, MBB, Yale, *Science* 1996.



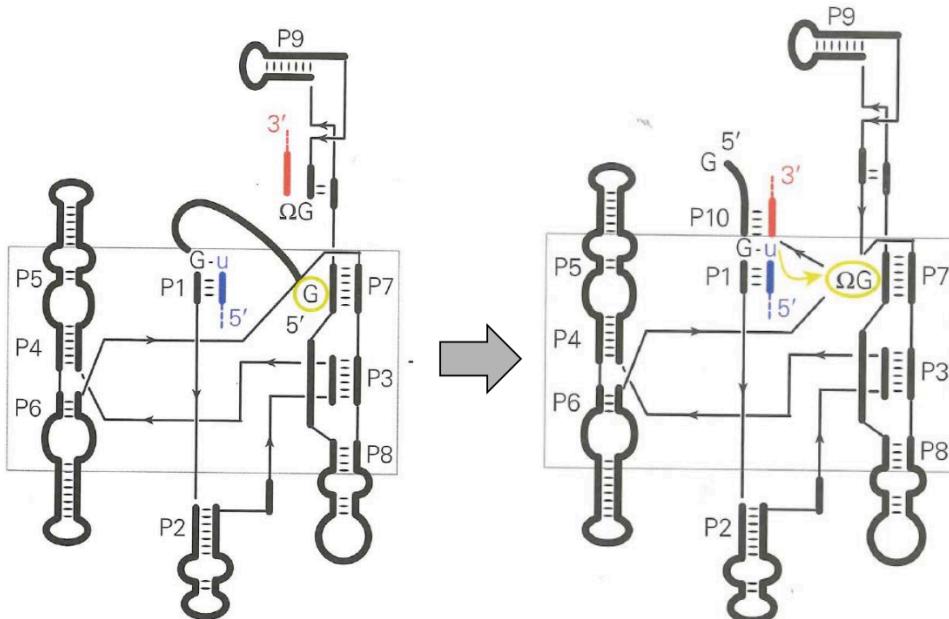
**Figure 16.58 The self-splicing reaction.** The 3'-hydroxyl group of an exogenous guanosine (blue) is the attacking nucleophile in the first step of group I self-splicing. The bridging 3' oxygen (red) is the leaving group, and the new 3' end of the 5' exon becomes the attacking nucleophile during the second step of splicing (not shown). Refer to Figure 16.59 for a schematic view of the complete splicing reaction.

### G. Group I Intron Splicing: Step 1

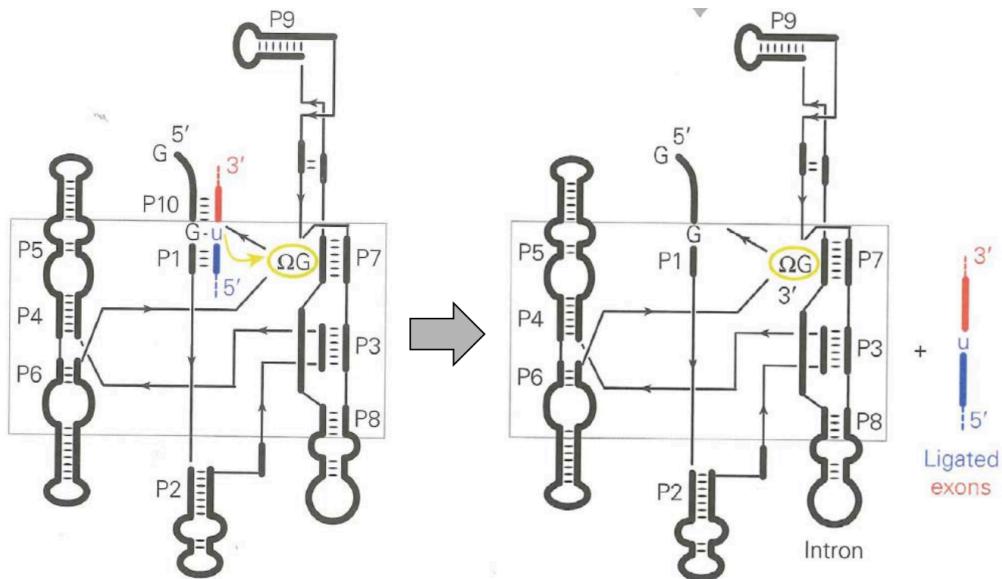
- a. G binds to a specific spot. Its 3'-OH attacks the phosphate between the 5'-exon and intron.
- b. 5'-exon is covalently cleaved off, but remains base-paired to the intron (P1).



- c. The intron undergoes conformational rearrangement
  - i. aG is replaced by omega-G (last nucleotide of intron)
  - ii. This allows 3'-exon to base-pair with the exon directly above the 5'-exon.



- d. 3'-OH on the 5'-exon attacks the phosphate between omega-G and 3'-exon.
  - i. Omega-G is released during ligation of 5' and 3' exons.

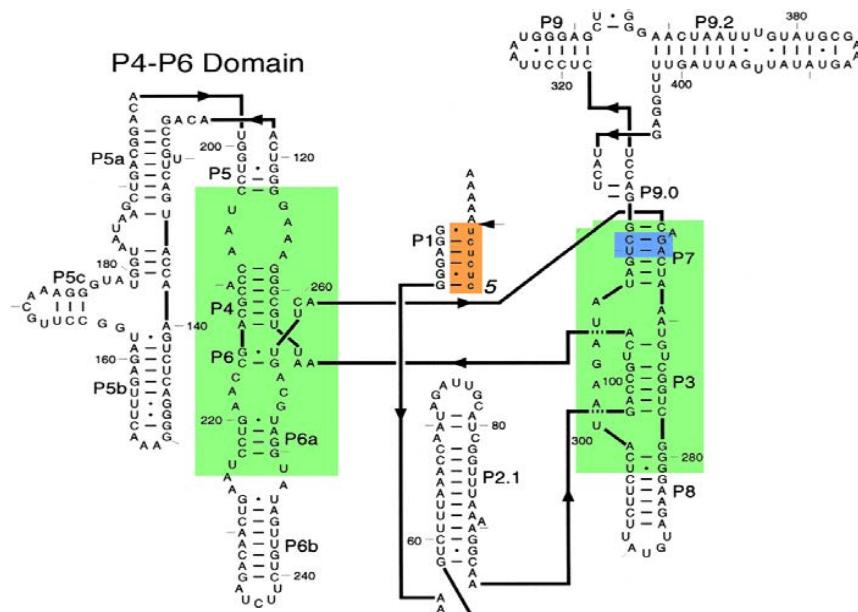


#### H. The A Minor Motif in RNA Tertiary Structure

- Adenosines are often conserved in ssRNA; infrequent mutations
- These As are primary mediators of long-range tertiary interactions in RNA; allow helical packing within an RNA tertiary structure.
- Four categories of A-mediated tertiary interactions based on A-position relative to base pair.
- Type I and II A-minor motifs are the most biologically relevant.
- Type I is unusually stable due to extensive van der Waals surface (hydrophobic adenosine edge and minor groove edge of base-pair).

#### I. “True” Enzymatic Activity

- True enzymes are left unmodified and capable of multiple turnover catalysis.
- Group I intron was converted by removing the loop connecting the 5' exon to the intron, and removing the 3' exon.
- 5'-exon-AAAAAA is added in trans, and may be cleaved by G addition.

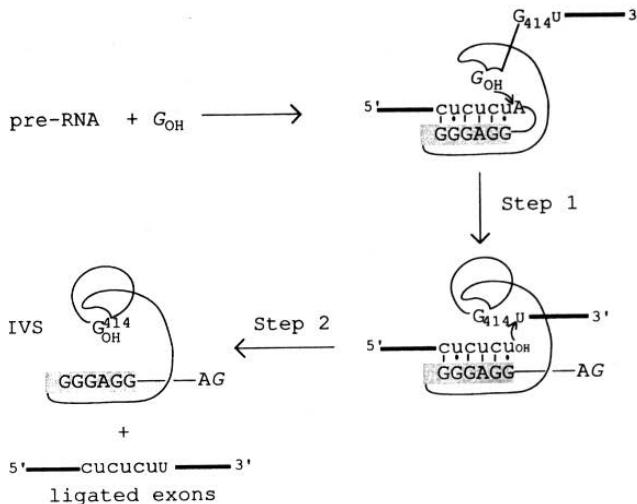


## J. Reaction Pathways of the Splicing and Enzymatic Form of the Group I Intron

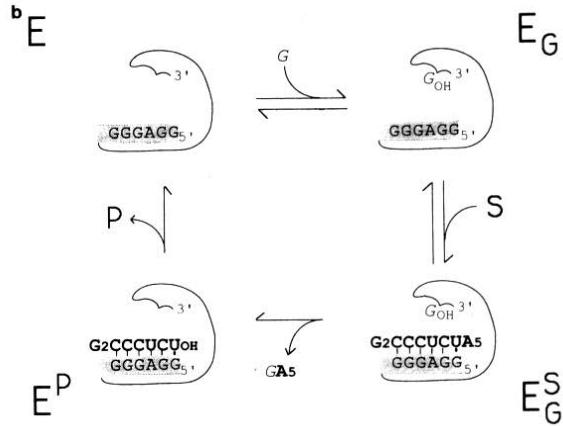
### a. Ribozyme Form

- i. Analogous to step 1 of splicing
- ii. Reaction obeys classical enzymatic kinetics (e.g. is saturable)

### Normal Reaction



### Multiple Turnover Reaction



## K. Mechanistic Enzymology: Comparison of Polymerase and Ribozyme Active Sites

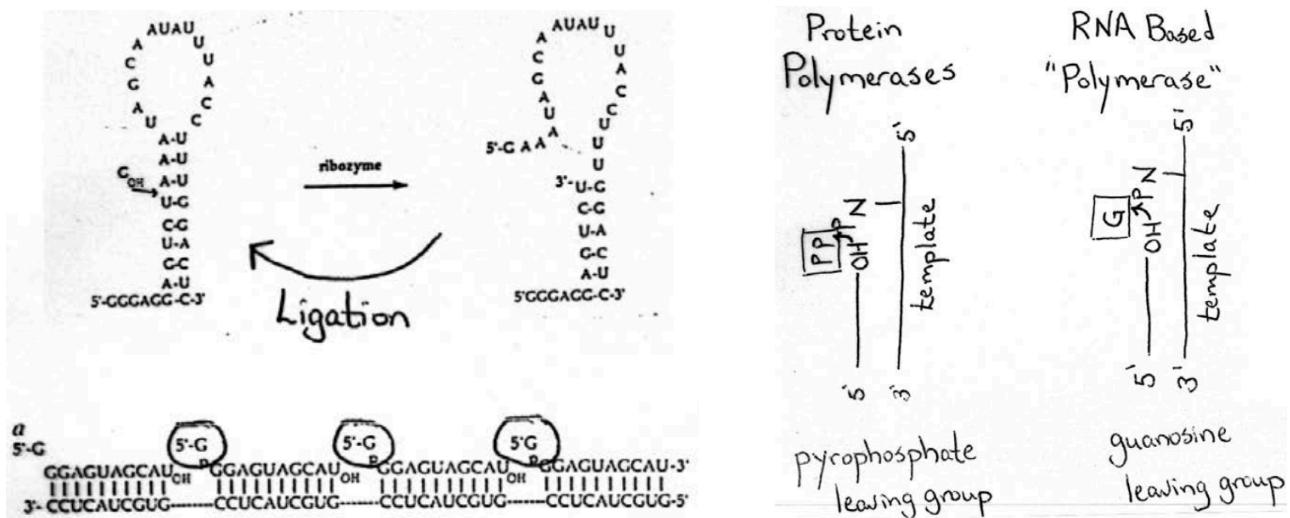
- a. Both are two-metal mechanisms involving a nucleophile (3'-OH) from an incoming nucleotide (dNTP or 3'-exon), that knocks off a leaving group (PP<sub>i</sub> / G).
- b. Two active-site metal ions coordinated to negatively charged body (aspartate/RNA backbone).

## L. Group I Intron as an RNA Pol

- a. Ribozyme without 5'-exon can bind substrate in trans.
- b. Ligation (where a phosphodiester linkage forms and kicks off G) is the reverse of splicing step 1 (where G attacks and breaks a phosphodiester linkage).
- c. RNA World hypothesis: RNA serves as both information carrier and catalytic molecule.

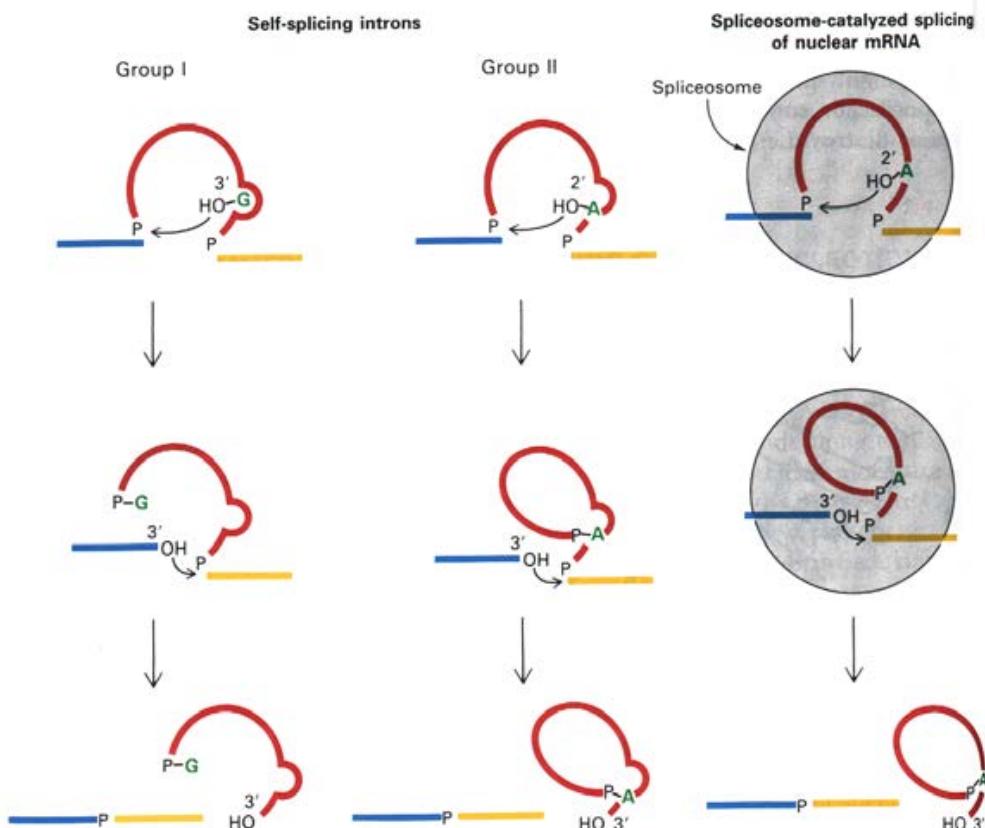
## M. Endonuclease RNase P is a ribozyme

- a. RNA component alone can catalyze reaction with high salt (M1 protein is dispensable)
- b. Coordinates H<sub>2</sub>O for hydrolysis of phosphate on pre-tRNA.
- c. Two metal mechanism.



#### N. Group II Self-Slicing Introns

- Step 1: uses an internal A's 2'-OH, not exogenous G's 3'-OH, to release the 5'-exon. Since the A is internal, this creates a lariat intermediate.
- Step 2: 3'-OH on 5'-exon attacks the phosphate attached to the 3'-exon, which releases the lariat intron and joins both exons.
- The 5'-exon is held by complementary base-pairing to noncontiguous sequences.  
Intron/Exon binding sites: IBS and EBS 1 and 2.



**Figure 33-47**

Comparison of self-splicing and spliceosome-catalyzed splicing. The exons being joined are shown in blue and yellow, and the attacking unit in green. The catalytic site is formed by the intron itself (red) in group I and II splicing. In contrast, the splicing of nuclear mRNA precursor is catalyzed by snRNPs in a spliceosome. [After P.A. Sharp. *Science* 235(1987):769.]



# Lecture 11: Protein Synthesis I

## A. The Ribosome is a Polypeptidyl Transferase

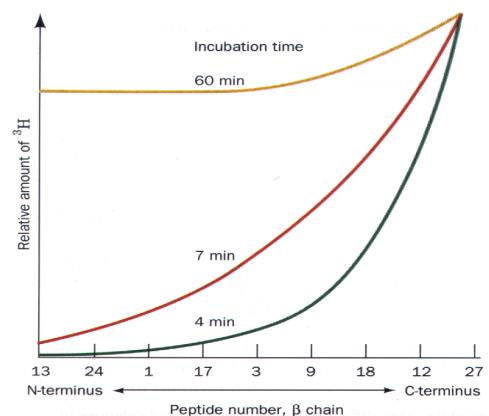
- Translation requires GTP, soluble proteins (incl. synthetases and charged aa-tRNAs), ribosomes, and mRNA.
- Ribosome is a RNA-protein complex of 2,500 kD, 250 Å across.
- 20,000 ribosomes per bacterium (1/4 total mass) incorporate 20 AA/sec
- Polyribosomes simultaneously translate the same mRNA.

## B. Direction of Protein Synthesis

- $|AAA| \dots |AAA|AAC|$  created Lys...Asn.
- Protein synthesis is directional (**5'-3' becomes N-C**)
- This allows translation to immediately follow transcription (as the 5' end comes out).

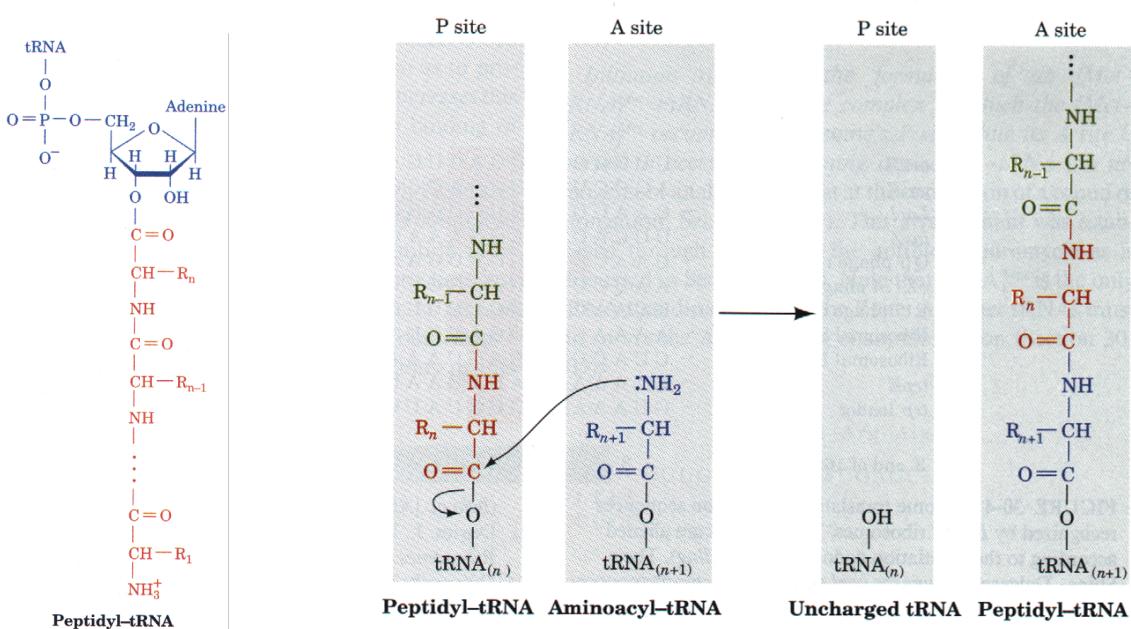
## C. Dintzis experiment

- $^3\text{H}$  is added, but not long enough for the ribosome to finish a hemoglobin chain.
- ONLY full-length proteins purified and fragmented.
- When incubating briefly, only C-terminus is labeled, because  $^3\text{H}$  was tacked onto the end.
- When incubating longer, everything (incl. N-terminus) is labeled because fragments that start with radiolabeled AAs have time to grow to full length.



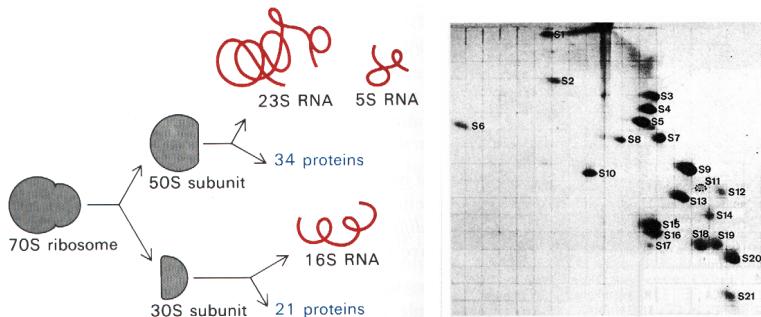
## D. Peptidyl Transfer Reaction

- High salt releases nascent peptide from actively translating ribosomes.
- Peptides COO- terminus is attached to tRNA CCA-terminus at the 3'-OH.
- P-site binds peptidyl-tRNA, and A-site binds aminoacyl-tRNA.
- Below: aa-tRNA's nucleophilic amino group attacks the ester linkage connecting the peptide to the tRNA, forming an amide/peptide bond.



## E. Composition of the Ribosome

a. Ribosome is 2/3rds RNA



- b. Ribosome can be reconstituted from purified RNA and protein
  - c. The proteins of the large subunit are L#, and the ones in the small subunit are S#. The number (#) is based on migration location in a 2D protein gel.

#### F. Soluble Translation Factors: Initiation Factor (IF), Elongation Factor (EF), Release Factor (RF)

- a. IF-1 Assists IF-3 binding
  - b. IF-2 Binds initiator tRNA and GTP
  - c. IF-3 Releases mRNA/tRNA from old 30S subunit, aids new mRNA binding
  - d. EF-Tu Binds aa-tRNA + GTP
  - e. EF-Ts Displaces GDP from EF-Tu
  - f. EF-G Promotes translocation through GTP-binding and hydrolysis
  - g. RF-1 Recognizes UAA and UAG stop codons
  - h. RF-2 Recognizes UAA and UGA stop codons
  - i. RF-3 Stimulates RF-1/RF-2 release via GTP hydrolysis
  - j. RRF Together with EF-G, induces ribosomal dissociation to subunits.

## G. Translation Initiation

- a. Protein synthesis begins at AUG near the 5'-end
  - b. Shine-Dalgarno sequence: purine (A/G) rich
    - i. Upstream of the AUG
    - ii. Recognized by the 3' end of the 16S rRNA: pyrimidine (U/C) rich.

Ribosomal S12	- AAA ACC CAG GAG CUA UUU A AUG GCA ACA -
Ribosomal L10	- CU AC CAG GAG CAA AG CUA A AUG GCU UUA -
<i>trpE</i>	- CAAA AUU AGAG AAU AA CAA AUG CAA ACA -
<i>trp leader</i>	- GUAAA AAGGG UAU CGAC AAUG AAAG CA -

3' end of 16S rRNA                    3' HO AUUCCUCCACUAG - 5'

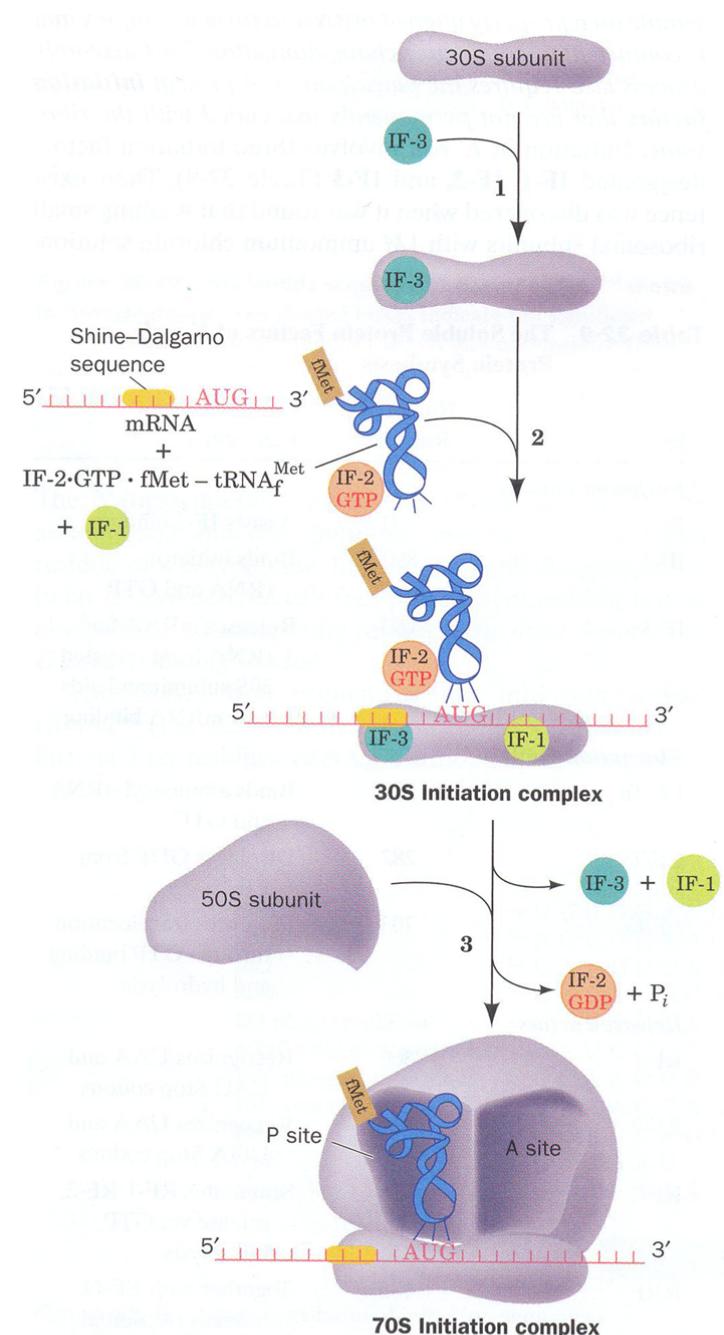
- c. Prokaryotes use Shine-Dalgarno to differentiate initiation/internal Met, because many mRNAs are polycistronic.
  - d. Eukaryotes use the 5' cap instead.
  - e. Form ribosomal complex, cut 5' of Shine-Dalgarno recognition segment in 16S rRNA with colicin E<sub>3</sub>, dissociation by SDS, run rRNAs and Shine Dalgarno complex on a gel.
  - f. Initiation begins with fMet-tRNA, not Met-tRNA

## H. Translation Initiation

- 70S ribosome dissociates into 30S and 50S by IF-1 and IF-3.
- GTP-bound IF-2 binds fMet-tRNA
- IF-2/tRNA binds 30S (+/- mRNA)
- IF-3 release → 50S binding
- IF-2 hydrolyzes GTP to GDP → release of IF-1 and IF-2.
- Result: fMet-tRNA is bound to AUG in the P site, and A-site is unoccupied.

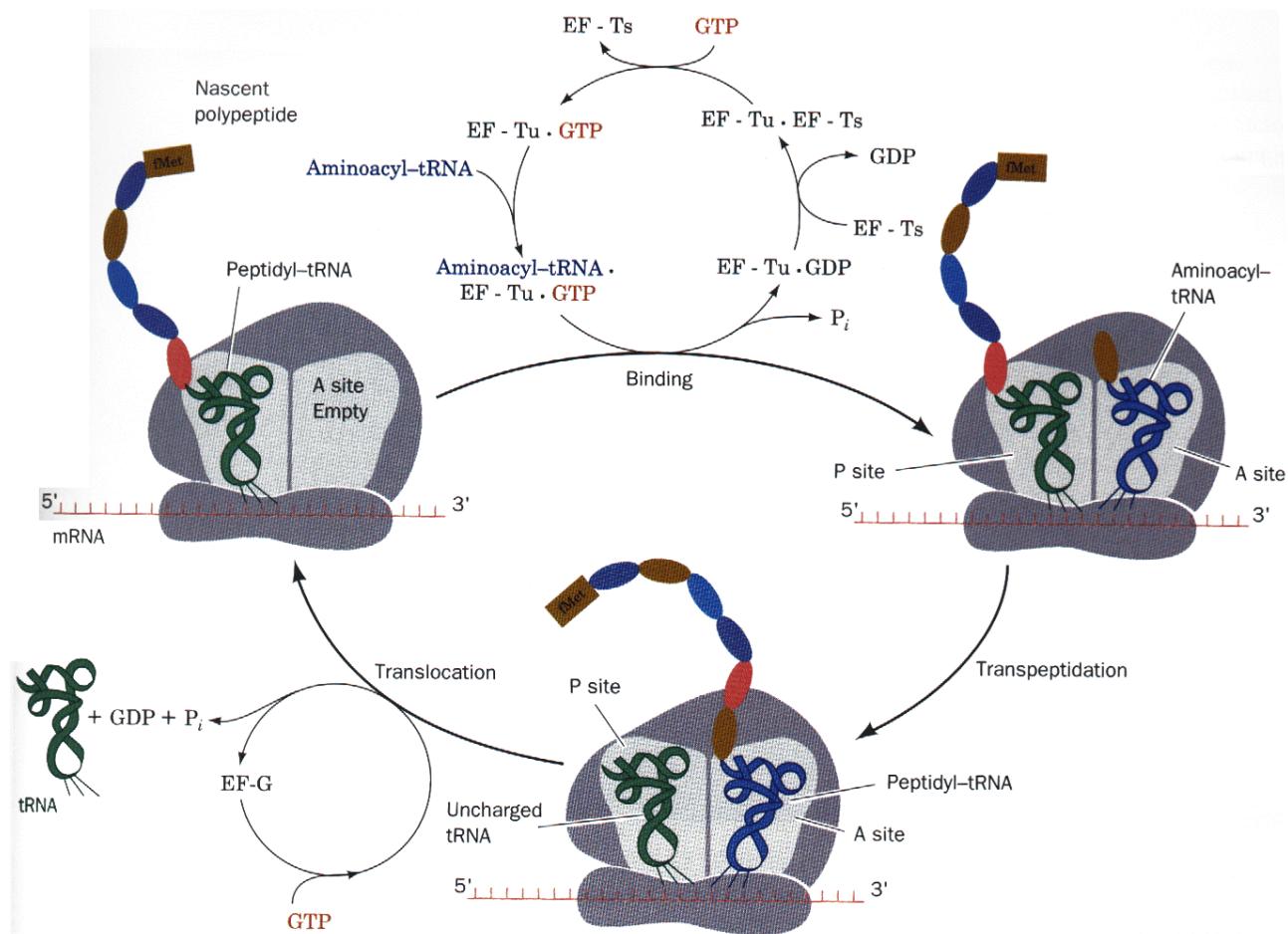
## I. Initiator tRNA

- Only fMet-tRNA can initiate, because it is recognized by IF-2.
- EF-Tu recognizes Met-tRNA.
- Transformylase formylates Met → fMet after it attaches to  $\text{tRNA}_f^{\text{Met}}$ .
- Formyl group (and often Met) are removed from nascent proteins.
- Binding affinity differences between Met-tRNA and fMet-tRNA include unpaired bases at acceptor stem and other base pairing/extrabase.



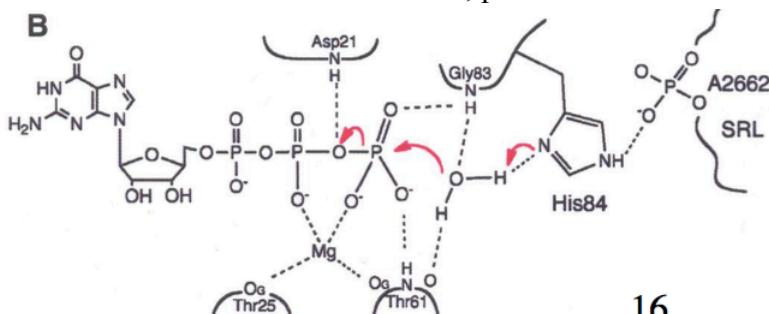
### J. Translation Elongation

- Decoding & Binding: The aa-tRNA + EF-Tu + GTP complex binds to the A site (if the anticodon is complementary).
- Transpeptidation: peptide bond formed by nucleophilic displacement (no energy required).
- Translocation: uncharged tRNA moved from P to E, peptidyl-tRNA moved from A to P. Requires EF-G and GTP hydrolysis.
- EF-Tu + tRNA and EF-G have similar 3D structures, consistent with similar binding positions in the ribosome.



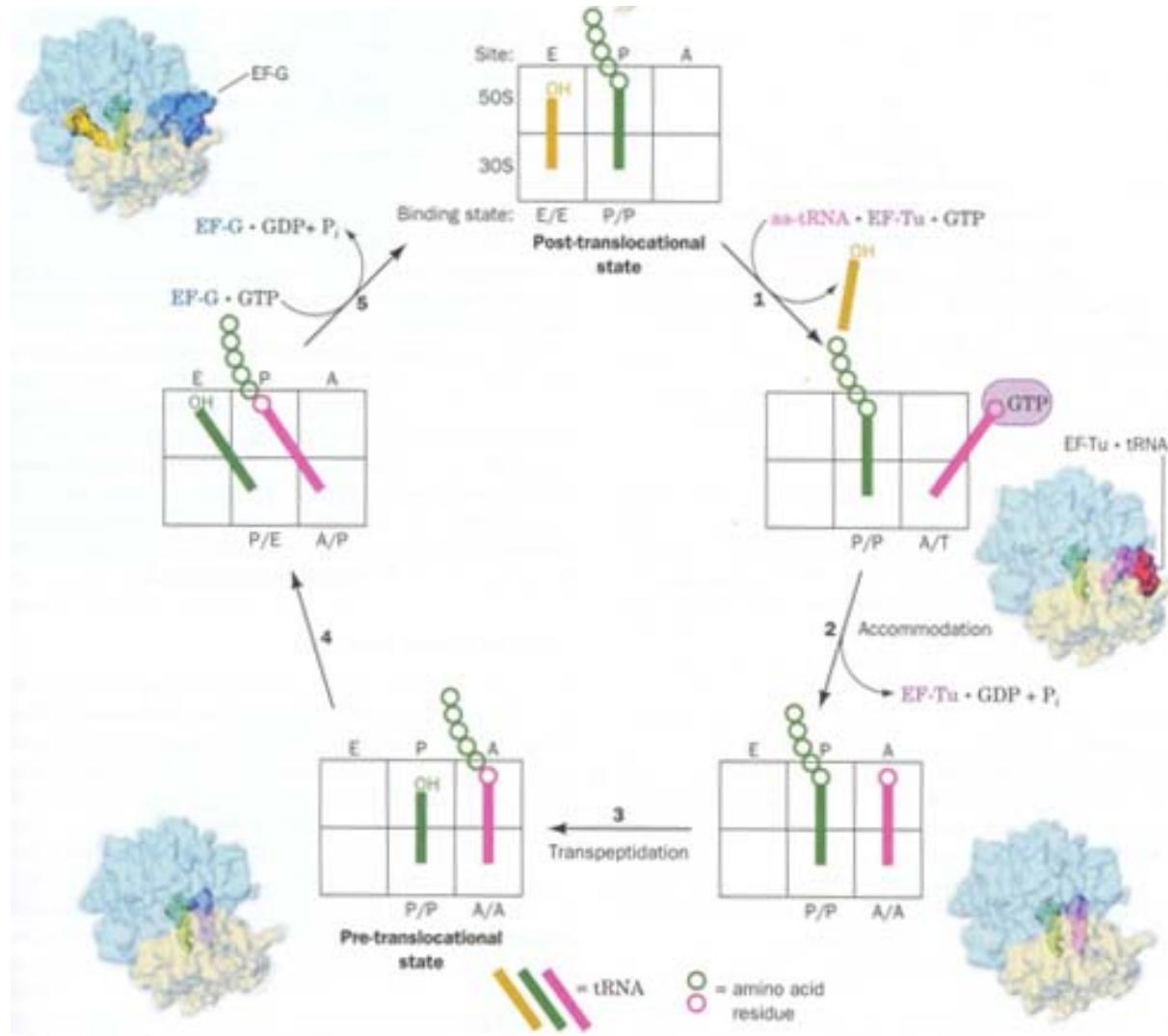
### K. GTP hydrolysis “on” the Ribosome

- EF-Tu, EF-G, IF-2, and RF-3 are all GTPases
- Conserved-A<sub>2662</sub> on 23S rRNA (sarcin-ricin loop) positions conserved-His on GTPases to activate H<sub>2</sub>O to attack GTP gamma-phosphate. [A --- His --- H<sub>2</sub>O → GTP]
- Ricin removes the A residue; potent inhibitor of translation.



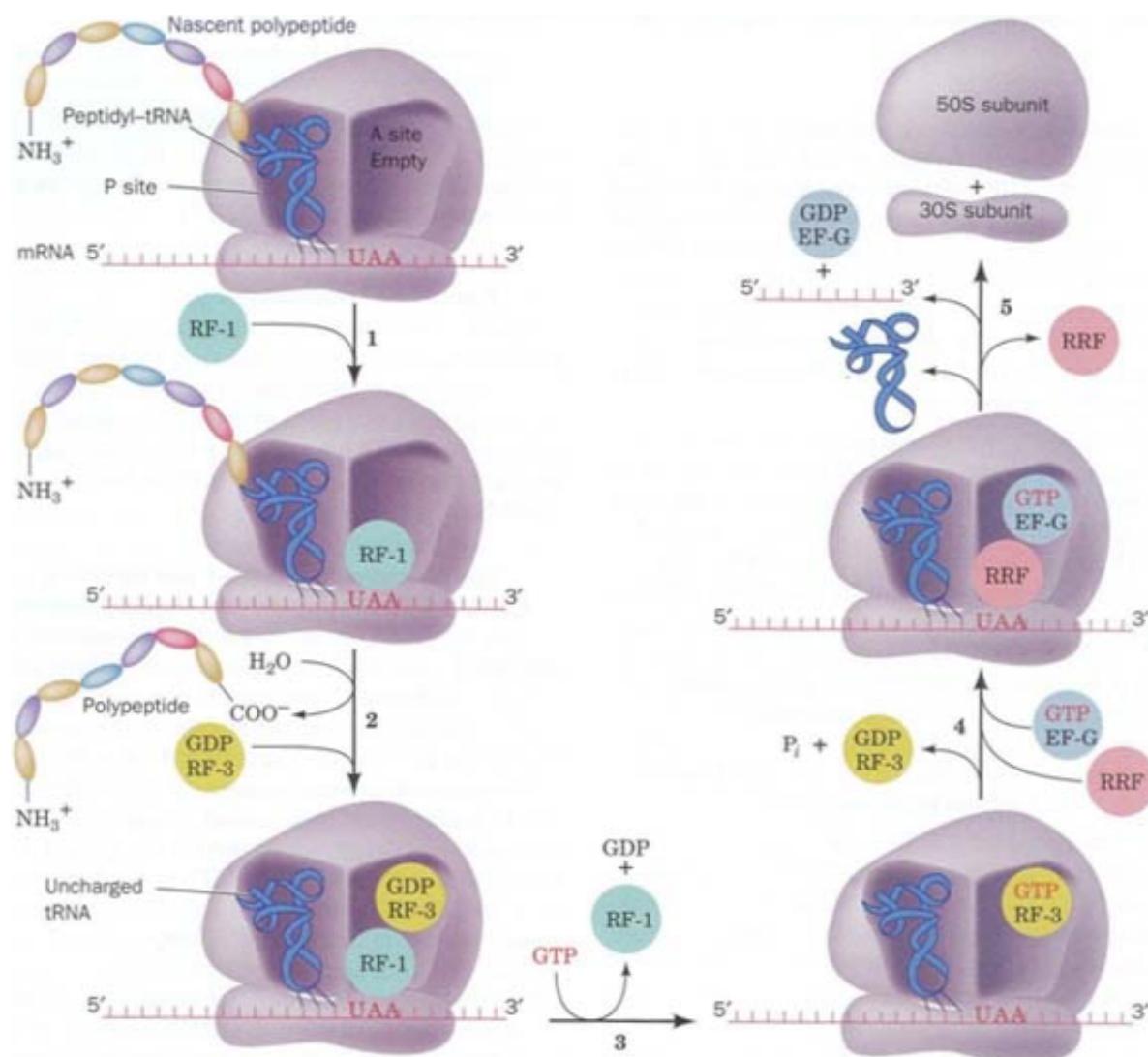
## L. tRNA Translocation

- Hybrid sites model: peptidyl transfer goes through intermediates
- New peptidyl tRNA intermediate is half in the P site and half in the A site
- Newly deacylated tRNA is half in the E site and half in the P site
- EF-G moves completes the movement to the E and P sites.



## M. Translation Termination

- RF-1 or 2 binds to the empty A site
  - RF-1 recognizes UAA/UAG
  - RF-2 recognizes UAA/UGA
- Peptide transferred to H<sub>2</sub>O (rather than aa-tRNA)
- RF-3/GDP binds to the A site alongside RF.
- GTP replaces GDP and releases RF.
- RF-3 hydrolyzes GTP, and the resulting RF-3/GDP leaves.
- RRF (ribosome recycling factor) and EF-G/GTP bind to the A site.
- EF-G-bound GTP is hydrolyzed →
  - Uncharged tRNA leaves the P site.
  - RRF and EF-G leave the A site.
  - 30S, 50S, and mRNA come apart.

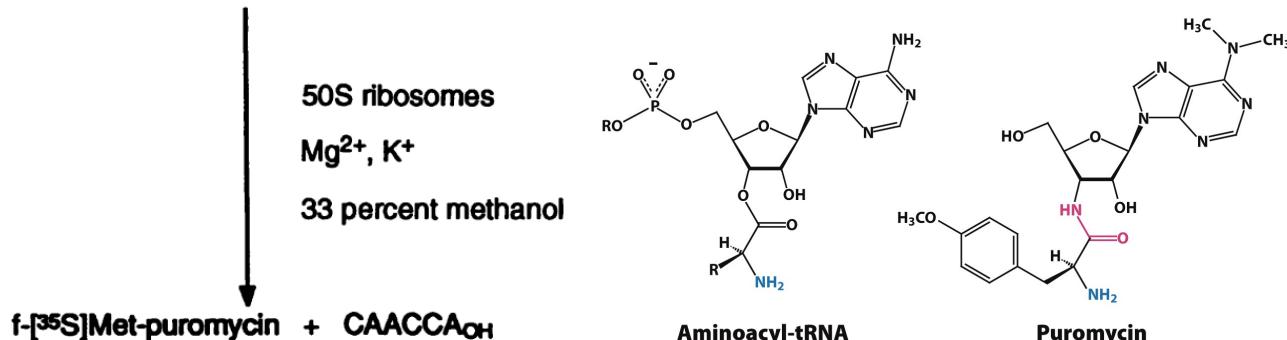


## Lecture 12: Protein Synthesis II

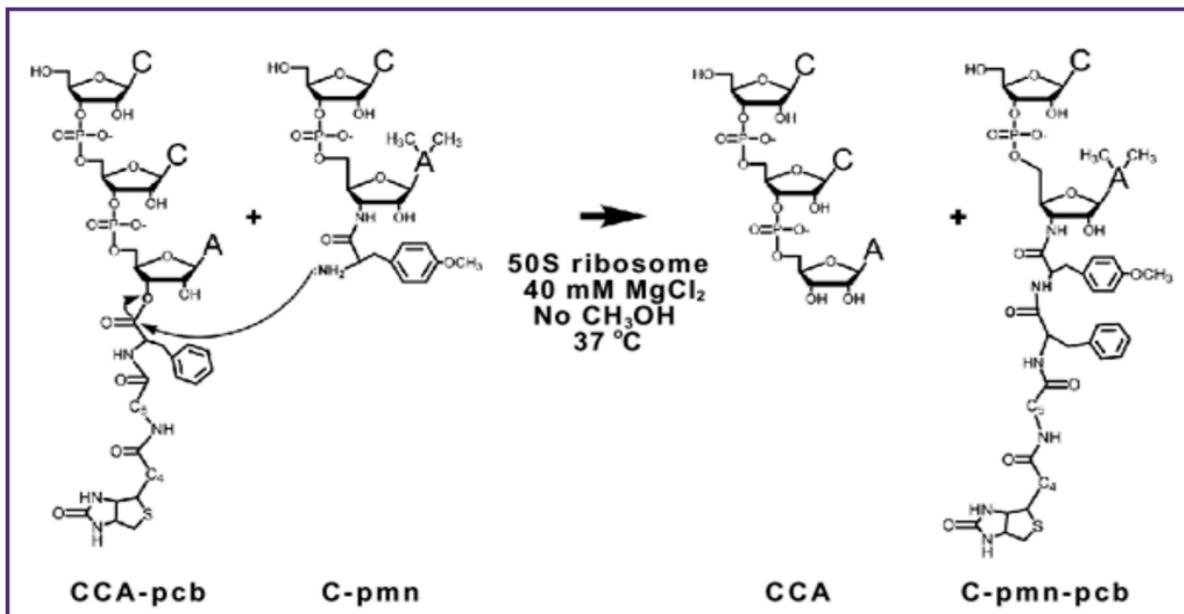
### A. Evidence of Ribosomal RNA Active Site

- a. Fragment reaction: model assay for peptide bond formation.
  - i. CAACCA (f-[<sup>35</sup>S]Met) – severed tRNA acceptor stem oligo + labeled fMet AA.
  - ii. Puromycin: mimics aa-tRNA.

**CAACCA (f-[<sup>35</sup>S]Met) + puromycin**



- b. Extraction of protein from [<sup>35</sup>S]-Met-labeled ribosomes in *T. aquaticus*.
- c. During successive treatment of SDS, Proteinase K, and phenol extraction:
  - i. [<sup>35</sup>S] Protein % went down from 100, 12, 7, 5.
  - ii. Peptidyl transferase activity stayed constant.
- d. Treatment with RNase T1, carbomycin, or chloramphenicol all destroyed peptidyl transferase activity (no band for fMet-puro).
- e. **What are the controls? Not adding any reagent.**
- f. **What level of ambiguity remains?**

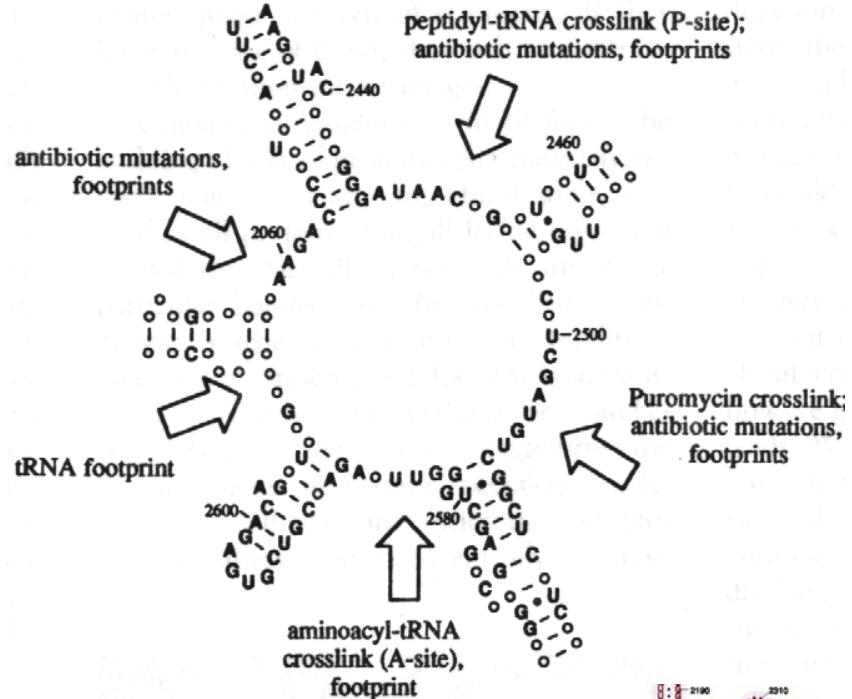
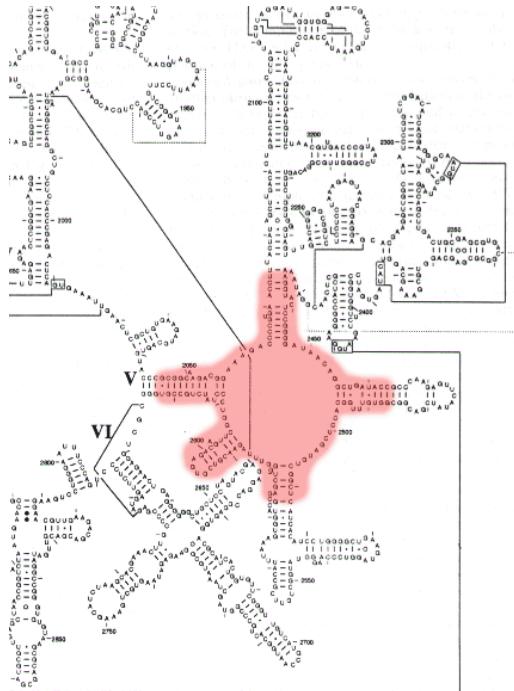


**Figure 1. Schematic of the modified fragment assay.**

The substrates are shown on the left. CCA-phenylalanine-caproic acid-biotin (CCA-pcb) and C-puromycin (C-pmn) undergo a ribosome-dependent reaction in which a peptide bond is formed between the  $\alpha$ -amino group of C-pmn and the carbonyl ester of the phenylalanine moiety of CCA-pcb, yielding the two products: C-puromycin-phenylalanine-caproic acid-biotin (C-pmn-pcb) and a deacylated CCA.

## B. E. Coli 23S rRNA and the Peptidyl Transferase Active Site

- a. Evidence that 23S rRNA Domain V is part of the peptidyl transferase active site
  - a. **Footprinting** and crosslinking of aa-tRNAs
  - b. Some mutations in this segment give antibiotic resistance
  - c. Antibiotics crosslink here.



## C. tRNA Binding to 23S rRNA

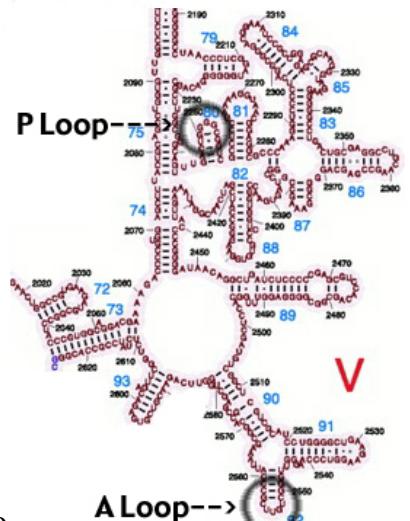
- a. Photocrosslinking, chemical footprinting, and mutagenesis identified site of tRNA CCA-3' binding in the A and P sites.
- b. A loop: binds A-site tRNA
  - i. G<sub>2553</sub> base pairs with middle C<sub>75</sub> in CCA-3'
  - ii. A<sub>76</sub> makes A-minor pair with the A loop.
- c. P loop: binds P-site tRNA
  - i. G<sub>2252</sub> base pairs with first C<sub>74</sub> in CCA-3'
  - ii. C<sub>75</sub> pairs with another G
  - iii. A<sub>76</sub> makes A-minor pair with the P loop.

## D. T. Steitz: Structure of 50S Ribosomal Subunit at 2.4 Å resolution

- a. 50S interior face (left) is RNA-rich; its exterior face (right) is protein-rich
- b. Subunit interaction by RNA-RNA contact.

## E. RNA in the 50S

- a. rRNA is traditionally in 6 domains, but it's monolithic in the actual structure.
- b. Conserved sequences concentrated in active site (others for tertiary structure)
- c. Expansion sequences are extended RNA regions that are NOT conserved.



## F. Protein in the 50S

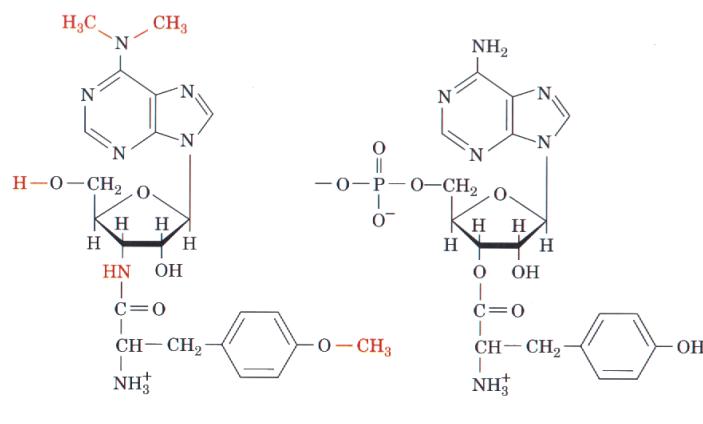
- a. All ribosomal proteins have globular bodies
- b. 13/30 have long unstructured protrusions in the low-protein interior that contact rRNA.  
Riboproteins act as “mortar” to stabilize the 3D structure of RNA “bricks.”
- c. Most proteins interact with multiple rRNA domains.
- d.  $\frac{1}{2}$  of the 23S rRNA nucleotides contact proteins by van der Waals.
- e. 50S interior is mostly basic AA, and **50S exterior is mostly acidic AA**.  
The interior basicity (+) helps stabilize the negative DNA backbone.

## G. Antibiotics and Ribosomal Function

- a. Most antibiotics selectively act on bacterial ribosome function and bind rRNA
- b. We can study translation by using antibiotics to block it at different stages.
- c. Antibiotics can inhibit proper function of: (1) peptidyl transferase, (2) translocation, (3) elongation factors, (4) mRNA reading, (5) chain initiation, (6) binding of aa-tRNAs.
- d. Puromycin (aa-tRNA analog) is especially useful.
  - i. Chain termination of growing peptide: peptidyl transfer attaches puromycin to the growing chain, but it cannot translocate because it has no tRNA body to be moved.
  - ii. Fragment reaction: simple assay for ribosomal function.

**TABLE 30-9. SOME RIBOSOMAL INHIBITORS**

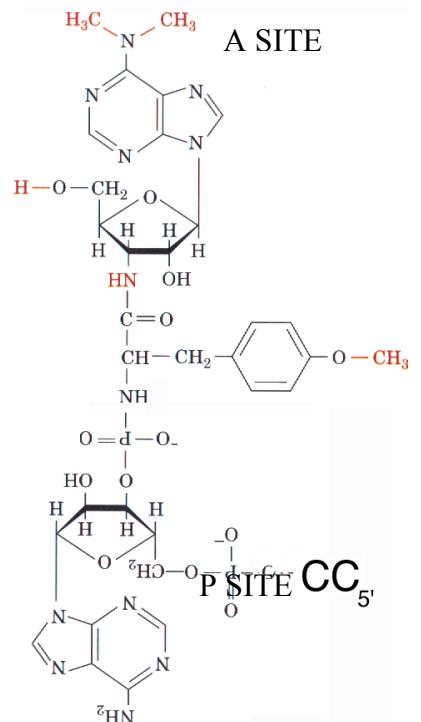
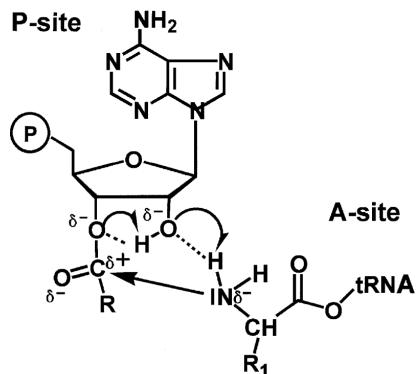
Inhibitor	Action
Chloramphenicol	Inhibits peptidyl transferase on the prokaryotic large subunit
Cycloheximide	Inhibits peptidyl transferase on the eukaryotic large subunit
Erythromycin	Inhibits translocation by the prokaryotic large subunit
Fusidic acid	Inhibits elongation in prokaryotes by binding to EF-G·GDP in a way that prevents its dissociation from the large subunit
Puromycin	An aminoacyl-tRNA analog that causes premature chain termination in prokaryotes and eukaryotes
Streptomycin	Causes mRNA misreading and inhibits chain initiation in prokaryotes
Tetracycline	Inhibits the binding of aminoacyl-tRNAs to the prokaryotic small subunit
Diphtheria toxin	Catalytically inactivates eEF-2 by ADP-ribosylation
Ricin/abrin	Poisonous plant proteins that catalytically inactivate the eukaryotic large subunit



**FIGURE 30-47.** Puromycin (*left*) resembles the 3' terminus of tyrosyl-tRNA (*right*).

## H. 50S Active Site and Catalysis

- a. Most useful antibiotic: puromycin-dACC
  - i. Phosphoramidate linkage resembles the peptidyl tetrahedral intermediate/transition-state.
  - ii. Position within structure helps identify active site
- b. The active site is surrounded exclusively by RNA (closest proteins > 18 Å away).
- c. tRNA CCA-3' in both A-site and P-site base-pairs with 23S.
- d. Ribosomal catalysis relies on entropic stabilization: aligning A-site NH<sub>2</sub> to attack P-site carbonyl.
- e. Proton shuttle mechanism: needs P-site tRNA-A<sub>76</sub> 2'-OH.
- f. A bound H<sub>2</sub>O stabilizes the intermediate oxyanion.
- g. NO metal ions involved.



## I. Exit Channel

- a. Active site in a deep cleft.
- b. Bottom of cleft is a 20Å diameter tunnel that runs through the subunit.
  - i. Allows nascent peptide to move from active site to exit site.
  - ii. Some antibiotics bind in and block the tunnel (molecular constipation).
  - iii. Visualized using heavy-atom cluster molecules.

## J. Antibiotics

- a. Streptomycin: aminoglycoside
  - i. Low concentration: misreads the message.
  - ii. High concentration: prevents chain initiation.
- b. Chloramphenicol
  - i. Binds near A-site and competes with aa-tRNA for binding.
  - ii. Inhibits peptidyl transferase activity of 50S.
- c. Tetracycline
  - i. Binds 30S subunit, blocks aa-tRNA binding.
- d. Erythromycin and Fusidic Acid
  - i. Bind at mouth of peptide exit tunnel- constipation.

## K. X-ray Crystallography of 30S

- a. Structure solved to 3 Å.
- b. Almost complete absence of proteins at 30S internal face.

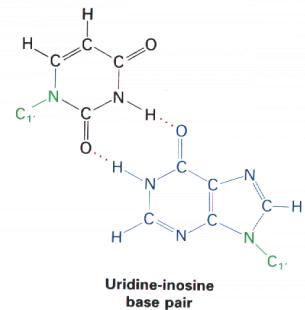
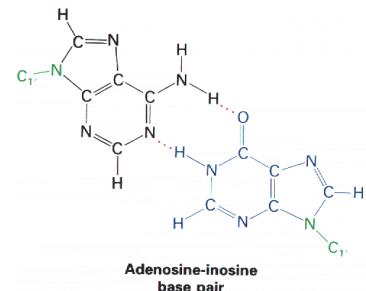
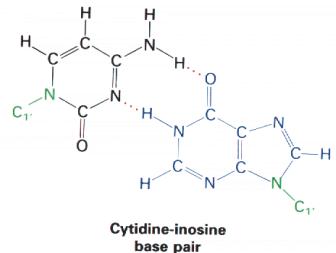
## L. Wobble Hypothesis

- a. Fewer than 61 tRNAs required! ( $4^3 - 3$  stop codons).
  - Each tRNAs may recognize more than 1 codon.
- b. 1<sup>st</sup> and 2<sup>nd</sup> bases are always the same; 3<sup>rd</sup> base has “wobble”
- c. Inosine at anticodon-1 allows C/U/A at codon-3
- d. G at anticodon-1 allows A/G at codon-3

<i>First base of anticodon</i>	<i>Third base of codon</i>
C	G
A	U
U	A or G
G	U or C
I	U, C, or A

#### M. Decoding by the 30S

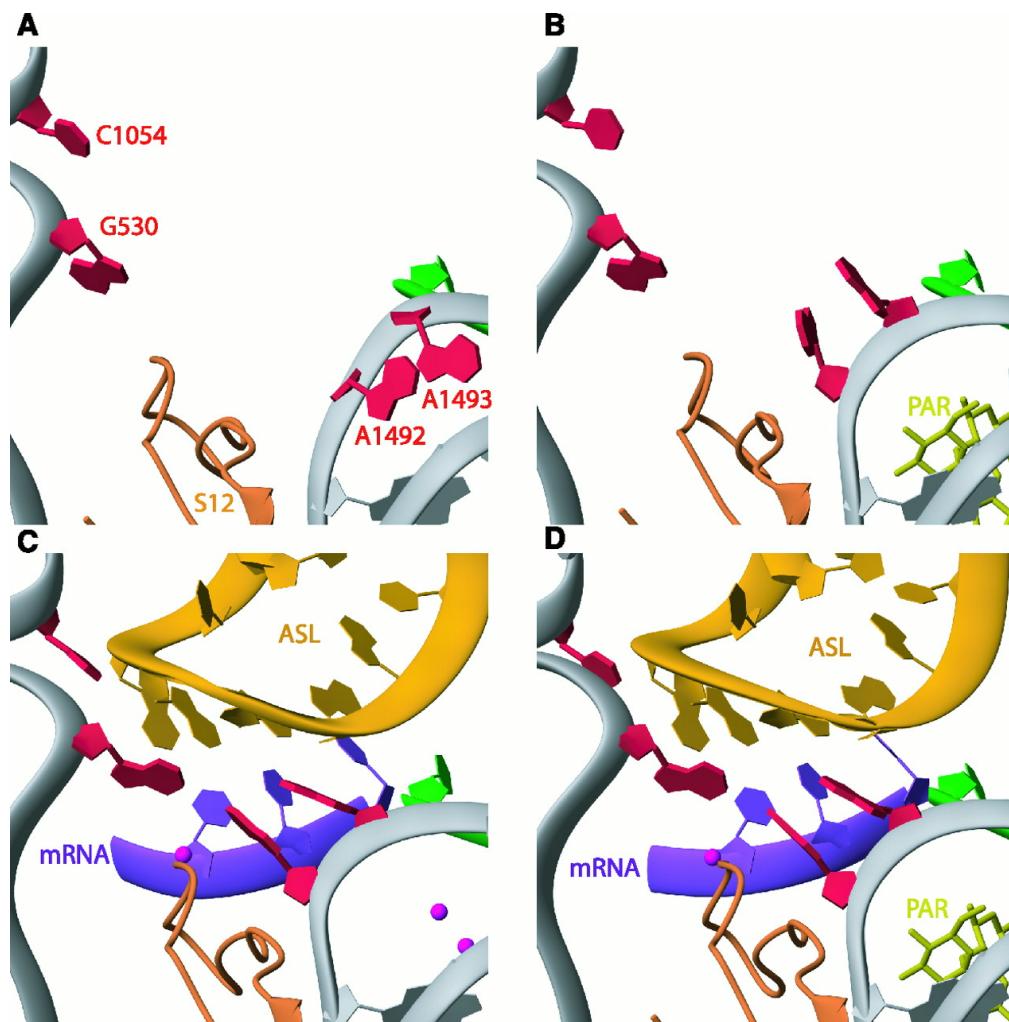
- a. 30S monitors codon-anticodon pairing in the A site.
- b. A-site and P-site tRNAs are bound; E-site tRNA is unbound.  
Sharp kink in the mRNA between A and P that creates distance from the E-site tRNA.
- c. Selection involves:
  - i. Initial recognition step
  - ii. EF-Tu hydrolysis of GTP
  - iii. Proofreading step
- d. Binding of correct (cognate) tRNA in A-site triggers 30S changes:
  - i. A<sub>1492</sub> and A<sub>1493</sub> flip out of an internal loop.
  - ii. G<sub>530</sub> moves from syn to anti.
- e. Codon-1: recognized by A<sub>1493</sub> via A-minor motif; requires WC base-pairing.  
(Same in 23S rRNA recognizing tRNA A<sub>76</sub>-3' in the A/P-sites).
- f. Codon-2: recognized by A<sub>1492</sub> and G<sub>530</sub> together; requires WC base-pairing.  
This is supported by Ser<sub>50</sub> in protein S<sub>12</sub>. Relevant mutations affect translational fidelity.
- g. Codon-3 (wobble): not closely monitored, which allows alternate base-pairing geometries.  
mRNA contacts the ribosome more than the tRNA.



N. Paromomycin: antibiotic that reduces ribosomal fidelity

- Look at G<sub>530</sub>, A<sub>2492</sub>, A<sub>1493</sub>
- Panel A (– paromomycin, – A-site cognate tRNA) – none activated
- Panel C (– paromomycin, + A-site cognate tRNA) – all activated correctly
- Panel B (+ paromomycin, – A-site cognate tRNA) – A,A activated prematurely
- Panel D (+ paromomycin, + A-site cognate tRNA) – all activated, even when incorrect

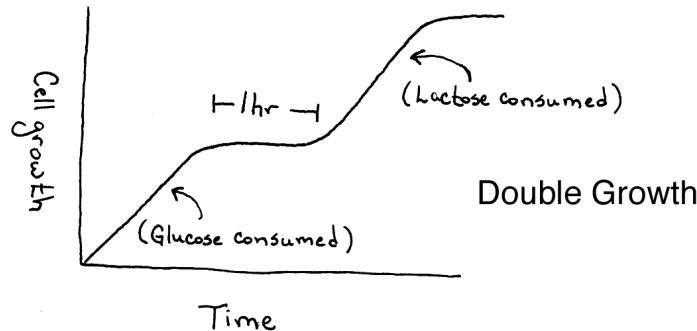
Paromomycin reduces translation fidelity by inducing structural changes that facilitate binding of incorrect tRNAs.



## Lecture 13: Prokaryotic Gene Expression I

### A. Monod and Lactose Biochemistry

- a. E. coli growing on glucose/lactose media have a biphasic growth curve.



- b.  $\beta$ -galactosidase: lactose  $\rightarrow$  glucose + galactose.

- c. Monod and Conn made artificial galactosides for (in)activation/monitoring lac induction.

- i. Original lac Efficient substrate, diminishing inducer
- ii. Types 1 & 2 Efficient substrates, no induction
- iii. Type 3 Efficient substrate, no induction, yellow color
- iv. Type 4 & 6 (IPTG) Resistant to enzyme, strong inducer
- v. Type 5 Resistant to enzyme, antagonist

Types 4-6 have a sulfur substitution that prevents cleavage.

- d. Monod 1951

- i. Grew auxotrophs on minimal media until the required AA runs out.
- ii. Added inducer- no enzyme.
- iii. Added AA- enzyme.
- iv. Conclusion: Enzyme induction requires synthesis (not pre-existing enzyme)
- v. Confirmation with pulse chase: radioactive AA found in new enzymes.

- e. Monod 1952

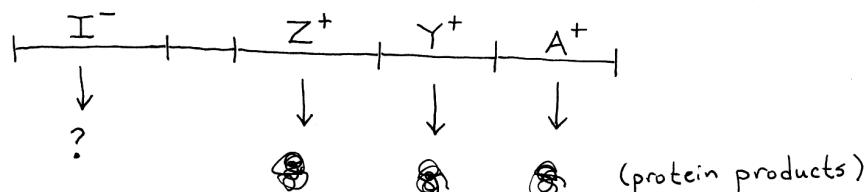
- i. Addition of IPTG (to E. coli growing on glucose medium) induced  $\beta$ -gal activity after 3 min.
- ii. Proved that the inducer is sufficient to produce B-gal.

- f. Cohen and Monod 1954

- i. Constitutive mutation for  $\beta$ -gal expression also constitutively expresses permease.
- ii. This suggests that Y and Z are induced/controlled together.

### B. Inducer (I) of $\beta$ -gal Activity

- a. Genetic map of E. coli places I near Z and Y



- b. In a constitutive I- mutant, all 3 are always on (no need for IPTG).

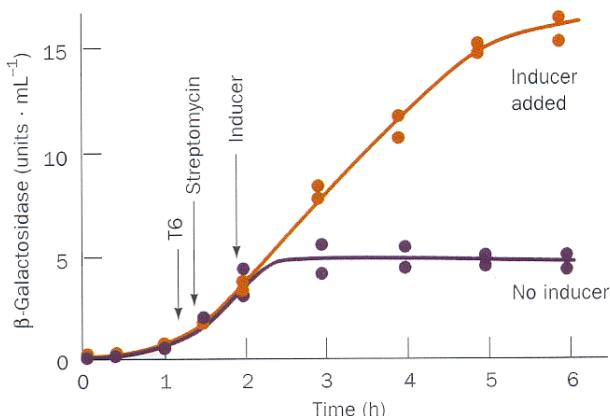
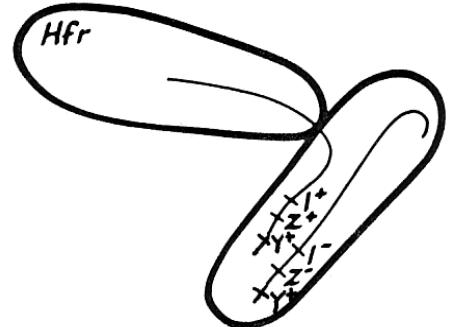
c. Models for the nature of I

1. I is a site needed for control by the inducer (**I must be DNA**).
2. I codes for an activator gp, activated by IPTG to turn on ZYA.
- I- codes for an always-active mutant.
3. I codes for a repressor gp that turns off ZYA unless IPTG stops it.
- I- codes for an inactive mutant that fails to stop ZYA.

C. Pardee, Jacob, Monod (PaJaMo) Experiment

- a. Bacterial conjugation: Hfr delivers 2<sup>nd</sup> chromosome into Hfr- to create merodiploids/partial diploids.
- b. Observe dominant/recessive nature in enzyme induction.
- c. Females were given resistance to streptomycin and T6, which killed off males to ensure that the assay only measures activity in Hfr- cells.
- d. Why do we want Y+ on the Hfr- I-/Z-/Y+?

You want there to be lac available to assay for Z activity. I would guess that conjugational recombination is not super often, so you don't want to count on both Y and Z being fixed.



**FIGURE 29-4.** The demonstration of the existence of the *lac* repressor through the appearance of  $\beta$ -galactosidase in the transient merozygotes (partial diploids) formed by mating  $I^+Z^+$  Hfr donors with  $I^-Z^-F^-$  recipients. The  $F^-$  strain was also resistant to both **bacteriophage T6** and **streptomycin**, whereas the Hfr strain was sensitive to these agents. Both types of cells were grown and mated in the absence of inducer. After sufficient time had passed for the transfer of the *lac* genes, the Hfr cells were selectively killed by the addition of T6 phage and streptomycin. In the absence of inducer (lower curve),  $\beta$ -galactosidase synthesis commenced at around the time that the *lac* genes had entered the  $F^-$  cells but stopped after  $\sim 1$  h. If inducer was added shortly after the Hfr donors had been killed (upper curve), enzyme synthesis continued unabated. This demonstrates that the cessation of  $\beta$ -galactosidase synthesis in uninduced cells is not due to the intrinsic loss of the ability to synthesize this enzyme but to the production of a repressor specified by the  $I^+$  gene. [After Pardee, A.B., Jacob, F., and Monod, J., *J. Mol. Biol.* 1, 173 (1959).]

D. Additional

- a. If I is a DNA sequence, then what??? Need cis-trans assay (next page).
- b. No mixing of cytoplasm during conjugation; the gp of I is in the cytoplasm.
- c. If I is an activator:
  - i. I- is dominant over I+ because a rogue activator can't be stopped by a discriminating one.
  - ii. After conjugation,  $\beta$ gal+ no matter what.
- d. If I is a repressor:
  - i. I+ is dominant over I-, because a working repressor can step in for the broken one.
  - ii. After conjugation,  $\beta$ gal depends on IPTG.
- e. Hfr I-/Z- x F- I+/Z+ gives no  $\beta$ -gal because the there will always be repressor floating around from the active I+.

E. Repressor must bind IPTG

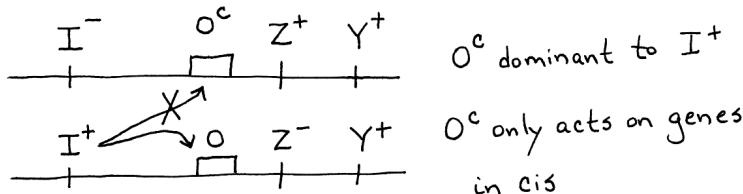
- a. I<sup>S</sup> mutants (super-repressors) cannot be inactivated by IPTG.

b.  $I^S$  dominant over  $I + \text{IPTG}$  because nothing will pop these guys off.

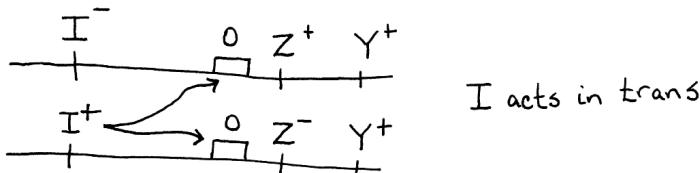
F. Repressor must act at DNA level on an operator

a.  $O^C$  is a constitutive operator that's insensitive to repressor

i.  $O^C$  acts in cis: not cytoplasmic (no gene product).



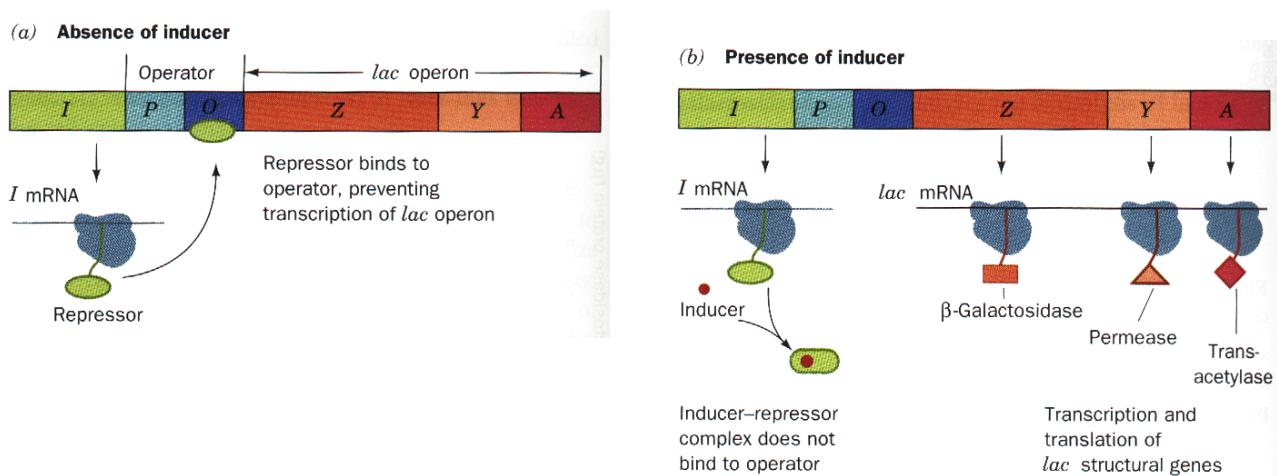
ii.  $I^+$  acts in trans: cytoplasmic gp and can affect other chromosomes.



iii. Dominant to  $I^+$ , because working repressor won't stop these guys.

b. Operon model of gene regulation: coordinated units of expression (operons) have an operator and gene set that are controlled by distant regulatory genes.

G. Model for Negative Regulation of Lac Operon



H. The Lac Repressor

a. Muller-Hill and Gilbert 1966 isolated lac repressor using labeled  $^{14}\text{C}$ -IPTG

i. Tetrameric protein (1 IPTG per subunit).

ii. Affinity for DNA ( $K_d$ ) is  $10^9$  stronger at the operator.

iii. Two domains: DNA-binding domain and IPTG-binding domain.

iv. DNA-binding rate of  $10^{10} \text{ M}^{-1}\text{s}^{-1}$  much faster than diffusion-limited maximum of  $10^7 \text{ M}^{-1}\text{s}^{-1}$ . This suggests a sliding mechanism (1D search instead of 3D).

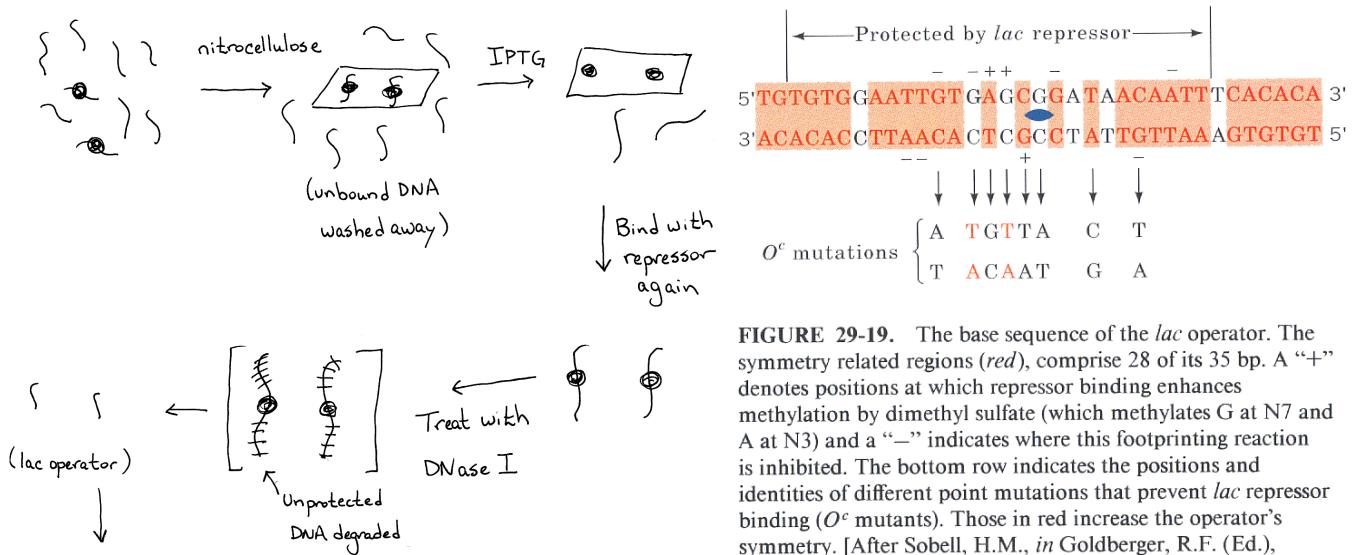
I. Isolating the Lac Operator (DNA segment)

a. Mix purified Lac repressor + sonicated DNA.

b. Bind the mixture with nitrocellulose (holds protein).

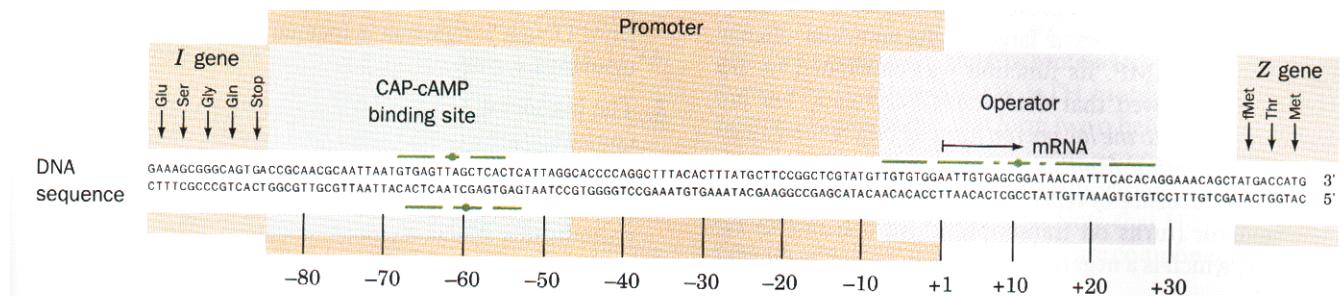
c. Add IPTG. The bound repressor will release DNA.

- d. Add Lac repressor to bind DNA and digest unbound/unprotected segments with DNase I.
- e. Sequence the resulting DNA fragment (convert to RNA first; we're 1960's era savages).
- f. Sequence has symmetric positions (dimeric binding) and  $O^c$  mutation sites (binding sites)



**FIGURE 29-19.** The base sequence of the *lac* operator. The symmetry related regions (red), comprise 28 of its 35 bp. A “+” denotes positions at which repressor binding enhances methylation by dimethyl sulfate (which methylates G at N7 and A at N3) and a “-” indicates where this footprinting reaction is inhibited. The bottom row indicates the positions and identities of different point mutations that prevent *lac* repressor binding ( $O^c$  mutants). Those in red increase the operator’s symmetry. [After Sobell, H.M., in Goldberger, R.F. (Ed.),

#### J. Lac Operator/Promoter Region from lacI to lacZ

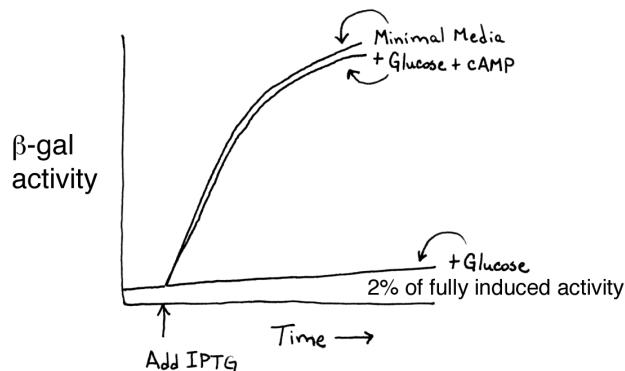
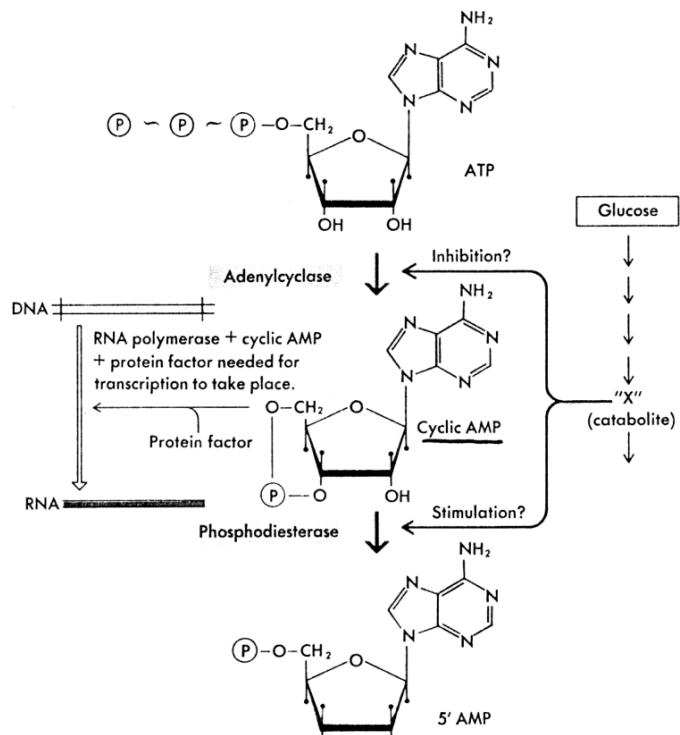
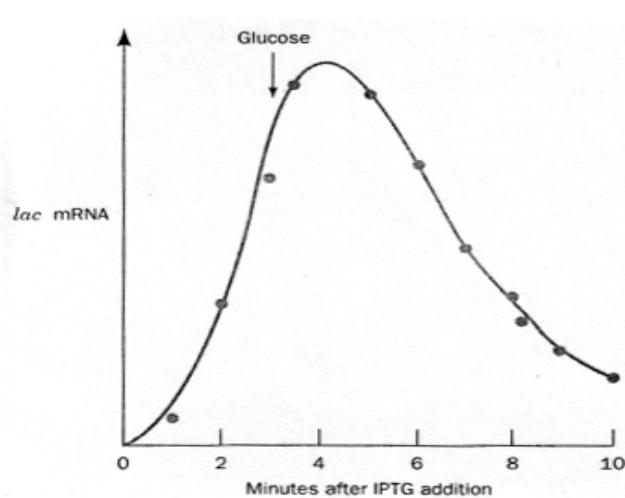


- a. Operator: positions [-7, +28]
- b. RNAP binds between [-20, +20]
- c. Lac repressor and RNAP can bind *simultaneously*.
- d. Mechanism: Lac repressor increases the specific affinity of RNAP for the promoter, so that it cannot leave and enter elongation.

## Lecture 14: Prokaryotic Gene Expression II

### A. Catabolite Repression and Positive Gene Regulation

- a. Lac repressor control alone does NOT fully explain the double growth curve.
- b. Glucose inhibits/decreases lac expression.



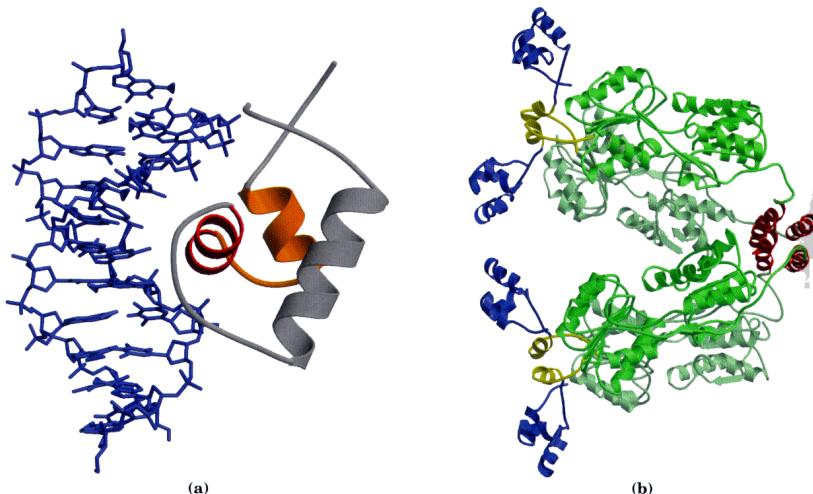
### c. Effects of cAMP

- a. cAMP (5'-3' cyclic AMP) is low when glucose is present.
- b. Addition of cAMP overcomes glucose repression (it must be the ultimate effector).
- c. cAMP binds CAP (catabolite activator protein)
- d. CAP/cAMP binds promoters to activate 100+ glucose-sensitive operons, like lac, arabinose, and galactose.
- d. Thus, lac operon responds to the presence of lactose and the absence of glucose.

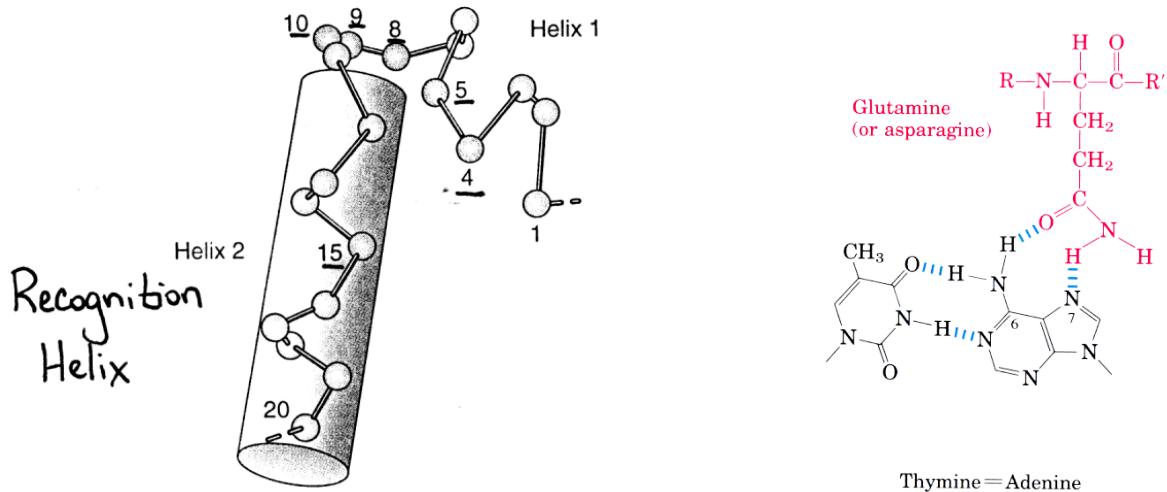
Condition	Result	CAP site	Promoter	Operator	$\beta$ -galactosidase gene
1. No cyclic AMP; no inducer	No transcription because repressor blocks RNA polymerase (and CAP is not bound)				
2. Cyclic AMP present; without inducer	No transcription because repressor still blocks RNA polymerase	cAMP CAP		Repressor	IPTG
3. No cyclic AMP; inducer present (e.g., IPTG)	Little transcription because CAP is not bound				
4. Cyclic AMP and inducer present	RNA polymerase binds and initiates lactose operon mRNA synthesis.			RNA polymerase	mRNA

### B. Lewis 1996 Lac Repressor Structure with DNA and IPTG

- a. Helix-turn-helix structural motif: common for prokaryotic protein-DNA interaction.
- b. Full dimeric lac repressor. Tips are HTH motifs, connected by allosteric helices.

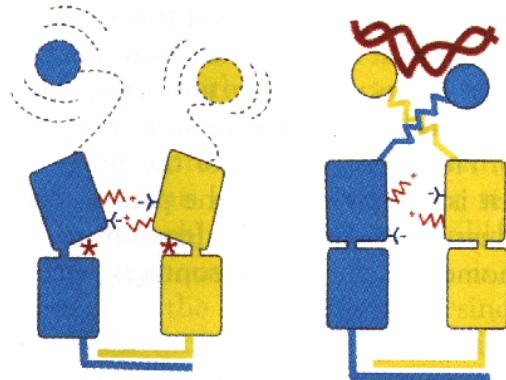


- c. The recognition helix (leftmost, red) interacts with DNA in the major groove. Recognition helices in dimeric proteins are 34 Å apart to bind adjacent major grooves. ( $3.4 \text{ \AA/nt} * 10 \text{ nt turns} = 34 \text{ \AA/turn}$ ).
- d. DNA binding can be disrupted by direct mutations at the interface, or in the support structure that changes the 34 Å distance or other conformational necessities.
- e. Conserved residues
  - i. 4 hydrophobic AAs stabilize helix-helix interactions.
  - ii. 2 AAs create the bend between helices.
- f. The other AAs provide DNA specificity. Many strategies; no reliable rules for specificity.



### C. Lac Repressor during IPTG-Binding

- Lac repressor binding bends DNA by  $\sim 60^\circ$ .
- IPTG-binding pulls apart and misaligns the 34A-apart alpha-helices at the major grooves.
- Asterisks: IPTG.
- Subdomain N-termini electrostatically cross-interact to stabilize the broken dimer. These interactions cease when IPTG dissociates.
- Where are the mutations?
  - $I^S$  mutations should be in the IPTG/lac binding site.
  - $I^-$  mutations should be in the lac promoter binding site.
  - $O^C$  mutations should be at the DNA-helix interface, removing specificity.



### D. CAP during cAMP-Binding

- cAMP stimulates large conformational changes in CAP.
- CAP binds DNA as a dimer in consecutive major grooves (HTH motif).
- DNA is bent by  $\sim 90^\circ$  at the binding site.
- The kinks close the major groove and widen the minor groove.
- CAP recruits RNAP by stabilizing its interaction with promoter DNA.

### E. Model for Lac Repressor Interaction with the Lac Operator by DNA Looping

#### Binding of lac repressor tetramer bends DNA

Lac Repressor: bottom

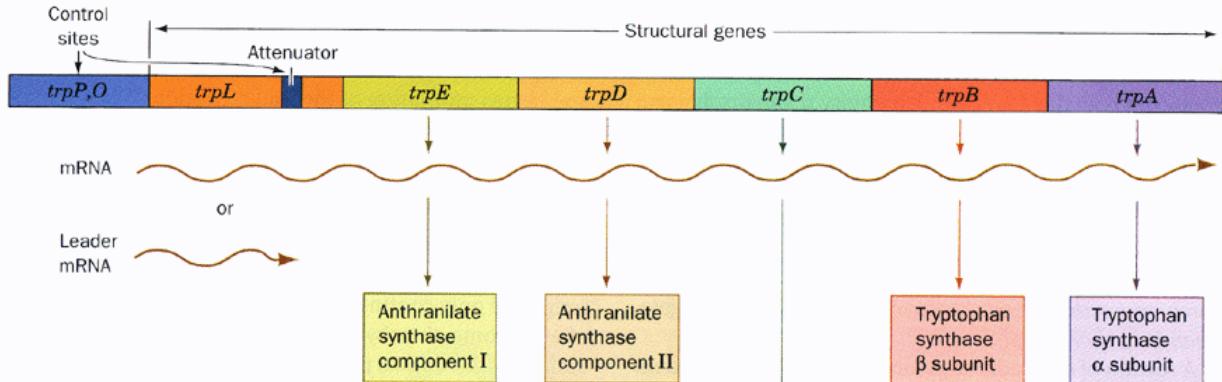
Lac Operators: 2 separate ends of DNA

Promoter: N and SW loop



## F. Transcriptional Regulation of the Trp Operon

- a. Trp biosynthesis enzymes (*trpE-trpA*) are all on a single mRNA transcript
- b. Trp repressor downregulates transcription 70x.
- c. Trp repressor binding to *trp* aligns the recognition helices (HTH motif).



## G. Trp Repressor Binding to DNA Operator

- a. Recognizes sequence indirectly (no direct H-bonding to bases)
  - i. Via shape of phosphate backbone (some solvent-mediated).
  - ii.  $\text{H}_2\text{O}$  mediates polar contacts with bases as an “honorary” side chain.

## H. Trp Repression

- a. Charles Yanofsky: studied TrpA and TrpB Expression
 

$\text{trpR}^+$ and $\text{trp}^+$	most repression (1)
$\text{trpR}^-$ and $\text{trp}^+$	some repression (70)
$\text{trpR}^-$ and $\text{trp}^-$	no repression (600)

 Conclusion:  $\text{trpR}^+$  and  $\text{trp}^+$  both limit expression.

\*NOTE:  $\Delta[\text{gene}]$  refers to a mutation in [gene]

- b. Mutations in 3 genes leads to no repression under conditions of  $\text{trpR}^-$  and  $\text{trp}^+$ 

$\Delta\text{trpL}$ (post-operator region)	no repression (550)
$\text{trpT}_{\text{TS}}$ (trp tRNA)	no repression (500)
$\text{trpRS}_{\text{TS}}$ (trp tRNA synthetase)	no repression (550)

 Conclusion: The trp-mediated repression pathway involves translation and the *trpL* region.

- c. mRNA Levels
  - i.  $\text{trpR}^-/\text{trp}^+$   $\text{trpL} \gg \text{trpE}$  (8x)
  - ii.  $\text{trpR}^-/\text{trp}^-$   $\text{trpL} == \text{trpE}$

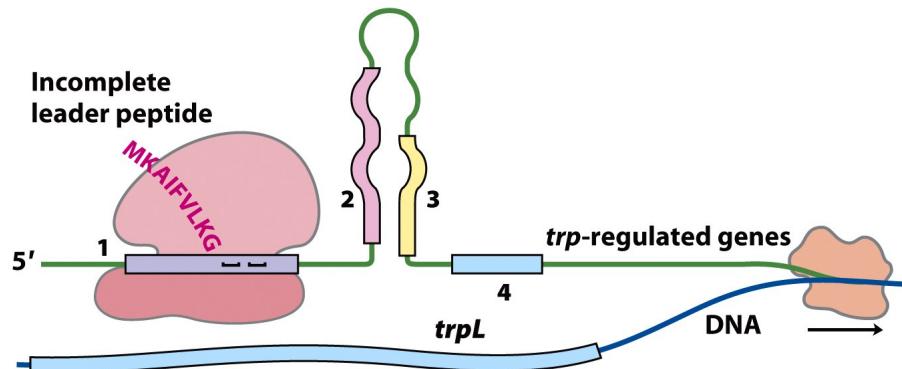
Conclusion: The trp-mediated repression pathway involves mRNA abundance.

- d. Transcriptional control by termination at the *trpL* leader peptide, which has TRP-TRP.

## I. Mechanism

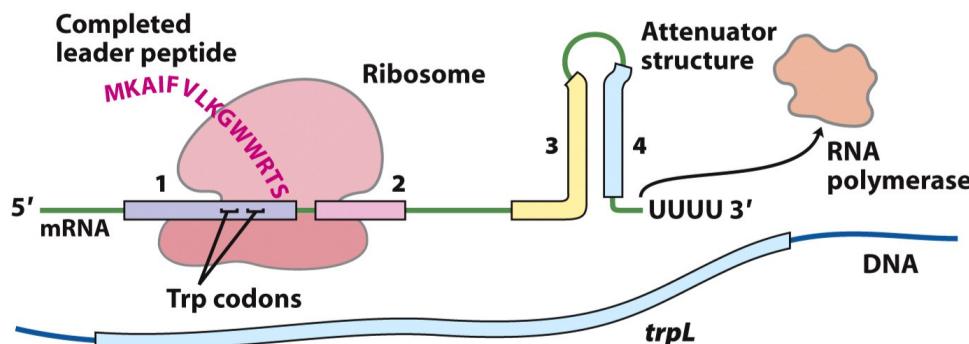
- a. 2 possible secondary structures
  - i. 1-2, and 3-4 terminator
  - ii. 2-3 antiterminator (1 and 4 unpaired)
- b. Low trp
  - i. Ribosome pauses at trp-trp codons in (1), because not enough are available.
  - ii. (2) forms a paired structure with (3).

This stops it from forming the attenuator structure with (4).



c. High trp

- Enough trp is available for the ribosome to quickly translate (1) and get to (2) before (3) is transcribed. After the ribosome passes, (2) pairs with (1).
- (3) is now free to pair with (4) to form a terminator-like attenuator.



J. Attenuation as a General Means of Transcriptional Control

- AA-operons commonly use the attenuation mechanism of containing lots of that AA.

**Table 31-3 Amino Acid Sequences of Some Leader Peptides in Operons Subject to Attenuation**

Operon	Amino Acid Sequence <sup>a</sup>
<i>trp</i>	Met-Lys-Ala-Ile-Phe-Val-Leu-Lys-Gly-TRP-TRP-Arg-Thr-Ser
<i>pheA</i>	Met-Lys-His-Ile-Pro-PHE-PHE-PHE-Ala-PHE-PHE-PHE-Thr-PHE-Pro
<i>his</i>	Met-Thr-Arg-Val-Gln-Phe-Lys-HIS-HIS-HIS-HIS-HIS-HIS-Pro-Asp
<i>leu</i>	Met-Ser-His-Ile-Val-Arg-Phe-Thr-Gly-LEU-LEU-LEU-Asn-Ala-Phe-Ile-Val-Arg-Gly-Arg-Pro-Val-Gly-Gly-Ile-Gln-His

- Advantages:

- Economical
- Sensitive to several AAs
- Links AA synthesis to demand, not supply.
- ONLY works in prokaryotes, where transcription and translation are linked.



# Lecture 15: Regulation of Bacteriophage Lambda

## A. Lambda Phage Cycle

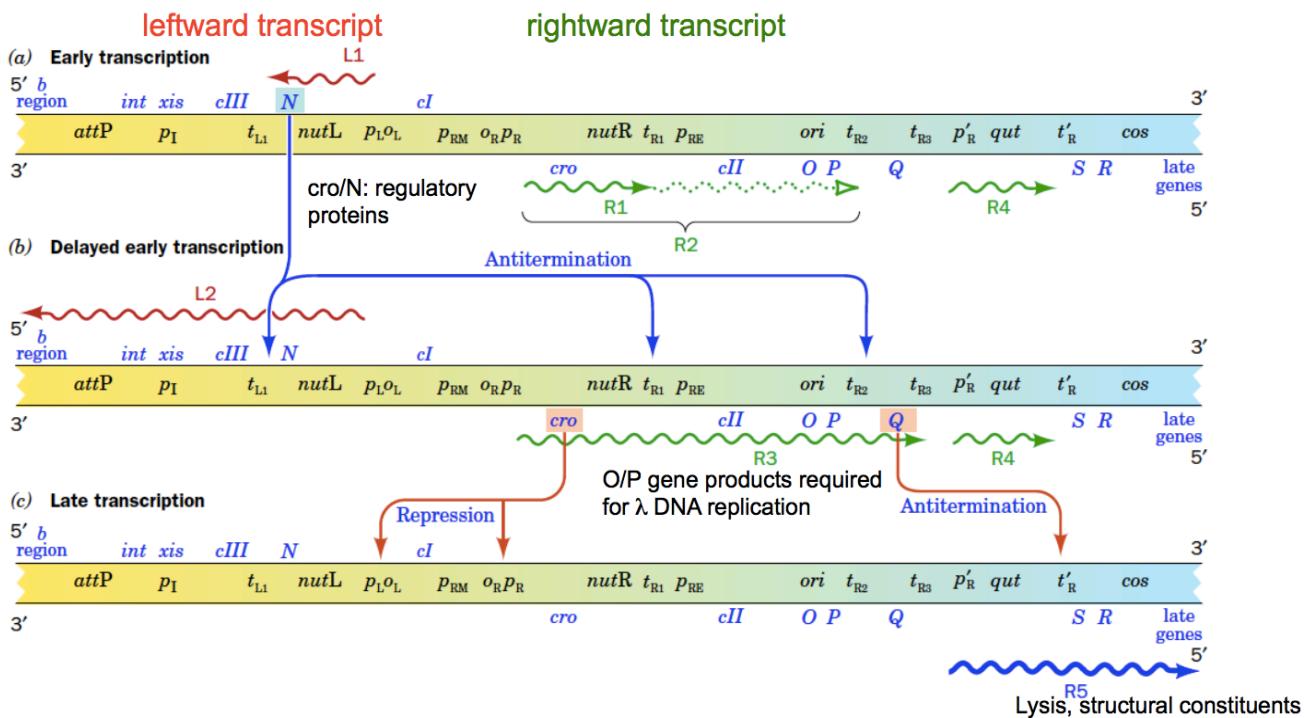
- Phage adsorbs to cell (via tail fibers on maltose transporter) and injects its DNA.
- Linear phage DNA circularizes
- Lysogenic mode: phage DNA stably integrated into host chromosome.
  - Induction: DNA damage (e.g. UV) induces excision → lytic mode.  
This can (rarely) occur spontaneously (every  $10^5$  divisions).
  - Lytic mode: makes DNA/proteins for ~100 progeny phages; lyses host.
- Terminology: prophage (DNA) is integrated into a lysogen (host)
- Good model for gene regulation and viral latency.

## B. Genome organization

- Genes highly organized into structural, regulatory, etc.
- The chromosome circularizes/linearizes at the *cos* site.

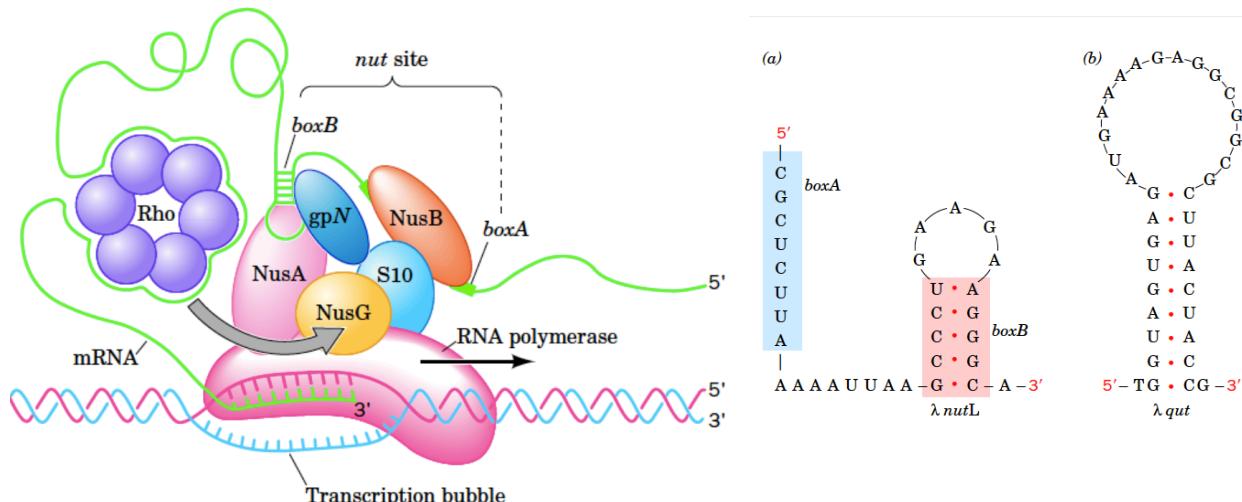
## C. The lytic mode: 3-phase transcriptional program

- Early
  - Active  $P_L$  (left) makes gpN and stopped by  $t_{L1}$ .
  - Active  $P_R$  (right) makes gpCro and stopped by  $t_{R1}$ .
  - Active  $P'_R$  (right-2) makes nothing and stopped by  $t'_R$ .
- Delayed early
  - gpN antiterminates  $t_{L1}$ ,  $t_{R1}/t_{R2}$ , turning on CII, CIII, O, P, Q.
  - gpO and gpP involved in lambda DNA replication.
- Late:
  - Cro blocks  $P_L$   $P_{RM}$ , and  $P_R$  → blocks lysogeny
  - gpQ antiterminates  $t'_R$  → turning on S and R (lysis, structural constituents).



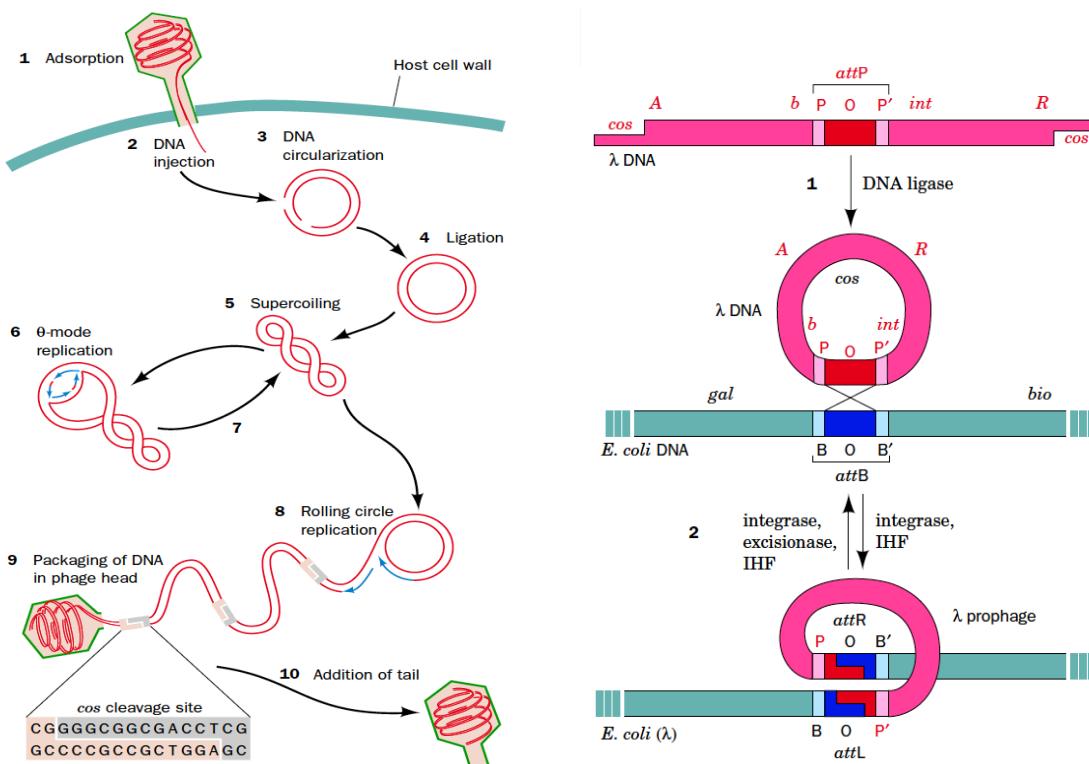
#### D. Anti-termination sites and mechanism

- a. nut = N utilization; qut = Q utilization
  - b. gpN + Nus proteins (host factors) form an antitermination complex on the nut site that ALSO binds transcribing RNAP.
  - c. Two possible ways that it antiterminates:
    - i. This complex inhibits RNAP from pausing at a termination site.  
This stops **rho factor** from catching up and releasing the transcript.
    - ii. gpN modulates NusG-rho interaction to inhibit transcript release.



#### E. Cellular decision: lytic or lysogenic?

- a. Lysogens express ONLY the cl gene (lambda repressor), which stops superinfection.
  - b. Induction dependent on RecA-mediated **SOS response** (often by UV).
  - c. Gene products O and P (delayed early transcription) find ori and initiate DNA replication.

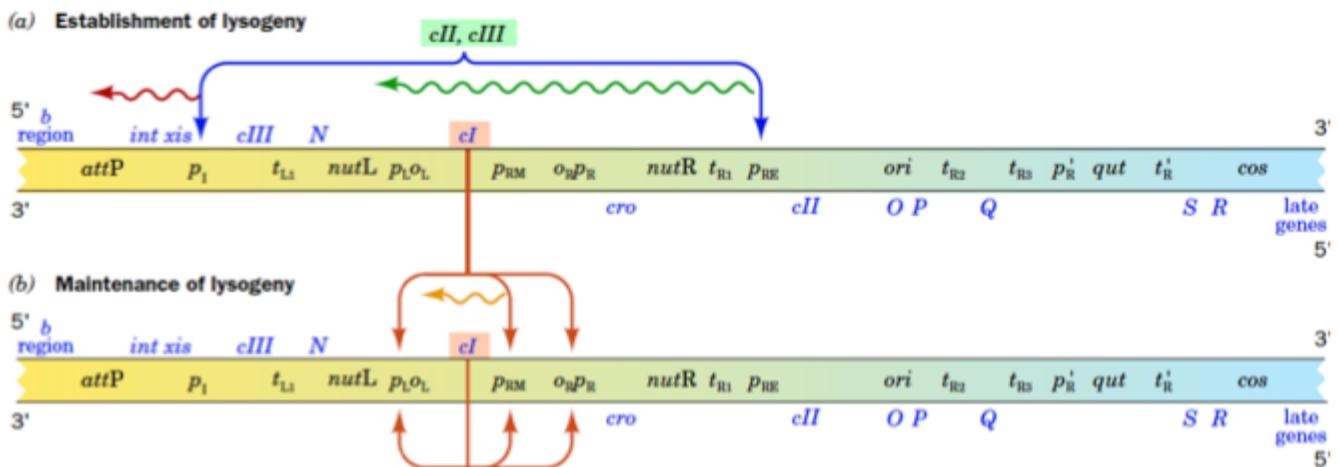


## F. Lambda Assembly

- Two parallel pathways for head and tail assembly.
- Discovered by conditional lethal mutants (allow you to block assembly at specific stages)
  - ts: temperature-sensitive mutants (amber mutations)
  - Suppressor tRNA rescue nonsense mutations until you remove it.
  - Intermediates accumulate and can be inspected by electron microscopy.
- In vivo* genetic complementation: cross infection of 2 mutants to see if their mutations are on the same gene.
- In vitro* complementation: mixing incrementally more purified cell extracts from 2 different mutants that rescue each other to identify protein products.

## G. Lysogenic Transcriptional Reprogramming

- cI repressor binds to operators  $O_R$  and  $O_L$  to turn off everything.
- $P_{RM}$  starts with AUG! How does this affect the protein level?
- After CII and CIII are made in delayed early transcription. The following 2 replace late transcription.



- The two cI mRNAs have different translational efficiencies
  - $P_{RM}$  starts with AUG
  - $P_{RE}$  works 20x better; starts with [Shine-Dalgarno]—AUG
  - $P_{RE}$  starts it; needs to be sensitive.  $P_{RM}$  doesn't care.

## H. Autogenous Regulation of cI transcription by cI at $O_R$

- $P_{RM}$  lies within and overlaps with  $O_R$ .

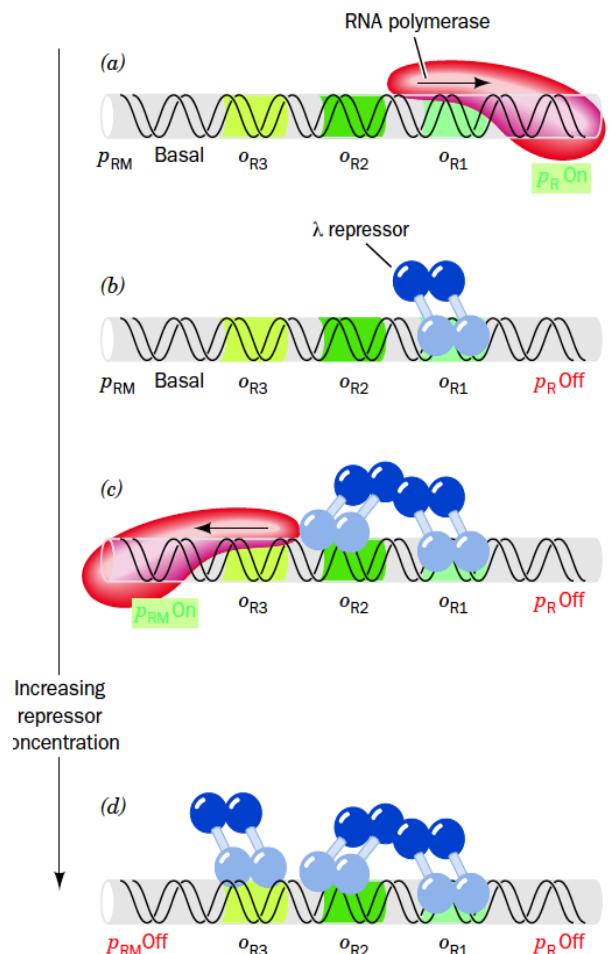
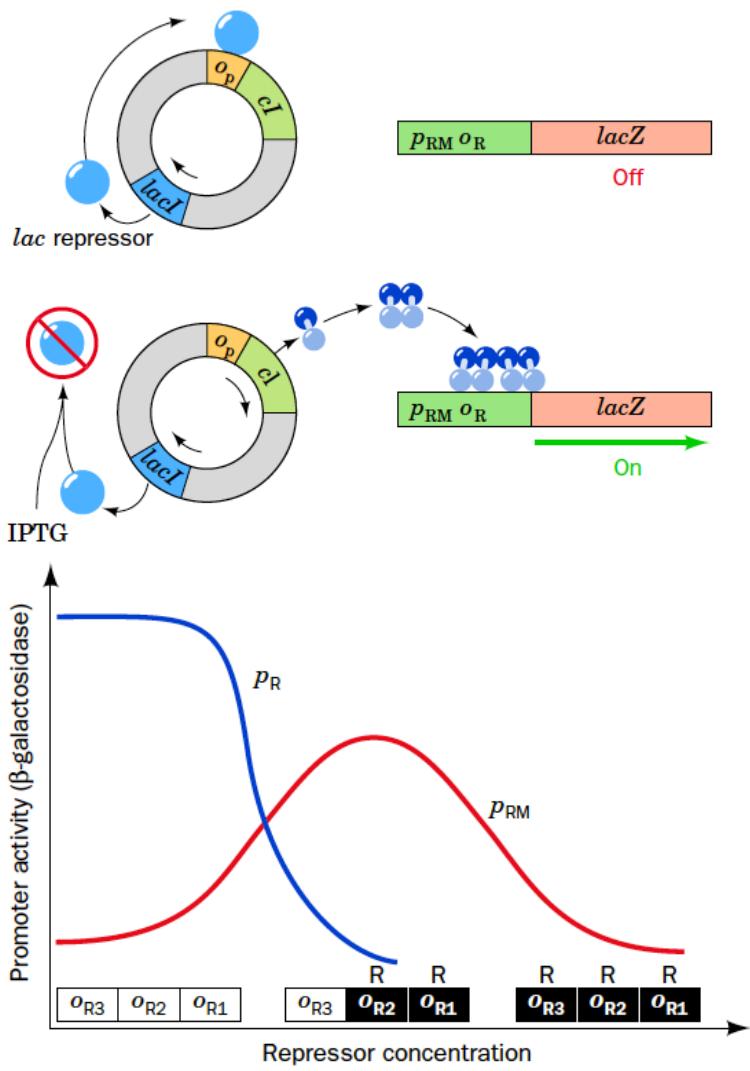
$O_R$ :



- $O_R$  is ~80 bp with 2 divergent promoters and 3 binding sites for lambda Cl repressor.
- Cl repressor is a dimer with a helix-turn-helix DNA-binding motif.

## I. Mechanism for cI Regulation of $P_{RM}$ and $P_R$

- a. Finding this is difficult because:
  - i. The regulated gp is the regulator!
  - ii. No readily assayable product.
- b. Reporter Gene Assay
  - i. Plasmid (left): lac operator-promoter ( $O_p$ ) controls the *cI* gene. *lacI* creates repressor, which can be blocked by IPTG.
  - ii. Prophage (right): contains promoter  $P_{RM}$  fused to *lacZ*. Level of  $\beta$ -galactosidase ( $\text{gp-lacZ}$ ) thus reflects  $P_{RM}$  activity.
  - iii. Similar experiments replaced *cI* with *cro* repressor and  $P_{RM}$  with  $P_R$  promoter.

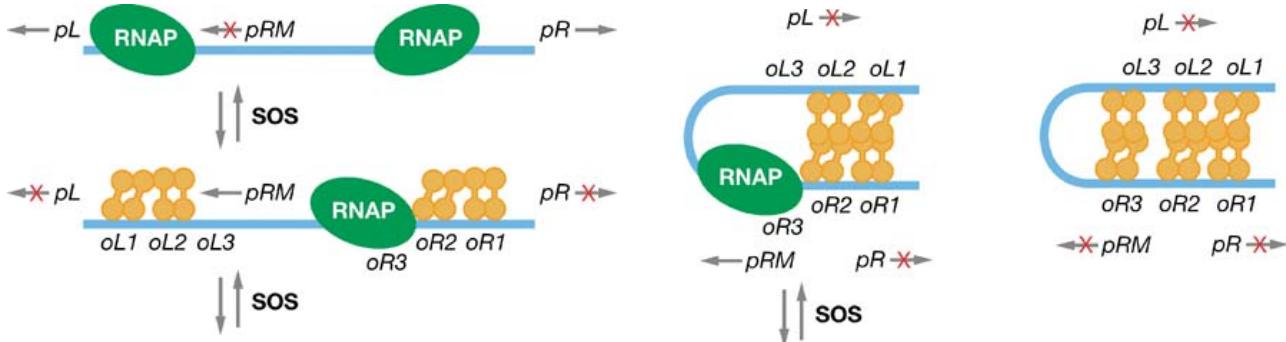


#### A. Promoter activity vs. [*cI*] repressor]

- a. No repressor: Active  $P_R$  and  $P_{RM}$ .
- b. 1<sup>st</sup> repressor on  $O_{R1}$ : Blocks (-)  $P_R$  (10x higher affinity for  $O_{R1}$ )
- c. 2<sup>nd</sup> repressor on  $O_{R2}$ : Boosts (+)  $P_{RM}$  (activates RNAP by contacting sigma)
- d. 3<sup>rd</sup> repressor on  $O_{R3}$ : Blocks (-)  $P_{RM}$  (via RNAP loading); lysogen levels of *cI*.
- e. *cI*- $O_{R1}$  enhances *cI* binding to  $O_{R2}$  – almost always simultaneously occupied.
- f. This keeps [*cI*] in a narrow range for stable lysogeny.  
Too high will release early, and too low will prevent induction when desired.

### B. DNA Looping Between O<sub>L</sub> and O<sub>R</sub> Creates a Cooperative Network

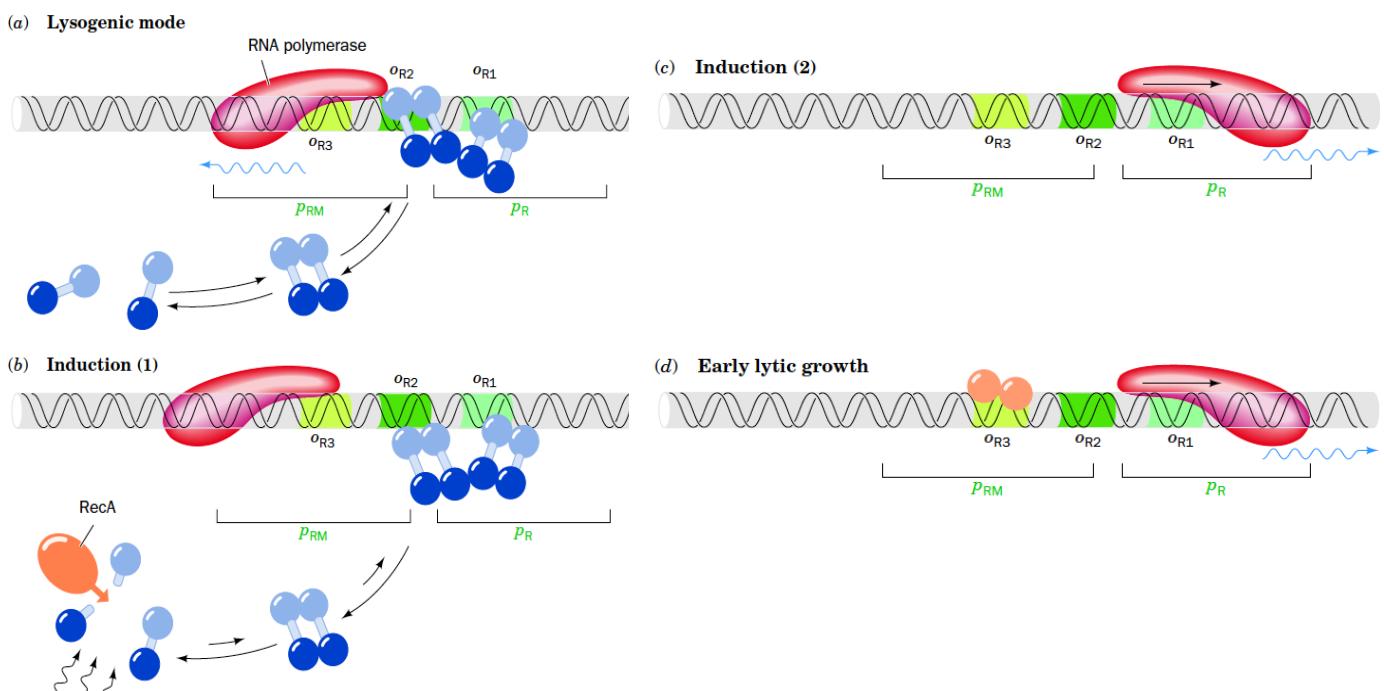
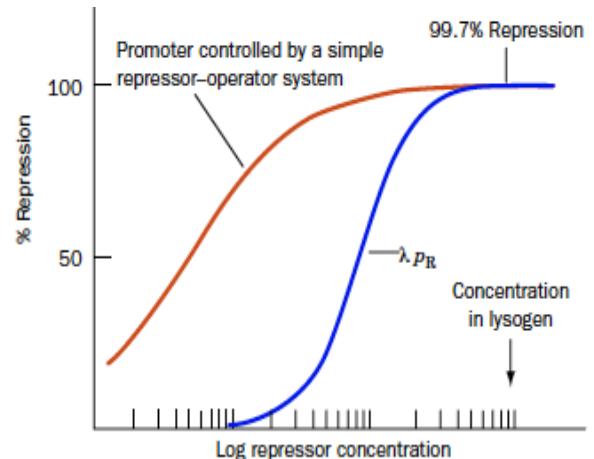
- a. cI binds to O<sub>L1</sub> and O<sub>L2</sub> as well, and the dimers associate in a long-range DNA loop.
- b. cI then binds cooperatively to O<sub>L3</sub> and O<sub>R3</sub>; allows balanced control of P<sub>L</sub> and P<sub>R</sub>.



- c. Based on cooperativity between O<sub>R1</sub> and O<sub>R2</sub>.
- d. Repression curves for lambda P<sub>R</sub> (right) and simple system, like lac operon (left).
- e. Lambda system is more sensitive to changes in [cI repressor].

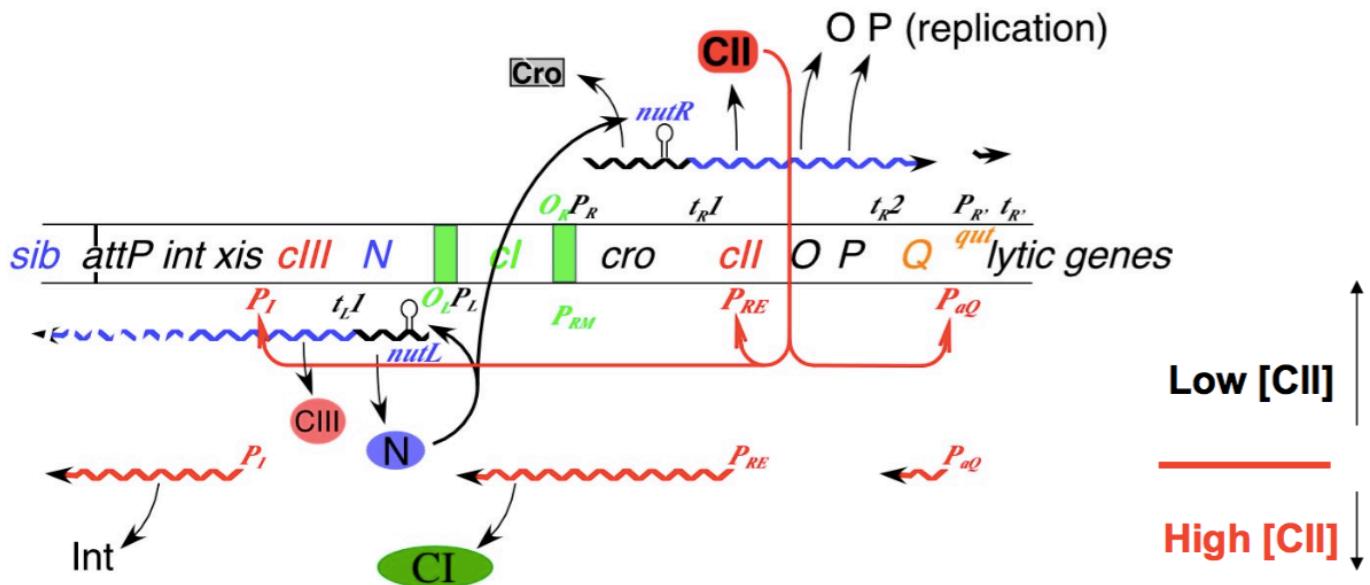
### C. Induction Mechanism

- a. Lysogeny: cI binds O<sub>R</sub> at C-terminus  
cI activates RNAP at N-terminus
- b. DNA damage: RecA stimulates self-cleavage of cI into monomers.
- c. Induction: Early transcription from P<sub>R</sub> and P<sub>L</sub>.
- d. Cro binds O<sub>R3</sub> and blocks cI synthesis → irreversible lytic cycle



#### D. Infection and the Lysis-Lysogeny Decision

- Low [CII] → Lysis, High C[II] → Lysogeny
- Protease FtsH normally degrades CII
  - CIII inhibits FtsH
  - High [cAMP] inhibits FtsH
- High M.O.I. →  $P_R$  and  $P_L$  are ON → N antiterminator →  $P_R$  CII and  $P_L$  CIII are active.
- Starvation conditions → high [cAMP], inhibits FtsH, stabilizes CII.
- CII is a transcriptional activator
  - Binds and activates -35 regions of  $P_I$ ,  $P_{RE}$ , and  $P_{aQ}$ .
  - $P_{RE}$  transcript efficiently translated into CI. WHY? Shine-Dalgarno.



#### E. Summary

- Initial lytic/lysogenic decision dependent on cII and cIII
- Normally, protease activity high, which breaks down cII → lytic cycle
- With limited nutrients, protease activity low, stabilizing cII → lysogenic cycle
- Phage is dormant until better times, when more resources are available.

## Lecture 16: Cloning & Genomic Analysis

---

### A. Molecular Cloning Propagates and Amplifies DNA

- a. Determine the sequence
- b. Make probes for experiments
- c. Express RNA or protein for structural/functional studies
- d. Make mutants for functional analysis
- e. Analyze gene expression in vivo
- f. Make transgenic animals
- g. Genome engineering
- h. Develop tests for diagnosis and treatment of disease.

### B. Key Enzymes of Molecular Biology

- a. Taq DNA Pol: synthesizes new DNA strands (thermal stable)
- b. Polynucleotide kinase: phosphorylates 5'-OH of polynucleotide.
- c. Terminal transferase: adds homopolymer tails to 3'-end of a duplex
- d. Alkaline Phosphatase: removes terminal phosphates on 5' or 3' end.

### C. Type II restriction enzymes

- a. Recognize and cleave dsDNA at specific 4-8 bp palindromic targets.
- b. Can be staggered ends (C2 symmetry) or blunt ends.
- c. Almost every 6-bp palindromic sequence (64 possibilities specified by 3 nt) is cut by at least 1 restriction enzyme.

### D. DNA Ligase

- a. Joins blunt ends and complementary sticky ends of unrelated DNA.
- b. Powered by ATP → AMP + PP<sub>i</sub> or NAD<sup>+</sup> → NMN<sup>+</sup> + AMP (species dependent).
- c. E. coli mechanism
  - i. NAD<sup>+</sup> adenyl group covalently transferred onto Lys residue (phosphoamide adduct)
  - ii. The adenyl group is transferred onto the 5'-(P)-terminus of the DNA nick via 5'-PP<sub>i</sub> linkage.
  - iii. Enzyme catalyzes the formation of the phosphodiester linkage by attack of the 3'-OH on the 5'-(P), which seals the nick and releases AMP.

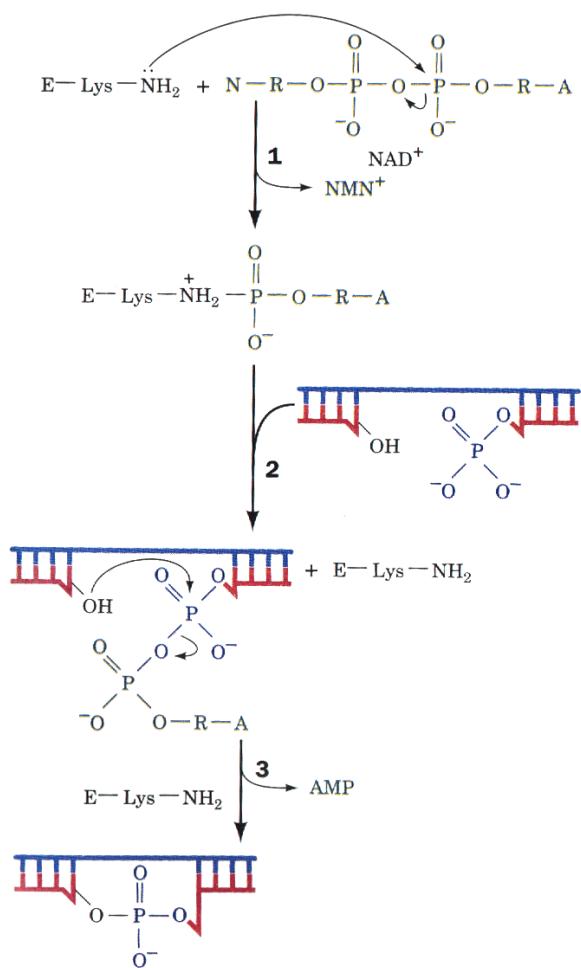
### E. Cloning and DNA Amplification

- a. Plasmid (extrachromosomal DNA) and source DNA are cut with the same restriction enzyme.
- b. Specificity of the annealing reaction allows specific insertions in a plasmid vector.
- c. Transformation and propagation of plasmid within host bacteria will amplify DNA.

### F. Cloning cDNA in Eukaryotes

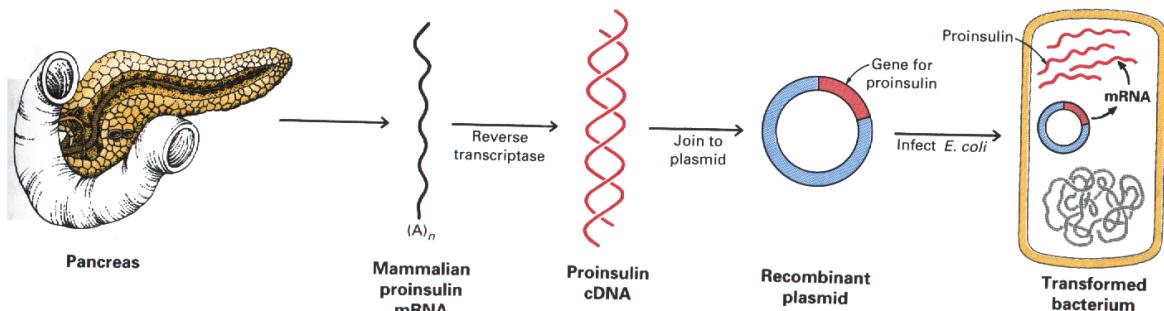
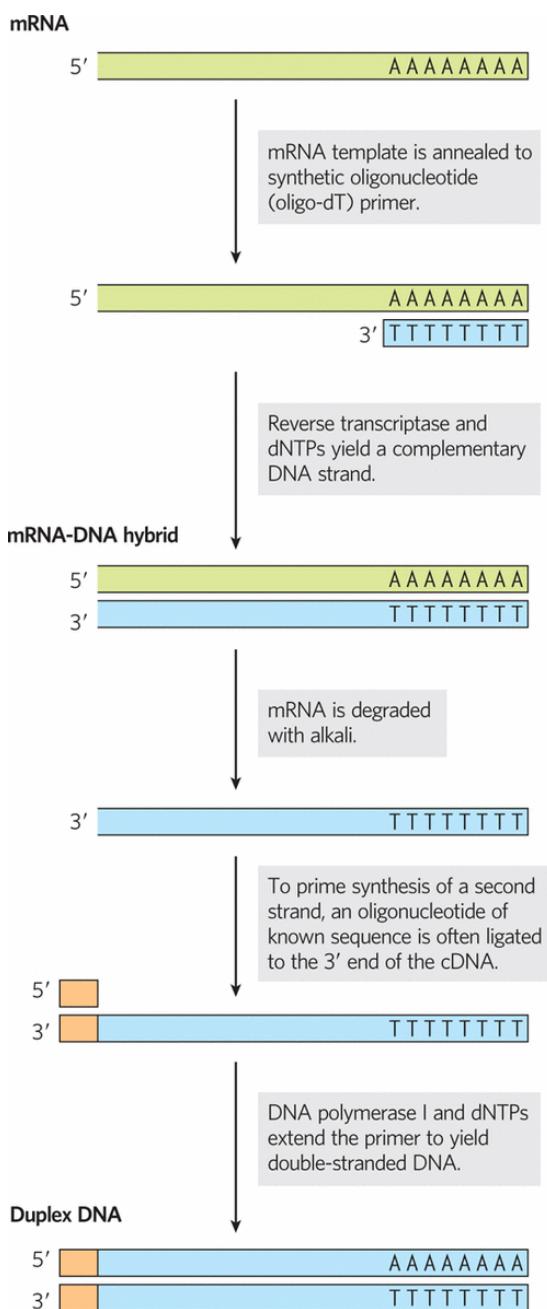
- a. Eukaryotic mRNAs are polyadenylated at 3'-ends.
- b. cDNA is made using poly-T primer + RT: this gives a weighted representation of all expressed genes in a sample.
- c. Each cDNA can be “subcloned” into a plasmid vector, and the net result is a cDNA library (of bacteria).

## DNA Ligation



**FIGURE 31-16.** The reactions catalyzed by *E. coli* DNA ligase. In eukaryotic and T4 DNA ligases, NAD<sup>+</sup> is replaced by ATP so that PP<sub>i</sub> rather than NMN<sup>+</sup> is eliminated in the first reaction step. The numbered steps are described in the text.

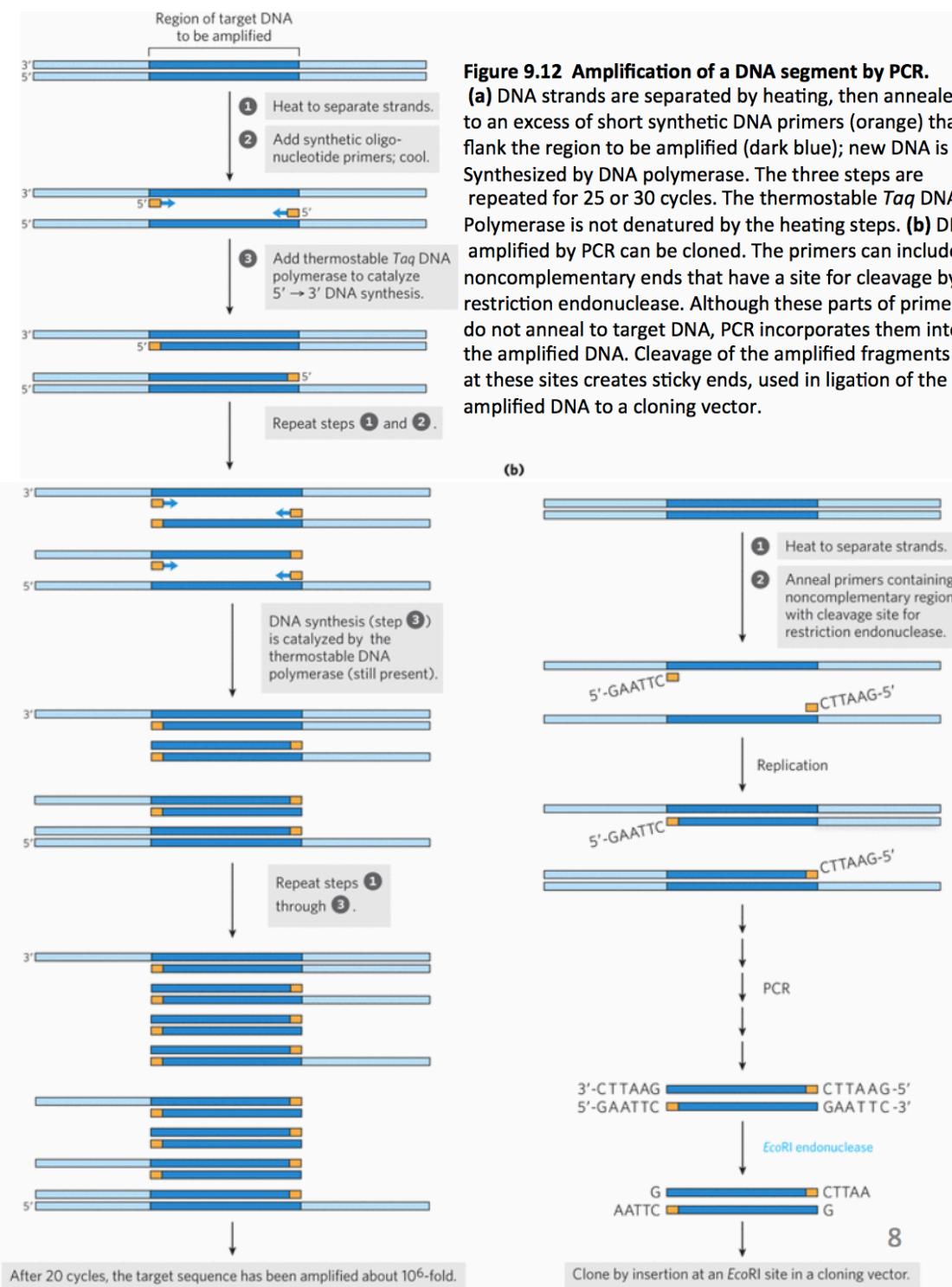
## cDNA libraries





## G. DNA Amplification by PCR

- Much simpler and more powerful than cloning (though less accurate).
- Developed in mid-1980s, after DNA oligo synthesis could prepare 15-100 nt sequences.
- Begin with very small amounts of genomic DNA or cDNA.
- Primer oligos anneal to opposing strands with 3' ends available.
- Thermostable DNA Pol extends primers to full oligos, and the strands are melted.
- Process repeats 25-30 times, with DNA amounts doubling each cycle.
- Product is DNA between and including the primers (other regions are lost).



**Figure 9.12 Amplification of a DNA segment by PCR.**

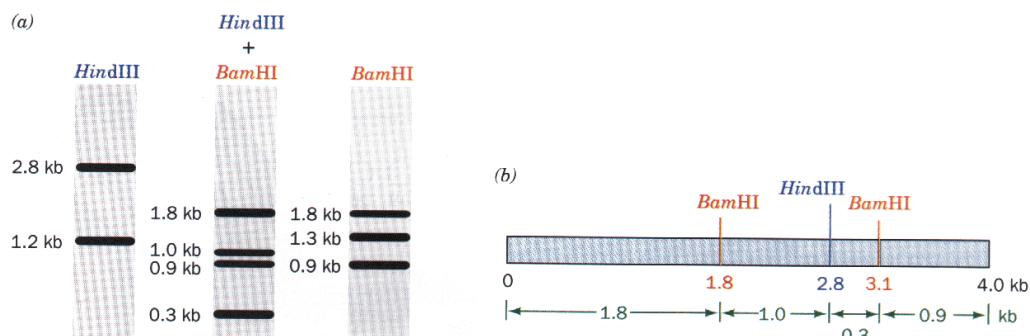
**(a)** DNA strands are separated by heating, then annealed to an excess of short synthetic DNA primers (orange) that flank the region to be amplified (dark blue); new DNA is synthesized by DNA polymerase. The three steps are repeated for 25 or 30 cycles. The thermostable *Taq* DNA Polymerase is not denatured by the heating steps. **(b)** DNA amplified by PCR can be cloned. The primers can include noncomplementary ends that have a site for cleavage by a restriction endonuclease. Although these parts of primers do not anneal to target DNA, PCR incorporates them into the amplified DNA. Cleavage of the amplified fragments at these sites creates sticky ends, used in ligation of the amplified DNA to a cloning vector.

## H. Molecular Cloning Details

- Alkaline phosphatase (e.g. calf-intestinal phosphatase, CIP) removes 5'-phosphates to prevent self-ligation. Otherwise, you'll transform bacteria without the insert.
- Transformation by heat shock (usu  $\text{Ca}^{2+}$  competent cells) or electroporation.
- Cloning vectors may be designed to include:
  - Cloning vectors contain *ori* to allow stable replication.
  - Antibiotic resistance,  $\beta$ -gal, or other screening properties.
    - $\beta$ -gal breaks down X-gal into blue product.
  - Polylinker site (several restriction enzyme targets in a row) for insertions.
  - Promoter sequence to drive expression of genes cloned into the plasmid.

## I. Restriction Mapping

- Allows for physical map (by nt), instead of genetic one (by recombination frequency).
- Location of restriction enzyme target can be found by running digested pieces on a gel and looking at their sizes.
- Ambiguities can be resolved by multiple digestion.



## J. Southern Blot Hybridization

- Multiple fragments, esp. from chromosomal DNA, can appear as a smear of DNA.
- Cut DNA, run on a gel, transfer to nitrocellulose, denature (into ssDNA) and cross-link to the membrane by UV.
- Probe with complementary labeled ssDNA ( $^{32}\text{P}$ ).

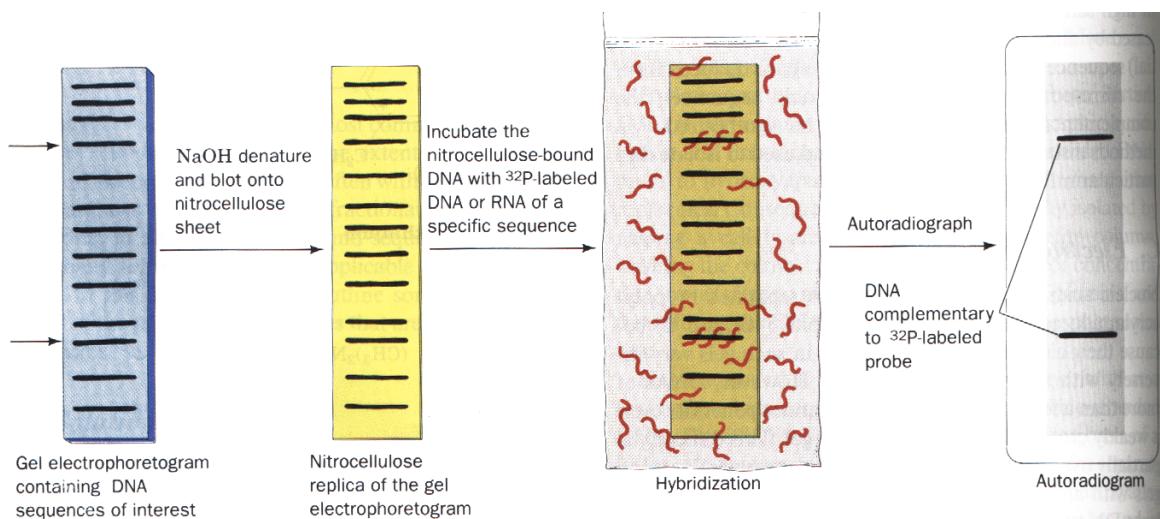


FIGURE 28-30. The detection of DNAs containing specific base sequences by the Southern transfer technique.

### K. Restriction Fragment Length Polymorphism (RFLP)

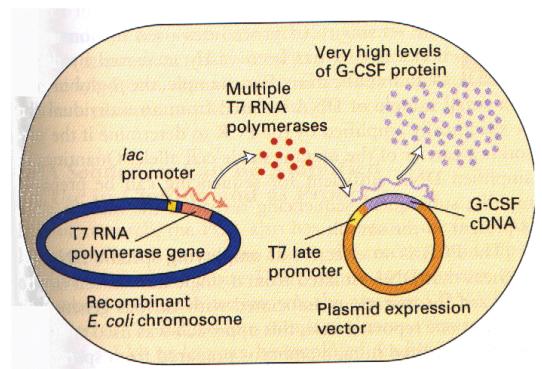
- Cut with enzyme and run fragments on a gel. Different target sites lead to different bands.
- RFLPs of highly linked genes are likely inherited together and can be used as diagnostic tools. (ideally <5% recombination frequency)
- PCR can amplify genomic DNA for RFLP

### L. Random aside on Huntington's

- CAG repeat expansion results in lengthened poly-Gln sequence → misfolding → amyloid fibers.

### M. Gene Overexpression

- Direct Expression: clone the gene into a vector with promoter and all. Usually by replacement.
- T7 RNAP: clone gene of interest after T7 late promoter and dual-express T7 RNAP (which may be in a different plasmid/controlled by another inducer – IPTG).

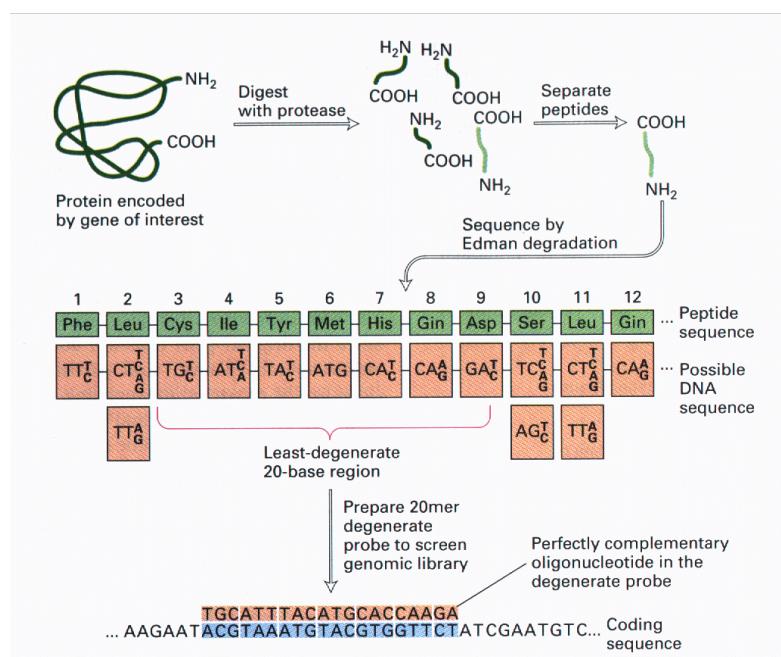


### N. Fusion Proteins

- Add a gene coding for an epitope tag directly adjacent to your gene of interest.
- This may be GFP, protein A, (His)<sub>6</sub>, B-gal, Glutathione-S-transferase (GST), Maltose-binding protein, Chitin binding domain, etc.
- IP may generate multiple bands (interacting proteins), which are identified by mass spec.

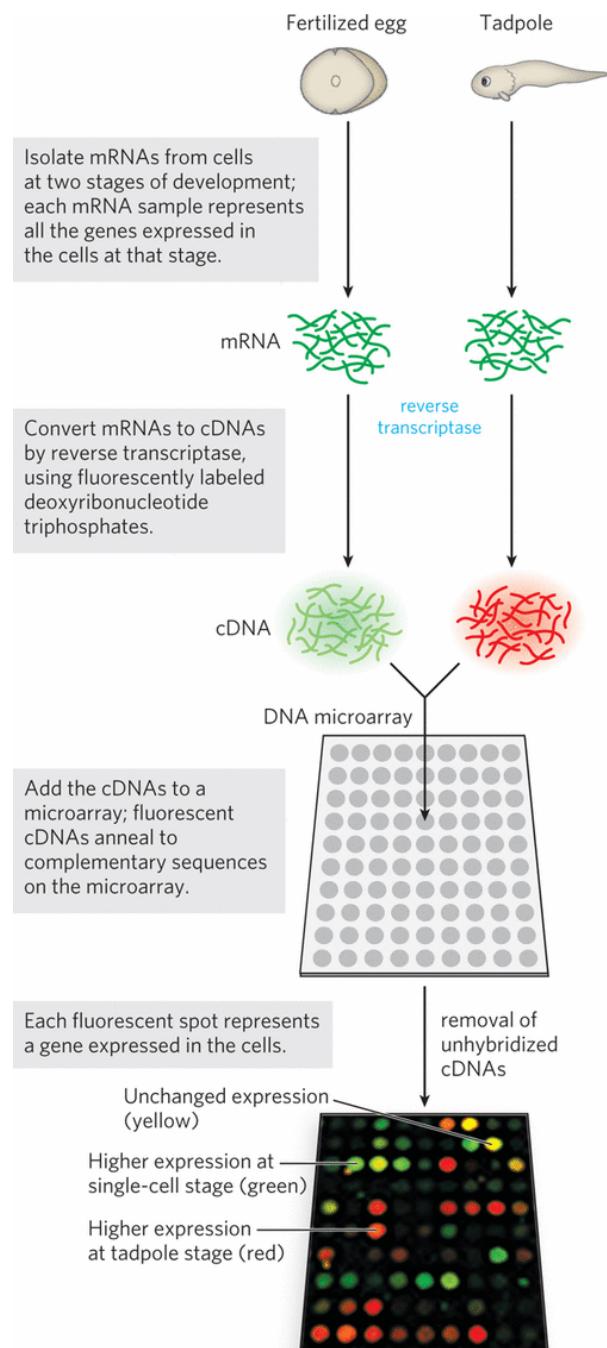
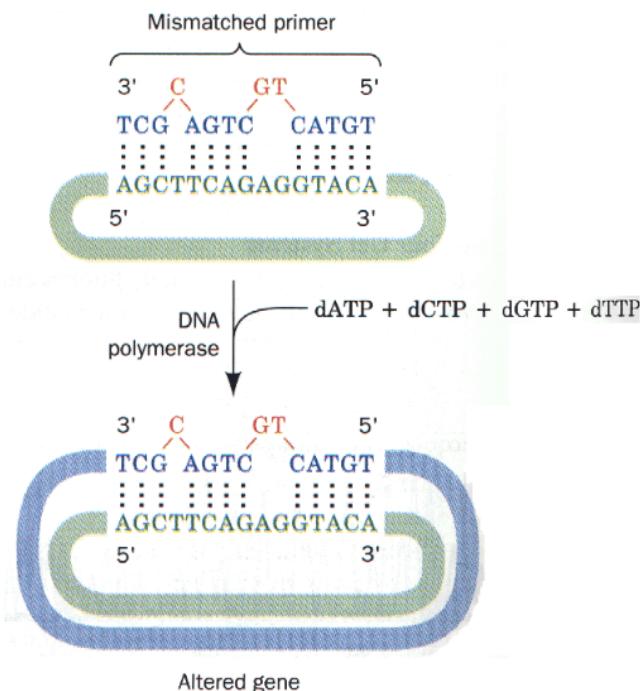
### O. Oligonucleotides

- Degenerate oligos (nonspecific)
  - Collect protein, digest with protease, sequence peptides by Edman degradation.
  - Construct a plausible DNA sequence and use the least degenerate region to prepare 20-mer degenerate probes to screen the genome.



b. Site-Directed Mutagenesis

- i. Synthesize oligo with desired mutations.
- ii. Hybridize to DNA encoding original gene (acts as a primer).
- iii. Extend by DNA Pol I to generate the mutant gene.
- iv. Insert into host to yield mutant DNA.
- v. More information: how do you separate/amplify it???

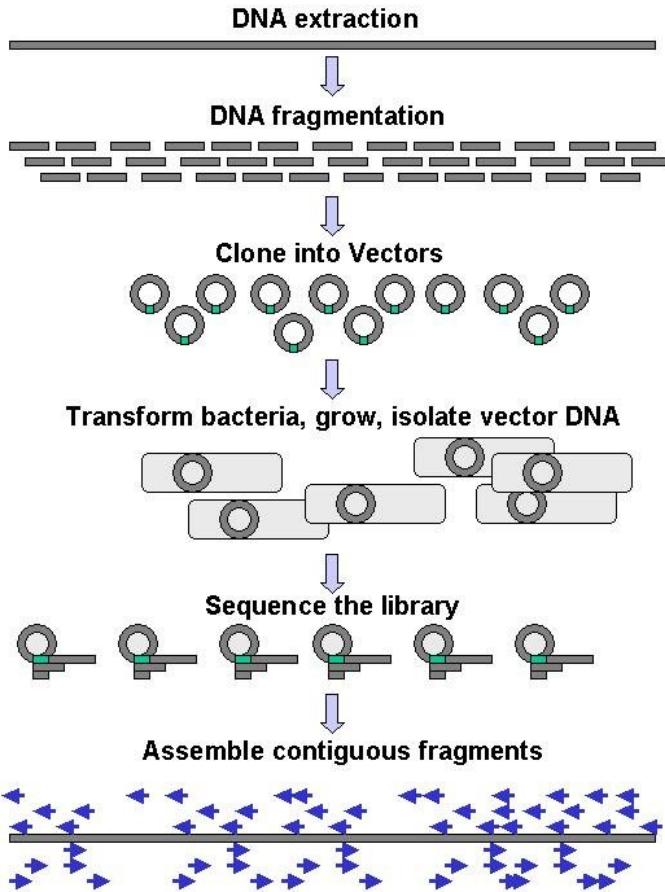


P. Microarrays and Gene Expression

- a. DNA chip has known sequences at specific positions on a grid.
- b. RNA → cDNA, then tagged with chromophore (green/red).
- c. Mix and anneal labeled cDNA to the chip.

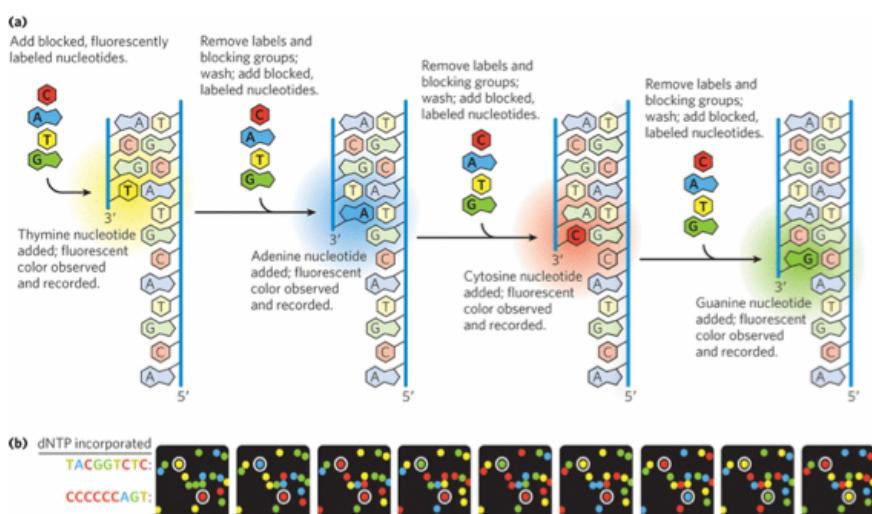
**Q. Human Genome Project**

- NIH: government and labs
- Celera: collection of automated DNA sequencers.
- Sequencing of ordered arrangement of cloned segments
- Shotgun sequencing and reassembly of overall sequence by computer alignment.



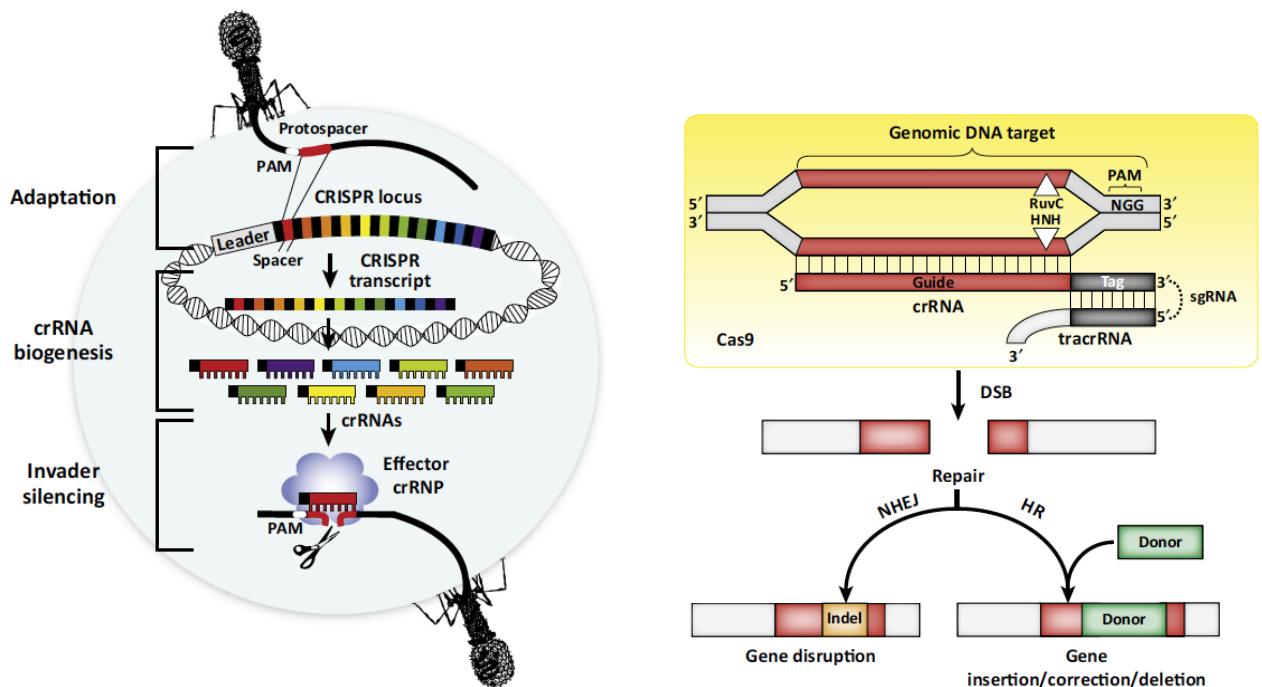
**R. Next Generation Sequencing (AKA Massively Parallel Sequencing, Deep Sequencing)**

- RNA-Seq: convert RNA to DNA, then use NGS.
- Whole genomes can be sequenced and aligned to a scaffold.



## S. Genome Editing with CRISPR-Cas

- a. Natural Process to silence viruses and foreign genetic elements
  - i. Invader DNA is integrated into the host CRISPR locus near the leader.
  - ii. The CRISPR locus contains copies of a short direct repeat sequence (black) that separates invader-derived sequences.
  - iii. Cas genes adjacent to these loci (not shown).
  - iv. PAM (photospacer adjacent motifs) are next to photospacers (viral sequences for CRISPR integration).
  - v. CRISPR locus transcripts processed to multiple crRNAs.
  - vi. crRNAs associate with 1 of 6 Cas proteins to form crRNA-Cas effector complexes.
  - vii. crRNP recognizes foreign genetic elements and cleaves in the hybrid region.
- b. RNA-programmable genome editing by Type II Cas9 CRISPR-Cas systems
  - i. Cas9 endonuclease has two nucleases that cleave dsDNA to make blunt ends.
  - ii. crRNA contains a guide region that directs cleavage.
  - iii. Enzyme requires a short PAM adjacent to the region of homology as an auxiliary target identification signal.
  - iv. Cas9 also requires a 2<sup>nd</sup> RNA (trans-activating, or tracrRNA) that base-pairs with repeat-derived tag of the crRNA.
  - v. crRNA + tracrRNA are covalently linked into a single-guide RNA (sgRNA)
  - vi. Cas-9 induced DSB undergo repair by non-homologous end-joining (NHEJ) or homologous recombination.
  - vii. NHEJ repair results in indels (short insertions or deletions) and can be used to selectively disrupt or inactivate targets.
  - viii. When provided with homologous donor DNA, HR repair can affect sequence replacement to insert, modify, or delete sequences at the site.





# Lecture 17: Eukaryotic Genome I

## A. Special Eukaryotic Characteristics

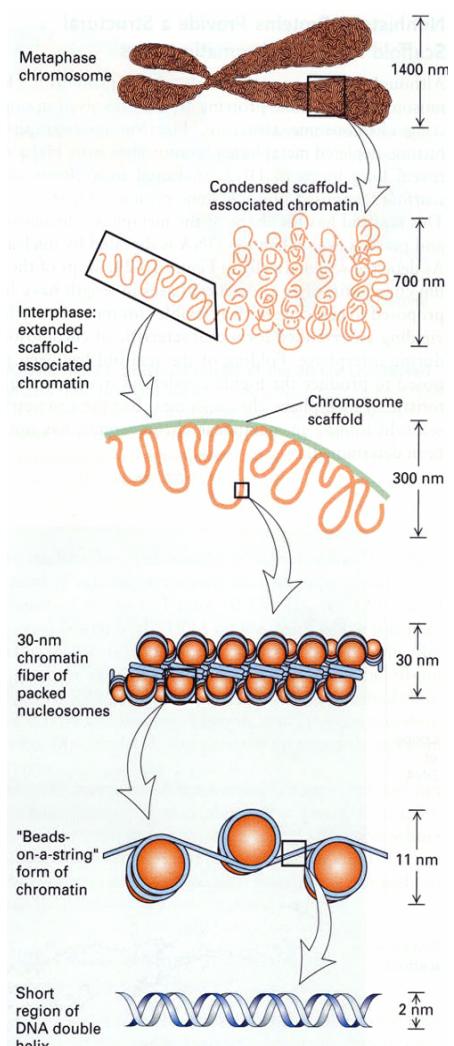
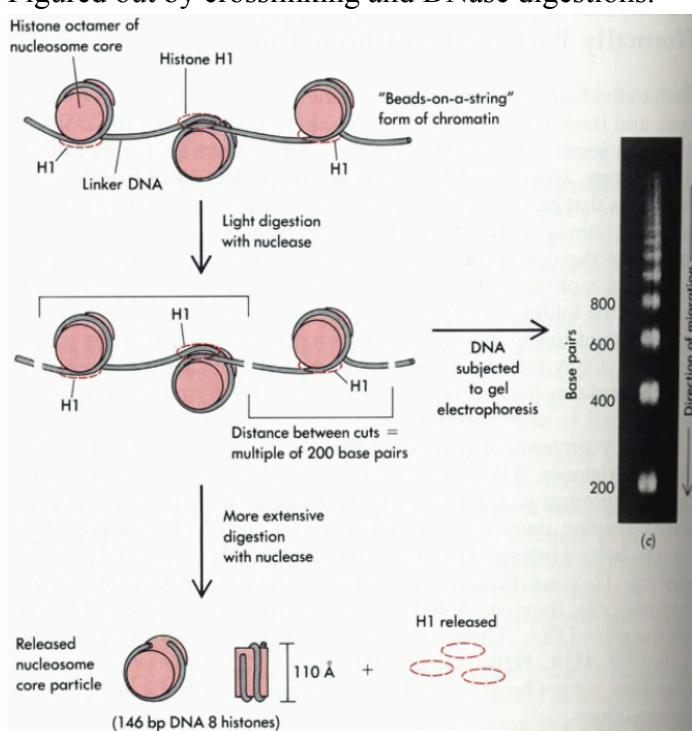
- a. Membrane-bound organelles (organelles, nucleus, mitochondria, ER, lysosomes).
- b. Fibrous systems and cytoskeleton
- c. Huge amount of DNA
  - i. C-value paradox: DNA per haploid cell is not proportional to # of genes.
- d. DNA in multiple linear chromosomes
  - i. Pulsed-field gel electrophoresis: alternatively pulses current at mirror angles.  
Can separate very large DNA molecules, which take longer to change direction.
  - ii. Chromatin: G1, S, and G2.  
Chromosomes: M.

## B. Nucleosome and Higher Order Chromatin Structure

- a. 46 chromosomes with ~3 billion bp = 2 meters.
- b. Nucleosomes: DNA wraps around histones (100 Å).
- c. Nucleofilament: helical array of nucleosomes (300 Å).  
Also: 30-nm chromatin-fiber.
- d. Metaphase chromosome: filaments arrayed on protein scaffold. TopoII is a major component of the scaffold.
- e. SMC (structural maintenance of chromosomes) proteins: cohesion and condensing.

## C. Nucleosomes are DNA-protein beads on a string

- a. Protein octamer with 2x (H2A, H2B, H3, and H4).
- b. An additional H1 organizes ~200 bp.
- c. Figured out by crosslinking and DNase digestions.



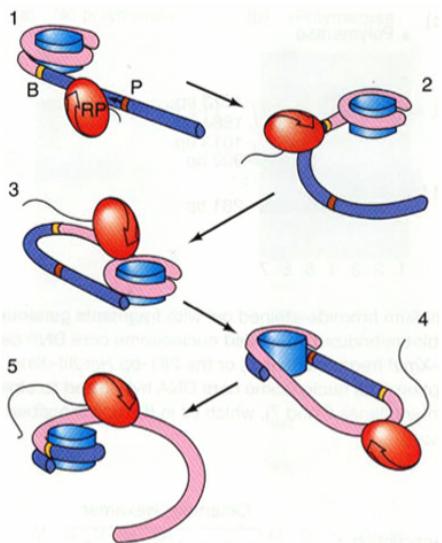
- d. Histone proteins are highly conserved:
  - i. Lys/Arg rich exterior: Basicity (+) binds DNA (-).
  - ii. Hydrophobic core for protein folding.
  - iii. DNA wraps around histones, -1 superhelical twist per nucleosome.
  - iv. 146 bp DNA wrapped around the core particle and 20-90 bp of linker DNA (between nucleosomes).

D. Nucleosome structure: solved by crystallography.

- a. H2A-H2B and H3-H4 heterodimers
- b. N-terminal histone tails are sites for modification.  
Normally wrapped in minor groove; unwrapped by acetylation.

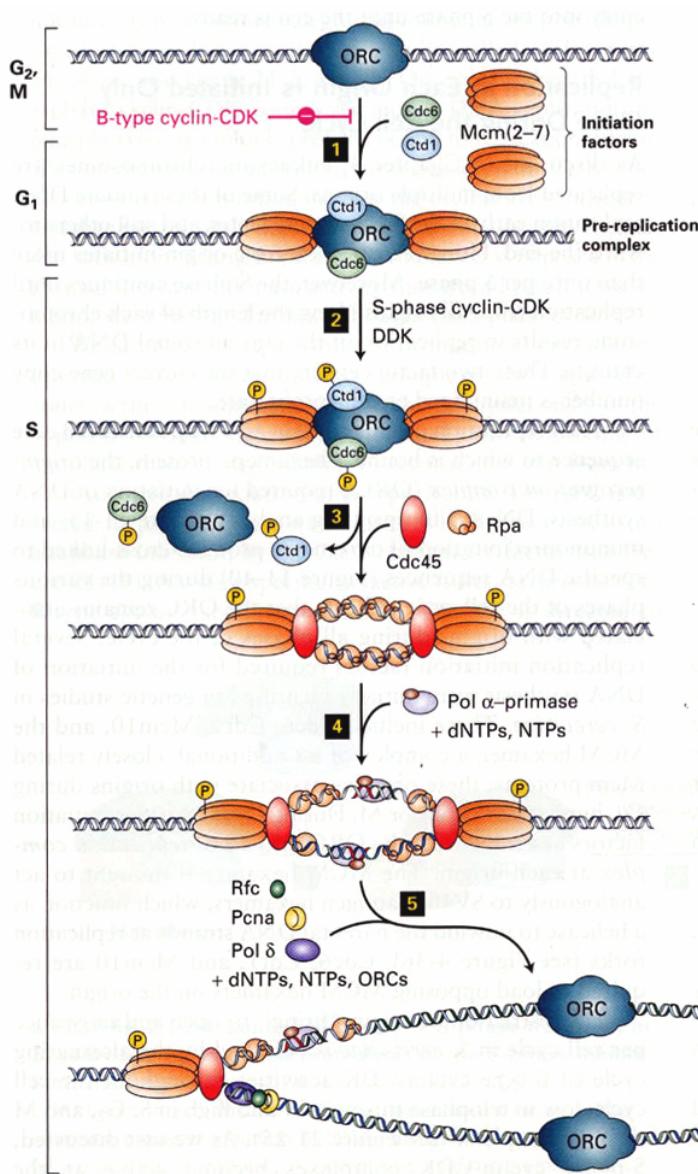
E. Nucleosomes and transcription

- a. 300 Å nucleofilament: H1 is on the inside, chromatin is inactive.
- b. Active chromatin is more sensitive to nucleases because it's less tightly wound.
- c. Hypersensitive sites near the beginning of active genes.
- d. Spooling model for nucleosome repositioning.



F. How are the giant eukaryotic DNA replicated?

- a. Multiple origins of bidirectional replication.
- b. Several different polymerases (for leading v. lagging).
- c. Controlled expression of CDKs helps coordinate replication so that each segment of DNA is replicated only once per cycle.



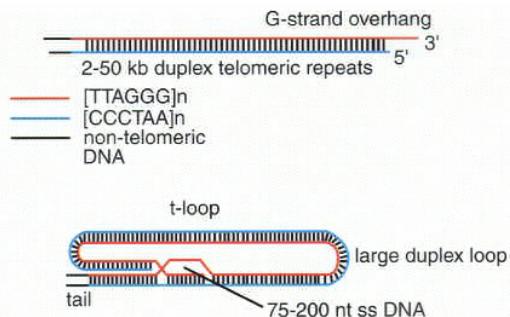
**► FIGURE 21-26 Model for assembly of the pre-replication complex and its regulation by S-phase cyclin-CDK complexes in *S. cerevisiae*.** Step 1: During early G<sub>1</sub>, unphosphorylated replication initiation factors assemble on an origin-recognition complex (ORC) bound to a replication origin to generate a pre-replication complex. Step 2: In the S phase, S-phase cyclin-CDK complexes and DDK phosphorylate components of the pre-replication complex. Step 3: This leads to binding of Cdc45, activation of the Mcm helicases, which unwind the parental DNA strands, and release of the phosphorylated Cdc6 and Ctd1 initiation factors. RPA binds to the resulting single-stranded DNA. Step 4: A complex of DNA polymerase  $\alpha$  (Pol  $\alpha$ ) and primase initiates the synthesis of daughter strands. Step 5: DNA polymerase  $\delta$  plus its accessory factors PCNA and Rfc elongate daughter strands initiated by Pol  $\alpha$ -primase. ORC binds to the origin sequence in the daughter double-stranded DNA, but the phosphorylated initiation factors cannot assemble a pre-replication complex on it. B-type cyclin-CDK complexes maintain the initiation factors in a phosphorylated state throughout the remainder of S, G<sub>2</sub>, and early anaphase (top). These factors cannot assemble into new pre-replication complexes until the B-type cyclins are degraded following their polyubiquitylation by the APC in late anaphase. Recent results indicate that additional proteins not shown function at step 3. While the phosphorylation of several pre-replication complex components is represented, all the critical components whose phosphorylation by an S phase cyclin-CDK or by DDK is required for initiation have not been determined.

### G. How are the ends of linear chromosomes replicated?

- The 3'-primer is removed following replication, leaving 3' overhangs.
- Telomeres are G-rich repetitive sequences at ends. These vary across species.
- Tetrahymena telomeres placed into yeast were extended with the yeast terminal sequence.
- Tetrahymena telomerase is an RNP!
  - Telomerase extends 3'-ends without a template.
  - Reverse transcribes built-in RNA template (1.5 telomeric repeats).

## H. Telomeric Structures and Binding Proteins

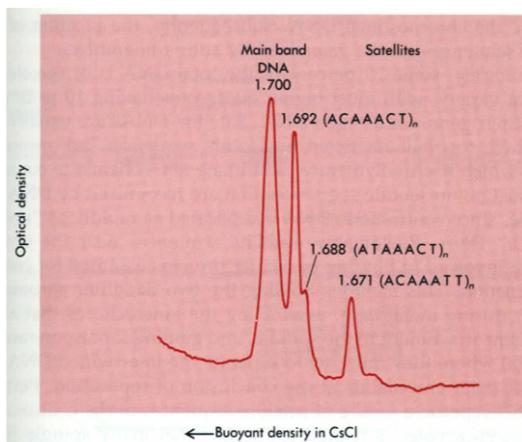
- dsDNA ends are recombinogenic and not stably inherited.
- All euk chromosomes have 3'-terminal overhangs with GT rich telomeric repeat.
- ss-polyG terminal sequences prevent end-joining by:
  - G-structures (quartets, tetraplexes) to knot the ends.
  - Recruit proteins to cap and prevent end-joining.
- T-loops: ss-G-rich stand invades telomeric duplex to cap the end.



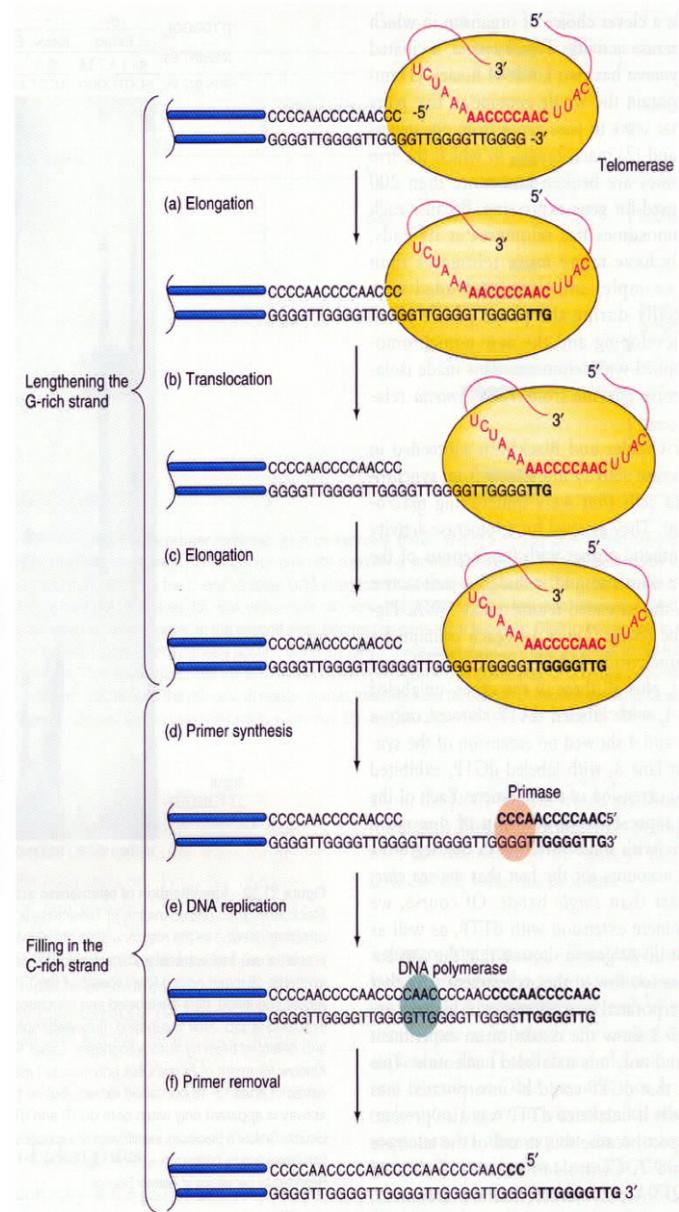
I. Telomerase activity is off in most normal cells, but appears in cancer cells.

## J. Repeated Sequences at Centromeres

- Centromeres: where spindle fibers attach to metaphase chromosomes.
- Repeated sequences from centromeres appear as satellite bands (slightly fragmented total DNAs separated in CsCl gradients).



- In situ hybridization (with labeled cRNA) located these satellite sequences at centromeres.



### K. rRNA/tRNA Genes Repeated

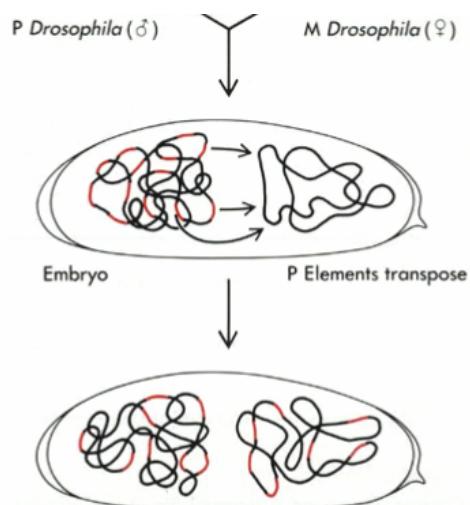
- a. rDNA repeats are clustered at several hundred copies.
- b. Protein-coding genes are single-copy:
  - i. Amplification at transcription and translation.
  - ii. Exception: histone proteins (need lots of them!).
- c. Fibroin and silkworms
  - i.  $10^5$  times more protein produced than mRNA transcribed.
  - ii. Isolated fibroin mRNA by synthesizing an oligo primer based on repeating AAs.
  - iii. Degenerate oligos and screen genome; hybridize to Northern blot.

### L. Types and Origin of Genes in the Human Proteome

- a. Humans have only 23,000 genes, comparable to yeast and worms.
- b. Greater complexity from alternative splicing.
- c. Most human genes are not unique to humans.
- d. Many are from other sources; some from bacteria via horizontal gene transfer.

### M. Transposons Inhabit Euk Genomes

- a. P-element is a cut-and-paste Drosophila transposon leading to “hybrid dysgenesis” (hybrids messed up because P elements disrupt vital genes).
- b. Expressed only in germline cells because of transposase requires tissue-specific splicing.
- c. P-element transformation:
  - i. Gene of interest inserted between ends of P element
  - ii. Inject P element into M (non-P) eggs or embryos, along with transposase.
  - iii. Insertions at multiple sites in Drosophila chromosomes.



### N. Are repetitive sequence elements selfish DNA?

- a. Throughout time, wild drosophila increasingly possess P elements.

### O. Human genome has lots of mobile elements.

- a. LINES (long interspersed repeat elements) include L1 retrotransposon.
- b. SINES (short interspersed repeat elements) include Alu sequences.

### P. Y chromosome is unusually young.

- a. Constant rate of mutation within repetitive elements (because no selection pressure).
- b. Molecular clock: comparing sequences from 2 sources to that of a common ancestor can give age estimates.
- c. Y chromosome is largely repeated elements; has a higher mutation rate than the X chromosome; looks unusually young.

### Q. Alu and L1 sequences can be mutagenic:

- a. Alu-alu recombination triggers chromosomal rearrangement.
- b. De novo insertions into genes.



# Lecture 18: Eukaryotic Genome II

## A. Eukaryotic mRNA post-transcriptional modifications:

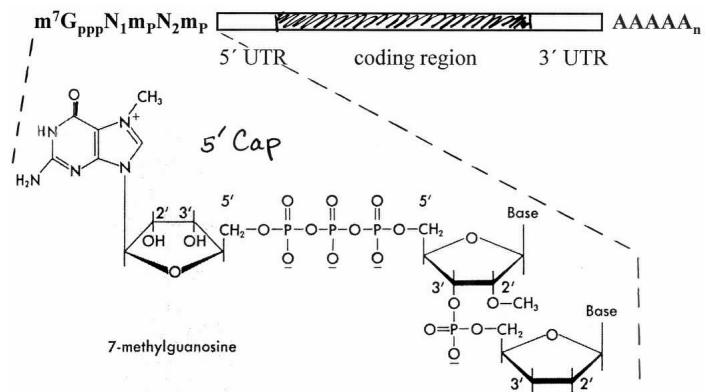
- 5' and 3' UTRs
- 5'-cap: 7meG (prevents degradation)
- 3'-polyA tails: ~200 nt added by polyA polymerase (PAP).
  - Isolate mRNA with oligo-dT.
  - Template cDNA synthesis.

## B. Heterogeneous nuclear (hn)RNA/pre-mRNA

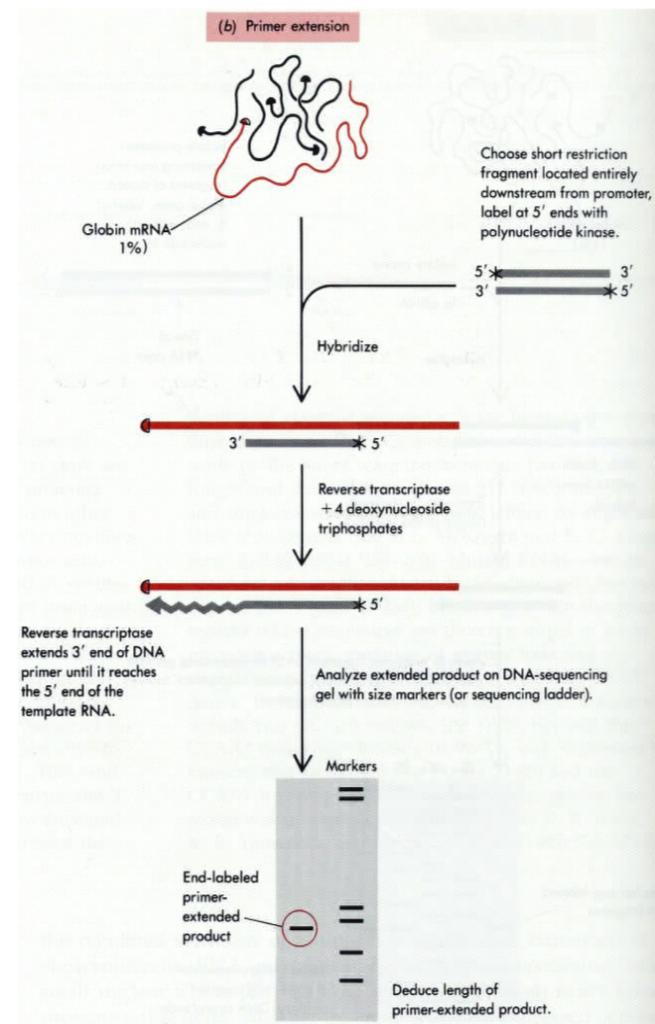
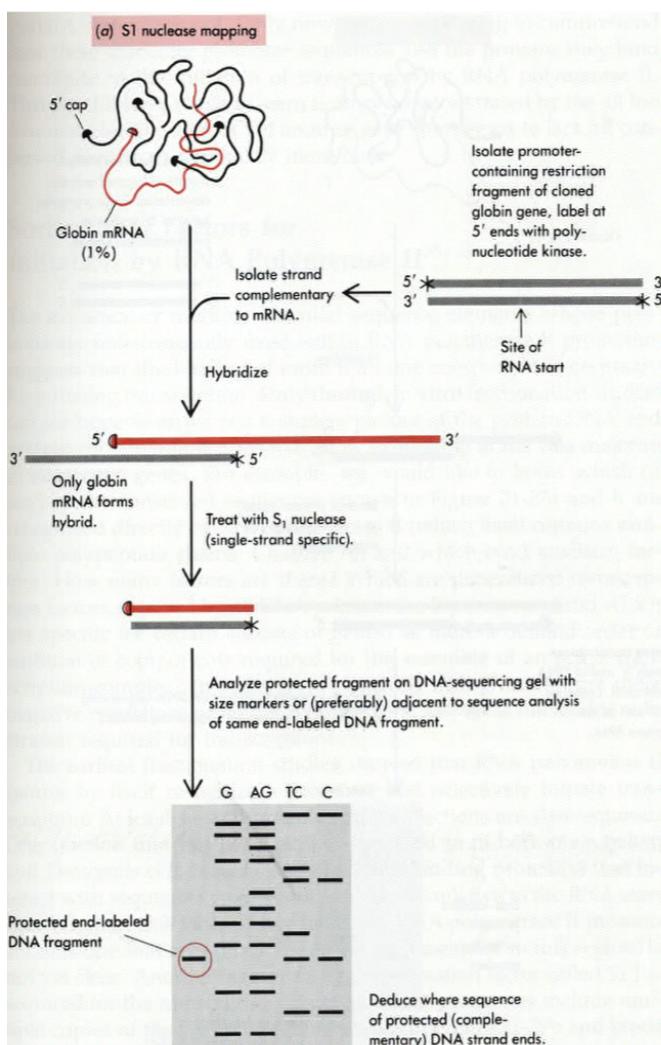
- 10x the size of mRNA.
- Have 5'-caps and polyA tails.
- Short-lived: half life of 3 min.
- 10% labeled hnRNA (exons) appears in cytoplasmic mRNA.

## C. Mapping Ends and Intron-Exon Boundaries in mRNA

- S1 nuclease mapping



## Primer Extension



b. S1 nuclease mapping

- Isolate 5'-labeled restriction fragment that contains the mRNA 5'-start.
- Hybridize to mRNA, digest with S1.
- Can also locate mRNA 3'-end by labeling the 3'-end of the probe instead.
- Requires detailed knowledge of gene structure and sequence data.

c. Primer extension

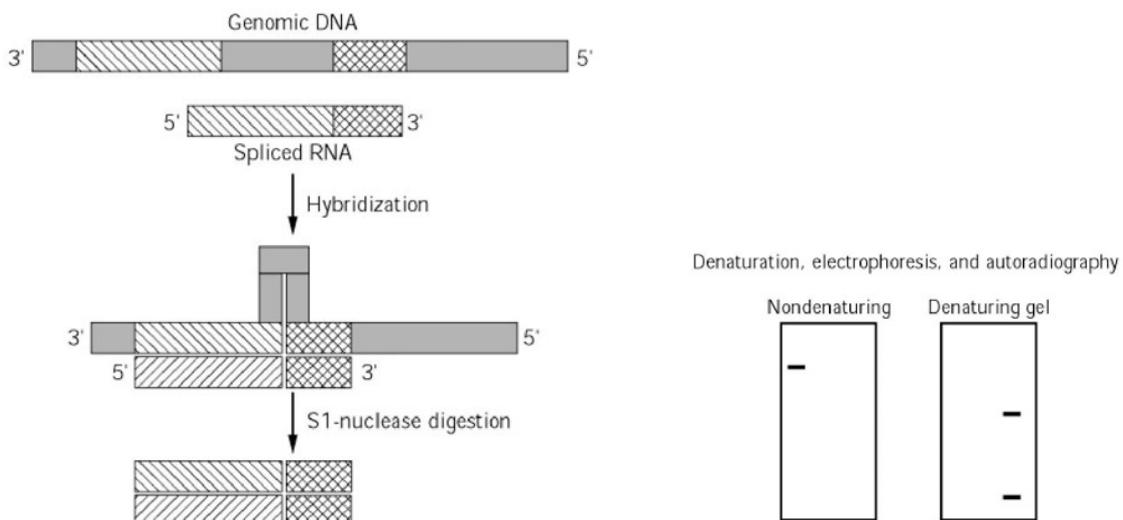
- Isolate 5'-labeled restriction fragment downstream of start site.
- Hybridize to mRNA, extend with RT.

d. Final gel

- Length of ds segment = distance of mRNA start site from labeled 5'-end.
  - Use size markers or sequencing ladder (higher resolution).
- e. RNase mapping is a variant of S1 mapping where a labeled RNA is used (instead of DNA).

D. S1 nuclease mapping of exons

- Hybridize mRNA to genomic DNA; the introns loop out and are digested by S1.
- Denature/melt the strands and run the DNA exons.

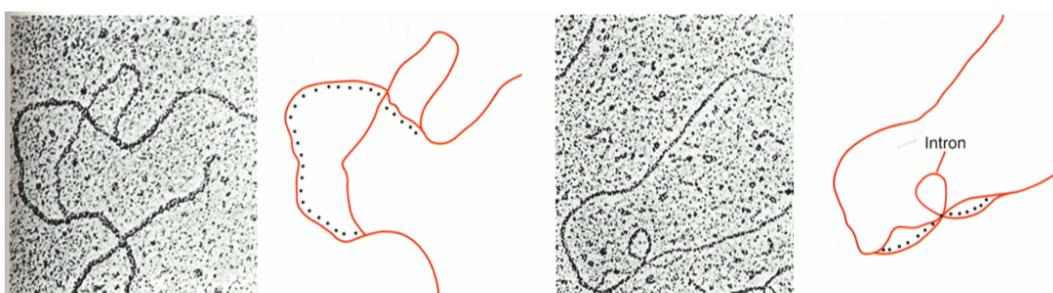


E. R-Loop Mapping Detects Introns

- Cloned gene DNA is incubated with mRNA in formamide (denaturant) at DNA T<sub>m</sub>.
- mRNA displaces coding strand, forming an R loop.
- DNA-RNA is more stable because of the A-form has better base stacking.

F. Proving that introns are transcribed into RNA

- R loops of spliced mRNA: introns are forced to loop out.
- R loops of hnRNA: introns bp naturally.



### G. Intron-Exon Structure of Eukaryotic Genes

- a. R loops can give an idea of intron size and location.
- b. Northern blotting to find pre-mRNA:
  - i. Isolate RNA-polyA, run on gel, transfer to blot, probe with labeled cDNA.
  - ii. Top band is primary transcript.
  - iii. Nuclear RNA bands are splicing results.

### H. Exon and Intron Sizes in the Genome

- a. Most organisms have small exons (~150 nt).
- b. Humans have smaller exons and much bigger introns.

### I. Things of interest

- a. Exons sometimes correspond to protein-folding domains.
- b. **Intron positions are conserved- shows that parsing the gene according to protein structure is selected for.**
- c. Ratio of non-coding to coding DNA rises with developmental complexity.

### J. When did Introns originate?

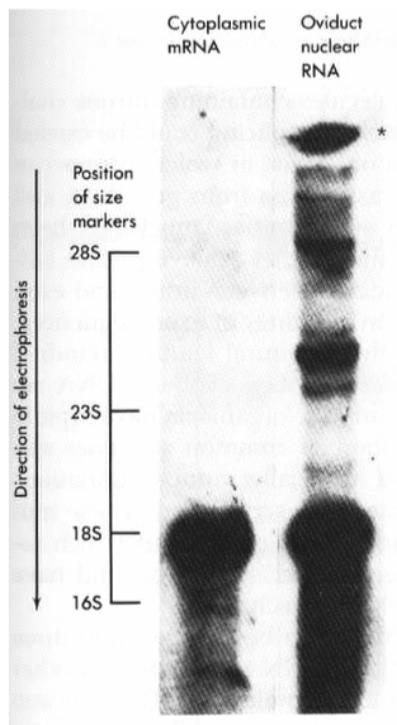
- a. Theory 1: early genomes had introns, and rapidly replicating organisms got rid of them (++generation times).
- b. Theory 2: introns have jumped into genes later.
- c. **Evidence for Early Introns**
  - i. Exon shuffling: same exon sometimes appears in different proteins because of recombination between exons (within introns).
  - ii. Exon sliding: new splice junctions can add new “exon” material.

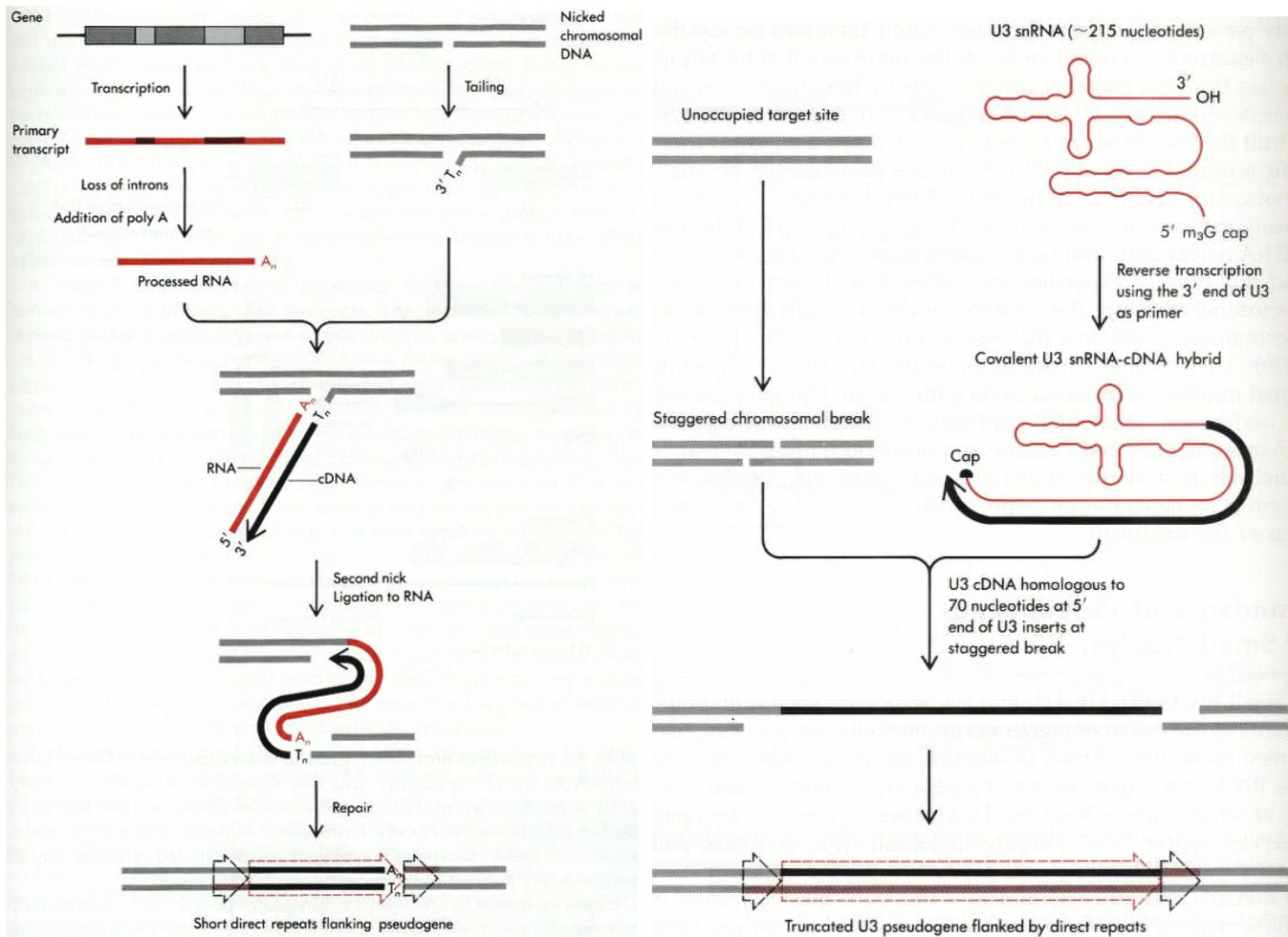
### K. Creation of Pseudogenes

- a. Processed pseudogene
  - i. Nicked chromosomal DNA is tailed (with T).
  - ii. The RNA polyA pairs at the nick and reverse transcribes cDNA.
  - iii. Second nick allows incorporation (RNA ligated to DNA, gaps repaired).
  - iv. Staggered nicks create short direct repeats.
  - v. Results in spliced replica of protein-coding gene; often has polyA sequence.
- b. RNA pseudogene
  - i. The RNA (e.g. U3 snRNA) is reverse transcribed using the 3'-end as a primer.
  - ii. The cDNA inserts at a staggered double nick, and gaps repaired with direct repeats.
  - iii. Only makes a partial replica (5'-end of RNA).
- c. Processed pseudogenes are highly abundant in mammals (10X more than real genes!!!)
- d. No promoters on processed pseudogenes, but has polyA tail. Alu has promoter (RPol III). RNA pseudogenes have neither promoters nor polyA.

### L. Trinucleotide repeat expansion and disease

- a. DNA Pol slips
  - i. Copies too much: new strand loop, increases repeats.
  - ii. Copies too little: old strand loop, decreases repeats.
- b. May lead to messed up splicing or protein structure.



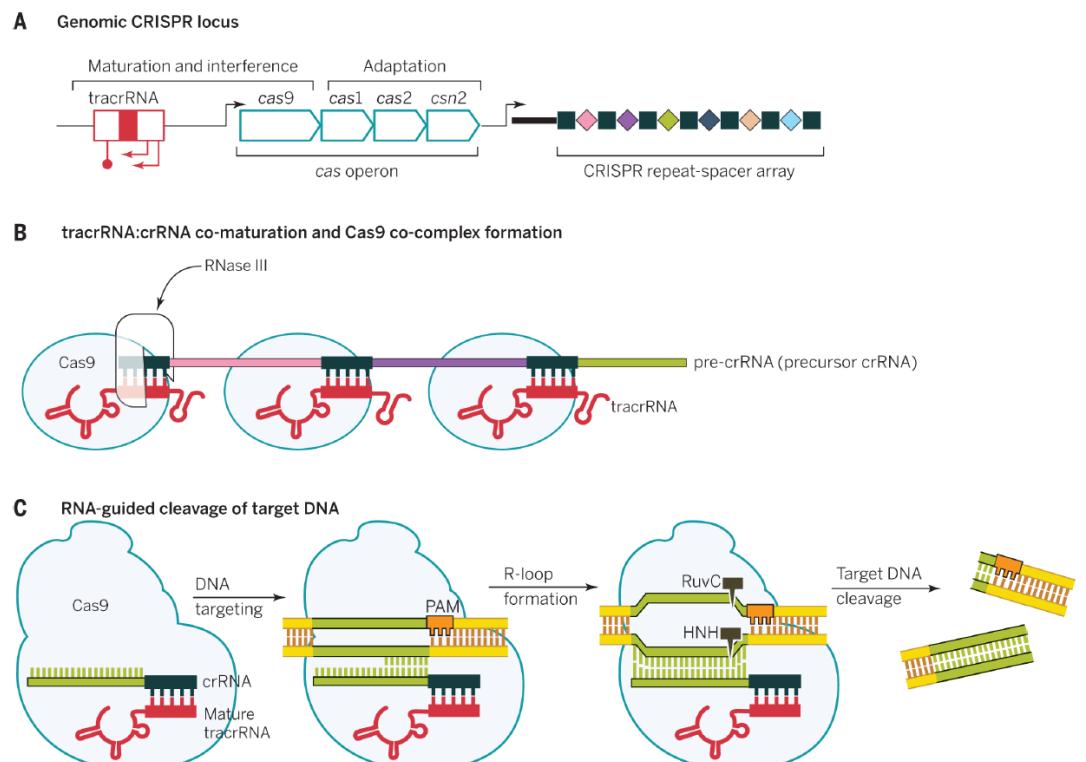


#### M. CRISPR: clustered regularly interspaced palindromic repeat

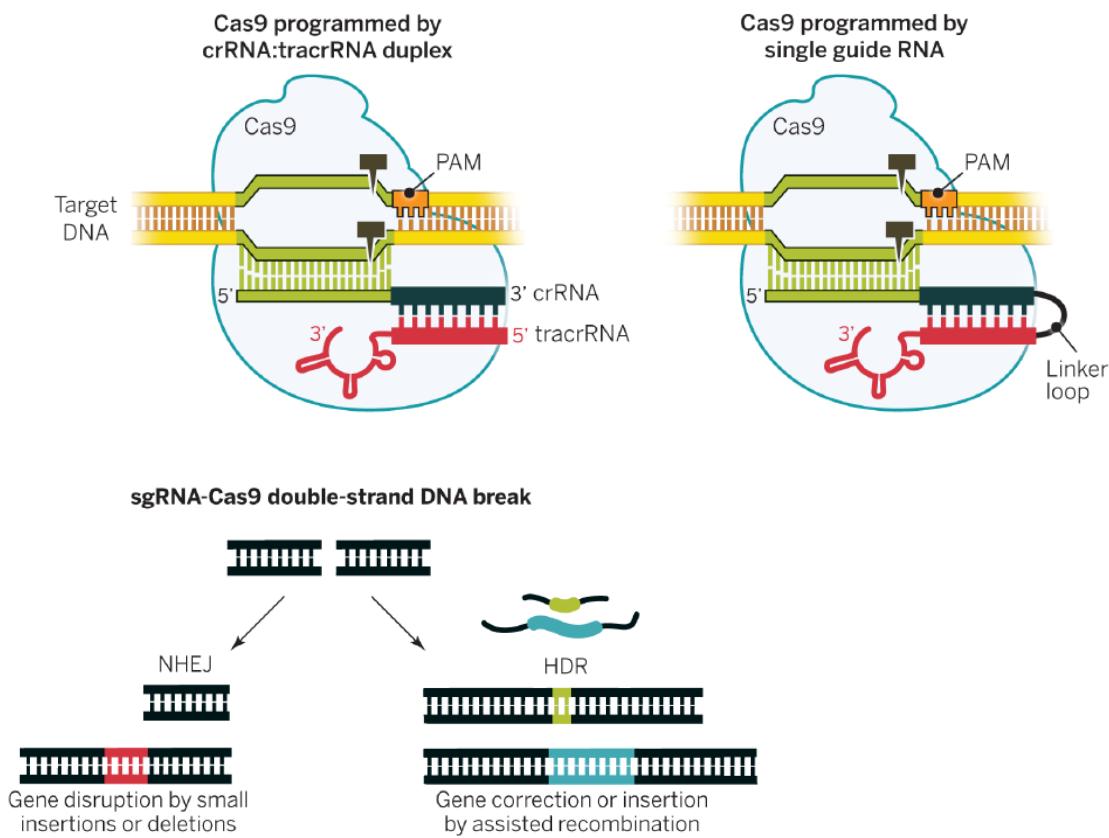
- CRISPR 30-40 bp “spacers” are derived from DNA of previous invaders.
- Intervening 30-40bp “repeats” are self.
- Resulting CRISPR (cr)RNAs guide Cas proteins (DNA endonucleases) to their targets.
- Many CRISPR/Cas systems - most are RNA guided and DNA-cleaving.  
The crRNA hybridizes with DNA to form an R-loop.
- The PAM site (usu. NGG) is required for targeting by Cas9; absent in self DNA!

#### N. CRISPR for therapeutics

- Designed single guide RNA (sgRNA) find identify complementary dsDNA next to a PAM.
- sgRNA contains RNA to recruit Cas9.
- Cleavage produces a break; which is repaired in 2 ways:
  - Homology directed repair (HDR): replaces a target sequence to revert unwanted mutations, generate new alleles, insert useful domains, and introduce transgenes.
  - Non-homologous end-joining (NHEJ): leaves small insertions/deletions to inactivate the gene.

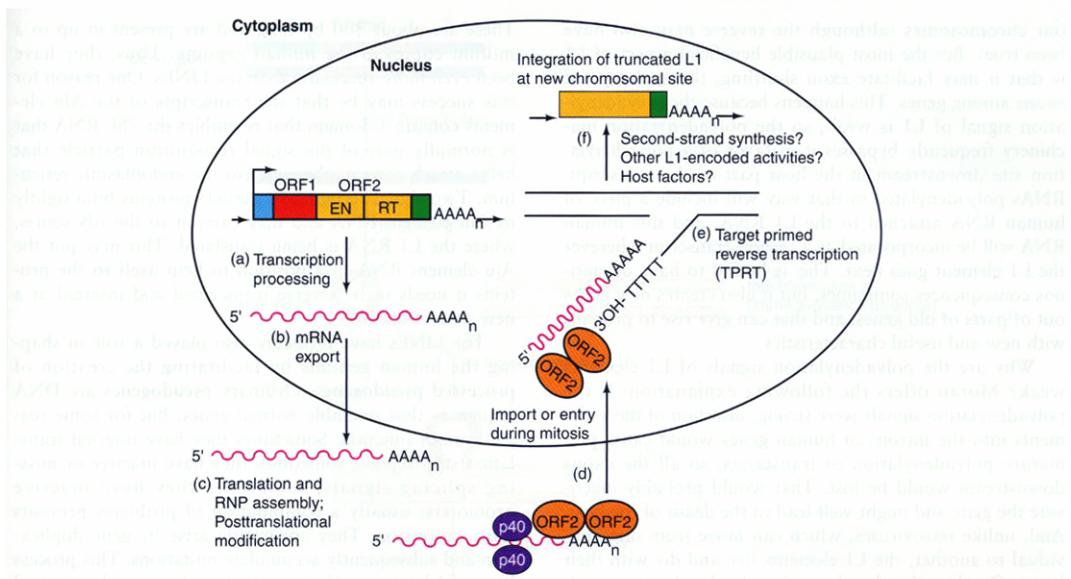


**Fig. 2. Biology of the type II-A CRISPR-Cas system.** The type II-A system from *S. pyogenes* is shown as an example. (A) The cas gene operon with tracrRNA and the CRISPR array. (B) The natural pathway of antiviral defense involves association of Cas9 with the antirepeat-repeat RNA (tracrRNA: crRNA) duplexes, RNA co-processing by ribonuclease III, further trimming, R-loop formation, and target DNA cleavage. (C) Details of the natural DNA cleavage with the duplex tracrRNA:crRNA.



## O. Too much DNA in Eukaryotes SUMMARY

- Human genome is only 1.5% coding!
- Introns make genes 1000-fold larger.
- Repeated gene families exist for tRNAs, rRNAs, snRNAs, and histones.
- Many pseudogenes resemble real genes, but do not encode gene products.
- Repeated sequence elements at centromeres and telomeres build structures critical for chromosome maintenance.
- Selfish DNA is half the human genome!
  - Includes transposons and retrotransposons (Alu and L1 sequences)
  - 100 L1s in the human genome.
  - Codes for RT and endonuclease. Nicks and initiates target-primed reverse transcription.



**Figure 23.40 A model for L1 transposition.** (a) The L1 element is transcribed and processed. (b) The transcript is exported from the nucleus. (c) The mRNA is translated to yield the ORF1 product (p40) and the ORF2 product, with endonuclease and reverse transcriptase activities. These proteins associate with the mRNA. (d) The ribonucleoprotein reenters the nucleus. (e) The endonuclease nicks the target DNA (anywhere in the genome), and the reverse transcriptase uses the new DNA 3'-end to prime synthesis of the reverse transcript. (f) In a series of unspecified steps, the second L1 strand is made and the whole element is ligated into the target DNA. (Source: Reprinted from Kazazian and Moran, "The impact of L1 retrotransposons on the human genome," *Nature Genetics* 19:19–24, 1998. Copyright 1998, with permission from Nature Publishing Group.)

# Lecture 19: Eukaryotic Transcription

## A. Eukaryotic v. Prokaryotic Transcription

- a. Euk RNAP cannot transcribe on their own; multiple proteins involved.
- b. Euk have 3 RNAPs.
- c. Prok have 1 RNAP, regulated by switching sigma factors (recognize different promoters).

## B. Gene regulation in mammalian cells

- a. Regulation in processing and stability.
- b. 5-10k housekeeping genes.
- c. Differential expression of 100-1000s of specific genes.

## C. James Darnell: Nuclear run-on experiment.

- a. Nascent transcripts in isolated nuclei are elongated with pulse of labeled NTPs.
- b. Extract RNAs and hybridize to nitrocellulose membranes with fixed cDNAs.  
Wash off lone RNAs, and visualize with autoradiography.
- c. Shows how gene transcription is different across cell types.
- d. Don't forget positive (housekeeping) and negative (nonmatch) controls!

## D. Bob Roeder: discovered 3 Euk RPol

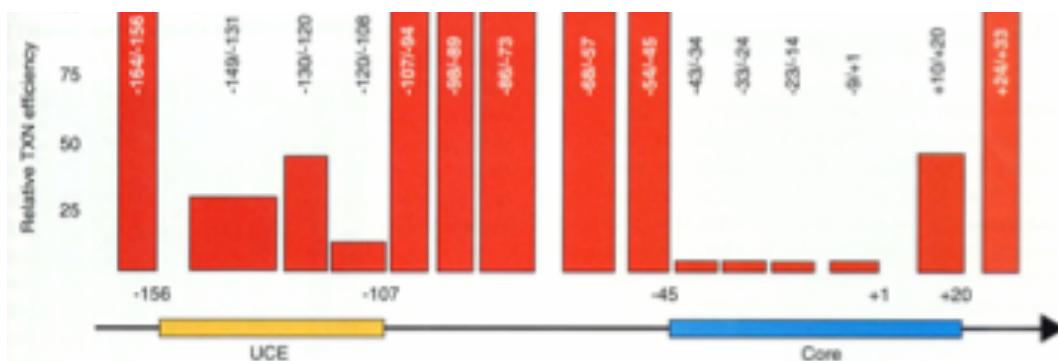
- a. Took cell extracts and separated into fractions by anion exchange chromatography.
- b. Fractions were assayed for ability to transcribe boiled DNA.
- c. RPol I: pre-rRNAs (nucleolus).
- d. RPol II: mRNA precursors and snRNA.
- e. RPol III: tRNA and 5S rRNA.

## E. RPol I-III have 12-16 subunits

- a. Largest subunits homologous to the bacterial RNAP  $\beta'$  (polymerase activity).  
In Pol II, this subunit has the CTD.
- b. 5 common subunits: others are homologous to  $\beta$ , 2x alpha, and omega.
- c. TBP: TATA-binding protein, for initiation.

## F. RPol I

- a. Nucleolus: rRNA is made, processed, and assembled.
- b. Most abundant RPol; recognizes 1 promoter and 1 terminator.

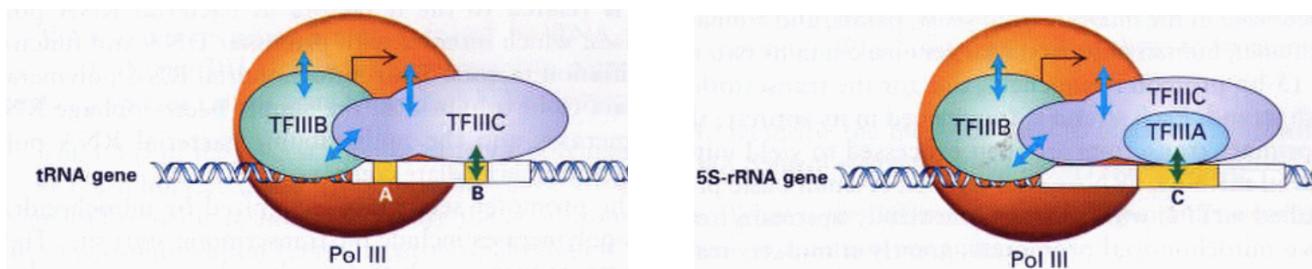


- c. Upstream control element (UCE) and core element are important for transcription.
  - i. UCE makes it optimal, but core is absolutely required.

- d. SL1 (selectivity factor 1):
  - i. Complex of 4 proteins (incl. TBP)
  - ii. Upstream binding factor (UBF): acts with SL1 to stimulate RPol I transcription.
  - iii. Strengthens RPol I-UCE binding, and opens a DNase-sensitive site in core element.
- e. Promoter elements and transcription factors are species-specific.
- f. RPol I is highly efficient; clearing the promoter is rate limiting.

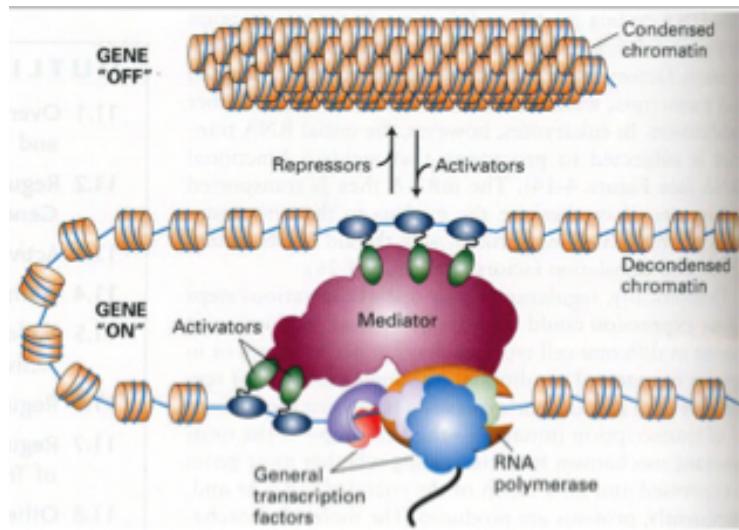
#### G. RPol III

- a. Transcribes small ncRNAs: pre-tRNAs, 5S rRNA, U6 snRNA, and Alu.
- b. Most promoters for Pol III are transcribed. Some upstream (U6 promoter), others internal.
- c. Multiprotein transcription factors involved:
  - i. TFIIIB contains TBP. Binds upstream of the start site. Absolutely required.  
Unable to bind DNA itself (like RNAP sigma).
  - ii. A/B/C boxes: internal promoter elements.
  - iii. tRNA genes: A/B boxes recruit TFIIIC recruits TFIIIB recruits RPol III.
  - iv. rRNA genes: A/C boxes recruit TFIICIA recruits TFIIIC...
- d. Termination at Pol III occurs at polyT runs of TTTT... residues.  
Transcripts must end with ...UUUU-OH-3'

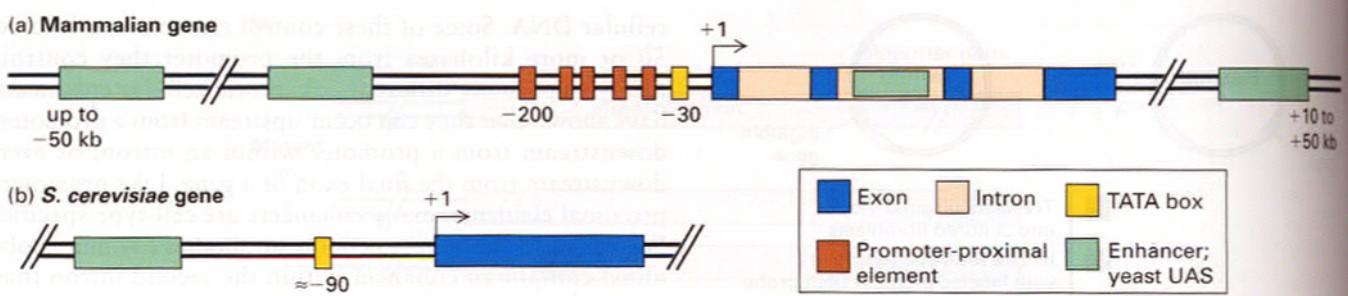


#### H. RPol II

- a. Transcribes mRNAs and snRNAs
- b. General transcription factors (basal TFs) bind at/near start site.  
Minimal set for site-specific initiation by Pol II (e.g. TFIIA/B/D/E/F/H).
- c. Transcriptional activators: bind distant sites
  - i. Modular: DNA-binding, activation, and flexible protein domains in any order.
  - ii. Grouped by DNA-binding domain: e.g. homeobox, HLH, Zn finger, etc.
- d. Control Elements
  - i. Promoter: sequences close to the start site recognized by Pol II and TFs.
  - ii. TATA box: 25-30 bp upstream of start site.
  - iii. Weak sequence conservation at start sites.
  - iv. Promoter-proximal elements:
    1. Bind transcriptional activators (like SP1).
    2. May occur downstream of start site!
  - v. Enhancer sequences:
    1. Interact with gene-specific transcriptional activators.
    2. Can be >10 kB away, within introns, upstream or downstream.
    3. Silencers are enhancers that downregulate a gene.



TFIIA	Stabilizes TBP and TAF binding
TFIIB	Stabilizes TBP binding; recruits RNAP II; influences start site selection
TFIID	Recognizes TATA box; recruits
TBP	TFIIA and TFIIB; has positive and negative regulatory functions
TAFs	
TFIIIE	An $\alpha_2\beta_2$ heterotetramer; recruits TFIIH and stimulates its helicase activity; enhances promoter melting
TFIIIF	Facilitates promoter targeting; stimulates elongation
TFIIH	Contains an ATP-dependent helicase that functions in promoter melting and clearance



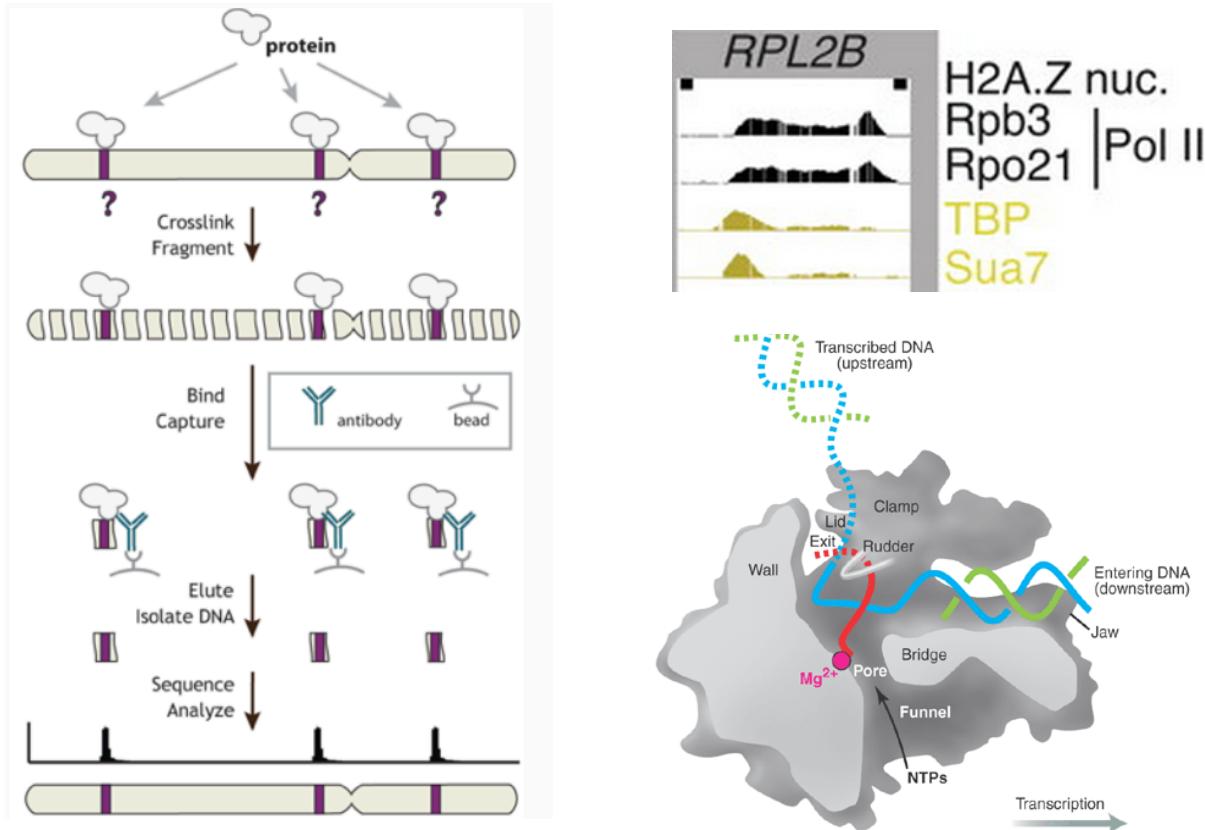
### I. Assembling RPol II Complex

- TAF: TBP-associated factor
- TFIID: contains TBP and TAFs
  - Contacts the TATA box (with TBP) and the start site (with TAFs).
  - Some promoters lack a TATA box and engage TBP via other proteins instead of direct DNA binding.
- TBP: binds TATA box
  - Bends the DNA to open the minor groove and destabilize the helix. Allows transcription from promoter on naked DNA. TFIID required *in vivo*.
  - Protein-DNA interaction via  $\beta$ -sheet.
  - Binds as monomer with pseudo-dimeric symmetry.
- Assembly Steps
  - TFIID binds TATA box with TBP.
  - ??? TFIIB binds downstream of the TATA box and stabilizes TBP. Defines direction of transcription.
  - Pol II, TFIIF, TFIIIE bind.
  - TFIIH contains ATP-dependent DNA helicase; creates transcription bubble. Also: kinase for CTD.

### J. Activation of RPol II by Basal Factors

- RPol II CTD: unstructured; multiple tandem repeats of X-Ser-X-X-Ser-X-Ser.
- CTD is phosphorylated on Ser by kinases on TFIIH.  
This stabilizes RPol II's interactions with TFs.

K. Chromatin Immunoprecipitation (ChIP) identifies the DNA-binding sites of proteins.



L. Structures of elongating yeast RPol II

- a. DNA bent 90°.
- b. Cramer and Kornberg solved yeast RPol II
- c. Trigger loop and Mg<sup>2+</sup> coordinate dNTP
- d. Bridge promotes RNA-DNA base-pairing and translocates the active site.
- e. Rudder directs nascent RNA to exit channel.

M. Enhancers

- a. 50-200 bp in length.
- b. Multiple binding sites for distinct transcriptional activators, which interact synergistically.

N. Enhanceosome Model

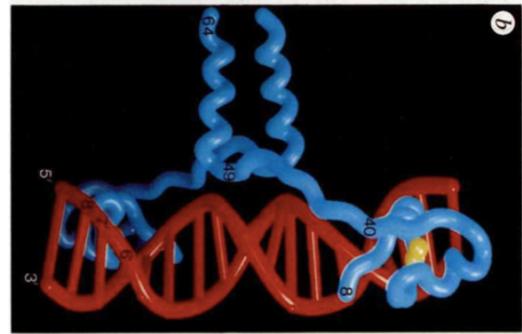
- a. Protein complex assembled at enhancers.
- b. Framework proteins stabilize bent DNA and protein-protein interactions.
- c. Modular Structure of Transcriptional Activators
  - i. DNA-binding domains: many motifs.
  - ii. Multimerization domains: where proteins join.
  - iii. Activation domains:
    1. Interact with basal transcriptional machinery.
    2. Less structured, residue redundancy.
    3. Acidic residues (Asp/Glu)  
(negative noodles, acid blobs).

O. Types of DNA-binding domains

- a. Helix-loop-helix
- b. Homeodomain (helix-turn-helix binding)
- c. Zn finger
- d. Basic zipper (bZIP)

P. GAL4 is a Zn Finger Homodimer with 3 modules

- a. Zn-binding region: binds DNA.
- b. Linker region: extends along minor groove.
- c. Dimerization element: short parallel coiled-coil.

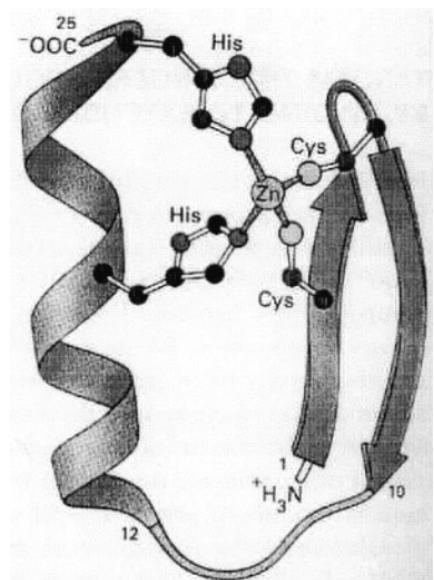


Q. Zn Finger is a protein-folding motif

- a. Zn finger motif: uses a-helical to contact and recognize DNA major groove.
- b.  $Zn^{2+}$  coordinates x2 His and x2 Cys tetrahedrally.
- c. TFIIB: 9 tandem Zn-finger repeats.  
His and Cys protrude from 2 antiparallel beta sheets and an alpha helix.
- d. Recognizes 3 bp each (mostly A/G), often repeated.  
Arg helps find G. NOT palindromic!

R. Mechanism of Transcriptional Activators

- a. Stimulate cooperative assembly of initiation complex.
- b. Histone acetyltransferase (HAT) complexes: recruited via protein-protein interactions with activation domains.
- c. Acetylation frees up nucleosome structure, facilitating promoter recognition by RPol II.



S. Histone Code Hypothesis

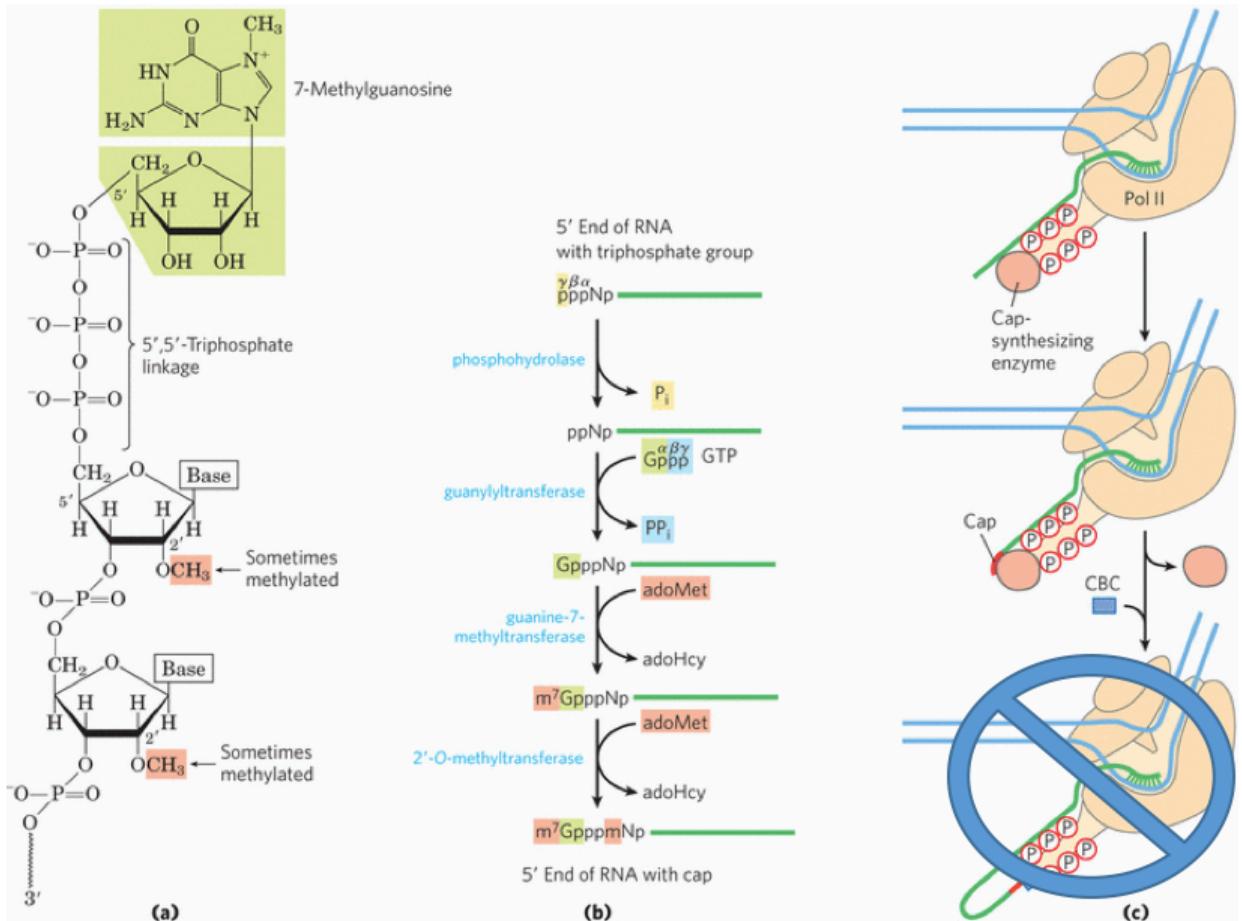
- a. Histone code: histone tails modifications that affect chromatin structure and transcription.
- b. Phos-CTD recruits methyltransferases.
- c. Methylation states may recruit TFIID (activation) or inhibit acetylation (suppression).



# Lecture 20: Eukaryotic RNA Processing

## A. 5'-capping of pre-mRNA

- Prevents digestion, helps translation.
- 7'-methylation of cap itself, then (often) 2'-methylation of nts 1/2
- m7G joined by 5'-5'-triphosphate linkage.
- Synthesis by enzymes tethered to RPol II CTD (co-transcriptional).



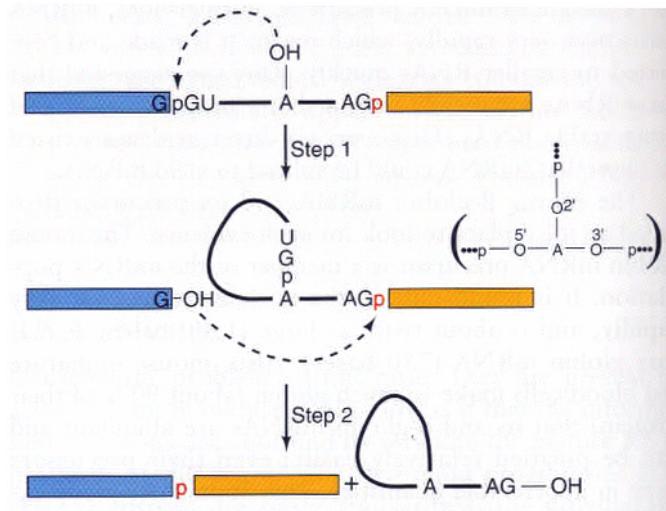
## B. RNA Splicing in Euks

- No introns in vertebrate pre-rRNA

Final Product	RNA Polymerase Synthesizing Pre-RNA	Location of Signals Important for Splicing	Distinguishing Characteristics
rRNA	I	Intron and splice sites	RNA catalysis, guanosine cofactor
mRNA	II	Splice sites, intron branch point	Lariat formation, involvement of small nuclear RNPs
tRNA	III	Exons	ATP-activated enzyme intermediates
Organelle RNA (mitochondrial, chloroplast)	Organellar		
Group I Introns* (rRNA, tRNA, mRNA)		Intron and splice sites (structure of intron same as rRNA above)	RNA catalysis, guanosine cofactor, intron-encoded splicing proteins
Group II Introns* (mRNA)		Intron and splice sites (sequences at splice sites similar to mRNA above)	RNA catalysis, lariat formation, intron-encoded splicing proteins

### C. Pre-mRNA Splicing Signals

- a. Consensus sequences at the ends of pre-mRNA introns:
  - i. Most interior nts are dispensable (to 70 nt minimum).
  - ii. Chimeric introns are spliced; splice sites are conserved (plant/animal).
  - iii. Mammals use “cryptic” splice sites (similar to the consensus sequences) when the natural splice site is disrupted. Yeast splicing just stops.
- b. Lariat intermediate
  - i. Internal intron A-OH attacks 5'-exon splice site (release of 5'-exon).
  - ii. Exon 1 G-OH attacks 3'-exon splice site (releases 3'-exon, connects to 5'-exon).

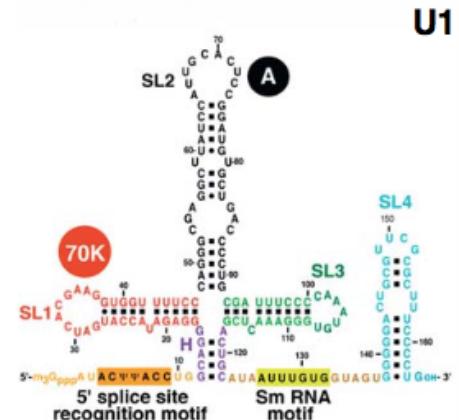


### D. Splicing Mutations and Disease

- a. 15% of all genetic diseases
- b. β-Thalassaemias ( $\beta$ -globin mutation/deficit):
  - i. Mutation of conserved G at 5'-intron forces use of cryptic sites.
  - ii. Mutation within intron creates new 5'-splice site (uncovers 3'-cryptic splice sites).

### E. snRNPs: small nuclear ribonucleoproteins

- a. snRNA complexes with 7 Sm proteins (except U6) via the Sm binding site, along with other specific proteins.
- b. Sm proteins are autoantigens in lupus, which provides antibodies against snRNPs.
- c. snRNA interacts with pre-mRNA via short absolutely conserved sequences.
- d. U6 is most highly conserved: catalytic core of the spliceosome.
- e. U4 delivers U6 to the spliceosome and leaves.
- f. snRNAs exist at  $10^6$  copies/cell.



## F. Steitz Lab: Evidence that snRNPs Involved in Splicing

### a. RNase Protection (“Bind-and-Chew”)

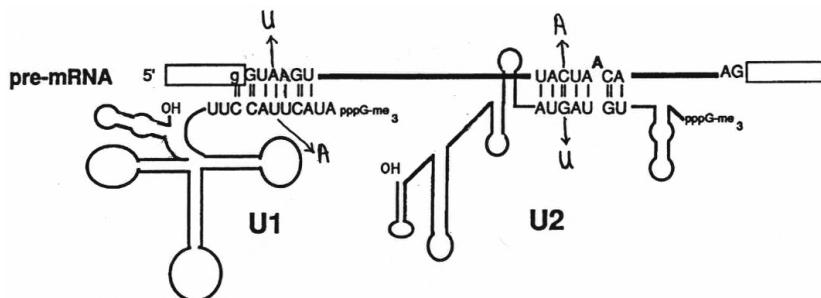
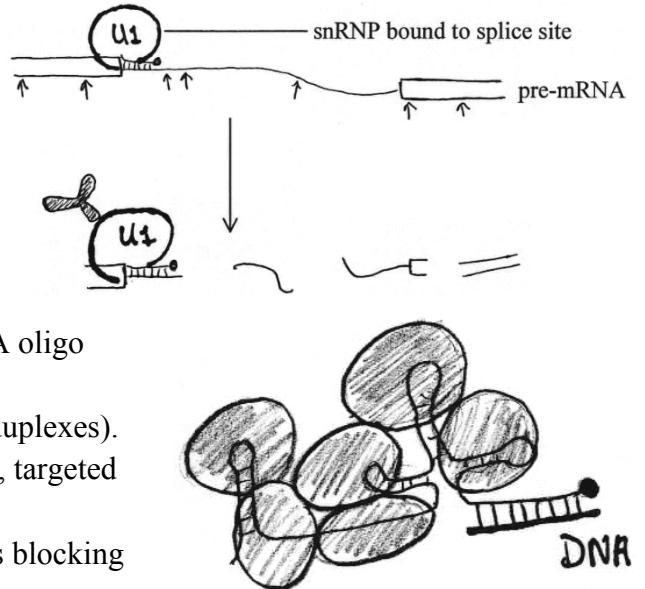
- Add  $^{32}\text{P}$ -labeled pre-mRNA to splicing extract.
- Digest with RNase; snRNPs protect bound fragment.
- Anti-snRNP-antibody selects fragment for analysis.

### b. RNase H Experiments

- Incubate splicing complex with DNA oligo complementary to snRNA.
- Add RNase H (cleaves DNA-RNA duplexes).
- Add splicing substrate. If no activity, targeted snRNP is essential.
- Worries from 2° structure or proteins blocking hybridization.

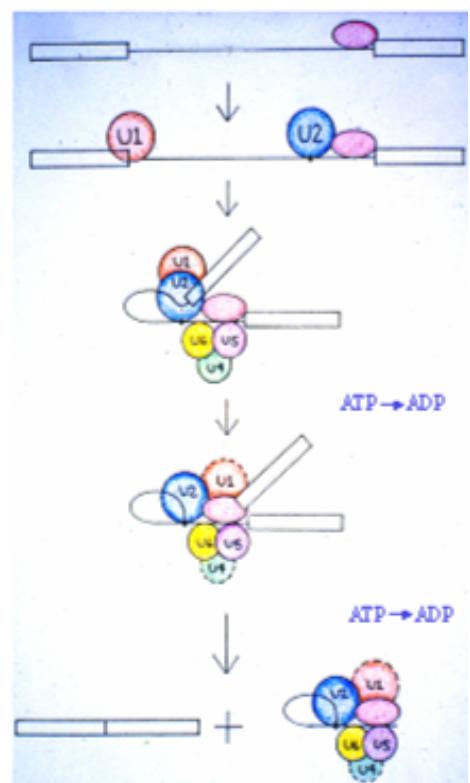
### c. Genetic Suppression Experiments

- Make knockout point mutations in the 5'-splice site or branch site.
- Engineer U1 or U2 compensatory mutations. Assay splicing
- Identifies binding interactions by base-pairing.
- Identifies essential U-RNAs (1, 2, 4, 6).
- snRNP interactions with pre-mRNA explain minimum intron size (~70 nt).



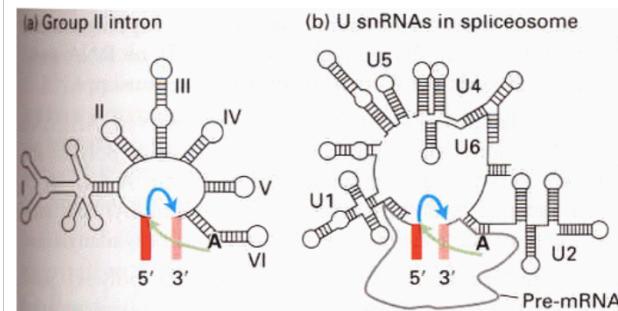
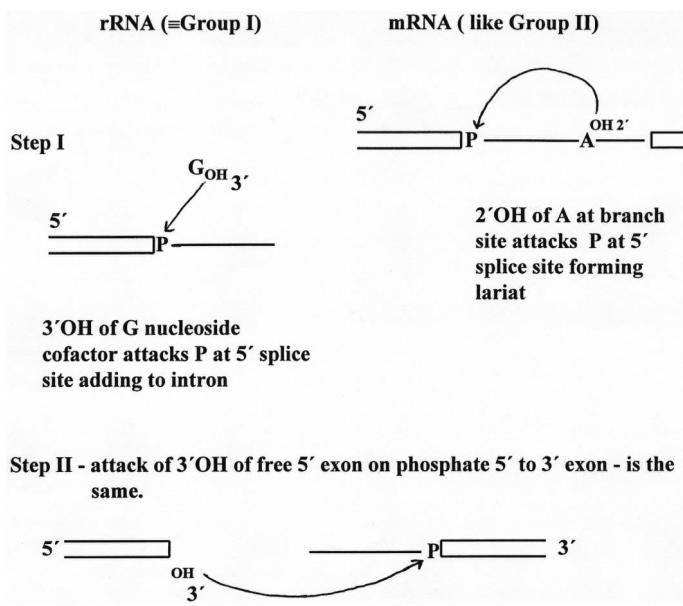
## G. Spliceosome Formation

- Active spliceosome is 60S (ribosome-sized) and contains 50-100 proteins in addition to snRNP proteins.
- Splicing factor U2AF binds intron poly-C/T upstream of 3'-splice site.
- U1/U2 bind at the 5'-splice site and branch site.
- U4/U6-U5 bind as tri-snRNP.
  - U4 separates and leaves
  - U6 replaces U1 at 5'-splice site.
  - U5 aligns exons for ligation.
- After exon ligation, the snRNPs recycle.
- ATP powers conformational rearrangements, but contributes no phosphates at any step.



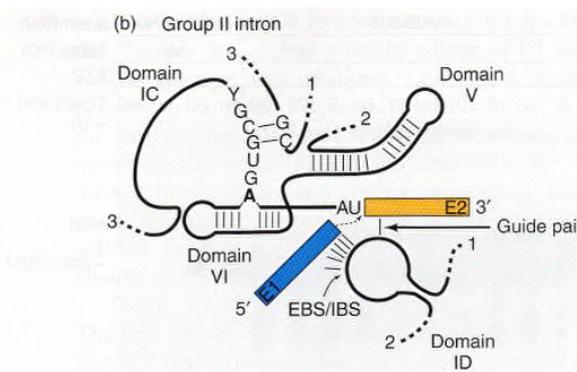
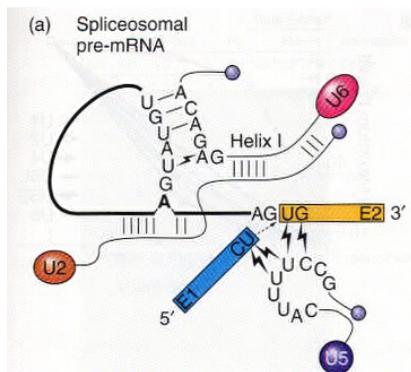
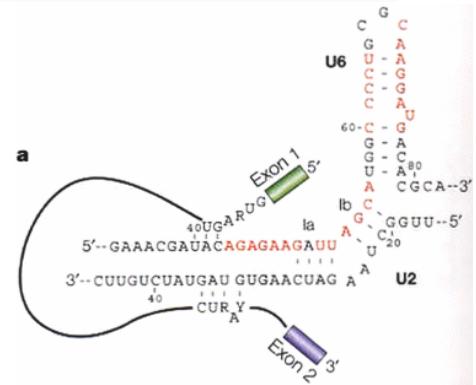
## H. Spliceosome as Ribozyme?

- a. Each nt attack and displacement is a transesterification, with 2 events in splicing.
- b. Resembles Group II mechanism.



## I. Dynamic Changes in RNA-RNA Interactions in the Spliceosome

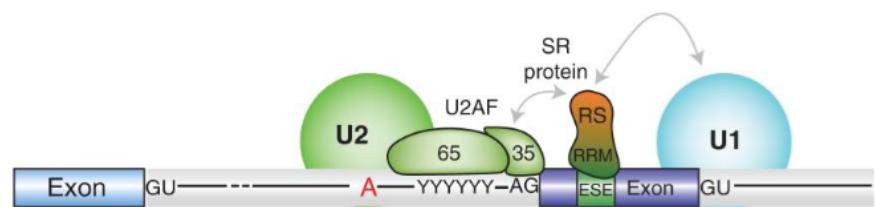
- a. Step 1
  - i. U1 snRNP identifies 5'-splice site, but replaced by U6 prior to step 1.
  - ii. U6 binds U2 to bring the 5'-splice site and branch site together for step 1.
- b. Step 2
  - i. Parallels with Group II suggest RNA catalysis.



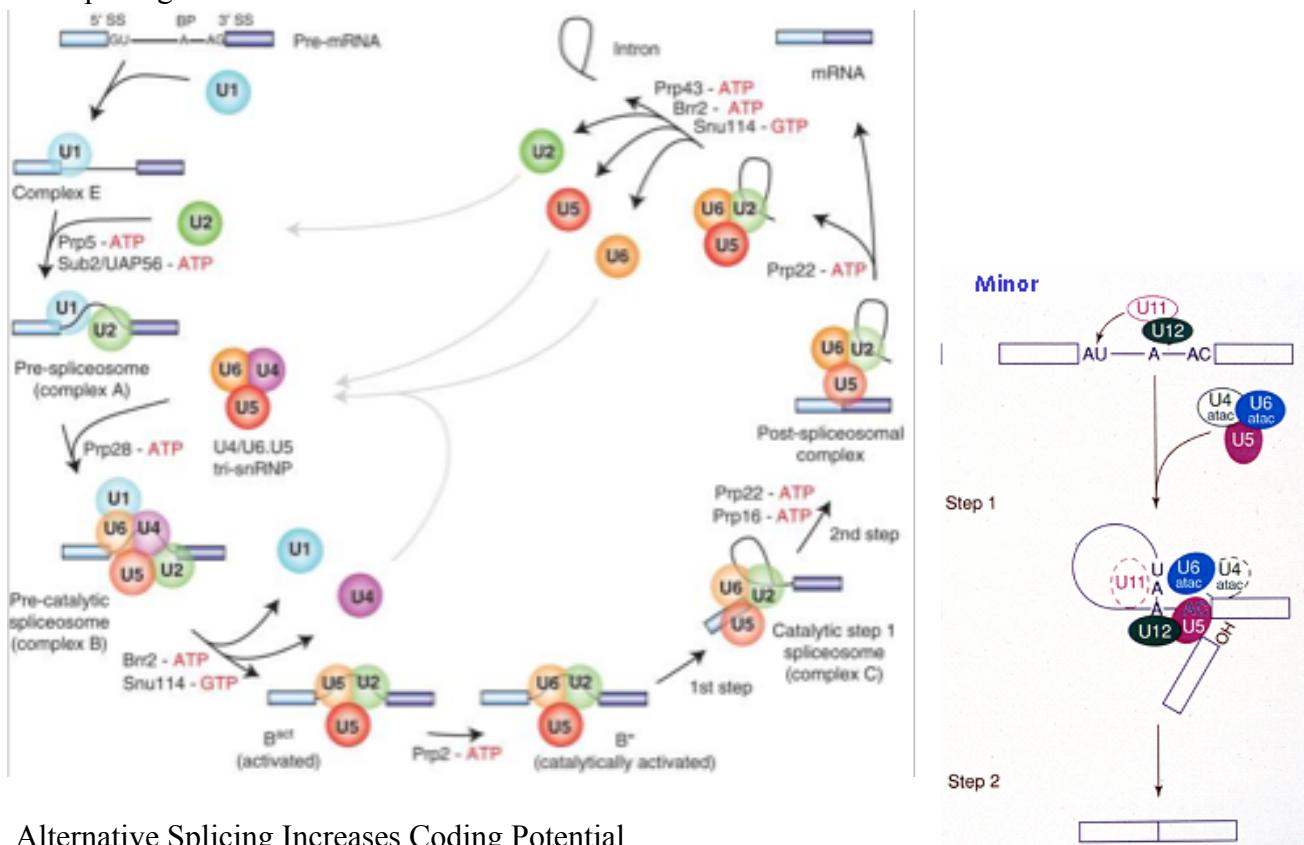
## J. Role of Spliceosomal Proteins

- a. Exon and Intron Definition
  - i. Assist U1/U2 snRNP binding.
  - ii. Mammals: SR proteins (splicing factors) have Ser/Arg-rich CTDs.
  - iii. ISE: Intronic splice enhancer, ESE: Exonic splice enhancer

- iv. ~20 SR proteins bind to ISEs/ESEs and help snRNPs find the splice sites.
- v. SR proteins also assist mRNA nuclear export.
- b. Promote RNA-RNA exchanges
  - i. DEAD/DEAH Box Proteins: ATPases, often helicases.



### K. Full Splicing Mechanism



### L. Alternative Splicing Increases Coding Potential

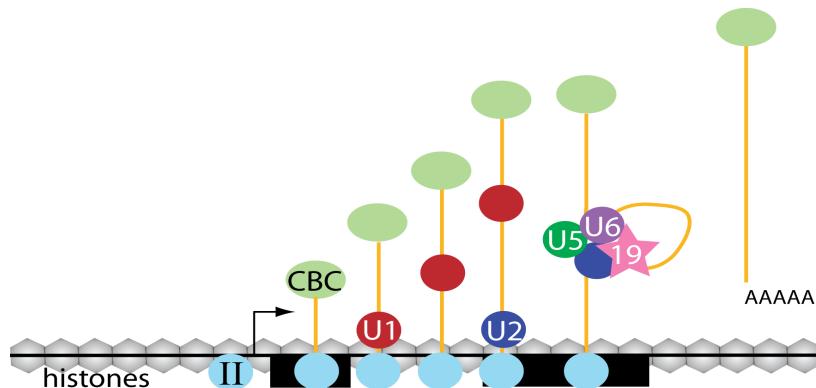
- a. 95% of human genes are alternatively spliced.
- b. Often by choosing each exon from a set of possibilities and splicing out the others.
- c. Drosophila sex determination by alternative splicing  
(male transcript introduces stop codons or adds chunks).
- d. Caused by: SR proteins, hnRNPs (heterogeneous nuclear RNPs), and other proteins bind to pre-mRNA to alter spliceosome assembly.

### M. Second or "Minor" Spliceosome (U12-dependent)

- a. Different consensus sequence but similar mechanism.
- b. Very old, present across plants and animals.
- c. Rare: 0.3% of human introns.
- d. snRNPs (U11, U12, U4atac, U6atac) resemble major snRNPs, but only 1% as abundant.
- e. U5 common to both spliceosomes.

## N. Pre-mRNA Splicing Occurs Co-Transcriptionally

### a. Processing steps associated with phases of transcription.



Transcription:	Initiation	Elongation	Termination
Processing:	Capping	Splicing	Polyadenylation

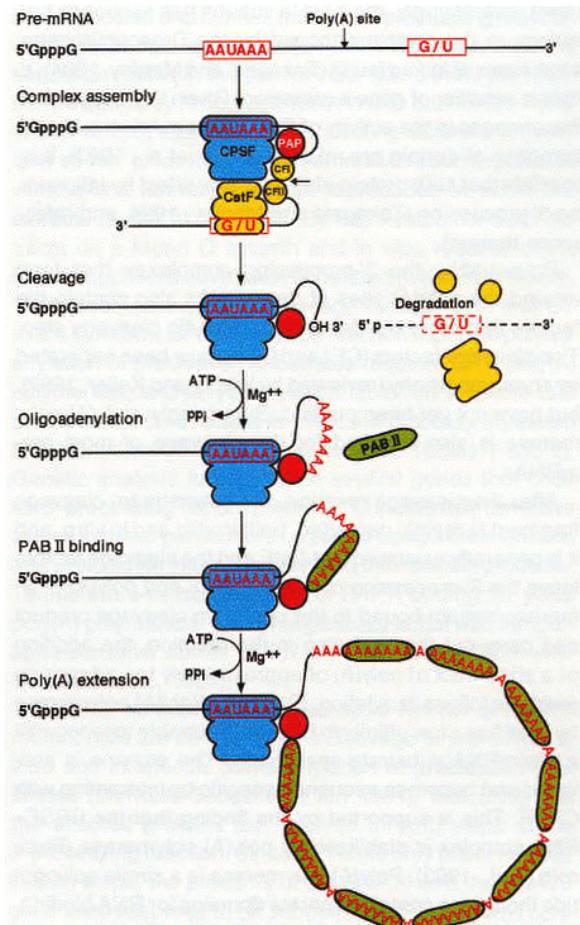
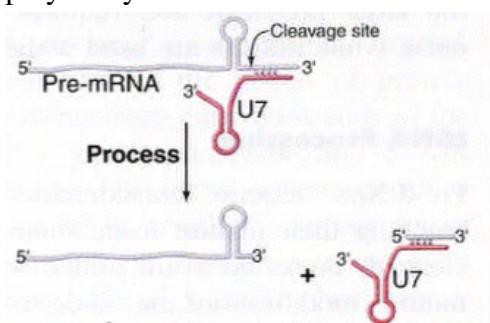
## O. Vertebrate mRNA 3'-End Formation

### a. Polyadenylation

- i. Only protein factors!
- ii. Pre-mRNA 3'-end includes AAUAAA signal and GU-rich region.
- iii. Complex assembly and cleavage.
- iv. First, ~10 polyA pre-tail.
- v. PAP (polyA polymerase) elongates to ~200 nts
- vi. PAB (polyA binding protein) decorates the strand 25-nt at a time.
- vii. Polyadenylation coordinated with termination/splicing via RPol II CTD.

### b. Histone mRNA

- i. No polyA tails and no introns, but still processed by 3'-cleavage.
- ii. U7 snRNP binds with an element 3'- to the cleavage site and cuts it.
- iii. Involves similar factors as polyadenylation.

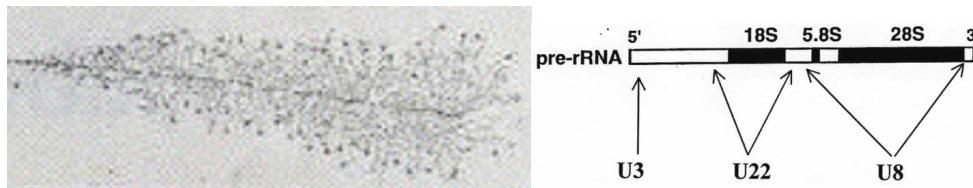


P. Summary of Nuclear Pre-mRNA Processing

<b>Modification</b>	<b>catalysis</b>	<b>how specified</b>	<b>function</b>
5' end capping	3 enzymes	Pol II CTD (trans)	stability splicing translation
Splicing	RNA (spliceosome)	RNA sequences (cis)	stability makes ORF(s) export
Polyadenylation	2 enzymes	RNA and Pol II (cis and trans)	stability alternative ORFs translation

Q. Nucleolus Inhabited by snoRNPs (lots)

- a. snoRNPs (small nucleolar RNPs) are involved in pre-rRNA processing.
- b. 3 snoRNPs involved in cleavage events.
- c. U3 snoRNP attaches near 5'-end and recruits proteins to form 5'-terminal knobs.



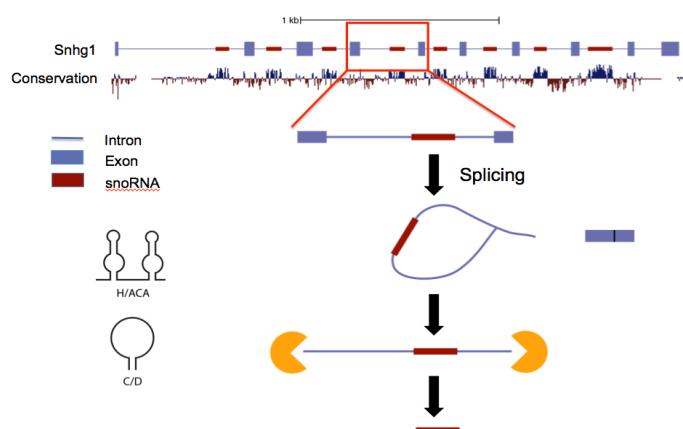
- d. U22 snoRNP excises 18S
- e. U8 snoRNP half-excises 5.8S and 28S.
- f. No intron in vertebrate pre-rRNA!

R. SnoRNPs Guide Modification of rRNA

- a. SnoRNPs base-pair to pre-rRNA and guide modifications at conserved positions.
- b. Box C/D snoRNAs direct 2'-O-methylation.
- c. Box H/ACA snoRNAs direct pseudouridylation by dyskerin.
- d. ~100 of each, so ~200 guide snoRNPs predicted.
- e. Each class uses different proteins.

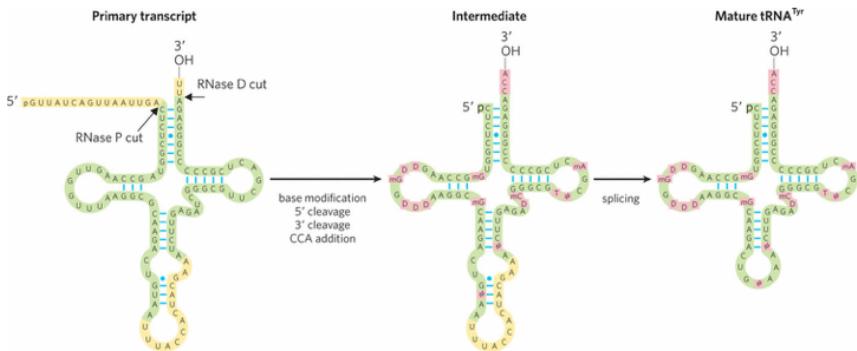
S. Most snoRNAs Encoded in Introns

- a. Introns from housekeeping protein-coding genes (riboproteins, nucleolar proteins, translation factors).
- b. SnoRNP proteins bind to snoRNA sequence during splicing.
- c. Lariat intron is debranched
- d. Exonucleases trim the ends and are stopped by snoRNP proteins.
- e. SnoRNP migrates to nucleolus.



T. Euk pre-tRNAs have introns with protein-mediated splicing.

- a. RNase P and D cut and define the acceptor stem.
- b. Base modification and CCA addition
- c. Splicing of the anticodon loop.



### U. RNA Editing Alters Primary Gene Transcripts

a. Mammalian cells:

- i. C → U deamination (protein-catalyzed)
- ii. A → I deamination (dsRNA-catalyzed; ADAR: adenosine deaminase RNA)

b. Protozoa:

- i. Site-specific insertion/deletion of Us in mitochondria.
- ii. Involves guide-RNA-directed endonucleases (opening cut), TUTases (insertions), and 3'-U-exonucleases (deletions).

# Lecture 21: Eukaryotic Translation

## A. Overview of Euk Translation

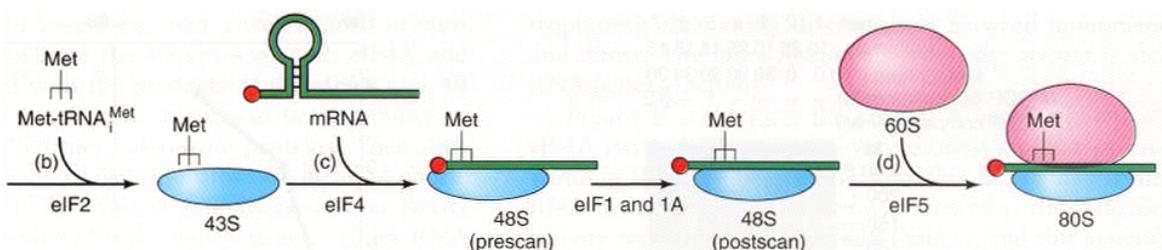
- Physically separated/compartmentalized from transcription.
- Cytoplasmic mRNA with short polyA tails are masked (not translated). Allows storage of maternal mRNAs and neural learning.
- Euk mRNA is monocistronic, 1 ORF.
- Initiation
  - Initiation sequences are different.
  - 5'-cap and polyA tail are necessary.
  - Requires ATP as well as GTP.
- Termination involves 1 RF for all 3 stop codons, and a stimulatory GTPase factor.
- Compartmental sorting of proteins (mRNA are translated where you want the protein).
- 80S Euk ribosomes are larger and more protein-rich.
  - 40S subunit (18S rRNA)
  - 60S subunit (28S, 5.8S, 5S rRNAs)
- Proteins may undergo post-translational cleavage and modification (even splicing!)

## B. Euk Translation Initiation

- No Shine-Dalgarno, but has 5'-UTR of 3 to 1000+ nts.
- Usually selects 1<sup>st</sup> AUG, but sequence context is important. Exceptions: viral proteins, proto-oncogenes/growth factors with IRES (internal ribosome entry sites).
- Kozak sequence: ACCAUGG is optimal
- Special Met-tRNA<sub>i</sub> required, but no formyl group.
- Scanning from 5'-end
  - Requires 5'-cap, 40S, and eIFs to start, ATP to scan.
  - Unfolds secondary structure.
  - Stops when Met-tRNA<sub>i</sub> pairs with AUG of an ORF.
  - 60S attaches and eIFs fall off.

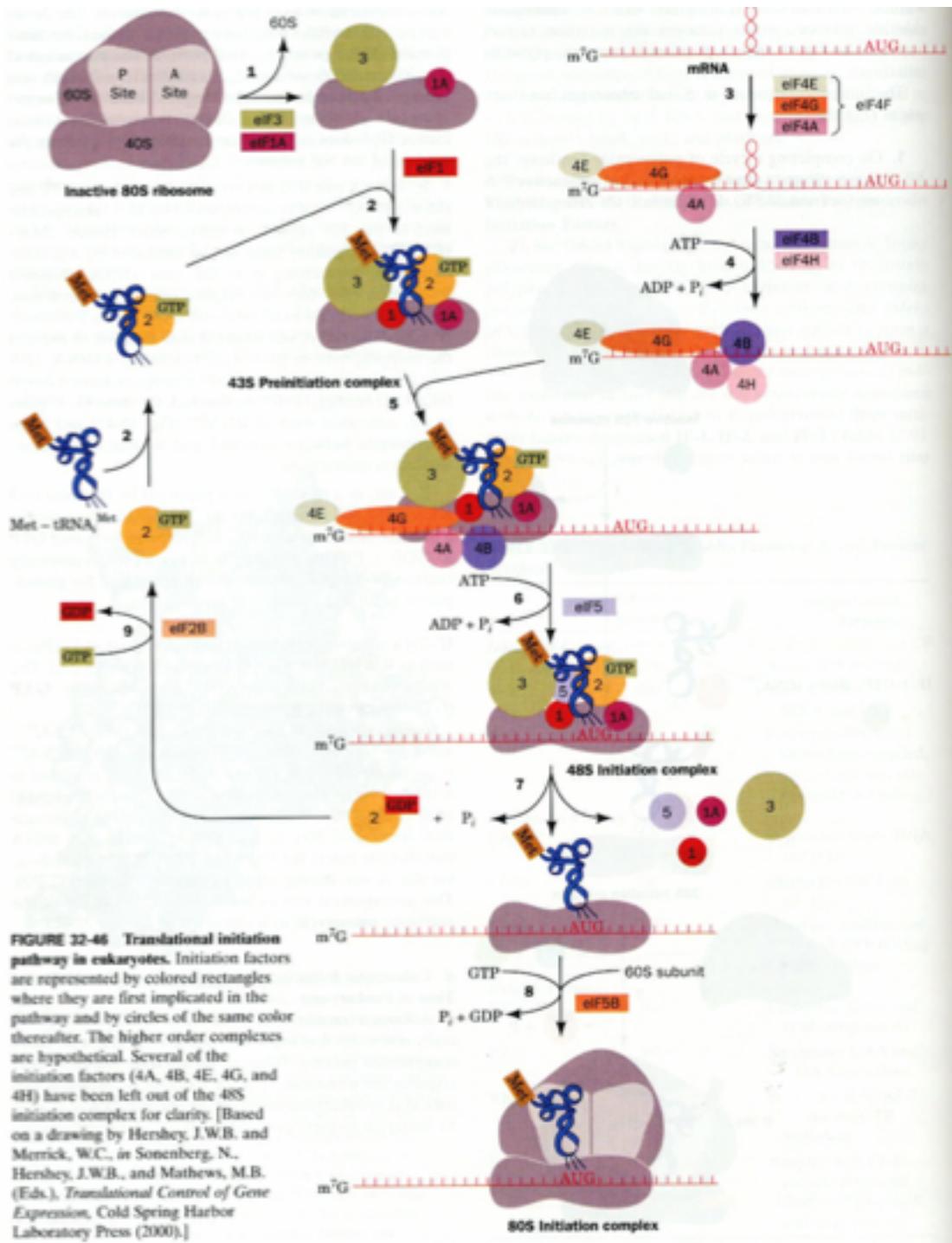
## C. Euk Initiation Factors (eIFs)

- eIF2: binds Met-tRNA<sub>i</sub> and GTP. Brings it to the 40S subunit to form 43S pre-primed complex which scans from cap.
- eIF4A is DEAD box helicase, unwinds 2° structure.
- GTP hydrolysis releases eIF2.
- Scanning can be regulated by upstream AUGs (as in GCN4) and 2° structure (as in iron).
- eIF2B (GEF) is the guanosine exchange factor: releases GDP from eIF2.  
Analog of Ts acting on Tu-GDP.



#### D. Euk Translation Initiation Cycle

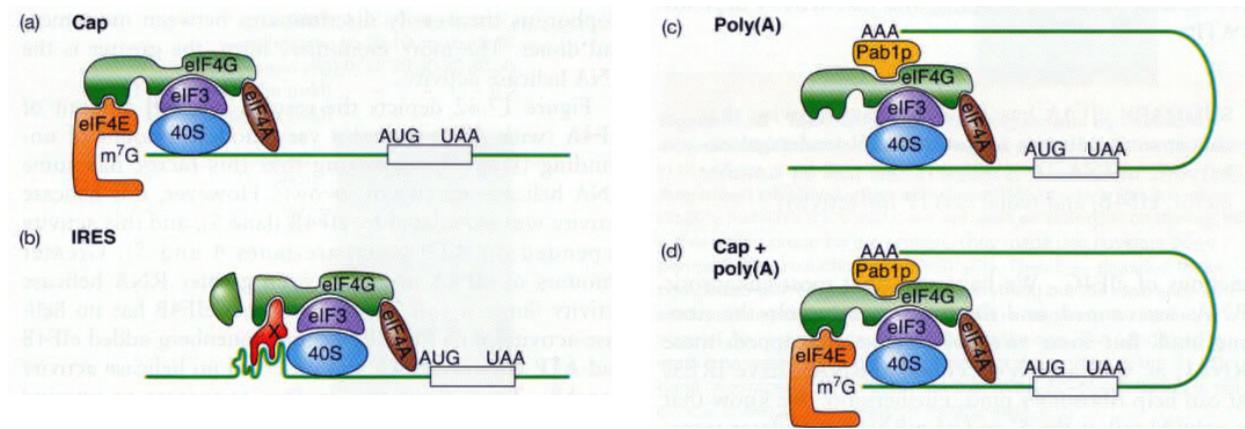
- Phos-eIF2 sequesters eIF2B, stopping translation.
- Need to know: eIF- 2, 2B, 4A, 4E, 4G.



**FIGURE 32-46** Translational initiation pathway in eukaryotes. Initiation factors are represented by colored rectangles where they are first implicated in the pathway and by circles of the same color thereafter. The higher order complexes are hypothetical. Several of the initiation factors (4A, 4B, 4E, 4G, and 4H) have been left out of the 48S initiation complex for clarity. [Based on a drawing by Hershey, J.W.B. and Merrick, W.C., or Sonenberg, N., Hershey, J.W.B., and Mathews, M.B. (Eds.), *Translational Control of Gene Expression*, Cold Spring Harbor Laboratory Press (2000).]

#### E. Interactions at mRNA Ends

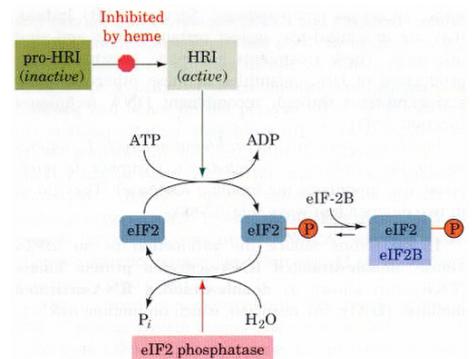
- Polysomes: multiple ribosomes translate a single mRNA.
- Circular mRNAs (linking cap and polyA) helps recycle ribosomes.
- eIF-4G as adaptor:
  - Binds eIF4E (cap), PAB (tail), and eIF3 (40S subunit).
  - Binds IRES proteins, allowing initiation at internal AUGs.



- d. Cap: recruits 40S to mRNA via eIF4E.
- e. IRES: recruits 40S to mRNA via RNA-binding protein X.
- f. PolyA: adaptor to connect 40S to polyA

#### F. Regulation of Initiation 1: Heme

- a. Heme + globins = hemoglobin.
- b. Reticulocytes are immature erythrocytes with a ton of globin mRNAs.
- c. Heme required for translation of globin mRNAs.
  - i. Heme inactivates an eIF2 kinase with 2 names:  
HRI (heme-regulated inhibitor)  
HCR (heme-controlled repressor)
  - ii. This sequesters eIF2B and stops GEF activity, stops protein synthesis.
  - iii. HRI also activated by heat shock, to prevent misfolded proteins.

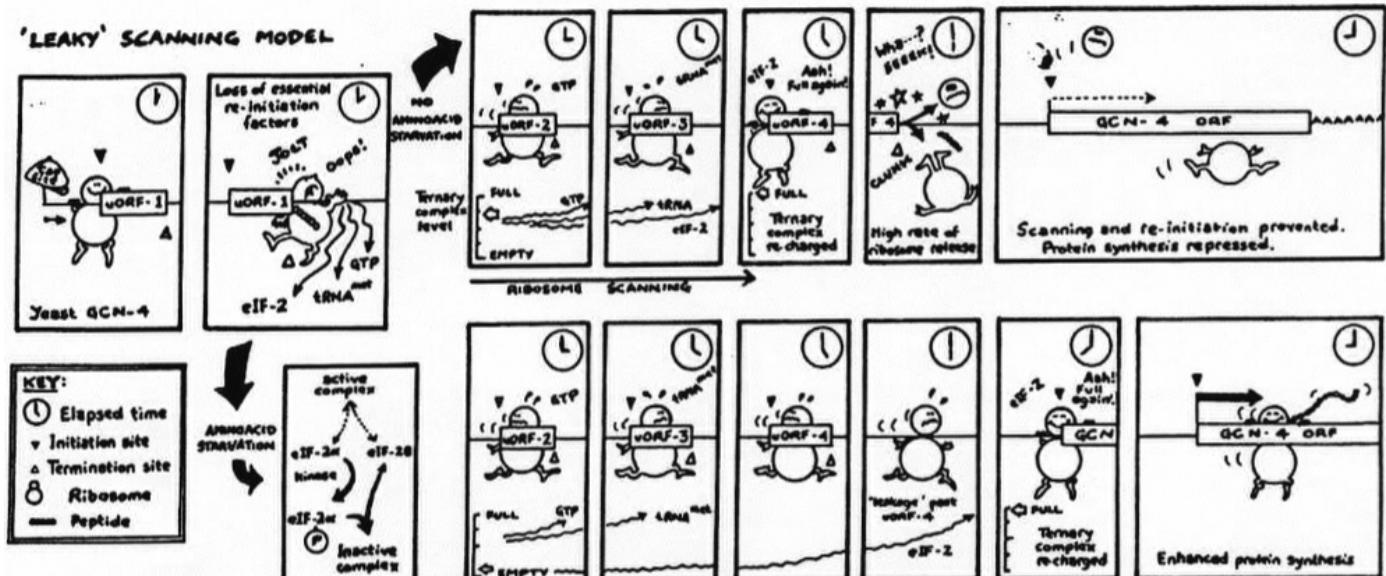


#### G. Regulation of Initiation 2: Interferon (glycoprotein)

- a. dsRNA activates interferon secretion, which signals antiviral state in neighbors.
- b. PKR (protein kinase dependent on dsRNA): phosphorylates α-subunit of eIF2 if dsRNA appears in the newly antiviral cell.
- c. Viral response: adenovirus encodes an imperfect dsRNA dummy that competes for binding to PKR but does not activate it.

#### H. Regulation of Initiation 2: GCN4 (yeast)

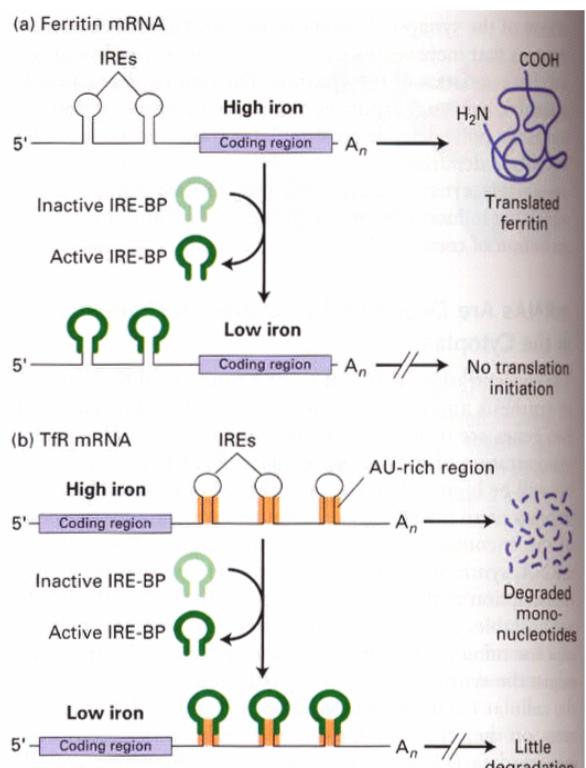
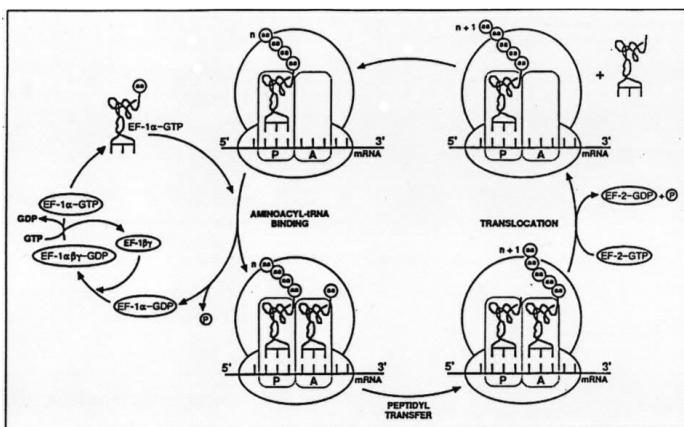
- a. GCN4 is a transcriptional activator of AA biosynthesis genes when limiting.
- b. Translational control: transcribed at a constant rate, regardless of supply.
- c. Mechanism
  - i. 5'-UTR of UCN4 mRNA contains 4 AUGs, each with a short ORF.
  - ii. Mutating/deleting these AUGs (or removing the whole 5'-UTR) leads to constitutive activation.
  - iii. Uncharged tRNAs bind and activate GCN2 kinase to phosphorylate eIF2 and reduce availability of TC (eIF2 + Met-tRNA<sub>i</sub> + GTP)
  - iv. ORF 1 is translated, and the 40S keeps scanning. If it has TC/eIF2 available, it will restart at ORF 3 or 4 and drop. Otherwise, it will scan past and arrive at the GCN4.



### I. Translational Control by Iron

- IREs (Iron Response Elements): ~10 bp conserved RNA stem loops bind IRE-BPs.
- IRE-BP is aconitase!
- Fe and IREs compete for binding.  
High Fe = Free IRE, low Fe = bound IRE.
- Ferritin (iron storage protein):  
BP can bind 5'-UTR IREs to stop scanning and translation.
- Transferrin (iron uptake protein):  
BP can bind 3'-UTR IREs to stabilize mRNA.
- AREs (AU-rich elements): found in 3'-UTRs; trigger degradation.

### J. Euk Elongation

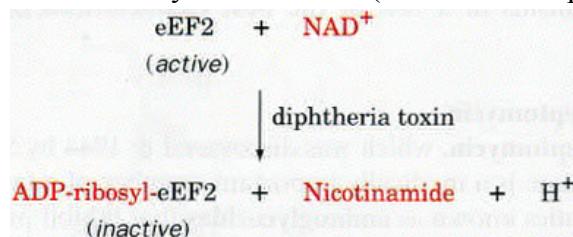


Factor	Function	Comparable Prokaryotic Factor*
EF1 $\alpha$	Binds aminoacyl-tRNA to ribosomes	EF-Tu
EF1 $\beta$	Assists the recycling of EF1 $\alpha$	EF-Ts
EF2	Translocation	EF-G
RF	Chain release	RF1 RF2

- a. Additional: RF3 binds GTP to stimulate termination.
- b. Control of elongation by
  - i. Phosphorylation
  - ii. ADP-ribosylation of eEF2 by Diphtheria toxin.

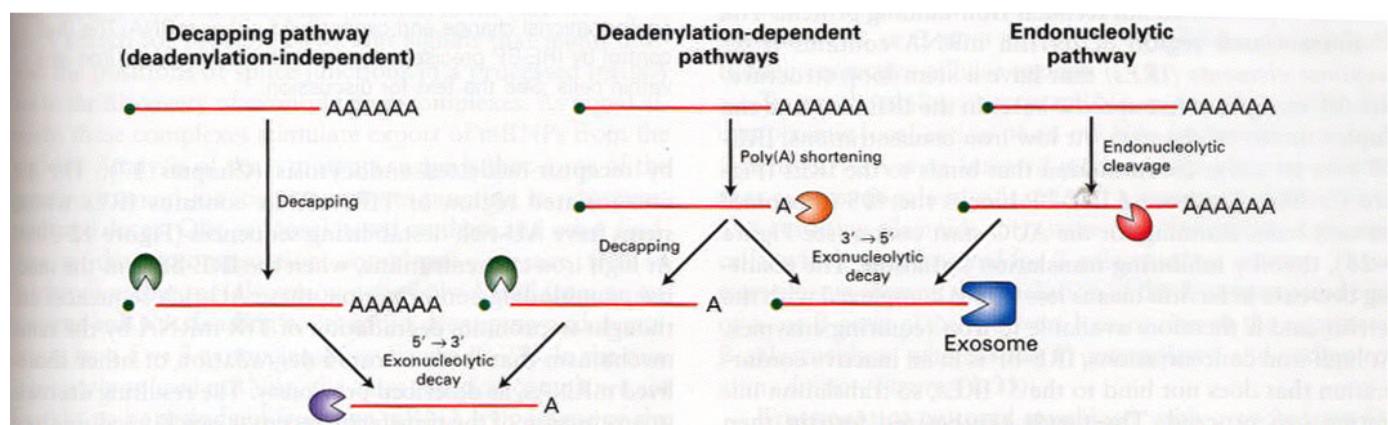
#### K. Inhibition of Elongation: Diphtheria Toxin

- a. ADP-ribosyl transferase: adds group to eEF2-His, inactivates factor.
- b. Targets unique diphthamide modification in euks (not in EF-G).
- c. Enzymatic activity: 1 molecule can poison all eEF2s in a cell.
- d. Reversible by nicotinamide (L' Chatelier's principle).



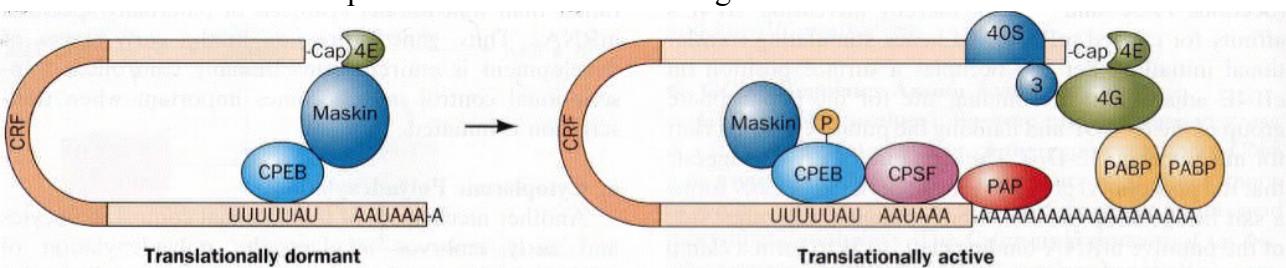
#### L. PolyA Tail

- a. PolyA lengthening increases translation.
- b. PolyA trimming triggers decay. Other degradation pathways also exist.



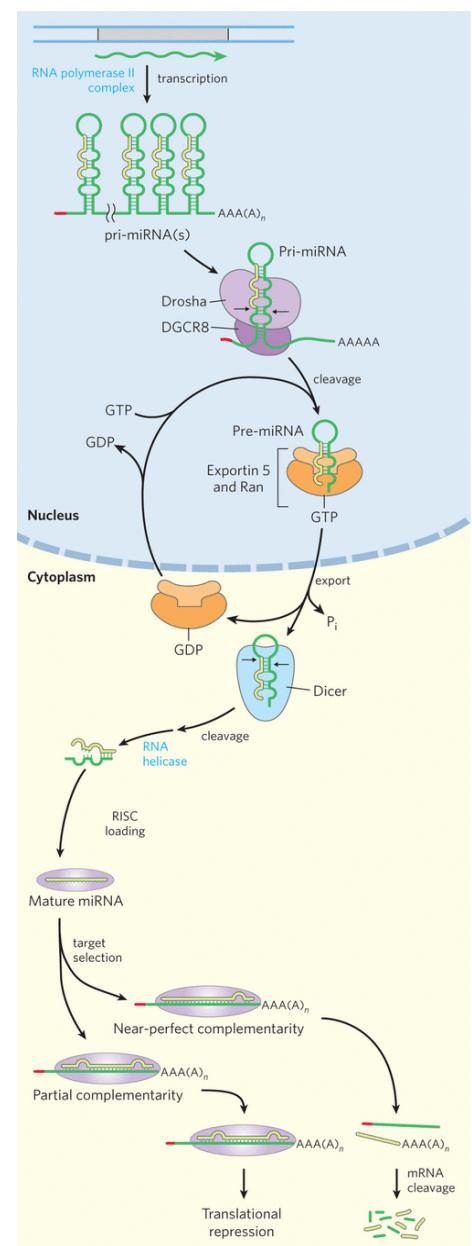
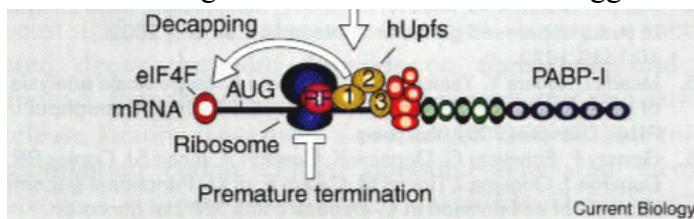
- c. Maternal mRNAs masked by proteins and short polyA.

- i. CPEB (cytoplasmic polyadenylation element binding protein) binds Maskin and mRNA CPE. Maskin binds eIF4E-cap to stop eIF4G from initiating translation.
- ii. Phos-CPEB recruits PAP to extend the tail. PABP binds the new tail and recruits eIF4G to displace maskin. Translation begins.



## M. mRNA Stability

- a. AREs (AU-rich elements) – usually AUUUA
  - i. Target mRNA for decay.
  - ii. Present in growth factor and proto-oncogene 3'-UTRs.
  - iii. AREs can be transferred to other mRNAs to reduce half-life.
  - iv. Mechanism: factors recognize ARE and recruit exosome to for 3'-5' degradation.
  - v. Counter: other factors (e.g. IRE-BP) can competitively bind the ARE.
- b. NMD (nonsense-mediated decay)
  - i. Also called mRNA surveillance.
  - ii. Degrades mRNA to avoid producing a truncated protein.  
This helps make nonsense mutation diseases recessive!
  - iii. Exon-junction complex (EJC): upstream of exon-exon junctions after splicing.
  - iv. When ribosome terminates prematurely, interaction with EJC factors leads to mRNA decapping.
  - v. Inserting an intron into the 3'-UTR triggers mRNA decapping and degradation.



## N. RNA interference (RNAi)

- a. Overexpression attempts knocked out endogenous expression. dsRNA triggers degradation of complementary mRNA.
- b. Dicer: reduces dsRNA to ~22 nt siRNAs, which are loaded onto RISC.
- c. RISC: contains Argonaute (Ago) RNase H-like endonuclease that cleaves target mRNA.

## O. Micro RNA (miR)

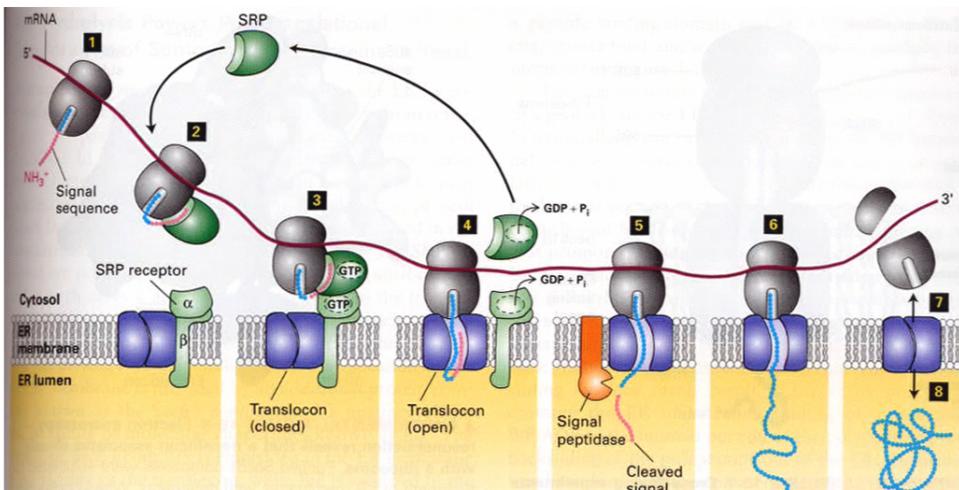
- a. Function
  - i. Perfect complementarity with target: RISC cleaves the target.
  - ii. Imperfect complementarity with target: represses translation.
- b. Biogenesis
  - i. Drosha cleaves pri-miRNA to form pre-miRNA.
  - ii. Dicer cleaves and loads miRNA into RISC.
  - iii. Low dicer associated with macular degeneration.
  - iv. Drosha and Dicer are RNase III enzymes: pseudo dimers; 2 catalytic domains per protein.

### P. The Signal Hypothesis

- a. How do membrane and secreted proteins cross bilayers?
- b. Signal sequence
  - i. Targets to proper location in the cell.
  - ii. Hydrophobic core preceded by 1+ basic residues.
  - iii. Usually at N-terminus.
  - iv. Cleaved off once protein crosses membrane.

### Q. Signal Recognition Particle (SRP)

- a. Transports nascent polypeptides across ER membranes.
- b. RNP complex of 7SL RNA and 6 proteins.
- c. Recognizes emerging signal sequence and binds ribosome to arrests translation.
  - i. Sequesters nasty enzymes meant for export.
  - ii. **??? Increases targeting fidelity.**
- d. SRP receptor in the membrane guides SRP to the translocon.
  - i. GTP hydrolysis triggers release of SRP and continuation of protein synthesis.
  - ii. Signal peptidase cleaves the signal.



### R. Controlled Protein Degradation by the Ubiquitin Pathway

- a. **Ubiquitin**: most conserved euk protein.
- b. **E3 (ubiquitin ligase)** links many ubiquitins to internal Lys of condemned proteins.
- c. **Poly-Ub protein** is directed to proteasome (multi-protease complex).
- d. ATP hydrolysis required at many steps.
- e. **N-end rule**: E3 recognizes N-terminal residues, which determine protein half-life (can be stabilizing or destabilizing).

### S. Protein Splicing

- a. Present in all kingdoms of life.
- b. Fast, autocatalytic, occurs via branched intermediate.
- c. **Inteins** (intron analog) are 100-2000 AA long, with conserved AA at N and C termini.
- d. **C-terminal exteins** contain conserved Ser.
- e. Some inteins are DNA endonucleases, **like those encoded by Group I RNA introns???**

## Lecture 22: Tumor Viruses & Oncogenes

---

## Lecture 23: Development

---

## Lecture 24: Immunology

---