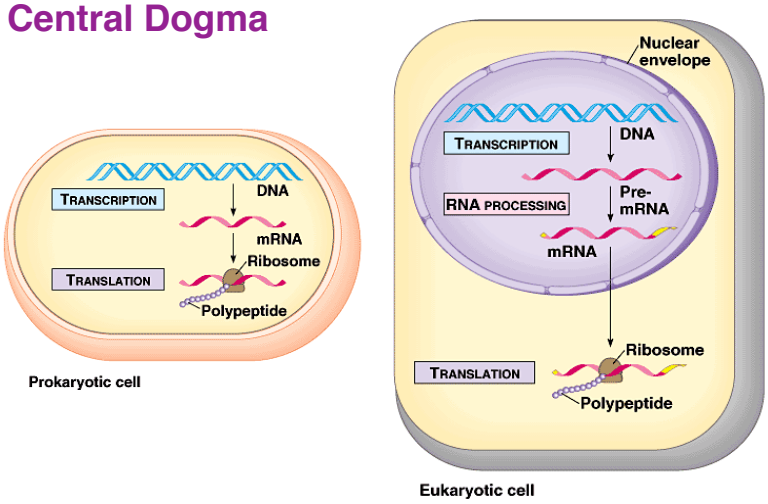
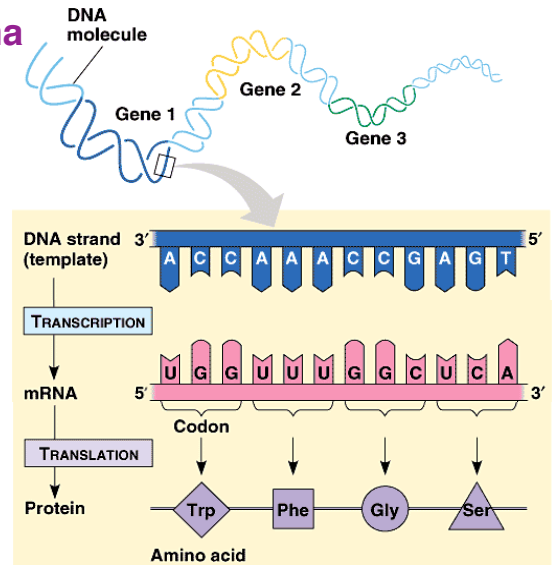
Central Dogma



Central Dogma



Translation (mRNA -> protein)

mRNA code for amino acids

1. Codons as Triplet code
2. Redundancy
3. Open reading frames
4. Start and stop codons
5. Mistakes in translation
6. Code is universal

Mechanism of Translation by tRNA and Ribosomes

Translation (mRNA -> protein)

1. The triplet code: DNA & RNA code for amino acid sequence using a triplet code:

3 bases = 1 codon

There are 20 amino acids, but only 4 bases (A,T,C,G)

So 1 base code -> 4 amino acids

2 base code -> 4 x 4 = 16 amino acids

3 base code -> 4 x 4 x 4 = 64 amino acids

2. Redundancy: With 3 base codons, have 44 triplets left over; the triplet code is redundant:

2 - 4 triplets code for each amino acid.

How the code was broken:

Put simple artificial RNA into an *in vitro* translation system, and see what peptides come out.

UUUUUUUUU → **Phe-Phe-Phe** **so UUU = Phe**

AAAAAAAAA → **Lys-Lys-Lys** **so AAA = Lys**

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

3. Open reading frames:

In the mRNA, there are no spaces to distinguish one codon from the next, so we have to look for the **open reading frame**
= the grouping of stretches of RNA so it makes "sense" to ribosomes as a series of codons.

Top strand:
5'-anddadehtdahohwnamtafehtrofnargodehtdnatacehtetagoddereht-3'

Bottom strand:
5'-qxthereddogatethecatandthedogranforthefatmanwhohadthedaddna-3'

Given a piece of DNA/RNA that codes for a protein, there are 6 possible open reading frames, but only the longest ORF makes sense:

Top strand:
5'-anddadehtdahohwnamtafehtrofnargodehtdnatacehtetagoddereht-3'

3 ORFs using the top strand:

and dad eht dah ohw nam taf eht rof nar god eht dna tac eht eta god der eht xq

a ndd ade htd aho hwn amt afe htr ofn arg ode htd nat ace the tag odd ere htx q

an dda deh tda hoh wna mta feh tro fna rgo deh tdn ata ceh tet ago dde reh txq

Bottom strand:
5'-qxthereddogatethecatandthedogranforthefatmanwhohadthedaddna-3'

3 ORFs using the bottom strand:

qxt her edd oga tet hec ata ndt hed ogr anf ort hef atm anw hoh adt hed add na

q xth ere ddo gat eth eca tan dth edo gra nfo rth efa tma nwh oha dth eda ddn a

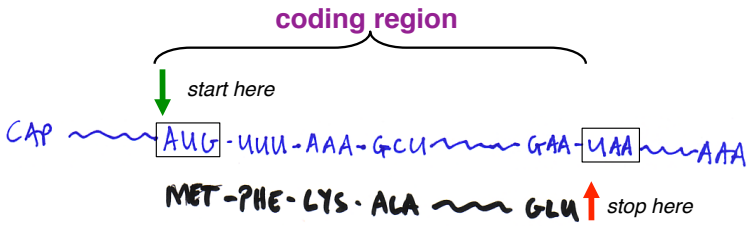
qx the red dog ate the cat and the dog ran for the man who had the dad dna

4. Start and Stop codons

There are special code words for start translation and stop translation.

AUG = start (= methionine, so all proteins start with met)

UAA, UAG, UGA = stop



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

Rattus norvegicus insulin 1 (Ins1), mRNA

NCBI Reference Sequence: NM_019129.3

FASTA Graphics

```
LOCUS       NM_019129               462 bp    mRNA    linear   ROD 22-OCT-2011
DEFINITION  Rattus norvegicus insulin 1 (Ins1), mRNA.
ACCESSION   NM_019129 REGION: 1..462
```

```
  1 aaccctaagt gaccagctac aatcatagac catcagcaag caggtcattg ttccaacatg
 61 gccctgtgga tgcgcttcct gcccctgctg gccctgctcg tcctctggga gcccaagcct
121 gcccaggctt ttgtcaaaca gcacctttgt ggtcctcacc tggtggaggc tctgtacctg
181 gtgtgtgggg aacgtggttt cttctacaca cccaagtccc gtcgtgaagt ggaggacccg
241 caagtgccac aactggagct gggtggaggc ccggaggccg gggatcttca gaccttggca
301 ctggaggttg cccggcagaa gcgtggcatt gtggatcagt gctgcaccag catctgctcc
361 ctctaccaac tggagaacta ctgcaactga gtccaccact ccccgcccac ccctctgcaa
421 tgaataaagc ctttgaatga gcaccaaaaa aaaaaaaaaa aa
```

```
source          1..463
                /organism="Rattus norvegicus"
                /mol_type="mRNA"
                /db_xref="taxon:10116"
                /chromosome="1"
                /map="1q54-q55"
gene            1..463
                /gene="Ins1"
                /note="insulin 1"
                /db_xref="GeneID:24505"
                /db_xref="RGD:2915"
misc_feature    10..12
                /gene="Ins1"
                /note="upstream in-frame stop codon"
STS             47..346
                /gene="Ins1"
                /standard_name="RH94799"
                /db_xref="UniSTS:120046"
CDS             58..390
                /gene="Ins1"
                /note="insulin-1"
                /codon_start=1
                /product="insulin-1 preproprotein"
                /protein_id="NP_062002.1"
                /db_xref="GI:9506815"
                /db_xref="GeneID:24505"
                /db_xref="RGD:2915"
                /translation="MALWMRFLPLLALLVLWEPKPAQAFVKQHLCGPHLVEALYLVCG
                ERGFFYTPKSRREVEDPQVPQLELGGGPEAGDLQTLALEVARQKRGIVDQCCTSICSL
                YQLENYCN"
sig_peptide     58..129
                /gene="Ins1"
                /inference="COORDINATES: ab initio prediction:SignalP:4.0"
proprotein      130..387
                /gene="Ins1"
                /product="proinsulin-1"
mat_peptide     130..219
```

5. Mistakes in the DNA code

Point mutations:

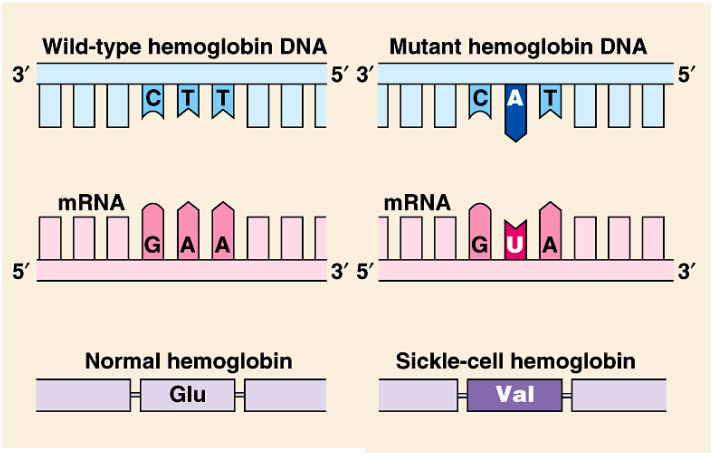
- **substitution -- the wrong base is transcribed**
- **insertion -- an extra base is inserted**
- **deletion -- a base is deleted**

Effects of mutations:

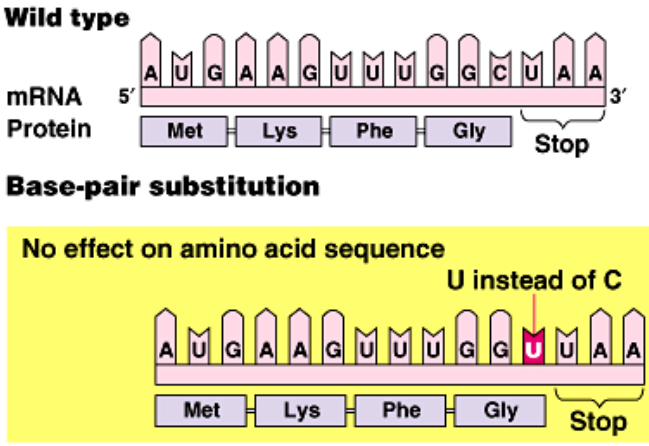
- **no effect (i.e. substitution in the 3rd base)**
- **missense**
- **nonsense**
- **frameshift -> missense or nonsense**

Note: mutations outside the coding region are more likely to have no effect...

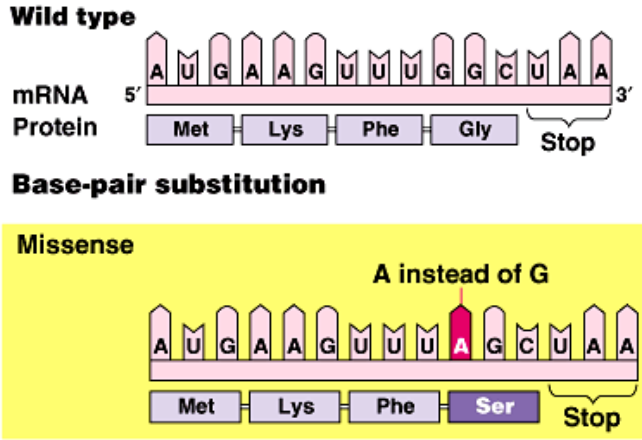
Point mutation in hemoglobin



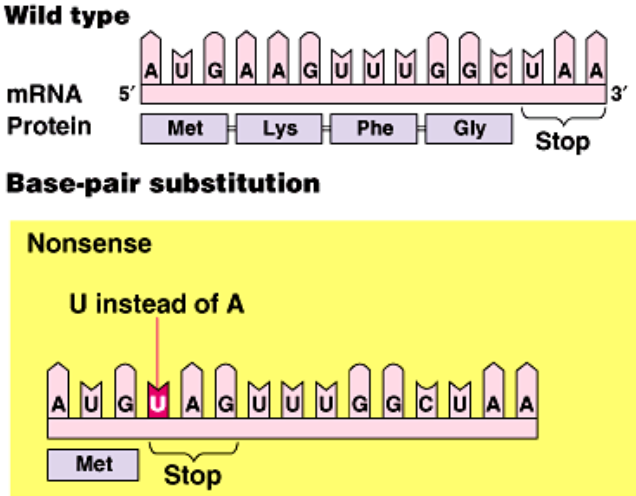
Base-pair Substitution



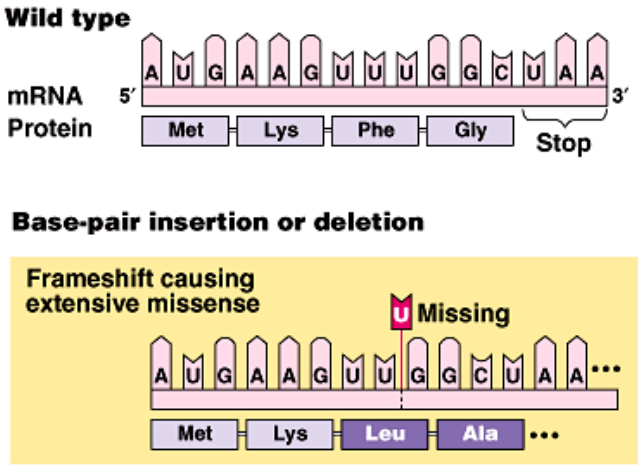
Base-pair Substitution



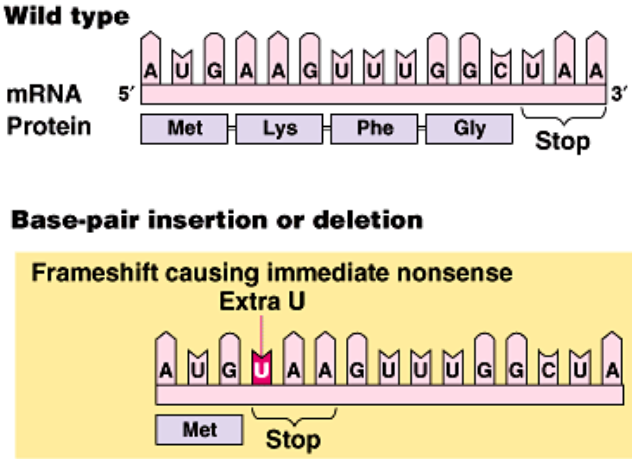
Base-pair Substitution



Base-Pair Insertion or Deletion



Base-Pair Insertion or Deletion



Base-Pair Insertion or Deletion

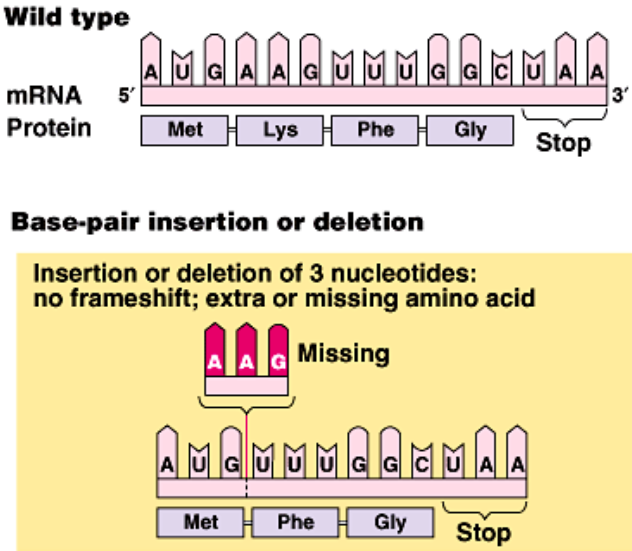
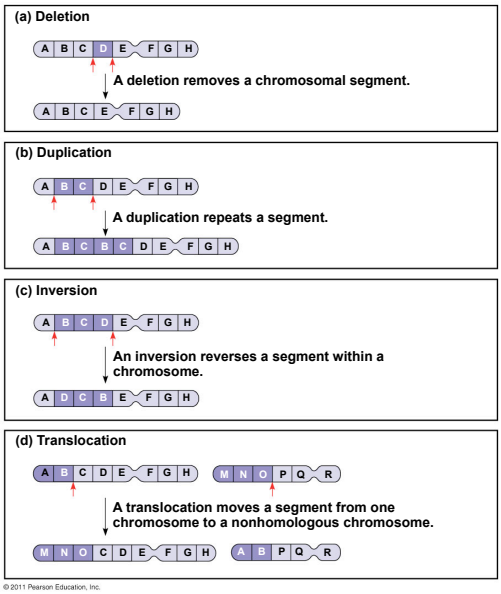
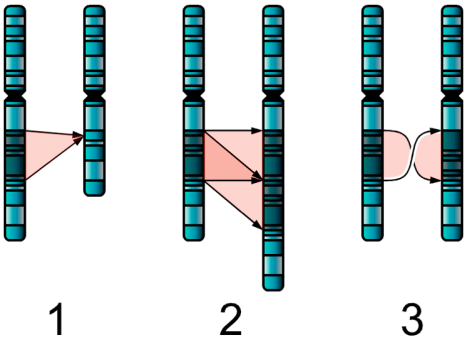


Figure 15.14



Chromosomal aberrations (large scale mutations)



The three major single chromosome mutations; deletion (1), duplication (2) and inversion (3).

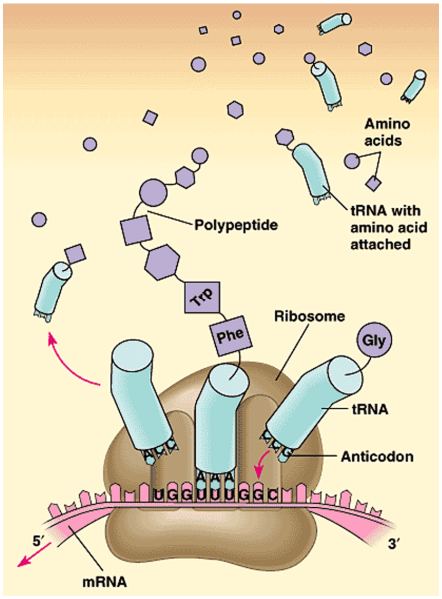
Down syndrome, usually is caused by an extra copy of chromosome 21 (trisomy 21).

6. Universal code

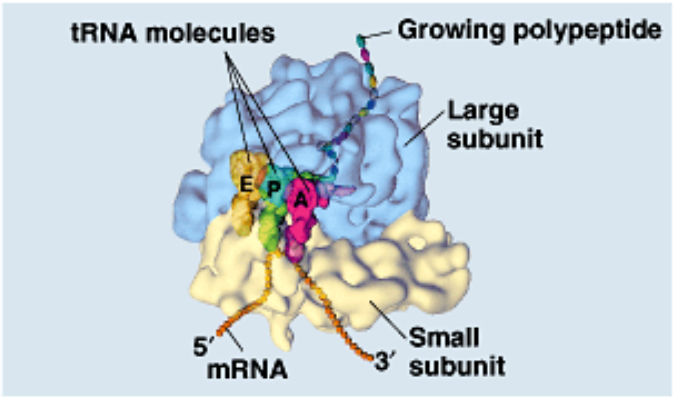
The triplet code is universal (almost) -- so all organisms can translate each other's genes.

So can use bacteria to make insulin...

Mechanics of Translation

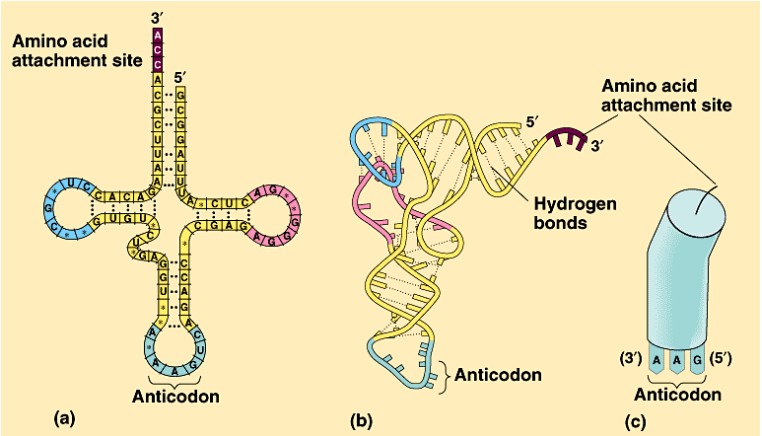


Ribosomes

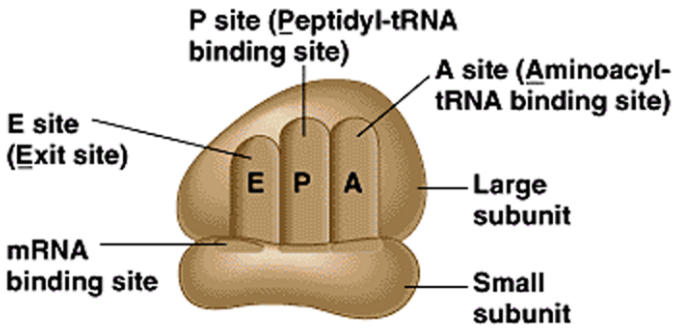


(a) Computer model of functioning ribosome

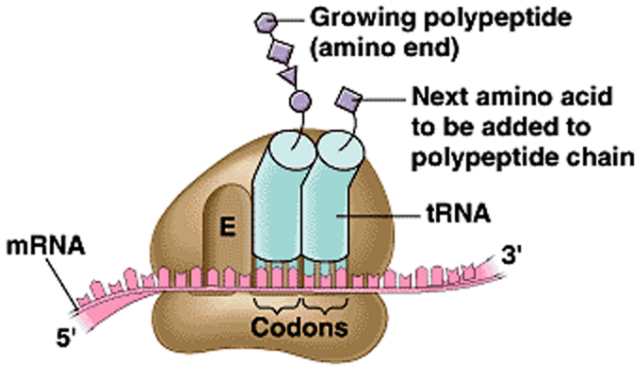
Transfer RNA (tRNA)



Ribosomes



Ribosomes



Translation

Initiation

1. Small ribosomal subunit binds to AUG start codon
2. Initiator tRNA (= Met-tRNA) binds to AUG.
3. Large ribosomal subunit binds Initiator tRNA in the P-site.

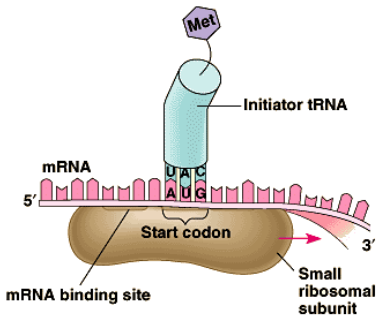
Elongation

4. Codon recognition: Incoming tRNA enters A site and matches its anticodon to next codon on the mRNA.
5. Peptide bond formation: growing polypetide chain is transferred to the new amino acid
6. Translocation: old tRNA in P site floats away, new tRNA moves from A site to P site, and ribosome moves down the mRNA by 1 codon.

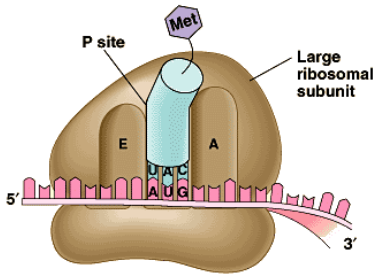
Termination

7. When the ribosome reaches the stop codon, the release factor enters the A site, cuts the polypeptide off the last tRNA, and the ribosome disassembles.

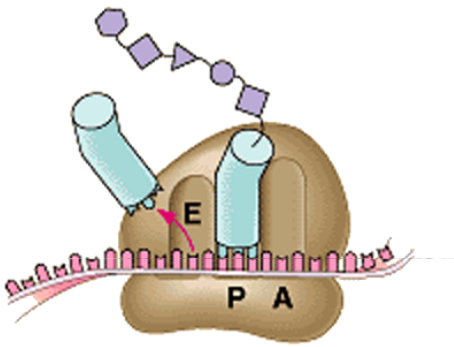
Translation: Initiation



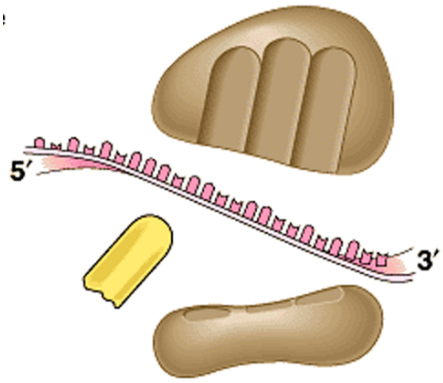
Translation: Initiation



Translation: Elongation

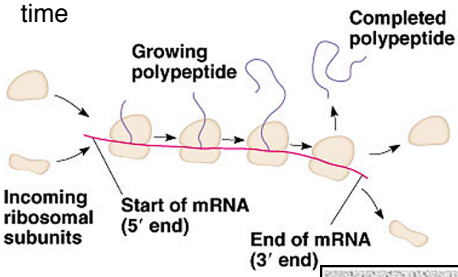


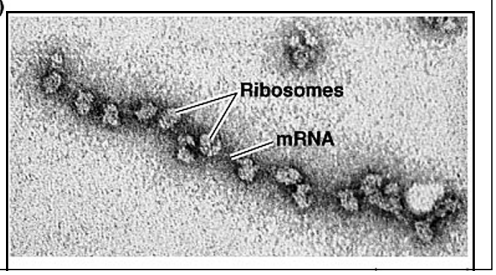
Translation: Termination



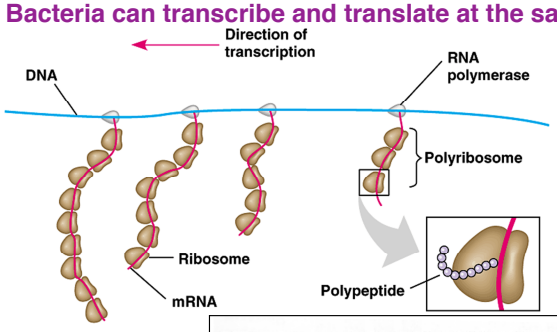
Polyribosomes

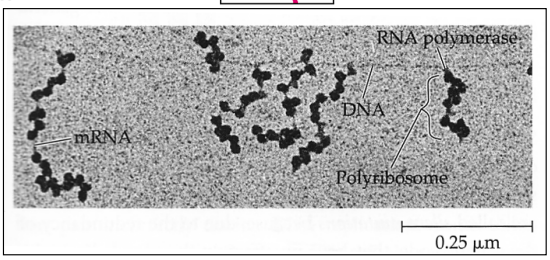
Multiple ribosomes can scan one mRNA strand at the same time



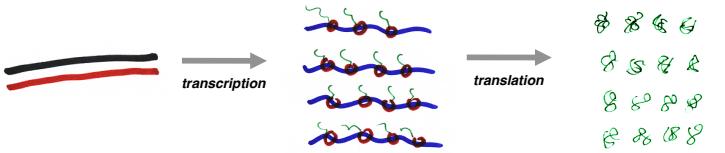


Bacteria can transcribe and translate at the same time:





Amplification



1x gene 1000x mRNA 100,000x protein

Transcription and translation represent amplification steps -- from one copy of gene to zillions of copies in protein.

Features of Translation:

1. **Uses tRNA to shuttle amino acids to ribosome**
   so there are at least 3 types of RNA in the cell:
   snRNA, tRNA, and mRNA

2. **All proteins start with Methionine because start codon is ATG (although Met can be cut off during protein processing)**
   so could label new proteins with radiolabeled Met

3. **Mutation could screw up start or stop codons**

4. **Translation is one way to control gene expression:**
   keep mRNA for longer/shorter times, block ribosomes from binding, etc.

5. **Target of antibiotics: block bacterial protein translation**

Antibiotics

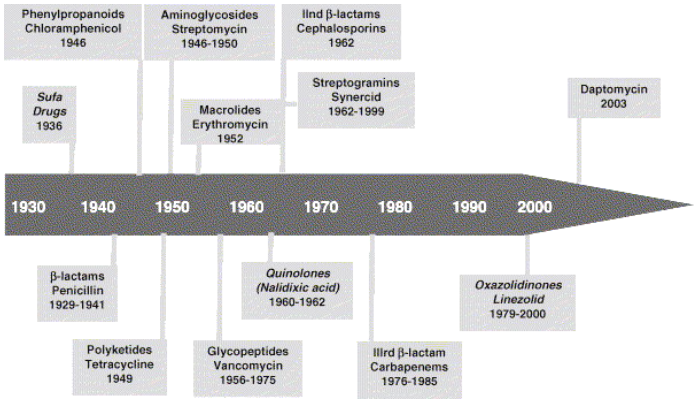
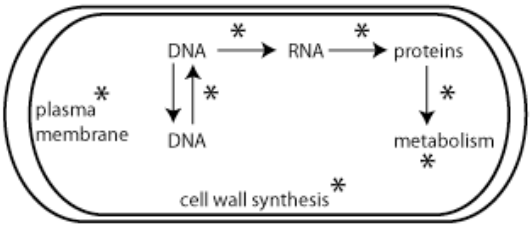


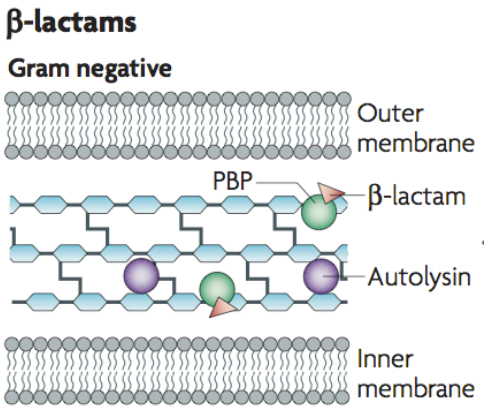
Fig. 2. Timeline of discovery novel classes of antibiotics and introduction in clinic.
Compounds listed in normal fonts are natural product derived and in italics are derived from synthetic origin.

Antibiotics

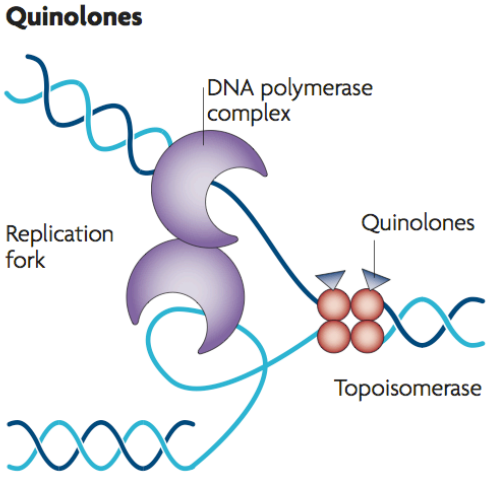


Antibiotics act at all of these points

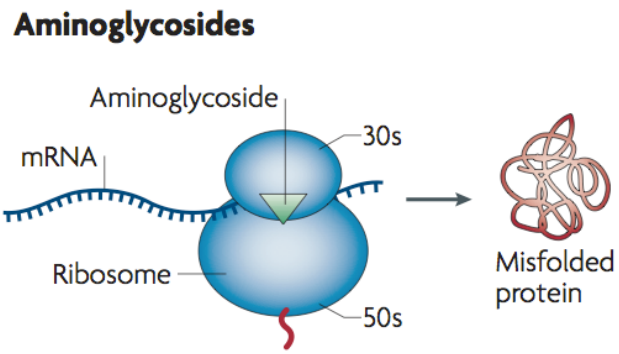
Pencillin et al.: Block bacterial cell wall synthesis at Pencillin binding proteins



DNA synthesis blockade



Interference with Translation



Nature Reviews Microbiology **8**, 423-435 (June 2010)

http://www.nature.com/nrmicro/journal/v8/n6/abs/nrmicro2333.html