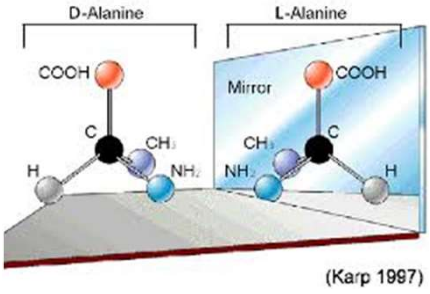
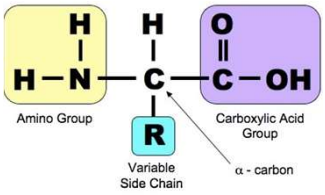
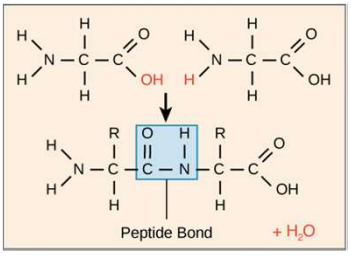
Amino Acids

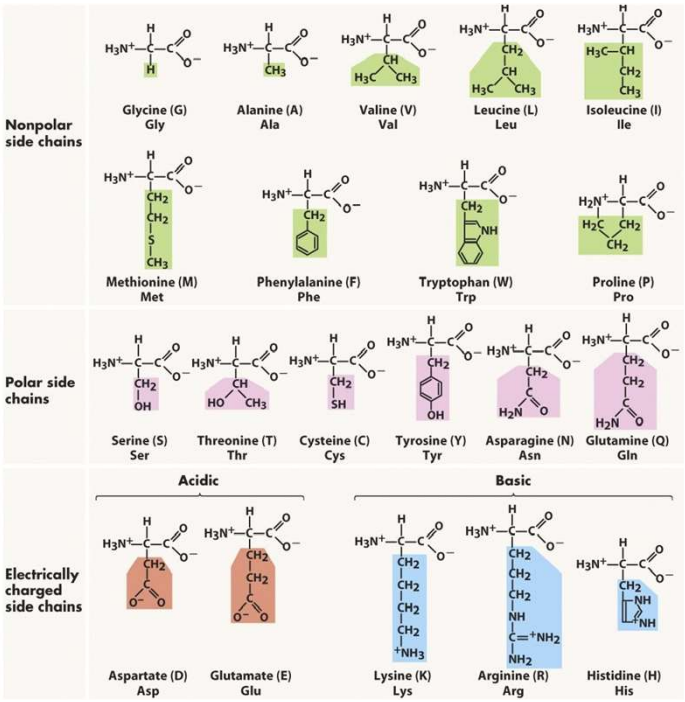


(Karp 1997)



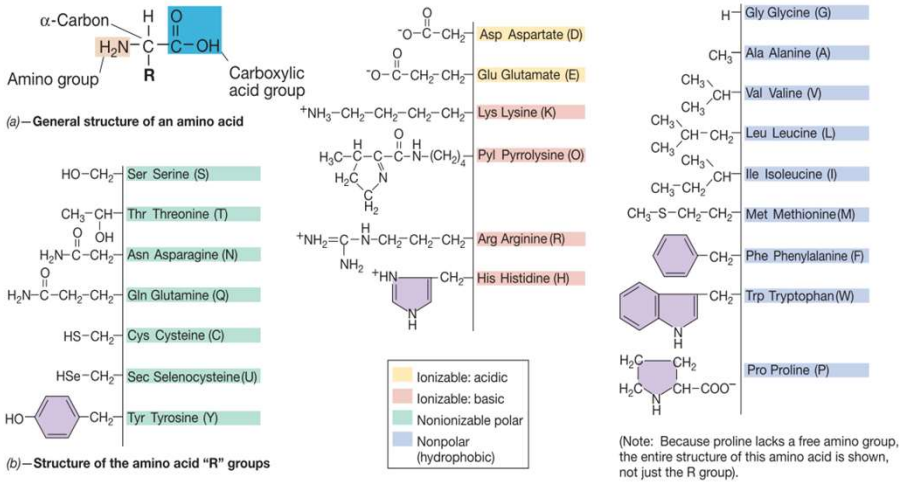


- ✓ Most natural proteins have L-amino acids
- ✓ Side chain R containing –COOH groups are acidic amino acids (aspartic acid, glutamic acid
- ✓ Those with hydrophobic side chains→ Nonpolar amino acids
- ✓ Cysteine contains sulfhydryl group→ two cysteine can form disulphide linkage



Structure of Genetically Encoded Amino Acids

- ✓ **Structure of the 22 genetically encoded amino acids**
- ✓ An average protein contains 300 amino acids → Enormous number of distinct proteins
- ✓ Polypeptide: C Terminus: Carboxylic acid & N Terminus: Amino group



Protein Synthesis: Translation

- Process involves formation of a polypeptide using a nucleic acid as template

- **Codon: An RNA triplet of three bases;** encodes a specific amino acid. The genetic code is RNA since, it is mRNA that is actually translated
- There are ~ 64 possible codons (4 bases taken three at a time = 4^3).
- Specific codons are used for starting and stopping translation

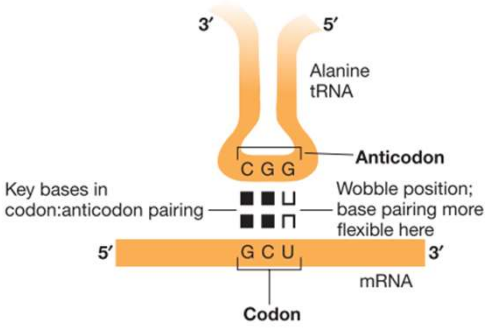
First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU, UUC Phe; UUA, UUG Leu	UCU, UCC, UCA, UCG Ser	UAU, UAC Tyr; UAA Stop; UAG Stop	UGU, UGC Cys; UGA Stop; UGG Trp	U C A G
C	CUU, CUC, CUA, CUG Leu	CCU, CCC, CCA, CCG Pro	CAU, CAC His; CAA, CAG Gln	CGU, CGC,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

Properties of the Genetic Code

- 64 codons in mRNA codes for 22 amino acids; Consequently, amino acids can be encoded by more than 1 codon
  - **Knowing codon at a given location unambiguously identifies the corresponding amino acid, the reverse is not true**
- *Degenerate code*: When many codons code for single amino acid
- A codon is recognized by specific base pairing with a complementary sequence of three bases called the **anticodon**, which is part of tRNAs;
- **A specific t-RNA would be required to recognize each codon**; *E coli* has separate t-RNAs for the 5 Leucine codons.
- Out of codons that encode the same amino acid→ organism-specific **codon bias seen;** some codons are preferred over others

The Wobble Concept

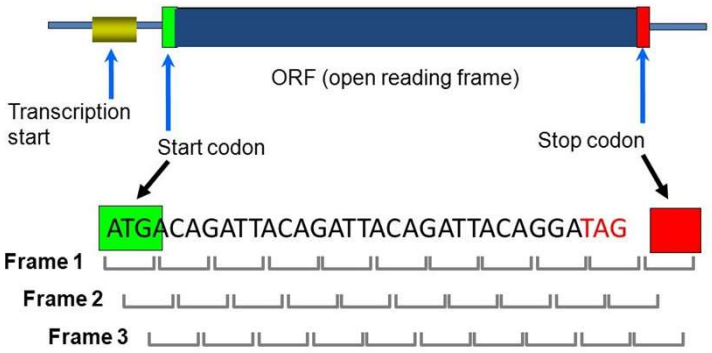
Wobble: tRNA molecules form standard base pairs at the first two positions of the codon while tolerating irregular base pairing at the third position



Some tRNAs can recognize more than one codon. There are two lysine codons in *E. coli*, there is only one lysyl tRNA, whose anticodon can base-pair with either AAA or AAG.

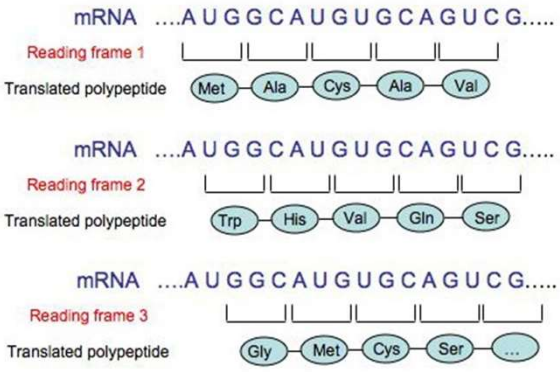
Open Reading Frames

- **Open reading frames (ORFs)**: If an mRNA can be translated, it contains a start codon (typically AUG) followed by codons and then a stop codon in the same reading frame as the start codon.
- Only ORFs long enough to encode a functional polypeptide are accepted as true coding sequences, while a few protein hormones and regulatory peptides are much shorter



Start and Stop Codons

- In bacteria mRNA is translated beginning with its **start codon** AUG which encodes a chemically modified methionine: *N-formylmethionine*
- AUG at the *beginning* encodes *N*-formylmethionine, while *within* the coding region encodes methionine. Two different tRNAs are involved
- *Archaea* and *Eukarya* insert a regular methionine as the first amino acid in a polypeptide
- It is essential for ribosome to find the correct start codon to begin translation, if it does not, the whole reading frame of the mRNA will be shifted and thus an entirely different protein will be made

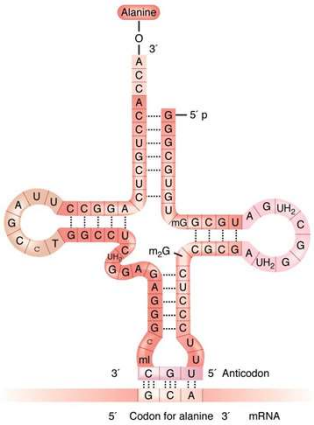


Start and Stop Codons

- **Reading frame fidelity: interactions between mRNA and rRNA within the ribosome**
- In prokaryotes, **rRNA recognizes a specific AUG on the mRNA as a start codon with the aid of an upstream sequence in the mRNA called the *ribosome-binding site* (RBS) or Shine–Dalgarno sequence**
- Some mRNAs can use other start codons, such as GUG, but still they incorporate ***N*-formylmethionine as the initiator amino acid**
- Few codons (**Stop codons, UAA, UAG, and UGA**) do not encode any amino acid, and they signal the termination of translation of a protein-coding sequence on the mRNA
- Stop codons are also called **nonsense codons**
  - Certain microorganism use certain nonsense codons to encode amino acids **(Example: animal mitochondria use the codon UGA to encode tryptophan instead of using it as a stop codon).**
  - Both selenocysteine and pyrrolysine (unusual and rare amino acids) are encoded by stop codons (UGA and UAG, respectively). **This switch is controlled by a recognition sequence just downstream of the now coding stop codon**

Transfer RNA

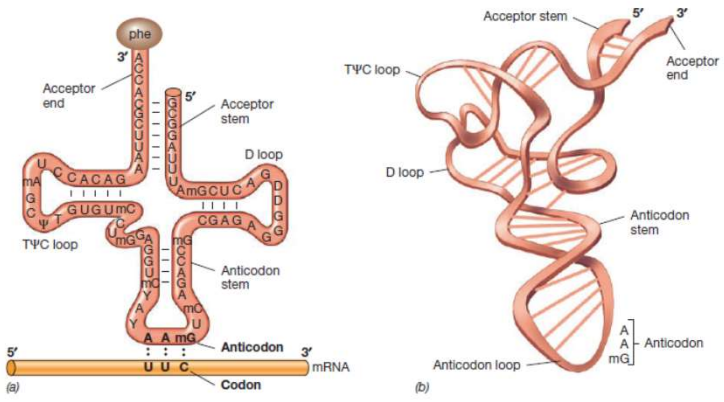
- A transfer RNA carries the **anticodon** that base-pairs with the codon on mRNA
- Each tRNA is specific for the amino acid that corresponds to its own anticodon **(its *cognate* amino acid)**
- **Aminoacyl-tRNA synthetases: Enzymes that ensure that each tRNA receives its correct amino acid**
  - For each amino acid, a separate aminoacyl-tRNA synthetase exists that specifically binds to both the amino acid and the tRNA with corresponding anticodons
  - they must recognize both a specific tRNA and its cognate amino acid
  - There are about 60 different tRNAs in bacterial cells and 100– 110 in mammalian cells



Structure of tRNA

- tRNA: Short, single-stranded molecules that contain extensive secondary structure; 73–93 nucleotides long
- Certain bases and secondary structures are constant for all tRNAs, whereas other parts are variable
- They contain some modified purine and pyrimidine bases: pseudouridine (ψ), inosine, dihydrouridine (D), ribothymidine, methyl guanosine, dimethyl guanosine, and methyl inosine
- The mature active tRNA **also contains extensive double-stranded regions within the molecule formed by internal base pairing** → tRNA can have a structure of cloverleaf
- Some regions of tRNA secondary structure are named after the modified bases found there **(for example, the TψC and D loops) or after their functions (for example, the anticodon loop and acceptor stem).**

Structure of tRNA



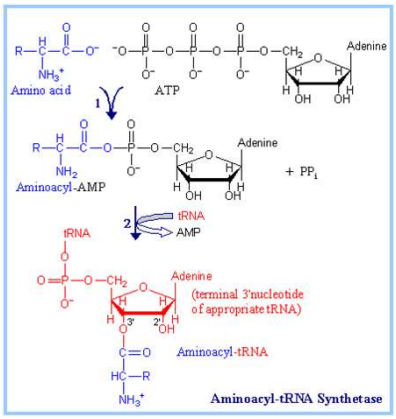
The Anticodon & Amino Acid Binding Site

- *Anticodon*: the group of 3 bases that recognizes the codon on the mRNA
- Other portions of the tRNA interact with both the rRNA and ribosomal protein, non-ribosomal translation protein & the **aminoacyl-tRNA synthetase enzyme**
- At the **3'end (acceptor stem)** of all tRNAs are three unpaired nucleotides (CCA), which is essential for its function
- In most organisms the 3' CCA is not encoded in the tRNA gene on the chromosome, each nucleotide is added one by one by a protein called *CCA adding enzyme*, using CTP and ATP as substrates
- **The cognate amino acid is then covalently attached to the terminal adenosine of the CCA end of its corresponding tRNA by an ester linkage to the ribose sugar→ This is an activated tRNA**

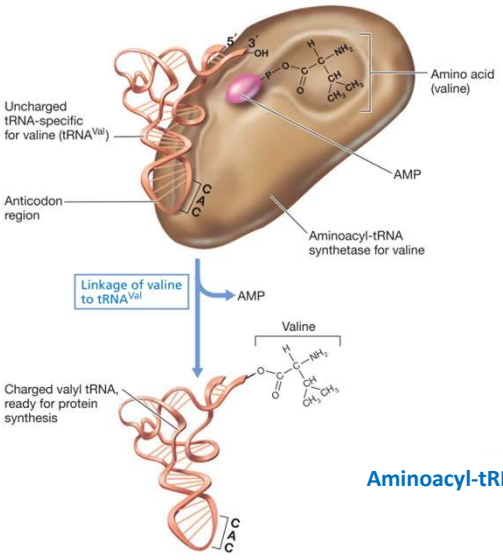
Recognition, Activation & Charging

- The anticodon of the tRNA is important for recognition by the synthetase
- The other key recognition nucleotides are often part of the acceptor stem or D loop of the tRNA

- The **aminoacyl-AMP intermediate** formed normally remains bound to the tRNA synthetase until collision with the appropriate tRNA molecule
- The pyrophosphate (PPi) formed in the first reaction is split by a pyrophosphatase, giving two molecules of inorganic phosphate
- After activation and charging the aminoacyl-tRNA leaves the synthetase until it is bound by a ribosome where actual polypeptide synthesis occurs



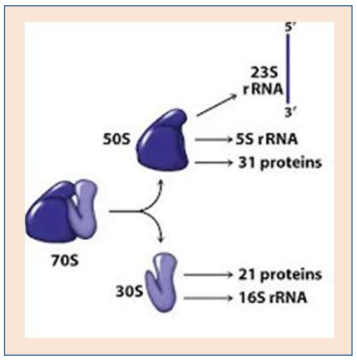
Charging of tRNA



Aminoacyl-tRNA synthetase enzyme

Ribosomes in Prokaryotes

- **Ribosomes** are the sites of protein synthesis
- Each ribosome consists of two subunits, 30S & 50S subunits that yield the intact 70S ribosomes
- The S-values are *Svedberg units*→ sedimentation coefficients of the subunits or intact ribosomes when subjected to centrifugal force in an ultracentrifuge
- The ribosome subunits alternately associate and dissociate and also interact with many other proteins
- Several proteins (translation factors) that are essential for ribosome function interact with the ribosome at various stages of translation



Initiation of Translation

- **Initiation complex forms consisting of the 30S subunit, mRNA, formylmethionine tRNA, and several initiation proteins called IF1, IF2, and IF3, and GTP**
- Next, a 50S ribosomal subunit is added to the initiation complex to form the active 70S ribosome. At the end, the ribosome separates again into 30S and 50S subunits
- **Upstream of the start codon, there are three to nine nucleotides referred as the ribosome-binding site (RBS). It helps bind the mRNA to the ribosome**. The 3' end of the 16S rRNA has complementary base sequences and binds here
- Base pairing between these two molecules holds the ribosome– mRNA complex securely together in the correct reading frame
- Polycistronic mRNA has multiple RBS sequences, one upstream of each coding sequence
- Translational initiation always begins with a special initiator aminoacyl-tRNA binding to the start codon, AUG
- After polypeptide completion, the formyl group is removed and the completed protein will be methionine. In many proteins this methionine is removed by a specific protease

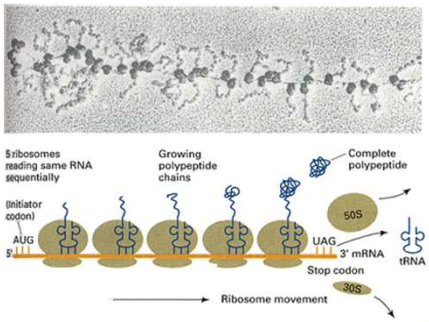
Elongation, Translocation & Termination

- The ribosome contains other sites where the tRNAs interact, located primarily on the 50S subunit
- **A site: the acceptor site on the ribosome is where the incoming charged tRNA first attaches**
- **P site: The peptide site on the ribosome is the site where the growing polypeptide chain is held by the previous tRNA**
- **The growing polypeptide chain moves to the tRNA at the A site as a new peptide bond is formed**
- Several nonribosomal proteins are required: Elongation factors, EF-Tu and EF-Ts and GTP
- Following elongation, the tRNA holding the polypeptide is translocated from A to P, thus opening the A for another charged tRNA, assisted by EF-G and one GTP for each translocation event
- **At each translocation step the ribosome advances three nucleotides, exposing a new codon at the A site**

Elongation, Translocation & Termination

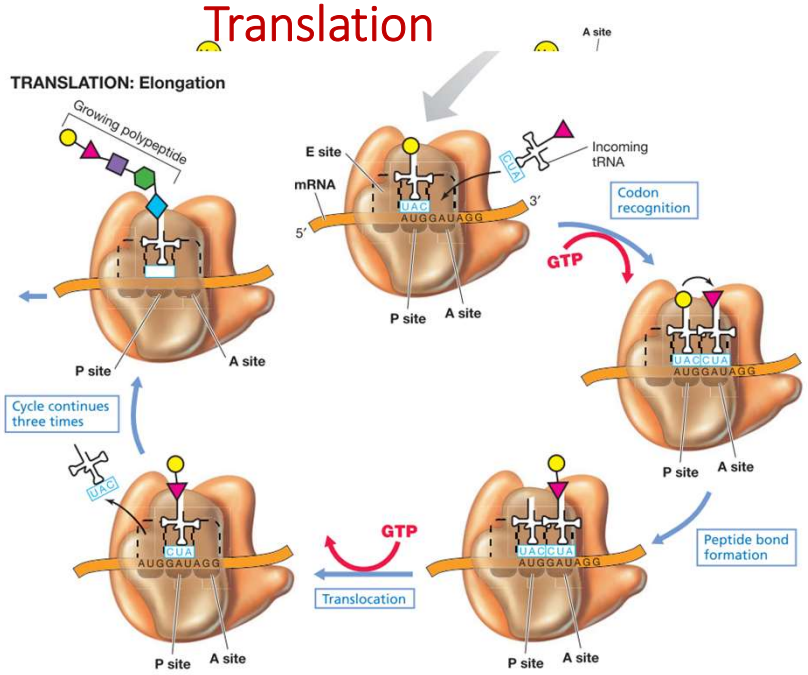
- **Translocation pushes the now empty tRNA to a third site, E site**. From E site tRNA is actually released from the ribosome.
- The ribosome must move *exactly* one codon at each step.

- Several ribosomes can translate a single mRNA molecule simultaneously, forming a complex called a *polysome.*
- Each ribosome in a polysome complex makes a complete polypeptide, thus increasing the speed and efficiency of translation.



- Protein synthesis terminates when the ribosome reaches a stop codon (nonsense codon). *Release factors* (RFs) recognize the stop codon and cleave the attached polypeptide from the final tRNA, releasing the finished product.

Translation



Role of Ribosomal RNA

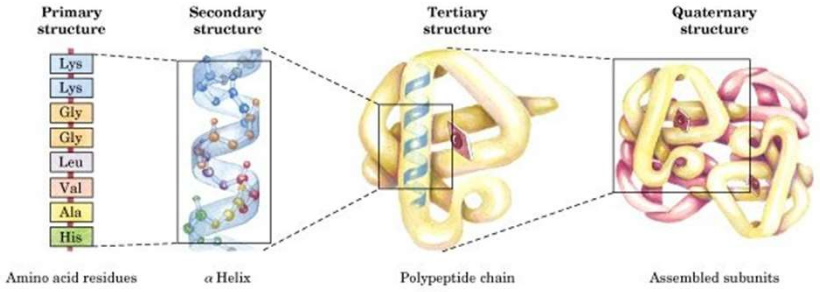
- 16S rRNA cause initiation through base pairing with the RBS on the mRNA
- The codons in the A and P sites, in the mRNA is held in position by binding to 16S rRNA and ribosomal proteins
- rRNA helps in ribosome subunit association, as well as in positioning tRNA in the A and P sites on the ribosome
- tRNAs are also bound to the ribosome by interactions of the anticodon stem–loop of the tRNA with specific sequences within 16S rRNA
- The acceptor end of the tRNA base-pairs with sequences in 23S rRNA
- **The actual formation of peptide bonds is catalyzed by rRNA**. The peptidyl transferase reaction occurs on the 50S subunit of the ribosome and is catalyzed by the 23S rRNA
- The 23S rRNA helps in translocation, and the EF proteins interact specifically with 23S rRNA

Freeing Trapped Ribosomes

- **Trapped ribosome:** If a ribosome reaches the end of an mRNA molecule and there is no stop codon, the ribosome cannot be released from the mRNA
- In bacteria, a small RNA molecule, *tmRNA*, frees stalled ribosomes
- tmRNA mimics both a tRNA (it carries the amino acid alanine) and mRNA (contains a short stretch of RNA that can be translated)
- When tmRNA collides with a stalled ribosome, it binds alongside the defective mRNA
- The tmRNA contains a stop codon that allows release factor to bind and disassemble the ribosome
- The protein made is defective and is subsequently degraded. This is accomplished by a short sequence of amino acids encoded by tmRNA and added to the end of the defective protein; the sequence is a signal for a specific protease to degrade the protein

Protein Folding

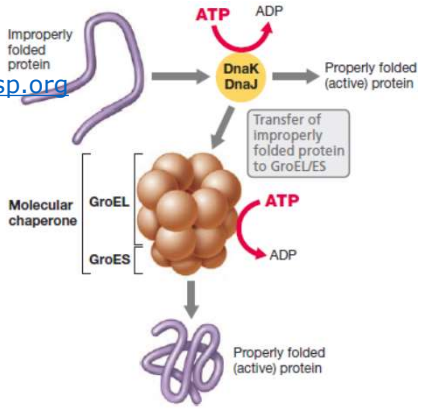
- For a protein to function properly it must fold correctly after it is synthesized and end up in the correct location
- Many proteins exist inside the cell, some must be transported to the periplasm or into the inner or outer membrane to facilitate processes such as ion, sugar, and electron transport
- Proteins such as toxins and extracellular enzymes (exoenzymes) must be secreted outside the cell



Chaperonins Assisted Protein Folding

- **Chaperonins** / Molecular chaperones: Assist folding in other polypeptides
  - **Chaperonins in *E. coli:* Proteins DnaK,DnaJ, GroEL, and GroES**
- The chaperonins themselves do not become part of the assembly, they only assist in folding
- They also help to prevent improper aggregation of proteins
- Chaperonins can also refold proteins that have partially denatured in the cell
- Chaperonins are one type of *heat shock protein*, and their synthesis is greatly accelerated when a cell is stressed by excessive heat
- The heat shock response is done to refold its partially denatured proteins for reuse before proteases destroy them

bschorr@aeesp.org



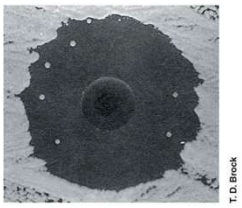
Mutation

- **A mutation is an in*heritable* change in the base sequence of the genome → a change that is passed on to the progeny cells**
- Although rate of spontaneous mutation is low, due to the fast growth rate of bacteria mutations can accumulate fast
- **Recombination involves physical exchange of DNA between *genetic elements* (chromosomes)**
- Mutation brings about very small amount of genetic change in a cell as compared to genetic recombination
- A strain of any cell or virus carrying a change in nucleotide sequence is called a **mutant**. A mutant differs from its parental strain in its **genotype (the nucleotide sequence of the genome)**
- **Phenotype (the observable properties of the mutant)** may also be altered → *mutant phenotype*
- **Wild-type strain**: strain isolated from nature

Selectable vs Non-selectable Mutant

- *Selectable*, confer some advantage on organisms possessing them under certain environmental conditions, so the progeny of the mutant cell are able to outgrow and replace the parent
  - ✓ Example is drug resistance: An antibiotic-resistant mutant can grow in the presence of an antibiotic that inhibits or kills the parent
  - ✓ It is relatively easy to detect and isolate selectable mutants (**Selection)**
- *Non-selectable*, they may lead to a very clear change in the phenotype of an organism
  - ✓ Example: Color loss in a pigmented organism. Non-pigmented cells usually have neither an advantage nor a disadvantage
  - ✓ Such mutations can be examined by observing a large numbers of colonies and looking for the "different" ones (**Screening**)

Selectable vs Non-selectable Mutant



(a)

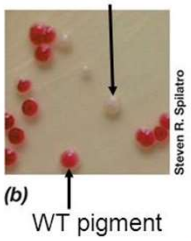
Selectable mutants: Antibiotic resistance colonies can be detected around a zone of clearance created by the inhibition of a sensitive bacterium

(b)

Nonselectable mutants: *Aspergilus nidulans* produces different interchangeable spontaneously.

Colonies of mutants of a species of ***Halobacterium***

Nonselectable mutants in *Serratia marcescens*



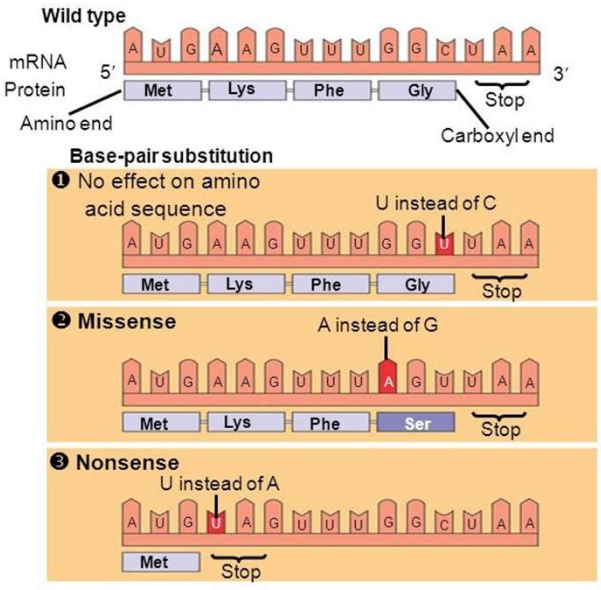
Induced or Spontaneous Mutations

- **Induced mutations:** those that are due to agents in the environment and made deliberately by humans
  - ✓ They can result from exposure to radiation (cosmic rays) → radiation alters the structure of bases in DNA
  - ✓ A variety of chemicals including oxygen radicals can chemically modify DNA; oxygen radicals converts guanine into 8-hydroxyguanine
- **Spontaneous mutations:** those that occur without external intervention
  - ✓ The bulk of spontaneous mutations result from occasional errors in the pairing of bases by DNA polymerase during DNA replication

Base-Pair substitutions

- Mutations that change only one base pair are called **point mutations**
  - Point mutation in a gene encoding a polypeptide→ amino acid sequence in the polypeptide will change → change in phenotype of the cell
  - Error in the DNA is transcribed into mRNA, and the erroneous mRNA is translated to yield a polypeptide
  - The genetic code is degenerate, hence not all mutations in the base sequence encoding a polypeptide will change the polypeptide
- **Silent mutation:** mutation which does not affect the sequence of the encoded polypeptide (no effect on phenotype)
  - Almost always it occurs in the third base of the codon
  - arginine and leucine can also have silent mutations in the first position

Base-Pair Substitutions

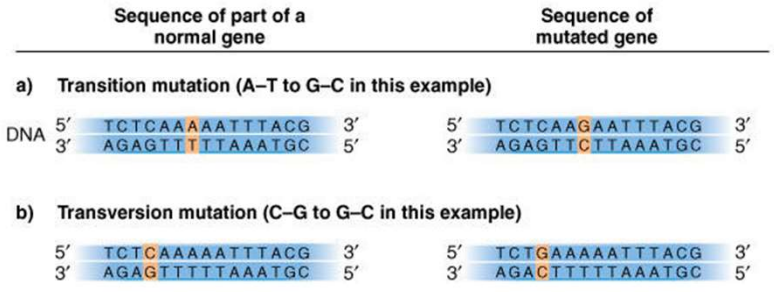


Base-Pair Substitutions

- **Missense mutation: Changes in the first or second base of the codon more often lead to significant changes in the amino acid sequence of the polypeptide**
  - If change is at a critical location in the polypeptide chain, the protein may be inactive or may have reduced activity
    - depends on where the substitution lies in the polypeptide chain and on how it affects protein folding and activity; mutations in the active site of an enzyme are more likely to destroy activity
  - Not all missense mutations necessarily lead to nonfunctional proteins
- **Nonsense mutation: formation of a nonsense (stop) codon. This results in premature termination of translation, leading to an incomplete polypeptide**
  - Unless the nonsense mutation is very near the end of the gene, the product is considered *truncated* or incomplete
  - Truncated proteins are completely inactive or they lack normal activity

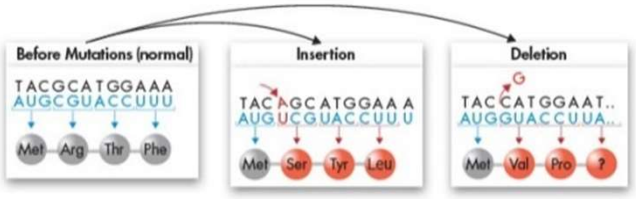
Transitions and Transversions

- Transitions are mutations in which one purine base (A or G) is substituted for another purine, or one pyrimidine base (C or T) is substituted for another pyrimidine
- Transversions are point mutations in which a purine base is substituted for a pyrimidine base, or vice versa



Frameshifts: Insertion or Deletion

- Any deletion or insertion of a single base pair results in a shift in the reading frame- **frameshift mutations**
  - Single base insertions / deletions change the primary sequence of the encoded polypeptide
  - Insertion or deletion of three base pairs adds or removes a whole codon. This results in addition or deletion of a single amino acid in the polypeptide sequence
- Many large insertion mutations are due to the insertion of specific identifiable DNA sequences called *transposable elements*



Reversions and Mutation Rates

- Point mutations are typically reversible, a process known as **reversion**. **A revertant is a strain in which the original phenotype is restored by a second mutation**
  - ✓ In ***same site revertants***, the mutation that restores activity is at the same site as the original mutation
  - ✓ If the back mutation is not only at the same site but also restores the original sequence, it is called a ***true revertant***
  - ✓ In ***second-site revertants***, the mutation is at a different site in the DNA. Second-site mutations can restore a wild-type phenotype if they function as ***suppressor mutations*—mutations that compensate for the effect of the original mutation**
- For most bacteria, errors in DNA replication occur at a frequency of 10^{-6} to 10^{-7} per thousand bases during a single round of replication
- Eukaryotes with very large genomes tend to have replication error rates about 10-fold lower than typical bacteria

Mutation

- DNA viruses, especially those with very small genomes, may have error rates 100-fold to 1000-fold higher than those of cellular organisms
- RNA viruses have even higher error rates due to less proofreading and the lack of RNA repair mechanisms
- **Single base errors during DNA replication are more likely to lead to missense mutations than to nonsense mutations since most single base substitutions yield codons that encode other amino acids**
- The next most frequent type of codon change is single base change leads to a silent mutation
- **A given codon can be changed to any of 27 other codons by a single base substitution, and on average, about two of these will be silent mutations, about one a nonsense mutation, and the rest will be missense mutations**
- There are also some DNA sequences, typically sequences containing short repeats, that are hot spots for mutations because the error frequency of DNA polymerase is relatively high there

Mutagenesis

A variety of chemical, physical, and biological agents can increase the mutation rate and agents are called **mutagens**

Agent	Action	Result
Base analogs		
5-Bromouracil	Incorporated like T; occasional faulty pairing with G	AT pair → GC pair; occasionally GC → AT
2-Aminopurine	Incorporated like A; faulty pairing with C	AT → GC; occasionally GC → AT
Chemicals reacting with DNA		
Nitrous acid (HNO_2)	Deaminates A and C	AT → GC and GC → AT
Hydroxylamine (NH_2OH)	Reacts with C	GC → AT
Alkylating agents		
Monofunctional (for example, ethyl methane sulfonate)	Puts methyl on G; faulty pairing with T	GC → AT
Bifunctional (for example, nitrogen mustards, mitomycin, nitrosoguanidine)	Cross-links DNA strands; faulty region excised by DNase	Both point mutations and deletions
Intercalative dyes		
Acridines, ethidium bromide	Inserts between two base pairs	Microinsertions and microdeletions
Radiation		
Ultraviolet	Pyrimidine dimer formation	Repair may lead to error or deletion
Ionizing radiation (for example, X-rays)	Free-radical attack on DNA, breaking chain	Repair may lead to error or deletion