Dept. of Computer Science and Engineering
University of Rajshahi
www.ru.ac.bd

Dr. Shamim Ahmad

Median polish Algorithm

The median polish is an exploratory data analysis procedure proposed by the statistician John Tukey.

Data Normalization

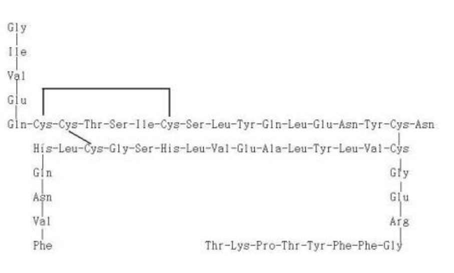
Protein synthesis

- Messenger RNA (mRNA) molecules direct the assembly of proteins on ribosomes.
- Transfer RNA (tRNA) molecules are used to to deliver amino acids to the ribosome
- Ribosomal RNA (rRNA) then links amino acids together to form proteins.

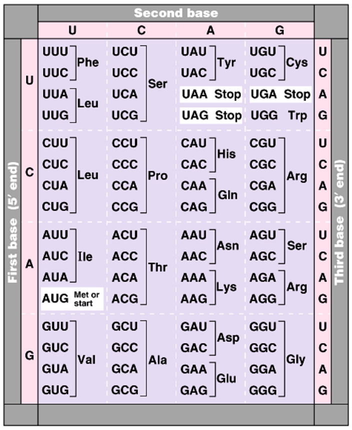
Insulin

Molecular Formula

$C_{256}H_{387}N_{65}O_{79}S_6$



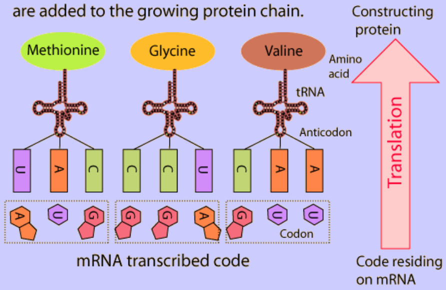
<u>Codon</u> = 3 letter section of mRNA that codes for one amino acid



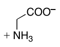
First base (5′ end)	Second base U	Second base C	Second base A	Second base G	Third base (3′ end)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
U	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
U	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
C	CUC Leu	CCC Pro	CAC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

Amino Acid	SLC	DNA codons
Isoleucine	I	ATT, ATC, ATA
Leucine	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	V	GTT, GTC, GTA, GTG
Phenylalanine	F	TTT, TTC
Methionine	M	ATG
Cysteine	C	TGT, TGC
Alanine	A	GCT, GCC, GCA, GCG
Glycine	G	GGT, GGC, GGA, GGG
Proline	P	CCT, CCC, CCA, CCG
Threonine	T	ACT, ACC, ACA, ACG
Serine	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Y	TAT, TAC
Tryptophan	W	TGG
Glutamine	Q	CAA, CAG
Asparagine	N	AAT, AAC
Histidine	H	CAT, CAC
Glutamic acid	E	GAA, GAG
Aspartic acid	D	GAT, GAC
Lysine	K	AAA, AAG
Arginine	R	CGT, CGC, CGA, CGG, AGA, AGG
Stop codons	Stop	TAA, TAG, TGA

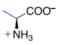
Amino acids corresponding to the codons are added to the growing protein chain.



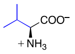
α-Amino acids
are classified by the properties of their sidechains.*Nonpolar*:



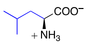
Glycine *(Gly, G)*



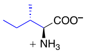
(S)-(+)-Alanine *(Ala, A)*



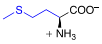
(S)-(+)-Valine *(Val, V)*



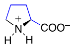
(S)-(–)-Leucine *(Leu, L)*



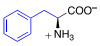
(2S,3S)-(+)-Isoleucine *(Ile, I)*



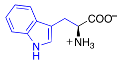
(S)-(–)-Methionine *(Met, M)*



(S)-(–)-Proline *(Pro, P)*

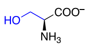


(S)-(–)-Phenylalanine *(Phe, F)*



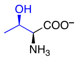
(S)-(–)-Tryptophan *(Trp, W)*

Polar but non-ionizable:



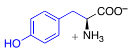
(S)-(–)-Serine *(Ser, S)*

pKa ~ 13



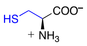
(2S,3R)-(–)-Threonine *(Thr, T)*

pKa ~ 13



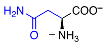
(S)-(–)-Tyrosine *(Tyr, Y)*

pKa ~ 10.1

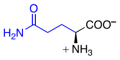


(R)-(–)-Cysteine *(Cys, C)*

pKa ~ 8.2

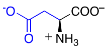


(S)-(–)-Asparagine *(Asn, N)*



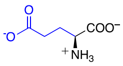
(S)-(+)-Glutamine *(Gln, Q)*

Acidic:



(S)-(+)-Aspartic Acid *(Asp, D)*

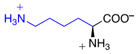
pKa ~ 3.6



(S)-(+)-Glutamic Acid *(Glu, E)*

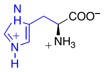
pKa ~ 4.2

Basic:



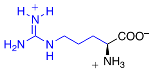
(S)-(+)-Lysine *(Lys, K)*

pKa ~ 10.5



(S)-(–)-Histidine *(His, H)*

pKa ~ 6.0

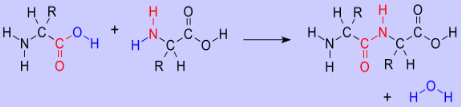


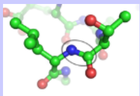
(S)-(+)-Arginine *(Arg, R)*

pKa ~ 12.5

Peptide bond

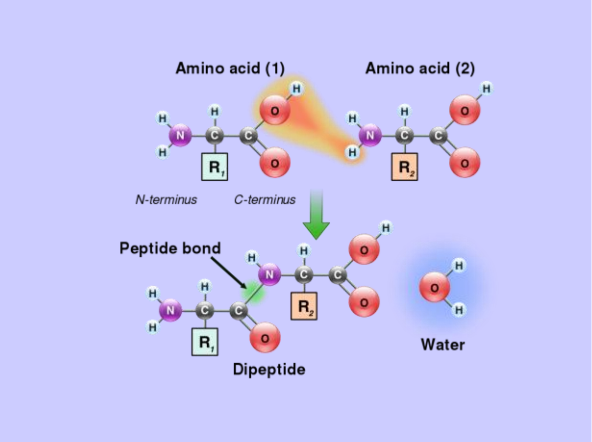
Covalent Chemical Bond

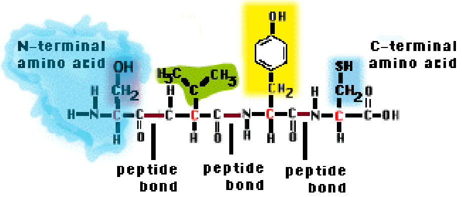


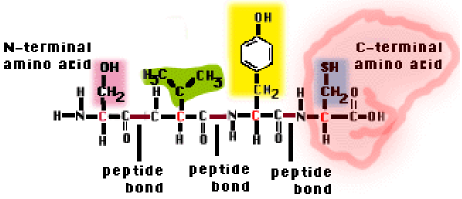


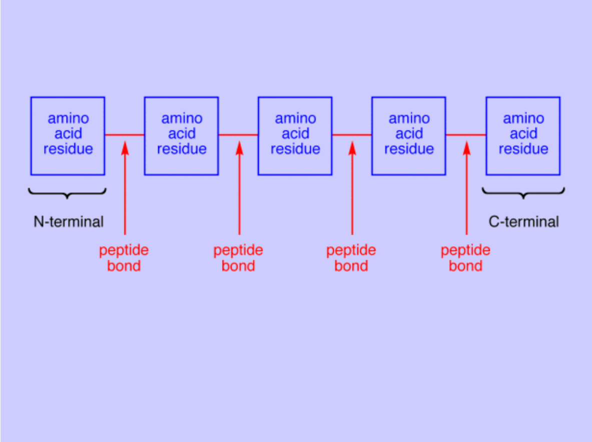
- The formation of the peptide bond consumes **energy**
- Which, in living systems, is derived from **ATP**

- A peptide bond can be broken by **hydrolysis** (the addition of water).
- They can release **8–16 kilojoule/mol** of free energy








amino acid residue

amino acid residue

amino acid residue

amino acid residue

amino acid residue

N-terminal

C-terminal

peptide bond

peptide bond

peptide bond

peptide bond

Definitions of the components:

Part 1 – Overview of PTMs

2. **Protein translation:** The process by which the mRNA template is read by ribosomes to synthesize the corresponding protein molecule on the basis of the three letter codons, which code for specific amino acids.

3. **Cytosol:** A cellular compartment that serves as the site for protein synthesis.

4. **Signal sequence:** A sequence that helps in directing the newly synthesized polypeptide chain to its appropriate intracellular organelle. This sequence is most often cleaved following protein folding and PTM.

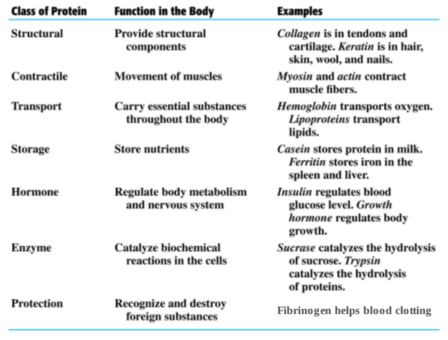
Definitions of the components:

Part 1 – Overview of PTMs

5. **Endoplasmic reticulum:** A membrane-bound cellular organelle that acts as a site for post-translational modification of the newly synthesized polypeptide chains.

6. **Cleaved protein:** The protein product obtained after removal of certain amino acid sequences such as N- or C-terminal sequences, signal sequence etc.

Function of proteins



Class of Protein	Function in the Body	Examples
Structural	Provide structural components	*Collagen* is in tendons and cartilage. *Keratin* is in hair, skin, wool, and nails.
Contractile	Movement of muscles	*Myosin* and *actin* contract muscle fibers.
Transport	Carry essential substances throughout the body	*Hemoglobin* transports oxygen. *Lipoproteins* transport lipids.
Storage	Store nutrients	*Casein* stores protein in milk. *Ferritin* stores iron in the spleen and liver.
Hormone	Regulate body metabolism and nervous system	*Insulin* regulates blood glucose level. *Growth hormone* regulates body growth.
Enzyme	Catalyze biochemical reactions in the cells	*Sucrase* catalyzes the hydrolysis of sucrose. *Trypsin* catalyzes the hydrolysis of proteins.
Protection	Recognize and destroy foreign substances	Fibrinogen helps blood clotting

Proteins

- Proteins account for 50% of the dry weight of the human body.

- Unlike lipids and carbohydrates, proteins are not stored, so they must be consumed daily.

- Current recommended daily intake for adults is 0.8 grams of protein per kg of body weight (more is needed for children).

- Dietary protein comes from eating meat and milk.

Proteins

100,000 different proteins in human body

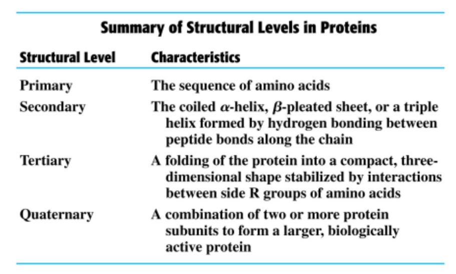
Fibrous proteins:

Insoluble in water – used for structural purposes (Keratin & Collagen).

Globular proteins:

More or less soluble in water – used for nonstructural purposes.

Summary of protein Structure

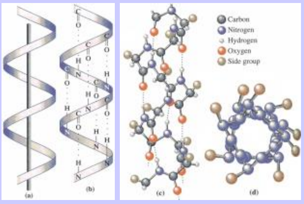


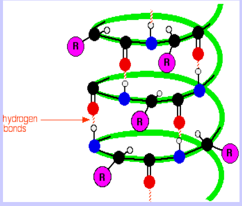
Summary of Structural Levels in Proteins

Structural Level	Characteristics
Primary	The sequence of amino acids
Secondary	The coiled α-helix, β-pleated sheet, or a triple helix formed by hydrogen bonding between peptide bonds along the chain
Tertiary	A folding of the protein into a compact, three-dimensional shape stabilized by interactions between side R groups of amino acids
Quaternary	A combination of two or more protein subunits to form a larger, biologically active protein

Alpha Helix

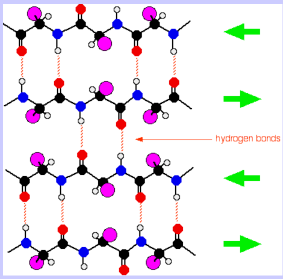
- The **Alpha Helix** The alpha helix (α-helix) structure resembles a coiled helical spring, with the coil configuration maintained by hydrogen bonds between N – H and C= O groups of every fourth amino acid

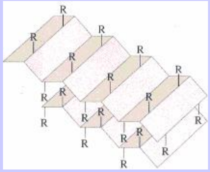




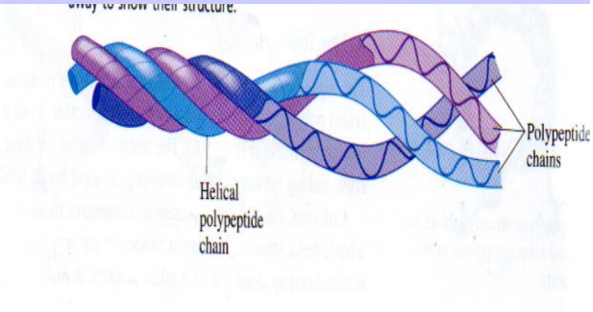
Beta pleated sheet

- The **beta pleated sheet** (β-pleated sheet) secondary structure involves amino acid chains that are almost completely extended.





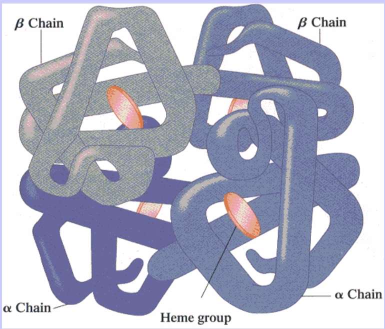
Tertiary structure



Quaternary structure

- **Quaternary structure** is the highest level of protein organization. It is found only in proteins that have **structures involving two or more polypeptide chains that** are independent of each other — that is, are not **covalently bonded to each other.** These multichain proteins are often called oligomeric proteins. The quaternary structure of a protein involves the associations among the separate chains in an oligomeric protein.

Hemoglobin



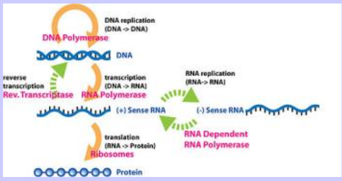
The Central Dogma was first stated by Francis Crick in 1958:

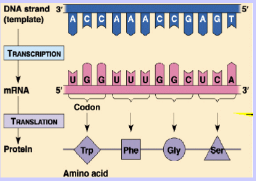
- "This states that once 'information' has passed into protein it cannot get out again.
- In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible,
- But transfer from protein to protein, or from protein to nucleic acid is impossible.
- Information means here the precise determination of sequence, either of **bases** in the **nucleic acid** or of **amino acid residues** in the **protein.**"

Restated in Nature paper published in 1970

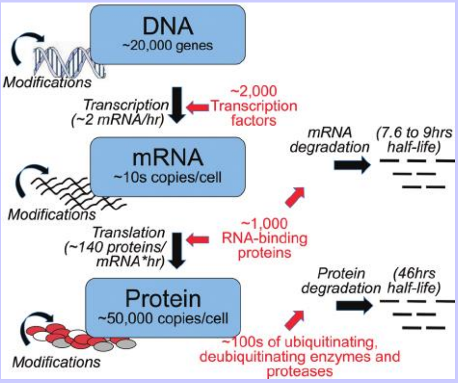
> The central dogma of <u>molecular biology</u> deals with the detailed residue-by-residue transfer of **sequential information**. It states that such information cannot be transferred back from protein to either protein or nucleic acid.
>
> — Francis Crick

Central Dogma





Central Dogma



General Transfers

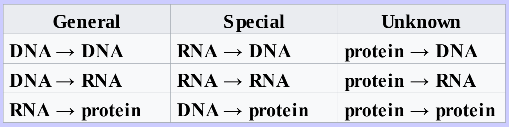
- DNA can be copied to DNA (DNA replication)
- DNA information can be copied into mRNA (transcription)
- Proteins can be synthesized using the information in mRNA as a template (translation).

Special Transfer

- RNA being copied from RNA (RNA replication, **Many Virus**)
- DNA being synthesised using an RNA template (**reverse transcription, HIV Virus)**
- Proteins being synthesised directly from a DNA template without the use of mRNA (**In Test Tube,**
- **E. coli Virus.Ribosome, Mouse.DNA)**)

Unknown Transfers

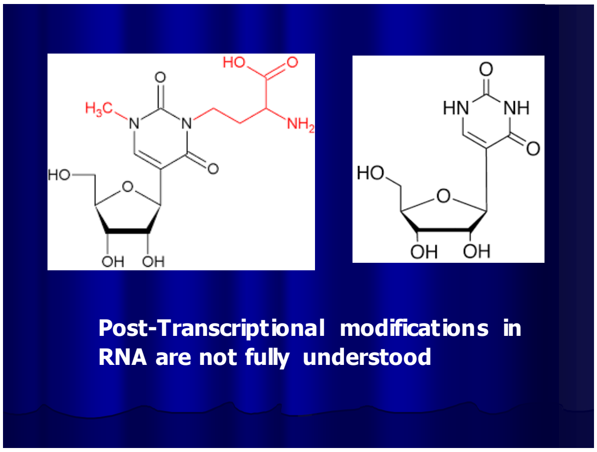
- A protein being copied from a protein
- Synthesis of RNA using the primary structure of a protein as a template
- DNA synthesis using the primary structure of a protein as a template.



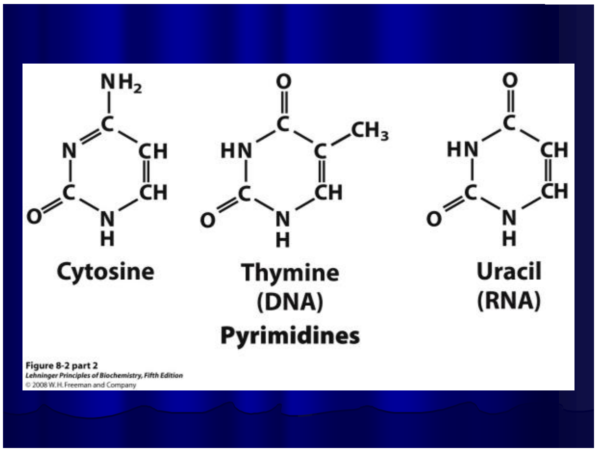
General	Special	Unknown
DNA → DNA	RNA → DNA	protein → DNA
DNA → RNA	RNA → RNA	protein → RNA
RNA → protein	DNA → protein	protein → protein

Post-translational modification (PTM)

- Refers to the covalent and generally enzymatic modification of proteins
- Post-translational modifications can occur on the amino acid side chains or at the protein's C- or N- termini.
- They can extend the chemical repertoire of the 20 standard **amino acids** by modifying an existing **functional group** or introducing a new one



Post-Transcriptional modifications in RNA are not fully understood



Cytosine

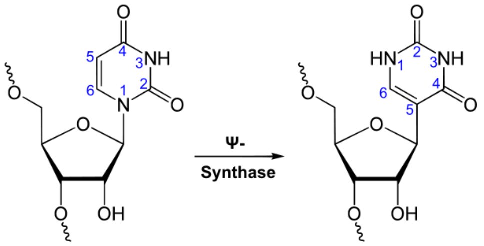
Thymine (DNA)

Uracil (RNA)

Pyrimidines

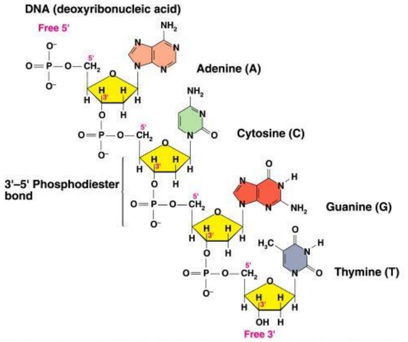
Figure 8-2 part 2
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

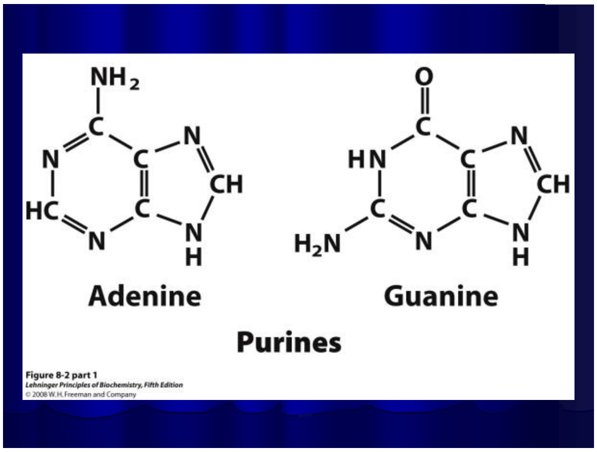
Post-Transcriptional modifications



Example of DNA Primary Structure

- In DNA, A, C, G, and T are linked by 3'-5' ester bonds between deoxyribose and phosphate



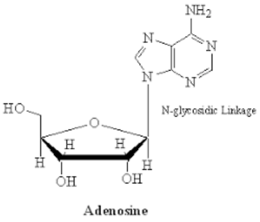


Adenine **Guanine**

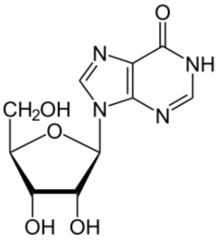
Purines

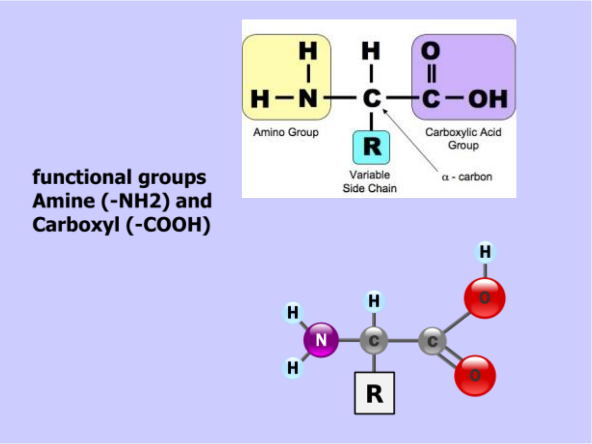
Figure 8-2 part 1
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

Inosine



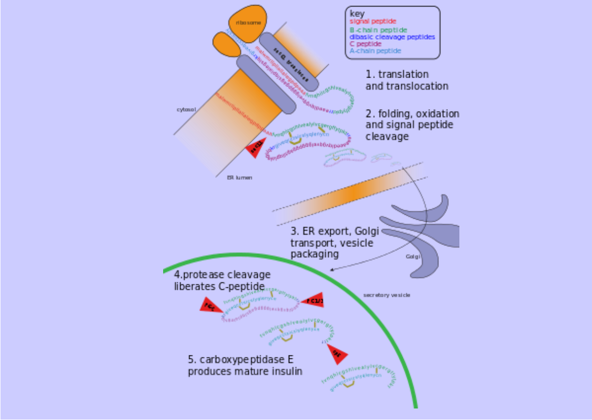
Adenosine





functional groups
Amine (-NH2) and
Carboxyl (-COOH)

Post-translational modification (PTM)



1. translation and translocation
2. folding, oxidation and signal peptide cleavage
3. ER export, Golgi transport, vesicle packaging
4.protease cleavage liberates C-peptide
5. carboxypeptidase E produces mature insulin

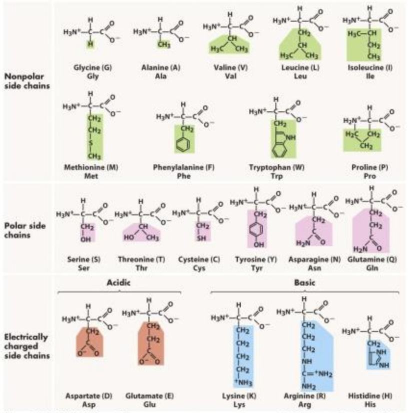


Figure 3-5 Biological Science, 2/e