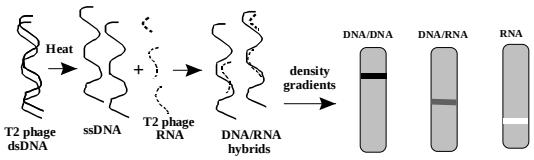
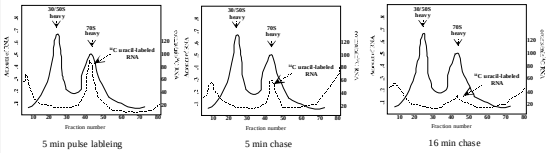
Lecture 3

What is a gene, how is it regulated, how is RNA made and processed, how are proteins made and what are their structures?

How does RNA fit in; its complementary to DNA

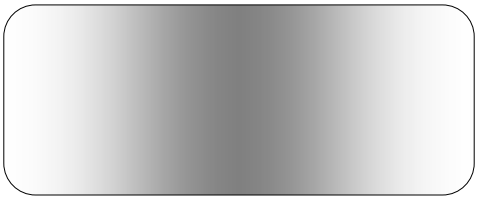


How do we know DNA makes RNA

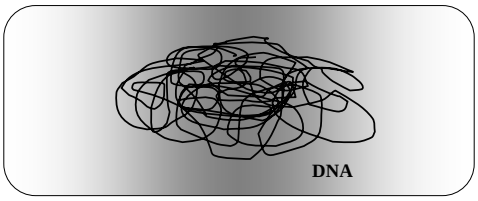


5 min pulse lableing | 5 min chase | 16 min chase

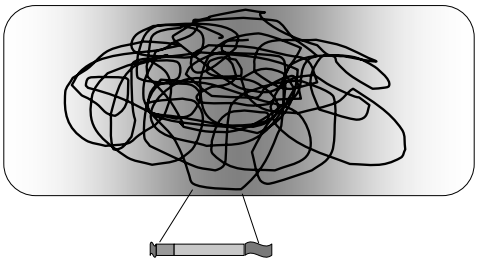
The simple bacterium *E. coli*



Its genome is composed of approximately 5 million base pairs of DNA

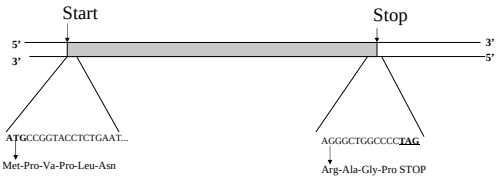


Within its genome are a few thousand genes

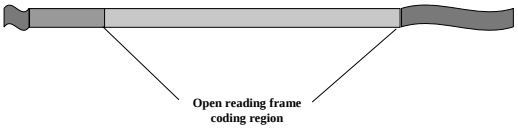


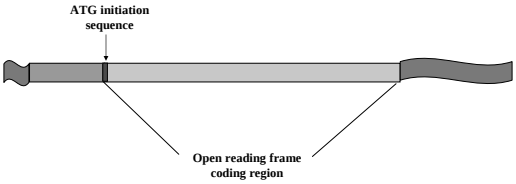
Defining a gene

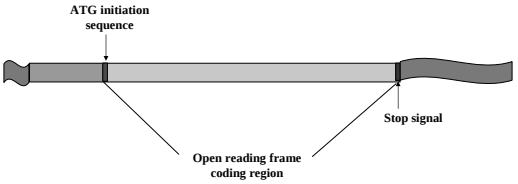
Within a bacterial gene, the information for a protein is found in a continuous sequence, Beginning with ATG and ending with a STOP codon



…but what defines a gene?







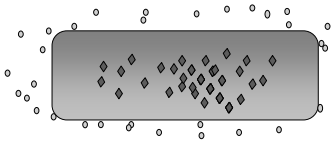
How we understand how genes work

- Jacob and Monod defined the lac operon
- Brenner and others determined that an unstable RNA molecule (messanger RNA) was the intermediary between DNA and protein.

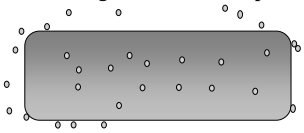
Jacob and Monod made many mutants that could not live on lactose

- These were two types
  - – those that could be complemented by another wild type gene
  - – and those that could not be complemented by the presence of a wild type gene

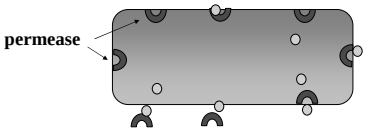
In the presence of glucose ♦ lactose ○ cannot enter the cell



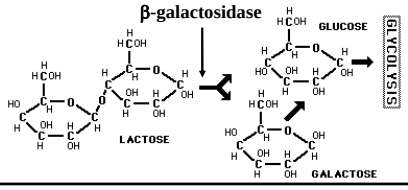
…but deprive the cells of glucose and lactose pours in



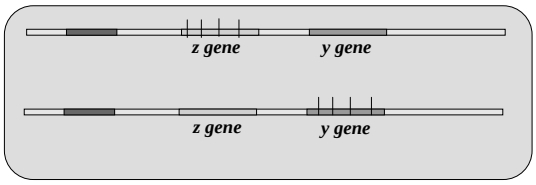
Permease transports lactose into cells



β-galactosidase converts lactose to glucose

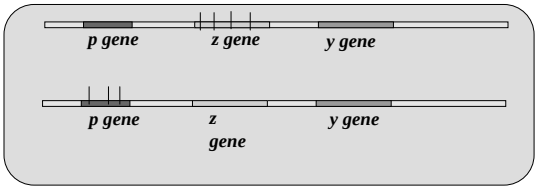


Mutations in the z or y gene could be complemented by the presence of a second wild type gene. These gene products are described as working in trans.



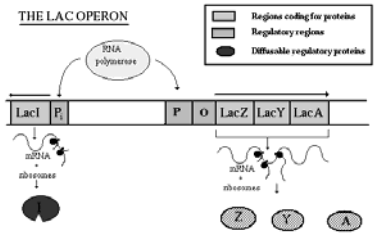
The *z* gene codes for b-galactosidase and *y* codes for permease

However, mutations in the gene that they called *p* could <u>not</u> be complemented by the presence of a wild type gene and so these mutants are know as cis-acting.

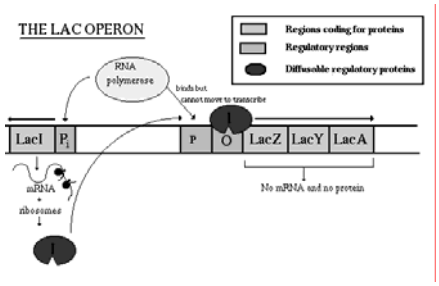


What J and M had done was discovered a promoter. A sequence within a gene that acted to regulate the expression of the gene.

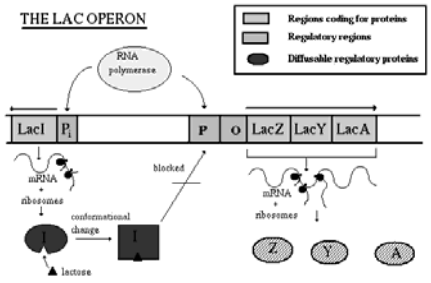
Lac Operon



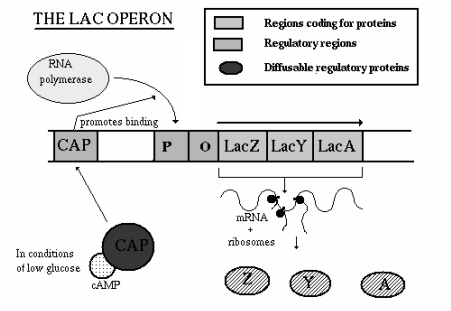
Negative regulation

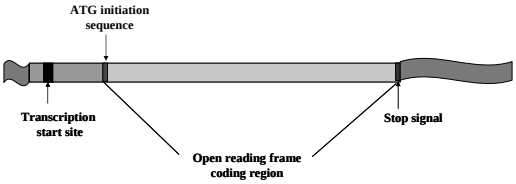


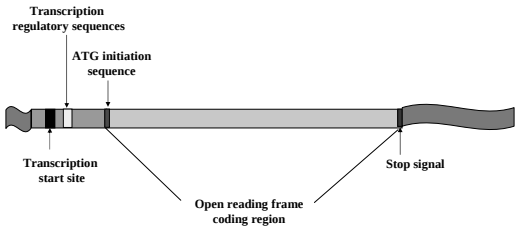
Lac on

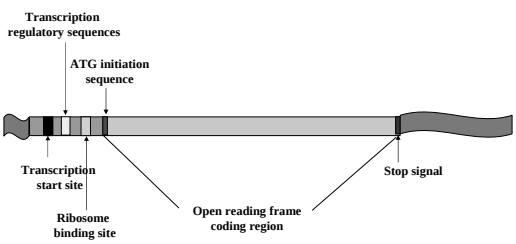


Influence of glucose (positive regulation)

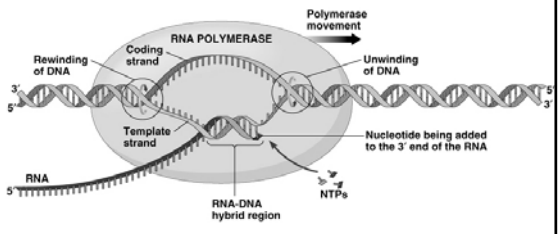






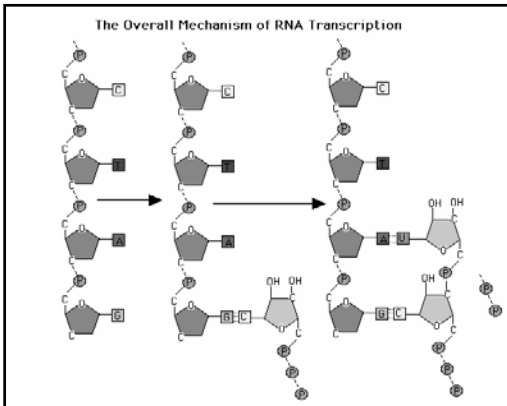


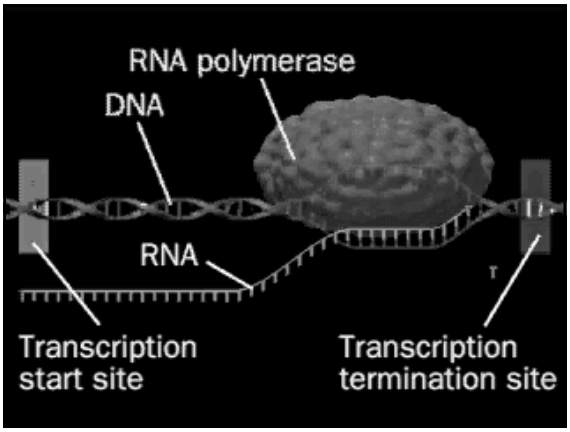
How do we access a specific packet of information (a gene) within the entire genome?



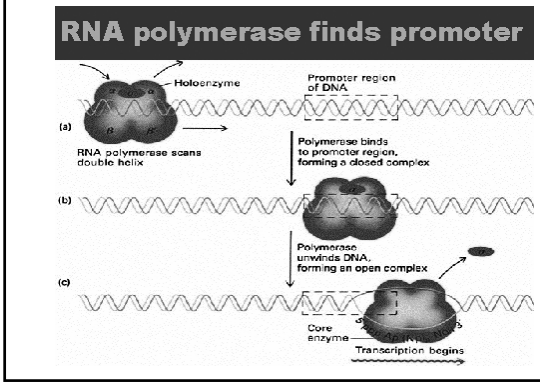
How does RNA Polymerase work?

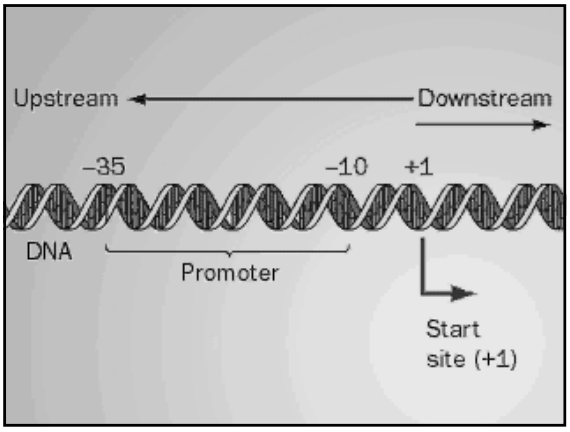
Reaction mechanism:
(NMP)n + NTP (NMP)n+1 PPi





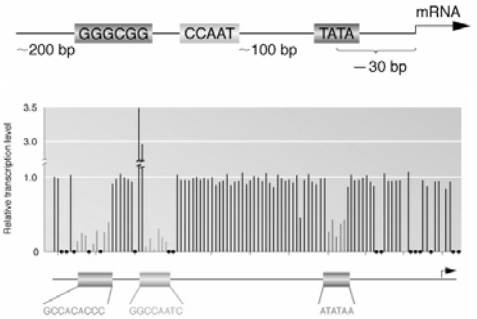
How do we know where to start transcription?

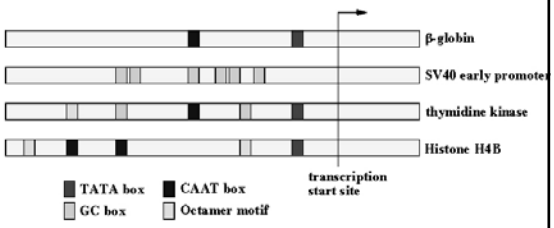




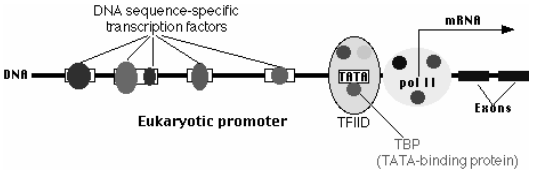
Eukaryotic promoters are more complex. The basal promoter refers to those sequences just upstream of the gene

How do we know that these are important regions of the promoter?

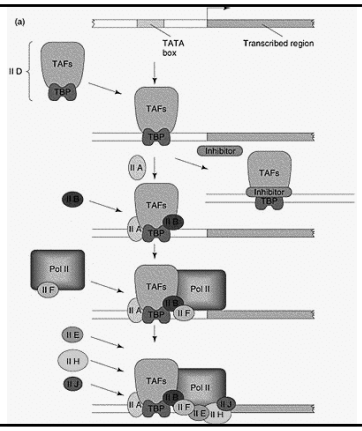


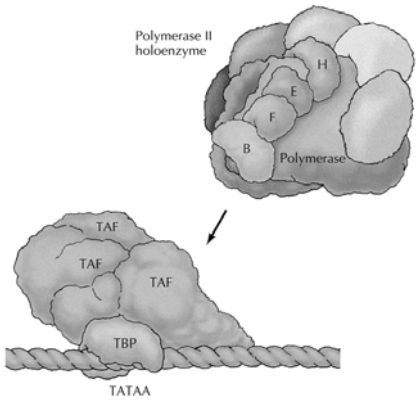


Eukaryotic Promoter



First a complex of proteins assemble at the TATA box including RNA polymerase II. This is the initiation step of transcription.

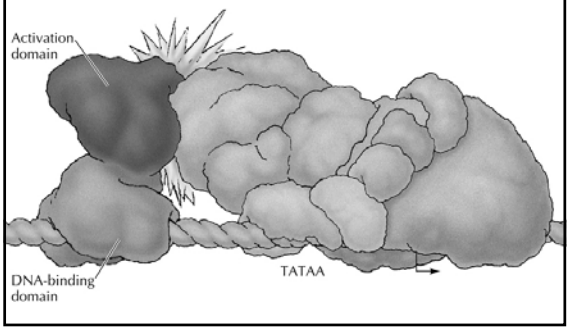




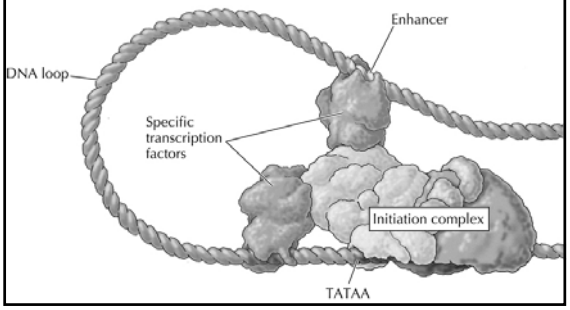
Regulation of transcription

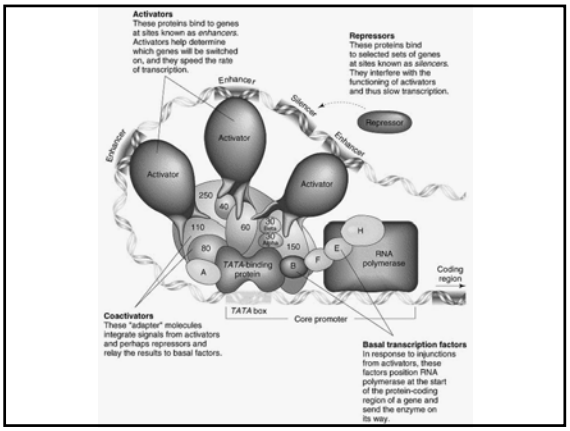
The next step in gene expression is its regulation. This is much more complicated then initiation and is still far from being completely understood

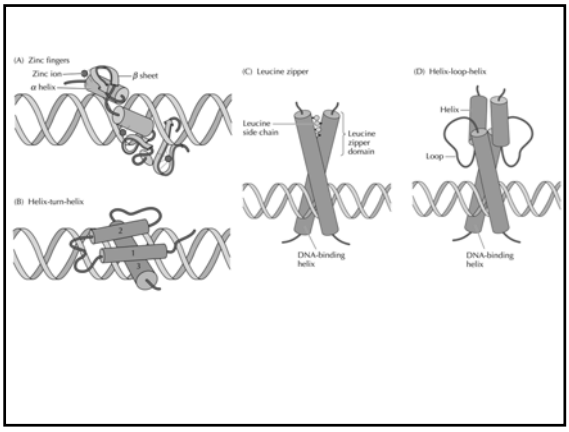
Transcription initiation from the TATA box requires upstream transcription factors



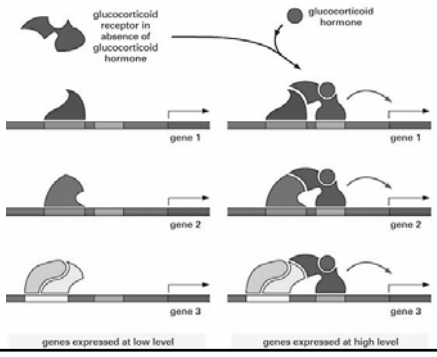
Regulatory proteins binding to the promoter and enhancer regions of are thought to interact by DNA folding

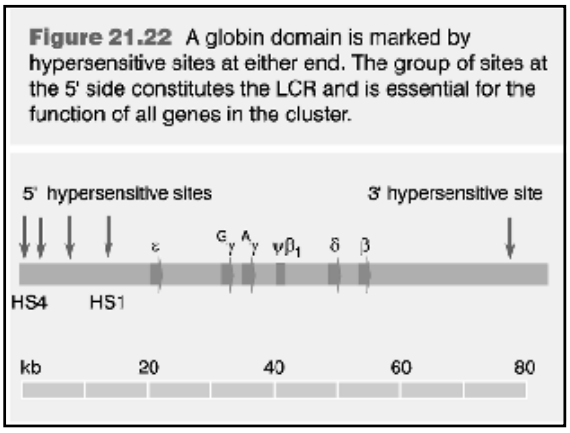






Regulation of promoter region





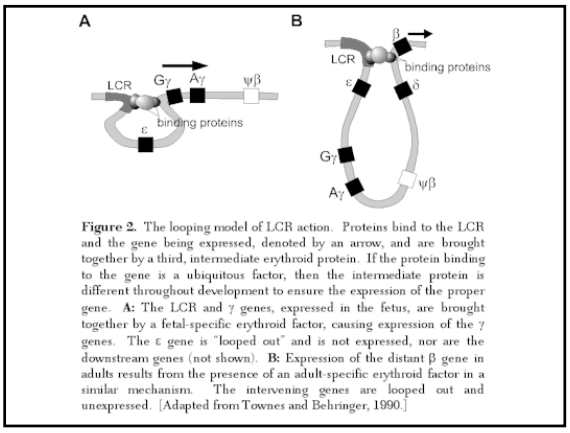
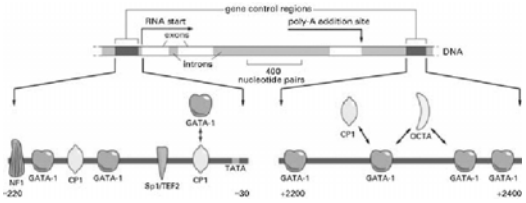
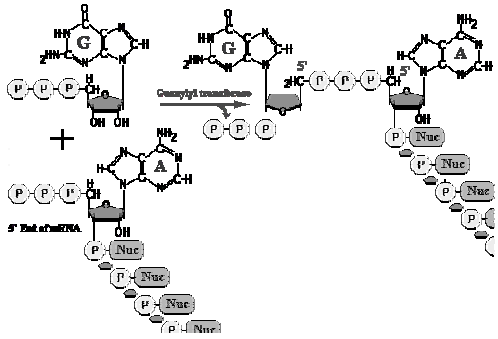


Figure 2. The looping model of LCR action. Proteins bind to the LCR and the gene being expressed, denoted by an arrow, and are brought together by a third, intermediate erythroid protein. If the protein binding to the gene is a ubiquitous factor, then the intermediate protein is different throughout development to ensure the expression of the proper gene. **A:** The LCR and γ genes, expressed in the fetus, are brought together by a fetal-specific erythroid factor, causing expression of the γ genes. The ε gene is "looped out" and is not expressed, nor are the downstream genes (not shown). **B:** Expression of the distant β gene in adults results from the presence of an adult-specific erythroid factor in a similar mechanism. The intervening genes are looped out and unexpressed. [Adapted from Townes and Behringer, 1990.]

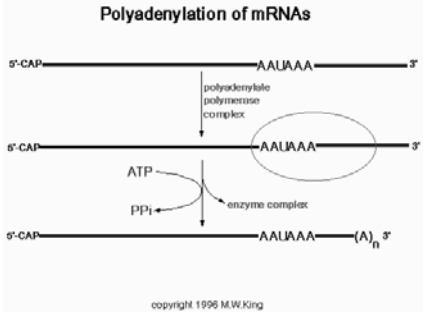
Globin gene complex



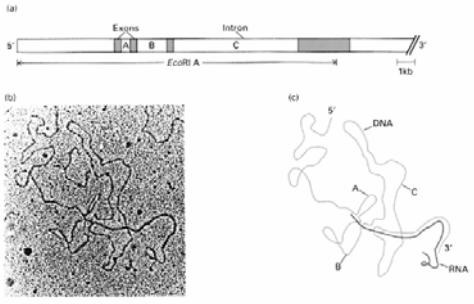
Adding a 5' cap



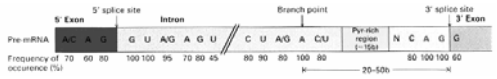
Termination and polyadenylation



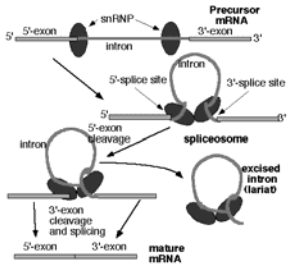
Pre-mRNA has introns

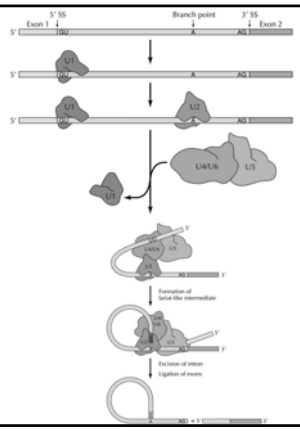


The splicing complex recognizes semiconserved sequences



Introns are removed by a process called splicing





snRNPs in splicing

- Donor site Branchpoint Acceptor site
- [----Exon----]GU----------A-----(Py)n--AG[----Exon----]
- U1 SnRNP recognises donor site by direct base pairing.
- U2 SnRNP recognises branch point by direct base pairing and subsequently also base pairs at the 5' end of the intron.
- U4&U6 (base paired together) are both in same SnRNP. U6 base pairs with U2.
- U5 SnRNP interacts with both donor and acceptor sites by base pairing, bringing together 2 exons.

Complexity of genes

- Splicing in some genes seems straightforward such as globin

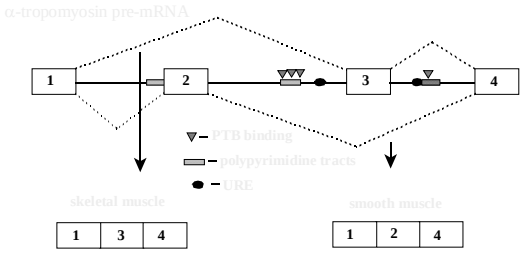
For other genes splicing is much more complex

- Fibrillin is a protein that is part of connective tissue. Mutations in it are associated with Marfan Syndrome (long limbs, crowned teeth elastic joints, heart problems and spinal column deformities. The protein is 3500 aa, and the gene is 110 kb long made up of 65 introns.
- Titin has 175 introns.
- With these large complex genes it is difficult to identify all of the exons and introns.

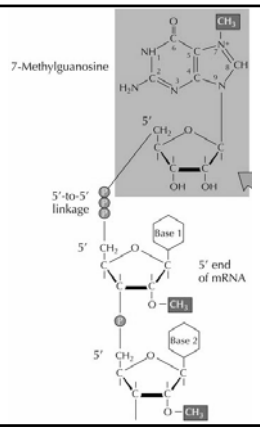
Alternative RNA splicing

- Shortly after the discovery of splicing came the realization that the exons in some genes were not utilized in the same way in every cell or stage of development. In other words exons could be skipped or added. This means that variations of a protein (called isoforms) can be produced from the same gene.

Alternative splicing of a tropomyosin



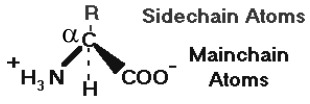
There are 3 forms of polypyrimiding tract binding protein (PTB) PTB1, PTB2 and PTB4. Binding of PTB4 to the polypyrimidine suppresses splicing while binding of PTB1 promotes splicing. In smooth muscle exon 3 of a-tropomyosin is not present. Thus, PTB4 is expressed in smooth muscle while PTB1 is not.



Proteins and Enzymes

The structure of proteins
How proteins functions
Proteins as enzymes

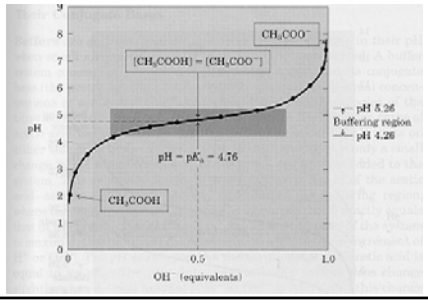
The R group gives and amino acid its unique character



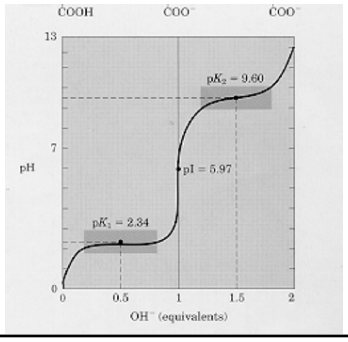
Dissociation constants

$$HA \rightleftharpoons H^+ + A^-$$

Titration curve of a weak acid



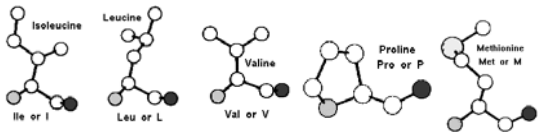
Titration curve of glycine



Properties of Amino Acids

Alaphatic amino acids

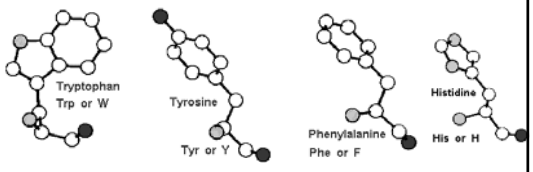
only carbon and hydrogen in side group



Honorary member

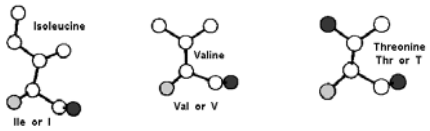
Strictly speaking, aliphatic implies that the protein side chain contains only carbon or hydrogen atoms. However, it is convenient to consider Methionine in this category. Although its side-chain contains a sulphur atom, it is largely non-reactive, meaning that Methionine effectively substitutes well with the true aliphatic amino acids.

Aromatic Amino Acids



A side chain is aromatic when it contains an aromatic ring system. The strict definition has to do with the number of electrons contained within the ring. Generally, aromatic ring systems are planar, and electons are shared over the whole ring structure.

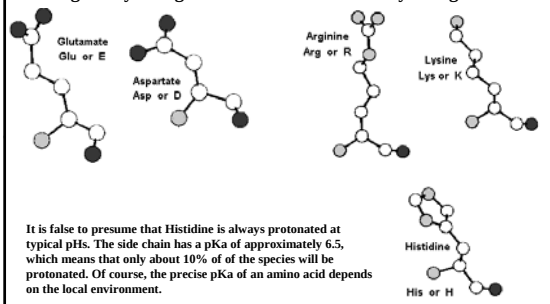
Amino acids with C-beta branching



Whereas most amino acids contain only one non-hydrogen substituent attached to their C-beta carbon, C-beta branched amino acids contain two (two carbons in Valine or Isoleucine; one carbon and one oxygen in Theronine) . This means that there is a lot more bulkiness near to the protein backbone, and thus means that these amino acids are more restricted in the conformations the main-chain can adopt. Perhaps the most pronounced effect of this is that it is more difficult for these amino acids to adopt an alpha-helical conformation, though it is easy and even preferred for them to lie within beta-sheets.

Charged Amino Acids

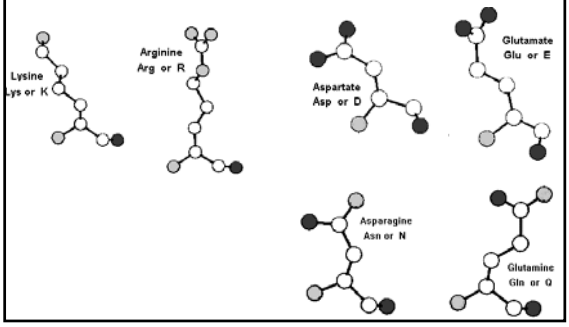
Negatively charged **Positively charged**



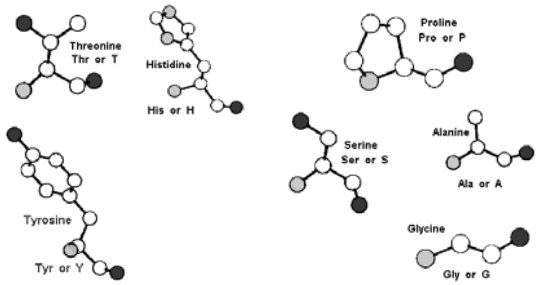
It is false to presume that Histidine is always protonated at typical pHs. The side chain has a pKa of approximately 6.5, which means that only about 10% of of the species will be protonated. Of course, the precise pKa of an amino acid depends on the local environment.

Partial positive charge

Polar amino acids

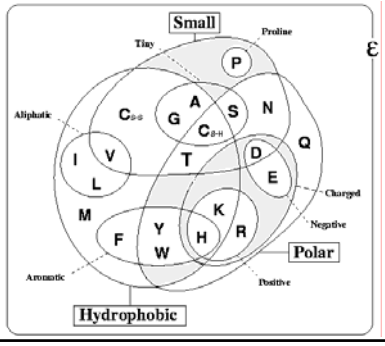


Somewhat polar amino acids



Polar amino acids are those with side-chains that prefer to reside in an aqueous (i.e. water) environment. For this reason, one generally finds these amino acids exposed on the surface of a protein.

Amino acids overlap in properties



How to think about amino acids

- *Substitutions*: Alanine generally prefers to substitute with other small amino acid, Pro, Gly, Ser.
- *Role in structure*: Alanine is arguably the most boring amino acid. It is not particularly hydrophobic and is non-polar. However, it contains a normal C-beta carbon, meaning that it is generally as hindered as other amino acids with respect to the conforomations that the backbone can adopt. For this reason, it is not surprising to see Alanine present in just about all non-critical protein contexts.
- *Role in function*: The Alanine side chain is very non-reactive, and is thus rarely directly involved in protein function. However it can play a role in substrate recognition or specificity, particularly in interactions with other non-reactive atoms such as carbon.

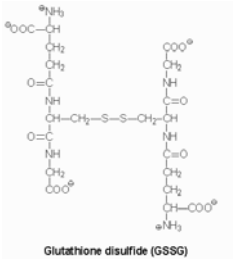
Tyrosine

- *Substitutions*: As Tyrosine is an aromatic, partially hydrophobic, amino acid, it prefers substitution with other amino acids of the same type (see above). It particularly prefers to exchange with Phenylalanine, which differs only in that it lacks the hydroxyl group in the ortho position on the benzene ring.
- *Role in function:* Unlike the very similar Phenylalanine, Tyrosine contains a reactive hydroxyl group, thus making it much more likely to be involved in interactions with non protein atoms. Like other aromatic amino acids, Tyrosine can be involved in interactions with non-protein ligands that themselves contain aromatic groups via stacking interactions.
- A common role for Tyrosines (and Serines and Threonines) within intracellular proteins is phosphorylation. Protein kinases frequently attach phosphates to Tyrosines in order to fascilitate the signal transduction process. Note that in this context, Tyrosine will rarely substitute for Serine or Threonine, since the enzymes that catalyse the reactions (i.e. the protein kinases) are highly specific (i.e. Tyrosine kinases generally do not work on Serines/Threonines and vice versa)

Cysteine

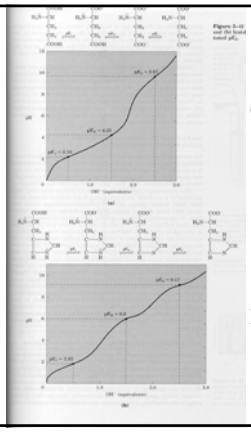
- *Substitutions*: Cysteine shows no preference generally for substituting with any other amino acid, though it can tolerate substitutions with other small amino acids. Largely the above preferences can be accounted for by the extremely varied roles that Cysteines play in proteins (see below). The substitutions preferences shown above are derived by analysis of all Cysteines, in all contexts, meaning that what are really quite varied preferences are averaged and blurred; the result being quite meaningless.
- *Role in structure*: The role of Cysteines in structure is very dependent on the cellular location of the protein in which they are contained. Within extracellular proteins, cysteines are frequently involved in disulphide bonds, where pairs of cysteines are oxidised to form a covalent bond. These bonds serve mostly to stabilise the protein structure, and the structure of many extracellular proteins is almost entirely determined by the topology of multiple disulphide bonds

Cystine andGlutathione



Glutathione disulfide (GSSG)

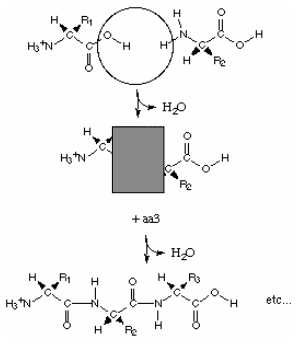
Glutathione (GSH) is a tripeptide composed of g-glutamate, cysteine and glycine. The sulfhydryl side chains of the cysteine residues of two glutathione molecules form a disulfide bond (GSSG) during the course of being oxidized in reactions with various oxides and peroxides in cells. Reduction of GSSG to two moles of GSH is the function of glutathione reductase, an enzyme that requires coupled oxidation of NADPH.



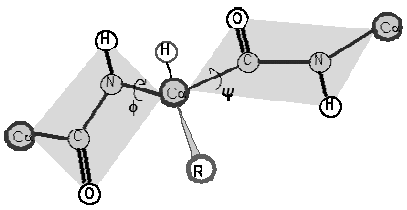
Glutamic acid

Histidine

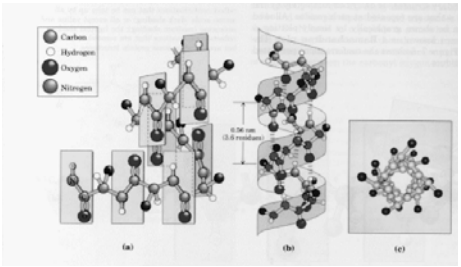
The peptide bond



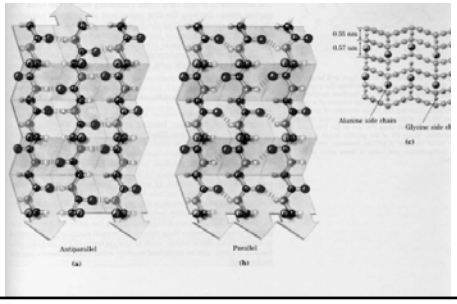
There is free rotation about the peptide bond



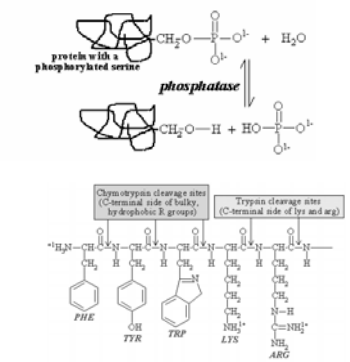
Proteins secondary structure, alpha helix



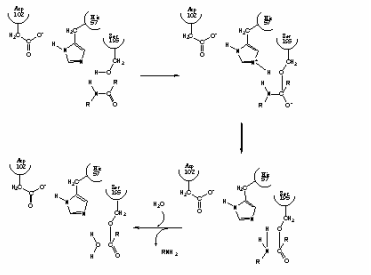
Secondary structure, beta pleated sheet



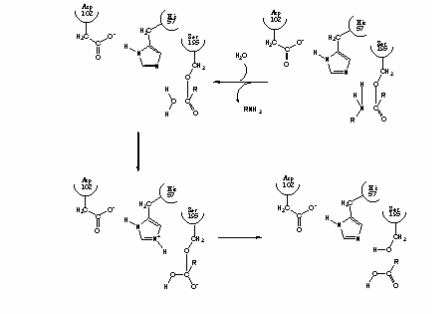
How enzymes work



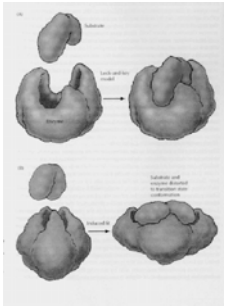
Catalytic Mechanism of Serine Proteases - 1



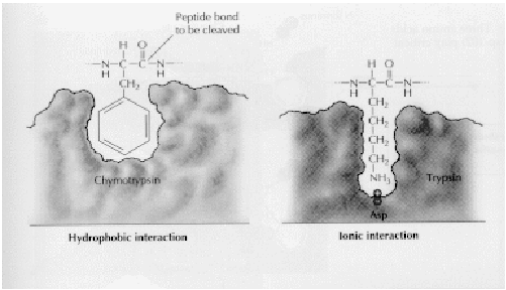
Catalytic Mechanism of Serine Proteases - 2



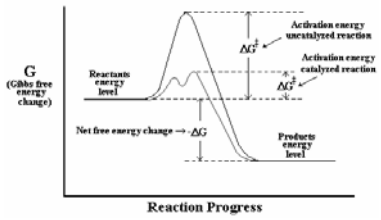
Lock and key



Specific interactions at active site



Enzymes lower the energy of activation



Transition State Stabilization

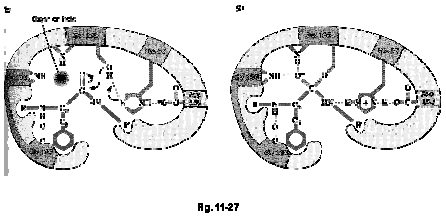


Fig. 11-27

How chymotrypsin works