Exam notes for bio250 semester one

Lecture One:

<u>Prokaryotes</u>: Contain no nuclei (mostly single-celled), bacteria plus archaea.
<u>Origin of Mitochondria</u>: Mitochondria were single-celled prokaryotes, which were engulfed by eukaryotes. This developed into a symbiotic relationship. The prokaryotic bacteria produced ATP in exchange for shelter and food. Mitochondria have their own genome, which consists of mitochondrial DNA.
<u>Eukaryotes</u>: Can be quite diverse. Single-celled eukaryotes include protists and yeast.
A <u>gene</u> is defined as the segment of DNA sequence corresponding to a single protein (or to a single catalytic or structural RNA molecule for those genes that produce RNA but not protein).

Lecture Two:

<u>DNA</u>: stands for deoxyribonucleic acid. <u>RNA</u>: stands for ribonucleicacid. Components of nucleotides:
<u>Pentose sugar</u>: Five-carbon sugar in ring form. In RNA sugar has an OH group at carbon two. In DNA this OH group is missing.
<u>Bases</u>: Bases are adenine (A), cytosine (C), guanine (G), and Thymine (T). In RNA thymine is replaced by Uracil (U). G and A are <u>purines</u>. T, U and C are <u>pyrimidines</u>. Purines have two fused rings. T has a methyl group, which U lacks.
<u>Nucleoside</u>: Base plus sugar. <u>Nucleotide</u>: Base plus sugar plus phosphate. DNA is synthesized from dNTP's, or deoxyribonucleotidetriphosphates. Phosphodiester linkages join nucleotides together. A pairs with T (U), which has two hydrogen bonds. G pairs with C and forms three hydrogen bonds and is therefore more stable. Hydrogen bonds, Van der Waals forces and hydrophobic interactions stabilize DNA. At the 5 prime end is the phosphate group and at that the 3 prime end is the OH group from the sugar. The strands in the DNA double helix are anti-parallel. DNA is unzipped by binding proteins in the cell or by heat in the lab. This process is called denaturation. Cooling allows annealing or renaturation of the double helix.

Lecture Three:

The genetic code is <u>degenerate</u>. Three nucleotides make one <u>codon</u>, which encodes for one amino acid. 4^3 means there are 64 possible condons but

there are only 20 amino acids plus the stop codon. <u>Protein functions</u>: regulation (kinases), signalling (receptors), transport (ion channels), catalysts (enzymes), movement (motor proteins) and structure (cytoskeletal proteins).

<u>Amino acids</u> can be basic (K, R, H), acidic (D, E), Uncharged polar (N, Q, S, T, Y) or non-polar (G, A, V, I, L, W, F, P, M, C). Mutations can separate one amino acid from another. Minimum number of mutations to go from one amino acid to another is dependent on the amino acids in question. Groups of amino acids with the same properties tend to be clustered in the codon table. Codons of amino acids with similar properties tend to have fewer mutational steps between them.

A <u>peptide bond</u> links amino acids through a condensation reaction, which results in the loss of water. An OH group is lost from the COOH group and hydrogen is lost from the NH3 group. The side chain (R) is connected to the alpha carbon. An <u>alpha helix</u> has 3.6 amino acids per turn and has hydrogen bonds between turns (intramolecular) between amino acid n and n+4. In <u>beta sheets</u> hydrogen bonds are formed between different strands (intermolecular). The carbonyl oxygen is hydrogen bonded to the amide hydrogen. The folded protein is held together by hydrophobic interactions, non-covalent bonds and disulfide linkages.

The <u>Src</u> protein has 4 different domains for specialized functions. Src signals pathways and it contains regulatory and catalytic domains.

<u>Hemoglobin</u> is composed of 4 subunits, 2 alpha and 2 Beta subunits. Sickle-cell anemia is caused by a mutation in the beta subunits. Hemoglobin transports oxygen from the lungs into the tissues. Myoglobin is a related protein, which is composed of one subunit and is found in muscle tissue. Myoglobin has a higher affinity for oxygen.

Lecture 4 (DNA Replication in Prokaryotes):

<u>DNA replication</u> is semi-conservative. The origin of replication tends to be A-T rich because A-T only has two hydrogen bonds and it is easier to break. The origin is recognized by initiator proteins. Bacteria have a single origin of replication (plasmid). Eukaryotes have multiple origins of replication (along chromosomes).

<u>ARS</u>: Stands for autonomously replicating sequences, this is often how a plasmid is replicated in bacteria. This means that it independently replicates itself without an outside force.

<u>DNA synthesis</u> always occurs in the 5 prime to 3 prime direction. The template is read in the 3 prime to 5 prime direction. Therefore, one strand is not continuously replicated. DNA synthesis in the 5 to 3 prime direction

helps in correcting errors. When a new nucleotide is being added to the 3 prime end, a pyrophosphate is released, which is then converted to two inorganic phosphates. This is what gives the energy to drive the reaction. If you wanted to synthesize DNA in the 3 to 5 prime direction, then the high-energy tri-phosphate that reacts to give the phosphodiester bond would be contained on the end of the chain and not on the nucleotide. This would not work for error correction because if you needed to eliminate a nucleotide and replace it with another then you would lose the high-energy tri-phosphate from the end of the chain and then you could not add a new nucleotide to the 5 prime end because there would be no energy to power the reaction. Therefore, even though it takes more time, DNA is always synthesized in the 5 to 3 prime direction. This means that there is a replication fork in DNA synthesis. Circular genomes contain more than one replication fork.

Steps and or factors needed for replication (prokaryotic):
Firstly initiator proteins bind, then there is unwinding by helicase and binding of primase, then the sliding clamp binds and holds DNA polymerase onto the DNA and then DNA ligase seals the Okazaki fragments.

DNA polymerase: Attaches nucleoside tri-phosphates to growing chain at the 3 prime end.

Primers and Primase: A special primer is needed at the start of replication in order to bind DNA polymerase. On the leading strand only one initial primer is needed but on the lagging strand several primers are required as many short Okazaki fragments are being synthesized. DNA primase is the enzyme used to synthesize short RNA primers on the lagging strand using ribonucleoside tri-phosphates. Once the lagging strand is completely sythesized, the old RNA primers are erased and replaced by DNA by a special DNA repair system. RNA primers are probably used instead of DNA primers because it gives the system an extra chance for error correction when RNA primers are removed and replaced by DNA.

DNA ligase: DNA ligase seals the Okazaki fragments by joining the 3 prime end of the new DNA fragment to the 5 prime end of the previous one. This reaction requires the use of ATP.

DNA helicase: This enzyme pry's apart double stranded DNA as it moves along the strand. In unzips at 1000 nucleotides per second and moves along in the 5 to 3 prime direction. This enzyme is made up of 6 subunits.

Single-stranded Binding Proteins (SSBP): These stabilize single-stranded DNA. They prevent hydrogen bonds and hairpins and kinks from forming. Initiator proteins bind to origin of replication (are highly regulated) and help helicase to bind. These are ATP dependent.

Sliding Clamp: This holds DNA polymerase onto the DNA, but rapidly releases it when a region of double-stranded DNA is encountered. On the

lagging strand DNA polymerase needs to synthesize next Okazaki fragment quickly so DNA polymerase needs to be able to fall off easily. However, the clamp also ensures that DNA polymerase does not fall off of the leading strand easily as this needs to be continuously replicated. The assembly of the clamp around DNA requires the hydrolysis of ATP by a clamp loader.
Primosome: In prokaryotes, the primase enzyme is linked directly to the helicase to form a unit on the lagging strand called the primosome. The helicase unwinds the DNA as the primosome makes RNA primers on the lagging strand.
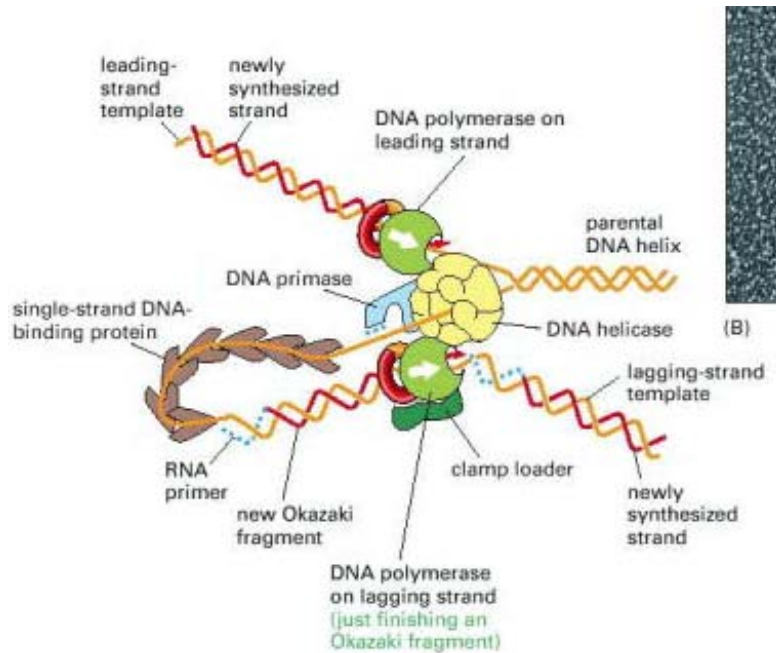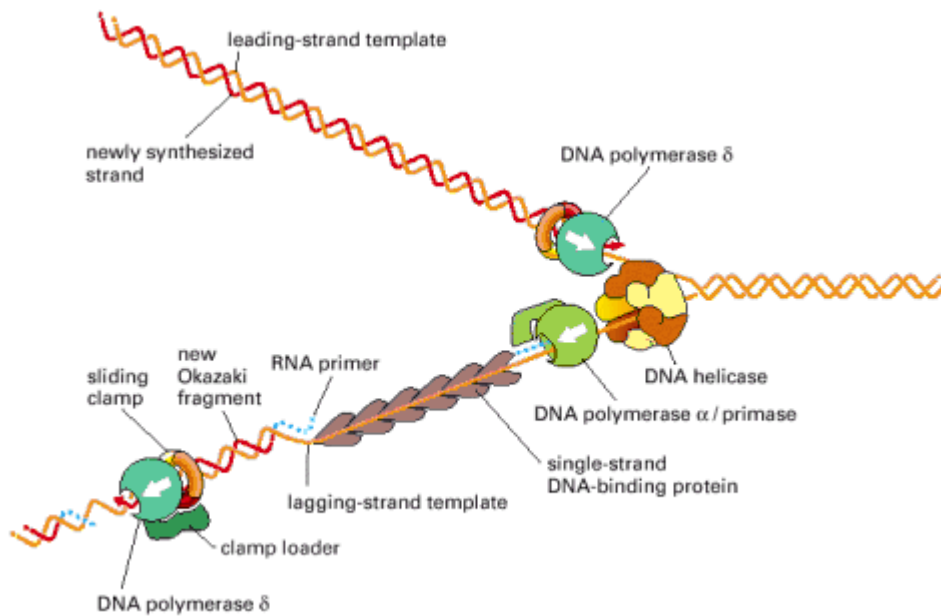

Summary of Prokaryotic DNA replication:
The leading strand is synthesized continuously from a single RNA primer.
The lagging strand is synthesized discontinuously from many RNA primers.
Okazaki fragments consist of RNA primers plus DNA. DNA replaces RNA primers and the fragments are sealed by ligase.
DNA replication always proceeds in the 5 to 3 prime direction, creating a replication fork.
The helicase plus the primase makes a primosome.
The lagging strand contains the predominant helicase.

Lecture five (Eukaryotic DNA replication):

leading-strand template

newly synthesized
strand

DNA polymerase δ

new
Okazaki
fragment

RNA primer

sliding
clamp

DNA helicase

DNA polymerase α / primase

single-strand
DNA-binding protein

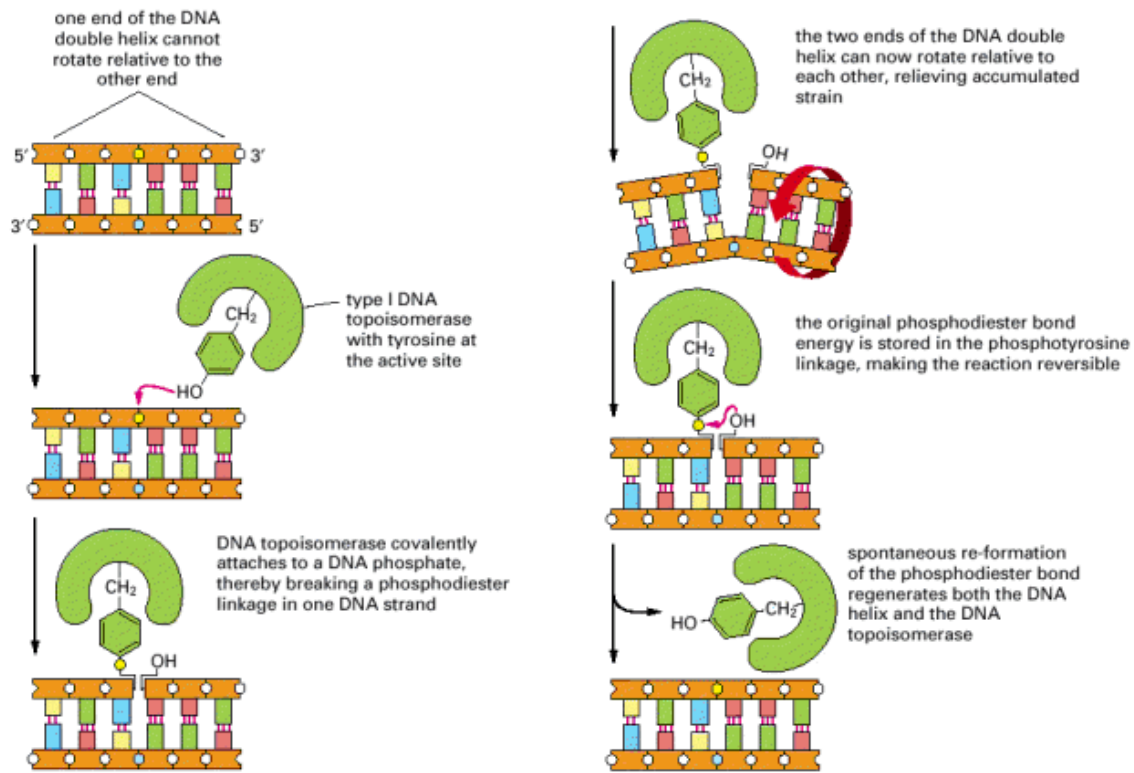lagging-strand template

clamp loader

DNA polymerase δ

In eukaryotic DNA replication helicase is not associated with primase. Primase is a subunit of a different type of DNA polymerase, which initiates synthesis of Okazaki fragments. Okazaki fragments are a whole degree smaller than they are in bacteria. During DNA replication there is a shortening of the lagging strand at the 5 prime end. This shortening could result in the loss of information so repetitive sequences are added to the 3 prime end of the parent strand.

Telomeres: are the ends of chromosomes with repetitive sequences. An RNA template is incorporated into an enzyme called Telomerase, which makes DNA copies and adds them to the end of the chromosome. This enzyme resembles another enzyme called reverse transcriptase. The sequences added are G-rich and they are added to the 3 prime end. Single stranded DNA is susceptible to attack and so it bends back and forms a hairpin loop. In some cells, Telomeres shorten each time the cell divides (in cells without Telomerase). The telomeres shortening can sometimes be used to count cell divisions and trigger the withdrawal of the cell from cell division. This may prevent some forms of cancer.
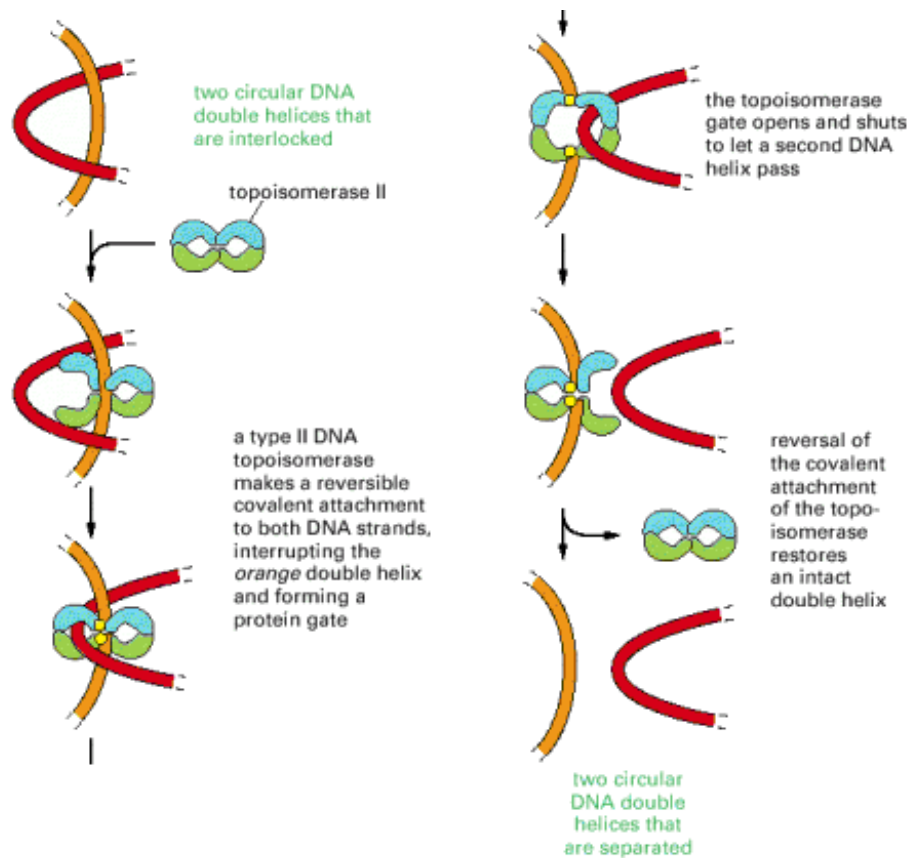
The Unwinding Problem: Unwinding DNA is energetically expensive. Supercoils that form that are in the same direction of the twist of the double helix are termed positive. Supercoils on the opposite direction are termed negative. Replication of DNA introduces supercoils in the positive direction. It is difficult to rotate the entire chromosome arm, it's a relatively fixed object.

Topoisomerase: Breaks and reforms phosphodiester bonds.

Topoisomerase I: Breaks single-stranded DNA, which allows for rotation around backbone. This results in negative supercoils, which are important just in front of the replication fork.

one end of the DNA double helix cannot rotate relative to the other end

the two ends of the DNA double helix can now rotate relative to each other, relieving accumulated strain

type I DNA topoisomerase with tyrosine at the active site

the original phosphodiester bond energy is stored in the phosphotyrosine linkage, making the reaction reversible

DNA topoisomerase covalently attaches to a DNA phosphate, thereby breaking a phosphodiester linkage in one DNA strand

spontaneous re-formation of the phosphodiester bond regenerates both the DNA helix and the DNA topoisomerase
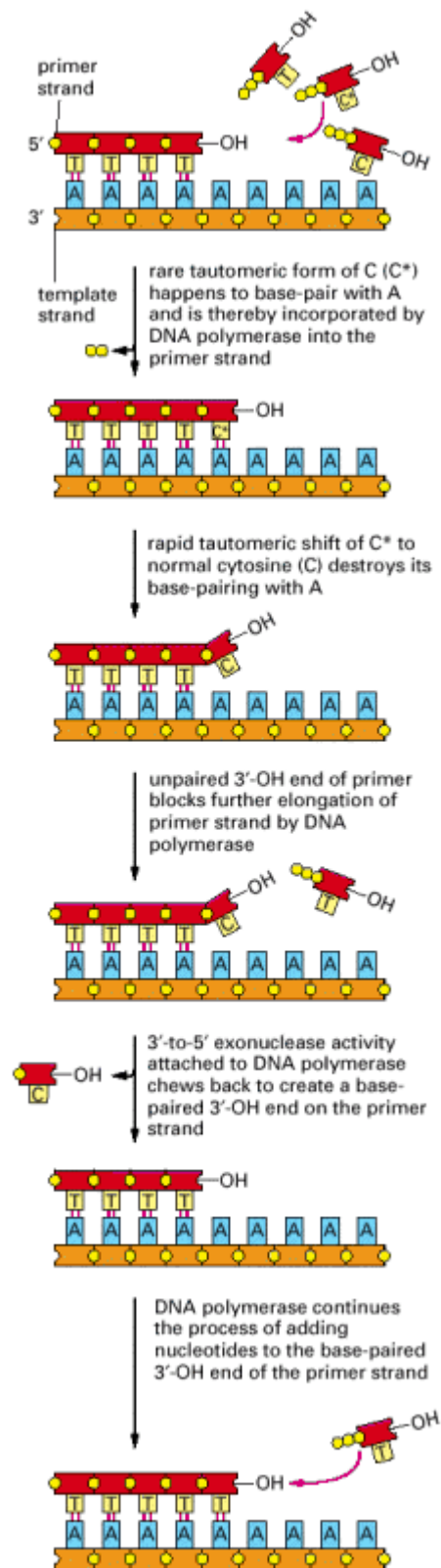
Topoisomerase II: Untangles two different helices if they get tangled together. It breaks double-stranded DNA, which allows for helices to pass through one another.

two circular DNA double helices that are interlocked

topoisomerase II

a type II DNA topoisomerase makes a reversible covalent attachment to both DNA strands, interrupting the *orange* double helix and forming a protein gate

the topoisomerase gate opens and shuts to let a second DNA helix pass

reversal of the covalent attachment of the topo-isomerase restores an intact double helix

two circular DNA double helices that are separated

DNA Repair Mechanisms: RNA polymerases have an error rate of about 1 in 10^4 nucleotides. DNA polymerases have an error rate of about 1 in 10^9 nucleotides. Therefore the human genome is only changed by 3 nucleotides each cell division. High fidelity is maintained by proofreading. Proofreading is done by 3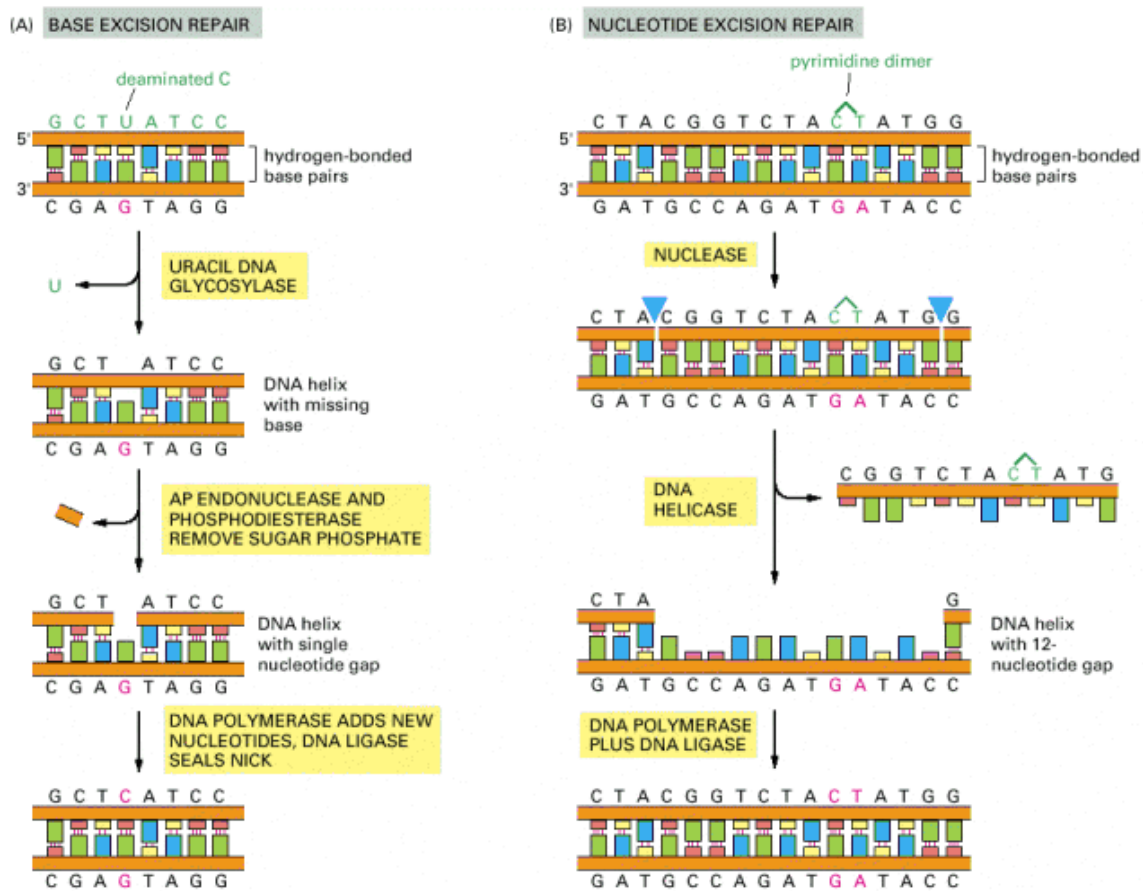 to 5 prime exonuclease and strand-directed mismatch repair. Exonuclease: Is part of DNA polymerase, it chews back misincorporated nucleotides.

primer strand

5′ ──OH

3′

template strand

rare tautomeric form of C (C*) happens to base-pair with A and is thereby incorporated by DNA polymerase into the primer strand

──OH

rapid tautomeric shift of C* to normal cytosine (C) destroys its base-pairing with A

──OH

unpaired 3′-OH end of primer blocks further elongation of primer strand by DNA polymerase

──OH

3′-to-5′ exonuclease activity attached to DNA polymerase chews back to create a base-paired 3′-OH end on the primer strand

──OH

DNA polymerase continues the process of adding nucleotides to the base-paired 3′-OH end of the primer strand

──OH

Defects in repair mechanisms can lead to many human diseases including breast, ovarian, colon and skin cancers. DNA can be damaged by oxidation,

methylation, heat and UV. DNA can be spontaneously damaged by depurination and/or deamination. Depurination affects purines (guanine and adenine). It results in the loss of a purine base. Deamination results in the loss of an amine group. This affects A, C and G. Loss of an amine in cytosine gives uracil. Deamination gives unnatural bases (for DNA) except for the deamination of 5-methyl cytosine, which gives rise to Thymine. In this case, where conversion is to a naturally occurring base, you can no longer distinguish which is the original naturally occurring base. Therefore, it is very dangerous.

There are two types of strand-directed mismatch repair. One type is base excision repair. This type targets a single nucleotide. The other type is called nucleotide excision repair. This targets many nucleotides.
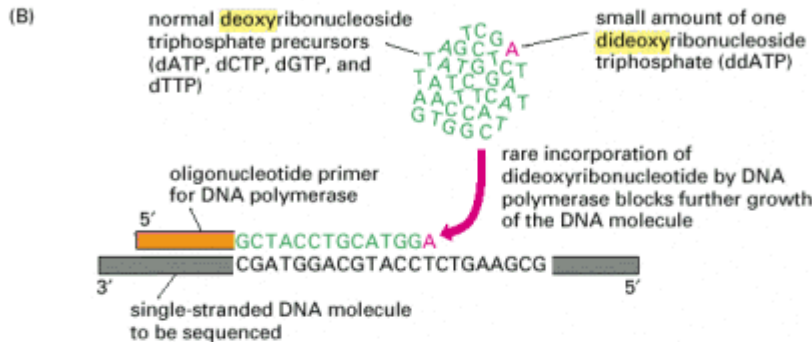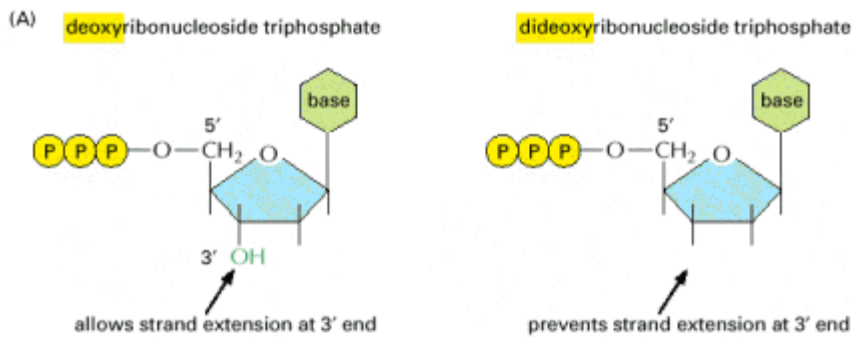

Exonucleases make corrections during DNA synthesis and the strand-directed mismatch repair makes corrections after synthesis. The strand-directed mismatch repair system requires double stranded DNA.

Lecture Six (DNA Sequencing):

Sequencing DNA is easier than sequencing proteins. The difficulties with sequencing proteins are that lots of protein is needed, its less stable and the code is degenerate (more than one DNA sequence can code for an amino acid). Also, the protein sequence can be inferred from the DNA sequence anyway.

Dideoxy Method of DNA Sequencing: This method uses DNA polymerase, dNTP's (DNA nucleotides) and ddNTP's (dideoxynucleotides) and primers. The first step in this procedure is to anneal the primer, then extend our sequence using pool of dNTP's, the added ddNTP's and DNA polymerase. The last step is to run the gel. DdNTP's lack an OH group, which prevents strand extension at the 3 prime end. We begin with 4 test tubes each containing a primer (common to all), that will be elongated by DNA polymerase. Each test tube also contains dNTP's needed to elongate the strand. However, one test tube also contains ddATP, one contains ddCTP, one contains ddGTP and the last one contains ddTTP. If these are incorporated into the strand, elongation will stop. We then run the contents of the four test tubes down the same gel, which will separate our DNA fragments by size. We can then read our sequence from the bottom of the gel to the top.

(A)

deoxyribonucleoside triphosphate

dideoxyribonucleoside triphosphate

base

P P P —O—CH₂ O
                5′

3′ OH

allows strand extension at 3′ end

base

P P P —O—CH₂ O
                5′

prevents strand extension at 3′ end

(B)

normal deoxyribonucleoside
triphosphate precursors
(dATP, dCTP, dGTP, and
dTTP)

small amount of one
dideoxyribonucleoside
triphosphate (ddATP)

```
    TCG
  AGCTA
TATGTCT
TATCGA
AACTTCAT
GTGCCAT
  TGGCT
```

rare incorporation of
dideoxyribonucleotide by DNA
polymerase blocks further growth
of the DNA molecule

oligonucleotide primer
for DNA polymerase

5′
GCTACCTGCATGGA
CGATGGACGTACCTCTGAAGCG
3′   5′

single-stranded DNA molecule
to be sequenced

(C)

5′ GCATATGTCAGTCCAG 3′ ⎤ double-stranded
3′ CGTATACAGTCAGGTC 5′ ⎦ DNA

labeled primer

5′ GCAT 3′
3′ CGTATACAGTCAGGTC 5′ single-stranded
                         DNA

+ excess dATP
         dTTP
         dCTP
         dGTP

+ ddATP + DNA polymerase	+ ddTTP + DNA polymerase	+ ddCTP + DNA polymerase	+ ddGTP + DNA polymerase
GCAT A	GCAT AT	GCAT ATGTC	GCAT ATG
GCAT ATGTCA	GCAT ATGT	GCAT ATGTCAGTC	GCAT ATGTCAG
GCAT ATGTCAGTCCA	GCAT ATGTCAGT	GCAT ATGTCAGTCC	GCAT ATGTCAGTCCAG

3′
G
A
C
C
T
G
A
C
T
G
T
A
5′

A   T   C   G

DNA sequence reading directly from the bottom of the gel upward, is

ATGTCAGTCCAG
1   12

For this method of DNA sequencing, the ratio of ddNTP's is low compared to dNTP's because if there are too many ddNTP's, the chances of them being incorporated into the strand are too high and the resulting fragments will be too short. The gel needs to be able to resolve differences in DNA fragment sizes of only one base, because we need to read the sequence one base at a time. The four test tube contents must be loaded in different lanes of the gel or else its impossible to determine which nucleotide terminated elongation. Only one type of ddNTP can be added per reaction tube or DNA replication can be terminated at more than one nucleotide, which confounds results. DdNTP's are not found in real cells because they would result in loss of genetic information.

DNA Sequencing vs. Replication: In sequencing a DNA primer is used instead of an RNA primer. Heat is used to separate the double-stranded DNA rather than enzymes and since heat separates the strands there is no replication fork.

Improvements in Sequencing Methods: Instead of radioactively labelling ddNTP's, they are now labelled with fluorescent dyes. Also now only one reaction tube is needed, not four (now each ddNTP can be labelled a different colour). Tiny capillaries are used instead of large slab gels. Also, there are many improvements in automation. You do not have to load sample, run gel, collect data, or read sequence by hand. Many genomes have been sequenced, and one amazing discovery has been the diversity of protists. The genome is an organisms entire DNA sequence, organelles. A large genome size does not mean more genes as there is non-coding DNA. Genome sequencing provides information about potential proteins, disease genes, and evolutionary histories and relationships. It can also indirectly give information on gene function, it helps to investigate function. The ensembl database has the human genome project. It has information on known or predicted genes, single nucleotide polymorphisms (mutations), regions of simple repetitive sequence and homologies. Issues for maintaining databases include: need to collect data (sequence info), need a web interface (to make data available), need to identify sequences (annotation), need to keep database up to date and need to coordinate it with other databases. The genome project is not complete. Genomes have been completed for archaea, bacteria and viruses. Gene function remains mostly unknown.

Lecture Seven (What/Where is a Gene):

A gene is a sequence of DNA that in transcribed into RNA. In general one gene encodes for one protein, but remember that there are two types of genes. Protein coding genes (encode for mRNA) and RNA coding genes (code for tRNA, rRNA ect.).

How to get Info on mystery sequence of DNA: First you want to perform a blast search. This finds similar sequences (to your sequence) in the blast database. Blast is an algorithm that uses short stretches of sequence similarity to find related genes in a database. The algorithm is fast and efficient. However, blast searches often turn up sequences that are entirely unrelated because only a short stretch of similarity is needed. It is too computationally expensive to look for long sequences. The blast search gives you an e-value and a score. A higher score and a lower e-value is a better match. These give you relative measures (comparisons only). Next, you will want to do a sequence alignment to see if your sequence aligns to a family of sequences. The software program for this is called ClustalW. ClustalW aligns sequences together by inserting gaps. The input file is a raw sequence data file and the output file is a multiple sequence alignment. The alignment is accomplished by an algorithm that clusters the most similar sequences with eachother. Changing ClustalW's parameters will change the sequence alignment. If ClustalW outputs an alignment this does not mean that the input sequences are related to eachother. Gaps in the alignment represent insertions or deletions of parts of the sequence.

Accuracy of Alignments: Alignments of similar sequences are more accurate. You can look for conserved motifs to judge accuracy of alignment (for example locations of prolines and/or disulfide bridges tend to be conserved, so you should expect these to line up in a sequence alignment). To assess sequence similarity you count the differences between two sequences at a time and calculate the percent difference. Sequence difference is the percent of sites different and sequence identity is the percent of sites that are identical. To know if a protein is not related to another you need other information such as structure. Also, if sequence identity is less than 30 percent, then proteins are probably not related.

Lecture Eight (continuation of seven):

The next thing we do is a phylogenetic analysis. The computer program for this is called Phylogenetic Analysis Using Parsimony or PAUP*. The input requirement is a multiple sequence alignment. Needs to be able to calculate the percent similarities and percent differences. Calculation is based on the idea that all sequences are derived from a single ancestral sequence. The output is a tree of phylogenetic relationships that shows in what order they

derived from eachother. The algorithm used by PAUP* minimizes the number of evolutionary steps along each lineage. It gives the tree with the highest probability of having occurred. The format required for analysis is NEXUS. You need to check the probability of the tree with statistical analysis. Need to do more extensive phylogenetic analysis using many more sequences.

Next you need to identify protein domains. This is done by a program called, Simple Modular Architecture Research Tool or SMART. It detects protein domains (finds structural or functional or both motifs) in an amino acid sequence.

(Check out slides 14-15 of lecture 8 to get smart results from example presented in class)

Usually when performing SMART you have some prior knowledge. For example, you may know the origin of tissue, usually the search is directed targeting certain genes/gene families using primers. Protein domains are found by the SMART algorithm through pattern recognition.

If by now you have identified your sequence, you want to determine if other animals have this gene. To investigate if other animals have the gene in their genome you use a Southern Blot. To investigate if they have the mRNA you use a northern blot and to investigate if they have the protein you would use a western blot.

To Perform a Southern Blot: First you cut the genome with restriction enzymes, which cut at specific locations. Then you electrophorese on agarose gel and then transfer the DNA onto a membrane. Next, you hybridize a labelled probe to the separated DNA, then detect the position of the labelled bands through use of X-ray film. For this you need highly specific restriction enzymes.
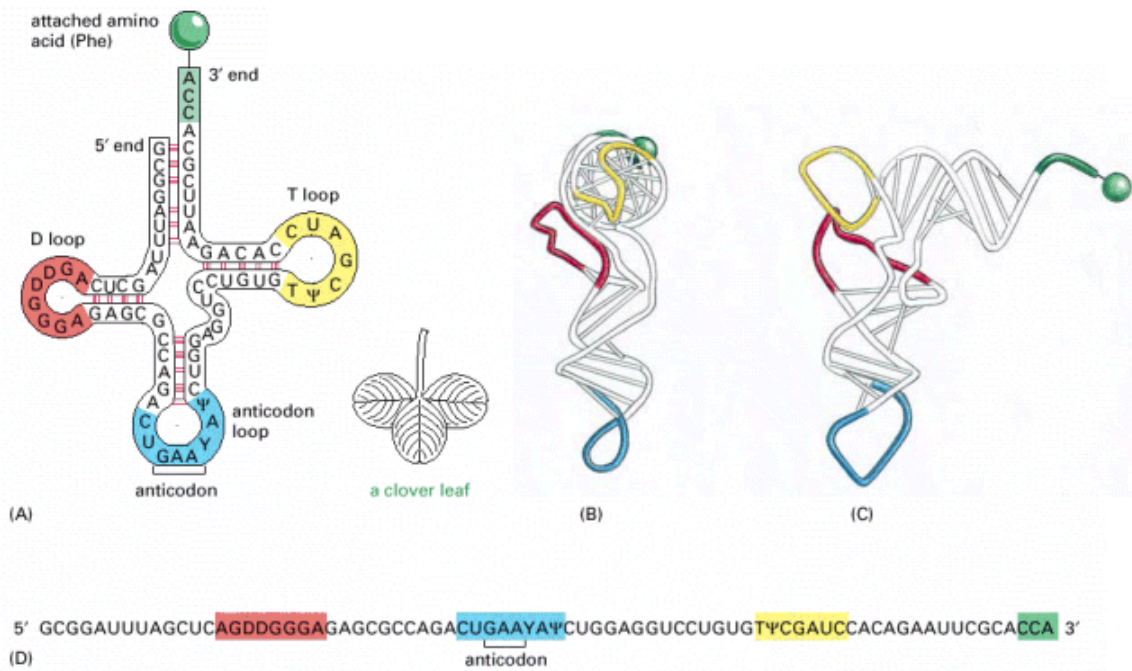
Role of Restriction Enzymes: They have specific recognition sequences and specific cut sites. Also have sticky and blunt ends. They leave 5 prime or 3 prime overhang on sequences.
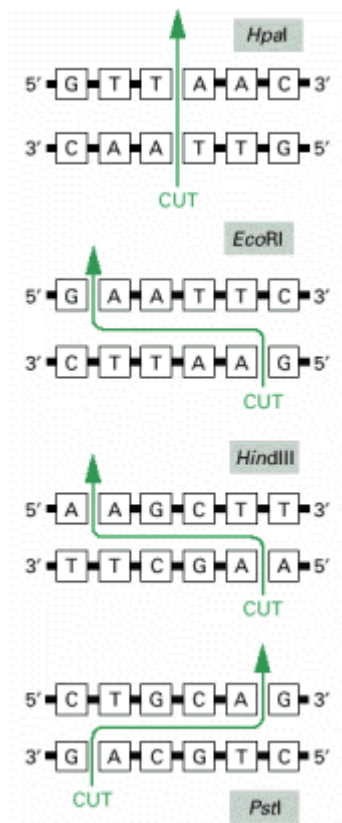
They are needed for southern blots to chop up genomic DNA. The probe for the southern blot is a gene of interest, the sequence that you are looking for. Different restriction enzymes cut at different sites.
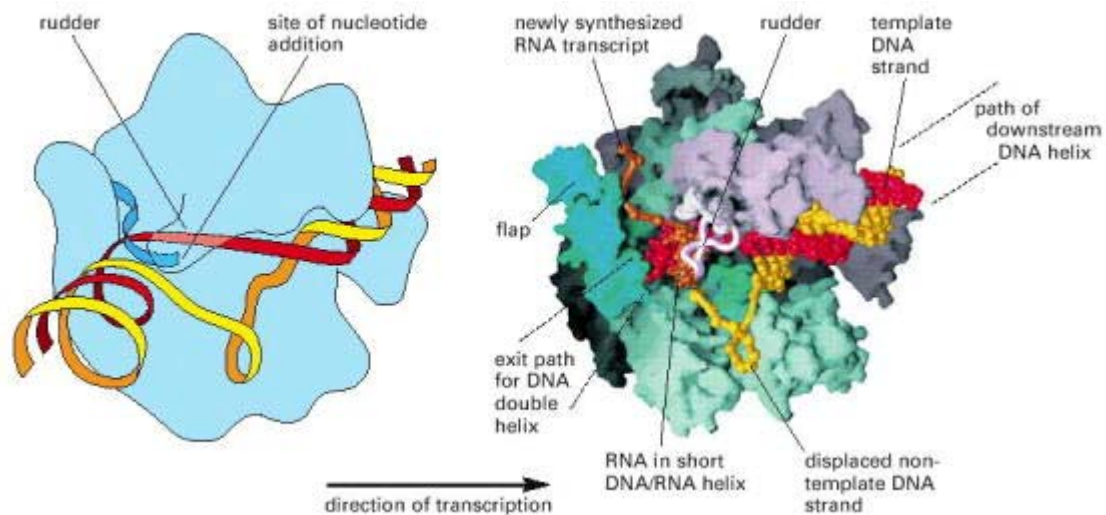
Lecture Nine (Prokaryotic Transcription):

RNA is transcribed into single strands and can exist in a variety of secondary structures such as loops. RNA is also capable of becoming double-stranded through base pairing. RNA secondary structure is important for the function of tRNA's, ribozymes/riboswitches and transcription termination. TRNA has an anti-codon loop, which recognizes its codon for an amino acid that it will then carry. It also has a D-loop and a T-loop. The amino acid is attached to the 3 prime end of the RNA chain. The structure is shaped like an upside down 3 leafed clover.

attached amino
acid (Phe)

3′ end

5′ end

T loop

D loop

C U A
G

GACAC
CUGUG T Ψ C

anticodon
loop

anticodon

a clover leaf

(A)

(B)

(C)

5′ GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUGAAYAΨCUGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCA 3′
(D)

anticodon

Riboswitches: Regulate gene expression. They rely on RNA secondary structure. An example is that of metabolites causing conformational change in RNA structure. When the metabolite is absent the structure of the RNA is such that transcription proceeds and proteins are expressed. When the metabolite is present it causes a conformation change in the RNA and transcription halts and proteins are not expressed.

RNA Polymerase: Catalyzes the formation of phosphodiester bonds that link the nucleotides together to from a linear chain. The "rudder" on the RNA polymerase helps to separate the RNA from the DNA.

rudder

site of nucleotide
addition

newly synthesized
RNA transcript

rudder

template
DNA
strand

path of
downstream
DNA helix

flap

exit path
for DNA
double
helix

RNA in short
DNA/RNA helix

displaced non-
template DNA
strand

direction of transcription

Transcription cycle:

1. Sigma factor (detachable subunit of RNA polymerase, reads promoter sequence on DNA) binds promoter
2. Sigma factor unwinds DNA so transcription can begin
3. Initial RNA synthesis occurs, very small strand of 10-15 bases. Initial start is very slow, and there is discarding of many small oligonucleotides.
4. Sigma factor releases
5. Structural change allows for fast RNA elongation with RNA polymerase
6. Termination (helped by hairpin loop) and release
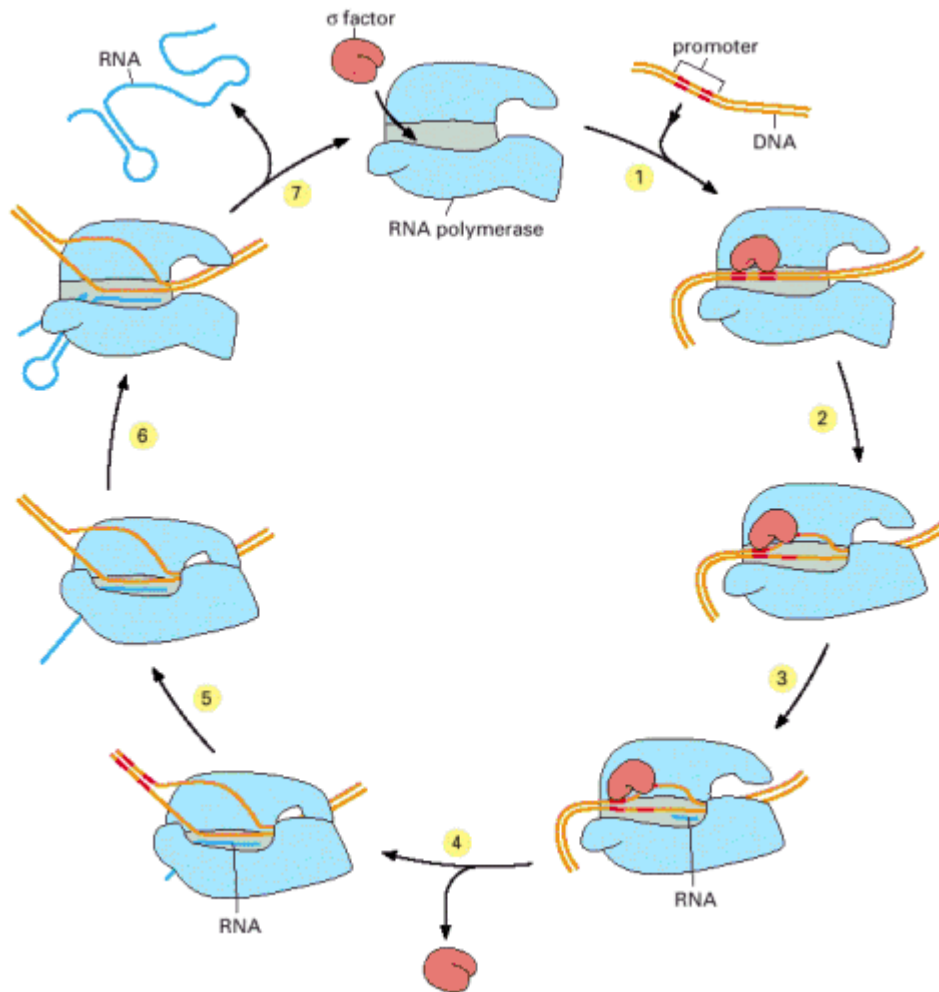
The RNA polymerase holoenzyme consists of a core plus various sigma factors. RNA polymerase does not require ATP to unwind DNA, it works by conformational changes to structures with more favourable energy. RNA is synthesized in the 5 to 3 prime direction. The initial steps of RNA synthesis are inefficient. Characteristics of RNA termination signals are hairpin loops and A-T rich sequences. The hairpin opens a flap that is holding the RNA transcript and this helps dissociate the transcript from the polymerase. An

exception to the rule about not needed ATP is the bacterial RNA polymerase closed complex. This uses ATP to initiate transcription.
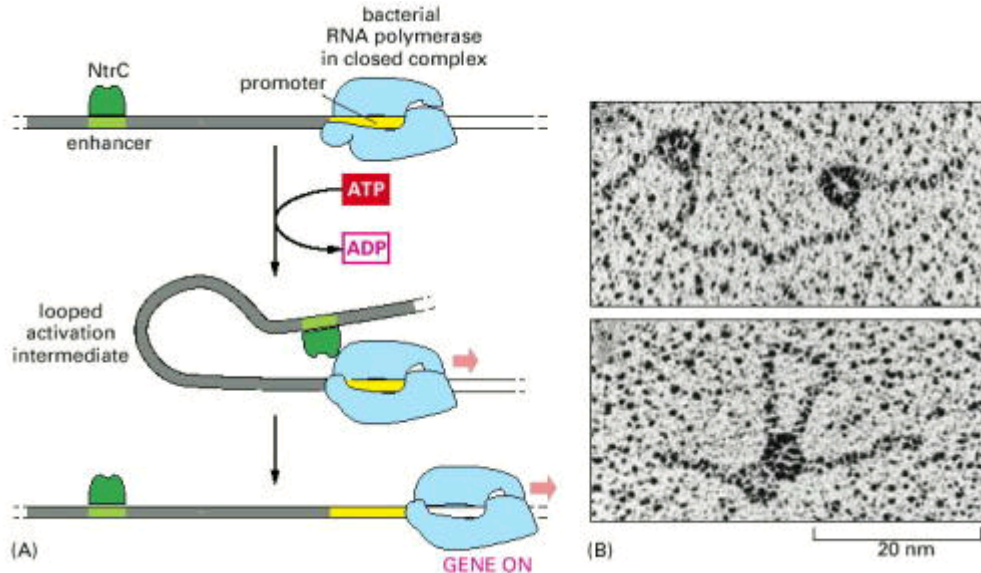
Promoters: are consensus sequences. They vary with organism but a lot of promoters are similar. In most cases, the promoter binds transcription factors. Promoter strength varies therefore promoter sequence also varies. The numbering for the start point of transcription is +1, site one. Promoter sequences are assymmetric because orientation determines which is the template strand. There is variation in sigma subunits. Most genes are recognized by sigma70. SEE TABLE 7-2. An RNA polymerase can bind to one sigma subunit. There is variation in sigma subunits due to regulation of sets of genes. Different sigma subunits have different consensus sequences.
Elongation of RNA transcript: The template is DNA. RNA is linked by phosphodiester bonds. The DNA-RNA hybrid is held together by base pairing. The RNA dissociates from the DNA as it leaves the polymerase, thus RNA's are single stranded.
Northern Blot: A procedure used to determine if a specific RNA transcript is present.
Steps:
1. Isolate RNA
2. Run Gel
3. Transfer to membrane
4. Hybridize the probe to membrane
5. Use X-ray film to detect fluorescent bands
The probe used is complementary to the RNA of interest. There is no need to cut the RNA with restriction endonucleases as is done with southern blotting

because the RNA is already in pieces. If successful the northern blot will tell you the size of the RNA transcript of interest. Northern blotting is usually used to detect mRNA.

LECTURE TEN (Eukaryotic Transcription):

In Eukaryotic transcription the mRNA transcript must be modified in the nucleus before let out in the cytoplasm. Start with primary mRNA transcript then remove the introns (non-coding sequences), then add 5 prime cap and 3 prime poly-A tail. In prokaryotes a single transcript can code for many genes. In eukaryotes a single transcript with introns, 5 prime cap and poly-A tail encodes one protein?
Eukaryotes possess many RNA polymerases:
RNA polymerase I (RNAPI) is found in the nucleolus and transcribes genes coding for some rRNA's.
RNAPII transcribes genes coding for all mRNA's and some snRNA's.
RNAPIII transcribes genes coding for tRNA's 5s-rRNA's and some snRNA's.
The phosphorylation of the C-terminal tail of RNAPII activates it and results in the binding of RNA processing proteins.
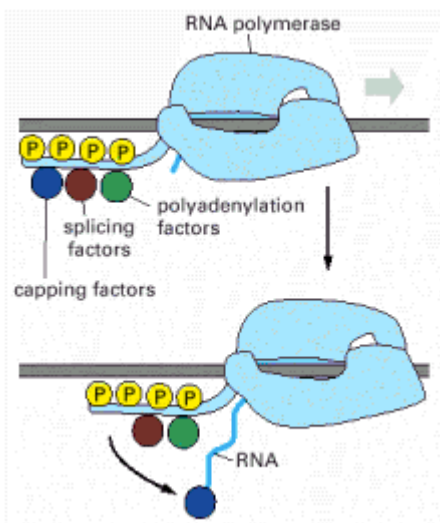
Subunits of RNA polymerases: RNAP's are complex structures with many subunits, some subunits are common to all three RNAP's. Some subunits resemble the subunits of bacterial RNAP's. The 3-D structure of RNAP in yeast and bacteria is very similar, suggesting that this enzyme is highly conserved.
Eukaryotic vs. bacterial RNAP's: eukaryotic RNAP's require proteins to help position them at the promoter, called transcription factors. These

transcription factors fulfil a similar role to the sigma subunit of bacterial RNAP's. Eukaryotic RNAP's need to deal with chromosomal structure and so they need more than one transcription factor.

Promoters: sequences recognized by transcription factors, which are required to initiate transcription. TATA box is a highly conserved promoter. It is found 25-36 base pairs upstream from start of transcription. Highly transcribed genes contain the TATA box, as it is a very strong promoter.
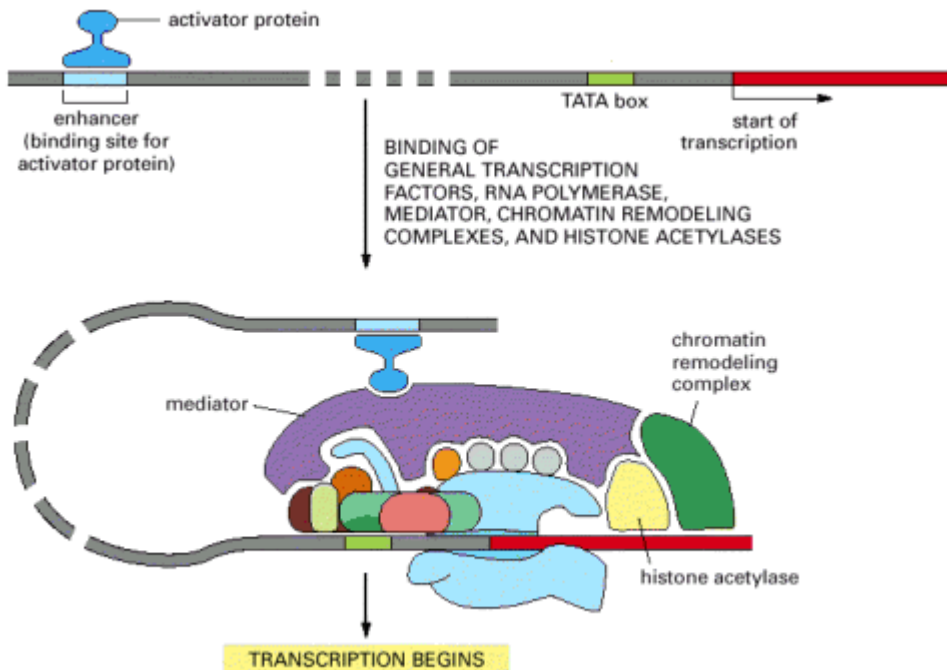
Steps in Initiation of Transcription:
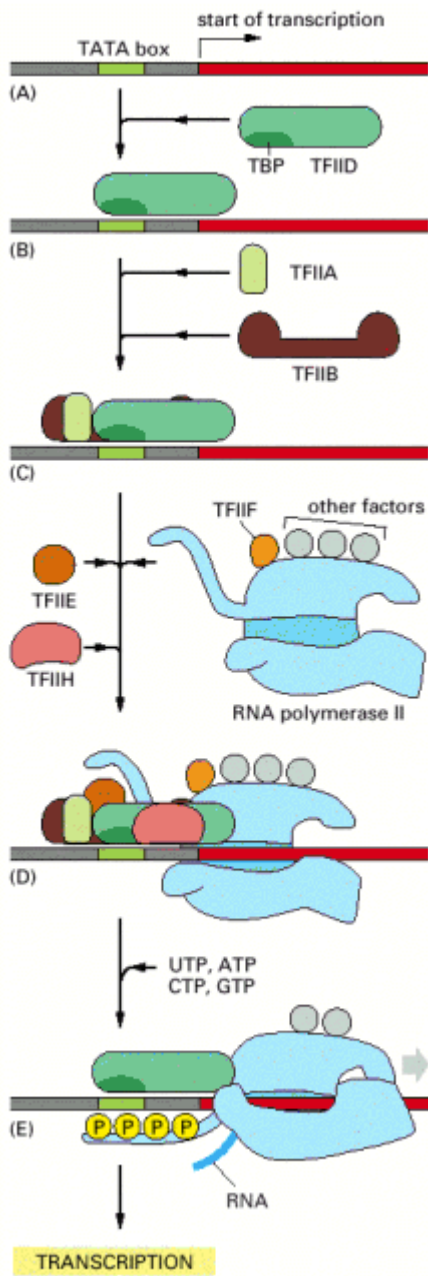
1. TBP (TATA binding protein), a subunit of TFIID (transcription factor D), binds to the TATA box promoter in the minor groove, bending the DNA
2. This attracts other transcription factors, which help to orient and bind RNAPII to the DNA
3. The helicase activity of TFIIH uses ATP to pry apart DNA strands at transcription start site.
4. TFIIH also phosphorylates the C-terminal tail of RNAPII, activating it so that transcription can begin.
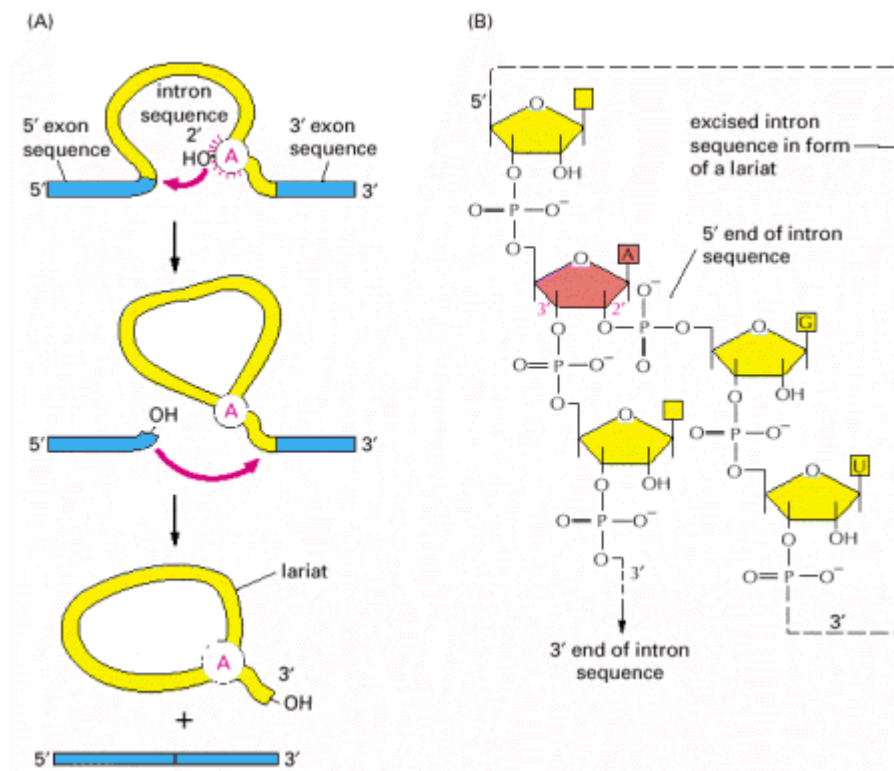
RNAPII C-terminal tail: In RNAPII only, the carboxyterminal domain of the largest subunit has a stretch of 7 amino acids that is repeated multiple times. The sequence is Tyr-Ser-Pro-Thr-Ser-Pro-Ser. The yeast enzyme has 26 repeats. The human enzyme has 52 repeats. This region is essential for viability.

Additional factors for transcription initiation: General transcription factors are also needed for initiation including an activator, an enhancer site, chromatin remodelling complexes, DNA loop and histone aceytlases. These will apparently be explained later as they have significance in regulation.

RNAPII is activated by phosphorylation (chemical addition of phosphate groups on Y/S). Over 100 subunits are involved in the initiation of eukaryotic transcription.

RNA processing: 5 prime RNA capping helps to protect the mRNA from degradation. The cap is added by removing a phosphate group, adding GMP (backwards) and then adding a methyl group to G base.

(A)

intron sequence
5' exon sequence
2'
HO A
3' exon sequence
5'
3'

OH A
5'
3'

lariat
A
3'
OH
+
5'
3'

(B)

5'
excised intron sequence in form of a lariat

5' end of intron sequence

3' 2'

3' end of intron sequence

MRNA splicing: Removes introns.
1. Adenosine attacks 5 prime slice site
2. 3 prime end of one exon reacts with 5 prime end of next exon to release intron. The extra OH group is extremely important for covalent bond. This OH group, which is only found in RNA nucleotides in necessary for the formation of the lariat structure in intron splicing.

There are certain sequences required for intron removal. For these see FIGURE 6-28 or slide 21 of lecture 10. Y represents C or U and R represents G or A.
snRNAP's plus proteins can base pair and recognize consensus sequences. Pre-mRNA splicing is performed by the spliceosome (large assembly of RNA and protein molecules that performs pre-mRNA splicing in eukaryotic cells).

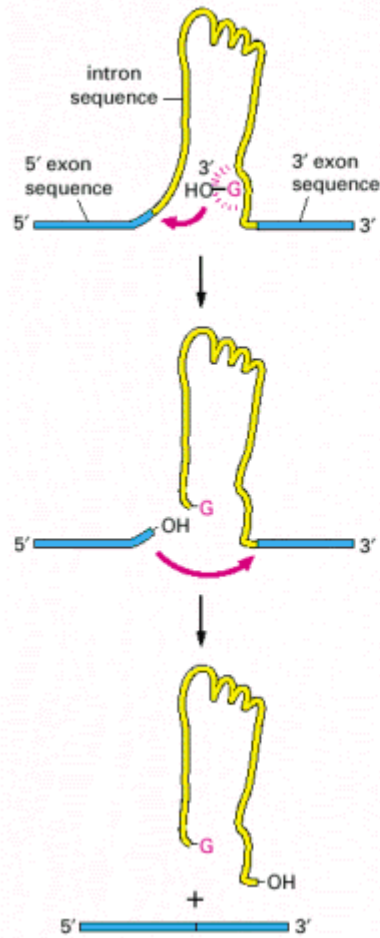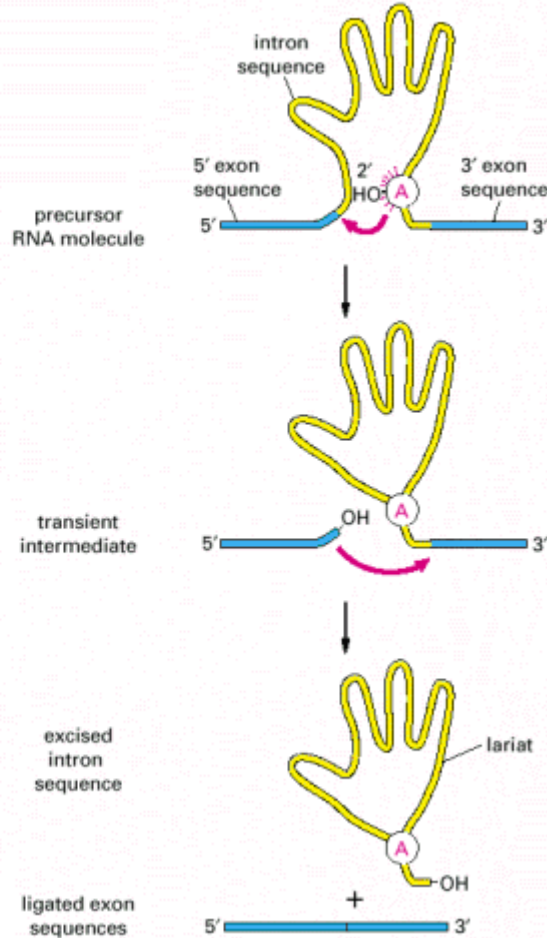**Lecture Eleven (Processing RNA to make mature mRNA):**

Types of RNA splicing:
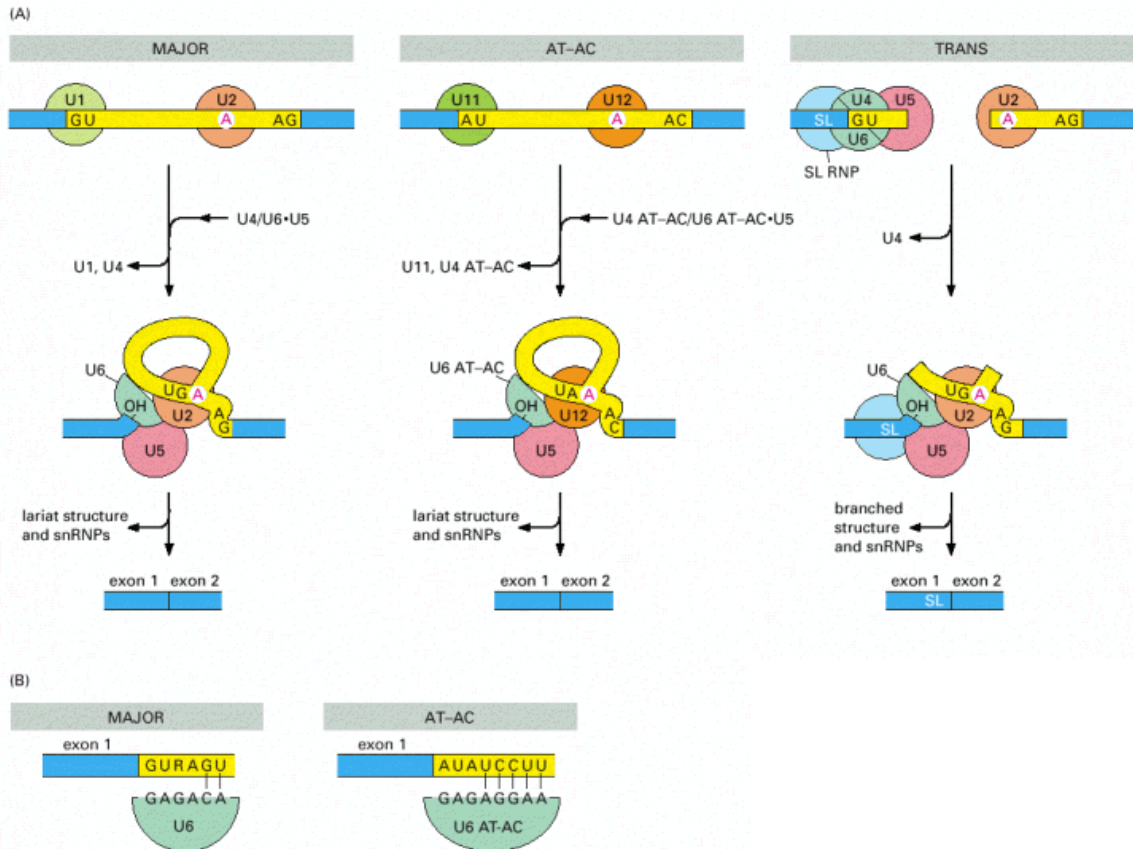
Major type is found in most systems, consensus site for slicing is GU and AG. See part A of diagram. AT-AC type is found in 0.1 percent of systems. The consensus sites are slightly different, AU instead of GU and AC instead of AG. Trans type is found in worms and trypanosome. It's the splicing together of two completely separate transcripts. Fuses same exon one onto multiple genes.

Self-splicing introns: 3-D structure of intron encourages catalysis. Group one self-slicing introns involves G and group two involves A. The spliceosome probably evolved from type two self-splicing introns. This mechanism of self-splicing is extremely inefficient without spliceosomes and helper proteins.
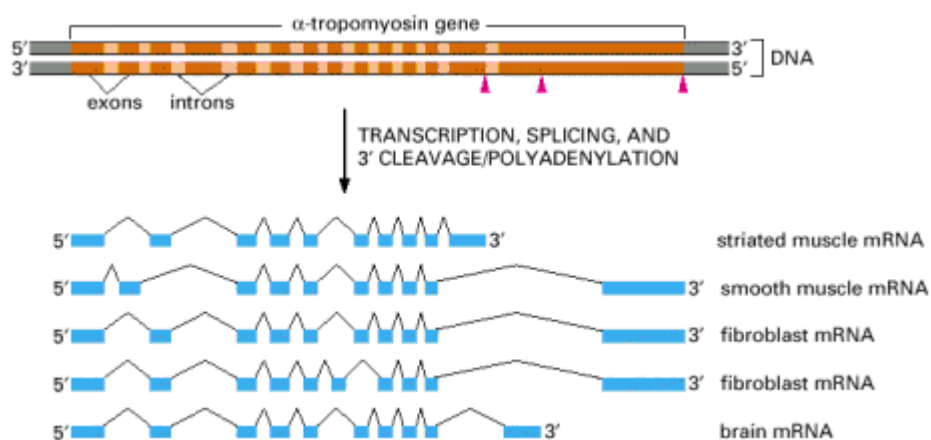
intron sequence

intron sequence

5′ exon sequence   3′ exon sequence

5′ exon sequence   3′ exon sequence

precursor RNA molecule

transient intermediate

excised intron sequence

lariat

ligated exon sequences

Alternative splicing can generate variation, which can lead to multiple functions.

α-tropomyosin gene

DNA

exons   introns

TRANSCRIPTION, SPLICING, AND 3′ CLEAVAGE/POLYADENYLATION

striated muscle mRNA

smooth muscle mRNA
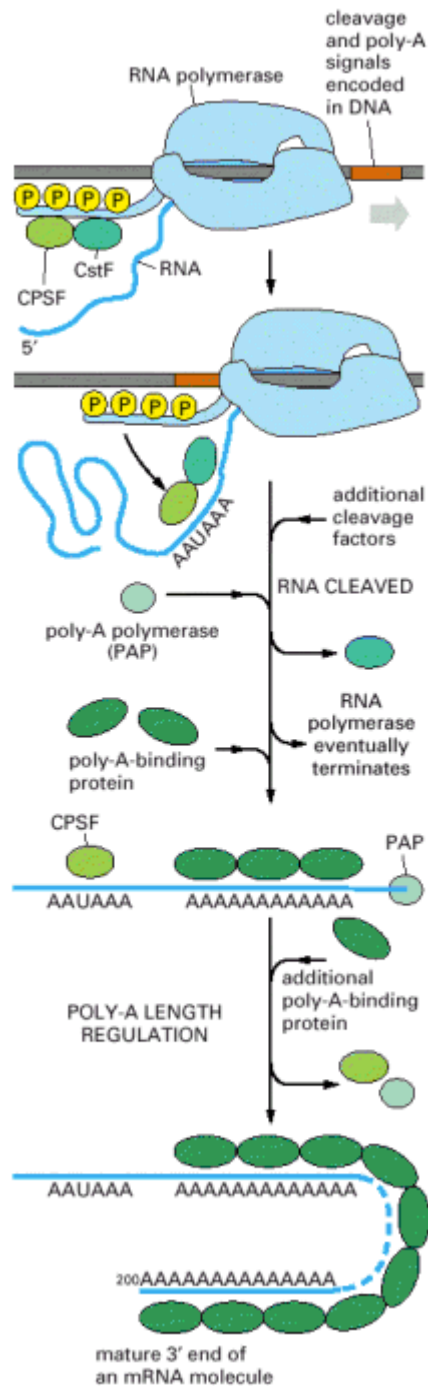
fibroblast mRNA

fibroblast mRNA

brain mRNA

Reasons for introns: Possible evolution of new proteins, alternative splicing. Finding genes is difficult because of introns (interruptions in coding

sequence, difficult to tell where splice site is—alternative splicing confounds things further.) Incorrect splicing leads to frameshifts in the coding sequence and truncated proteins that do not function properly.

Processing of 3 prime end:
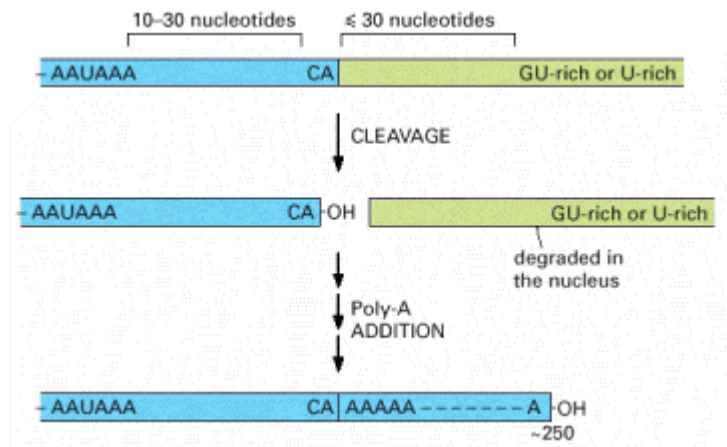
Blue section is 10-30 nucleotides. AAUAAA binds CPSF. The diagram shows the most common version of the sequence of nucleotides.

As RNA polymerase transcribes the DNA it encounters a cleavage poly-A signal. FIGURE 6-38 shows many green poly-A binding proteins.

cleavage and poly-A signals encoded in DNA

RNA polymerase

CstF   RNA

CPSF

5′

additional cleavage factors

RNA CLEAVED

poly-A polymerase (PAP)

RNA polymerase eventually terminates

poly-A-binding protein

CPSF   PAP

AAUAAA   AAAAAAAAAAAA

POLY-A LENGTH REGULATION

additional poly-A-binding protein

AAUAAA   AAAAAAAAAAAA

200AAAAAAAAAAAAAA

mature 3′ end of an mRNA molecule

Poly-A polymerase differs from other RNA polymerases because a template is not required. Cleavage and poly-A specificity factor binds AAUAAA. Cleavage stimulation factor F binds GU rich element. CPSF and CstF are located at the C-terminal of RNAPII. Poly-A polymerase adds approximately 200 A's to the 3 prime end of the transcript. The addition of a poly-A tail increases mRNA stability.

Functions for ribonucleoprotein (RNA's, proteins produced in the nucleolus) synthesis: Produces ribosomes, mRNA is only a few percent of the total RNA transcribed in the cell. 80 percent is rRNA. A typical mammalian cell contains 10 million ribosomes. Also produces telomerases, which are composed of proteins and RNA's.
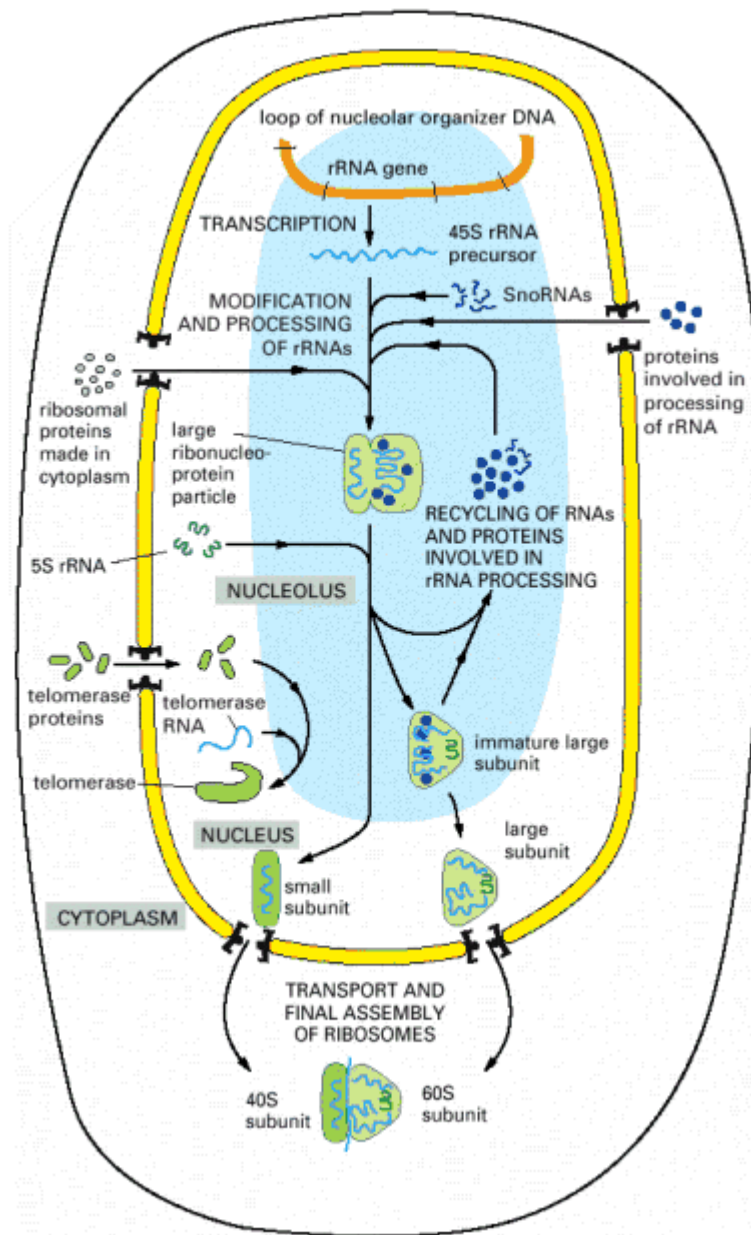
RNA WORLD:
Before cells contained DNA, they only had RNA. An RNA like molecule is thought to have catalysed its own synthesis (this does not happen now). This is one explanation to the paradox of DNA coding for protein but needing protein to transcribe DNA (which came first…). For an RNA world to occur

the molecule must be able to serve both as a template and a catalyst. Present day RNA is probably too complicated to catalyse its own synthesis correctly because there are too many competing reactions. A possible other candidate for a template/enzyme is PNA, it could have catalyzed its own reaction but does not exist now.
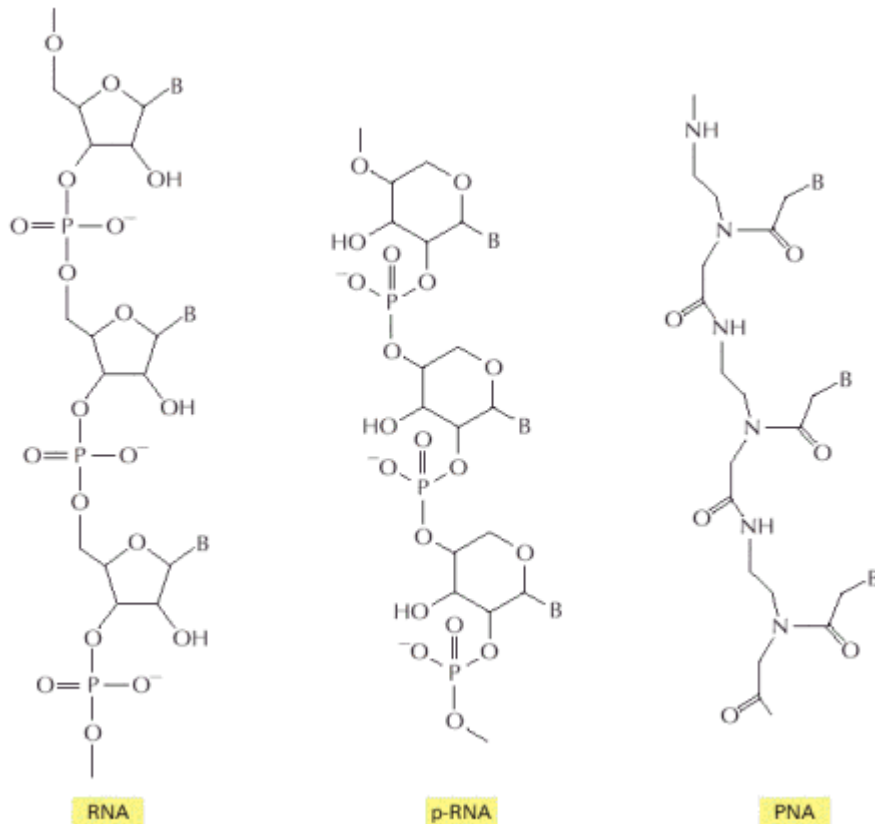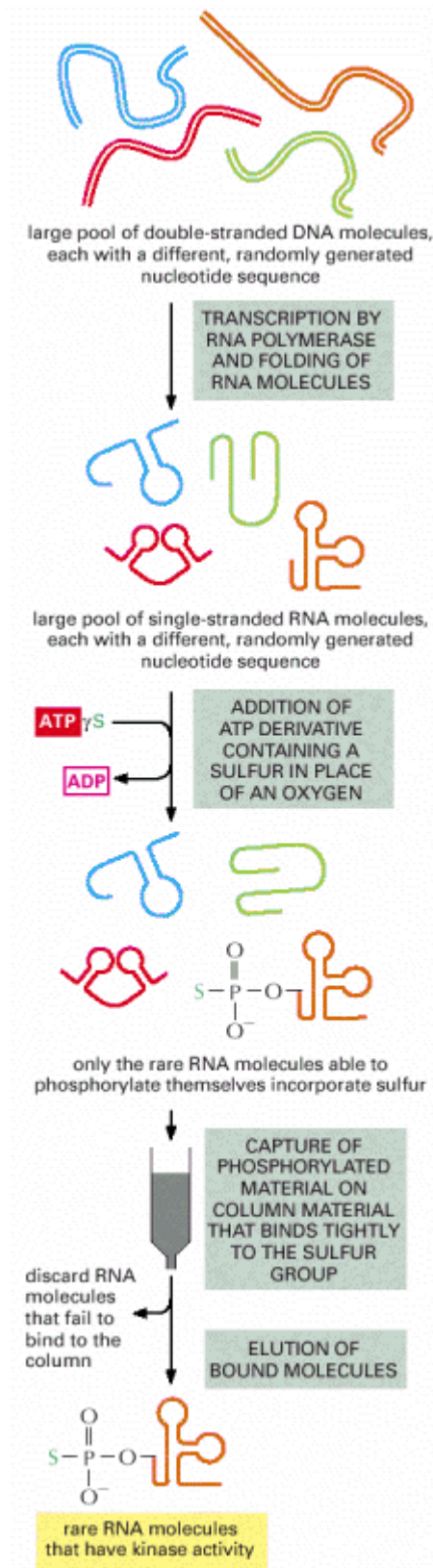

RNA   p-RNA   PNA

FIGURE 6-93 shows possible early RNA like molecules. The first one is RNA and it can base pair. The second one is p-RNA, which has fewer competing sites. The third one is PNA, modern day nucleic acid base on peptide backbone—can still base pair.

Ribozymes are RNA acting as enzymes. They can make peptide bonds (active site composed of RNA), perform RNA splicing, DNA cleavage and much more. You can use various experiments to test for rare RNA molecules that possess a specified catalytic activity. Do an in vitro selection of synthetic ribozymes.

large pool of double-stranded DNA molecules, each with a different, randomly generated nucleotide sequence

TRANSCRIPTION BY RNA POLYMERASE AND FOLDING OF RNA MOLECULES

large pool of single-stranded RNA molecules, each with a different, randomly generated nucleotide sequence

ATP γS

ADP

ADDITION OF ATP DERIVATIVE CONTAINING A SULFUR IN PLACE OF AN OXYGEN

$$S-P-O$$

only the rare RNA molecules able to phosphorylate themselves incorporate sulfur

CAPTURE OF PHOSPHORYLATED MATERIAL ON COLUMN MATERIAL THAT BINDS TIGHTLY TO THE SULFUR GROUP

discard RNA molecules that fail to bind to the column

ELUTION OF BOUND MOLECULES

$$S-P-O$$

rare RNA molecules that have kinase activity

In this figure ATP with S instead of O is placed in a test tube with RNA molecules. Only those RNA molecules that are able to phosphorylate

themselves will be able to incorporate the sulfur. Can get rid of the other RNA molecules by running everything down a column that contains a matrix with something that specifically bind sulfur and eluting out all other RNA's.

Lecture Twelve (Molecular Cloning Methods):

PCR:
Ingredients needed are: A template, primers, Taq, dNTP's, MgCl2, and buffer. First you must denature the template, then anneal the primers to the DNA and let elongation occur, then repeat this many times. Sometimes the second and the third steps are combined. The optimal temperature for the process is about 74°C. Hybridization of primers onto both separated strands occurs so you end up with two fragments of double stranded DNA for the one fragment that you started with.
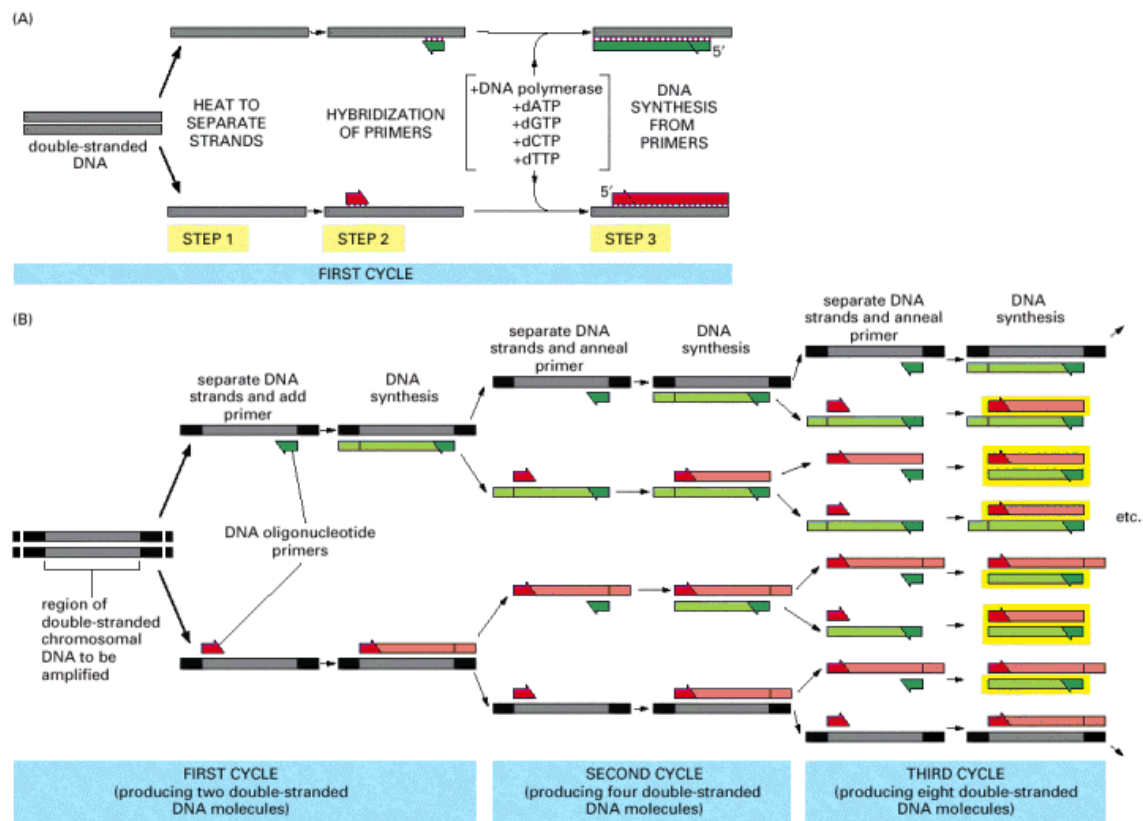
Because you are annealing a primer to each strand of the template you get exponential amplification of the DNA. You get 2 to the power of whatever cycle you are on DNA's. So if you are on cycle 30 you get 2^30 DNA fragments. A typical PCR takes 2-3 hours.

Choosing primers: Because DNA polymerase can add subunits only to the 3 prime end of the primer, the primer has to be situated upstream (more 5

prime than) from the sequence to be copied. To denature the template DNA the temperature should be raised to 92-94ºC. Then to anneal primers temperature must be lowered, then use optimal temperature for the third step. The final product of PCR reaction includes primer sequences but these are not products of replication and therefore are not subject to error correction. They just need to anneal to the template.

Isolating a gene from a species: you want to avoid using degenerate primers. The genetic code is degenerate remember, more than one codon is used to code for each amino acid. A thermostable polyermase is needed for PCR because it has to make it through the denaturation step. More stringent PCR annealing conditions include having the correct temperature and including Mg to avoid mispriming. Mg affects how primers bind and their specificity. It also affects the function of the primer. For the codons there are different levels of degeneracy. There is 1,2,3,4 and 6-fold degeneracy. For example 6-fold degenerate means there are six different condons that code for the same amino acid.

A degenerate primer: TGGGACGGCCAG, TGGGATGGTCAA, TGGGATGGCCAG and TGGGACGGGCAA all code for the same protein sequence—W,D,G,Q. Primer can be re-written as TGG GAY GGN CAR. Where Y stands for C/T, N stands for any nucleotide and R stands for G/C. The fold degeneracy of this primer is 16-fold.

Steps in degenerate primer design:
1. Download related gene sequences from other organisms using Genbank.
2. Align sequences using ClustalW.
3. Select primer sites.

Some factors to consider when choosing primers: Good primers to choose are from highly conserved regions of sequence. Want to minimize primer degeneracy. It is easier to amplify smaller fragments. Want reverse compliment for reverse primers. Want to choose gene portion of interest (may not be interested in the whole gene). Need to think about introns (if isolating gene from genomic DNA).

Isolate DNA template from organism: Steps:
1. Lyse cells
2. Bind DNA to membrane
3. Wash other cellular components through and discard
4. Elute DNA from membrane.

PCR results from agarose gel: On your gel one column is a DNA ladder, which is used as a size standard. The next column beside that is for your cDNA from tissue, the next column is the positive control, then the negative control and then the DNA ladder again. From the gel you can interpret results to have information about size and whether you have the gene of

interest or a contaminate. Negative results are interpreted using the positive control. To determine the size of your unknown sequence you need to run a size standard (or DNA ladder) down the gel. Positive control help to interpret negative PCR results (ie. No band). Negative controls help you tell if your PCR result is a result of contamination. If you get a PCR product, sequencing it will tell you if it is your gene of interest.

Cloning: Steps:
1. Ligate amplified fragment into pre-cut plasmid DNA.
2. Transform bacterial cells with vector.
3. Grow cells on media with antibiotic.

Plasmids are naturally occurring cDNA found in bacteria in addition to its single circular chromosome. They usually provide some benefit to bacterium—antibiotic resistance. Modified plasmids called cloning vectors are often used in the laboratory.

Lecture Thirteen (Translation):

Two essential steps ensure fidelity in translation. TRNA synthetase attaches correct amino acid to tRNA. Base pairing in translation—base pairing is correct for peptide bond. The synthetase includes a site for editing if the wrong amino acid is added. This is called hydrolytic editing.

Ribosomes: Are composed of over 50 proteins plus rRNA's. Structure includes the large ribosomal subunit and the small ribosomal subunit. There are three sites on the ribosome the E-site, the P-site and the A-site. Only two sites are occupied at one time. Ribosomes are located in the endoplasmic reticulum and the cytosol.

Protein Synthesis:
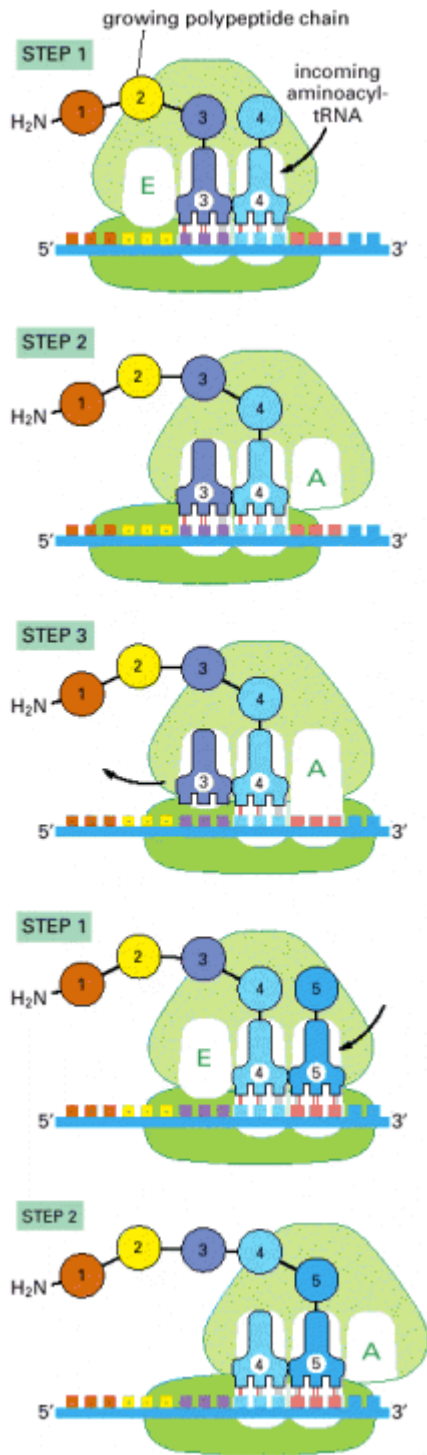
The A-site is occupied by the incoming aminoacyl tRNA. In the P-site the amino acid has formed a peptide bond with the peptide chain. Ribosome moves down so that the E-site is filled and the A-site is free then the cycle repeats.
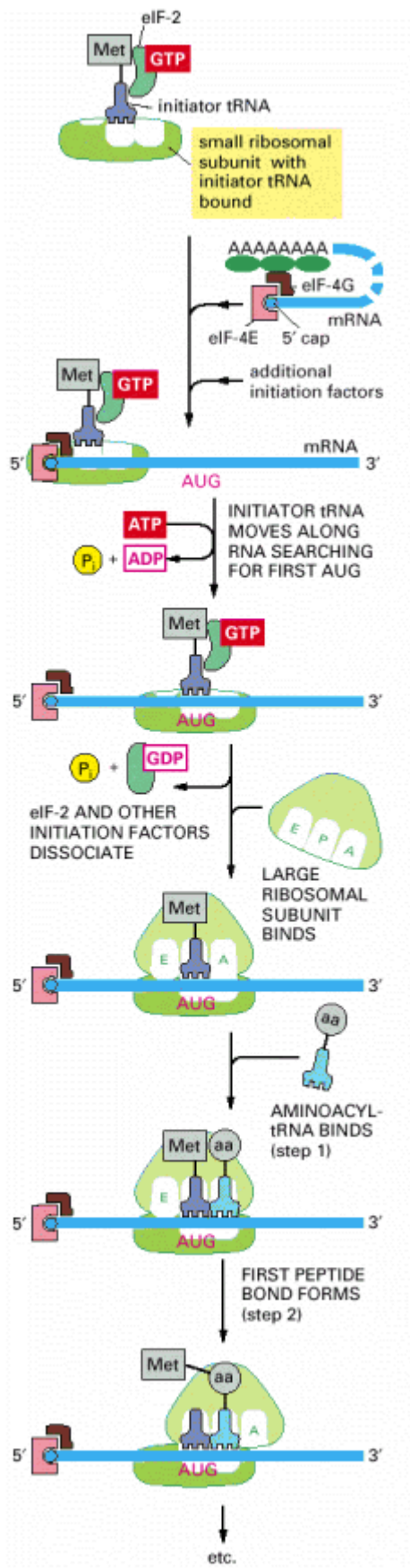
Quality control: Elongation factors that occur in bacteria slow formation of peptide bond until correct pairing of condon/anti-codon occurs. Second

elongation factor helps ribosome move forward on chain so that a new amino acid can enter. Ribosomes can perform protein synthesis without the help of elongation factors. The role of the elongation factors is to speed up the process making it more efficient and to perform error-checking. These are mediated by GTP hydrolysis. EF-Tu binds tRNA. EF-G helps ratchet forward one codon.
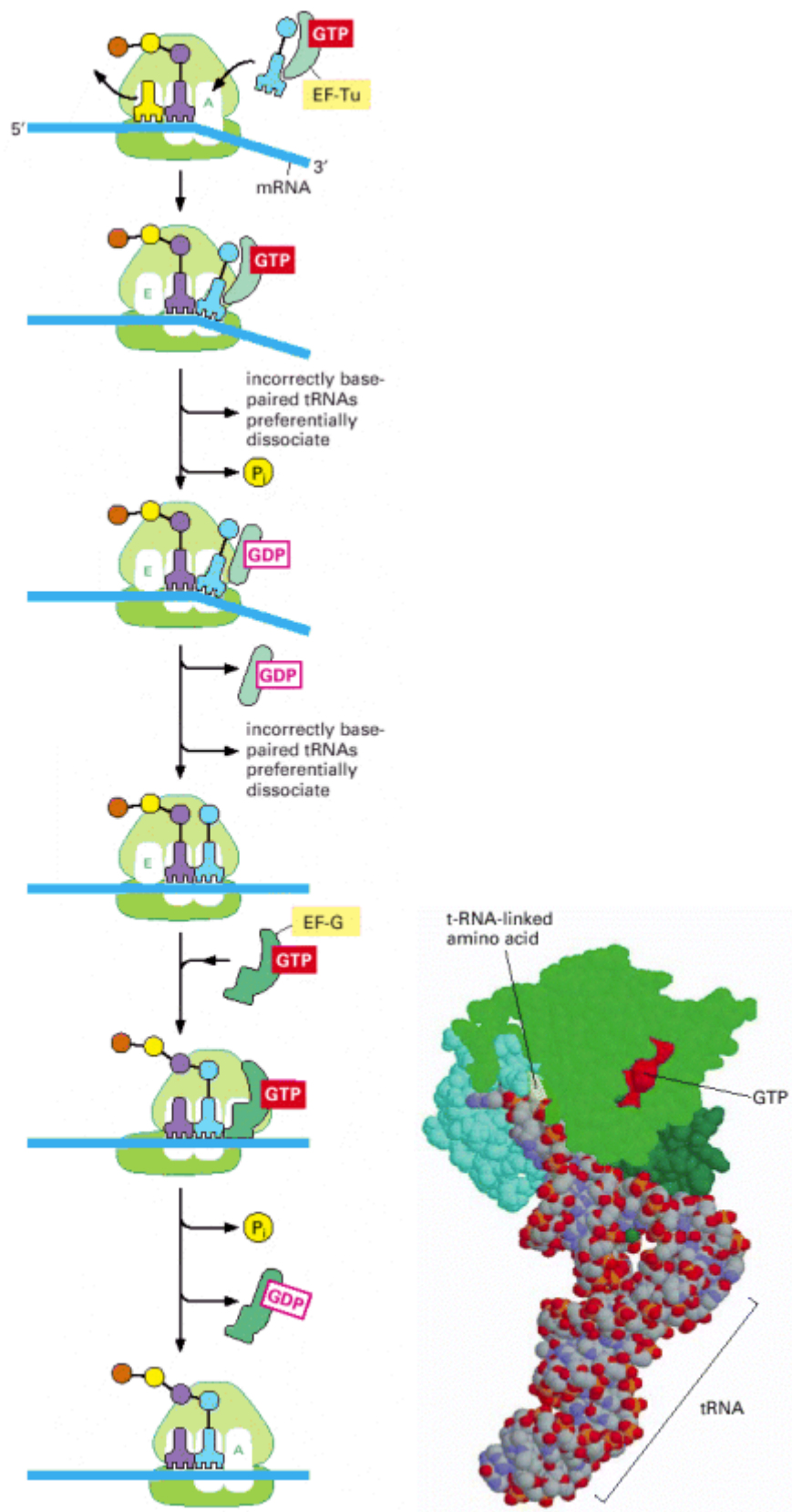
The large rRNA subunit of bacteria has 6 domains.

<u>Initiation of translation</u>: Initiator tRNA always encodes a methionine. All proteins start with methionine.

eIF-2

Met GTP

initiator tRNA

small ribosomal subunit with initiator tRNA bound

AAAAAAAA

eIF-4G

eIF-4E 5′ cap

mRNA

Met GTP

additional initiation factors

5′ mRNA 3′

AUG

ATP

P_i + ADP

INITIATOR tRNA MOVES ALONG RNA SEARCHING FOR FIRST AUG

Met GTP

5′ 3′

AUG

P_i + GDP

eIF-2 AND OTHER INITIATION FACTORS DISSOCIATE

E P A

LARGE RIBOSOMAL SUBUNIT BINDS

Met

5′ E A 3′

AUG

aa

AMINOACYL-tRNA BINDS (step 1)

Met aa

5′ E 3′

AUG

FIRST PEPTIDE BOND FORMS (step 2)

Met aa

5′ A 3′

AUG

etc.

Translation initiation is important for determining the reading frame. Initiation requires initiator tRNA and eIF's (eukaryotic initiation factors). In prokaryotes the Shine-Dalgarno sequence tells start of translation because they do not have 5 prime cap to tell start of synthesis. In eukaryotes the signal for the start of translation is the intact cap and tail.

Termination of transcription is signalled by the stop codon (UAA, UAG, UGA). The termination of translation is aided by release factors. They bind to any ribosome with a stop codon positioned in the A-site. The release factor performs a reaction that frees the carboxyl end of the growing polypeptide chain from its attachment to a tRNA molecule. Since only this attachment normally holds the polypeptide chain to the ribosome, the completed protein chain is immediately released to the cytoplasm. Release factors are an example of molecular mimicry (where one macromolecule resembles the shape of a chemically unrelated molecule). Release factors (which are made entirely of protein) bear an uncanny resemblance to the shape and charge distribution of a tRNA molecule.
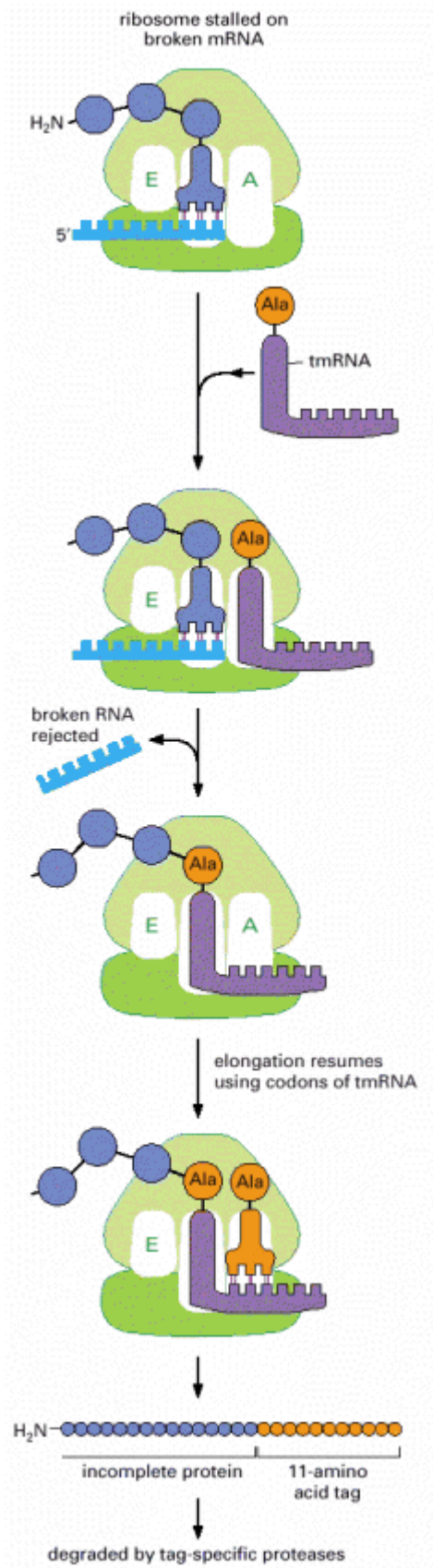
Protein synthesis is generally slow. To make synthesis faster many ribosomes are used—polyribosomes. Typical spacing between ribosomes is 80 nucleotides.

Quality control in eukaryotes:

Before translation can begin intact ends are needed on the mRNA molecule. TRNA synthetase has error correction site and can take out misincorporated amino acid—charging tRNA's. Within the ribosome there is GTP hydrolysis.

Quality Control in prokaryotes:

Used when ribosome stuck on broken mRNA molecule. Uses tmRNA hybrid molecule that has no anti-codon but has attached amino acid on the end. It attaches 11 amino acid tag (alanine) onto the end of the polypeptide to give signal to delete protein.

ribosome stalled on
broken mRNA

H_2N

E   A

5'

Ala

tmRNA

Ala

E

broken RNA
rejected

Ala

E   A

elongation resumes
using codons of tmRNA

Ala Ala

E

H_2N—

incomplete protein   11-amino
acid tag

degraded by tag-specific proteases

REVIEW: RIBOSWITCHES

They usually occur in mRNA upstream of coding sequence. They are used to regulate gene expression by regulating transcription or regulating translation.
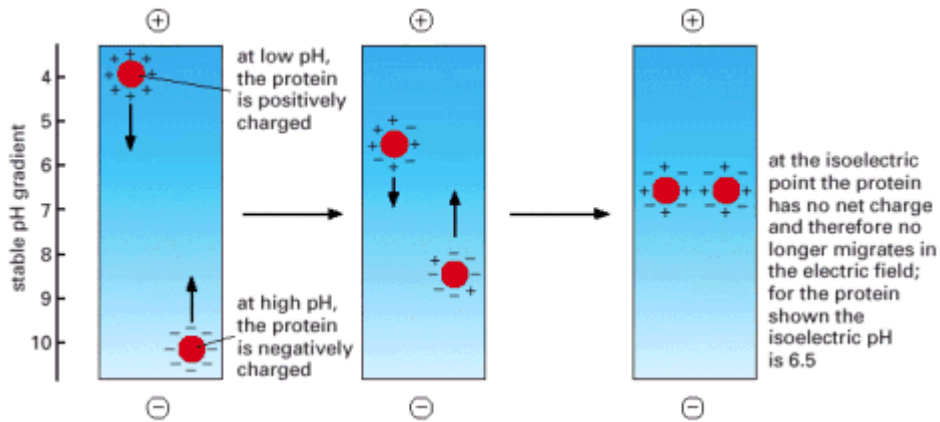
Prof. Campbell Lecture Nine (Proteome):

Differences in the proteome (the complete set of proteins produced by the transcriptome at any one time) define cells, tissues and organisms. The proteome is dynamic and changes under certain conditions (different cell, tissue, condition, time of day etc.) due to different patterns of gene expression. The study of the proteome is termed proteomics.
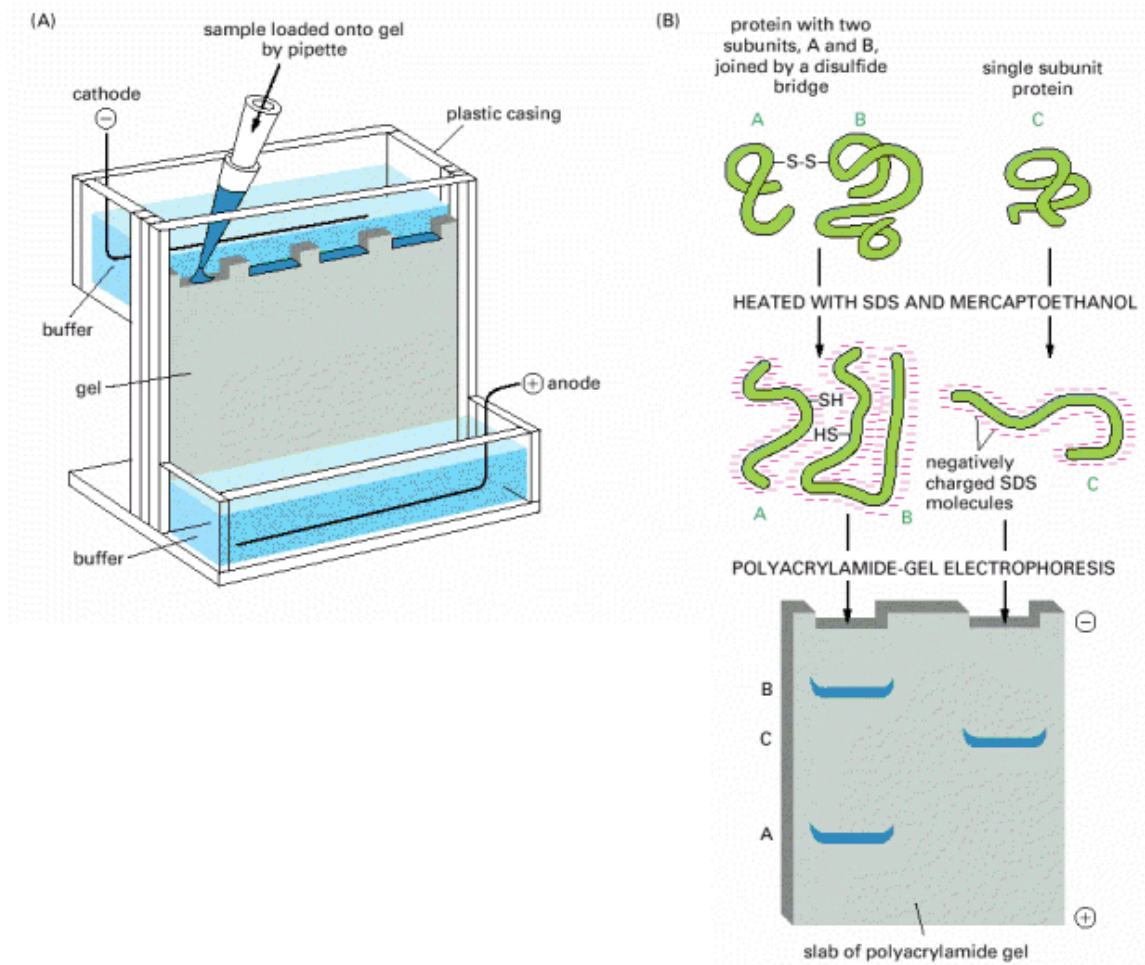Differences in the proteome can be detected using 2-D polyacrylamide gel electrophoresis. In this method the proteins are first separated on the basis of their isoelectric points by isoelectric focusing from left to right (on the gel) using a stable pH gradient. If the pH is above the isoelectric point the protein will have a negative charge. If the pH is below the isoelectric point of the protein it will have a positive charge. At the isoelectric point the protein is neutral and will stop migrating in the gel. The proteins are then differentiated based on molecular weight (from the top to the bottom of the gel) doing electrophoresis in the presence of SDS. The SDS denatures the protein, turning it inside out and this makes sure the mass/charge ratio of all the proteins is fixed so that they are separated solely on size. This method is called 2-D because you separate from left to right and from top to bottom.
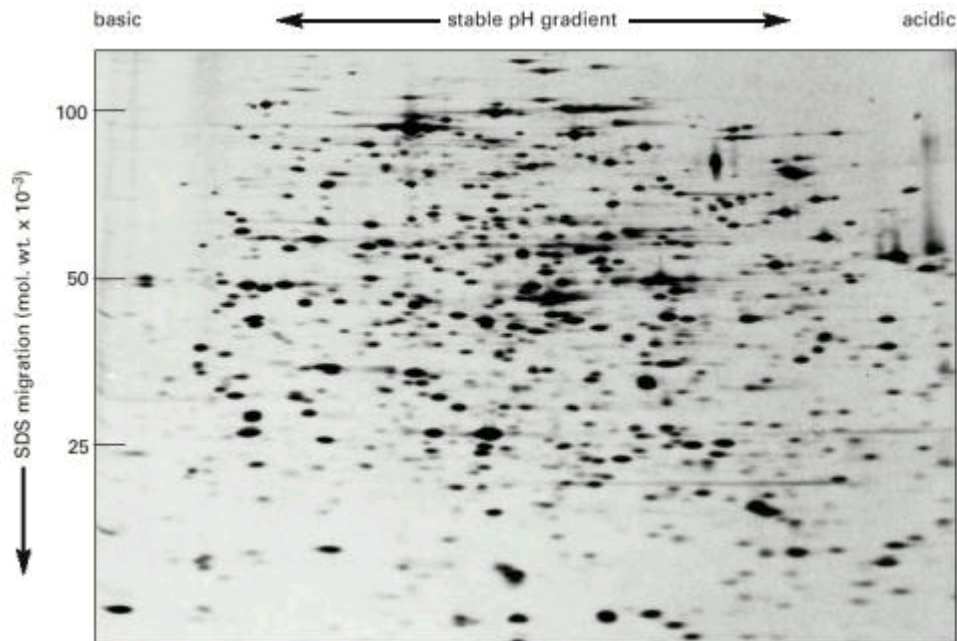
The first dimension in the 2-D Page is Isoelectric focussing. Every protein has a characteristic isoelectric point, the pH at which the protein has no net charge and therefore does not migrate in an electric field. In isoelectric focusing, proteins are separated electrophoretically in the polyacrylamide gel in which a gradient of pH has been established by a mixture of special buffers. Each protein moves to a position in the gradient that corresponds to its isoelectric point and stays there creating a band in the gel.

stable pH gradient

at low pH, the protein is positively charged

at high pH, the protein is negatively charged

at the isoelectric point the protein has no net charge and therefore no longer migrates in the electric field; for the protein shown the isoelectric pH is 6.5

The second dimension in 2-D page is the SDS-page. In the second step, the gel containing the proteins is again subjected to electrophoresis but in a direction that is at a right angle to the direction that is used in the first step (up and down instead of left to right). This time SDS is added and the proteins are separated according to their size.

(A) cathode; sample loaded onto gel by pipette; plastic casing; buffer; gel; anode; buffer

(B) protein with two subunits, A and B, joined by a disulfide bridge; single subunit protein; A B C; –S-S–; HEATED WITH SDS AND MERCAPTOETHANOL; SH; HS; A B; negatively charged SDS molecules; C; POLYACRYLAMIDE-GEL ELECTROPHORESIS; B; C; A; slab of polyacrylamide gel

Proteins can be rapidly characterized by combining digestion (with enzymes such as chymotrypsin or trypsin), mass spectrometry and bioinformatics. First you cut a single peptide spot from your gel that was used in the 2-D page. This protein is then digested with trypsin. The peptide fragments obtained from this are loaded into the mass spectrometer and their masses are measured. Sequence databases are then searched to find the gene that encodes a protein whose calculated tryptic digest profile matches these values.
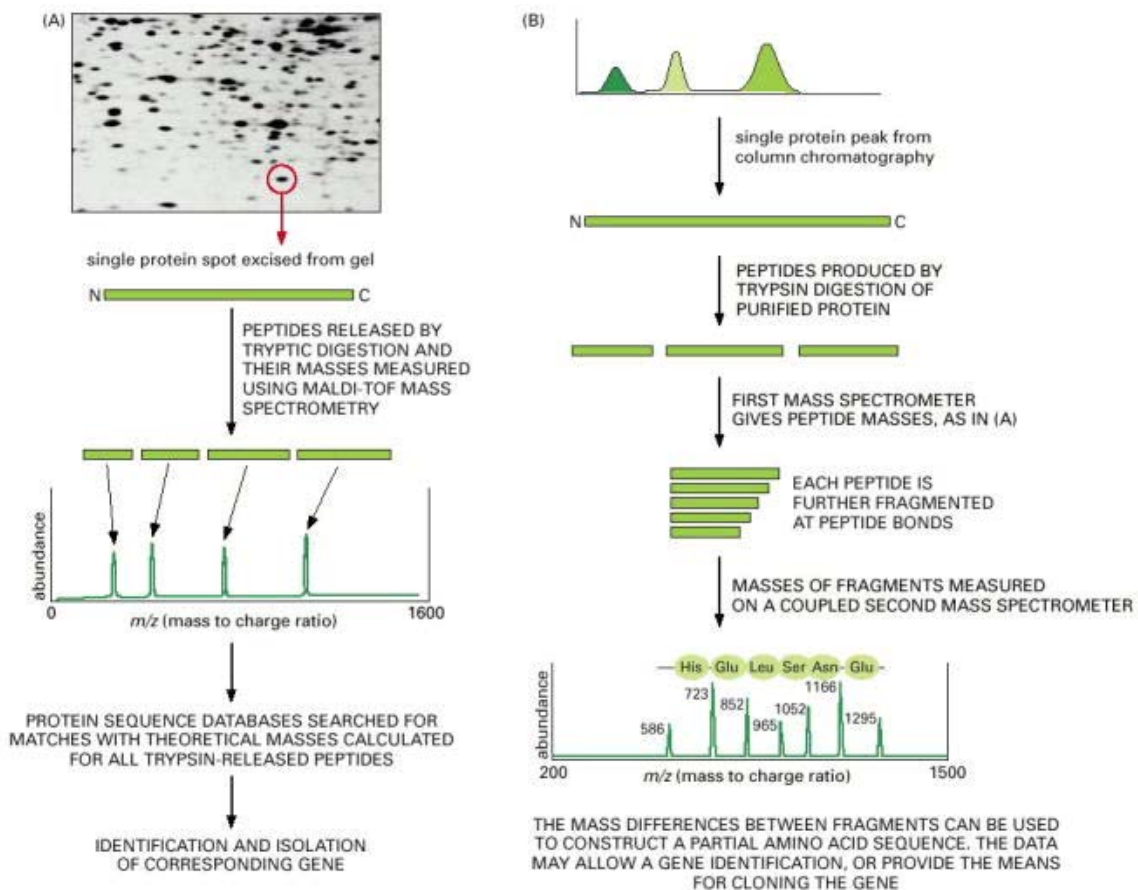
SEE FIGURE 8-20 A

Mass spectrometry can also be used to determine directly the amino acid sequence of peptide fragments. In this case, we again cut out a protein from our 2-D page gel and digest it with trypsin. The masses of the resulting fragments are again determined by mass spectrometry as in the first case. To determine the exact sequence, each peptide is again fragmented, primarily by cleaving its peptide bonds. This treatment generates a nested set of peptides, each differing in size by one amino acid. These fragments are fed into a second coupled mass spectrometer and their masses are determined. The difference in masses between two closely related peptides can be used to deduce the missing amino acid. By repeated applications of this procedure, a partial amino acid sequence of the original protein can be determined.

SEE FIGURE 8-20 B (top)

Characterizing a protein with unknown function (from gene to function):
The first step is to clone the gene for your protein. The genomic version of the gene is large with lots of non-coding DNA and bacterial cells lack the machinery for splicing. The mRNA version of the gene contains the code for

the protein without the introns but can not be cloned in RNA form. Therefore you need a complementary copy in DNA form or cDNA. Getting cDNA: This is done by extracting the mRNA from cells and then making a complementary DNA (cDNA) copy of each mRNA molecule present. This reaction is catalysed by the reverse transcriptase enzyme of retroviruses, which synthesizes a DNA chain on an RNA template. Need to design primers complementary to the 5 and 3 prime ends of the coding sequence for our unknown protein (poly-T primer on 3 prime end). The single-stranded DNA molecules synthesized by the reverse transcriptase are converted to double stranded DNA molecules by DNA polymerase, and these molecules are inserted into a plasmid or virus vector and cloned. The RNA that was originally paired with the DNA was degraded using Rnase H. After cloning the gene, sequence it (using our sequence methods from Chang's lectures) then from the sequence predict the amino acid sequence and then check blast to make sure that the sequence is the one you wanted.
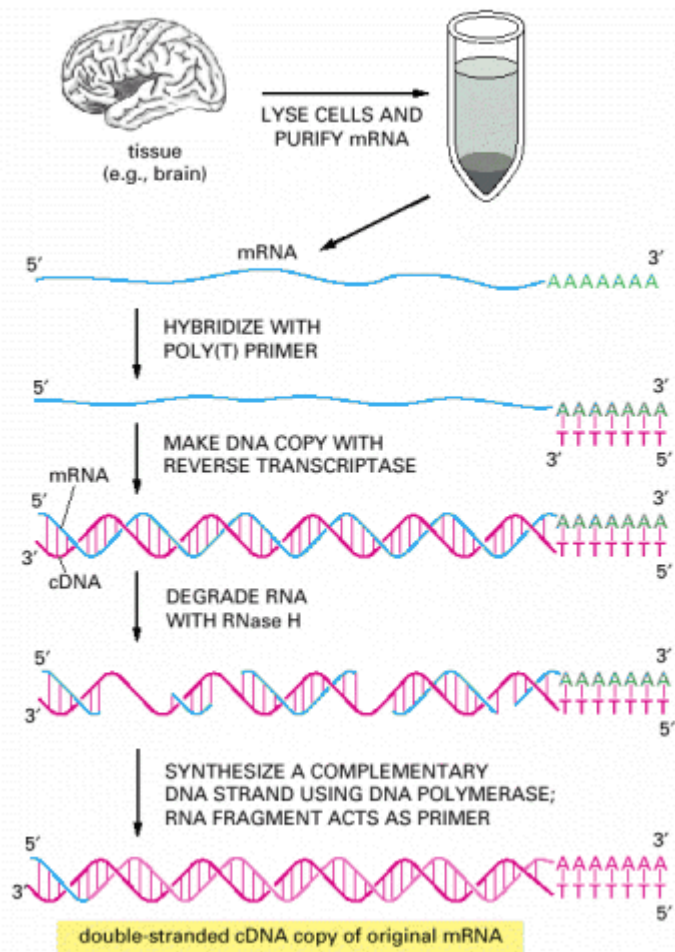
Step 2- Express your protein in bacteria.

Use a plasmid vector that has been engineered to contain a highly active promoter, which causes unusually large amounts of mRNA to be produced from an adjacent protein coding gene inserted into the plasmid vector (the gene inserted into the plasmid is "our" gene). The plasmid can then be inserted into bacterial cells where the gene can be efficiently transcribed and translated into protein. Now we will have lots of "our" protein.

STEP 3 – Characterise your protein's structure and function (if possible).
- Purify your protein from the bacterial mixture.
- Conduct structural analyses – NMR spectroscopy or X-ray crystallography.
- Attempt to assay activity of protein - enzyme? EMSA?

Where is my protein expressed/localised?
Step 4 – Raise antibodies against your protein
- Purify your protein and use to immunise rabbits to generate antibodies. Antibodies are made in billions of different forms each with different binding sites that recognise specific target molecules. Your protein is the antigen (Ag).
- Test antibodies for specificity using immunoprecipitation.
- Put protein extract including your protein into a test tube and add Ab raised against your protein.
- The Ab should bind to your protein. Centrifuge the test tube, which removes other proteins that aren't bound by antibodies. Examine immunoprecipitate using SDS-PAGE to make sure that your protein was precipitated. Or you can test antibodies using Western Blot. In this method you separate proteins in the extract using SDS-PAGE. Transfer proteins to solid support (nitrocellulose or nylon membrane). Now allow Ab against your protein **(primary antibodies)** to bind to the membrane. Add secondary Ab that was raised using rabbit antibodies as an Ag and which is coupled with a "marker".

Western Blot: The western blot is also known as an immunoblot. Southern blot is a DNA blot probed with DNA. Northern blot is an RNA blot probed with DNA or RNA. A western blot is a protein blot probed with antibodies. In the case above, using the marker secondary antibodies, the western blot is going to show only our protein.

Now we want to use our tested antibodies to determine the cellular (which cells contain the protein) localization of our protein. First we take thin sections of fixed tissues and incubate with primary antibodies raised against your protein. Then incubate with secondary antibodies tagged with a

"marker". Then visualize the maker on the secondary antibodies. The technique is called immunolocalization, and more specifically either immunohistochemistry (tissues) or immunoctyochemistry (cells).

STEP 5: Subcellular (where in the cell) localisation determination using reporter genes. Recall: reporter genes: Proxy for the endogenous gene. It reports on timing and localization of expression driven by a specific promoter. Preferably non-invasive and non-destructive. Several common reporters: Green fluorescent protein (GFP), Luciferase (luc)—light, Beta-Glucuronidase (GUS)—blue precipitate. For example: You can fuse GFP to your protein and then inject this into a live cell to examine protein localization. You make your fusion protein by introducing a promoter plus the cDNA encoding your protein plus the cDNA encoding the reporter gene (for example GFP) into a plasmid. The plasmid then transcribes your gene fused to the reporter gene and this is then translated into a fusion protein of your protein and the reporter protein. The fusion protein is then injected into a live cell (such as a guard cell) and you can see where in the cell the protein is localized by looking for green fluorescence.