CHAPTER 11

EXPRESSION OF GENETIC INFORMATION: FROM TRANSCRIPTION TO TRANSLATION

OBJECTIVES

- Define the relationship between genes and polypeptides.
- Describe the flow of information through cells and the cell components that participate.
- Elaborate on what is known about the fine structure of the gene in prokaryotes and eukaryotes.
- Point out the differences between transcription in prokaryotes and eukaryotes.
- Outline the process by which rRNA and tRNA are processed to their mature forms.
- Define heterogeneous nuclear RNAs and describe the process by which they are converted to mature mRNA.
- Describe the structure of mRNA, the 5' cap and the poly(A) tail.
- Summarize some of the experiments that led up to the discovery of the split gene.
- Emphasize the proposed importance of the split gene to evolution.
- Summarize the processing of hnRNA to mature mRNA, emphasizing the removal of introns and splicing together of exons.
- Summarize the discovery of ribozyme involvement in RNA processing and protein synthesis.
- Explain the mechanisms by which ribozymes carry out catalysis.
- Describe the properties of the genetic code, its theoretical underpinnings and the codon assignments.
- Define the role of tRNAs in decoding the genetic code.
- Summarize the steps in all stages of translation: tRNA charging, initiation, elongation and termination.

LECTURE OUTLINE

The Relationship Between Genes and Proteins

I. Archibald Garrod (Scottish physician, 1908) – reported that certain rare inherited diseases were caused by the absence of specific enzymes

A. Alcaptonuria - urine gets dark upon exposure to air; they lack enzyme that oxidizes homogentisic acid

1. Homogentisic acid is formed during the breakdown of the amino acids phenylalanine & tyrosine

2. Excess homogentisic acid is excreted into urine & oxidizes in air, turning black

B. He called such diseases inborn errors of metabolism - ignored for decades even though he had discovered connection between a genetic defect, a specific enzyme, and a specific metabolic condition

II. George Beadle & Edward Tatum (Caltech, 1940s) – resurrected the idea that genes direct the production of enzymes

A. Used Neurospora, a tropical bread mold, which grows in a very simple medium containing a single carbon source (e.g., a sugar), inorganic salts, & biotin (a B vitamin)

1. It was presumed to make all its required metabolites & they reasoned that an organism with such broad synthetic capacity should be very sensitive to enzymatic deficiencies

2. These deficiencies should be easily detected with the right experimental protocol

B. Beadle & Tatum's protocol – started with irradiation, which destroys the ability to make some enzymes & screen them for mutations that caused cells to lack a particular enzyme

1. Irradiate mold spores, creating individual cell populations with mutated genes

2. Screen them for mutation by growing spores on minimal medium that lacked essential compounds known to be synthesized by this organism (Neurospora)

3. If spore cannot grow in minimal medium, but can in medium supplemented with certain coenzyme (e.g., coenzyme A's pantothenic acid), it has enzymatic deficiency halting this compound's synthesis

4. Found specific metabolic deficiencies resulting from such enzyme deficiencies

5. Suggested that a gene carries the information for the construction of a particular enzyme

C. Beadle & Tatum's results – irradiated >1000 cells

1. Two cells proved unable to grow on minimal medium: one needed pyridoxine (vitamin B_6) & the other needed thiamine (vitamin B_1)

2. Eventually, progeny of ~100,000 irradiated spores were tested & dozens of mutants were found

3. Each mutant had a gene defect causing enzyme deficiency, preventing cells from catalyzing a particular metabolic reaction —> clearly a gene carries information for making a particular enzyme

D. This conclusion became known as the "one gene - one enzyme" hypothesis

1. Since enzymes often have >1 chain each of which is encoded by its own gene, the hypothesis was modified to one gene - one polypeptide hypothesis; still good approximation of basic gene function

2. One gene now shown to be able to be spliced differently to generate a variety of related polypeptides so one gene – one polypeptide hypothesis had to be modified

III. Vernon Ingram (Cambridge Univ., 1956) - reported on molecular consequence of sickle cell anemia mutation; cleaved normal & sickle cell hemoglobin into specific pieces with trypsin (~30 fragments)

A. Did paper chromatography to distinguish fragments from the two samples of hemoglobin —> one fragment in sickle cell preparation migrated differently from those in the normal hemoglobin prep

B. He sequenced the small fragment that was different in the 2 preps – the difference is valine substitution in the mutant sickle cell hemoglobin for glutamic acid in the normal molecule

1. This difference was apparently responsible for all of the symptoms associated with the disease

2. Ingram had demonstrated that a mutation in a single gene had caused a single substitution in the amino acid sequence of a single polypeptide

An Overview of the Flow of Information Through the Cell

I. The RNA intermediate that carries genetic information to cytoplasm is messenger RNA [mRNA]; it is the intermediate between a gene & its polypeptide

A. Existence of mRNA was first demonstrated in 1961 by François Jacob & Jacques Monod (Pasteur Inst., Paris), Sydney Brenner (Univ. of Cambridge) & Matthew Meselson (Calif. Inst. of Tech.)

B. mRNA is a complementary copy of 1 of the 2 DNA strands in gene (by process called **transcription**)

1. Since its sequence is complementary to one of the DNA strands of the gene from which it is transcribed, it carries the same information as the gene itself
2. Allows separation of information storage (gene is in nucleus as part of huge, immobile DNA molecule) & utilization (info imparted to smaller, more mobile RNA that passes into cytoplasm)
3. Once in cytoplasm, mRNA can serve as template to direct incorporation of amino acids in a particular order encoded by DNA nucleotide sequence & mRNA derived from it
4. mRNA use allows cell to greatly amplify synthetic output; one DNA molecule can serve as template in formation of many mRNA molecules, each of which can be used to form many polypeptides
5. Gene stays safe, while a working mRNA copy serves as template directing polypeptide synthesis; mRNA is relatively unstable; its half-life is short, typically measured in hours

C. mRNA code read during **translation** (protein synthesis), a complex process involving dozens of different components, including ribosomes in cytoplasm (large & small subunits)

1. Ribosomes are nonspecific translation machinery components that can be programmed to translate info encoded by any mRNA; bacterial ribosomes used to make proteins encoded by human mRNAs
2. Ribosomes are made of protein & ribosomal RNA (rRNA; plays structural & catalytic role); rRNAs are transcribed from one of the DNA strands of a gene
3. rRNAs do not function in informational capacity, but are adapted to recognize & bind other molecules, provide structural support & catalyze chemical reaction that covalently links amino acids
4. Ribosome subunits join when protein synthesis starts; separate when it ends

II. Transfer RNAs (tRNAs) – a third major class of RNA that is required during protein synthesis; they help convert DNA language to that of amino acids

III. Both rRNAs & tRNAs owe their activity to their complex secondary & tertiary structures, unlike DNA, which has a similar overall structure regardless of the source

A. Many RNAs fold into complex 3D shapes, which differ markedly from one type of RNA to another

1. RNAs carry out a diverse array of functions because of their different shapes
2. RNA folding follows certain rules & is determined by formation of complementary base pair regions unlike proteins, the folding of which is driven by withdrawal of hydrophobic residues into interior
3. Base-paired regions typically form double-stranded & double helical stems, which are connected to single-stranded regions (loops)
4. Unlike DNA (has only standard Watson-Crick base pairs), RNAs often have nonstandard mismatched base pairs & modified nitrogenous bases (serve as recognition sites for proteins & other RNAs)

B. tRNAs & rRNAs have long half-life in cell (typically measured in days) unlike short-lived mRNAs

IV. Eukaryotic cells makes a host of other RNAs, which also play vital roles in cellular metabolism

A. Small nuclear RNAs (snRNAs)

B. Small nucleolar RNAs (snoRNAs)
C. Small interfering RNAs (siRNAs)

D. Micro RNAs (miRNAs) – burst onto the scene in the past few years; our genome encodes hundreds of these tiny micro RNAs, yet we have very little idea what any of them do

Transcription: The Basic Process

I. Transcription done by **DNA-dependent RNA polymerases** in both prokaryotes & eukaryotes; incorporate nucleotides to make RNA from DNA **template**; DNA provides information for RNA strand synthesis

A. The enzymes are commonly known as RNA polymerases

B. They incorporate the nucleotides into the growing RNA chain one at a time; its sequence is complementary to one of the strands of DNA

II. RNA synthesis begins with the association of the RNA polymerase with the DNA template – proteins have evolved to recognize & bind specific nucleotide sequences in a strand of nucleic acids

A. RNA polymerases bind to sites on DNA called **promoters** before starting transcription; they cannot recognize promoters on their own, but require the help of other proteins (**transcription factors)**

1. Transcription factors (TFs) are particularly important in initiation of eukaryotic gene transcription

B. Promoter also contains information specifying which strand of DNA to transcribe & the site at which transcription begins

III. Polymerase moves along DNA in 3' —> 5' direction (toward 5' end) laying down complementary RNA in 5' —> 3' direction, unwinding DNA temporarily as it travels along DNA template

A. Catalyzes reaction: $RNA_n + NPPP \longrightarrow RNA_{n+1} + PP_i$ ($\longrightarrow 2\ P_i$)

1. The ribonucleoside triphosphate precursors (NPPPs) are hydrolyzed into nucleoside monophosphates as they are polymerized into a covalent RNA chain

B. Reactions leading to nucleic acid & protein synthesis are inherently different from those of intermediary metabolism

1. Such reactions must be essentially irreversible unlike many reactions of intermediary metabolism, which may be close enough to equilibrium that a considerable reverse reaction can be measured

2. To prevent reverse reaction (there is virtually none), nucleic acid synthesis is coupled to exergonic pyrophosphate hydrolysis; done by pyrophosphatase; this releases a large amount of free energy

3. Hydrolysis of pyrophosphate produces two inorganic phosphates (P_I), making overall reaction essentially irreversible

4. A nucleotide is incorporated into a RNA strand if it forms proper Watson-Crick base pair with its opposite number in the DNA strand being transcribed

C. DNA helix reforms behind polymerase after it passes a particular stretch of DNA; only a few (~9) RNA nucleotides stay attached to DNA as DNA-RNA hybrid just behind site of polymerase operation

D. Bacterial RNA polymerase can incorporate 50 - 100 nucleotides/sec; most genes in cell are transcribed simultaneously by numerous polymerases

IV. Enzyme must stay attached to DNA over long stretches of template to make the prodigiously long mRNAs that they do, but be loose enough to move; the enzyme is **processive**

A. It is hard to study processivity using biochemical methodologies that tend to average out differences between individual protein molecules so they used methods like those used to study cytoskeletal motors

B. Movement is similar to cytoskeletal motor kinesin; studies confirm this – attach a single RNA polymerase molecule to the surface of a glass cover slip

1. Allow RNA polymerase to transcribe DNA with fluorescent bead covalently linked to one of its ends & monitor bead movement under a fluorescence microscope
2. The bead is free in solution & its range of movement is proportional to the length of the DNA between the polymerase & the bead
3. As polymerase transcribes template, connecting DNA strand is shortened & bead movement is restricted
4. Allows determination of transcription rate of individual polymerase & whether it transcribes DNA in a steady or discontinuous movement
5. Investigators can trap bead by focused laser beam & vary the force until it is just sufficient to stop the polymerase from continuing transcription
6. These studies show that polymerase moves with a force >2X that of myosin interacting with actin

C. Energy required to drive the movement comes from NPPP precursor hydrolysis as they are incorporated into growing RNA chain

V. Enzymes do not necessarily move in steady, continuous fashion even though they contain relatively powerful motors; may pause at certain sites along template for varying periods

A. Sometimes, stalled polymerase must digest away 3' end of new transcript & resynthesize missing part before it can move; a number of elongation factors enhance enzyme's abilities to pass roadblocks

B. Both activities (digestion & polymerization) done by active site of polymerase

Transcription: Prokaryotes

I. *E. coli* has a single type of RNA polymerase - 5 subunits tightly associated to form core enzyme

A. Purify core enzyme, add to solution of bacterial DNA & ribonucleoside triphophates —> enzyme binds to DNA & synthesizes RNA (but not the same RNA as in cell)

1. It starts synthesis at inappropriate sites (random sites it would normally have ignored *in vivo*)

B. Purify accessory polypeptide sigma (σ) factor & add it to mix before polymerase attaches to DNA

1. Transcription starts at right places (selected specific locations, not random ones)
2. σ attachment increases core affinity for DNA promoter sites & decreases affinity for DNA in general
3. Once transcription begins, σ leaves & core enzyme continues synthesis

C. Archaea (the other major prokaryote group along with bacteria) also possess a single RNA polymerase, but it has very different subunit composition, more closely resembling a eukaryotic RNA polymerase

II. Bacterial promoters are located in region of DNA strand just preceding the RNA synthesis initiation site

A. The nucleotide at which transcription is initiated is called +1; the preceding nucleotide is –1
B. Those portions of DNA preceding initiation site (toward template 3' end) are said to be upstream from that site
C. Those portions of DNA succeeding initiation site (toward template 5' end) are said to be downstream from that site

III. Similar DNA sequences were seen in association with genes in roughly same location from gene to gene in bacteria; sequences are called **consensus sequences** & are found just upstream from many bacterial genes
A. A consensus sequence is the most common version of such a conserved DNA sequence, but some variation in sequence can occur from gene to gene
B. DNA sequences just upstream from a large number of bacterial genes have 2 short stretches of DNA that are similar from one gene to another (-35 region & -10 region)

IV. The first stretch is centered ~35 bases upstream from initiation site; typically has TTGACA sequence
A. The TTGACA sequence (the –35 element) is called a consensus sequence, indicating that it is the most common version of a conserved sequence but that some variation occurs from gene to gene
B. The –35 region in promoter is recognized by a σ factor associated with the RNA polymerase
C. Bacteria have a variety of different σ factors that recognize different versions of promoter sequence - the s^{70} factor is known as the housekeeping σ factor; it initiates transcription of most genes
1. Alternative σ factors start transcription of a small number of specific genes that participate in a common response
2. When *E. coli* cells are exposed to a sudden rise in temperature, a new σ factor is made that recognizes a different promoter; it leads to coordinated transcription of a battery of heat-shock genes
3. The products of these genes protect cell proteins from thermal damage

V. The second conserved sequence is found at ~10 bases upstream from the initiation site
A. It has the TATAAT consensus sequence (**Pribnow box**)
B. Responsible for identifying the precise nucleotide at which transcription starts (named after discoverer)

VI. X-ray crystallography analysis of bacterial RNA polymerase shows molecule shaped like a crab claw with a pair of mobile pincers (or jaws) enclosing a positively charged internal channel that holds double helix
A. As σ factor interacts with promoter, jaws of enzyme grip the downstream DNA duplex, which resides within the channel
B. Once bound to the promoter, the enzyme separates (melts) the 2 DNA strands in the start site region
1. Strand separation makes the template strand accessible to the enzyme's active site, which resides at the back wall of the channel
C. Transcription initiation appears to be difficult since RNA polymerase typically makes several unsuccessful attempts to assemble an RNA transcript
D. Once it has laid down 10 - 12 nucleotides in growing transcript, enzyme undergoes a major conformational change stabilizing it & converting it into a transcriptional elongation complex
1. The transcriptional elongation complex can move processively along the DNA of a gene.
E. Elongation complex formation is generally followed by release of σ factor & promoter DNA

VI. Termination of transcription - specific nucleotide sequence signals termination when it is reached

A. Sometimes, a protein called **rho** is needed for termination of bacterial transcription - rho has enzymatic activity that may peel the 3' end of RNA transcript away from the DNA to which it is bound

B. Usually the polymerase stops transcription on its own when it reaches a terminator sequence & releases the completed RNA chain without requiring additional factors

Transcription and RNA Processing in Eukaryotic Cells: General Information

I. Eukaryotic cells have 3 distinct RNA polymerases (yeast have same as in mammalian cells), each responsible for synthesizing a different group of RNAs; no prokaryotes found with multiple RNA polymerases

A. RNA polymerase I - synthesizes the larger rRNAs (28S, 18S, 5.8S)

B. RNA polymerase II - synthesizes mRNAs & most small nuclear RNAs (snRNAs & snoRNAs)

C. RNA polymerase III - synthesizes various low MW RNAs (the various tRNAs, 5S rRNA & U6 snRNA)

II. X-ray crystallographic structure of yeast polymerase II enhanced understanding of eukaryotic transcription – Roger Kornberg et al., Stanford University

A. Yeast enzyme has 7 more subunits than its bacterial counterpart, but the fundamental core structure of the two enzymes & their basic mechanism of transcription are virtually identical

B. Some of the additional subunits of eukaryotic polymerases are thought to play key roles in the interaction of the enzyme with other proteins

C. A major distinction between transcription in prokaryotes & eukaryotes is the requirement in eukaryotes for a large variety of accessory proteins or transcription factors (TFs)

1. They play a role in virtually every aspect of transcription process from binding polymerase to DNA template, to initiation of transcription, to its elongation & termination

2. They are crucial for the operation of all 3 types of eukaryotic RNA polymerases

III. All 3 major RNA types (mRNA, tRNA, rRNA) are processed before they are mature; they are derived from precursor RNA molecules that are considerably longer than the final RNA product

A. Primary (1°) transcript (pre-RNA) is the initial precursor RNA & it is equivalent in length to the full length of the DNA transcribed; it is then edited

1. Corresponding segment of DNA from which 1° transcript is transcribed is called **transcription unit**

2. 1° transcripts do not exist within the cell as naked RNA but become associated with proteins even as they are synthesized

3. 1° transcript is typically short-lived, having a fleeting existence; it is processed into smaller, functional RNAs by a series of "cut-&-paste" reactions

B. Processing requires variety of small RNAs (90 – 300 nucleotides long) & their associated proteins (small nuclear RNAs [snRNAs])

Transcription and RNA Processing in Eukaryotic Cells: rRNAs & tRNAs

I. Ribosomal RNAs - >80% of cell RNA, so rDNA sequences repeat hundreds of times to supply so many transcripts; each ribosome consists of several rRNA molecules together with dozens of ribosomal proteins

A. This moderately repetitive rDNA is typically clustered in one or a few genome regions; human genome has 5 rDNA clusters, each on a different chromosome; their function is to produce ribosomes

1. In (nondividing) interphase cell, the rDNA clusters are gathered together as ≥1 irregularly-shaped nuclear structures called nucleoli (singular - nucleolus) that function in producing ribosomes

2. Nucleoli disappear at cell division (mitosis) & reappear in daughter cell nuclei around chromosome regions having rRNA genes; chromosome regions containing rDNA are called **nucleolar organizers**

B. Bulk of nucleolus is composed of ribosomal subunits that give it granular appearance

1. Embedded in granular mass are ≥1 rounded cores made mostly of fibrillar material
2. Fibrillar core is thought to consist of rDNA templates & nascent rRNA transcripts

II. Synthesizing rRNA precursor - study amphibian eggs because they are large & have so many nucleoli

A. Amphibian oocytes (up to 2.5 mm dia) selectively amplify rDNA a lot & increase number of nucleoli (hundreds); amplification needed by fertilized egg to make ribosomes for embryonic development

1. Mammalian oocytes are also large (typical for oocytes) but only 100 μm in diameter
2. Due to their many nucleoli egg cells/oocytes, especially amphibian eggs, each actively making rRNA, are ideal for studying rRNA synthesis & processing

B. Oscar Miller, Jr., U. of Virginia, (late 1960s) - gently dispersed nucleoli fibrillar cores of oocytes —> revealed presence of large circular fiber; in EM, these fibers resembled a chain of Christmas trees

1. Saw several distinct rDNA genes situated one after another along single DNA chain (tandem repeat)

2. Cell concentrates all of its rRNA synthesizing & processing machinery in single nuclear region

3. Each of the ~100 fibers in Christmas tree is nascent rRNA caught in act of elongation (can see polymerase I); each is a branch of the Christmas tree

4. Shorter fibrils are near rDNA gene transcription initiation site, longer ones at end; promoter is just upstream from transcription initiation site

5. Length of DNA from shortest to longest RNA fibrils corresponds to a single transcription unit; the longer the fibril is, the closer the transcript is to completion

6. High RNA polymerase molecule density along each transcription unit (~1 every 100 DNA bps) reflects high rRNA synthesis rate in these oocyte nucleoli

7. The RNA fibrils contain clumps & associated particles (consisting of RNA & protein) that convert rRNA precursors into final rRNA products & assemble them into ribosomal subunits

8. Nontranscribed spacers between rDNA (& tRNA & histone) genes & most tandemly repeated genes

III. Processing of rRNA precursor – eukaryotic ribosomes have 4 distinct rRNAs: 3 rRNAs in the large subunit (28S, 5.8S, 5S in humans); 1 in the small (18S in humans) subunit

A. S value (or Svedberg unit) refers to sedimentation coefficient of RNA; the larger the number, the more rapidly the molecule moves through a field of force during centrifugation

1. For a group of chemically similar molecules, S value correlates with size (larger S value, larger size)
2. 28S, 18S, 5.8S & 5S RNAs are ~5,000, 2,000, 160 & 120 nucleotides long, respectively

B. 28S, 5.8S & 5S rRNAs are carved by various nucleases from a single 1° transcript (pre-rRNA) at specific sites; 5S rRNA is synthesized from a separate RNA precursor outside nucleolus

C. Pre-rRNA has 2 peculiarities not seen in other RNA transcripts: many methylated nucleotides & pseudouridine residues; all of these modifications are posttranscriptional (after nascent RNA synthesis)

1. In humans, before a pre-rRNA is first cleaved, >100 CH_3 groups have been added to ribose groups in the molecule & ~95 uridine residues have been altered chemically to pseudouridine
2. Altered nucleotides are at specific positions & clustered in parts of pre-rRNA that have been conserved during vertebrate evolution
3. All of the nucleotides altered in the pre-rRNA are found in the final molecules, while unaltered sections are discarded during processing
4. Functions of groups are unclear but they may protect parts of pre-RNA from enzymatic cleavage, promote folding into final 3-D structure and/or promote rRNA interactions with other molecules
5. Mutations in enzyme that converts uridine to pseudouridine are linked to dyskeratosis, a rare, fatal disease characterized by skin abnormalities, bone marrow failure & elevated cancer susceptibility

D. Label pre-rRNA with ^{14}C-methyl group from labeled methionine (in most cells, it is CH_3 donor)

1. Methyl group is transferred enzymatically from methionine to pre-rRNA nucleotides
2. After brief labeling period in cultured mammalian cells, ^{14}C seen in 45S RNA ~13,000 nucleotides in length), 45S RNA then cleaved into smaller molecules; trimmed down to 28S, 18S & 5.8S rRNAs
3. Mature 28S, 18S & 5.8S rRNAs have in total ~7,000 nucleotides; slightly > half of pre-rRNA

E. Pulse-chase with ^{14}C-methionine reveals intermediates: 45S peaks in nucleolar RNA fraction after 10 min, 45S gone from nucleolus after ~1 hr —> largely replaced by 32S peaks in nucleolar material

1. The 32S RNA is one of 2 major products made from 45S 1° transcript; it is seen as distinct peak in nucleolar fraction from 40 - 150 min —> 32S is precursor to the mature 28S rRNA
2. The other major product leaves the nucleolus quite rapidly & is seen in cytoplasm as mature 18S rRNA (within 40 min it is in cytoplasmic fraction)
3. After 2 or more hours, nearly all of radioactivity has left nucleolus & most has accumulated in cytoplasmic 28S & 18S rRNAs
4. Radiolabel is also in 4S RNA peak of cytoplasm; represents CH_3 groups transferred to small tRNAs

IV. The role of the small, nucleolar RNAs (snoRNAs) in pre-rRNA processing; they are packaged with particular proteins to form snoRNPs (small, nucleolar ribonucleoproteins) & help in pre-rRNA processing

A. snoRNPs associate with rRNA precursor before it is fully transcribed

1. The first RNP particle to attach to pre-rRNA transcript contains U3 snoRNA & >2 dozen different proteins; it is so big that it can be seen in EM as ball bound to outer end of each nascent RNA fibril
2. It binds to precursor 5' end of transcript & catalyzes removal of transcript 5' end
3. Some other enzymatic cleavages are thought to be catalyzed by the exosome, which is an RNA-degrading machine that consists of nearly a dozen different exonucleases

B. U3 & several other snoRNAs were identified many years ago because of their relatively large quantities (~10^6 copies/cell) discovered long ago

C. More recently, a new class was discovered - lower concentration (~10^4 copies/cell); divided into 2 groups based on their function & similarities in nucleotide sequence

1. Box C/D snoRNAs – determine which pre-rRNA nucleotides have their ribose moieties methylated
2. Box H/ACA snoRNAs – determine which uridines are converted to pseudouridines

D. Both groups have relatively long stretches (10-21 nucleotides) that are complementary to parts of rRNA transcript

1. Example of principle that single-stranded nucleic acids with such complementary nucleotide sequences can form double-stranded hybrids
2. They bind to specific portions of pre-rRNA to form an RNA-RNA duplex & guide an enzyme (a methylase or pseudouridylase) within the snoRNP to modify a particular pre-rRNA nucleotide
3. ~200 different snoRNAs exist, one for each pre-rRNA pseudouridylated or ribose-methylated site; if gene for one snoRNA is deleted, one pre-rRNA nucleotide is not modified enzymatically
4. These RNAs are encoded within intervening sequences of other genes

V. Ribosomal subunit assembly also done in nucleolus; 2 protein types associate with RNA as it's processed

A. Ribosomal proteins that will eventually remain in ribosomal subunits

B. Accessory proteins that have transient interaction with rRNA intermediates & are needed for processing (enzymes that cleave rRNA precursors & proteins that protect sites from cleavage)

1. These include > a dozen RNA helicases that unwind regions of double-stranded RNA
2. These enzymes are presumably involved in the many structural rearrangements that occur during ribosome formation, including the association & dissociation of the snoRNAs

VI. 5S rRNA synthesis & processing (~120 bases long) - part of prokaryote & eukaryote large ribosomal subunit

A. In eukaryotes, 5S rRNA is encoded by large number of identical genes that are separate from other rRNA genes & located outside nucleolus; organized in tandem array with spacers like 45S rRNA genes

1. Transcribed by RNA polymerase III - 5' end of 1° transcript is identical to mature 5S rRNA; but the 3' end contains extra nucleotides that are removed during processing
2. After synthesis, 5S rRNA is transported to nucleolus to participate in ribosome subunit assembly

B. Polymerase III binds to promoter within transcribed part of target gene rather than upstream - unusual

1. Introduce altered 5S rRNA genes into host —> see if cell transcribes it (if it can act as template)
2. Remove whole 5' flanking region —> still transcribed starting at normal initiation site
3. Delete central part of gene from (from about nucleotide 50-80 of 120 bp gene) —> polymerase would not transcribe the DNA or even bind to it
4. Put internal promoter elsewhere in genome —> get transcription at new site by polymerase III

C. Polymerase III transcribes several different RNAs

1. It binds to internal promoter when transcribing pre-5S RNA or pre-tRNA, but binds to an upstream promoter when transcribing the precursors for several others, including U6 snRNA

VII. Transfer RNAs - ~50 tRNAs in plant & animal cells, each encoded by DNA sequences repeated a number of times within genome

A. Repeat number varies with organism (yeast - ~275 total tRNA genes, fruit flies - ~850, humans - ~1,300)

1. Genes are found in small clusters scattered throughout genome
2. A single cluster typically contains multiple copies of different tRNA genes
3. DNA sequence encoding a given tRNA is typically found in >1 cluster

4. DNA within cluster (tDNA) consists largely of nontranscribed spacer sequences with tRNA coding sequences situated at irregular intervals in tandemly repeated arrangement

B. Transcribed by polymerase III; the promoter is within the coding sequence, not in 5' flanking region

1. 1° transcript of tRNA is bigger than final product; pieces on both 5' & 3' sides of the precursor (& sometimes an interior piece) must be trimmed away
2. The enzyme ribonuclease P is an endonuclease involved in pre-tRNA processing
3. Ribonuclease P - present in both bacterial & eukaryotic cells; consists of both RNA & protein subunits; ribonuclease P RNA subunit catalyzes cleavage of pre-tRNA substrate

C. All mature tRNAs have the triplet CCA sequence at their 3' end (added enzymatically after tRNA processing); play key role in protein synthesis

Transcription and RNA Processing in Eukaryotic Cells: Background Information on mRNAs

I. If eukaryotic cells are incubated for short time (30 min) in ^{3}H-uridine or ^{32}P-phosphate & killed right after, most of radiolabel is incorporated into large group of RNA molecules with following properties:

A. They have large molecular weights (up to ~80S or 50,000 nucleotides)

B. As a group, they are represented by RNAs of diverse (heterogeneous) nucleotide sequence, **and**

C. They are found only in the nucleus

II. Because of these properties, these RNAs are called **heterogeneous nuclear RNAs** (**hnRNAs**)

A. Pulse cells in ^{3}H-uridine or ^{32}P-phosphate briefly & then chase in unlabeled medium for an hour or more before cells are killed & the RNA extracted —> label in large nuclear RNAs drops sharply

1. The label appears instead in much smaller mRNAs found in cytoplasm
2. These experiments first begun by James Darnell, Jr. & Klaus Scherrer (MIT) suggested that large, rapidly labeled hnRNAs were precursors to smaller cytoplasmic mRNAs; confirmed over past 40 yrs

B. Most of the RNA in the cell is present as 18S & 28S rRNA, along with tRNAs & snRNAs that stay at the top of the tube after centrifugation, as detected by absorbance of UV light

1. Neither hnRNA nor mRNA constitute a significant fraction of cellular RNA
2. However, while they are only a small percentage of total cellular RNA in eukaryotes, they constitute a large percentage of the RNA being synthesized by that cell at any given moment

C. hnRNAs & mRNAs are degraded after relatively brief time periods so they do not make up a big portion of total cellular RNA & are correspondingly hard to detect by fractionation & absorbance

1. This is particularly true of the hnRNAs with their half-lives of only a few minutes; they are processed into mRNAs (or completely degraded) even as they are being synthesized

2. In contrast, rRNAs & tRNAs have half-lives measured in days or weeks & thus gradually accumulate to become the predominant species in the cell
3. mRNA half-lives vary depending on particular species, range from ~15 minutes to period of days

III. Machinery used in mRNA transcription – RNA polymerase II & general transcription factors (GTFs)
A. RNA polymerase II synthesizes all eukaryotic mRNA precursors; it is composed of 12 different subunits & is remarkably conserved from yeast to mammals

B. Initiation of transcription by polymerase II occurs in cooperation with a number of general transcription factors (GTFs); their precise roles remain to be determined
1. Called "general" TFs because the same ones are required for accurate transcription of a diverse array of genes in a wide variety of different organisms
C. Polymerase II promoters lie to 5' side of each transcription unit
1. In the vast majority of genes, a critical portion of the promoter lies between 24 & 32 bases upstream from transcription initiation site
2. This region often contains a consensus sequence that is either identical or very similar to the oligonucleotide 5'-TATAAA-3' & is known as the TATA box
3. The TATA box of the DNA is the site of assembly of a preinitiation complex that contains the GTFs & the polymerase; this complex must assemble before gene transcription can be initiated

IV. Assembly of preinitiation complex - first step is binding of a protein, the TATA-binding protein (TBP) that specifically recognizes the TATA box of eukaryotic promoters
A. Thus, a purified eukaryotic polymerase, as in prokaryotic cells, is not able to recognize a promoter directly & cannot initiate accurate transcription on its own
1. TBP is present as a subunit of a much larger protein complex called TFIID (***t***ranscription ***f***actor for polymerase ***II***, fraction ***D***)
2. The TBP DNA-binding domain consists of a 10-stranded β sheet curved into a saddle-shaped structure that sits astride the DNA
B. TBP is actually a universal TF that mediates binding of all 3 eukaryotic RNA polymerases; it is present as one of the subunits of 3 different proteins
1. As subunit of TFIID, TBP promotes binding of RNA polymerase II
2. As a subunit of the proteins SL1 or TFIIIB, TBP promotes binding of RNA polymerases I & III, respectively
3. A large number of promoters lack a TATA box, but are capable of binding TBP
C. X-ray crystallography shows that binding of TBP to a polymerase II promoter causes a dramatic distortion in DNA conformation
1. TBP inserts itself into the minor groove of the double helix, bending the DNA molecule >80° at the site of DNA-protein interaction
2. Bound DNA develops a distinct kink along its length & DNA duplex becomes unwound over span of ~8 base pairs

V. TFIID binding sets stage for assembly of complete preinitiation complex
A. May occur in stepwise manner or in single step with binding of large, preassembled complex (holoenzyme) containing the polymerase & the remaining GTFs
B. 3 GTFs interact with promoter on DNA (TBP of TFIID, TFIIA & TFIIB) & provide platform for binding of huge, multisubunit RNA polymerase with its attached TFIIF

1. Once RNA polymerase-TFIIF is in position, another pair of GTFs (TFIIE & TFIIH) joins the complex & transforms the polymerase into an active, transcribing machine

C. TFIIH is the only GTF known to have enzymatic activities
   1. One of its subunits functions as a protein kinase to phosphorylate RNA polymerase
   2. 2 other subunits act as DNA unwinding enzymes (helicases); the DNA helicases separate DNA strands of promoter; allows polymerase access to template strand

D. Once transcription starts, certain of the GTFs (including TFIID) may be left behind at promoter site, while others are released from the complex
   1. As long as TFIID stays bound to the promoter, more RNA polymerases may be able to attach to the promoter site & initiate additional rounds of transcription without delay

E. Carboxyl-terminal domain (CTD) of the largest RNA polymerase II subunit has an unusual structure
   1. It consists of a sequence of 7 amino acids (-Tyr-Ser-Pro-Thr-Ser-Pro-Ser-) that is repeated over & over; in mammals, the CTD consists of 52 repeats of this heptapeptide
   2. Of the 7 amino acids, serines 2 & 5 are prime candidates for phosphorylation by protein kinases
   3. The RNA polymerase II that assembles into the preinitiation complex is not phosphorylated, whereas when it is transcribing RNA, it is heavily phosphorylated
   4. All of the added phosphate groups are found in CTD; RNA polymerase II is thought to be phosphorylated just after the initiation of transcription
   5. CTD phosphorylation can be catalyzed by at least 4 different kinases, including the kinase activity of TFIIH
   6. Polymerase phosphorylation may act as a trigger that uncouples the enzyme from GTFs and/or promoter DNA, allowing enzyme to escape from preinitiation complex & move down DNA template
   7. A polymerase involved in elongation may associate with a number of large accessory proteins, including ELL & SII
   8. According to some estimates, an elongating polymerase is part of a huge complex (>50 components & a total molecular mass of >3 million daltons)
   9. Probably, the template moves through an immobilized transcriptional machine rather than the machinery moving along the template

F. RNA polymerase II & its GTFs are sufficient to promote a low, basal level of transcription from most promoters *in vitro*; a variety of specific TFs bind to DNA regulatory regions & do the following:
   1. Determine whether or not a preinitiation complex assembles at a particular promoter **and/or**
   2. Determine the rate at which polymerase initiates new rounds of transcription from that promoter

VI. The structure of mRNAs - mRNAs share certain properties:
   A. They contain a continuous nucleotide sequence encoding a specific polypeptide
   B. They are found in the cytoplasm
   C. They are attached to ribosomes when they are translated
   D. Most have a significant noncoding, nontranslated segment that does not direct assembly of amino acids
      1. ~25% of each globin mRNA consists of noncoding, nontranslated regions
      2. Noncoding portions are found at both the 5' & 3' ends of a mRNA & contain sequences that have important regulatory roles
   E. Eukaryotic mRNAs have special modifications at their 5' & 3' termini that are found neither on prokaryotic messages nor on tRNAs, or rRNAs

1. 3' end of nearly all eukaryotic mRNAs – has a string of 50 - 250 adenosine residues that form a the poly(A) tail; histones are an exception
2. 5' end – has methylated guanosine cap

VII. Split genes: An introduction to an unexpected finding

A. Precursor-product relationship between hnRNAs & mRNAs was established fairly early, but major sticking point was the difference in size between the 2 RNA populations
1. The hnRNAs are several times the size of mRNAs; processing needed as with rRNA precursors

B. Early rRNA processing studies showed that mature RNAs were carved from larger precursors
1. Large segments were removed from both the 5' & 3' sides of various rRNA intermediates to yield the final, mature rRNA products
2. It was thought a similar pathway might account for processing of hnRNAs to mRNAs, since both contain 5' methylguanosine caps & 3' poly(A) tails & specific mRNA sequences found in hnRNAs
3. But mRNAs constitute such a diverse population that it was impossible to follow the steps in the processing of a single mRNA species; the problem was solved by an unexpected discovery

C. Until 1977, it was assumed that a continuous linear mRNA nucleotide sequence was complementary to a continuous DNA sequence in one strand of the DNA of a gene

VIII. Split genes: Experimental evidence

A. Philip Sharp et. al. (MIT) & Richard Roberts, Louise Chow et. al. (Cold Spring Harbor Labs, NY) - found mRNAs were transcribed from DNA segments separated from one another along template
1. These protein-coding DNA segments are separated from one another along the template strand
2. Found in adenovirus transcription studies; a pathogen that can infect a variety of mammalian cells
3. Found a number of different adenovirus mRNAs with same 150-200-nucleotide 5' terminus; thought it might represent a repeated stretch of nucleotides near promoter region of each gene **but……..**
4. 5' leader sequence is not complementary to a repeated sequence & not even complementary to a continuous stretch of nucleotides in the template DNA
5. Instead, the 5' leader is transcribed from 3 distinct & separate DNA segments
6. DNA regions between these blocks (**intervening sequences**) are missing in corresponding mRNA
7. Not peculiar to viruses, because soon seen in nonviral, cellular genes

B. Alec Jeffreys & Richard Flavell (1977, U. of Amsterdam) - saw intervening sequence (~600 bases) within β-globin gene (split gene); soon found in other genes; now apparent to be more rule than exception
1. Exons – those parts of the gene that contribute to the mature RNA product
2. Introns - intervening sequences

C. Split genes are widespread in eukaryotes, although introns of simpler eukaryotes (yeast, nematodes) tend to be fewer in number & smaller in size than those of more complex plants & animals
1. Introns are found in all types of genes, including those that code for tRNAs, rRNAs & mRNAs

D. Hypothesized that introns must be removed from the 1° transcript corresponding to the entire transcription unit to make mature mRNA

1. If this were true, then segments corresponding to introns should be found in 1° transcript
2. This would also explain why hnRNA molecules are so much larger than the mRNAs they produce

3. The sizes of a few pre-mRNAs had been determined - globin sequence found in nuclear RNA molecule sedimenting at 15S; while final mRNA is 10S

E. Shirley Tilghman, Philip Leder, et al. (NIH) - R-loop formation seen in EM; determined physical relationship between 15S & 10S globin RNAs & provided information on split gene transcription

1. Single-stranded, complementary DNA strands bind specifically to one another; so can single-stranded DNA & RNA molecules if sequences are complementary (DNA-RNA hybridization)
2. Formed DNA-RNA hybrid between globin gene-containing DNA fragment & 15S RNA —> got continuous, double-stranded, DNA-RNA hybrid & could see it in EM
3. Form hybrid between globin gene fragment & mature 10S globin mRNA —> in DNA coding region center, single-stranded loop bulged out; corresponds to introns (no complement in 10S mRNA)

F. Also found same thing in ovalbumin (protein found in hen's eggs) gene (DNA - mRNA hybrid) —> 7 distinct loops form; corresponded to 7 introns (~3 times as much DNA as that in this gene's 8 exons)

1. Subsequent studies show that individual exons average ~150 nucleotides
2. Individual introns average ~3,500 nucleotides; explains why hnRNAs are much longer than mRNAs
3. Example: human dystrophin gene extends for roughly 100 times the length needed to code for its corresponding message
4. Example: type I collagen gene contains >50 introns; average human gene contains ~9 introns
5. All of the above provided strong evidence for the proposition that mRNA formation in eukaryotes occurs by removal of internal sequences of ribonucleotides from much larger pre-mRNAs

Transcription and RNA Processing in Eukaryotic Cells: Processing

I. RNA polymerase II assembles a 1° transcript that is complementary to the DNA of the entire transcription unit; the transcript is then processed in nucleus to produce mature mRNA that is transported to cytoplasm

A. As transcription proceeds, transcripts associate with proteins & larger particles (proteins & ribonucleoproteins) that contain agents responsible for converting transcript to mature mRNA

B. The conversion process requires the addition of a 5' cap & a 3' poly(A) tail to the transcript ends & removal of any intervening introns

II. 5' methylguanosine cap forms very soon after RNA synthesis begins; 5' end initially has triphosphate from first nucleoside triphosphate incorporated in RNA at initiation site of RNA synthesis

A. Once the 5' end of an mRNA precursor is synthesized, several enzymes act on this end of molecule

1. First, the last of the three phosphates is removed by an enzyme, leaving behind diphosphate

2. Then, GMP is added in inverted orientation so guanosine 5' end faces 5' end of RNA chain thus the first 2 nucleosides are joined by an unusual 5'-5' triphosphate bridge

3. Next, the terminal inverted, guanosine is methylated at position 7 on guanine base while nucleotide on triphosphate bridge internal side is methylated at ribose 2' position (methylguanosine cap)

B. These enzymatic modifications at 5' end of 1° transcript occur very quickly, while the RNA is still in its very early stages of synthesis; various internal nucleotides may or may not be methylated as well

C. May serve several functions: prevents exonuclease digestion of mRNA 5' end, aids in transport of mRNA out of nucleus, important role in initiation of mRNA translation

III. The poly(A) tail – 3' end of mRNA contains a string of adenosine residues that forms a tail; some proteins lack it, e.g., histone mRNAs, not required for synthesis, processing, transport or translation

A. Tail invariably starts ~20 bases downstream from AAUAAA, a recognition site in 1° transcript for the assembly of a protein complex that carries out processing reactions at the mRNA 3' end

1. Studies suggest that the assembled poly(A) processing complex is physically associated with the RNA polymerase as it synthesizes the 1° transcript

2. Included among proteins of processing complex is an endonuclease that cleaves the pre-mRNA just downstream from the recognition site

3. After nuclease cleavage, poly(A) polymerase adds ~250 adenosines without the need of template

B. The poly(A) tail together with an associated protein protects the mRNA from premature degradation by exonucleases

C. Poly(A) tail also allows mRNA purification by affinity chromatography - pass cellular RNA mix through column with bound synthetic poly(T) —> poly(A) RNAs (mRNAs) stick, the rest go through

1. This removes the mRNAs from solution, while more abundant tRNAs, rRNAs & snRNAs that lack the poly(A) tail pass through along with the aqueous solvent

IV. Splicing the primary transcript: removal of intervening sequences or introns (**RNA splicing**)

A. Requires break at 5' & 3' intron ends (splice sites) & covalent joining of adjacent exons (ligation) situated on either side of the splice site

1. Splicing process must be absolutely precise because a single base error (addition or loss) at any of the splice junctions changes the reading frame & results in mRNA mistranslation

2. Hundreds of exon-intron junctions were examined in eukaryotes from yeast to insects to mammals, revealing conserved sequence of ancient evolutionary origin at exon-intron junction

3. Most common conserved sequence at eukaryotic exon-intron borders in mammalian pre-mRNA is G/GU at 5' intron end (5' splice site) & AG/G at 3' end (3' splice site)

4. These sequences are in vast majority of eukaryotic pre-mRNAs **but……..**

B. ~1% of introns have AT & AC dinucleotides at 5' & 3' intron ends (rather than GU & AG), respectively

1. These AT/AC introns are processed by a different type of spliceosome (U12 spliceosome)

2. It contains a U12 snRNA instead of the U2 snRNA of the major spliceosome

3. U12 spliceosomes are absent from yeast & nematodes, but present in plants, insects & vertebrates

C. In addition, adjacent regions of intron contain preferred nucleotides that play an important role in splice site recognition; in contrast, intron interior is highly divergent

D. These sequences are necessary for splice-site recognition, but they are not sufficient

1. Introns typically run for 1000s of nucleotides & often contain segments that match the above consensus sequences without them being recognized as splicing signals; the cell ignores them
2. Additional clues that allow splicing machinery to distinguish between exons & introns are thought to be provided by specific sequences (exonic splicing enhancers or ESEs) situated within the exons
3. Changes in DNA sequence within either a splice site or an ESE can lead to inclusion of an intron or the exclusion of an exon
4. There is an estimate that up to 15% of inherited human disease results from mutations that alter pre-mRNA splicing, like certain types of human thalassemia

E. Remarkable & unexpected RNA catalytic abilities led to an understanding of splicing mechanism

1. Thomas Cech et al. (1982, U. of Colorado) found the first evidence of RNA catalysis in pre-rRNA that could splice itself in the ciliated protozoan *Tetrahymena* (**RNA enzymes** or **ribozymes**)
2. Experiments revealed the existence of RNA enzymes or ribozymes & changed the thinking of biologists about the relative roles of RNA & protein in the mechanism of RNA splicing

V. Two types of intron splicing mechanisms have been described – group I & group II introns

A. Group I introns (*Tetrahymena* pre-rRNA is an example) - most common in fungal/plant mitochondria, plant chloroplasts & in nuclear RNA of lower eukaryotes, like *Tetrahymena*

1. Group I intron sequences are quite variable, but their 3D structures are very similar
2. This complex folded structure along with well-placed patches of conserved nucleotides allows introns to splice themselves out of larger RNA

B. Group II introns - also self-splicing; seen in fungal mitochondria, plant chloroplasts & a wide variety of bacteria

1. Group II introns fold into complex structure very different from Group I introns; they undergo self-splicing by passing through intermediate stage (lariat, like cowboy rope for catching calves)
2. First step is cleavage of 5' splice site followed by formation of lariat by means of covalent bond between 5' end of intron & an adenosine residue near the 3' end of intron
3. The subsequent cleavage of 3' splice site releases lariat & allows cut ends of exons to be covalently joined or ligated

VI. Steps in intron removal from animal cell pre-mRNAs is very similar to those followed by Group II introns - difference is intron can't splice itself; it needs many snRNAs (small nuclear RNAs) & their associated proteins

A. As each large hnRNA molecule is transcribed, it associates with a variety of proteins to form an hnRNP (heterogeneous nuclear ribonucleoprotein), which is substrate for processing reactions that follow

1. Our understanding of RNA splicing steps achieved largely through studies of cell-free extracts
2. They accurately splice pre-mRNAs *in vitro*

B. Processing occurs as each intron of pre-mRNA becomes associated with macromolecular complex called a **spliceosome**

1. Each spliceosome consists of a variety of proteins & a number of distinct ribonucleoprotein particles (snRNPs; pronounced snurps), since they are composed of snRNAs bound to specific proteins

2. Spliceosomes are not present within nucleus in a prefabricated state, but they are assembled as their component snRNPs bind to the pre-mRNA

C. Once spliceosome machinery is assembled, snRNPs carry out the reactions that cut the introns out of the transcript & paste the ends of the exons together

D. Taken together, intron removal requires several snRNP particles: U1 snRNP, U2 snRNP, U5 snRNP & U4/U6 snRNP (contains U4 & U6 snRNAs bound together)

VII. In addition to its snRNA, each snRNP contains a dozen or more proteins

A. One family, the Sm proteins, is present in all of the snRNPs; they bind to one another & to a conserved site on each snRNA (except U6 snRNA) to form the core of the snRNP

1. Sm proteins were first identified because they are the targets of antibodies produced by patients with the autoimmune disease systemic lupus erythematosus

B. The other proteins of the snRNPs are unique to each particle

VIII. Multiple rearrangements among RNA molecules that occur during spliceosome assembly are probably mediated by ATP-consuming, RNA helicases present within snRNPs

A. RNA helicases unwind double-stranded RNAs, like the U4-U6 duplex; this allows the displaced RNAs to bind new partners as happens with U2-U6 interaction

1. Spiceosomal helicases are also thought to strip RNAs from bound proteins, including U2AF protein
2. At least 8 different helicases are implicated in the splicing of yeast pre-mRNAs

B. The following facts suggest that snRNAs are the catalytically active snRNP components, not the proteins

1. Pre-mRNAs are spliced by the same pair of chemical reactions that occur as group II introns splice themselves
2. The snRNAs needed for splicing pre-mRNAs closely resemble parts of the group II introns

C. The spliceosome is most likely a ribozyme, a proposal that has received much experimental support

1. Of the various snRNAs that participate in RNA splicing, U6 is most likely to act as a ribozyme, making both cuts in the pre-mRNA required for intron removal

D. The proteins are thought to serve supplementary roles like:

1. Maintaining the proper 3D structure of the snRNA
2. Driving changes in snRNA conformation
3. Transporting spliced mRNAs to the nuclear envelope
4. Selecting the splice sites to be used during the processing of a particular pre-mRNA

IX. The exonic splicing sequences (ESEs) that play a key role in exon recognition by splicing machinery serve as binding sites for a family of RNA-binding proteins called SR proteins

A. SR proteins are so named due to their large number of arginine [R] - serine [S] dipeptides & are thought to form interlacing networks that span intron/exon borders & help recruit snRNPs to splice sites

1. Positively charged SR proteins may also bind electrostatically to negatively charged phosphate groups that are added to the CTD of the polymerase as transcription is initiated

B. Thus, assembly of splicing machinery at an intron occurs in conjunction with the synthesis of an intron by the polymerase

1. The CTD is thought to recruit a wide variety of processing factors
2. In fact, most of RNA processing machinery apparently travels with polymerase as part of giant mRNA factory

C. Most genes have a number of introns & splicing reactions must occur repeatedly on single 1° transcript

1. Evidence suggests that introns are removed in a preferred order, generating specific processing intermediates whose size lies between that of 1° transcript & mature RNA

Evolutionary Implications of Split Genes and RNA Splicing

I. Discovery that RNAs are capable of catalyzing chemical reactions chemical reactions has had an enormous impact on our view of biological evolution

A. Ever since DNA's discovery as genetic material the big evolutionary question has been: Which came first - proteins (catalysts) or DNA (information storage)?

1. Dilemma arose from the seemingly nonoverlapping functions of these 2 types of macromolecules
2. Ribozyme discovery in early 1980s showed that one type of molecule - RNA - could do both

B. Led to suggestion that RNA did both at the start of life & that DNA & proteins were probably absent at an early stage in the evolution of life

1. During this period, RNA molecules performed double duty: they served as genetic material & catalyzed chemical reactions including those required for RNA replication
2. It was an RNA world, but at later stage in evolution, RNA's functions, catalysis & information storage, were taken over by protein & DNA, respectively
3. RNA was left to function as a go-between in the flow of genetic information
4. Many researchers believe that splicing may be an example of a legacy from an ancient RNA world

C. Group II introns found in purple bacteria & cyanobacteria (close relatives of chloroplast/mitochondria ancestors) supports idea that group II introns may be the source that gave rise to pre-mRNA introns

1. Introns were originally present in endosymbiotic organelles residing in primitive eukaryotic cells
2. Over time, introns left organelle DNA & invaded nuclear DNA, like mitochondrial/chloroplast genes
3. Once in nucleus, the introns may have moved place to place by transposition; some introns of modern nuclear genes can still act like mobile genetic elements
4. Since they could self-splice, the presence of introns in the middle of gene would not have been a problem; sequences corresponding to introns would have just excised themselves from 1° transcript
5. Over time, catalytic portions of introns moved from the interior of protein-coding genes to separate genome locations; RNAs encoded by these "new" genes kept participating in splicing process
6. Eventually, the RNAs encoded by these genes evolved into snRNAs whose catalytic activity became dependent on proteins
7. Together, the snRNAs & proteins evolved into the snRNP components of the spliceosome
8. As this occurred, the internal intron nucleotides no longer had function —> this explains their variable length & divergent nucleotide sequences

II. Presence of introns may have created added cellular burden, since cells had to remove intervening sequences from their transcripts, but they still have some virtues

A. RNA splicing is one of the steps along path to mRNA formation that is subject to cellular regulation
   1. Many 1° transcripts can be processed by ≥2 pathways so that a sequence that acts as an intron in one pathway becomes an exon in an alternate pathway
   2. Because of this **alternate splicing**; the same gene can code for >1 polypeptide

B. It was also found that snoRNAs (needed for rRNA processing) are encoded by introns rather than exons
   1. Most of these snoRNAs are within the introns of genes coding for polypeptides involved in ribosome assembly & function (ribosomal proteins, translation factors)
   2. When such genes are transcribed & introns excised, some are processed into snoRNAs, not discarded
   3. Several genes are known in which roles of introns & exons are reversed - gene transcribed into 1° transcript whose introns are processed into snoRNAs, while exons are degraded —> yield no mRNA

III. Presence of introns may also have had a major impact on biological evolution

A. Proteins often contain sections that are homologous to parts of several other proteins
   1. Most probably, genes like this are encoded by genes that are composites of parts of other genes
   2. Reflects probable movement of genetic modules among unrelated genes (**exon shuffling**)
   3. Greatly facilitated by presence of introns, which act like inert spacer elements between exons
   4. Genetic rearrangements require breaks in DNA, which can occur within introns without causing dangerous mutations that might impair the organism

B. Over a long time, exons can be shuffled independently in various ways, allowing a nearly infinite number of combinations in search for new & useful coding sequences
   1. With exon shuffling, evolution need not occur only by slow accumulation of point mutations
   2. May allow evolution to jump forward in quantum leaps (get new proteins in a single generation)

IV. Creating new ribozymes in the laboratory - make new RNA catalysts to explore RNA catalytic potential

A. To date, only a few reactions have been found to be catalyzed by naturally occurring RNAs:
   1. Cleavage & ligation of phosphodiester bonds needed for RNA splicing
   2. Formation of peptide bonds during protein synthesis

B. Are the above the only kinds of reactions that RNAs can catalyze or has the catalytic repertoire been sharply restricted by the evolution of more efficient protein enzymes?

C. Some researchers are creating catalytic RNAs from "scratch" without preconceived design to explore catalytic potential of RNA – could prove principle that RNAs could evolve through natural selection
   1. Let automated DNA-synthesizing machines assemble DNAs with random nucleotide sequences
   2. Transcription of these DNAs produces RNA population with correspondingly random sequences
   3. Select RNAs from population by virtue of their particular properties (molecular or test tube evolution in lab)

D. One study selected synthetic RNAs for their ability to bind specific ligands (nucleoside triphosphate, amino acid) via affinity chromatography (link ligands to beads, pack them in chromatography column)

1. Pass RNAs through column & those that bind ligand stick to column; others (most) pass through
2. Once isolated by binding properties, increase their numbers under conditions that introduce random substitutions (mutations) into their base sequence
3. Select again with more stringent conditions; with each round, binding affinity for ligand gets higher
4. Assume that some RNAs that bind tightly may have catalytic abilities that could modify the ligand

E. In one study, researchers initially selected for RNAs that bound to specific amino acids & later for a subpopulation that would transfer a specific amino acid onto 3' end of targeted tRNA
1. This is same basic reaction as that carried out by aminoacyl tRNA synthetases that attach amino acids to tRNAs prior to protein synthesis
2. It is speculated that amino acids may have been used initially as adjuncts (cofactors) to enhance catalytic reactions carried out by ribozymes
3. Over time, ribozymes presumably evolved that were able to string specific amino acids together to form small proteins, which were more versatile catalysts than their RNA predecessors
4. Ribosomes, which synthesize proteins, are really ribozymes at heart – strong support for scenario

F. As proteins took over a greater share of the workload in the primitive cell, the RNA world was gradually transferred into an RNA-protein world
1. Later, RNA was presumably replaced by DNA as the genetic material, which led to a DNA-RNA-protein world
2. DNA evolution may have required only 2 types of enzymes: ribonucleotide reductase to convert ribonucleotides into deoxyribonucleotides & reverse transcriptase to transcribe RNA to DNA
3. The fact that RNA catalysts do not appear to be involved in either DNA synthesis or transcription supports the idea that DNA was the last of the three to appear on the scene
4. At some point, a code evolved to allow the genetic material to specify the sequence of amino acids to be incorporated into a given protein

Small Non-Coding RNAs and RNA Interference

I. RNA interference or dsRNA-mediated interference (RNAi) – an extraordinary phenomenon discovered in mid-1990s in studies on a wide variety of organisms, including plants, nematodes & fruit flies

A. Double-stranded RNAs (dsRNAs) were taken up by cells where they induced a response leading to the selective destruction of mRNAs having the same sequence as the added dsRNA

B. Example: you want to stop the production of the enzyme phosphorylase in the cells of a nematode worm so that the effect of this enzyme deficiency on the worm's phenotype can be determined
1. Accomplish it simply by soaking the worm in a solution of dsRNA sharing the same sequence as the target mRNA
2. This is similar in effect, although entirely different in mechanism, to the formation of knockout mice that lack a particular gene encoding a particular protein

C. RNAi was first demonstrated by Andrew Fire (Carnegie Inst. of Washington) & Craig Mello (Univ. of Mass.) in 1998; it can be shown to occur in virtually every type of eukaryotic organism
1. It is thought to have evolved as a type of "genetic immune system" to protect organisms from the presence of foreign or unwanted genetic material

2. RNAi probably evolved as a mechanism to block viral replication and/or to suppress the movements of transposons within the genome
3. Both the above potentially dangerous processes typically involve dsRNA intermediate formation —> cells recognize dsRNA as undesirable since they are not made by cell's normal genetic activities

D. How can the presence of dsRNA within a cell block formation of a particular protein? – answer comes from studies on RNA interference carried out on cell-free extracts
1. The dsRNA that starts the response in extracts is first cleaved into small (21–23 nucleotide), double-stranded fragments (small-interfering RNAs [siRNAs]) by a particular type of ribonuclease (Dicer)
2. Following their separation, each of the individual strands of siRNA becomes associated with a group of proteins to form a complex that binds a complementary mRNA
3. Once the siRNA in the complex has bound to the target mRNA, the bound mRNA is cleaved by an associated ribonuclease
4. The siRNA acts like a specific guide to direct the activity of an associated enzyme toward a complementary target RNA
5. This is not unlike the way snoRNAs act as guides directing the activity of methylation enzymes to complementary sites on rRNAs
6. Each siRNA can orchestrate the destruction of many copies of the mRNA, thereby blocking the synthesis of the encoded protein

E. Until 2001, RNAi could only be reliably demonstrated in non-mammalian cells
1. When dsRNA is added to mammalian cells in culture or injected directly into mammal's body, it usually starts a global response that inhibits protein synthesis in general
2. This happens instead of stopping specific protein synthesis & is thought to have evolved in early mammals as a means of protecting cells from infection by viruses
3. To get around the large-scale global response, researchers began to use very small dsRNAs
4. They found that treatment of mammalian cells with dsRNAs that were 21 nucleotides long did not trigger global inhibition of protein synthesis
5. These smaller dsRNAs were equivalent in size to the siRNAs produced as intermediates during RNA interference in other organisms
6. They were able to inhibit the synthesis of the specific protein encoded by an mRNA having a matching nucleotide sequence; proteins encoded by other mRNAs were not affected
7. By using dsRNAs of the size produced by the ribonuclease involved in RNAi, researchers were able to induce the degradation of specific mRNAs in mammalian cells
8. This technique has become a major tool helping geneticists to learn more about the function of newly identified genes
9. One can knock out the activity of the gene in question, using dsRNAs to destroy the mRNAs it generates & look for cellular deficiencies resulting from the absence of the encoded protein

II. Micro-RNAs: Hundreds of RNAs of unknown function

A. In 1993, a small RNA was discovered in the nematode *C. elegans* that was nearly complementary to segments in the 3'-untranslated region of specific mRNAs produced by the organism
1. During larval development, this small RNA binds to the complementary mRNA, blocking translation of the message, which apparently triggers a transition to the next developmental stage
2. Broader importance of observation was not appreciated until 2000, when it was shown that one small worm RNA (a 21-nucleotide species called *let-7*) is highly conserved throughout evolution
3. Humans encode several RNAs that are either identical or nearly identical to *let-7*

B. In recent years, it has become evident that both plants & animals produce hundreds (perhaps thousands) of tiny RNAs called micro RNAs (miRNAs); due to their small size, they were undetected for decades
   1. As in nematodes, miRNAs are synthesized only at certain times during development or in certain tissues of a plant or animal & are presumed to play a regulatory role
   2. Sizes of miRNAs (at roughly 20 – 25 nucleotides in length) place them in the same size range as the siRNAs involved in RNAi

C. Micro RNAs are produced by the same processing machinery responsible for siRNA formation
   1. Each miRNA is carved from a double-stranded, fold-back RNA precursor that serves as a substrate for the Dicer ribonuclease
   2. Big difference between siRNA & miRNA - siRNA derived from transposable element, ds viral product or dsRNA from researcher, while miRNA is encoded by conventional genome segment

D. Mutations in one of Dicer genes in plant *Arabidopsis* lead to a range of abnormal developmental phenotypes, suggesting miRNAs may play role in turning genes on and/or off during development
   1. Many of the miRNAs likely function as translational inhibitors, just like the ones originally discovered in nematodes
   2. Recent plant, yeast & protist studies hint at wide range of other potential functions, including gene transcription suppression, selective mRNA cleavage & induction of genomic rearrangements

E. Noncoding RNAs are probably key players in regulation of complex gene expression networks
   1. Analysis of RNAs made by an extensive array of mouse tissues suggests that thousands of different non-coding RNAs are made by these animals – they will be a major area of study

Encoding Genetic Information

I. Once DNA structure was discovered in 1953, it became obvious that the sequence of amino acids in a polypeptide was specified by the sequence of nucleotides in the DNA of a gene

A. It seemed unlikely that DNA could serve as a direct, physical template for protein assembly

B. It was presumed instead that information stored in the nucleotide sequence was present in some type of genetic code

C. The discovery of mRNA as an intermediate in information flow from DNA to protein raised questions about how the code might be constructed; among them was the one below:
   1. How can the ribonucleotide alphabet (A,C,G,U) code for the 20 words in the amino acid sequences?

II. Properties of the genetic code

A. George Gamow, physicist – proposed that each amino acid is encoded by 3 sequential nucleotides (triplet codon); the code words for amino acids were called codons
   1. In order to code unambiguously for 20 amino acids needed at least 3 letter codons (64 of them); 4 (4^1) possible one-letter words, 16 (4^2) possible 2-letter words & 64 (4^3) possible 3-letter words
   2. The triplet nature of the code was verified by some insightful genetic experiments by Francis Crick, Sydney Brenner & their colleagues at Cambridge University

B. Gamow also proposed code is overlapping, but he was wrong
   1. If triplet code was overlapping, each nucleotide would be part of 3 codons; if it is nonoverlapping, each nucleotide would be part of only one codon

2. Look at mutant proteins (sickle cell hemoglobin) which change only 1 base —> affects only 1 amino acid (if code overlaps 3 consecutive aminos should be affected) —> thus, code is nonoverlapping

C. There are 64 possible triplet codons but only 20 amino acids, therefore the code is degenerate (in fact, it is highly degenerate); what do the extra 44 codons do? – nearly all 64 codons specify amino acids
   1. Some aminos coded for by >1 codon; 3 codons are termination codons (punctuation) that code for no amino acids; they cause the reading of the message to stop
   2. Crick predicted degeneracy; noted great range in base composition among DNAs of various bacteria
   3. A difference in G + C content (20-74%) of various organisms had been noted, while amino acid composition of proteins varied little overall (so same aminos encoded by different base sequences)

III. Identifying the codons – by 1961, the code's general properties were known, but no coding assignments of specific triplets had been made

A. Marshall Nirenberg & Heinrich Matthaei - made artificial genetic messages & determined protein encoded in cell-free protein synthesis system
   1. Cell-free protein synthesis system was bacterial extract, 20 amino acids & materials needed for protein synthesis, like ribosomes & various soluble factors
   2. First message was a poly(U) (a polyribonucleotide consisting exclusively of uridine) & it coded for polyphenylalanine (a polymer of phenylalanine), so UUU codes for phenylalanine
   3. These two and others made more synthetic mRNAs & eventually deciphered the code

B. Codon assignments are essentially universal (in all living organisms) regardless of cell type (or organelle); there are exceptions (first seen in mitochondrial mRNAs); examples from human mitochondria below:
   1. UGA is read as tryptophan rather than stop; AUA is read as methionine rather than isoleucine; AGA & AGG are read as stop rather than arginine
   2. More recently, exceptions found here & there, in nuclear codons of protists & fungi
   3. Code similarities in mitochondria & nuclei outweigh differences
   4. It is evident that minor deviations have evolved as secondary changes from standard genetic code
   5. Thus, all known organisms present on Earth today share a common evolutionary origin

C. Codon assignments are distinctly nonrandom; codons coding for same amino acid are generally similar & tend to be clustered within a particular portion of the coding chart
   1. Clustering reflects similarity in codons that specify the same amino acid
   2. Thus, spontaneous mutations causing single base changes in a gene often do not cause a change in the amino acid sequence of the corresponding protein
   3. Another safeguard is that chemically similar amino acids are coded for by similar codons
   4. For example, hydrophobic amino acid codons are similar & clustered on the coding chart so in the case of a single base mutation, a hydrophobic amino acid often replaces another hydrophobic amino
   5. In addition, the greatest similarity between amino acid-related codons usually occurs in first 2 nucleotides of triplet; while the greatest variability occurs in the third nucleotide
   6. Example: glycine is encoded by 4 codons, all of which begin with the nucleotides GG

Decoding the Codons: The Role and Structure of the Transfer RNAs

I. Protein synthesis is called translation because the language of nucleic acids must be converted to the language of amino acids – decoding is accomplished by transfer RNAs, which act like adaptors
   A. Each tRNA is linked to a specific amino acid (an aa-tRNA)
   B. Each tRNA is also able to recognize a particular codon in the mRNA
   C. Interaction between successive codons in mRNA & specific aa-tRNAs leads to synthesis of polypeptide with an ordered amino acid sequence

II. Crick proposed a nucleic acid adaptor that would decode the genetic message through base pairing
   A. Soon after, tRNA was found, fulfilled his prediction; each tRNA recognizes mRNA codon & an amino acid; Robert Holley (Cornell, 1965) sequenced yeast alanine-tRNA (77 nucleotides, 10 modified)
   B. Now more tRNAs have been purified & sequenced & they have a number of distinct similarities that have been found in all of them
      1. All have relatively small, similar numbers of nucleotides (73 - 93) & are thus roughly same length
      2. All have a significant number of unusual bases made by altering normal base posttranscriptionally
      3. All have base sequences in one part of molecule that are complementary to those in other parts
      4. Thus, all of them have the potential to fold in a similar way to form cloverleaf-like structure (in 2 dimensions) with base-paired stems & unpaired loops
      5. Amino acid carried by the tRNA is always attached to A (adenosine) at 3' end of molecule
      6. Unusual bases mostly concentrated in the loops where they disrupt H bond formation in these regions; also serve as potential recognition sites for various proteins
   C. tRNAs fold into a unique & defined tertiary structure; X-ray diffraction has shown tRNAs are constructed of 2 double helices arranged in the shape of an L
      1. Form L-shaped 3° structure by weak interactions between certain areas of the molecule; invariant bases are responsible; they are found at comparable sites in all tRNA molecules
      2. Common shapes reflect fact that they all take part in the similar series of reactions of protein synthesis
      3. While similar in shape, each tRNA has unique features; thus, each one can be distinguished from the others allowing attachment of an amino acid to the appropriate (cognate) tRNA
   D. The middle tRNA loop has an anticodon (a stretch of 3 sequential nucleotides) that H bonds to the complementary mRNA codon, allowing proper amino acid placement
      1. Loop is always made of 7 nucleotides (the middle 3 of which constitute the anticodon); anticodon is located at one end of L-shaped tRNA opposite from the end at which the amino acid is attached
      2. The greatest similarities among codons specifying the same amino acid occur in first 2 nucleotides of the triplet, whereas the greatest variability in these same codons is the third nucleotide of triplet
      3. Base at third position in codon not as important; same tRNA can recognize >1 codon
      4. In 16 codons ending in U, if you change U to C —> you specify same amino; change third site from A to G —> also usually has no effect on amino acid determination
   E. The interchangeability of the base at the third position led Francis Crick to propose that the same tRNA may be able to recognize >1 codon
      1. Crick's wobble hypothesis – the steric requirement between the anticodon of the tRNA & the codon of the mRNA is very strict for first two positions, but more flexible at third position

2. Thus, 2 codons specifying the same amino acid & differing only at third position should use the same tRNA in protein synthesis – once again Crick was correct

F. Rules governing the wobble at codon 3rd position (proof – purified tRNAs can bind to more than one codon for a particular amino acid; example: due to wobble, 6 leucine codons require only 3 tRNAs)

1. U of the anticodon can pair with A or G of mRNA
2. G of anticodon can pair with U or C of mRNA
3. I (inosine, modification of guanine in original tRNA) of anticodon pairs with U, C or A of mRNA

III. Amino acid activation - each tRNA **must** attach to correct (cognate) amino acid

A. Aminoacyl-tRNA synthetases (aaRS) covalently link amino acids to 3' end of cognate tRNA

1. Each amino acid is recognized by specific aaRS which charges appropriate tRNAs for amino acid (any tRNA whose anticodon recognizes one of the various codons specifying that amino acid)
2. aaRSs surprisingly heterogeneous in structure although it was thought they would be similar

B. aaRSs are excellent examples of the specificity of protein-nucleic acid interactions

1. There must be common traits in all tRNAs specifying a given amino acid to allow a single aaRS to recognize all of these tRNAs
2. At the same time, there must be traits that allow aaRSs to discriminate against all of the tRNAs for the other amino acids

IV. Techniques used to learn tRNA structural variations that account for an aaRS's ability to select or reject tRNAs as substrates:

A. aaRS 3D structure determination by X-ray crystallography - find which tRNA sites make direct contact with the aaRS

1. The acceptor stem & the anticodon (the 2 ends of the tRNA) are particularly important for recognition by most of these enzymes

B. Find tRNA changes leading the tRNA to be aminoacylated by noncognate aaRS

1. A specific base pair in a $tRNA^{Ala}$ (the G-U base pair involving the 3rd G from the 5' end of the molecule) is the primary determinant of its interaction with alanyl-tRNA synthetase
2. Insertion of this specific base pair (G-U) into acceptor stem of a $tRNA^{Phe}$ or a $tRNA^{Cys}$ is sufficient to cause these tRNAs to be recognized by alanyl-tRNA synthetase & to be aminoacylated with alanine

V. aaRSs carry out a two-step reaction:

A. ATP + amino acid —> aminoacyl-AMP + PP_i

1. ATP energy activates amino acid by formation of an adenylated amino acid (aminoacyl-AMP), which is bound to enzyme & stays attached to aaRS initially after first step
2. Step 1 is the primary energy-requiring step in chemical reactions leading to polypeptide synthesis
3. Later events, like transfer of amino acid to tRNA (below) & then to the growing polypeptide chain are thermodynamically favored
4. PP_i made in first step is hydrolyzed to P_i later for more energy, further driving the overall reaction toward the formation of products
5. Energy is expended during protein synthesis, but it is not used in the formation of peptide bonds

B. aminoacyl-AMP + tRNA —> aminoacyl-tRNA + AMP

1. In second step, the aaRS transfers its bound amino acid to the 3' end of a cognate tRNA; the aaRS has a proofreading mechanism, which severs the amino acid from tRNA if the wrong one is attached
2. Not all aaRSs possess the above type of proofreading mechanism; some of them remove an incorrect amino acid by hydrolyzing the aminoacyl-AMP linkage after the first reaction step
3. The 2 amino acids most difficult to distinguish are valine & isoleucine, which differ by a single methylene group
4. The isoleucyl-tRNA synthetase employs both types of proofreading mechanisms to ensure accurate aminoacylation

VI. aaRSs play key role in translation of mRNA nucleotide sequence into amino acid sequence of polypeptide since they determine which amino acid is linked to a particular tRNA – Fritz Lipmann et al.

A. Lipmann & his colleagues chemically altered an amino acid after it was attached to its specific tRNA

1. Charged $tRNA^{Cys}$ with cysteine, then converted the attached cysteine to alanine & allowed these substituted amino acids to be incorporated during protein synthesis
2. Alanine was inserted in protein where cysteine should have been, so the unchanged anticodon determined the placement of the amino as it would have if amino acid were unaltered
3. The anticodon (recognition group of tRNA) was unaltered & continued to recognize the same mRNA nucleotides it would have if the chemical alteration of the amino acid had not been made

B. Thus, the amino acid itself plays no direct role in determining where it is incorporated in polypeptide

Translating Genetic Information: Background Information and Prokaryotic Initiation

I. Protein synthesis (**translation**) may be the most complex synthetic activity in a cell

A. It requires a number of cellular constituents:

1. All of the various tRNAs with their attached amino acids
2. Ribosomes
3. A mRNA
4. Numerous proteins having different functions
5. Cations
6. GTP

B. Complexity is not surprising given that protein synthesis requires incorporation of 20 different amino acids in precise sequence dictated by coded message written in language using different elements

C. Bacterial & eukaryotic translation are remarkably similar – bacterial translation is simpler & better understood; primary difference is that eukaryotes need more soluble (nonribosomal) protein factors

D. Both processes are divided into 3 distinct activities: initiation, elongation & termination of the chain

II. Ribosome starts translation at a specific mRNA site (the initiation codon) by attaching there

A. The initiation codon is AUG & by binding there the ribosome establishes the reading frame

B. The ribosome then moves ahead in consecutive 3 base blocks;

III. Initiation Step 1: Bringing the small ribosomal subunit to the initiation codon - mRNA binds to ribosome subunits separately

A. The small ribosomal subunit binds to the first (or one of the first) AUGs in mRNA (initiation codon)

1. GUG can be the initiation codon & serves as such in a few natural messages
2. When GUG is the initiation codon, N-formylmethionine is still used to form the initiation complex
3. When GUG is internal to the message & not on the end, the codon specifies valine

B. How does the small ribosomal subunit select the initial AUG as opposed to an internal one?

1. Bacterial mRNAs have the specific Shine-Dalgarno (S-D) sequence 5 - 10 nucleotides before initiation codon
2. S-D sequence is complementary to sequence near 3' end of 16S rRNA of small ribosomal subunit
3. Interaction between these complementary sequences on the mRNA & rRNA leads to attachment of a 30S subunit at the AUG codon

C. Several steps require initiation factors (IFs in prokaryotes; eIFs in eukaryotes) - prokaryotic cells require 3 initiation factors, IF1, IF2 & IF3, which bind 30S subunit & help it attach to mRNA

1. IF2 is a GTP-binding protein required for attachment of the first aminoacyl-tRNA
2. IF3 may prevent the large (50S) subunit from joining prematurely to the 30S subunit & also facilitate entry of initial aa-tRNA
3. IF1 facilitates attachment of 30S subunit to the mRNA & may prevent the aa-tRNA from entering the wrong site on the ribosome

IV. Step 2: Bringing the first aa-tRNA into ribosome - AUG is initiation codon & the only methionine codon

A. Methionine is always the first amino acid incorporated at N-terminus of a nascent polypeptide chain

1. In prokaryotes, this initial methionine bears a formyl group (N-formylmethionine)
2. Usually, methionine or N-formylmethionine is later removed enzymatically; ~50% of the time it remains in mature polypeptide
3. Cells have 2 distinct methionyl-tRNAs: one initiates ($tRNA_i^{Met}$); the other puts methionine in interior position of a polypeptide ($tRNA^{Met}$)
4. Mitochondria & chloroplasts initiate with N-formylmethionine (evidence for endosymbiosis)

B. Aminoacylated initiator tRNA enters the preinitiation complex at the P site by binding to both the AUG codon of mRNA & the IF2 initiation factor; IF1 & IF3 are released

V. Step 3: Assembling the complete initiation complex - happens after initiator tRNA is bound to the AUG codon & IF3 is displaced

A. Large subunit joins the complex & the GTP bound to IF2 is hydrolyzed

B. GTP hydrolysis probably drives ribosome conformational shift required for the release of IF2-GDP

Translating Genetic Information: Eukaryotic Initiation

I. Eukaryotic cells require at least 12 initiation factors comprising a total of >25 polypeptide chains

A. Several of these eIFs (eIF1, eIF1A, eIF3) bind to the 40S subunit, which prepares the subunit for binding to the mRNA

II. The initiator $tRNA_f^{Met}$ also binds the 40S subunit prior to its interaction with the mRNA
   A. The initiator tRNA enters the P site of the subunit in association with eIF2-GTP, which is homologous to the bacterial IF2-GTP
   B. Next, the small ribosomal subunit with its associated initiation factors & charged tRNA, which together form a 43S preinitiation complex, binds to 5' end of mRNA, bearing the methylguanosine cap

III. 43S complex is originally recruited to mRNA with the help of a cluster of initiation factors that are already bound to the mRNA; among them:
   A. eIF4E binds to the 5' cap of the eukaryotic mRNA
   B. eIF4A moves along the 5' end of the message removing any double-stranded regions that would interfere with the movement of the 43S complex along the mRNA
   C. eIF4G serves as linker between the 5' capped end & 3' polyadenylated end of the mRNA; it, in effect, converts a linear mRNA into a circular message
      1. The reasons for the circularization of eukaryotic mRNAs are not entirely clear

IV. After the 43S complex binds to mRNA 5' end, the complex then scans along message until it reaches a recognizable nucleotide sequence (typically 5'-CCACC**AUG**C-3'), containing AUG initiation codon
   A. Once the 43S complex reaches the AUG codon, eIF2-GTP is hydrolyzed
   B. eIF2-GDP & other associated eIFs are released & the large 60S subunit joins the complex to complete initiation
      1. Not all mRNAs are translated after attachment of the small ribosomal subunit to the 5' end of the message
      2. Many viral mRNAs & a relatively small number of cell mRNAs, most notably those used during mitosis, are translated by ribosome attaching to the mRNA at internal ribosome entry site (IRES)
      3. The IRES may be located at some distance from the 5' end of the message

Translating Genetic Information: The Role of the Ribosome

I. Ribosomes are molecular machines similar in some ways to molecular motors; during translation, ribosome undergoes repetitive cycle of mechanical changes driven by energy from GTP hydrolysis
   A. Unlike myosin or kinesin, which move along physical track, ribosome moves along an mRNA tape containing encoded information
   B. Ribosomes are programmable machines; the information stored in mRNA determines sequence of aminoacyl-tRNAs the ribosome uses in translation
   C. They are distinguished from many other cellular machines by the importance of their component which play major roles in translation
      1. Select tRNAs & bind protein factors
      2. Catalyze a key reaction, the formation of the peptide bond (polymerize amino acids)
      3. Ensure accurate translation

II. Early high-resolution EM imaging studies - ribosomes shown to have highly irregular structure with bulges, lobes, channels & bridges
   A. These studies provided evidence of major conformational changes occurring in both the small & large ribosomal subunits during translation

B. During 1990s, major advances made in crystallization of ribosomes & led to X-ray crystallographic studies by end of decade

C. Each ribosome has 3 sites for association with tRNAs; the sites receive each tRNA in successive steps of elongation cycle

1. A (aminoacyl) site - generally aminoacyl tRNAs enter here (except for $tRNA_i^{Met}$ probably)
2. P (peptidyl) site - tRNAs donate amino acid(s) of growing chain from here
3. E (exit) site - tRNA leaves from here after losing attached amino acid(s)

D. The tRNAs bind within these sites & span the gap between the 2 ribosomal subunits

1. The anticodon ends of the bound tRNAs contact the small subunit, which plays a key role in decoding the information contained in the mRNA
2. In contrast, the amino acid-carrying ends of bound tRNAs contact the large subunit, which plays a key role in catalyzing peptide bond formation

III. Recent high-resolution structural studies include the following:

A. Interface between the small & large subunits forms a relatively spacious cavity that is lined almost exclusively by RNA

1. The side of the small subunit that faces this cavity is lined along its length by a single, continuous, double-stranded RNA helix
2. The surfaces of the 2 subunits that face one another contain the binding sites for the mRNA & incoming tRNAs & are thus of key importance for the function of the ribosomes
3. The fact that these surfaces consist largely of RNA supports the proposal that primordial ribosomes were composed exclusively of RNA

B. The active site, where amino acids are covalently linked to one another, also consists of RNA

1. This catalytic portion of large subunit resides in a deep cleft, which protects the newly formed peptide bond from hydrolysis by the aqueous solvent

C. The mRNA is situated in a narrow channel that winds around the neck of the small subunit, passing through the A, P & E sites

D. A tunnel runs completely through the large subunit beginning at the active site - this tunnel may provide a passageway for translocation of the elongating polypeptide through the ribosome

E. Most of the proteins of the ribosomal subunits have multiple RNA-binding sites & are ideally situated to stabilize the complex tertiary structure of the rRNAs

Translating Genetic Information: Elongation

I. Step 1: Aminoacyl-tRNA selection - after initiation, charged initiator tRNA is in P site & A site is empty

A. Second aminoacyl-tRNA enters vacant A site in first elongation step

1. Before second aminoacyl-tRNA can effectively bind the exposed mRNA in A site, tRNA must bind with a protein elongation factor linked to GTP (EF-Tu or Tu - prokaryotes; eEF1α - eukaryotes)
2. EF-Tu is required to deliver aminoacyl-tRNAs to the A site of the ribosome
3. All aminoacyl-tRNA-Tu-GTPs can enter A site; only those with right anticodon that is complementary to the mRNA codon situated in the A site will be trapped there by ribosome

B. Studies show that the rRNA molecule of the small subunit plays a key role in recognizing a proper codon-anticodon interaction & thus ensuring the high accuracy of translation

1. Once the proper aminoacyl-tRNA-Tu-GTP is bound to the mRNA codon, the GTP is hydrolyzed

2. Then, the Tu-GDP complex is released, leaving the aa-tRNA bound in the ribosome's A site

3. Regeneration of Tu-GTP from released Tu-GDP requires another elongation factor, EF-Ts

II. Step 2: Peptide bond formation - occurs between amino acids on 2 tRNAs; after the above step, the 2 amino acids are juxtaposed in such a position that they can react with one another

A. Peptide bond now forms between these 2 amino acids

1. Amino group of aa-tRNA in A site reacts with carbonyl group of aa-tRNA in P site; displacing P site tRNA; fMet transferred to N-terminus of amino acid on tRNA in A site

2. tRNA attached to second codon in A site now has 2 amino acids (dipeptide) attached to it; tRNA in P site has none (deacylated)

B. Peptide bond formation occurs spontaneously without the input of external energy

1. Catalyzed by peptidyl transferase, a component of the large subunit of the ribosome

2. It is a ribozyme, but for many years was thought to be protein

III. Step 3: Translocation - requires GTP-bound elongation factor (EF-G - prokaryotes; eEF2 - eukaryotes) & GTP hydrolysis; converts chemical energy of GTP into mechanical energy for moving ribosome

A. After formation of the first peptide bond, the tRNA has one end still attached to its complementary codon on the mRNA & the other end attached to a dipeptide

1. The tRNA of the P site is now devoid of any linked amino acid

B. Next step, translocation, is characterized by dramatic shifts in position between the ribosome's 2 subunits

1. The ribosome & mRNA move relative to each other; in essence, the ribosome moves three nucleotides (one codon) along mRNA in 5'—>3' direction

2. Translocation is accompanied by the tRNA-dipeptide moving to P site from the A site still H bonded to second codon of the mRNA; the deacylated tRNA moves from the P site to the E site

C. Translocation is apparently driven by conformational changes in another GTP-bound elongation factor (EF-G in prokaryotes; eEF2 in eukaryotes) after hydrolysis of its bound GTP

1. Under certain conditions, translocation can occur *in vitro* in the absence of EF-G & GTP

2. Demonstrates that translocation, with all its complex conformational changes, is an intrinsic property of the ribosome, the rate of which is greatly accelerated by EF-G & GTP

IV. Step 4: Releasing the deacylated tRNA - uncharged tRNA leaves P site —> E site —> out of ribosome emptying the E site

A. For each elongation cycle, at least 2 molecules of GTP are hydrolyzed – one during aminoacyl-tRNA selection & one during translocation

1. Each elongation cycle takes ~0.05 sec, most of it probably spent sampling aa-tRNAs from cytosol

B. Once peptidyl-tRNA moves to P site by translocation, A site is once again open & new aminoacyl-tRNA moves in

1. Anticodon of new tRNA is complementary to third codon

2. Get new peptide bond (now 3 amino acids [tripeptide] on tRNA in A site), translocation & another open A site, binding of fourth tRNA, etc.

C. The elongation cycle continues until termination

V. Reading frame is fixed when ribosome binds to the initiation codon at the beginning of translation; some of most destructive mutations are frameshift mutations

A. Add or remove a single nucleotide from DNA —> get frameshift mutation so ribosome moves along in wrong reading frame, making a significant change in amino acid sequence

1. Incorrect sequence read from site of mutation through remainder of coding sequence

2. Result is entirely abnormal sequence of amino acids from point of mutation to end of protein

B. Some mRNAs have recoding signal that causes the ribosome to change its reading frame; alters reading frame by backing up (shift to –1 frame) or skipping ahead 1 base (to +1 frame)

Translating Genetic Information: Termination

I. There are no tRNAs whose anticodons are complementary to stop codons (3 of the 64 codons; UAA, UGA, UAG); protein released when one is reached

A. One minor exception has been found - incorporation of selenocysteine in ~12 mammalian proteins

1. Selenocysteine is rare amino acid containing the metal selenium (the 21st amino acid)

2. It has its own tRNA - tRNA^{Sec}, but lacks its own aa-tRNA synthetase

3. Selenocysteine is inserted into a very small number of polypeptides

B. This unique tRNA is recognized by seryl-tRNA synthetase that attaches serine to 3' end of tRNA^{Sec}; after attachment, serine is altered enzymatically to selenocysteine

C. Selenocysteine is encoded by the stop codon UGA, which is usually read as a termination signal

1. In a few cases, UGA is followed by a folded region of the mRNA that binds a special EF

2. The special EF causes ribosome to recruit a tRNA^{Sec} into the A site rather than a termination factor

D. Normally, presence of one of these codons stops elongation & releases the polypeptide associated with the last tRNA

II. Termination requires the presence of release factors; interact directly with stop codons

A. Bacteria have 3 release factors - RF1 recognizes UAA & UAG; RF2 recognizes UGA & UAA; RF3 is not codon-specific, but increases the activity of the other factors

B. Eukaryotes have 2 release factors (eRF1 & eRF3); work together & recognize all of the stop codons

C. The codon-recognizing release factors (RF1, RF2 & eRF1) enter the ribosome A site

1. A conserved tripeptide within the release factor is thought to interact directly with the stop codon in the A site, not unlike the way that a tRNA anticodon interacts with sense codon in that site

2. The ester bond linking the nascent polypeptide chain to the tRNA is then hydrolyzed & the completed polypeptide is released from the last tRNA

3. As with IFs & EFs, one of the release factors (RF3 or eRF3) is a G protein that binds GTP; the precise roles of these proteins is unclear

D. Once translation stops, both the deacylated tRNA & the release factor are then released

E. Once termination is complete, the ribosome separates from mRNA & dissociates into its large & small subunits; they will then be ready for more translation

III. mRNA surveillance: sometimes mutations arise that can create stop codons in the middle of a coding sequence within the gene; studied for decades

A. Get premature termination – such mutations are called **nonsense mutations**; they cause a variety of inherited disorders in humans (responsible for ~30% of them)

B. Cells possess an mRNA surveillance mechanism capable of detecting messages with premature termination codons

1. Usually, mRNAs containing such mutations are only translated once before they are recognized & selectively destroyed by a process called **nonsense-mediated decay** (NMD)
2. NMD protects the cell from producing nonfunctional, shortened proteins

C. How can a cell distinguish between a legitimate termination codon & a premature termination codon that has arisen through mutation or inaccurate splicing?

1. When a spliceosome removes an intron, a complex of proteins is deposited on the transcript ~22 nucleotides upstream from the newly formed exon-exon junction
2. This conglomerate of proteins is called the **exon-junction complex (EJC)** & it travels with the mRNA until it is translated
3. As mRNA undergoes its initial round of translation, the EJCs are thought to be displaced by the advancing ribosome
4. There is controversy about whether the initial round of translation occurs in the nucleus prior to mRNA export or just outside the nucleus in the cytoplasm
5. If mRNA had a premature termination codon, ribosome would stop at mutation site & then dissociate, leaving attached to mRNA any EJCs that were downstream of premature termination site
6. Thus, the presence of ≥1 EJC on translated mRNA serves as indelible mark identifying the mRNA as defective & starts a series of events leading to enzymatic destruction of abnormal message

D. NMD serves as another reminder of the opportunistic nature of biological evolution

1. Evolution has taken advantage of introns to facilitate exon shuffling & has used the process by which these gene inserts are removed to establish a mechanism of quality control
2. The mechanism of quality control ensures that only untainted mRNAs advance to a stage where they can be translated

Polyribosomes

I. Many ribosomes are usually attached to the same mRNA; can be seen in EM; this complex of mRNA & ribosomes is called a **polyribosome** or **polysome**

A. After first ribosome assembles on chain & moves a sufficient distance, another attaches & begins translation

B. Simultaneous translation of mRNA by many ribosomes greatly increases cell protein synthesis rate

C. Polysomes free in cytosol synthesize soluble proteins that stay in cell; also seen on cytosolic surface of ER where they make membrane, secretory and/or organelle proteins

II. Protein synthesis in prokaryotes begins to occur before transcription completed; not so in eukaryotes

A. mRNA is produced in 5' to 3' direction & serves as translation template in same direction - as soon as mRNA has begun to be made, the 5' end is available for attachment of ribosomes

B. The process was first visualized in EM by Oscar Miller, Jr. et al. (Oak Ridge National Laboratory)

C. In eukaryotes, mRNA must leave nucleus to be translated, so transcription (in nucleus) & translation (in cytoplasm) cannot occur simultaneously along same mRNA

The Human Perspective: Potential Clinical Applications of RNA Interference

I. Medical scientists are constantly looking for magic bullets, therapeutic compounds that fight a particular disease in a highly specific manner without toxic side effects

A. Viral infections & cancer are targets for a new type of molecular magic bullet

1. Viruses can ravage an infected cell because they synthesize mRNAs that code for viral proteins that disrupt cellular activities
2. Most cells that become cancerous contain mutations in certain genes (called **oncogenes**), causing them to produce mutant mRNAs that are translated into abnormal versions of cell proteins

B. What if a patient with such a disease could be treated with drug that destroyed/inhibited the specific mRNAs made by viral genome or mutant cancer cell while ignoring all the other cellular mRNAs?

C. RNA interference could perhaps offer such an opportunity

II. RNAi leads to the degradation of a specific mRNA by introducing to cells a double-stranded siRNA in which one of the strands is complementary to the mRNA being targeted

A. When cells are incubated with a 21 – 23 nucleotide synthetic siRNA, the molecules are taken up by the cells & incorporated into an mRNA-cleaving ribonucleoprotein complex

1. The mRNA-cleaving ribonucleoprotein complex attacks the complementary mRNA

B. Alternatively, RNAi can be induced within mammalian cells that have been genetically engineered to carry a gene with inverted repeats

1. When the gene is transcribed, the RNA product folds back on itself to form a hairpin-shaped (double-stranded) siRNA precursor that is processed into an active siRNA
2. Using these methods, siRNAs corresponding to HIV or poliovirus sequences have been produced within cultured cells
3. These siRNAs cause degradation of targeted viral RNA & protect the cells from viral infection
4. Depending on the particular targeted mRNA, cells can inhibit the entry of the virus into the cell or block its replication at various stages in the viral cycle
5. Similar approaches have been used to inhibit synthesis of disease-causing proteins in cultured cells, like proteins having expanded polyglutamine tracts or abnormal cancer-causing proteins

III. The major obstacle with these gene therapies is the difficulty in delivering siRNAs (or the vectors that encode them) to affected tissues in the body

A. The first major attempt to deliver siRNAs to a tissue within body (mouse liver) was very successful

1. Fulminant hepatitis is an inflammatory condition that leads to liver failure in persons suffering from various diseases, including infection by hepatitis viruses
2. Fulminant hepatitis can be induced in mice by treatment with a number of toxic agents
3. Studies on these mice indicate that the death of liver cells is triggered by a cell-surface receptor called Fas
4. Mice injected intravenously with an siRNA that targets & destroys *Fas* mRNA became relatively resistant to development of fulminant hepatitis

5. This happens even if the siRNA is delivered after administration of the toxic agent
6. This observation suggests that progression of the disease might be halted even after liver damage has been initiated

B. One advantage of delivery of siRNA through the bloodstream (as opposed to incorporation as part of a DNA vector) is that the effect of the added RNA is transient, lasting only a couple of weeks
   1. This is important in this particular case, because a permanent deficiency in Fas leads to serious complications for immune function

LECTURE HINTS

The Relationship Between Genes and Proteins

Discuss alcaptonuria as an example of inherited errors of metabolism. When you describe the most noticeable symptom of the disease to a class, it has a definite effect on the audience and gets their attention. You may also wish to bring up phenylketonuria and explain to them the meaning of the warning printed on the side of a can of aspartame-sweetened soda. Mention the Beadle and Tatum experiment as well. Talk about the result or, if there is time, describe the entire experiment in greater detail. By asking leading questions, get the class to work their way from the one gene - one enzyme hypothesis to the one gene - one protein and, finally, the one gene - one polypeptide hypothesis. You may also wish to mention that a significant number of genes can give rise to more than one polypeptide by using different exons via exon shuffling in the mRNA produced. Once again, since you've probably done it before, talk about the differences between normal and sickle cell hemoglobin.

Outline for the class the Central Dogma starting with replication and moving on to transcription and translation. You may want to point out the humor of Dogma in science (Central Dogma being a joke). Point out the difference between transcription and translation as well. In addition, emphasize the more recent discovery of reverse transcription. Ask the students what this means and relate reverse transcription to HIV and AIDS, if they do not immediately make the connection.

Analogy

The Monk Analogy

I stress the point that transcription involves converting the language of DNA to the language of RNA and liken this conversion to going from one dialect of a language to another. I often raise the image of the Xerox commercial in which the monk (scribe) is so happy that his job has been made easier by the purchase of a Xerox copier. Instead of transcribing his Bible, he now only needs to Xerox it. Translation, on the other hand, is the conversion of the language of nucleic acids to a completely different language, that of proteins (amino acids).

Transcription: The Basic Process

Outline the chemical reaction of transcription and the enzymes that participate in both prokaryotes and eukaryotes. Introduce the concept of promoters and describe the known structural characteristics of both prokaryotic and eukaryotic promoters. Briefly describe the

termination of transcription, both rho-dependent and -independent termination. Emphasize the importance of RNA secondary structure in the rho-independent type of termination.

Transcription and RNA Processing in Eukaryotic Cells

Mention the three RNA polymerases (I, II and III) found in eukaryotes and outline the differences in their function and products. Of primary importance in this section is the concept of RNA processing whether the product is mRNA, rRNA or tRNA. You may wish to discuss, in some detail, processing of rRNA and outline experiments like those in the text. I generally mention it briefly and spend more time on the processing of mRNAs. Some may wish to reserve detailed description of RNA processing for an advanced cell or molecular biology course.

Transcription and RNA Processing of mRNAs

Describe the high turnover rate in nuclear RNAs and the discovery of heterogeneous nuclear RNAs. You may wish to go into some of the evidence. Then move the discussion to a description of mRNA structure. Describe the structures and functions of the 5' methylated guanosine cap and 3' poly(A) tail.

Perhaps the most surprising and striking aspect of eukaryotic mRNA structure is the concept of the split gene and the consequent necessity for processing of hnRNA in part to remove the noncoding portions of the full transcript while splicing together the coding portions. Student understanding is aided by a discussion of the adenovirus, β-globin and/or especially the R-loop experiments, which clearly illustrate the genesis of this concept. I like to handle these experiments by proposing experiments and asking students what they would expect to see in the results if one thing or another were true. This is an effective way to enhance their understanding of the experiments. Define exons and introns and use the terminology in order to ensure that the class gets the concept. It is useful to stress the splicing reaction that joins together exons after the removal of the intron between them. Also, mention the spliceosome and the components of which it is composed, like snRNPs. Point out the case of the autoimmune disease lupus erythematosus that has been found to be related to an autoimmune reaction against these snRNPs. Describe the processing of mRNA and the self-splicing ability of Group I and II introns. Reiterate, at this point, the concept of ribozymes that was discussed briefly in the enzyme chapter and point out its relevance to the discussion of splicing. Acquaint students with the present thinking about the possible functions of the introns (listed in the text and above in the outline).

Encoding Genetic Information

Illustration

The Genetic Code: Gamow's Common Sense Proposal

First, define the problem - to represent unambiguously 20 amino acids by words in the language of nucleic acids (a sequence of nucleic acid bases). Explain George Gamow's idea that one should think of each nucleotide as a letter in the alphabet and think of the genetic code as being composed of words made from the letters of this alphabet. Gamow also suggested that Nature would only make the code complicated enough to code unambiguously for the 20 amino acids. Given that there are 4 bases

in RNA and DNA, Gamow pointed out that if the code consisted of one-letter words, only 3 amino acids could be represented unambiguously; the other 17 amino acids would be represented by the remaining code word. A code made up of two-letter words would do better, coding unambiguously for 15 amino acids (4 possibilities for the first letter times 4 possibilities for the second letter and 16 overall) but leaving five of the amino acids out in the cold coded for by the remaining code word. To illustrate the principle, ask students to list each possible two-letter word. If each word in the code consisted of three letters, there would be 64 possible words (4 x 4 x 4). This would be more than enough to code unambiguously for 20 amino acids. It would even allow synonyms so that some amino acids could be represented by more than one word and some words in this language could also be used as stop signals. Despite being right about the triplet codon, Gamow was wrong in his proposal of an overlapping code; the code has since then been shown to be nonoverlapping.

Spend some time explaining the reasoning behind the idea of the triplet codon of the genetic code. Take the students through Gamow's reasoning. Ask them leading questions; see if they understand the calculations for the possible number of codons with different-sized codons.

Also, point out the nonrandom nature of codon assignments and the functional advantage it confers. As always, examples help to illustrate the principle.

Decoding the Codons: The Role of the Transfer RNAs

Describe the role of tRNAs as the decoder of the genetic code. At one end, a tRNA possesses an anticodon, which is complementary to a codon in the mRNA. On the other end, it is attached to the amino acid represented by the codon with which its anticodon pairs.

Analogy

The Plug Adapter Analogy

I often liken tRNAs to three-prong plug adapters. These adapters became necessary when appliances began appearing with an extra prong on their plug for grounding. Unfortunately, most older houses only had outlets that could accept only two pronged plugs. This became a significant problem. The problem was solved by developing an adapter that could accept the third prong on one side while having only two prongs projecting from the other end. The adapter fit quite well into the wall sockets of older houses and allowed their occupants to make use of modern appliances without refitting all of the electrical outlets in the house. Transfer RNAs are similar. They accept the appropriate amino acid on the 3' end while the anticodon loop is able to match up with the correct codon and facilitate the placement of the right amino acid in the right place in a growing polypeptide chain. An amino acid alone would be unable to accomplish this feat of recognition. In this sense, the tRNA functions in a fashion similar to the three-prong adapter.

After describing tRNA function, speak briefly about aminoacyl-tRNA synthetases and their role in charging the tRNAs for protein synthesis. It is interesting to note that the attachment of AMP to the amino acid in the first step of this reaction supplies the energy necessary to drive protein assembly (peptide bond formation). More energy is supplied later by hydrolysis of another product of the first step in the overall reaction, pyrophosphate.

I sometimes draw a parallel between these enzymes and the glycosyltransferases so important in the construction of the oligosaccharide chains on secretory and membrane proteins, which recognize specific donor sugars and add the sugars to specific sugars onto the ends of the growing oligosaccharide chains on these glycoproteins.

Translating Genetic Information

I spend some time discussing the process of translation. I often do this in a laboratory session where more time can be spent. We recently added a credit to the Cell Biology course and are now able to cover protein synthesis in more detail. This has helped the professor in the Genetics course that follows Cell Biology quite a bit. It is clear that understanding this and some other processes is furthered by repetition. When discussing protein synthesis, I tend to leave out the soluble initiation and elongation factors. I mention their existence and tell the students that I will not emphasize their role. These factors and their roles unnecessarily clutter the students' minds and tend to hinder their basic understanding of the process. After they have acquired a basic comprehension of the process, the roles of these factors can be introduced and outlined in later courses (Genetics and Molecular Biology).

Two other aspects of protein synthesis should also be mentioned. First, emphasize that many ribosomes can travel down a single mRNA at the same time, forming a polyribosome (polysome). Ask the students to speculate on the practical reason for this. They will usually come up with enhanced efficiency of protein synthesis and an increase in the rate of the process. Second, point out a striking difference between prokaryotes and eukaryotes. Transcription and translation can occur upon the same piece of mRNA in prokaryotes. This is obviously not possible in eukaryotes. Ask the students why that is. They usually get the right answer, even if some prodding is necessary.

Illustration

The Translation Simulation

During a laboratory session, I have employed a translation simulation to demonstrate how the process takes place. There are a number of such simulations published or for sale (Carolina Biological has one for sale, I believe). I use one I developed along with two colleagues, Drs. Carol Hepfer and Susan DiBartolomeis. As in other simulations, students take the parts of components of the process and build a polypeptide from a mRNA they are given at the beginning of the simulation. The special feature of this simulation is that it can be performed without giving students much of an introduction to the process. Each student is given a detailed job description for their component in the process (A site, P site, release factor, aminoacyl-tRNA synthetase, etc.). The job descriptions are only given to the students who will have that particular job. Their job is explained to no one else. After the job descriptions have been distributed, we turn the students loose on the simulation. After some initial confusion and bumping into one another, they begin to understand their assignments. When they do the simulation a second time, everything goes much more smoothly. The point made by this particular simulation is that each component in the process knows what its job is but knows nothing about the responsibilities of the other components. Yet, amazingly, the system works. The simulation can also be used in the traditional way with an appropriate introduction to the process and the tasks of each individual component given to the whole class. It has worked well either way and is often mentioned by the students as a favorite lab exercise. They also do well when tested on the material in their final exam.