**Chapter 14**

**Overview: The Flow of Genetic Information**

The information content of DNA is in the form of specific sequences of nucleotides along the DNA strands.
The DNA inherited by an organism leads to specific traits by dictating the synthesis of proteins.
Gene expression, the process by which DNA directs protein synthesis, includes two stages called transcription and translation.
Proteins are the links between genotype and phenotype.
For example, Mendel's dwarf pea plants lack a functioning copy of the gene that specifies the synthesis of a key protein, gibberellin.
Gibberellins stimulate the normal elongation of stems.

**Concept 14.1 Genes specify proteins via transcription and translation**

**The study of metabolic defects provided evidence that genes specify proteins.**

In 1909, Archibald Gerrod was the first to suggest that genes dictate phenotype through enzymes that catalyze specific chemical reactions in the cell.
He suggested that the symptoms of an inherited disease reflect a person's inability to synthesize a particular enzyme.
He referred to such diseases as "inborn errors of metabolism."
Gerrod speculated that alkaptonuria, a hereditary disease, was caused by the absence of an enzyme that breaks down a specific substrate, alkapton.
Research conducted several decades later supported Gerrod's hypothesis.
Progress in linking genes and enzymes rested on the growing understanding that cells synthesize and degrade most organic molecules in a series of steps, a metabolic pathway.
In the 1930s, George Beadle and Boris Ephrussi speculated that each mutation affecting eye color in Drosophila blocks pigment synthesis at a specific step by preventing production of the enzyme that catalyzes that step.
However, neither the chemical reactions nor the enzymes that catalyze them were known at the time.
Beadle and Edward Tatum were finally able to establish the link between genes and enzymes in their exploration of the metabolism of a bread mold, Neurospora crassa.
They bombarded Neurospora with X-rays and screened the survivors for mutants that differed in their nutritional needs.
Wild-type Neurospora can grow on a minimal medium of agar, inorganic salts, glucose, and the vitamin biotin.
Beadle and Tatum identified mutants that could not survive on minimal medium, because they were unable to synthesize certain essential molecules from the minimal ingredients.
However, most of these nutritional mutants can survive on a complete growth medium that includes all 20 amino acids and a few other nutrients.
One type of mutant required only the addition of arginine to the minimal growth medium.
Beadle and Tatum concluded that this mutant was defective somewhere in the biochemical pathway that normally synthesizes arginine.
They identified three classes of arginine-deficient mutants, each apparently lacking a key enzyme at a different step in the synthesis of arginine.
They demonstrated this by growing these mutant strains in media that provided different intermediate molecules.
Their results provided strong evidence for the one gene–one enzyme hypothesis.
Later research refined the one gene–one enzyme hypothesis.
First, not all proteins are enzymes.
Keratin, the structural protein of hair, and insulin, a hormone, both are proteins and gene products.
This tweaked the hypothesis to one gene–one protein.
Later research demonstrated that many proteins are composed of several polypeptides, each of which has its own gene.
Therefore, Beadle and Tatum's idea has been restated as the **one gene–one polypeptide hypothesis.**

Some genes code for RNA molecules that play important roles in cells although they are never translated into protein.

**Transcription and translation are the two main processes linking gene to protein.**

Genes provide the instructions for making specific proteins.

The bridge between DNA and protein synthesis is the nucleic acid RNA.

RNA is chemically similar to DNA, except that it contains ribose as its sugar and substitutes the nitrogenous base uracil for thymine.

An RNA molecule almost always consists of a single strand.

In DNA or RNA, the four nucleotide monomers act like the letters of the alphabet to communicate information.

The specific sequence of hundreds or thousands of nucleotides in each gene carries the information for the primary structure of proteins, the linear order of the 20 possible amino acids.

To get from DNA, written in one chemical language, to protein, written in another, requires two major stages: transcription and translation.

During **transcription,** a DNA strand provides a template for the synthesis of a complementary RNA strand.

Just as a DNA strand provides a template for the synthesis of each new complementary strand during DNA replication, it provides a template for assembling a sequence of RNA nucleotides.

Transcription of many genes produces a **messenger RNA (mRNA)** molecule.

During **translation,** there is a change of language.

The site of translation is the **ribosome,** complex particles that facilitate the orderly assembly of amino acids into polypeptide chains.

Why can't proteins be translated directly from DNA?

The use of an RNA intermediate provides protection for DNA and its genetic information.

Using an RNA intermediate allows more copies of a protein to be made simultaneously, since many RNA transcripts can be made from one gene.

Also, each gene transcript can be translated repeatedly.

The basic mechanics of transcription and translation are similar in eukaryotes and prokaryotes.

Because bacteria lack nuclei, their DNA is not segregated from ribosomes and other protein-synthesizing equipment.

This allows the coupling of transcription and translation.

Ribosomes attach to the leading end of an mRNA molecule while transcription is still in progress.

In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs at ribosomes in the cytoplasm.

The transcription of a protein-coding eukaryotic gene results in pre-mRNA.

The initial RNA transcript of any gene is called a **primary transcript.**

**RNA processing** yields the finished mRNA.

To summarize, genes program protein synthesis via genetic messages in the form of messenger RNA.

The molecular chain of command in a cell is DNA RNA protein.

**In the genetic code, nucleotide triplets specify amino acids.**

If the genetic code consisted of a single nucleotide or even pairs of nucleotides per amino acid, there would not be enough combinations (4 and 16, respectively) to code for all 20 amino acids.

Triplets of nucleotide bases are the smallest units of uniform length that can code for all the amino acids.

With a **triplet code,** three consecutive bases specify an amino acid, creating $4^3$ (64) possible code words.

The genetic instructions for a polypeptide chain are written in DNA as a series of nonoverlapping three-nucleotide words.

During transcription, one DNA strand, the **template strand,** provides a template for ordering the sequence of nucleotides in an RNA transcript.

A given DNA strand can be the template strand for some genes along a DNA molecule, while for other genes in other regions, the complementary strand may function as the template.

The complementary RNA molecule is synthesized according to base-pairing rules, except that uracil is the complementary base to adenine.

Like a new strand of DNA, the RNA molecule is synthesized in an antiparallel direction to the template strand of DNA.

The mRNA base triplets are called **codons,** and they are written in the 5' 3' direction.

During translation, the sequence of codons along an mRNA molecule is translated into a sequence of amino acids making up the polypeptide chain.
During translation, the codons are read in the 5' 3' direction along the mRNA.
Each codon specifies which one of the 20 amino acids will be incorporated at the corresponding position along a polypeptide.
Because codons are base triplets, the number of nucleotides making up a genetic message must be three times the number of amino acids making up the protein product.
It takes at least 300 nucleotides to code for a polypeptide that is 100 amino acids long.
The task of matching each codon to its amino acid counterpart began in the early 1960s.
Marshall Nirenberg determined the first match: UUU coded for the amino acid phenylalanine.
He created an artificial mRNA molecule entirely of uracil and added it to a test tube mixture of amino acids, ribosomes, and other components for protein synthesis.
This "poly-U" translated into a polypeptide containing a single amino acid, phenylalanine, in a long chain.
AAA, GGG, and CCC were solved in the same way.
Other more elaborate techniques were required to decode mixed triplets such as AUA and CGA.
By the mid-1960s the entire code was deciphered.
Sixty-one of 64 triplets code for amino acids.
The codon AUG not only codes for the amino acid methionine, but also indicates the "start" of translation.
Three codons do not indicate amino acids but are "stop" signals marking the termination of translation.
There is redundancy in the genetic code but no ambiguity.
Several codons may specify the same amino acid, but no codon specifies more than one amino acid.
The redundancy in the code is not random. In many cases, codons that are synonyms for a particular amino acid differ only in the third base of the triplet.
To extract the message from the genetic code requires specifying the correct starting point.
This establishes the **reading frame;** subsequent codons are read in groups of three nucleotides.
The cell's protein-synthesizing machinery reads the message as a series of nonoverlapping three-letter words.
In summary, genetic information is encoded as a sequence of nonoverlapping base triplets, or codons, each of which is translated into a specific amino acid during protein synthesis.
**The genetic code must have evolved very early in the history of life.**
The genetic code is nearly universal, shared by organisms from the simplest bacteria to the most complex plants and animals.
In laboratory experiments, genes can be transcribed and translated after they are transplanted from one species to another.
This has permitted bacteria to be programmed to synthesize certain human proteins after insertion of the appropriate human genes.
Such applications are exciting developments in biotechnology.
Exceptions to the universality of the genetic code exist in certain unicellular eukaryotes and in the organelle genes of some species.
Some prokaryotes can translate stop codons into one of two amino acids not found in most organisms.
The evolutionary significance of the near universality of the genetic code is clear.
A language shared by all living things arose very early in the history of life—early enough to be present in the common ancestors of all modern organisms.
A shared genetic vocabulary is a reminder of the kinship that bonds all life on Earth.

**Concept 14.2 Transcription is the DNA-directed synthesis of RNA: a closer look**

Messenger RNA, the carrier of information from DNA to the cell's protein-synthesizing machinery, is transcribed from the template strand of a gene. To start off, **RNA polymerase** separates the DNA strands at the appropriate point and bonds the RNA nucleotides as they base-pair along the DNA template.

Like DNA polymerases, RNA polymerases can only assemble a polynucleotide in its 5' 3' direction.

Unlike DNA polymerases, RNA polymerases are able to start a chain from scratch; they don't need a primer.

Specific sequences of nucleotides along the DNA mark where gene transcription begins and ends.

RNA polymerase attaches and initiates transcription at the **promoter.**

In prokaryotes, the sequence that signals the end of transcription is called the **terminator.**

Molecular biologists refer to the direction of transcription as “downstream” and the other direction as “upstream.”

The stretch of DNA that is transcribed into an RNA molecule is called a **transcription unit.**

Bacteria have a single type of RNA polymerase that synthesizes all RNA molecules.

In contrast, eukaryotes have three RNA polymerases (I, II, and III) in their nuclei.

RNA polymerase II is used for mRNA synthesis.

Transcription can be separated into three stages: initiation, elongation, and termination of the RNA chain.

The presence of a promoter sequence determines which strand of the DNA helix is the template.

Within the promoter is the starting point for the transcription of a gene.
The promoter also includes a binding site for RNA polymerase several dozen nucleotides “upstream” of the start point.

In prokaryotes, RNA polymerase can recognize and bind directly to the promoter region.

In eukaryotes, proteins called **transcription factors** mediate the binding of RNA polymerase and the initiation of transcription.

Only after certain transcription factors are attached to the promoter does RNA polymerase II bind to it.

The completed assembly of transcription factors and RNA polymerase II bound to a promoter is called a **transcription initiation complex.**

A crucial promoter DNA sequence is called a **TATA box.**

RNA polymerase then starts transcription.

As RNA polymerase moves along the DNA, it untwists the double helix, 10 to 20 bases at time.

The enzyme adds nucleotides to the 3’ end of the growing strand.

Behind the point of RNA synthesis, the double helix re-forms and the RNA molecule peels away.

Transcription progresses at a rate of 60 nucleotides per second in eukaryotes.

A single gene can be transcribed simultaneously by several RNA polymerases at a time.

A growing strand of RNA trails off from each polymerase.

The length of each new strand reflects how far along the template the enzyme has traveled from the start point.

The congregation of many polymerase molecules simultaneously transcribing a single gene increases the amount of mRNA transcribed from it.

This helps the cell make the encoded protein in large amounts.

Transcription proceeds until after the RNA polymerase transcribes a terminator sequence in the DNA.

In prokaryotes, RNA polymerase stops transcription right at the end of the terminator.

Both the RNA and DNA are then released.

In eukaryotes, the pre-mRNA is cleaved from the growing RNA chain while RNA polymerase

II continues to transcribe the DNA.

Specifically, the polymerase transcribes a DNA sequence called the polyadenylation signal sequence that codes for a polyadenylation sequence (AAUAAA) in the pre-mRNA.

At a point about 10 to 35 nucleotides past this sequence, the pre-mRNA is cut from the enzyme.

The polymerase continues transcribing for hundreds of nucleotides.

Transcription is terminated when the polymerase eventually falls off the DNA.

**Concept 14.3 Eukaryotic cells modify RNA after transcription**
Enzymes in the eukaryotic nucleus modify pre-mRNA before the genetic messages are dispatched to the cytoplasm.
During RNA processing, both ends of the primary transcript are usually altered.
Certain interior parts of the molecule are cut out and the remaining parts spliced together.
At the 5' end of the pre-mRNA molecule, a modified form of guanine is added, the **5' cap.**
At the 3' end, an enzyme adds 50 to 250 adenine nucleotides, the **poly-A tail.**
These modifications share several important functions.
They seem to facilitate the export of mRNA from the nucleus.
They help protect mRNA from hydrolytic enzymes.
They help the ribosomes attach to the 5' end of the mRNA.
The most remarkable stage of RNA processing occurs during the removal of a large portion of the RNA molecule in a cut-and-paste job of **RNA splicing.**
Most eukaryotic genes and their RNA transcripts have long noncoding stretches of nucleotides.
Noncoding segments of nucleotides called intervening regions, or **introns,** lie between coding regions.
The final mRNA transcript includes coding regions, **exons,** which are translated into amino acid sequences, plus the leader and trailer sequences.
RNA splicing removes introns and joins exons to create an mRNA molecule with a continuous coding sequence.
This splicing is accomplished by a **spliceosome.**
Spliceosomes consist of a variety of proteins and several small nuclear ribonucleoproteins (snRNPs) that recognize the splice sites.
snRNPs are located in the cell nucleus and are composed of RNA and protein molecules.
Each snRNP has several protein molecules and a small nuclear RNA molecule (snRNA).
Each snRNA is about 150 nucleotides long.
The spliceosome interacts with certain sites along an intron, releasing the introns and joining together the two exons that flanked the introns.
snRNAs appear to play a major role in catalytic processes, as well as spliceosome assembly and splice site recognition.

The idea of a catalytic role for snRNA arose from the discovery of **ribozymes,** RNA molecules that function as enzymes.
In some organisms, splicing occurs without proteins or additional RNA molecules.
The intron RNA functions as a ribozyme and catalyzes its own excision.
For example, in the protozoan Tetrahymena, self-splicing occurs in the production of ribosomal RNA (rRNA), a component of the organism's ribosomes.
The pre-rRNA actually removes its own introns.
The discovery of ribozymes rendered obsolete the statement, "All biological catalysts are proteins."
The fact that RNA is single-stranded plays an important role in allowing certain RNA molecules to function as ribozymes.
A region of the RNA molecule may base-pair with a complementary region elsewhere in the same molecule, thus giving the RNA a specific 3-D structure that is key to its ability to catalyze reactions.
Introns and RNA splicing appear to have several functions.
Some introns play a regulatory role in the cell. These introns contain sequences that control gene activity in some way.
Splicing itself may regulate the passage of mRNA from the nucleus to the cytoplasm.
One clear benefit of split genes is to enable one gene to encode for more than one polypeptide.
**Alternative RNA splicing** gives rise to two or more different polypeptides, depending on which segments are treated as exons.
Sex differences in fruit flies may be due to differences in splicing RNA transcribed from certain genes.
Early results of the Human Genome Project indicate that this phenomenon may be common in humans, and may explain why we have a relatively small number of genes.
Proteins often have a modular architecture with discrete structural and functional regions called **domains.**
The presence of introns in a gene may facilitate the evolution of new and potentially useful proteins as a result of a process known as exon shuffling.
In many cases, different exons code for different domains of a protein.
The presence of introns increases the probability of potentially beneficial crossing over between genes.
Introns increase the opportunity for recombination between two alleles of a gene.
This raises the probability that a crossover will switch one version of an exon for another version found on the homologous chromosome.
There may also be occasional mixing and matching of exons between completely different genes.
Either way, exon shuffling can lead to new proteins through novel combinations of functions.

**Concept 14.4 Translation is the RNA-directed synthesis of a polypeptide: a closer look**

In the process of translation, a cell interprets a series of codons along an mRNA molecule and builds a polypeptide.

The interpreter is **transfer RNA (tRNA),** which transfers amino acids from the cytoplasmic pool to a ribosome.

A cell has all 20 amino acids available in its cytoplasm, either by synthesizing them from scratch or by taking them up from the surrounding solution.

The ribosome adds each amino acid carried by tRNA to the growing end of the polypeptide chain.

During translation, each type of tRNA links an mRNA codon with the appropriate amino acid.

Each tRNA arriving at the ribosome carries a specific amino acid at one end and has a specific nucleotide triplet, an **anticodon,** at the other.

The anticodon base-pairs with a complementary codon on mRNA.

If the codon on mRNA is UUU, a tRNA with an AAA anticodon and carrying phenylalanine will bind to it.

Codon by codon, tRNAs deposit amino acids in the prescribed order, and the ribosome joins them into a polypeptide chain.

The tRNA molecule is a translator, because it can read a nucleic acid word (the mRNA codon) and translate it to a protein word (the amino acid).

Like other types of RNA, tRNA molecules are transcribed from DNA templates in the nucleus.

Once it reaches the cytoplasm, each tRNA is used repeatedly, picking up its designated amino acid in the cytosol, depositing the amino acid at the ribosome, and returning to the cytosol to pick up another copy of that amino acid.

A tRNA molecule consists of a strand of about 80 nucleotides that folds back on itself to form a three-dimensional structure.

It includes a loop containing the anticodon and an attachment site at the 3' end for an amino acid.

If each anticodon had to be a perfect match to each codon, we would expect to find 61 types of tRNA, but the actual number is about 45.

The anticodons of some tRNAs recognize more than one codon.

This is possible because the rules for base pairing between the third base of the codon and anticodon are relaxed (called **wobble**).

At the wobble position, U on the anticodon can bind with A or G in the third position of a codon.

Wobble explains why the synonymous codons for a given amino acid can differ in their third base, but not usually in their other bases.

Each amino acid is joined to the correct tRNA by **aminoacyl-tRNA synthetase.**
The 20 different synthetases match the 20 different amino acids.

Each has active sites for only a specific tRNA-and-amino-acid combination.

The synthetase catalyzes a covalent bond between them in a process driven by ATP hydrolysis.

The result is an aminoacyl-tRNA or activated amino acid.

Ribosomes facilitate the specific coupling of the tRNA anticodons with mRNA codons during protein synthesis.

Each ribosome is made up of a large and a small subunit.

The subunits are composed of proteins and **ribosomal RNA (rRNA),** the most abundant RNA in the cell.

In eukaryotes, the subunits are made in the nucleolus.

After rRNA genes are transcribed to rRNA in the nucleus, the rRNA and proteins are assembled to form the subunits with proteins from the cytoplasm.

The subunits exit the nucleus via nuclear pores.

The large and small subunits join to form a functional ribosome only when they attach to an mRNA molecule.

While very similar in structure and function, prokaryotic and eukaryotic ribosomes have enough differences that certain antibiotic drugs (like tetracycline) can paralyze prokaryotic ribosomes without inhibiting eukaryotic ribosomes.

Each ribosome has a binding site for mRNA and three binding sites for tRNA molecules.

The **P site** holds the tRNA carrying the growing polypeptide chain.

The **A site** carries the tRNA with the next amino acid to be added to the chain.

Discharged tRNAs leave the ribosome at the **E** (exit) **site.**

The ribosome holds the tRNA and mRNA in close proximity and positions the new amino acid for addition to the carboxyl end of the growing polypeptide.

It then catalyzes the formation of the peptide bond.
As the polypeptide becomes longer, it passes through an exit tunnel in the ribosome's large unit and is released to the cytosol.
Recent advances in our understanding of the structure of the ribosome strongly support the hypothesis that rRNA, not protein, carries out the ribosome's functions.
RNA is the main constituent at the interphase between the two subunits and of the A and P sites.
It is the catalyst for peptide bond formation.
A ribosome can be regarded as one colossal ribozyme.
Translation can be divided into three stages: initiation, elongation, and termination.
All three phases require protein "factors" that aid in the translation process.
Both initiation and chain elongation require energy provided by the hydrolysis of GTP.
**Initiation** brings together mRNA, a tRNA with the first amino acid, and the two ribosomal subunits.
First, a small ribosomal subunit binds with mRNA and a special initiator tRNA, which carries methionine and attaches to the start codon.
The small subunit then moves downstream along the mRNA until it reaches the start codon, AUG, which signals the start of translation.
This establishes the reading frame for the mRNA.
The initiator tRNA, already associated with the complex, then hydrogen-bonds with the start codon.
Proteins called initiation factors bring in the large subunit so that the initiator tRNA occupies the P site.
**Elongation** involves the participation of several protein elongation factors, and consists of a series of three-step cycles as each amino acid is added to the proceeding one.
During **codon recognition,** an elongation factor assists hydrogen bonding between the mRNA codon under the A site with the corresponding anticodon of tRNA carrying the appropriate amino acid.
This step requires the hydrolysis of two GTP.
During **peptide bond formation,** an rRNA molecule catalyzes the formation of a peptide bond between the polypeptide in the P site with the new amino acid in the A site.
This step separates the tRNA at the P site from the growing polypeptide chain and transfers the chain, now one amino acid longer, to the tRNA at the A site.
During **translocation,** the ribosome moves the tRNA with the attached polypeptide from the A site to the P site.
Because the anticodon remains bonded to the mRNA codon, the mRNA moves along with it.
The next codon is now available at the A site.
The tRNA that had been in the P site is moved to the E site and then leaves the ribosome.
Translocation is fueled by the hydrolysis of GTP.
Effectively, translocation ensures that the mRNA is "read" 5' 3' codon by codon.

• The three steps of elongation continue to add amino acids codon by codon until the polypeptide chain is completed.

**Termination** occurs when one of the three stop codons reaches the A site.

A release factor binds to the stop codon and hydrolyzes the bond between the polypeptide and its tRNA in the P site.

This frees the polypeptide, and the translation complex disassembles.

Typically a single mRNA is used to make many copies of a polypeptide simultaneously.

Multiple ribosomes, **polyribosomes,** may trail along the same mRNA.

Polyribosomes can be found in prokaryotic and eukaryotic cells.

A ribosome requires less than a minute to translate an average-sized mRNA into a polypeptide.

During and after synthesis, a polypeptide coils and folds to its three-dimensional shape spontaneously.

The primary structure, the order of amino acids, determines the secondary and tertiary structure.

Chaperone proteins may aid correct folding.

In addition, proteins may require posttranslational modifications before doing their particular job.

This may require additions such as sugars, lipids, or phosphate groups to amino acids.

Enzymes may remove some amino acids or cleave whole polypeptide chains.

Two or more polypeptides may join to form a protein.

**Signal peptides target some eukaryotic polypeptides to specific destinations in the cell.**

Two populations of ribosomes, free and bound, are active participants in protein synthesis.

Free ribosomes are suspended in the cytosol and synthesize proteins that reside in the cytosol.

Bound ribosomes are attached to the cytosolic side of the endoplasmic reticulum.

They synthesize proteins of the endomembrane system as well as proteins secreted from the cell.

While bound and free ribosomes are identical in structure, their location depends on the type of protein that they are synthesizing.

Translation in all ribosomes begins in the cytosol, but a polypeptide destined for the endomembrane system or for export has a specific **signal peptide** region at or near the leading end.

This consists of a sequence of about 20 amino acids.

A **signal recognition particle (SRP)** binds to the signal peptide and attaches it and its ribosome to a receptor protein in the ER membrane.

The SRP consists of a protein-RNA complex.

After binding, the SRP leaves and protein synthesis resumes with the growing polypeptide snaking across the membrane into the cisternal space via a protein pore.

An enzyme usually cleaves the signal polypeptide.

Secretory proteins are released entirely into the cisternal space, but membrane proteins remain partially embedded in the ER membrane.

Other kinds of signal peptides are used to target polypeptides to mitochondria, chloroplasts, the nucleus, and other organelles that are not part of the endomembrane system.

In these cases, translation is completed in the cytosol before the polypeptide is imported into the organelle.

While the mechanisms of translocation vary, each of these polypeptides has a “ZIP code” that ensures its delivery to the correct cellular location.

Prokaryotes also employ signal sequences to target proteins for secretion.

**Concept 14.5 RNA plays multiple roles in the cell: a review**

The cellular machinery of protein synthesis and ER targeting is dominated by various kinds of RNA.

In addition to mRNA, these include tRNA; rRNA; and in eukaryotes, snRNA and SRP RNA.

A type of RNA called small nucleolar RNA (snoRNA) aids in processing pre-rRNA transcripts in the nucleolus, a process necessary for ribosome formation.

Recent research has also revealed the presence of small, single-stranded and double-stranded RNA molecules that play important roles in regulating which genes get expressed.

These types of RNA include small interfering RNA (siRNA) and microRNA (miRNA).

The diverse functions of RNA are based, in part, on its ability to form hydrogen bonds with other nucleic acid molecules (DNA or RNA).

It can also assume a specific three-dimensional shape by forming hydrogen bonds between bases in different parts of its polynucleotide chain.

DNA may be the genetic material of all living cells today, but RNA is much more versatile.
The diverse functions of RNA range from structural to informational to catalytic.

**Comparing gene expression in prokaryotes and eukaryotes reveals key differences**
Although prokaryotes and eukaryotes carry out transcription and translation in very similar ways, they do have differences in cellular machinery and in details of the processes.
Eukaryotic RNA polymerases differ from those of prokaryotes and require transcription factors.
They differ in how transcription is terminated.
Their ribosomes also are different.
One major difference is that prokaryotes can transcribe and translate the same gene simultaneously.
The new protein quickly diffuses to its operating site.
In eukaryotes, the nuclear envelope segregates transcription from translation.
In addition, extensive RNA processing is carried out between these processes.
This provides additional steps whose regulation helps coordinate the elaborate activities of a eukaryotic cell.
Eukaryotic cells also have complicated mechanisms for targeting proteins to the appropriate organelle.

**Point mutations can affect protein structure and function**
**Mutations** are changes in the genetic material of a cell (or virus).
These include large-scale mutations in which long segments of DNA are affected (for example, translocations, duplications, and inversions).
A chemical change in just one base pair of a gene causes a **point mutation.**
If these occur in gametes or cells producing gametes, they may be transmitted to future generations.
For example, sickle-cell disease is caused by a mutation of a single base pair in the gene that codes for one of the polypeptides of hemoglobin.
A change in a single nucleotide from T to A in the DNA template leads to an abnormal protein.
A point mutation that results in the replacement of a pair of complementary nucleotides with another nucleotide pair is called a **base-pair substitution.**
Some base-pair substitutions have little or no impact on protein function.
In silent mutations, altered nucleotides still code for the same amino acids because of redundancy in the genetic code.
Other changes lead to switches from one amino acid to another with similar properties.
Still other mutations may occur in a region where the exact amino acid sequence is not essential for function.
Other base-pair substitutions cause a readily detectable change in a protein.
These are usually detrimental but can occasionally lead to an improved protein or one with novel capabilities.
Changes in amino acids at crucial sites, especially active sites, are likely to impact function.
**Missense mutations** are those that still code for an amino acid but a different one.
**Nonsense mutations** change an amino acid codon into a stop codon, nearly always leading to a nonfunctional protein.
**Insertions** and **deletions** are additions or losses of nucleotide pairs in a gene.
These have a disastrous effect on the resulting protein more often than substitutions do.
Unless insertion or deletion mutations occur in multiples of three, they cause a **frameshift mutation.**
All the nucleotides downstream of the deletion or insertion will be improperly grouped into codons.
The result will be extensive missense, ending sooner or later in nonsense—premature termination.
Mutations can occur in a number of ways.
Errors can occur during DNA replication, DNA repair, or DNA recombination.
These can lead to base-pair substitutions, insertions, or deletions, as well as mutations affecting longer stretches of DNA.
These are called spontaneous mutations.
Rough estimates suggest that about 1 nucleotide in every $10^{10}$ is altered and inherited by daughter cells.
**Mutagens** are chemical or physical agents that interact with DNA to cause mutations.
Physical agents include high-energy radiation like X-rays and ultraviolet light.
Chemical mutagens fall into several categories.

Some chemicals are base analogues that may be substituted into DNA, but they pair incorrectly during DNA replication.
Other mutagens interfere with DNA replication by inserting into DNA and distorting the double helix.
Still others cause chemical changes in bases that change their pairing properties.
Researchers have developed various methods to test the mutagenic activity of different chemicals.
These tests are often used as a preliminary screen of chemicals to identify those that may cause cancer.
This makes sense because most carcinogens are mutagenic and most mutagens are carcinogenic.

**What is a gene? We revisit the question.**

The Mendelian concept of a gene views it as a discrete unit of inheritance that affects phenotype.
Morgan and his colleagues assigned genes to specific loci on chromosomes.
We can also view a gene as a specific nucleotide sequence along a region of a DNA molecule.
We can define a gene functionally as a DNA sequence that codes for a specific polypeptide chain.
All these definitions are useful in certain contexts.
Even the one gene–one polypeptide definition must be refined and applied selectively.
Most eukaryotic genes contain large introns that have no corresponding segments in polypeptides.
Promoters and other regulatory regions of DNA are not transcribed either, but they must be present for transcription to occur.
Our molecular definition must also include the various types of RNA that are not translated into polypeptides, such as rRNA, tRNA, and other RNAs.
This is our definition of a gene: A gene is a region of DNA whose final product is either a polypeptide or an RNA molecule.