

## How Genes Work

### Central Dogma of Biology and Operons

#### Transcription

- The information contained in DNA is stored in blocks called genes
  - the genes code for proteins
  - the proteins determine what a cell will be like
- The DNA stores this information safely in the nucleus where it never leaves
  - instructions are copied from the DNA into messages comprised of RNA
  - these messages are sent out into the cell to direct the assembly of proteins
- The path of information is often referred to as the **central dogma**

DNA → RNA → protein

- The use of information in DNA to direct the production of particular proteins is called **gene expression**, which takes place in two stages
  - **transcription** is the process when a messenger RNA (mRNA) is made from a gene within the DNA
  - **translation** is the process of using the mRNA to direct the production of a protein
- RNA is the same as DNA except that the sugars in RNA have an extra oxygen and T is replaced by another pyrimidine called **uracil** (U)
- The cell uses three kinds of RNA
  - messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA)
- A protein called **RNA polymerase** produces the mRNA copy of DNA during transcription
  - it first binds to one strand of the DNA at a site called the **promoter** and then moves down the DNA molecule and assembles a complementary copy of RNA
  - transcription ends when the RNA polymerase reaches a certain nucleotide sequence that signals it stop

#### Transcription

#### Translation

- To correctly read a gene, a cell must translate the information encoded in the DNA into the language of proteins
  - translation follows rules set out by the **genetic code**
  - the mRNA is “read” in three-nucleotide units called **codons**
    - each codon corresponds to a particular amino acid

- The **genetic code dictionary** was determined from trial-and-error experiments to work out which codons matched with which amino acids
- The genetic code is universal and employed by all living things

### **The genetic code (RNA codons)**

- Translation occurs in **ribosomes**, which are the protein-making factories of the cell
  - each ribosome is a complex of proteins and several segments of ribosomal RNA (rRNA)
  - ribosomes are comprised of two subunits
    - small subunit
    - large subunit
  - the small subunit has a short sequence of its rRNA exposed that is identical to the leader sequence that begins all genes
    - mRNA binds to the small subunit
- The large RNA subunit has three binding sites for transfer RNA (tRNA) located directly adjacent to the exposed rRNA sequence on the small subunit
  - these binding sites are called the A, P, and E sites
  - it is the tRNA molecules that bring amino acids to the ribosome to use in making proteins

### **A ribosome is composed of two subunits**

- The structure of a tRNA molecule is important to its function
  - it holds an amino acid attachment site at one end and a three-nucleotide sequence at the other end
  - this three-nucleotide sequence is called the **anticodon** and is complementary to 1 of the 64 codons of the genetic code
  - **activating enzymes** match the amino acids with their proper tRNAs

### **The structure of tRNA.**

- Once an mRNA molecule has bound to the small ribosomal subunit, the other larger ribosomal subunit binds as well, forming a complete ribosome
  - during translation, the mRNA threads through the ribosome three nucleotides at a time
  - a new tRNA holding an amino acid to be added enters the ribosome at the A site
- Before a new tRNA can be added, the previous tRNA in the A site shifts to the P site
- At the P site **peptide** bonds form between the incoming amino acid and the growing chain of amino acids

- The now empty tRNA in the P site eventually shifts to the E site where it is released

### **How translation works**

- Translation continues until a “**stop**” codon is encountered that signals the end of the protein
- The ribosome then falls apart and the newly made protein is released into the cell

### **Ribosomes guide the translation process**

#### **Architecture of the Gene**

- The prokaryotic gene is an uninterrupted stretch of DNA nucleotides that corresponds to proteins
- In contrast, the coding portions of the DNA nucleotide sequence are interrupted by non-coding sections of DNA
  - the coding portions are known as **exons** while the non-coding portions are known as **introns**
- When a eukaryotic cell first transcribes a gene, it produces a **primary RNA transcript** of the entire gene
  - the primary transcript is then processed in the nucleus
    - enzyme-RNA complexes cut out the introns and join together the exons to form a shorter mRNA transcript
    - the sequences of the introns (90% of typical human gene) are not translated
    - a 5' cap and a 3' poly-A tail are also added
- In humans, genes may be spliced together in different ways
  - by using different combinations of the same exons, different proteins can be created
  - this is termed **alternative splicing**
  - the 25,000 genes of the human genome appear to encode as many as 120,000 different mRNAs

### **How protein synthesis works in eukaryotes**

- Most eukaryotic genes exist in multiple copies called **multi-gene families**
  - the copies are different versions of the same gene
- Some DNA sequences are scattered randomly about on the chromosomes
  - they are unusual in that they are repeated thousands of times and can move around from one chromosome to another

- these segments are called **transposable sequences** or **transposons**

### Turning Genes Off and On

- Every cell must be able to regulate when particular genes are used
  - cells control gene expression by saying when individual genes are to be transcribed
  - in prokaryotes, genes can be turned off by the binding of a **repressor**, a protein that binds to the DNA and blocks access to the promoter
  - genes can be turned on by the **activator**, a protein that makes the promoter more accessible to RNA polymerase
- An **operon** is a segment of DNA containing a cluster of genes that are all transcribed as a unit
  - for example, the *lac* operon in the bacterium *Escherichia coli*
    - this operon contains genes that code for enzymes associated with breaking down the sugar lactose
    - in addition the operon contains regulatory elements: the operator and promoter
    - the presence of lactose affects a repressor protein and causes it to fall off the operator, making it possible for RNA polymerase to transcribe the operon

### How the *lac* operon works

- Activators work with repressors to control transcription
  - in the *lac* operon, the activator is known as CAP
    - CAP is required for RNA polymerase to function properly at the operon
    - CAP binds to the DNA when glucose and lactose levels are low
      - but transcription only proceeds when lactose is present

### Activators and repressors at the *lac* operon

- Eukaryotic genes have special sequences, called **enhancers**, that help guide RNA polymerase to the promoter at the beginning of the gene
  - even if this regulatory sequence is located far away from the gene it influences, the enhancer can have an effect because the DNA can bend to bring the regulatory and promoter sequences together

### How enhancers work