BASIC

NUCLEIC ACID STRUCTURE AND FUNCTION | **Lead Editor:** Bob Moss

# Translation: DNA to mRNA to Protein

**By: Suzanne Clancy, Ph.D. & William Brown, Ph.D. (*Write Science Right*) © 2008 Nature Education**

**Citation: Clancy, S. & Brown, W. (2008) Translation: DNA to mRNA to Protein. *Nature Education* 1(1):101**

How does the cell convert DNA into working proteins? The process of translation can be seen as the decoding of instructions for making proteins, involving mRNA in transcription as well as tRNA.

Aa Aa Aa

The genes in DNA encode protein molecules, which are the "workhorses" of the cell, carrying out all the functions necessary for life. For example, enzymes, including those that metabolize nutrients and synthesize new cellular constituents, as well as DNA polymerases and other enzymes that make copies of DNA during cell division, are all proteins.

In the simplest sense, expressing a gene means manufacturing its corresponding protein, and this multilayered process has two major steps. In the first step, the information in DNA is transferred to a messenger RNA (mRNA) molecule by way of a process called transcription. During transcription, the DNA of a gene serves as a template for complementary base-pairing, and an enzyme called RNA polymerase II catalyzes the formation of a pre-mRNA molecule, which is then processed to form mature mRNA (Figure 1). The resulting mRNA is a single-stranded copy of the gene, which next must be translated into a protein molecule.

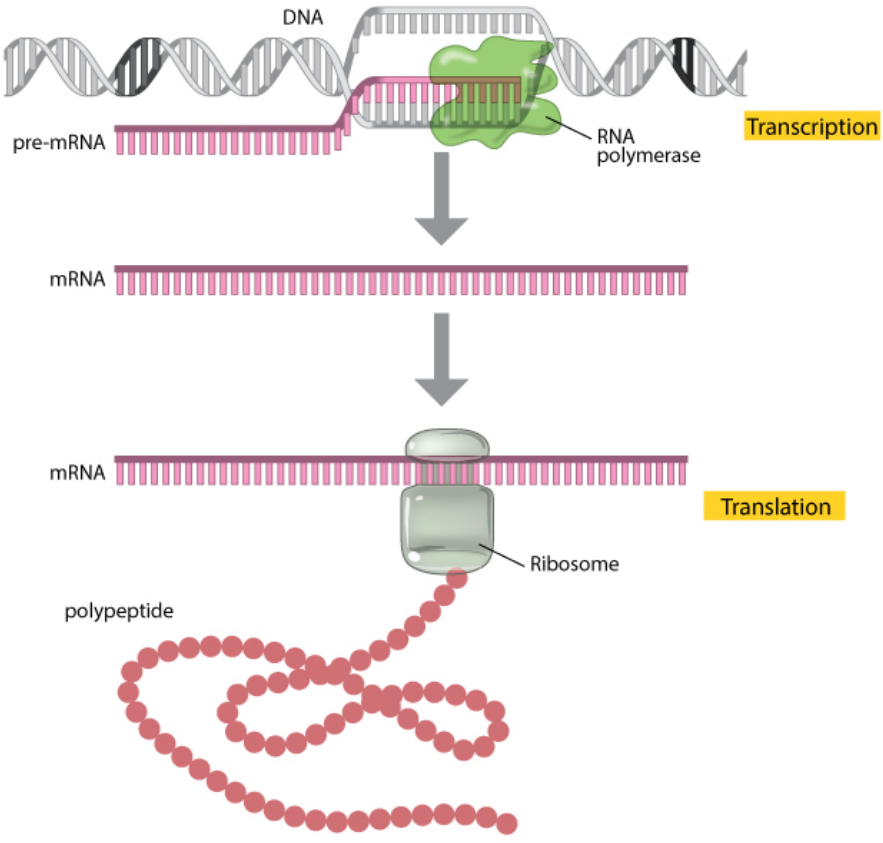


**Figure 1: A gene is expressed through the processes of transcription and translation.**
During transcription, the enzyme RNA polymerase (green) uses DNA as a template to produce a pre-mRNA transcript (pink). The pre-mRNA is processed to form a mature mRNA molecule that can be translated to build the protein molecule (polypeptide) encoded by the original gene.


Figure Detail

During translation, which is the second major step in gene expression, the mRNA is "read" according to the genetic code, which relates the DNA sequence to the amino acid sequence in proteins (Figure 2). Each group of three bases in mRNA constitutes a codon, and each codon specifies a particular amino acid (hence, it is a triplet code). The mRNA sequence is thus used as a template to assemble—in order—the chain of amino acids that form a protein.

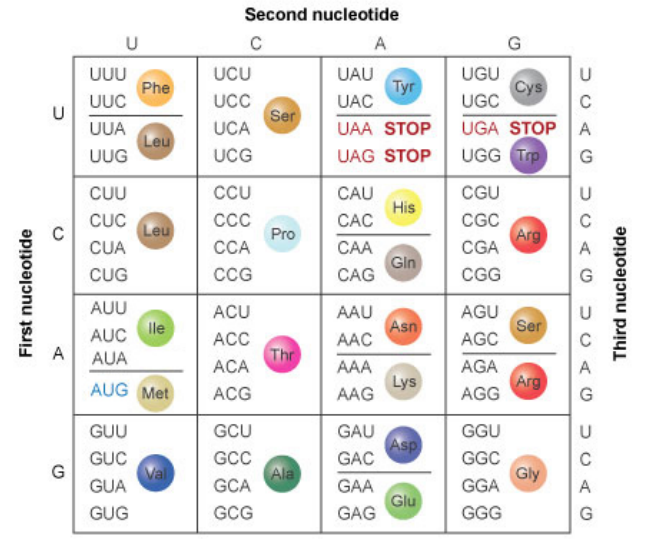


**Figure 2: The amino acids specified by each mRNA codon. Multiple codons can code for the same amino acid.**

The codons are written 5' to 3', as they appear in the mRNA. AUG is an initiation codon; UAA, UAG, and UGA are termination (stop) codons.



Figure Detail

But where does translation take place within a cell? What individual substeps are a part of this process? And does translation differ between prokaryotes and eukaryotes? The answers to questions such as these reveal a great deal about the essential similarities between all species.

## Where Translation Occurs

Within all cells, the translation machinery resides within a specialized organelle called the ribosome. In eukaryotes, mature mRNA molecules must leave the nucleus and travel to the cytoplasm, where the ribosomes are located. On the other hand, in prokaryotic organisms, ribosomes can attach to mRNA while it is still being transcribed. In this situation, translation begins at the 5' end of the mRNA while the 3' end is still attached to DNA.

In all types of cells, the ribosome is composed of two subunits: the large (50S) subunit and the small (30S) subunit (S, for svedberg unit, is a measure of sedimentation velocity and, therefore, mass). Each subunit exists separately in the cytoplasm, but the two join together on the mRNA molecule. The ribosomal subunits contain proteins and specialized RNA molecules—specifically, ribosomal RNA (rRNA) and transfer RNA (tRNA). The tRNA molecules are adaptor molecules—they have one end that can read the triplet code in the mRNA through complementary base-pairing, and another end that attaches to a specific amino acid (Chapeville *et al.*, 1962; Grunberger *et al.*, 1969). The idea that tRNA was an adaptor molecule was first proposed by Francis Crick, co-discoverer of DNA structure, who did much of the key work in deciphering the genetic code (Crick, 1958).

Within the ribosome, the mRNA and aminoacyl-tRNA complexes are held together closely, which facilitates base-pairing. The rRNA catalyzes the attachment of each new amino acid to the growing chain.

## The Beginning of mRNA Is Not Translated

Interestingly, not all regions of an mRNA molecule correspond to particular amino acids. In particular, there is an area near the 5' end of the molecule that is known as the untranslated region (UTR) or leader sequence. This portion of mRNA is located between the first nucleotide that is transcribed and the start codon (AUG) of the coding region, and it does not affect the sequence of amino acids in a protein (Figure 3).

So, what is the purpose of the UTR? It turns out that the leader sequence is important because it contains a ribosome-binding site. In bacteria, this site is known as the Shine-Dalgarno box (AGGAGG), after scientists John Shine and Lynn Dalgarno, who first characterized it. A similar site in vertebrates was characterized by Marilyn Kozak and is thus known as the Kozak box. In bacterial mRNA, the 5' UTR is normally short; in human mRNA, the median length of the 5' UTR is about 170 nucleotides. If the leader is long, it may contain regulatory sequences, including binding sites for proteins, that can affect the stability of the mRNA or the efficiency of its translation.

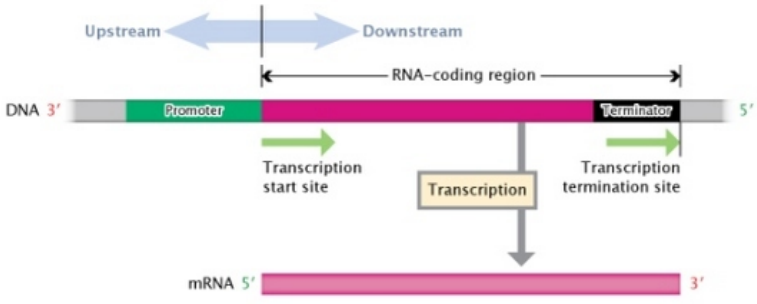


**Figure 3: A DNA transcription unit.**

A DNA transcription unit is composed, from its 3' to 5' end, of an RNA-coding region (pink rectangle) flanked by a promoter region (green rectangle) and a terminator region (black rectangle). Regions to the left, or moving towards the 3' end, of the transcription start site are considered \"upstream;\" regions to the right, or moving towards the 5' end, of the transcription start site are considered \"downstream.\"

## Translation Begins After the Assembly of a Complex Structure

The translation of mRNA begins with the formation of a complex on the mRNA (Figure 4). First, three initiation factor proteins (known as IF1, IF2, and IF3) bind to the small subunit of the ribosome. This preinitiation complex and a methionine-carrying tRNA then bind to the mRNA, near the AUG start codon, forming the initiation complex.

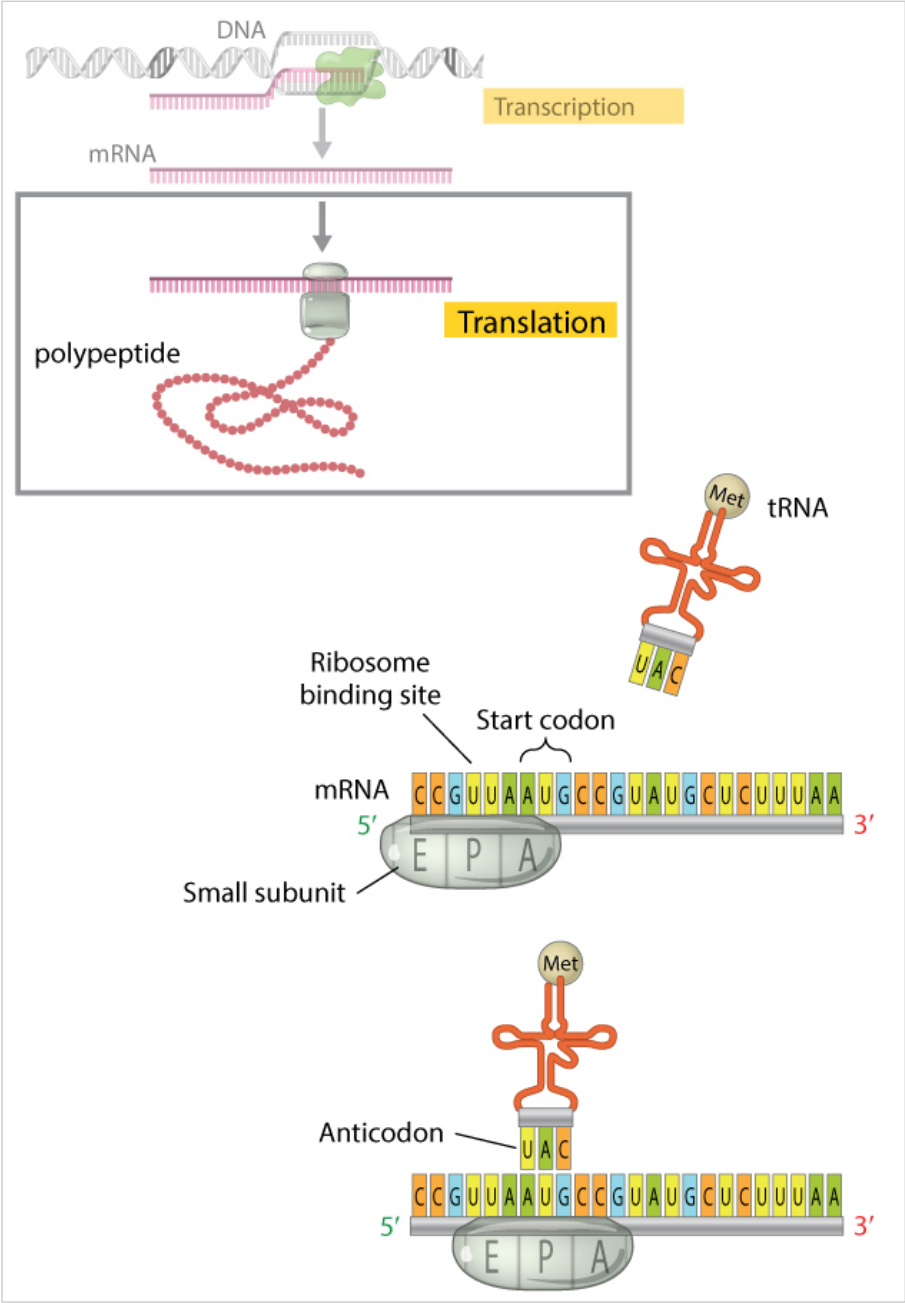


**Figure 4: The translation initiation complex.**
When translation begins, the small subunit of the ribosome and an initiator tRNA molecule assemble on the mRNA transcript. The small subunit of the ribosome has three binding sites: an amino acid site (A), a polypeptide site (P), and an exit site (E). The initiator tRNA molecule carrying the amino acid methionine binds to the AUG start codon of the mRNA transcript at the ribosome's P site where it will become the first amino acid incorporated into the growing polypeptide chain. Here, the initiator tRNA molecule is shown binding after the small ribosomal subunit has assembled on the mRNA; the order in which this occurs is unique to prokaryotic cells. In eukaryotes, the free initiator tRNA first binds the small ribosomal subunit to form a complex. The complex then binds the mRNA transcript, so that the tRNA and the small ribosomal subunit bind the mRNA simultaneously.


Figure Detail

Although methionine (Met) is the first amino acid incorporated into any new protein, it is not always the first amino acid in mature proteins—in many proteins, methionine is removed after translation. In fact, if a large number of proteins are sequenced and compared with their known gene sequences, methionine (or formylmethionine) occurs at the N-terminus of all of them. However, not all amino acids are equally likely to occur second in the chain, and the second amino acid influences whether the initial methionine is enzymatically removed. For example, many proteins begin with methionine followed by alanine. In both prokaryotes and eukaryotes, these proteins have the methionine removed, so that alanine becomes the N-terminal amino acid (Table 1). However, if the second amino acid is lysine, which is also frequently the case, methionine is not removed (at least in the sample proteins that have been studied thus far). These proteins therefore begin with methionine followed by lysine (Flinta *et al.*, 1986).

Table 1 shows the N-terminal sequences of proteins in prokaryotes and eukaryotes, based on a sample of 170 prokaryotic and 120 eukaryotic proteins (Flinta *et al.*, 1986). In the table, M represents methionine, A represents alanine, K represents lysine, S represents serine, and T represents threonine.

***Table 1: N-Terminal Sequences of Proteins***

| N-Terminal Sequence | Percent of Prokaryotic Proteins with This Sequence | Percent of Eukaryotic Proteins with This Sequence |
|---|---|---|
| MA* | 28.24% | 19.17% |
| MK** | 10.59% | 2.50% |
| MS* | 9.41% | 11.67% |
| MT* | 7.65% | 6.67% |

* Methionine was removed in all of these proteins

** Methionine was not removed from any of these proteins

Once the initiation complex is formed on the mRNA, the large ribosomal subunit binds to this complex, which causes the release of IFs (initiation factors). The large subunit of the ribosome has three sites at which tRNA molecules can bind. The A (amino acid) site is the location at which the aminoacyl-tRNA anticodon base pairs up with the mRNA codon, ensuring that correct amino acid is added to the growing polypeptide chain. The P (polypeptide) site is the location at which the amino acid is transferred from its tRNA to the growing polypeptide chain. Finally, the E (exit) site is the location at which the "empty" tRNA sits before being released back into the cytoplasm to bind another amino acid and repeat the process. The initiator methionine tRNA is the only aminoacyl-tRNA that can bind in the P site of the ribosome, and the A site is aligned with the second mRNA codon. The ribosome is thus ready to bind the second aminoacyl-tRNA at the A site, which will be joined to the initiator methionine by the first peptide bond (Figure 5).

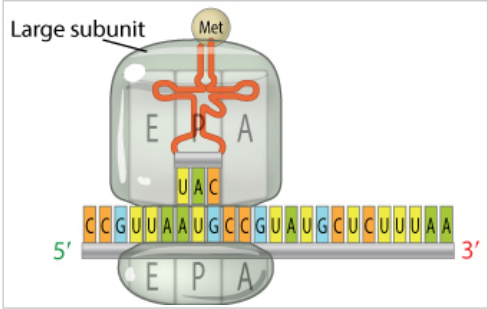


**Figure 5: The large ribosomal subunit binds to the small ribosomal subunit to complete the initiation complex.**
The initiator tRNA molecule, carrying the methionine amino acid that will serve as the first amino acid of the polypeptide chain, is bound to the P site on the ribosome. The A site is aligned with the next codon, which will be bound by the anticodon of the next incoming tRNA.


## The Elongation Phase

The next phase in translation is known as the elongation phase (Figure 6). First, the ribosome moves along the mRNA in the 5'-to-3'direction, which requires the elongation factor G, in a process called translocation. The tRNA that corresponds to the second codon can then bind to the A site, a step that requires elongation factors (in *E. coli*, these are called EF-Tu and EF-Ts), as well as guanosine triphosphate (GTP) as an energy source for the process. Upon binding of the tRNA-amino acid complex in the A site, GTP is cleaved to form guanosine diphosphate (GDP), then released along with EF-Tu to be recycled by EF-Ts for the next round.

Next, peptide bonds between the now-adjacent first and second amino acids are formed through a peptidyl transferase activity. For many years, it was thought that an enzyme catalyzed this step, but recent evidence indicates that the transferase activity is a catalytic function of rRNA (Pierce, 2000). After the peptide bond is formed, the ribosome shifts, or translocates, again, thus causing the tRNA to occupy the E site. The tRNA is then released to the cytoplasm to pick up another amino acid. In addition, the A site is now empty and ready to receive the tRNA for the next codon.

This process is repeated until all the codons in the mRNA have been read by tRNA molecules, and the amino acids attached to the tRNAs have been linked together in the growing polypeptide chain in the appropriate order. At this point, translation must be terminated, and the nascent protein must be released from the mRNA and ribosome.

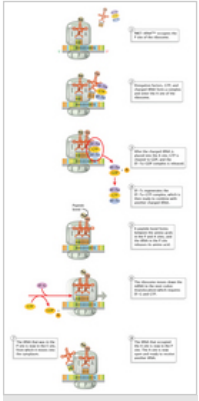

Figure 6

Figure Detail

## Termination of Translation

There are three termination codons that are employed at the end of a protein-coding sequence in mRNA: UAA, UAG, and UGA. No tRNAs recognize these codons. Thus, in the place of these tRNAs, one of several proteins, called release factors, binds and facilitates release of the mRNA from the ribosome and subsequent dissociation of the ribosome.

## Comparing Eukaryotic and Prokaryotic Translation

The translation process is very similar in prokaryotes and eukaryotes. Although different elongation, initiation, and termination factors are used, the genetic code is generally identical. As previously noted, in bacteria, transcription and translation take place simultaneously, and mRNAs are relatively short-lived. In eukaryotes, however, mRNAs have highly variable half-lives, are

subject to modifications, and must exit the nucleus to be translated; these multiple steps offer additional opportunities to regulate levels of protein production, and thereby fine-tune gene expression.

## References and Recommended Reading

Chapeville, F., *et al.* On the role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences* **48**, 1086–1092 (1962)

Crick, F. On protein synthesis. *Symposia of the Society for Experimental Biology* **12**, 138–163 (1958)

Flinta, C., *et al*. Sequence determinants of N-terminal protein processing. *European Journal of Biochemistry* **154**, 193–196 (1986)

Grunberger, D., *et al*. Codon recognition by enzymatically mischarged valine transfer ribonucleic acid. *Science* **166**, 1635–1637 (1969) doi:10.1126/science.166.3913.1635

Kozak, M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*. *Nature* **308**, 241–246 (1984) doi:10.1038308241a0 (link to article)

---. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986)

---. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* **15**, 8125–8148 (1987)

Pierce, B. A. *Genetics: A conceptual approach* (New York, Freeman, 2000)

Shine, J., & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**, 34–38 (1975) doi:10.1038/254034a0 (link to article)

**Outline** | **Keywords** | **Add Content to Group** FEEDBACK

## Explore This Subject

### APPLICATIONS IN BIOTECHNOLOGY

- Genetically Modified Organisms (GMOs): Transgenic Crops and Recombinant DNA Technology
- Recombinant DNA Technology and Transgenic Animals
- Restriction Enzymes
- The Biotechnology Revolution: PCR and the Use of Reverse Transcriptase to Clone Expressed Genes

### DNA REPLICATION

- DNA Damage & Repair: Mechanisms for Maintaining DNA Integrity
- DNA Replication and Causes of Mutation
- Genetic Mutation
- Genetic Mutation
- Major Molecular Events of DNA Replication
- Semi-Conservative DNA Replication: Meselson and Stahl

### JUMPING GENES

- Barbara McClintock and the Discovery of Jumping Genes (Transposons)
- Functions and Utility of *Alu* Jumping Genes
- Transposons, or Jumping Genes: Not Junk DNA?
- Transposons: The Jumping Genes

### TRANSCRIPTION & TRANSLATION

- DNA Transcription
- RNA Transcription by RNA Polymerase: Prokaryotes vs Eukaryotes
- Translation: DNA to mRNA to Protein
- What is a Gene? Colinearity and Transcription Units

### DISCOVERY OF GENETIC MATERIAL

- Barbara McClintock and the Discovery of Jumping Genes (Transposons)
- Discovery of DNA as the Hereditary Material using *Streptococcus pneumoniae*
- Discovery of DNA Structure and Function: Watson and Crick
- Isolating Hereditary Material: Frederick Griffith, Oswald Avery, Alfred Hershey, and Martha Chase

### GENE COPIES

- Copy Number Variation
- Copy Number Variation and Genetic Disease
- Copy Number Variation and Human Disease
- DNA Deletion and Duplication and the Associated Genetic Disorders
- Tandem Repeats and Morphological Variation

### RNA

- Chemical Structure of RNA
- Eukaryotic Genome Complexity
- Genome Packaging in Prokaryotes: the Circular Chromosome of *E. coli*
- RNA Functions
- RNA Splicing: Introns, Exons and Spliceosome
- RNA Transcription by RNA Polymerase: Prokaryotes vs Eukaryotes
- What is a Gene? Colinearity and Transcription Units