Converting DNA to PROTIENS

END-SEM PROJECT REPORT

Submitted by

B.SAI ABHISHEK - BL.EN.U4AIE21015
RUCHITH BALAJI - BL.EN.U4AIE21017
CHILAKURU HARI - BL.EN.U4AIE21038

In partial fulfillment for the award of the degree

Of

BACHELOR OF TECHNOLOGY

IN

INTELLIGENCE IN BIOLOGICAL SYSTEMS

AMRITA SCHOOL OF ENGINEERING, BANGALORE

AMRITA VISHWA VIDYAPEETHAM

BANGALORE – 560035

July -2022.

TABLE OF CONTENTS

ABSTRACT

This chapter discusses components of nucleotides, DNA, and protein synthesis. Genetic information is stored in DNA in the nucleus and mitochondria of cells. DNA consists of two strands of nucleotides on a phosphor-deoxyribose backbone.

The two strands form a double helix that is stabilized by the formation of hydrogen bonds among nucleotide bases on the two strands. Replication of DNA is based on making strands complementary to the two strands produced when the double-stranded DNA is unwound. The four bases include two purines and two pyrimidines. Protein synthesis occurs on the ribosomes, using messenger RNA (mRNA) as a template for transfer RNA.

The ribosomes themselves are large and complex structures consisting of ribosomal RNA (rRNA) and a number of proteins. The ribosomes consist of a 60S and 40S subunit. The 60S subunit has three rRNA and 49 other proteins, and the 40S subunit has an 18S rRNA and another 33 proteins. It is found that mRNA directs the synthesis of specific proteins by virtue of its sequence of codons.

Cells have developed elaborate methods for determining what parts of the DNA are transcribed into mRNA. These methods involve transcription factors—enhancers and repressors—and the histone code. The DNA in cells is wrapped around a complex of histone proteins, forming a nucleosome. It is observed that modification of the histones allows specific sections of the DNA to be either used to make proteins or silenced.

LIST OF FIGURES

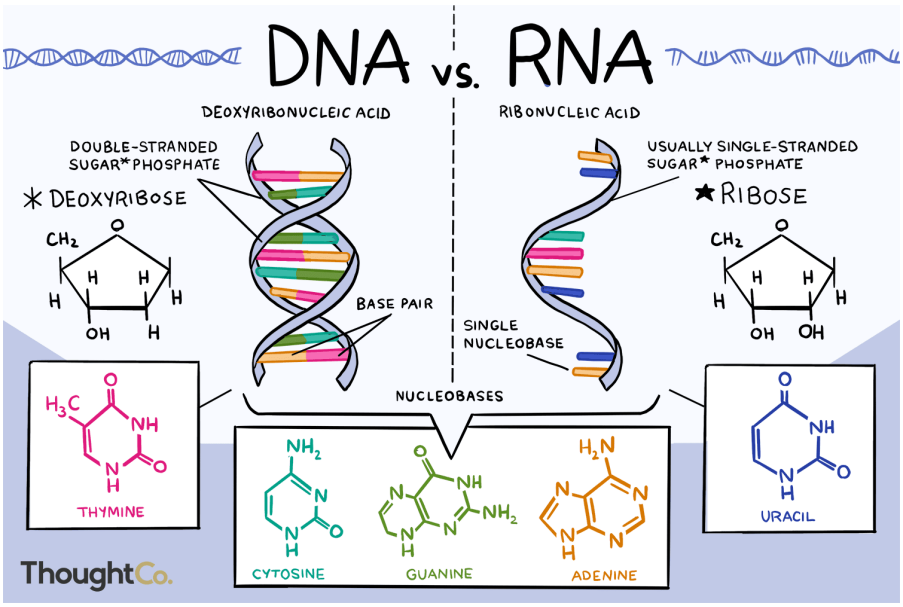
4.1 Introduction:

Life on earth is very diverse, from single-celled protozoans to complex multicellular plants and animals. But at the molecular level, all life is fundamentally made up of the same building blocks – DNA and RNA.

One of the primary differences between DNA and RNA is that DNA is double-stranded while RNA is single-stranded.

Messenger RNA (mRNA) analysis is the most commonly-used application for qPCR. Researchers reverse transcribe RNA, then use the cDNA produced as a template in qPCR reactions to detect and quantitate gene expression products

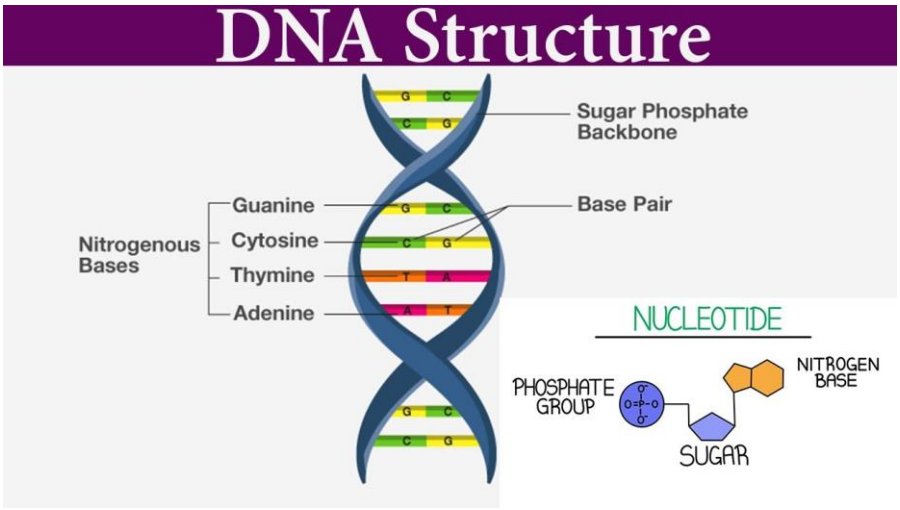
Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, but in nonprotein coding genes such as rRNA genes or tRNA genes, the product is a structural or housekeeping RNA. In addition, small non-coding RNAs (miRNAs, pi RNA) and various classes of long non-coding RNAs are involved in a variety of regulatory functions



4.2 DNA:

DNA, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms. Nearly every cell in a person's body has the same DNA. Most DNA is located in the cell nucleus (where it is called nuclear DNA), but a small amount of DNA can also be found in the mitochondria (where it is called mitochondrial DNA or mtDNA). Mitochondria are structures within cells that convert the energy from food into a form that cells can use.

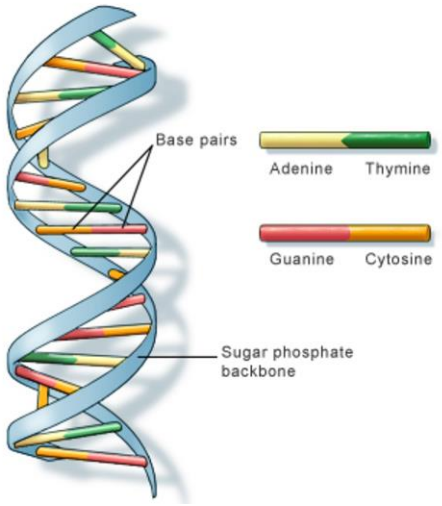
The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Human DNA consists of about 3 billion bases, and more than 99 percent of those bases are the same in all people. The order, or sequence, of these bases determines the information available for building and maintaining an organism, similar to the way in which letters of the alphabet appear in a certain order to form words and sentences.



DNA bases pair up with each other, A with T and C with G, to form units called base pairs. Each base is also attached to a sugar molecule and a phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are arranged in two long strands that form a spiral called a double helix.

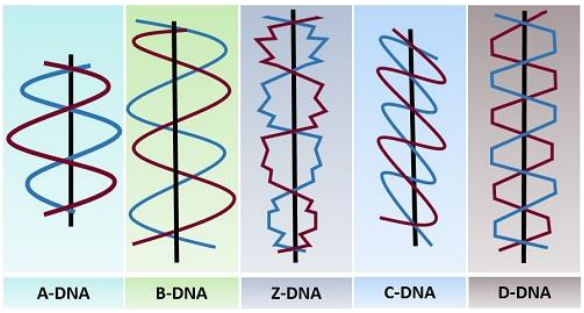
The structure of the double helix is somewhat like a ladder, with the base pairs forming the ladder's rungs and the sugar and phosphate molecules forming the vertical sidepieces of the ladder.

An important property of DNA is that it can replicate, or make copies of itself. Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of bases. This is critical when cells divide because each new cell needs to have an exact copy of the DNA present in the old cell.



Types of DNA

a) A – DNA
b) B – DNA
c) C – DNA
d) D – DNA
e) Z – DNA

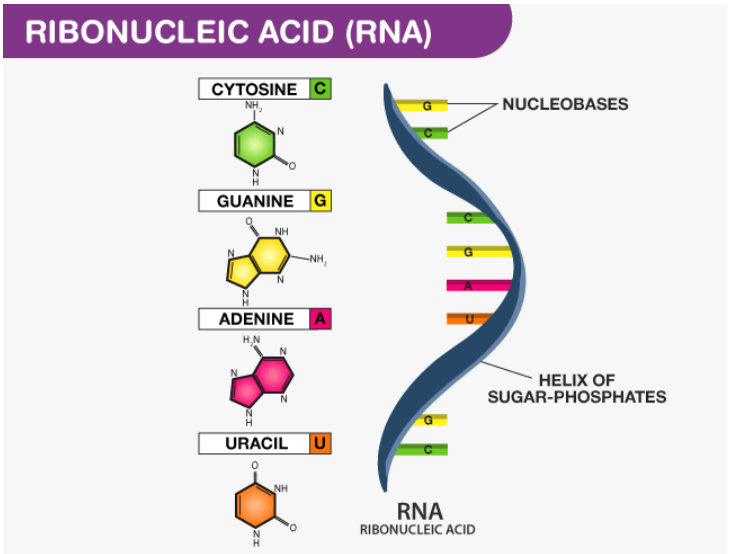


4.3 RNA

Ribonucleic acid (RNA) is a molecule that is present in the majority of living organisms and viruses. It is made up of nucleotides, which are ribose sugars attached to nitrogenous bases and phosphate groups. The nitrogenous bases include adenine, guanine, uracil, and cytosine. RNA mostly exists in the single-stranded form, but there are special RNA viruses that are double-stranded. The RNA molecule can have a variety of lengths and structures.

An RNA virus uses RNA instead of DNA as its genetic material and can cause many human diseases. Transcription is the process of RNA formation from DNA, and translation is the process of protein synthesis from RNA. The means of RNA synthesis and the way that it functions differs between eukaryotes and prokaryotes. Specific RNA molecules also regulate gene expression and have the potential to serve as therapeutic agents in human diseases.

RNA contains the sugar ribose, phosphates, and the nitrogenous bases adenine (A), guanine (G),cytosine (C), and uracil (U). DNA and RNA share the nitrogenous bases A, G, and C. Thymine is usually only present in DNA and uracil is usually only present in RNA.



Types Of RNA

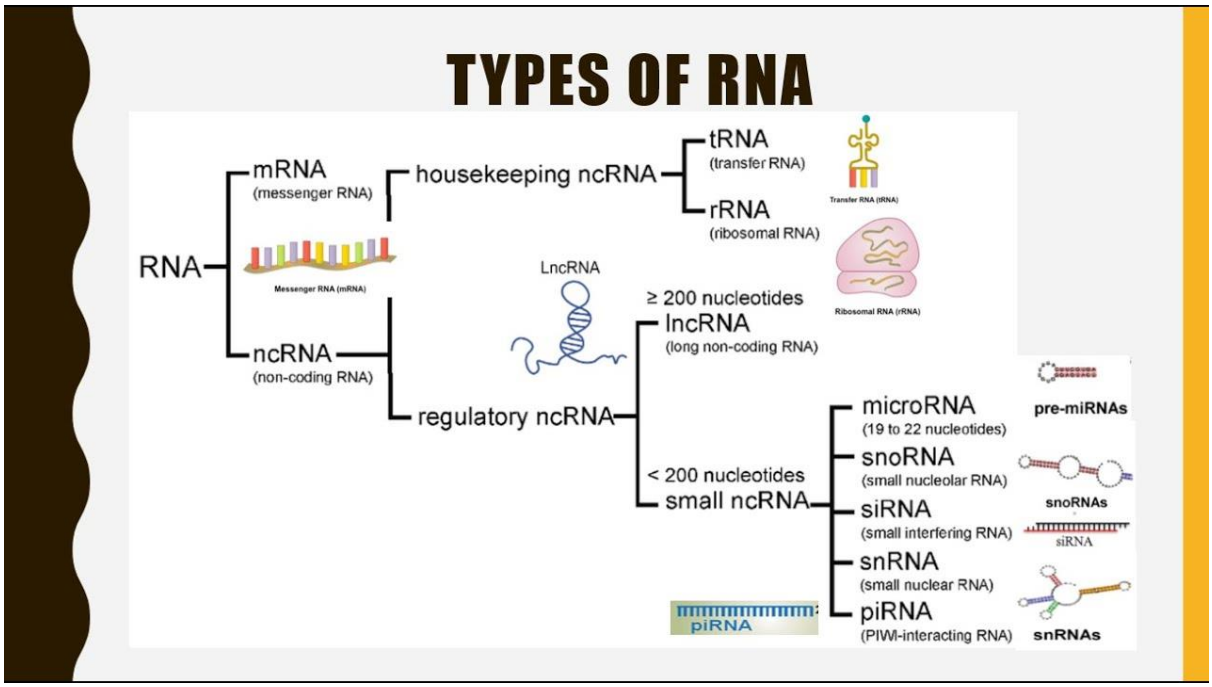
- Only some of the genes in cells are expressed into RNA. The following are the types of RNA wherein each type is encoded by its own type of gene:

➢ **tRNA**– The transfer RNA or the tRNA carries amino acids to ribosomes while translation

➢ **mRNA** – The messenger RNA or the mRNA encodes amino acid sequences of a polypeptide

➢ **rRNA** – The ribosomal RNA or the rRNA produces ribosomes with the ribosomal proteins that are organelles responsible for the translation of the mRNA.

➢ **snRNA** – The small nuclear RNA forms the complexes along with proteins which are utilized in RNA processing in the eukaryotes.

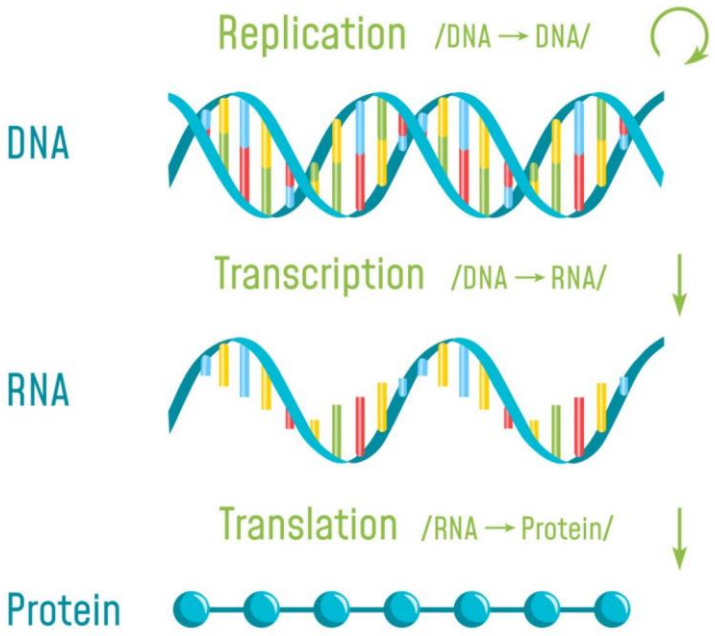


Role of Proteins

- To place these ideas in the proper context, remember that some proteins are enzymes that aid cells by catalyzing chemical reactions.
- These chemical reactions occur after the enzyme binds its substrate at the enzyme's active site.
- The enzyme's active site matches the substrate molecule in size, shape and chemical properties.
- Proteins play a critical role in how cells successfully meet the challenges of living.
- Cells use proteins to maintain their shape and to speed up important chemical reactions such as photosynthesis and respiration.
- A cell will not live long if it cannot reliably create the proteins that it needs for survival.

DNA to Proteins

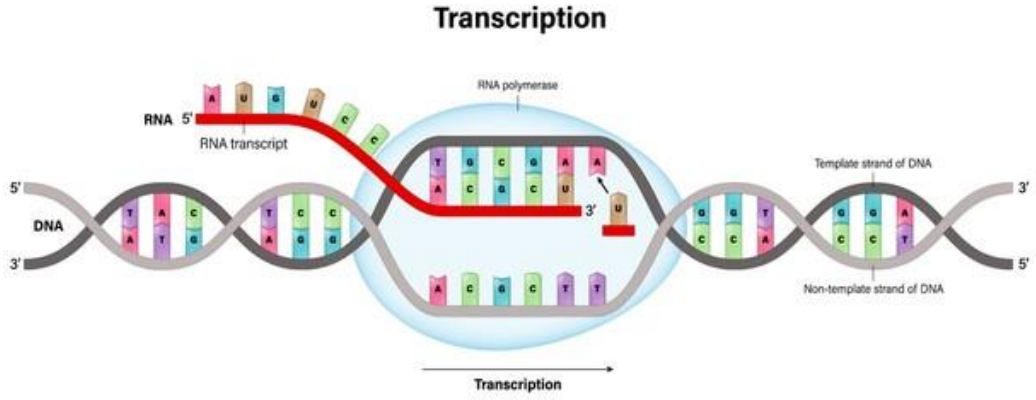
- ❖ Most genes contain the information needed to make functional molecules called proteins.
- ❖ The journey from gene to protein is complex and tightly controlled within each cell. It consists of two major steps: transcription and translation.
- ❖ Together, transcription and translation are known as gene expression.



4.4 DNA to RNA(Transcription)

Transcription is the process by which the information in a strand of DNA is copied into a new molecule of messenger RNA (mRNA). DNA safely and stably stores genetic material in the nuclei of cells as a reference, or template. Meanwhile, mRNA is comparable to a copy from a reference book because it carries the same information as DNA but is not used for long-term storage and can freely exit the nucleus. Although the mRNA contains the same information, it is not an identical copy of the DNA segment, because its sequence is complementary to the DNA template.

Transcription is carried out by an enzyme called RNA polymerase and a number of accessory proteins called transcription factors. Transcription factors can bind to specific DNA sequences called enhancer and promoter sequences in order to recruit RNA polymerase to an appropriate transcription site. Together, the transcription factors and RNA polymerase form a complex called the transcription initiation complex. This complex initiates transcription, and the RNA polymerase begins mRNA synthesis by matching complementary bases to the original DNA strand. The mRNA molecule is elongated and, once the strand is completely synthesized, transcription is terminated. The newly formed mRNA copies of the gene then serve as blueprints for protein synthesis during the process of translation.



- Consider the following DNA string as an example.

AATTTGCTTGGCAAATCGTATGCCTTGGG

- You can obtain the relevant RNA string as follows, by replacing all occurrences of T with U.

AAUUUGCUUGGCAAAUCGUAUGCCUUGGG

4.5 RNA to PROTEINS

Translation involves “decoding” a messenger RNA (mRNA) and using its information to build a polypeptide, or chain of amino acids. For most purposes, a polypeptide is basically just a protein (with the technical difference being that some large proteins are made up of several polypeptide chains).

The genetic code

In an mRNA, the instructions for building a polypeptide come in groups of three nucleotides called codons. Here are some key features of codons to keep in mind as we move forward:

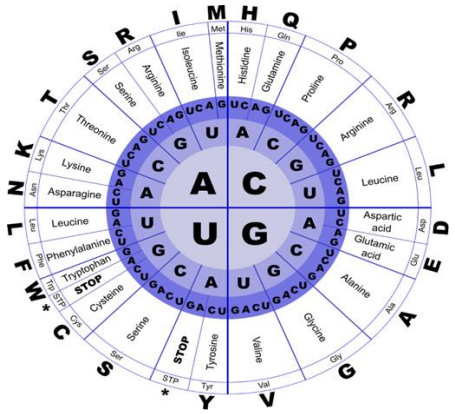
- There are 616161 different codons for amino acids
- Three “stop” codons mark the polypeptide as finished
- One codon, AUG, is a “start” signal to kick off translation (it also specifies the amino acid methionine)

These relationships between mRNA codons and amino acids are known as the genetic code.

Codons to amino acids

In translation, the codons of an mRNA are read in order (from the 5' end to the 3' end) by molecules called **transfer RNAs**, or **tRNAs**.

Each tRNA has an **anticodon**, a set of three nucleotides that binds to a matching mRNA codon through base pairing. The other end of the tRNA carries the amino acid that's specified by the codon.



Translation: Beginning, middle, and end

A book or movie has three basic parts: a beginning, middle, and end. Translation has pretty much the same three parts, but they have fancier names: initiation, elongation, and termination.

- **Initiation** ("beginning"): in this stage, the ribosome gets together with the mRNA and the first tRNA so translation can begin.

- **Elongation** ("middle"): in this stage, amino acids are brought to the ribosome by tRNAs and linked together to form a chain.

- **Termination** ("end"): in the last stage, the finished polypeptide is released to go and do its job in the cell.

- If you know an RNA sequence, you can translate it into the corresponding protein sequence by using the genetic code. This is the same way the cell itself generates a protein sequence.

- The genetic code for RNA (also called RNA codon table) shows how we uniquely relate a 4-nucleotide sequence (A, U, G, C) to a set of 20 amino acids.

4.6 CODE

```
def transcript(x) :
  l = list(x)
  for i in range(len(x)):
      if(l[i]=='G'):
          l[i]='C'
      elif(l[i]=='C'):
          l[i]='G'
      elif (l[i] == 'T'):
          l[i] = 'A'
      elif (l[i] == 'A'):
          l[i] = 'U'
      else:
          print('Invalid Input')

  print("Translated DNA : ",end="")
  for char in l:
      print(char,end="")
inputfile ="DNA_sequence_original.txt"
f = open(inputfile, "r")
x = f.read()

x = x.replace("\n", "")
x = x.replace("\r", "")
# function calling
Z =transcript(x)

print("")
rna = input("Enter the RNA sequence: ")

# RNA codon table
rna_codon = {
  "GCA":"A", "GCC":"A", "GCG":"A", "GCU":"A",
  "UGC":"C", "UGU":"C", "GAC":"D", "GAU":"D",
  "GAA":"E", "GAG":"E", "UUC":"F", "UUU":"F",
  "GGA":"G", "GGC":"G", "GGG":"G", "GGU":"G",
  "CAC":"H", "CAU":"H", "AUA":"I", "AUC":"I",
  "AUU":"I", "AAA":"K", "AAG":"K", "UUA":"L",
  "UUG":"L", "CUA":"L", "CUC":"L", "CUG":"L",
  "CUU":"L", "AUG":"M", "AAC":"N", "AAU":"N",
  "CCA":"P", "CCC":"P", "CCG":"P", "CCU":"P",
  "CAA":"Q", "CAG":"Q", "AGA":"R", "AGG":"R",
  "CGA":"R", "CGC":"R", "CGU":"R", "CGG":"R",
  "AGC":"S", "AGU":"S", "UCA":"S", "UCC":"S",
  "UCG":"S", "UCU":"S", "ACA":"T", "ACC":"T",
  "ACG":"T", "ACU":"T", "GUA":"V", "GUC":"V",
  "GUG":"V", "GUU":"V", "UGG":"W", "UAC":"Y",
  "UAU":"Y", "UAG":"!", "UAA":"!", "UGA":"!"
}
```

```
protein_string = ""

# Generate protein string
for i in range(0, len(rna)-(3+len(rna)%3), 3):
    if (rna_codon[rna[i:i+3]] == "!") :
        break
    protein_string += rna_codon[rna[i:i+3]]

# Print the protein string
print ("Protein String: ", protein_string)
```

4.7 OUTPUT

```
>>>
= RESTART: C:\Users\MY PC\OneDrive\Desktop\Ruchith_1ST YEAR\SEM2\ENDSEM PROJECTS
\IBS\DNA_PROTIEN.py
Translated DNA : UACUCGGGGGAAUGUUGUUGUUAAGAUGAUAGUUCGAAUCGUUGGCCGUGGUAGUAAUGUCGG
UGUUCGAUAGUGACUGAUAAUCGAACUUAACCGGAACUUGAUUUGUGUAAUCGAUAGUAAGGUUAAUAGAGUUUUGUUGU
AGUGGGGGCUCGCUGUCUCCGGCGGUGGUUCAUGAAGAAUUGAGUUCGUCGACGAAGUCGUGAUUAGGAUAAGAGUUCAU
GUUAGUUGCGAACUGUGAGUCCUUGAACCCUGUAAUGUGUUUAUUGUUUGUUAUGUAGUUUGUAGAAUGAUUGUUACCGU
GAUCGGUAUUUUAAUCCGGAACGUGGUUGUGUGAUGACUGAGGGUCUCCAAUAAGUUCCUUGUAGUUAUUGUUGUCGGGA
UUAAUAAUGAUGCACUGUUUUUGAUCGGGGAUAUCGUGAGUAAUAAUAAUGUUCGUUGUUGAAUAGGAUAUAUCAGAAUU
AUUGUUACCCCCAUAAUAGAUGGUAGCAUCCUCCUACCCCGCCGGAUUUGGUUUGGGUUUGGGCUUUUUAGUAUCGCAUG
AGGAGUUAGCGUGUGGAACCGACUUAUAGUUACCAAUGGGGUGAUUAUUGUUUUGAUGAGUAAAAUUUGGAACCGUAAAU
GUAGGAUUAUUGUUGUCGAUAUAAGGAGAGUGAUUAAAAUAGUUUUAGUUUUUGGAAUGUUCUAUGUAGGUGUGAUACCC
GUGAUAGGGGUUGUGAAUAUUAAAAUUGUUAUGAUUGUGAUAAUAGGGAUCCUCCUGACGGGGGUGAUUGUCCUAAGUAU
GGGUUCACUGAAUAAGAUGUUCUUAAUUGUCGUGUUUUGGAUGAUUAUCAUCGGUGUGAAGAUCGGGAUAGUCGUGAGAA
UUCGGAGAAGAAAAUGGAGGCUGAUUGUAUGUGUUACUGGGAAUGGUGUGGGGGGUUGUGGUGAUAUAAGUUUGUGUUUA
CUGCCAAAUUGAUAAAUUUAUUAUGUUGAGGGGAUAGGUAGAGGUAUCGGGAGAGUUGGGAUAAUAAUGGUUAAUGUGGG
GAGUAGUAUUAUAUCUUUGAAUCCUAUUGUAAUUUGGUUCCCGGAAGUUUCGGGUUUUAUCCUCACUUUUAGGGGAUCAG
AGACUAUUCUGGACACUAUAAGAUUUUGUGUAGAAGACUUACGUUGGGUCUGUG
Enter the RNA sequence: UACUCGGGGGAAUGUUGUUGUUAAGAUGAUAGUUCGAAUCGUUGGCCGUGGUAGUA
AUGUCGGUGUUCGAUAGUGACUGAUAAUCGAACUUAACCGGAACUUGAUUUGUGUAAUCGAUAGUAAGGUUAAUAGAGUU
UUGUUGUAGUGGGGGCUCGCUGUCUCCGGCGGUGGUUCAUGAAGAAUUGAGUUCGUCGACGAAGUCGUGAUUAGGAUAAG
AGUUCAUGUUAGUUGCGAACUGUGAGUCCUUGAACCCUGUAAUGUGUUUAUUGUUUGUUAUGUAGUUUGUAGAAUGAUUG
UUACCGUGAUCGGUAUUUUAAUCCGGAACGUGGUUGUGUGAUGACUGAGGGUCUCCAAUAAGUUCCUUGUAGUUAUUGUU
GUCGGGAUUAAUAAUGAUGCACUGUUUUUGAUCGGGGAUAUCGUGAGUAAUAAUAAUGUUCGUUGUUGAAUAGGAUAUAU
CAGAAUUAUUGUUACCCCCAUAAUAGAUGGUAGCAUCCUCCUACCCCGCCGGAUUUGGUUUGGGUUUGGGCUUUUUAGUA
UCGCAUGAGGAGUUAGCGUGUGGAACCGACUUAUAGUUACCAAUGGGGUGAUUAUUGUUUUGAUGAGUAAAAUUUGGAAC
CGUAAAUGUAGGAUUAUUGUUGUCGAUAUAAGGAGAGUGAUUAAAAUAGUUUUAGUUUUUGGAAUGUUCUAUGUAGGUGU
GAUACCCGUGAUAGGGGUUGUGAAUAUUAAAAUUGUUAUGAUUGUGAUAAUAGGGAUCCUCCUGACGGGGGUGAUUGUCC
UAAGUAUGGGUUCACUGAAUAAGAUGUUCUUAAUUGUCGUGUUUUGGAUGAUUAUCAUCGGUGUGAAGAUCGGGAUAGUC
GUGAGAAUUCGGAGAAGAAAAUGGAGGCUGAUUGUAUGUGUUACUGGGAAUGGUGUGGGGGGUUGUGGUGAUAUAAGUUU
GUGUUUACUGCCAAAUUGAUAAAUUUAUUAUGUUGAGGGGAUAGGUAGAGGUAUCGGGAGAGUUGGGAUAAUAAUGGUUA
AUGUGGGGAGUAGUAUUAUAUCUUUGAAUCCUAUUGUAAUUUGGUUCCCGGAAGUUUCGGGUUUUAUCCUCACUUUUAGG
GGAUCAGAGACUAUUCUGGACACUAUAAGAUUUUGUGUAGAAGACUUACGUUGGGUCUGUG
Protein String:  YSGECCC
>>>
```

4.8 CONCLUSION

- The final product of the transcription is mRNA which is directly used in the translation once it migrates to the cytoplasm.
- Whereas the final product of the translation is a chain of amino acid, although it does not directly form a protein.
- Transcription is a highly complex series of events that involves the interaction of numerous proteins and nucleic acids to result in an mRNA strand that can be translated into a protein.