LABORATORY 6:
CRACKING THE GENETIC CODE

Objectives

In this laboratory session you will examine some of the experiments used by scientists in the 1960s to crack the genetic code. The procedure involved synthesizing specific strands of messenger RNA (mRNA) and then examining what amino acids were produced by translation of these genetic instructions. This was a complex, puzzle-solving exercise that took many years and deservedly won the Nobel Prize. You will be using a computer simulation to reproduce these experiments and use the same systematic approach to construct your own version of the genetic code.

Preparation

Carefully read the introduction in order to prepare for this laboratory project and quiz. You should be familiar with the following concepts and techniques:

- transcription
- translation
- codons
- amino acid structure
- using cell extracts

INTRODUCTION

The Genetic Code

Our proteins are polymers of amino acids, arranged in an order specified by DNA sequences. The cell strings together a sequence of amino acids based on a DNA code comprised of only four **nucleobases: A (adenine), T (thymine), G (guanine)** and **C (cytosine)**. The central steps in this process are illustrated in *Figure 1*.

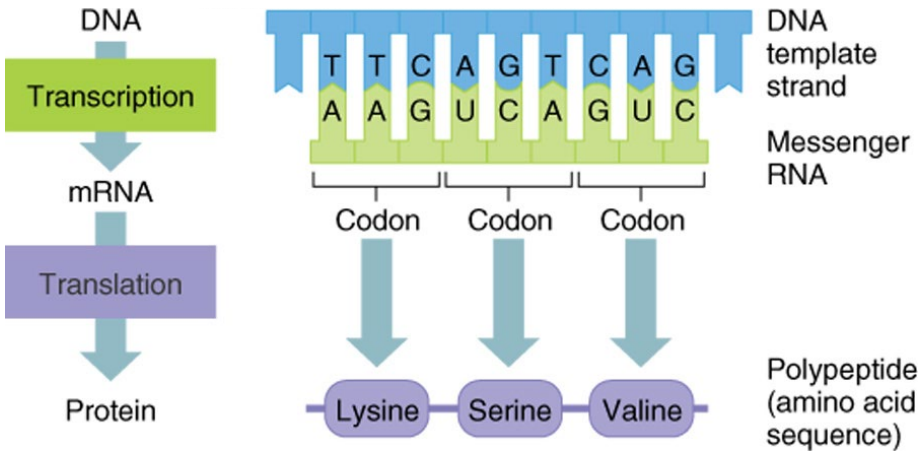


Figure 1: DNA specifies the sequence of a protein via the genetic code.

The first step involves making an RNA version of the DNA's sequence. This process is called **transcription** and uses one of the DNA strands to make a **complementary copy** that becomes a **messenger RNA (mRNA)**. The mRNA strand also has four bases: three of them (A, G, and C) are the same as DNA but mRNA contains **uracil (U)** instead of thymine (T). The term "complementary" refers to the fact that a specific base within the DNA strand (e.g., A) always pairs with a specific base within the mRNA strand (e.g., U). These **base pairing rules** ensure that the sequence of bases in mRNA is a faithful replica (although not an exact copy) of the sequence of bases in its parent DNA. *Figure 2* on the following page illustrates the structures of DNA and mRNA side-by-side for comparison.

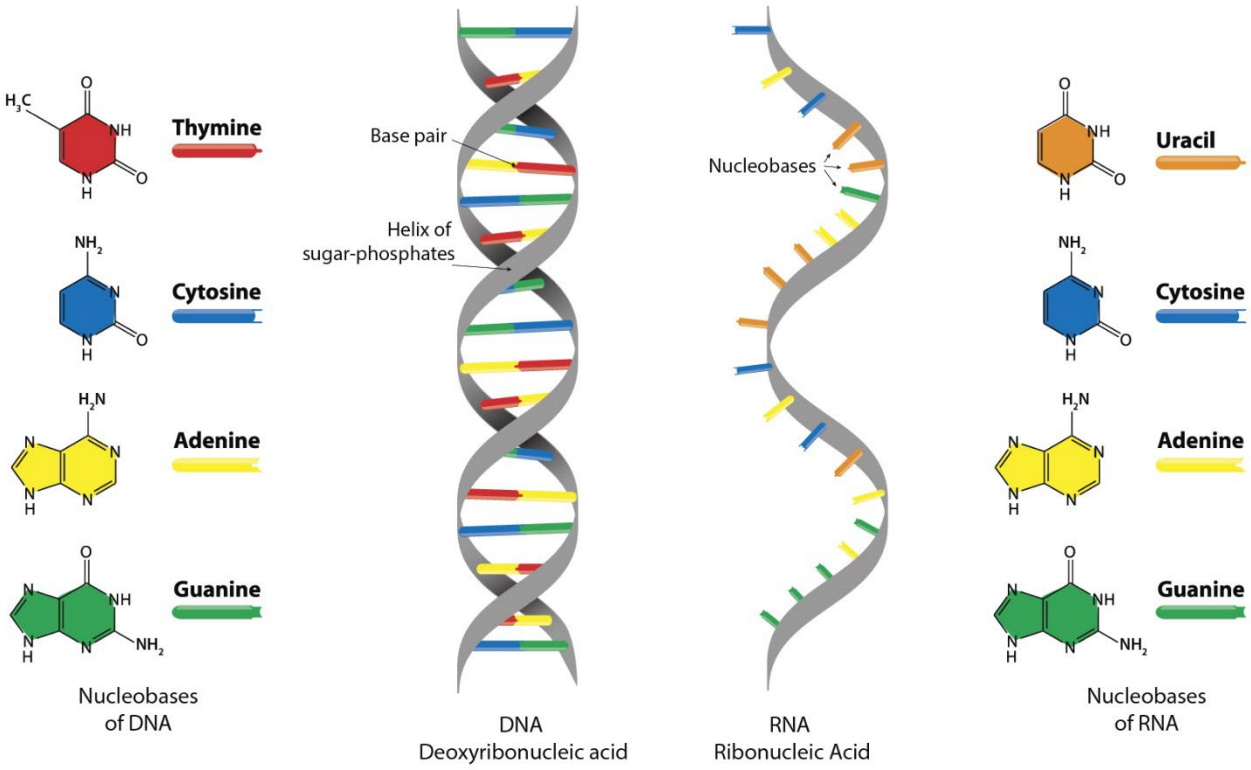
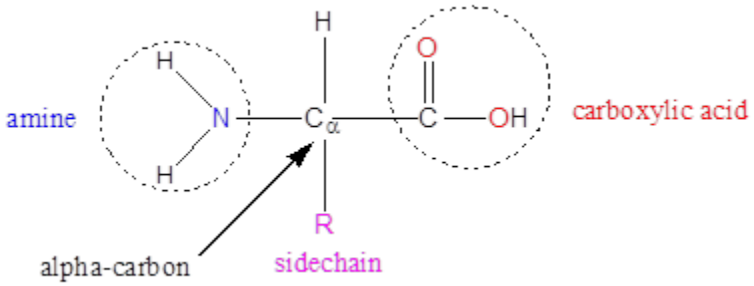


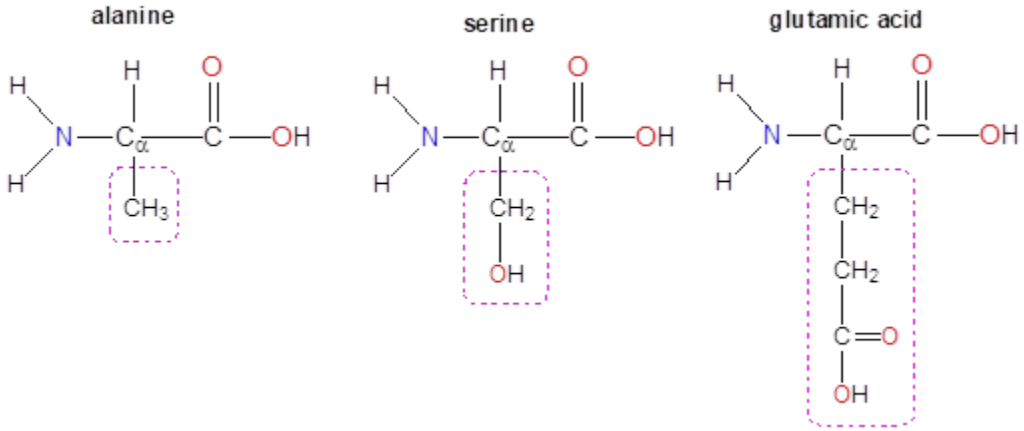
Figure 2: The structures of DNA and RNA including nucleobases.

The second step is called **translation** and is far more complex. The sequence of mRNA bases is used as a set of instructions to direct the synthesis of a **polypeptide**, composed of a specific sequence of **amino acids**. All proteins that perform essential functions within cells are polypeptide chains and differ in their sequence of amino acids. The relationship between the sequence of bases in mRNA and its corresponding amino acid is called the **genetic code**. We now know that it takes three mRNA bases – called a **codon** – to code for a particular type of amino acid.

There are twenty different types of amino acids commonly found within proteins and they share a general structure shown in *Figure 3(a)*. Each type of amino acids differs in the chemical composition of its **sidechain**, which is represented as **R**. These sidechains have varying structures and properties – some sidechains are simple hydrocarbons whereas other contain various functional groups. *Figure 3(b)* shows three different amino acids and highlights their sidechain structures.



(a)



(b)

Figure 3: (a) General structure of an amino acid. (b) The three amino acids shown differ in the chemical structures of their sidechains (highlighted in dashed boxes).

Table 1 lists all twenty amino acids according to their full name and three-letter symbol in alphabetical order. Although many of the names will be unfamiliar, they all represent variations on the chemical structure shown in *Figure 3(a)*. It is not necessary to memorize this list and the table can be used for reference during the laboratory exercise.

Table 1: Alphabetical listing of the 20 common amino acids

3 LETTER SYMBOL	FULL NAME
Ala	**alanine**
Arg	**arginine**
Asn	**asparagine**
Asp	**aspartic acid**
Cys	**cysteine**
Gln	**glutamine**
Glu	**glutamic acid**
Gly	**glycine**
His	**histidine**
Ile	**isoleucine**
Leu	**leucine**
Lys	**lysine**
Met	**methionine**
Phe	**phenylalanine**
Pro	**proline**
Ser	**serine**
Thr	**threonine**
Trp	**tryptophan**
Tyr	**tyrosine**
Val	**valine**

In order for scientists to "crack" the genetic code, they had to determine the rules that relate a particular set of three mRNA bases (a codon) to a particular amino acid. This was achieved through systematically making a variety of different mRNA sequences and examining what sequence of amino acids they produced via translation. In living cells this process occurs within the **ribosome**, a complex molecular factory for making proteins. With the aid of **transfer RNA (tRNA)**, the ribosome is able to form the amino acid chains that make up the structure of proteins. Scientists devised a way to mimic this process in the laboratory using "extracts" from bacterial cells that contains the necessary protein-synthesizing components. This enables translation to be carried out in a controlled environment and controlled without interference from other cellular processes. The mechanism behind protein synthesis in the cell is shown in *Figure 4* on the following page.

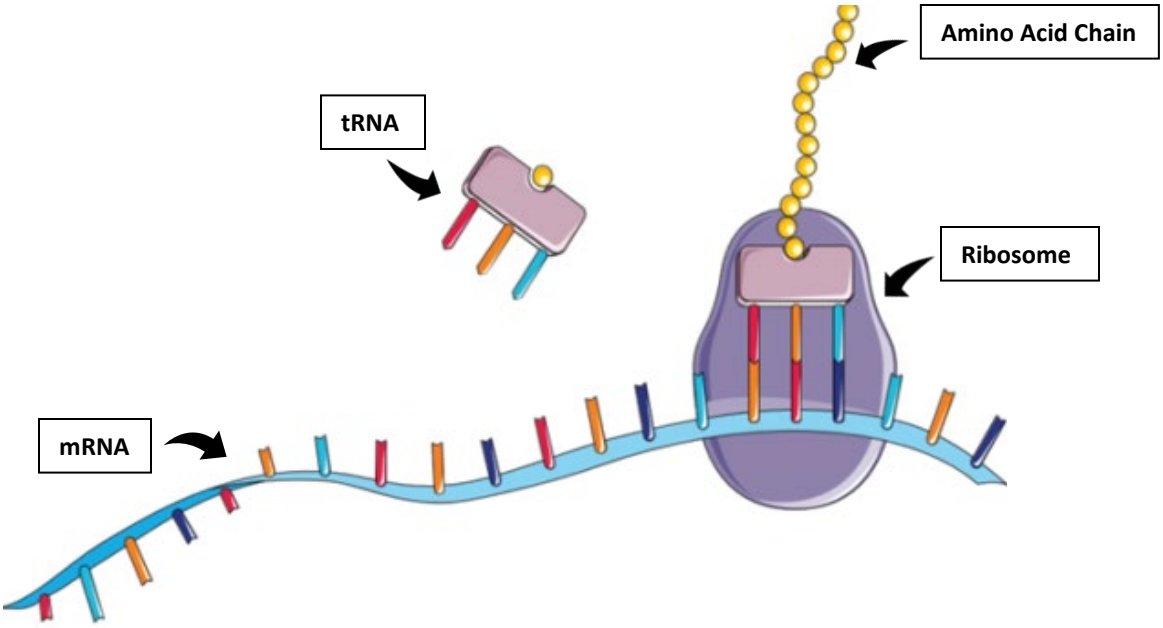


Figure 4: The process of protein creation from mRNA that occurs within the ribosome.

In this week's lab exercise, you will use a computer program to simulate the series of experiments that were used to crack the genetic code. Your simulations will take mere minutes, whereas the original experiments took many years. However, the basic principles of the process remain the same.

COVER SHEET
LABORATORY 6:
CRACKING THE GENETIC CODE

HUMAN GENETICS – CORE-UA 303 – Professor Fitch

Name: _____ Date: _____

Lab Partner Name(s): _____

_____

Laboratory Instructor: _____ Section: _____

CLEAN-UP CHECKLIST

It is essential that you clean up properly after the experiments are complete.

☐ Please ensure that the laptops are left with the Expasy Homepage open.

***NOTE**: Check with your lab instructor that your bench is <u>suitably clean</u> before leaving the lab.*

<u>Complete clean-up procedures</u> are worth <u>5 points</u> towards your lab score.

Lab Instructor Initials _____

GRADING	
Attendance:	_____ / 5
Quiz:	_____ / 10
Lab Project:	_____ / 30
Clean Up:	_____ / 5
Total:	_____ / 50

For TA use only

PROCEDURE AND DATA SHEETS

Part 1: Historical and Theoretical Background

During the late 1950s and early 1960s researchers were able to solve one of the major secrets of life: how genes encoded proteins. The problem the researchers were trying to solve was how a linear sequence of four nucleotides (A, G, C, and U) determined the amino acid sequence of proteins made up of twenty different amino acids. This simulation will allow you to reproduce some of the experiments that scientists used to determine how translation was accomplished. Using the web tool "Translate" from the Expasy Bioinformatics Resource Portal at https://web.expasy.org/translate/, your mission is to crack the "genetic code".

As having each nucleotide code for one amino acid would only allow for 4 different amino acids per protein (there are only 4 different kinds of nucleotides in mRNA), it was obvious to the researchers that there had to be some conversion between multiple nucleotides and each amino acid. Based on the assumption that nature has evolved to be as efficient as possible, researchers first deduced the minimum number of nucleotides in a "codon" that would be sufficient to provide sufficient information content to encode all 20 amino acids.

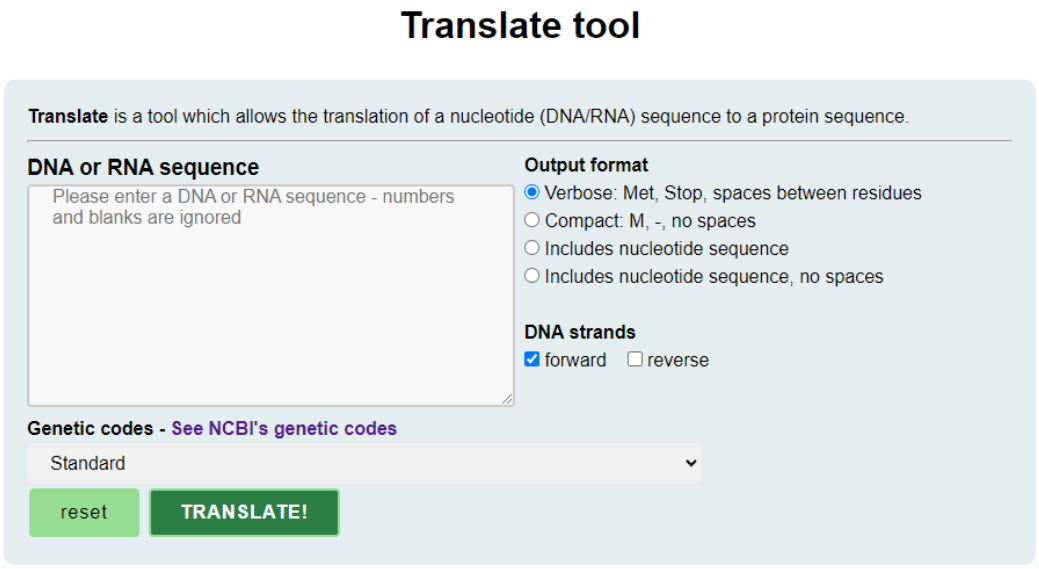
Question 1 [10 Points]: Would two nucleotides per codon be sufficient to provide enough different combinations/codons to encode all 20 amino acids? What about three-nucleotide codons? Four? Explain your reasoning.

A major step forward in figuring out the code was the discovery by Marshall Nirenberg in 1961 that a cell free extract made from *E. coli* cells could translate RNA added to the extract into proteins (without needing the AUG initiation codon required by intact cells). The newly synthesized protein compositions could be determined by measuring the incorporation of radioactive amino acids. In the first experiment he and his postdoctoral associate, Heinrich Matthaei (at the NIH at 3:00 AM, May 27, 1961) made poly-U RNA, using the enzyme polynucleotide synthetase, and found that it was translated into poly-phenylalanine (FFFFF...) using the cell-free extract. This was definitive proof that RNA could code for the synthesis of proteins and gave the first possible assignment of a nucleotide code to the amino acid it specified.

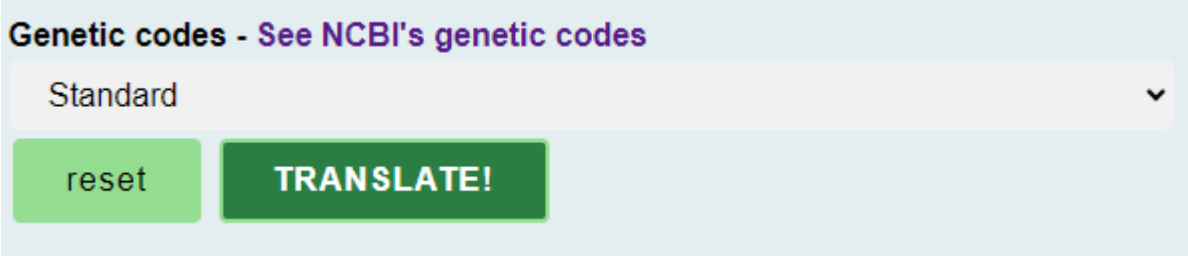
Reproduce this famous experiment using a computer simulation

First, make sure the Translate web site works for you:

1. Go to the following URL: https://web.expasy.org/translate/
2. Use the following settings:
    - For "Output format", select "Verbose: Met, Stop, spaces between residues".
    - For "DNA strands", select "forward" only (i.e., deselect "reverse").
    - For "Genetic codes", select the default "Standard".



3. Try entering a poly-U sequence containing at least 9 Uracil nucleotides (UUUUUUUUUU) into the text box under the header "DNA or RNA sequence
4. After entering your RNA sequence, click the TRANSLATE button.



5. The RNA sequence that you typed into the box is automatically reverse-transcribed to a DNA sequence, but that is okay! Just look at the Results of translation, which shows the results of different "reading frames".

Question 2 [30 Points]:

(a) What is a "reading frame"? Why are there three of them, labeled 1-3?

(b) This poly-U sequence actually represents the first experiment by Nirenberg and Matthaei. What do you deduce is the amino acid that corresponds to the triplet codon UUU?

(c) Design your simulation "experiment" to test what amino acid corresponds to AAA, CCC and GGG. What is your design, results and conclusions?

Part 2: Experiments to Infer the Code for at Least 3 Codons

While the Nirenberg experiments showed that RNA did determine the amino acids in the protein, they did not show how many bases were used for each codon, or whether the codes were overlapping or not: i.e., whether the second codon was read from the second or third bases [overlapping] or from the first base after the last base of the first codon [non-overlapping]. Nor could the experiments test if there could be bases in between the codons that did not encode anything (e.g., if the Gs in between AUC, AAC and ACA in the sequence AUCGGGCAACGGGACAGGGGG were not translated into amino acids, like using spaces to separate words in a sentence). Har Gobind Khorana (originally from a little village in Punjab that became a part of Pakistan in 1947—his biography is a fascinating and inspiring story!) developed a means to produce poly-dinucleotides and, later, poly-trinucleotides and poly-tetranucleotide sequences of DNA that could then be transcribed into RNA. These artificial RNA templates were translated in the cell free translation mix.

Question 3 [60 Points]: Your mission is to use a set of poly-dinucleotides (e.g. $(AC)_n$ or ACACAC...), poly-trinucleotides (e.g. $(AAC)_n$ = AACAACAACAAC...) and poly-tetranucleotides to "crack the code" for **three different codons** (that are also not AAA, CCC, GGG or UUU). Describe your experimental design, results and conclusions below.

(*Hint*: Using the results of experiments with poly-dinucleotides, e.g. $(AC)_n$, make a hypothesis for codon assignments. Then, try a similar poly-trinucleotide, e.g. $(CAC)_n$ to test your hypothesis. Did the results support or refute your codon assignment? Do you need to revise your hypothesis? Is there evidence that one of the amino acids can be encoded by more than one codon (i.e. the code is "degenerate")? What happens when you use a poly-tetranucleotide? Don't worry about any new amino acids that showed up in such an experiment, just use it to test your hypotheses from the previous experiments. From your set of experiments, infer conclusions with respect to the amino-acid assignments for 3 codons.)

Khorana, Holley and Nirenberg were awarded the Nobel Prize in Physiology or Medicine in 1968 for their work leading to the elucidation of the genetic code. Now you might have an idea for why they deserved it!