

## How Did They Figure Out The Genetic Code?

J. José Bonner; 1/10/01  
Modified 11/23/06

There are two scenarios here; some student groups should get one, other student groups should get the other. Then, challenge the students to learn what they can from these scenarios--as a competition. At the end, compare the results of the two groups, building a table of the Genetic Code. Then, give them the data from Set III, and add the results to the table.

After students have wrestled with the data, and after you have together built a table with a few of the codons identified, give them the entire table (shown below). The data from Sets I, II, and III give students the basic information: there are three bases per codon, sometimes more than one codon specifies a single amino acid, and some codons do not correspond to amino acids at all (but are, in fact, "stop" codons that trigger termination of translation). Typically, these three features of the Code are confusing. However, working through the data, and discovering these features themselves, will give students a sense of "ownership" of their discoveries and lessen their confusion. [They still will not know *why* it works this way--none of us do--but they will know *that* it does, and they will know the data upon which these conclusions are based.] When, in the next session, they begin to hear about the mechanism of translation, with tRNAs and ribosomes, they'll know there's a code, that it uses 3-base codons, and that the tRNAs provide a simple explanation for these observations.

		2nd position								
		U		C		A		G		
1st position	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	3rd position
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	
		UUA	Leu	UCA	Ser	UAA	STOP	UGA	STOP	
		UUG	Leu	UCG	Ser	UAG	STOP	UGG	Trp	
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	
		AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	
		AUA	Ile	ACA	Thr	AAA	Lys	AGA	Ser	
		AUG	Met	ACG	Thr	AAG	Lys	AGG	Ser	
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly		
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly		
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly		
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly		

3-letter "word"  
corresponding amino acid

## Set I

In science, nearly everything is a race among different laboratories, each one hoping to make the next breakthrough before anyone else does. Today, in early 1961, the race is about the Genetic Code. In just the last few years, we've learned the structure of DNA, and the fact that genetic information *must* be "stored" in DNA as the sequence of bases. We've learned that the cytoplasmic enzymes that build protein according to the DNA code do not "read" the DNA itself. There is some kind of intermediate. The best candidate for this intermediate, or "messenger" is RNA. Now, the question is: how does it work? What *is* the code? Whoever cracks the code first wins.

You are a researcher in Marshall Nierenberg's laboratory. You've figured out a way to determine the genetic code. You have:

- bacterial cells, which you can grind up into a kind of concentrated soup. We call this a "cell-free" system, because it contains all of the material that is inside cells, but has no complete (unbroken) cells in it.
- synthetic RNA molecules, which you have made in the lab with specific sequences.
- some radioactive amino acids (a mix of all 20)
- methods by which you can:
  - separate different proteins on the basis of their chemical properties
  - break proteins apart completely into individual amino acids
  - remove one amino acid at a time from the ends of proteins
  - separate and identify different amino acids

To figure out the code, you repeat the following experiment with each RNA molecule:

1. add one synthetic RNA to the cell-free system.
2. add the radioactive amino acid mix.
3. wait for the cell-free system to build protein, using your synthetic RNA as a "template."
4. separate different proteins from the reaction mix, and determine which (radioactive) amino acids were used to build protein.

With the first few RNA molecules that you use, you obtain the following data:

Experiment #	RNA used in reaction	Protein produced by cell-free system
1	UUUUUUUUUUUUUUUU...	Phe-Phe-Phe-Phe-Phe...
2	CCCCCCCCCCCCCCCC...	Pro-Pro-Pro-Pro-Pro...
3	UCUCUCUCUCUCUCU...	Leu-Ser-Leu-Ser-Leu...
4	UCCUCCUCCUCCUCC...	Ser-Ser-Ser-Ser-Ser... <i>and</i> Pro-Pro-Pro-Pro-Pro... <i>and</i> Leu-Leu-Leu-Leu-Leu...

Data Analysis for Set I:

1. What are all the possible codes for Phe?
2. What are all the possible codes for Pro?
3. How does experiment #3 help you decide among these choices?
4. In experiment #3, how many different RNA codes were used by the cell-free system?  
What are they?
5. How do you explain the results of experiment #4?
6. Make a table of the different RNA codes you have discovered--i.e., what RNA sequence codes for what amino acid.

## Set II

In science, nearly everything is a race among different laboratories, each one hoping to make the next breakthrough before anyone else does. Today, in early 1961, the race is about the Genetic Code. In just the last few years, we've learned the structure of DNA, and the fact that genetic information *must* be "stored" in DNA as the sequence of bases. We've learned that the cytoplasmic enzymes that build protein according to the DNA code do not "read" the DNA itself. There is some kind of intermediate. The best candidate for this intermediate, or "messenger" is RNA. Now, the question is: how does it work? What *is* the code? Whoever cracks the code first wins.

You are a researcher in Fred Hochstein's laboratory. You've figured out a way to determine the genetic code. You have:

- bacterial cells, which you can grind up into a kind of concentrated soup. We call this a "cell-free" system, because it contains all of the material that is inside cells, but has no complete (unbroken) cells in it.
- synthetic RNA molecules, which you have made in the lab with specific sequences.
- some radioactive amino acids (a mix of all 20)
- methods by which you can:
  - separate different proteins on the basis of their chemical properties
  - break proteins apart completely into individual amino acids
  - remove one amino acid at a time from the ends of proteins
  - separate and identify different amino acids

To figure out the code, you repeat the following experiment with each RNA molecule:

1. add one synthetic RNA to the cell-free system.
2. add the radioactive amino acid mix.
3. wait for the cell-free system to build protein, using your synthetic RNA as a "template."
4. separate different proteins from the reaction mix, and determine which (radioactive) amino acids were used to build protein.

With the first few RNA molecules that you use, you obtain the following data:

Experiment #	RNA used in reaction	Protein produced by cell-free system
1	GGGGGGGGGGGGGGG...	Gly-Gly-Gly-Gly-Gly...
2	CCCCCCCCCCCCCCC...	Pro-Pro-Pro-Pro-Pro...
3	GCGCGCGCGCGCG...	Ala-Arg-Ala-Arg-Ala...
4	GCCGCCGCCGCCGCC...	Ala-Ala-Ala-Ala-Ala... <i>and</i> Pro-Pro-Pro-Pro-Pro... <i>and</i> Arg-Arg-Arg-Arg-Arg...

Data Analysis for Set II:

1. What are all the possible codes for Gly?
2. What are all the possible codes for Pro?
3. How does experiment #3 help you decide among these choices?
4. In experiment #3, how many different RNA codes were used by the cell-free system?  
What are they?
5. How do you explain the results of experiment #4?
6. Make a table of the different RNA codes you have discovered--i.e., what RNA sequence codes for what amino acid.

### Data Set III

Experiment #	RNA used in reaction	Protein produced by cell-free system
1	UUUUUUUUUUUUUUU...	Phe-Phe-Phe-Phe-Phe...
2	AAAAAAAAAAAAAAAAA...	Lys-Lys-Lys-Lys-Lys...
3	AUAUAUAUAUAUAUA...	Ile-Tyr-Ile-Tyr-Ile-Tyr...
4	AAUAAUAAUAAUAAU...	Asn-Asn-Asn-Asn-Asn-... <i>and</i> Ile-Ile-Ile-Ile-Ile-Ile-...

#### Data Analysis for Set III:

This is as straightforward as Data Set I and Data Set II until we get to experiment #4. In sets I and II, experiment #4 gave three different proteins. Here it gives only two. What does this tell us?

#### Extension:

Now, perhaps, we are ready to look at the complete table of the Genetic Code. We understand that it was experiments like these that led Marshall Nierenberg's lab to figure out the Genetic Code. We don't need to go through all of the experiments, or understand all of the techniques; nonetheless, working through the logic and the data-analysis gives us a pretty good sense of what was involved.

#### Some interesting notes:

When we present protein synthesis in a lecture that describes what happens, students are left with many puzzles. One of the big ones is "why 3 bases *per* codon?" The teaching approach presented here *doesn't answer this question*. Rather, it illustrates the fundamental nature of science, and the generic answer to all such "why" questions. We can ask, by means of experimentation, what the genetic code is; we can obtain the answer to this question. We can describe what we find; we cannot address *why* we didn't find something else. In general:

**Science can address *how* the world works. It cannot address *why* it works that way.**

Once we know the genetic code, and the fact that ribosomes "read" the code with the help of tRNA molecules that form base pairs with the codons of the mRNA, we *can* address the next question students often ask. They often wonder why ribosomes "walk" down mRNA in steps of three bases. Here, answering *how* ribosomes function gives us the answer to this "why" question. Once a ribosome has ejected a "used" tRNA, no longer carrying an amino acid, the ribosome moves down the mRNA by re-adjusting its position relative to the tRNA that remains H-bonded to the mRNA. Essentially, the ribosome moves the width of one tRNA--one codon, which happens to be 3 bases. When we speak of ribosomes moving "3 bases at a time," we imply that ribosomes can count, and have some kind of mysterious intelligence. When we speak of ribosomes moving "one tRNA width" at a time, "counting" becomes unnecessary; it's just a clockwork mechanism. An unintelligent clockwork mechanism is easier to understand.