



# Cracking the genetic code: replicating a scientific discovery

Image courtesy of Caroline Davis2010; image source: flickr

By Jordi Domènech-Casal

In 1953, James Watson and Francis Crick discovered the structure of DNA, the molecule that carries our genetic information. In 1958, Crick postulated the central dogma of molecular biology: that the flow of information goes from DNA to RNA to protein. But the question remained: how did the four-letter alphabet of nucleotides in DNA (A, C, T and G) or its equivalent in RNA (A, C, U and G) encode the 20-letter alphabet of amino acids that build our proteins? What was the genetic code?

In 1961, Marshall W Nirenberg and Johann H Matthaei deciphered the first letter of the code, revealing that the RNA sequence UUU encodes the amino acid phenylalanine. Subsequently, Har Gobind Khorana showed that the repeating nucleotide sequence UCUCUCUCUCUC encodes a strand of amino acids reading serine-leucine-serine-leucine. By 1965, largely due to the work of Nirenberg and Khorana, the genetic code had been completely cracked. It revealed that each group of three nucleotides (known as codons) encodes a specific amino acid, and that the order of the codons determines the order of amino acids in (and, consequently, the chemical and biological properties of) the resulting protein.



REVIEW

- ✓ Biology
- ✓ Chemistry
- ✓ Maths
- ✓ Protein synthesis
- ✓ Ages 16–19

This article offers a strategy that helps teachers to explore, simply and accessibly, one of the most challenging aspects of science teaching: helping their students to appreciate and understand how science actually works. Acquiring knowledge requires scientists to ask good questions, design and carry out good experiments, and work together to address uncertainty. This is exactly what the students need to do in this activity to crack the genetic code.

I anticipate that teachers of disciplines other than biology (particularly maths and chemistry) would also find this article useful. It would also be a very good activity to use during a science fair.

Betina Lopes, Portugal

**i**  
**Background**

## How did Nirenberg and Khorana crack the genetic code?

Nirenberg and Khorana compared short sequences of the nucleic acid RNA and the resulting amino acid sequences (peptides). To do this, they followed the protocol that Nirenberg developed with Matthaei.

This involved artificially synthesising a specific sequence of RNA nucleotides and mixing it with extracts of *Escherichia coli* bacteria that contained ribosomes and other cellular machinery necessary for protein synthesis. The scientists then prepared 20 samples of the resulting mixture; to each sample, they added one radioactively labelled amino acid and 19 unlabelled amino acids, then allowed protein synthesis to occur. Each of the 20 samples contained a different radioactively labelled amino acid. If the resulting peptide was radioactive, it indicated that the radioactively labelled amino acid was included, confirming that the RNA nucleotide sequence coded for this amino acid at some point.

By repeating this experiment with different RNA sequences, more and more information could be gathered about the genetic code. After simple sequences such as UUUUUU and AAAAAA had been tested, further teams of scientists took up the challenge, analysing more complex RNA sequences, eventually allowing all 64 codons to be de-coded.

The genetic code itself is a crucial element of biology lessons, providing a molecular explanation of the actions of genes (for example, in mutation, evolution and gene expression). Furthermore, the way in which Nirenberg and Khorana cracked the genetic code – by comparing short sequences of RNA with the resulting amino acid sequences – can be re-run as an inquiry-based teaching activity at school. Using the sequences provided by the teacher, the students work in teams to:

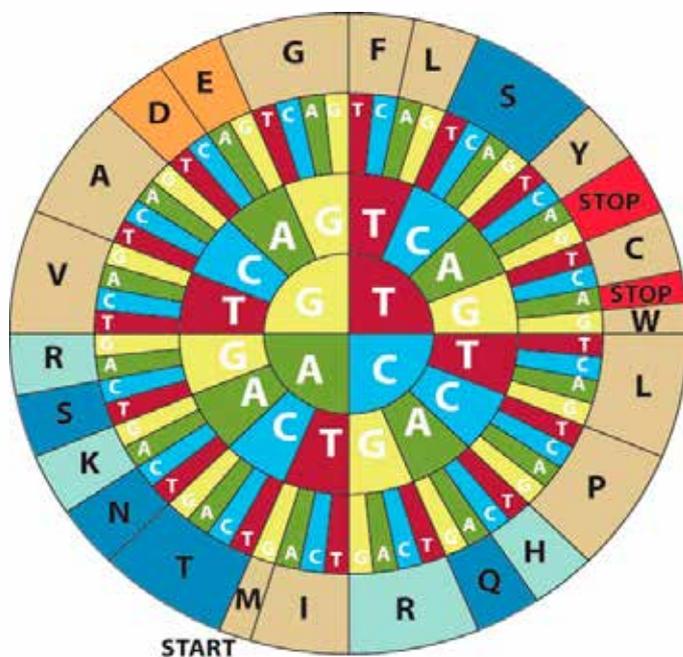
- Identify patterns
- Construct hypotheses and explanatory models
- Design experiments
- Reach conclusions from partial data
- Establish the strength of their conclusions
- Communicate and justify their conclusions in a scientific manner.

The activity thus offers a model for teaching the nature of scientific knowledge: a provisory consensus constructed by the community with conclusions of diverse strength based on partial evidence.

## Cracking the code in the classroom

This activity is suitable for 14- to 18-year-old students working in teams of 3–4, and takes about two hours, divided into four steps plus a final discussion. It is designed as an introduction to molecular biology, before you explain anything about the genetic code or the central dogma of molecular biology.

Students are asked to crack a code composed of different sequences of letters (A, C, T, G) using the messages that those sequences encode (e.g. AspHisTrp...). In each of the first three steps, each team is given a different set of letter sequences and corresponding messages. At each step, they will need to re-evaluate their conclusions from the previous steps, and modify their solution to the code.



Amino acid code			
A - Alanine	G - Glycine	M - Methionine	S - Serine
C - Cysteine	H - Histidine	N - Asparagine	T - Threonine
D - Aspartic Acid	I - Isoleucine	P - Proline	V - Valine
E - Glutamic acid	K - Lysine	Q - Glutamine	W - Tryptophan
F - Phenylalanine	L - Leucine	R - Arginine	Y - Tyrosine

Image courtesy of Cath Brooksbank EMBL

Figure 1: The genetic code. To decode a DNA sense codon, find the first letter of your sequence in the inner circle and work outwards to identify the corresponding amino acid. For example, CAT codes for H (histidine).

Explain that all the groups will be working to crack the same code, using different examples. Do not tell your students about the biological nature of the sequences (DNA and amino acids); they should focus on finding patterns and relationships.

Nirenberg and Khorana used RNA sequences to crack the code; in contrast, this activity uses DNA sequences (sense codons, 5' to 3'). The crux of the activity is the existence of the code rather than the details of transcription and translation, which can be addressed in subsequent lessons.

After each step, you may ask one student from each team to join a different team. (This mimics the dynamics of how scientific knowledge is acquired and shared, for example at conferences or through publications.)

Otherwise, teams may exchange information only when they are told to do so. (If one team gets stuck and discouraged, it can be more motivating to ask another team to help them rather than the teacher.)

## Materials

- Worksheets 1-4 for each team, which can be downloaded from the *Science in School* website<sup>w1</sup>. The sequence sets are different for each team.
- Figure 1 or a smartphone app for easily converting DNA codons to amino acids<sup>w2</sup>.

## Procedure

Allow at least 10-15 minutes for your students to discuss each step. When all the teams feel that they have obtained all the possible information from their sequences, move on to the next step.

**1. Detecting frames.** Give each team a copy of worksheet 1, which contains three sequences that do not contain synonym codons or stop codons. All the sequences begin with an ATG codon, encoding the amino acid methionine (Met).

Using the three sequences, the students should be able to establish that the code is based on triplets of letters and to make their first

Sequence	Message	Students discover that...
ATGTTAGGTAGTAAAGATGCT	MetLeuGlySerLysAspAla	The code is based on triplets and each triplet represents one of the three-letter elements, e.g. Met.
ATGCATGAAGCTATTTATGAT	MetHisGluAlaIleTyrAsp	
ATGGGTAGTGATGAAGCTTAT	MetGlySerAspGluAlaTyr	

Table 1: An example of worksheet 1

Sequence	Message	Students discover that...
ATGGTTTCGTACTACTGCGTCA	MetValSerTyrThrAlaSer	Some elements can be encoded by more than one triplet, e.g. Ser.
ATGCCGTACACATGTGTACACA	MetProTyrThrCysValThr	
ATGACGAGTGC GTTGTGCGAT	MetThrSerAlaLeuCysAsp	

Table 2: An example of one team's sequences for step 2

Sequence	Message	Students discover that...
TGTCATGCATCCGTCATCACTGAC	–	The ATG triplet determines the beginning of the message and the TGA triplet its end.
TGCGTGACTATGGACACAGTCGT	MetAspThrVal	
ATGTGTCGATGACTGATCATG	MetCysArg	
ATGTGCGTACACATTTGAGTC	MetCysValHisIle	
ATGCTGTACACATGATGCACAGT	MetLeuTyrThr	

Table 3: An example of one team's sequences for step 3

hypotheses about the meaning of some of these triplets.

**2. Building a model.** Give each team a copy of worksheet 2, which contains three new sequences, some of which include synonym codons.

The students should be able to confirm some of their hypotheses from step 1, while other hypotheses may be cast into doubt.

**3. Adjusting the model to new evidence.** Give each team a copy of worksheet 3, which contains new sequences presenting more complexity: some sequences lack the initial ATG codon, some have it further into the sequence, and some have a stop codon. These characteristics either result in

messages that are shorter than the seven-amino-acid sequences in the previous steps, or produce no message at all.

The worksheets for this step each contain two lists of sequences. You can choose whether to give your students all the sequences together (to make this step easier) or in two separate sub-steps (to make it harder).

Besides confirming the triplets that encode some amino acids, these sequences allow the students to identify the key roles of the methionine (start) and stop codons.

**4. Testing hypotheses and designing experiments.** The students should now be able to propose a partial

solution to the code. To test their hypotheses, give each team a copy of worksheet 4 and ask them to design an experiment. They should propose changes to four specific sequences that they were given in the previous steps, and note any change that they would expect to the message. You then give them the correct message, using figure 1 if necessary. Was the result what they expected? If not, what does that tell them? This mimics a rapid process of hypothesising, designing experiments and analysing results.

As a conclusion to the activity, each team should present their partial solution for the code to the rest of the class, justifying their conclusions. Those parts of the code that are accepted by the rest of the group – representing the scientific community – should be written on the blackboard. Controversial or unclear parts should also be noted. The result will be a consensual, partial genetic code.

Avoid confirming immediately if the code that your students have constructed is correct. Explain that in science, there is no book to compare your results to, and the only way to find if something is correct is by asking good questions, designing good experiments, and sharing information and ideas with your peers to gain consensus.

## Discussion

Ask your students to consider the following questions:

- How did you discover what you now know?
- Did you discuss your ideas with anyone during the activity? What did you discuss?
- How did you confirm whether your hypotheses were accurate?
- Did you reject any of your initial hypotheses? Which ones?
- How did you resolve any contradictory conclusions within or between teams?
- Were all your conclusions equally strong?

Image courtesy of Jordi Domènech-Casal Adaptations



Students comparing new data with their conclusions from previous steps

After the discussion, explain to your students that the sequences were DNA and amino acid sequences, and that they have just reproduced a key experiment in molecular biology. Your students should now be motivated to learn more about the genetic code and the central dogma of molecular biology, including how similar their activity was to the way in which the genetic code was really cracked.

You could recap the activity, reminding your students what they discovered for themselves:

- In step 1, that the genetic code is based on triplets of nucleotides (codons).
- In step 2, that the genetic code is redundant but not ambiguous: each codon encodes a single element (e.g. an amino acid), but some elements are encoded by more than one codon.

Image courtesy of Duncan Hull; image source: flickr



- In step 3, that the code includes start and stop codons to specify the beginning and the end of the encoded amino acid sequences.

(Note that the activity could generate the misconception that proteins are usually composed of six or seven amino acids, so this may need to be addressed.)

Explain that the way your students have been working, in collaborative and / or competitive teams, with the membership of the teams changing, and information being shared with other teams, reflects the way that scientists work in real life.

To make the activity easier, you could give your students more sequences in each step (e.g. the sequence sets for two teams). Alternatively, you could leave out step 3, and simply explain the role of the start and stop codons after the activity.

## Acknowledgements

Pedagogic reflections on the activity described in this article are part of the work of the language and context in science education (llenguatge i contextos en educació científica, LICEC) research group at the Autonomous University of Barcelona (reference 2014SGR1492), financed by the Spanish Ministry of Economics and Competitiveness (reference EDU2015-66643-C2-1-P).

## Web references

- w1 Download the worksheets for the activity from the *Science in School* website. See: [www.scienceinschool.org/2016/issue36/code](http://www.scienceinschool.org/2016/issue36/code)
- The worksheets are also available from the author's website: <https://sites.google.com/a/xtec.cat/hacking-the-code>
- w2 The Nobel Prize website has a table to translate the codons into amino acids. See: [www.nobelprize.org](http://www.nobelprize.org) or use the direct link: <http://tinyurl.com/jkqz9vz>

## Resources

This activity is part of the C3 science education project, which develops inquiry problem-based learning activities to cover the science curriculum. See: <https://sites.google.com/a/xtec.cat/c3/home>

Other English-language activities in the project address plate tectonics, mitosis and cancer, human evolution, phylogeny, genetic heredity, gene expression and ecosystem

Image courtesy of Daniel; image source: Flickr



Complementary nucleotide base pairs

dynamics. See: <https://sites.google.com/site/proyectandobiogeo>

See also:

Domènech-Casal J (2013) Hacking the code: una aproximació indagadora a l'ensenyament del codi genètic, o seguint les passes de Nirenberg i Khorana. *Ciències: revista del professorat de ciències de primària i secundària* **25**: 20-25

Domènech-Casal J (in press) *Proyectando BioGeo, un itinerario de trabajo por proyectos contextualizados basado en la indagación y la Naturaleza de la Ciencia. Alambique, Didáctica de las Ciencias Experimentales*

Nirenberg M, et al (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the USA* **53(5)**: 1161-1168

The article can be downloaded free of charge from the Pubmed Central website. See: [www.ncbi.nlm.nih.gov/pmc](http://www.ncbi.nlm.nih.gov/pmc) or use the direct link: <http://tinyurl.com/z8fyqx9>

Read the story of 'How the code was cracked' on the Nobel Prize website. See: [www.nobelprize.org](http://www.nobelprize.org) or use the direct link: <http://tinyurl.com/h4j90lf>

Image courtesy of Francisco Gonzalez; image source: Flickr



In 1968, Marshall W Nirenberg, Har Gobind Khorana and Robert W Holley were awarded the Nobel Prize in Physiology or Medicine 'for their interpretation of the genetic code and its function in protein synthesis'. Details of their work are described in the presentation speech. See: [www.nobelprize.org](http://www.nobelprize.org) or use the direct link: <http://tinyurl.com/z9vsyqs>

Francis Crick, James Watson and Maurice Wilkins were awarded the 1962 Nobel Prize in Physiology or Medicine 'for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material'. See: [www.nobelprize.org](http://www.nobelprize.org) or use the direct link: <http://tinyurl.com/hozaohv>

In 1970, Francis Crick described how the central dogma of molecular biology was developed.

Crick F (1970) Central dogma of molecular biology. *Nature* **227**: 561-563. doi:10.1038/227561a0

Many papers by Crick are freely available on the *Nature* website. See: [www.nature.com](http://www.nature.com)

Jordi Domènech-Casal has a PhD in biology and worked for eight years as a researcher in genetics and inorganic chemistry at the universities of Barcelona, Spain; Bologna, Italy; and Paris VII, France. He currently works as a science teacher at the Institut de Granollers, a secondary school in Barcelona, and as a teacher trainer and advisor in science education for the Catalan government and on the

European project 'Engaging science'. He is a member of the language and science education (*llenguatge i ensenyament de les ciències, LIEC*) research group at the Autonomous University of Barcelona, where his interests focus on the role of language and context in inquiry- and project-based learning activities in science education.

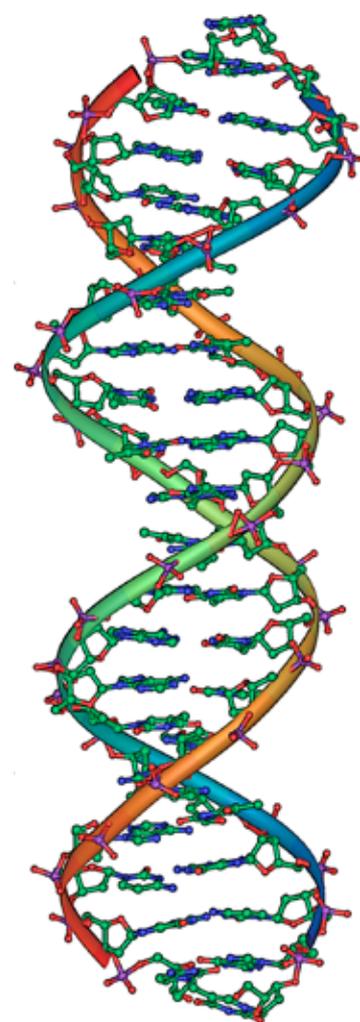


Image courtesy of mstroek / Wikimedia Commons