Elucidation of the Genetic Code

4 major advances helped figure out the code

1) Genetic mutations alter protein sequences
2) The demonstration of colinearity between genes and protein
3) The idea of triplet codons
4) Deciphering the genetic code (UUU= Phe)

What is the general nature of the code?

<table>
<thead>
<tr>
<th>mRNA (nucleotides)</th>
<th>4 different types (A,C,G,U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein (amino acids)</td>
<td>20 different types</td>
</tr>
</tbody>
</table>

- Clearly there cannot be a 1:1 correspondence in the code
- Therefore nucleotides must be read in groups!
- Would reading of pairs be enough? 4 X 4 = 16 different combinations
  - Not enough to specify 20 amino acids!
- How about reading in groups of 3? 4 X 4 X 4 = 64 combinations
  - More than enough to code for 20 amino acids
- Therefore the concept of a triplet codon (coding unit) was born!

(George Gamow – of the Big Bang Theory. Ralph Alphe with Hans Bethe)
How is the code organized: Overlapping? Punctuated?

- Is the code overlapping or non-overlapping?
- Is the code punctuated or non-punctuated (continuous)?

If code is overlapping:

Sickle cell anemia

Inherited in semidominant fashion in African Americans.

Affects red blood cells which have very few proteins

Linus Pauling

Sickle cell anemia affects hemoglobin

Hemoglobin run out on a protein gel

Figure 1. Fingerprints of hemoglobins A and S (improved method); photograph of ninhydrin-positive peptide spots on filter paper (BAGLIONI 1961)

Vernon Ingram

Single amino-acid substitutions occur in sickle cell hemoglobin!

Ingram, V. M. Genetics 2004;167:1-7
Single aminoacid changes in β-globin cause Sickle cell anemia

- Argues very strongly against an overlapping genetic code

Tryptophan synthetase gene: many mutations mapped and ordered.

Purified mutant protein and identified the position of every a.a. substitution!

Charles Yanofsky, 1964
The first experiments: translation of homopolymeric synthetic mRNAs

Hypothesis: a code based on RNA sequence specifies amino acid sequence
- Use the enzyme polynucleotide phosphorylase to make homopolymeric RNA

But is it a triplet code?
- This was shown by Francis Crick Sydney Brenner and colleagues (mid-1960s)

Experimental strategy: use a chemical mutagen to create 'point mutations' in a coding sequence (gene):

Either
- (i) Insert one additional base
- Or (ii) Delete a single base

After making the mutants, examine function of the proteins coded by the mutant sequences in isolation and in combinations created by recombination

Experimental proof that the code is a triplet code
- Mutants in which a single base had either been inserted or deleted

Imagine the sequence -CAT- repeated many times (CAU in RNA code):

CATCATCATCATCATCATCATCATCATCATCAT……..  

it would code for:

HIS HIS HIS HIS HIS HIS HISHISHISHISHISHISHISHISHISHISHIS……..

Imagine that a 'G' is inserted:

CATCATCATCATCATCATGATCATCATCATCATCATCATCATCATCATCAT……..  

it would now code for:

HIS HIS HIS HIS HIS HIS ALA SER SER SER……..

Result: All the amino acids 'downstream' of the inserted base are changed

Base insertion mutants indicate a continuous code (and more)

If the code is interrupted by punctuation:

As we shall see, insertion of a single base changes all the following amino acids

If the code is continuous:

Conclusion: The code is continuous, without spaces or internal punctuation

CONCLUSION: UUU is an mRNA codon for phenylalanine.  
AAA is an mRNA codon for lysine.  
CCC is an mRNA codon for proline.
Crick reasoned:

If the reading frame is based on triplets and 2 more single base additions are introduced into the mutant, the reading frame should be restored.

Imagine that we add an A and a C to the original mutant:

```
CATCATCATCATCATCATCAGCATCATCATCAT
```

...it would now code for:

```
HIS HIS HIS HIS GLN SER LEU METHIS HIS HIS HIS.....
```

We have restored the reading frame after MET and are back to decoding HIS, thus proving that the code consists of triplet codons.

Although the protein has been altered, some proteins can tolerate such small changes and function almost normally (pseudowild).

Summary of Crick’s results:

<table>
<thead>
<tr>
<th>Insertion</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>One bp insertion</td>
<td>Mutant</td>
</tr>
<tr>
<td>Two bp insertion</td>
<td>Mutant</td>
</tr>
<tr>
<td>Three bp insertion</td>
<td>Pseudowild</td>
</tr>
</tbody>
</table>

Likewise:

<table>
<thead>
<tr>
<th>Deletion</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>One bp deletion</td>
<td>Mutant</td>
</tr>
<tr>
<td>Two bp deletion</td>
<td>Mutant</td>
</tr>
<tr>
<td>Three bp deletion</td>
<td>Pseudowild</td>
</tr>
</tbody>
</table>

Also:

<table>
<thead>
<tr>
<th>Insertion and Deletion</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>One bp insertion and one deletion</td>
<td>Pseudowild</td>
</tr>
</tbody>
</table>

Conclusion: Since only triple mutants were pseudowild, the code must be organized as triplet codons.

The experimental assault on the code: Phase 2

The 2nd phase: translation of mixed composition synthetic mRNAs

Polynucleotide phosphorylase can also be used to make synthetic mRNA mixed co-polymers (e.g. poly(U,G))

Pioneered by Severo Ochoa (Nobel prize: 1959)

```
Poly (U,G): UGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGU
The 3rd phase: investigations using codon-sized RNAs (trinucleotides)

- Nirenberg and Leder cracked the entire code except STOP codons in 1964
- They showed that a synthetic RNA corresponding to a single codon (trinucleotide) could direct the binding of specific aminoacyl-tRNAs to ribosomes in vitro
- By using $^{14}$C-labelled amino acids with all possible trinucleotide codons they showed that 61 (of the 64 possible) codons could code for the 20 amino acids
- Therefore a given amino acid can be coded for by more than 1 codon

i.e. the genetic code is degenerate

Determination of all possible matches between codon and amino acid

The RNA triplet and a matching aminoacylated tRNA bind to the ribosome, where they form a large ‘ternary’ complex that can’t pass through a filter

- The Nirenberg and Leder strategy:
  - For each synthetic triplet RNA set up 20 reactions; only one contains a specific $^{14}$C-labelled amino acid
  - Ask which reaction causes radioactivity to be retained on the filter
  - That’s the one encoded by the triplet codon!

Nonsense mutations led to discovery of the STOP codons

The study of ‘Nonsense’ mutations led to the discovery of the codons that act as STOP codons

- Amber mutation causes premature termination of translation i.e. ‘STOP’ signal
- ‘STOP’ signal is suppressed

Result: A shortened, inactive form of the phosphatase protein is synthesized

- The full-length form of the phosphatase protein is synthesized

Conclusion: the triplet GUU specifies (codes for) the amino acid Valine
### Summary: features of the Genetic Code

- All the codons have meaning: 61 specify amino acids; the other 3 are ‘nonsense’ or STOP codons.
- The code is **unambiguous** – only one amino acid is indicated by each codon.
  - The code is **degenerate**: except for Trp and Met, each amino acid is encoded by two or more codons. 
    - e.g. GGA, GGC, GGG, GGT – each code for Glycine (4-fold degeneracy).
- Codons representing the same or similar amino acids are often similar in sequence.
- Where 2nd base is a pyrimidine (C, U): usually codes for a nonpolar amino acid.
- Where 2nd base is a purine (A, G): usually codes for a polar or charged amino acid.

### Verification of the code in vivo

- Compare amino acid sequences of wild-type and mutant proteins.
- The great majority can be explained by single base changes – exactly as predicted by the in vitro code.
- Of particular importance was evidence from analysis of Hemoglobin α-chain variants.
  - Normal: Val Lys Gly His Gly Lys
  - Mutant 1: Val Lys Asp His Gly Lys Hb Norfolk
  - Mutant 2: Val Lys Gly Tyr Gly Lys HbM Boston
- Changes in amino acid sequence are consistent with the Code:
  - Gly ➔ Asp: GGU ➔ GAU
  - His ➔ Tyr: CAU ➔ UAU

### Codon usage and variations in the genetic code

<table>
<thead>
<tr>
<th>Codon usage and variations in the genetic code</th>
<th>Codon usage and variations in the genetic code</th>
</tr>
</thead>
</table>

The Vertebrate Mitochondrial Code (species specific)

<table>
<thead>
<tr>
<th>Code 2</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGA</td>
<td>Ter</td>
</tr>
<tr>
<td>AGG</td>
<td>Ter</td>
</tr>
<tr>
<td>AUA</td>
<td>Met M</td>
</tr>
<tr>
<td>UGA</td>
<td>Trp W</td>
</tr>
</tbody>
</table>

The table shows codon usage and variations in the genetic code, including the vertebrate mitochondrial code and the standard codons for the 20 amino acids.