Last Tuesday...

Nirenberg and Matthaei: RNA is the template for protein synthesis (poly-U —> phenylalanine)

Thursday!

CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS*

By J. Heinrich Matthaei,† Oliver W. Jones, Robert G. Martin, and Marshall W. Nirenberg

NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, BETHESDA

Communicated by Richard Roberts, February 27, 1962
Francis Crick
Sydney Brenner

• 1927 - today

• Born in South Africa

• BS in Anatomy and Physiology

• MS in Cytogenetics

• PhD in Physical Chemistry from Oxford

• Joined Salk Institute in 1976

• Established *C. elegans* as model organism for developmental biology

• 2002 Nobel Prize Physiology or Medicine
Leslie Barnett

- 1920 - 2002
- Born in London
- BS in Dairying
- Worked with Brenner for most part of her life

GENERAL NATURE OF THE GENETIC CODE FOR PROTEINS

By Dr. F. H. C. CRICK, F.R.S., LESLIE BARNETT, Dr. S. BRENNER
and Dr. R. J. WATTS-TOBIN

Medical Research Council Unit for Molecular Biology,
Cavendish Laboratory, Cambridge

Bacteriophage T4

Escherichia coli

Figure from Brock Biology of Microorganisms
Cracking the Genetic Code

Viral Plaques

Phage assembly

1. Attachment
2. Entry of phage DNA and degradation of host DNA
3. Synthesis of viral genomes and proteins
4. Assembly
5. Release
Why Phage T4?

- Plaques are an easy screening system
- Allows investigation of rare events (trillions of tries in a single LB plate)
- **rII locus**: phenotypes allows genetic mapping (Benzer)

<table>
<thead>
<tr>
<th>rII genes</th>
<th>E. coli K-12 (K)</th>
<th>E. coli B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>non-leaky mutation (null)</td>
<td>×</td>
<td>larger/irregular plaques (r-plaque)</td>
</tr>
<tr>
<td>leaky mutation (partial function)</td>
<td>✓</td>
<td>larger/irregular plaques (r-plaque)</td>
</tr>
</tbody>
</table>

A - green; B - red
Why Phage T4?

B
null mutation
null mutation
B''
null mutation

Distance (B’ to B'') = \frac{\# \text{ plaques (K)}}{\# \text{ plaques (B)}}

Grow in B
No growth in K
Grow in B
Growth in K

Grow in B
No growth in K
The Genetic Code is not overlapping

Evidence comes from previous studies:

• Tobacco mosaic virus RNA: mutations in RNA change only 1 amino acid (Tsugita et. al)

• Abnormal human hemoglobins shows only single amino acid changes (Watson et. al)
The Genetic Code is not overlapping

- How to find the code (reading frame) in a non-overlapping arrangement?

  1) The comma hypothesis

     \text{CODE, CODE, CODE, CODE, CODE, CODE, CODE, CODE}

  2) The comma-free hypothesis

     \text{CODE CODE CODE CODE CODE CODE CODE CODE CODE}
     \text{ODEC ODEC ODEC ODEC ODEC ODEC ODEC ODEC}
     \text{DECO DECO DECO DECO DECO DECO DECO DECO DECO}

  3) The fixed start hypothesis

     \text{CODE CODE CODE CODE CODE CODE CODE CODE CODE}

\[\text{CODE} \rightarrow \text{CODE, CODE, CODE, CODE, CODE, CODE, CODE, CODE} \]
Experimental System

- FC0 mutant in the B1 segment of the “B cistron”

A

B1 B

- Non-leaky (null) mutant
- No growth in K
- Growth (r-plaques) in B

- FC0 mutant was produced by proflavin treatment

H₂N

N

NH₂

- Adds or deletes a base
- Generates mostly non-leaky mutants
- Second mutation (suppressor) restores wild-type phenotype
- Suppressor mutations by themselves are “non-leaky r” (null) mutants
Suppressors of suppressors

<table>
<thead>
<tr>
<th>WT</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC0</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Suppressor of FC0</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Suppressor of Suppressor</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
</tbody>
</table>

Mutation: None, +, + and -, - and +
Phenotype: WT, null, WT
rII mutations

Collection of ~ 80 mutants

- All mutants are non-leaky r (null)
- All mutants (except FC0) were occurred spontaneously
- Non-spontaneous (proflavin) mutations are similar to spontaneous mutations (line h)
### Generating double mutants

- $+ \text{ and } +$: r phenotype
- $- \text{ and } -$: r phenotype

#### Table 1. Double Mutants having the r Phenotype

<table>
<thead>
<tr>
<th></th>
<th>With -</th>
<th>+ With +</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC (1 + 21)</td>
<td>FC (0 + 58)</td>
<td>FC (40 + 57)</td>
</tr>
<tr>
<td>FC (23 + 21)</td>
<td>FC (0 + 38)</td>
<td>FC (40 + 58)</td>
</tr>
<tr>
<td>FC (1 + 23)</td>
<td>FC (0 + 40)</td>
<td>FC (40 + 55)</td>
</tr>
<tr>
<td>FC (1 + 9)</td>
<td>FC (0 + 55)</td>
<td>FC (40 + 54)</td>
</tr>
<tr>
<td>FC (1 + 9)</td>
<td>FC (0 + 54)</td>
<td>FC (40 + 38)</td>
</tr>
</tbody>
</table>

#### Diagram:

- **WT**: A B C A B C A B C A B C A B C A B C A B C A B C A B C
- **+ AND +**: A B C A A B C A B C A B C A B C A B C A B C A B C A B C
- **- AND -**: A B C A B C A B C B C A B C A B C A B C A B C A B C

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>WT</td>
</tr>
<tr>
<td>+ and +</td>
<td>null</td>
</tr>
<tr>
<td>- and -</td>
<td>null</td>
</tr>
</tbody>
</table>
Generating double mutants

- + and -: Will all combinations be WT?

"Unacceptable codons might exits":
- nonsense
- end-of-chain
- complications in protein structure

"Convention" on frameshifts:

Figure 3
Generating double mutants

Forward shifts are more acceptable in this region.
Reverse shifts may generate "unacceptable" triplets (stop codon) — "Unacceptable region"
Generating double mutants

Table 2. DOUBLE MUTANTS OF THE TYPE (+ WITH -)

<table>
<thead>
<tr>
<th></th>
<th>FC 41</th>
<th>FC 0</th>
<th>FC 40</th>
<th>FC 42</th>
<th>FC 58*</th>
<th>FC 63</th>
<th>FC 38</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC 1</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC 86</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>FC 9</td>
<td>r</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>FC 82</td>
<td>r</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>FC 21</td>
<td>r</td>
<td>W</td>
<td></td>
<td></td>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>FC 88</td>
<td>r</td>
<td>r</td>
<td></td>
<td></td>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>FC 87</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td></td>
<td>r</td>
<td></td>
<td>W</td>
</tr>
</tbody>
</table>
Generating triple mutants

Table 3. **Triple Mutants having a Wild or Pseudo-wild Phenotype**

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>WT</td>
</tr>
<tr>
<td>+ and +</td>
<td>null</td>
</tr>
<tr>
<td>+ and + and +</td>
<td>WT</td>
</tr>
</tbody>
</table>

Genetic code is a triplet (or less likely a multiple of 3)
The start site hypothesis

Function of B cistron

✓
✓
✓
X
✓
✓ or X
Is the code degenerate?

- 4 common base pairs organized in triplets = $4^3$ combinations = 64 triplets
- 20 common amino acids
- If the genetic code is not degenerate: there is 44 nonsense amino acids
Is the code degenerate?

- 4 common base pairs organized in triplets = $4^3$ combinations = 64 triplets
- 20 common amino acids
- If the genetic code is not degenerate: there is 44 nonsense amino acids
- The genetic code must be degenerate
- “how many triplets code amino-acids and how many have other functions we are unable to say”
The genetic code is of the following general type:

- A group of three bases (or, less likely, a multiple of three bases), codes one amino acid
- The code is non-overlapping
- Sequence of bases is read from a fixed point
- The code is probably degenerate

"the coding problem is wide open for experimental attack, and in fact many laboratories, including our own, are already working on the problem"
Har Gobind Khorana
University of Wisconsin, Madison
Mastered synthesis of RNA

CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS*

By J. Heinrich Matthaei,† Oliver W. Jones, Robert G. Martin, and Marshall W. Nirenberg

NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, BETHESDA

Communicated by Richard Roberts, February 27, 1962
# Cracking the Genetic Code

## TABLE 3

Amino Acid Incorporation into Protein Stimulated by Randomly Mixed Polynucleotides

<table>
<thead>
<tr>
<th>Polynucleotide</th>
<th>UA</th>
<th>UC</th>
<th>UG</th>
<th>UAC</th>
<th>UGC</th>
<th>UGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base-ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>0.87</td>
<td>0.39</td>
<td>0.76</td>
<td>0.834</td>
<td>0.341</td>
<td>0.675</td>
</tr>
<tr>
<td>A</td>
<td>0.13</td>
<td>0.61</td>
<td>0.24</td>
<td>0.050</td>
<td>0.152</td>
<td>0.291</td>
</tr>
<tr>
<td>A</td>
<td>0.116</td>
<td>0.502</td>
<td>0.034</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Probability of triplet relative to phenylalanine

<table>
<thead>
<tr>
<th>(UUU) = 100%</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>UUA</td>
<td>13</td>
<td>157</td>
<td>32</td>
<td>6.0</td>
<td>46.2</td>
<td>43</td>
</tr>
<tr>
<td>UAA</td>
<td>2.2</td>
<td>244</td>
<td>10.6</td>
<td>0.4</td>
<td>21.0</td>
<td>19</td>
</tr>
<tr>
<td>AAA</td>
<td>0.3</td>
<td>382</td>
<td>3.4</td>
<td>0.02</td>
<td>1.0</td>
<td>8.1</td>
</tr>
<tr>
<td>UUC</td>
<td>13.9</td>
<td>147</td>
<td>5.1</td>
<td>0.26</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>UCC</td>
<td>1.9</td>
<td>218</td>
<td>0.8</td>
<td>2.2</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>0.3</td>
<td>322</td>
<td>0.01</td>
<td>8.4</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>UAC</td>
<td>0.8</td>
<td>68.1</td>
<td>2.2</td>
<td>0.1</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>AAC</td>
<td>0.05</td>
<td>31.7</td>
<td>0.1</td>
<td>0.01</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>0.12</td>
<td>101</td>
<td></td>
<td>0.01</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

### Amino Acid

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Phenylalanine</th>
<th>Arginine</th>
<th>Alanine</th>
<th>Serine</th>
<th>Proline</th>
<th>Tyrosine</th>
<th>Isoleucine</th>
<th>Valine</th>
<th>Leucine</th>
<th>Cysteine</th>
<th>Tryptophan</th>
<th>Glycine</th>
<th>Methionine</th>
<th>Glutamic acid</th>
<th>Lysine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>1.1</td>
<td>3.2</td>
<td>285</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>4.9</td>
<td>0.6</td>
<td>1.5</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>4.7</td>
<td>0.6</td>
<td>1.5</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.76</td>
<td>4.7</td>
<td>0.6</td>
<td>1.5</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>0.24</td>
<td>0.4</td>
<td>0</td>
<td>0.0</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.24</td>
<td>0.3</td>
<td>12</td>
<td>0.6</td>
<td>1.2</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.1</td>
<td>0.6</td>
<td>1.2</td>
<td>0.0</td>
<td>0.02</td>
<td>0.02</td>
<td>0.05</td>
<td>0.6</td>
<td>0.44</td>
<td>0.116</td>
</tr>
</tbody>
</table>

Composition of coding units:

- UUU...
- UCG...
- UGC...
- UUC + UCG...
- UCC...
- UAA...
- AAA...
- UAC + UUG...
- UUG...
- UUG...
- UGA...
- UGA...
- UAA... (?)
Cracking the Genetic Code

Add ribosomes

Labeled Ser tRNA + synthetic trinucleotide

Pour through filter

Ribosome

No radioactivity trapped in filter

Labeled Leu tRNA + synthetic trinucleotide

Radioactivity trapped in filter
Cracking the Genetic Code

1966

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU</td>
<td>UCU</td>
<td>UAU</td>
<td>UGU</td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
<td>Stop</td>
<td>Stop</td>
</tr>
<tr>
<td>C</td>
<td>CUU</td>
<td>CCU</td>
<td>CAU</td>
<td>CGU</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
<td>His</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pro</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>AUU</td>
<td>ACU</td>
<td>AAU</td>
<td>AGU</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
<td>Asn</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>GUU</td>
<td>GCU</td>
<td>GAU</td>
<td>GGU</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
<td>Asp</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cracking the Genetic Code

1968 Nobel Prize for Deciphering the Genetic Code

Marshall Nirenberg

Robert Holley

Har Gobind Khorana
Will Society Be Prepared?

"Science has reached a new frontier. A revolution far greater… then the atomic or hydrogen bomb"

- New York Times, 1961

“might lead to methods of tampering with life, of creating new diseases, of controlling minds, of influencing heredity, even perhaps in certain desired directions.”

- Arne Tiselius, Nobel Laureate in Chemistry
Thank you!