

Can you Find the Gene?

An exercise in DNA sequence analysis
Kenneth R. Miller, Professor of Biology, Brown University

A group of workers has isolated a fragment of DNA containing the structural gene for the coat protein of a bacterial virus, bacteriophage MS2. The protein is known to be **94 amino acids** in length. They have published the data from one DNA strand in the region of the gene:

```

5' -
TAGAGCCCTCAACCGGAGTTT
GAAGCATGGCTTCTAACTTTA
CTCAGTTCGTTCTCGTCGACA
ATGGCGGAAC TGGCGACGTGA
CTGTCGCCCAAGCAACTTCG
CTAACGGGGTTCGCTGAATGGA
TCAGCTCTAACTCGCGTTCAC
AGGCTTACAAAGTAACCTGTT
GGCGTTCGTA CTTAAATATGG
AACTAACCATTC CAATTTTTCG
CTACGAATTC CGACTGCGAGC
TTATTGTTAAGGCAATGCAAG
GTCTCCTAAAAGATGGA AACC
CGATTCCCTCAGCAATCGCAG
CAA ACTCCGGCATCTACTAAT
AGACGCCGGCCATTCAAACAT
GAGGATTACCC-3'
    
```

For your convenience, the sequence is printed at the left in rows of 21 bases each. A total of 347 bases are in the sequence.

These folks have done the hard part - they've found a piece of DNA containing the gene. Now all you have to do is the following:

- Find which strand of DNA (this one or the complementary one) serves as template for mRNA synthesis.
- Write out the mRNA synthesized from that template.
- Decode the message using the genetic code table, and write out the amino acid sequence in the protein product produced by this gene.

Here's an approach that you may use to solve the problem. You have been given the sequence of one strand of a segment of DNA:

```

5' -TAGAGCCC.....TTACCC-3'
    
```

Step 1: Use base-pairing rules to determine the sequence of the complimentary strand:

```

5' -TAGAGCCC.....TTACCC-3'
3' -ATCTCGGG.....AATGGG-5'
    
```

Step 2: Use the same base-pairing rules to determine the mRNA sequence that would be synthesized using either of these two strands as template:

```

DNA  → 5' -TAGAGCCC.....TTACCC-3'
RNA  → 3' -AUCUCGGG.....AAUGGG-5'
    
```

```

DNA  → 3' -ATCTCGGG.....AATGGG-5'
RNA  → 5' -UAGAGCCC.....UUACCC-3'
    
```

Once you've got the two possible mRNA sequences, you are ready to figure out which one is the actual mRNA. How can you do this? Well, the actual mRNA should have an initiation codon, followed by 93 other codons, followed by a "stop" codon.

The identification of the correct mRNA (and decoding the message) is complicated by the fact that each of the two possible mRNAs can be read in a different **reading frame**:

```

-AUC-UCG-GGA-AUG-GG...
-AU-CUC-GGG-AAU-GGG...
-A-UCU-CGG-GAA-UGG-G...
    
```

So how do we know which reading frame is correct? Simple. Using a particular reading frame, scan the sequence for "stop" codons. The correct scan will show an "open reading frame," with no stop codons, over an unbroken stretch of nearly 400 bases. But scans made out-of-frame will hit a "stop" signal roughly once every 20 codons.

Now that you have the correct strand and reading frame, you can turn to the genetic code table and decode the

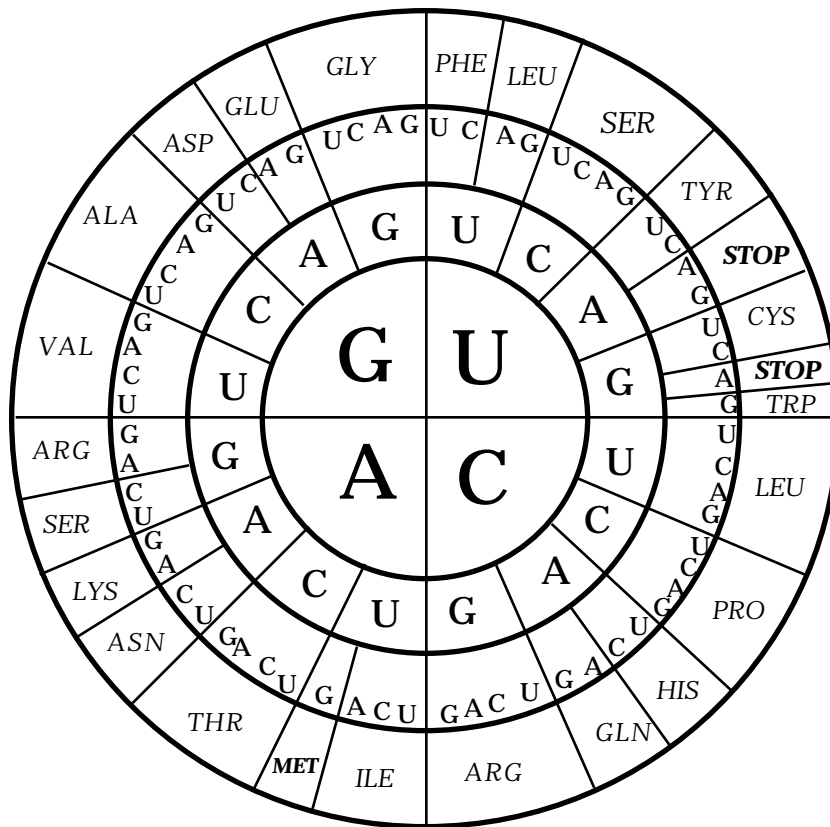
5' -
 TAGAGCCCTCAACCGGAGTTT
 GAAGCATGGCTTCTAACTTTA
 CTCAGTTCGTTCTCGTCGACA
 ATGGCGGAACTGGCGACGTGA
 CTGTCGCCCAAGCAACTTCG
 CTAACGGGGTTCGCTGAATGGA
 TCAGCTCTAACTCGCGTTCAC
 AGGCTTACAAAGTAACCTGTT
 GCGGTTTCGTAATAAATATGG
 AACTAACCATTCCAATTTTCG
 CTACGAATTCCGACTGCGAGC
 TTATTGTTAAGGCAATGCAAG
 GTCTCCTAAAAGATGGAAACC
 CGATTCCCTCAGCAATCGCAG
 CAAACTCCGGCATCTACTAAT
 AGACGCCGGCCATTCAAACAT
 GAGGATTACCC - 3'

**Bases
 #1-126**

**Bases
 #127-252**

**Bases
 # 253-347**

The Genetic Code



A Circular Genetic Code Table?

Absolutely. A circular arrangement of the code is more quickly understood by students and is much easier to use.



The circular Genetic Code table is just one of many innovations in the new **BIOLOGY** by Miller & Levine.

Additional copies of this problem and copies of its **solution** are available at the web site we've established for students and teachers using our textbooks: