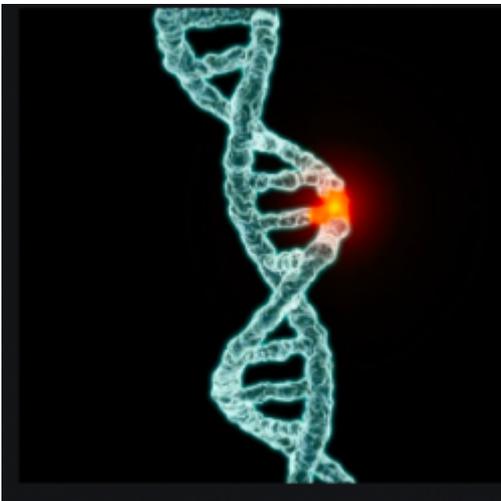


# The Idiot's Guide to Viral Mutation



By Josh Bloom — January 4, 2021

*We need new coronavirus variants like a duodenal ulcer, but they're here – something any virologist would have said was inevitable. Here's a lesson on how mutation works. Plus an explanation of what those crazy letters and numbers mean that you see in the news.*



Mutation. Image: Wikipedia Commons

[1]

As if we don't have enough problems.

COVID-19 is way more than bad enough, so pretty much the last thing we needed was the emergence of mutated forms of the damn thing. But it was unavoidable. Why? Because of natural selection. This is how viruses thrive – by spontaneously "improving" themselves, which can make them more contagious, more (or also less) virulent (deadly), or any of these.

The clever little devils also mutate in response to a drug or vaccine, again ensuring that they thrive. The mechanism is the same, but in this case, selective pressure rather than random errors is the driving force behind mutation. Let's say that a person infected with Virus A is treated with an antiviral drug that prevents replication of Virus A. The drug works by stopping one of the essential steps (these are called targets) required to build new virus particles. Is this the end of Virus A? Hell, no.

The virus doesn't give up. It is programmed to make multiple errors in its DNA or RNA synthesis, and eventually, a new variant or strain will emerge that is not inhibited by the drug. Virus 1A then

becomes the predominant circulating strain **(1,2)**. Since the COVID variants began emerging well before the first vaccines were administered, it is safe to say that the mutations we're now learning about were spontaneous, not caused by resistance to any therapy.

When new variants emerge, they need to be identified by a name or number. That's why we are seeing all those bizarre-looking numbers and letters attached to the virus. But it's not really complicated. The numbers and letters actually make a lot of sense. Here's why:

By using technology beyond the scope of this article, it is now quite easy to determine the sequence of the virus's genetic components. Don't let this next diagram scare you away. Even though it looks like a nightmare, it's really quite simple. More or less.

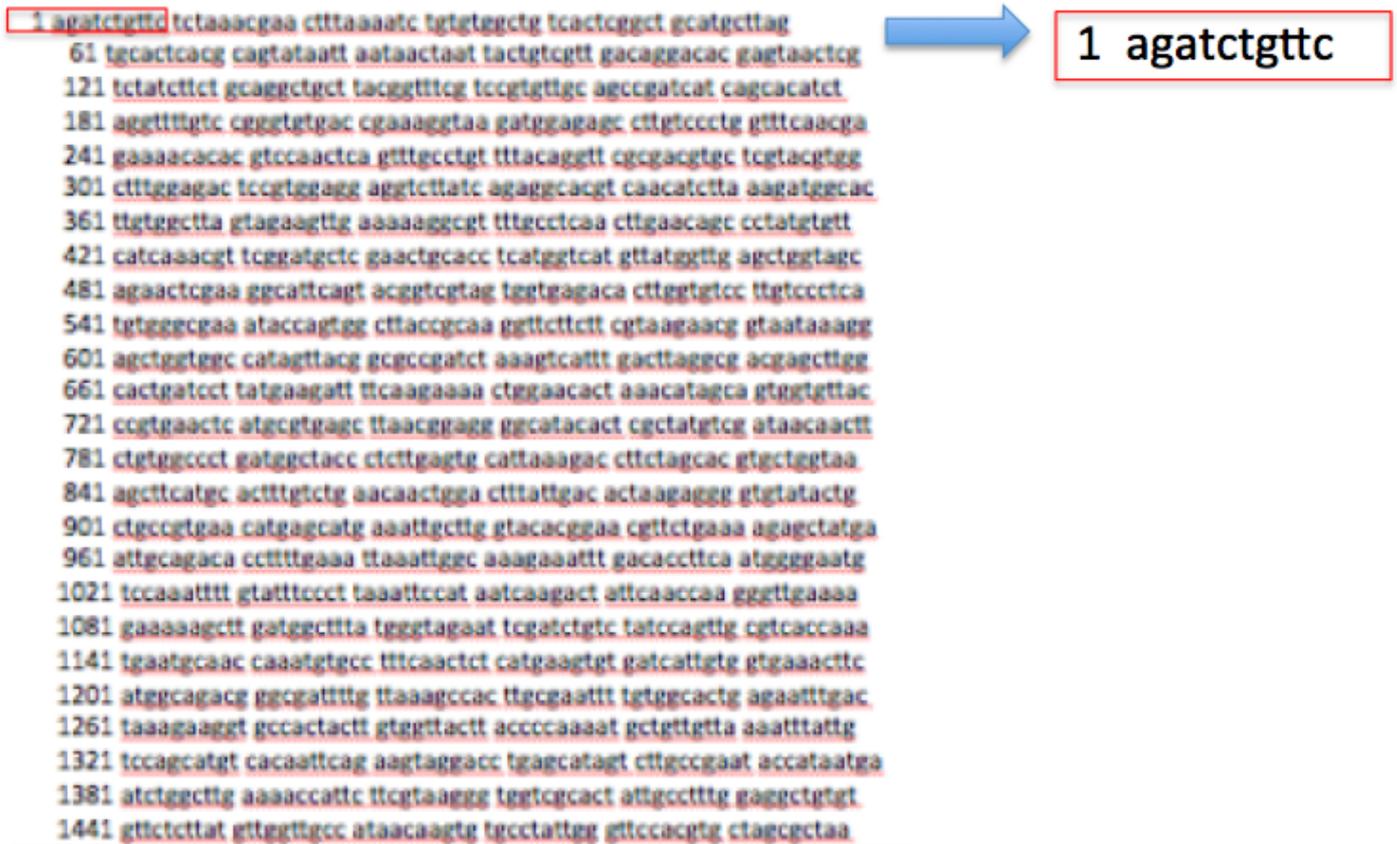
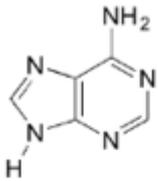
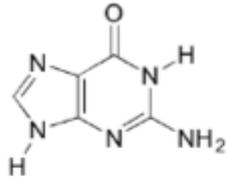


Figure 1. The first 1500 of the 29,764 nucleotides that make up the SARS-CoV-2 genome. The first 10 units (red box) are enlarged on the right. There will be a quiz. Source: [NIH](#) [2]

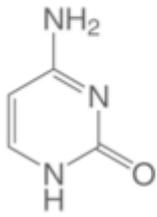
That Rorschach Test above is actually quite logical. It is the order that the nucleic acids (the genetic material) of SARS-CoV-2 are hooked together. There are only four different chemicals called nitrogenous bases that make up the 29,764 nucleotide genome of the virus. These are connected in seemingly random order. Below are the four bases and their structures. **(3)**.



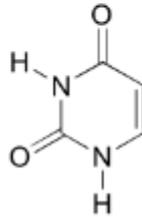
**Adenine (A)**



**Guanine (G)**



**Cytosine (C)**



**Uracil (U)**

**ACUUCG**

*(Left) The 4 nitrogenous bases that make up RNA. (Right) A shortcut for writing out a six-nucleic fragment of RNA. This fragment consists of the adenine–cytosine–uracil–uracil–cytosine–guanine bonded together by phosphate groups on sugars. TMI for this article.*

### **So What?**

The relevance of these four nucleic acids to the structure of SARS-CoV-2 is not immediately obvious. To understand what's going on you have to know a (very) little about the genetic code (sorry). I'll keep it simple. The purpose of genes is to make proteins in viruses, humans, and all other life forms. But it's the way this works that is fascinating.

Nature (chemistry, really) figured out how to assemble the 20 amino acids that make proteins in the right order, and it's very cool. Three-piece fragments of DNA/RNA (these are called codons) determine which amino acid is selected to go next into the protein's growing chain (another way of saying this is that UUU codes for phenylalanine). More on this below.

### **It Gets (A Little) Worse**

The coronavirus's genome consists of 27,764 nucleic acids, each of which is assigned a number, for example, A1445, G556, U 11,001. The chain begins with A-1 and ends with C-29,764. Once the genomic sequence is determined, it is easy to figure out which proteins make up the virus (the genetic code). You just have to know a little trick. Each codon inserts a different amino acid into the growing protein strand. It is known which codons are responsible for "selecting" the next amino acid; this is the genetic code.

### **Codon Amino acid (4)**

GGA - Glycine

UGG - Tryptophan

UUU - Phenylalanine

*(Figure 2). The genetic code tells us how DNA/RNA assemble proteins. A sequence in the genome that contains three uracils (U) in a row will insert phenylalanine in that position. A sequence containing UGG will select tryptophan. This process goes on ad-nauseum until the*

protein is completed.

So it should be (sort of) obvious how protein synthesis works. Three-piece fragments of DNA or RNA determine which amino acid is "grabbed" by the mechanism that links them together to form proteins. Changes in codons are responsible for mutations to take place.

## Genetic material makes proteins

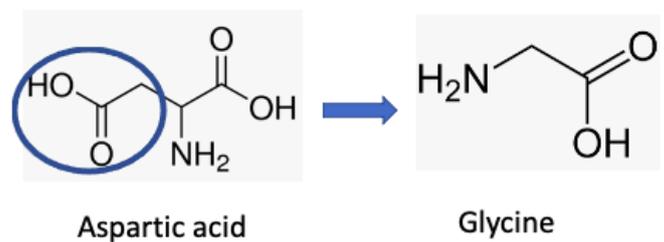
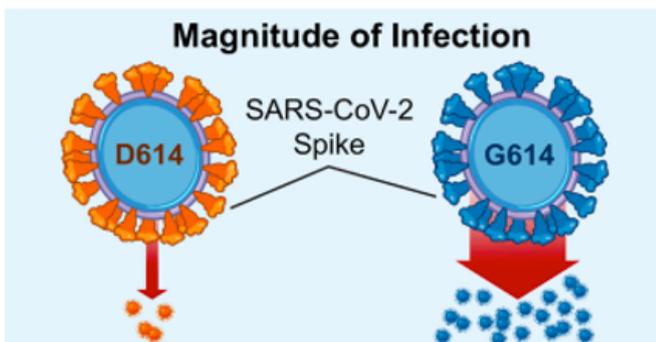
As with nucleic acids, amino acids must also be assigned a number depending on their position in the chain. Here is a diagram of the first 80 amino acids in the coronavirus.

1    MESLVLVGVNEKTHVQLSLPVLQVRDVLVVRGFGDSVEEALSEAREHLKNGTCGLVELEKGVLPQLEQPYVFIKRS DALSTN    80

Each of the 20 amino acids has both a **three-letter and one-letter abbreviation** [3]. Some make sense; some don't. For example, phenylalanine is appropriately abbreviated "phe" but is known by the single letter "F." Lysine is "lys," but also "K."

So, beginning at the first amino acid of the protein sequence of SARS-CoV-2, we have methionine (M), glutamic acid (E), serine(S), leucine (L), valine (V)...

This is where the confusing numbers come in, but now that you're all protein chemistry experts, it's a piece of cake. One variant that seems to be more infectious **has been identified** [4] as D614G. It means that at the 614th amino acid (in the spike protein), D (aspartic acid) has been replaced by G (glycine). Of the **1255 amino acids** [5] in the spike protein, simply changing ONE of them gives a variant that is more infective.



*When aspartic acid at position 614 is replaced by glycine the virus becomes more infectious. One small change (blue circle) in one of thousands of amino acids has a profound impact on the virus. Fascinating and also terrifying. Source: Cell [4]*

## Why did the change happen?

To answer this, we need to go back to the genetic code. The three nucleic acid codon that makes aspartic acid is Guanine-Adenine-Uracil (GAU). For glycine, it's GGU. Adenine has been replaced by guanine during the assembly of the RNA, and this tiny change – one stinking nucleic acid out of 27,764 – is incorporated by error, and all of a sudden, we have a more contagious virus. Scary? Yes. Fascinating? Also yes.

But (and this is probably small comfort) now when you see N501Y, you'll know that in the spike, asparagine has been replaced by tyrosine, and in K417N, lysine has replaced asparagine.

Now you know a little about some of the confusing terminology in the news and where it comes from. Perhaps you'll be a little smarter. That is unless you're still walking around without a mask.

## NOTES

(1) Variants have small genetic differences. Strains have large differences.

(2) Mutations that cause viruses to become resistant to drugs are the bane of antiviral researchers. This is why successful drugs for HIV and HCV infection contain more than one drug. Drug cocktails decrease resistance.

(3) The nucleic acids in RNA are ACGU. In DNA, they are ACTG. Uracil (U) and thymine (T) differ by only one methyl group.

(4) It's not quite this simple. Most amino acids have multiple codons (built-in redundancies). For example, both UAU and UAC will incorporate (code for) tyrosine and CCU, CCC, CCA, and CCG; all encode for proline.

---

COPYRIGHT © 1978-2016 BY THE AMERICAN COUNCIL ON SCIENCE AND HEALTH

---

**Source URL:** <https://www.acsh.org/news/2021/01/04/idiots-guide-viral-mutation-15251>

### Links

[1] [https://commons.wikimedia.org/wiki/File:Mutation\\_of\\_Genes.jpg](https://commons.wikimedia.org/wiki/File:Mutation_of_Genes.jpg)

[2] <https://www.ncbi.nlm.nih.gov/nuccore/MW422255>

[3] [https://molbiol-tools.ca/Amino\\_acid\\_abbreviations.htm](https://molbiol-tools.ca/Amino_acid_abbreviations.htm)

[4] [https://www.cell.com/cell/pdf/S0092-8674\(20\)30820-5.pdf](https://www.cell.com/cell/pdf/S0092-8674(20)30820-5.pdf)

[5] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC387703/>