Ancestry of the genetic code

The origin of life is the origin of heredity as well

S ANANTHANARAYANAN

From 1859 when Charles Darwin proposed the theory of evolution to the discovery of DNA and the genetic code, in the 1950s, the mystery of life and heredity has been laid bare down the centuries. At its core is the code, built of three-letter words, using a four-character alphabet, which helps rebuild millions of proteins, to enable living things to do what sets them apart – to reproduce.

The code is a mathematically elegant construction – it is precise, economical and error-protected – an end product more efficient than any variant that we can suggest. It is universal and unchanged, from the simplest, single-celled organism to the greatest of mammals. By what stages could this code have arisen? Masayori Inouye, Risa Takino, Yojiro Ishida and Keiko Inouye from the Rutgers-Robert Wood Johnson Medical School, New Jersey, propose a new look at the question in the journal, *Proceedings of the National Academy of Sciences*.

In the same way that the most inspiring concept of an architect cannot be realised unless she prepares a blueprint, an organism, no matter how efficient, cannot have a second generation unless it contains within itself the blueprint of its own construction. Living things are essentially their cells and the set of proteins, which cells produce and control the way other cells of the organism behave. The cells of all living things hence contain a blueprint, in the form of a long (very long – billions of units long) ticker tape that carries the code for the proteins. The *DNA molecule* is the tape, and the code for the proteins are bits of DNA, called the *genes*. And the genes are built up of three-letter words of an alphabet of four kinds of chemical groups, the letters, called the bases.

Now, the structure of proteins has got optimised to consist of a chain, often a very long chain, of components from a set of just 20 different units, called *amino acids*. Within the DNA, each group of three letters, formed out of the four letters that are available, is called a *codon* and is the template for creation of an amino acid. The box on this page shows how many three-letter words we can form with four alphabets, and it works out to be 64. If the word had only two letters, there would be only 16 ways that it could be formed, which is not enough to describe 20 amino acids. We hence need at least three letters in the word, and if 64 is a lot more than 20, well, three codons have special uses, but the remaining 61 provide alternate forms for the most frequent amino acids -- as an insurance to avoid errors when the code in the DNA is transcribed!

That living organisms are able to implement this mathematically elegant system, using just chemical combinations within the organisms' cells, shows the great power of the process of evolution and raises a question of how it may have come about. One theory is that the first amino acids were born from the elements in the stormy and energetic environment of early Earth. Amino acids that have been created in laboratory simulations, and traces found in meteorites, suggest that there may have been 10 amino acids at the start of life, and these grew into 10 more, stabilising at the efficient number of 20. The work done by the authors of the paper, however, finds that there may have been seven amino acids to start with, and more than one route for their development.

The four letters, or chemical groups, which form the codons are – U for uracil, C for cytosine, A for adenine and G for guanine. The picture shows how the 20 amino acids (and three "stop" codons to separate the genes) are formed by combining U, C, A and G. Significantly, we see two amino acids are encoded by only one codon, there are eight coded by two codons, just one coded by three codons, five coded by four codons and three coded by six codons. The number of redundant forms, however, does not generally correspond to the abundance of the amino acids, the paper says. For example, among the three amino acids coded by six codons, (green) arginine and serine are not the most frequently found. It is hence likely that the different forms came about by different processes.

In the case of leucine and arginine, the codons share bases in such a way that one codon can transform to another with a change of only one base. This, however, is not true in the case of serine. Here, we have four codons that start with "UC" and two more that start with "AG". It would hence take a change of two bases for a codon in one group to reach a form in the other. Further, the paper notes, single base changes, in the first or second place, leads to six different amino acids that are unrelated to serine. The authors hence suggest that the origin of the two forms which start with "AG" was different from the origin of forms that start with "UC".

To seek evidence of this suggestion, the authors analyse 4,225 protein coding genes of *E. coli*, a common intestinal bacterium. What they find is that although there are, in serine, theoretically two "AG" codons to four "UC" codons, the occurrence is not in the ratio of 1:2, but is as high as 3:4. The "AG" codons are thus used disproportionately more often, and again, within the "AG" codons, it is more often the "AGC" codon. And then, there are differences in where the two forms of serine occur or are used.

This fits in, the paper says, with the idea that more analysis brings forward, that "AGC" was evolutionarily one of the most primitive codons for serine, itself having descended from a form for GGC, for glycine. The analysis leads to the hypothesis that the codon for first amino acid had the form "GG" and from this the first seven amino acids arose. The remaining 13 arose from these seven, but the alternate form, "AG" of serine came through an independent route.

More work on the genomes of other bacteria and other life forms, and the roles that the two forms of serine play, could further illuminate the path by which they came to be, the paper says.

The writer can be contacted at response@simplescience.in

HOW MANY WORDS CAN WE CREATE?

With four alphabets at our disposal, we can choose the first of the three letters in any of four ways. For each choice that we make, the second letter can again be chosen in four ways. There are hence 4 x 4 = 16 ways to choose the first two letters. Now, for the third letter, again, we have four choices. The total number of three-letter words we can form is thus, 4 x 4 x 4 = 64.

U,C,A,G in the 1st place	C in the 2nd place	A in the 3rd place	G in the 4th place
U	UUU phenylalanine	UCU serine	UAU tyrosine	UGU cysteine
	UUC phenylalanine	UCC serine	UAC tyrosine	UGC cysteine
	UUA leucine	UCA serine	UAA stop	UGA stop
	UUG leucine	UCG serine	UAG stop	UGG tryptophan
C	CUU leucine	CCU proline	CAU histidine	CGU arginine
	CUC leucine	CCC proline	CAC histidine	CGC arginine
	CUA leucine	CCA proline	CAA glutamine	CGA arginine
	CUG leucine	CCG proline	CAG glutamine	CGG arginine
A	AUU isoleucine	ACU threonine	AAU asparagine	AGU serine
	AUC isoleucine	ACC threonine	AAC asparagine	AGC serine
	AUA isoleucine	ACA threonine	AAA lysine	AGA arginine
	AUG methionine	ACG threonine	AAG lysine	AGG arginine
G	GUU valine	GCU alanine	GAU aspartic acid	GGU glycine
	GUC valine	GCC alanine	GAC aspartic acid	GGC glycine
	GUA valine	GCA alanine	GAA glutamic acid	GGA glycine
	GUG valine	GCG alanine	GAG glutamic acid	GGG glycine

U is uracil C is cytosine A is adenine G is guanine

PLUS POINTS

Cosmic brain

The universe is similar to a huge human brain, scientists have found.

A new study investigated the differences and similarities between two of the most complex systems in existence, though at entirely different scales – the cosmos and its galaxies, and the brain and its neuronal cells.

They found that while the scale is clearly different, the structure is remarkably similar. In some cases, the two systems seemed more alike to each other than they did to the parts that make them up. It suggests that vastly different physical processes can lead to very similar complex and organised structures.

For example, the human brains work because of the network of nearly 70 billion neurons that together make it up. The universe is thought to have at least 100 billion galaxies.

In each system, they are assembled together in a complex web or network, spread out in long filaments and nodes that link them up. Those spreading nodes are familiar to pictures of both the universe and the brain, and account for some of the superficial similarities in images.

But in each system, those threads only make up about 30 per cent of the mass. In each, some 70 per cent of the mass is actually made up of parts that appear to be passive – the brain's water, and the universe's dark energy.

To dig further into those similarities, researchers compared the way those galactic networks form with sections of the brain. They looked to understand how the matter was spread across the two very different networks.

"We calculated the spectral density of both systems. This is a technique often employed in cosmology for studying the spatial distribution of galaxies", said Franco Vazza, an astrophysicist at the University of Bologna who worked on the study with University of Verona neurosurgeon Alberto Feletti, "Our analysis showed that the distribution of the fluctuation within the cerebellum neuronal network on a scale from one micrometre to 0.1 millimetres follows the same progression of the distribution of matter in the cosmic web but, of course, on a larger scale that goes from five million to 500 million light-years".

They also examined the ways that the webs of neutrons and galaxies connect up -- once again finding noticeable similarities, with the systems seeming more similar to each other than to their component parts. To do so, they compared the average number of connections between each of the nodes, and how they cluster.

"Once again, structural parameters have identified unexpected agreement levels. Probably, the connectivity within the two networks evolves following similar physical principles, despite the striking and obvious difference between the physical powers regulating galaxies and neurons", said Feletti.

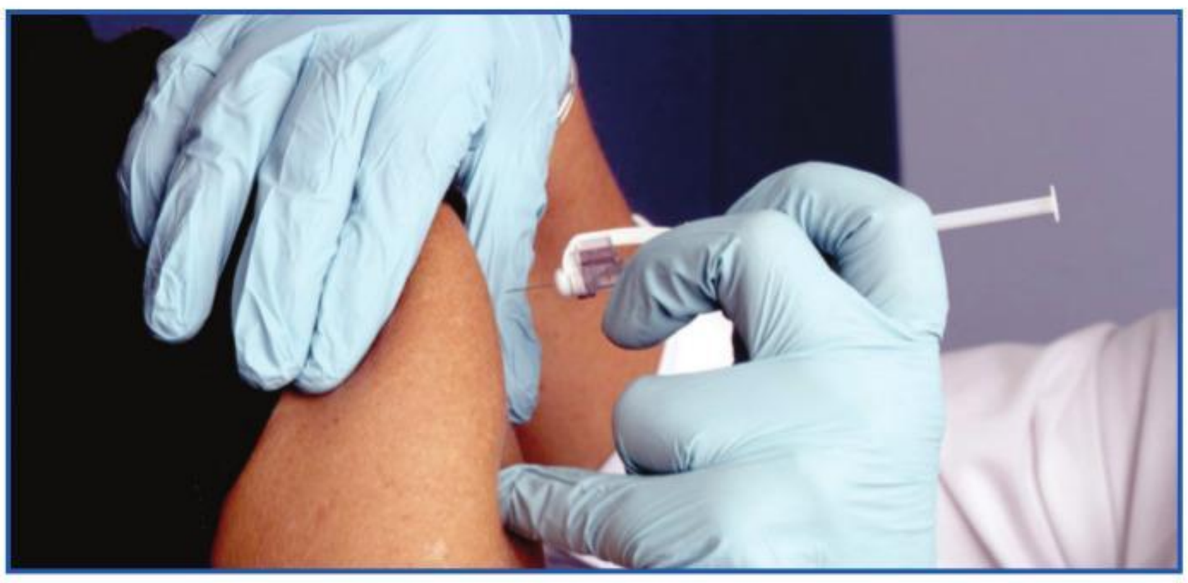
A paper describing the findings, "The quantitative comparison between the neuronal network and the cosmic web", is published in the journal *Frontiers of Physics*.

Farewell Arecibo

The renowned Arecibo telescope in Puerto Rico will be dismantled after 57 years of service due to the rupture of cables that have led to the threat of collapse, the US National Science Foundation announced last Thursday. Two cables supporting the 900-ton instruments for the telescope above a radio dish 1,000 feet in diameter broke on 10 August and 6 November.

Engineers are concerned other cables could also break at any time, making any attempt at repair excessively dangerous. The telescope is one of the largest in the world and has been a tool for many astronomical discoveries. For nearly six decades, the Arecibo Observatory has served as a beacon for breakthrough science and what a partnership with a community can look like. Using the hashtag "WhatAreciboMeansToMe", messages of sadness at the news spread on *Twitter* from both professional and amateur astronomers who have used the telescope for their work in observing the cosmos for decades.

–THE DAILY STAR/ANN



HOLDS MUCH PROMISE

Here's how mRNA vaccines from Pfizer and Moderna work, why they're a breakthrough and why they need to be kept so cold

SANJAY MISHRA

As the weather cools, the number of infections of the Covid-19 pandemic are rising sharply. Hamstrung by pandemic fatigue, economic constraints and political discord, public health officials have struggled to control the surging pandemic. But now, a rush of interim analyses from pharmaceutical companies Moderna and Pfizer/BioNTech have spurred optimism that a novel type of vaccine made from messenger RNA, known as mRNA, can offer high levels of protection by preventing Covid-19 among people who are vaccinated.

Although unpublished, these preliminary reports have exceeded the expectations of many vaccine experts, including mine. Until early this year, I worked on developing vaccine candidates against Zika and dengue. Now I am coordinating an international effort to collect reports on adult patients with current or previous cancers who have also been diagnosed with Covid-19.

Promising preliminary results

Moderna reported that during the phase three study of its vaccine candidate mRNA-1273, which enrolled 30,000 adult US participants, just five of the 95 Covid-19 cases occurred among the vaccinated, while 90 infections were identified in the placebo group. This corresponds to an efficacy of 94.5 per cent. None of the infected patients who received the vaccine developed severe Covid-19, while 11 (12 per cent) of those who received the placebo did.

Similarly, the Pfizer-BioNTech vaccine candidate, BNT162b2, was 90 per cent effective in preventing infection during the phase three clinical trial, which enrolled 43,538 participants, with 30 per cent in the US and 42 abroad.

How does mRNA vaccine work?

Vaccines train the immune system to recognise the disease-causing part of a virus. Vaccines traditionally contain either weakened viruses or purified signature proteins of the virus.

But an mRNA vaccine is different, because rather than having the viral protein injected, a person receives genetic material – mRNA – that encodes the viral protein. When these genetic instructions are injected into the upper arm, the muscle cells translate them to make the viral protein directly in the body.

This approach mimics what the Sars-CoV-2 does in nature – but the vaccine mRNA codes only for the critical fragment of the viral protein. This gives the immune system a preview of what the real virus looks like without causing disease. This preview gives the immune system time to design powerful antibodies that can neutralise the real virus if the individual is ever infected.

While this synthetic mRNA is genetic material, it cannot be transmitted to the next generation. After an mRNA injection, this molecule guides the protein production inside the muscle cells, which reaches peak levels for 24 to 48 hours and can last for a few more days.

Why is making an mRNA vaccine so fast?

Traditional vaccine development, although well studied, is very time-consuming and cannot respond instantaneously against novel pandemics such as Covid-19.

For example, for seasonal flu, it takes roughly six months from identification of the circulating influenza virus strain to produce a vaccine. The candidate flu vaccine virus is grown for about three weeks to produce a hybrid virus, which is less dangerous and better able to grow in hens' eggs. The hybrid virus is then injected into a lot of fertilised eggs and incubated for several days to make more copies. Then the fluid containing virus is harvested from eggs, the vaccine viruses are killed, and the viral proteins are purified over several days.

The mRNA vaccines can leapfrog the hurdles of developing traditional vaccines such as producing non-infectious viruses or producing viral proteins at medically demanding levels of purity.

MRNA vaccines eliminate much of the manufacturing process because rather than having viral proteins injected, the human body uses the instructions to manufacture viral proteins itself. Also, mRNA molecules are far simpler than proteins. For vaccines, mRNA is manufactured by chemical rather than biological synthesis, so it is much quicker than conventional vaccines to be redesigned, scaled up and mass-produced.

In fact, within days of the genetic code of the Sars-CoV-2 virus becoming available, the mRNA code for a candidate vaccine testing was ready. What's most attractive is that once the mRNA vaccine tools become viable, mRNA can be quickly tailored for other future pandemics.

What are the problems with mRNA?

MRNA technology isn't new. It was shown a while back that when synthetic mRNA is injected into an animal, the cells can produce a desired protein. But the progress remained slow. That's because mRNA is not only notoriously unstable and easy to degrade into smaller components, it is also easily destroyed by the human body's immune defences, which make delivering it to the target very inefficient.

But beginning in 2005, researchers figured out how to stabilise mRNA and package it into small particles to deliver it as a vaccine. The mRNA Covid-19 vaccines are expected to be the first using this technology to be approved by the Food and Drug Administration, US. After a decade of work, the mRNA vaccines are now ready for evaluation. Physicians will be watching for unintended immune reactions, which can be both helpful and detrimental.

Why keep mRNA supercold?

The most important challenge for development of a mRNA vaccine remains its inherent instability, because it is more likely to break apart above freezing temperatures.

Modification of the mRNA building blocks and development of the particles that can cocoon it relatively safely have helped the mRNA vaccine candidates. But this new class of vaccine still requires unprecedented freezer conditions for distribution and administration.

What are the refrigeration requirements?

The Pfizer-BioNTech mRNA vaccine will need to be optimally stored at minus 94 degrees Fahrenheit and will degrade in around five days at normal refrigeration temperatures of slightly above freezing.

In contrast, Moderna claims its vaccine can be maintained at most home or medical freezer temperatures for up to six months for shipping and longer-term storage. Moderna also claims its vaccine can remain stable at standard refrigerated conditions, of 36 to 46 degrees Fahrenheit, for up to 30 days after thawing, within the six-month shelf life.

Not surprisingly, Pfizer is also developing shipping containers using dry ice to address shipping constraints.

The writer is project coordinator and staff scientist, Vanderbilt University Medical Center, Vanderbilt University, US. This article first appeared on www.theconversation.com