

George Gamow and the Genetic Code

Vidyanand Nanjundiah



The author received his PhD in physics from the University of Chicago. His research interests lie in the areas of developmental biology and evolution.

On February 28, 1953, in a pub in Cambridge, Francis Crick was telling everyone who cared to listen that he and James Watson had just discovered the secret of life. The April 25 issue of the journal *Nature* carried the same news in the form of their first, and most famous, paper, "A Structure for Deoxyribose Nucleic Acid". In it they announced that DNA, the molecular basis of heredity, was a right-handed double helix. It consisted of two intertwined, anti-parallel helical strands. Each strand was a long molecule made up of subunits which contained a sugar, deoxyribose, a phosphate group, and one of the four bases adenine (A), guanine (G), thymine (T) and cytosine (C). The two strands specified each other; they were 'complementary'. This was because they were held together by hydrogen bonds formed between adenine and thymine (A-T) and between guanine and cytosine (G-C). On May 30 there was a follow-up by Watson and Crick in the same journal, entitled "Genetical Implications of the Structure of Deoxyribonucleic Acid". It was seen by Luis Alvarez and brought by him to the attention of George Gamow, then visiting the University of California at Berkeley.

Gamow, a physicist who had emigrated from the Soviet Union, was 49 years old at the time and famous for his contributions to quantum mechanics and nuclear physics, especially for a brilliant explanation of α -particle radioactivity. He was also well known for deducing that if our universe had a hot and dense beginning (much later to be called the 'big bang'), it would contain an observable trace in the form of black body radiation with a characteristic temperature. In both cases he had taken accepted physical laws for granted and applied them to unusual situations. Biology was not all that new to him either, and he had written a Mr. Tompkins book on the subject. The 'Tompkins' and other similar books dealing with physics mark him as one of the great scientific popularisers of all time. They earned him the

Keywords

Coding problem, information, molecular biology.



Kalinga Prize in 1956. Unlike those books, the one on biology has been forgotten. Coming as it did slightly before the double helix turned biology upside down, its chances of success could hardly have been worse. Gamow must have been introduced to biology much earlier, though. He was friends with Max Delbrück, one of the founders of molecular biology. Years before, when Delbrück, then a physicist, had gone to Copenhagen to serve the obligatory apprenticeship with Niels Bohr, it was Gamow, his senior by a few years, who took him in hand.

To return to our story, on the 8th of July Gamow addressed a letter to Watson and Crick. He introduced himself as “a physicist, not a biologist’ who was “very much excited by [their] May 30 article”. He went on to add, “If your point of view is correct, and I am sure it is at least in its essentials, each organism will be characterized by a long number written in quadrucal (?) system”.... “For example, the animal will be a cat if Adenine is always followed by Cytosine in the DNA chain...”.

Gamow’s letter touched off an extraordinary enterprise in the history of biology, the search for a purely *formal* (meaning arbitrary) set of rules, a Genetic Code, that would relate the hereditary information carried in DNA to the stuff that built bodies, proteins. Audacious though it seemed at the time, the proposal was not without precedent. In the influential book *What Is Life?*, published in 1944, Erwin Schrödinger, one of the founders of quantum mechanics, had alluded to a “hereditary code-script” that could specify the difference between “...a rhododendron, a beetle, a mouse or a woman”. Schrödinger used the explicit example of the Morse code of dots and dashes to explain how a small number of symbols could encode an enormous number of messages. But in this and subsequent forays into the ‘coding problem’, as it came to be christened, Gamow went beyond what Schrödinger had said. Firstly, he asserted that one could work out the code solely from a knowledge of the sequence of bases in a DNA molecule and the sequence of amino acids in the protein that it encoded; the details of the underlying chemistry were unimportant. Secondly, he simplified the prob-

Gamow's letter touched off an extraordinary enterprise in the history of biology, the search for a purely *formal* (meaning arbitrary) set of rules, a Genetic Code, that would relate the hereditary information carried in DNA to the stuff that built bodies, proteins.

Schrödinger used the explicit example of the Morse code of dots and dashes to explain how a small number of symbols could encode an enormous number of messages.



The coding problem galvanized the infant field of molecular biology.

lem even more by reducing it to a numerical exercise. There were four bases in DNA, he pointed out, and 20 amino acids in proteins. (The figure of 20 was a bold guess that turned out to be correct, though the actual 20 are not the same as those that Gamow listed.) How could a language with four letters, the language of DNA, be translated into a language with twenty words, the language of proteins?

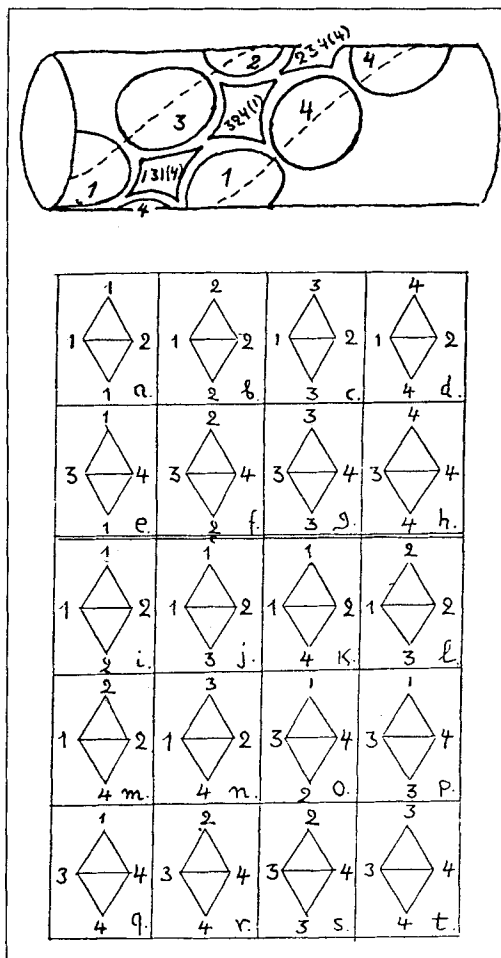
The coding problem galvanized the infant field of molecular biology. Elementary combinatorial reasoning showed that the code would have to make use of at least three of the four bases at a time. (Because $4 \times 4 = 16$, which is less than 20, a 2-letter code would be insufficient). Given a 3-letter code, the problem was soon posed as one of going from DNA to protein in two steps. To begin with, DNA led to an intermediate, RNA, which was formed by using the same rules for pairing bases as between complementary strands of DNA. Then, starting from a series of sequences of three RNA bases, one had to get to a protein sequence. (Amazingly, this had occurred to Alexander Dounce, a chemist, well before the announcement of the double helix. Dounce also anticipated – again, correctly – that enzymes would be required to do the job.) The second part was the real puzzle. By stripping the problem down to what seemed to be its mathematical essentials, Gamow succeeded in attracting many of the brightest scientific minds of the day to give it a try. Among them were Feynman, Delbrück, von Neumann and Teller. Besides Gamow himself, Crick contributed ingenious solutions. Together with Watson, he was responsible for drawing up a list of the correct, or canonical, twenty amino acids –namely, of those amino acids that were encoded by DNA. His assumption, since shown to be largely correct, was that the other amino acids that are found in proteins are formed via chemical modifications of the canonical twenty. It was Crick who postulated that the genetic code was universal, that is, the same in all species. This too is largely true. However, as far as the code itself went, all the theories and models turned out to be off the mark. What should be noted is that none of them was trivial. In fact, one of Crick's

By stripping the coding problem down to what seemed to be its mathematical essentials, Gamow succeeded in attracting many of the brightest scientific minds of the day to give it a try.



solutions to the coding problem, known as the comma-less code, has been called 'the most elegant theory in biology that was wrong'.

Let me try to give a flavour of the types of coding hypotheses that caused much excitement. Gamow proposed a 'diamond code', so called because it depended on the diamond or trapezoid-shaped cavity formed between four nucleotide bases in DNA. If three successive bases on one strand are numbered 1, 2 and 3, two opposite corners of the diamond were supposed to be formed by 1 and 3', and the remaining two corners by 2 and 2', where 3' and 2' are complementary to 3 and 2. The order of the bases along a strand was not significant. An amino acid was supposed to fit snugly, meaning in a stereospecific fashion, into the diamond. The code depended on a direct interaction of amino acids and DNA. It was a triplet code: only three bases were important (1, 2 and 3), since the fourth (2') was automatically specified by the rules of base-pairing. Because successive amino acids on a protein shared two of the three bases that encoded them, this was an 'overlapping' code. (Before looking at the figure, the reader is invited to verify that this scheme does yield the magic number of 20 amino acids.) How did the diamond code fare? To begin with, there was an attractive simplicity to the notion that amino acids could bind to DNA in this fashion. The distance between two bases in DNA happens to be approximately equal to that between successive amino acids in a protein. Thus, it appeared plausible that DNA might specify proteins with something like a one-to-one matching between the cavities defined by base triplets and amino acids. The diamond code came with an added bonus, because the resulting protein would be as large as it could be: it would contain about as many amino acids as there were bases in a strand of DNA. However, the stereochemistry of binding was difficult to understand – in par-



Schematic representation of Gamow's 'diamond-code'. The top picture represents the diamond or trapezoid-shaped cavity formed between four bases in DNA. The lower figure represents a coding scheme for 20 amino acids that he proposed. (Note that the numbers used in the text are different.) (From G Gamow, *Nature*, Vol.173, p.318, 1954)

Gamow's contribution to the solution of the coding problem was to have set off the chase and maintained its momentum.

particular because the sequences 12(2')3 and 32(2')1 had to encode the same amino acid. The triplet aspect was a plus point but it was part of the code by assumption. The third feature, the overlapping nature of code, was to prove its downfall. Brenner proved that all overlapping codes were impossible. The reason was that they strongly constrained which amino acid could be a neighbour of which. For example, if the base sequence on a DNA strand was 12345, successive amino acids would be encoded by the bases 123, 234 and 345. This meant that once the amino acid corresponding to 123 was fixed, the next amino acid *had* to be one whose code began with 23. Brenner showed that proteins simply did not obey these constraints.

The genetic code was finally deciphered with the help of experiments which involved a combination of smart thinking and luck. Biochemical reasoning played a role in many of them. Unexpectedly, the code was degenerate: many base triplets or 'codons' stood for the same amino acid. Looking back, Gamow's contribution to the solution of the coding problem was to have set off the chase and maintained its momentum. In this he brought to bear a number of traits that were significant. Firstly, it did not bother him that he was an outsider trying to stick his neck into an unknown field. Secondly, he took the double helix for granted. He did not bother himself with whether the structure was right or wrong, the number of details which went into defining it, or how it could be verified, but asked instead, Where do we go from here? Thirdly, he had the ability to distinguish between those facts which were important for building a theory and those which could safely be ignored. Most importantly, Gamow possessed an infectious, almost manic, enthusiasm that touched everyone and made the breaking of the code a uniquely social enterprise. He founded the whimsically named RNA tie club, with twenty full members (one for each amino acid) and four associate members (one for each base). Club members were to communicate with each other about progress on the coding problem. One idea to emerge from the theoretical attack on the genetic code initiated by Gamow was a stroke of genius; it came

Gamow possessed an infectious, almost manic, enthusiasm that touched everyone and made the breaking of the code a uniquely social enterprise.



by way of an informal communication by Crick to the RNA tie club. In it he put forward the hypothesis that there had to be a small molecule, an adaptor, that acted as an intermediary between an RNA triplet and an amino acid.

A lasting contribution of the work on the coding problem was that it made the concept of *information* central to biology. There is an extreme, and controversial, version of the information theory approach to DNA, known as genetic determinism. It states that living creatures carry implicit representations of themselves in their DNA, as Gamow implied about cats in his letter to Watson and Crick. Today, most biologists would feel uncomfortable with this viewpoint.

Is that the end of the story? Well, not quite. There are open questions pertaining to the genetic code. For instance, we still do not know for certain whether the code is really arbitrary, a 'frozen accident' in Crick's words. An alternative hypothesis is that there is something about the nature of the adaptor molecule, or about the base triplet–amino acid link, that makes the code that we have an automatic consequence of the physics and chemistry of molecular interactions. A second question is related to the first and derives from the observation that the code is not truly universal. It is known, for example, that the same triplet of bases can 'mean' different things in different organisms. Therefore the code can change. This leads one to ask: Like so many other properties of living organisms, is the genetic code too a product of evolution by natural selection?

Suggested Reading

- [1] Brian Hayes, The Invention of the Genetic Code, *American Scientist*, Vol. 86, No. 1, pp. 814, Jan. Feb. 1998. <http://www.americanscientist.org/template/AssetDetail/assetid/20831;jsessionid=aaagLO6nDe-mbQ>
- [2] Stephen J Freeland and Laurence D Hurst, Evolution Encoded, *Scientific American*, pp. 84-91, April 2004. <http://www.sciam.com/issue.cfm?issueDate=Apr-04>
- [3] Apoorva Patel, Why genetic information processing could have a quantum basis, *Journal of Biosciences*, Vol. 26, No.2, pp. 145-151, June 2001. <http://www.ias.ac.in/jbiosci>
- [4] Martin Mulligan, The Genetic Code, www.mun.ca/biochem/courses/3107/Lectures/Topics/Genetic_code.html
- [5] Anand Sarabhai, After DNA at the MRC, *Journal of Biosciences*, Vol. 28, No.6, pp. 665-669, December 2003. <http://www.ias.ac.in/jbiosci/dec2003/contents.htm>

A lasting contribution of the work on the coding problem was that it made the concept of *information* central to biology.

Address for Correspondence

Vidyanand Nanjundiah
Indian Institute of Science and
Jawaharlal Nehru Centre for
Advanced Scientific Research
Bangalore 560 012, India.
Email: vidya@ces.iisc.ernet.in