CellPress

Correspondence

On Francis Crick, the genetic code, and a clever kid

Bob Goldstein

A few years ago, Francis Crick's son told me a story that I can't get out of my mind. I had contacted Michael Crick by email while digging through the background of the researchers who had cracked the genetic code in the 1960s. Francis had died in 2004, and I was contacting some of the people who knew him when he was struggling to decipher the code. Francis didn't appear to struggle often — he is known mostly for his successes — and, as it turns out, this one well-known struggle may have had a clue sitting just barely out of sight.

After co-discovering the structure of DNA in 1953 [1–3], Francis Crick spent more than a decade trying to decipher the code hidden in DNA [4]. The central problem, as Crick saw it, boiled down to "how a sequence of four things (nucleotides) can determine a sequence of twenty things (amino acids)" [5].

Crick knew that simple pairs of nucleotides couldn't encode the 20 core amino acids because pairs could only encode 16 possibilities (four possible nucleotides in the first position, times four possible nucleotides in the second position). So a sequence of three nucleotides — 4x4x4 or 64 possibilities — was likely to be a minimal word size in DNA.

Scientists' early attempts to solve the code relied on cryptography-based approaches because no nucleic acid sequences were known in the 1950s. Crick and many others aimed to build a cipher — a table of the DNA sequences that coded for each amino acid. In theory, if they knew the secret code for each amino acid and were then given a page of nucleic acid sequence written out, they would be able to write out the protein it encodes.

Among Crick's imagined solutions was one in which a correct reading frame in the DNA might be specified if only certain three-nucleotide-long words in the DNA made sense. This was an appealing solution because, by some simple math stemming from Crick's assumptions, it happened that there would be exactly 20 possible 'sense' words, exactly matching the number of core amino acids [6].

Sydney Brenner, who worked with Crick on the coding problem, confirmed to me the allure of solutions that happened to produce exactly 20 words. "20 became the magic number", Brenner said. "If it gives 20, the assumption was, you see, well, there has to be something in it" (interview with the author, 15 September 2014). However, there were few clues that could help Crick and Brenner distinguish which of the theories that produced exactly 20 words might be correct, if any.

Five years into codebreaking, Crick expressed exasperation toward even his own then-favorite theory, writing, "Thus we have deduced the magic number, twenty … Nevertheless, I must confess that I find it impossible to form any considered judgment of this idea. It may be complete nonsense, or it may be the heart of the matter. Only time will show."

Time showed that the number 20 was a complete red herring. Biochemists revealed that the real code has not 20 but 64 three-lettered words, and loads of redundancy. Most of the 20 amino acids are encoded by two to six different codons. As early as 1955, Crick had imagined redundant codes [7], yet he still clung to solutions that produced exactly 20 words.

I had contacted Michael Crick upon learning that Francis had an interest in secret codes even before he had developed an obsession for DNA. In 1950, when Michael was a 10-year-old boy at boarding school, Francis mailed his son a birthday gift: a history of cryptography entitled *Codes and Ciphers*. Michael laid on his dorm room bed and eagerly read the book. Then he devised his own secret code: a simple substitution code in which the letters of the alphabet and the numbers 0–9 were replaced with symbols.

Michael brought his secret code home with him when he returned for Christmas to the family's tiny apartment in Cambridge. Francis and a visiting mathematician friend, Georg Kreisel, challenged Michael, telling him that they could crack his code if given a page of text written out in the code.

Michael wrote out a page of text for them. Francis and Georg worked together in the living room to solve the code. "At first they were rather cocky," Michael told me, "but they got more and more frustrated and finally gave up after about two days."

The trick that Michael had used: he had hidden high-frequency letters (letters such as E, T, and S) with redundancy — he used *multiple* symbols to encode each of these letters [8]. As a result, if his father and Kriesel had looked for the most frequently written symbols and guessed that these might encode each of the frequently used letters, their efforts would have been in vain. Before the structure of DNA, and years before DNA codebreaking efforts had begun, Francis Crick's 10-year-old child had stumped him in the same way that the genetic code would later stump him — using redundancy.

I was curious to find out when the shoe had dropped for Francis. Had Michael told his father his trick while Francis was working on the DNA code? Or long after it had become clear to Francis that redundancy had made his search for precisely 20 word codes a fruitless path? I envisioned Francis smacking himself in the head and wondering if he would have approached the coding problem differently had redundancy been a more prominent possibility in his mind. So I asked Michael how his father had reacted when he revealed his trick.

"I never told either my father or Kreisel anything — it being my intention to keep the code secret."

REFERENCES

1. Watson, J.D., and Crick, F.H.C. (1953). A structure for deoxyribose nucleic acid. Nature *171*, 737–738.
2. Wilkins, M.H.F., Stokes, A.R., and Wilson, H.R. (1953). Molecular structure of deoxypentose nucleic acids. Nature *171*, 738–740.
3. Franklin, R.E., and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740–741.
4. Cobb, M. (2015). Life's Greatest Secret: The Race to Crack the Genetic Code (New York: Basic Books).
5. Crick, F.H.C., Griffith, J.S., and Orgel, L.E. (1957). Codes without commas. Proc. Natl. Acad. Sci. USA *43*, 416–421.
6. Crick, F.H.C. (1958). On protein synthesis. Symp. Soc. Exp. Biol. *12*, 138–163.
7. Crick, F.H.C. (1955). On degenerate templates and the adaptor hypothesis. An unpublished note for the RNA Tie Club.
8. Gambino, M. (2013). Document Deep Dive: Francis Crick Explains the 'Secret of Life'. Smithsonian online, September 26, 2013. https://www.smithsonianmag.com/history/document-deep-dive-francis-crick-explains-secret-life-180947946/.

Biology Department, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
E-mail: bobg@email.unc.edu