Establishing the Triplet Nature of the Genetic Code

Charles Yanofsky[1,*]
[1]Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA
*Correspondence: yanofsky@cmgm.stanford.edu
DOI 10.1016/j.cell.2007.02.029

In 1961, Crick, Barnett, Brenner, and Watts-Tobin (Crick et al., 1961) designed an elegant experimental strategy to determine the nature of the genetic code. Remarkably, they reached the correct conclusion despite the absence of technology to analyze and compare DNA and protein sequences.

The initial description of the linear duplex structure of DNA by James Watson and Francis Crick in the early 1950s was truly a monumental advance. At that time, technology did not exist for isolating a gene, determining its nucleotide sequence, or relating such a sequence to the amino acid sequence of the corresponding protein. Messenger RNA had not

In this Essay, Charles Yanofsky discusses his favorite pioneering science paper and points out what we can learn from the scientific accomplishments of the past.

been discovered, and very little was known about protein synthesis. It was evident that there were many different proteins in the cells of each organism, and it was becoming apparent that most proteins consist of a linear sequence of amino acids. George Beadle and Edward Tatum (Beadle and Tatum, 1941) in their pioneering studies with the fungus *Neurospora* had suggested a 1:1:1 relationship between gene, enzyme, and biochemical reaction. But how the nucleotide sequence of each gene was related to the amino acid sequence of its encoded protein remained a major unanswered question.

Setting the Stage
In their landmark 1961 *Nature* paper entitled "General Nature of the Genetic Code for Proteins," Francis Crick, Leslie Barnett, Sydney Brenner, and Rich-

ard Watts-Tobin (Crick et al., 1961) finally solved the riddle. They concluded correctly that the genetic code is a triplet code, the code is degenerate, triplets are not overlapping, there are no commas (although introns were subsequently discovered), and each nucleotide sequence is read from a specific starting point. This paper has long been one of my favorites because at the time it appeared I did not think the existing knowledge and experimental procedures were sufficient to allow anyone to deduce the general nature of the genetic code. Indeed, in 1961, the only analytical procedure that could be used to order the presumed nucleotide sequence of a gene was fine structure genetic mapping using mutants with alterations in that gene. The mutationally altered sites mapped by this procedure were presumed to be the sites of nucleotide changes. Fortunately for Crick and his coworkers, Seymour Benzer in the late 1950s had developed an elegant assay using mutations of the rII region of the T4 phage (which has two adjacent genes, called cistrons *A* and *B*) (Benzer, 1959, 1960). With this system, Benzer provided the first detailed fine structure map of a genetic region, in this case from the phage genome. Despite the inability to sequence DNA, Crick and his colleagues became convinced that they could use mutagenesis and genetic recombination with

the T4 rII system to map altered sites in this genetic region and to establish the general nature of the genetic code.

Benzer's map of the T4 rII region was consistent with the conclusion that each gene is comprised of a linear sequence of nucleotides, each of which could undergo a heritable change that could alter the protein product. That polypeptides also consist of linear sequences of amino acids had been established in the early 1950s by Fred Sanger (Sanger and Tuppy, 1951) and others. Then, in the late 1950s, Vernon Ingram showed that a mutant human hemoglobin polypeptide had but a single amino acid change (Ingram, 1958). The logical conclusion drawn from these and related studies was that each gene consists of a unique linear sequence of nucleotides, and that each sequence is translated into a unique linear sequence of amino acids. If this interpretation was correct, then there had to be a "genetic code" relating the nucleotide sequence of each gene to the amino acid sequence of its encoded protein.

Crick had been concerned with the genetic code for some time and had published several papers describing his thoughts on this subject (Crick et al. 1957; Crick, 1958). He proposed the "adaptor" hypothesis—that specific molecules (subsequently identified as tRNAs) were responsible for "translating" a specific nucleotide sequence into a specific amino acid (Crick, 1958). Sydney Brenner had also given the genetic code seri-

Figure 1. Determining the Triplet Nature of the Genetic Code
Shown is a suppressor mutant analysis using the rII *B* region of the genome of phage T4. The FC0 mutant is presumed to have a single nucleotide addition (+ A) to the wild-type sequence. The FC0 suppressed mutant has an additional change, a single nucleotide deletion (– C). The suppressed mutant's phenotype is wild-type (rII+). Combining two single nucleotide additions in rII gives an rII– mutant phenotype. However, combining three nucleotide additions gives an rII+ wild-type phenotype. Given that functional rII proteins are produced in triple addition (+) mutants, and also in triple deletion (–) mutants, the genetic code must be a triplet code with three nucleotides coding for each single amino acid.

ous thought. He had compared the known amino acid sequences of proteins and concluded that the genetic code could not be overlapping (one of several possibilities) because each amino acid could be located adjacent to each of the other 19 amino acids (Brenner, 1957).

A Landmark Paper
In the initial section of their 1961 paper, Crick et al. described what they believed were reasonable alternatives regarding the nature of the genetic code. Two of these potential "codes"—overlapping and nonoverlapping—were presented in their Figure 1. The strategy they devised for determining which proposed code was correct was based on their belief that one class of mutagens, the acridine dyes (such as proflavin), caused single base pair additions or deletions in the DNA (Brenner et al., 1958, 1961b). This was not the prevailing view at that time but was supported by their data (Brenner et al., 1961b). Consistent with this interpretation came the observation that acridine-induced mutations in DNA had an unusual characteristic: they were non-leaky (that is, the mutation led to complete loss of function of the gene). This suggested that these mutations possibly prevented proper translation of the coding region downstream of the site of the mutation. Crick and

his colleagues further reasoned that any mutation caused by a single base pair addition might be suppressed by a nearby mutational change of the opposite type, a single base pair deletion. This second change would restore the "natural" reading frame, downstream of the site of the second mutation. Thus, only the polypeptide segment specified by the DNA sequence between the sites of the two mutational changes would be altered.

They tested these predictions using a proflavin-induced T4 rII mutant that they designated FC0. They arbitrarily assumed that this mutant arose as a result of a single base pair addition (termed "+"). This mutation mapped to a site in a specific segment of the *B* cistron of the T4 rII region, which Benzer had shown was not essential for rII B function. The significance of selecting this genetic region for their analyses was the expectation that any altered amino acid sequence specified by the DNA segment between + (base pair addition) and – (base pair deletion) changes, or vice versa, in the selected B region would not be deleterious. In fact, when they performed their initial analyses with FC0 selecting for mutations restoring rII+ function, they observed many second-site suppressor changes; these mapped to neighboring sites on either side of the FC0 mutation site. They

then separated these suppressor changes from the FC0 mutation by genetic recombination and observed that each suppressor change, when present alone in rII *B*, also had a mutant phenotype (see Figure 1). These separated suppressor changes were generally non-leaky, consistent with the expectation that each was due to a base pair deletion.

The investigators next isolated suppressors of these suppressors; these should be base pair additions, given that they reversed the mutant phenotype of the previously isolated and presumed "deletion" suppressors. Combining mutational changes in the various sets, they observed that there were only two classes: those that appeared to have a single base pair addition, and those that appeared to have a single base pair deletion. Thus any single mutational change in one class, when combined with a single mutational change in the same class, would be expected to produce a mutant phenotype, whereas when combined with a neighboring mutational change in the second class, the rII+ phenotype might be observed. This is precisely what they found: most suppressor mutations were located at sites relatively close to the site of the primary "suppressed" mutation. It appeared that misreading resulting from a single base pair addition could only be corrected by

deletion of a neighboring single base pair deletion. This suggested the presence of a triplet nucleotide code and that some coding triplets could not be translated, perhaps serving as translational stops. The authors also observed that some of their presumed +/− combinations had a pseudo-wild-type phenotype, consistent with the expectation that some of the amino acid sequences specified by the DNA segment between the sites of the two mutational changes would contain an amino acid that was only partly functionally acceptable at that position.

They next performed an additional analysis, focusing on their principal objective, determining the general nature of the genetic code. They reasoned that if the genetic code is a triplet code then combining three + mutational changes, or three − mutational changes, would produce a phage with a wild-type or pseudo-wild-type phenotype. This is exactly what they found (see Table 3 in Crick et al., 1961). In these triple mutants, one amino acid was presumably added to or deleted from the B protein. These findings provided compelling evidence in support of their conclusion that the genetic code is a triplet code.

Crick, Barnett, Brenner, and Watts-Tobin designed one additional test of the logic of their interpretations, exploiting what appeared to be a gene fusion. They introduced presumed + and − mutational changes into the A cistron of a special deletion mutant called 1589 in which the nondeleted segment of the rII A cistron was predicted to be fused to the remaining portion of the B cistron. However, this deletion mutant retained the function of the B protein suggesting that the remaining section of B cistron DNA was most likely translated in its proper reading frame, producing a fused A–B polypeptide retaining B function. Normally, mutations in either the A cistron or B cistron have no effect on expression of the other cistron, consistent with the view that rII A and rII B are separate genes. Introducing + or − mutational changes into the remaining portion of rII A of the 1589 deletion mutant, as predicted, eliminated

rII B activity (see Figure 5 in Crick et al., 1961). However, when + and − changes were combined in the rII A region of 1589, B activity was restored. These findings were consistent with the authors' conclusion that the nucleotide sequence of a gene is translated linearly, beginning at a fixed starting point and proceeding until a translational stop signal is encountered.

Next, they performed additional experiments to test whether the DNA-protein coding ratio could be six nucleotides per amino acid rather than three nucleotides per amino acid, which they favored based on their data. All of their findings were consistent with the conclusion that each acridine-induced mutation involved a single base pair addition or deletion, and not two base pair additions or deletions. However they recognized that these studies did not completely rule out the six nucleotide possibility. They also argued that their results were most consistent with the code being degenerate. Given that there are 64 (4 × 4 × 4) different triplet sequences, and only 20 amino acids, 44 of these triplets would presumably encode stop sequences, or would serve some other function, if the code were not degenerate. If 44 triplets are stop codons, the authors believed that they would have experienced greater difficulty restoring rII function by combining + and − mutational changes. As many of their presumed addition and deletion mutational changes were rescued by suppressor mutations, a stop codon could not have been introduced between the two. The most likely explanation was that the genetic code is highly degenerate.

The experimental analyses described in their paper are extremely convincing. In the next-to-last paragraph of their article, they mention the announcement by Marshall Nirenberg at the 1961 Biochemical Congress in Moscow that he and Matthaei had added an RNA polyuridylic acid to an in vitro cell-free protein synthesis system and had produced polyphenylalanine. This implied that some sequence of uridines codes for phenylalanine, and, more importantly,

that synthetic RNAs could be translated into proteins using this cell-free system. Crick et al. also referred to the classic paper by Nirenberg and Matthaei (1961) published several months before their paper appeared in *Nature*. In this paper, Nirenberg and Matthaei mention in a "Note added in proof" that "polycytidylic acid specifically mediates the incorporation of L-proline into a TCA-precipitable product." Crick et al.'s 1961 paper is outstanding because they deduced correctly the general nature of the genetic code despite their inability to decipher the genetic code by direct experimental analysis.

Final Comment

It was obvious that the approaches introduced by Nirenberg in 1961 would not only establish the nature of the genetic code, but would permit the identification of the nucleotide triplets coding for each amino acid. Therefore, one might ask whether the studies on the nature of the genetic code described in Crick et al.'s 1961 paper have had any significant scientific impact. Would it have been wiser for Crick and coworkers to suspend their genetic studies on the nature of the code and wait until existing technology had advanced to a stage when they could apply it themselves in solving this important problem? Of course the answer is no—their analyses were brilliant, and their conclusions were invaluable additions to our knowledge of the informational content of DNA. Their results immediately changed my thoughts regarding the genetic code, gene-protein colinearity, and suppression, all subjects that my group was actively studying at that time. I suspect their conclusions had a great impact on many scientists engaged in research in which the general nature of the genetic code was of concern.

The 1961 paper by Crick et al. is an outstanding example of the use of thought and logic to solve basic biological problems. In my opinion, it is a superb paper to assign to students in courses because it illustrates how combining knowledge and wisdom can provide answers to important scientific questions.

Note

The experiments presented in Crick et al.'s 1961 paper have been described in great detail by Jeffrey H. Miller in his book *Discovering Molecular Genetics* (Miller, 1996). Another outstanding molecular geneticist, Jonathan Beckwith, has written an introduction to this 1961 Crick paper in the book *Microbiology, A Centenary Perspective* (Beckwith, 1999). A recently published book entitled *Francis Crick: Discoverer of the Genetic Code*, written by Matt Ridley, focuses on Francis Crick's life and career (Ridley, 2006).

REFERENCES

Beadle, G., and Tatum, E.L. (1941). Proc. Natl. Acad. Sci. USA *27*, 499–506.

Beckwith, J. (1999). In Microbiology, A Centenary Perspective. editors: Joklik, W.K., Ljungdahl, L.G., O'Brien, A.D., von Graevenitz, A., and Yanofsky, C. (Washington, DC: ASM Press), p. 384.

Benzer, S. (1959). Proc. Natl. Acad. Sci. USA *45*, 1607–1620.

Benzer, S. (1960). Proc. Natl. Acad. Sci. USA *47*, 403–415.

Brenner, S. (1957). Proc. Natl. Acad. Sci. USA *43*, 687–694.

Brenner, S., Benzer, S., and Barnett, L. (1958). Nature *182*, 983–985.

Brenner, S., Barnett, L., Crick, F.H.C., and Orgel, A. (1961b). J. Mol. Biol. *3*, 121–124.

Crick, F.H.C. (1958). Symp. Soc. Exp. Biol. *12*, 138–163.

Crick, F.H.C., Griffith, J.S., and Orgel, L.E. (1957). Proc. Natl. Acad. Sci. USA *43*, 416–421.

Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). Nature *192*, 1227–1232.

Ingram, V. (1958). Biochim. Biophys. Acta *28*, 539–545.

Miller, J.H. (1996). Discovering Molecular Genetics (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Nirenberg, M.W., and Matthaei, J.H. (1961). Proc. Natl. Acad. Sci. USA *47*, 1588–1602.

Ridley, M. (2006). Francis Crick: Discoverer of the Genetic Code (New York: Atlas Books).

Sanger, F., and Tuppy, H. (1951). Biochem. J. *49*, 481–490.