Revisiting a Breathtaking Publication in the History of Molecular Biology

Richard Losick

Department of Molecular and Cellular Biology, *Harvard University, Cambridge, MA 02420, USA*

Correspondence to : *losick@mcb.harvard.edu,*
https://doi.org/10.1016/j.jmb.2025.169415
Edited by M. Gottesman

The birth of molecular biology in the second half of the twentieth century was driven by a series of seminal publications that transformed our understanding of the nature of the genetic material and how living things work [1]. Among these advances none was more breathtaking in its simplicity and depth of insight than the 1961 report of Francis Crick, Leslie Barnett, Sydney Brenner, and Richard J. Watts-Tobin on the "*General Nature of the Genetic Code for Proteins*" [2]. The authors' elegant experiments and theoretical analysis led them to correctly surmise that the code is triplet (or a multiple thereof), non-overlapping and degenerate. Here we revisit this classic paper because we believe its interpretation has been oversimplified over the years and should be seen in the light of subsequent DNA sequence analysis of the mutants generated during the investigation into the nature of the genetic code.

The authors took advantage of the high-resolution genetic system of Seymour Benzer [3]. Benzer exploited the inability of mutants of the *rIIA* and *rIIB* genes (cistrons) of phage T4 to form plaques on a lawn of *E. coli* cells of strain K12 (a phage λ lysogen) while retaining the capacity to form plaques on a lawn of strain B. Indeed, the ability of *rII* mutants to form plaques on strain K12 was reduced by greater than 10^{8th} compared to plaque formation on strain B. This high sensitivity enabled Benzer to observe recombination frequencies as low as 0.013, which showed that the mutations were close together. But the fact that he never observed lower frequencies, even though he could have detected frequencies that were orders of magnitude lower, showed that the genetic material is granular. Benzer had, in effect, pushed genetics to the level of nucleotides. Importantly, this capacity to observe recombination between closely linked mutations made it possible for Crick et al. [2] to construct the double and triple mutants of the *rIIB* gene that is at the heart of their historic publication. (This seemingly arcane interplay between T4 and its host is now understood as an early example of what is recognized as vast repertoire of immunity/anti-immunity systems operating in the microbial world.)

The starting point for the investigation of Crick et al. [2] was the *rIIB* mutation P13, which had been generated with the acridine mutagen proflavin and which Crick renamed FC0 [4]. Many of the experiments with FC0 and suppressor mutations generated with it were carried out by Crick himself, the one time the greatest theoretician of molecular biology did experiments with his own hands.

Unlike mutagens that induce base pair substitution mutations, which can result partially functional mutant proteins, proflavin, a flat molecule that intercalates between base pairs, induces non-leaky mutations. Crick et al. [2] inferred that such a mutagen would induce base pair insertions and deletions and that such mutations would shift the reading frame, altering the coding sequence downstream of the mutation. Indeed, such mutations are now referred to as frameshift mutants. The authors arbitrarily assumed that FC0 was an insertion, which they designated by the symbol "+". But as they noted, nothing would have changed in their inferences had FC0 been a deletion of one or more base pairs. Starting with FC0 they isolated suppressor mutations, which restored growth on K12. These suppressor mutations were designated as a type "–", which they hypothesized restored the correct reading frame. Among these type – suppressors was FC1, which was used to isolate additional suppressors, which were desig-

nated as type +. The FC family of mutations, suppressors and suppressors-of-suppressors were located in the left-hand region of *rIIB*, which as Benzer had shown, was not essential for function and hence could tolerate amino acid substitutions as long as the original reading frame was restored downstream. Thus, and whereas combining a + with a − restored the wildtype phenotype (plaque formation on a lawn of *E. coli* K12), a combination of two + type mutations or two − type mutations did not. In other words, *rIIB* could tolerate a short stretch of codon changes between nearby + and − mutations but combining a + with a + or a − with a – would not restore the proper reading frame downstream and hence *rIIB* function would remain severely impaired.

To generalize their results from the simplified model they used to explain their experimental design, they employed modulo arithmetic. According to this formalism, the symbol + represents +**m** modulo **n** and − represents −**m** modulo **n**, where **m** is an integral number of base pairs, positive or negative, and **n** is the number of base pairs per codon. In modulo arithmetic, **m** modulo **n** is the remainder when **m** is divided by **n**. Thus, for example, if **m** is +1 or +2 and **n** is 3, then **m** modulo **n** would represent the insertion of one or two base pairs. Likewise, −**m** could be −1 or −2, representing the deletion of one or two base pairs. Meanwhile, **n** could be 3 or 4 or 5 bases.

The formalism was needed to ensure that the + and − symbols were not literally taken to indicate the insertion or deletion of a single base pair. Nonetheless, Crick et al. [2] went on to state that "*. . .we have convincing evidence that the coding ratio is in fact 3 or a multiple of 3. . ..This we have obtained by constructing triple mutants of the form (+ with + with +) or (− with − with −)*." This evidence is presented in Table 3 ("Triple Mutants Having a Wild or Pseudo-Wild Phenotype" [2]), the most striking and impactful (but also over interpreted, e.g. [5]) of the figures and tables in their publication. Table 3 showed that combining FC0 with two additional + mutants to create a triple mutant (+ and + and +), restored the wildtype phenotype. Likewise, combining FC1, a – type, with two additional – types to create a triple (− and − and −) similarly restored a wildtype phenotype.

Crick, Barnett, Brenner and Watts-Tobin [2] carried out their experimental and theoretical analysis before DNA sequencing had been invented more than a decade later. However, by 1987, Larry Gold and his colleagues [6] had applied DNA sequencing technology to determine the nature of several of the mutants in the Crick et al. [2] publication. Most noteworthy, but overlooked, was the finding that the type – mutation FC1 was an insertion of two base pairs (+AC) rather than a deletion of one [6]. Meanwhile, the type – mutations FC21 and FC23, which were also sequenced, were determined to be single base pair deletions (−A and −T, respectively) [6]. Thus, in Table 3 [2], the combination of FC1 with two single base-pair deletions would have canceled out the AC insertion. That is, +AC −A −T (+2 −1 −1) simply equals zero. Thus, there was no change in the length of the DNA over the region covered by the insertions and deletions and hence no net insertion or deletion of coding units. Despite the **m** modulo **n** formalism in which **m** would be −1 for all three mutations, the observed wildtype phenotype would have been consistent with a code of three, four or five bases. That is, given that the overall sum of the base pair insertions of the three mutations is zero, **0** modulo **n** would be 0 whether **n** is 3, 4, or 5.

Figure 1 summarizes these considerations as a cartoon. Panel a represents a triplet code as repetitions of ABC with the reading frame indicated in bold. Panel b shows the consequence of introducing three −1s (a −A, a −B and a −C). As can be seen, the proper reading fame (bold) is restored after the third −1. If, however, the code is quadruplet (represented by ABCD in panel c), then the introduction of three −1s does not restore the ABCD-reading frame. Now let's consider the case in panel d in which a +2 (+*AB*) is combined with two −1s (a −B and a −C). As can be seen, the ABC-reading frame is restored downstream after the second −1. But as shown in panel e, this is also the case when the open-reading frame is quadruple (ABCD). In other words, the combination of a +2 with a −1 and a −1 restores the downstream open reading frame whether the code is triplet or quadruple.

Figure 2 revisits the FC1 FC 21 FC 23 triple mutant in light of our present-day knowledge of the genetic code and the sequence results of Gold and co-workers [6]. As explained in the legend, the FC1 insertion of two base pairs (+AC) shifts the reading frame to codons for the eleven amino acids indicated by single letter abbreviations in the second line. Next, the FC21 deletion of an A creates a codon for tryptophan (W) shown in the third line. Finally, the FC23 deletion of a T restores the original reading frame (top line). Thus, the effect of combining a +2 base pair insertion with two single base pair deletions is fully consistent with a triplet (n = 3) code. However, and as this commentary argues, combining the +AC insertion with two single base pair deletions would have also restored the correct reading frame if the code (**n**) was 4 or 5.

Nonetheless, and despite these considerations, Crick et al. [2] were right but for a different reason based on the double mutants of Table 2, which presents the results of combining type – mutations with type + mutations. Specifically, when FC1 (a − type mutation) was separately combined with five FC + type mutations (0, 38, 40, 41, 58), the resulting double mutants exhibited wildtype phenotypes. Among these five + type mutations only the sequence of FC38 was determined (a single T insertion) [6]; thus

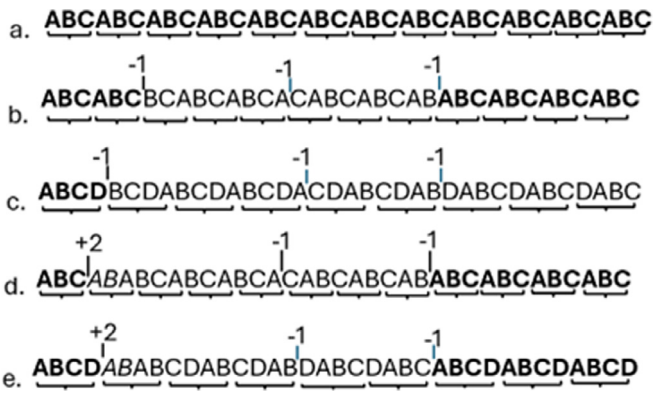

Figure 1. The cartoon compares the effects of −1 −1 −1 and +2 −1 −1 triple mutations on the reading frame when the code is triplet (depicted as ABC) and quadruple (ABCD). Panel a is a series of triple codons in the same reading frame (bold letters). Panel b is a series of triple codons in which −1 −1 −1 restores the original reading frame (bold). Panel c is a series of quadruple codons in which −1 —1 −1 fails to restore the original reading frame. Panel d is a series of triple codons in which +2 −1 −1 restores the original reading frame (bold). Panel e is a series of quadruple codons in which +2 −1 −1 restores the original reading frame (bold).

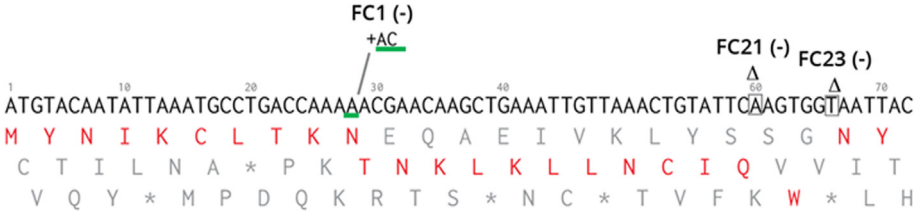

Figure 2. The figure is based on the sequence determinations of Gold and co-workers [6] and was kindly provided by Sean Eddy. Highlighted in green is the site of insertion of the FC1 mutation +AC. Note that in proceeding from left to right, the +AC insertion shifts the reading frame from the codons labeled in red (using the single letter abbreviation for amino acids) in the top line to those in the second line. Meanwhile, the FC21 mutation −A shifts the reading frame to the tryptophan codon (W) in the bottom line. Finally, the FC23 mutations −T restores the wildtype reading frame as indicated by asparagine (N) and tyrosine (Y) codons in the top line.

FC1 (+AC) combined with FC38 (+T) restored the reading frame and did so by inserting an extra codon in *rIIB*. Ergo, Crick et al. [2] were correct in inferring that the code is triplet but not based on the phenotype of the FC1 21 23 triple mutant of Table 3 alone but rather in combination with the phenotype of the double mutant of Table 2.

In light of this finding, it would be of the utmost interest to know the sequence of FC0. Unfortunately, FC0 is not among the mutants sequenced by Gold and colleagues [6]. Could it have been a two base pair deletion rather than a one base pair insertion? [Recall that FC0 (P13) was generated with an acridine, and acridines are known to cause multi base pair insertions and deletions [7]. If so, and once again, Table 3 alone would not have been the basis for inferring that the code is triplet. Rather, the creation of a triple mutant in which such a hypothetical two base pair deletion was combined with two single base pair insertions would have resulted in no net alteration to the reading frame (−2 + 1 + 1 = 0). Sadly, efforts to track down the FC0 have so far failed.

Brenner himself stated that "*…this was a real 'house of cards' theory. You had to buy everything. You couldn't take one fact and let it stand by itself and say the rest could go. Everything was so interlocked. You had to buy the plus and minuses, you had to buy the barriers, you had to buy the triplet phase, and all these went together. It was the whole that explained it, and if you attacked any one part of it the whole thing fell apart. So it was an all or nothing theory* [8]" Nonetheless, at the end of the day, the theory was spot on, and Crick et al. [2] correctly predicted that the genetic code is triplet and nonoverlapping based on experiments of the utmost simplicity.

DATA AVAILABILITY

No data was used for the research described in the article.

DECLARATION OF COMPETING INTEREST

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The author thanks Sean Eddy for extensive discussions. The author also thanks Adam Strandberg and Ari Hilibrand for advice on modular arithmetic and Craig Hunter and Polina Kehayova for helpful advice on the manuscript.

Received 11 July 2025;
Accepted 28 August 2025;
Available online 3 September 2025

References

[1]. Judson, H., (1979). The Eighth Day of Creation: Makers of the Revolution in Biology. Simon & Schuster, New York.

[2]. Crick, F.H., Barnett, L., Brenner, S., Watts-Tobin, R.J., (1961). General nature of the genetic code for proteins. *Nature* **192**, 1227–1232. https://doi.org/10.1038/1921227a0. PMID 13882203. S2CID 4276146.

[3]. Benzer, S., (1955). Fine structure of a genetic region in bacteriophage. *PNAS* **41**, 344–354. https://doi.org/10.1073/pnas.41.6.344. PMC 528093. PMID 16589677.

[4]. Barnett, L., Brenner, S., Crick, F.H., Shulman, R.G., Watts-Tobin, R.J., (1967). Phase-shift and other mutants in the first part of the *rIIB* cistron of bacteriophage T4. *Philos. Trans. R. Soc. London Ser. B, Biol. Sci.* **252**, 487–560 https://www.jstor.org/stable/2416779.

[5]. Yanofsky, C., (2007). Establishing the triplet nature of the genetic code. *Cell* **128**, 815–818.

[6]. Shinedling, S., Gayle, M., Pribnow, P., Gold, L., (1987). Mutations affecting translation of the bacteriophage T4 *rIIB* gene cloned in *Escherichia coli. Mol. Gen Genet.* **207**, 224–232.

[7]. Hoffman, G.R., Deschênes, S.M., Manyin, T., Fuchs, R., (1996). Mutagenicity of acridines in a reversion assay based on tetracycline resistance in plasmid pBR322 in *Escherichia coli. Mutat. Res.* **351**, 33–43.

[8]. Brenner, S., (2001). My Life in Science. As Told to Lewis Wolpert. BioMed Central Limited, p. 97.