

# The RNA code: Nature's Rosetta Stone

C. Thomas Caskey<sup>a,1</sup> and Philip Leder<sup>b,1</sup>

<sup>a</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; and <sup>b</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115

Marshall Nirenberg and Heinrich Matthaei initiated their biochemical approach to elucidating the genetic code in 1959, 6 years following the discovery of the double helix structure of DNA (1). At the time, geneticists were using T4 bacteriophage mutagenesis/reversion of plaque morphology toward the

same goal by inferring that three mutations were needed to specify an amino acid (2). The genetic code challenge was to link the nucleic acid sequence composed of four bases to translation of 20 amino acids. Nirenberg presented his breakthrough announcement twice at the 1961 International Congress of

Biochemistry (ICB) in Moscow: the first presentation to a small audience and the second, at the urging and support of Francis Crick, to the entire congress.

The first biochemical innovation was the development of a simple microbial whole cell extract, which synthesized radioactive proteins programmed by RNA. After demonstrating naturally occurring RNA could stimulate *in vitro* protein synthesis, Nirenberg and Matthaei used synthetic RNA homopolymers to stimulate homopeptides synthesis *in vitro*. The first code association occurred when RNA made only of uracil (poly U) was found to direct the synthesis of only polyphenylalanine. This was the ICB 1961 announcement that was published in detail in the now classic paper in PNAS (3). In a series of papers that followed, many of them also in PNAS, additional code associations were made by using RNA polymers of mixed composition to program radioactive peptide synthesis (1 radiolabeled and 19 unlabeled amino acids). This spectacularly simple approach to the RNA code fostered a highly competitive race, rather than a collaborative environment, with Severo Ochoa, a renowned scientist in RNA polymer enzymology. In the pursuit of breaking the genetic code, Nirenberg's fellow National Institutes of Health investigators (Maxine Singer, Leon Hepple, and Bob Martin) put on hold their research programs to collaborate and accelerate the discovery of the genetic code because they shared in the excitement and understood well the overriding significance of this pursuit. Their collective effort with the Nirenberg fellows (O. W. Jones, J. Trupin, F. Rottman, and C. O'Neal), rapidly determined the code's base content but not the sequence order (4).

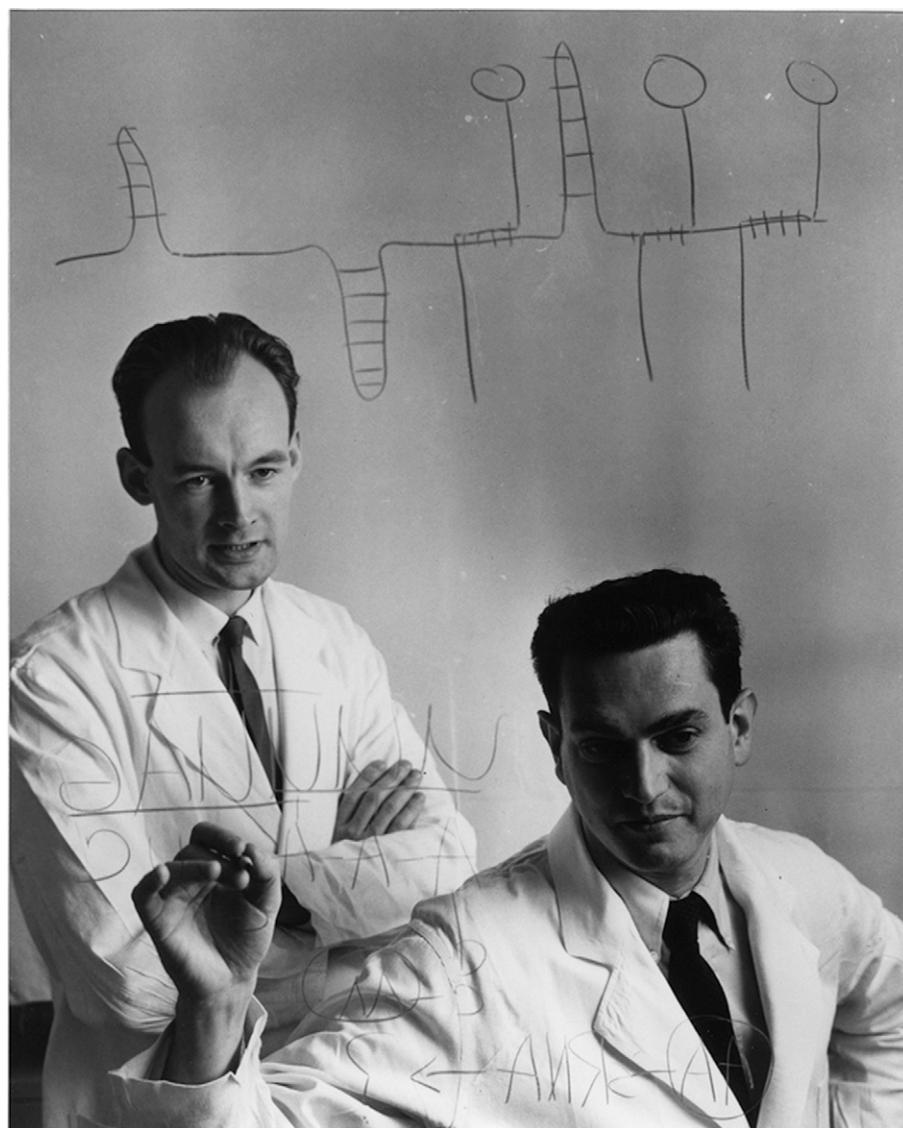
There were limits regarding what one could learn from the poly U-type experiment. It remained, for example, to show directly that a code word consisted of three

Author contributions: C.T.C. and P.L. wrote the paper.

The authors declare no conflict of interest.

This article is part of the special series of PNAS 100th Anniversary articles to commemorate exceptional research published in PNAS over the last century. See the companion article, "The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides" on page 1588 in issue 10 of volume 47, and see the Inner Workings article on page 5760.

<sup>1</sup>To whom correspondence may be addressed. E-mail: tcaskey@bcm.edu or leder@receptor.med.harvard.edu.



Heinrich Matthaei (Left) and Marshall Nirenberg (Right) collaborated in the now classic experiment demonstrating protein synthesis from synthetic poly-U that provided the initial clue to breaking the genetic code. This photograph was taken circa 1962 as they continued to decipher additional elements of the code. (Courtesy of Marshall W. Nirenberg from the National Library of Medicine).

bases and that the order of bases within that triplet conveyed specificity. Of course, this had to be done quickly because a second competition was now fueled with Professor Gobind Khorana, an extraordinary chemist at the University of Wisconsin, who was ready with the material and intellectual resources to test each of the possible 64 triplet codons for their activity. The identification of the triplet codons was made by a second in vitro assay (P. Leder and S. Peska), which measured the binding of radioactive aminoacyl transfer RNA (tRNA) to ribosomes by a triplet (not doublet) synthetic RNA molecule (codon) (5). This radioactive complex bound to nitrocellulose membranes creating a simple and rapid means of codon/amino acid assignments. This approach broke the genetic code. The challenge for the Nirenberg Laboratory was how to synthesize the 64 triplets composed of four bases (A, G, U, and C). This was enabled by the commercial availability of 16 RNA doublets. Triplets were synthesized by several methods (three enzymologic and one chemical). The enzymes used were polynucleotide phosphorylase (P. Leder), T1 nuclease, and RNase A (M. Bernfield), and in addition, organic synthesis was utilized (M. Wilcox and R. Brimacombe). All methods added the third 3' base to the 16 doublets, yielding 64 triplets. The code was confirmed to be triplet and discovered to be degenerate (6) (several triplets encoding a single amino acid). Independent research on tRNA isoaccepting molecules provided the evidence for the wobble hypothesis. Francis Crick proposed a single tRNA's anticodon had the capacity to form alternative base pairs, enabling recognition of more than one codon (7).

The elucidation of the three codons for peptide chain termination required an additional in vitro assay: release of a radioactive pseudo peptide (*N*-formylmethionine) from preformed [f-met-tRNA·AUG·ribosome] complexes (8). The formyl-methionine (pseudo) peptide due to *N*-blocked (terminus) cleavage from the tRNA required one of three codons, UAA, UGA, or UAG, thus completing the assignment of 64 codons to either an amino acid or termination signal. The ter-

mination codon recognition molecules were discovered to be proteins, termed peptide release factors (RF). The bacterial RF1 recognized UAA and UAG, whereas RF2 recognized UAA and UGA. A single mammalian RF protein recognized all three codons (E. Scolnick, J. Goldstein, A. Beaudet, and T. Caskey) (9) Peptide cleavage from the

## Is there a basic chemical reason for [the genetic code], or is it to some degree a matter of historical chance? My personal belief is that there is an underlying meaning for this and that it will be found.

—Marshall Nirenberg, *On the Translation of the Genetic Code*, 1968

tRNA required the ribosomal peptidyl transferase (10).

Using these biochemical in vitro approaches, Nirenberg deciphered the code within 7 years, from his first experiments in 1959 to the 64 codon amino acid and termination assignments. The code was found to be universal for nearly all living things: from microbes to eukaryotes and to plants, fungi,

and animals (11). These results provided the Rosetta Stone for many of the major discoveries that have followed and that would not have been possible without deciphering how nucleic acid sequences are translated to protein sequences. The discovery of fragmented genes, RNA splicing, RNA editing, and viral (HIV) concatenated proteins followed because the coding sequence for proteins was revealed, and its interruption by noncoding sequences could now be recognized. The interpretation of automated DNA sequences of whole genomes relies on the Rosetta Stone of the universal genetic code. The transformational utility for the study of all genomes, from viral to human, is established. New disease gene associations abound. Gene targeted diagnostics have become commonplace. New therapeutics derived from gene sequences, such as erythropoietin, have reached US Food and Drug Administration approval (12).

Dr. Nirenberg was a bold and risk-taking experimentalist (13, 14). He took seriously the task to “get it right” by his careful and critical research. Marshall Nirenberg, Ghobind Khorana, and Robert R. Holley shared the Nobel Prize Award in Physiology and Medicine in 1968 for this body of research. Dr. Nirenberg's Nobel Award was the first for the National Institutes of Health.

1 Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. *Nature* 171(4356):737–738.

2 Crick FH, Barnett L, Brenner S, Watts-Tobin RJ (1961) General nature of the genetic code for proteins. *Nature* 192:1227–1232.

3 Nirenberg MW, Matthaei JH (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci USA* 47:1588–1602.

4 Martin RG, Matthaei JH, Jones OW, Nirenberg MW (1962) Ribonucleotide composition of the genetic code. *Biochem Biophys Res Commun* 6:410–414.

5 Nirenberg M, Leder P (1964) RNA codewords and protein synthesis: The effect of trinucleotides upon the binding of sRNA to ribosomes. *Science* 145(3639):1399–1407.

6 Trupin JS, et al. (1965) Codewords and protein synthesis, VI: On the nucleotide sequences of degenerate codeword set for isoleucine, tyrosine, asparagine, and lysine. *Proc Natl Acad Sci USA* 53:807–811.

7 Crick FHC (1966) Codon-anticodon pairing: The wobble hypothesis. *J Mol Biol* 19(2):548–555.

8 Caskey CT, Tompkins R, Scolnick E, Caryk T, Nirenberg M (1968) Sequential translation of trinucleotide codons for the initiation and termination of protein synthesis. *Science* 162(3849):135–138.

9 Scolnick EM, Caskey CT (1969) Peptide chain termination. V. The role of release factors in mRNA terminator codon recognition. *Proc Natl Acad Sci USA* 64(4):1235–1241.

10 Caskey CT, Beaudet AL, Scolnick EM, Rosman M (1971) Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc Natl Acad Sci USA* 68(12):3163–3167.

11 Marshall RE, Caskey CT, Nirenberg M (1967) Fine structure of RNA codewords recognized by bacterial, amphibian, and mammalian transfer RNA. *Science* 155(3764):820–826.

12 Leary WE (1989) New anemia drug approved by U.S. for kidney disease *NY Times* June 2, Section D, p 4.

13 Caskey CT (2010) Obituary: Marshall Warren Nirenberg (1927–2010). *Nature* 464(7285):44.

14 Leder P (2010) Retrospective. Marshall Warren Nirenberg (1927–2010). *Science* 327(5968):972.