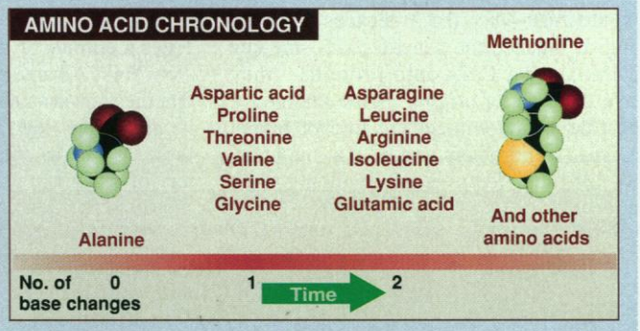
The First Codon and Its Descendants

NEW YORK—An expert in signal processing, Edward Trifonov excels in gleaning information from complex patterns. While others probe the overall origin of the genetic code, which specifies how the sequence of nucleotide base "letters" spell out the amino acids that make up proteins (see main text, p. 329), this biophysicist at the Weizmann Institute for Science in Rehovot, Israel, has applied his skills to determining which of the code's words came first.

"There are traces in [modern] sequence of the distant past," Trifonov said here last month at a New York Academy of Sciences meeting on Molecular Strategies in Biological Evolution. By looking for common features in the messenger RNA (mRNA) molecules that carry genetic messages from genes to the cell's protein factories, he concluded that the first word—a triplet of bases, or codon, that codes for a single amino acid—was GCU; that word stands for the bases guanine, cytosine, and uracil and codes for the amino acid alanine. He then went on to calculate how GCU might have evolved into the current set of 61 codons that specify the 20 amino acids. Trifonov has "come up with a unifying view of the origin of the genetic code," says Giorgio Bernardi, a molecular biologist at the Jacques Monod Institute in Paris, who considers Trifonov's results quite plausible.

Trifonov looked at mRNA for clues to the early code, because many researchers think that RNA predated the DNA of modern genomes. He noticed a hidden pattern of GCU repeats in mRNA sequences in many organisms: The GCU repeats are spaced in such a way as to align with matching bases in the ribosome, the structure that translates mRNA into protein. Matching bases attract, so the GCU repeats seem to help bind the mRNA to the ribosome. Because GCU is so common and plays so basic a role in translation, "it could also represent ancient RNA code," Trifonov thought.



Word gains. Genomic "words" for early amino acids evolved from single base changes; additional base changes led to more complex amino acids.

Then he and his colleague Thomas Bettecken, now at Magdeburg University in Germany, realized that RNA made up of this triplet might have gained an edge over other potential codons early in evolution, by interacting more easily with other nearby molecules. The DNA equivalent of the codon is GCT (T for thymine), a triplet that can cause a glitch in the cell's DNA-copying machinery and result in excess copies in daughter cells—a property that can disrupt gene function in diseases, including myotonic dystrophy and Huntington's disease. Assuming that GCU had the same property when RNA was the genetic material, this triplet would be more likely than others to produce longer RNA molecules that could fold in multiple ways to recognize and interact with amino acids and other molecules. "This exceptional property of expandability would give an advantage to that [triplet]," says Trifonov.

Once they had identified the potential first codon, the team made a series of one-nucleotide changes in the triplet (GCU to UCU, for example, or to GAU) to come up with six more codons. Because these could have evolved from GCU in a single step, they may have specified the next generation of amino acids. "The earliest changes were one-letter changes in GCU," Trifonov explains. "[Two-]letter changes were all later."

Then last year Trifonov and Bettecken compared their results to two other lists of potentially ancient amino acids, from origin-of-life experiments that came up with amino acids from the primordial soup, and chemical studies that identified amino acids with relatively simple structures. "If you put [these results] together, they overlap," Trifonov reported.

Trifonov himself emphasizes that this history is just a theory, and other researchers say no one may ever know for sure if he's right. Still, "these are nice arguments," says James Shapiro, a bacterial geneticist at the University of Chicago. "They all fit together and are very satisfying."

–ELIZABETH PENNISI

chosen based on their affinity for an amino acid are allowing them to test the idea. Several years ago, Michael Yarus of the University of Colorado, Boulder, noticed that in his experiments, the RNA strands that were best at binding a given amino acid tended to contain codons for that amino acid. But because the three-base codons often show up at random, the data were inconclusive.

Now evolutionary biologist Laura Landweber and graduate student Rob Knight of Princeton University have done a more careful analysis, looking specifically at where the amino acid arginine binds to random RNA strands generated in several researchers' experiments. If there is no real affinity, they reasoned, codons for arginine will appear as often in the regions where the amino acid does not bind as in regions that arginine homes in on. They found, instead, that while arginine codons made up 30% of the nonbinding RNA sites—the expected percentage, given that arginine has many possible codons—they made up 72% of the sequences in the binding regions. That suggests, says Landweber, that it's no accident that these codons specify arginine.

SOURCE: ADAPTED FROM E. TRIFONOV

The arginine evidence is intriguing, says evolutionary biologist Leslie Orgel of the Salk Institute in La Jolla, California. "But it's premature to draw any very strong conclusions" from data on the affinities of a single amino acid, he says. Researchers are delighted, however, that experimenters are now tackling the question. "Previously we had to rely solely on theory," says Lehman, but "if [Landweber's analysis] holds up, it will provide a convincing body of evidence" that basic chemical forces helped to shape the code.

Once the code was born, a different kind of pressure, the need to minimize errors, might have refined it. While some researchers have argued that any changes to the code over its 3.5-billion-year history would have been like switching the keys on a typewriter, leading to hopelessly garbled proteins, others argued that the existing code is so good at its job that it must have been shaped by natural selection. For example, in 1991, evolutionary biologists Laurence Hurst of the University of Bath in England and David Haig of Harvard University showed that of all the possible codes made from the four bases and the 20 amino acids, the natural code is among the best at minimizing the effect of mutations. They found that single-base changes in a codon are likely to substitute a chemically similar amino acid and therefore make only minimal changes to the final protein.

Now Hurst's graduate student Stephen Freeland at Cambridge University in England has taken the analysis a step farther by taking into account the kinds of mistakes that are most likely to occur. First, the bases