Four and a half billion years ago, our planet was a jagged, rocky landscape rich in minerals, swirling with carbon dioxide and other gases, and bombarded with meteorites. Somehow, around this time, the chemistry at work in this harsh environment yielded molecules that started to self-replicate. Thus began life on Earth.

Most researchers studying early life believe that the first self-replicating molecule to take hold was RNA, an idea known as the RNA-world hypothesis. But eventually came proteins—the actors that carry out almost all cellular functions—and DNA, the blueprint that encodes them. It’s the proteins—and their building blocks, amino acids—that have always intrigued biochemist A. Keith Dunker, an emeritus professor at Indiana University. He knew that researchers had long speculated that the early genetic code was much simpler than it is today, perhaps just 2 or 3 DNA nucleotides coding for 12 or 14 amino acids rather than 4 nucleotides coding for 20 amino acids. Since the discovery of the structure of DNA in 1953, many researchers have speculated about how that code evolved. “But no one was discussing why these last amino acids were added,” Dunker says. “What was the selective advantage?”

Sometime in the late 1990s he read a paper about the amino acid tryptophan. The authors argued that tryptophan was the last of the 20 amino acids to evolve. Tryptophan is the largest amino acid, and with its indole ring, it has a relatively complex chemical structure, so its late appearance is perhaps unsurprising. But the assertion caught Dunker’s eye.

Dunker, then at Washington State University, had recently begun studying intrinsically disordered proteins, or IDPs. Rather than hold a distinct conformation, IDPs wiggle and flap like strands of slightly sticky cooked spaghetti, generally forming numerous but weak bonds with other proteins or nucleic acids. Work by Dunker and others was beginning to contradict the dogma that a protein’s structure determines its function.

To study disorder in proteins, Dunker and his colleagues had begun digging through protein databases to compare how frequently each of the 20 naturally occurring amino acids appears in structured proteins and in disordered ones. They turned these prevalence rates into a scale ranking...
A plausible idea also exists for how disordered proteins may have contributed to early life, says Vladimir N. Uversky, a biochemist at the University of South Florida. According to the RNA-world hypothesis, life emerged from primitive but self-replicating RNA molecules that could catalyze chemical reactions. RNA on its own does not stay stably folded; it needs a scaffold. Some researchers posit that disordered proteins, which can be loose and floppy on their own but firm while in contact with molecules they bind with, may have served the purpose of interacting with RNA to stabilize it.

That idea aligns with what’s seen in modern cells—for example, with ribosomes, which are made of RNA interacting with proteins. “You can’t make a ribosome without positively charged polypeptides,” Dunker says. Peptides penetrate deep into the center of the ribosome, binding to the phosphate groups on RNA to hold the whole structure together. But outside the context of the ribosome, these peptides are disordered. Dunker suggests that this stabilization role is so crucial that RNA may have co-evolved with disordered proteins. “I don’t think there was ever an RNA world,” Dunker says. “It was an RNA-IDP world.”

**Primordial Proteins**

Today, researchers know that 30–50% of all protein sequences found in mammals are disordered, meaning that they don’t fold into stable 3-D structures. But that doesn’t mean those sequences have no function; studies suggest that IDPs and structured proteins perform complementary jobs in cells. “Disorder can do things that structure can’t and vice versa,” Dunker says. Unlike structured proteins, IDPs tend to be multitaskers. They act like molecular Velcro, binding and releasing multiple targets rather than engaging tightly with single ones. This positions them to act as cellular signaling hubs that orchestrate complex cellular events such as gene transcription and signal transduction. They also play a crucial role in forming transient, membraneless structures called biomolecular condensates, which may act as reaction chambers in the cell.

Dunker isn’t the only one who thinks disordered proteins may have played a unique and crucial role in the origin of life. When the building blocks of biochemistry were forming on Earth, “disorder was a given—there was no other option,” NASA biophysicist and astrobiologist Andrew Pohorille says.

The very nature of proteins predicts this, he says. Most structured proteins that are not membrane proteins have a hydrophobic core and a hydrophilic exterior. To achieve that, “you need length,” Pohorille says. “You can’t just do it with a few amino acids.” The smallest member of today’s known protein repertoire that has a hydrophobic core runs about 34 amino acids long, but it is unlikely that peptide sequences of that length could have existed from the get-go. “There must have been shorter proteins that were not well ordered, yet still performed functions,” he says.

What’s more, Pohorille and others speculate that in this very early period, the machinery for translating DNA into proteins was likely to be basic and error prone. The high mutation rate in amino acid sequences was perhaps advantageous because it served as an engine for generating proteins with novel functions. Because changes to an amino acid sequence often affect how a protein folds, disordered proteins “tend to be more robust in the face of mutations,” again supporting the possibility of their early presence, Pohorille says.
random loop sequences and tried to see if a bit of test tube evolution could coax it to take on a specific function: joining RNA strands. To simulate evolutionary pressure in the lab, he synthesized trillions of these randomized protein variants and screened them for the ability to catalyze the ligation of two RNA molecules. Any proteins that showed such activity became the next generation.

After only a few generations, the proteins evolved into several enzymes that accelerated the reaction more than 2 million-fold, says Seelig, who is now an associate professor of biological sciences at the University of Minnesota Twin Cities. To his surprise, the enzyme was much more dynamic than natural proteins and included a large disordered loop. The experiment doesn’t prove anything about how the very earliest proteins evolved, he says—that would require a time machine—but it does propose a possible scenario of how evolution could make something from very little. And the fact that this something turned out to be so flexible supports a role for disordered proteins on early Earth.

More recently, synthetic biologist Klára Hlouchová of Charles University set out to see what kinds of polypeptides emerge when completely random sequences of amino acids are strung together. Her team first generated thousands of random sequences, each 100 amino acids long, and then expressed several of them in bacteria. The idea was to simulate how early amino acid sequences may have been generated. “We think that early on, when the translation machinery was still not fixed, there was probably a lot of randomness in how peptides and proteins originated,” she says.

Hlouchová says that many of the sequences expressed quite well, meaning they didn’t turn into insoluble clumps that would be toxic to cells or be chopped up by proteases, as researchers have traditionally assumed. And proteins that resembled IDPs because they had fewer local interactions between amino acid residues expressed better. Her team is now conducting a similar study that more directly mimics early life by narrowing the pool of amino acids to 10 widely thought to be abundant in the prebiotic environment and then characterizing these proteins’ properties.

Knowing the minimal requirements for protein function isn’t just about re-creating the past, Hlouchová says. It also adds to the tool kit of protein engineering and synthetic biology. Scientists in protein engineering have long thought that they needed to design new proteins from fixed structural scaffolds because enzymes need those folds to start with, she says. “We now see that this is probably not the prerequisite.”

### An Aromatic Evolution

All this evidence for disorder in early life on Earth, though circumstantial, circles back to Dunker’s late 1990s tryptophan inspiration that the evolution of amino acids toward greater complexity ran parallel to that of proteins evolving toward greater structure. He believes that one key feature that characterized the earliest proteins—and supports the idea that they were disordered—is that their amino acids lacked aromatic rings. Researchers can’t date exactly when aromatic amino acids evolved, but most believe that all 20 amino acids had evolved by about 3.5 billion years ago, when the hypothetical last universal common ancestor, or LUCA, existed. Computational and theoretical studies, however, hint that two of the three aromatic amino acids, tryptophan and tyrosine, may have evolved later, Hlouchová says.

![L-Tryptophan](https://dx.doi.org/10.1021/acscentsci.0c01499)

Tryptophan, relatively large and complex, was one of the last amino acids to evolve and is the most protein-structure-promoting amino acid.

Regardless, the notion that the evolution of aromatic amino acids tracks with greater structure is supported by analyses of databases of modern proteins to explore aromatics’ possible role. “So far, we haven’t been able to find a single enzyme that doesn’t have an aromatic core to stabilize the active site in its specific structure,” Dunker says. “The aromatics are what make it really tight and strong rather than gooey.”

Last year, Uversky and his colleagues published an analysis of 817 proteomes from all domains of life catalogued in the Protein Data Bank. Proteins lacking aromatic residues and cysteine, also a structure-promoting amino acid, are rare, but the ones that exist tend to be disordered, taking shape by interacting with nucleic acids rather than forming their own firm, folded structure. “If you don’t have aromatic residues, you probably don’t have much of an option to be stable,” Uversky says.

Of course, it’s impossible to know what really happened at the dawn of biological life. But if there was a time when aromatic amino acids didn’t yet exist, what did proteins look like? Dunker and Hlouchová teamed up to try to simulate early Earth in a test tube by pulling all the aromatic amino acids—phenylalanine, tyrosine, and tryptophan—out of a protein called dephospho-coenzyme A kinase (DPCK).
and replacing them with leucine to see if it could still do its job without them. DPCK is a key enzyme in synthesizing coenzyme A, a highly abundant compound that is involved in fatty acid synthesis and cellular metabolism, processes highly conserved in evolution. Aromatics make up about 10% of DPCK’s composition.

The researchers found that the modified enzyme maintains its secondary structure—that is, the local interactions between residues—but loses its overall conformation. In that state, it still performs the first step in its repertoire (hydrolyzing ATP) but becomes less efficient at the second (sticking a phosphate onto coenzyme A). “This suggests that you can have functional proteins from even the early amino acids—the nonaromatics—even though some of their activities would be impaired,” Hlouchová says.

The fact that enzymatic activity can exist in the absence of structure aligns with the idea that early life could have chugged along for a while without structured proteins, Dunker says. And the corollary is that the later-evolving amino acids may have been key to developing structure. That possibility may answer the question bugging him since he read those early papers: Why had those last amino acids evolved? “It’s not something I set out to prove,” he says. “The answer just popped into my head: the formation of structure.”

Alla Katsnelson is a freelance contributor to Chemical & Engineering News, the weekly newsmagazine of the American Chemical Society.