Vol. 20, e200026 (2023)
doi: 10.2142/biophysico.bppb-v20.0026

Biophysics and Physicobiology
https://www.jstage.jst.go.jp/browse/biophysico/

*Review Article (Invited)*

# Why we are made of proteins and nucleic acids: Structural biology views on extraterrestrial life

Shunsuke Tagami

*RIKEN Center for Biosystems Dynamics Research, Yokohama, Kanagawa 230-0045, Japan*

Received March 31, 2023; Accepted May 29, 2023;
Released online in J-STAGE as advance publication June 2, 2023
Edited by Haruki Nakamura

**Is it a miracle that life exists on the Earth, or is it a common phenomenon in the universe? If extraterrestrial organisms exist, what are they like? To answer these questions, we must understand what kinds of molecules could evolve into life, or in other words, what properties are generally required to perform biological functions and store genetic information. This review summarizes recent findings on simple ancestral proteins, outlines the basic knowledge in textbooks, and discusses the generally required properties for biological molecules from structural biology viewpoints (e.g., restriction of shapes, and types of intra- and intermolecular interactions), leading to the conclusion that proteins and nucleic acids are at least one of the simplest (and perhaps very common) forms of catalytic and genetic biopolymers in the universe. This review article is an extended version of the Japanese article, On the Origin of Life: Coevolution between RNA and Peptide, published in SEIBUTSU BUTSURI Vol. 61, p. 232-235 (2021).**

**Key words:** origin of life, astrobiology

***Significance***

One of the most important goals of the current space exploration missions is to find another life form (or its vestiges). What molecules should we look for to achieve the purpose? Can we expect extraterrestrial lives also have proteins and nucleic acids? Should we take completely different molecules into consideration? In this review, the essentially required properties for biomolecules (on the Earth and other planets) are discussed from structural biology viewpoints.

## Introduction

If a complex organism (as complex as the simplest bacterium we know on the Earth) were to emerge on another planet, then molecules to store genetic information and perform various chemical and physical functions must also have evolved. On the Earth, nucleic acids serve as information storage, and proteins play the main catalytic and mechanical roles. As no other examples of life have been discovered yet at the time of writing, we do not know if the terrestrial system (the combination of nucleic acids and proteins) is the only style in the universe or if there are totally different biomolecular systems. Still, we might be able to estimate the generally required properties for biological molecules in the universe, by carefully inspecting our own system. Constraints in the structures of biomolecules would be exceptionally pivotal, as their functions are fundamentally realized by their shapes and surfaces. In this review, the previous experiments to reconstruct ancient simple proteins are summarized, and then the properties of ideal biomolecules in the universe are discussed from structure biology viewpoints.

Corresponding author: Shunsuke Tagami, RIKEN Center for Biosystems Dynamics Research, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ORCID iD: https://orcid.org/0000-0002-1720-3627, e-mail: shunsuke.tagami@riken.jp

**Reconstruction of Primitive Proteins**

Numerous catalytic molecules are required for cellular survival. If the emergences of such molecules are miraculously rare, then we would be alone in this universe. However, if they are rather common events, then our universe might be full of life. Modern proteins on the Earth have overwhelmingly complicated sequences and structures, suggesting the difficulty of emerging by chance from non-living chemistry. This section outlines the experiments to demonstrate that folded proteins could have emerged in a plausible step-by-step evolutionary process from simple prebiotic peptides.

**Secondary Structure Assemblies**

Peptides with secondary structures might have been intermediates between randomly synthesized prebiotic peptides and folded proteins [1-3]. For example, in 1975, Brack and Orgel reported that a simple peptide with alternating valine and lysine residues can self-assemble into large β-sheet structures [4]. The β-sheet structures were likely formed by simply orienting the lysine residues on one side to form a hydrophilic surface, and the valine residues on the other side to form a hydrophobic surface, which was likely further utilized to bind the hydrophobic sides of other β-sheets (Figure 1A). α-Helical assemblies were also formed by simple peptides with periodic hydrophobic patterns, which define the interaction surfaces with other peptide molecules (Figure 1B) [5]. Thus, primitive secondary structures and their assemblies can be formed by very simple sequences [6]. Furthermore, such peptides might have propagated by self-ligation reactions in peptide assemblies and become enriched on the ancient Earth [7-11].

Peptide assemblies with secondary structures might also have played essential roles in co-evolution with RNA [1-3,12]. Our group has recently demonstrated that a peptide with hydrophobic and cationic moieties (P43: AKKVWIIMGGS) can form insoluble assemblies containing β-amyloid structures that accumulate RNA on their surfaces [13]. Indeed, longer RNAs bound more stably on the P43 aggregates even in high salt conditions. Furthermore, such peptides might also have supported RNA synthesis. RNA polymerase ribozyme (RPR) is an artificial RNA enzyme with an RNA-dependent RNA polymerase activity and regarded as a model molecule for the self-replicating RNA on the primordial Earth [14,15]. The aggregates of P43 and its sequence-simplified versions (e.g., $K_2V_6$: KKVVVVVV) enhanced the activity of RPR on their surfaces, although their effects were largely dependent on the buffer conditions [13]. Such β-amyloid peptides might have worked as selection platforms to concentrate longer RNAs from environments and support the functions of primordial ribozymes (Figure 1C).

**Ancient Proteins with Smaller Sequence Spaces**

There still seems to be a huge gap between simple peptides with secondary structures and modern proteins with tertiary structures. As modern proteins are polymers composed of 20 different amino acids, even a relatively small protein with 100 amino acid residues would have $20^{100}$ (~$10^{130}$) possible sequences. It is almost unthinkable that functional sequences were efficiently searched from such a huge sequence space on the prebiotic Earth. In the early evolutionary stages of life, proteins might have emerged from a much simpler sequence library. For example, if an ancient protein could have emerged as a short peptide with 20 residues and five different amino acid types, the required sequence space would be only $5^{20}$ (~$10^{14}$), and its total mass could be less than 1 μg. To demonstrate that proteins with defined tertiary structures and functions can emerge from such smaller sequence spaces, the reconstructions of ancestral proteins with 1) shorter peptides and 2) fewer amino acid types have been performed, as described below.

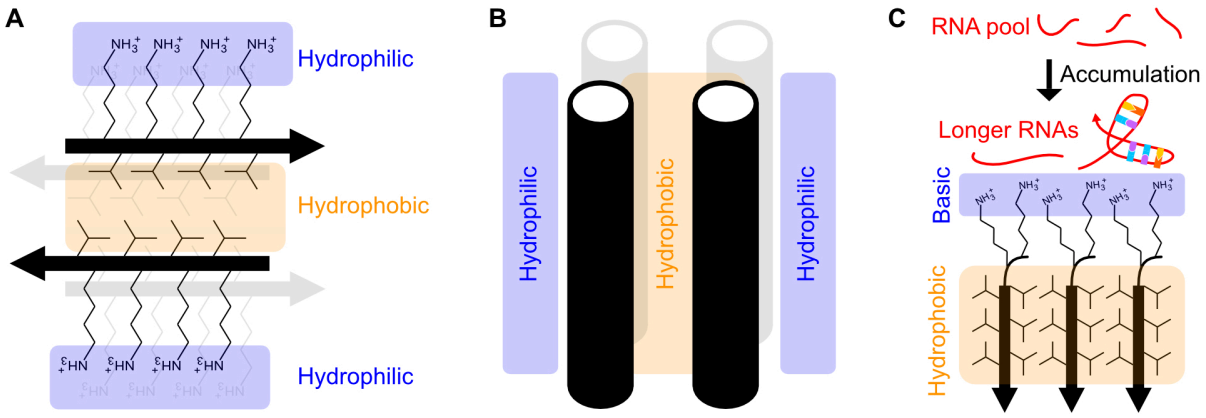



**Figure 1** Schematic structures of simple peptide assemblies with secondary structures. Hydrophilic and hydrophobic parts are colored blue and orange, respectively.

The structures of some primitive proteins were probably created by the self-assembly of shorter peptides (10–50 amino acids) [16-19]. For example, the protein folds shown in the top panels of Figure 2 have internal pseudo-symmetries [20-25]. They are assumed to be descendants of ancient homo-oligomeric peptides, which evolved into monomeric proteins by gene fusion and then gradually lost the perfect symmetry by mutations in each repeating unit. Applying various protein engineering techniques, these pseudo-symmetric folds have been reconstructed as short homo-oligomeric peptides, or polypeptides with identical repeats (Figure 2, bottom) [26-37].

Attempts to create proteins using fewer amino acid types have also been reported, resulting in the exclusion of 7 to 13 amino acid types for several proteins [36-45]. For example, an ancestral nucleoside diphosphate kinase (NDK), designed by using only 13 amino acids, maintained high thermal stability and enzymatic activity [44,45]. Even a further engineered NDK variant with only ten amino acid types could fold into the proper tertiary structure, although the activity was lost [44,45].

In a few cases, reconstructions by short peptides and smaller amino acid repertories were simultaneously achieved. For example, the β-trefoil fold was reconstructed as a polypeptide with three 42-amino-acid repeats containing only 12 amino acid types (Figure 2G) [36]. Another example is the DPBB fold conserved in various proteins, including the catalytic domains of cellular RNA polymerases. It was reconstructed by the homodimerization of a 43-amino-acid peptide containing only seven amino acid types (GAVDEK and R) (Figure 2H) [37]. Interestingly these seven amino acid types can be coded by GNN and ARR (R = A or G) in the modern genetic code. The five amino acids coded by GNN (GAVDE) are simple prebiotic amino acids. The other two (K and R) are positively charged and essential in interactions with nucleic acids. Thus, this might reflect an ancient amino acid repertory encoded by a primitive genetic code.

**Structure Formation with Simplified Hydrophobic Cores**

The formation of hydrophobic cores is essential for protein folding. Although simple β-amyloids or α-helical assemblies can be formed relatively easily [1-6], the formation of tertiary structures seems to be much more complicated. For protein stability, it is generally considered important to arrange the hydrophobic side chains without gaps in the hydrophobic cores. However, precisely packing multiple amino acid residues within the cores in various shapes would be exceptionally difficult for prebiotic chemistry or a primitive translation system with a lot of errors. To solve this problem, efforts to reconstruct folded proteins with simplified hydrophobic cores have been performed.

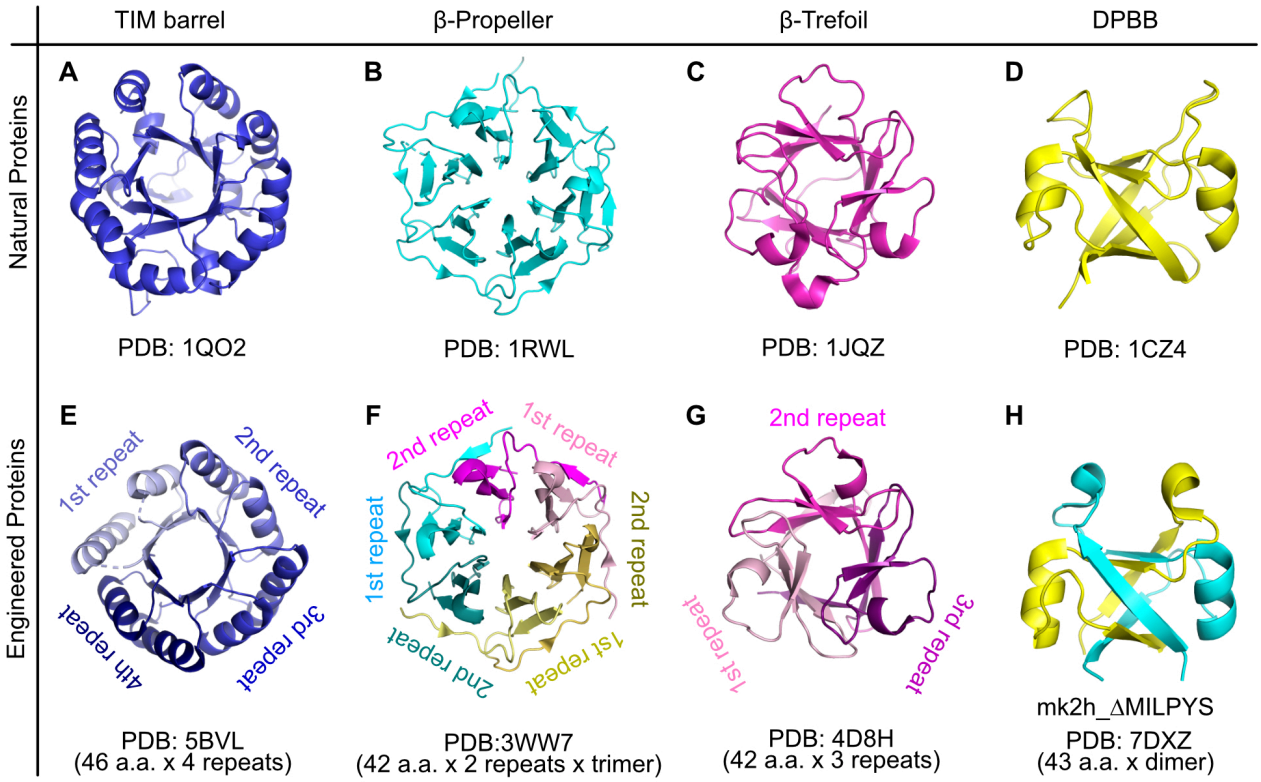



**Figure 2** Structures of natural pseudo-symmetric proteins (A–D) and engineered symmetric proteins (E–H). Different monomers in oligomeric designs are colored with different colors. Multiple repeats in the monomer units are colored with different brightness.

In the early days of protein engineering experiments, the hydrophobic cores of some proteins could reportedly be simplified by enriching one or two amino acid types [38]. A helix bundle protein (Rop) was reconstructed with a simplified hydrophobic core containing only alanine and leucine residues placed in a zigzag pattern (Figure 3A) [46,47]. Hydrophobic cores with more complicated folds (e.g., SrcSH3, T4 lysozyme, phage 434 Cro repressor) could also be extensively replaced by single amino acid types with relatively large structures (leucine, isoleucine, or methionine) (Figure 3B–D) [39,48-50]. These hydrophobic side chains have many possible rotamers (side chain conformations) and would fit within various shapes to fill the hydrophobic cores without gaps.

However, the enrichment of such flexible side chains, especially the long linear one (methionine), would also be disadvantageous for protein stability, as they would cause large entropic losses when fixed inside hydrophobic cores [49]. This is probably the reason why long non-branched hydrophobic amino acids, like nor-valine and nor-leucine, were excluded from the proteinaceous amino acid repertory, despite their availabilities on the ancient Earth [54,55].

Recent examples of ancestral and de novo protein designs demonstrated that it is even unnecessary to fill hydrophobic cores with large side chains. The ancient DPBB fold mentioned above contains only the two shortest hydrophobic amino acid types (alanine and valine) (Figure 2H, Figure 3E) [37]. Koga's group reported a very thermostable de novo protein with a hydrophobic core mostly composed of valine residues (Tm = 106 ºC, Figure 3F) [56]. The structures of these ancestral and de novo proteins contain significant unfilled volumes within their hydrophobic cores. Interestingly, their overall structures were not compacted, as compared to the non-simplified variants with larger hydrophobic residues in the cores. These results indicated that the hydrophobic cores of some proteins can be formed without neat optimization, and even without large side chain fittings (by Leu, Ile, or Met) or zigzag patterning (Ala/Leu). In such cases, the formation and stability of the tertiary structures are likely to be more dependent on the designs of stable main chain structures (e.g., patterns of secondary structures and loops) [56-59].

Ancient protein evolution might have started with main chain topologies or folds that were stable even without neatly packed cores. Hydrophobic cores enriched with short side chains could have been more realistic, as simpler amino acids were probably plentiful on the ancient Earth. Furthermore, the presence of unfilled spaces in such cores would have been advantageous, as they could easily tolerate almost unavoidable contaminations with different amino acids by prebiotic chemistry or primitive translation systems. Such adaptability for random or bulky mutations is even seen in modern proteins [60,61].

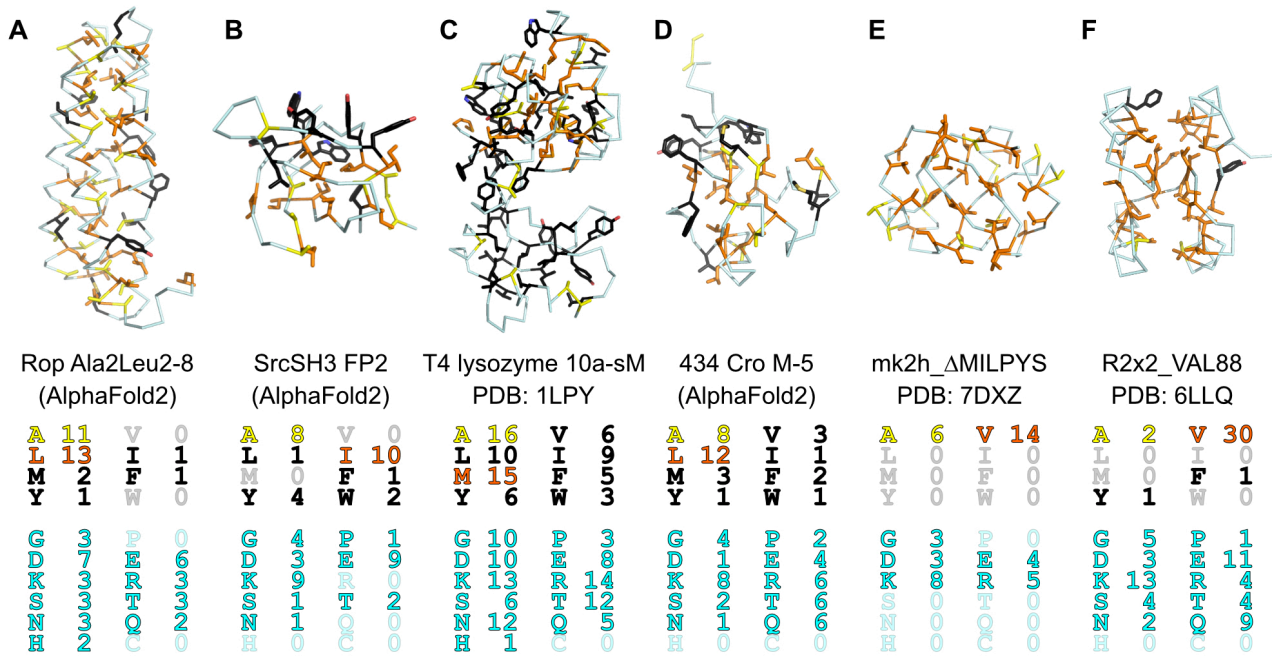



**Figure 3** Engineered proteins with simplified hydrophobic cores. Experimental or predicted structures of engineered proteins with simplified hydrophobic cores are shown with their amino acid compositions. Protein sequences are adopted from each reference or PDB. Structural models of Rop Ala2Leu2-8, SrcSH3 FP2, and 434 Cro M-5 were predicted by Alphafold2 [51-53]. Structures of Rop Ala2Leu2-8 and mk2h_ΔMILPYS are shown as homodimers. The C-terminal tag was removed from the model and amino acid composition of R2x2_VAL88. Hydrophobic side chains are shown as stick models. Alanine residues are colored yellow. Another enriched hydrophobic amino acid in each design is colored orange. Other hydrophobic residues and non-hydrophobic residues are colored black and light cyan, respectively.

**Evolution from Simple Peptides to Folded Proteins**

Simple peptides are considered to have existed on the prebiotic Earth [62]. For example, lysine-rich peptides and their analogs could have been synthesized by prebiotic chemistry [63-66], and mutually stabilized with nucleic acids [67]. As mentioned above, simple KV-rich peptides can form β-amyloid assemblies (Figure 1A) [4], capture RNA molecules, and support RNA synthesis by a model ribozyme (Figure 1C) [13]. These results suggest that such simple prebiotic peptides could have been sufficiently functional to support an RNA-based primitive life system on the ancient Earth.

Such peptides might have been conjugated and evolved into ancient (poly-) nucleotide-binding motifs with a few secondary structures [68-70]. P-loop and Helix-hairpin-Helix (HhH) are examples of such ancient motifs. P-loop and its flanking secondary structures have been suggested to be the evolutionary seed of a wide variety of nucleotide-binding enzymes (Figure 4A) [71-73]. Such peptides containing the β-(P-loop)-α element reportedly showed rudimentary helicase and adenylate-kinase activities [74,75]. HhH is a well-conserved nucleic-acid binding motif that adopts a pseudo-dimeric structure (Figure 4B), and a reconstructed ancient HhH motif can form a homodimer and assemble into macromolecular coacervates with RNA [76,77]. Furthermore, the DPBB fold conserved in the active site of modern RNA polymerase might have emerged by fusion between ancient RNA-binding peptides. The sequence of the ancestral DPBB design with seven amino acid types is highly KV-rich (eight lysine and fourteen valine residues in the 43 a.a. peptide) (Figure 2H, Figure 3E). Perhaps there were only a few steps between KV-rich RNA-binding β-amyloid peptides and β-barrel proteins [13,37].

Koga and colleagues successfully designed various protein folds by first linking secondary structure fragments with ideal loops for each topology, and then optimizing the side chains [56-59]. Similar processes might have happened in the early evolution of folded proteins. Protein folds with simple hydrophobic cores might have first emerged by fusion events between short peptides without sequence optimization, as wide range of sequences could be tolerated in reconstruction of simple protein folds [37,78]. As hydrophobic interactions do not have specificity or directionality, "*hydrophobicity is nearly a sufficient criterion for the construction of a functional core*", which "*would greatly reduce the initial hurdle on the evolutionary pathway to novel enzymes*" [60]. Thus, the emergence of complicated protein structures might have been facilitated by relatively simple step-by-step procedures.

In contrast, nearly half of the amino acids in the modern genetic code are hydrophobic (8 out of 20 a.a. types), indicating the importance of their diversity. In very early protein evolution, a primitive genetic code might have encoded prebiotic hydrophobic amino acids (Ala, Val, Leu, Ile) without precise discrimination, as primitive hydrophobic cores could have formed by incorporating them almost randomly [37,78]. Then, through survival competitions, these amino acids might have been assigned to different codons, increasing the structural accuracies (and thus stabilities and activities) of the encoded proteins. Finally, paying higher synthetic costs, the genetic code might have started to incorporate larger hydrophobic amino acids (Met, Phe, Tyr, Trp). Such side chains likely have stronger aggregation tendencies [79], but also function, for example, in protein folding in lipid bilayers [80].

**General Requirements for Biological Molecules in the Universe**

Now, the emergence of proteins on Earth seems not to have been so miraculous (even if it still had been a very rare event). What kinds of molecules would then be selected as counterparts of proteins and nucleic acids if a life system emerged and survived on another planet (or moon) somewhere in the universe? Would they be something (very) similar to ours, or totally different? In this section, I scrutinize the generally required properties of biological molecules responsible for chemical/mechanical functions and information storage, from a structural biology perspective.

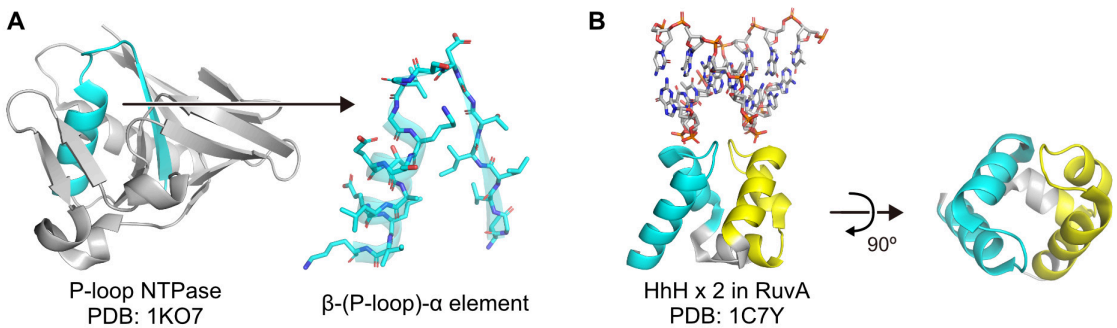



**Figure 4** Structures of ancient (poly-) nucleotide-binding motifs. (A) Structure of the P-loop NTPase domain from HPr kinase and the closeup view of the β-(P-loop)-α element. The β-(P-loop)-α element is colored cyan. (B) Structure of the pseudo-dimeric HhH motifs in a Holliday junction binding protein, RuvA. The two HhH motifs are colored cyan and yellow, respectively.

**Molecular Bonds and Interactions**

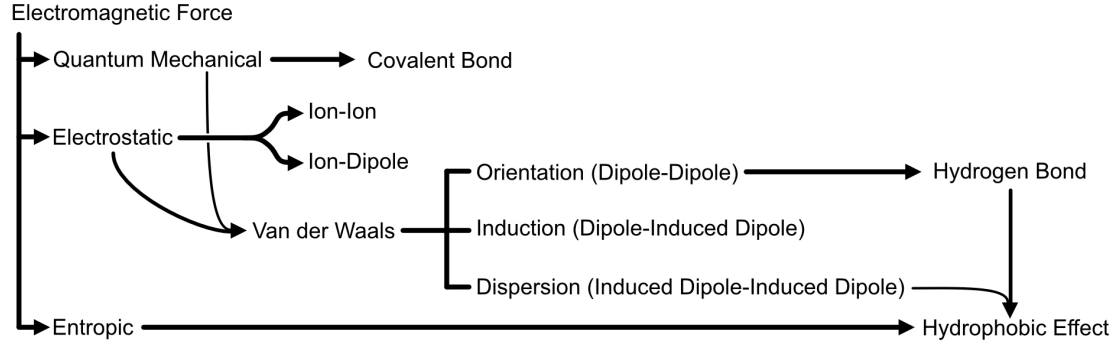
First, the properties of representative molecular bonds and interactions are summarized in Figure 5, as they are the basis of the discussion in the following subsections (If you are confident about your textbook knowledge on this topic, you may skip this subsection. If you would prefer more detailed explanations, the recommended textbook is listed in the references [81]).

All bonds and interactions inside or between biomolecules are fundamentally caused by electromagnetic forces. They are categorized into different groups depending on their apparent mechanisms (e.g., quantum mechanical, electrostatic, entropic) and properties (e.g., strength, specificity, directionality). Although such categorizations are somewhat ambiguous and occasionally cause confusion, they are still beneficial for a quick understanding of the nature of molecular bonds and interactions.

Covalent bonds have quantum mechanical properties, are formed by chemical reactions, and result in the synthesis of new molecules connected by shared electron orbits. Contrarily, other interactions result in physical contact between molecules without electron sharing. Electrostatic interactions involving charged atoms are further categorized into ion-ion and ion-dipole interactions. Van der Waals interactions are partially electrostatic and partially quantum mechanical, and contain three components: orientation (dipole-dipole), induction (dipole-induced dipole), and dispersion (induced dipole-induced dipole). While the interactions between ions and permanent dipoles can be attractive or repulsive depending on the involved molecules, the dispersion force is ubiquitous and always attractive. Although the dispersion force is usually weak, it can become stronger when electrons are only weakly caught in their orbits and easily disturbed by surrounding electrostatic fields (e.g., in large atoms and π bonds).

The hydrogen bond is roughly a special case of dipole-dipole interactions where an almost naked proton bridges two electronegative atoms (donor and acceptor) in a straight line. Because of the very small radius of the proton and large dipoles of the donor and acceptor molecules, the attracting Coulomb forces between the atoms are much stronger than in usual dipole-dipole interactions. Water has a tetrahedral molecular structure and forms three-dimensional networks of hydrogen bonds. To accommodate non-polar molecules in water, enough cavity spaces need to be formed against such strong hydrogen bond networks. Reorientations of water molecules surrounding non-polar molecules are also induced when they are solubilized in water, resulting in large entropy losses. Thus, non-polar molecules aggregate together in water to minimize the total non-polar surface and free as many water molecules as possible (hydrophobic effect). Although the resulting structures of non-polar complexes seem to be assembled by the non-polar dispersion force, it cannot be the driving force of the hydrophobic effect because the dispersion forces between non-polar molecules and between non-polar and water molecules are not very different. Thus, the hydrophobic interaction is mainly entropic and thus a long-range force, which can drive molecular assembly and folding in polar environments.

A




B

| | Bond/Interaction | Pairing specificity | Directionality | Formation mechanism |
|---|---|---|---|---|
| i) | Covalent Bond | Low | High | Chemical Reaction |
| ii) | Ion-Ion | High | Low | Physical Contact |
| iii) | Hydrogen Bond | High | Medium–High | Physical Contact |
| iv) | Hydrophobic Effect | Low | Low | Physical Contact |

**Figure 5** Molecular bonds and interactions. (A) Types of molecular bonds and interactions. (B) Properties of representative bonds and interactions in biomolecules.

Figure 5B summarizes the properties of four representative bonds/interactions in biomolecules, especially focusing on pairing specificity and directionality (angle constraints). Covalent bonds can be formed between relatively free combinations of atoms. However, their spontaneous formations in biomolecules are observed only in rare cases (e.g., disulfide bonds between two cysteine residues), since they usually require specific catalytic reactions. Covalent bonds also have very strong directionality. Furthermore, double bonds have no rotational freedom and confer structural rigidity. Ion bonds require specific pairs (positive and negative) to be formed and have no directionality. Hydrogen bonds require specific pairs (donor and acceptor) to be formed and have relatively strong directionality. No other interactions can simultaneously have such pairing specificity and directionality, which confer the unique discriminating ability to hydrogen bonds. Contrarily, hydrophobic interactions do not require any pairing specificity or directionality. They just form between non-polar molecules in polar solvents. I will discuss which of these interactions would be ideal to realize the specific properties of ideal biological molecules, especially for various chemical/mechanical functions and information storage, in the following subsections.

**Molecules for Chemical and Mechanical Functions: Most Likely Proteins**

If we suppose that some complex organisms (as complex as the last universal common ancestor, LUCA, on the Earth) could emerge on a planet, then numerous functional molecules would be required (like a few hundred proteins coded by LUCA) (Figure 6 (i)) [82]. To make the synthesis of such a large variety of molecules feasible, they would be composed of polymers with more or less uniform main chains and variable side chains (Figure 6 (ii)). Such polymers can be synthesized by a single polymerization system, and still achieve the huge diversity of the products.

To materialize numerous chemical and mechanical functions, the polymers also need to adopt various structures, as such functions are fundamentally accomplished by the shapes and surfaces of molecules (Figure 6 (iii)). In particular, high structural diversity at very small scales (sub-angstrom to a few nanometers) is essential to perform various chemical reactions in many metabolic pathways. In addition, if the polymers had defined structures, then their chains must have folded around cores (Figure 6 (iv)). What interactions would then be used to form such cores?

If numerous different shapes and cores had formed, then their emergence must have been easy enough to happen frequently (Figure 6 (v)). From this point of view, interactions with strict rules like covalent/ion/hydrogen bonds are unsuitable. Covalent bonds would not form easily or fit in different shapes either. It would also be impossibly complicated

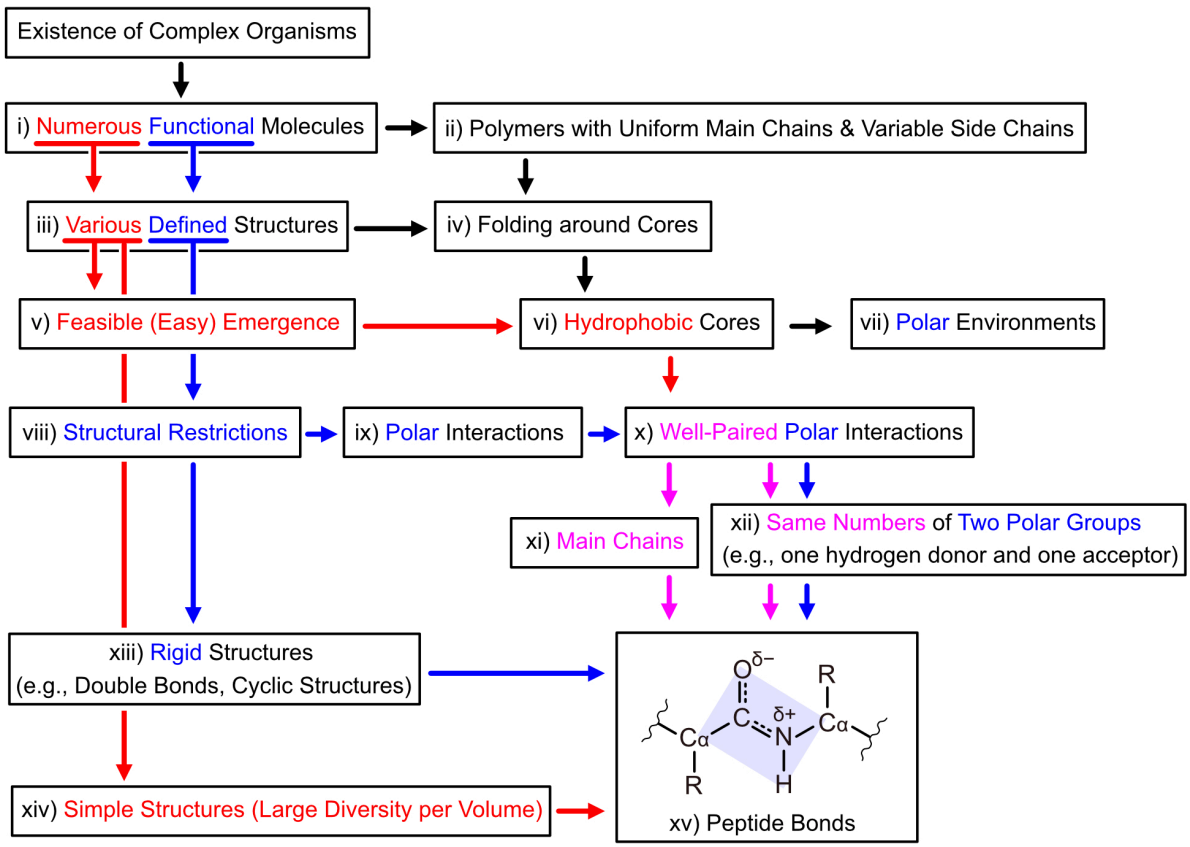



**Figure 6** Required conditions for ideal functional biopolymers.

to solve hundreds of different 3D puzzles in such small scales by placing multiple charged or polar side chains interacting in the right combinations and right angles. Accordingly, we need to choose a force that does not have pairing specificity or directionality. Although the dispersion force meets this requirement, it would not be strong enough to form a core in non-polar environments. Thus, the only choice that allows a large variety of structures without meticulous design would be hydrophobic interactions in polar environments (Figure 6 (vi, vii)). As we have seen for the terrestrial proteins, the hydrophobic cores can form without a lot of optimizations and produce a huge variety of protein folds (The difficulty of the polar core formation might also be indicated by the structures of transmembrane proteins. Even in the lipid bilayers where the hydrophobic effect is negligible, it is obvious that transmembrane proteins do not fold around polar cores, but still mainly use hydrophobic interfaces. Although the folding mechanism of transmembrane proteins is not as well understood as that of soluble proteins, it is likely supported by several mechanisms, such as interactions between cationic/aromatic residues with the phosphate head groups of lipid bilayers [80]).

However, if the polymers were completely hydrophobic, they would just form irregular aggregates in the polar environments. Thus, here are the apparently contradicting demands: freedom to allow variety and restriction to define structures (Figure 6 (viii)). Such structural restrictions should be introduced by interactions with stricter rules while keeping the easy formation mechanisms. Thus, polar interactions (ionic or hydrogen bonds) must also be introduced to fold the polymers into well-defined 3D structures (Figure 6 (ix)).

But how can polar elements become embedded in hydrophobic cores? They must be introduced in well-paired (canceled) forms (Figure 6 (x)). In the case of our functional polymers (i.e., proteins), the introduction of a single polar residue in the hydrophobic core usually destabilizes the protein structure. In contrast, we sometimes find a pair of charged residues (salt bridge) even in a hydrophobic core. However, such well-placed pairs of polar side chains would not frequently occur by random mutations in hydrophobic cores and could not be the primary mechanism to introduce structural regularity into biopolymers in the universe. Thus, the polar interactions in hydrophobic cores are much more likely to be achieved by the main chains with regular repeating units (Figure 6 (xi)).

To embed the main chains in hydrophobic structures, all polar parts in their repeating units should ideally be canceled by pairing with each other. In other words, the numbers of two pairing groups in the main chain should be the same (Figure 6 (xii)). For example, in one of the simplest cases, the main chain would have one hydrogen donor and one acceptor in its repeating unit, which would pair perfectly without leftovers.

An additional way to balance the freedom and regularity in the polymer structures might be the introduction of partial rigidity in the main chains (Figure 6 (xiii)). For example, double bonds or cyclic structures would restrict bond rotations and make the polymer tend to adopt regular conformations. If evolutionary competitions selected more suitable molecules, then the biopolymers of surviving organisms might also have main chains with such optimized rigidity.

Finally, even after fulfilling the above conditions, the monomer units of the biopolymers must be kept simple (Figure 6 (xiv)). They need to be synthesized by non-biological chemistry or by very primitive life systems. Also, not only diversity of sequence but also diversity per volume would be essential to realize various functions. If the monomer units were too large (e.g., 50 Å), then the polymer would never have optimum binding surfaces for multiple substrates.

Summing up the above arguments, functional biopolymers should fold into various shapes and have hydrophobic cores (Figure 6 (iii, vi)). They must have main chains with the same numbers of two groups to form polar interactions (Figure 6 (xii)), ideally with partially rigid structures (Figure 6 (xiii)). Their monomer unit should also be as small as possible (Figure 6 (xiv)). What molecules satisfy such requirements? Living on the Earth, we know one example, poly-α-amino acids (i.e., proteins) (Figure 6 (xv)).

The repeating unit of poly-α-amino acids has one donor and one acceptor of a hydrogen bond, which are used in extensive intramolecular interactions (self-complementary) [83]. Peptide bonds also have a double bond property because of their resonant structures. These bonds have an almost fixed plane, while still leaving partial rotational freedom around the α-carbon (φ and ψ angles) [84]. Thus, the requirements for ideal functional biopolymers are met by the six atoms in the peptide bond (Figure 6 (xv)). This also seems to be at least one of the simplest polymers to fulfill the above conditions. Considering that amino acids are quite common in the universe and peptide bonds can easily form even without life [62], we would expect some significant population (maybe the majority) of extraterrestrial lives to use proteins to perform chemical and mechanical functions. Therefore, the conclusion here is the same as the one from a seminal review by Weber and Miller: "*If life were to arise on another planet, we would expect that the catalysts would be poly-alpha-amino-acids...*" [54].

**Molecules for Information Storage and Transfer: XNA**

What molecules would store the genetic information of extraterrestrial lives? The simplest way we know to save and convey information is a sequence of letters (Figure 7 (i)). Molecules for information storage would most likely materialize this concept. For this purpose, again, polymers with uniform main chains (sequencing) and different side chains (letters) would be the easiest choice (Figure 7 (ii)). What kind of interactions would they use? What kind of shapes would they adopt?

Information precision and high distinctiveness would be the most important properties required for letters in genetic polymers (Figure 7 (iii)). Such letters would somehow be materialized by bonds or interactions with high specificities. Considering that genetic information should not only be stored but also transmitted, the letters are likely read and written through interactions that can be formed, dissociated, and reused easily. In the bonds and interactions listed in Figure 5B, the hydrogen bond is apparently the best candidate to meet these requirements, as it has pairing specificity, directionality, and can be formed by physical contact (Figure 7 (iv)).

In contrast to functional biopolymers with hydrophobic cores and various structures, the genetic polymers with hydrogen-bonded letters would have much more regular structures. As I argued above, placing multiple polar side chains in various 3D puzzles would be a highly complicated problem, since we cannot place them randomly, unlike the cases of non-polar side chains for hydrophobic interactions. Thus, only finite interaction patterns between limited components would arise. In other words, a few different "pairs" of side chains would work as letters in the genetic polymers (Figure 7 (v)).

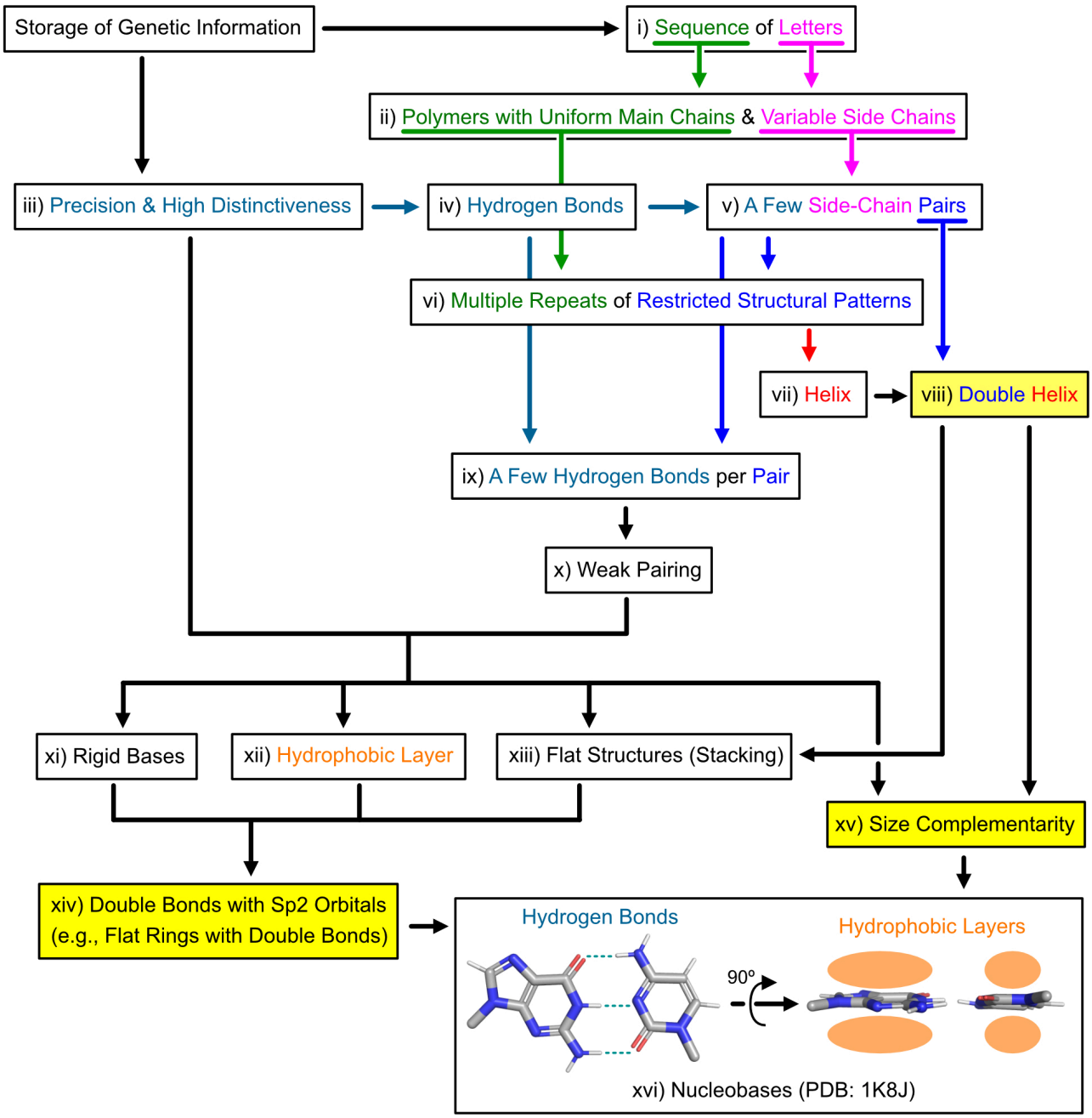



**Figure 7** Required conditions for ideal genetic biopolymers.

It seems reasonable to assume that the structures of these few side chain pairs would be somewhat similar to each other, except for the discriminators of letters (hydrogen bonds), because they need to meet the same functional and structural requirements (also see below for a more detailed discussion about the structures of "base" pairs). Such more or less uniform side chain pairs must be connected by uniform main chains and repeated multiple times to encode genetic information (Figure 7 (vi)). If we connect and repeat a uniform structure in a uniform way (a linear, head-to-tail polymer), it would be a helix (Figure 7 (vii)) [85]. As a genetic polymer contains paired side chains, it would be a double helix (Figure 7 (viii)).

What structures, then, should the paired side chains have? To discriminate different letters, they should have at least a few hydrogen bonds. Larger numbers of the pairing bonds might increase the stability and specificity of the pairs, thus contributing to the precision of information storage and transfer. However, it would be more difficult to synthesize larger side chains, and finding the complementary pairs would also be challenging. For example, the probability for randomly ordered donors and acceptors from two side chains to form ten perfectly-coupled hydrogen bonds would be 0.1% ($1/2^{10}$). Realistically, the pairing side chains in genetic materials would only have a few hydrogen bonds per pair (Figure 7 (ix)).

However, this estimation raises another problem. In the discussion about functional biopolymers above, I estimated that such biopolymers (and life) would emerge in polar environments to allow various hydrophobic cores to form (Figure 6 (vii)). In such polar environments, polar interactions become weak (Figure 7 (x)). A few hydrogen bonds are not sufficient to stably couple side chains, as such polar regions can also form hydrogen bonds with surrounding water molecules. To overcome this problem and to guarantee stable pairing between letters, a few features must be present in the side chains: 1) rigidity, 2) hydrophobicity, and 3) flat conformation.

First, the discriminators of the letters (donors and acceptors of hydrogen bonds) should be fixed on rigid basement structures, or "bases" (Figure 7 (xi)). If the pairing side chains have flexible structures and the hydrogen bond donors and acceptors also fluctuate, then the chances of having the proper pairing would drastically decrease. The hydrogen bonding between fluctuating structures would also cause large entropic losses by fixing them in the paired conformations. If the pairing side chains have rigid bases to fix the hydrogen bond donors and acceptors, then such problems can be avoided.

Second, the polar discriminators of the genetic letters should ideally be sequestered in a hydrophobic layer to enhance the recognition abilities of hydrogen bonds (Figure 7 (xii)). Such hydrophobic parts would also be present in the side chain bases. As the surrounding environments are supposed to be polar, the genetic polymer should also have an outer hydrophilic layer. Thus, the round slice of the genetic polymer would have the hydrogen-bonded pairings at the center of the double helix, the side chain bases as the middle hydrophobic layer, and the outer hydrophilic layer likely formed by the main chain.

This ideal architecture of genetic polymers leads to an additional requirement for the structures of the side chain bases. As they are repeated and trapped in the inner space of the double helix, they must be piled up neatly. Thus, flat structures would be favored (Figure 7 (xiii)). If the bases of the side chains had bumpy structures (like the chair and boat conformations of cyclohexane), then there would be unfavorable unfilled spaces when stacked in a helical structure. In contrast, the neatly piled-up structure of the flat hydrophobic bases would contribute to the stability of the overall double-helical structure of the genetic polymer by interactions between stacking bases. These stacking interactions could be further enhanced if they contain strongly induced dipoles, for example, in π bonds.

Summing up the above arguments, the optimal side chains in genetic polymers would have rigid and flat base structures with hydrophobic moieties (ideally containing strongly induced dipoles). Fortunately, these conditions can be solved by a simple solution: double bonds with sp2 orbitals (Figure 7 (xiv)). Double bonds confer rigidity to the molecular structures by restricting the rotation around them. Sp2 contains three hybrid orbitals in a plane. The hydrogen bond donors and acceptors would also stick out in the same plane, supporting their interactions in the proper straight-line conformation. Furthermore, the π electrons in double bonds would have strongly induced dipoles and enhanced hydrophobic stacking between bases. If the ring structures are formed by multiple double bonds and sp2 orbitals, then they would be especially suitable, since they are very rigid, flat, and rich in π electrons.

Additionally, to support weak recognitions by a few hydrogen bonds, it would be more desirable if the pairing bases have size complementarity (i.e., the small one forms a pair with the large one), which would be a strong discriminator when trapped in the limited space inside the double helix (Figure 7 (xv)). Considering the parsimonious aspect of evolution, this size complementarity would be achieved by the minimum required structures: a one-ring base forms a pair with a two-ring base.

On the Earth, we know one example of side chains meeting the above conditions for the ideal genetic polymers: nucleobases (Figure 7 (xvi)). They pair by a few hydrogen bonds at the center of the double helix structures of DNA and RNA [86]. They also have rigid and flat ring structures with several double bonds (including resonance structures). They are piled up neatly inside the double helix, stack with each other by their rich π electrons to form the hydrophobic middle layer, and stabilize the paired helix [87-94]. The size complementarity between one- and two-ring structures is also present in nucleobases. Thus, although the existence of different letters with similarly ideal or exceptional properties cannot be excluded (e.g., a hydrophobic shape-complementary pair of artificial nucleobases) [95], the structures of the terrestrial

genetic letters seem to be at least one of the simplest forms in the universe to meet such a number of conditions for ideal genetic letters.

In contrast to the side chains, various choices might be allowed for the main chains of the genetic polymers, as they are likely to be the outer layer and would have fewer structural restrictions. The main chains of the terrestrial genetic polymers are composed of sugars and phosphate. The sugars can be different even in our life system (ribose and deoxyribose), and can also alter the overall conformation of the nucleic acids (A-form and B-form). Artificial nucleic acids (or xeno nucleic acids, XNA) with different sugars (e.g., HNA) or even with non-sugar counterparts have also been developed and shown to store and convey information [96,97]. Thus, ribose and deoxyribose might have been selected by chemical availability or stability (including the stability of their double helix) under some specific conditions on the Earth.

Phosphate might also have been chosen for its chemical properties. It is not a structurally essential component as XNA without phosphate can also form base-paired double helices (e.g., PNA) [96-98]. With its negative charge, phosphate can repulse the attacking $OH^-$ in the aqueous environment, making the nucleic acids relatively stable [99]. Another chemical reason might be the strategy to achieve the synthesis of nucleic acids without uncontrolled degradation. Polymerization (dehydration) of monomer parts of nucleic acids (NMPs/dNMPs) does not occur spontaneously in aqueous environments. Thus, the substrates of the polymerase reactions must somehow become activated. Among a lot of possible activated substrates, polymerases do not use the simplest ones (NDPs/dNDPs), but the doubly activated ones (NTPs/dNTPs), paying higher synthetic costs. The reason for this choice is to prevent the reverse reaction [100,101]. After the incorporation of an NMP/dNMP in the elongating strand, pyrophosphate (PPi) is released and further degraded into two phosphates, which prevents PPi from re-entering the catalytic site of polymerases to trigger the reverse reaction (pyrophosphorolysis). If polymerases were to use NDPs/dNDPs as substrates, then the leaving group of the polymerization reaction would be phosphate, which would not be further degraded and inevitably cause the reverse reaction.

The doubly activated substrates can also function in another chemical equilibration (NTP ↔ NDP + Pi) in the same environment. Cells likely utilize this for reversible reactions that should be controlled in temporal energy/nutrition richness. Using the same substrates, non-reversible reactions including genetic polymer synthesis and reversible reactions in response to the fluctuating environments can be simultaneously performed with different chemical equilibrations [101]. The few repeating phosphates in a nucleotide are likely one of the simplest ways to realize such an elaborate control system for the complicated traffic of numerous chemical reactions in the cell. Furthermore, lipid membranes assembled by hydrophobic interactions can prevent phosphorylated cellular ingredients (e.g., nucleic acids and numerous metabolic intermediates) from leaking out, keeping the integrity of the cell [99,102]. If such phosphate chemistry systems were also adapted on another planet, then their genetic polymer might be something similar to ours.

Another possible reason for the main chain choice of nucleic acids on the Earth might have been related to the enzymatic functions of RNA. Most natural ribozymes perform phosphoryl transfer reactions, such as self-cleavage and its reverse ligation [103]. The mechanism is essentially an $S_N2$-type nucleophilic attack on the phosphodiester bond by a 2′ oxygen or a hydroxide ion, and sometimes catalyzed by $Mg^{2+}$ ions chelated between nearby phosphate groups. In other words, the RNA structure seems to be optimized for such self-editing reactions, while proteins catalyze diverse metabolic reactions. This might indicate RNA originally emerged as a selfish self-replicator [104,105].

**Conclusion: Co-evolution between Two Polymers**

In this review, I have discussed the structures of ideal biopolymers for elaborate life systems, considering the properties of representative bonds and interactions. This process was almost like reading a biochemistry textbook in reverse and discovering how our biopolymers elegantly meet the required conditions.

To perform their numerous chemical and mechanical functions, functional biopolymers must fold around hydrophobic cores and adopt various structures. They also need to have relatively rigid main chains with one hydrogen bond donor and one acceptor, to introduce regularity into otherwise random hydrophobic structures. Proteins are probably the simplest molecules to fulfill such conditions. They might also be the functional polymers of many extraterrestrial lives.

In contrast, genetic polymers should store and transmit information precisely. They need to form double helices with flat, stacking side chains paired by a few hydrogen bonds. DNA and RNA have ideal side chains, the nucleobases. Their main chains might have been selected for more chemistry-based reasons. The genetic polymers of extraterrestrial life might have some variety, especially in their main chains. Still, some of them might be very similar or even the same as ours.

It is worth mentioning that all of the above arguments suppose the existence of highly evolved lives. Simpler life-like entities composed of non-ideal molecules might have been present before the emergence of more elaborate life systems, or in special environments. Even on the Earth, nucleic acids or peptides might have played both catalytic and genetic roles in very primitive life systems, without cooperating with each other [1,104,105]. For instance, nucleic acids can catalyze chemical reactions (sometimes using modified nucleobases) [106-108], and peptides might also undergo self-templated replication [7-11].

However, ideal genetic polymers like DNA and RNA cannot be as functional as proteins, as they are all thumbs. Their side chains are inevitably large and have limited varieties, which would result in much less diversity per volume. In contrast, ideal catalytic polymers like proteins have various side chains, including small hydrophobic ones, which would not be replicated precisely. Therefore, the co-evolution between the ideal genetic polymers and the ideal catalytic polymers would be highly advantageous (molecular mutualism) [109] and might also be the only way to generate complicated life systems like ours.

## Conflicts of Interest

The author declares no conflicts of interest.

## Author Contribution

S.T. wrote the manuscript.

## Data Availability

The evidence data generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Acknowledgements

I was supported by JSPS (22H01346). I thank Sota Yagi for fruitful discussions.

## References

[1] Greenwald, J., Kwiatkowski, W., Riek, R. Peptide amyloids in the origin of life. J. Mol. Biol. 430, 3735–3750 (2018). https://doi.org/10.1016/j.jmb.2018.05.046
[2] Despotovic, D., Tawfik, D. S. Proto-proteins in protocells. ChemSystemsChem 3, e2100002 (2021). https://doi.org/10.1002/syst.202100002
[3] Tagami, S., Li, P. The origin of life: RNA and protein co-evolution on the ancient earth. Dev. Growth Differ. 65, 167–174 (2023). https://doi.org/10.1111/dgd.12845
[4] Brack, A., Orgel, L. E. β structures of alternating polypeptides and their possible prebiotic significance. Nature 256, 383–387 (1975). https://doi.org/10.1038/256383a0
[5] DeGrado, W. F., Wasserman, Z. R., Lear, J. D. Protein design, a minimalist approach. Science 243, 622–628 (1989). https://doi.org/10.1126/science.2464850
[6] DeGrado, W. F., Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. J. Am. Chem. Soc. 107, 7684–7689 (1985). https://doi.org/10.1021/ja00311a076
[7] Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K., Ghadiri, M. R. A self-replicating peptide. Nature 382, 525–528 (1996). https://doi.org/10.1038/382525a0
[8] Takahashi, Y., Mihara, H. Construction of a chemically and conformationally self-replicating system of amyloid-like fibrils. Bioorg. Med. Chem. 12, 693–699 (2004). https://doi.org/10.1016/j.bmc.2003.11.022
[9] Rubinov, B., Wagner, N., Rapaport, H., Ashkenasy, G. Self-replicating amphiphilic β-sheet peptides. Angew. Chem. Int. Ed. Engl. 48, 6683–6686 (2009). https://doi.org/10.1002/anie.200902790
[10] Nanda, J., Rubinov, B., Ivnitski, D., Mukherjee, R., Shtelman, E., Motro, Y., et al. Emergence of native peptide sequences in prebiotic replication networks. Nat. Commun. 8, 434 (2017). https://doi.org/10.1038/s41467-017-00463-1
[11] Rout, S. K., Friedmann, M. P., Riek, R., Greenwald, J. A prebiotic template-directed peptide synthesis based on amyloids. Nat. Commun. 9, 234 (2018). https://doi.org/10.1038/s41467-017-02742-3
[12] Frenkel-Pinter, M., Samanta, M., Ashkenasy, G., Leman, L. J. Prebiotic peptides: Molecular hubs in the origin of life. Chem. Rev. 120, 4707–4765 (2020). https://doi.org/10.1021/acs.chemrev.9b00664
[13] Li, P., Holliger, P., Tagami, S. Hydrophobic-cationic peptides modulate RNA polymerase ribozyme activity by accretion. Nat. Commun. 13, 3050 (2022). https://doi.org/10.1038/s41467-022-30590-3
[14] Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., Bartel, D. P. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. Science 292, 1319–1325 (2001). https://doi.org/10.1126/science.1060786

[15] Wochner, A., Attwater, J., Coulson, A., Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. Science 332, 209–212 (2011). https://doi.org/10.1126/science.1200752
[16] Blaber, M., Lee, J., Longo, L. Emergence of symmetric protein architecture from a simple peptide motif: Evolutionary models. Cell. Mol. Life Sci. 69, 3999–4006 (2012). https://doi.org/10.1007/s00018-012-1077-3
[17] Höcker, B. Design of proteins from smaller fragments—Learning from evolution. Curr. Opin. Struct. Biol. 27, 56–62 (2014). https://doi.org/10.1016/j.sbi.2014.04.007
[18] Alva, V., Lupas, A. N. From ancestral peptides to designed proteins. Curr. Opin. Struct. Biol. 48, 103–109 (2018). https://doi.org/10.1016/j.sbi.2017.11.006
[19] Vrancken, J. P. M., Tame, J. R. H., Voet, A. R. D. Development and applications of artificial symmetrical proteins. Comput. Struct. Biotechnol. J. 18, 3959–3968 (2020). https://doi.org/10.1016/j.csbj.2020.10.040
[20] Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., Wilmanns, M. Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. Science 289, 1546–1550 (2000). https://doi.org/10.1126/science.289.5484.1546
[21] Good, M. C., Greenstein, A. E., Young, T. A., Ng, H.-L., Alber, T. Sensor domain of the mycobacterium tuberculosis receptor Ser/Thr protein kinase, PknD, forms a highly symmetric β propeller. J. Mol. Biol. 339, 459–469 (2004). https://doi.org/10.1016/j.jmb.2004.03.063
[22] Murzin, A. G., Lesk, A. M., Chothia, C. β-Trefoil fold: Patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors J. Mol. Biol. 223, 531–543 (1992). https://doi.org/10.1016/0022-2836(92)90668-A
[23] Brych, S. R., Blaber, S. I., Logan, T. M., Blaber, M. Structure and stability effects of mutations designed to increase the primary sequence symmetry within the core region of a β-trefoil. Protein Sci. 10, 2587–2599 (2009). https://doi.org/10.1110/ps.ps.34701
[24] Coles, M., Diercks, T., Liermann, J., Gröger, A., Rockel, B., Baumeister, W., et al. The solution structure of VAT-N reveals a 'Missing Link' in the evolution of complex enzymes from a simple βαββ element. Curr. Biol. 9, 1158–1168 (1999). https://doi.org/10.1016/S0960-9822(00)80017-2
[25] Castillo, R. M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T. L., Murzin, A. G. A six-stranded double-psi beta barrel is shared by several protein superfamilies. Structure 7, 227–236 (1999). https://doi.org/10.1016/s0969-2126(99)80028-8
[26] Huang, P. S., Feldmeier, K., Parmeggiani, F., Velasco, D. F., Hocker, B., Baker, D. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. 12, 29–34 (2016). https://doi.org/10.1038/nchembio.1966
[27] Höcker, B., Lochner, A., Seitz, T., Claren, J., Sterner, R. High-resolution crystal structure of an artificial $(\beta\alpha)_8$ -barrel protein designed from identical half-barrels. Biochemistry 48, 1145–1147 (2009). https://doi.org/10.1021/bi802125b
[28] Voet, A. R. D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., et al. Computational design of a self-assembling symmetrical β-propeller protein. Proc. Natl. Acad. Sci. U.S.A. 111, 15102–15107 (2014). https://doi.org/10.1073/pnas.1412768111
[29] Smock, R. G., Yadid, I., Dym, O., Clarke, J., Tawfik, D. S. De novo evolutionary emergence of a symmetrical protein is shaped by folding constraints. Cell 164, 476–486 (2016). https://doi.org/10.1016/j.cell.2015.12.024
[30] Noguchi, H., Addy, C., Simoncini, D., Wouters, S., Mylemans, B., Van Meervelt, L., et al. Computational design of symmetrical eight-bladed β-propeller proteins. IUCrJ 6, 46–55 (2019). https://doi.org/10.1107/S205225251801480X
[31] Afanasieva, E., Chaudhuri, I., Martin, J., Hertle, E., Ursinus, A., Alva, V., et al. Structural diversity of oligomeric β-propellers with different numbers of identical blades. eLife 8, 49853 (2019). https://doi.org/10.7554/eLife.49853
[32] Mylemans, B., Laier, I., Kamata, K., Akashi, S., Noguchi, H., Tame, J. R. H., et al. Structural plasticity of a designer protein sheds light on β-propeller protein evolution. FEBS J. 288, 530–545 (2021). https://doi.org/10.1111/febs.15347
[33] Lee, J., Blaber, M. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. Proc. Natl. Acad. Sci. U.S.A. 108, 126–130 (2011). https://doi.org/10.1073/pnas.1015032108
[34] Broom, A., Doxey, A. C., Lobsanov, Y. D., Berthin, L. G., Rose, D. R., Howell, P. L., et al. Modular evolution and the origins of symmetry: Reconstruction of a three-fold symmetric globular protein. Structure 20, 161–171 (2012). https://doi.org/10.1016/j.str.2011.10.021
[35] Terada, D., Voet, A. R. D., Noguchi, H., Kamata, K., Ohki, M., Addy, C., et al. Computational design of a symmetrical β-trefoil lectin with cancer cell binding activity. Sci. Rep. 7, 5943 (2017). https://doi.org/10.1038/s41598-017-06332-7
[36] Longo, L. M., Lee, J., Blaber, M. Simplified protein design biased for prebiotic amino acids yields a foldable,

halophilic protein. Proc. Natl. Acad. Sci. U.S.A. 110, 2135–2139 (2013). https://doi.org/10.1073/pnas.1219530110
[37] Yagi, S., Padhi, A. K., Vucinic, J., Barbe, S., Schiex, T., Nakagawa, R., et al. Seven amino acid types suffice to create the core fold of RNA polymerase. J. Am. Chem. Soc. 143, 15998–16006 (2021). https://doi.org/10.1021/jacs.1c05367
[38] Plaxco, K. W., Riddle, D. S., Grantcharova, V., Baker, D. Simplified proteins: Minimalist solutions to the 'Protein Folding Problem.' Curr. Opin. Struct. Biol. 8, 80–85 (1998). https://doi.org/10.1016/S0959-440X(98)80013-4
[39] Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., et al. Functional rapidly folding proteins from simplified amino acid sequences. Nat. Struct. Biol. 4, 805–809 (1997). https://doi.org/10.1038/nsb1097-805
[40] Walter, K. U., Vamvaca, K., Hilvert, D. An active enzyme constructed from a 9-amino acid alphabet. J. Biol. Chem. 280, 37742–37746 (2005). https://doi.org/10.1074/jbc.M507210200
[41] Müller, M. M., Allison, J. R., Hongdilokkul, N., Gaillon, L., Kast, P., van Gunsteren, W. F., et al. Directed evolution of a model primordial enzyme provides insights into the development of the genetic code. PLoS Genet. 9, e1003187 (2013). https://doi.org/10.1371/journal.pgen.1003187
[42] Akanuma, S., Kigawa, T., Yokoyama, S. Combinatorial Mutagenesis to Restrict Amino Acid Usage in an Enzyme to a Reduced Set. Proc. Natl. Acad. Sci. U.S.A. 99, 13549–13553 (2002). https://doi.org/10.1073/pnas.222243999
[43] Akanuma, S., Yokobori, S., Nakajima, Y., Bessho, M., Yamagishi, A. Robustness of Predictions of Extremely Thermally Stable Proteins in Ancient Organisms. Evolution 69, 2954–2962 (2015). https://doi.org/10.1111/evo.12779
[44] Shibue, R., Sasamoto, T., Shimada, M., Zhang, B., Yamagishi, A., Akanuma, S. Comprehensive reduction of amino acid set in a protein suggests the importance of prebiotic amino acids for stable proteins. Sci. Rep. 8, 1227 (2018). https://doi.org/10.1038/s41598-018-19561-1
[45] Kimura, M., Akanuma, S. Reconstruction and characterization of thermally stable and catalytically active proteins comprising an alphabet of - 13 amino acids. J. Mol. Evol. 88, 372–381 (2020). https://doi.org/10.1007/s00239-020-09938-0
[46] Munson, M., Regan, L., O'Brien, R., Sturtevant, J. M. Redesigning the hydrophobic core of a four-helix-bundle protein. Protein Sci. 3, 2015–2022 (1994). https://doi.org/10.1002/pro.5560031114
[47] Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., et al. What makes a protein a protein? hydrophobic core designs that specify stability and structural properties. Protein Sci. 5, 1584–1593 (1996). https://doi.org/10.1002/pro.5560050813
[48] Gassner, N. C., Baase, W. A., Matthews, B. W. A test of the "Jigsaw Puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc. Natl. Acad. Sci. U.S.A. 93, 12155–12158 (1996). https://doi.org/10.1073/pnas.93.22.12155
[49] Gassner, N. C., Baase, W. A., Mooers, B. H. M., Busam, R. D., Weaver, L. H., Lindstrom, J. D., et al. Multiple methionine substitutions are tolerated in T4 lysozyme and have coupled effects on folding and stability. Biophys. Chem. 100, 325–340 (2002). https://doi.org/10.1016/S0301-4622(02)00290-9
[50] Desjarlais, J. R., Handel, T. M. De novo design of the hydrophobic cores of proteins. Protein Sci. 4, 2006–2018 (1995). https://doi.org/10.1002/pro.5560041006
[51] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with alphafold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2
[52] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. Protein complex prediction with alphafold-multimer. bioRxiv (2021). https://doi.org/10.1101/2021.10.04.463034
[53] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M. ColabFold: Making protein folding accessible to all. Nat. Methods 19, 679–682 (2022). https://doi.org/10.1038/s41592-022-01488-1
[54] Weber, A. L., Miller, S. L. Reasons for the occurrence of the twenty coded protein amino acids. J. Mol. Evol. 17, 273–284 (1981). https://doi.org/10.1007/BF01795749
[55] Makarov, M., Sanchez Rocha, A. C., Krystufek, R., Cherepashuk, I., Dzmitruk, V., Charnavets, T., et al. Early selection of the amino acid alphabet was adaptively shaped by biophysical constraints of foldability. J. Am. Chem. Soc. 145, 5320–5329 (2023). https://doi.org/10.1021/jacs.2c12987
[56] Koga, R., Yamamoto, M., Kosugi, T., Kobayashi, N., Sugiki, T., Fujiwara, T., et al. Robust Folding of a de novo designed ideal protein even with most of the core mutated to valine. Proc. Natl. Acad. Sci. U.S.A. 117, 31149–31156 (2020). https://doi.org/10.1073/pnas.2002120117
[57] Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T., et al. Principles for designing ideal protein structures. Nature 491, 222–227 (2012). https://doi.org/10.1038/nature11600
[58] Lin, Y.-R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A. F., Montelione, G. T., et al. Control over overall

shape and size in de novo designed proteins. Proc. Natl. Acad. Sci. U.S.A. 112, E5478–E5485 (2015). https://doi.org/10.1073/pnas.1509508112

[59] Koga, N., Koga, R., Liu, G., Castellanos, J., Montelione, G. T., Baker, D. Role of backbone strain in de novo design of complex α/β protein structures. Nat. Commun. 12, 3921 (2021). https://doi.org/10.1038/s41467-021-24050-7

[60] Axe, D. D., Foster, N. W., Fersht, A. R. Active barnase variants with completely random hydrophobic cores. Proc. Natl. Acad. Sci. U.S.A. 93, 5590–5594 (1996). https://doi.org/10.1073/pnas.93.11.5590

[61] O'Brien, R., Driscoll, P. C., Davis, B., Ladbury, J. E., Wynn, R., Plaxco, K. W., et al. The adaptability of escherichia coli thioredoxin to non-conservative amino acid substitutions. Protein Sci. 6, 1325–1332 (1997). https://doi.org/10.1002/pro.5560060621

[62] Kitadai, N., Maruyama, S. Origins of building blocks of life: A review. Geosci. Front. 9, 1117–1153 (2018). https://doi.org/10.1016/j.gsf.2017.07.007

[63] Kitadai, N., Nishiuchi, K., Nishii, A., Fukushi, K. Amorphous silica-promoted lysine dimerization: A thermodynamic prediction. Orig. Life Evol. Biosph. 48, 23–34 (2018). https://doi.org/10.1007/s11084-017-9548-z

[64] Frenkel-Pinter, M., Haynes, J. W., Martin, C, M., Petrov, A. S., Burcar, B. T., Krishnamurthy, R., et al. Selective incorporation of proteinaceous over nonproteinaceous cationic amino acids in model prebiotic oligomerization reactions. Proc. Natl. Acad. Sci. U.S.A. 116, 16338–16346 (2019). https://doi.org/10.1073/pnas.1904849116

[65] Canavelli, P., Islam, S., Powner, M. W. Peptide ligation by chemoselective aminonitrile coupling in water. Nature 571, 546–549 (2019). https://doi.org/10.1038/s41586-019-1371-4

[66] Thoma, B., Powner, M. W. Selective synthesis of lysine peptides and the prebiotically plausible synthesis of catalytically active diaminopropionic acid peptide nitriles in water. J. Am. Chem. Soc. 145, 3121–3130 (2023). https://doi.org/10.1021/jacs.2c12497

[67] Frenkel-Pinter, M., Haynes, J. W., Mohyeldin, A. M., C, M., Sargon, A. B., Petrov, A. S., et al. Mutually stabilizing interactions between proto-peptides and RNA. Nat. Commun. 11, 3137 (2020). https://doi.org/10.1038/s41467-020-16891-5

[68] Fetrow, J. S., Godzik, A. Function driven protein evolution. a possible proto-protein for the RNA-binding proteins. Pac. Symp. Biocomput. 485–496 (1998).

[69] Söding, J., Lupas, A. N. More than the sum of their parts: On the evolution of proteins from peptides. BioEssays 25, 837–846 (2003). https://doi.org/10.1002/bies.10321

[70] Romero Romero, M. L., Rabin, A., Tawfik, D. S. Functional proteins from short peptides: Dayhoff's hypothesis turns 50. Angew. Chemie Int. Ed. 55, 15966–15971 (2016). https://doi.org/10.1002/anie.201609977

[71] Laurino, P., Tóth-Petróczy, Á., Meana-Pañeda, R., Lin, W., Truhlar, D. G., Tawfik, D. S. An ancient fingerprint indicates the common ancestry of rossmann-fold enzymes utilizing different ribose-based cofactors. PLoS Biol. 14, e1002396 (2016). https://doi.org/10.1371/journal.pbio.1002396

[72] Romero Romero, M. L., Yang, F., Lin, Y.-R., Toth-Petroczy, A., Berezovsky, I. N., Goncearenco, A., et al. Simple yet functional phosphate-loop proteins. Proc. Natl. Acad. Sci. U.S.A. 115, E11943–E11950 (2018). https://doi.org/10.1073/pnas.1812400115

[73] Longo, L. M., Jabłońska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., et al. On the emergence of p-loop ntpase and rossmann enzymes from a beta-alpha-beta ancestral fragment. eLife 9, 64415 (2020). https://doi.org/10.7554/eLife.64415

[74] Vyas, P., Trofimyuk, O., Longo, L. M., Deshmukh, F. K., Sharon, M., Tawfik, D. S. Helicase-like functions in phosphate loop containing beta-alpha polypeptides. Proc. Natl. Acad. Sci. U.S.A. 118, e2016131118 (2021). https://doi.org/10.1073/pnas.2016131118

[75] Vyas, P., Malitsky, S., Itkin, M., Tawfik, D. S. On the origins of enzymes: Phosphate-binding polypeptides mediate phosphoryl transfer to synthesize adenosine triphosphate. J. Am. Chem. Soc. 145, 8344–8354 (2023). https://doi.org/10.1021/jacs.2c08636

[76] Longo, L. M., Despotović, D., Weil-Ktorza, O., Walker, M. J., Jabłońska, J., Fridmann-Sirkis, Y., et al. Primordial emergence of a nucleic acid-binding protein via phase separation and statistical ornithine-to-arginine conversion. Proc. Natl. Acad. Sci. U.S.A. 117, 15731–15739 (2020). https://doi.org/10.1073/pnas.2001989117

[77] Seal, M., Weil-Ktorza, O., Despotović, D., Tawfik, D. S., Levy, Y., Metanis, N., et al. Peptide-RNA coacervates as a cradle for the evolution of folded domains. J. Am. Chem. Soc. 144, 14150–14160 (2022). https://doi.org/10.1021/jacs.2c03819

[78] Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., Hecht, M. H. Protein design by binary patterning of polar and nonpolar amino acids. Science 262, 1680–1685 (1993). https://doi.org/10.1126/science.8259512

[79] Tretyachenko, V., Vymětal, J., Neuwirthová, T., Vondrášek, J., Fujishima, K., Hlouchová, K. Modern and prebiotic amino acids support distinct structural profiles in proteins. Open Biol. 12, 220040 (2022).

https://doi.org/10.1098/rsob.220040
[80] Corin, K., Bowie, J. U. How physical forces drive the process of helical membrane protein folding. EMBO Rep. 23, e53025 (2022). https://doi.org/10.15252/embr.202153025
[81] Israelachvili, J. N. Intermolecular and surface forces (Elsevier, Burlington, 2011). https://doi.org/10.1016/C2009-0-21560-1
[82] Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., et al. The physiology and habitat of the last universal common ancestor. Nat. Microbiol. 1, 16116 (2016). https://doi.org/10.1038/nmicrobiol.2016.116
[83] Runnels, C. M., Lanier, K. A., Williams, J. K., Bowman, J. C., Petrov, A. S., Hud, N. V., et al. Folding, assembly, and persistence: The essential nature and origins of biopolymers. J. Mol. Evol. 86, 598–610 (2018). https://doi.org/10.1007/s00239-018-9876-2
[84] Ramachandran, G. N., Sasisekharan, V. Conformation of polypeptides and proteins. Adv. Protein Chem. 23, 283–437 (1968). https://doi.org/10.1016/S0065-3233(08)60402-7
[85] Cantor, C. R., Schimmel, P. R. Biophysical chemistry - Part I: The Conformation of Biological, Macromolecules (W. H. Freeman and Company, New York, 1980).
[86] Watson, J. D., Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature 171, 737–738 (1953). https://doi.org/10.1038/171737a0
[87] Herskovits, T. T. Nonaqueous solutions of DNA: Factors determining the stability of the helical configuration in solution. Arch. Biochem. Biophys. 97, 474–484 (1962). https://doi.org/10.1016/0003-9861(62)90110-8
[88] Sínanoĝlu, O., Abdulnur, S. Hydrophobic stacking of bases and the solvent denaturation of DNA. Photochem. Photobiol. 3, 333–342 (1964). https://doi.org/10.1111/j.1751-1097.1964.tb08156.x
[89] Frank-Kamenetskii, M. D. Biophysics of the DNA molecule. Phys. Rep. 288, 13–60 (1997). https://doi.org/10.1016/S0370-1573(97)00020-3
[90] Yakovchuk, P., Protozanova, E., Frank-Kamenetskii, M. D. Base-stacking and base- pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res. 34, 564–574 (2006). https://doi.org/10.1093/nar/gkj454
[91] Feng, B., Sosa, R. P., Mårtensson, A. K. F., Jiang, K., Tong, A., Dorfman, K. D., et al. Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. Proc. Natl. Acad. Sci. U.S.A. 116, 17169–17174 (2019). https://doi.org/10.1073/pnas.1909122116
[92] Lindman, B., Medronho, B., Alves, L., Norgren, M., Nordenskiöld, L. Hydrophobic interactions control the self-assembly of DNA and cellulose. Q. Rev. Biophys. 54, e3 (2021). https://doi.org/10.1017/S0033583521000019
[93] Privalov, P. L., Crane-Robinson, C. Forces maintaining the DNA double helix. Eur. Biophys. J. 49, 315–321 (2020). https://doi.org/10.1007/s00249-020-01437-w
[94] Poater, J., Swart, M., Bickelhaupt, F. M., Fonseca Guerra, C. B-DNA structure and stability: The role of hydrogen bonding, π–π stacking interactions, twist-angle, and solvation. Org. Biomol. Chem. 12, 4691–4700 (2014). https://doi.org/10.1039/C4OB00427B
[95] Lee, K. H., Hamashima, K., Kimoto, M., Hirao, I. Genetic alphabet expansion biotechnology by creating unnatural base pairs. Curr. Opin. Biotechnol. 51, 8–15 (2018). https://doi.org/10.1016/j.copbio.2017.09.006
[96] Duffy, K., Arangundy-Franklin, S., Holliger, P. Modified nucleic acids: Replication, evolution, and next-generation therapeutics. BMC Biol. 18, 112 (2020). https://doi.org/10.1186/s12915-020-00803-6
[97] Asanuma, H., Kamiya, Y., Kashida, H., Murayama, K. Xeno Nucleic Acids (XNAs) having non-ribose scaffolds with unique supramolecular properties. Chem. Commun. 58, 3993–4004 (2022). https://doi.org/10.1039/D1CC05868A
[98] Rasmussen, H., Kastrup, J. S., Nielsen, J. N., Nielsen, J. M., Nielsen, P. E. Crystal structure of a Peptide Nucleic Acid (PNA) duplex at 1.7 Å resolution. Nat. Struct. Biol. 4, 98–101 (1997). https://doi.org/10.1038/nsb0297-98
[99] Westheimer, F. H. Why nature chose phosphates. Science 235, 1173–1178 (1987). https://doi.org/10.1126/science.2434996
[100] Kornberg, A. On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions. in Horizons in Biochemistry (Kasha, M., Pullman, B., eds.) pp. 251–264 (Academic Press, New York, 1962).
[101] Wimmer, J. L. E., Kleinermanns, K., Martin, W. F. Pyrophosphate and irreversibility in evolution, or why ppi is not an energy currency and why nature chose triphosphates. Front. Microbiol. 12, 759359 (2021). https://doi.org/10.3389/fmicb.2021.759359
[102] Davis, B. D. On the importance of being ionized. arch. Biochem. Biophys. 78, 497–509 (1958). https://doi.org/10.1016/0003-9861(58)90374-6
[103] Alonso, D., Mondragón, A. Mechanisms of catalytic RNA molecules. Biochem. Soc.Trans. 49, 1529–1535 (2021). https://doi.org/10.1042/BST20200465
[104] Gilbert, W. Origin of life: The RNA world. Nature 319, 618 (1986). https://doi.org/10.1038/319618a0

[105] Joyce, G. F. RNA evolution and the origins of life. Nature 338, 217–224 (1989). https://doi.org/10.1038/338217a0
[106] Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., Cech, T. R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell 31, 147–157 (1982). https://doi.org/10.1016/0092-8674(82)90414-7
[107] Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S. The RNA moiety of ribonuclease p is the catalytic subunit of the enzyme. Cell 35, 849–857 (1983). https://doi.org/10.1016/0092-8674(83)90117-4
[108] Müller, F., Escobar, L., Xu, F., Węgrzyn, E., Nainytė, M., Amatov, T., et al. A prebiotically plausible scenario of an RNA–peptide world. Nature 605, 279–284 (2022). https://doi.org/10.1038/s41586-022-04676-3
[109] Lanier, K. A., Petrov, A. S., Williams, L. D. The central symbiosis of molecular biology: Molecules in mutualism. J. Mol. Evol. 85, 8–13 (2017). https://doi.org/10.1007/s00239-017-9804-x