life

Review

Xeno Amino Acids: A Look into Biochemistry as We Do Not Know It

Sean M. Brown *,†, Christopher Mayer-Bacon and Stephen Freeland †

Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore, MD 21250, USA
* Correspondence: sbrown16@umbc.edu
† These authors contributed equally to this work.

Abstract: Would another origin of life resemble Earth's biochemical use of amino acids? Here, we review current knowledge at three levels: (1) Could other classes of chemical structure serve as building blocks for biopolymer structure and catalysis? Amino acids now seem both readily available to, and a plausible chemical attractor for, life as we do not know it. Amino acids thus remain important and tractable targets for astrobiological research. (2) If amino acids are used, would we expect the same L-alpha-structural subclass used by life? Despite numerous ideas, it is not clear why life favors L-enantiomers. It seems clearer, however, why life on Earth uses the shortest possible (alpha-) amino acid backbone, and why each carries only one side chain. However, assertions that other backbones are physicochemically impossible have relaxed into arguments that they are disadvantageous. (3) Would we expect a similar set of side chains to those within the genetic code? Many plausible alternatives exist. Furthermore, evidence exists for both evolutionary advantage and physicochemical constraint as explanatory factors for those encoded by life. Overall, as focus shifts from amino acids as a chemical class to specific side chains used by post-LUCA biology, the probable role of physicochemical constraint diminishes relative to that of biological evolution. Exciting opportunities now present themselves for laboratory work and computing to explore how changing the amino acid alphabet alters the universe of protein folds. Near-term milestones include: (a) expanding evidence about amino acids as attractors within chemical evolution; (b) extending characterization of other backbones relative to biological proteins; and (c) merging computing and laboratory explorations of structures and functions unlocked by xeno peptides.

Keywords: astrobiology; amino acid; review; xenobiology; peptide biochemistry; abiogenesis; evolution

1. Introduction

A key question for astrobiology is whether life originating elsewhere in the universe would share similar biochemistry to that of life on Earth. Here, we narrow the challenging focus of that question to the topic of amino acids.

A foundational step of early biological evolution was to establish a genetically encoded 'alphabet' comprising 20 different amino acids, often known as the canonical set. Since then, the greatest deviations in ~3.5 billion years have been the addition of a 21st amino acid (Selenocysteine, Sec) within some lineages of bacteria [1], archaea [2] and eukaryotes [3]; and a 22nd (Pyrrolysine, Pyl), in two of these three domains (archaea [4] and bacteria [5]). This clear process of evolutionary extension from 20 to 22 [6] complements evidence that the canonical set of 20 is itself an outcome of biological evolution rather than a chemical prerequisite for life to begin. And yet, it is the canonical alphabet of 20, a foundation of biological, biochemical, and biomedical research, where knowledge has accumulated.

Xenobiology is an "*emergent technoscience . . . based on unusual biochemistries*" [7]. In this sense, we refer below to amino acids from beyond life's standard genetic code as "xeno" amino acids. The very architecture built to facilitate contemporary biological research

constrains how little we know about possibilities for biochemistry based on xeno amino acids. It is still new biotechnology to develop laboratory protocols for manipulating and analyzing biological proteins beyond the genetically encoded 20 amino acids (e.g., compare [8] with [9]). The data and tools of bioinformatics remain mostly built around an assumption that any site within a biological protein can exist in one of 20 states. However, such a fundamental feature of life on Earth offers a tempting potential for developing tractable, focused ideas about agnostic biosignatures. Whatever can be established about the likelihood of life elsewhere in the universe using amino acids or, better yet, about the characteristics of a "life-sustaining set," is a direct and significant contribution to current astrobiology.

Here, we review amino acids (Box 1) by revisiting and expanding three questions first introduced by Weber and Miller [10]: (1) Why does life on Earth use amino acids, rather than some other class of molecule? (2) Why does it use L-*α*-amino acids rather than other structural sub-classes? and (3) Why does the post-LUCA genetic code comprise 20-22 specific side chains? But whereas Weber and Miller approached the topic as chemists—summarizing what some call "bottom-up" thinking [11]—we approach the topic from biology, reasoning "top-down" as we work backward from life as we know it. In addressing each question, we focus on what is known and what is unknown about "xeno" amino acids—those from beyond the genetically encoded alphabet of 20.

Box 1. The Role of Amino Acids in Terrestrial Biology.

Everything alive today constructs metabolism primarily as a network of genetically encoded proteins. Each protein is a polymerized sequence of amino acids. In 1972, Christian Anfinsen was awarded the Nobel Prize in Chemistry for demonstrating that a protein's primary sequence (i.e., which members of the amino acid alphabet are joined together and in what order) determines how a linear polymer folds into a three dimensional conformation (Anfinsen, 1973). Since LUCA, all life on Earth genetically encodes 20 of these amino acids, although some lineages are evolving to add selenocysteine and/or pyrrolysine. Each of these 20 (+2) genetically encoded amino acids (**A**) is defined by a constant backbone (**B**): an amine (-NH2) at one end, a carboxyl (-COOH) at the other, and an "alpha" carbon atom between these two functional groups. This carbon atom carries a variable side chain (R-group), and differences between these side-chains distinguish each amino acid. Proteins are formed when a covalent, peptide bond links the carboxyl group (-COOH) of one amino acid to the amino group (-NH2) of another (**C**). The resulting thread-like backbone of every protein contains a series of rotatable bonds (ϕ and Ψ) within this peptide chain (**D**). The Phi (ϕ) and Psi (Ψ) angles around each alpha carbon define any given protein's 3-dimensional structure (**E**).
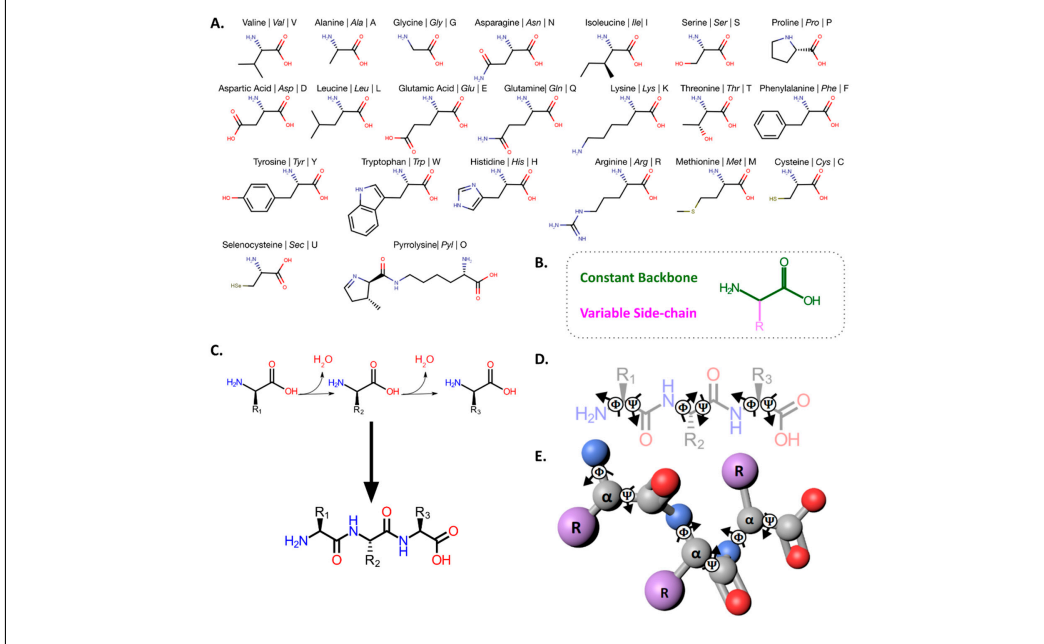
2. Would a Xeno Biochemistry Use Amino Acids?

An exploration of amino acids' relevance to xeno biochemistry can usefully begin with one, simple observation: the standard or canonical alphabet of 20 genetically encoded amino acids is extremely good at what it does. Protein-based metabolism constructed with this one set of molecular building blocks has diversified so successfully that contemporary scientists remain actively engaged in finding environmental limits to life on Earth (e.g., [12]: Table 4). Genetically encoded proteins sustain biology, for example, in polar volcanoes [13], nuclear-contaminated sites [14], and "toxic" acid-mine drainage [15]. A current understanding is that "*water activity appears to be the single key parameter controlling the biospace of Earth's life, and numerous other parameters limiting life (e.g., temperature and salinity) are, in fact, acting on the availability of water*" [12]. Indeed, life on Earth is now recognized to flourish across a range of conditions that overlap significantly with extraterrestrial environments (*ibid*: Figure 2). The standard amino acid alphabet has even proven sufficient to sustain life for three years on the exterior of the International Space Station [16] and evidence is growing that Earth life could travel between different planetary bodies [17]!
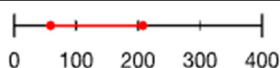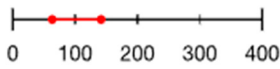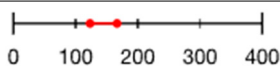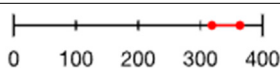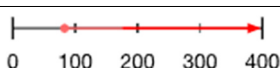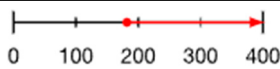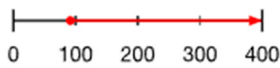
Of course, amino acids' impressive potential for constructing polymer catalysts is merely consistent with, not evidence for, their likely use within a xeno biochemistry. An important, complementary question follows: would an independent origin of life "discover" this type of organic molecule? In the mid-20th century, "spark tube" experiments produced the first suggestion of an affirmative answer by simulating physicochemical conditions thought to represent a prebiotic planet Earth [18,19]. A circulating mixture of water, methane, hydrogen, and ammonia provided with energy in the form of heat and an electric spark was shown to produce amino acids, among many other organic compounds. This direct connection between the abiotic universe and fundamental biochemistry inspired an entire literature which, in short, reveals that variations in both energy source and reactants change only the quantity of amino acids produced, and the diversity of side chains, not their presence/absence as a class of chemical structure (see [20] for a thorough review, but with plentiful, continuing research, e.g., [21–23]).

Since the 1970s, this direct connection between the abiotic universe and biochemistry has been affirmed by analysis of carbonaceous meteorites [24], the organic chemistry of which provides a natural analog to laboratory simulations [25–27]. Indeed, advances in instrumentation reveal an increasingly diverse repertoire of amino acids among the organic compounds found both within newly discovered meteorites (e.g., [28,29]) and reanalysis of those studied previously (e.g., [30]). Recent advances in space sciences are now removing the need to wait for meteorites to fall to Earth as earlier this year; fifteen different amino acids were identified in situ on the Ryugu asteroid [31] formed by multiple reaction pathways [32]. Echoing the broad findings of prebiotic simulations, it is not the presence/absence of amino acids that changes in these meteorites, but rather "*The abundances [that] vary significantly [according to] different degrees of secondary alteration processes including thermal and aqueous alteration*" [33]. Contemporary science is increasingly clear that abiotic organic chemistry synthesizes amino acids almost anywhere that sufficient energy melts ice into water in the presence of organic carbon and nitrogen.

This cosmic ubiquity distinguishes amino acids from other fundamental components of biochemistry (Figure 1). In addition to proteins, life as we know it comprises genetic material in the form of polymerized nucleotide sequences and is encapsulated within lipid membranes. Neither lipids nor nucleotides form easily under prebiotic conditions. Certainly, fatty acids occur and could potentially play a role as forerunners to lipids [34,35], and nucleobases, a subcomponent of nucleotides, also occur [36,37] but prebiotic synthesis of nucleotides themselves is far more controversial [38–40]. Broadly speaking, the difference can be understood from the number and types of atoms involved: amino acids in general, and those produced by abiotic synthesis in particular, comprise fewer atoms than lipids or nucleotides (Table 1). Not only do larger molecules imply the need for more atoms to find and react with one another in the absence of any guiding enzyme, but the addition of each new heavy atom brings exponentially expanding structural combinations [41]. Thus ribose

($C_5H_{10}O_5$) is formed by one of the oldest organic syntheses known to science [42], but in the absence of catalysis [43], total synthesis yield divides between countless other structures that share a similar chemical formula [44,45]. Such simple generalizations of course ignore many sophisticated considerations, most notably reaction pathway dynamics and a role for non-biological catalysts, but their usefulness is supported by noting that nucleobases and fatty acids fall within a molecular weight range similar to that of abiotically plausible amino acids, while nucleotides and lipids do not (Table 1).

Table 1. Amino acids are smaller than other components of life's biochemistry.

	Heavy Atoms	Molecular Weight (g/mol)	Chemical Elements
Coded Amino Acids (ACDEFGHIKLMNPQRSTVWY)	5–15		CHONS
'Prebiotic' Amino Acids (ADEGILPSTV)	5–8		CHONS
Nucleobases	8–11		CHONP
Nucleotides *Nucleobase + Ribose + PO$_4$*	23–24		CHONP
Fatty Acids	≥5		CHO
Propionic acid §	5	74	
Decanoic acid †	12	172	
Lipids	≥12		CHO
Triformin *	12	176	
Sugars *Monosaccharides*	≥6		CHO
Triose ‡	6	90	
Ribose	10	150	

* Smallest triglyceride; ‡ smallest monosaccharide; † prebiotically plausible membrane former [35]; § smallest fatty acid.

Beyond mere atom counts, amino acids distinguish themselves by chemical composition from the other fundamental components of biochemistry. All components comprise just six chemical elements (C, H, N, O, P, and S). Excluding only the noble gasses (He, Ne, and Ar), four of these six 'biochemical' elements (C, H, O, and N) are the most abundant atoms in the universe, and are sufficient to produce 18 of the 20 genetically encoded amino acids. The remaining two amino acids require only the addition of sulfur, which follows close behind in terms of abundance [46]. In contrast, both nucleotides and biological membrane lipids incorporate phosphorus, which is generally far less abundant than C, H, O, N, or S. Again, this argument ignores a host of more sophisticated considerations, such as microenvironments that may have delivered phosphorus to an origin of life [47]. Overall it is clear, however, that prebiotic synthesis of RNA remains persistently more challenging than that of amino acids. The world's leading research here continues to search for whatever it is that all previous efforts have missed. Either the right kind of mineral surface was needed to catalyze the pathways which form and derivatize ribose [48], or no minerals are required because nucleobases lacked any backbone in life's earliest stages [49], or some combination of physicochemical conditions, not yet tried, is the missing answer. Answers here are potentially endless: recent examples include photochemistry [50] or the sort of cyclical self-purification of RNA from within a more heterogeneous polymer [51] that we describe below for the case of amino acids versus hydroxy acids [52]. In this context, it is noteworthy that at least some of the ingenious chemistry developing here is overtly

motivated by the perception that an RNA world paradigm provides "*a mandate for chemistry to explain how RNA might have been generated prebiotically on the early Earth*" [53], and must thus be balanced against serious arguments that RNA might instead be a product of early biological evolution rather than a prerequisite [38,39,54].
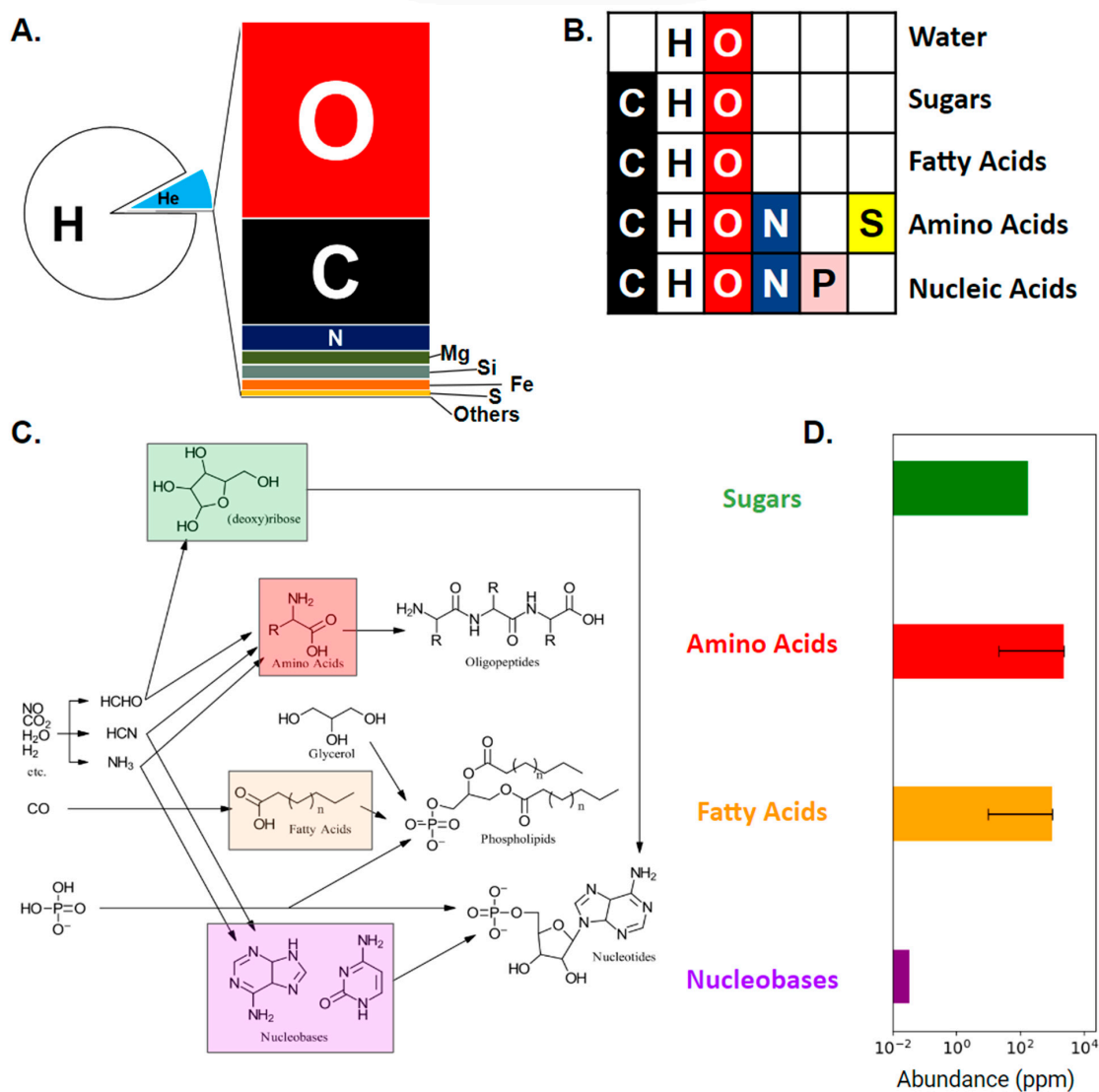
Figure 1. Life's fundamental biochemistry comprises just six chemical elements (carbon, nitrogen, hydrogen, oxygen, sulfur, and phosphorus). (**A**) The atomic composition of the Milky Way Galaxy [55] is primarily dominated by hydrogen and helium, but the remaining portion is dominated by oxygen, carbon, and nitrogen. (**B**) Carbon, nitrogen, hydrogen, oxygen, sulfur, and phosphorus are distributed between five classes of important biomolecules. Nitrogen occurs in what are arguably the two most important—genetic information (nucleic acid) and the structural and catalytic molecules that interact to produce metabolism (proteins). (**C**) Simplified abiotic synthetic pathways of life's biochemical building blocks (adapted from [56]). (**D**) These fundamental building blocks are found in meteorites (shown in log scale). Sugars are found up to 180 parts per million (ppm) in the Murchison (carbonaceous chondrite, CM) meteorite [57]. Amino acids, the most abundant, can be found up to 21 ppm within CM Chondrites and 2400 ppm within CR Chondrites [26,27]. Fatty acids can be found up to 1000 ppm and 10 ppm within CM and CR Chondrites, respectively [58]. Nucleobases are found least abundantly up to 34 parts per billion (ppb) in the Murchison meteorite [59]; nucleotides have never been detected within extraterrestrial material.

Looking beyond components of fundamental biochemistry, many other organics form under plausible prebiotic conditions. Some, such as sulfonic and hydroxy acids, are fully capable of forming polymers [60,61]. While the functional potential of polymers made from these alternatives is underexplored, especially their potential to form catalytic enzyme analogs, amino acids already show some unexpected advantages to an origin of life. The esters that connect hydroxy acids, and the thioesters which connect sulfonic acids are, for example, less stable to hydrolysis than the peptide bonds which link amino acids [62]. This difference in stability contributes to a self-purification of depsipeptides (heteropolymers comprising a mixture of amino acids and hydroxy acids) in an environment that cycles through wet and dry conditions (Figure 2 and [52]). The most likely abiotic (hetero)polymers show the potential to develop into peptide sequences during chemical evolution. Spontaneous self-purification towards amino acid enrichment of depsipeptides is particularly relevant to extraterrestrial life, because it suggests how even different starting points for polymer-based catalysis might converge upon amino acids over time.
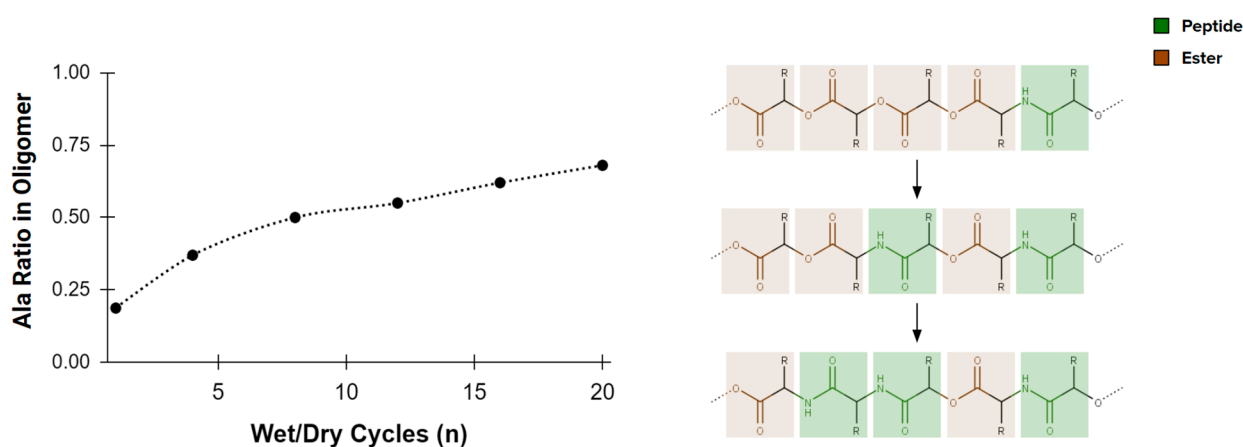
Figure 2. Heteropolymers comprising a mixture of amino acids and hydroxy acids (depsipeptides), exposed to wet-dry cycling, become enriched in amino acids (adapted from [52]). This enrichment, in part, is due to the stability difference between peptide (C-N) and ester (C-O-C) bonds, suggesting the eventual convergence of an amino acid homopolymer (peptide) over time.

In summary, amino acids emulate the same properties that motivate current explorations for extraterrestrial life to "*follow the water*" [63,64]. While we cannot rule out life-sustaining possibilities of other solvents such as methane [65] or supercritical carbon dioxide [66], the unique biophysics of water (e.g., [67,68]) combines with its cosmic abundance [69] to justify its central role within biochemistry. The same themes of ready availability and unusually useful physicochemical properties are true of amino acids: they form almost unstoppably within the abiotic organic chemistry that occurs in the presence of liquid water (e.g., [33]). They can spontaneously enrich within heteropolymer sequences (e.g., [52]). Once polymerized, they display an amazing versatility to perform catalytic and structural roles in environments which overlap significantly with those identified for extraterrestrial environments (e.g., [12]). Thus, while certainties elude any scientific inquiry that looks beyond biochemistry as we know it, amino acids are excellent candidates with which to logically extend the current search for extraterrestrial life around water.

3. Would a Xeno Biochemistry Use Monosubstituted L-α-Amino Acids?

3.1. α-Amino Acids versus Longer Backbones

Viewing amino acids as building blocks for biopolymers offers further explanatory power to turn from "L-" to "α-" as a feature of genetically encoded amino acids. Alpha amino acids are those in which a single carbon atom is situated between the C- and N-termini (Box 1). However, the number of carbon atoms here can be larger, with α-amino acids as the simplest structural subclass within a theoretically infinite series: β-, γ-, δ-,

etc. (Figure 3). Each carbon atom added to the backbone adds two different positions at which a side chain could attach. Thus, whereas an α-carbon atom can be only mono- or di-substituted, the addition of a β-carbon atom permits up to four side chains, γ- up to six, and so on. The addition of each backbone carbon atom, a potential new chiral center, thus increases structural possibilities exponentially. For example, the six possible carbon side chain attachment sites in a γ-amino acid imply that two different side chains could occupy any of 30 ($_6P_2$) different permutations and six different side chains could be arranged in 720 different ways ($_6P_6$).
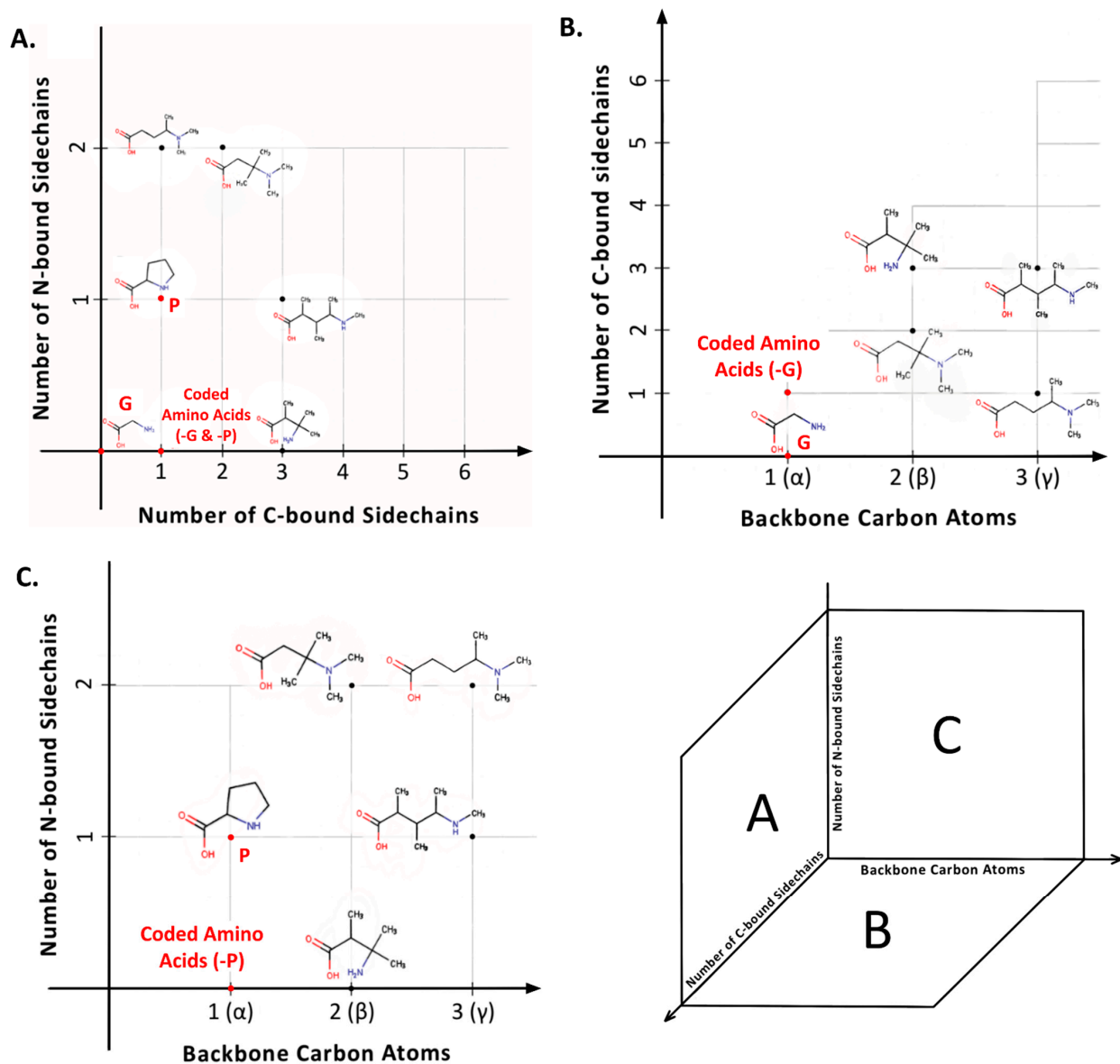
Figure 3. The universe of amino acid structures. (**A**) Distribution of amino acids based on the number of C-bound side chains vs. N-bound side chains (genetically encoded amino acids highlighted: alanine, proline, and glycine). (**B**) With each additional C-atom in the backbone, the number of possible C-bound side chain attachment sites increase by 2. The coded amino acids, except glycine, are merely a point in this possible space. (**C**) While C-bound side chain attachment sites are theoretically infinite, the backbone nitrogen can only attach two side chains while retaining its neutral valence. Here, the genetically encoded amino acids, except proline, exist in simply one point within the possible space.

Just as was the case for chiral alternatives, longer amino acid backbones are far more than a theoretical possibility. The same laboratory simulations and meteorite analyses that indicate the prebiotic plausibility of D-enantiomers and disubstituted α-amino acids also reveal the presence of β-, γ-, and δ-amino acids ([20]: Table 1). Further echoing the situation described above for alternatives to the L-enantiomer, amino acids with additional backbone carbon atoms also occur throughout present-day biology [70]. Once again, themes of cell signaling and allelochemistry (defensive and offensive toxicity) surface amidst a broad repertoire of functions. For example, the simplest γ-amino acid, gamma-aminobutyric acid (GABA), functions as both an important neurotransmitter within animals and a cell signaling molecule in plants [71,72] whereas β-Methylamino-L-alanine, or BMAA, is a powerful neurotoxin to mammals produced by cyanobacteria [73] and plants [74]. Clearly, it is not beyond the reach of biological evolution to work with longer backbones or indeed a heterogeneous mixture of different backbone types.

From the perspective of peptide folding, however, longer backbones produce less stable secondary structures. Each carbon–carbon bond within the amino acid backbone permits rotation that increases the flexibility of a peptide chain [75–77] (also see Figure 3). But while the lack of backbone rigidity has long been noted as one simple reason why natural selection would favor α-amino acids for the genetic code [10], synthetic biology demonstrates that β-amino acid peptide structures are possible [78–80]. Thus, arguments for the exclusion of β-, γ-, and δ-amino acids from the genetic code again resemble those for homochirality: a more robust explanation than strict biophysical constraint is evolutionary optimization. If this is an evolutionary outcome, then perhaps one might need to look no further than the higher energy cost for any cell working with a more diverse repertoire of building blocks. For example, to polymerize a mixture of "α- and β-amino acids, four enzymes would probably be required. One enzyme would be needed for each of the four combinations of substrates: α, α; α, β; β, α and β, β" [10]. Even simpler, the additional carbons in the amino acid backbone would result in a more expensive metabolism, because it would logically take more energy to synthesize, manipulate, move, and even degrade the more massive proteins that are built with β-, γ-, and δ-amino acids.

A current understanding of why Earth's biology genetically encodes L-α-amino acids acknowledges that other options, including D-enantiomers, disubstituted amino acids, and longer backbones, were plausibly available to life's origin and evolution. Against this background, explanations for life as we know it have retained a central theme since their inception that there might be some advantage to alpha, monosubstituted amino acids for linear polymers, which fold into complex three-dimensional shapes. What has changed over time has been a retreat from "hard" statements about impossibility of other chemical structures in favor of "softer" statements about preferential attributes of those found in the canonical alphabet. Whereas comparisons between amino acids and other types of molecules (e.g., sulfonic or hydroxy acids) locate these advantages within physics (the strength of the peptide bond), comparisons between different backbones add to this signs of evolutionary influence, and explanatory power for side chains currently favor biological evolution. For example, other backbone types are capable of forming polymers, but would seem likely to cost more energy, both directly and indirectly. Thus, while no direct evidence shows that primordial life used another type of molecular backbone for structural and catalytic polymers (if anything, it would seem more likely that it could have used a mixture of different building blocks), increasingly clear reasons suggest why evolution would favor the streamlining of any such alphabet into the single type of repeating structure encountered within genetically encoded L-α-amino acids: whether or not life on Earth began with L-α-amino acids, it can be seen why it would have evolved to this state. Since the reasoning involved comes from physics and chemistry, it would seem unsurprising to discover a similar outcome within an independent origin of life (a xeno biochemistry), so long as something like natural selection has caused replicating systems to distinguish themselves from the abiotic universe [81]. To expect otherwise, science would need to

identify specific physical conditions under which functional advantages of larger and/or more heterogeneous molecules would outweigh their cost.

3.2. L- vs. D-Stereochemistry

Were a xeno biochemistry to use amino acids, innumerable options exist quite different from life as we know it. In terms of chemical nomenclature, the genetically encoded alphabet is dominated by L-α-amino acids, where "L" and "α" each denote subsets of a far larger set of amino acid structures. A useful next level of inquiry can therefore ask why the subunits for genetically encoded proteins of life on Earth are restricted to (i) L-enantiomers of (ii) α-amino acids.

Would a Xeno-Biochemistry Use L-Amino Acids?

The widespread designation "L-" specifies one of the two mirror-image conformations possible for a single side-chain bound to a single (α) carbon atom situated between the C- and N-termini (Box 1; Figure 3). Because these termini are different, the two positions at which a side chain can bond to the intervening carbon atom form different (non-superimposable) three-dimensional molecules. L- versus D-enantiomers refer to the relative configuration when the α-carbon is viewed from the -N terminus with the -COOH group pointing upwards. Another perspective defines absolute position, *S*- versus *R*-, based on atomic number of constituent atoms [82] but is rarely encountered in amino acid literature because 19 of the 20 genetically encoded amino acids are L-amino acids. The 20th, glycine, is a unique, achiral exception to this pattern. In addition to the single hydrogen present on all 20 (monosubstituted) amino acids, glycine has a second hydrogen atom attached to the α-carbon, whereas all others have a side chain (R-group). From the perspective of the set used to construct genetically encoded proteins, glycine is therefore most usefully perceived as the point-of-origin (zero) for the L-series, or indeed any series: effectively the absence of a side chain.

Without catalysts, such as protein enzymes, undirected chemical syntheses of amino acids generally produce equal amounts of L- and D-enantiomers. Excepting some rare reports of L-enantiomeric excess [28], this racemic mixture is exactly what is observed as the norm in meteorites [28,83] and the results of prebiotic simulation experiments [84,85] (reviewed in [86]). D-amino acids are, furthermore, synthesized and used throughout contemporary biology (reviewed in [87,88]). The venom of the desert grass spider *Agelenopsis aperta*, for example, is known to contain D-serine at position 46 in omega-agatoxin IVB [89] while D-Serine functions as a signaling molecule mediating NMDA receptor activity in mammalian brains [90]. Biological use of D-amino acids does not extend, however, to genetic decoding. D-enantiomers are generally toxic to cells to the degree that many prokaryotes and eukaryotes have evolved detoxification enzymes, which control their concentration within the cell, sometimes by converting D-amino acids into L-equivalents [91,92]. A common cause of this toxicity is that D-amino acids interfere with "normal" (ribosomal) genetic decoding and organisms which incorporate D-amino acids into peptides usually do so through specialized, "*non-ribosomal peptide synthesis.*" [93]. For example, D-alanine and D-glutamate are incorporated into cell wall structure by all bacterial cell walls that are known to contain peptidoglycan [94]. But the ribosome, and thus its tolerance for the molecules into which it translates genetic messages, is a product of biological evolution. We may conclude that D-enantiomers were both available to life's origins and throughout its subsequent evolution: for post-LUCA genetic decoding, they are selected against for similar reasons that would disadvantage a UK motorist attempting to drive on the left side of roads in the USA. But was there any reason why the L-enantiomer was the direction in which this situation would resolve if played out a second time by another instance of biological evolution?

Alluded to above, one possible answer involves "*nonbiological enantiomeric enrichment processes prior to the emergence of life*" [28]. The bias reported from some simulations of interstellar and circumstellar astrochemistry illustrates this [95]. However, the causal mech-

anism(s) for this or any other source of L-enantiomeric bias remains unclear. The literature of suggestions inspired by physics alone includes multiple ideas for direct molecular interactions with light [96–99] and equally diverse explication of the long-recognized [100] role for crystal formation [101]. While both families of explanation could perhaps merge into one, a more general set of ideas relating to symmetry-breaking [102,103], shifting perspective to chemistry, finds an equally diverse set of competing and overlapping suggestions to explain enantioenrichment as a result of reaction pathways instead (e.g., Formose [104]; Strecker [105]). Thus, until further evidence resolves the current lack of consensus (reviewed in [106–108]), it is perhaps more helpful to notice where the vying explanations agree: all tend to involve L-enantiomer *enrichment*, a bias, rather than complete absence of the D-enantiomer. This, in turn, implies that life's homochirality arose through some sort of evolutionary feedback, whether physical, chemical [109] or, shifting to a third perspective, biological (e.g., [11,110]). Viewed in this light, the unknown variable at present is the *extent* of evolutionary feedback versus a foundational bias laid by physics and/or chemistry [111,112]

Recognizing a role for evolutionary feedback also provides reasons to retreat from rigid answers that once seemed clear regarding sub-questions within the topic of "L- versus D-". For example, beneath the question of which amino acid enantiomer is genetically encoded lies the simpler question: why are the genetically encoded amino acids homochiral rather than heterochiral? Cleaves [20] summarizes a longstanding view (e.g., [113,114]) that *"the exclusive use of one isomer allows for the formation of regular secondary structural motifs."* The idea here is that helices and sheets, the foundations of a three-dimensional protein structure, are stabilized through intramolecular interactions that would be obstructed by a haphazard mixture of L- and D-amino acids (Figure 4). However, recent evidence shows that heterochiral proteins, *"though less stable than their homochiral analogues, exhibit structural requirements (folding, substrate binding and active sites) suitable for promoting early metabolism"* [112] (see also [115]). Thus, a better explanation than a biophysical necessity, that life can only originate using homochiral building blocks, is that evolution favored a homochiral set of amino acids for efficiency much as increasing size and speed of traffic caused nations around the world to decide which side of the road travels in which direction.

Using both L- and D-versions of each amino acid could also be metabolically inefficient. At first sight, it doubles the amino acid alphabet size with which an organism decodes genes to work with both enantiomers. The biological cost/benefits of homochirality, however, depend on where one focuses. From the perspective of building protein structure, different enantiomers are indeed effectively different amino acids. From the perspective of biosynthesis, however, the existence of well-characterized enzymes which flip the chirality of the alpha-carbon complicates any cost calculations. Bacterial systems, for example, use racemases to convert L-amino acids into their D-forms (e.g., [116]). Any cost argument for homochirality thus differs from "cost" arguments made about alpha backbones versus longer backbones (e.g., [10]).

In this sense, it seems clearer why replicating systems would be drawn to homochiral amino acids as monomeric building blocks than why it would use the L- versus D-enantiomers. Consensus wisdom has long held that "mirror-life" would function perfectly well [100,117,118] and, at an extreme, evolutionary competition could have led to the eradication of a fully functional 'D-life' [119].
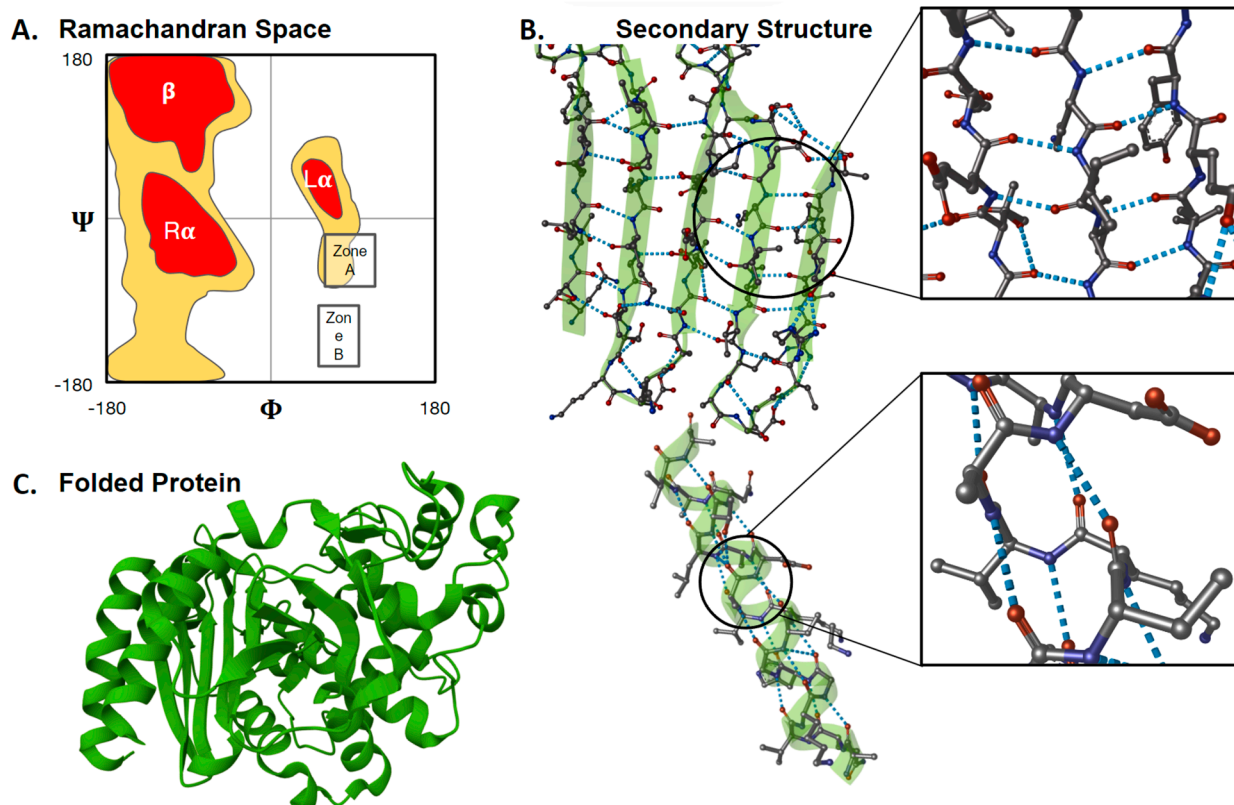
Figure 4. Xeno amino acids potentially change the biochemistry of protein folding (**A**) Combinations of peptide torsion angles (ϕ and Ψ: see Box 1) summarize peptide/protein secondary structure (adapted from [120]). Zones A and B are torsion-angle pairs disallowed by biophysics under strict steric considerations yet observed empirically upon careful investigation for specific combinations of side chains (discussed in [121]). This provides one clue that the current, well-established map of secondary structure could shift or become unrecognizable if rebuilt for xeno amino acids. (**B**) The dominant secondary structures of life on Earth, α-helices and β-sheets (first described by [122]), form and are stabilized by hydrogen bonds (blue dotted lines). These hydrogen bonds usually involve atoms in the amino acid backbone. The bonds and, therefore, protein structures would be obstructed or altered if peptides comprised a heterogeneous enantiomeric mixture, more than one sidechain per amino acid or longer backbones (i.e., β-, γ-, δ-) (**C**) Multiple secondary structures within a single polymerized amino acid sequence combine to form the larger, folded tertiary structure. It follows from (**A**–**C**) that protein structures could be unpredictably different if the fundamental building blocks (amino acids) were changed. Images were created using Mol* Viewer [123] with PDB 4LV0 AmpC beta-lactamase in complex with m-aminophenyl boronic acid [124].

3.3. Monosubstitution

A continuing focus on evolutionary feedback rather than biophysical necessity addresses a subtly different sub-question implied by L-homochirality: why do all genetically encoded amino acids use only one of the two possible side chain attachment points presented by the α-carbon (i.e., why are they monosubstituted)? Like a mixture of L- and D-chiralities, the presence of two side chains on the α-carbon (α,α-disubstituted) has long been recognized to obstruct secondary structure formation (e.g., [125] and Figure 4B). However, while "*the secondary structure of [disubstituted amino acid] peptides [are] especially restricted,*" [126], subsequent evidence again shows that structure formation is possible [127]. A more robust explanation for a monosubstituted alphabet is what McKay [128] calls the Lego Principle: "*Biological processes, in contrast to abiotic mechanisms, do not make use of the range of possible organic molecules. Instead, biology is built from a selected set...General arguments*

of thermodynamic efficiency. . .suggest that this selectivity is required for biological function and is a general result of natural selection."

4. Would a Xeno Biochemistry Use Different Side-Chains?

The third and final amino acid attribute that deserves careful consideration for alternative biochemistries is the set of 20 side chains used within the standard genetic code. Both backbones and side chains contribute to producing protein structure, but differently so. Whereas it is the unvarying features of the L-α-backbone that matter (notwithstanding the special cases of glycine and proline above), it is the differences between side chains that are important. In 1972, the Nobel Prize in Chemistry was awarded for finding that "*at least for a small globular protein in its standard physiological environment, the native structure is determined only by the protein's amino acid sequence*" [129]. This finding ended some fundamentally misled ideas about protein folding (e.g., the cyclol hypothesis [130]; see [131]) with the knowledge that the set of 20 genetically encoded side chains define what protein shapes and functions can be genetically encoded by life [132,133]. Thus, while the "Lego" principle accounts neatly for an unvarying, L-α-backbone (Figure 4), any explanation of side chains must introduce new ideas to explain diversity.
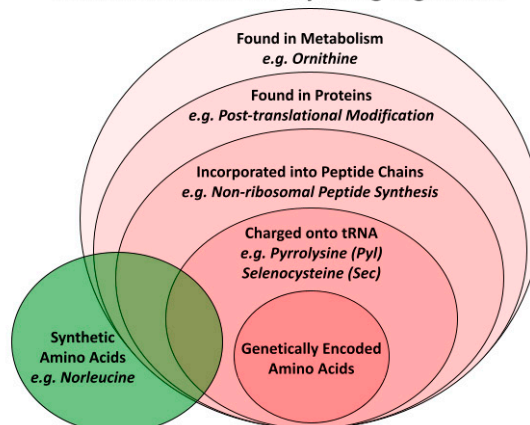
Like the discussion of backbones presented above, plenty of plausible alternatives exist to the set of 20 side chains genetically encoded by life as we know it. Once again, these options are informed by prebiotic chemistry (both simulations and meteorite analysis; Figure 5A), and by their widespread use within biology (Figure 5B). Indeed, early glimpses of diversity [134,135] reflected limitations of instrumentation more than chemical reality for both abiotic [20,33] and biological amino acids [136,137]. Complementing these naturally occurring alternatives is plentiful experimental evidence that other side chains can still function within the genetic code, even after 3.5 billion years of evolution. In recent years, synthetic biologists have engineered more than 250 different amino acid side chains into protein synthesis [138]. Indeed, the subfield of non-coded amino acids (ncAA's) is developing so fast that the total of 250 is out of date and any alternative suggested here would be obsolete within months (e.g., [139,140]). Such technological progress aligns well with the widespread use of xeno amino acids in specialized versions of peptide synthesis [141,142] to suggest the imminent delivery of human-engineered alternative amino acid alphabets [143].

Given clear evidence for a multiplicity of alternatives, it is useful to remember that side chain diversity directly defines the corresponding universe of shapes and functions. Indeed, a major challenge for current research is the theoretically infinite diversity of side chains made possible by organic chemistry. Although imposing a maximum side chain size (e.g., by volume, number of atoms, etc.) constrains the set into a finite number, any such number is problematically large (Figure 5D). There are, for example, approximately 5.6×10^5 isomers of the side chain for Tryptophan [41], the largest of the coded amino acids by volume, before adding the cumulative side chains smaller than this, and/or those encompassed by slightly different atomic composition. Since synthetic biology has already successfully incorporated L-α-amino acids far larger than Tryptophan, and far more chemically diverse than anything seen in the genetic code, into "natural" (ribosomal) genetic decoding (Figure 5B), perhaps the single clearest idea for xeno side chains at present is that those used by post-LUCA life on Earth are not the only set of chemical structures capable of linking into functional biopolymers. The more interesting question is: what shaped this particular evolutionary outcome?
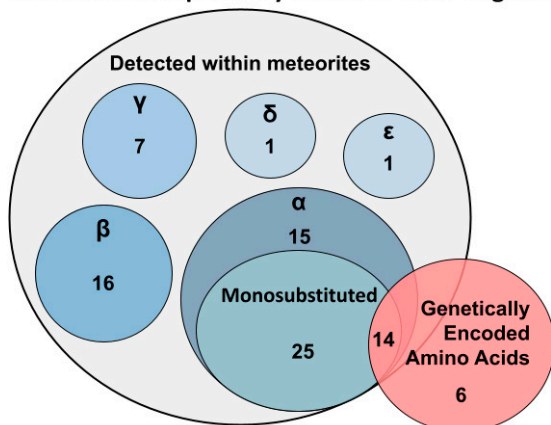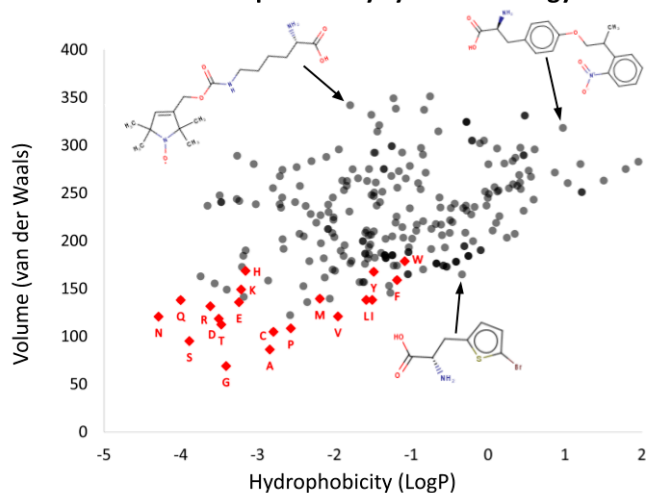
Figure 5. Like amino acid backbones, many side chains are possible beyond the 20 found within the standard genetic code. (**A**) Abiotic synthesis: various other side chains are found in meteorites and produced by prebiotic simulations (see [144]: supplementary data); (**B**) biological: many other side chains are used by living organisms (see [145]); (**C**) synthetic biology has successfully incorporated hundreds of alternative side chains into protein synthesis, including numerous far larger than anything found within the standard genetic code, shown as red diamonds [146–148] (also see [138]). (**D**) Theoretical: the addition of each carbon atom increases exponentially (combinatorially) the number of chemical structures that are possible (see: [41]).

Given amino acid side chains' importance in defining protein structure and the clear potential for alternatives, surprisingly little research has addressed the consequences of building proteins with other side chains. The initial success of prebiotic simulations and their alignment with meteorite analyses certainly inspired a small, early sub-literature considering amino acids from beyond the genetic code, usually in the form of a "deep dive" into one particular amino acid (e.g., Norleucine [149]; Ornithine [150], but see also [10] for a review). However, in the later years of the 20th century, focus narrowed to how the 20 amino acid "meanings" became incorporated into the standard genetic code rather than looking beyond. Certainly, to explore amino acids from beyond the genetic code is costly in terms of both time and money, but considerable time and money were spent investigating the 20 (see AAIndex [151]), so why did this activity not look beyond the molecules of the central dogma? No single reason clearly explains why but, with hindsight, several contributing factors may be inferred.

From the perspective of biology, the discovery that side chains steer protein folding emerged within a larger framework, represented by four other Nobel prizes [152–155]. Together, the work awarded by these prizes describes how all life on Earth converts genetic information into protein-based metabolism. Within this "central dogma of molecular biology" [156], amino acids build biological proteins because they are programmed to do so by genetic information. A considerable literature thus developed to discuss how 20 amino acids became assigned to 64 different genetic code words (codons) (see, for example, [135,157–160]. Indeed, the genetic code remains central to biology such that in 2022 alone PubMed reported 64 new publications using the keyword terms *genetic code* and *evolution*.

From the perspective of prebiotic chemistry, the syntheses that accounted so easily for some of the amino acids found within the standard genetic code gave way to unexpected difficulties in accounting for the rest [135]. Synthetic, organic chemists researching life's origins thus diverted efforts towards accounting for missing members of the 20 rather than exploring side chains that lie beyond.

From the perspective of protein structural biochemistry, the satisfyingly simple insight that side chains steer protein folding proved frustratingly difficult to model or predict with detailed physicochemistry. Levinthal [161] famously captured the essence of the problem by pointing out the overwhelming number of possible conformations into which polymers built from an alphabet of 20 different side chains could potentially fold. Simple, pragmatic urgency of making progress in solving the "protein folding problem" [162] replaced asking equivalent questions about other possible side chains, and another relevant research community focused on the 20 rather than looking beyond.

Finally, from the perspective of "origins" research, a sixth Nobel Prize was awarded for the discovery of catalytic RNA [163]. This extension of the Central Dogma led directly to the declaration of the RNA world hypothesis [164–166], which was taken by many to imply that amino acids entered an evolved, RNA-based biology (e.g., [167–170]). Under such thinking, the set of amino acids found within the genetic code is one that can be synthesized by metabolism rather than made available by prebiotic chemistry. The resulting shift in perspective is seen by comparing two influential review articles, separated by three decades. Whereas Weber and Miller [10] used expertise in prebiotic chemistry to discuss which amino acid side chains would have been available to life's origin, by 2017, Doig explained that "*If protein synthesis arose from the RNA World… life was already biochemically sophisticated and the environment was substantially modified from the conditions prevailing during abiogenesis. Arguments based on prebiotic conditions are thus not especially helpful in rationalizing amino acid selection.*" [171] While interpretations of the RNA-world hypothesis continue to diversify [39], the idea that the 20 genetically encoded amino acids reflect the evolutionary expansion of simpler, earlier code continues to gain multidisciplinary consensus (Figure 6).
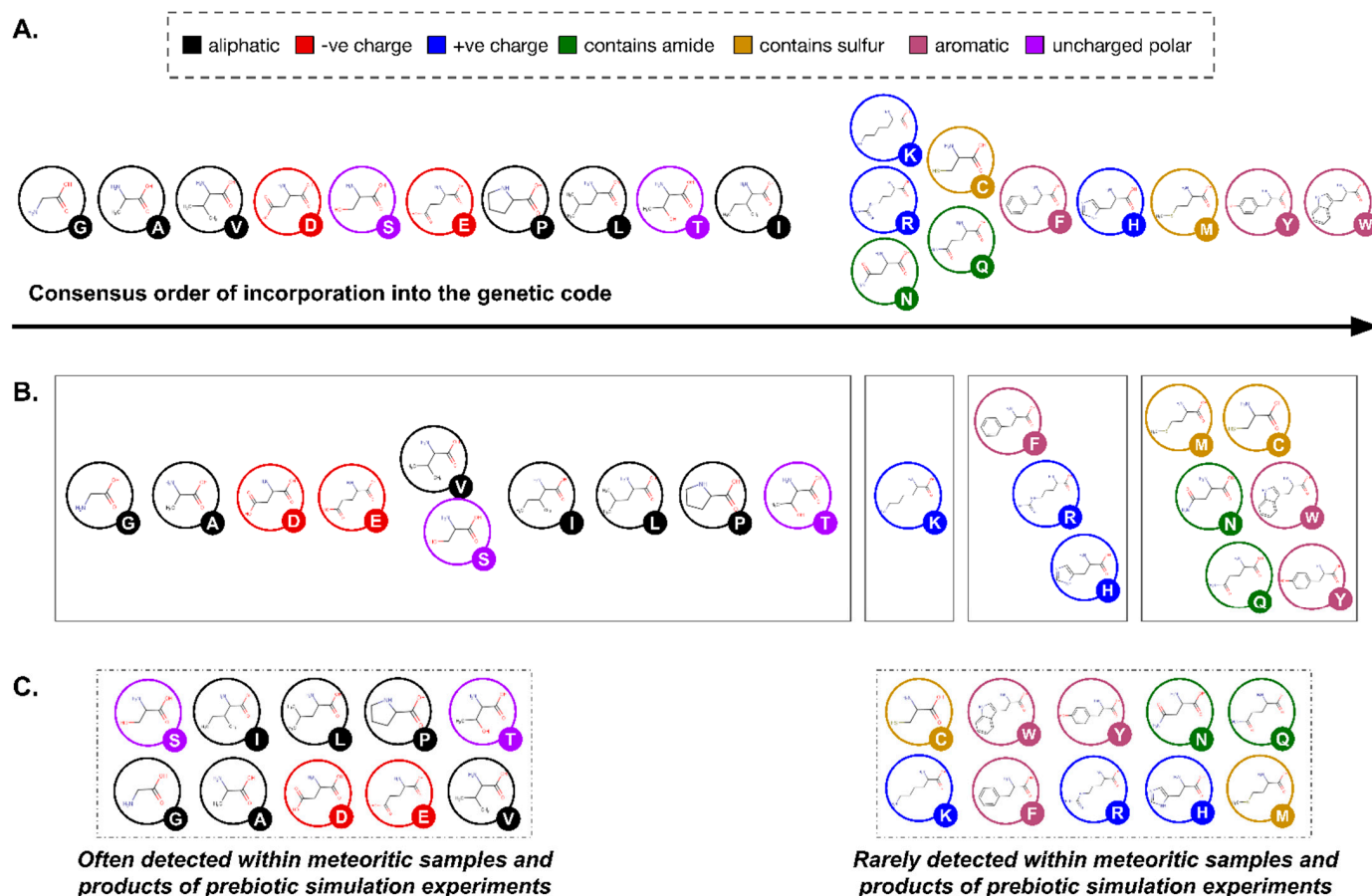
Figure 6. A comparison of three major syntheses of scientific literature concerning the antiquity of amino acids within the standard genetic code. All agree that the canonical alphabet of 20 amino acids evolved from an earlier genetic code involving fewer amino acids. (**A**) Trifonov [172] analyzed 40 peer-reviewed publications about the evolution of the genetic code to calculate a detailed chronology by which the set of 20 became established. (**B**) Higgs and Pudritz [173] analyzed a similar amount of different literature to arrive at broadly similar conclusions. (**C**) Cleaves [20] focused on the literature of prebiotic chemistry alone (meteorites, spark tube experiments, and HCN polymerization) to agree with both. Adapted from [144].

Jumping ahead to the 21st century, the past decade has witnessed a resurgence of interest in looking beyond the genetically encoded alphabet of 20 side chains. At present, around 25 peer-reviewed publications contribute directly to this literature (Figure 7), and can be understood as deriving from three distinct research communities, each now equally relevant to xeno biochemistry: **De Novo Protein Design, Prebiotic Chemistry**, and **Molecular Evolutionary Biology**. In general terms, these different communities are worth distinguishing because, prior to a shared mutual interest in xeno amino acid side chains, their approaches connect only by going back further to foundational authors who wrote with great influence around the discovery of the central dogma of molecular biology. It is then the emerging, new synthesis of these three subfields which promises exciting new progress.
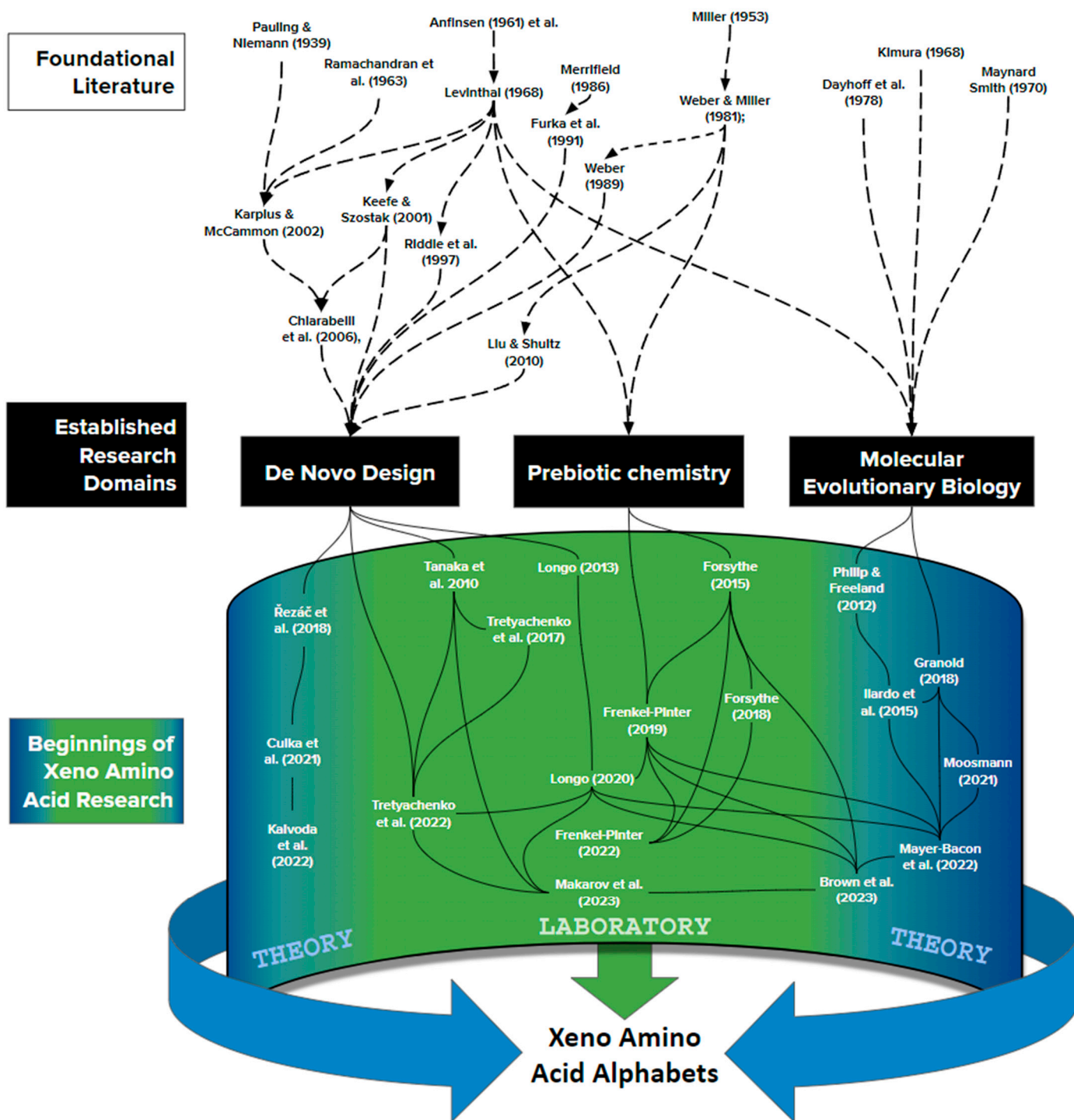
Figure 7. A literature map of foundational and recent (>2010) publications [10,18,52,60,78,131,132, 144,147,161,174–198] converging on direct exploration of alternative amino acid alphabets, traced back through diverse subdisciplines to foundational works. This literature involves both theory and experiment. Whereas experimental work is already starting to integrate the 3 named subfields (de novo protein design, prebiotic chemistry, and molecular evolution), relevant theory is at present siloed between two largely unconnected edges: sophisticated subatomic modeling of alternative side chains and biologically inspired design of xeno alphabets.

4.1. Clues from De Novo Protein Design: Altering the Functional Units of Life as We Know It

De novo protein design builds from Anfinsen's [174] demonstration that a protein's three-dimensional structure is produced by the specific sequence of amino acid side chains.

Woolfson [199] characterizes three approaches that inform both the protein folding problem and provide a powerful foundation for adapting to xeno alphabet thinking.

(i) **Amino acid alphabet simplification:** It has long been speculated that an alphabet comprising fewer than 20 amino acids could build functional proteins (e.g., [200]), and research working with reduced amino acid alphabets, or Minimal Protein Design, "*uses straightforward chemical principles such as patterning of polar (p) and hydrophobic (h) amino-acid residues to direct the folding and assembly of secondary structures…*" [199]. Riddle [195] first demonstrated empirically that a random sequence of amino acids drawn from a reduced subset of that canonical twenty can exhibit structure and function. Tanaka [196] then scaled up this observation by comparing three peptide libraries constructed from random sequences using alphabets of different lengths. Within this methodological framework, Longo et al. [187] then built a "*foldable halophilic protein*" from an alphabet primarily reduced to those which are prebiotically plausible. From here, it was a tractable and clear step to introduce xeno side chains [186]. Tretyachenko et al. [197] further advanced this approach by introducing high-throughput sequencing. Interestingly, as the composition of prebiotically plausible amino acids increased to 100%, new folding principles started to emerge [201].

Most recently of all, Makarov et al. [188] has started to introduce xeno amino acid side chains within a reduced alphabet framework in order to compare canonical versus non-canonical side chains. Presently, it remains to be seen just how many and how deep the new protein-folding principles that come from building with xeno side chains are.

(ii) **Rational peptide design** constructs peptides that sample a targeted region of protein sequence space using "*sequence-to-structure relationships garnered from biochemical, bioinformatics or empirical studies*" [199]. The invention of solid-state synthesis [190] permitted researchers for the first time to synthesize protein sequences efficiently without involving life's molecular machinery for genetic decoding: a freedom powerful enough to earn yet another Nobel prize in chemistry [202]. Furka's subsequent addition [179] of a "mix and split strategy" added the power of combinatorial chemistry to this approach (for a recent review see [203]). The underlying and significant advantage here for studying xeno amino acids is a lack of dependence on life as we know it. This potential is, however, only now being realized (e.g., [188]), and is currently limited to small oligopeptides. Indeed, no one has yet worked with an entirely xeno amino acid alphabet. The challenge is as much mathematical as biochemical: increasing the length of a peptide by each amino acid increases exponentially the possible sequence space. A peptide of length 100, for example, drawn from an alphabet of 20 amino acids can be any of ~10^{130} possible sequences, enormously more than the number of atoms estimated to comprise the entire universe.

(iii) **Computational modeling** uses biophysics to understand protein folding in silico by generating and evaluating "*full atomistic models for many different sequences for a given design target … ahead of experimental studies*" [199]. The "protein folding problem" was born when it was noticed that physics somehow sorts through pragmatically infinite conformational possibilities to produce Anfinsen's lauded outcomes ("Levinthal's Paradox" [161]). Progress in understanding how biophysics does so came in 1994 when the Critical Assessment of Protein Structure Prediction (CASP) coalesced diverse approaches into an annual competition. By 2005, the overview of results was that "*current major challenges are refining comparative models [as they fast approach] experimental accuracy.*" In other words, all the best protein fold prediction algorithms share one idea in common: begin by finding a protein structure already known to science that is similar to the one under current scrutiny. In contrast, predictions built using first principles of physics and chemistry were relegated to "*handle parts of comparative models not available from a template*" [204]. Hope for merging these two approaches came from Rosetta [205], which showed that accurate protein structure (~300 amino acids) can be predicted from concatenated peptide fragments five amino acids long (5-mers). It was a return to homology modeling, however, when DeepMind's AlphaFold 2 [206] effectively solved the protein folding problem for most natural proteins, but only by taking the basis for its predictions into a black box.

To know how xeno alphabets will fold, it is logically necessary to take the advice from CASP and focus on what we understand about the physics of "standard" protein folding, however challenging. Here, molecular dynamics (reviewed in [183]) paved the way for thinking about biomolecules in terms of "conformational dynamics" [207], which means comparing the Gibbs free energy of different possible three-dimensional conformations as a guide to understanding the stable one which biological polymers find. This "ab initio" approach still cannot predict protein structure with anything like the power and accuracy of AlphaFold 2, but recent advances in quantum mechanical models (as seen in [176,182]) are fast building the framework for future prediction of oligopeptides incorporating, or entirely comprising of xeno amino acids.

To Wolfson's three subfields, a careful discussion of xeno amino acids may usefully add a fourth: **Alternative genetic codes.** A central point of the central dogma is that life decodes polymerized sequences of nucleotides (genes) into polymerized sequences of amino acids (proteins) (Box 1), and yet it has been traditional to study one type of biopolymer or the other. RNA and protein research have populated the pages of different journals, the authors of these two studies have gathered at different conferences, developed different specialized terminologies, and generally checked many boxes for being considered as different academic disciplines [208]. Wolfson, a distinguished protein researcher writing for the protein community, focuses on amino acid alphabet simplification, rational peptide design, and computational protein modeling. Alternative genetic codes add something different simply by exploring protein structure and functions from the perspective of genetics, where a host of tools were developed in the wake of the central dogma to study and manipulate genetic material. For example, when Keefe and Szostak [184] first explored the frequency at which folds and functions occur within protein sequence space using (mRNA) phage display, their artificial selection of RNA sequences under controlled mutation rates deliberately emulated the power of natural selection to find new, functional proteins. Twenty years later, this approach has matured to bring into view, among much else, user-defined ("programmable") genetic codes [209,210]. This potential meets impressive progress by synthetic biologists, who have engineered more than 250 different amino acid side chains into protein synthesis [138]. At least some of this effort is with an eye toward systems in which "*. . . a [semi-synthetic organism] is now, for the first time, able to efficiently produce proteins containing multiple, proximal ncAAs*" [143].

4.2. Clues from Prebiotic Chemistry: Bridging the Gap between Life and the Non-Living Universe

From the perspective of prebiotic chemistry, the two parallel and intertwined perspectives with which we introduced this review remain directly relevant to understanding xeno amino acids: chemical simulations in the laboratory [211] and direct analysis of prebiotic environments [33].

Instead of working forward from prebiotic chemistry, others have tried to work backward from the post-LUCA genetic code. Jukes [150] was among the first to provide a specific candidate sidechain (ornithine) as a possible forerunner to the genetically encoded amino acid arginine. Additional contributions over the next couple of decades were relatively sparse and led a thorough review to conclude their probable irrelevance for reasons of biophysics: ornithine peptides, for example, "*are unstable because internal lactamization*" [10], where lactamization refers to a carbon-bound, linear, side chain bending around to also bond with the backbone amine, producing a cyclical structure in the subclass of amino acids to which proline belongs (Figure 3B). In this sense, it was the growth of the RNA world hypothesis that clarified "*Arguments based on prebiotic conditions [alone] are thus not especially helpful in rationalizing amino acid selection.*" [171]. In other words, once evolution by natural selection is at work on the alphabet, it is fully capable of introducing amino acids that are prebiotically implausible. Moving into the 21st century, growing acceptance of the idea that terrestrial life genetically encodes up to 22 amino acids [212] provides another kind of empirical evidence that the genetically encoded alphabet can and does evolve. Wong and Bronskill [135] first overtly introduced the idea that an explanation for

the standard alphabet would have to include both prebiotic chemistry and subsequent biological evolution. And yet, until the 21st century, the 20 still dominated all knowledge about amino acids in proteins.

Recent work has begun to explore xeno side chains in comparison with canonical amino acids. Fenkle-Pinter et al. [178], for example, have concluded that canonical side chains seem predisposed to form polymers more readily than non-canonical alternatives on biophysical grounds. By comparing the readiness with which lysine, arginine, and histidine form peptide bonds with each other, as opposed to with analogs from beyond the standard alphabet (ornithine, 2,4-diaminobutyric acid, and 2,3-diaminopropionic acid), this work concludes that "*the proteinaceous amino acids exhibit more selective oligomerization [suggesting] a chemical basis for the selection of Lys, Arg, and His over other cationic amino acids.*" While the results and indeed the question are pioneering important new information, it is for now puzzling that the particular amino acids studied are ones which other disciplines, from meteoritics to molecular evolution, agree entered the code only after enzyme-based metabolism had removed any semblance of proteins forming by competing to oligomerize (Figure 6).

4.3. Clues from Molecular Evolutionary Biology: Natural Selection Guiding Alphabet Design

From the perspective of evolutionary biology, the discovery of life's central dogma defined for the first time specific, universal parameters with which life evolves at a molecular level. This replaced generations of mathematical modeling built creatively and cleverly on the limited knowledge that genes are particulate (non-blending), occur on chromosomes, and mutations change them: foundational rules that were famously criticized as "*bean bag [population] genetics*" (see [213] for a review).

In the aftermath of the central dogma, three pioneers of evolutionary theory developed the new potential for molecular detail from different directions. Motoo Kimura [185] used the new molecular knowledge to translate older, powerful population genetics that emphasized the role of chance (genetic drift) relative to natural selection. He noticed quickly that the central dogma implies significant "selectively neutral" evolution through, for example, redundancy in the genetic code [214]. In contrast, Maynard Smith [189] pioneered how to think clearly about natural selection at the molecular level, picturing adaptive walks through sequence space. Margaret Dayhoff [177] complemented both approaches by using computing to summarize and then analyze empirical patterns of molecular evolution. She extracted quantitative statements about the patterns by which amino acids substituted for one another over time, largely corroborating Kimura's thinking as the strongest signal within molecular evolution.

The combined insights of these individuals and the work they subsequently inspired combined to move molecular evolution into a much more central role within biological and biomedical research. For example, bioinformatics is built around evolutionary ideas such as homology and phylogeny. It turns out these ideas are important for research that ranges from predicting protein folds to finding and understanding the role of protein-coding genes. However, here, just as we noted more generally in the introduction for all of biology, the resulting focus on one canonical alphabet of 20 possible amino acids has produced a contemporary science that is surprisingly blind to how molecular evolution would change if a different set of side chains were involved. It is, for example, easy to imagine why xeno amino acids could bring new physicochemistry of protein folding if we focus on the "standard" amino acid cysteine. A specific and unique characteristic of cysteine, one which contributed directly to Anfinsen's Nobel prize-winning work, is that two instances of cysteine at very different places within a single protein sequence can form disulfide bridges with one another as the protein folds into a three-dimensional shape. If the genetic code lacked cysteine then nothing like disulfide bridges would exist among the other 19 to inform us (or a machine learning algorithm) of their possible existence and role in protein folding. Less extreme but more widespread than new covalent bonds, "*side-chain and backbone interactions [within 'natural' protein sequences] may provide the energetic*

compensation necessary for populating [hitherto unrecognized] region of φ–ψ space" [121]. If sidechain physicochemistry of the 20 can still expand understanding of sequence/structure relationships, then it would seem unwise to expect an indefinitely large and diverse set of xeno side chains not to alter these relationships further.

Almost everything we know about each member of the canonical alphabet is relative to the other 19. The challenge is knowing where to focus within a vast set of possible side chains (see [41]) and an equally vast set of possible biophysical properties (e.g., [215]).

One sustained attempt to circumvent this limitation draws inspiration from the evolutionary methodology of optimality theory [216]. When we wish to understand an aspect of the living world, we may ask what about it is unusual and plausibly the result of natural selection? Doing so quantitatively, in the context of plausible alternatives, begins a framework for scientific exploration of evolutionary cause(s). The idea is not that the initial hypothesis is correct, but rather it is a way for the researcher to enter an iterative cycle of comparing predictions against observed reality so as to inform a new, better prediction. Retesting, with iteration ad infinitum, inevitably leads to an improved understanding of evolutionary causes, both the specific selection pressures involved and the unexplained role attributable to random genetic drift [216]. Traditionally, this approach has been used with organismal phenotypes, especially behavior, such as the time dung flies spend mating [32], or what size mussel a shore crab chooses to crack open for food [217]. The optimality approach has been adapted, however, to molecular fundamentals: first to the size and content of the genetic alphabet [218,219], then the distribution of amino acid "meanings" within the genetic code itself [220,221] and now to the amino acids as one possible set among many [192,222].

A primary challenge for amino acids is to define and quantify features of their chemical structures upon which natural selection could plausibly have acted, not only for the canonical twenty but also for xeno alternatives. Careful biophysical measurements of amino acids from beyond the standard alphabet form an excellent example of where current science offers little data. It is not that it is difficult to identify biophysical aspects of protein folding consistent with what we observe in nature. However, correlation does not imply causation, and there is currently little evidence with which to test this understanding other than a single, pioneering database of short, human-engineered peptides which each contain ~one xeno or non-coded amino acid (ModPep: see [223]). By the 21st century, however, computational chemistry was creating algorithms that could predict fundamental biophysical properties of molecular structures, and these were shown to be fully capable of estimating accurately the properties of molecules the size and complexity of L-α amino acids [224]. From here, investigations of the ways in which the genetically encoded amino acids distinguish themselves from xeno alternatives needed only the addition of one more, elegantly simple idea from the 21st century: chemistry space [225,226]. Chemistry space thinking is that any carefully defined, measurable biophysical property of a chemical structure may be thought of as a coordinate, such that measuring several properties of one molecule begins to define its chemical structure as a point within a multidimensional space. Equivalent measurements for other molecules define a cloud of points, wherein proximity means similarity, and distance means dissimilarity. Any statistical and/or geometrical concepts to compare points can be used to test quantitative hypotheses. The application of such thinking to organic chemical structures quickly revolutionized pharmacology, particularly the drug discovery industry [227].

By 2011 analysis of the chemistry space occupied by amino acids detected a highly unusual distribution by comparing the genetically encoded 20 with xeno alternatives. The two specific physicochemical properties involved (volume and hydrophobicity) are known to guide protein folding [228,229]. Subsequent work has expanded this evidence in both depth [144,222] and breadth [230], and has even detected strong, non-random patterns in additional amino acid properties [180,191]. The most recent expansion of this work has now identified for the first time specific examples of entirely xeno amino acids that match the statistical profile established by life since LUCA [175]. An exciting next step will be

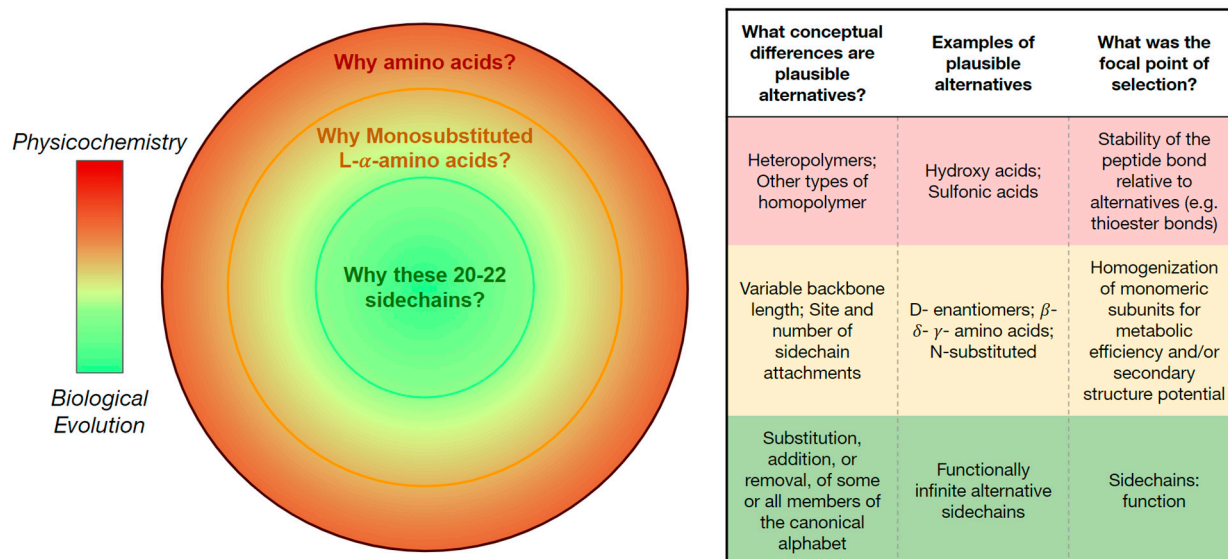to empirically test whether such alphabets exhibit some sort of identifiable advantage for protein folding.

5. Discussion

This review synthesizes current knowledge regarding three, overlapping questions: (1) Would xeno biochemistry use amino acids? (2) Would it use monosubstituted L-*α*-amino acids? and (3) Would it use different side chains? Below we summarize answers to each, along with examples of tractable near-future milestones of particular relevance to astrobiology.

5.1. Would Xeno Biochemistry Use Amino Acids?

One set of 20 amino acids has allowed life on Earth to inhabit an impressive diversity of environments. Indeed, conditions now recognized as supporting life on Earth overlap considerably with those identified for other planetary bodies in the solar system [12,231]. On another front, simulations and meteorite analyses agree that amino acids form readily under a wide range of abiotic and prebiotic conditions. Other organic polymers (e.g., hydroxy acids) that also form readily, polymerize with bonds that are less stable to hydrolysis than the peptide bonds which link amino acids. Depsipeptides thus spontaneously self-purify towards greater amino acid enrichment under wet–dry cycling. In other words, organic chemistry offers good reasons why any life might be expected to "encounter" amino acids (Figure 8: outer shell), chemistry provides good reasoning why it might use them and biology shows us how resilient and versatile would be the result.

Current evidence suggests variable influences shape the canonical amino acid alphabet

What conceptual differences are plausible alternatives?	Examples of plausible alternatives	What was the focal point of selection?
Heteropolymers; Other types of homopolymer	Hydroxy acids; Sulfonic acids	Stability of the peptide bond relative to alternatives (e.g. thioester bonds)
Variable backbone length; Site and number of sidechain attachments	D- enantiomers; *β*-*δ*- *γ*- amino acids; N-substituted	Homogenization of monomeric subunits for metabolic efficiency and/or secondary structure potential
Substitution, addition, or removal, of some or all members of the canonical alphabet	Functionally infinite alternative sidechains	Sidechains: function

Figure 8. Why does life on Earth use one precise set of 20 L-*α*-monosubstituted-amino acids? As the focus of this question narrows from amino acids as a class of chemicals to the 20 specific side chains used by post-LUCA life, the probable role of physicochemical constraint diminishes relative to that of biological evolution.

5.2. Would a Xeno Biochemistry Use Monosubstituted L-*α*-Amino Acids?

Without catalysis, undirected chemical syntheses of amino acids generally produce equal amounts of L- and D-enantiomers. Even in the rare cases where L-enantiomeric excess has been detected, it is bias rather than absence of the D-enantiomer. At present, the most plausible inference is that genetically encoded homochirality arose through at least some evolutionary feedback, whether chemical or biological (see [232] for an exploration

of the difference). Such thinking introduces a new perspective beyond physicochemical arguments for using amino acids at all. Of course, even biological evolution could imply natural selection or genetic drift. So far evidence is stronger for natural selection and the current debate about L-chiral and α-amino acids is therefore better characterized as negotiating the relative power of biophysical constraint versus natural selection (Figure 8, middle shell).

Candidates to drive selection are not hard to identify, at least in some ways. Using only one stereoisomer permits the foundation of all protein structure as we know it: protein secondary structure is stabilized through intramolecular interactions that would be obstructed for a polymer comprising heterogeneous stereoisomers (Figure 4). It is entirely reasonable, then, that biological evolution selected a homochiral set of amino acids for efficiency. However, multiple overlapping interpretations of efficiency are easy to think of, and it remains to be determined which form the best guide to understanding how life on Earth has turned out. Similar thinking addresses the number of side chains per amino acid. Peptides comprising disubstituted amino acids can produce secondary structure, but they would inevitably require more energy for synthesis, transport, manipulation, and degradation, due to their greater mass. This is the logic of the "Lego Principle": "*General arguments of thermodynamic efficiency . . . suggest that selectivity [to a reduced set of molecular building blocks] is required for biological function and is a general result of natural selection*" [128].

From here, current understanding of a third feature for amino acids proceeds easily. The genetically encoded set of amino acids are all α-amino acids. Again, prebiotic simulations and meteorite analyses clearly indicate that amino acids with longer backbones (β-, γ-, and δ-amino acids, etc.; Figure 3) were available throughout life's origin, and the widespread use of these longer-backbone amino acids within current biology demonstrate their continued availability throughout evolution. It has long been recognized that each carbon–carbon bond within these longer amino acid backbones presents a new site of possible rotation and that the resulting increase in flexibility for peptides, reducing structural stability. Fifty years ago, it seemed clear that this biophysical constraint accounted for evolution's "choice" of alpha amino acids. Once again, subsequent research has demonstrated that longer backbones and/or even a more heterogeneous diversity of backbone lengths can produce viable and even biomedically relevant protein structures (reviewed in [233]). It is at least as compelling then to suggest that evolution, and specifically natural selection, would have favored the lower energy budget of working with the smallest, least massive backbone as a universal feature of its monomeric building blocks.

5.3. Would a Xeno Biochemistry Use Different Side Chains?

As was true for alternative backbones, we now know that many other side chains can function within the genetic code, even after 3.5 billion years of evolution [234]. De Novo Protein Design is teaching us how to create and analyze peptides and proteins that incorporate xeno amino acids, providing tools and techniques with which to move into this uncharted territory [188]. The main contribution of prebiotic chemistry, in the process of connecting the abiotic universe to life, has been providing empirical evidence that alpha amino acid backbones cost less to make and use [20]. Molecular Evolutionary Biology has started to develop specific predictions for alternative amino acid alphabets by identifying and emulating quantifiable, biophysical properties of the encoded 20 [138].

In summary: Current knowledge about alternatives to the L-α-amino acids used by life indicates that as the focus narrows from amino acids as a chemical class to the 20 specific side chains genetically encoded by post-LUCA biology, the influence of biophysics diminishes relative to that of biological evolution (Figure 8). That being said, much remains unknown. Whereas Weber and Miller [10] concluded that "*we would expect that the catalysts would be poly-alpha-amino acids and that about 75% of the amino acids would be the same as on the earth*", in 2023, we suggest instead that emerging ideas, technologies, and datasets are positioned to make such estimate possible within the next decade.

5.4. What Tractable Questions Would Represent Progress for Xeno Amino Acid Science?

From this conclusion, we identify 3 overlapping near-term goals that will expand our understanding of biochemistry as we do not know it.

Regarding alternatives to amino acids as monomeric building blocks, the current frontier is characterizing their potential to polymerize. Looking ahead, it is rapidly becoming both tractable and important to expand current work (e.g., [235]) that characterizes structures and functions of individual examples into general, systematic statements about how such xeno polymers differ from proteins.

Similarly, the design and synthesis of amino acid polymers with longer backbones (β-, γ-, δ-, etc.), mixed chirality, or multiple side chains are increasingly well understood to form structures and functions (e.g., [78]). From here, it will be exciting to see systematic characterization that quantifies the structural and functional range of such molecules relative to L-α-amino acid polymers.

Finally, for side chains, momentum is growing for biomedical research that engineers noncanonical amino acids into otherwise natural proteins [234], and for work that mixes canonical and xeno side chains from an origins perspective (e.g., [178,188]). From here, a natural new milestone will be to see these approaches meld into rational design of entirely xeno proteins and even xeno alphabets. For this to happen, one of several innovations needed is for two branches of theory to connect: sophisticated biophysical calculation of structure applied to xeno alphabets designed to emulate biology's canonical alphabet. Both approaches seem likely to learn and grow from each other.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Böck, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. Selenocysteine: The 21st Amino Acid. *Mol. Microbiol.* **1991**, *5*, 515–520. [CrossRef] [PubMed]
2. Rother, M.; Krzycki, J.A. Selenocysteine, Pyrrolysine, and the Unique Energy Metabolism of Methanogenic Archaea. *Archaea* **2010**, *2010*, e453642. [CrossRef] [PubMed]
3. Kivenson, V.; Paul, B.G.; Valentine, D.L. An Ecological Basis for Dual Genetic Code Expansion in Marine Deltaproteobacteria. *Front. Microbiol.* **2021**, *12*, 680620. [CrossRef] [PubMed]
4. Sun, J.; Evans, P.N.; Gagen, E.J.; Woodcroft, B.J.; Hedlund, B.P.; Woyke, T.; Hugenholtz, P.; Rinke, C. Recoding of Stop Codons Expands the Metabolic Potential of Two Novel Asgardarchaeota Lineages. *ISME Commun.* **2021**, *1*, 1–14. [CrossRef] [PubMed]
5. Brugère, J.-F.; Atkins, J.F.; O'Toole, P.W.; Borrel, G. Pyrrolysine in Archaea: A 22nd Amino Acid Encoded through a Genetic Code Expansion. *Emerg. Top. Life Sci.* **2018**, *2*, 607–618. [CrossRef] [PubMed]
6. Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Söll, D. Pyrrolysine Is Not Hardwired for Cotranslational Insertion at UAG Codons. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 3141–3146. [CrossRef]
7. Budisa, N.; Kubyshkin, V.; Schmidt, M. Xenobiology: A Journey towards Parallel Life Forms. *ChemBioChem* **2020**, *21*, 2228–2231. [CrossRef]
8. Young, T.S.; Schultz, P.G. Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J. Biol. Chem.* **2010**, *285*, 11039–11044. [CrossRef]

9. Opuu, V.; Simonson, T. Enzyme redesign and genetic code expansion. *Protein Eng. Des. Sel.* 2023; *ahead of print*. [CrossRef]
10. Weber, A.L.; Miller, S.L. Reasons for the Occurrence of the Twenty Coded Protein Amino Acids. *J. Mol. Evol.* **1981**, *17*, 273–284. [CrossRef]
11. Preiner, M.; Asche, S.; Becker, S.; Betts, H.C.; Boniface, A.; Camprubi, E.; Chandru, K.; Erastova, V.; Garg, S.G.; Khawaja, N.; et al. The Future of Origin of Life Research: Bridging Decades-Old Divisions. *Life* **2020**, *10*, 20. [CrossRef]
12. Merino, N.; Aronson, H.S.; Bojanova, D.P.; Feyhl-Buska, J.; Wong, M.L.; Zhang, S.; Giovannelli, D. Living at the Extremes: Extremophiles and the Limits of Life in a Planetary Context. *Front. Microbiol.* **2019**, *10*, 780. [CrossRef] [PubMed]
13. Bendia, A.G.; Araujo, G.G.; Pulschen, A.A.; Contro, B.; Duarte, R.T.D.; Rodrigues, F.; Galante, D.; Pellizari, V.H. Surviving in Hot and Cold: Psychrophiles and Thermophiles from Deception Island Volcano, Antarctica. *Extremophiles* **2018**, *22*, 917–929. [CrossRef] [PubMed]
14. Fredrickson, J.K.; Zachara, J.M.; Balkwill, D.L.; Kennedy, D.; Li, S.-M.W.; Kostandarithes, H.M.; Daly, M.J.; Romine, M.F.; Brockman, F.J. Geomicrobiology of High-Level Nuclear Waste-Contaminated Vadose Sediments at the Hanford Site, Washington State. *Appl. Environ. Microbiol.* **2004**, *70*, 4230–4241. [CrossRef]
15. Baker, B.J.; Banfield, J.F. Microbial Communities in Acid Mine Drainage. *FEMS Microbiol. Ecol.* **2003**, *44*, 139–152. [CrossRef]
16. Kawaguchi, Y.; Shibuya, M.; Kinoshita, I.; Yatabe, J.; Narumi, I.; Shibata, H.; Hayashi, R.; Fujiwara, D.; Murano, Y.; Hashimoto, H.; et al. DNA Damage and Survival Time Course of Deinococcal Cell Pellets during 3 Years of Exposure to Outer Space. *Front. Microbiol.* **2020**, *11*, 2050. [CrossRef] [PubMed]
17. Danko, D.C.; Sierra, M.A.; Benardini, J.N.; Guan, L.; Wood, J.M.; Singh, N.; Seuylemezian, A.; Butler, D.J.; Ryon, K.; Kuchin, K.; et al. A Comprehensive Metagenomics Framework to Characterize Organisms Relevant for Planetary Protection. *Microbiome* **2021**, *9*, 82. [CrossRef]
18. Miller, S.L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science* **1953**, *117*, 528–529. [CrossRef] [PubMed]
19. Miller, S.L.; Urey, H.C. Organic Compound Synthesis on the Primitive Earth. *Science* **1959**, *130*, 245–251. [CrossRef]
20. Cleaves, H.J., II. The Origin of the Biologically Coded Amino Acids. *J. Theor. Biol.* **2010**, *263*, 490–498. [CrossRef]
21. Kebukawa, Y.; Chan, Q.H.S.; Tachibana, S.; Kobayashi, K.; Zolensky, M.E. One-Pot Synthesis of Amino Acid Precursors with Insoluble Organic Matter in Planetesimals with Aqueous Activity. *Sci. Adv.* **2017**, *3*, e1602093. [CrossRef]
22. Magrino, T.; Pietrucci, F.; Saitta, A.M. Step by Step Strecker Amino Acid Synthesis from Ab Initio Prebiotic Chemistry. *J. Phys. Chem. Lett.* **2021**, *12*, 2630–2637. [CrossRef]
23. Pietrucci, F.; Aponte, J.C.; Starr, R.; Pérez-Villa, A.; Elsila, J.E.; Dworkin, J.P.; Saitta, A.M. Hydrothermal Decomposition of Amino Acids and Origins of Prebiotic Meteoritic Organic Compounds. *ACS Earth Space Chem.* **2018**, *2*, 588–598. [CrossRef]
24. Kvenvolden, K.A.; Lawless, J.G.; Ponnamperuma, C. Nonprotein Amino Acids in the Murchison Meteorite. *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 486–490. [CrossRef]
25. Burton, A.S.; Stern, J.C.; Elsila, J.E.; Glavin, D.P.; Dworkin, J.P. Understanding Prebiotic Chemistry through the Analysis of Extraterrestrial Amino Acids and Nucleobases in Meteorites. *Chem. Soc. Rev.* **2012**, *41*, 5459–5472. [CrossRef] [PubMed]
26. Glavin, D.P.; Callahan, M.P.; Dworkin, J.P.; Elsila, J.E. The Effects of Parent Body Processes on Amino Acids in Carbonaceous Chondrites. *Meteorit. Planet. Sci.* **2010**, *45*, 1948–1972. [CrossRef]
27. Pizzarello, S.; Shock, E. The Organic Composition of Carbonaceous Meteorites: The Evolutionary Story Ahead of Biochemistry. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002105. [CrossRef]
28. Elsila, J.E.; Aponte, J.C.; Blackmond, D.G.; Burton, A.S.; Dworkin, J.P.; Glavin, D.P. Meteoritic Amino Acids: Diversity in Compositions Reflects Parent Body Histories. *ACS Cent. Sci.* **2016**, *2*, 370–379. [CrossRef]
29. Simkus, D.N.; Aponte, J.C.; Elsila, J.E.; Hilts, R.W.; McLain, H.L.; Herd, C.D.K. New Insights into the Heterogeneity of the Tagish Lake Meteorite: Soluble Organic Compositions of Variously Altered Specimens. *Meteorit. Planet. Sci.* **2019**, *54*, 1283–1302. [CrossRef]
30. Koga, T.; Naraoka, H. A New Family of Extraterrestrial Amino Acids in the Murchison Meteorite. *Sci. Rep.* **2017**, *7*, 636. [CrossRef]
31. Naraoka, H.; Takano, Y.; Dworkin, J.P. Soluble Organic Molecules in Samples of the Carbonaceous Asteroid (162173) Ryugu. *Science* **2023**, *379*, eabn9033. [CrossRef]
32. Parker, G.A. The Reproductive Behaviour and the Nature of Sexual Selection in *Scatophaga stercoraria* L. (Diptera: Scatophagidae): II. The Fertilization Rate and the Spatial and Temporal Relationships of Each Sex Around the Site of Mating and Oviposition. *J. Anim. Ecol.* **1970**, *39*, 205–228. [CrossRef]
33. Aponte, J.C.; Elsila, J.E.; Hein, J.E.; Dworkin, J.P.; Glavin, D.P.; McLain, H.L.; Parker, E.T.; Cao, T.; Berger, E.L.; Burton, A.S. Analysis of Amino Acids, Hydroxy Acids, and Amines in CR Chondrites. *Meteorit. Planet. Sci.* **2020**, *55*, 2422–2439. [CrossRef] [PubMed]
34. Deamer, D. The Role of Lipid Membranes in Life's Origin. *Life* **2017**, *7*, 5. [CrossRef] [PubMed]
35. Todd, Z.R.; Cohen, Z.R.; Catling, D.C.; Keller, S.L.; Black, R.A. Growth of Prebiotically Plausible Fatty Acid Vesicles Proceeds in the Presence of Prebiotic Amino Acids, Dipeptides, Sugars, and Nucleic Acid Components. *Langmuir* **2022**, *38*, 15106–15112. [CrossRef] [PubMed]
36. Okamura, H.; Becker, S.; Tiede, N.; Wiedemann, S.; Feldmann, J.; Carell, T. A One-Pot, Water Compatible Synthesis of Pyrimidine Nucleobases under Plausible Prebiotic Conditions. *Chem. Commun.* **2019**, *55*, 1939–1942. [CrossRef] [PubMed]

37. Oró, J. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **1961**, *191*, 1193–1194. [CrossRef] [PubMed]

38. Engelhart, A.E.; Hud, N.V. Primitive Genetic Polymers. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002196. [CrossRef]

39. Fine, J.L.; Pearlman, R.E. On the Origin of Life: An RNA-Focused Synthesis and Narrative. *RNA* **2023**, rna.079598.123. [CrossRef]

40. Orgel, L.E. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123. [CrossRef]

41. Meringer, M.; Cleaves, H.J.; Freeland, S.J. Beyond Terrestrial Biology: Charting the Chemical Universe of α-Amino Acid Structures. *J. Chem. Inf. Model.* **2013**, *53*, 2851–2862. [CrossRef] [PubMed]

42. Boutlerow, M.A. Formation Synthétique d'une Substance Sucrée. *CR Acad. Sci.* **1861**, *53*, 145–147.

43. Ricardo, A.; Carrigan, M.A.; Olcott, A.N.; Benner, S.A. Borate Minerals Stabilize Ribose. *Science* **2004**, *303*, 196. [CrossRef] [PubMed]

44. Decker, P.; Schweer, H.; Pohlamnn, R. Bioids: X. Identification of Formose Sugars, Presumable Prebiotic Metabolites, Using Capillary Gas Chromatography/Gas Chromatography—Mas Spectrometry of n-Butoxime Trifluoroacetates on OV-225. *J. Chromatogr. A* **1982**, *244*, 281–291. [CrossRef]

45. Shapiro, R. Prebiotic Ribose Synthesis: A Critical Analysis. *Orig. Life Evol. Biosph.* **1988**, *18*, 71–85. [CrossRef]

46. Asplund, M.; Grevesse, N.; Sauval, A.J. The new solar abundances—Part I: The observations. In *Communications in Asteroseismology*; Verlag der Österreichischen Akademie der Wissenschaften: Wien, Austria, 2006. [CrossRef]

47. Pasek, M.A. Rethinking Early Earth Phosphorus Geochemistry. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 853–858. [CrossRef]

48. Jerome, C.A.; Kim, H.-J.; Mojzsis, S.J.; Benner, S.A.; Biondi, E. Catalytic Synthesis of Polyribonucleic Acid on Prebiotic Rock Glasses. *Astrobiology* **2022**, *22*, 629–636. [CrossRef]

49. Schuster, G.B.; Cafferty, B.J.; Karunakaran, S.C.; Hud, N.V. Water-Soluble Supramolecular Polymers of Paired and Stacked Heterocycles: Assembly, Structure, Properties, and a Possible Path to Pre-RNA. *J. Am. Chem. Soc.* **2021**, *143*, 9279–9296. [CrossRef]

50. Green, N.J.; Xu, J.; Sutherland, J.D. Illuminating Life's Origins: UV Photochemistry in Abiotic Synthesis of Biomolecules. *J. Am. Chem. Soc.* **2021**, *143*, 7219–7236. [CrossRef]

51. Kim, S.C.; O'flaherty, D.K.; Giurgiu, C.; Zhou, L.; Szostak, J.W. The Emergence of RNA from the Heterogeneous Products of Prebiotic Nucleotide Synthesis. *J. Am. Chem. Soc.* **2021**, *143*, 3267–3279. [CrossRef]

52. Forsythe, J.G.; Yu, S.; Mamajanov, I.; Grover, M.A.; Krishnamurthy, R.; Fernández, F.M.; Hud, N.V. Ester-Mediated Amide Bond Formation Driven by Wet–Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angew. Chem. Int. Ed.* **2015**, *54*, 9871–9875. [CrossRef]

53. Anastasi, C.; Buchet, F.F.; Crowe, M.A.; Parkes, A.L.; Powner, M.W.; Smith, J.M.; Sutherland, J.D. RNA: Prebiotic Product, or Biotic Invention? *Chem. Biodivers.* **2007**, *4*, 721–739. [CrossRef] [PubMed]

54. Freeland, S. Undefining Life's Biochemistry: Implications for Abiogenesis. *J. R. Soc. Interface* **2022**, *19*, 20210814. [CrossRef] [PubMed]

55. Croswell, K. *The Alchemy of the Heavens*; Oxford University Press: Oxford, UK, 1996.

56. Kitadai, N.; Maruyama, S. Origins of Building Blocks of Life: A Review. *Geosci. Front.* **2018**, *9*, 1117–1153. [CrossRef]

57. Furukawa, Y.; Chikaraishi, Y.; Ohkouchi, N.; Ogawa, N.O.; Glavin, D.P.; Dworkin, J.P.; Abe, C.; Nakamura, T. Extraterrestrial Ribose and Other Sugars in Primitive Meteorites. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24440–24445. [CrossRef] [PubMed]

58. Lai, J.C.-Y.; Pearce, B.K.; Pudritz, R.E.; Lee, D. Meteoritic Abundances of Fatty Acids and Potential Reaction Pathways in Planetesimals. *Icarus* **2019**, *319*, 685–700. [CrossRef]

59. Oba, Y.; Takano, Y.; Furukawa, Y.; Koga, T.; Glavin, D.P.; Dworkin, J.P.; Naraoka, H. Identifying the Wide Diversity of Extraterrestrial Purine and Pyrimidine Nucleobases in Carbonaceous Meteorites. *Nat. Commun.* **2022**, *13*, 2008. [CrossRef]

60. Frenkel-Pinter, M.; Jacobson, K.C.; Eskew-Martin, J.; Forsythe, J.G.; Grover, M.A.; Williams, L.D.; Hud, N.V. Differential Oligomerization of Alpha versus Beta Amino Acids and Hydroxy Acids in Abiotic Proto-Peptide Synthesis Reactions. *Life* **2022**, *12*, 265. [CrossRef]

61. Guo, R.; McGrath, J.E. 5.17—Aromatic Polyethers, Polyetherketones, Polysulfides, and Polysulfones. In *Polymer Science: A Comprehensive Reference*; Matyjaszewski, K., Möller, M., Eds.; Elsevier: Amsterdam, The Netherlands, 2012; pp. 377–430. [CrossRef]

62. Robinson, B.A.; Tester, J.W. Kinetics of Alkaline Hydrolysis of Organic Esters and Amides in Neutrally-Buffered Solution. *Int. J. Chem. Kinet.* **1990**, *22*, 431–448. [CrossRef]

63. Irion, R. Astrobiologists Try to "Follow the Water to Life". *Science* **2002**, *296*, 647–648. [CrossRef]

64. Schwieterman, E.W.; Kiang, N.Y.; Parenteau, M.N.; Harman, C.E.; DasSarma, S.; Fisher, T.M.; Arney, G.N.; Hartnett, H.E.; Reinhard, C.T.; Olson, S.L.; et al. Exoplanet Biosignatures: A Review of Remotely Detectable Signs of Life. *Astrobiology* **2018**, *18*, 663–708. [CrossRef]

65. McKay, C.P.; Smith, H.D. Possibilities for Methanogenic Life in Liquid Methane on the Surface of Titan. *Icarus* **2005**, *178*, 274–276. [CrossRef]

66. Budisa, N.; Schulze-Makuch, D. Supercritical Carbon Dioxide and Its Potential as a Life-Sustaining Solvent in a Planetary Environment. *Life* **2014**, *4*, 331–340. [CrossRef]

67. Finney, J.L. Water? What's so Special about It? *Phil Trans. R. Soc. Lond. B* **2004**, *359*, 1145–1165. [CrossRef]

68. Lynden-Bell, R.M.; Morris, S.C.; Barrow, J.D.; Finney, J.L.; Harper, C. (Eds.) *Water and Life: The Unique Properties of H_2O*; CRC Press: Boca Raton, FL, USA, 2010. [CrossRef]

69. Mottl, M.J.; Glazer, B.T.; Kaiser, R.I.; Meech, K.J. Water and Astrobiology. *Geochemistry* **2007**, *67*, 253–282. [CrossRef]
70. Lelais, G.; Seebach, D. Beta2-Amino Acids-Syntheses, Occurrence in Natural Products, and Components of Beta-Peptides1,2. *Biopolymers* **2004**, *76*, 206–243. [CrossRef] [PubMed]
71. Li, L.; Dou, N.; Zhang, H.; Wu, C. The Versatile GABA in Plants. *Plant Signal. Behav.* **2021**, *16*, 1862565. [CrossRef] [PubMed]
72. Sigel, E.; Steinmann, M.E. Structure, Function, and Modulation of GABA(A) Receptors. *J. Biol. Chem.* **2012**, *287*, 40224–40231. [CrossRef] [PubMed]
73. Cox, P.A.; Banack, S.A.; Murch, S.J.; Rasmussen, U.; Tien, G.; Bidigare, R.R.; Metcalf, J.S.; Morrison, L.F.; Codd, G.A.; Bergman, B. Diverse Taxa of Cyanobacteria Produce β-N-Methylamino-l-Alanine, a Neurotoxic Amino Acid. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5074–5078. [CrossRef] [PubMed]
74. Vega, A.; Bell, E.A. α-Amino-β-Methylaminopropionic Acid, a New Amino Acid from Seeds of Cycas Circinalis. *Phytochemistry* **1967**, *6*, 759–762. [CrossRef]
75. Kiss, L.; Mándity, I.M.; Fülöp, F. Highly Functionalized Cyclic β-Amino Acid Moieties as Promising Scaffolds in Peptide Research and Drug Design. *Amino Acids* **2017**, *49*, 1441–1455. [CrossRef] [PubMed]
76. Legrand, B.; Maillard, L.T. α,β-Unsaturated γ-Peptide Foldamers. *ChemPlusChem* **2021**, *86*, 629–645. [CrossRef]
77. Nagata, M.; Watanabe, M.; Doi, R.; Uemura, M.; Ochiai, N.; Ichinose, W.; Fujiwara, K.; Sato, Y.; Kameda, T.; Takeuchi, K.; et al. Helix-Forming Aliphatic Homo-δ-Peptide Foldamers Based on the Conformational Restriction Effects of Cyclopropane. *Org. Biomol. Chem.* **2023**, *21*, 970–980. [CrossRef]
78. Forsythe, J.G.; English, S.L.; Simoneaux, R.E.; Weber, A.L. Synthesis of β-Peptide Standards for Use in Model Prebiotic Reactions. *Orig. Life Evol. Biosph.* **2018**, *48*, 201–211. [CrossRef] [PubMed]
79. Fülöp, F.; Martinek, T.A.; Tóth, G.K. Application of Alicyclic β-Amino Acids in Peptide Chemistry. *Chem. Soc. Rev.* **2006**, *35*, 323–334. [CrossRef]
80. Steer, D.; Lew, R.; Perlmutter, P.; Smith, A.; Aguilar, M.-I. Beta-Amino Acids: Versatile Peptidomimetics. *Curr. Med. Chem.* **2002**, *9*, 811–822. [CrossRef]
81. Levin, S.R.; Scott, T.W.; Cooper, H.S.; West, S.A. Darwin's Aliens. *Int. J. Astrobiol.* **2019**, *18*, 1–9. [CrossRef]
82. Burton, A.S.; Berger, E.L. Insights into Abiotically-Generated Amino Acid Enantiomeric Excesses Found in Meteorites. *Life* **2018**, *8*, 14. [CrossRef]
83. Cronin, J.R.; Pizzarello, S. Amino Acids in Meteorites. *Adv. Space Res.* **1983**, *3*, 5–18. [CrossRef]
84. Ashe, K.; Fernández-García, C.; Corpinot, M.K.; Coggins, A.J.; Bučar, D.-K.; Powner, M.W. Selective Prebiotic Synthesis of Phosphoroaminonitriles and Aminothioamides in Neutral Water. *Commun. Chem.* **2019**, *2*, 23. [CrossRef]
85. Strecker, A. Ueber Die Künstliche Bildung Der Milchsäure Und Einen Neuen, Dem Glycocoll Homologen Körper—Strecker—1850—Justus Liebigs Annalen Der Chemie—Wiley Online Library. *Justus Liebigs Ann. Der Chem.* **1850**, *75*, 27–45. [CrossRef]
86. Masamba, W. Petasis vs. Strecker Amino Acid Synthesis: Convergence, Divergence and Opportunities in Organic Synthesis. *Molecules* **2021**, *26*, 1707. [CrossRef] [PubMed]
87. Grishin, D.V.; Zhdanov, D.D.; Pokrovskaya, M.V.; Sokolov, N.N. D-Amino Acids in Nature, Agriculture and Biomedicine. *All Life* **2020**, *13*, 11–22. [CrossRef]
88. Sasabe, J.; Suzuki, M. Distinctive Roles of D-Amino Acids in the Homochiral World: Chirality of Amino Acids Modulates Mammalian Physiology and Pathology. *Keio J. Med.* **2019**, *68*, 1–16. [CrossRef]
89. Heck, S.D.; Siok, C.J.; Krapcho, K.J.; Kelbaugh, P.R.; Thadeio, P.F.; Welch, M.J.; Williams, R.D.; Ganong, A.H.; Kelly, M.E.; Lanzetti, A.J.; et al. Functional Consequences of Posttranslational Isomerization of Ser46 in a Calcium Channel Toxin. *Science* **1994**, *266*, 1065–1068. [CrossRef] [PubMed]
90. Wolosker, H.; Dumin, E.; Balan, L.; Foltyn, V.N. D-Amino Acids in the Brain: D-Serine in Neurotransmission and Neurodegeneration. *FEBS J.* **2008**, *275*, 3514–3526. [CrossRef]
91. D'Aniello, A.; D'Onofrio, G.; Pischetola, M.; D'Aniello, G.; Vetere, A.; Petrucelli, L.; Fisher, G.H. Biological Role of D-Amino Acid Oxidase and D-Aspartate Oxidase. Effects of D-Amino Acids. *J. Biol. Chem.* **1993**, *268*, 26941–26949. [CrossRef]
92. Pollegioni, L.; Piubelli, L.; Sacchi, S.; Pilone, M.S.; Molla, G. Physiological Functions of D-Amino Acid Oxidases: From Yeast to Humans. *Cell. Mol. Life Sci.* **2007**, *64*, 1373–1394. [CrossRef]
93. Maruyama, C.; Hamano, Y. The Assembly-Line Enzymology of Nonribosomal Peptide Biosynthesis. *Methods Mol. Biol.* **2023**, *2670*, 3–16. [CrossRef]
94. Vollmer, W.; Blanot, D.; de Pedro, M.A. Peptidoglycan Structure and Architecture. *FEMS Microbiol. Rev.* **2008**, *32*, 149–167. [CrossRef]
95. Evans, A.C.; Meinert, C.; Giri, C.; Goesmann, F.; Meierhenrich, U.J. Chirality, Photochemistry and the Detection of Amino Acids in Interstellar Ice Analogues and Comets. *Chem. Soc. Rev.* **2012**, *41*, 5447–5458. [CrossRef]
96. Davankov, V.A. Inherent Homochirality of Primary Particles and Meteorite Impacts as Possible Source of Prebiotic Molecular Chirality. *Russ. J. Phys. Chem.* **2009**, *83*, 1247–1256. [CrossRef]
97. Jorissen, A.; Cerf, C. Asymmetric Photoreactions as the Origin of Biomolecular Homochirality: A Critical Review. *Orig. Life Evol. Biosph.* **2002**, *32*, 129–142. [CrossRef] [PubMed]
98. Kawasaki, T.; Sato, M.; Ishiguro, S.; Saito, T.; Morishita, Y.; Sato, I.; Nishino, H.; Inoue, Y.; Soai, K. Enantioselective Synthesis of Near Enantiopure Compound by Asymmetric Autocatalysis Triggered by Asymmetric Photolysis with Circularly Polarized Light. *J. Am. Chem. Soc.* **2005**, *127*, 3274–3275. [CrossRef] [PubMed]

99. Ozturk, S.F.; Sasselov, D.D. On the Origins of Life's Homochirality: Inducing Enantiomeric Excess with Spin-Polarized Electrons. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2204765119. [CrossRef] [PubMed]

100. Pasteur, M.L. Translations: On the Origin of Ferments. New Experiments Relative to so-Termed Spontaneous Generation. *J. Cell Sci.* **1860**, *s1–s8*, 255–259. [CrossRef]

101. Weissbuch, I.; Lahav, M. Crystalline Architectures as Templates of Relevance to the Origins of Homochirality. *Chem. Rev.* **2011**, *111*, 3236–3267. [CrossRef] [PubMed]

102. Noorduin, W.L.; Bode, A.A.C.; van der Meijden, M.; Meekes, H.; van Etteger, A.F.; van Enckevort, W.J.P.; Christianen, P.C.M.; Kaptein, B.; Kellogg, R.M.; Rasing, T.; et al. Complete Chiral Symmetry Breaking of an Amino Acid Derivative Directed by Circularly Polarized Light. *Nat. Chem.* **2009**, *1*, 729–732. [CrossRef]

103. Takahashi, J.; Kobayashi, K. Origin of Terrestrial Bioorganic Homochirality and Symmetry Breaking in the Universe. *Symmetry* **2019**, *11*, 919. [CrossRef]

104. Breslow, R.; Cheng, Z.-L. L-Amino Acids Catalyze the Formation of an Excess of D-Glyceraldehyde, and Thus of Other D Sugars, under Credible Prebiotic Conditions. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5723–5725. [CrossRef]

105. Wagner, A.J.; Zubarev, D.Y.; Aspuru-Guzik, A.; Blackmond, D.G. Chiral Sugars Drive Enantioenrichment in Prebiotic Amino Acid Synthesis. *ACS Cent. Sci.* **2017**, *3*, 322–328. [CrossRef]

106. Blackmond, D.G. The Origin of Biological Homochirality. *Cold Spring Harb. Perspect. Biol.* **2019**, *11*, a032540. [CrossRef]

107. Percec, V.; Leowanawat, P. Why Are Biological Systems Homochiral? *Isr. J. Chem.* **2011**, *51*, 1107–1117. [CrossRef]

108. Toxvaerd, S. Origin of Homochirality in Biosystems. *Int. J. Mol. Sci.* **2009**, *10*, 1290–1299. [CrossRef]

109. Bryliakov, K.P. Chemical Mechanisms of Prebiotic Chirality Amplification. *Research* **2020**, *2020*, 5689246. [CrossRef] [PubMed]

110. Ruiz-Mirazo, K.; Briones, C.; de la Escosura, A. Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. *Chem. Rev.* **2014**, *114*, 285–366. [CrossRef] [PubMed]

111. Bonner, W.A. The Origin and Amplification of Biomolecular Chirality. *Orig. Life Evol. Biosph.* **1991**, *21*, 59–111. [CrossRef] [PubMed]

112. Sallembien, Q.; Bouteiller, L.; Crassous, J.; Raynal, M. Possible Chemical and Physical Scenarios towards Biological Homochirality. *Chem. Soc. Rev.* **2022**, *51*, 3436–3476. [CrossRef] [PubMed]

113. Brack, A.; Spach, G. β-Structures of Polypeptides with L- and D-Residues. *J. Mol. Evol.* **1979**, *13*, 35–46. [CrossRef] [PubMed]

114. Nanda, V.; Andrianarijaona, A.; Narayanan, C. The Role of Protein Homochirality in Shaping the Energy Landscape of Folding. *Protein Sci.* **2007**, *16*, 1667–1675. [CrossRef] [PubMed]

115. Weil-Ktorza, O.; Fridmann-Sirkis, Y.; Despotovic, D.; Naveh-Tassa, S.; Levy, Y.; Metanis, N.; Longo, L.M. Functional Ambidexterity of an Ancient Nucleic Acid-Binding Domain. *bioRxiv* **2023**. [CrossRef]

116. Genchi, G. An overview on D-amino acids. *Amino Acids* **2017**, *49*, 1521–1533. [CrossRef] [PubMed]

117. Fan, C.; Deng, Q.; Zhu, T.F. Bioorthogonal Information Storage in L-DNA with a High-Fidelity Mirror-Image Pfu DNA Polymerase. *Nat. Biotechnol.* **2021**, *39*, 1548–1555. [CrossRef]

118. Wang, M.; Jiang, W.; Liu, X.; Wang, J.; Zhang, B.; Fan, C.; Liu, L.; Pena-Alcantara, G.; Ling, J.-J.; Chen, J.; et al. Mirror-Image Gene Transcription and Reverse Transcription. *Chem* **2019**, *5*, 848–857. [CrossRef]

119. Green, M.M.; Jain, V. Homochirality in Life: Two Equal Runners, One Tripped. *Orig. Life Evol. Biosph.* **2010**, *40*, 111–118. [CrossRef]

120. Elsliger, M.-A.; Wilson, I.A. 1.8 Structure Validation and Analysis. In *Comprehensive Biophysics*; Egelman, E.H., Ed.; Elsevier: Amsterdam, The Netherlands, 2012; pp. 116–135. [CrossRef]

121. Kalmankar, N.V.; Ramakrishnan, C.; Balaram, P. Sparsely Populated Residue Conformations in Protein Structures: Revisiting "Experimental" Ramachandran Maps. *Proteins* **2014**, *82*, 1101–1112. [CrossRef]

122. Linderstrã¸m-Lang, K. *Proteins and Enzymes*; Stanford University Press: Redwood City, CA, USA, 1952.

123. Sehnal, D.; Bittrich, S.; Deshpande, M.; Svobodová, R.; Berka, K.; Bazgier, V.; Velankar, S.; Burley, S.K.; Koča, J.; Rose, A.S. Mol* Viewer: Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures. *Nucleic Acids Res.* **2021**, *49*, W431–W437. [CrossRef]

124. London, N.; Miller, R.M.; Krishnan, S.; Uchida, K.; Irwin, J.J.; Eidam, O.; Gibold, L.; Cimermančič, P.; Bonnet, R.; Shoichet, B.K.; et al. Covalent Docking of Large Libraries for the Discovery of Chemical Probes. *Nat. Chem. Biol.* **2014**, *10*, 1066–1072. [CrossRef]

125. Fretheim, K.; Iwai, S.; Feeney, R.E. Extensive Modification of Protein Amino Groups by Reductive Addition of Different Sized Substituents. *Int. J. Pept. Protein Res.* **1979**, *14*, 451–456. [CrossRef]

126. Tanaka, M. Design and Synthesis of Chiral α,α-Disubstituted Amino Acids and Conformational Study of Their Oligopeptides. *Chem. Pharm. Bull.* **2007**, *55*, 349–358. [CrossRef]

127. Wang, X.; Li, J.; Hayashi, Y. Highly Sterically Hindered Peptide Bond Formation between α,α-Disubstituted α-Amino Acids and N-Alkyl Cysteines Using α,α-Disubstituted α-Amidonitrile. *J. Am. Chem. Soc.* **2022**, *144*, 10145–10150. [CrossRef]

128. McKay, C.P. What Is Life—And How Do We Search for It in Other Worlds? *PLoS Biol.* **2004**, *2*, e302. [CrossRef] [PubMed]

129. Anfinsen, C.B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230. [CrossRef] [PubMed]

130. Frank, F.C. Energy of Formation of Cyclol Molecules. *Nature* **1936**, *138*, 242. [CrossRef]

131. Linus, P.; Carl, N. The Structure of Proteins. *J. Am. Chem. Soc.* **1939**, *61*, 1860–1867. [CrossRef]

132. Chiarabelli, C.; Vrijbloed, J.W.; Thomas, R.M.; Luisi, P.L. Investigation of de Novo Totally Random Biosequences, Part I. *Chem. Biodivers.* **2006**, *3*, 827–839. [CrossRef]

133. Chiarabelli, C.; Vrijbloed, J.W.; De Lucrezia, D.; Thomas, R.M.; Stano, P.; Policelli, F.; Ottone, T.; Papa, E.; Luisi, P.L. Investigation of de Novo Totally Random Biosequences, Part II: On the Folding Frequency in a Totally Random Library of de Novo Proteins Obtained by Phage Display. *Chem. Biodivers.* **2006**, *3*, 840–859. [CrossRef]

134. Uy, R.; Wold, F. Posttranslational Covalent Modification of Proteins. *Science* **1977**, *198*, 890–896. [CrossRef]

135. Wong, J.T.-F.; Bronskill, P.M. Inadequacy of Prebiotic Synthesis as Origin of Proteinous Amino Acids. *J. Mol. Evol.* **1979**, *13*, 115–125. [CrossRef]

136. Fekkes, D. Automated Analysis of Primary Amino Acids in Plasma by High-Performance Liquid Chromatography. *Methods Mol. Biol.* **2012**, *828*, 183–200. [CrossRef]

137. Flissi, A.; Ricart, E.; Campart, C.; Chevalier, M.; Dufresne, Y.; Michalik, J.; Jacques, P.; Flahaut, C.; Lisacek, F.; Leclère, V.; et al. Norine: Update of the Nonribosomal Peptide Resource. *Nucleic Acids Res.* **2020**, *48*, D465–D469. [CrossRef] [PubMed]

138. Mayer-Bacon, C.; Agboha, N.; Muscalli, M.; Freeland, S. Evolution as a Guide to Designing Xeno Amino Acid Alphabets. *Int. J. Mol. Sci.* **2021**, *22*, 2787. [CrossRef] [PubMed]

139. Andrews, J.; Gan, Q.; Fan, C. "Not-so-Popular" Orthogonal Pairs in Genetic Code Expansion. *Protein Sci.* **2023**, *32*, e4559. [CrossRef]

140. Lee, J.; Schwieter, K.E.; Watkins, A.M.; Kim, D.S.; Yu, H.; Schwarz, K.J.; Lim, J.; Coronado, J.; Byrom, M.; Anslyn, E.V.; et al. Expanding the Limits of the Second Genetic Code with Ribozymes. *Nat. Commun.* **2019**, *10*, 5097. [CrossRef] [PubMed]

141. Dell, M.; Dunbar, K.L.; Hertweck, C. Ribosome-Independent Peptide Biosynthesis: The Challenge of a Unifying Nomenclature. *Nat. Prod. Rep.* **2022**, *39*, 453–459. [CrossRef]

142. Reimer, J.M.; Haque, A.S.; Tarry, M.J.; Schmeing, T.M. Piecing Together Nonribosomal Peptide Synthesis. *Curr. Opin. Struct. Biol.* **2018**, *49*, 104–113. [CrossRef]

143. Feldman, A.W.; Dien, V.T.; Karadeema, R.J.; Fischer, E.C.; You, Y.; Anderson, B.A.; Krishnamurthy, R.; Chen, J.S.; Li, L.; Romesberg, F.E. Optimization of Replication, Transcription, and Translation in a Semi-Synthetic Organism. *J. Am. Chem. Soc.* **2019**, *141*, 10644–10653. [CrossRef]

144. Mayer-Bacon, C.; Meringer, M.; Havel, R.; Aponte, J.C.; Freeland, S. A Closer Look at Non-Random Patterns within Chemistry Space for a Smaller, Earlier Amino Acid Alphabet. *J. Mol. Evol.* **2022**, *90*, 307–323. [CrossRef]

145. Freeland, S. "Terrestrial" Amino Acids and Their Evolution. In *Amino Acids, Peptides and Proteins in Organic Chemistry*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2009; pp. 43–75. [CrossRef]

146. Dumas, A.; Lercher, L.; Spicer, C.D.; Davis, B.G. Designing Logical Codon Reassignment—Expanding the Chemistry in Biology. *Chem. Sci.* **2014**, *6*, 50–69. [CrossRef]

147. Liu, C.C.; Schultz, P.G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444. [CrossRef]

148. Nödling, A.R.; Spear, L.A.; Williams, T.L.; Luk, L.Y.; Tsai, Y.-H. Using Genetically Incorporated Unnatural Amino Acids to Control Protein Functions in Mammalian Cells. *Essays Biochem.* **2019**, *63*, 237–266. [CrossRef] [PubMed]

149. Anfinsen, C.B.; Corley, L.G. An Active Variant of Staphylococcal Nuclease Containing Norleucine in Place of Methionine. *J. Biol. Chem.* **1969**, *244*, 5149–5152. [CrossRef] [PubMed]

150. Jukes, T.H. Arginine as an Evolutionary Intruder into Protein Synthesis. *Biochem. Biophys. Res. Commun.* **1973**, *53*, 709–714. [CrossRef] [PubMed]

151. Kawashima, S.; Pokarowski, P.; Pokarowska, M.; Kolinski, A.; Katayama, T.; Kanehisa, M. AAindex: Amino Acid Index Database, Progress Report 2008. *Nucleic Acids Res* **2008**, *36*, D202–D205. [CrossRef] [PubMed]

152. Beadle, G.W.; Tatum, E.L.; Lederberg, J. The Nobel Prize in Physiology or Medicine 1958. Available online: https://www.nobelprize.org/prizes/medicine/1958/summary/ (accessed on 29 October 2023).

153. Crick, F.H.; Watson, J.D.; Wilkins, M.H. The Nobel Prize in Physiology or Medicine 1962. Available online: https://www.nobelprize.org/prizes/medicine/1962/summary/ (accessed on 29 October 2023).

154. Delbrück, M.; Hershey, A.D.; Luria, S.E. The Nobel Prize in Physiology or Medicine 1969. Available online: https://www.nobelprize.org/prizes/medicine/1969/press-release/ (accessed on 29 October 2023).

155. Holley, R.W.; Khorana, H.G.; Nirenberg, M.W. The Nobel Prize in Physiology or Medicine 1968. Available online: https://www.nobelprize.org/prizes/medicine/1968/summary/ (accessed on 29 October 2023).

156. Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227*, 561–563. [CrossRef]

157. Crick, F.H. Codon—Anticodon Pairing: The Wobble Hypothesis. *J. Mol. Biol.* **1966**, *19*, 548–555. [CrossRef]

158. Knight, R.D.; Freeland, S.J.; Landweber, L.F. Selection, History and Chemistry: The Three Faces of the Genetic Code. *Trends Biochem. Sci.* **1999**, *24*, 241–247. [CrossRef]

159. Koonin, E.V.; Novozhilov, A.S. Origin and Evolution of the Universal Genetic Code. *Annu. Rev. Genet.* **2017**, *51*, 45–62. [CrossRef]

160. Wong, J.T.-F. A Co-Evolution Theory of the Genetic Code. *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1909–1912. [CrossRef]

161. Levinthal, C. How to Fold Graciously. *Mössbaun Spectrosc. Biol. Syst. Proc.* **1969**, *67*, 22–24.

162. Kryshtafovych, A.; Venclovas, Č.; Fidelis, K.; Moult, J. Progress over the First Decade of CASP Experiments. *Proteins* **2005**, *61* (Suppl. S7), 225–236. [CrossRef] [PubMed]

163. Altman, S.; Cech, T.R. The Nobel Prize in Chemistry 1989. Available online: https://www.nobelprize.org/prizes/chemistry/1989/summary/ (accessed on 29 October 2023).

164. Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319*, 618. [CrossRef]

165. Joyce, G.F. The Antiquity of RNA-Based Evolution. *Nature* **2002**, *418*, 214–221. [CrossRef]

166. Štorchová, H.; Gesteland, R.F.; Atkins, J.F. The RNA World. *Biol. Plant* **1994**, *36*, 358. [CrossRef]
167. Benner, S.A.; Ellington, A.D.; Tauer, A. Modern Metabolism as a Palimpsest of the RNA World. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7054–7058. [CrossRef] [PubMed]
168. Freeland, S.J.; Knight, R.D.; Landweber, L.F. Do Proteins Predate DNA? *Science* **1999**, *286*, 690–692. [CrossRef]
169. Rivas, M.; Fox, G.E. Ancestry of RNA/RNA Interaction Regions within Segmented Ribosomes. *RNA* **2023**, *29*, 1388–1399. [CrossRef]
170. Yarus, M. The Genetic Code and RNA-Amino Acid Affinities. *Life* **2017**, *7*, 13. [CrossRef]
171. Doig, A.J. Frozen, but No Accident—Why the 20 Standard Amino Acids Were Selected. *FEBS J.* **2017**, *284*, 1296–1305. [CrossRef]
172. Trifonov, E.N. Consensus Temporal Order of Amino Acids and Evolution of the Triplet Code. *Gene* **2000**, *261*, 139–151. [CrossRef]
173. Higgs, P.G.; Pudritz, R.E. A Thermodynamic Basis for Prebiotic Amino Acid Synthesis and the Nature of the First Genetic Code. *Astrobiology* **2009**, *9*, 483–490. [CrossRef] [PubMed]
174. Anfinsen, C.B.; Haber, E.; Sela, M.; White, F.H., Jr. The Kinetics of Formation of Native Ribonuclease during Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1309–1314. [CrossRef] [PubMed]
175. Brown, S.M.; Voráček, V.; Freeland, S. What Would an Alien Amino Acid Alphabet Look Like and Why? *Astrobiology* **2023**, *23*, 536–549. [CrossRef] [PubMed]
176. Culka, M.; Kalvoda, T.; Gutten, O.; Rulíšek, L. Mapping Conformational Space of All 8000 Tripeptides by Quantum Chemical Methods: What Strain Is Affordable within Folded Protein Chains? *J. Phys. Chem. B* **2021**, *125*, 58–69. [CrossRef] [PubMed]
177. Dayhoff, M.O.; Schwartz, R.M.; Orcutt, B.C. 22 A Model of Evolutionary Change in Proteins. *Atlas Protein Seq. Struct.* **1978**, *5*, 345–352.
178. Frenkel-Pinter, M.; Haynes, J.W.C.M.; Petrov, A.S.; Burcar, B.T.; Krishnamurthy, R.; Hud, N.V.; Leman, L.J.; Williams, L.D. Selective Incorporation of Proteinaceous over Nonprosteinaceous Cationic Amino Acids in Model Prebiotic Oligomerization Reactions. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 16338–16346. [CrossRef] [PubMed]
179. Furka, Á.; Sebestyén, F.; Asgedom, M.; Dibó, G. General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493. [CrossRef]
180. Granold, M.; Hajieva, P.; Toşa, M.I.; Irimie, F.-D.; Moosmann, B. Modern Diversification of the Amino Acid Repertoire Driven by Oxygen. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 41–46. [CrossRef]
181. Ilardo, M.; Meringer, M.; Freeland, S.; Rasulev, B.; Ii, H.J.C. Extraordinarily Adaptive Properties of the Genetically Encoded Amino Acids. *Sci. Rep.* **2015**, *5*, 9414. [CrossRef]
182. Kalvoda, T.; Culka, M.; Rulíšek, L.; Andris, E. Exhaustive Mapping of the Conformational Space of Natural Dipeptides by the DFT-D3//COSMO-RS Method. *J. Phys. Chem. B* **2022**, *126*, 5949–5958. [CrossRef]
183. Karplus, M.; McCammon, J.A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646–652. [CrossRef] [PubMed]
184. Keefe, A.D.; Szostak, J.W. Functional Proteins from a Random-Sequence Library. *Nature* **2001**, *410*, 715–718. [CrossRef] [PubMed]
185. Kimura, M. Evolutionary Rate at the Molecular Level. *Nature* **1968**, *217*, 624–626. [CrossRef] [PubMed]
186. Longo, L.M.; Despotović, D.; Weil-Ktorza, O.; Walker, M.J.; Jabłońska, J.; Fridmann-Sirkis, Y.; Varani, G.; Metanis, N.; Tawfik, D.S. Primordial Emergence of a Nucleic Acid-Binding Protein via Phase Separation and Statistical Ornithine-to-Arginine Conversion. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 15731–15739. [CrossRef] [PubMed]
187. Longo, L.M.; Lee, J.; Blaber, M. Simplified Protein Design Biased for Prebiotic Amino Acids Yields a Foldable, Halophilic Protein. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2135–2139. [CrossRef]
188. Makarov, M.; Rocha, A.C.S.; Krystufek, R.; Cherepashuk, I.; Dzmitruk, V.; Charnavets, T.; Faustino, A.M.; Lebl, M.; Fujishima, K.; Fried, S.D.; et al. Early Selection of the Amino Acid Alphabet Was Adaptively Shaped by Biophysical Constraints of Foldability. *J. Am. Chem. Soc.* **2023**, *149*, 5320–5329. [CrossRef]
189. Maynard Smith, J. Natural Selection and the Concept of a Protein Space. *Nature* **1970**, *225*, 563–564. [CrossRef] [PubMed]
190. Merrifield, B. Solid Phase Synthesis. *Science* **1986**, *232*, 341–347. [CrossRef]
191. Moosmann, B. Redox Biochemistry of the Genetic Code. *Trends Biochem. Sci.* **2021**, *46*, 83–86. [CrossRef]
192. Philip, G.K.; Freeland, S.J. Did Evolution Select a Nonrandom "Alphabet" of Amino Acids? *Astrobiology* **2011**, *11*, 235–240. [CrossRef]
193. Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95–99. [CrossRef] [PubMed]
194. Řezáč, J.; Bím, D.; Gutten, O.; Rulíšek, L. Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set. *J. Chem. Theory Comput.* **2018**, *14*, 1254–1266. [CrossRef] [PubMed]
195. Riddle, D.S.; Santiago, J.V.; Bray-Hall, S.T.; Doshi, N.; Grantcharova, V.P.; Yi, Q.; Baker, D. Functional Rapidly Folding Proteins from Simplified Amino Acid Sequences. *Nat. Struct. Mol. Biol.* **1997**, *4*, 805–809. [CrossRef]
196. Tanaka, J.; Doi, N.; Takashima, H.; Yanagawa, H. Comparative Characterization of Random-Sequence Proteins Consisting of 5, 12, and 20 Kinds of Amino Acids. *Protein Sci.* **2010**, *19*, 786–795. [CrossRef] [PubMed]
197. 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.* **2022**, *12*, 220040. [CrossRef] [PubMed]
198. Weber, A.L. Thermal Synthesis and Hydrolysis of Polyglyceric Acid. *Orig. Life Evol. Biosph.* **1989**, *19*, 7–19. [CrossRef] [PubMed]

199. Woolfson, D.N. A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. *J. Mol. Biol.* **2021**, *433*, 167160. [CrossRef]
200. Brack, A.; Orgel, L.E. β Structures of Alternating Polypeptides and Their Possible Prebiotic Significance. *Nature* **1975**, *256*, 383–387. [CrossRef]
201. Tretyachenko, V.; Vymětal, J.; Bednárová, L.; Kopecký, V.; Hofbauerová, K.; Jindrová, H.; Hubálek, M.; Souček, R.; Konvalinka, J.; Vondrášek, J.; et al. Random Protein Sequences Can Form Defined Secondary Structures and Are Well-Tolerated In Vivo. *Sci. Rep.* **2017**, *7*, 15449. [CrossRef]
202. Merrifield, B. The Nobel Prize in Chemistry 1984. Available online: https://www.nobelprize.org/prizes/chemistry/1984/summary/ (accessed on 29 October 2023).
203. Furka, Á. Forty Years of Combinatorial Technology. *Drug Discov. Today* **2022**, *27*, 103308. [CrossRef]
204. Moult, J. A Decade of CASP: Progress, Bottlenecks and Prognosis in Protein Structure Prediction. *Curr. Opin. Struct. Biol.* **2005**, *15*, 285–289. [CrossRef] [PubMed]
205. Simons, K.T.; Bonneau, R.; Ruczinski, I.; Baker, D. Ab Initio Protein Structure Prediction of CASP III Targets Using ROSETTA. *Proteins* **1999**, *37* (Suppl. S3), 171–176. [CrossRef]
206. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]
207. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585–590. [CrossRef]
208. Krishnan, A. *What Are Academic Disciplines? Some Observations on the Disciplinarity vs. Interdisciplinarity Debate*; Working Paper; National Centre for Research Methods: Bristol, UK, 2009.
209. Chin, J.W. Reprogramming the Genetic Code. *EMBO J.* **2011**, *30*, 2312–2324. [CrossRef] [PubMed]
210. de la Torre, D.; Chin, J.W. Reprogramming the Genetic Code. *Nat. Rev. Genet.* **2020**, *22*, 169–184. [CrossRef]
211. Bada, J.L.; Cleaves, H.J. Ab Initio Simulations and the Miller Prebiotic Synthesis Experiment. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E342. [CrossRef]
212. Atkins, J.F.; Gesteland, R. The 22nd Amino Acid. *Science* **2002**, *296*, 1409–1410. [CrossRef]
213. Rao, V.; Nanjundiah, V.J.B.S. Haldane, Ernst Mayr and the Beanbag Genetics Dispute. *J. Hist. Biol.* **2011**, *44*, 233–281. [CrossRef]
214. Jukes, T.H.; Kimura, M. Evolutionary Constraints and the Neutral Theory. *J. Mol. Evol.* **1984**, *21*, 90–92. [CrossRef]
215. Ghosh, S.; Pal, J.; Cattani, C.; Maji, B.; Bhattacharya, D.K. Protein Sequence Comparison Based on Representation on a Finite Dimensional Unit Hypercube. *J. Biomol. Struct. Dyn.* **2023**, 1–15. [CrossRef]
216. Parker, G.; Smith, J. Optimality theory in evolutionary biology. *Nature* **1990**, *348*, 27–33. [CrossRef]
217. Elner, R.W. The Mechanics of Predation by the Shore Crab, *Carcinus maenas* (L.), on the Edible Mussel, *Mytilus edulis* L. *Oecologia* **1978**, *36*, 333–344. [CrossRef] [PubMed]
218. Szathmáry, E. Four Letters in the Genetic Alphabet: A Frozen Evolutionary Optimum? *Proc. Biol. Sci.* **1991**, *245*, 91–99. [CrossRef]
219. Szathmáry, E. Why Are There Four Letters in the Genetic Alphabet? *Nat. Rev. Genet.* **2003**, *4*, 995–1001. [CrossRef]
220. Freeland, S.J.; Hurst, L.D. The Genetic Code Is One in a Million. *J. Mol. Evol.* **1998**, *47*, 238–248. [CrossRef] [PubMed]
221. Omachi, Y.; Saito, N.; Furusawa, C. Rare-Event Sampling Analysis Uncovers the Fitness Landscape of the Genetic Code. *PLoS Comput. Biol.* **2023**, *19*, e1011034. [CrossRef] [PubMed]
222. Mayer-Bacon, C.; Freeland, S.J. A Broader Context for Understanding Amino Acid Alphabet Optimality. *J. Theor. Biol.* **2021**, *520*, 110661. [CrossRef]
223. Singh, S.; Singh, H.; Tuknait, A.; Chaudhary, K.; Singh, B.; Kumaran, S.; Raghava, G.P.S. PEPstrMOD: Structure Prediction of Peptides Containing Natural, Non-Natural and Modified Residues. *Biol. Direct.* **2015**, *10*, 73. [CrossRef]
224. Lu, Y.; Freeland, S. Testing the Potential for Computational Chemistry to Quantify Biophysical Properties of the Non-Proteinaceous Amino Acids. *Astrobiology* **2006**, *6*, 606–624. [CrossRef]
225. Dobson, C.M. Chemical Space and Biology. *Nature* **2004**, *432*, 824–828. [CrossRef]
226. Lipinski, C.; Hopkins, A. Navigating Chemical Space for Biology and Medicine. *Nature* **2004**, *432*, 855–861. [CrossRef]
227. Lipinski, C.A. Rule of Five in 2015 and beyond: Target and Ligand Structural Limitations, Ligand Chemistry Structure and Drug Discovery Project Decisions. *Adv. Drug Deliv. Rev.* **2016**, *101*, 34–41. [CrossRef]
228. Lins, L.; Brasseur, R. The hydrophobic effect in protein folding. *FASEB J.* **1995**, *9*, 535–540. [CrossRef]
229. Vascon, F.; Gasparotto, M.; Giacomello, M.; Cendron, L.; Bergantino, E.; Filippini, F.; Righetto, I. Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1774–1789. [CrossRef]
230. Ilardo, M.A.; Freeland, S.J. Testing for Adaptive Signatures of Amino Acid Alphabet Evolution Using Chemistry Space. *J. Syst. Chem.* **2014**, *5*, 1. [CrossRef]
231. Cockell, C.S.; Bush, T.; Bryce, C.; Direito, S.; Fox-Powell, M.; Harrison, J.P.; Lammer, H.; Landenmark, H.; Martin-Torres, J.; Nicholson, N.; et al. Habitability: A Review. *Astrobiology* **2016**, *16*, 89–117. [CrossRef] [PubMed]
232. Wong, M.L.; Cleland, C.E.; Arend, D.; Bartlett, S.; Cleaves, H.J.; Demarest, H.; Prabhu, A.; Lunine, J.I.; Hazen, R.M. On the Roles of Function and Selection in Evolving Systems. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2310223120. [CrossRef]
233. Cabrele, C.; Martinek, T.A.; Reiser, O.; Berlicki, B. Peptides Containing β-Amino Acid Patterns: Challenges and Successes in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57*, 9718–9739. [CrossRef] [PubMed]

234. Hickey, J.L.; Sindhikara, D.; Zultanski, S.L.; Schultz, D.M. Beyond 20 in the 21st Century: Prospects and Challenges of Non-Canonical Amino Acids in Peptide Drug Discovery. *ACS Med. Chem. Lett.* **2023**, *14*, 557–565. [CrossRef] [PubMed]

235. Chen, F.; Teniola, O.R.; Laurencin, C.T. Biodegradable Polyphosphazenes for Regenerative Engineering. *J. Mater. Res.* **2022**, *37*, 1417–1428. [CrossRef] [PubMed]