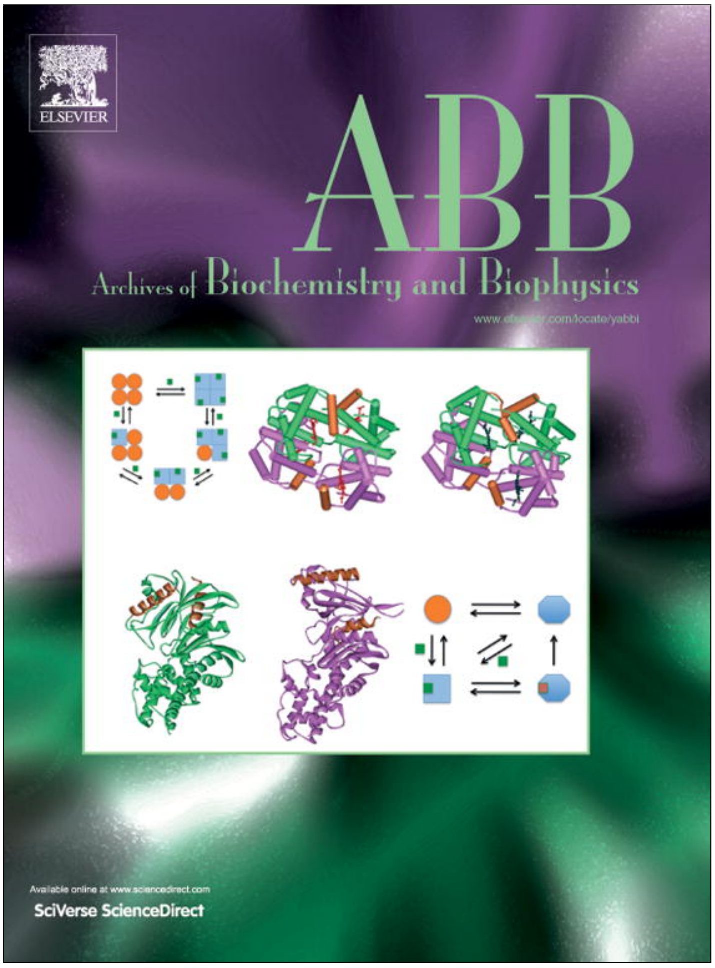
ELSEVIER
ABB
Archives of Biochemistry and Biophysics
www.elsevier.com/locate/yabbi
Available online at www.sciencedirect.com
SciVerse ScienceDirect

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Archives of Biochemistry and Biophysics 526 (2012) 16–21

Contents lists available at SciVerse ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Review

Protein design at the interface of the pre-biotic and biotic worlds

Liam M. Longo, Michael Blaber *

Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL 32306-4300, United States

ARTICLE INFO

Article history:
Received 31 May 2012
Available online 6 July 2012

Keywords:
Biogenesis
Proteogenesis
Pre-biotic
Abiotic
Protein evolution
Protein folding

ABSTRACT

"Proteogenesis" (the origin of proteins) is a likely key event in the unsolved problem of biogenesis (the origin of life). The raw material for the very first proteins comprised the available amino acids produced and accumulated upon the early earth via abiotic chemical and physical processes. A broad consensus is emerging that this pre-biotic set likely comprised Ala, Asp, Glu, Gly, Ile, Leu, Pro, Ser, Thr, and Val. A key question in proteogenesis is whether such abiotically-produced amino acids comprise a "foldable" set. Current knowledge of protein folding identifies properties of complexity, secondary structure propensity, hydrophobic–hydrophilic patterning, core-packing potential, among others, as necessary elements of foldability. None of these requirements excludes the pre-biotic set of amino acids from being a foldable set. Moreover, nucleophile and metal ion/mineral binding capabilities also appear present in the pre-biotic set. Properties of the pre-biotic set, however, likely restrict foldability to the acidophilic/halophilic environment.

Biogenesis is one of the grand unsolved problems in science, and likely requires the combined knowledge of biology, cosmology, chemistry, geology and physics to solve. "Proteogenesis" (the origin of proteins) is a key part of biogenesis principally because proteins are uniquely capable of performing the diverse chemistry necessary to maintain living systems. A consensus opinion is emerging that a limited set of α-amino acids was present on the pre-biotic earth, produced or delivered by abiotic chemical and physical processes. Such pre-biotic amino acids provided the raw material for the very first polypeptides (i.e., proteogenesis) prior to the emergence of any biosynthetic pathway. Advances in the area of protein folding can shed light upon, or frame important unanswered questions about, the "proteogenic potential" of the pre-biotic set of amino acids. For example, does the pre-biotic set of amino acids contain members that can promote formation of the three fundamental types of protein secondary structure (i.e., α-helix, β-strand, and reverse-turn)? Does it contain members that can support hydrophobic–hydrophilic patterning essential for defined structure in aqueous solution? Does it contain the necessary complexity to enable efficient folding pathways of simple globular proteins (i.e., is the pre-biotic set of amino acids intrinsically capable of providing a solution to Levinthal's paradox [1])? In short, *a major unanswered question in proteogenesis is whether the pre-biotic set of amino acids comprises a "foldable" set*. This review is intended to evaluate the consensus pre-biotic set of amino acids in terms of current knowledge of protein folding and design. It will be argued that the pre-biotic set of amino acids likely comprises a foldable set, one that is especially suited to proteogenesis within an acidophile/halophile environment.

General aspects of biogenesis and the pre-biotic amino acids

The earth's hydrosphere is postulated to have formed ~4.3–4.2 Gya in the Hadean period due to out-gassing of water, sulfur dioxide, carbon dioxide, hydrogen and other volatiles from volcanic eruptions [2]. The earliest fossil record of microorganisms (filamentous *archae*) occurs ~3.5 Gya in the early Archean period [3]. Thus, a ~700 million year period likely comprises the processes of biogenesis, emergence of the last universal common ancestor (LUCA)[1], and subsequent evolution of filamentous *archae* (Fig. 1). Organic compounds produced by abiotic processes, which served as the raw material for biogenesis, are postulated to have included the products from atmospheric (lightning) discharge, hydrothermal vent chemistry, as well as organics delivered by comets and meteorites during the late heavy bombardment (LHB; 3.8–4.1 Gya). Sampling of the present atmosphere or hydrothermal vents to identify pre-biotic organics is not feasible due to widespread "contamination" of the present earth by biotic molecules and the dramatic chemical changes such molecules have caused upon both the hydrosphere and mineralogy [4]. However, Miller–Urey spark discharge experiments [5–7] and related experiments simulating hydrothermal vent chemistry [8–10] are experimental approaches to elucidate plausible abiotic chemical synthesis products present in the pre-biotic earth.

* Corresponding author. Address: 1115 West Call St., Tallahassee, FL 32306-4300, United States. Fax: 850 645 5781.
E-mail address: michael.blaber@med.fsu.edu (M. Blaber).

[1] *Abbreviations usesd:* LUCA, last universal common ancestor; LHB, late heavy bombardment.

0003-9861/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.abb.2012.06.009

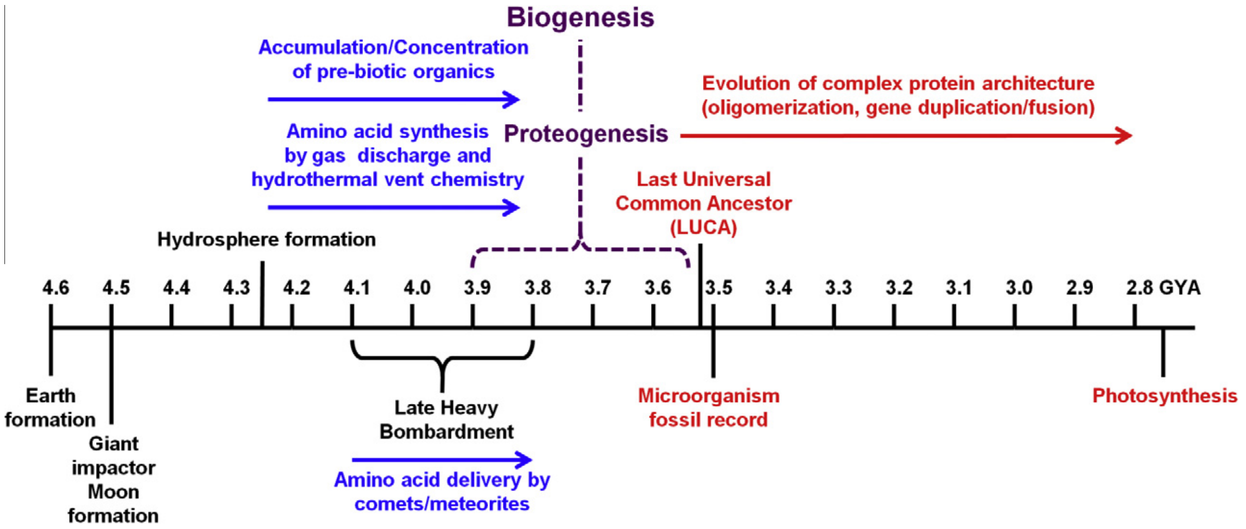
Biogenesis
Accumulation/Concentration of pre-biotic organics
Amino acid synthesis by gas discharge and hydrothermal vent chemistry
Proteogenesis
Evolution of complex protein architecture (oligomerization, gene duplication/fusion)
Last Universal Common Ancestor (LUCA)
Hydrosphere formation
4.6
4.5
4.4
4.3
4.2
4.1
4.0
3.9
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8 GYA
Earth formation
Giant impactor Moon formation
Late Heavy Bombardment
Amino acid delivery by comets/meteorites
Microorganism fossil record
Photosynthesis

Fig. 1. A timeline of key events in the Earth's formation and biogenesis (the gap between the inanimate (blue) and animate (red)). The formation of proteins (proteogenesis) is a key element of biogenesis.

Table 1
Sources of pre-biotic α-amino acids (relative levels) from the analysis of comets/meteorites, Miller–Urey-type spark discharge and hydrothermal syntheses experiments.

	Comet/Meteorite					Spark discharge				Hydrothermal		Summary			
A.A.	Wild 2 [14]	Murchison [13]	Murchison [11]	Murray [11]	Yamato [12]	[15]	[7]	[6]	with FeS/H_2S [107]	[8]	[17]	Comet/ meteorite	Spark discharge	Hydrothermal	Consensus [18,19]
Ala	+	++	++	++	++	+++	+++	+++	+++	++	++	++	+++	++	+++
Cys															
Asp	+	++	+	++	+	++	+	++	+	++		+	+	+	++
Glu	+	++	++	++	++	++	++	++	+++	+		++	++	+	+
Phe								+					+		
Gly	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
His															
Ile		+			+		+		+	+		+	+	+	+
Lys															
Leu		+			+		+		++			+	+		++
Met															
Asn															
Pro		++	+	+		+						+	+		+
Gln															
Arg															
Ser	+				+		++	++	+++	+++	+	+	++	++	+
Thr					+		++					+	+		+
Val		+	+	+	++	++	+++	++	++			+	++		++
Trp															
Tyr															

Furthermore, present day analyses of pristine comet and meteorite samples can provide data on the organic compounds synthesized in deep space and delivered to the earth's surface during the LHB [11–14]. Such diverse analyses identify a limited, but remarkably consistent, set of α-carboxylic and α-amino acids, postulated to be produced via Strecker-type synthesis [15–17], with less evidence for nitrogenous bases. Thus, raw material for biogenesis appears much more supportive of proteogenesis (i.e., a "protein first" view of biogenesis) than nucleogenesis. The analyses of potential sources of abiotic organics identify a consensus set of 10 racemic α-amino acids [18,19] including Ala, Asp, Glu, Gly, Ile, Leu, Pro, Ser, Thr, and Val (using the three-letter code) (Table 1). An early report on the incubation of anhydrous ammonia with hydrogen cyanide at room temperature identified the presence of Lys, Arg, and His α-amino acids (in addition to Asp, Thr, Ser, Glu, Gly, Ala, Val, Ile, and Leu) [20]; however, the identification of basic amino acids has not been observed in the broader range of studies of abiotic synthesis (Table 1). A key question for biogenesis is the proteogenic potential of the pre-biotic set of amino acids.

Does the pre-biotic set of amino acids contain the necessary information to encode a foldable polypeptide?

Protein folding is widely held to occur on a funnel-shaped energy landscape [21] in which the native state (i.e., folded) structure is a limited ensemble of closely related conformations that represent the unique, global minimum of the Gibbs energy surface [22]. Mechanistically, movement along this energy surface is thought to proceed either by "zipping and assembly" [23], "nucleation–condensation" [24,25], or a combination of the two. Cooperativity, the observation that proteins often fold in a concerted "all-or-nothing" fashion, is considered a hallmark of protein folding. Folding cooperativity results from a delicate balance between distal and proximate interactions within the folded structure, thereby suggesting a "well-funneled" energy landscape [26]. Thus, a rigorous definition for the proteogenic potential of a given set of amino acids requires the ability to cooperatively fold into a unique conformation without becoming kinetically trapped.

Sequence complexity

A diversity of interactions is thought to be required to support native-like protein folding [27] and suggests that some minimum alphabet size of amino acids is required for foldability. Proteins with high α-helical content have proven amenable to significant reduction in sequence complexity: the DHP_1 protein ("designed helical protein 1") was constructed using a set of 7 different amino acids [28]; the Sauer group's QLR proteins (based upon Gln, Leu, and Arg residues) contain a total of 9 different amino acids [29]; and a functional form a AroQ chorismate mutase was constructed also using a total of 9 different amino acids [30]. Efforts to simplify α/β proteins have also been successful: Akanuma and coworkers redesigned a functional orotate phosphoribosyltransferase using only 13 amino acids for all 218 positions [31]. Baker and coworkers were able to redesign the SH3 domain using a set of 14 amino acids (but predominately enriched for a 5 amino acid alphabet) [32]. Theoretical approaches support the above results: Romero and coworkers used Shannon's entropy as a formal measure of sequence complexity and concluded that the lower bound for a foldable alphabet size was approximately 10 different types of amino acids [33]. Both Murphy and coworkers [34] and Wang and coworkers [35] assessed the information content of reduced amino acid alphabets by their ability to identify homologs in protein database searches and concluded that ∼10 amino acids is the minimum alphabet size that did not suffer from a significant loss of physicochemical information. Taken together, these results indicate redundancy within the set of 20 common amino acids, and suggest a minimal set of about 7–13 amino acids with appropriate physicochemical properties is sufficient to achieve a foldable set for a wide variety of protein architectures. With 10 amino acids, the pre-biotic set of amino acids lies within the proposed limits of amino acid alphabet complexity for foldability. Therefore, on the grounds of sequence complexity alone, the pre-biotic set of amino acids may potentially comprise a foldable set; the key question being whether the physicochemical properties of the amino acids in the pre-biotic set are appropriate for protein folding, and if so, what would be the general properties of resultant polypeptides?

Secondary structure propensity

As highlighted in Table 2, there exists a broad representation within the pre-biotic set of amino acids of specific residues that favor the formation of each fundamental type of protein secondary structure, including α-helices (Ala, Leu, Glu) [36], as well as the helix capping box residues (Ser, Asp, Thr and Glu) which have been shown to have a profound effect on helix stability and fraying [37], β-strands (Ile, Thr, Val) [38], and reverse-turns (Asp, Gly, Pro) [39,40]. Although the role of secondary structure propensity is considered modest when compared to hydrophobic–hydrophilic patterning and core packing effects, the importance of secondary structure optimization becomes pronounced in proteins with marginal thermodynamic stability or when considered across the protein as a whole [41]. For example, the observation that the pre-biotic set of amino acids contains the consensus sequence (Asx–Pro–Asx–Gly) for Type-1 β-turns (the most common type of reverse-turn) is notable. Turn secondary structure can serve as critical folding nuclei, especially for proteins with high β-sheet content [42–44] and the favorable energy contribution from turns has been shown to be critical for efficient folding [45]. Furthermore, reverse-turn secondary structure is essential to the formation of globular protein architecture since the α-helix and β-strand are essentially linear (i.e., fibrous) structural elements. Therefore, amino acids within the pre-biotic set appear capable of forming each of the fundamental secondary structure elements, including structural features known to nucleate protein folding, as well as enable globular architecture.

Hydrophobic–hydrophilic patterning

Patterning of hydrophobic–hydrophilic residues is a critical determinant of protein structure in aqueous solution. It is well established that maximization of solvent entropy by burial (i.e. desolvation) of hydrophobic side chains is a fundamental driving force for protein folding [46]. Analysis of protein secondary structure elements has revealed characteristic hydrophobic–hydrophilic patterning motifs [47]. Indeed, studies that enforce α-helical patterning schemes but use residues with high β-sheet propensities observe signatures of α-helical structure, suggesting that hydrophobic–hydrophilic patterning is the primary determinant of the general protein architecture [48,49]. Computational efforts echo this result: the success of simplified models of protein folding, most notably the hydrophobic–hydrophilic lattice model pioneered by Dill [50], is dependent on the fact that simplifying protein folding into just hydrophobic and hydrophilic residues can capture a fundamental aspect of the folding reaction. Detailed analyses of amino acid interactions further justify this observation: Wang and Wang observed that the dominant eigenvector - and thus, the major interaction term - of the Miyazawa–Jernigan interaction matrix [51] corresponds roughly to an index of hydrophobicity [52,53]. It appears essential therefore that for any set of amino acids to comprise a foldable set it must support hydrophobic-hydrophilic patterning. As shown in Table 2, residues that are rated as being highly hydrophobic (Ala, Ile, Leu, Val), hydrophilic (Ser, Thr), and charged (Glu, Asp) are all members of the pre-biotic set of amino acids. Therefore, the pre-biotic set appears

Table 2
Physicochemical properties of the pre-biotic set of amino acids (consensus from Table 1).

	Hydrophobic	Hydrophilic	Charged(acidic)	Charged(basic)	α-helix propensity[36]	β-strand propensity[38]	Reverse turn propensity[108,40]	Nucleophile potential
Ala	+				+			
Asp			+				+	+
Glu			+					+
Gly							+	
Ile	+					+		
Leu	+				+			
Pro							+	
Ser		+						+
Thr		+				+		+
Val	+					+		

intrinsically capable of supporting hydrophobic–hydrophilic patterning essential to the design of foldable polypeptides in aqueous solution. It should be noted, however, that reduction of the amino acid alphabet to just two residues, a generic H (hydrophobic) and P (polar, or hydrophilic), appears insufficient to reproduce many of the properties considered to be native-like. Simplified "proteins" generated with these models often have marginally funneled rugged energy landscapes and fail to exhibit signatures of cooperative folding [27,54,55]. Computational studies show that, in general, larger alphabets with a greater diversity of interactions are needed to reproduce native-like folding characteristics [27]. Thus, hydrophobic–hydrophilic patterning is necessary, but not sufficient, to code for foldable proteins.

Hydrophobic core packing

Although the pre-biotic set of amino acids contains the majority of hydrophobic amino acids present in the set of 20 common amino acids, it is notably devoid of the large aromatic residues (i.e., Phe, Trp, and Tyr), suggesting that pre-biotic core-packing efficiency may be reduced (i.e., may contain significant packing defects) compared to typical evolved protein cores [56–58]. Evidence from core repacking and sequence minimization studies, however, shows that core-packing arrangements for foldable proteins can be accomplished in the absence of the large aromatic residues. Walter and coworkers were able to select for a metabolically competent form of AroQ chorismate mutase without inclusion of aromatic residues within the core [30]. Indeed, numerous α-helical proteins, including DHP_1 [28], QLR proteins [29], Rop [59], α4 [60] and phage 434 Cro [61] have been constructed with minimized core complexity, often without aromatic residues. Similar design efforts with β-sheet proteins have enjoyed less success, although some encouraging results have been reported [62,63]. Thus, although there has been significant debate about the importance of structural complementarity within the hydrophobic core [64–66], a general picture has emerged in which structural complementarity is often stabilizing but is not strictly required for the acquisition of complex architecture (and defects in core packing may play an important role in protein dynamics [58,67]). A notable example was provided by Matthews and coworkers [68] in which T4 Lysozyme was found to accommodate 10 methionine substitutions in the hydrophobic core, indicating that the exact identity of a hydrophobic residue is often less critical than hydrophobic–hydrophilic patterning. Successful core repacking of a variety of proteins supports the suggestion that the core has flexible design requirements [69–71]. Concentrated salt solutions, e.g., 1.0–4.0 M, are known to stabilize hydrophobic interactions by an excluded volume effect [72] and solution conditions of high salt can therefore relax the requirement for precise structural complementarity in core packing, thereby promoting effective folding behavior despite core-packing defects. Thus, although proteins comprised of the pre-biotic set of amino acids are not capable of the full range of hydrophobic packing interactions observed in extant proteins, repacking studies suggest that the hydrophobic residues available to pre-biotic protein design will be sufficient to code for native-like proteins, especially in the context of a stabilizing, high salt environment.

Salt bridges and electrostatic potential

The pre-biotic set of amino acids is generally acknowledged as being devoid of basic residues [18,19] (see also Table 1), and *de facto* cannot provide salt bridges (excluding potential interactions with the N-terminus) - potentially the strongest type of non-covalent interaction. Due to the low dielectric of the hydrophobic core region, buried salt bridges can be major contributors to evolved protein stability, while surface exposed salt bridges appear less significant (due to the substantially higher dielectric constant of solvent) [73,74]. Despite the significant contribution to protein stability provided by buried salt bridges, hydrophobic substitutions have been shown theoretically and experimentally to effectively compensate [75,76]. Separate from the lack of salt-bridges, the most significant electrostatic challenge facing pre-biotic protein folding potential is the large negative charge bias that is generated in the absence of basic residues. In particular, at neutral pH, pre-biotic polypeptides would be expected to carry a highly negative charge bias and protein folding would be disfavored due to a sharp increase in like-charge density upon collapse. However, while almost all extant proteomes exhibit a biphasic distribution of pI values (with one peak at pI ~5.0 and another at pI ~10.0) the halophile proteome is unique in containing only a single pI distribution centered at pI ~4.5 [77,78] due to a general lack of basic amino acids. Halophile proteins are soluble and stable in high salt solutions due to selective binding of hydrated salt cations by a high surface density of carboxylate groups (provided by the acidic amino acids Asp and Glu) as well as salt stabilization of hydrophobic core-packing interactions [79]. Moreover, acidophile conditions (e.g. pH 2–5) could further support the folding of pre-biotic proteins via a significant reduction in overall net charge due to substantial protonation of carboxylic acid groups. Thus, the halophile and acidophile environments appear favorable for the folding of polypeptides generated using the pre-biotic set.

Previous pre-biotic protein design efforts

Several experimental studies to elucidate the folding potential of the pre-biotic set of amino acids have been reported, although this is an area of research that is in its infancy. In an effort to assess the foldability of pre-biotic proteins, Doi and coworkers generated a random sequence library utilizing Ala, Gly, Val, Asp, and Glu, a subset of the pre-biotic set of amino acids [18]. Although folded proteins were not detected from this random library, the authors did note that a significant number of the constructs were nonetheless soluble. The buffer conditions were not identified in this report; however, if neutrality and low salt are assumed, such pre-biotic proteins would approximate linear polyanions, which are characteristically observed to be soluble but unstructured. In related studies, Brack and coworkers reported that the addition of salts could induce the formation of secondary structure, both α-helical and β-sheet, in polydisperse peptides composed primarily of Glu and Leu [80,81]. Consequently, these results cannot be used to exclude the possibility of pre-biotic protein folding in an acidophile and/or halophile environment. Chakrabartty and coworkers found that a simple peptide (the "KIA7" peptide) composed primarily of Lys, Ile and Ala, adopts substantial α-helical structure in solution [82–84], and concluded that the pre-biotic set of amino acids had significant design potential. This result supports α-helix forming potential for the pre-biotic set of amino acids. There is some question, however, as to the inclusion of Lys (indeed, any basic amino acid) as a member of the pre-biotic set (see Table 1); moreover, in plants and bacteria Lys is synthesized from Asp and Arg is synthesized from Glu [85]. Therefore, conclusions based upon pre-biotic protein design studies utilizing basic residues or aromatic residues as key structural elements are to be considered with some caution.

Overall, the considerations of sequence complexity, secondary structure propensities, hydrophobic–hydrophilic patterning potential, hydrophobic core-packing potential, and electrostatic properties do not identify any issue that would preclude the pre-biotic set of amino acids from comprising a foldable set; in fact, current evidence points to quite the opposite: *a remarkably broad potential for the pre-biotic set of amino acids as a protein design "tool* *kit*". However, one possible restriction in this regard is the exclusively acidic nature of the pre-biotic set of amino acids, as well as the lack of large aromatic hydrophobic residues, potentially necessitating an acidophile and/or halophile environment to promote protein folding. Of additional note, the pre-biotic set of amino acids contains several residues (Ser, Thr, Asp, and Glu) that can serve as nucleophiles, and can thus provide chemical functionality.

Conclusions, unsolved problems, and future directions

The knowledge gap about biogenesis is being bridged from both the left hand and right hand sides in Fig. 1. From the right hand side, recent "top–down" studies of protein structure have focused upon the role of symmetry in the evolution of complex protein architecture [86]. The majority of fundamental protein superfolds exhibit various forms of rotational symmetry (C*n*; *n* typically 2–8) in their tertiary structure [87] and this has been postulated to be the result of gene duplication and fusion in their evolution from smaller polypeptides [88–90]. Fragmentation studies of proteins with symmetric folds have yielded comparatively simple polypeptides (~35–50 amino acids) with the ability to oligomerize and recapitulate the complex symmetric protein architecture [89,91–96]. These studies tend to support a specific evolutionary model (the "conserved architecture" model [94,86]) whereby organisms having a simple genome can nonetheless achieve complex protein architectures via oligomeric assembly of simple peptide motifs. However, a general lack of examples of such oligomeric assembly in extant organisms suggests that this hypothesized evolutionary stage precedes the LUCA [94,93]; thus, such studies potentially probe evolutionary processes earlier than ~3.5 Gya. The properties of such archaic polypeptides are a matter of conjecture, but key features would likely include an appropriate chemical patterning (i.e., hydrophobic–hydrophilic [97,98]) to enable higher order protein folding. Among the major unsolved problems is how such patterning was achieved and how condensation reactions to create a peptide bond in aqueous solution could have occurred without high-energy amino acid intermediates.

Progress from the left hand side of the biogenesis knowledge gap (as regards proteogenesis) includes the demonstration that high-salt conditions can promote condensation reactions yielding peptide bonds from amino acids (known as "salt-induced peptide formation" or SIPF) [99]. Under high-salt conditions the metal cation can be unsaturated in its H-bond interactions with water, thereby driving condensation reactions (such as peptide bond formation). A related issue is the need to concentrate sparse amino acids in the pre-biotic environment to promote condensation into peptides. While hydrothermal vents have been identified as sources of abiotic organic synthesis reactions, efflux into the oceans would result in dilution. In contrast, since amino acids are non-volatile compounds, evaporation would result in their concentration (and simultaneously, increase salt concentration to promote SIPF). It is feasible that regions of the hydrosphere that produced key pre-biotic organics (e.g., hydrothermal vents), were separate from regions where oligomerization or synthesis into more complex molecules took place (e.g., evaporative lakes). An intimate role for minerals in the process of biogenesis is becoming increasingly compelling. Mineral "evolution" and biogenesis have been postulated as concurrent interrelated events [4]. Mineral surfaces can serve to adsorb and concentrate organic compounds, and can chemically activate peptides and amino acids thus promoting peptide bond formation [100,101]. Furthermore, such adsorption and chemical activity can include stereo-selective binding and deamination of specific amino acid isomers [102]. Mineral crystals are regular (periodic) arrangements of constituent molecules; thus, they can potentially serve as templates for patterning of chemically different amino acids on their surface, and conversely, such peptides can promote specific crystal nucleation of minerals [103]. Thus, in the "grey zone" between the inanimate and animate in biogenesis, specific minerals and early biopolymers might form complexes that template reciprocal propagation. Peptides that selectively bind minerals characteristically utilize the carboxylic acid residues Glu and Asp [103–106]; and intriguingly, certain extant mineral-binding polypeptide sequences are almost exclusively comprised of amino acids from the consensus pre-biotic set [104]. Thus, another property potentially intrinsic to the pre-biotic set of amino acids, and also possibly key for biogenesis, is mineral-binding functionality.

The above discussion of progress in filling in the gaps in understanding of biogenesis suggests tremendous progress is being made, through efforts of scientists in diverse disciplines, in solving this major unsolved problem. *There appears to be no fundamental limitation to protein folding potential that can be identified for the pre-biotic set of amino acids* and apparent environmental restrictions to enable foldability (i.e., high salt/low pH) may actually serve to frame detailed hypotheses regarding key proteogenic processes and environments. Proteogenesis therefore appears to be one of the most promising avenues with which to understand biogenesis, and elucidating the deterministic properties of the pre-biotic set of amino acids in proteogenesis appears feasible and likely have a major impact upon our understanding of the potential for biogenesis elsewhere in the universe.

References

[1] Levinthal C How to fold graciously. In: DeBrunner JTP, Munck E (eds) Mossbauer Spectroscopy in Biological Systems, 1969. University of Illinois Press, pp 22–24.
[2] A.J. Cavosie, J.W. Valley, S.A. Wilde, E.I.M. Facility, Earth Planet. Sci. Lett. 235 (2005) 663–681.
[3] J.W. Schopf, Science 260 (1993) 640–646.
[4] R.M. Hazen, D. Papineau, W. Bleeker, R.T. Downs, J.M. Ferry, T.J. McCoy, D.A. Sverjensky, H. Yang, Am. Mineral. 93 (2008) 1693–1720.
[5] S.L. Miller, Science 117 (1953) 528–529.
[6] A.P. Johnson, H.J. Cleaves, J.P. Dworkin, D.P. Glavin, A. Lazcano, J.L. Bada, Science 322 (2008) 404.
[7] E.T. Parker, H.J. Cleaves, J.P. Dworkin, D.P. Glavin, M. Callahan, A. Aubrey, A. Lazcano, J.L. Bada, Proc. Nat. Acad. Sci. U.S.A. 108 (2011) 5526–5531.
[8] R.J.-C. Hennet, N.G. Holm, M.H. Engel, Naturwissenschaften 79 (1992) 361–365.
[9] N.G. Holm, E. Andersson, Astrobiol. 5 (2005) 444–460.
[10] T.M. McCollom, J.S. Seewald, Chem. Rev. 107 (2007) 382–401.
[11] J.R. Cronin, C.B. Moore, Science 172 (1971) 1327–1329.
[12] A. Shimoyama, C. Ponnamperuma, K. Yanai, Nature 282 (1979) 394–396.
[13] M.H. Engel, B. Nagy, Nature 296 (1982) 837–840.
[14] D.P. Glavin, J.P. Dworkin, S.A. Sandford, Meteorit. Planet. Sci. 43 (2008) 399–413.
[15] Y. Wolman, W.J. Haverland, S.L. Miller, Proc. Nat. Acad. Sci. U.S.A. 69 (1972) 809–811.
[16] N.R. Lerner, E. Peterson, S. Chang, Geochim. Cosmochim. Acta 57 (1993) 4713–4723.
[17] C. Huber, G. Wachtershauser, Science 314 (2006) 630–632.
[18] N. Doi, K. Kakukawa, Y. Oishi, H. Yanagawa, Protein Eng. Des. Sel. 18 (2005) 279–284.
[19] G.D. McDonald, M.C. Storrie-Lombardi, Astrobiol. 10 (2010) 989–1000.
[20] C.N. Matthews, R.E. Moser, Nature 215 (1967) 1230–1234.
[21] P.E. Leopold, M. Montal, J.N. Onuchic, Proc. Nat. Acad. Sci. U.S.A. 89 (1992) 8721–8725.
[22] C.B. Anfinsen, Science 181 (4096) (1973) 223–230.
[23] K.A. Dill, S.B. Ozkan, M.S. Shell, T.R. Weikl, Annu. Rev. Biophys. 37 (2008) 289–316, http://dx.doi.org/10.1146/annurev.biophys.37.092707.153558.
[24] A.R. Fersht, Curr. Opin. Struct. Biol. 7 (1) (1997) 3–9.
[25] A.R. Fersht, Proc. Nat. Acad. Sci. U.S.A. 92 (24) (1995) 10869–10873.
[26] J.N. Onuchic, P.G. Wolynes, Curr. Opin. Struct. Biol. 14 (1) (2004) 70–75, http://dx.doi.org/10.1016/j.sbi.2004.01.009.
[27] P.G. Wolynes, Nat. Struct. Biol. 4 (11) (1997) 871–874.
[28] C.E. Schafmeister, S.L. LaPorte, L.J. Miercke, R.M. Stroud, Nat. Struct. Biol. 4 (12) (1997) 1039–1046.
[29] A.R. Davidson, K.J. Lumb, R.T. Sauer, Nat. Struct. Biol. 2 (1995) 856–864.
[30] K.U. Walter, K. Vamvaca, D. Hilvert, J. Biol. Chem. 280 (2005) 37742–37746.
[31] S. Akanuma, T. Kigawa, S. Yokoyama, Proc. Nat. Acad. Sci. U.S.A. 99 (2002) 13549–13553.

[32] D.S. Riddle, J.V. Santiago, S.T. Bray-Hall, N. Doshi, V.P. Grantcharova, Q. Yi, D. Baker, Nat. Struct. Biol. 4 (1997) 805–809.
[33] P. Romero, Z. Obradovic, A.K. Dunker, FEBS Lett. 462 (3) (1999) 363–367.
[34] L.R. Murphy, A. Wallqvist, R.M. Levy, Protein Eng. 13 (2000) 149–152.
[35] K. Fan, W. Wang, J. Mol. Biol. 328 (2003) 921–926.
[36] C.N. Pace, J.M. Scholtz, Biophys. J . 75 (1998) 422–427.
[37] E.T. Harper, G.D. Rose, Biochemistry 32 (30) (1993) 7605–7609.
[38] A.G. Street, S.L. Mayo, Proc. Nat. Acad. Sci. U.S.A. 96 (16) (1999) 9074–9076.
[39] K. Gunasekaran, C. Ramakrishnan, P. Balaram, Protein Eng. 10 (1997) 1131–1141.
[40] E.G. Hutchinson, J.M. Thornton, Protein Sci. 3 (12) (1994) 2207–2216.
[41] M.H. Cordes, A.R. Davidson, R.T. Sauer, Curr. Opin. Struct. Biol. 6 (1) (1996) 3–10.
[42] M. Jager, H. Nguyen, J.C. Crane, J.W. Kelly, M. Gruebele, J. Mol. Biol. 311 (2) (2001) 373–393, http://dx.doi.org/10.1006/jmbi.2001.4873.
[43] A.M. Marcelino, L.M. Gierasch, Biopolymers 89 (5) (2008) 380–391, http://dx.doi.org/10.1002/bip. 20960.
[44] M. Petrovich, A.L. Jonsson, N. Ferguson, V. Daggett, A.R. Fersht, J. Mol. Biol. 360 (4) (2006) 865–881, http://dx.doi.org/10.1016/j.jmb.2006.05.050.
[45] J. Lee, V.K. Dubey, L.M. Longo, M. Blaber, J. Mol. Biol. 377 (2008) 1251–1264.
[46] K.A. Dill, Biochemistry 29 (1990) 7133–7155.
[47] M.W. West, M.H. Hecht, Protein Sci. 4 (10) (1995) 2032–2039, http://dx.doi.org/10.1002/pro.5560041008.
[48] H. Xiong, B.L. Buckwalter, H.M. Shieh, M.H. Hecht, Proc. Nat. Acad. Sci. U.S.A. 92 (14) (1995) 6349–6353.
[49] G. Bellesia, A.I. Jewett, J.E. Shea, Protein Sci. 19 (1) (2010) 141–154, http://dx.doi.org/10.1002/pro.288.
[50] K.A. Dill, Biochemistry 24 (6) (1985) 1501–1509.
[51] S. Miyazawa, R.L. Jernigan, J. Mol. Biol. 256 (3) (1996) 623–644, http://dx.doi.org/10.1006/jmbi.1996.0114.
[52] J. Wang, W. Wang, Nat. Struct. Biol. 6 (1999) 1033–1038.
[53] H.S. Chan, Nat. Struct. Biol. 6 (11) (1999) 994–996.
[54] N.E. Buchler, R.A. Goldstein, Proteins 34 (1) (1999) 113–124.
[55] E.I. Shakhnovich, Fold Des. 3 (3) (1998) R45–58.
[56] D. Shortle, W.E. Stites, A.K. Meeker, Biochemistry 29 (35) (1990) 8033–8041.
[57] A.E. Eriksson, W.A. Baase, B.W. Matthews, J. Mol. Biol. 229 (3) (1993) 747–769.
[58] A.E. Eriksson, W.A. Baase, X.J. Zhang, D.W. Heinz, M. Blaber, E.P. Baldwin, B.W. Matthews, Science 255 (5041) (1992) 178–183.
[59] M. Munson, R. O'Brien, J.M. Sturtevant, L. Regan, Protein Sci. 3 (11) (1994) 2015–2022, http://dx.doi.org/10.1002/pro.5560031114.
[60] L. Regan, W.F. DeGrado, Science 241 (4868) (1988) 976–978.
[61] J.R. Desjarlais, T.M. Handel, Protein Sci. 4 (10) (1995) 2006–2018, http://dx.doi.org/10.1002/pro.5560041006.
[62] M.T. Jumawid, T. Takahashi, T. Yamazaki, H. Ashigai, H. Mihara, Protein Sci. 18 (2009) 384–398.
[63] G.A. Lazar, J.R. Desjarlais, T.M. Handel, Protein Sci. 6 (6) (1997) 1167–1178, http://dx.doi.org/10.1002/pro.5560060605.
[64] F.H.C. Crick, Acta Crystallogr. 6 (8–9) (1953) 689–697.
[65] W. Kauzmann, Adv. Protein Chem. 14 (1959) 1–63.
[66] C. Chothia, M. Levitt, D. Richardson, J. Mol. Biol. 145 (1) (1981) 215–250.
[67] A. Morton, B.W. Matthews, Biochemistry 34 (1995) 8576–8588.
[68] N.C. Gassner, W.A. Baase, B.W. Matthews, Proc. Nat. Acad. Sci. U.S.A. 93 (1996) 12155–12158.
[69] D.D. Axe, N.W. Foster, A.R. Fersht, Proc. Nat. Acad. Sci. U.S.A. 93 (11) (1996) 5590–5594.
[70] E.P. Baldwin, O. Hajiseyedjavadi, W.A. Baase, B.W. Matthews, Science 262 (1993) 1715–1718.
[71] W.A. Lim, R.T. Sauer, J. Mol. Biol. 219 (1991) 359–376.
[72] M.T. Record Jr., W. Zhang, C.F. Anderson, Adv. Protein Chem. 51 (1998) 281–353.
[73] S. Dao-pin, U. Sauer, H. Nicholson, B.W. Matthews, Biochemistry 30 (1991) 7142–7153.
[74] D. Sali, M. Bycroft, A.R. Fersht, J. Mol. Biol. 220 (1991) 779–788.
[75] Z.S. Hendsch, B. Tidor, Protein Sci. 3 (2) (1994) 211–226, http://dx.doi.org/10.1002/pro.5560030206.
[76] C.D. Waldburger, J.F. Schildbach, R.T. Sauer, Nat. Struct. Biol. 2 (2) (1995) 122–128.
[77] S.P. Kennedy, W.V. Ng, S.L. Salzberg, L. Hood, S. DasSarma, Genome Res. 11 (2001) 1641–1650.
[78] A. Oren, F. Larimer, P. Richardson, A. Lapidus, L.N. Csonka, Extremophiles 9 (2005) 275–279.
[79] H. Eisenberg, M. Mevarech, G. Zaccai, Adv. Protein Chem. 43 (1992) 1–62.
[80] M. Bertrand, A. Brack, Origins Life Evol. Biosphere 27 (5–6) (1997) 585–595.
[81] M. Bertrand, D. Sy, A. Brack, J. Pept. Res. 49 (3) (1997) 269–272.
[82] J. Lopez de la Osa, D.A. Bateman, S. Ho, C. Gonzalez, A. Chakrabartty, D.V. Laurents, Proc. Nat. Acad. Sci. U.S.A. 104 (38) (2007) 14941–14946.
[83] D.W. Frost, C.M. Yip, A. Chakrabartty, Biopolymers 80 (1) (2005) 26–33, http://dx.doi.org/10.1002/bip. 20188.
[84] C.L. Boon, D. Frost, A. Chakrabartty, Biopolymers 76 (3) (2004) 244–257, http://dx.doi.org/10.1002/bip. 20074.
[85] J.H. McClendon, Earth Sci. Rev. 47 (1999) 71–93.
[86] M. Blaber, J. Lee, Curr. Opin. Structural Biol. (2012), in press, http://dx.doi.org/10.1016/j.sbi.2012.05.008.
[87] J.M. Thornton, C.A. Orengo, A.E. Todd, F.M. Pearl, J. Mol. Biol. 293 (2) (1999) 333–342.
[88] A.D. McLachlan, J. Mol. Biol. 64 (2) (1972) 417–437.
[89] D. Lang, R. Thoma, M. Henn-Sax, R. Sterner, M. Wilmanns, Science 289 (5484) (2000) 1546–1550.
[90] J. Soding, A.N. Lupas, BioEssays 25 (2003) 837–846.
[91] I. Yadid, D.S. Tawfik, J. Mol. Biol. 365 (2007) 10–17.
[92] S. Akanuma, A. Yamagishi, J. Mol. Biol. 382 (2008) 458–466.
[93] M. Richter, M. Bosnali, L. Carstensen, T. Seitz, H. Durchschlag, S. Blanquart, R. Merkl, R. Sterner, J. Mol. Biol. 398 (2010) 763–773.
[94] J. Lee, M. Blaber, Proc. Nat. Acad. Sci. U.S.A. 108 (2011) 126–130.
[95] J. Lee, S.I. Blaber, V.K. Dubey, M. Blaber, J. Mol. Biol. 407 (2011) 744–763.
[96] I. Yadid, D.S. Tawfik, Protein Eng. Des. Sel. 24 (1–2) (2011) 185–195.
[97] S. Kamtekar, J.M. Schiffer, H. Xiong, J.M. Babik, M.H. Hecht, Science 262 (1993) 1680–1685.
[98] S. Roy, M.H. Hecht, Biochemistry 39 (2000) 4603–4607.
[99] B.M. Rode, Peptides 20 (1999) 773–786.
[100] J. Bujdak, A. Eder, Y. Yongyai, K. Faybikova, B.M. Rode, J. Inorg. Biochem. 61 (1996) 69–78.
[101] E. Schreiner, N.N. Nair, D. Marx, J. Am. Chem. Soc. 130 (2008) 2768–2770.
[102] B. Siffert, A. Naidja, Clay Miner. 27 (1992) 109–118.
[103] D.B. DeOliveira, R.A. Laursen, J. Am. Chem. Soc. 119 (1997) 10627–10631.
[104] M.E. Bolander, M.F. Young, L.W. Fisher, Y. Yamada, Proc. Nat. Acad. Sci. U.S.A. 85 (1988) 2919–2923.
[105] R.A. Peckauskas, Biopolymers 15 (1976) 569–581.
[106] S. Shimizu, B. Sabsay, A. Veis, J.D. Ostrow, R.V. Rege, L.G. Dawes, J. Clin. Invest. 84 (1989) 1990–1996.
[107] A.D. Keefe, S.L. Miller, G. McDonald, J. Bada, Proc. Nat. Acad. Sci. U.S.A. 92 (1995) 11904–11906.
[108] K. Guruprasad, S. Rajkumar, J. Biosci. 25 (2) (2000) 143–156.