



Prebiotic protein design supports a halophile origin of foldable proteins

Liam M. Longo and Michael Blaber*

Department of Biomedical Sciences, College of Medicine, Florida State University, Tallahassee, FL, USA

*Correspondence: michael.blaber@med.fsu.edu

Edited by:

R. Thane Papke, University of Connecticut, USA

Reviewed by:

Dominique Madern, Institut de Biologie Structurale, France

Johann P. Gogarten, University of Connecticut, USA

Greg Fournier, Massachusetts Institute of Technology, USA

Keywords: halophile, abiogenesis, proteogenesis, prebiotic, protein folding

There are significant challenges in forming testable hypotheses regarding abiogenesis (i.e., the origin of life); for example, the original environment on the early Earth during the process of abiogenesis is a matter of debate [although it was significantly different from the current environment (Oparin, 1952; Hazen et al., 2008)]. Furthermore, the process of abiogenesis occurred over a time scale that is impractical to replicate as a laboratory experiment. More difficult still is the likelihood that current life forms are far removed from the earliest “living systems”—which may well have utilized entirely different initial energetic, biochemical, and “genetic” systems. Despite such difficulties, there are potentially testable hypotheses regarding the origin of *important classes* of biomolecules from abiotic processes. A key biomolecule that emerged in abiogenesis is the foldable polypeptide, which ultimately evolved to provide essentially all of the important biochemical and structural machinery in living systems. Each of the generally-acknowledged abiotic chemical processes present during abiogenesis, including atmospheric spark discharge, hydrothermal vent chemistry, as well as deep-space synthesis and delivery of organic material via comet and asteroid bombardment, can produce a subset of the 20 common α -amino acids [for a summary see Longo and Blaber (2012)]. Such prebiotic amino acids would have provided the raw material for the earliest polypeptides (i.e., “proteogenesis”); thus, the properties of such amino acids and polypeptides are of special interest. As with all things related to abiogenesis,

the set of prebiotic amino acids available for proteogenesis has been a matter of debate; however, a compilation of broad and diverse analyses is arguably converging upon a consensus set of 10 prebiotic α -amino acids (Figure 1).

The alphabet size and chemical properties of the prebiotic α -amino acids are critical parameters as regards the capability to form foldable polypeptides. The rules of protein folding are not fully understood; however, some essential requirements of amino acids to promote folding are known. The tertiary structure of folded proteins is an assemblage of common secondary structure elements—including α -helix, β -strand, and reverse-turns. Thus, a foldable set of amino acids includes representatives with a high propensity to form each of the common secondary structure elements. Additionally, soluble globular proteins typically fold so as to sequester hydrophobic side chains within the protein interior, and this forms a significant energetic contribution to the overall stability of the folded protein; thus, a foldable set of amino acids contains both hydrophobic and hydrophilic members. Finally, functional considerations require that among the amino acids is a representative that can act as a nucleophile, and thereby provide useful chemical activity to an otherwise benign structural scaffold. While folding requirements are demanding, it is clear that the extant set of 20 common α -amino acids is redundant in this regard. Thus, the question of the *minimum* α -amino acid alphabet necessary to enable protein foldability has also been investigated, with a proposed minimum

alphabet size of 9–10 amino acids (Romero et al., 1999; Murphy et al., 2000). Thus, with regard to set size, the prebiotic α -amino acid alphabet is located on the very cusp of foldable potential—a precarious position indeed, as it would provide essentially no redundancy in the requirements for protein foldability. Viewed in these terms, the prebiotic set is *remarkable* in containing all necessary elements for foldability—including high-propensity amino acids for each of the basic types of protein secondary structure, hydrophobic and hydrophilic amino acids, as well as several nucleophilic amino acids [for a discussion of such properties see Longo and Blaber (2012)]. However, the characteristics of a protein comprised of the prebiotic set of α -amino acids present a *stark deviation from the majority of extant proteins*, since aromatic residues, key contributors to the critical hydrophobic effect that drives protein folding, are absent in the prebiotic alphabet. Furthermore, there are no basic amino acids in the prebiotic set (McDonald and Storrie-Lombardi, 2010), thus restricting protein design to exclusively acidic polypeptides—limiting the presence of salt bridge interactions and resulting in acidic pI.

A number of successful studies of simplified protein design have been reported whereby foldable proteins have been constructed from a reduced α -amino acid alphabet, and relevance for proteogenesis have been described. However, such studies have focused exclusively upon achieving minimization of the *alphabet size*, without regard to the *prebiotic relevance* of the included amino acid alphabet.

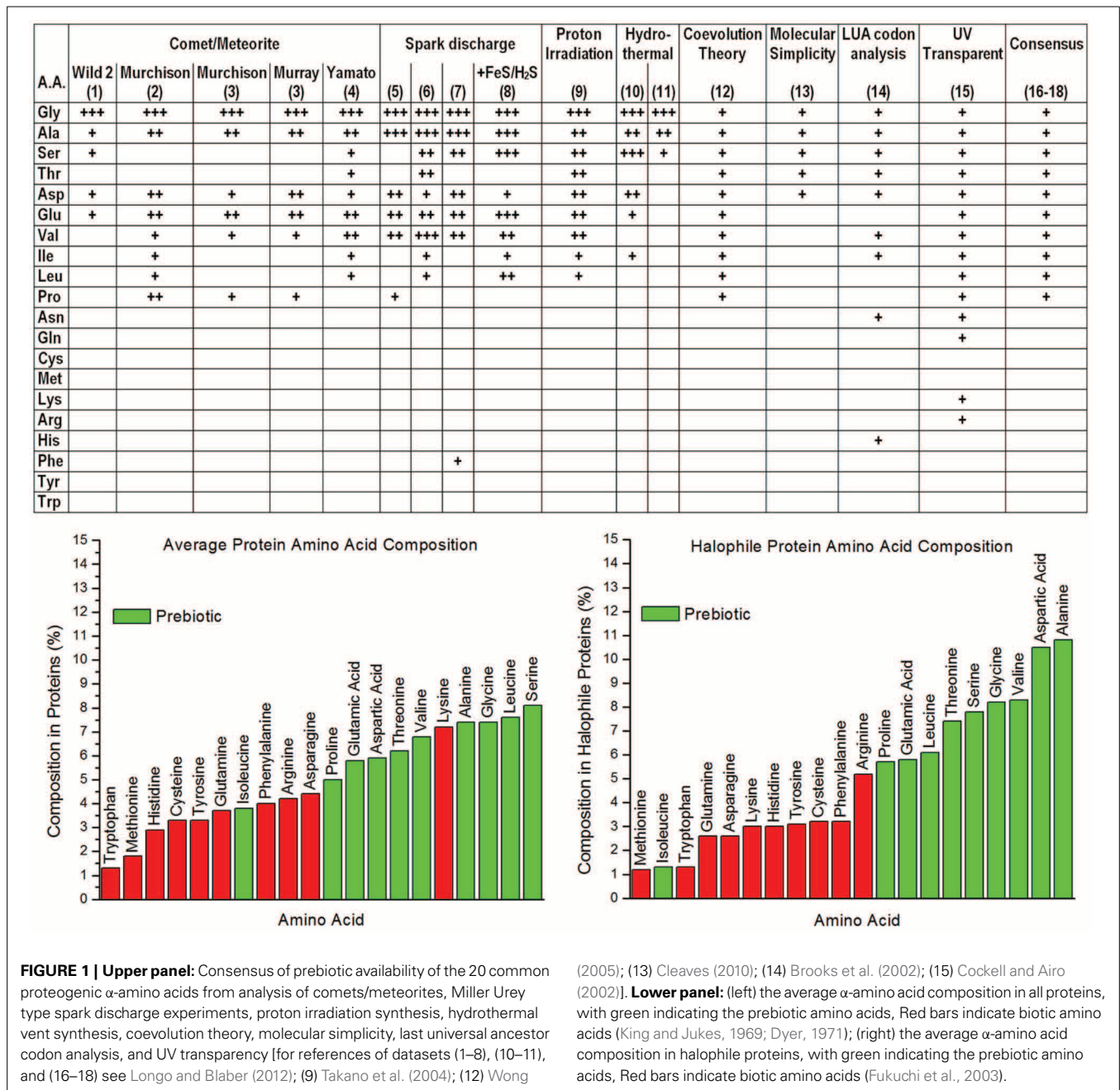


FIGURE 1 | Upper panel: Consensus of prebiotic availability of the 20 common proteogenic α -amino acids from analysis of comets/meteorites, Miller Urey type spark discharge experiments, proton irradiation synthesis, hydrothermal vent synthesis, coevolution theory, molecular simplicity, last universal ancestor codon analysis, and UV transparency [for references of datasets (1–8), (10–11), and (16–18) see Longo and Blaber (2012); (9) Takano et al. (2004); (12) Wong

(2005); (13) Cleaves (2010); (14) Brooks et al. (2002); (15) Cockell and Airo (2002)]. **Lower panel:** (left) the average α -amino acid composition in all proteins, with green indicating the prebiotic amino acids, Red bars indicate biotic amino acids (King and Jukes, 1969; Dyer, 1971); (right) the average α -amino acid composition in halophile proteins, with green indicating the prebiotic amino acids, Red bars indicate biotic amino acids (Fukuchi et al., 2003).

Thus, without exception, such minimal foldable proteins have depended upon critical aromatic amino acids within the core, as well as stabilizing salt bridges (dependent upon basic amino acids), to achieve a stable structure—no minimal protein design has utilized a plausible prebiotic alphabet. Thus, while minimal alphabets can yield foldable polypeptides, the foldable potential of the set of prebiotic amino acids has not been explored with the necessary rigor.

To determine the folding potential of the set of prebiotic α -amino acids our lab evaluated the consequences of enriching for the prebiotic set in a designed β -trefoil protein. The β -trefoil is a common protein architecture that has been the subject of much study as regards its evolutionary emergence from a simple 42-mer peptide motif (Lee and Blaber, 2011; Lee et al., 2011). Two “primitive” versions of the β -trefoil protein were subsequently designed; in primitive version 1 (PV1) the

amino acid alphabet was reduced to 13 unique amino acids with an overall prebiotic composition of 74%; in primitive version 2 (PV2) the amino acid alphabet was reduced to 12 unique amino acids with an overall prebiotic composition of 79%. Notably, PV2 is devoid of any aromatic amino acids. The hydrophobic core of the β -trefoil architecture involves a substantial number of positions (21 total; or ~17% of total positions), and with PV2 this important region of the protein is

entirely prebiotic and comprised of only 3 different amino acids (Leu, Ile, and Val). In addition to reduced core hydrophobicity the PV1 and PV2 proteins exhibit a substantial increase in acidic property (due to the exclusively acidic nature of the prebiotic set of amino acids). A high negative surface charge density is a characteristic feature of halophilic proteins, enabling them to remain soluble in high salt via carboxylate binding of solvated metal cations. Additionally, halophile proteins are characterized as having reduced hydrophobicity, and denaturation under low salt conditions. This property of prebiotic amino acid composition and compatibility with a high salt environment is understood principally in terms of the biophysics of protein solubility in high salt enabled by surface acidic charge that binds hydrated salt cation and the effect of salt in stabilizing the hydrophobic effect. In low salt buffer PV2 is only fractionally folded even at its temperature of maximum stability. However, high salt stabilizes the PV2 protein, shifting its melting temperature into the region of high-mesophile/low-thermophile stability and exhibiting >99% fractional native state; thus, by the criteria of efficient foldability PV2 is an *obligate halophilic protein* (Longo et al., 2013). While high salt also stabilizes PV1, it is not essential for folding stability as the aromatic residues in the core result in efficient hydrophobic packing within the β -trefoil architecture. The crystal structure of the PV2 protein showed a substantial acidic surface charge characteristic of halophile proteins, and distinctly different from the initial mesophile β -trefoil protein from which PV2 was derived. Thus, by several defining criteria the enrichment of prebiotic amino acids in creating the PV2 protein had produced a halophile protein (Longo et al., 2013). The PV2 protein, however, is not 100% prebiotic in its amino acid composition, and further work to achieve an entirely prebiotic foldable protein is needed to support the hypotheses put forth in this opinion article.

A reasonable postulate of abiogenesis is that some residual aspect of the process may still be identifiable in extant organisms. Protein machinery in extant

organisms can be profoundly complex—as can be seen in molecular assemblies such as ATPase, ribosomes, cilia, the proteasome, pyruvate dehydrogenase complex, myosin, and others; however, such complex protein assemblages are built up from *remarkably simple α -amino acids that can be synthesized by abiotic chemical processes*. The amino acid composition of proteins is enriched for the prebiotic set, with 64% of amino acids being prebiotic (**Figure 1**); however, the composition of halophile proteins shows a substantially greater enrichment (72%) of prebiotic amino acids (**Figure 1**). Thus, it is compelling to speculate that this signature is a legacy of abiogenesis—in that *the properties of the halophile environment are highly compatible with foldable polypeptides derived from available prebiotic α -amino acids*. The halophile environment is typically thought of as a curious niche that mesophiles adapted into; however, it has also been proposed as an appropriate environment for abiogenesis and proteogenesis (Dundas, 1998; Rode, 1999). Studies of the folding potential of the set of prebiotic α -amino acids suggest that the halophile environment was a potentially ideal cradle for the proteogenic process in abiogenesis.

REFERENCES

- Brooks, D. J., Fresco, J. R., Lesk, A. M., and Singh, M. (2002). Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. *Mol. Biol. Evol.* 19, 1645–1655. doi: 10.1093/oxfordjournals.molbev.a003988
- Cleaves, H. J. I. (2010). The origin of the biologically coded amino acids. *J. Theor. Biol.* 263, 490–498. doi: 10.1016/j.jtbi.2009.12.014
- Cockell, C. S., and Airo, A. (2002). On the plausibility of a UV transparent biochemistry. *Orig. Life Evol. Biosph.* 32, 255–274. doi: 10.1023/A:1016507810083
- Dundas, I. (1998). Was the environment for primordial life hypersaline? *Extremophiles* 2, 375–377. doi: 10.1007/s007920050081
- Dyer, K. F. (1971). The quiet revolution: a new synthesis of biological knowledge. *J. Biol. Educ.* 5, 15–24. doi: 10.1080/00219266.1971.9653663
- Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M., and Nishikawa, K. (2003). Unique amino acid composition of proteins in halophilic bacteria. *J. Mol. Biol.* 327, 347–357. doi: 10.1016/S0022-2836(03)00150-5
- Hazen, R. M., Papineau, D., Bleeker, W., Downs, R. T., Ferry, J. M., McCoy, T. J., et al. (2008). Mineral evolution. *Am. Mineral.* 93, 1693–1720. doi: 10.2138/am.2008.2955
- King, J. L., and Jukes, T. H. (1969). Non-Darwinian evolution. *Science* 164, 788–798. doi: 10.1126/science.164.3881.788
- Lee, J., and Blaber, M. (2011). Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc. Natl. Acad. Sci. U.S.A.* 108, 126–130. doi: 10.1073/pnas.1015032108
- Lee, J., Blaber, S. I., Dubey, V. K., and Blaber, M. (2011). A polypeptide “building block” for the β -trefoil fold identified by “top-down symmetric deconstruction.” *J. Mol. Biol.* 407, 744–763. doi: 10.1016/j.jmb.2011.02.002
- Longo, L., Lee, J., and Blaber, M. (2013). Simplified protein design biased for pre-biotic amino acids yields a foldable, halophilic protein. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2135–2139. doi: 10.1073/pnas.1219530110
- Longo, L. M., and Blaber, M. (2012). Protein design at the interface of the pre-biotic and biotic worlds. *Arch. Biochem. Biophys.* 526, 16–21. doi: 10.1016/j.abb.2012.06.009
- McDonald, G. D., and Storrie-Lombardi, M. C. (2010). Biochemical constraints in a protobiotic Earth devoid of basic amino acids: the “BAA(-) World.” *Astrobiology* 10, 989–1000. doi: 10.1089/ast.2010.0484
- Murphy, L. R., Wallqvist, A., and Levy, R. M. (2000). Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng.* 13, 149–152. doi: 10.1093/protein/13.3.149
- Oparin, A. I. (1952). *The Origin of Life*. New York, NY: Dover.
- Rode, B. M. (1999). Peptides and the origin of life. *Peptides* 20, 773–786. doi: 10.1016/S0196-9781(99)00062-5
- Romero, P., Obradovic, Z., and Dunker, A. K. (1999). Folding minimal sequences: the lower bound for sequence complexity of globular proteins. *FEBS Lett* 462, 363–367. doi: 10.1016/S0014-5793(99)01557-4
- Takano, Y., Ohashi, A., Kaneko, T., and Kobayashi, K. (2004). Abiotic synthesis of high-molecular-weight organics from an inorganic gas mixture of carbon monoxide, ammonia, and water by 3 MeV proton irradiation. *Appl. Phys. Lett.* 84, 1410–1412. doi: 10.1063/1.1646757
- Wong, J. T.-F. (2005). Coevolution theory of the genetic code at age thirty. *Bioessays* 27, 406–425. doi: 10.1002/bies.20208

Received: 09 November 2013; accepted: 19 December 2013; published online: 06 January 2014.

Citation: Longo LM and Blaber M (2014) Prebiotic protein design supports a halophile origin of foldable proteins. *Front. Microbiol.* 4:418. doi: 10.3389/fmicb.2013.00418

This article was submitted to *Extreme Microbiology*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Longo and Blaber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.