

EXPERIMENTAL TEST OF PLAUSIBLE REDUCED AMINO ACID SETS IN PRIMORDIAL PROTEINS

Satoshi Akanuma

Faculty of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan

akanuma@waseda.jp

(Received: September, 23, 2020 Accepted: October, 28, 2020)

Abstract

Most extant organisms synthesize proteins using the 20 kinds of amino acids that are coded by the standard genetic code table. Currently, these amino acids are incorporated from the environment or synthesized via biosynthetic pathways. However, primordial proteins should have been synthesized only from prebiotically available amino acids. It is not clear, at the moment, whether all 20 coded amino acids were present in sufficient concentrations in the prebiotic environment. Therefore, it could be possible that primordial proteins were synthesized from significantly fewer than 20 kinds of amino acid. Herein, I review previous attempts to synthesize proteins from a limited set of the 20 amino acids and discuss about the possibility of simpler constituents for primordial proteins.

Keywords: Earliest protein, Miller experiment, Murchison meteorite, origin of life, prebiotic amino acids, reduced amino acid alphabet

原始タンパク質における少数種アミノ酸組成の実験的検証

赤沼哲史

早稲田大学人間科学学術院

〒359-1192 埼玉県所沢市三ヶ島 2-579-15

akanuma@waseda.jp

はじめに

生命の起源については多くの証拠が、RNA が遺伝情報を含み、機能分子としても働いたとする RNA ワールド仮説を支持している¹⁻³。リボヌクレオチドの非生物的合成^{4,5}、RNA リボザイムによって触媒される RNA 複製^{6,7}、生体触媒に依存しない RNA 複製⁸、およびクロスキラル RNA ポリメラーゼリボザイム^{9,10}の発見も、進化の最も早い段階では RNA ベースの生命が存在したことを支持している。しかし、地球上のすべての現存生物では、ほとんどの生物学的プロセスにおいてタンパク質が中心的な役割を果たしている。したがって、RNA ワールドの拡張、すなわち、RNA ワールドにタンパク質がどのように出現したのかが次の重要な課題である。

原始環境におけるアミノ酸の可用性

現存するタンパク質の基本的な構造単位は 20 種類の L-アミノ酸 (標準アミノ酸) であり、標準遺伝暗号表によって指定されている。ほとんどの場

合、現存する生物はこれらのアミノ酸を環境から取り込むか、生合成経路を経て細胞内で合成し、タンパク質合成に利用している。しかし、生物によるアミノ酸生合成経路が確立する以前の原始的なタンパク質合成では、原始環境で非生物的に合成されたアミノ酸だけが利用されたはずである^{11,12}。現在のところ、原始地球上に 20 種類の標準アミノ酸がすべて存在していたかは定かではない。例えば、原始地球大気を実験室内で模倣し行われた放電実験 (ミラーの実験) の生成物中には、20 種類の標準アミノ酸のうち 10 種類 (アラニン、アスパラギン酸、グルタミン酸、グリシン、イソロイシン、ロイシン、プロリン、セリン、トレオニン、バリン) が観察された^{13,14}。この実験が行われた当時は、原始大気は水素、メタン、アンモニアなどの還元的ガスを主成分としたと考えられていたが、現在では、水蒸気や二酸化炭素、窒素などを含むより酸化的大気であったと考えられるようになった。酸化的大気中では非生物的に生成される有機化合物の量が大きく減少するが、それでも、ミラーの実験で生成が確認された 10 種類の標準アミノ酸は、非生物的に合成されやすいアミノ酸であると考えられることは可能である。

一方、隕石中の有機物組成は、宇宙における有機物組成の縮図とも言える。すなわち、隕石中から見つかった有機化合物は、宇宙および初期地球に存在した多様な有機化合物を反映していると推測できる¹⁵。マーチソン隕石は 1969 年にオーストラリアに落下した炭素質コンドライトに分類される隕石であり、多くの有機物を含んでいた。特に、アミノ酸については 70 種類以上がこれまでに検出されているが、その多くが非タンパク質性のアミノ酸であり、20 種類の標準アミノ酸のうちマーチソン隕石から検出されたのは、ミラーの実験で生成が確認された 10 種類のアミノ酸からセリンとトレオニンを除いた 8 種類だけであった^{16,17}。ただし、マーチソン隕石にセリンとトレオニンが存在しないのは、隕石が熱や放射線、その他の破壊的な要因に曝されていたためかもしれないとの考えもある¹¹。Tawfik らは、最も古い酵素の最初の機能はリン酸の結合であり、グリシンに加えてセリンとトレオニンがリン酸との結合に重要なアミノ酸であると述べている¹⁸。したがって、標準アミノ酸のうちマーチソン隕石から見つかった 8 種類のアミノ酸にセリンとトレオニンを加えた 10 種類のアミノ酸 (すなわち、ミラーの実験の生成物中から検出された 10 種類のアミノ酸) が、しばしば「プレバイオティックアミノ酸」と呼ばれている。

このように、原始地球では現在のタンパク質合成系において遺伝的にコード化されている 20 種類の標準アミノ酸の一部だけが十分な量で存在していたと現時点では考えられる。したがって、原始的なタンパク質は、20 種類よりもかなり少ないアミノ酸で形成されていたと考えるのが妥当である。

プレバイオティックアミノ酸による原始タンパク質合成の妥当性

原始タンパク質のアミノ酸組成は、いくつかの理論的な研究から推定されてきた。Trifonov は、様々な論文報告を参考に、20 種類の標準アミノ酸が以下の順序でタンパク質合成系に取り込まれたと推定した¹⁹。グリシン/アラニン、バリン/アスパラギン酸、プロリン、セリン、グルタミン酸/ロイシン、トレオニン、アルギニン、アスパラギン、リジン、グルタミン、イソロイシン、システイン、ヒスチジン、フェニルアラニン、メチオニン、チロシン、トリプトファン。前述したミラーの実験で生成が確認された 10 種類のアミノ酸のうち、イソロイシンを除く 9 種類のアミノ酸がこの順序の前半を占めている。

Keller らは、プレバイオティックアミノ酸が自己組織化した脂肪酸膜に結合し、膜の安定化をもたらすことを明らかにした²⁰。さらに、これらのアミノ酸が細胞膜表面に局所的に存在することによって、タンパク質の形成を助けた可能性があることを提唱した。Gáspári らは、いくつかのアミノ酸サブセットをコードするランダムな DNA 配列を翻訳して合成される仮想タンパク質の構造的嗜好性を解析した²¹。その結果、10 種類のプレバイオティックアミノ酸をコードした DNA 配列からつくられた仮想タンパク質が、現存タンパク質と類似した構造特性を持つタンパク質をコードできると予測された。Solis は、シングルドメインタンパク質の大規模なデータセットを用いたコンピュータ解析により、10 種類のプレバイオティックアミノ酸からなる配列が、他のすべての組み合わせの 10 種類のアミノ酸からなる配列と比べて、タンパク質の骨格構造形成に有利であることを明らかにした²²。

このような研究から、20 種類の標準アミノ酸が揃う以前には、前述した 10 種類のプレバイオティックアミノ酸から構成された原始タンパク質が機能していたのではないかと推測された。

タンパク質工学実験による少数種類アミノ酸組成の探索

過去 30 年の間に、20 種類に満たない少数種類のアミノ酸だけからタンパク質を再構成、あるいは、新規に合成する試みがおこなわれてきた (表 1)。1980 年代の終わりから 1990 年代にかけて、特定の機能を持たないヘリカルバンドル構造が、5~7 種類のアミノ酸だけから新規合成できることが示された²³⁻²⁵。さらに、比較的小さなタンパク質を持つ生物学的機能が、アミノ酸の種類数を制限しても保持されることも示された。例えば、機能を保持した SH3 ドメイン変異体が、その 70% のアミノ酸部位を 5 種類のアミノ酸だけが占めるように再

構成された²⁶。さらに、ヘリカルな構造を持つ AroQ コリスマ酸ムターゼが 9 種類のアミノ酸から再構成されたが、触媒活性を保持していた²⁷。さらに、9 種類のアミノ酸から再構成された AroQ コリスマ酸ムターゼに 2 種類のアミノ酸を追加することによって、安定性および触媒活性が改善された²⁸。前述のタンパク質と比べると比較的大きく、より複雑なトポロジーを持つタンパク質を用いた研究としては、213 残基からなる大腸菌由来オロト酸ホスホリボシルトランスフェラーゼを、代謝機能を失わせることなく 13 アミノ酸種から再構成した例が挙げられる²⁹。

これらの研究は、現在の 20 種類の標準アミノ酸よりもはるかに少ない種類のアミノ酸セットから安定なタンパク質構造、あるいは、機能を保持したタンパク質を合成できることを明確に示している。しかし、これらの研究のすべてにおいて、少数種類アミノ酸セットは任意に選択されたものである。一方、Blaber らは、ミラーの実験で生成された 10 種類のアミノ酸 (プレバイオティックアミノ酸) に偏って構築された β -トレフォイル蛋白質について報告している³⁰。彼らが再構成した β -トレフォイル蛋白質のうち、最も単純化されたアミノ酸組成を持つ変異体では、全体の 79% のアミノ酸部位がプレバイオティックアミノ酸で占められ、高塩濃度下で安定な立体構造を形成した。この結果から、彼らは高塩濃度環境がプレバイオティックアミノ酸によって構成されたアミノ酸配列の折り畳みに寄与すると主張した。

祖先型配列から遡る原始アミノ酸組成の探索

近年急速に増えている生物のゲノムデータを比較することで、過去のタンパク質配列を予測し、復元することが可能となった。著者らの研究グループでは、祖先生物のヌクレオシド二リン酸キナーゼ (NDK) 配列を複数推定し、復元した。復元したすべての祖先 NDK は高い耐熱性を持ち、全生物共通祖先が超好熱菌であったことを示す最初の実験による証拠となった^{31,32}。復元した NDK の一つである Arc1 は変性温度が 114°C で、現存生物が持つ NDK と同程度の大きさのリン酸基転移活性を持つ。そこで、著者らは Arc1 から系統的にアミノ酸を欠損させる実験をおこなった。Arc1 はシステインを欠いており 19 アミノ酸種から構成される。Arc1 に含まれるどれか 1 種類のアミノ酸を他のアミノ酸に置換し欠損させた、すなわち 18 種類のアミノ酸から構成された Arc1 変異体を 19 個作製し、その耐熱性と触媒活性を調べた。その結果から、アラニン、フェニルアラニン、イソロイシン、リジン、ロイシン、メチオニン、グルタミン、セリン、トレオニン、トリプトファンを欠損が容易なアミノ酸種、その他のアミノ酸を欠損が困難なアミノ酸種であると判別した。さらに、欠損が容易と判定したアミノ酸を系統的かつ網羅的に複数種類同時に欠損させることによって、異なる 13 種類のアミノ酸だけから再構成された 2 つの変異体 Arc1-13FI と Arc1-13KS が 70°C を超える変性温度と、元の Arc1 と比べると著しく小さいものの、検出可

表 1. アミノ酸組成が単純化された代表的なタンパク質

タンパク質	アミノ酸の種類と割合 ^a	特徴	文献
4ヘリックスバンドル	全 108 残基が 7 アミノ酸 (A, E, G, K, L, Q, S)	X-線結晶構造解析により構造を決定	(25)
SH3 ドメイン	全長の 70%が 5 アミノ酸 (A, E, G, I, K)	ペプチド結合活性を保持	(26)
コリスミ酸ムターゼ			
9-CM	全 92 残基が 9 アミノ酸 (D, E, N, K, F, I, L, M, R)	触媒活性あり	(27)
11-CM	9-CM に T と V を追加	9-CM と比べ触媒活性と安定性が向上	(28)
オロト酸ホスホリボシルトランスフェラーゼ	全 212 残基が 13 アミノ酸 (A, D, E, F, G, K, L, P, R, S, T, V, Y)	188 残基を 9 アミノ酸 (A, D, G, L, P, R, T, V, Y) が占める。触媒活性あり	(29)
β -トレフォイル蛋白質	全 126 残基が 12 アミノ酸 (D, E, G, I, L, N, P, Q, R, S, T, V)	全長の 79%をプレバイオティックアミノ酸 (D, E, G, I, L, P, S, T, V) が占める。高塩濃度存在下で安定な立体構造を形成	(30)
祖先型 NDK 改変体			
Arc1-13KS	全 138 残基が 13 アミノ酸 (A, D, E, F, G, H, I, L, N, P, R, V, Y)	変性中点温度 74°C、検出可能な大きさの触媒活性あり	(33)
Arc1-13FI	全 138 残基が 13 アミノ酸 (A, D, E, G, H, K, L, N, P, R, S, V, Y)	変性中点温度 74°C、検出可能な大きさの触媒活性あり	(33)
Arc1-10KS	全 138 残基が 10 アミノ酸 (A, D, E, F, G, I, L, P, R, V)	変性中点温度 85°C、触媒活性は検出されなかった	(33)
Arc1-10FI	全 138 残基が 10 アミノ酸 (A, D, E, G, K, L, P, R, S, V)	大腸菌を宿主とした発現系では発現せず	(34)
Arc1-10KS+KHNY	Arc1-10KS の活性中心に K, H, N, Y を追加	変性中点温度 85°C、検出可能な大きさの触媒活性あり	(34)
Arc1-10FI+HNY	Arc1-10FI の活性中心に H, N, Y を追加	変性中点温度 76°C、検出可能な大きさの触媒活性あり	(34)

^aアミノ酸は 1 文字表記で表している。

能なりン酸基転移活性を示すことを明らかにした (表 1)³³。すなわち、13 アミノ酸種類あれば、活性を持った安定な NDK を再構成できることを明らかにした。さらに Arc1-13FI と Arc1-13KS からヒスチジン、アスパラギン、チロシンを欠損させた Arc1-10FI と Arc1-10KS を合成した。Arc1-10FI は大腸菌を宿主とした標準的なタンパク質発現系では合成が確認されなかった³⁴。一方、Arc1-10KS は検出可能な触媒活性が見られなかったが、Arc1-13KS と比べて 10°C 以上耐熱性が向上した (表 1)³³。この結果から、ヒスチジン、アスパラギン、チロシンは触媒活性に重要であるが、安定性にはそれほど寄与しないことが分かった。Arc1-10KS では、全アミノ酸部位の 81%をミラーの実験で生成された 10 種類のプレバイオティックアミノ酸が占めているが、Arc1-10KS は低塩濃度 (50 mM KCl) の緩衝液中でも、前述したプレバイオティックアミノ酸に偏ったアミノ酸組成を持つよう再構成された β -トレフォイル蛋白質よりもはるかに安定であった。したがって、高塩濃度はプレバイオティックアミノ酸が多数を占めるタンパク質の安定な立体構造形成に必ずしも重要ではなさそうである。

次に、触媒活性に重要な活性中心の 3 つのアミノ酸を Arc1-10FI に回復させた Arc1-10FI+HNY と、

活性中心の 4 つのアミノ酸を Arc1-10KS に回復させた Arc1-10KS+KHNY を再構成したところ、どちらも検出可能な大きさのリン酸基転移活性を示した (表 1)³⁴。Arc1-10FI+HNY と Arc1-10KS+KHNY では、プレバイオティックアミノ酸がそれぞれ 82%と 81%を占めており、プレバイオティックアミノ酸に偏ったアミノ酸組成を持つタンパク質であっても、小さいながらも触媒活性を持つ安定な立体構想を形成できることを示した。

今後の課題

現在のところ、原始地球環境にタンパク質を合成するのに十分な濃度で存在していたと考えられているのは、ミラーの実験で生成が確認された、プレバイオティックアミノ酸としてしばしば参照される 10 種類のアミノ酸だけである。しかし、他のアミノ酸も前生物環境で何らかの方法で合成されていた可能性を排除することはできない。実際、Sutherland らはミラーの実験の生成物とは異なる、非生物的に合成された可能性のあるアミノ酸セットを提唱している³⁵。また、隕石中からは現在のタンパク質合成に使われる標準アミノ酸以外のアミノ酸が多く検出されており、このような非標準アミノ酸が原始タンパク質で利用されていた可能性

もある。例えば、プレバイオティックアミノ酸の特徴の一つに塩基性アミノ酸を含まないことが挙げられるが³⁶、Tawfikらは最初の核酸結合タンパク質は、現存タンパク質では使われていない塩基性アミノ酸であるオルニチンを含む、短くて単純な配列から生じた可能性を指摘している³⁷。

10種類のアミノ酸からなるプレバイオティックアミノ酸セットのもう一つの特徴が、現存タンパク質の触媒残基として高頻度で見られるヒスチジンが含まれていないことである。この事実から、生命の初期には、アミノ酸は、安定な立体構造を形成する可溶性タンパク質の生成を可能にするように選択され、一方で、RNAなどの他の分子が触媒機能に関与していたことが提案されている^{12,33}。Francisは、原始遺伝暗号表は既存のものよりもはるかに単純であり、その後、タンパク質の機能を向上させるために新しいアミノ酸が遺伝暗号表に追加されたのではないかと指摘している³⁸。今後は、RNAワールドに誕生した最初のタンパク質のアミノ酸組成だけでなく、最初のタンパク質が獲得した機能についても明らかにする必要がある。

謝辞

ここで紹介した著者の研究は、基盤研究 (B) (17H03716)、挑戦的研究 (萌芽) (19K21903)、NINS アストロバイオロジーセンター・サテライト研究 (AB302004、AB312004、AB022003) などの支援によって行われました。

引用文献

- Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157 (1982).
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930 (2000).
- Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239-242 (2009).
- Becker, S. et al. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76-82 (2019).
- Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209-212 (2011).
- Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci USA* **113**, 9786-9791 (2016).
- Kim, S. C., O'Flaherty, D. K., Zhou, L., Lelyveld, V. S. & Szostak, J. W. Inosine, but none of the 8-oxo-purines, is a plausible component of a primordial version of RNA. *Proc Natl Acad Sci USA* **115**, 13318-13323 (2018).
- Sczepanski, J. T. & Joyce, G. F. A cross-chiral RNA polymerase ribozyme. *Nature* **515**, 440-442 (2014).
- Tjhung, K. F., Sczepanski, J. T., Murtfeldt, E. R. & Joyce, G. F. RNA-Catalyzed Cross-Chiral Polymerization of RNA. *J Am Chem Soc*, doi:10.1021/jacs.0c05635 (2020).
- Cleaves, H. J., 2nd. The origin of the biologically coded amino acids. *J Theor Biol* **263**, 490-498 (2010).
- Doig, A. J. Frozen, but no accident - why the 20 standard amino acids were selected. *FEBS J* **284**, 1296-1305 (2017).
- Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528-529 (1953).
- Weber, A. L. & Miller, S. L. Reasons for the occurrence of the twenty coded protein amino acids. *J Mol Evol* **17**, 273-284 (1981).
- 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* **41**, 5459-5472 (2012).
- Cronin, J. R. & Pizzarello, S. Amino acids in meteorites. *Adv Space Res* **3**, 5-18 (1983).
- Cronin, J. R. Origin of organic compounds in carbonaceous chondrites. *Adv Space Res* **9**, 59-64 (1989).
- Longo, L. M., Petrović, D., Kamerlin, S. C. L. & Tawfik, D. S. Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes. *Proc Natl Acad Sci USA* **117**, 5310-5318 (2020).
- Trifonov, E. N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **261**, 139-151 (2000).
- Cornell, C. E. et al. Prebiotic amino acids bind to and stabilize prebiotic fatty acid membranes. *Proc Natl Acad Sci USA* **116**, 17239-17244 (2019).
- Angyan, A. F., Ortutay, C. & Gaspari, Z. Are proposed early genetic codes capable of encoding viable proteins? *J Mol Evol* **78**, 263-274 (2014).
- Solis, A. D. Reduced alphabet of prebiotic amino acids optimally encodes the conformational space of diverse extant protein folds. *BMC Evol Biol* **19**, 158 (2019).
- Regan, L. & DeGrado, W. F. Characterization of a helical protein designed from first principles. *Science* **241**, 976-978 (1988).
- Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680-1685 (1993).
- Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. & Stroud, R. M. A designed four helix bundle protein with native-like structure. *Nat Struct Biol* **4**, 1039-1046 (1997).
- Riddle, D. S. et al. Functional rapidly folding

- proteins from simplified amino acid sequences. *Nat Struct Biol* **4**, 805-809 (1997).
27. Walter, K. U., Vamvaca, K. & Hilvert, D. An active enzyme constructed from a 9-amino acid alphabet. *J Biol Chem* **280**, 37742-37746 (2005).
 28. Muller, M. M. et al. Directed evolution of a model primordial enzyme provides insights into the development of the genetic code. *PLoS Genet* **9**, e1003187 (2013).
 29. Akanuma, S., Kigawa, T. & Yokoyama, S. Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc Natl Acad Sci USA* **99**, 13549-13553 (2002).
 30. 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* **110**, 2135-2139 (2013).
 31. Akanuma, S. et al. Experimental evidence for the thermophilicity of ancestral life. *Proc Natl Acad Sci USA* **110**, 11067-11072 (2013).
 32. Akanuma, S., Yokobori, S., Nakajima, Y., Bessho, M. & Yamagishi, A. Robustness of predictions of extremely thermally stable proteins in ancient organisms. *Evolution* **69**, 2954-2962 (2015).
 33. Shibue, R. et al. Comprehensive reduction of amino acid set in a protein suggests the importance of prebiotic amino acids for stable proteins. *Sci Rep* **8**, 1227 (2018).
 34. Kimura, M. & Akanuma, S. Reconstruction and Characterization of Thermally Stable and Catalytically Active Proteins Comprising an Alphabet of ~13 Amino Acids. *J Mol Evol* **88**, 372-381 (2020).
 35. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat Chem* **7**, 301-307 (2015).
 36. McDonald, G. D. & Storrie-Lombardi, M. C. Biochemical constraints in a protobiotic earth devoid of basic amino acids: the "BAA(-) world". *Astrobiology* **10**, 989-1000 (2010).
 37. Longo, L. M. et al. Primordial emergence of a nucleic acid-binding protein via phase separation and statistical ornithine-to-arginine conversion. *Proc Natl Acad Sci USA* **117**, 15731-15739 (2020).
 38. Francis, B. R. Evolution of the genetic code by incorporation of amino acids that improved or changed protein function. *J Mol Evol* **77**, 134-158 (2013).