

REDUCED AMINO ACID SET PROTEINS SUGGEST A ROLE OF PREBIOTIC AMINO ACIDS IN PRIMITIVE PROTEIN SYNTHESIS

Rei Shibue¹, Satoshi Akanuma^{1,2}

¹Faculty of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan

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

Tel, +81-4-2947-6727; fax, +81-4-2947-6811;

email, akanuma@waseda.jp

(Received: July, 26, 2018 Accepted: September, 2, 2018)

Abstract

Contemporary organisms commonly use the genetically encoded 20 amino acids to synthesize proteins. However, earlier protein synthesis was plausibly much simpler and utilized a subset of the standard proteinogenic amino acids. We reconstructed ancestral nucleoside diphosphate kinases (NDKs) by combining the computational inference of ancestral sequence and a whole gene synthesis technique. One of the reconstructed NDKs, Arc1, is an extremely thermally stable protein with the unfolding midpoint temperature of 114°C and the magnitude of its specific activity is comparable with those of extant NDK. Arc1 lacks cysteine and therefore consists of 19 amino acid letters. To address simpler constituents for primitive proteins, we comprehensively reduce the size of the amino acid alphabet in Arc1 to reduced sets. We found that many but not all of the reduced-set amino acids were consistent with those plausibly abundant in primitive environment. We also found that the remaining inconsistent amino acids were important for activity but not for stability. Our results are compatible with the hypothesis that amino acids available in the prebiotic environment were first used for creating stable protein structures and their functions were later improved by recruiting other amino acids.

Keywords: Ancestral protein, origin of life, origin of translation, primitive amino acid repertoire, primitive proteins, protein simplification, reduced amino acid alphabet

背景

地球上で生息するすべての生物は、核酸重合体を遺伝物質として持ち、ほとんどの場合、タンパク質が機能分子として細胞内外で働いている。DNA や RNA のような核酸の重合体は、タンパク質のアミノ酸配列を指定する情報を持ち、タンパク質は DNA の複製や転写による RNA 合成に関わっている。したがって、生命の起源に関連して、「核酸が先か？タンパク質が先か？」という問いの解明が長年の課題であった。しかし、非生物 RNA 重合 [1,2]、リボザイムによって触媒された RNA 複製 [3]、自身の鏡像異性体を合成するリボザイムの発見により [4]、タンパク質の誕生に先じた RNA ワールド仮説が支持を得てきた。次の課題となるのが RNA ワールドにどのようにしてタンパク質が誕生したかである。

初期のタンパク質合成系に使われたアミノ酸種類についての議論は、遺伝暗号の起源と進化に関する議論と密接に関連する。DNA の二重らせんモデルの提唱者の一人である Crick は、現在の標準遺伝暗号表の成立は偶然に依ることが大きいとする「偶然凍結説」を唱えた [5]。それに対して、標準遺伝暗号表の成立に合理的な説明を与えようとしたいくつかの理論が提唱されてきたが、それらは必ずしも一致しない [6-10]。しかし、それらの理論は共通して、初期の遺伝暗号表は 20 種類よりも少ないアミノ酸だけを指定していて、進化の過程で徐々に指定されるアミノ酸が増加して、最終的に現在の 20 種類に至ったと説明している。初期の地球環境を模した環境を実験室内で再現し、放電実験をおこなったミラーの実験における生成物や、隕石の分析から示されたアミノ酸組成から、標準遺伝暗号表で指定された 20 種のアミノ酸の一部だけが前生物的環境下でも存在していたと推定されている [11-16]。さらに、どのようにして 20 種類のアミノ酸が選ばれたのかについても議論がおこなわれてきた [17-20]。

しかし、たとえ初期のタンパク質が 20 種類未満のアミノ酸だけから合成されていたとしても、そのタンパク質は、何らかの機能を示す程度には安定な構造を形成したはずである。現在までも、必ずしも 20 種類のアミノ酸が揃っていても、天然様立体構造を形成すること、あるいは、触媒機能を持ったタンパク質を合成することが可能であることが示されてきた [21-24]。しかし、これらの研究で少数種アミノ酸から合成されたタンパク質の多くは触媒機能を持たなかったり、安定な立体構造を形成しなかったりした。さらに、ほとんどの場合、少数種アミノ酸は恣意的に選ばれてきた。

現在、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。筆者らは、この技術を利用して祖先生物のヌクレオシド二リン酸キナーゼ (NDK) 配列を推定・復元した [25,26]。復元した祖先型 NDK の一つである Arc1 は、変性中点温度が 114°C と極端に熱安定なタンパク質であり、その結晶構造から同一のサブユニット 6 個が会合した六量体構造が明らかになっている [25]。Arc1 はシステインを含まず、19 種類のアミノ酸から構成されている。本稿では、系統的に選択された少数種アミノ酸からタンパク質を合成した筆者らの最新の研究 [27] について紹介する。すなわち、筆者らは、安定な立体構造形成能や触媒活性を失わせることなく、Arc1 を構成するアミノ酸種を 20

種類から徐々に減らしていくことによって、安定で触媒活性を持つNDKを再構成するのに必要な少数アミノ酸種類を明らかにした。さらに、地球化学的な研究から前生物環境下に存在したと推定されているアミノ酸種と比較した。

アミノ酸種を欠損させた Arc1 改変体の安定性と触媒活性

筆者らは、最初にシステインを除く 19 種類のアミノ酸のうちの 1 種を欠損させた 19 種類の Arc1 改変体を再構成した。Arc1 はシステインを持たないので、いずれの改変体も 18 アミノ酸種で構成されている。アミノ酸を欠損する際には、現存の 309 個の NDK 配列の多重アライメントを参考にし、欠損させるアミノ酸をその座位に最も多く見られる別のアミノ酸種に置換した。筆者らは、ある座位に最も高頻度で出現するアミノ酸種は、NDK の進化においてその座位に最も適しているアミノ酸として選択されてきたと考えた。加えて、高頻度で出現するアミノ酸は、タンパク質の安定性にも寄与することが多いことも知られている [28]。元のアミノ酸が多重アライメント中に含まれる現存 NDK において完全に保存されている場合は、側鎖の物理化学的性質が最も似ているアミノ酸に置換した。メチオニンの欠損に際しては、N 末端残基は考慮しなかった。Table 1 に示したように、グリシンまたはグルタミン酸を欠損させた改変体（すなわち、システインとグリシンを持たないか、あるいは、システインとグルタミン酸を持たない 18 種類のアミノ酸で構成された改変体）は可溶性タンパク質として得られなかったため、それ以上の解析をおこなわなかった。恐らく、グリシンとグルタミン酸は Arc1 のフォールディングや熱力学的安定性に大きく寄与している、安定な立体構造形成に重要なアミノ酸であると思われる。残りの 17 改変体の熱安定性を解析するため、それぞれの改変体について、タンパク質中の α ヘリックス含量の指標となる 222 nm の楕円率の温度変化を測定した。さらに、観測された熱変性曲線から、変性中点温度を求めた。それぞれの改変体について、70°C における比活性も求めた（ただし、バリンを欠損させた改変体の比活性は 50°C で測定した）。NDK はヌクレオシド三リン酸の γ リン酸基をヌクレオシド二リン酸に転移する反応を触媒する。Table 1 にまとめたように、いくつかのアミノ酸種を欠損させると Arc1 の安定性や比活性が大きく低下した。特にバリンの欠損は、安定性と活性の両方に影響した。アスパラギン酸、ヒスチジン、アスパラギン、アルギニンのうちの 1 種を欠損させても安定性には大きな変化は見られなかったが、比活性は大きく低下した。プロリンまたはチロシンの欠損は、Arc1 の触媒活性を低下させた。それに対して、残りの 10 アミノ酸種のうちの 1 種を欠損させても、Arc1 の安定性や活性に大きくは影響しなかった。次に、これらの 10 アミノ酸種の中から複数種のアミノ酸を欠損させた改変体を作製することにした。

複数アミノ酸種を欠損させた改変体の作製

筆者らは、欠損させても Arc1 の安定性と比活性にあまり影響しなかった 10 アミノ酸種のうち、非極性アミノ酸であるアラニン、フェニルアラニン、イソロイシン、ロイシン、メチオニンを対象として複数アミノ酸種を同時に欠損させた改変体の作製と解析をおこなった。非極性側鎖は NDK の触媒

活性に直接は関わっていないと考えたからである。アラニン、フェニルアラニン、イソロイシン、メチオニンのうちの 2 種を同時に欠損させた改変体は、どの組み合わせで欠損させても安定性と比活性を大きくは劣化させなかった (Table 2)。しかし、ロイシンを含む組み合わせで欠損させた場合は、特に比活性への影響が大きい傾向が見られた (Table 2)。さらに、アラニン、フェニルアラニン、イソロイシン、メチオニンのうちの 3 種を同時に欠損させた改変体を作製し、それらの変性温度と比活性を求めたところ、フェニルアラニン、イソロイシン、メチオニンを欠損させた Arc1-16 は変性中点温度が 78°C と比較的高い熱安定性を維持し、比活性も 260 U/mg と大きくは低下しなかった (Table 2)。

筆者らはさらに、Arc1-16 からグルタミン、トレオニン、トリプトファンを欠損させた Arc1-13 を作製し、解析した。その結果、Arc1-13 は変性中点温度が 74°C で、50°C における比活性は 12 U/mg であった (Table 2)。

次に、別のアミノ酸の組み合わせでの欠損の影響を調べるため、リジン、グルタミン、セリン、トレオニン、トリプトファンを同時に欠損させた改変体 Arc1-14 を合成した。熱変性測定から、Arc1-14 の変性中点温度は 81°C であることが示され、70°C での比活性は 8.6 U/mg であった (Table 2)。筆者らは、Arc1-14 からさらに他のアミノ酸を欠損させることを試み、Arc1-14 からメチオニンを欠損させた Arc1-13M の変性中点温度が 74°C で、50°C における比活性が 0.15 U/mg であることを明らかにした (Table 2)。したがって、異なる組み合わせの 13 アミノ酸種だけから構成された 2 つの Arc1 改変体 (Arc1-13 と Arc1-13M) が、70°C を超える変性温度と検出可能なリン酸基転移活性を持つことが分かった。

前生物環境下に比較的豊富に存在したと推定されているアミノ酸種との比較

Table 3 で、元の Arc1 と Arc1-13、Arc1-13M のアミノ酸組成を比較した。この表から分かるように、Arc1-13 と Arc1-13M を構成するそれぞれの 13 アミノ酸種のうち、11 アミノ酸種が両改変体で共通している。そこで筆者らは、両者に共通する 11 アミノ酸種を NDK 合成における「必須アミノ酸」と呼ぶことにした。次に、必須アミノ酸を、ミラーの実験あるいはマーチソン隕石中から検出されたアミノ酸と比較した。原始地球環境を模倣した条件下で放電による有機物合成をおこなったミラーの実験では、標準遺伝暗号表で指定された 20 アミノ酸種のうちの 10 アミノ酸が検出された (Fig. 1) [11,15]。さらに、そのうちの 8 アミノ酸種がマーチソン隕石からも見つかった (Fig. 1) [13,18]。隕石中から検出されるアミノ酸を含む有機物は、宇宙に由来すると考えることができる [29]。マーチソン隕石中のアミノ酸組成が前生物的環境下でのアミノ酸組成を反映していると仮定すると、これらの 8 アミノ酸種は原始地球環境に比較的豊富に存在したと考えられる。生物によるアミノ酸合成経路誕生以前に存在した原始タンパク質は、環境中のアミノ酸が重合することで合成されたはずである [6]。したがって、ミラーの実験とマーチソン隕石の両方から検出されたアミノ酸は、原始タンパク質が合成された際に利用可能であったと考えられる。

Table 1. Unfolding temperatures and specific activities at 70°C of Arc1 and its variants, each devoid of one amino acid letter.

Variant	Eliminated amino acids	Unfolding midpoint temperature (°C)	Specific activity (U/mg)
Arc1	C	114	2200
Arc1-A	A, C	105	360
Arc1-D	C, D	105	0.64
Arc1-E	C, E	— ^a	— ^a
Arc1-F	C, F	103	1100
Arc1-G	C, G	— ^a	— ^a
Arc1-H	C, H	121	0.35
Arc1-I	C, I	99	1200
Arc1-K	C, K	106	67
Arc1-L	C, L	103	550
Arc1-M	C, M	96	1200
Arc1-N	C, N	118	3.4
Arc1-P	C, P	95	1.7
Arc1-Q	C, Q	113	1500
Arc1-R	C, R	109	3.8
Arc1-S	C, S	104	280
Arc1-T	C, T	108	370
Arc1-V	C, V	57	3.4 ^b
Arc1-W	C, W	113	2600
Arc1-Y	C, Y	114	43

^aAppeared to be insoluble.

^bSpecific activity of the variant devoid of valine was determined at 50°C.

Table 2. Unfolding temperatures and specific activities at 70°C of simplified variants devoid of multiple amino acid letters.

Variant	Eliminated amino acids	Unfolding midpoint temperature (°C)	Specific activity (U/mg)
Arc1-AF	A, C, F	90	400
Arc1-AI	A, C, I	87	320
Arc1-AL	A, C, L	91	500
Arc1-AM	A, C, M	85	64
Arc1-FI	C, F, I	84	410
Arc1-FL	C, F, L	100	30
Arc1-FM	C, F, M	94	520
Arc1-IL	C, I, L	93	14
Arc1-IM	C, I, M	87	550
Arc1-LM	C, L, M	78	6.8
Arc1-16	C, F, I, M	78	260
Arc1-13	C, F, I, M, Q, T, W	74	12 ^a
Arc1-14	C, K, Q, S, T, W	81	8.6
Arc1-13M	C, K, M, Q, S, T, W	74	0.15 ^a

^aSpecific activity was determined at 50°C.

ミラーの実験とマーチソン隕石から共通して検出された 8 アミノ酸のうち、イソロイシンを除く 7 アミノ酸が NDK 合成における「必須アミノ酸」と一致した (Fig. 1)。他の研究グループからも、ミラーの実験から検出された 10 アミノ酸種を主要な構成成分としたアミノ酸配列が、高塩濃度下で安定な β -trefoil 構造へと折りたたまれることが報告

されている [30]。したがって、少数種アミノ酸による安定で活性を持ったタンパク質の再構成実験は、原始地球環境に豊富に存在したアミノ酸の推定を試みた地球化学的な研究を補うことができるはずである。

Table 3. Amino acid compositions of Arc1 and its simplified variants composed of 13 amino acid letters.

Amino acid	Arc1	Arc1-13	Arc1-13M
Ala	11	12	19
Cys	0	0	0
Asp	6	6	8
Glu	16	16	18
Phe	7	0	7
Gly	11	11	12
His	3	3	3
Ile	10	0	11
Lys	11	11	0
Leu	7	22	14
Met ^a	7	0	0
Asn	3	3	4
Pro	6	6	7
Gln	1	0	0
Arg	10	11	19
Ser	8	12	0
Thr	4	0	0
Val	14	20	14
Trp	1	0	0
Tyr	2	5	2

^aThe N-terminal methionine residues are not taken into account.

しかし、NDK 合成の「必須アミノ酸」のうち、ヒスチジン、アスパラギン、アルギニン、チロシンは、ミラーの実験とマーチソン隕石のどちらからも検出されていない (Fig. 1)。したがって、これ

らの4アミノ酸は原始地球ではあまり合成されていなかったかもしれない [18]。この不一致の解釈のひとつは、原始タンパク質は非酵素的化学反応をサポートはしたが、直接触媒には関与していな

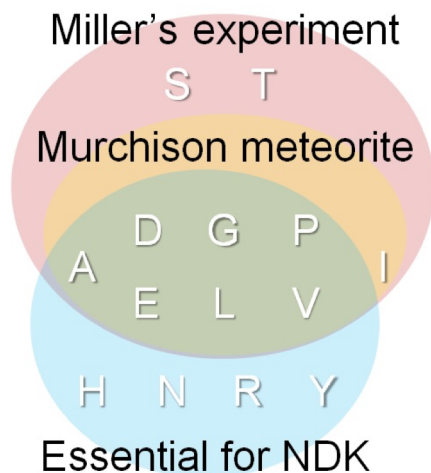


Fig. 1. Venn diagram comparing the essential amino acids for active NDK (cyan) with amino acids that were found in the products of the Miller's experiment (magenta) and Murchison meteorite (orange).

かった可能性である。タンパク質の初期進化では、タンパク質合成に使われるアミノ酸は触媒を促進する目的としては選択されず、むしろ、安定な立体構造形成を達成する目的で選ばれていたという仮説が提唱されている [31]。この仮説では、RNAや補酵素のようなタンパク質以外の分子、あるいは金属が直接の触媒の担い手であったと考えている。実際、原始地球環境に豊富に存在したと推定されているアミノ酸の多くは、機能的な側鎖構造を持っていない非極性アミノ酸である。ここまで紹介してきた筆者らの研究からも、ヒスチジン、アスパラギン、アルギニン、チロシンは触媒活性には重要であるが、タンパク質の安定性にはあまり寄与していないことが示された (Table 1)。さらに筆者らは、Arc1-13M からヒスチジン、アスパラギン、チロシンを欠損させた Arc1-10 を作製したが、Arc1-10 は Arc1-13 と比べて変性中点温度が 11°C 回復した。しかし、Arc1-10 は検出可能なリン酸基転移活性を示さなかった。この結果も、ヒスチジン、アスパラギン、チロシンが触媒活性には重要である一方、安定性にはほとんど寄与していないことを示した。したがって、これらのアミノ酸は、ミラーの実験やマーチソン隕石から見つかったアミノ酸よりも後にタンパク質合成に使われるようになり、タンパク質の機能の多様化や向上に寄与したと考えられる。しかし、Sutherland らは、アルギニンとアスパラギンの前駆体の非生物的合成経路を最近報告した [32]。さらに、ヒスチジンやイミダゾリドの非生物的合成についての報告もある [33-35]。したがって、ヒスチジン、アスパラギン、アルギニン、チロシンが原始地球上で何らかの化学反応によって合成され、最初からタンパク質合成に使われていた可能性を完全に否定できるわけではないことを最後に付け加えたい。

謝辞

本研究の一部は、科研費 15H01068、17H03716、17H05237、および、大学共同利用機関法人自然科学研究機構アストロバイオロジーセンター・サテライト研究 (AB302004) の助成を受けたものである。

引用文献

1. Powner, M. W., Gerland, B. and Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459, 239-242 (2009).
2. DeGuzman, V., Vercoutere, W., Shenasa, H. and Deamer, D. Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. *J. Mol. Evol.* 78, 251-262 (2014).
3. Wochner, A., Attwater, J., Coulson, A. and Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332, 209-212 (2011).
4. Szczepanski, J. T. and Joyce, G. F. A cross-chiral RNA polymerase ribozyme. *Nature* 515, 440-442 (2014).
5. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* 38, 367-379 (1968).
6. Wong, J. T. A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. USA* 72, 1909-1912 (1975).
7. Eigen, M. and Schuster, P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64, 541-565 (1977).
8. Wachtershauser, G. Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* 52, 452-484 (1988).
9. Higgs, P. G. A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. *Biol. Direct.* 4, 16 (2009).
10. Johnson, D. B. and Wang, L. Imprints of the genetic code in the ribosome. *Proc. Natl. Acad. Sci. USA* 107, 8298-8303 (2010).
11. Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* 117, 528-529 (1953).
12. Kvenvolden, K. *et al.* Evidence for extraterrestrial amino-acids and hydrocarbons in the Murchison meteorite. *Nature* 228, 923-926 (1970).
13. Cronin, J. R. and Pizzarello, S. Amino acids in meteorites. *Adv. Space Res.* 3, 5-18 (1983).
14. Cleaves, H. J., Chalmers, J. H., Lazcano, A., Miller, S. L. and Bada, J. L. A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. *Orig. Life Evol. Biosph.* 38, 105-115 (2008).
15. Johnson, A. P. *et al.* The Miller volcanic spark discharge experiment. *Science* 322, 404 (2008).
16. Bada, J. L. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev.* 42, 2186-2196 (2013).
17. Weber, A. L. and Miller, S. L. Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol.* 17, 273-284 (1981).
18. Cleaves, H. J., 2nd. The origin of the biologically coded amino acids. *J. Theor. Biol.* 263, 490-498 (2010).
19. Philip, G. K. and Freeland, S. J. Did evolution select a nonrandom "alphabet" of amino acids? *Astrobiology* 11, 235-240 (2011).
20. Ilardo, M., Meringer, M., Freeland, S., Rasulev, B. and Cleaves, H. J., 2nd. Extraordinarily adaptive properties of the genetically encoded amino acids. *Sci. Rep.* 5, 9414 (2015).
21. Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. and Stroud, R. M. A designed four helix bundle protein with

- native-like structure. *Nat. Struct. Biol.* 4, 1039-1046 (1997)
22. Akanuma, S., Kigawa, T. and 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).
 23. Walter, K. U., Vamvaca, K. and Hilvert, D. An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* 280, 37742-37746 (2005).
 24. 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).
 25. Akanuma, S. *et al.* Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. USA* 110, 11067-11072 (2013).
 26. Akanuma, S., Yokobori, S., Nakajima, Y., Bessho, M. and Yamagishi, A. Robustness of predictions of extremely thermally stable proteins in ancient organisms. *Evolution* 69, 2954-2962 (2015).
 27. Shibue, R., Sasamoto, T., Shimada, M., Zhang, B., Yamagishi, A. and Akanuma, S. Comprehensive reduction of amino acid set in a protein suggests the importance of prebiotic amino acids for stable proteins. *Sci. Rep.* 8, 1227 (2018)
 28. Steipe, B. Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes. *Methods Enzymol.* 388, 176-186 (2004).
 29. Burton, A. S., Stern, J. C., Elsil, J. E., Glavin, D. P. and Dworkin, J. P. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* 41, 5459-5472 (2012).
 30. Longo, L. M., Lee, J. and 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. Doig, A. J. Frozen, but no accident - why the 20 standard amino acids were selected. *FEBS J.* 284, 1296-1305 (2017).
 32. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. and Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* 7, 301-307 (2015).
 33. Shen, C., Yang, L., Miller, S. L. and Oro, J. Prebiotic synthesis of imidazole-4-acetaldehyde and histidine. *Orig. Life Evol. Biosph.* 17, 295-305 (1987).
 34. Shen, C., Yang, L., Miller, S. L. and Oro, J. Prebiotic synthesis of histidine. *J. Mol. Evol.* 31, 167-174 (1990).
 35. Vazquez-Salazar, A. *et al.* Can an imidazole be formed from an alanyl-seryl-glycine tripeptide under possible prebiotic conditions? *Orig. Life Evol. Biosph.* 47, 345-354 (2016).