

令和 5 年 6 月 26 日現在

機関番号：32689

研究種目：基盤研究(A) (一般)

研究期間：2019～2021

課題番号：19H00985

研究課題名(和文)生命に現在の20種類の標準アミノ酸は必要か: 遺伝暗号改変による理工学アプローチ

研究課題名(英文) Is the set of 20 Canonical amino acids essential for Life: from viewpoints of alternative genetic codes

研究代表者

木賀 大介 (Kiga, Daisuke)

早稲田大学・理工学術院・教授

研究者番号：30376587

交付決定額(研究期間全体)：(直接経費) 35,100,000円

研究成果の概要(和文)：本研究の目的は、代表者独自の遺伝暗号の改変を活かした合成生物学アプローチによって、生命を構成するために20種類のアミノ酸全てが必要か、という疑問を追求することにある。その過程で、計算機による予測のサポートを受けたタンパク質の人工進化を行った。変異体の活性予測を行う計算手法として、祖先配列推定法と、変分オートエンコーダーという深層学習手法、2つの方法を比較した。その結果、解糖系の全てのステップについて、トリプトファンを持たずとも十分な活性を持つ酵素群を揃えることができた。また、工学応用の例として、硫黄原子を含まないIGFPについて、蛍光が観測可能な変異体を創出した。

研究成果の学術的意義や社会的意義

天然のタンパク質合成に用いられる20種類のアミノ酸からトリプトファンを除いたセットで生物の中心代謝系である解糖系の活性を構築できたことは、生命の起原と初期進化のシナリオに物質的証拠を提示したことになり、学術的意義が大きい。また、酸化耐性を発揮させるために硫黄原子を持たずとも十分な活性があるタンパク質を創出できたことは、今後のタンパク質産業への貢献が強く期待できる。

研究成果の概要(英文)：The purpose of this synthetic-biology study was to pursue the question of whether all 20 amino acids are necessary to constitute life. For the pursuit, protein variants were created by artificial evolution using engineered genetic codes with the support of computational prediction. For predicting the activity of protein mutants, we compared two methods: the ancestral sequence reconstruction method and a deep learning method called variational autoencoder. As a result, we obtained enzymes with sufficient activity without tryptophan for all steps of the glycolytic pathway. As an example of engineering application, we also created an active mutant of GFP without sulfur atoms.

研究分野：合成生物学

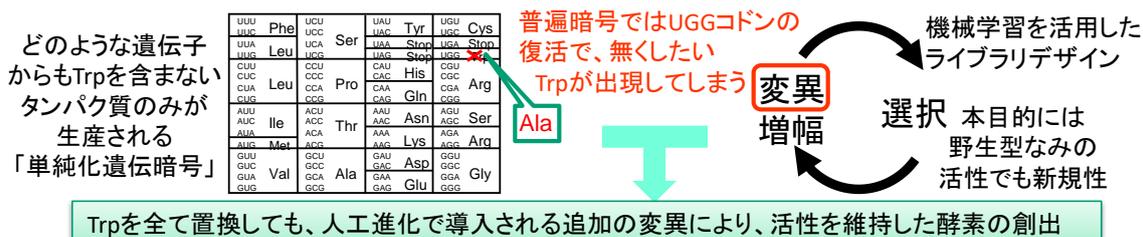
キーワード：タンパク質 機械学習 合成生物学 遺伝暗号 進化分子工学

### 1. 研究開始当初の背景

なぜ生命は現在の 20 種類のアミノ酸セットを採用したのだろうか？単純なカタチから複雑なカタチへと変化する生命進化の本質を考慮すれば、普遍遺伝暗号を共有する生命の共通祖先の確立以前、遺伝暗号の成立時から 20 というアミノ酸の種類が定まっていたとは考えにくい。

このような問いについて考察する一助として、例えば、Trp を持たない酵素群によって生命活動の鍵となる反応系を構築することを着想した。このような系の構築により、そもそも生命システムを構築するために 20 種類のアミノ酸全てが必要なのだろうか、という疑問に答えることができる。

19 種類以下のアミノ酸のみで成立する「単純化」遺伝暗号表を構築できることが、代表者のこれまでの成果により示されていた(下図左)(Nucleic Acids Res 2012, ACS SynBiol 2014)。さらに、種々のタンパク質についても、19 種類以下のアミノ酸のみで構築されるタンパク質の創出が可能なが示されており、この類の創出研究は研究代表者の新技術によって進歩してきた(ACS SynBiol 2018)。



タンパク質に含まれるアミノ酸の種類は、第一義的には遺伝暗号によって規定される。遺伝暗号表の起源と進化の仮説については、この最適化の過程がある時点で停止したという Crick による偶然凍結説があり(参考 1)、さらに近年では、アミノ酸数の少なかった遺伝暗号表にアミノ酸が加わってきた、という議論がなされていた。この議論の鍵となる酵素が、20 種類のアミノ酸それぞれに対応して存在し、そのアミノ酸と tRNA とを結合させる専用のアミノアシル tRNA 合成酵素である(以下 aaRS)。換言すれば、あるアミノ酸が専用の酵素を持つことが、そのアミノ酸が遺伝暗号に採用されるために重要な条件である。そこで、この酵素のアミノ酸配列や立体構造の比較から、遺伝暗号表の進化を考察することも行なわれている。例えば、TyrRS と TrpRS は、20 種類の aaRS の中で構造が類似しているペアである。種々の生物のこの 2 種の酵素の配列比較によって祖先型酵素配列を推定すると、これらの共通祖先酵素に Trp が含まれていなかった可能性が高いことも示されている。また、Trp を代謝中間体とする他の遺伝暗号中のアミノ酸が無い、化学進化で Trp が作成されにくいなど、種々の研究から、遺伝暗号表に含まれるアミノ酸の種類が少なかった時代があり、Trp は遺伝暗号に最終段階で加わったと考える研究者が多かった(参考 2)。

遺伝暗号に含まれるアミノ酸の種類が現在より少ないということは、19種類以下のアミノ酸のみで生命システムが構築されていた、ということである。生命システムの要素であるタンパク質については、研究代表者の遺伝暗号改変以前の研究において、他研究者が遺伝暗号表の改変を伴わずにアミノ酸の種類を減らしたタンパク質の特性を研究してはきたが(参考3)、これらの研究で得られた”タンパク質”の活性は天然に比べて極めて低いものであった。

なぜならば、天然のタンパク質が淘汰・増幅・変異を繰り返してダーウィン進化の過程を経た産物であることに對し、これまでの研究では、特定のコドンのみを使用するようにデザインされたライブラリからタンパク質を単離したのみで、再度の変異導入を要する人工進化を行っていなかったからである。それは、普遍遺伝暗号を使用する限りは、人工進化の過程で生じる遺伝子への変異によって、実験系から除去したはずのアミノ酸を指定するコドンへの塩基置換が生じてしまうという、既存の実験手法の限界に由来する。

一方研究代表者は、特定のアミノ酸群のみを含んだ新規な「単純化」遺伝暗号システムを世界にさきがけて構築し、さらにこれを活用した進化工学により、アミノ酸の種類が限定され、かつ活性の高いタンパク質を創出することに成功した。

(参考 1)Crick, J Mol Bio 38(1968)367-79. (参考 2) Trifonov. J Biomol Struct Dyn. 22(2004) 1-11. (参考 3) Akanuma et al., PNAS, 99(2002), 13549-53.

### 2. 研究の目的

本研究の目的は、研究代表者独自の遺伝暗号の改変を活かした合成生物学アプローチによって、生命を構成するために 20 種類のアミノ酸全てが必要か、という疑問を追求することにある。本研究ではこの根源的な疑問に答えるため、トリプトファンなど特定のアミノ酸を持たなくとも、生命の根幹をなす代謝系が動作することを示すことを目指した。さらに、この研究技術を工学的に活用する一例として、酸化耐性の獲得を目指して、含硫アミノ酸を持たなくとも活性を保

持するタンパク質を得ることも目的とした。

### 3. 研究の方法

Trp、または含硫アミノ酸(Cys, Met)を持たないタンパク質の取得は、リジンを持たずとも活性のあるインターフェロンの創出において(ACS SynBio1 2018)、すでに研究代表者が確立させた独自技術を用いた。種々の相同タンパク質のアミノ酸配列を参照し進化学を行うことで、10 残基以上あった Lys を全て置換しても、さらに別の部位の変異と併せて活性が野生型以上となった。

タンパク質の人工進化は、目標とする特性を持つタンパク質変異体を得るために有効な実験方法であるが、特定のアミノ酸を含まないタンパク質を得ることは、研究代表者の技術なしには困難であった。なぜならば、20 種類のアミノ酸を指定する標準遺伝暗号表を用いた実験系を用いると、一度は欠損させたはずのアミノ酸種が、人工進化中の変異によって再びタンパク質に戻ってしまう可能性が高いためである。

一方、研究代表者が開発してきた、普遍暗号における Met や Cys のコドンが Ala、Ser に対応する改変暗号を使用すると、どのようなミスセンス変異が生じて、酸化の弱点となるこれら含硫アミノ酸を持たないタンパク質が生産される。その結果、人工進化中に変異を導入しても、これら含硫アミノ酸を持たないまま、タンパク質のアミノ酸配列を進化させることができる。本独自手法により、酸化耐性タンパク質を創出する人工進化でのこの困難は解消している。同様の手法で Trp を含まないタンパク質の人工進化も可能である。

特定のアミノ酸を持たないタンパク質を得る本研究では、野生型配列データ群を活用した機械学習を適用することで生物実験を効率良く進めた。本来、機械学習は、データセットの内側を補間する内挿において圧倒的な強みをもつ一方、外挿によって機能向上を達成することは困難である。しかし、本研究のアプローチでは、野生型配列群を補間する内挿により、「弱点となるアミノ酸を全て置換しても野生型並みの活性を保持したタンパク質の開発」に繋がる。なぜならば、必要とされる酵素活性は、近年 DNA 配列データが膨大に蓄積されている野生型酵素と同等で良く、野生型以上である必要はない。さらに、Met・Cys という弱点を含まない 18 種類のアミノ酸からなる配列は、20 種類アミノ酸のサブセットである。このため、本研究での機械学習による活性予測は、機械学習による判別機の内挿問題となり、高い精度を期待できた。

### 4. 研究成果

解糖系の全てのステップについて、トリプトファンを持たずとも十分な活性を持つ酵素群を揃えることができた。そのために、まずは、データベース中にはトリプトファンを持たない、または 1 個のみ持つ解糖系酵素があるため、これら大腸菌で発現する配列として遺伝子合成した。合成した DNA を活用して、大腸菌の解糖系 1 酵素の欠損株の相補活性、または、精製酵素の試験管内活性について評価したが、数個の酵素については活性を見出すことはできなかった。その場合、対応する大腸菌の酵素は複数個のトリプトファンのみをもっていたが、これら全て置換した結果、弱いながらも活性を見出すことができた。そこで、これら大腸菌酵素変異体にさらなる置換を導入した結果、欠損株の相補について、野生型酵素に近い生育速度回復を達成することができた。

工学応用の例として、硫黄原子を含まない GFP について、蛍光が観測可能な変異体を創出した。換言すると、全てのシステインおよびメチオニン残基を置換しても蛍光を持つ GFP を単離している。この単離のための人工進化では、ある段階までは標準暗号表を使ったタンパク質合成を行ったが、蛍光を持つ変異体の多くは復帰変異を持ってしまい、活性型変異体の単離効率が極めて低かった。そこで、どのような塩基配列の mRNA からこれら含硫アミノ酸を持たないペプチドが生産される単純化暗号表を用いた人工進化を行った。その結果、復帰変異の障害なしに、硫黄原子を含まない GFP が蛍光を発することを示すことができた。

本研究では、同一の活性を持つと推測される野生型タンパク質配列データ群を活用した活性予測器として、祖先配列推定法と、変分オートエンコーダーという深層学習手法、2つの方法を比較した。前者は、系統樹を考慮しつつも残基間相互作用を考慮しない方法である。一方、後者は、残基間相互作用を考慮しつつも系統樹の情報を直接は考慮しない方法である。その結果、2つの値はおおむね正の相関を示しつつ、祖先配列としての尤度が高い一方で機械学習の活性予測値が低い配列、その逆の配列、尤度も活性予測値も共に高い配列、が存在することがわかった。このことは、機械学習を古典的な祖先配列推定と組み合わせることで、二重のふるいをかけられることを示している。続いて、どの座位やどのアミノ酸への置換がどちらのツールに影響が出やすいのかを確認した。その結果、20 種類のアミノ酸の中で他とは性質が大きく異なるプロリン(イミノ酸)への置換は祖先配列推定において過剰に高く評価され、トリプトファン(大きな芳香環を側鎖にもつ)への置換は祖先配列推定で過剰に低く評価されることが示され、予測での使用に注意を要することが分かった。これら計算機手法を、本研究の生物実験で構築すべき変異体の選択に活用した。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 2件）

1. 著者名 Beal Jacob, Goi Moreno Angel, Myers Chris, Hecht Ariel, Vicente Mara del Carmen, Parco Maria, Schmidt Markus, Timmis Kenneth, Baldwin Geoff, Friedrichs Steffi, Freemont Paul, Kiga Daisuke, Ordozgoiti Elena, Rennig Maja, Rios Leonardo, Tanner Kristie, Lorenzo V?ctor, Porcar Manuel	4. 巻 21
2. 論文標題 The long journey towards standards for engineering biosystems	5. 発行年 2020年
3. 雑誌名 EMBO reports	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.15252/embr.202050521	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 木賀 大介	4. 巻 93
2. 論文標題 遺伝暗号による制約を突破する翻訳システムの改変と非標準アミノ酸の合成・重合	5. 発行年 2021年
3. 雑誌名 生化学	6. 最初と最後の頁 281 ~ 282
掲載論文のDOI（デジタルオブジェクト識別子） 10.14952/SEIKAGAKU.2021.930281	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 木賀 大介	4. 巻 93
2. 論文標題 遺伝暗号改変を活用した人工進化からみるmagic20の意味	5. 発行年 2021年
3. 雑誌名 生化学	6. 最初と最後の頁 359 ~ 365
掲載論文のDOI（デジタルオブジェクト識別子） 10.14952/SEIKAGAKU.2021.930359	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計16件（うち招待講演 10件/うち国際学会 3件）

1. 発表者名 高木有隣, 橋本真奈, 西田暁史, 柘植謙爾, 木賀大介
2. 発表標題 Trpを持たない酵素群によって構成される解糖系の構築に向けた活性測定
3. 学会等名 生物物理学会関東支部会
4. 発表年 2022年

1. 発表者名 木賀大介 宮崎和光 山村雅幸
2. 発表標題 ありえた生命のかたちの設計と実装をDXする
3. 学会等名 ラボオートメーション研究会 (招待講演)
4. 発表年 2022年

1. 発表者名 Daisuke Kiga
2. 発表標題 Simplified genetic code for directed evolution of proteins without specific amino acid species
3. 学会等名 BuildACell (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 木賀大介
2. 発表標題 合成生物学とバイオセキュリティ
3. 学会等名 バイオエコノミーの現状 セミナー (シリーズ) 応用編 (招待講演)
4. 発表年 2022年

1. 発表者名 木賀大介
2. 発表標題 合成生物学に期待される役割
3. 学会等名 生命倫理学会 (招待講演)
4. 発表年 2021年

1. 発表者名 木賀大介
2. 発表標題 ありえた生体高分子ネットワークを 創出するBioDOSの構築
3. 学会等名 「データ駆動・AI駆動を中心としたデジタルトランスフォーメーションによる生命科学研究の革新バイオDX」キックオフシンポジウム（招待講演）
4. 発表年 2021年

1. 発表者名 満富健太,木賀大介
2. 発表標題 組換えと翻訳エラーの併用が進化の局所解トラップを緩和することの計算機実験による検証
3. 学会等名 第 10回日本生物物理学会関東支部会（オンライン）
4. 発表年 2021年

1. 発表者名 宮地亮多,木賀大介
2. 発表標題 単純化遺伝暗号を用いた 酸化耐性GFPの人工進化
3. 学会等名 第 10回日本生物物理学会関東支部会（オンライン）
4. 発表年 2021年

1. 発表者名 榎本利彦,村上泰平,網蔵和晃,松本航,○木賀大介
2. 発表標題 普遍遺伝暗号を用いた 人工進化の限界： 硫黄抜き酸化耐性GFPを例に
3. 学会等名 第 10回日本生物物理学会関東支部会（オンライン）
4. 発表年 2020年

1. 発表者名 木賀大介
2. 発表標題 種々の遺伝暗号それぞれが発揮できるより広い機能性生体高分子の進化戦略
3. 学会等名 第42回日本分子生物学会年会（福岡）（招待講演）
4. 発表年 2019年

1. 発表者名 Takato Kameyama, Risa Igarashi, Shigeharu Mukai, Hideaki Kanehara, Masaki Aota, Daisuke Kiga, Noboru Murata
2. 発表標題 Influence Analysis of Amino Acid Residues on Protein Functions using Attention-based Neural Networks
3. 学会等名 Chem-Bio Informatics Society(CBI) Annual Meeting 2019Chem-Bio Informatics Society(CBI) Annual Meeting 2019（国際学会）
4. 発表年 2019年

1. 発表者名 木賀大介
2. 発表標題 合成生物学・遺伝子工学とデュアルユース 大規模遺伝子工学が普及した際のリスク
3. 学会等名 第19回日本バイオセーフティ学会（招待講演）
4. 発表年 2019年

1. 発表者名 Daisuke Kiga
2. 発表標題 Directed Evolution under genetic codes with reduced-size alphabet or with low fidelity
3. 学会等名 OIST（招待講演）
4. 発表年 2019年

1. 発表者名 木賀大介
2. 発表標題 原始生命の誕生
3. 学会等名 武田セミナー（第16回）（招待講演）
4. 発表年 2019年

1. 発表者名 木賀大介
2. 発表標題 生命の部品を組み合わせることに由る理学と工学
3. 学会等名 都立医学研セミナー（招待講演）
4. 発表年 2019年

1. 発表者名 Daisuke Kiga
2. 発表標題 Novel Directed Evolution Using Engineered Genetic Code
3. 学会等名 5 th Core-to-Core International Symposium "3D Lab-Exchange Program"（国際学会）
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 島園 進 (著, 編集), 四ノ宮 成祥 (著, 編集), 木賀 大介 (著), 須田 桃子 (著), 原山 優子 (著)	4. 発行年 2022年
2. 出版社 専修大学出版局	5. 総ページ数 182
3. 書名 合成生物学は社会に何をもたらすか	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	村田 昇  (Murata Noboru)  (60242038)	早稲田大学・理工学術院・教授    (32689)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関