Synthesis assistance of transmembrane domains is a fundamental function of the genetic code table assignment

Genshiro Esumi
Pediatric Surgery, University of Occupational and Environmental Health, Kitakyushu, Japan

In the genetic code table, nucleic acid sequences on genes correspond to amino acid sequences on proteins. While the genetic code table assignment is common among almost all current organisms, there has never been a clear explanation for its reason. This paper presents three examinations to provide some background on this issue.

First, principal component analysis of the amino acid compositions of all human proteins showed that membrane proteins with two or more transmembrane domains were separated from those without, dominantly by their second principal component and partially by their first principal component. Their eigenvectors indicated that these membrane proteins tend to contain some amino acids in higher amounts. These are phenylalanine, tyrosine, isoleucine, methionine, tryptophan, cystine, valine, and leucine. Many of them are hydrophobic. Moreover, they all correspond to codons containing uracil in their first or second letter in the genetic code table.

Second, principal component analysis of the nucleic acid composition of the genes for all human proteins revealed that the first and second principal components of the nucleic acid compositions strongly correlated with the above first and second principal components of the amino acid compositions, respectively. Furthermore, I found two correlations. First, the first principal component of the nucleic acid compositions strongly correlates with their GC (guanine and cytosine) content. The second is that the second principal component strongly correlates with the ratio of thymine to adenine + thymine, respectively.

Third, making a scatter plot of all human protein genes by the ratio of thymine to adenine + thymine and their GC content, I found that membrane proteins strongly correlate with genes with more thymine than adenine, regardless of their GC content.

Membrane proteins account for 30% of all proteins, and their sound synthesis must be one of the highest priority issues of life. Meanwhile, these membrane proteins contain transmembrane domains consisting of 20~ consecutive hydrophobic amino acid residue structures. We have believed that organisms must achieve their proteins by accumulating random mutations during their evolution. However, the results of this analysis indicate that even random genetic mutations can make a much more efficient generation of amino acid compositions that form transmembrane domains only by using gene segments with more thymine than adenine. These results suggest that organisms use the assignment characteristics of the genetic code table to synthesize their transmembrane domains.

The synthesis assistance of transmembrane domains could be a fundamental function of the current genetic code table. Therefore, I assume that the current genetic code tables have continued to be selected by such beneficial functions and are in a state of convergence.

Keywords: genetic code table, transmembrane domains, amino acid composition, nucleic acid composition, principal component analysis

E-mail: esumi@clnc.uoeh-u.ac.jp

*The author has no conflicts of interest relevant to the content of this article.

膜貫通ドメイン合成支援は遺伝暗号表配列の重要な機能である

江角 元史郎
産業医科大学病院 小児外科

遺伝暗号表は遺伝子上の核酸配列とタンパク質のアミノ酸配列を対応させる対応表である。この遺伝暗号表の配列は現存するすべての生物においてほぼ共通であることが知られているが、なぜそのようになっているかは現在まで明確には説明されていない。

今回、ヒトの全タンパク質のアミノ酸組成を主成分分析したところ、主としてその第二主成分、次いでその第一主成分によって膜貫通ドメインを複数持つ膜タンパク質とそうでないタンパク質が別れることが明らかになった。固有ベクトルから、これらの膜タンパク質には、フェニルアラニン、チロシン、イソロイシン、メチオニン、トリプトファン、シスチン、バリン、ロイシンの含有が多いと推測されたが、大半は疎水性アミノ酸であり、また同時に全て、遺伝暗号表において、コドンの 1 文字目もしくは 2 文字目に核酸 U を含むコドンに対応したアミノ酸であった。(U は遺伝子上の T に対応する。)

次に、ヒトの全タンパク質の遺伝子の核酸組成を主成分分析したところ、その第一、第二主成分がそれぞれ上記アミノ酸組成の第一主成分、第二主成分に強く相関することが明らかとなった。さらに、この核酸の第一、第二主成分はそれぞれ、その核酸組成の AT 含量(GC 含量に連動)、T と A の存在比にもそれぞれ強く相関していることも判明した。

さらに、ヒト遺伝子をその核酸組成の T と A の比、および、AT 含量でプロットしたところ、上記膜貫通ドメインを持つタンパク質群は、A よりも T が多い遺伝子と強く相関していた。

一般に膜タンパク質は全タンパク質の 3 割を占めると言われており、膜タンパク質の安定的な合成は生命の至上課題の一つであると推測される。これらの膜タンパク質には共通して膜貫通ドメインが含まれるが、この膜貫通ドメインは疎水性の高いアミノ酸残基が一定数連続して配列する特殊な構造をしている。一般に、生物が進化の中で特定の機能性タンパク質を獲得していくためには、ランダムな変異を積み重ねた結果として偶然に望むタンパク質が生成されることを期待する必要があるとされてきた。しかし、今回の解析結果から推測すると、遺伝子の核酸組成において A よりも T が多い部分においては遺伝子変異がランダムであっても膜貫通ドメインに多いアミノ酸がコードされる確率が格段に高くなっていると考えられた。以上より、生物は遺伝暗号表の配列の特徴を利用して膜貫通ドメインの合成をアシストしている可能性があると考えられた。

膜貫通ドメインの合成支援は遺伝暗号表配列の重要な機能の一つであり、現在の遺伝暗号表はこの機能も含めた様々な機能によって選択され続け、その結果として収束状態にあると考えられた。

キーワード:遺伝暗号表, 膜貫通ドメイン, アミノ酸組成, 核酸組成, 主成分分析

背景

遺伝暗号表は遺伝子上の核酸配列とタンパク質のアミノ酸配列を対応させる対応表である。この遺伝暗号表の配列は現存するすべての生物においてほぼ共通であることが知られているが、なぜそのようになっているかは現在まで明確には説明されていなかった[1]。

筆者は過去の報告で細胞内において最も高合成コストなアミノ酸が集中するのは膜タンパク質であることを示しており、特にそのようなアミノ酸は膜貫通ドメインに集中していると推測していた [2]。膜貫通ドメインは全ての生物細胞の膜タンパク質中に普遍的に存在しておりその存在量も少なくない。今回、ヒトゲノム上の全タンパク質のアミノ酸組成と対応する核酸組成の統計解析を行ったところ、この膜貫通ドメインが核酸組成の局所的特徴に対応して生成されていることが新たに推測された。この推測について、以下に 3 つの解析とその結果を示し説明する。

解析① ヒトの全タンパク質のアミノ酸組成の主成分分析と膜貫通ドメイン

対象と方法

ヒトの全タンパク質のアミノ酸組成を、NCBI のサイトに掲載されたヒトの全タンパク質の CDS データ(n=123410)より作成した[3]。これは CDS データの核酸配列を標準遺伝暗号表に基づいてアミノ酸配列に変換し、これらの各タンパク質のアミノ酸残基数をカウントし組成データを作成することで行った。また、NCBI の CDS データを確認すると、RefSeq ID が付記されていた。次に UniProt のサイトに掲載されているヒトのタンパク質のデータ(n=205050)を参照したところ、これらには RefSeq ID と膜貫通ドメインのタイプ、数の情報が掲載されていた。まず、UniProt のデータを用いて RefSeq ID と膜貫通ドメイン情報の対応表を作成した[4]。さらに、NCBI のサイトの CDS データと UniProt のサイトのタンパク質データについて RefSeq ID を用いて突合を行った。その結果、NCBI のデータ上の 60163 のタンパク質について UniProt 上の ID との突合が可能であったため、今回の解析ではこれらのタンパク質（タンパク質の遺伝子配列、アミノ酸組成と膜貫通ドメイン情報のセット）を対象とした。

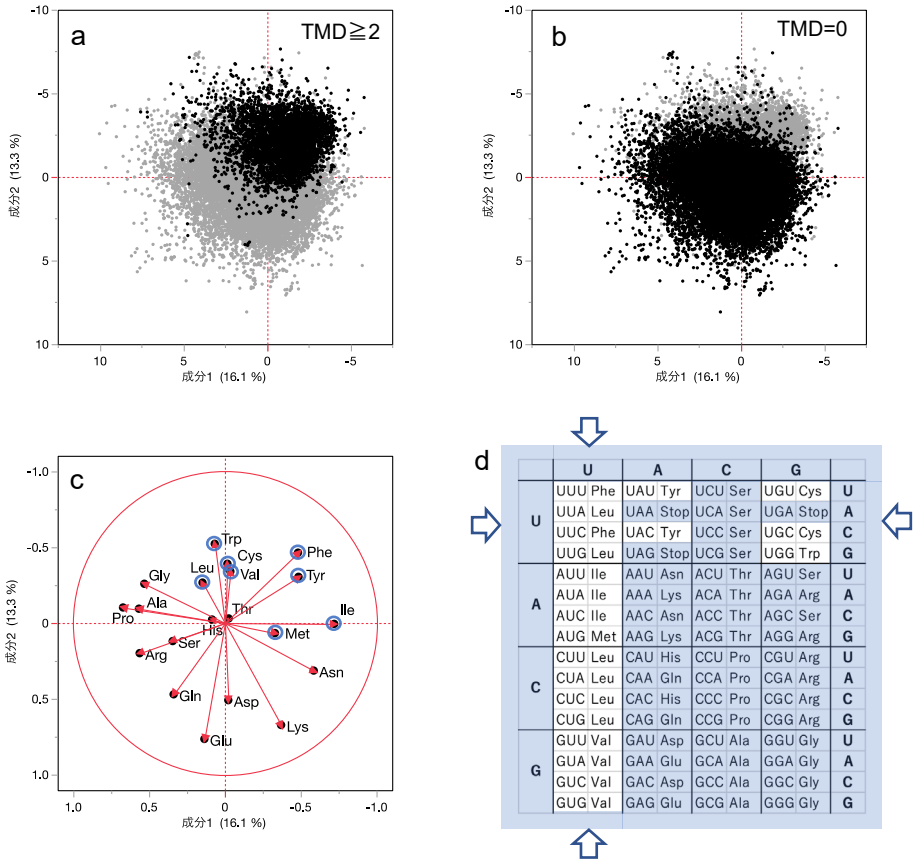
計算したタンパク質のアミノ酸組成データの統計解析には、主成分分析を用いた。組成データ解析においては通常の統計解析手法が使えないことが示されているため、今回はその対策として有心対数比変換を行ってから主成分分析を行った[5]。有心対数比変換においては 0 となる組成項目があると変換できないが、タンパク質には一部のアミノ酸が含まれていないものが含まれている。このため、各タンパク質のアミノ酸数集計で 0 になったアミノ酸についてはこれを 0.5 に変えてから組成データとして変換を行った。

以下の解析を含め、本研究においては、データの配列の集計・変換に Microsoft® Excel for Mac v16.61 (Microsoft Corporation, Redmond, WA, USA)を使用し、また、主成分分析を含む統計解析とグラフの作成には JMP® 16.2.0 (SAS Institute Inc., Chicago, IL, USA)を使用した。

結果

突合したヒトの全タンパク質（n=60163）のアミノ酸組成の主成分分析の結果を示す。（Figure 1）

Figure 1



	U	A	C	G	
U	UUU Phe	UAU Tyr	UCU Ser	UGU Cys	U
	UUA Leu	UAA Stop	UCA Ser	UGA Stop	A
	UUC Phe	UAC Tyr	UCC Ser	UGC Cys	C
	UUG Leu	UAG Stop	UCG Ser	UGG Trp	G
A	AUU Ile	AAU Asn	ACU Thr	AGU Ser	U
	AUA Ile	AAA Lys	ACA Thr	AGA Arg	A
	AUC Ile	AAC Asn	ACC Thr	AGC Ser	C
	AUG Met	AAG Lys	ACG Thr	AGG Arg	G
C	CUU Leu	CAU His	CCU Pro	CGU Arg	U
	CUA Leu	CAA Gln	CCA Pro	CGA Arg	A
	CUC Leu	CAC His	CCC Pro	CGC Arg	C
	CUG Leu	CAG Gln	CCG Pro	CGG Arg	G
G	GUU Val	GAU Asp	GCU Ala	GGU Gly	U
	GUA Val	GAA Glu	GCA Ala	GGA Gly	A
	GUC Val	GAC Asp	GCC Ala	GGC Gly	C
	GUG Val	GAG Glu	GCG Ala	GGG Gly	G

ヒトの全タンパク質のアミノ酸組成を主成分分析したところ、主としてその第二主成分、次いで第一主成分によって膜貫通ドメインを複数持つ膜タンパク質（TMD≧2）とそうでないタンパク質（TMD=0）が別れることが明らかになった。（Figure 1a,b） プロット配置の方向、およびこれらの固有ベクトルの方向から、プロットされた膜タンパク質には、フェニルアラニン、チロシン、イソロイシン、メチオニン、トリプトファン、シスチン、バリン、ロイシンの含有が多いと推測された。（Figure 1c,） これらの大半は疎水性アミノ酸であり、また同時に全てのアミノ酸が遺伝暗号表においてコドンの 1 文字目もしくは 2 文字目に U（ウラシル）を含むコドンに対応したアミノ酸であった。（Figure 1d,矢印）

※一般に遺伝暗号表では 4 つの核酸を UCAG の順に並べて表示されているが、本論文では後述の内容にもある通り遺伝子の最大バリエーションが AT 含量（GC 含量）であることを考慮し、核酸の順番を UACG の順とした遺伝暗号表を採用した。（Figure 1d）

解析② ヒトの全タンパク質遺伝子の核酸組成の主成分分析と膜貫通ドメイン

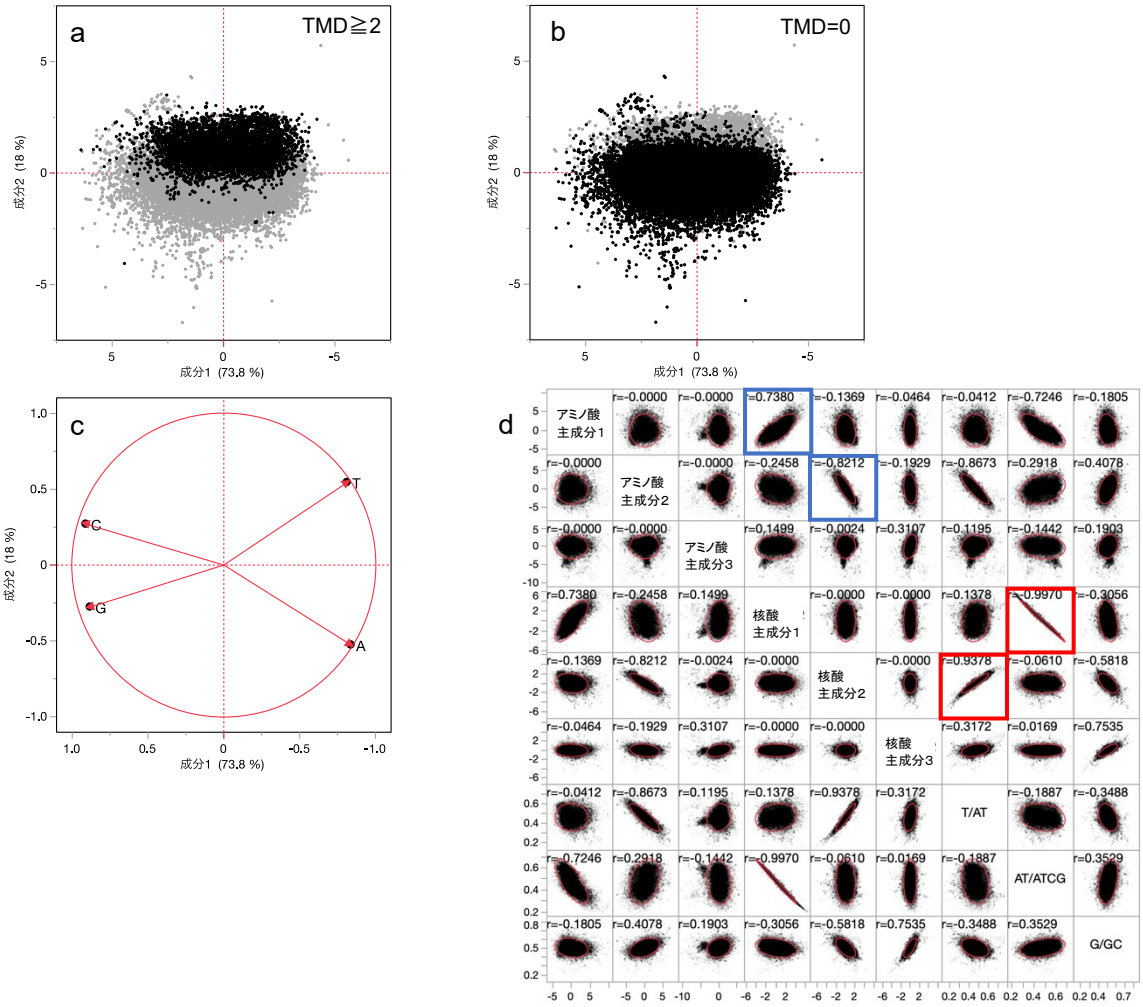
対象と方法

ヒトの全タンパク質遺伝子の核酸組成は、解析①で対象としたタンパク質（n=60163）について、その CDS のタンパク質に対応した核酸配列に含まれる核酸数を集計して作成した。核酸組成のアデニン、グアニン、シトシン、チミンにおいては要素が 0 となるタンパク質は含まれていなかった。これらについて、解析①と同様に有心対数比変換を行った後主成分分析を行った。

結果

突合したヒトの全タンパク質（上記）の核酸組成の主成分分析の結果を示す。（Figure 2）

Figure 2



ヒトの全タンパク質の遺伝子の核酸組成の主成分分析の第一、第二主成分によって各タンパク質をプロットしたところ、(解析①と同様に、)膜貫通ドメインを二つ以上持つタンパク質(TMD≧2)と持たないタンパク質(TMD=0)が別れてプロットされた。(Figure 2a,b) これらの主成分の固有ベクトルを確認したところ、第一主成分において AT と CG が反対のベクトルとなっていることから、AT/ATCG(全核酸 A、T、C、G の和に対するアデニンとチミンの核酸含量の和 AT の比)が、そして、第二主成分方向には T と A、C と G がそれぞれ反対向きのベクトルになっているため、T/AT(アデニンとチミン含量の和に対するチミン含量の比)と、G/GC(グアニンとシトシンの含量の和に対するグアニン含量の比)がそれぞれ相関する可能性が示唆された。(Figure 2c) この検証のためこれらの相互相関を解析したところ、まず、アミノ酸組成と核酸組成の第一主成分、第二主成分同士が強く相関していた。(Figure 2d, 青枠) さらに、核酸の第一、第二主成分と AT/ATCG、T/AT がそれぞれ非常に強く相関していた。(Figure 2d, 赤枠)

解析③ ヒトタンパク質遺伝子の核酸組成と膜貫通ドメイン

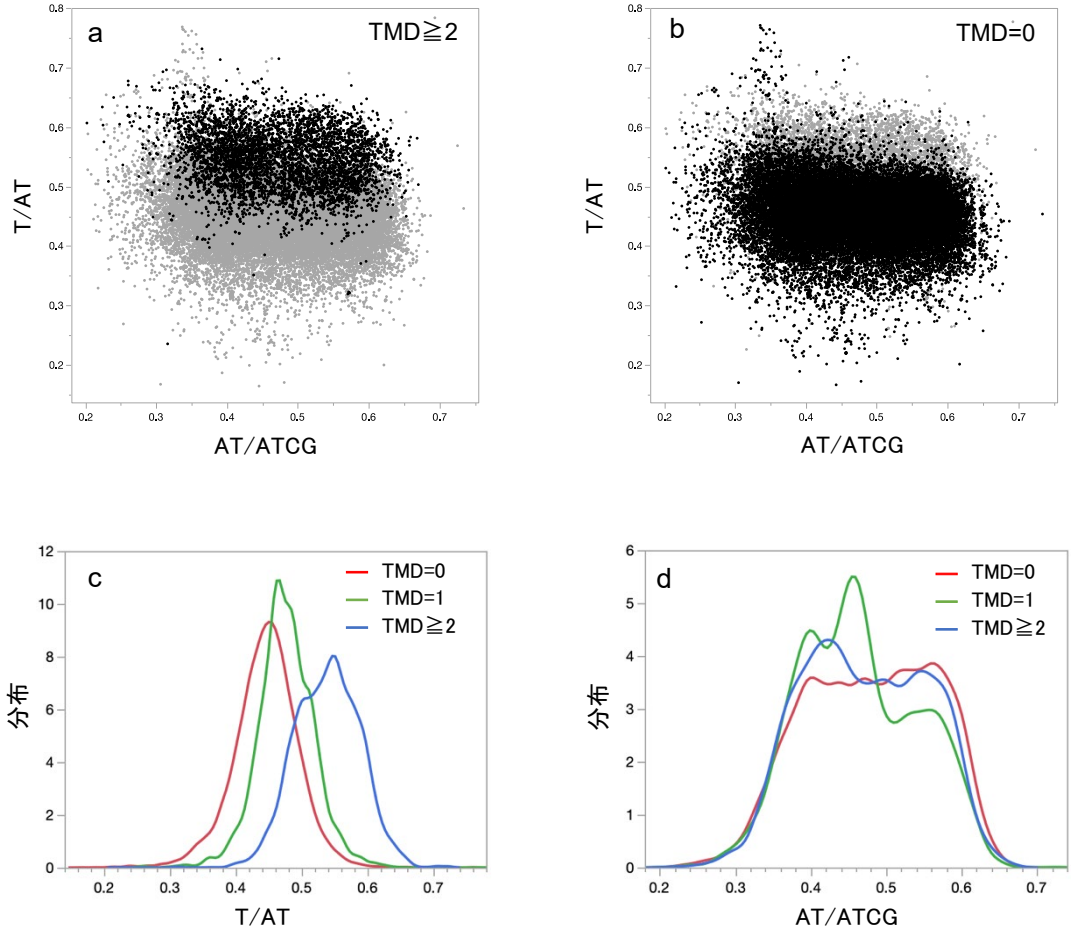
対象と方法

解析②によって、核酸組成の主成分分析の第一、第 2 主成分が AT/ATCG と T/AT で近似されうることが推測されたため、これら変数を用いて散布図に対象のタンパク質（n=60163）をプロットした。

結果

AT/ATCG を横軸、T/AT を縦軸にとって、対象の全タンパク質をプロットした結果を示す。（Figure 3）

Figure 3



各タンパク質のプロットの結果、解析①、②と同様に、膜貫通ドメインを複数持つタンパク質は膜貫通ドメインを持たないタンパク質とは別れてプロットされていた。（Figure 3a,b） それぞれの軸について分布を検討したところ、膜タンパク質のプロットは T/AT については、0.5 以上、つまり、A よりも T が多いことが相関していると推測された。（Figure 3c） また、この相関は、AT/ATCG との間には認められなかった。（Figure 3d）

考察

本論文では 3 つの解析について、その結果を示した。

1 つ目の解析では、ヒトのタンパク質のアミノ酸組成の主成分分析の第一、第二主成分が複数の膜貫通ドメインを持つタンパク質と持たないタンパク質を分けていることを示した。それぞれの固有ベクトルより、膜貫通ドメインにおいては、特定のアミノ酸群がより多く使われている可能性が示された。また、このアミノ酸群は、遺伝暗号表上で U（ウラシル）と相関している可能性、つまり遺伝子上の T（チミン）と相関している可能性が示された。

2 つ目の解析では、ヒトのタンパク質をコードする遺伝子の核酸組成の主成分分析を行った。核酸組成の第一主成分、第二主成分は、それぞれ、アミノ酸組成の第一主成分、第二主成分と相関していた。これは、ヒトのタンパク質におけるアミノ酸組成の多様性は、まず第一にその遺伝子の核酸組成により生じるということを示すと考えられた。また、核酸組成の主成分の固有ベクトルと核酸比の相関解析より、ヒトの核酸組成の多様性は、その第一主成分に相関する AT/ATCG と、その第二主成分に相関する T/AT によって大まかに表現できる可能性も示された。

3 つ目の解析において、実際に AT/ATCG と T/AT によってタンパク質をプロットしたところ、解析①、解析②で認められた膜貫通ドメインを複数持つ膜タンパク質（TMD≧2）と膜貫通ドメインを持たないタンパク質（TMD=0）を境界している要素は、核酸第一主成分に相関しより寄与率が大きいと考えられる AT/ATCG ではなく、むしろ核酸第 2 主成分に相関した T/AT の値であった。この結果より、膜タンパク質（TMD≧2）は、A よりも T が多い遺伝子と相関している可能性が考えられた。

以上 3 つの解析の結果、解析①より、遺伝子上に A よりも T が多ければ膜貫通ドメインのアミノ酸組成をコードする確率が上がると推測され、また解析②③より、複数の膜貫通ドメインを持つタンパク質（TMD≧2）をコードする遺伝子には A よりも T が多い傾向がある、ということが示された。

では、個々の膜貫通ドメインにおいては、本当に A よりも T が多いのだろうか？これについて別に解析を行った結果を示す。（Figure 4）

Figure 4

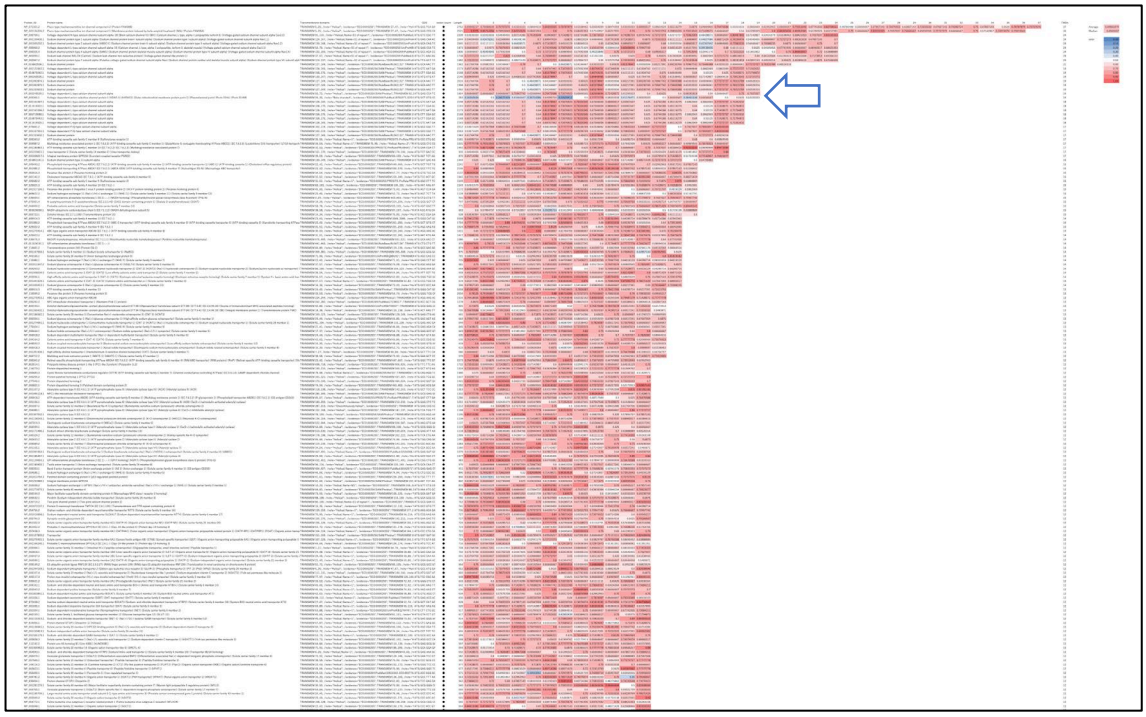


Figure 4 は突合・整合できたタンパク質の情報（n=4178）について、それぞれの持つ膜貫通ドメインの数が多い順に並べ、それぞれのタンパク質が持っている個々の膜貫通ドメインの遺伝子配列についてそれぞれの核酸 A と T の数をカウントし、その T/AT の値に応じて各セルを着色した表である。T/AT が 0 に近くなると青、0.5 で白、1 に近くなると赤となるように着色したところ、ほとんどすべての膜貫通ドメインが赤色になった。個々のドメインについて確認していくと、α-ヘリックス型の膜貫通ドメインについてはほぼ全て赤く染まっているが、画像中に矢印で示されたβ-バレル型の膜貫通ドメインを持つタンパク質では青い（T よりも A が多い）ドメインが目立った。（Figure 4 矢印） A よりも T が多いのはわずかに含まれるβ-バレル型ではなく、多数を占めるα-ヘリックス型の膜貫通ドメインであると考えられた。解析した膜貫通ドメイン全体における T/AT は、平均値、中央値とも 0.70 であり、膜貫通ドメインにおいては T と A の数の比が 7:3 程度となることが多いと考えられた。また、1 回膜貫通型タンパク質の膜貫通ドメインのみで集計したところ、平均値、中央値とも 0.75 であった。今回集計した 36 回膜貫通型から 1 回膜貫通型まで、全ての膜貫通ドメインは全体に赤い色となっており、T/AT は一様に高いと考えられた。（Supplemental Figure）

※Figure 4 のオリジナルについては、そのファイルサイズが大きいため、Supplemental Figure として別に添付した。

以上より、全ての膜貫通ドメインは T/AT が高くなっていることが示された。

それでは、T/AT が高い部分は必ず膜貫通ドメインになるのだろうか。これについて更に別の解析を行った結果を示す。(Figure 5)

Figure 5

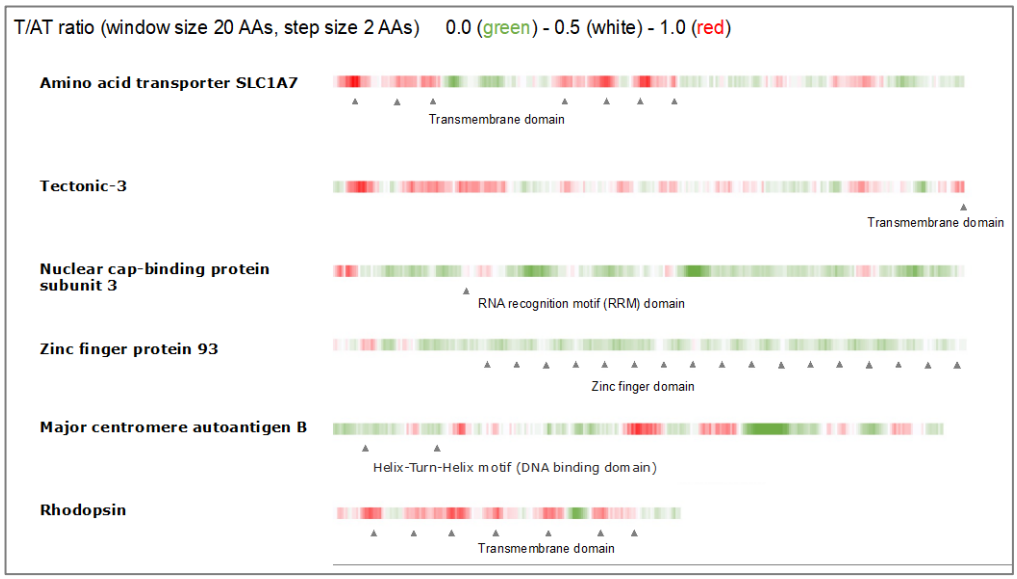


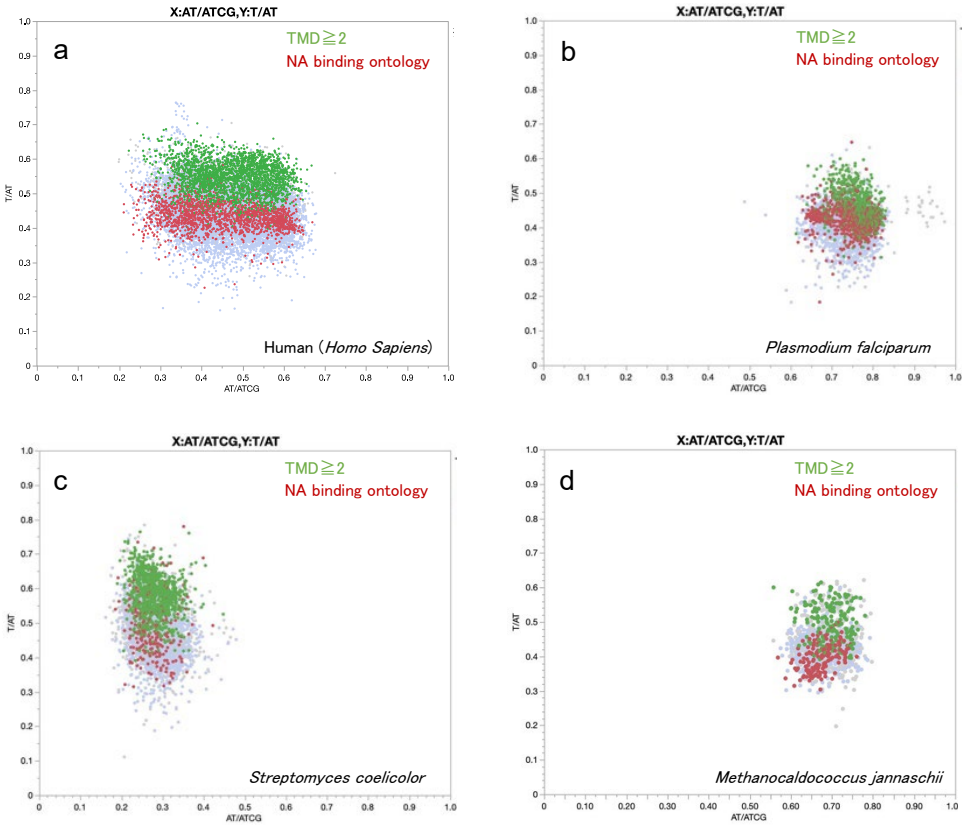
Figure 5 は、ヒトのタンパク質から同程度のサイズ(600 アミノ酸残基程度)のタンパク質を複数選び、その遺伝子中の T/AT を部分ごとに計算し着色したものである。連続する 20 残基の中に含まれる A および T の個数より T/AT を計算して着色表示しており、2 残基ごとにその対象をずらして計算している。計算された値に応じて、0 に近ければ緑、0.5 で白、1 に近ければ赤となるように着色した。各膜貫通ドメインの開始位置にマーク(▲)したところ、最上段のアミノ酸トランスポーターSLC1A7 では 7 つの膜貫通ドメインがすべて、最も T/AT が高い部位に一致して存在していることが明らかとなった。一方で、2 段目の Tectonic-3 蛋白質は 1 回膜貫通型であるが、同部位よりも T/AT が高い部位が別に存在していた。これらより、T/AT が高い(A よりも T が多い)ことは、膜貫通ドメインに相関する条件ではあるが、絶対条件ではないことが示された。また、膜貫通ドメインを形成するためには、ドメインが疎水性の残基で構成されることの他に、ドメイン以外の部分に親水性の残基が多い(少なくない)ことも必要であると考えられる。つまり、膜タンパク質においては、T/AT が多いだけでなく、そのコントラストが重要なのであろうと推測された。

Figure 5 においては、他の機能性ドメインについても検討を行っている。3、4、5段目のタンパク質はそれぞれ核酸に結合する RRM motif、Zinc finger domain、Helix-Turn-Helix motif を持つタンパク質であり、これらについてもそれらの該当部位にマーク(▲)を表示した。マーク位置の T/AT を確認するとその値は 0.5 よりも 0 に近い(T よりも A が多い)様子であった。核酸と結合するためには塩基性のアミノ酸残基(リシン、ヒスチジン、アルギニン、トリプトファン)が必要であるが、このうちリシン、ヒスチジンとアルギニンの一部コドンが遺伝暗号表における A に相関する位置にあることも無関係ではないと推測された。(Figure 1d 参照)

（最集段の Rhodopsin も確認のため描画したが、この膜貫通ドメインも T/AT 高値であった。）

さらに追加の検討として、ヒト以外の生物についての検討、および、上記、核酸結合ドメインの T/AT についての検討を行った。（Figure 6）

Figure 6



様々な GC 含量（AT/ATCG とほぼ同義）をとる生物 4 種類について、縦軸を T/AT、横軸を AT/ATCG として、膜貫通ドメインを二つ以上持つタンパク質を緑、核酸結合の Gene Ontology があるものを赤、それ以外を薄い水色でプロットした。（Figure 6a-d）

Figure 6a はヒト（*Homo sapiens*）、6b はマラリア原虫（*Plasmodium falciparum (isolate 3d7)*）、6c は放線菌（*Streptomyces coelicolor (strain ATCC BAA-471/ A3(2) /M145)*）、6d は超好熱性メタン菌（*Methanocaldococcus jannaschii (strain ATCC 43067)*）のタンパク質のプロットである。

今回検討した 4 種のプロットを検討した結果、いずれの種においても AT/ATCG の値に関わらず、T/AT が膜タンパク質（TMD≧2）や核酸結合能力のあるタンパク質（NA binding gene ontology）と相関している可能性が示された。もちろん今回検討したのは一部の生物種のみであるが、これらの生物種は最も極端な GC 含量をとる生物として選択した種であり、それ以外の生物はこれらの生物の中間の値を取る可能性が高い。これらを踏まえると、生物は T/AT を用いてこれらのタンパク質の合成をサポートしている可能性が高いと考えられた。

過去の報告において、Thymine の割合が多い遺伝子からは膜貫通ドメインが生成されやすい、と結論した論文はすでにあったが[6]、遺伝暗号表が T よりも A が多いことで膜貫通ドメインを生成する支援をしている可能性に言及したのは、検索した限り本論文が最初である。

シャルガフの法則（第 2 則）では、遺伝子上の A と T の核酸比は全体としてバランスする傾向があるとされている[7]。しかし局所的に見れば A と T の存在比にはゆらぎが存在し、生命はそのゆらぎを増幅することで機能性のあるタンパク質合成を支援していると推測した。もしそれが正しいのであれば、遺伝子上に認められる A と T の存在比のゆらぎは、より有利な表現型と相関することで自然選択において選択され続けて現在まで引き継がれているのかもしれない。また、A よりも T が多い遺伝子の DNA 上の相補鎖は T よりも A が多い配列となるが、Figure 5 の Zinc Finger Protein 93 では、膜貫通ドメインと同じようなサイズ（スケール）で T よりも A が多い部分が連続しており、これらは相互に遺伝子配列のゆらぎの根源を示している可能性もあると考えられた。これらは見方を変えれば、進化のさざなみによって刻まれた、遺伝子上の波打ち際の縞模様のようなものなのかもしれない。

一方で、Figure 6 において、AT/ATCG の変動は T/AT とタンパク質ドメインの作り分けには全く関与していなかった。筆者は以前の論文において、遺伝暗号表配列は、遺伝子の GC 含量（AT/ATCG = AT 含量と連動）の変動の影響を相殺し、安定したアミノ酸組成のタンパク質を生成することができるような配列になっており、これが遺伝暗号表の機能の一つと考えられることを示している[8]。これもあわせて考えると、「遺伝子配列から AT/ATCG の影響を受けずに T/AT によって機能的タンパク質ドメインの作り分けをできること」それ自体が、遺伝暗号表の配列が実現している基本的な機能であると推測した。

遺伝暗号の起源については、エラーや変異に対する堅牢性によって選択されてきたという考え方が一般的である[9]。しかし、生命の本質がランダムな変異からの生存しうる多様性の創出であることを考えると、堅牢性という現存機能の保存・維持ではなく、積極的な変異からの新機能の創出のほうが注目されるべきであると考える。その視点に立つと、遺伝暗号表の配列はランダムな変異から機能性のあるタンパク質を得る上で効率が最も良い方向に進化してきたという考えにも妥当性があるように思われる。むしろ、A よりも T が多いだけで膜貫通ドメインの生成が要易になるのであれば、これも一つの堅牢性と言えるのではないかと考えた。

まとめ

一般に膜タンパク質はゲノム上の全タンパク質の 3 割を占めると言われており、膜タンパク質の安定的な合成は生命の至上課題の一つであると推測される。そして、これらの膜タンパク質には共通して膜貫通ドメインが含まれるが、この膜貫通ドメインは疎水性の高いアミノ酸残基が一定数連続して配列する特殊な構造をしている。一般に、生物が進化の中で特定の機能性タンパク質を獲得していくためには、ランダムな変異を積み重ねた結果として偶然に望むタンパク質が生成されることを期待する必要があるとされてきたが、このように統制された構造をランダムに作り上げることは容易ではないと推測される。しかし、今回の解析結果から推測すると、遺伝子の核酸組成において A よりも T が多い遺伝子配列が利用できさえすれば、たとえ遺伝子変異がランダムであっても結果として膜貫通ドメインを形成するアミノ酸が生成される確率が格段に高くなることが明らかとなった。以上より、生物は遺伝暗号表の配列の特徴を利用して膜貫通ドメインの合成をアシストしている可能性があると考えられた。

結論

膜貫通ドメインの合成支援は遺伝暗号表配列の重要な機能の一つであり、現在の遺伝暗号表はこの機能も含めた様々な機能によって選択され続け、その結果として収束状態にあると推測した。

引用文献・サイト

1. Wnętrzak, M., Błażej, P., & Mackiewicz, P. (2019). Optimization of the standard genetic code in terms of two mutation types: Point mutations and frameshifts. *Bio Systems*, *181*, 44–50. https://doi.org/10.1016/j.biosystems.2019.04.012
2. 江角　元史郎. (2022). 必須アミノ酸の起源　細胞外マトリックス仮説. *Jxiv.* https://doi.org/10.51094/jxiv.121
3. National Center for Biotechnology Information (NCBI). (2022). Genome assembly GRCh38.p14. *National Library of Medicine (NIH) website.* https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.40/
4. UniProt consortium. (2022). Homo sapiens (species) Related UniProtKB entries. *UniProt website.* https://www.uniprot.org/uniprotkb?query=(taxonomy_id:9606)
5. 太田　亨, 新井　宏嘉. (2006). 組成データ解析の問題点とその解決方法. *地質学誌, 112*(3), 173-187. https://doi.org/10.5575/geosoc.112.173
6. Vakirlis, N., Acar, O., Hsu, B., Castilho Coelho, N., van Oss, S. B., Wacholder, A., Medetgul-Ernar, K., Bowman, R. W., Hines, C. P., Iannotta, J., Parikh, S. B., McLysaght, A., Camacho, C. J., O'Donnell, A. F., Ideker, T., & Carvunis, A.-R. (2020). De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. *Nature Communications*, *11*(1), 781. https://doi.org/10.1038/s41467-020-14500-z
7. Fariselli, P., Taccioli, C., Pagani, L., & Maritan, A. (2021). DNA sequence symmetries from randomness: The origin of the Chargaff's second parity rule. *Briefings in Bioinformatics*, *22*(2), 2172–2181. https://doi.org/10.1093/bib/bbaa041
8. Esumi, G. (2022). Synonymous codon usage and its bias in the bacterial proteomes primarily offset GC content variation to maintain optimal amino acid compositions. *Jxiv.* https://doi.org/10.51094/jxiv.99
9. Wichmann, S., & Ardern, Z. (2019). Optimality in the standard genetic code is robust with respect to comparison code sets. *Bio Systems*, *185*, 104023. https://doi.org/10.1016/j.biosystems.2019.104023