

Cistron の塩基組成の推定計算法 と分子進化機構への応用

大阪大学医学部遺伝学教室・院生

宇佐美 謙治

Statistical analysis of the base composition of cistrons
which correspond to protein molecules, using data on the
amino acid composition of protein.

Kenji USAMI

Department of Genetics, Faculty of Medicine, Osaka University
Joancho-33, Kitaku, Osaka, Japan

1. 緒言

「古典進化学」が各種生物の形態・機能といった定性的な表現型の比較を出発点としているのに対し、「分子進化学」は各種生物の生体分子の比較に立脚点を持つ。（この立場は分子遺伝学及び集団遺伝学の理論と結合することにより、進化を定量的に取り扱う事を可能にした。）

ところが今日までの分子進化の知識体系は殆んどが蛋白質のアミノ酸置換を中心に展開され、機能の相同な蛋白質分子の一次構造を各種生物間で比較する事に重点を置いてきたと言える。しかるにこうした蛋白質の組成・配列の差違の裏には常に遺伝子の関与があるはずである。従って次の目標は、特定の蛋白質を code する cistron の組成・配列の比較であらねばならない。ところが実験技術の発達にも拘らず cistron の塩基配列の決定は蛋白質ほどの成果を得てはおらず、せいぜい平均 GC 含量を細菌の分類に利用するか、置換部位で codon の変化を考察するか、後述する様に、全 DNA 中の GC, AT 量と全蛋白質中の各アミノ酸量の相関性を求める程度に止まらざるを得なくなっている。cistron の塩基配列についての実験事実が余りにも少ない今、分子進化

学は大きな壁に直面していると言えよう。しかし私達はこの現状の下でも進化に貢献する遺伝子変化を追求せねばならない。そこで筆者は既に配列の確定した蛋白質のアミノ酸組成から、その蛋白質を code している cistron の塩基組成を推定する計算式を提出した。ここでは genetic code が均等に使用されると仮定して計算式を導出し、実例をもって式の検定をし、その上幾種もある codon の利用が偏倚的である可能性を提示した。さらに応用として、幾らかの蛋白質群について、cistron の塩基組成に立脚点をおいた系統進化への寄与を企った。この計算式は上述の如く、分子進化に対応する遺伝子変化の機構の研究に一つの突破口を提供すると共に、1 アミノ酸に対する複数個の codon の使われ方についての知見をもたらすであろう。さらに現在、国内外で盛んに取組まれている核酸の塩基配列決定の実験に純度を推定したり GC 含量の最高最低値を推定し、それらの実験に一応の目安をつけるものとして意味を持っている。

II. 方法と検証¹⁷⁾

(i) 方法： 計算式の導出の前に核酸と蛋白質との相関性に関連した分子進化の研究について述べよう。かつて末岡^{1)-2), 19)} は塩基含量と genetic code との関係を見い出す試みとして、細菌 DNA の塩基分析結果を細菌全蛋白質のアミノ酸の割合と比較し、両者の間に、次の様な相互関係がある事を示した。① GC 含量は Ala, Arg, Gly, Pro の各含量と正相関する。② AT 含量は Ile, Lys, Asp + Asn, Glu + Gln, Tyr, Phe の各含量と正相関する。③ His, Val, Leu, Thr, Ser, Met は GC, AT のどの含量とも相関しない。(この知見は DNA の塩基組成と蛋白質のアミノ酸含量との間の直接関係を示すこれ迄の唯一の例である。) この種の研究は、その後、Genetic Code が確立すると codon とアミノ酸組成との相関性を求める集団遺伝学的研究に引き継がれた。³⁾⁻⁵⁾ しかしこれらの知見はいづれも、cistron と蛋白質との相関性を示すものではなかった。しかも最近 DNA のかなりの部分が蛋白質合成に参与しないという知見²²⁾ が数多く示され、従来、考えられていた程には gene の数は多くない事が確実になりつつある。この Non-functional DNA 部位の意義はともかく、蛋白質と遺伝子との関係を問題にする際、私達は Functional な遺伝子をまづ俎上に持ってこなければならない。

そこで筆者は既に配列の確定している蛋白質のアミノ酸組成から当該蛋白質を code している cistron の塩基組成を推定する計算法を提出し、蛋白質と cistron の間の相関性を明示しようとした。類似の試みが既に太田・木村⁶⁾によりなされていた。氏らの報告に比して、筆者は計算法の検証を配列決定の完了したMS2 coat protein で行ったこと、重点の1つが codon 利用の偏よりの可能性を求めている点にある事、それ故に、最も codon 利用の偏よりに影響を及ぼすと思われる codon の3字目の塩基量推定式が提出された点、そして生化学者が好んで用いる GC 含量に重点をおいた事さらには、式の利用に際して、Computer 技術を必要としない点、又、遠い目標を遺伝子変化から見た系統樹づくりと分子進化速度の計算に置いている点で、氏らの報告とことになっている。^{*}

それでは、式の導出法を示そう。まず仮定としてどの codon も均等に使用されているとする。そこで例えば、codon 1字目にUをもつアミノ酸に着目すると、code 表 (Table.1) から次のアミノ酸が抽出される。

Phe, Leu, Ser, Tyr, Cys, Trp

今、ある蛋白質(又は相同蛋白質群)のアミノ酸組成が得られたとする。この時、上述のアミノ酸の実数を [] で示すと

[Phe] [Leu] [Ser] [Tyr] [Cys] [Trp]

ここで、Leu, Arg, Ser等の「分散コード」を持つアミノ酸に対し、次の規定を設定する。

Leu _I (UUA, UUG)	Leu _{II} (CUX)
Arg _I (CGX)	Arg _{II} (AGA, AGG)
Ser _I (UCX)	Ser _{II} (AGU, AGC)

*) 「木村資生氏の承諾を得て、以下使用される記号は、氏らの論文で使用されたものに統一した。」

Table 1. THE GENETIC CODE

1st ↓ 2nd →	U	C	A	G	↓ 3rd
U	PHE PHE LEU LEU	SER SER SER SER	TYR TYR Ochre Amber	CYS CYS Opal TRP	U C A G
C	LEU LEU LEU LEU	PRO PRO PRO PRO	HIS HIS GLN GLN	ARG ARG ARG ARG	U C A G
A	ILE ILE ILE MET	THR THR THR THR	ASN ASN LYS LYS	SER SER ARG ARG	U C A G
G	VAL VAL VAL VAL	ALA ALA ALA ALA	ASP ASP GLU GLU	GLY GLY GLY GLY	U C A G

よって codon 1 字目の U の実数は次式で与えられる。

$$\Sigma [U1] = \{ [Phe] + [Leu]_1 + [Ser]_1 + [Try] + [Cys] + [Trp] \} \quad ①$$

次に codon 1 字目の U の出現頻度を求めよう。それには、① 式の右辺を全アミノ酸実数 (N) で割ればよく、左辺の [U1] を U1 と置換し、これを codon 1 字目の U の出現頻度とする。ここに② 式の両辺は出現頻度の算出式となり、① ② 式は等値

である。

$$U1 = \frac{1}{N} \{ [Phe] + [Leu]_I + [Ser]_I + [Tyr] + [Cys] + [Trp] \} \times 100 \quad (2)$$

ここで「分散コード」をもつアミノ酸についての補正式を出す。題意の仮定より、均等に codon が使用されているとすれば、

$$\left\{ \begin{array}{l} [Ser]_I = \frac{4}{6} [Ser] \\ [Ser]_{II} = \frac{2}{6} [Ser] \\ \\ [Arg]_I = \frac{4}{6} [Arg] \\ [Arg]_{II} = \frac{2}{6} [Arg] \\ \\ [Leu]_I = \frac{2}{6} [Leu] \\ [Leu]_{II} = \frac{4}{6} [Leu] \end{array} \right\} \cdots \cdots (3)$$

従って③ 式を② 式に代入し

$$U1 = \frac{1}{N} \{ [Phe] + \frac{2}{6} [Leu] + \frac{4}{6} [Ser] + [Tyr] + [Cys] + [Trp] \} \times 100 \quad (4)$$

以下、同様に、合計 12 個の方程式が求められる。さらに、GC 含量を計算する為に次式を導いた。

$$GC = (G1 + G2 + G3) + (C1 + C2 + C3) \text{ として}$$

$$GC = \frac{1}{3N} \{ [Met] + 2[Trp] + \frac{1}{2} ([Lys] + [Phe] + [Tyr] + [AsN]) \}$$

略号： U(Uracil) C(Cytosine) G(Guanine) A(Adenine)

X(Uracil, Cytosine, Guanine, Adenine)

$$\begin{aligned}
 & + \frac{3}{2} ([\text{Val}] + [\text{Asp}] + [\text{Glu}] + [\text{Gln}] + [\text{His}] + [\text{Thr}] \\
 & \qquad \qquad \qquad + [\text{Cys}] + [\text{Ser}]) \\
 & + \frac{5}{2} ([\text{Ala}] + [\text{Gly}] + [\text{Pro}]) + \frac{1}{6} (7[\text{Leu}] + 13[\text{Arg}]) \\
 & + \frac{1}{3} [\text{Ile}] \} \times 100
 \end{aligned}$$

全方程式は以下に要約した。

$$\begin{cases}
 \text{U1} = 1/\text{N} ([\text{Phe}] + 4/6[\text{Ser}] + 2/6[\text{Leu}] + [\text{Tyr}] + [\text{Cys}] + [\text{Trp}]) \times 100. \\
 \text{C1} = 1/\text{N} (4/6[\text{Leu}] + [\text{Pro}] + [\text{His}] + [\text{Gln}] + 4/6[\text{Arg}]) \times 100. \\
 \text{A1} = 1/\text{N} ([\text{Ile}] + [\text{Thr}] + [\text{Asn}] + [\text{Lys}] + 2/6[\text{Ser}] + 2/6[\text{Arg}] + [\text{Met}]) \times 100. \\
 \text{G1} = 1/\text{N} ([\text{Val}] + [\text{Ala}] + [\text{Asp}] + [\text{Glu}] + [\text{Gly}]) \times 100.
 \end{cases}$$

$$\begin{cases}
 \text{U2} = 1/\text{N} ([\text{Phe}] + [\text{Leu}] + [\text{Ile}] + [\text{Met}] + [\text{Val}]) \times 100. \\
 \text{C2} = 1/\text{N} (4/6[\text{Ser}] + [\text{Pro}] + [\text{Thr}] + [\text{Ala}]) \times 100. \\
 \text{A2} = 1/\text{N} ([\text{Tyr}] + [\text{His}] + [\text{Gln}] + [\text{Asp}] + [\text{Asn}] + [\text{Lys}] + [\text{Glu}]) \times 100. \\
 \text{G2} = 1/\text{N} ([\text{Cys}] + [\text{Trp}] + [\text{Arg}] + [\text{Gly}] + 2/6[\text{Ser}]) \times 100.
 \end{cases}$$

$$\begin{cases}
 \text{U3} = 1/\text{N} \{ 1/2 ([\text{Phe}] + [\text{Tyr}] + [\text{Cys}] + [\text{His}] + [\text{Asn}] + [\text{Asp}]) \\
 \quad + 1/3 ([\text{Ile}] + [\text{Ser}]) + 1/4 ([\text{Pro}] + [\text{Thr}] + [\text{Val}] + [\text{Ala}] + [\text{Gly}]) \\
 \quad + 1/6 ([\text{Leu}] + [\text{Arg}]) \} \times 100. \\
 \text{C3} = 1/\text{N} \{ 1/2 ([\text{Phe}] + [\text{Tyr}] + [\text{Cys}] + [\text{His}] + [\text{Asn}] + [\text{Asp}]) \\
 \quad + 1/3 ([\text{Ile}] + [\text{Ser}]) + 1/4 ([\text{Pro}] + [\text{Thr}] + [\text{Val}] + [\text{Ala}] + [\text{Gly}]) \\
 \quad + 1/6 ([\text{Leu}] + [\text{Arg}]) \} \times 100. \\
 \text{A3} = 1/\text{N} \{ 1/2 ([\text{Gln}] + [\text{Glu}] + [\text{Lys}]) + 1/3 ([\text{Ile}] + [\text{Leu}] + [\text{Arg}]) \\
 \quad + 1/4 ([\text{Pro}] + [\text{Thr}] + [\text{Val}] + [\text{Ala}] + [\text{Gly}]) + 1/6 [\text{Ser}] \} \times 100. \\
 \text{G3} = 1/\text{N} \{ [\text{Trp}] + [\text{Met}] + 1/2 ([\text{Gln}] + [\text{Glu}] + [\text{Lys}]) + 1/3 ([\text{Leu}] + [\text{Arg}]) \\
 \quad + 1/4 ([\text{Pro}] + [\text{Thr}] + [\text{Val}] + [\text{Ala}] + [\text{Gly}]) + 1/6 [\text{Ser}] \} \times 100.
 \end{cases}$$

$$\begin{aligned}
G1 + C1 = GC1 &= 1/N \{ [Val] + [Ala] + [Asp] + [Glu] + [Gln] + [Gly] \\
&\quad + [Pro] + [His] + 4/6 ([Leu] + [Arg]) \} \times 100. \\
G2 + C2 = GC2 &= 1/N \{ [Ser] + [Pro] + [Thr] + [Ala] + [Cys] + [Trp] \\
&\quad + [Arg] + [Gly] \} \times 100. \\
G3 + C3 = GC3 &= 1/N \{ [Trp] + [Met] + 1/2 ([Gln] + [Glu] + [Lys] + [Phe] \\
&\quad + [Tyr] + [Cys] + [His] + [Asn] + [Asp] + [Pro] + [Thr] \\
&\quad + [Val] + [Ala] + [Gly]) + 1/3 ([Ile] + [Ser] + [Leu] \\
&\quad + [Arg]) + 1/6 ([Leu] + [Arg] + [Ser]) \} \times 100. \\
G1 + G2 + G3 + C1 + C2 + C3 = GC &= 1/3N \{ 2[Trp] + [Met] \\
&\quad + 1/2 ([Lys] + [Phe] + [Tyr] + [Asn]) \\
&\quad + 3/2 ([Val] + [Asp] + [Glu] + [Gln] + [His] + [Thr] \\
&\quad + [Cys] + [Ser]) \\
&\quad + 5/2 ([Ala] + [Gly] + [Pro]) \\
&\quad + 1/6 (7[Leu] + 13[Arg]) + 1/3[Ile] \} \times 100.
\end{aligned}$$

ii) 検証：この式の有効性は、今後 cistron の塩基配列が、その産物の蛋白質の一次構造と対応できる例が集積して初めて明確になるのであろうが、現実のところ、検証の要求に耐え得る実例は Phage MS2 の Coat Protein Cistron⁷⁾ のみである。従って検証は、この Cistron を使用してなされた。(Table 2. 3., Fig. 1. 2)

この検証から次の事が明らかとなる。1) codon 1, 2字目の各塩基含量に関して計算値と実測値は近接し、特に2字目に関しては誤差を含まない。一方3字目は前2者ほどの近接は示さない。(Fig. 1) 2) GC 含量に関して GC1, GC2, GC3, GC のいずれも計算値と実測値とが極めて近接している。(Fig. 2) 3) U, A の各全量は計算値・実測値の一致が見られるが、G, C は一致度は低く、G, C 交互で相補的である。(G の実測値は C の計算値と、そして C の実測値は G の計算値と近似している。)*)

Table 2

Base composition (%) in the coat protein cistron of phage MS2 RNA as obtained by direct biochemical analysis and as calculated by statistical analysis of the amino acid composition of the coat protein of coliphage MS2.

codon		U	C	A	G
1 st.	Expected	17.829 %	14.987 %	31.525 %	35.659 %
	Observed	17.054 %	17.054 %	30.233 %	35.659 %
	χ^2	0.615 (0.90 > p > 0.80)			
2 nd.	Expected	27.132 %	29.199 %	27.132 %	16.537 %
	Observed	27.132 %	29.457 %	27.132 %	16.279 %
	χ^2	0.000 (p = 1.00)			
3 rd.	Expected	26.227 %	26.227 %	23.256 %	24.290 %
	Observed	28.682 %	34.109 %	20.155 %	17.054 %
	χ^2	6.352 (0.10 > p > 0.05)			

	Obs.	Exp.	χ^2
GC 1	52.713 %	50.646 %	0.154 (0.95 > p > 0.90)
GC 2	45.736 %	45.736 %	
GC 3	51.163 %	50.517 %	
GC	49.871 %	48.966 %	

Table 3

Molar proportion of bases in the coat protein cistron of coliphage MS2 RNA

	U (%)	C (%)	A (%)	G (%)	$\frac{C}{G}$	$\frac{A}{U}$	$\frac{A+G}{C+U}$	$\frac{A+C}{G+U}$	$\frac{A+U}{G+C}$
obs.	24.289	26.873	25.840	22.997	1.169	1.064	0.955	1.115	1.005
cal.	23.729	23.471	27.304	25.495	0.921	1.151	1.119	1.032	1.042
χ^2	0.308	3.415	0.631	2.613	—	—	—	—	—

U : 0.90 > p > 0.80 C : 0.20 > p > 0.10
 A : 0.80 > p > 0.70 G : 0.30 > p > 0.20

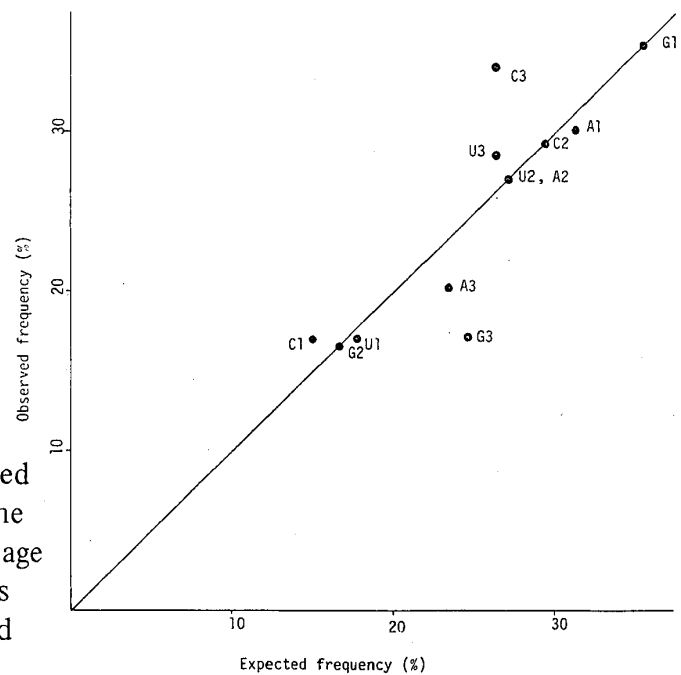


Fig. 1. Correlation between the observed and expected frequencies of bases in the complete coat protein cistron of coliphage MS2 RNA. The straight line represents an idealized equality of expectation and observation.

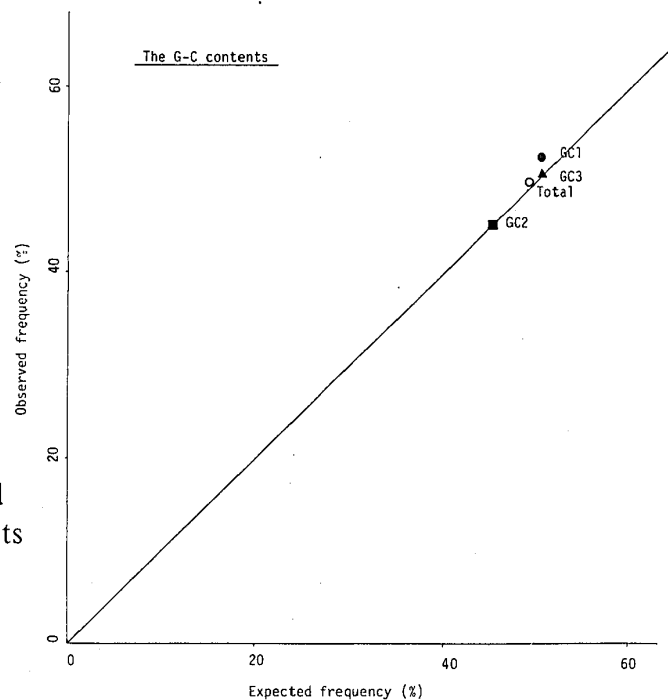


Fig. 2. Correlation between the observed and expected frequencies of G-C contents of complete coat protein cistron in coliphage MS2 RNA. The straight line represents an idealized equality of expectation and observation.

Ⅲ. 考 察

上述の結果から次の事が考察される。(1) 3 字目の計算値と実測値間の誤差の原因は、計算式が codon の均等使用の仮定に立っているのに対して、検証の材料 MS2 coat protein cistron 自身が codon を不均等に使用している矛盾点にある。^{*}) それでは 3 字目にのみ差が出て 1, 2 字目になぜ値の一致を見たのであろうか? それは、code 表から推察出来るように 1, 2 字目はいわば「共通項」とも言うべきもので、指定範囲が広く、少数の「分散コード」のアミノ酸を除き、どのように不均等な codon の使用があっても、指定される塩基は 1 種である事による。一方 3 字目はいわば「選択項」とも言うべきもので指定範囲は限定されるので、不均等に codon が使用されれば、計算値は非使用の codon の方へ「空費」されるからである。そこで cistron の塩基配列が決定された時、使用 codon が均等使用の範囲からどの程度隔っているかを定量的に表示する応用面がこの計算式から導びかれる。(2) 次の問題は、果して codon の不均等使用が一般的なのか、(一般化できれば、計算式の変更は余儀なくされる) それとも MS2 の場合は特殊で均等使用が普通であるのかどうかという点である。この件に関しては余りにも実験事実が乏しく、我々は集団遺伝学や情報理論の力を借ねばならない。それでは codon 使用に偏倚性 (Bias) をもたらす可能な機構にどのようなものが考えられるだろうか? 第 1 に t-RNA と mRNA との対向である。この対向についての要約を示すと; ① 情報コードは縮退していて 1 種のアミノ酸は 1 種以上の code を持つ。② この codon の多様性は特定アミノ酸に対し各々特異性の異なる複数個の t-RNA が存在する事に依る。③ 1 つの t-RNA は 1 つ以上の codon に対向できる anticodon を持つ。これは codon-anticodon 間の“Wobble pairing”を形成する事で可能となる。④ 生物種の差違により、ある特定の 1 アミノ酸に対応する複数個の t-RNA 同志は比較的な量に差がある。ある生物では該当の t-RNA が複数の指定 codon の 1 部を全く利用しないか、稀にしか利用しない。この性質は種特異的である。¹²⁾⁻¹³⁾ (Table 4) ⑤ t-RNA の比較的量は生理状態の変化により変動する場合もある。¹⁴⁾ 今、この内で

^{*}) 「この不均等利用に対する理由づけとして、Grantham⁸⁾ 及び Ball⁹⁾ の報告がある。」

Table. 4.

Species dependent differences in response of AA-tRNA to trinucleotide codons.¹³⁾

codon		t-RNA		
		E. coli	Xenopus laevis	Guinea pig liver
Arg	AGG	±	++++	+++
	CGG	±	++++	++++
Met	UUG	++	±	±
Ala	GCG	++++	±	++
Ile	AUA	±	++	++
Lys	AAG	±	++++	++++
Ser	UCG	++++	±	++
	AGU	±	+++	+++
	AGC	±	+++	+++
Cys	UGA	±		+++

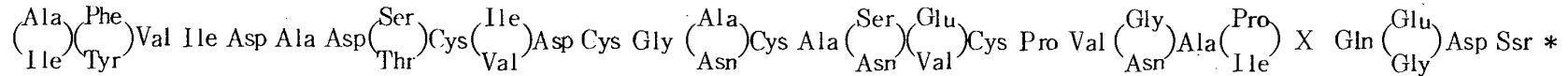
The following scale indicates the approximate response of AA-tRNA to a nucleotide relative to the responses of the same AA-tRNA preparation to all other trinucleotides for that amino acid (except Gly-tRNA which was assayed only with GGU and GGC).

++++	70 - 100 %
+++	50 - 70 %
++	20 - 50 %
±	0 - 20 %

問題にすべきは④に示される例である。もしもこの種の研究成果が多数集積すれば計算式はより緻密な式に改良されるだろう。(ちなみに、現在迄に、16 アミノ酸に対する49種の t-RNA が分離され、一次構造が明らかになっているので、codeが全生物で普遍的であるという前提の下に、Wobble pairing を含めて anticodon と codon を対応させて見ると表5に示すようになる。(Table 5). 対向の可能性があるとは言え Wobble pairing の対象となる対向がどのような使用分布をとるかについての実験事

Ferredoxin

- 1 5 10 15 20 25
- (1) Ala Tyr Lys Ile Ala Asp Ser Cys Val Ser Cys Gly Ala Cys Ala Ser Glu Cys Pro Val Asn Ala Ile Ser Gln Gly Asp Ser (f1)
- Ile Phe Val Ile Asp Ala Asp Thr Cys Ile Asp Cys Gly Asn Cys Ala Asn Val Cys Pro Val Gly Ala Pro Val Gln Glu* (f2)
- (2) Ala Phe Val Ile Asn Asp Ser Cys Val Ser Cys Gly Ala Cys Ala Gly Glu Cys Pro Val Ser Ala Ile Thr Gln Gly Asp Thr (f3)
- Gln Phe Val Ile Asp Ala Asp Thr Cys Ile Asp Cys Gly Asn Cys Ala Asn Val Cys Pro Val Gly Ala Pro Asn Gln Glu * (f4)
- (3) Ala Tyr Val Ile Asn Asp Ser Cys Ile Ala Cys Gly Ala Cys Lys Pro Glu Cys Pro Val Asn Ile Gln Gln Gly Ser (f5)
- Ile Tyr Ala Ile Asp Ala Asp Ser Cys Ile Asp Cys Gly Ser Cys Ala Ser Val Cys Pro Val Gl Ala Pro Asn Pro Glu Asp * (f6)



Ferredoxin

(1) *Clostridium pasteurianum*(2) *C. butyricum*(3) *Micrococcus aerogenes*

実が必要であろう。) 不均等をもたらす第2の可能性はNucleotideの生合成の難易にある。MS2についてGrantham⁸⁾, Ball⁹⁾が示した如く, Pyrimidine (C, U, T), Purine (G, A)の生合成の難易が自然淘汰に直結するという事が確実であれば, codon 使用に影響力を持った機構となる可能性をもつ。次に触られねばならぬ点は, DNA上の塩基分布が分節的である事である。山岸²³⁾⁻²⁵⁾は H_g^{++} がAT対と, A_g^+ がGC対と特異的に配位結合する性質を利用して, DNA断片の浮遊密度の差を大きくし, GC含量差の僅少な断片の分画に成功した。特定の遠心条件で断片の長さを最小1 cistron 相当の長さ(～1000ヌクレオチド対)にまで小さくしていった結果, *E. coli* DNAの塩基分布は不均一で分節的であると結論した。なお分節的となる理由として氏は, *M. lysodeikticus*, *Cl. perfringens* を使用しての別の実験的根拠²⁶⁾から, codonの3字目に塩基使用の偏りがあることによるという仮説を提出している。この知見に関連してcodon使用に偏りを生ぜしめる可能性のある事柄は

Table. 6

The frequency of occurrence of trinucleotide codon in H_{AA}^M maximum.¹⁰⁾

	T	C	A	G	
T	0.0299	0.0002	0.0217	0.0230	T
	0.0123	0.0046	0.0121	0.0201	C
	0.0136	0.0152	0.0122	0.0364	A
	0.0234	0.0050	0.0011	0.0323	G
C	0.0030	0.0075	0.0298	0.0037	T
	0.0000	0.0141	0.0143	0.0031	C
	0.0038	0.0232	0.0270	0.0054	A
	0.0168	0.0036	0.0204	0.0000	G
A	0.0306	0.0089	0.0283	0.0133	T
	0.0098	0.0171	0.0152	0.0118	C
	0.0189	0.0302	0.0263	0.0245	A
	0.0475	0.0000	0.0219	0.0130	G
G	0.0157	0.0071	0.0270	0.0135	T
	0.0030	0.0127	0.0146	0.0113	C
	0.0108	0.0229	0.0261	0.0206	A
	0.0240	0.0038	0.0192	0.0124	G

AT, GC rich DNA の存在である。このように DNA 塩基の不均等分布を示す DNA 断片が真の 1 遺伝子単位であるとすれば、(長さではなく、実際の機能している DNA 部位として)、その結果として codon の使用に偏った分布をもたらすであろう。又長谷川ら¹⁰⁾は分子進化を多様な蛋白質を生産する過程とみなし、情報理論の立場から塩基組成を推計した。その結果 DNA の塩基組成, codon に出現頻度の偏りがある事を示した (Table 6), この結果と MS2 での塩基分布 (Table 7) とを比較するとかなりの一致を見る。さらに Fitch¹¹⁾は種々の RNA phage の塩基配列から, codon 使用の Non-randomness の性質のある事を示している。このように codon の使用頻度については、まだ推察の域を出ず、実験例が殆んどない状態である。研究方法の検討が必要であろう。

Table. 7

Frequency of occurrence of codons found in the MS2 coat protein cistron,

	U	C	A	G	
U	$\begin{array}{l} 1 \\ 3 \\ 1 \\ 0 \end{array} \left\{ \begin{array}{l} \text{Phe} \\ \\ \text{Leu} \end{array} \right.$	$\begin{array}{l} 3 \\ 2 \\ 2 \\ 2 \end{array} \left\{ \begin{array}{l} \\ \text{Ser} \end{array} \right.$	$\begin{array}{l} 0 \\ 4 \\ - \\ - \end{array} \left\{ \begin{array}{l} \text{Tyr} \\ \\ \end{array} \right.$	$\begin{array}{l} 1 \\ 1 \\ - \\ 2 \end{array} \left\{ \begin{array}{l} \text{Cys} \\ \\ \text{Trp} \end{array} \right.$	$\begin{array}{l} \text{U} \\ \text{C} \\ \text{A} \\ \text{G} \end{array}$
C	$\begin{array}{l} 2 \\ 2 \\ 2 \\ 0 \end{array} \left\{ \begin{array}{l} \\ \text{Leu} \end{array} \right.$	$\begin{array}{l} 2 \\ 1 \\ 2 \\ 1 \end{array} \left\{ \begin{array}{l} \text{Pro} \end{array} \right.$	$\begin{array}{l} 0 \\ 0 \\ 1 \\ 5 \end{array} \left\{ \begin{array}{l} \text{His} \\ \\ \text{Gln} \end{array} \right.$	$\begin{array}{l} 3 \\ 1 \\ 0 \\ 0 \end{array} \left\{ \begin{array}{l} \text{Arg} \end{array} \right.$	$\begin{array}{l} \text{U} \\ \text{C} \\ \text{A} \\ \text{G} \end{array}$
A	$\begin{array}{l} 4 \\ 4 \\ 0 \\ 2 \end{array} \left\{ \begin{array}{l} \text{Ile} \\ \\ \text{Met} \end{array} \right.$	$\begin{array}{l} 4 \\ 4 \\ 0 \\ 1 \end{array} \left\{ \begin{array}{l} \text{Thr} \end{array} \right.$	$\begin{array}{l} 4 \\ 6 \\ 5 \\ 1 \end{array} \left\{ \begin{array}{l} \text{Asn} \\ \\ \text{Lys} \end{array} \right.$	$\begin{array}{l} 0 \\ 4 \\ 0 \\ 0 \end{array} \left\{ \begin{array}{l} \text{Ser} \\ \\ \text{Arg} \end{array} \right.$	$\begin{array}{l} \text{U} \\ \text{C} \\ \text{A} \\ \text{G} \end{array}$
G	$\begin{array}{l} 4 \\ 4 \\ 3 \\ 3 \end{array} \left\{ \begin{array}{l} \text{Val} \end{array} \right.$	$\begin{array}{l} 5 \\ 2 \\ 6 \\ 1 \end{array} \left\{ \begin{array}{l} \text{Ala} \end{array} \right.$	$\begin{array}{l} 1 \\ 3 \\ 2 \\ 3 \end{array} \left\{ \begin{array}{l} \text{Asp} \\ \\ \text{Glu} \end{array} \right.$	$\begin{array}{l} 3 \\ 3 \\ 2 \\ 1 \end{array} \left\{ \begin{array}{l} \text{Gly} \end{array} \right.$	$\begin{array}{l} \text{U} \\ \text{C} \\ \text{A} \\ \text{G} \end{array}$

IV. 応用¹⁸⁾

以上の検証によって提示された計算式は、使用出来るものとみなしてもよいので、分子進化機構への導入を企った。(紙面の都合の為、詳細は別の機会にする。) ① Respiratory Proteins (Cyt. C. 26 種; Cyt. C₃, Cyt. C₅₅₁, Cyt. b₅ 各 1 種): 著者はアミノ酸置換の少ないという理由と、一次構造が広範な種について判明しているという理由で最もよく分子進化機構の研究に使用される「呼吸蛋白」を選び cistron の比較に基づく系統樹作成及び、塩基置換ないしは GC 含量の差に基づく分子進化速度を測定しようと企図し、基礎資料を提示した。(Table 8), (木村氏²⁷⁾⁻²⁸⁾ はアミノ酸置換に基づいて、有名な、分子進化速度の計算式を提示している。それによると Cyt. C. は 0.3 pauling を示し、分子進化速度は、他の蛋白群に比し、おそい方に属すると結論している。ところで、もしも分子進化速度が塩基の変化を基にして与えられれば、アミノ酸置換を基にした値よりも差違を拡大して与えることが出来るのではなかろうか? なぜならば、コードは“同義語”を保有するので、アミノ酸の変化よりも、実際の塩基変化の方が多いため。著者はこの立場から分子進化速度の計算法を提示しようとしている。) 又、推定 GC 含量が当該生物の実測の DNA 平均 GC 含量とどの程度差を有するかを求めた。(Table 9) その結果、これらの推定 GC 含量は平均の実測 GC 含量にかなり近似していて、Cyt. C. Cistron は、実測の DNA 平均 GC 含量を代表する DNA 域である事が推察される。この知見と、GC 含量が一定値(高等動物の場合は 42%)に収束する事、¹⁾⁻²⁾ 推定 GC 含量がほぼ平均値 42% に近似する点、変化の少ない蛋白質である点は、考慮されねばならない。なお、現在、植物の Cyt. C. についての系統樹作成²⁰⁾²¹⁾ がとりくまれている。(植物は化石が少ない為、分類が動物ほどには進んでいない現状にある。) 上述の点を含め、cistron に基づいた分子進化機構の研究が必要である。 ② Coat Proteins (TMV, 3 種; f2, MS2, R17, ZR, fd, ZJ 2): RNA phage は血清学的性質や塩基組成分析、蛋白質一次構造等の類似性から、これら生物を分類する試みが初まっている。著者は cistron に基点を置いた比較分類を成すべく推定塩基含量を示した。(Table 10) ③ Ferredoxin. (Cl. pasteurianum, Cl. butyricum, M. aerogenes): 松原ら²⁹⁾ の研究によりこの蛋白質群の分子進化はかなり詳細なまでに研究されている。この蛋白質群は分子内

Table 8

Estimated base frequencies (%) in the first ,

Protein	Total number amino acids	U1	C1	A1
Cytochrome C				
Human	104	13.782	13.782	40.705
Bovine	104	13.141	14.743	38.462
Pig	104	13.141	14.743	38.462
Horse	104	12.500	14.743	41.026
Dog	104	13.141	14.743	38.462
Rabbit	104	13.141	13.782	38.462
California gray Whale	104	13.141	13.782	38.462
Kangaroo	104	12.180	13.141	41.026
King Penguin	104	14.743	13.141	40.385
Chicken	104	15.064	13.782	39.423
Pekin Duck	104	14.423	13.782	38.141
Snapping Turtle	104	12.820	12.820	39.743
Rattlesnake	104	14.743	13.462	41.026
Bullfrog	104	16.346	13.782	37.180
Turna Fish	103	16.182	13.916	35.922
Fuget sound Dogfish	104	16.026	17.949	36.218
Pacific lamprey	104	16.666	15.705	35.897
Silkworm Moths	107	15.888	15.888	31.776
Tobacco Horn Worm Moth	107	15.264	17.757	31.464
Fruit Fly	107	13.395	18.068	32.087
Screw-worm Fly	107	14.019	17.446	32.087
Wheat	112	18.154	19.346	29.464
<i>Neurospora crassa</i>	107	16.199	12.773	35.514
Baker's Yeast (2sp.)	108	16.975	16.049	36.419
<i>Candida krusei</i>	109	16.514	18.043	29.663
<i>Rhodospirillum rubrum</i>	112	16.964	10.119	33.631
Cytochrome C3				
<i>Desulfovibrio vulgaris</i>	107	16.511	14.953	31.152
Cytochrome C551				
<i>Pseudomonas fluorescens</i>	82	12.601	18.699	23.577
Cytochrome B5				
Bovine	85	16.079	17.647	27.451

second and third positions of the RNA codons

Base frequency (percent)								
G1	U2	C2	A2	G2	U3	C3	A3	G3
31.731	22.115	17.628	42.308	17.949	22.515	22.515	26.842	28.125
33.654	20.192	17.949	43.269	18.589	22.035	22.035	27.485	28.446
33.654	20.192	17.949	43.269	18.589	22.035	22.035	27.485	28.446
31.731	20.192	19.231	44.231	16.346	21.714	21.714	27.805	28.766
33.654	20.192	19.872	41.346	18.589	22.035	22.035	27.485	28.446
34.615	21.154	17.949	43.269	17.628	22.996	22.996	26.523	27.485
34.615	21.154	17.949	42.308	18.589	22.275	22.275	27.244	28.206
33.654	21.154	18.269	42.308	18.269	22.917	22.917	26.923	27.243
31.731	20.192	18.911	42.308	18.589	23.238	23.238	26.603	26.923
31.731	21.154	17.949	42.308	18.589	23.077	23.077	26.603	27.244
33.654	20.192	19.231	42.308	18.269	22.916	22.916	26.603	27.564
34.615	20.192	19.551	42.308	17.949	21.714	21.714	27.964	28.606
30.769	21.154	20.192	40.385	18.269	22.757	22.757	26.122	28.365
32.692	20.192	17.308	41.346	21.154	22.996	22.996	26.683	27.324
33.981	20.388	19.094	40.777	19.741	23.301	23.301	25.405	27.994
39.808	20.192	20.513	40.385	18.911	22.115	22.115	27.724	28.044
31.731	22.115	18.589	41.346	17.949	22.275	22.275	27.563	27.885
36.449	19.626	22.430	38.318	19.626	24.844	24.844	24.844	25.467
35.514	20.561	22.741	38.318	18.380	25.000	25.000	24.688	25.311
36.449	21.495	22.119	37.383	19.003	23.676	23.676	26.168	26.480
36.449	21.495	22.119	37.383	19.003	23.988	23.988	25.857	26.168
33.036	16.964	30.060	36.607	16.369	25.670	25.670	23.884	24.777
35.514	19.626	21.495	38.318	20.561	25.312	25.312	24.065	25.311
30.556	19.444	20.062	41.667	18.827	24.537	24.537	24.692	26.235
35.780	17.431	27.523	35.780	19.266	24.236	24.236	24.388	27.140
39.286	20.536	26.190	41.964	11.310	24.255	24.255	24.702	26.786
37.383	14.019	21.495	45.794	18.692	26.636	26.636	21.963	24.766
45.122	21.951	28.049	34.146	15.854	22.662	22.662	25.509	29.167
38.824	20.000	20.785	44.706	14.509	24.020	24.020	26.176	25.785

Table 9

Estimated G-C contents (%) in the first, second and third position in RNA codon, and total G-C contents, comparing with the experimental data of mean G-C contents per total DNA in a cell.

Protein	Base frequency (percent))				experimental data* (total DNA)
	GC1	GC2	GC3	GC	
Cytochrome C					
Human	45.513	35.577	50.640	43.910	39.4 (Liver)
Bovine	48.397	36.538	50.481	45.139	42.0 (Pancreas)
Pig	48.397	36.538	50.481	45.139	41.0 (Liver)
Horse	46.474	35.577	50.480	44.177	43.0 (Spleen)
Dog	48.397	38.461	50.481	45.780	41.9 (Testis)
Rabbit	48.397	35.577	50.481	44.818	44.4 (Liver)
California gray Whale	48.397	36.538	50.481	45.139	
Great grey Kangaroo	46.795	36.538	50.160	44.498	
King Penguin	44.872	37.500	50.161	44.178	
Chicken	45.513	36.538	50.321	44.124	41.7 (Liver)
Pekin Duck	47.436	37.500	50.480	45.139	
Snapping Turtle	47.435	37.500	50.320	45.085	
Rattlesnake	44.231	38.461	51.122	44.605	
Bullfrog	46.474	38.462	50.320	45.085	43.7 (Liver), 42.3 (Ovaries)
Tuna Fish	47.897	38.835	51.295	46.009	
Puget sound Dogfish	47.757	39.424	50.159	45.780	
Pacific Lamprey	47.436	36.538	50.160	44.711	
Silkworm Moths	52.337	42.056	50.311	48.235	
Tobacco horn worm Moth	53.271	41.121	50.311	48.234	
Fruit Fly	54.517	41.122	50.156	48.598	39.8 (Whole), 42.3 (egg)
Screw-worm Fly	53.895	41.122	50.156	48.391	
Wheat germ	52.282	46.429	50.447	49.753	42.4-46.6 (germ)
<i>Neurospora creassa</i>	48.287	42.056	50.623	46.989	53.0-55.0, 53.7
Baker's Yeast (2sp.)	46.605	38.889	50.772	45.422	34.1-41.0
<i>Candida krusei</i>	53.823	46.789	51.376	50.663	38.0-39.6
<i>Rhodospirillum rubrum</i>	49.405	37.500	51.041	45.982	60.0-67.0
Cytochrome C3					
<i>Desulfovibrio vulgaris</i>	52.336	40.187	51.402	47.975	
Cytochrome C551					
<i>Pseudomonas fluorescens</i> , p6009	63.821	43.903	51.829	53.184	59.4-63.6, 63.0
Cytochrome B5					
Bovine	56.471	35.294	49.805	77.190	42.0 (Pancreas)

* Experimental data* were chosen from the data compiled on Handbook of Biochemistry (1970). It shows the mean G-C contents per total DNA in a cell.

Table 10

Estimated base frequencies (%) in the first, second and
third positions of the RNA codon.

Protein	Total number amino acids	Base frequency (percent)											
		U1	C1	A1	G1	U2	C2	A2	G2	U3	C3	A3	G3
Coat Protein													
TMV <i>Vulgare</i>	158	19.409	21.097	29.114	30.380	27.215	30.802	25.316	16.666	26.687	26.687	23.313	23.313
TMV <i>OM</i>	158	19.409	21.097	28.481	31.013	27.215	30.169	25.949	16.666	26.793	26.793	23.101	23.313
TMV <i>Dahlemense</i>	158	20.253	21.941	28.692	29.114	27.848	29.536	27.215	15.401	26.003	26.003	23.471	24.525
Bacteriophage f2	129	18.088	15.504	31.525	34.884	27.132	29.199	27.132	16.537	26.357	26.357	23.515	23.773
Bacteriophage MS2	129	17.829	14.987	31.525	35.659	27.132	29.199	27.132	16.537	26.227	26.227	23.256	24.290
Bacteriophage R17	129	17.829	14.988	32.300	34.884	27.132	29.199	27.132	16.537	26.227	26.227	23.256	24.290
Bacteriophage ZR	129	17.829	14.988	32.300	34.884	27.132	29.199	27.132	16.537	26.227	26.227	23.256	24.290
Bacteriophage fd	49	19.049	6.802	29.251	44.898	28.571	31.973	26.531	12.924	25.000	25.000	24.320	25.682
Bacteriophage ZJ2	49	19.049	6.802	27.210	46.939	28.571	31.973	26.531	12.924	25.000	25.000	24.320	25.682

Cistron の塩基組成の推定計算法と分子進化機構への応用

に反復構造を有し、現分子に至る迄に遺伝子重複を主因として進化して来たと思われる。著者は松原らとは独立に、反復構造を示し、この構造から原始 Ferredoxin の一次構造を推定し、さらにその上で、反復分子 (f1, f2 f6), 全分子, 原始型分子の各 GC 含量を計算した。遺伝子の重複過程は GC 含量そのものに大きな変動を伴っていないと思われる。(Table 11) ④ Clupeine (Z, YI, YII) : Black ら¹⁵⁾ 及び Fitch¹⁶⁾ により、この分子の原始型は Ala-(Arg)_{4~5} であり、分子進化の過程で数回の重複・置換・挿入を受け現分子になる可能性が示されている。(Fig. 3. 4.)。著者は彼らの示す過程で GC 含量がどのような変化を示すかを計算した。その結果 (i) 各過程で GC 含量の変動は僅少である。(この結果は、分子レベルの進化は、形態レベルの進化ほどには著明ではないという木村らの指摘に一致する。) (ii) YI は Z, YII と比べ、進化過程で早く分離したと考えられる。(この結果は Black ら¹⁵⁾ の結論と一致している。) (Table 12)

Table 11. Estimated G-C contents (%) of Ferredoxin cistron
and the proposed repeating fragments cistrons.

Ferredoxin	N	GC1	GC2	GC3	GC
<i>Clostridium pasteurianum</i>					
total	55	54.545	52.727	48.484	51.919
fragment 1 (fr 1)	28	50.000	60.714	48.808	53.174
fragment 2 (fr 2)	27	59.259	44.444	48.148	50.617
<i>Clostridium butyricum</i>					
total	55	56.364	52.727	48.787	52.626
fragment 3 (fr 3)	28	53.571	60.714	48.809	54.365
fragment 4 (fr 4)	27	59.259	44.444	48.765	50.823
<i>Micrococcus aerogens</i>					
total	54	53.704	53.704	48.148	51.852
fragment 5 (fr 5)	26	50.000	50.000	48.077	49.359
fragment 6 (fr 6)	28	57.143	57.143	49.286	54.524

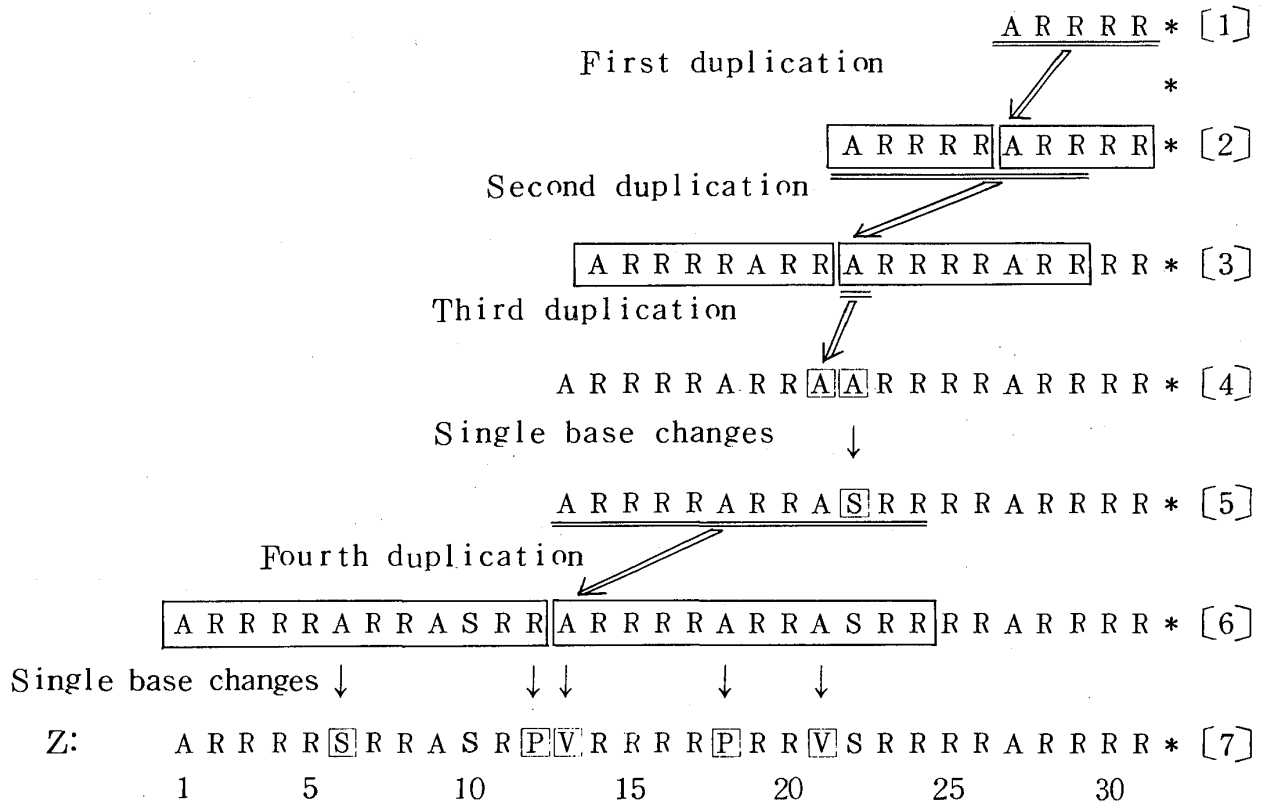


Fig. 3 A proposed model for the evolution of clupeine Z (by Black and Dixon)¹⁵⁾. The segment of polypeptide underlined corresponds to that portion of the structural gene which is partially duplicated and the resulting new polypeptide sequence is enclosed in a box. Single arrows represent single base changes while the double arrows show the partial gene duplication.

V. 要 約

(1) 配列確定のなされた蛋白質のアミノ酸組成から、その蛋白質を code する cistron の塩基組成を推定する計算式を提出した。② この計算式の検証に塩基配列の確定した MS2 coat protein cistron が使用された。③ その結果、codon 1, 2 字目は実測値との一致を示した。3 字目については前者ほどの一致が示されなかった。又 GC 含量については実測値と極めて密接した値を示し、最も実用できる。④ 元来、偏倚的な codon 使用をしている MS2 を使用しての検証にも拘らず、1, 2 字目が一

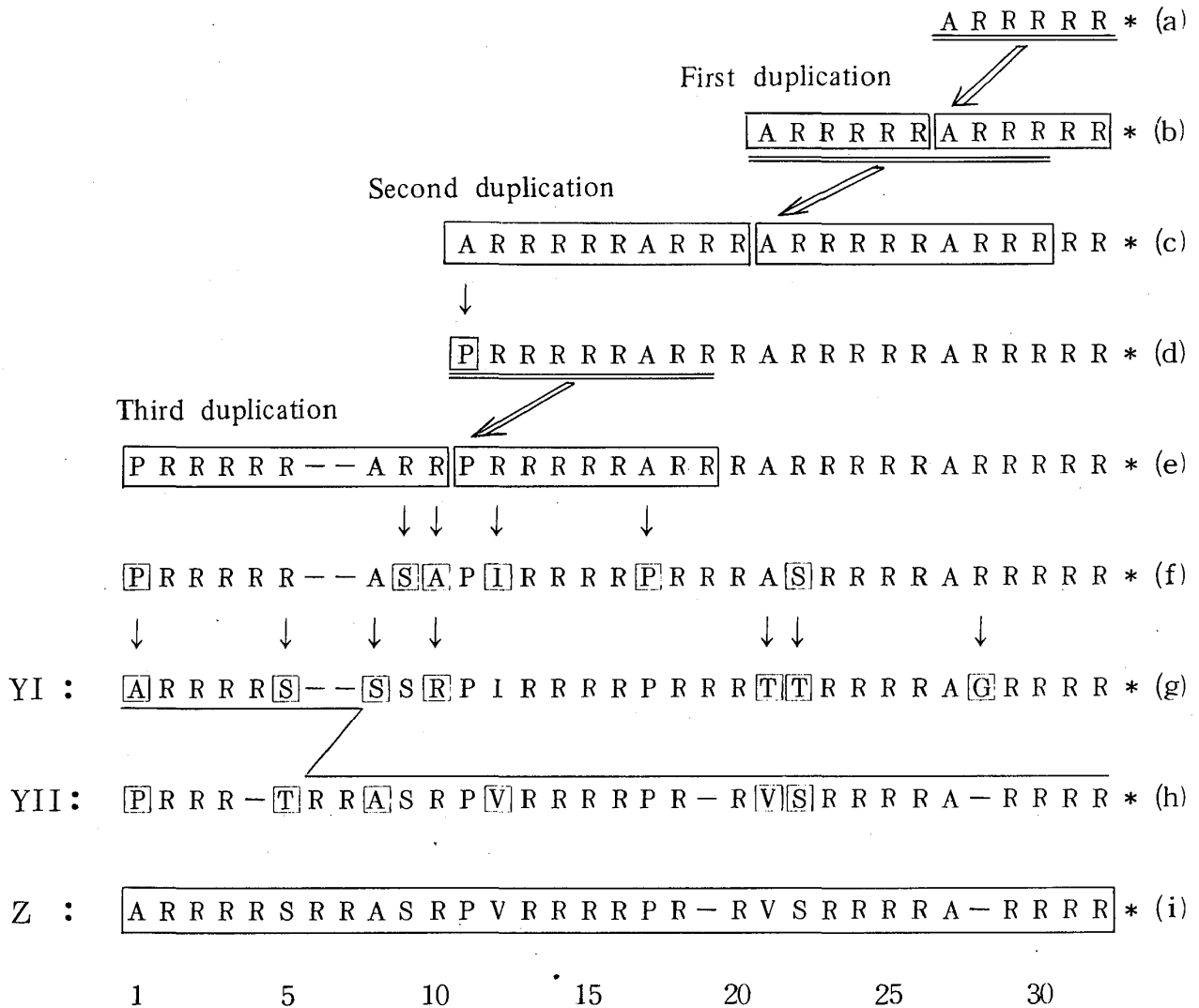


Fig. 4 A possible history of clupeines depicting clupeine Z as a crossover product (by Fitch)¹⁶. The segment of polypeptide underlined corresponds to that portion of the structural gene which is partially duplicated and the resulting new polypeptide sequence is enclosed in a box. Single arrows represent single base changes except to replace R to A after the third duplication, while the double arrows show the partial gene duplication. The single line between YI and YII indicates the amino-acids that compose the Z sequence.

Table 12. Estimated G-C contents (%) of cistrons
which code for clupeines and their ancestral
peptides.

Clupeine	N	GC1	GC2	GC3	GC
sequence 1 [1]	5	73.3333	100.000	49.9999	74.4444
sequence 2 [2]	10	73.3333	100.000	49.9999	74.4444
sequence 3 [3]	18	74.0740	100.000	49.9999	74.6913
sequence 4 [4]	19	75.4385	100.000	49.9999	75.1461
sequence 5 [5]	19	70.1754	100.000	50.0000	73.3918
sequence 6 [6]	31	69.8924	100.000	50.0000	73.2974
Z [7]	31	67.7419	93.548	50.0000	70.4300
sequence a (a)	6	72.2222	100.000	49.9998	74.0740
sequence b (b)	12	72.2222	100.000	49.9998	74.0740
sequence c (c)	22	72.7272	100.000	50.0000	74.2424
sequence d (d)	22	68.1818	100.000	50.0000	72.7272
sequence e (e)	31	73.1182	100.000	49.9999	74.3727
sequence f (f)	31	66.6666	96.774	49.4623	70.9677
YI (g)	31	59.1397	96.774	49.4622	68.4587
YII (h)	31	67.7766	93.333	49.9996	70.3699
Z (i)	31	67.7419	93.548	50.0000	70.4300

致し、3字目が一致しにくい理由を述べた。⑤ 考察として、1 アミノ酸の有する複数個の code が実際に均等使用されているか否かを、不均等な利用のある可能性に立ってその条件を示した。⑥ 応用として、respiratory proteins ; Coat proteins の“cistron Phylogenic Tree”作成及び分子進化速度の計算の準備をした。⑦ さらに応用として、Ferredoxin につき、分子内反復構造を提示し、それから原始型 Ferredoxin の構造を推定、GC 含量の変動の有無について検討した。⑧ さらに clupeine について、既に提出されている原子型 clupeine から現分子に至る迄の進化過程を重複等に伴う GC 含量の変化の有無で調べた。現在のところ、重複には GC 含量の大きな変動を伴わないという結論を持っている。⑨ 又、提示された計

宇佐美謙治

算式の応用として, *cistron* の塩基配列が決定された時, 使用 *codon* が均等使用の範囲からどの程度隔っているかを定量的に表示する事が出来る。

(謝辞: 本稿を書くにあたり, 影ながら研究を支援して下さいった吉川秀男博士及び, 資料の提供と論文作成に参画された 関得一郎博士, さらに統計処理を援助して下さいった 広吉寿樹博士に感謝する。又この「分子進化機構に関するシンポジウム」に招待して下さいった, 寺本英博士及び, 論文作成の際に議論に加わって下さいった山岸秀夫博士に深謝する。さらに, 記号等の使用を快諾して下さいった木村資生博士に感謝する。

本稿に納めた資料はすべて上述シンポジウムにて発表したものである。なお紙面の都合で, 十分に記述出来なかった分は, シンポジウムでの討論からの成果をふまえてあらためて原報^{17) 18)} として発表する。)

参 考 文 献

- 1) Sueoka, N. : Correlation between base composition of Deoxyribonucleic acid and Amino acid composition of protein. *Proc. N. A. S.*, **47**, 1141-1149 (1961).
- 2) Sueoka, N. : Compositional correlation between Deoxyribonucleic acid and Protein. *Cold Spring Harbor Symp. on Quant. Biol.*, **26**, 35-43 (1961).
- 3) Kimura, M. : Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetic Res.*, **11**, 247-269 (1968).
- 4) King, J. L. and Jukes, T. H. : Non-Darwinian Evolution. *Science*, **164**, 788-798 (1969).
- 5) Kikkawa, H. and Usami, K. : Optimization of the genetic code - the correlation between the frequency of amino acid in protein and the degeneracy of amino acid code. (unpublished). (Kikkawa has reported partially it in the Memorial Lecture for his Retirement, on Mar. 20, 1971.)
- 6) Ohta, T. and Kimura, M. : Statistical Analysis of the base composition of genes using data on the amino acid composition of proteins. *Genetics*, **64**, 387-395 (1970).
- 7) Min Jou, M., Haegeman, G., Ysebaert, M. and Fiers, W. : Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, **237**, 82-88 (1972).

- 8) Grantham, R. : Codon base randomness and composition drift in coliphage Nature New Biology, **237**, 265 (1972).
- 9) Ball, L. A. : Codon base randomness in bacteriophage MS2. Nature New Biology, **243**, 78 (1973).
- 10) Hasegawa, M. and Yano, T. : Informational entropy of DNA and evolution. in "Origin of life and evolution – the experimental approach –" (separate volume of "Protein, Nucleic acid and Enzyme"), 134–141, (ed. by Ueyama, K.) (Kyoritsu Shuppan Co. 1972.)
- 11) Fitch, W. M. : Considerations regarding the regulation of gene transcription and messenger translation. J. Mol. Evol., **1**, 185–207 (1972).
- 12) Söll, D., Jones, D. S., Ohtsuka, E., Faulkner, R. D., Lohrmann, R., Hayatsu, H., Cherayil, J. D., Hampel, A. and Bock, R. M. : Specificity of s RNA for recognition of codons as studied by the ribosomal binding technique. J. Mol. Biol., **19**, 556–573 (1966).
- 13) Nirenberg, M., Caskey, T., Marshall, R., Brimacombe, R., Kellogg, D., Doctor, B., Hatfield, D., Levin, J., Rottman, F., Pestka, S., Wilcox, M. and Anderson, F. : The RNA code and Protein Synthesis. Cold Spring Harbor Symp. on Quant. Biol., **31**, 11–24 (1966).
- 14) Doi, R. H. and Kaneko, I. : Transfer RNA patterns of *Bacillus subtilis* during sporulation and growth. Cold Spring Harbor Symp. on Quant. Biol., **31**, 581–582 (1966).
- 15) Black, K. A. and Dixon, G. H. : Evolution of protamin – a further example of partial gene duplication. Nature, **216**, 152–154 (1967).
- 16) Fitch, W. M. : Evolution of clupeine-Z, a probable crossover product. Nature New Biology, **229**, 245–247 (1971).
- 17) Usami, K. : Attempt to estimate the base composition of cistrons which correspond to proteins, using data on the amino acid composition of protein. (I) Statistical Method and Verification. (in preparation).
- 18) Usami, K. : Statistical analysis of the base composition of cistrons which correspond to protein molecules, using data on the amino acid composition of protein. (I) Statistical method and Verification. (II) An attempt to application of this method to the molecular evolution. (in preparation).

- 19) Sueoka, N. : On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. - N. A. S.*, **48**, 582-592 (1962).
- 20) McLaughlin, P. J. and Davhoff, M. O. : Eukaryoto evolution - a view based on Cytochrome c sequence data. *J. Mol. Evol.*, **2**, 99-116. (1973).
- 21) Fitch, W. M. : Is the fixation of observable mutations distributed randomly among the three nucleotide positions of the codon? *J. Mol. Evol.*, **2**, 123-136 (1973)
- 22) Crick, F. : General model for the chromosomes of high organisms. *Nature*, **234**, 25-27 (1971). Vogel, F. : *Nature*, **201**, 847 (1964).
Muller, H. J. : in "Heritage from Mendel" (ed. by Brink, R. A.), 419 (University of Winsconsin Press, 1967).
Crow, F. and Kimura, M. : "An introduction to population genetics theory" (Harper and Row, New York, 1970).
Ohno, S. : *Nature*, **234**, 134, (1971). Editorial : *Kagaku*, **43**, 455 (1973).
- 23) Yamagishi, H. : Nucleotide distribution in the DNA of *Escherichia coli*. *J. Mol. Biol.*, **49**, 603-608 (1970).
- 24) Yamagishi, H. : Non-random distribution of nucleotides in DNA. *Seibutsu Butsuri*, **10**, 223-235 (1970).
- 25) Yamagishi, H. : Regularity of Nucleotide distribution and Duplication of DNA. *Biological Science (Seibutsu Kagaku)*, **25**, 86-89 (1973).
- 26) Yamagishi, H. : Nucleotide distribution in bacterial DNA's differing in G + C Content. *J. Mol. Evol.* (in press).
- 27) Kimura, M. : The rate of molecular evolution considered from the standpoint of population genetics. *Proc. N. A. S.*, **63**, 1181-1188 (1969).
- 28) Kimura, M. : On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**, 18-25 (1971).
- 29) Matsubara, H. : Evolution of Ferredoxin. *Chemistry and Biology (Kagaku to Seibutsu)*, **10**, 700-708 (1972).
Tanaka, M., Nakashima, T., Benson, A., Mower, H. F. and Yasunobu, K. T. B. B.R. C., **16**, 422 (1964).
Fitch, W. M. : *J. Mol. Biol.*, **16**, 17 (1966).
Matsubara, H., Jukes, T. H. and Cantor, C. R. : *Brookhaven Symposia in Biology*, **21**, 201 (1968).