5G-01

Codon Code とアミノ酸についての考察

和田　平司　　松岡　保江　　三角田　秀実　　福田　美和　　林　郁枝

On the Amino Acid for Codon Code

Heiji Wada　Yasue Matuoka　Hidemi Misumida　Miwa Fukuda
Ikue Hayashi

(所属なし)

あらまし　Codon Code における４つのヌクレオチドから作られるアミノ酸の数は２０種類である。
開始 Codon AUG と終始 Codon UAG と UGA と UAA の４種類がある。
我々は何故６４種類の異なるアミノ酸ができないのか考察したので報告する。

キーワード　Codon Code　アミノ酸　符号理論　結合子

1.はじめに

４つのヌクレオチドから Codon Code は３つのヌクレオチドの結合によりアミノ酸ができる。

理論的には６４通りのアミノ酸ができることになるが。異なるアミノ酸は 20 種類である。

４つのヌクレオチド U，G，C，A は U(ウラシル)、G(グアニン)、C(シドシン)、A(アデニン)を 2bit のデジタル情報に当てはめると Codon Code と符号理論が導ける。

その Codon Code と符号理論を結びつけると 20 種類のアミノ酸しか一誤り訂正符号ができない。そのことを考察したので報告する。

2. 本論

2－1）64 通りのアミノ酸が理論的にはできる。

$$\left.\begin{array}{l} U=X=\text{“11”(ウラシル)} \\ G=V=\text{“10”(グアニン)} \\ C=I=\text{“01”(シトシン)} \\ A=\emptyset=\text{“00”(アデニン)} \end{array}\right\} \quad (1)$$

（1）に４つのヌクレオチドとギリシャ・ローマ数と２bit の状態について示したものである。

理論的には $4^3=4\times4\times4=64$

即ち６４通りのアミノ酸ができる。

2－２）Codon Code と符号理論符号長を n とする符号の符号理論について

k ＝を情報点数、mを検査点数 n を符号長とすると

$n=ik=im$　　…（２）

（２）なる符号理論が導ける。

i＝３とし、k ＝２とすると n ＝６なる符号長の符号が形ちづくられる。

$n=i\cdot(k_1x^1+k_0x^0)$

$=i\cdot(m_1x^1+m_0x^0)$　　…（３）

(2)なる符号が出来る。

2－３）符号距離について

$d\leqq n-k=n-m$　…（4）

このことについて　Codon code では

$d\leqq 6-2=4$ の符号が作られる。

2－4) 20 種類のアミノ酸について、符号距離から考えると d ＝3 の場合

ＵＧＡ、UAG、ＵＣＡ、UAC、ＧＧＧ、ＣＣＣ、ＣＧＧ、GCC の一誤り訂正及び検出しかできない。

６種のグループ×３で 18 種類のアミノ酸ができる。と CCC、GGG の２つで 20 種類の此のうち、開始と終始の Codon があり､20 種類の中に開始と終始は含まれている。

よってここに 20 種類のアミノ酸しかできない。

逆にこのことを言い換えれば我々の提唱する符号理論はCodon Codeの原理を的確に証明したことになる。

３．考察

3－1）符号距離について

一誤り訂正符号において符号距離は重要な意味を持つ。

此の場合、d ≤4 であるから、d が 4、3、2、1 について誤り検出一誤り訂正ができることを意味するが、

a) d ＝4 の場合の符号は

UUA、UCC、UGC、UCG、UGG となる。

ｂ) d ＝3 の場合の符号は

CCC、GGG、CGG、GCC、UGA、UAG、UCA、UAC

ｃ) d ＝2 の場合の符号は

ACC、AGG、AGC、ACG、AUA である。

ｄ) d ＝ 1 の場合の符号は誤り訂正はできない。符号行列が作れない。

e) d ＝0 の場合の符号は 0 である。

ここで、一誤り訂正及び一誤り検出に限定すると d ＝4 と d ＝2 の符号距離は 2 だから重ならない。但し 2 誤りの場合は重なる。

d ＝3 と d ＝1 の符号距離は 2 だから重ならない。

3－2） 符号で d ＝4 と d ＝2 を用いるか符号で d ＝3 と d ＝1 を用いるか？

A グループ　d ＝4 と d ＝ 2 では 15 個の符号が作られる。

B グループ d ＝3 では 20 個の符号が作られる。

ここで一誤り訂正、検出の場合は b グループの 18 個の少ない方を用いる此の 18 個でできるアミノ酸の種類は、6 ×3＝18 と 2 種類の CCC、GGG の Codon Code を足すと 20 種類

3－3） 開始 Codon と終始 Codon は開始 Codon＝AUG の一種類、終止 Codon＝UAG、UGA、UAA よって b グループは d ＝3 のグループであり 4 種類は含まれている。

∴18＋２＝20 種類

アミノ酸が 20 種類しかできない理由がここにあり、説明できた。

４．結論

我々は Codon Code において何故アミノ酸の種数が 20 種類しかできないかについて検討した。

その結果 Codon Code の一誤り訂正に必要な符号距離から検討した結果 d ＝3 の符号では 18 個の符号のグループが考えられる。ＣＣＧ、ＧＧＣ、ＵＧＡ、ＵＡＧ、ＵＣＡ、ＵＡＣとＣＣＣとＧＧＧの Codon で作られる種類しか一誤り検出及び訂正しかできない。20 種しかアミノ酸はできない。

ＣＣＧ、ＧＧＣ、ＵＧＡ、ＵＡＧ、ＵＣＡ、ＵＡＣの６種類×３とＣＣＣとＧＧＧの 2 種類で 18＋2＝20 種類である。

開始 Codon と終始 Codon は此の 20 種類の中に含まれている。よって符号距離の d ≤3 のアミノ酸のグループの種類しか 20 種類のアミノ酸しかない。普通に符号は d ≤4、d ≤3、d ≤2、で作られるが

種類は 20 種類である。

残された課題として符号における復号列が最小になる符号を作るにはどうすればいいか残されている。

参考文献

（１）　和田、三角田ら“ギリシャ・ローマ数の結合子について”FIT2018,第 17 回情報科学技術フォーラム、PP249~250（2018 年）

（２）　和田、大庭ら“IVPITEL は雨に乗って遣ってくる—ギリシャ・ローマ数について—,”第 83 回情報処理学会全国大会 PP 1 ～149、150（2021 年）

（３）　和田、林ら“IVPITEL は郭公の翼を広げいてる（其の一）－ギリシャ、ローマ数の符号理論－,FIT2022 情報科学技術フォーラム、PP117~121（2022 年）

（４）　若山正人“電子情報通信学会誌”vol160、№７、PP1280~1284（2017 年）

（５）　東京教育大学数学教育研究会“少年少女と算数の教育”小林書店（1975 年）