

HIV-1の情報理論的解析

HIRAMATSU, Toyokazu / TADA, Hideki / 平松, 豊一 / 多田,
秀樹

(出版者 / Publisher)

法政大学工学部

(雑誌名 / Journal or Publication Title)

Bulletin of the Faculty of Engineering, Hosei University / 法政大学工学部
研究集報

(巻 / Volume)

43

(開始ページ / Start Page)

13

(終了ページ / End Page)

19

(発行年 / Year)

2007-03

(URL)

<https://doi.org/10.15002/00003745>

HIV-1の情報理論的解析

AN INFORMATION THEORETIC ANALYSIS OF HIV-1

多田秀樹*, 平松豊一**

Hideki TADA and Toyokazu HIRAMATSU

The V3 loop is a region in the envelope protein of HIV-1 whose polypeptide(amino acid sequence) extremely tends to change. Certain amino acid pairs in the V3 loop do not have only independent mutations but covariant mutations. We apply information theory for an analysis of the V3 loop in order to draw a certain law of change of polypeptide of the V3 loop.

Key Words : V3 loop, Mutual Information, Specific Information, Covariant Mutation flow

1. はじめに

現在、世界中で深刻な社会問題を引き起こしているウイルスの一つに HIV-1 いわゆるエイズウイルスがある。感染予防のためのワクチンの開発、および実用化を妨げている理由は、HIV-1 が免疫応答の中核にあるヘルパー T 細胞を破壊してしまう特性をもつこと、優れた動物モデルが無いこと、など多岐にわたるが、なかでも抗原構造が著しい変異性を持つことが、大きな障害の一つとして挙げられる。とくにアミノ酸配列のなかの V3 loop と呼ばれる領域は、劇的な変異性を持つことで知られ、機能上、免疫学上の両面において、重要な研究対象となっている。

V3 loop には、実際さまざなアミノ酸配列が現れるが、それらがまったく独立かつ不規則な変化のすえに生じているとはむしろ考えにくく、特定の位置ないしアミノ酸の変異については、生物現象特有の有機的な結びつきが隠れていることも十分考えられる。そこで、本論文では、幾つかのアミノ酸が何らかの関連を持ちながら変化しているのではないかという仮説のもとに、相関の強さを表す尺度として新たに共変指数を導入し、情報理論を応用した解析を試みる。

エントロピー(情報量)とは、確率分布をもつ事象 i_0 が与えられた時、 i_0 の「無秩序さ・曖昧さ・不確実さ」を表す尺度として知られる。また、多次元相互情報量は、確率分布を持ついくつかの事象 i_0, \dots, i_n が与えられた時、 i_0, \dots, i_n が共通してもつエントロピーを表す。仮に、多次元相互情報量が正の値をとり、ある一つの事象 i_j ($0 \leq j \leq n$) についての情報が明らかになったとすると、相互情報量は各事象が共通して持つエントロピーの

ため、他の事象では相互情報量に相当するエントロピーの減少が見られる。すなわち、 i_j は他の事象に関する情報を持っていることとなり、各事象 i_0, \dots, i_n には何らかの相関があるとみなされる。しかしながら、多次元相互情報量のみを用いて位置間に働く相関の強さを考えようとすると、多次元相互情報量は各事象に対する情報量の大きさの大小に依存してしまう場合も多く、相関の強さを表す尺度として妥当ではない。そこで本論文では、新たに相関の強さを表す尺度として共変指数を定義し、V3 loop 内の位置間に働く相互依存性を探り、共変グループの特定とアミノ酸ネットワークの解明を目的とし実験を行った。

2. 基本事項

(1) タンパク質

多くの生物の遺伝情報は、4種類の塩基(アデニン、チミン、シトシン、グアニン)からなる DNA という形で細胞内に蓄えられている。DNA は、転写をおこし RNA となり、これがコドン(3塩基)単位でアミノ酸に翻訳されることによって必要なタンパク質が合成されている。

表1 アミノ酸コード表

A	Alanine	R	Arginine
N	Asparagine	D	Aspartic acid
C	Cysteine	Q	Glutamine
E	Glutamic acid	G	Glycine
H	Histidine	I	Isoleucine
L	Leucine	K	Lysine
M	Methionine	F	Phenylalanine
P	Proline	S	Serine
T	Threonine	W	Tryptophan
Y	Tyrosine	V	Valine

*大学院システム工学専攻

**工学部

しかし、今回取り扱う HIV-1 は、生物界で唯一逆転酵素をもつレトロウィルスと呼ばれるものに属し、遺伝情報は RNA の形で蓄えられている。

タンパク質は、一本の鎖状につながった 20 種類のアミノ酸が複雑な立体構造を成して形成されていて、タンパク質ごとに構成アミノ酸の種類・数・結合順は異なる。この構造及び機能を解析することはポストゲノムと呼ばれ、多くの研究機関が取り組んでいる問題でもある。表 1 に 20 種類のアミノ酸とそのアミノ酸コードを示す。

(2) HIV-1 と V3 loop

HIV-1 は直径 110nm、約 9500 塩基からなる RNA 型エンベロープウィルスである。ウィルス粒子の外側を構成するエンベロープには、糖タンパク質 gp120 と糖二重層を貫く糖タンパク質 gp41 があり、gp120 内に存在する V3 loop 領域は非常に変化の富んだ領域(様々な配列が現れる)であり、機能上・免疫学上の両面において重要とされ多くの研究の対象となっている。

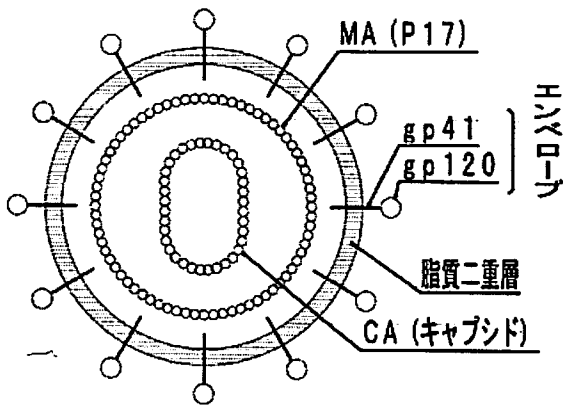


図 1: HIV-1 の構造

V3 loop 領域は、両末端を Cysteine(C) でジスルフィド結合をしたループ構造をしていて、抗体や T 細胞を中和する効力を持つ。

図 2 は V3 loop の一例の抽象的なモデル図である。アルファベットはそれぞれアミノ酸コードに対応し、それらの外に添えてある数字は解析を行う上での便宜上の位置(site)を表している。また、V3 loop の 14~17 の位置におけるアミノ酸配列が GPGQ であればアフリカ分離株、GPGR であれば欧米分離株と呼ばれる。異なる配列モデルの場合、アミノ酸の配列は異なり、本論文では、異なる位置にあるアミノ酸が、何らかの関連をもって同時に変化することを共変と呼ぶ。

図 2 の各位置における、20 種のアミノ酸が現れる確率(生起確率)を、データベース(National Center for Biotechnology Information(NCBI)[7])に実在する約

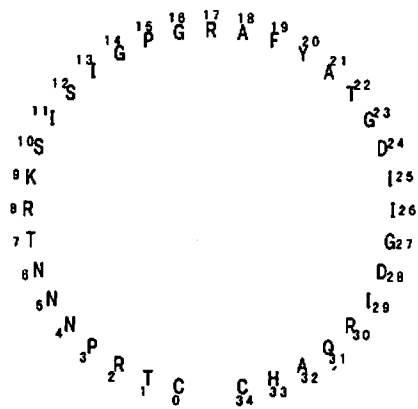


図 2: V3 loop のモデル図

17000 例の V3 loop の配列から求め、各位置における生起確率を用いて、アミノ酸の変化量をエントロピーの概念を用いて定量的に表現する。

(3) アライメント (alignment)

アライメントとは、配列の類似する部分、または同一となる部分を縦に揃えて並べ合わせる操作のことを言い、アミノ酸配列の類似性解析の基本的なものの一つである。二本のアミノ酸配列 HEAGAWGHEE と EGWHEAE のアライメントを行うと以下ようになる。

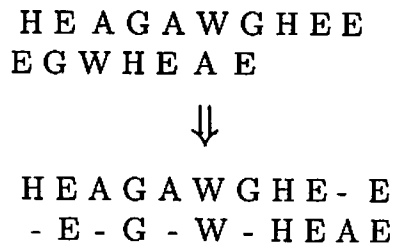


図 3 アライメント

同じアミノ酸や性質の似たアミノ酸が縦に、同じ位置になるように、ところどころハイフン(-)が挿入されている。生化学的にはこれらのギャップ(gap)は遺伝子の欠損等によりアミノ酸が出現していないことを意味する。縦方向に揃った代表的文字で構成される配列をコンセンサス配列(consensus sequence)と呼び、この中に見られるパターンがアライメントされた配列群を特徴付けるものと判断できるとき、このパターンをモチーフ(motif)と呼ぶ。また、ギャップの入り方の相対的に少ない領域を保存部位(conservative site)と言い、保存性の高い部位は配列の中でもタンパク質の構造や機能の実現の上で重要な部分であると推測できる。重要な配列部分は進化の過程の中で保存されているため、生物の同種のタンパク質をアライメントで調べると保存部位を発見できる。対象となる配列が 2 本の場合はペアワイズアライメント

(pairwise alignment) と呼び、基本的には配列と配列の類似性を求めることに使われる。また、3本以上の配列を扱う場合はマルチプルアライメント (multiple alignment) と呼んで区別され、ペアワイズアライメントよりも効果的に配列の共通性を見出せる。

3. エントロピーと多位置間相互情報量

(1) エントロピーと相互情報量

V3 loop のそれぞれの位置では様々なアミノ酸が現れる場合もあれば、逆にほぼ決まったアミノ酸が現れる場合もある。そこで V3 loop の位置 i ($0 \leq i \leq 34$) に現れるアミノ酸を $A^{(i)}$ で表す。ここで、 A は 20 種類のアミノ酸コードに対応するアルファベット 1 文字が表記される。あるアミノ酸の生起確率を $P(A^{(i)})$ として、位置 i におけるアミノ酸の変化の不確実度は、エントロピー

$$H(i) = - \sum_{A^{(i)}} P(A^{(i)}) \log P(A^{(i)})$$

で表される。エントロピー $H(i)$ は非負である。また、現れるアミノ酸が多様であるほど高い値を示す。ここで、位置 i に $A^{(i)}$ 、位置 j に $A^{(j)}$ が現れる生起確率を $P(A^{(i)}, A^{(j)})$ ($i \neq j$) とすると

$$H(i, j) = - \sum P(A^{(i)}, A^{(j)}) \log P(A^{(i)}, A^{(j)})$$

を結合エントロピーと呼ぶ。

アミノ酸が現れる位置 i と j をペアとして捉え、相互情報量を考えると、両位置間の共変の有無をあらわす事ができる。(相互情報量 $M(i, j)$ は非負である。)

$$\begin{aligned} M(i, j) &= H(i) - H(i|j) \\ &= H(j) - H(j|i) \\ &= H(i) + H(j) - H(i, j). \end{aligned} \quad (1)$$

ただし、 $H(i|j)$ は条件付エントロピーとする。

$$H(i|j) = - \sum P(A^{(i)}) P(A^{(i)}|A^{(j)}) \log P(A^{(i)}|A^{(j)}).$$

ここで、2 位置 i, j におけるアミノ酸が、完全に独立な変化をするか、変化がないときに、 $M(i, j)$ の値は最小値 0 をとる。位置 i, j の相関が強くなるほど $M(i, j)$ の値は大きくなる。

(2) 多位置間相互情報量

2 位置間における共変の数値化の試みは [1] に見られるが、そこで指数として用いられているのは相互情報量である。本論文の目的は、多位置間の共変を考察することであり、したがって、相互情報量ではなく多次元相互情報量を用いる。以下、多次元相互情報量のかわりに多位置間相互情報量と呼ぶ。相互情報量を、3 位置間以上に

拡張した場合、多位置間相互情報量は負の値も取り得ることに注意する。負の相互情報量とは情報理論の観点からみると、異なるいくつかの情報源から得られる情報により、情報を得た側のエントロピーが増大したときにおこりうる。一方、これを共変という観点から見ると、負の相互情報量が現れる位置間では共変を阻害するように働く位置が含まれていることを意味するのではないかと考えられる。

任意の m 位置 i_0, \dots, i_{m-1} での多位置間相互情報量が全て正の値をとり、かつ、位置 i_m を加えた $m+1$ 位置間での多位置間相互情報量のはじめて負の値をとったとき、位置 i_m は共変グループに含まれていないと考えることができる。

位置 i_0, \dots, i_m にアミノ酸列 $A^{(i_0)}, \dots, A^{(i_m)}$ が現われる結合事象の生じる確率は

$$\begin{aligned} P(A^{(i_0)}, A^{(i_1)}, \dots, A^{(i_m)}) \\ = P(A^{(i_m)}) P(A^{(i_0)}, A^{(i_1)}, \dots, A^{(i_{m-1})} | A^{(i_m)}). \end{aligned}$$

結合エントロピーは一般に以下の式で表される。

$$\begin{aligned} H(i_0, i_1, \dots, i_m) \\ \stackrel{\text{def}}{=} - \sum_{A^{(i_j)}} P(A^{(i_0)}, A^{(i_1)}, \dots, A^{(i_m)}) \\ \times \log P(A^{(i_0)}, A^{(i_1)}, \dots, A^{(i_m)}). \end{aligned}$$

また、位置 i_m にアミノ酸 $A^{(i_m)}$ が現われたとき位置 i_0, \dots, i_{m-1} にアミノ酸列 $A^{(i_0)}, \dots, A^{(i_{m-1})}$ の現われる条件付確率を $P(A^{(i_0)}, \dots, A^{(i_{m-1})} | A^{(i_m)})$ とすると、条件つきエントロピーは以下の様に表され、右辺の和は $0 \leq j \leq m$ について各々 20 種類のアミノ酸 $A^{(i_j)}$ を渡る。

$$\begin{aligned} H(i_0, i_1, \dots, i_{m-1} | i_m) \\ \stackrel{\text{def}}{=} - \sum P(A^{(i_m)}) P(A^{(i_0)}, \dots, A^{(i_{m-1})} | A^{(i_m)}) \\ \times \log P(A^{(i_0)}, \dots, A^{(i_{m-1})} | A^{(i_m)}) \\ = H(i_0, i_1, \dots, i_m) - H(i_m). \end{aligned}$$

任意の複数の位置 $i_0, i_1, \dots, i_{m-1}, i_m$, ($0 \leq i_j \leq 34$) での多位置間相互情報量は、 i_0, i_1, \dots, i_{m-1} 位置間の相互情報量と位置 i_m における条件付き相互情報量を用いて以下の様に表される。

$$\begin{aligned} M(i_0, i_1, \dots, i_{m-1}, i_m) \\ = M(i_0, i_1, \dots, i_{m-1}) - M(i_1, i_2, \dots, i_{m-1} | i_m). \end{aligned}$$

$m = 1$ のときの相互情報量 (1) は良く知られたもので

ある。[4]

$$\begin{aligned}
 M(i_0, i_1) &= H(i_0) - H(i_0|i_1) \\
 &= - \sum_{A^{(i_0)}} P(A^{(i_0)}) \log P(A^{(i_0)}) \\
 &\quad + \sum_{A^{(i_0)}, A^{(i_1)}} P(A^{(i_1)}) P(A^{(i_0)}|A^{(i_1)}) \\
 &\quad \times \log P(A^{(i_1)}|A^{(i_2)}). \quad (2)
 \end{aligned}$$

われわれはこの関係式を一般化して、 $m \geq 2$ に対しても相互情報量及び条件付相互情報量を考える。

$m = 2$ のとき

$$M(i_0, i_1, i_2) = M(i_0, i_1) - M(i_0, i_1|i_2).$$

右辺の $M(i_0, i_1|i_2)$ は条件付相互情報量であって、(2) を条件付確率で置き換えることで定義される。

$$\begin{aligned}
 M(i_0, i_1|i_2) &\stackrel{\text{def}}{=} - \sum_{A^{(i_0)}, A^{(i_2)}} P(A^{(i_2)}) P(A^{(i_0)}|A^{(i_2)}) \log P(A^{(i_0)}|A^{(i_2)}) \\
 &\quad + \sum_{A^{(i_0)}, A^{(i_1)}, A^{(i_2)}} P(A^{(i_2)}) P(A^{(i_1)}|A^{(i_2)}) P(A^{(i_0)}|A^{(i_1)}|A^{(i_2)}) \\
 &\quad \times \log P(A^{(i_0)}|A^{(i_1)}|A^{(i_2)}). \quad (3)
 \end{aligned}$$

$P(A^{(i_0)}|\dots|A^{(i_n)}) = \frac{P(A^{(i_0)}, \dots, A^{(i_n)})}{P(A^{(i_0)}) \dots P(A^{(i_n)})}$ を用いて (3) の右辺を整理すると、

$$\begin{aligned}
 M(i_0, i_1|i_2) &= H(i_0, i_2) - H(i_2) - H(i_0, i_1, i_2) + H(i_1, i_2) \quad (4)
 \end{aligned}$$

が得られる。 $M(i_0, i_1, i_2) = M(i_0, i_1) - M(i_0, i_1|i_2)$ に (1)、(4) を代入すると

$$\begin{aligned}
 M(i_0, i_1, i_2) &= \sum_{j=0}^2 H(i_j) - \sum_{0 \leq j < k \leq 2} H(i_j, i_k) + H(i_0, i_1, i_2).
 \end{aligned}$$

同様に、 $M(i_0, i_1, \dots, i_m)$ は $M(i_0, i_1, \dots, i_{m-1})$ と条件付確率を考えた $M(i_0, \dots, i_{m-1}|i_m)$ より求められ、多次元相互情報量は結合エントロピーを用いて一般に以下の様に表される。

$$\begin{aligned}
 M(i_0, \dots, i_m) &= \sum_{j_0=0}^m H(i_{j_0}) - \sum_{0 \leq j_0 < j_1 \leq m} H(i_{j_0}, i_{j_1}) \\
 &\quad + \sum_{0 \leq j_0 < j_1 < j_2 \leq m} H(i_{j_0}, i_{j_1}, i_{j_2}) - \dots \\
 &\quad \dots + (-1)^{n-1} \sum_{0 \leq j_0 < j_1 < \dots < j_{(m-1)} \leq m} H(i_{j_0}, i_{j_1}, \dots, i_{j_{(m-1)}}) \\
 &\quad + (-1)^n H(i_0, i_1, \dots, i_m).
 \end{aligned}$$

4. 共変指数

各位置における 1 次エントロピーをふまえた上で多位置

間の相互情報量を考えた時、1 次エントロピーの低い位置間の相互情報量の値よりも 1 次エントロピーの高い位置間の相互情報量の値の方が高い値をとり易くなる。しかしながら、1 次エントロピーの低い位置においても共変を起こしている可能性は十分に考えられ、相互情報量の値をそのままグループの特定に至る共変の強さを表す尺度として用いるのは妥当ではないと考えられる。そこで、次式を共変指数 K として定義する。

$$K(i_0, \dots, i_m) = \frac{M(i_0, \dots, i_m)}{m+1} \sum_{j=0}^m \frac{1}{H(j)}. \quad (4)$$

ここで、 m 位置間以下での相互情報量は全て正の値を示すもののみを考える。これは 3.2 節でも前述したが、仮に位置 $(0 \leq m' < m)$ の任意の位置間での m' 位置間相互情報量が全て正の値をとり、位置 m' を加えた $m'+1$ 位置間での多変数相互情報量で、はじめて負の値をとったとすると、この位置 m' は共変グループに含まれていないと考えることができる。

(4) 式は、各位置の持つ 1 次エントロピーに対する $m+1$ 位置間相互情報量の平均の割合を示し、共変指数の値が 1 に近いほど $m+1$ 位置間の共変の度合いは強いものとして考えられる。したがって、実用上は共変指数の閾値を適切に定めておくことが望ましい。具体的な手順を以下に示す。

m 位置共変グループ特定までのながれ

1. マルチプルアライメントを行って、配列長の同じアミノ酸配列データを作成する。
2. m 位置までの同時確率データの計算をする。
3. 2. で得られた同時確率データから、2 位置間から m 位置間までの全ての相互情報量を計算する。
 $(m$ 位置間までの相互情報量のなかに、一つでも負の値をとるものがあれば、共変なし)
4. 3. で得られた相互情報量から、2 位置間から m 位置間までの全ての共変指数を計算する。
 $(2 \sim m$ 位置間までの共変指数のうち、一つでもその値が閾値未満のとき、共変なし)
5. $m+1$ 位置での共変指数を考えた時、全ての位置における共変指数が閾値未満のとき、 m 位置からなる共変グループが存在すると考えられる。
5. データ群の作成

V3 loop に関するアミノ酸配列のデータは NCBI のホームページから約 17000 例取り出す事ができる。これにいったん PAPIA システム [8] によるアライメントをかけて Cysteine から始まり Cysteine に終る長さ 35 前後の配列のみを取り出し、実験用のデータとした。さらに、

HIV-1のV3 loopにおけるアミノ酸配列は、位置14~17付近がGPGR(欧米分離株)のもととGPGQ(アフリカ分離株)のもとに分類されるため、この2種にデータ群をGPGR(8340例)とGPGQ(4033例)に大別した。このようにして得られたデータを最新版 clustal W にかけて、マルチプルアライメントを行った。さらに今回の実験では、ギャップの挿入の見られた位置は考察の対象からははずして、アミノ酸のみで構成される、位置のみを比較の対象とすることにして、配列長を35とした。従来の clustal W ver. 1.82 には、最長配列×配列数 ≤ 10000 という処理上の制限があるため、一度に何千本もの配列のアライメントを行うことが不可能であったが、今回、これらの処理を国立遺伝学研究所にお願いすることで実現した。

6. 実験結果

各位置におけるエントロピーを計算して、アミノ酸がどの程度変化しているのかを求める。また、2~4位置間での共変指数を算出する事により、アミノ酸が共変している可能性が高いと思われる位置を示す。

(1) 位置*i*におけるエントロピー

V3 loop 領域における各位置のエントロピーの値を図4に示す。V3 loop 領域は両末端位置0と34にCysteine(C)を持つ。また、位置14 ≤ *i* ≤ 17は、GPGR型、GPGQ型のみを実験の対象としているため、それらの位置についてはエントロピーの値は0となる。また、エントロピーの値が相対的に低い位置(2, 3, 5~8, 27, 29, 30, 32)は保存性があり、エントロピーの高い位置(4, 9~12, 18, 21, 24)は変異性に富んだ部位であると思われる。

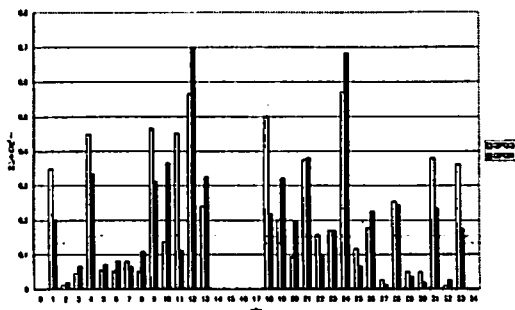


図4: エントロピー $H(i)$

(2) 2位置(*i, j*)における共変指数

図5、図6は2位置間での共変指数のグラフである。両位置におけるアミノ酸が完全な共変を示す時は最大値1、完全に独立な変動、もしくは変動が無い時には最小値0

を示す。両型共に2位置間共変指数は、V3 loop の中間部分にあたる位置4~24に他への依存度が大きいと思われる位置が多く現れる。

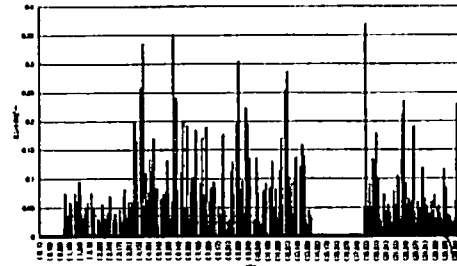


図5: GPGQ型: 共変指数 $K(i, j)$

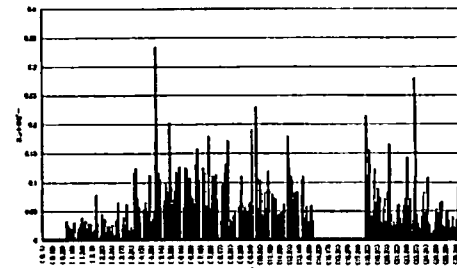


図6: GPGR型: 共変指数 $K(i, j)$

(3) 3位置(*i, j, k*)における共変指数

表2 共変指数 $K(i, j, k)$

site(GPGQ)	共変指数	site(GPGR)	共変指数
(4, 18, 21)	0.22516	(10, 18, 19)	0.09635
(9, 18, 21)	0.17517	(5, 7, 26)	0.09522
(6, 7, 11)	0.16845	(5, 7, 10)	0.08287
(4, 9, 21)	0.15701	(10, 12, 18)	0.07499
(12, 18, 21)	0.15215	(4, 5, 7)	0.05755
(9, 21, 31)	0.15112	(5, 6, 26)	0.05650
(9, 21, 33)	0.14668	(10, 18, 31)	0.05612
(6, 7, 21)	0.13935	(18, 23, 24)	0.05568
(21, 31, 33)	0.12926	(5, 22, 28)	0.05090
(4, 12, 21)	0.12912	(18, 19, 31)	0.05025
(6, 7, 26)	0.12611	(7, 8, 10)	0.05020
(9, 31, 33)	0.11978	(5, 6, 22)	0.04636
(18, 21, 31)	0.11328	(7, 8, 26)	0.04411
(6, 11, 21)	0.10546	(7, 10, 18)	0.04402
(4, 9, 18)	0.09917	(12, 18, 19)	0.04377
(4, 21, 31)	0.09337	(20, 23, 24)	0.04351
(4, 12, 18)	0.09289	(5, 7, 8)	0.04219
(18, 21, 33)	0.09226	(5, 10, 26)	0.04171
(9, 12, 21)	0.09052	(7, 18, 21)	0.04154
(12, 21, 33)	0.09040	(7, 18, 31)	0.04113
⋮	⋮	⋮	⋮

3位置間での共変指数の値が高いものから順に20組をGPGQ・GPGR型について表2に示す。また、図7、図

8 は共変指数が高く、3 位置の間で共変していると思われる位置グループを実線で結んだものである。GPGQ・GPGR 型ともに、位置 14~17 の左側のせまい位置間での 3 位置共変指数が高く見られ、また右側の位置では離れたいくつかの決まった位置が、3 位置での共変に関与していると思われる。

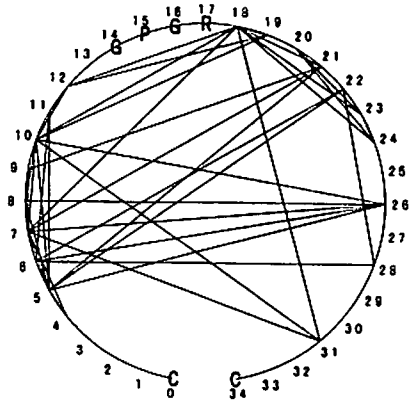


図 7: 3 位置共変 (GPGR 型)

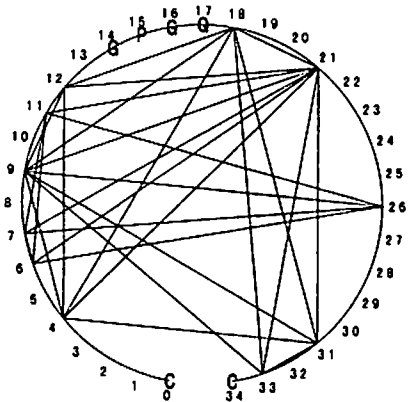


図 8: 3 位置共変 (GPGQ 型)

(4) 4 位置 (i, j, k, l) における共変指数

3 位置間での共変指数で特に値の高かった GPGR 型の位置 (5,7,26)、(10,18,19)、GPGQ 型の位置 (4,9,18)、(6,7,11) について、これらの 3 位置と V3 loop 内の任意の位置における 4 位置での共変指数をそれぞれ図 9~図 12 に示す。図 9 より、GPGR 型、3 位置 (5,7,26) においては位置 5, 8, 10, 22 が、図 10 より (10,18,19) においては位置 4, 7, 12, 31 が 4 位置目の共変グループ候補としてあげられる。また、GPGQ 型の 3 位置 (4,9,18) においては、位置 21 が他の位置に比べ強い相関が見られる。同様に、(6,7,11) は位置 11, 18, 21 との強い相関が見られる。それぞれの 4 位置では、共変指数が他と比べ高い値をとる位置が存在することから、さらに 4 位置での共変指数

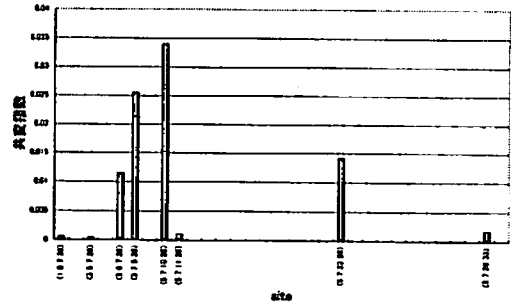


図 9: GPGR 型 : 共変指数 $K(5, 7, 26, i)$

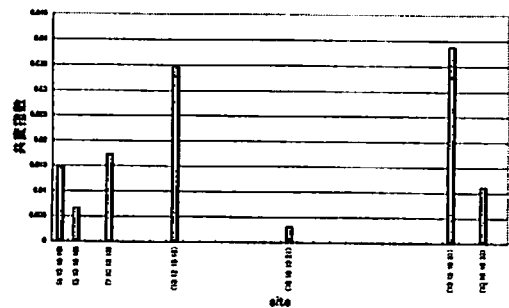


図 10: GPGR 型 : 共変指数 $K(10, 18, 19, i)$

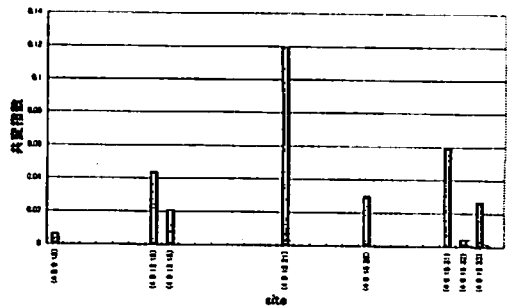


図 11: GPGQ 型 : 共変指数 $K(4, 9, 18, i)$

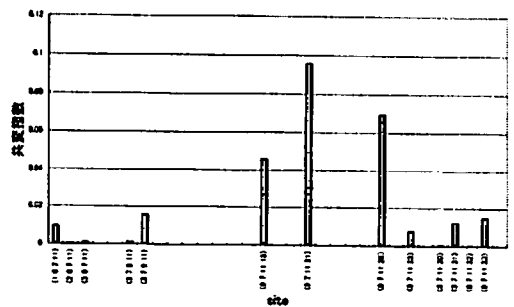


図 12: GPGQ 型 : 共変指数 $K(6, 7, 11, i)$

の高い値を示した位置との5位置の共変指数を考えることで共変している可能性のある位置を絞り込み、共変グループの特定を行うことができると考える。この際、求める5位置での共変指数において、任意の3~4位置の共変指数はすべて閾値以上の値をとり、さらに5位置の共変指数が閾値以上のものを、共変しているものとみなす。

7. むすび

本論文では、V3 loop内のアミノ酸配列の位置に何らかの相互依存が存在し、V3 loopの機能もしくは構造に関与するという仮定のもとに、多次元相互情報量と新たに共変指数を導入することで、位置間に働くネットワークの解明を試みた。今回得られた数値データでは、V3 loop内にいくつかのネットワークが存在することを予想させるものとなったが、さらに生物学、農学、薬学、医学面からの研究の進展を待って、その真偽を明らかにしたい。また、生体内の酵素などのタンパク質は、互いに依存しあいその働きが制御されている事はすでに知られ、新薬の開発等に活かされている。今回導入した手法をアミノ酸配列の変異の激しいウイルス等に応用することで、その研究に役に立つことを期待したい。

謝辞

本研究において、アミノ酸配列のアライメントデータを国立遺伝学研究所から提供していただく際に仲介の労をとっていただいた獣医畜産大学の関田英太郎助教授(当時)に、また、貴重な助言を多く頂きました東京理科大学の大矢雅則教授に深く感謝いたします。

参考文献

- [1] B.T.M. Korber, R.M.Farber, D.H.Wolpert, A.S.Lapedes, "Covariation of mutation in the V3 loop of human immunodeficiency virus type 1 envelope protein," An information theoretic analysis, Proc.Natl.Acad.Sci.USA, vol.90, pp.7176-7180, August 1993.
- [2] B. アルバート他 中村桂子他訳, 細胞の分子生物学第3版, 教育社,1995.
- [3] M. Oya, "Information Theoretial Treatment of Genes," Trans. IEICE, E72-5, pp.556-560, May 1989.
- [4] 嵩忠雄, 情報と符号の理論入門, 昭晃堂,1989.
- [5] 中山晃治, 田辺文雄, 多田秀樹, 関田英太郎, 平松豊一, 西村滋人, "エイズウイルスにおけるV3ループの情報理論的解析," 第25回情報理論とその応用シンポジウム予稿集, pp.763-766, 2002.
- [6] 多田秀樹, 関田英太郎, "HIV-1とエントロピー," 研究会「符号と暗号の代数的数理」, 数理解析研究所講究録 1420, pp.18-27, 2005.
- [7] National Center for Biotechnology Information, <http://www4.ncbi.nlm.nih.gov/Entrez/>
- [8] Parallel Protein Information Analysis System(PAPIA), http://www.cbrc.jp/papia/cgi/mul_queryJ.pl