

タンパク質適応進化の時空間モデル

渡部 輝明¹・岸野 洋久²

(受付 2011 年 8 月 16 日；改訂 11 月 30 日；採択 12 月 2 日)

要 旨

タンパク質はアミノ酸配列を変異させることで環境に適応していくが、遺伝子上で起こった突然変異は環境に特有の選択圧のもとで淘汰されていく。そのため選択圧はタンパク質表面の領域に依存すると同時に、進化の過程において変化する環境にも依存していくものと考えられる。そのことから選択圧の時空間的な揺らぎを明らかにすることがタンパク質適応進化について理解するための重要な役割を果たすと考えられる。本稿ではまず A 型インフルエンザウイルスの 35 年に渡る変異の過程での選択圧の変化を観る。ヘマグルチニンタンパク質の宿主細胞受容体結合領域における抗体との結合親和性がどのように変化したかを観ていくことでタンパク質にかかっていた選択圧の変化をうかがい知る。そして SARS コロナウイルスを用いて正常細胞、感染細胞、ウイルス粒子、抗体の 4 者からなる宿主内動態の変化を解析し、変異ウイルスの宿主適応度を解析する。そして集団遺伝学の理論から変異ウイルスが宿主集団に定着する確率を求める。結合領域での幾つかの変異が高い確率で固定され得ることが判った。

キーワード：分子進化、選択圧、空間分布、固定確率。

1. はじめに

本稿で扱うインフルエンザ (Flu) ウイルスや SARS (severe acute respiratory syndrome：重症急性呼吸器症候群) コロナウイルスは、宿主細胞に侵入するために宿主細胞受容体にウイルス粒子膜上のタンパク質を結合させる。この受容体結合 (receptor-bound：RB) タンパク質は、その変異において拮抗する選択圧を受けていると考えられる。宿主へ適応したウイルスは RB タンパク質による宿主細胞受容体への結合能を維持し続ける必要があるが、その一方で生体防御機構 (免疫系) からの RB タンパク質への攻撃をかわしていかなければならない。免疫系からの主な攻撃は抗体による結合である。抗体に結合された RB タンパク質は、宿主細胞受容体への結合が阻害されてその役割を果たすことが出来ない。抗体の結合を不可能にするように RB タンパク質は変異するが、このとき宿主細胞受容体への結合能も犠牲になる可能性がある。このような状況における選択圧によって淘汰された変異は、タンパク質の機能にどのような変化をもたらすのかは自明なことではない。またウイルス変異の過程のどのタイミングで選択圧が多様化をもたらす働きをしたのかを知ることは、ウイルス進化の機構解明に大きな貢献をもたらすと考えられる。そしてウイルスにどの位の頻度で生じる変異がどの程度の割合で固定されるのかを予測することは、対ウイルス薬剤の開発に貴重な情報を提供することになる。

¹ 高知大学 医学部附属医学情報センター：〒783-8505 高知県南国市岡豊町小蓮

² 東京大学 農学生命科学研究科：〒113-8657 東京都文京区弥生 1-1-1

2. インフルエンザヘマグルチニンタンパク質の変異

Flu ウイルスの RB タンパク質は、ヘマグルチニン(HA)と呼ばれるタンパク質である。HA はホモ 3 量体の筒状のタンパク質であり、中心部は α ヘリックスによって構成され、先端にはシアル酸結合部位がある。宿主細胞受容体はこの結合部位に結合する。各モノマーは HA1 と HA2 のサブユニットから構成されている。我々は 1968 年から 2003 年までの間にヒトから分離された 253 株の A 型 Flu ウイルスの HA タンパク質の配列を解析した (Watabe et al., 2007)。HA1 の全アミノ酸配列を最尤法 (PHYLIP; Felsenstein, 2005) で解析して得た系統樹のトポロジーを用いて、その制約下で parsimony の最適性基準に従い結合領域のアミノ酸置換を最大節約法 (PAUP*; Swofford, 2003) により再構築した。ここで結合領域とは HA と抗体との結合に関与する領域を指し、抗体の任意のアミノ酸残基と α 炭素原子間が 1nm (10Å) 以下にある残基によって構成される領域として定義した。 α 炭素原子とはアミノ酸残基の最も主鎖に近い炭素原子を指す。HA との複合体構造が X 線結晶構造解析により測定されている抗体は 4 つほど Protein Data Bank (PDB) に登録されており (HC19, HC63, HC45 and BH151)、それらとの複合体について解析した。そのうちの 2 つは宿主細胞受容体との結合領域を覆う形で結合が構成されており (HC19 and HC63)、残る 2 つはそこから外れた領域で結合する (HC45 and BH151)。我々は再構築したアミノ酸置換を基に 1968 年から 2003 年に渡る HA と抗体との結合能の変化を解析した。

2.1 タンパク質複合体の結合能評価

タンパク質複合体の結合能評価はその結合定数の計算を以て行う。結合定数 K_a はタンパク質 A と B の複合体状態の構造が与えられた時のアミノ酸配列の分布と、個別の構造が与えられた時のそれぞれのアミノ酸配列の分布の積の比に比例することが示されている (Watabe et al., 2007) :

$$(2.1) \quad K_a \propto \frac{P(\text{seq}_A, \text{seq}_B | \text{str}_{A+B})}{P(\text{seq}_A | \text{str}_A) P(\text{seq}_B | \text{str}_B)}$$

一般にはアミノ酸配列が与えられた時の構造の分布 $P(\text{str} | \text{seq})$ を分子動力学などの手法を用いて計算するが、計算時間が膨大であることからタンパク質変異のシミュレーションや進化過程の追跡などには不向きである。そのため我々はバイズの定理を用いて (2.1) 式を得て結合定数の計算を行う。このときアミノ酸配列変異に対してタンパク質の主鎖構造は不変であることを仮定している。各配列分布 $P(\text{seq} | \text{str})$ は配列に関して以下のように分解が可能である :

$$(2.2) \quad P(a_1, \dots, a_L | \mathbf{X}) \cong \prod_i P(a_i | \mathbf{X}) \prod_{i < j} \frac{P(a_i, a_j | \mathbf{X})}{P(a_i | \mathbf{X}) P(a_j | \mathbf{X})}$$

ここで a_i は配列長が L のアミノ酸配列の i 番目のアミノ酸を表している。 \mathbf{X} はアミノ酸残基の α 炭素原子の位置を表している、 $\mathbf{X} = (x_1, \dots, x_L)$ 。ここで x_i は i 番目の残基の α 炭素原子の空間座標を表している、 $x_i = (x_i, y_i, z_i)$ 。(2.2) 式では 3 つ以上のアミノ酸残基間の相関は低いことを仮定して、それらに対応する高次の項を無視している。更に構造をパラメータ化する。1 アミノ酸分布については $P(a_i | \mathbf{X}) \cong P(a_i | \nu_i)$ とし、 ν_i は i 番目の残基の α 炭素原子を中心とした半径 1nm の球状領域内の α 炭素原子数を表している。2 アミノ酸分布については $P(a_i, a_j | \mathbf{X}) \cong P(a_i, a_j | \nu_i, \nu_j, r_{ij}, n_{ij})$ とし、 r_{ij} は i 番目と j 番目の残基の α 炭素原子間の距離を表し、 n_{ij} は $|i - j|$ を表している。このように構造を 1 アミノ酸、2 アミノ酸を特徴付けるパラメータによって表現し、更にそれぞれのパラメータにカテゴリーを導入する。そのことで PDB に蓄積されている構造データから、カテゴリー化された構造毎に 1 アミノ酸、2 アミノ酸の頻度分布を計算することが出来る (Watabe et al., 2006)。求められた頻度分布を用いて (2.1)

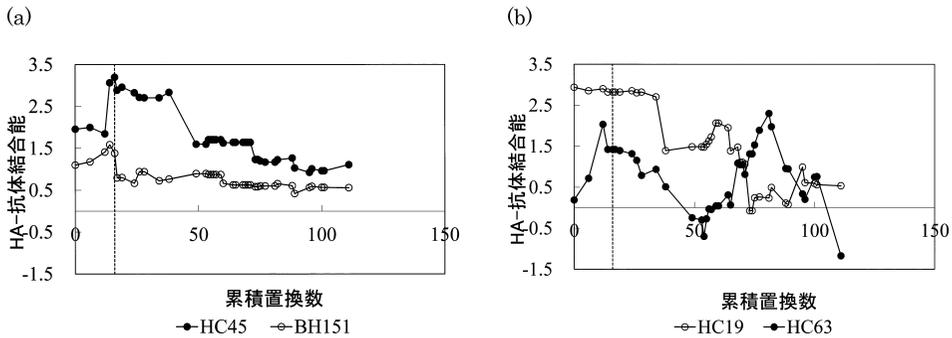


図 1. 系統樹の“幹”に沿ったアミノ酸累積置換数(横軸)と HA-抗体結合能(縦軸)の変化. 置換数が 20 弱の辺りで引いてある点線は 1968 年付近の分岐点を示している.

式を計算することが出来る.

図 1 に HA1 の全アミノ酸配列に対して再構築したアミノ酸置換の累積置換数を横軸にとった HA-抗体結合能の変遷を示した. 図 1 (a) では宿主細胞受容体との結合領域から外れた領域で HA と結合する抗体についての解析結果を示している. 35 年に渡る結合能の変化は比較的緩やかなものとなっている. 一方, 図 1 (b) では宿主細胞受容体との結合領域を覆うように HA と結合する抗体について示している. 結合能の変化は起伏に富んでおり, HC63 抗体においては一度大きく失った結合能が再び元の状態まで回復し, 再度失っていく様子が示されている. この違いは抗体が HA へ与える脅威の差であると考えられる. 宿主細胞受容体を覆ってしまう抗体に対しては, その結合能を失わせるべく結合領域においてアミノ酸置換を積極的に行う. しかし宿主細胞受容体との結合能を維持する必要があるため, 状態を回復するようなアミノ酸置換もある時期から行うと考えることが出来る. ここで特筆すべきは“状態の回復”がアミノ酸配列の復元により実現されているわけではなく, アミノ酸配列と祖先配列との距離は隔たって行く一方にあることである. タンパク質間の結合能における冗長性が見て取れる.

3. アミノ酸変異の固定確率

我々は更に感染現象におけるウイルス側と宿主側の分子間での複雑な相互作用を考慮し, RB タンパク質に生じた変異がどの程度の割合で固定されるのかを解析した (Watabe and Kishino, 2010). 正常細胞, 感染細胞, ウイルス, そして抗体の 4 者で宿主内動態を表現する方程式系が 1996 年に Nowak らによって提案されている (Nowak and Bangham, 1996). しかし彼らの方程式系において相互作用の強さを表すパラメータを決定する方法はなく, 解析に任意性が残る結果となっていた. 我々は相互作用の強さを分子間の結合親和性と関連させて, RB タンパク質の変異に伴う結合親和性の変化を方程式系のパラメータに反映させた.

3.1 ウイルス動態の数理モデル

数理モデルは正常細胞 (X), 感染細胞 (Y), ウイルス (V), そして抗体 (W) の 4 者の変数が

らなり、以下の微分方程式に従う：

$$(3.1) \quad \begin{aligned} \dot{X} &= \lambda - dX - \beta XV \\ \dot{Y} &= \beta XV - aY \\ \dot{V} &= kY - uV - \beta XV - qVW \\ \dot{W} &= rV - hW - qVW \end{aligned}$$

正常細胞は λ で一定に産生され、 βXV で感染細胞となる。感染細胞は kY でウイルス粒子を産生する。正常細胞、感染細胞、ウイルス粒子の寿命はそれぞれ $1/d$, $1/a$, $1/u$ である。これらのことからウイルス粒子の基本再生産率は $R_0 = \beta\lambda k / (adu)$ で与えられる。ここで基本再生産率とはウイルス粒子1つ当たりから結果的に産生されるウイルス粒子の数を指して言う。また抗体はウイルス粒子量に比例した率 rV で産生され、その寿命は $1/h$ で表される。産生された抗体はウイルス粒子を捕らえ、減少率 qVW で減少させる。各パラメータはこれまでに発表されている実験事実を基に決定し、(3.1)式の連立微分方程式を数値計算の手法を用いて解く。

変異ウイルスは感染直後に出現すると仮定し、(3.1)式の数理モデルを以下の様に拡張する。変異ウイルス出現まで ($0 \leq t < \hat{t}$)、拡張された方程式は添え字を除いて拡張前の方程式と同じである：

$$\begin{aligned} \dot{X} &= \lambda - dX - \beta_1 XV_1 \\ \dot{Y}_1 &= \beta_1 XV_1 - aY_1 \\ \dot{V}_1 &= kY_1 - uV_1 - \beta_1 XV_1 - q_1 V_1 W \\ \dot{W} &= rV_1 - hW - q_1 V_1 W \end{aligned}$$

初めに侵入したウイルス粒子の増殖過程で変異ウイルスが出現することを仮定し、その出現時間 \hat{t} をウイルス粒子の寿命程度 ($1/u$) にとる。変異ウイルス出現以後 ($t \geq \hat{t}$) の動態を以下の連立微分方程式で記述する：

$$(3.2) \quad \begin{aligned} \dot{X} &= \lambda - dX - \beta_1 XV_1 - \beta_2 XV_2 \\ \dot{Y}_i &= \beta_i XV_i - aY_i \\ \dot{V}_i &= kY_i - uV_i - \beta_i XV_i - q_i V_i W \\ \dot{W} &= rV_1 + rV_2 - hW - q_1 V_1 W - q_2 V_2 W \end{aligned}$$

添え字 i は 1 か 2 をとる。変異ウイルスは出現に続き宿主細胞に感染し、感染した細胞 (Y_2) は変異ウイルス粒子 (V_2) を産生する。

3.2 宿主内での適応度

本稿ではRBタンパク質に変異が起きたものを変異ウイルスとして扱い、それ以外の変異については考慮しない。そのため変異ウイルスは宿主細胞受容体との結合能、及び抗体との結合能にのみ野生型との差を有し、数理モデルにおいてはパラメータの組 (β, q) で特徴付けられる。(結合能とパラメータとの関連は詳細を後述する。)ウイルス集団 $(\beta, q) = (x, y)$ について、その宿主内での適応度を変数 V の時間積分によって定義する：

$$w(x, y) = \int_0^{\tau} V(x, y; t) dt$$

宿主間の伝染力は宿主内のウイルス量に直接影響されていることを仮定している。感染宿主が医療機関による医療処置を受けるまで ($t < \tau$) の期間で時間積分を行い有効なウイルス量とした。

パラメータ (β, q) による 2 次元平面上で、ウイルス適応度ランドスケープは w の値の変化で与えられる。高い値の w を持つウイルスは新たな宿主へ効率よく感染出来ると考えられる。

3.3 変異ウイルスの宿主集団への定着

変異ウイルスの宿主集団への固定確率を計算するためには選択的優位性と初期頻度を定義する必要がある。選択的優位性は変異ウイルスと野生型の適応度の比で定義される：

$$s = \frac{w(x', y')}{w(x, y)} - 1$$

ここで $(\beta, q) = (x', y')$ は変異ウイルスを特徴づけている。頻度が p で与えられる変異ウイルスの固定確率は以下の微分方程式を境界条件 $u(0) = 0$ 及び $u(1) = 1$ のもとで解くことにより与えられる：

$$\frac{\nu(p)}{2} \frac{d^2 u(p)}{dp^2} + m(p) \frac{du(p)}{dp} = 0$$

$m(p)$ と $\nu(p)$ はそれぞれ以下の様にと与えられる：

$$m(p) = \frac{sp(1-p)}{1+sp}$$

$$\nu(p) = \frac{p(1-p)}{N}$$

N は宿主集団の規模を表している。全ての宿主はワクチン接種を受けているとし、変異ウイルスが出現するまでは野生型のウイルスに感染しているものとする。宿主個々が感染と治癒を経て随時入れ替わっていると、変異ウイルスが宿主集団に定着するまでは集団の規模は一定であると仮定している。変異ウイルスの固定確率は以下の様に得られる：

$$(3.3) \quad u(p) \equiv u(p; s) = \frac{1 - e^{-(2N-1)\log(1+sp)}}{1 - e^{-(2N-1)\log(1+s)}}$$

3.4 変異ウイルスの初期頻度

変異ウイルスは宿主内で野生型のウイルスと競合する関係にある。この競り合いの結果によって変異ウイルスが新たな宿主へ感染し得るかどうかが決まる。(3.2)式において野生型ウイルス (V_1) と変異ウイルス (V_2) についての時間積分を考える：

$$w_1(x, y) = \int_0^\tau V_1(x, y; t) dt$$

$$w_2(x', y') = \int_0^\tau V_2(x', y'; t) dt$$

変異ウイルスが出現した宿主以外の宿主 ($(N-1)$ 人) では野生型のウイルス (適応度： $w(x, y)$) のみが存在している。変異ウイルスの頻度は w_2 と w_1 の比でのみ考えるのではなく、全ての宿主でのウイルス量の総和との比で考える：

$$\frac{w_2(x', y')}{(N-1)w(x, y) + w_1(x, y) + w_2(x', y')}$$

この頻度に従い次の世代で変異ウイルスに感染する宿主の数の期待値は次のように表すことが出来る：

$$\eta = N \frac{w_2(x', y')}{(N-1)w(x, y) + w_1(x, y) + w_2(x', y')}$$

この期待値から次の世代の宿主数 (k) の分布はポアソン分布を用いて以下の様にと与えられる：

$$(3.4) \quad P(\eta; k) = \frac{e^{-\eta} \eta^k}{k!}$$

(3.3)式と(3.4)式を組み合わせると、変異ウイルスの固定確率が以下の様に得られる:

$$U(\eta, s) = \sum_{k=1}^N P(\eta; k) u\left(\frac{k}{N}; s\right)$$

3.5 パラメータ設定と構造データ

ワクチン接種者を100万人規模で想定し、その内の1%が感染しているとする($N = 10,000$). 正常細胞の寿命を100日($1/d = 100$ days)とし(Nowak and Bangham, 1996), 抗体の寿命は20日($1/h = 20$ days)としている(Insel and Looney, 2004). ウイルス粒子の寿命はウイルスにより大きく異なるがここでは2時間($1/u = 2h$)とし, 感染細胞については2日($1/a = 2$ days)としている(Watabe and Kishino, 2010).

SARS コロナウイルスのRBタンパク質と宿主細胞受容体(ACE2)の複合体構造データはX線結晶構造解析により測定されており, PDBで公開されている(Li et al., 2005, PDB code: 2AJF). 同様に抗体との複合体も2種類公開されており, 抗体80Rとの複合体(Hwang et al., 2006, PDB code: 2GHW)と抗体m396との複合体(Prabakaran et al., 2006, PDB code: 2DD8)を用いる.

3.6 宿主内動態と適応度ランドスケープ

図2(a)にウイルスの宿主内動態を示している. 基本再生産率が $R_0 = 100$ であり, パラメータ q について3つの値を適用している. それぞれの値で正常細胞の最低量が $Xd/\lambda|_{\min} = 0.99, 0.90, 0.50$ のように決まっている. (3.1)式の第一式でウイルス量(V)を0とした場合に得られる解($X_0 = \lambda/d$)は, ウイルス感染が無い状態での正常細胞量を表している. そのため正常細胞量(X)を X_0 で割った量 Xd/λ は, ウイルス感染が無い状態での正常細胞量に対する比に相当している. 感染初期ではウイルス量は対数尺度で線形に増加しており, 抗体の中和能力(q)には依存していない. 図2(b)に変異ウイルスと野生型とが競り合っている様子が示されている. 野生型の基本再生産率が $R_0 = 100$ で変異ウイルスでは $R_0 = 200$ の場合に相当し, パラメータ q は図2(a)において正常細胞の最低量が $Xd/\lambda|_{\min} = 0.99$ の場合で示している. 変異ウイルスが野生型を抑制している様子がわかる.

(3.1)式で示されている数理モデルを解き, 適応度ランドスケープを得た(図3, 抗体80Rとの場合を示した). 得られた適応度ランドスケープは数理モデルに含まれるパラメータの値によらず共通の特徴を持っている. 適応度は変数 V の時間積分で定義されているが, 感染期間を考慮して時間 τ で打ち切っている(ここでは $\tau = 10$ days). 更に低い基本再生産率($R_0 \sim 1$)ではウイルス動態は感染初期において抗体の中和能力には依存していない. そのため基本再生産率が低い領域(β の値の小さい領域)では適応度ランドスケープは抗体の中和能力(q の値)には依

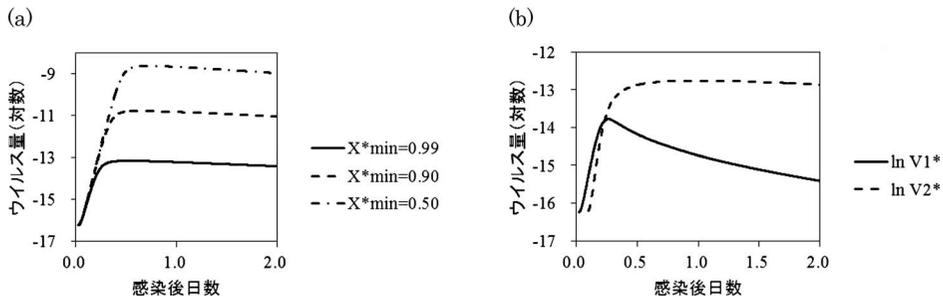


図2. 感染後のウイルス量の変化.

存していない。

3.7 複合体構造データを基にした変異のシミュレーション

先にも述べたように本稿ではウイルスの変異をRBタンパク質に起きた変異に限って考える。野生型ウイルスのRBタンパク質と宿主細胞受容体との結合能 $K_a^{CR(Wild)}$ は複合体構造データ (2AJF) を用いて (2.1) 式で計算できる。同様に抗体 (80R または m396) との結合能 $K_a^{AB(Wild)}$ も複合体構造データ (2GHW または 2DD8) を用いて計算できる。そしてRBタンパク質の結合領域においてアミノ酸残基を1残基変異させて再び複合体の結合能を計算する、 $K_a^{CR(Mutant)}$ 及び $K_a^{AB(Mutant)}$ 。我々はこれら野生型結合能と変異型結合能の比が3.1節の数理モデルでウイルスを特徴付けるパラメータ (β, q) における比と等しいことを仮定して数理モデルにおける変異ウイルスをシミュレートする:

$$\frac{\beta^{(Mutant)}}{\beta^{(Wild)}} = \frac{K_a^{CR(Mutant)}}{K_a^{CR(Wild)}}, \quad \frac{q^{(Mutant)}}{q^{(Wild)}} = \frac{K_a^{AB(Mutant)}}{K_a^{AB(Wild)}}$$

選択強度の分布を観るため任意に1000通りの1残基変異を起こした。それぞれの適応度ランドスケープ上での位置と宿主集団への固定確率をグレースケールで図3に示した。基本再生産率が $R_0 = 100$ で正常細胞の最低量が $Xd/\lambda|_{\min} = 0.99$ となる位置での $(\beta, q) = (x, y)$ を野生型としてシミュレートしている。幾つかの変異は1に近い固定確率を示している。

3.8 適応進化に寄与する主なアミノ酸残基

図4に宿主集団に高い確率で定着し得る変異を起こすアミノ酸残基を示した。丸印により囲まれている白い球は該当するアミノ酸残基の α 炭素原子の位置を示している。19通りの変異における固定確率の平均が0.4以上の残基を示しており、幾つかの変異では0.9を超える固定確率を示した。シミュレーションにおける野生型は図3の場合と同様である。図4(a)は抗体80Rとの場合を示しており、黒色で示した構造が抗体80Rの抗原結合部位である。薄い灰色で示した宿主細胞受容体ACE2の構造を重ね合わせて示している。図4(b)は抗体m396との場合を示している。

高い固定確率で宿主集団に定着し得る変異を起こす残基において突然変異率を考慮すること

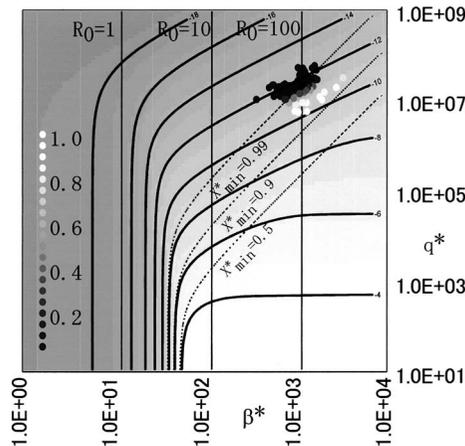


図3. 数理モデルにより得られた適応度ランドスケープ。結合領域における1残基置換による変異ウイルスの位置を固定確率 (0: 黒 ~ 1: 白) と共に示した。

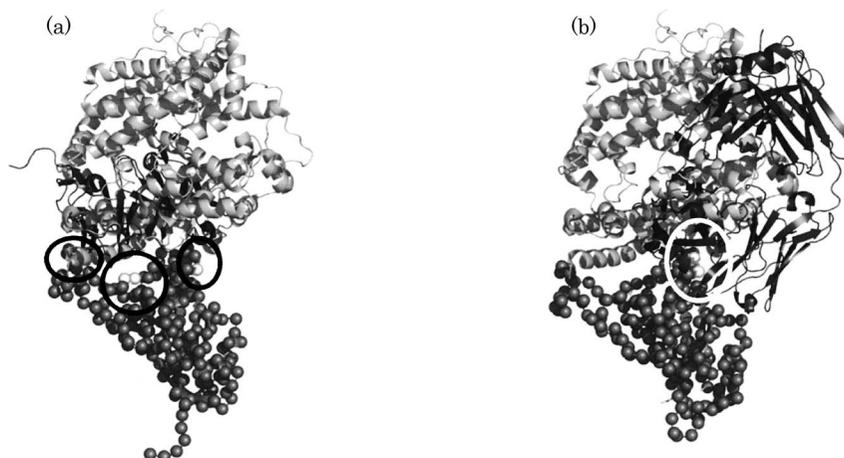


図 4. 宿主集団に高い確率で定着し得る変異を起こすアミノ酸残基を丸印で囲み示した.

で、宿主集団に定着する変異がどの程度の時間的な尺度で起こり得るのかを見積もることが出来る。そのことはワクチン開発において有用な知見を与えるものと考えられる。

4. まとめ

Flu ウイルスや SARS コロナウイルスが宿主細胞受容体へ侵入する最初の段階で機能する RB タンパク質について解析した。RB タンパク質はウイルス粒子の膜表面上に存在し、感染初期に直接機能するため宿主免疫系からの攻撃に常にさらされている。このタンパク質にかかる淘汰圧をまずは A 型 Flu ウイルスの 35 年に渡る変異の過程において調べた。RB タンパク質の宿主細胞受容体結合領域における抗体との結合親和性の変化を通してタンパク質にかかっていた選択圧の変化が見て取れた。抗体の脅威の度合いが大きい場合、選択圧は強くなり結合親和性を急激に変化させていたことが判った。そして SARS コロナウイルスを用いて正常細胞、感染細胞、ウイルス粒子、抗体の 4 者からなる宿主内動態の変化を解析し、変異ウイルスの宿主適応度を解析した。得られた適応度をもとに変異ウイルスが宿主集団に定着する確率を計算したところ、0.9 を超える高い確率で定着する変異が起き得ることが判った。これにアミノ酸残基毎の突然変異率を導入すれば、既存ウイルスと置き換わる変異ウイルスの出現時期を予測することも可能である。

参 考 文 献

- Felsenstein, J. (2005). *PHYLIP (Phylogeny Inference Package) Version 3.6*, Department of Genome Sciences, University of Washington, Seattle, Washington.
- Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W. A. and Liddington, R. C. (2006). Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R, *Journal of Biological Chemistry*, **281**, 34610–34616.
- Insel, R. A. and Looney, R. J. (2004). The B-lymphocyte system: Fundamental immunology, *Im-*

- munologic Disorders in Infants & Children*, 5th ed. (eds. E. R. Stiehm, H. D. Ochs and J. A. Winkelstein), 53–84, Elsevier, Philadelphia, Pennsylvania.
- Li, F., Li, W., Farzan, M. and Harrison, S. C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor, *Science*, **309**, 1864–1868.
- Nowak, M. A. and Bangham, C. R. M. (1996). Population dynamics of immune responses to persistent viruses, *Science*, **272**, 74–79.
- Prabakaran, P., Gan, J., Feng, Y., Zhu, Z., Choudhry, V., Xiao, X., Ji, X. and Dimitrov, D. S. (2006). Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody, *Journal of Biological Chemistry*, **281**, 15829–15836.
- Swofford, D. L. (2003). *PAUP* Version 4.0: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Sinauer Associates, Sunderland, Massachusetts.
- Watabe, T. and Kishino, H. (2010). Structural considerations in the fitness landscape of a virus, *Molecular Biology and Evolution*, **27**, 1782–1791.
- Watabe, T., Kishino, H., Okuhara, Y. and Kitazoe, Y. (2006). Fold recognition of the HIV-1 V3 loop and flexibility of its crown structure during the course of adaptation to a host, *Genetics*, **172**, 1385–1396.
- Watabe, T., Kishino, H., de Oliveira Martins, L. and Kitazoe, Y. (2007). A likelihood-based index of protein-protein binding affinities with application to influenza HA escape from antibodies, *Molecular Biology and Evolution*, **24**, 1627–1638.

Time Course and Spatial Distribution of Selection Pressure on Virus Proteins and Their Adaptation to Host Population

Teruaki Watabe¹ and Hirohisa Kishino²

¹Center of Medical Information Science, Kochi University

²Graduate School of Agriculture and Life Science, University of Tokyo

Proteins manage their adaptations to environments and/or gains of functions by substituting amino acid sequences. Therefore, mutations on a protein-coding gene are subject to the selection pressure of their environment. The strength and character of selection pressure may vary among the spatial regions of the protein structure and the temporal domains on evolutionary process. Thus, revealing the spatio-temporal fluctuation of the selection pressure gives us greater knowledge about the adaptive evolution of the protein. In this work, we first followed an evolutionary process of the influenza A hemagglutinin for 35 years and examined its long-term adaptation to its host population. By monitoring changes in the binding ability of hemagglutinin to antibodies, we can determine the changes in the selection pressure on the hemagglutinin. Second, we developed a mathematical model that describes the population dynamics of viruses, antibodies, and normal/infected cells within a host. The coefficients describe the binding affinity between the virus and the induced antibody and that between the virus and its receptor. We estimated the effect of a mutation in a binding region on the binding affinity. Using population genetic theory, we evaluated the probability that a mutant is fixed in a host population. We simulated the adaptive evolution of coronavirus, the etiological agent of severe acute respiratory syndrome, and showed that some mutations in the binding region may have high fixation probabilities in the vaccinated host population.