

厚生労働科学研究費補助金  
肝炎等克服緊急研究事業

ウイルス肝炎感染防止体制の確立に関する総合研究

平成19年度 総括・分担研究報告書

主任研究者 山口一成  
平成20(2008)年3月

## 研究組織

### 主任研究者

山口一成 国立感染症研究所・血液・安全性研究部

### 分担研究者

高橋孝喜 東京大学・医学部附属病院・輸血部

半田 誠 慶応義塾大学・医学部・輸血・細胞療法部

田所憲治 日本赤十字中央血液研究所

高松純樹 名古屋大学・医学部附属病院・輸血部

大戸 斉 福島医科大学・医学部附属病院・輸血・移植免疫部

紀野修一 旭川医科大学・臨床検査・輸血部

浜口 功 国立感染症研究所・血液・安全性研究部

古田里佳 大阪赤十字血液センター・研究部

水谷哲也 国立感染症研究所・ウイルス一部

## 目次

### I. 総括研究報告

共通 Motif 集合 primer 予測アルゴリズム (CoCoMo) を用いた HIVHCV および HBV 用 degenerate primer の設計 (山口一成)	5
図 1: CoCoMo における PCR 位置モチーフ選定アルゴリズムの流れ図	23
表 1: 塩基配列分析事例における 3mer 頻度	24
表 2: 塩基配列分析事例における 3mer の位置	24
表 3: 塩基配列分析事例における相互に相補的な 3mer の位置	24
表 4: 塩基配列分析事例における PCR 位置に存在する 3mer	25
表 5: 塩基配列分析事例における相互に相補的な 3mer の組み合わせと仮想 PCR での利用 PCR 数	25
表 6: 塩基配列分析事例における相互に相補的な 3mer のうち一方の削除	25
表 7: 塩基配列分析事例における選択された 3mer による仮想 PCR	26
図 2: 鋳型作成およびその後の TAP-PCR の流れ	26
図 3: 合成鋳型	27
表 8: モデル実験での鋳型作成用 primer	28
表 9: TAP-PCR 用鋳型作成に用いた primer とその確認に用いた primer	28
表 10: TAP-PCR 用鋳型作成に用いた鋳型と primer の組み合わせ	29
表 11: TAP-PCR 用 primer	30
表 12: 鋳型作成用 primer	30
表 13: TAP-PCR 用 primer	31
図 4: コロナウイルス科の遺伝的近縁性相関図	32
表 14: ウイルスへの適用実験に用いた primer	33
図 5: PCR で作成された TAP-PCR 用鋳型の両端	35
図 6: 作成された鋳型の正確性	35
図 7: degenerate region の degeneracy の限界	35
図 8: TAP-PCR の感度の限界と TAP region の長さの影響	36
図 9: primer 全長 25mer での TAP-PCR	37
図 10: 同一の primer による TAP-PCR	37
図 11: SARS コロナウイルス感染細胞由来 cDNA を鋳型とした TAP-PCR	38
図 12: HIV-1 ゲノムデータの social network analysis	39
表 15: 第 1 二次グループ内共通 degenerate primer	40

表 16: 第 2 二次グループ内共通 degenerate primer	40
表 17: 第 3 二次グループ内共通 degenerate primer	41
表 18: GenBank から入手した HCV ウイルス塩基配列の一次グループ化	42
表 19: HCV ゲノム塩基配列のサブグループ間を網羅する degenerate primer	48
図 13: HCV 共通 degenerate primer による HCV cDNA 増幅	49

## II. 分担研究報告

1. ウイルス遺伝子の非特異的増幅方法の確立 (水谷哲也)	50
図 14: Sigma 社の非特異的ウイルス核酸増幅反応	54
図 15: 改良した核酸増幅反応	55
2. ウイルス核酸検出のための DNA チップ作製と検出反応条件の検討 (浜口功、半田誠)	56
図 16: DNA チップの作製と HCV、HBV 核酸の検出	59
3. 国内技術を基盤としたウイルス検出用新規核酸増幅法の開発 (古田里佳、田所憲治)	60
4. 肝炎ウイルスと院内感染及び潜伏肝炎ウイルス斉活性化の研究 (紀野修一、大戸斉、高橋孝喜、高松純樹)	63
図 17: 平成 19 年度「輸血業務に関する総合的アンケート」	70
III. 研究成果の刊行に関する一覧表	85
IV. 研究成果の別刷	87

厚生労働科学研究費補助金（肝炎等克服緊急研究事業）

総括・分担研究報告書

ウイルス肝炎感染防止体制の確立に関する総合研究（H19-肝炎一般-003）

研究課題：共通 Motif 集合 primer 予測アルゴリズム (CoCoMo) を用いた HIV、HCV  
および HBV 用 degenerate primer の設計

主任研究者： 山口一成 国立感染症研究所 血液・安全性研究部

協力研究者： 遠藤大二 酪農学園大学獣医学部

### 研究要旨

ウイルスの実際の塩基置換は、完全にランダムではなく、ある傾向を持つことが予想されており、また、ゲノムの塩基配列そのものについても一定の志向性があることが知られている。6-10 塩基程度のパターン (motif) の頻度を調べた場合、動物種ごとに出現頻度の傾向が存在していることが明らかにされている。また 6-12 塩基程度の短い配列の出現頻度の傾向は、HIV、HCV および HBV においても認められた。本研究では、ウイルス内の 6-12 塩基の塩基配列パターンの傾向を primer 設計に応用するプログラム群を開発し、HIV、HCV および HBV での degenerate primer の設計への適用を試みた。

#### A. 研究目的

近年、血清中のウイルスを高感度および低コストで検出するための方法として、核酸の検出方法に基づく NAT 法が用いられている。この方法の核となるのは、ウイルス RNA または DNA 断片を検出するための PCR 法である。したがって、高い効率で確実に対象ウイルスを検出するための PCR 法の開発は、現行での検出方法の

感度をさらに高めるために必要といえる。

PCR 法は、対象遺伝子上の二か所に結合する合成ヌクレオチド (オリゴマー) から DNA 合成が開始される現象を利用して 100-5000 塩基程度の遺伝子断片を特異的に増幅させる方法であるが、遺伝子上での位置の特異性を担保するために、ヌクレオチドが 18-30 個連結した (18~30 塩基の) のオリゴマーが用いられる。このオ

リゴマーはDNA合成の開始点となることから、PCR法に用いる場合には primer とされる。primer によるDNA合成の開始は、primer が鋳型DNAと二重鎖を形成する個所から開始されるが、正確に塩基配列が一致する箇所は全ゲノム上でも、想定された一か所であることが予想されている。通常のPCR法では合成開始点の正確性は、PCR法での遺伝子の検出には有利に働くが、鋳型上に塩基配列の置換が存在するウイルス等の場合には、塩基の不一致が混入する可能性が高く、正確な一致を前提としない primer の使用方法が求められる。第一の選択肢としては、特定の位置に二種以上の塩基が混合した状態で合成されたオリゴマーを primer として用いる方法が用いられる。たとえば、二種の鋳型の塩基配列が

ATGGGCAGGGA

ATGGCCAGGGA

であった場合、5番目の位置の塩基をGとCの混合塩基として

ATGG{CG}AGGGA

という配列の primer が使用される。このような primer は塩基が縮重していることから、degenerate primer と呼ばれる。

HIV、HCV および HBV など、多様な塩基置換を生じているウイルスを鋳型として PCR 法を実施する場合には、この degenerate primer を使用することにより、多くの変異株を同時に検出することが可能となる。

検出の範囲を広めるための degenerate primer を設計する場合、既知の塩基配列をデータベース上から取得し、配列を整列(アライメント)させ、株間での塩基置換または変異が少ない領域を目視した上、primer 合成配列が決定される。実際には、多くの領域が候補に挙がるため、多数の候補配列から実験的に primer を絞り込む作業が行われる。このような primer 設計作業は、primer が鋳型に結合する際の物理化学的な性質への primer 上の degenerated 塩基の位置の影響や、degenerated 塩基を最小にとどめるための配列の選択など、複雑な因子が多数関係するため、単純なプログラムでの解決が難しい。しかしながら、既知の配列は日常的に報告数が増加しており、実際の変異株の多様性が把握できるようになった一方で、塩基配列が 200 本以上になった場合、配列の整列結果と人間の目視に基づいて primer 設計を進めることは難しくなっている。

primer の設計などの技術についての研究は1980年代から行われており、近年は、数学的アルゴリズム研究者による primer 設計の検討が盛んになっている。数学的アルゴリズム研究者は、多数の鋳型塩基配列に共通する degenerate primer の設計について検討し、多くの場合、可能な primer のすべてを検討する方法では莫大な時間がかかることを考察している。Weirauchらは、degenerate primer の選択方法を下記の三種の問題に整理し、それぞれにつ

いて現実的な計算方法を提案している。

1) 各 degenerate primer が最大数の鋳型を増幅可能にするための primer の塩基長および degeneracy (縮重塩基の多様性の程度) を求める。

2) 各 degenerate primer が最小の degeneracy を取るための primer の塩基長および最小の degeneracy を求める

3) 複数の degenerate primer が必要となる場合、最少数の primer により全鋳型を増幅するための primer の選択。

これらアルゴリズムに基づいて、primer を設計するプログラムも開発されている。しかしながら、上記の手法での決定される primer は、非常に多数になり、実際に使用する研究者が設計手法について理解することがきわめて難しいため、しか、最適な primer を選定するためには、その問題を満たす primer 候補を挙げた上、多数回の実験を実施するしかない。現実的には、整列に基づいて primer 候補を絞り込む場合と作業量が変わらないこととなる。

ウイルスの実際の塩基置換は、完全にランダムではなく、ある傾向を持つことが予想されており、また、ゲノムの塩基配列そのものについても一定の志向性があること知られている。このような傾向は、単純な4種の塩基の構成では見られないが、6-10塩基程度のパターン(motif)の頻度を調べた場合、動物種ごとに出現頻度の傾向が存在していることが明らかにされている。分担者はRNAについても

同様にウイルスと宿主の間にはモチーフの出現頻度の相違があることを見出した。このような、6-12塩基程度の短い配列の出現頻度の傾向は、HIV, HCV および HBV においても認められた。従来提唱されている primer 設計アルゴリズムでは、CODEHOP などのようにアミノ酸の出現頻度と縮重の影響を利用する方法が知られているが、ウイルスのように十分なゲノムサイズがない場合には、アミノ酸配列の限定は最適な選択肢を示すとは限らず、宿主と類似した配列を選択してバックグラウンドの原因になる可能性もある。本研究では、ウイルス内の6-12塩基の塩基配列パターンの傾向を primer 設計に応用するプログラム群を開発し、HIV, HCV および HBV での degenerate primer の設計への適用を試みた。このような primer の設計アルゴリズムは既知の分離株間で共通する motif を集めることから CoCoMo 法(Collection of common motifs)と称し、また、実際の PCR 法は3'末端の配列のみが鋳型と正確に整列可能であることから TAP-PCR 法と称することとする。

## B. 研究方法

### 1. primer 設計プログラム群の開発環境

#### 1) ハードウェアおよびオペレーティングシステム

CoCoMo アルゴリズムを実施するためのコンピュータとしては、64bit CPU

(Athlon64)および2 G バイトのメモリーを搭載した DOS/V 企画パーソナルコンピュータを用いた。プログラムを稼働するためのオペレーティングシステムとしては、Linux (SUSE10.2 または Ubuntu7.10) を用いた。

## 2) 開発言語およびデータベース

開発言語としては、Ruby 1.86 を、同言語のバイオインフォマティクス用の拡張モジュールとしては bioruby を、データベースとの接続モジュールとしては MySQL/Ruby を使用した。鋳型上の 6-12 塩基のモチーフの出現頻度を累計するための gas ソフトをコンパイルおよび稼働するためには、gcc C コンパイラを使用した。一組の塩基配列についての相同性の算出には、NCBI が提供している blast プログラム群のうち bl2seq を使用した。

元遺伝子データ、処理過程の一時データ、予測 primer 塩基配列、位置情報および方向等の付随情報を格納するためのデータベースには MySQL 5.0 を使用した。データベース設定の際には、maximum packet size を 200 MB とすることにより、大量のデータ挿入を可能にした。

塩基配列間の相同性全体を図示するためには、social network 分析用ソフトの Pajek を使用した。

## 3) 開発用検証用データの入手

開発中の小規模データおよび各ウイルスの既知の全遺伝子データは、EMBL または GenBank よりウイルス種ごとにダウ

ンロードし、Ruby で加工後 MySQL に格納して使用した。

## 2. CoCoMo アルゴリズムおよび実行用プログラム群の構成

CoCoMo アルゴリズムでは、基本的な primer 選定方針を提示し、実際の設計においては、各ステップ用のプログラム群を個別に準備した。これは、当該アルゴリズムが対象塩基配列の傾向に依存しているため、実施中の結果に応じて、実施条件や使用プログラムの一部を変更する必要が生じたため、一貫したプログラムとして準備せず、各ステップを実行後にデータを参照して次ステップのプログラムと条件を選択したためである。

### 1) CoCoMo アルゴリズム

CoCoMo は、degenerate primer を設計するために下記の基本的には下記の処理を実施する。

#### ① 検索後ダウンロードされた遺伝子データのデータベースへの格納

ウイルス種別の遺伝子データを収集する際には、ウイルスゲノムプロジェクト上の整備されたゲノムデータのみでは、HCV でも 90 個程度のみがリス手アップされているのみで、既知の遺伝子データを網羅しているとは言えない。そのため、遺伝子データとしては、EMBL または GenBank データベース上で、種名としてウイルス種を特定した検索を実施し、探索された塩基配列データをすべてダウンロ



ードした。特に GenBank は、2005 年ごろから 50 MB を超える大量の検索データのダウンロードも許可しているため(分担者の実施から推測、未確認)、各種ウイルスデータをダウンロードする際には、主として用いた。

② 格納された遺伝子データのサイズおよび相同性を用いた類似塩基配列群(一次グループ)への分類

格納されたデータは、前述のように、ゲノムや特定の抗原遺伝子のみではなく、名称の異なる様々な断片を含んでいる。しかしながら、CoCoMo では、塩基配列間で共通する motif の頻度を参照して primer を設計するため、設計に先だって、塩基配列データを特定遺伝子領域やゲノム全長に分ける必要が生じた。GenBank の規約上では、このような分類は塩基配列に付随する Annotation 情報から読み取ることができることになっているが、塩基配列の登録者は必ずしも同一の遺伝子名称やゲノム全長との記述をしているわけではないため、Annotation による分類は十分な実用性を示さなかった。そのため、塩基配列の分類のため、塩基配列を長さが大きい順に並べ、順次それよりも短い配列と相同性とサイズの類似性を基準として塩基配列を分類した。通常、サイズの差を 30%以内で、50 塩基以上の相同領域が存在する条件を設定することにより、ゲノムや主要抗原に関する塩基配列を分類し、同一グループにまとめることが可能であった。

③ primer 設計対象塩基配列群の選定と、塩基配列相互の組み合わせについての相同性の算出

②においてグループ化された塩基配列のうち、ゲノムに相当する塩基配列グループと、250 塩基以上のサイズで、100 以上の塩基配列から構成されるグループについては、塩基配列相互の相同性を算出した。

④ 対象塩基配列相互の相同性の程度と相互の相同性についての social network analysis による二次グループに分類

Social network analysis は、相互に結合されるノードの親密性やグループ分けを行う手法で、塩基配列間でも一定以上の相同性の存在やその相同性の数値によってグループ化が可能になる。相同性やその数値によって分類が可能。

⑤ 二次グループ数が 20 程度の場合には、各グループから代表のデータを選別して、グループ代表による代表グループを作出

PCR の位置にあるオリゴマーを選択するソフトウェアは、その第一段階で大量の計算容量を必要とする。HCV のゲノム程度の大きさの塩基配列の場合には、対象塩基配列数を 20 程度以内に抑える必要がある。そのため、いったん形成されたグループ内でのバランスをとりつつ、塩基配列数を減らすための方法として、相同性の高いウイルス群から代表を選択するという方法を選んだ。

⑥ 一旦、可能な 6 塩基モチーフをすべて想定した上、対象塩基配列上で PCR の位置となる 6~10 塩基 motif を選択する (図 1)。まず、レパートリー数が 4,096 種であるため、primer を予測するための計算の実行可能性が十分に高いことから、モチーフの基本を 6 塩基とした。6 塩基の配列のみでは、オリゴマーを合成した場合でも primer として働くことはできないが、PCR primer と同様に対象塩基配列上に向かい合わせの位置に設定されるような 6 塩基 motif を選択することにより、その外側の塩基配列を連結することにより、degenerate primer を設計した。degenerate primer 設計時の 3' 端のモチーフは長いほど 3' 端の配列が安定して長くなるため、6 塩基の予測終了後、motif を伸長することにより、PCR の位置にある motif を選択した。

motif の選択においては、PCR を想定する二種の motif が対象塩基配列上のどの場所でも相補的に重複した位置に存在した場合、最終的に作成した degenerate primer でも相補的に primer が自己結合してしまう可能性が高くなる。そのため、6 塩基 motif の選択はそのような重複 motif のうちの片方を削除する形で進めた。以下に 5' -ATGGATGTGGCAT-3' (相補鎖: 5' -ATGCCACATCCAT-3') という単純な配列を鋳型とした 3mer primer の選択を事例として説明する。

まず、センス鎖、相補鎖の両方において 5' 側から 3mer ずつ読み、その総数を

数えた (表 1)。

次に存在位置と方向を 3mer ごとに記録した。表 1 の事例について 3mer の位置を算出すると表 2 のようになる。

続けて 2 塩基以上が重なる組み合わせの 3mer を全てリストアップした。事例では、ATG と CAT、ATG と CCA などが相補的な 3mer の組み合わせとしてリストアップされた (表 3)。

続けて、相補的な 3mer の組み合わせのうち、片方を候補群から削除するため、仮想的に PCR を実施した。具体的には、表 1 で算出された対象塩基配列上の位置を基に生成する可能性のある PCR 産物を全て想定し、各オリゴマーについて、primer として寄与する PCR 産物の数を積算した。

3mer の事例で、オリゴマー間の間隔が 3 塩基以上開いている場合に PCR の位置にあるとして、表 4 のような PCR 産物の生成が予想される。下表中では ATG が 6 件の PCR 産物に利用されている。

3mer での事例では、一件の塩基配列データを想定しているため、その一件の中での仮想 PCR における利用回数を基準に 3mer の選択をした (表 5)。

上記の事例の相補的 3mer のうち、PCR に利用される頻度の少ない側を消去すると表 6 となる。

上記の事例のうち、右側の表の GTG-CCA の相補的組み合わせでは、CCA が仮想 PCR での利用頻度が高いが、CCA は、ATG-CCA 間での比較で削除されているため、GTG

は削除する必要がなくなる。ただし左側の表で GTG-CAC 間での比較で削除されることになるため、GTG および CCA の双方が削除されることとなる。この相補的な motif の除去は、相補的な組み合わせが存在しなくなるまで実施する。

3mer の事例では、一回目の削除で全ての相補的組み合わせが無くなったが、実際の事例では、多数回この作業が繰り返された。想定している事例では、一方の 3mer を候補から削除した後、仮想 PCR は下表のように変更され、PCR がおきることが予想される数は 12 件に減少した (表 7)。

3mer の事例では、ATG-ATG および ATG-TGC の組み合わせについて 5 件の PCR が予想されるため、これらが選択されることとなる。実際の実施時には、完全に対象ゲノム上での重複がない 6 塩基 motif が選択された。

⑦最終的に選択された motif の 5' 端の配列を調べ、実際の対象配列上に存在する塩基の多様性をすべて満たす degenerate primer を作出した。

## 2) 塩基配列の格納および分類プログラム群

上記の CoCoMo アルゴリズムを実際に処理するためには、各ステップごとにプログラムを準備した。

## 3. 長鎖合成オリゴマーを用いた設計方法の検証実験

6-10 塩基のオリゴマーを 3' 端に存在

する primer として設計を行う場合、3' 端のオリゴマーの長さが増幅能力に関する関係を検討する必要がある。そのため、長鎖の合成オリゴマー利用したモデル実験を実施した。

### 1) 鋳型の作成

TAP-PCR 法用の primer 設計条件を導出する実験には、両端に degeneracy をもたせた合成オリゴマーを鋳型として用いた。ただし、全ての場合の鋳型を合成すると、その数が多量であるため莫大な費用がかかる。そこで、まず核となる部分を合成オリゴマーとして準備し、その合成オリゴマーを鋳型としてそれぞれの degeneracy をもたせた primer で PCR をかけることで目的の鋳型を作成した (図 2)。PCR による鋳型作成後は 2 本鎖になるが、図では 1 本鎖のみ示す)。このとき、核となる鋳型の末端で、primer が結合する種間共通配列を TAP 配列、またその primer によって接続される degeneracy をもった領域のことを degenerated 配列と呼び、同様に TAPprimer の種間共通配列を TAP region、またその 5' 側の degeneracy をもった領域を degenerate region と呼ぶこととした。

### 2) core 鋳型の合成

まず、ssRNA ウイルスの Influenza A virus (A/Puerto Rico/8/34(H1N1)) segment 1 と Newcastle disease virus から、共通の 8 塩基の配列 (5' -AACCCAGC-3'、5' -CCTGGTCA-3' ) を in

silico MolecularCloning(R) Standard Editionで検索した。5' -AACCCAGC-3' の配列は H1N1 では 151-158 に、NDV では 1599-1606 と 4568-4575 に存在した。また 5' -CCTGGTCA-3' の配列は H1N1 では 472-479 に、NDV では 12958-12965 と 14367-14374 に存在した。

実験に使いやすい長さにするため、H1N1 および NDV より共通の 8 塩基には含まれる配列を選び出した。すなわち共通の 8 塩基の間を両端から同じ塩基数ずつ切り出し H1N1 は 50bp と 150bp とし、NDV は 4568-4575 と 12958-12965 の間から切り出し 100bp とし、それらの配列を core 配列として用いた。また、その後の実験で共通部分を延長するため、NDV の共通部分の内側には H1N1 の共通部分の両端の配列をそれぞれ 8 塩基ずつ接続させた。

以上のようにして合成した core 鋳型を、短いものから順に coreInfl50、coreNDV100、coreInfl150 と命名した (図 3)。

### 3) TAP-PCR 用鋳型の作成

次にこれらを核として両端に degeneracy をもたせた鋳型を PCR で作成した。鋳型としては、coreInfl50、coreNDV100、coreInfl150 を用いた。primer としては、共通部分の 5' 側にそれぞれ 8 塩基の配列を接続したのを用いた。これは coreInfl150 で用いる primer (H1N1\_f、H1N1\_r) を基準とし、coreInfl50 と coreNDV100 は、この 5' 側

の 8 塩基中に 1~5 塩基の degeneracy をもたせた primer を用いた (表 8)。鋳型が coreInfl50 の場合、primer は Infl50\_f-1~f-5、Infl50\_r-1~r-5 を用いた。同様に coreVDV100 の場合は NDV\_f-1~f-5、NDV\_r-1~r-5 を用い、coreInfl150 の場合は H1N1\_f、H1N1\_r を用いた。この PCR は以下のようにして行なった。すなわち 2 × Go Taq Green Master Mix (Promega) 12.5 μl、DDW 11.1 μl、Template (10pmol/μl) 1 μl、Primer (50pmol/μl) 各 0.2 μl を混合し、94°C 2 分 × 1、[94°C 1 分、31°C 1 分、72°C 3 分] × 13、[94°C 1 分、60°C 1 分、72°C 3 分] × 38、72°C 10 分 × 1 で反応させた。31°C は共通の 8 塩基の配列のアニーリング温度である。サーマルサイクラーは、i Cycler (Bio-Rad Laboratories)、My Genie®32 Thermal Block (BIONEER)、T GRADIENT (Whatman Biometra) の 3 種類を用いた。それぞれの PCR 産物を Infl50(1)~(5)、NDV(1)~(5)、Infl150 と命名した。

### 4) TAP-PCR 用鋳型の精製

PCR で作成した鋳型には、PCR 反応による不純物が含まれている可能性があるため、TAP-PCR への利用に先んじてこれらの産物を SpinClean™ PCR Purification Kit (Messenger of biotechnology) で精製した。

### 5) 作成された鋳型の確認

上記の作業により鋳型の両端に

degenerate 配列を接続したが、これを PCR 法により確認した。

確認に用いる鋳型は Inf1150 とし、TAP-PCR 用鋳型を作成するために鋳型としては coreInf1150 を用い、primer は Inf1150\_f-25-9 と Inf1150\_r-25-9 (表 9) を用いて PCR を行なった。この PCR 産物を d-Inf1150 と命名した。

ここで作成した d-Inf1150 が coreInf1150 の両端に degenerated 配列を接続したものであることを確認するために、片方に特異的な primer をそれぞれ用いた PCR を行なった。まず鋳型としては、coreInf1150 と d-Inf1150 をそれぞれ用いた。primer としては Inf1150\_degenerated-f と Inf1150\_degenerated-r、および Inf1150\_TAP-f と Inf1150\_TAP-r を用いた (表 9)。Inf1150\_degenerated は degenerated 配列に、Inf1150\_TAP は TAP 配列にプライミングする primer である。PCR 終了後、各サンプルを 2.5%アガロースゲル上で 5 $\mu$ l ずつ電気泳動し、バンドの有無を確認した。

## 6) degenerate region の degeneracy の限界の検討

鋳型は前述のように作成し、両端に degeneracy をもたせた鋳型 [Inf150(1)~(5)、NDV(1)~(5)、Inf1150] を用いた。primer は degenerate region の degeneracy がそれぞれ 1~5 塩基の TAP primer を用いた。鋳型は 3 種類を混合し

て用いた。すなわち、Inf150(1)、NDV(1)、Inf1150 の混合に対し primer は Inf1-NDV\_F-1、Inf1-NDV\_R-1 を用いた (表 10)。degeneracy が 2~5 塩基の鋳型と primer に対しても同様にして混合して PCR を行なった。

## 7) 感度の限界の検討

次に、primer の TAP region の塩基数を増加させ、degeneracy の影響を検討した。同時に、それぞれの塩基数において鋳型を限界希釈することで、感度の限界を検討した。

鋳型と primer は degenerate region の degeneracy が 4 塩基のものを用いた。鋳型は Inf150(4)、NDV(4)、Inf1150 を 1 $\mu$ l あたり 10<sup>6</sup>~10<sup>1</sup> コピーに希釈して、それぞれ同じ濃度のものを 3 種類混合して用いた。primer は TAP region が 8~12 塩基のものを用いた (表 11)。

### i) degenerate region の長さの検討

実際のウイルス群では、degenerated 配列が長く、その中に degeneracy をもった塩基が点在している場合もあると思われる。よって、短い TAP region と長い degenerate region で構成される全長 25mer の primer で同様に鋳型の作成と TAP-PCR を行ない、degenerate region の長さや primer 長の影響を調べた。

まず、TAP-PCR 用鋳型の作成を行なった。鋳型としては coreInf150、coreNDV150、coreInf1150 を用いた。鋳型が coreInf150 の場合、primer は Inf150\_f-25-8 と

Inf150\_r-25-8、また Inf150\_f-25-9 と Inf150\_r-25-9 を用いた。primer が Inf150\_f-25-8 と Inf150\_r-25-8 のときの PCR 産物は Inf150 (25-8) と命名した。他の鑄型に対しても同様に PCR を行い、その鑄型と primer の組み合わせ、またその反応で生成される TAP-PCR 用鑄型の名称は表 12 に示す。

次に TAP-PCR を行なった。鑄型は上記で作成したものを  $1\mu\text{l}$  あたり  $10^6\sim 10^1$  コピーに希釈し 3 種類を混合して用いた。primer は TAP region が 8mer と 9mer のものを用いた (表 6)。反応は、 $2\times$ Go Taq Green Master Mix (Promega)  $12.5\mu\text{l}$ 、DDW  $9.1\mu\text{l}$ 、Template 各  $1\mu\text{l}$ 、Primer ( $50\text{pmol}/\mu\text{l}$ ) 各  $0.2\mu\text{l}$  を混合し、 $94^\circ\text{C}$  2 分 $\times$ 1、 $[94^\circ\text{C}$  1 分、 $A^\circ\text{C}$ ※ 1 分、 $72^\circ\text{C}$  30 秒] $\times$ 38、 $72^\circ\text{C}$  10 分 $\times$ 1 で反応させた (※それぞれの primer のアニーリング温度は、TAP region が 8mer のものは  $53.9^\circ\text{C}$ 、9mer のものは  $62.5^\circ\text{C}$ )。PCR 終了後、各サンプルを 2.5%アガロースゲル上で  $5\mu\text{l}$  ずつ電気泳動し、バンドの有無を確認した。

#### ii) 完全一致の primer での TAP-PCR

ここまで degeneracy をもった鑄型と primer の組み合わせで PCR を行なってきたため、次に degeneracy が PCR に与える影響を検討した。まず degeneracy をもたないように 3 種の core 鑄型全てに共通の primer で TAP-PCR 用鑄型を作成し、さらにその primer で TAP-PCR を行なった。

TAP-PCR 用鑄型を作成するために、鑄型としては coreInf150、coreNDV150、

coreInf150 を用いた。primer は Inf150\_f-25-9 と Inf150\_r-25-9 を用いた。

次に TAP-PCR を行なった。鑄型は上記で作成したものを  $1\mu\text{l}$  あたり  $10^6\sim 10^1$  分子に希釈し 3 種類を混合して用いた。primer は Inf150\_f-25-9 と Inf150\_r-25-9 を用いた。反応は、 $2\times$ Go Taq Green Master Mix (Promega)  $12.5\mu\text{l}$ 、DDW  $9.1\mu\text{l}$ 、Template 各  $1\mu\text{l}$ 、Primer ( $50\text{pmol}/\mu\text{l}$ ) 各  $0.2\mu\text{l}$  を混合し、 $94^\circ\text{C}$  2 分 $\times$ 1、 $[94^\circ\text{C}$  1 分、 $63^\circ\text{C}$  1 分、 $72^\circ\text{C}$  30 秒] $\times$ 38、 $72^\circ\text{C}$  10 分 $\times$ 1 で反応させた。PCR 終了後、各サンプルを 2.5%アガロースゲル上で  $5\mu\text{l}$  ずつ電気泳動し、バンドの有無を確認した。

#### 8) SARS コロナウイルスを用いた CoCoMo 設計プログラムの実効性の検証

primer の対象とするウイルスは SARS corona virus、Bat SARS corona virus、Civet SARS corona virus とした。これらのウイルスはコロナウイルス科において比較的近縁であることから、適用が容易な事例として用いられた (図 4)。実験は感染症の危険性から、感染症研究所の村山分室の P4 施設で実施された。

#### SARS corona virus 感染細胞の作成

SARS corona virus 感染細胞の作成には Vero E6 細胞と、Frankfurt 1 株として分離された SARS corona virus を用いた。

Vero E6 細胞は通常 75cm<sup>2</sup> のフラスコ内で、0.2mM/ml の L-glutamine、100U/ml のペニシリン、10 μg/ml のストレプトマイシン、5%(v/v) の牛胎子血清 (FBS) を添加した DMEM (Sigma-Aldrich) 中で経代培養され、5% CO<sub>2</sub> 大気中に 37°C で維持された。細胞は 25cm<sup>2</sup> フラスコにいったん分けられ、100% 集密に達するまで培養された。その培養液はウイルス感染の前に、2% の FBS を含んだ DMEM に置き換えられた。その後、SARS corona virus Frankfurt 1 株の感染は、感染効率 (m. o. i) が 10 の細胞内で行なわれた。

RNA の抽出には、Isogen (Nippon gene) を用いて定法に従った。すなわち、Isogen に浸漬した細胞 200 μl に Isogen 600 μl を加え、ボルテックスをかけた後、12,000rpm で 15 分遠心し、上精を採取した。上精に 500 μl のイソプロピルアルコールを加え、13,000rpm で 10 分遠心、沈殿を 500 μl の 75% エタノールで洗浄し、13,000rpm で 5 分間遠心した。上精を捨てた後、ペレットを完全に乾燥させ、RNase free water (DDW) 30 μl に溶出し、RNA として用いた。

RT には、M-MLV Reverse Transcriptase (SIGMA-Aldrich ジャパン) および Superscript III (INVITROGEN) を用い、それぞれ 1 本鎖、2 本鎖 cDNA 合成を行った。方法は各製品の説明書に従った。すなわち、Random Primer もしくは NR9 6 1 μl、10mM dNTP mix 1 μl、RNA RNA 溶液 1 μl、DDW 7 μl を氷上で混合した後、

70°C で 10 分間インキュベートした。その後反応物を氷上で急冷し、10×M-MLV Reverse Transcriptase Buffer 2 μl、続いて 2 本鎖目の cDNA を合成するため、M-MLV Reverse Transcriptase 1 μl、RNase Inhibitor (20 U / μl) 1 μl、DDW 6 μl を氷上で混合した。37°C で 50 分間インキュベートし、1 本鎖 cDNA を合成した。

M-MLV Reverse Transcriptase で合成した 1 本鎖 cDNA 全量 20 μl、DDW 83.5 μl、5×Second-Strand Reaction Buffer 30 μl、10mM dNTP mix 3 μl、E. coli DNA Ligase (10U/μl) 1 μl、E. coli DNA Polymerase I (3.5 U / μl) 11.5 μl、E. coli RNase H (2 U / μl) 1 μl を混合し、16°C で 2 時間インキュベートした。その後 T4 DNA Polymerase (5 U / μl) 2 μl を加え、16°C で 5 分間反応を続けた。反応物を氷上に置き、0.5 M EDTA 10 μl を混合し、2 本鎖 cDNA を合成した。合成終了後、フェノールおよびクロロホルム抽出により反応を停止し、続けてエタノール沈殿を行った。すなわち、合成終了後の溶液に同量のフェノールを加えた後ボルテックスをかけ、12,000rpm で 20 分間遠心後、水層を新たなチューブに分取した。クロロホルムについても同様の操作を施した後、溶液の 1/10 量の 3M 酢酸ナトリウムと 0.1 μl のグリコーゲン (Roche Applied Science) を加え、さらに 2.5 倍量の 99.5% エタノールを加えて転倒混和した。室温で 13,000rpm、15 分間遠心後、チュ

ープの底に核酸の沈殿を確認し、上精を完全に廃棄した。さらに 70%エタノールを 1 ml 加えた後同様の操作を行い、ペレットを風乾させ、30  $\mu$ l の DDW に溶解させた。DDW に溶解させた 2 本鎖 cDNA を SpinClean PCR Purification Kit (M-Biotec) で精製した。

鑄型としては上記のように作成した SARS coronavirus 感染細胞由来 cDNA を用いた。primer は SARS coronavirus、Bat SARS coronavirus、Civet SARS coronavirus の 3 種に対応するものを、TAP region を 10mer として設計した (表 14。増幅産物サイズは SARS コロナウイルスに対応するものである。)。PCR は定法に基づいて行ない、PCR 終了後、各サンプルを 2.5%アガロースゲル上で 5  $\mu$ l ずつ電気泳動し、バンドの有無を確認した。

### iii) HIV-1 および HCV 検出用 degenerate primer の設計

前述 2. 2) に示したプログラム群を利用して、種内の変異株全体をカバーする primer をデザインした。

HIV-1 については、取得されたゲノム塩基配列数が多数であり、かつ多様であったため、social network analysis で形成されたグループごとにグループ内の鑄型を共通して増幅する primer を設計した。

HCV については、social network analysis で形成された一次グループの代表をグループとしてまとめなおした場合も、代表全体をカバーする degenerated primer が設計されたため、それらの

degenerated primer を種全体をカバーする primer とみなした。

## C. 研究結果

### 1. TAP-PCR 試験用鑄型の作成

coreInfl150、coreNDV100、coreInfl150 を鑄型とし、5' 側の degenerate region が 1~5 塩基の degeneracy をもつ primer でそれぞれ PCR を行うことにより、TAP-PCR 用の鑄型を作成した。最初に作成された degenerate region の degeneracy の限界の検討のために作成された鑄型の両端の配列を図 5 に示す。

また、coreInfl150、および PCR によりその両端に degenerated 配列を接続した d-Infl150 を鑄型とし、Infl150\_degenerated および Infl150\_TAP を primer として用いて PCR を行った (図 6)。Infl150\_TAP は core 鑄型の末端にプライミングする primer であり、その PCR 産物は 166bp となる。また Infl150\_degenerated は core 鑄型の両端に接続された degenerated 配列にプライミングする primer であり、その PCR 産物は 198bp となる。primer が Infl150\_degenerated のとき、鑄型に coreInfl150 を用いた時はバンドが検出されなかったが、d-Infl150 を用いた時は検出された。また、Infl150\_TAP を用いたときは鑄型が coreInfl150、d-Infl150 のどちらでもバンドは検出された。この結果から鑄型 DNA が期待通り作出された事が示唆された。



## 2. degenerate region の degeneracy の限界

3 種類の core 鋳型と、8mer の TAP region の 5' 側に 1~5 塩基の degeneracy をもつ degenerate region を接続させた primer で TAP-PCR を行なうことにより、degenerate region の degeneracy の限界を検討した。degenerate region が 8mer であるため、Inf150 は 82mer、NDV は 132mer、Inf1150 は 182mer となる。この PCR 産物を 2.5%アガロースゲル上で電気泳動したところ、degeneracy が 5mer まで Inf150、NDV、Inf1150 の 3 本のバンドが明瞭であった (図 7)。以降の実験では感度の限界を検討するために鋳型の量を減らしていくため、比較的弱めの degeneracy を用いることとし、TAP primer の degenerate region の degeneracy の限界を  $3^4=81$  と設定した。

本実験での増幅産物のバンドのうち、NDV は degeneracy の量の影響をほとんど受けなかったが、Inf150 と Inf1150 は degeneracy が大きくなっていくにしたがってバンドが薄くなっていった。これは、NDV は TAP 領域の配列が、Inf1 より比較的 PCR がかかりやすい配列であったことが原因であると考えられる。

## 3. TAP-PCR 法の感度の限界

degeneracy が  $3^4$  の鋳型と、8mer から 12mer の TAP region の 5' 側に、degeneracy が  $3^4$  でかつ 8mer の

degenerate region を接続させた primer を用いて限界希釈 PCR を行なうことにより、検出感度の限界の検討を行った。すなわち、1 反応当たり鋳型 DNA が  $10^6 \sim 10^1$  コピーになるように 1/10 ずつ希釈していき、同じ濃度のものを 3 種類混合したものを鋳型として用いた (図 8)。11mer では negative control である DW でもバンドが検出された。これは実験の過程で primer の中に鋳型が混入してしまったことが原因であると考えられる。

## 4. TAPprimer の長さの影響

ウイルスゲノムにおいて、degenerated 配列が長く、その中に degeneracy をもつた塩基が点在している場合を想定し、短い TAP region と長い degenerate region で構成される全長 25mer の primer で同様に鋳型の作成と TAP-PCR を行ない、適切な degenerate region の長さや primer 長を検討した。一方、コロナウイルス科で多種検出 primer を設計したところ、degenerated 配列の degeneracy は比較的低いことが判明した。よってここからは degeneracy を  $3^3=27$  に下げたものを用いた (図 9)。

PCR の結果、TAP region の長さにも関わらず、より少量の鋳型でもバンドが検出された。しかし、degenerate region が 8mer のものは Inf1150 のバンドが全く検出されず、9mer のものは Inf150 と Inf1150 のバンドがほとんど検出されなかった。

## 5. 完全一致の primer での TAP-PCR

前記の 25mer の primer では 3 本のバンドのバランスが非常に悪かった。そこで、同一の primer で鋳型を作成し、さらにその primer で TAP-PCR を行なうことによって、その原因を検討した (図 10)。PCR の結果、 $10^4$  コピーまでバンドが明瞭であった。しかし Inf150 のバンドは薄く、増幅効率が低いことが示唆された。

## 6. SARS ウイルスを用いた実証試験

CoCoMo アルゴリズムで設計した primer を実際のウイルスに用いることで、primer の自動化の有効性を検討した。鋳型としては SARS コロナウイルス感染細胞を用いてそれに RT-PCR を行なったことから、感染細胞のゲノムが混入している状態での TAP-PCR の有効性もまた同時に検討した。PCR の結果、レーン 2 で PCR 産物が得られなかった事以外は全て予想されたサイズの増幅産物が見られた (図 11)。

## 7. HIV-1 検出用 degenerate primer の設計

データベースで公開されている 1083 件の HIV-1 ゲノム配列を収集し、primer 設計が可能なグループ分けをした。その際 Social network analysis に用いている Pajek ソフトの機能上データ数が 200 件以下でなければ実施できないため、実施相溶性 95%以上をグループとしてまとめ、1083 件の配列を 196 の配列グループにま

とめた。続けて、選択された 196 の配列について、相溶性が 80%以上の関係のみを Social network analysis によって描出したところ、概略的に 3 種の二次グループに分かれることが示唆された (図 12)。これらの二次グループについて、グループ内での degenerate primer を設計した (表 15~17)。各グループについて最上位の primer について PCR を実施したところ良好な結果を得た。HCV 検出用 degenerate primer の設計

GenBank で HCV のデータを検索したところ、28,988 件の塩基配列の登録があり、全てをダウンロード、MySQL データベースに格納後、サイズおよび相溶性に基づいてグループに分類したところ、241 のグループに分かれた (表 18)。全ゲノムのサイズと一致するグループには 235 件の塩基配列が分類されたため、このグループについて、相互の相溶性を求めた後、social network 分析に基づいて 9 個のグループに分割した。続けて、それぞれのグループから代表配列選択して代表グループを形成した後、代表グループを網羅する primer (HCV ゲノム共通 degenerate primer) を設計した (表 19)。それらの一部を primer として、標準 HCV ウイルス由来 cDNA を鋳型として PCR を実施したが、確実な増幅は 12 件中の 1 件のみで観察された。

## D. 考察

## 1 CoCoMo アルゴリズムの適用範囲

本研究では、CoCoMo アルゴリズムによって SARS コロナウイルス、HIV-1 および HCV ウイルスを検出するための primer を設計した。いずれの場合についても、基本的には、①対象塩基配列の入手、分類および対象塩基配列グループの選定→②塩基配列間の相同性と social network 解析に基づくグループ化→③対象塩基配列上の 6 塩基 motif の選定→④PCR の位置にある相補的 6 塩基 motif の削除→④選定されたオリゴマーの 5' 領域 15 塩基の選択と共通 degenerated 配列の決定、という処理順序で degenerated primer を抽出することができた。これらのステップのうち、③以降は塩基配列グループによって大きな差はなく、一旦対象グループが決定されれば自動的に進められることが示唆された。ただし、相補オリゴマーの削除ステップには多大な計算資源が必要とされるため、実際には処理が可能な各塩基配列の長さや塩基配列数は限定されている。その限界内にグループを設定するための①での分類グループの選定と②での効率的なグループの設定が設計された primer の適用範囲と検出感度に大きく影響することが示唆された。

## 2. 合成オリゴマーを用いた検証試験からの結論

合成オリゴマーを用いた検証試験では、TAP region の延長にしたがって少量の鋳型でもバンドが検出されるようになり、

感度の上昇が見られた。A、T を 2°C、G、C を 4°C とすると、primer の TAP region のみの T<sub>m</sub> 値は 8mer から順に、26°C、28°C、32°C、34°C、36°C となった。そこから計算すると、degenerated region の T<sub>m</sub> 値は全てほぼ 17°C 前後と想定され、TAP-PCR 用 primer のためのアニーニング温度は、その長さの割には T<sub>m</sub> 値を低く設定すべきであることが示唆された。また、6-11 塩基のモチーフの領域の長さが増幅の感度に影響する傾向が見られた。ただし、TAP region が 12mer の場合は 11mer までに比べ感度が下がったが、この原因は不明である。これまでの結果では、効率の良い増幅のためには TAP region と degenerated region の長さにはバランスが重要であり、degenerated region 比して TAP region 長すぎる場合には必ずしも感度を上げないことが示唆された。

## 3. TAP-PCR に適切な primer 長の検討

図 8 での TAP region が 8mer、9mer での PCR と比較した場合、図 9 で degenerated region を延長すると、感度の上昇が見られ、鋳型が 102 コピーでもバンドが検出された。一方、特異性は下がり、8mer では Inf150 のバンドが見られず、9mer では Inf150 と Inf1150 のバンドがほとんど検出されなかった。また、図 10 で完全一致の鋳型と primer で PCR を行なったところ、さらに感度の上昇は見られ鋳型が 101 コピーでもバンドが検出さ

れた。しかしここでも特異性は低く、Inf150 のバンドはほとんど検出されなかった。

したがって、図9で見られた特異性の低下は degeneracy をもった primer を用いたことが原因ではなく、degenerated region の延長が原因であると考えられると思われる。よって、図8に示されるように、今回の鋳型では 11mer の TAP region と 8mer の degenerated region の組み合わせが適切であると判断された。また図7より、NDV は degeneracy の増加にも関わらずそのバンドの濃さが一定であるが、Inf1 は degeneracy の増加に従ってバンドが薄くなっていくことから、TAP-PCR の効率の差は鋳型の配列により影響されると思われる。よって、適切な primer は各対象ウイルスに対して検討する必要が示唆された。

#### 4. 実際のウイルスへの応用

SARS コロナウイルス感染細胞から鋳型を作成することで、感染細胞のゲノムが混入している状態での自動化して設計された primer の有効性を検討した。PCR の結果、33 サンプル中 1 サンプル以外は全て予想された増幅産物が得られた(図11)。このサンプルに PCR がかからなかった原因としては、おそらく設計と実験に用いた株の差や、または変異があった可能性が考えられる。実験全体としてはほぼ全てのサンプルに PCR がかかったことから、この primer 設計方法は実際のウイルスに

対しても有効であるということが示唆された。

#### 5. HIV-1 および HCV 用 degenerated primer の合成

HIV-1 および HCV に対しての primer 設計では、それまでの事例に比べて遺伝子データベース上に大量のデータが蓄積されているため、基本的には計算機上での primer 設計には有利な状況であると考えられた。ただし、実際の設計段階において、6 塩基 motif を基本とする CoCoMo アルゴリズムでは計算機の機能限界が問題となった。そのため、適切なグループを構成し、ある程度限定されているが一方では良くゲノムグループ全体の配列情報を代表しているグループの作成が degenerate primer 設計上は重要であることが示唆された。HIV-1 においては、塩基配列を限定する過程で、ゲノム全体の相同性比較から HIV-1 を 3 種のグループに分け、それぞれについて degenerate primer を設計した。この設計 primer は

比較的効率良い PCR 結果を出していることから、有効な方法と考えられる。しかしながら、現在 social network 分析に使用した Pajek が図示可能なデータが 200 程度に限定されている。HCV では、図示に合致する数に塩基配列データを限定することが難しかったため、図示に代わる方法として仮想空間上での分類を行うソフト(create\_group\_sns)を利用した。しかしながら、設計された degenerate