

学 位 論 文

哺乳類細胞で有効かつ標的遺伝子に特異的なsiRNA設計法の
確立とその抗ウイルスRNA干渉法への応用

Antiviral RNA interference based on the development of the highly
effective, target-specific siRNA design in mammalian cells

平成18年12月 博士(理学) 申請

東京大学大学院理学系研究科
生物化学専攻

内 藤 雄 樹

Abstract

RNA 干渉 (RNA interference; RNAi) とは, 細胞に導入した 2 本鎖 RNA (double-stranded RNA; dsRNA) が, 相同な塩基配列をもつメッセンジャー RNA (mRNA) を特異的に切断することによって遺伝子発現を抑制する現象である. RNAi の分子機構は, 生体内においてはウイルスやトランスポゾンに対する防御機構の一部と考えられているほか, 生体内で多数発現しているマイクロ RNA (microRNA; miRNA) による遺伝子発現の制御において重要な役割を担っている. また人為的に dsRNA を細胞内に導入することによって, 任意の遺伝子を簡便にノックダウンできることから, RNAi は遺伝子の機能解析に有効な手法として注目されている.

哺乳類細胞で RNAi 実験をおこなうには, 標的遺伝子と相同な, 化学合成した約 21 塩基対の短い 2 本鎖 RNA (short-interfering RNA; siRNA) または siRNA を発現するベクターを細胞に導入する方法が一般的に用いられる. しかし siRNA の配列はどのようなものでも細胞内で機能するわけではなく, 無作為に設計した siRNA の 7~8 割は RNAi 活性を示さない. さらに, siRNA の長さは約 21 塩基対と短いため, 標的遺伝子とはまったく無関係な遺伝子にも相同部分が存在する確率が高く, siRNA がそのような遺伝子を抑制してしまうことがある (オフターゲット効果). 本研究ではまず, 特にこれら 2 つの問題点を考慮した, 哺乳類細胞で有効かつオフターゲット効果の少ない siRNA の設計法を構築した. 哺乳類細胞で有効な siRNA の設計法は, 著者らを含む 3 つの研究グループから提唱されている方法の比較検討をおこない, もっとも良い著者らの方法 (Ui-Tei *et al.* 2004, *Nucleic Acids Res.* **32**, 936-948) を用いた. またオフターゲット効果が少ない siRNA を設計するために, 標的以外のすべての遺伝子に対して相同性が最小となる siRNA 配列を選択するという情報科学的手法を用いた. 21 塩基の相同性検索は, Yamada らのアルゴリズム (Yamada *et al.* 2005, *Bioinformatics* **21**, 1316-1324) を用いた. 本研究で構築した siRNA 設計システムによって, ヒトの全遺伝子をほぼ網羅できるゲノムワイドな siRNA 設計をおこなった.

一方, siRNA の標的は内在性遺伝子に限定されず, 細胞に感染したウイルス由来の RNA を標的とすることによってウイルスの増殖を抑制できることが示されている. 本研究で構築した siRNA 設計法は, 哺乳類細胞に感染するウイルスを標的とする場合にもそのまま適用できると考えられる. しかし HIV-1 や HCV など多くの RNA ウイルスは, 著しくゲノム多様性が高く, 1 種類の siRNA を単独で用いた場合, その siRNA に耐性をもつ変異ウイルスが容易に出現することが報告されている. そこで本研究では, 多様性の高いウイルスを対象として, その高度保存領域を同定し, それらの領域を標的とする siRNA の設計をおこない有効性を評価した.

Abstract

RNA interference (RNAi) is the process of double-stranded RNA (dsRNA) mediated, post transcriptional gene silencing. It is now widely used to knockdown gene expression in a sequence-specific manner, making it a powerful tool not only for studying gene function, but also for therapeutic applications including antiviral treatments.

In mammalian cells, RNAi can be induced using 21 nucleotide-length short-interfering RNA (siRNA) or siRNA expression vectors. However, the efficacy of each siRNA varies widely depending on its sequence; only a limited fraction of randomly designed siRNAs is functional. Moreover, off-target silencing effects arise when the introduced siRNA has sequence similarities with unrelated genes. In the present study, rational designing of highly functional, off-target minimized siRNAs is proposed. Highly functional siRNAs are selected using novel guidelines that were established through an extensive study of the relationship between siRNA sequences and their RNAi efficacy (Ui-Tei *et al.* 2004, *Nucleic Acids Res.* **32**, 936-948). Off-target effects are minimized by an exhaustive scan of the non-redundant transcriptome database using accelerated off-target search algorithm developed by Yamada *et al.* (Yamada *et al.* 2005, *Bioinformatics* **21**, 1316-1324). The present study also publicizes the entire list of highly effective, off-target minimized siRNA sequences for human and mouse genes listed in the RefSeq database.

siRNA mediated RNAi is also applicable for antiviral therapy by targeting viral sequences. However, in case of RNA viruses, designing functional siRNAs that target viral sequences is problematic because of their extraordinarily high genetic diversity. It is also known that siRNA-resistant viral mutants emerge rapidly due to their high mutation rate. In this study, I attempted to design the effective antiviral siRNAs that target HIV-1. Highly conserved regions within HIV-1 genomes were identified through an exhaustive computational analysis, and the functionality of siRNAs targeting the highest possible conserved regions was validated. Several promising candidates of antiviral siRNAs are presented, which effectively inhibited multiple subtypes of HIV-1 by targeting 5' UTR, protease, and integrase regions. The present study also publicizes the web-based online software system named siVirus (<http://siVirus.RNAi.jp/>), which provides functional, off-target minimized siRNAs that target highly conserved regions of divergent viral sequences including HIV, HCV, influenza virus, and SARS coronavirus, all of which pose enormous threats to global human health.

目 次

| | |
|---|----|
| 緒 論 | 1 |
| 第一章 | 9 |
| 1. 結果と考察 | 11 |
| 1-1. 哺乳類細胞で活性の高い siRNA 設計法の検討 | 11 |
| 1-2. ゲノムワイドな siRNA 設計への適用可能性 | 15 |
| 1-3. siRNA の標的切断活性を評価するアッセイ系の構築 | 16 |
| 1-4. オフターゲット効果が少ない siRNA 設計法の実験的検討 (1) | 21 |
| 1-5. オフターゲット効果が少ない siRNA 設計法の実験的検討 (2) | 24 |
| 1-6. siRNA 設計システムの構築 | 26 |
| 1-7. ヒト全遺伝子に対する siRNA 設計 | 28 |
| 1-8. 長い 2 本鎖 RNA から生成される siRNA のオフターゲット効果 | 30 |
| 2. 方 法 | 33 |
| 2-1. Ui-Tei らによる siRNA 設計法 | 33 |
| 2-2. Reynolds らによる siRNA 設計法 | 33 |
| 2-3. Amarzguioui らによる siRNA 設計法 | 33 |
| 2-4. siRNA | 33 |
| 2-5. 細胞の培養およびトランスフェクション | 34 |
| 2-6. ルシフェラーゼ遺伝子を標的とする RNAi | 34 |
| 2-7. 標的 mRNA 発現ベクター pTREC の構築 | 34 |
| 2-8. 標的切断アッセイ | 34 |
| 2-9. ビメンチン遺伝子を標的とする RNAi | 35 |
| 2-10. プライマーエクステンション法 | 35 |
| 2-11. seed マッチによるオフターゲット効果の検討 | 36 |
| 2-12. 共同研究に関する事項 | 36 |
| 第二章 | 37 |
| 1. 結果と考察 | 39 |
| 1-1. HIV-1 を標的とする siRNA 設計法の検討 | 39 |
| 1-2. siRNA の抗 HIV-1 効果の検証 | 42 |
| 1-3. 抗ウイルス siRNA 設計システムの構築 | 49 |

| | |
|-------------------------------|----|
| 2. 方 法 | 53 |
| 2-1. siRNA | 53 |
| 2-2. 細胞の培養およびトランスフェクション | 53 |
| 2-3. 標的切断アッセイ | 53 |
| 2-4. HIV-1 逆転写酵素活性の測定 | 53 |
| 2-5. siVirus ウェブサーバの構築 | 53 |
| 2-6. 共同研究に関する事項 | 53 |
| 文 献 | 55 |
| 巻末表 | 63 |
| Table S1 | 65 |
| Table S2 | 67 |
| Table S3 | 77 |
| Table S4 | 80 |
| 謝 辞 | 85 |
| | |
| (附 録) | |
| 配列表 | 1 |

緒論

緒 論

RNA 干渉 (RNA interference; RNAi) とは, 細胞に導入された 2 本鎖 RNA (double-stranded RNA; dsRNA) が, 相同な塩基配列をもつメッセンジャー RNA (mRNA) を特異的に切断することによって遺伝子発現を抑制する現象である. RNAi の分子機構は, 生体内においてはウイルスやトランスポゾンに対する防御機構の一部と考えられているほか, 生体内で多数発現しているマイクロ RNA (microRNA; miRNA) による遺伝子発現の制御において重要な役割を担っている. また人為的に dsRNA を細胞内に導入または発現させることによって, ヒト, マウス, ショウジョウバエ, 線虫などを含む多くの真核生物において簡便に任意の遺伝子をノックダウンできることから, RNAi は遺伝子の機能解析に有効な手法として大きな注目を集めている.

RNAi は, 1998 年に Fire らによる線虫を用いた実験ではじめて報告された (Fire *et al.*, 1998). Fire らは, dsRNA を線虫にインジェクションにより導入することで, dsRNA と相同な配列をもつ遺伝子が効率よく抑制されることを示した. 一方, その dsRNA を構成するセンス 1 本鎖 RNA またはアンチセンス 1 本鎖 RNA を単独で線虫にインジェクションしても, 標的遺伝子を抑制する効果は見られなかった. 同様の現象は線虫のみならず, ショウジョウバエ (Kennerdell *et al.*, 1998), ゼブラフィッシュ (Wargelius *et al.*, 1999), 哺乳類細胞 (Ui-Tei *et al.*, 2000) などにも見いだされ, RNAi の基本的な機構が多くの生物種に保存されていることが明らかになった.

細胞内に導入された dsRNA は, Dicer によって 21 ~ 23 塩基対の短い 2 本鎖 RNA (short-interfering RNA; siRNA) に分解される (Bernstein *et al.*, 2001; Ketting *et al.*, 2001). このとき siRNA の 5' 末端はリン酸基を有し, 3' 末端は 2 塩基の突出末端となっている (Fig. 1). siRNA は数種類のタンパク質と複合体 (RNA-induced silencing complex; RISC) を形成し, この複合体が標的 mRNA を認識し切断する. RISC を構成する主要なタンパクとして, Argonaute (Hammond *et al.*, 2001; Martinez

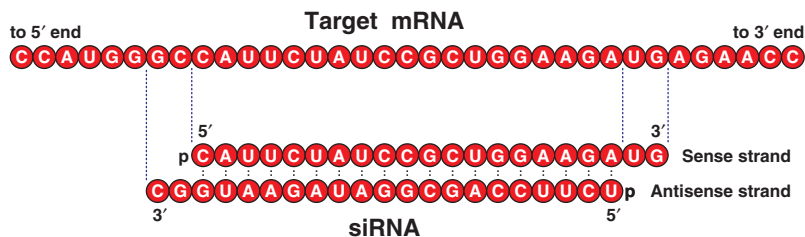


Figure 1. siRNA は 21 塩基対前後の短い 2 本鎖 RNA である. siRNA は 5' 末端にリン酸基を有し, 3' 末端は 2 塩基の突出末端となっている.

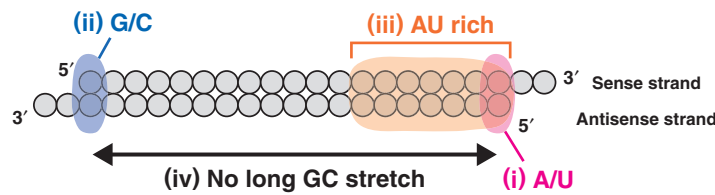


Figure 2. Ui-Tei らによる、哺乳類細胞で高い RNAi 活性を示す siRNA 配列の規則性.

et al., 2002; Doi *et al.*, 2003), Dicer, TRBP (Chendrimada *et al.*, 2005), PACT (Lee *et al.*, 2006) などが報告されている.

RNAi が発見された当初、哺乳類細胞に 30 塩基対以上の長い dsRNA を導入すると、遺伝子発現を非特異的に抑制するインターフェロン応答が誘導されるため (Grant *et al.*, 1995; Kaufman *et al.*, 1999), 哺乳類細胞で RNAi をおこなうことは難しいと考えられていた. しかし 2001 年に Elbashir らによって、約 21 塩基対の siRNA を用いればインターフェロン応答を引き起こすことなく RNAi を誘導できることが示されたため、ヒトやマウスを含む哺乳類細胞に対しても RNAi の利用が可能となった (Elbashir *et al.*, 2001a).

しかしながら、哺乳類細胞に対して siRNA を用いる場合、siRNA の配列によって RNAi 活性に大きな差異が見られることがわかった (Holen *et al.*, 2002). ホタルルシフェラーゼ遺伝子を標的とした実験では、無作為に設計した siRNA のうちの 2 ~ 3 割程度しか十分な RNAi 活性を示さず、それ以外の多くの siRNA は機能しなかった (Ui-Tei *et al.*, 2004). そこで Ui-Tei および著者らは、多数の siRNA を用いた系統的な解析をおこない、RNAi を効率よく誘導できる siRNA が示す配列上の規則性を検討した. その結果、(i) siRNA を構成するアンチセンス鎖の 5' 末端が A または U であり、(ii) センス鎖の 5' 末端が G または C であり、(iii) アンチセンス鎖の 5' 端の 7 塩基のうち 4 塩基以上が A または U であり、(iv) 配列中に G/C が 10 塩基以上連続しないような siRNA (Fig. 2 に i ~ iv を図示) は、Figure 3 に示すように哺乳類細胞において高い確率で標的遺伝子を抑制できた. 一方、(i) ~ (iii) の条件をすべて満たさない siRNA の活性はどれも非常に低かった. この配列規則性は、ヒトおよびマウスの内在性遺伝子を標的とした場合でも、またニワトリ胚における個体レベルの実験でも成り立っていた. さらに、siRNA を細胞に導入する場合だけでなく、DNA ベクターから siRNA を発現させた場合でもこの規則性が同様に成り立っていた. この規則性に基づいて siRNA を設計することにより、ノックダウンの成功率を飛躍的に高めることが可能となった.

一方、siRNA の長さは約 21 塩基対と短いため、標的遺伝子とはまったく無関係な遺伝子にも相同部分が存在する確率が高い. そのような場合、siRNA によって標的とは無関係な遺伝子の発

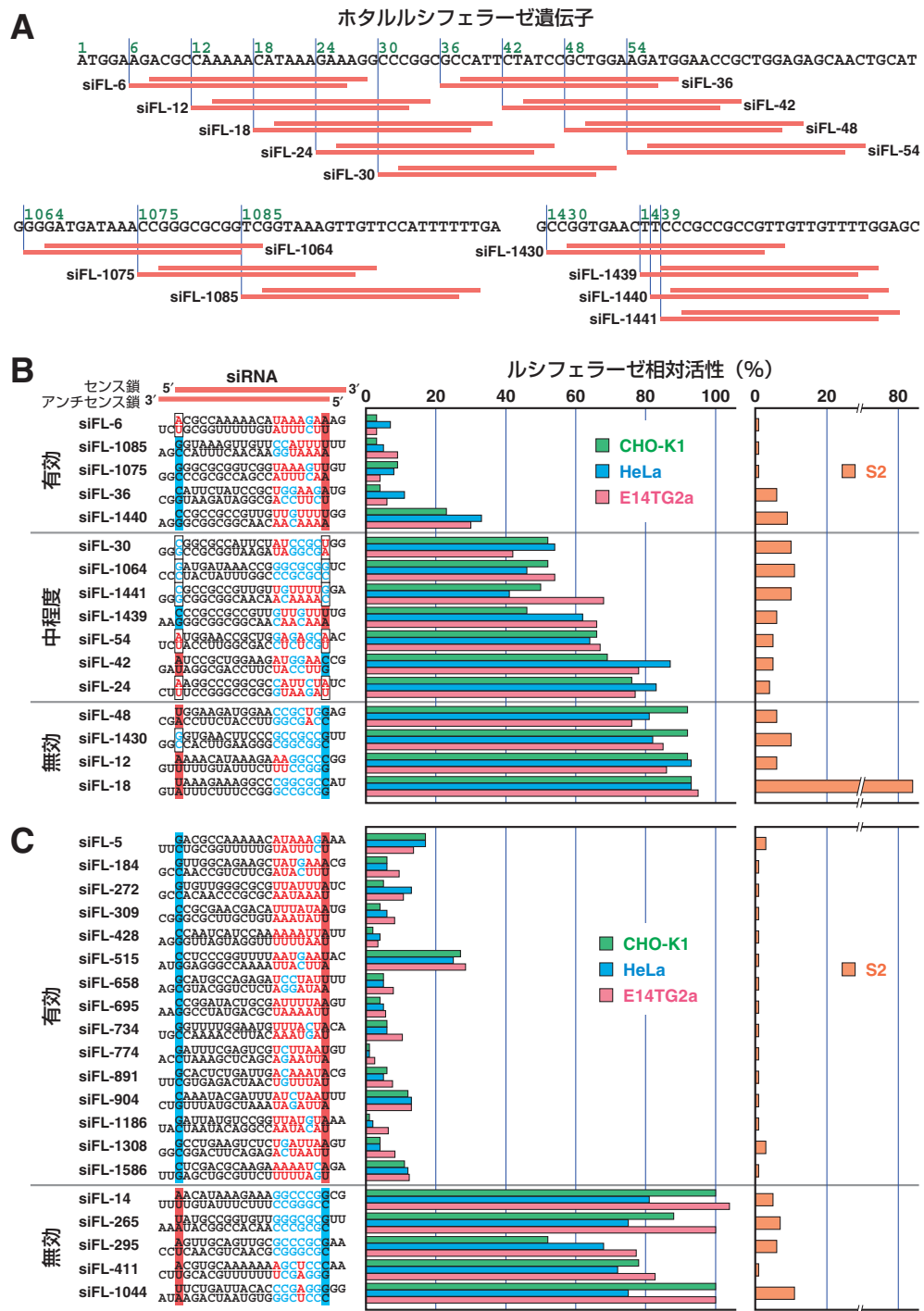


Figure 3. ホタルルシフェラーゼ遺伝子に対して設計した siRNA とその RNAi 効果 (Ui-Tei *et al.* 2004 より作成). 哺乳類細胞であるチャイニーズハムスターの CHO-K1, ヒトの HeLa, マウス ES 細胞の E14TG2a における結果および, ショウジョウバエの S2 細胞における結果を示した. (A) ホタルルシフェラーゼ遺伝子に対して無作為に設計した 16 種類の siRNA. (B) 16 種類の siRNA による RNAi 効果を, Dual-luciferase reporter assay system (Promega 社) で測定し, RNAi 効果の高い順番に並べた. (C) 有効と予測した siRNA (15 種類) と, 無効と予測した siRNA (5 種類) の RNAi 効果. siRNA の番号は, ホタルルシフェラーゼ遺伝子のコード領域における位置に対応している.

現まで抑制される可能性がある。これは siRNA のオフターゲット効果と呼ばれ、siRNA を用いた遺伝子機能解析や、siRNA を医薬に応用するうえでの深刻な課題のひとつと考えられている (Jackson *et al.*, 2004)。オフターゲット効果を回避するためには、標的以外のすべての遺伝子に対して相同性が最小となる siRNA 配列を選択することが望ましい。一般的には、設計した siRNA 配列を BLAST (Altschul *et al.*, 1990) で検索し、無関係な遺伝子がヒットしないものを選択すればよいとされている。しかし BLAST による検索は、siRNA のような短い配列の検索において見落としが多いという欠点があることが判った。たとえば、**Figure 4A** に示すアライメントは 19 塩基中 16 塩基が一致しているが、連続した 7 塩基以上の一致が存在しないため BLAST では相同性を検出できない。また **Figure 4B** のアライメントは 19 塩基中 18 塩基が一致しているが、BLAST の標準パラメータは 1 塩基の一致を +1 点、1 塩基の不一致を -3 点として相同性を評価するため、3' 端の 3 塩基の一致 (CTG) を見落とし、15 塩基中 15 塩基の一致という不完全な結果を返す。このように BLAST は短い配列の相同性検索に不向きであるため、BLAST で siRNA の特異性を確認する従来の方法ではオフターゲット効果のリスクが過小に評価される (Naito, Yamada *et al.*, 2004)。この問題に対し Yamada らは、siRNA のような短い配列の相同性検索を見落としなく高速に実行するアルゴリズムを開発した (Yamada *et al.*, 2005)。Yamada らの方法は、与えられた 19 塩基の配列に対し、ヒト全遺伝子のなかから 3 ミスマッチ以内で相同なすべての配列を、数ミリ秒で見落としなく検索可能である。

本研究では、このような siRNA 設計において考慮すべき 2 つの項目、すなわち、(1) 哺乳類細胞に対して有効な siRNA を選択し、かつ、(2) オフターゲット効果が低く特異性の高い siRNA を選択することが可能な siRNA 設計法の構築をおこなった。第 1 章では、この siRNA 設計システムの構築について述べ、さらにこのシステムを用いたヒト全遺伝子に対するゲノムワイドな siRNA 設計について論じる。

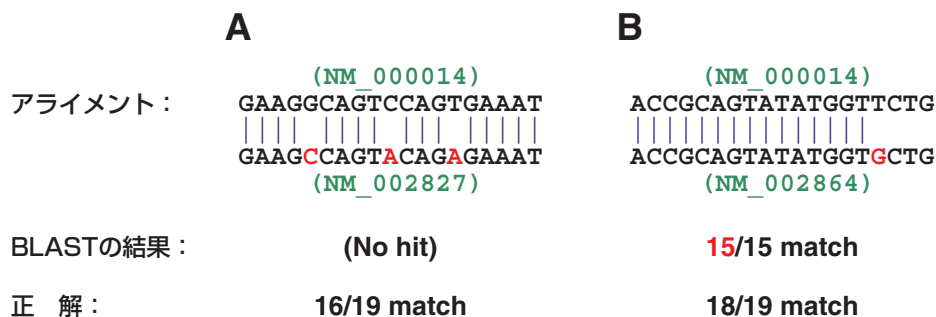


Figure 4. BLAST は短い配列の相同性を見落とす (Naito, Yamada *et al.* 2004 より作成)。

(A) 7 塩基以上の連続一致が存在しない場合は BLAST で相同性を検出できない。(B) 配列の末端付近にミスマッチがある場合、BLAST は最良のアライメントを見落とす。

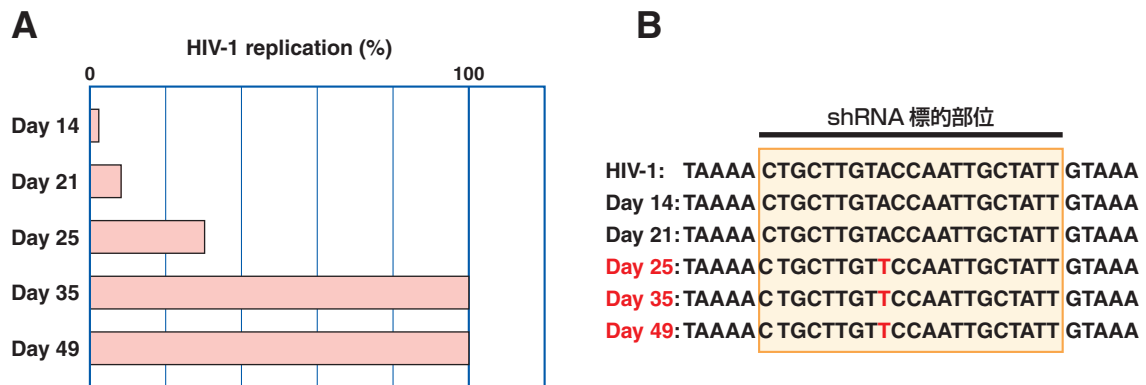


Figure 5. Boden *et al.*, 2003a より作成. (A) *tat* 遺伝子を標的とする shRNA を恒常的に発現する H9 細胞における, HIV-1 の増殖抑制効果. HIV-1 の増殖は p24 抗原を ELISA で定量することによって評価した. (B) 培地上清中の HIV-1 をシーケンスした結果. 25 日後より shRNA の標的部位に 1 塩基の置換を含む耐性ウイルスがあらわれた.

また, siRNA の標的は内在性遺伝子に限定されず, 細胞に感染したウイルス由来の RNA を標的とすることによってウイルスの増殖を抑制できることも示されている (Leonard *et al.*, 2006). 第 1 章に述べる siRNA 設計法は, 哺乳類細胞に感染するウイルスを標的とする場合にもそのまま適用できると考えられる. しかしながら, HIV-1 や HCV など多くの RNA ウイルスでは, 著しく高いゲノム多様性のために siRNA の設計が難しいばかりでなく, 1 種類の siRNA を単独で用いた場合, その siRNA に耐性をもつウイルスが容易に出現しうることが報告されている (Boden *et al.*, 2003a; Das *et al.*, 2004). Boden らの報告によれば, HIV-1 の *tat* 遺伝子を標的とする siRNA と同様の効果をもつヘアピン型 RNA (short hairpin RNA; shRNA) を恒常的に発現する細胞に HIV-1 を感染させたところ, 感染の初期においては HIV-1 の増殖を効率よく抑制できた. しかし感染の 25 日後には, shRNA の標的部位に 1 塩基の置換を含む shRNA 耐性ウイルスがあらわれた (Fig. 5). このような課題を克服するひとつの方法は, ウイルスの配列のなかで高度に保存された領域を標的とすることであると考えられる.

第 2 章では, ウイルスの配列のなかで高度に保存された領域を標的に, 有効な siRNA を設計する方法について論じる. 本研究ではまず, データベースより入手可能な約 500 種の HIV-1 ゲノム配列を詳細に解析し, siRNA の標的として好適な, 21 塩基の長さで保存度が高い領域を同定した. ついで, これらの領域を標的とする siRNA を, 第 1 章で述べる siRNA 設計法に基づき設計した. さらに, 設計した siRNA の HIV-1 増殖抑制効果を, 複数のサブタイプに属する 4 種の HIV-1 感染性分子クローンで検証した. その結果, 多くの siRNA が 4 種すべての HIV-1 の増殖を効果的に抑制でき, 本研究で構築した抗ウイルス siRNA 設計法の有効性を確認できた. 第 2 章の最後では, 本結果に基づく抗ウイルス siRNA 設計システム siVirus の構築およびウェブサーバの公開についても述べる.

なお、本研究の一部は既に下記に報告した。

- Ui-Tei, K., Naito, Y. & Saigo, K.** Guidelines for the selection of effective short-interfering RNA sequences for functional genomics. *Methods Mol. Biol.* **361**, 201-216 (2006).
- Ui-Tei, K., Naito, Y. & Saigo, K.** Essential notes regarding the design of functional siRNAs for efficient mammalian RNAi. *J. Biomed. Biotechnol.* **2006**, 65052 (2006).
- Naito, Y., Ui-Tei, K., Nishikawa, T., Takebe, Y. & Saigo, K.** siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. *Nucleic Acids Res.* **34**, W448-W450 (2006).
- Naito, Y., Yamada, T., Matsumiya, T., Ui-Tei, K., Saigo, K. & Morishita, S.** dsCheck: highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. *Nucleic Acids Res.* **33**, W589-W591 (2005).
- Naito, Y., Yamada, T., Ui-Tei, K., Morishita, S. & Saigo, K.** siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.* **32**, W124-W129 (2004).
- Ui-Tei, K., Ueda, R., Zenno, S., Takahashi, F., Doi, N., Naito, Y., Yamamoto, M., Hashimoto, N., Takahashi, K., Hamada, T., Tokunaga, T. & Saigo, K.** RNA interference induced by transient or stable expression of hairpin structures of double-stranded RNA in *Drosophila* and mammalian cells. *Mol. Biol.* **38**, 276-287 (2004).
- Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R. & Saigo, K.** Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.* **32**, 936-948 (2004).

第一章

第一章

哺乳類細胞で有効かつ標的遺伝子に特異的な siRNA 設計法の構築

1. 結果と考察

1-1. 哺乳類細胞で活性の高い siRNA 設計法の検討

哺乳類細胞では、無作為に設計した siRNA のうち 2 ～ 3 割程度しか十分な RNAi 活性を示さず、それ以外の多くの siRNA は機能しないことが知られている (Holen *et al.*, 2002). この問題に対して Ui-Tei および著者らは、哺乳類細胞で高い RNAi 活性を示す siRNA の配列規則性を見いだした (Ui-Tei *et al.*, 2004). また、他の研究グループからも、哺乳類細胞で有効な siRNA の設計法について、実験結果に基づく具体的な方法が提唱されている (Reynolds *et al.*, 2004; Amarzguioui *et al.*, 2004). 本研究ではまず、Ui-Tei ら、Reynolds ら、Amarzguioui らの 3 つの siRNA 設計法の比較評価をおこなった。

Ui-Tei ら、Reynolds ら、Amarzguioui らの各 siRNA 設計法は、いずれも siRNA の配列自体からその RNAi 活性を予測し、有効な siRNA を選択しようとするものである (Fig. 1-1). まず、siRNA を無作為に設計した場合に、それぞれの設計法を満たす siRNA がどのような割合で設計できる

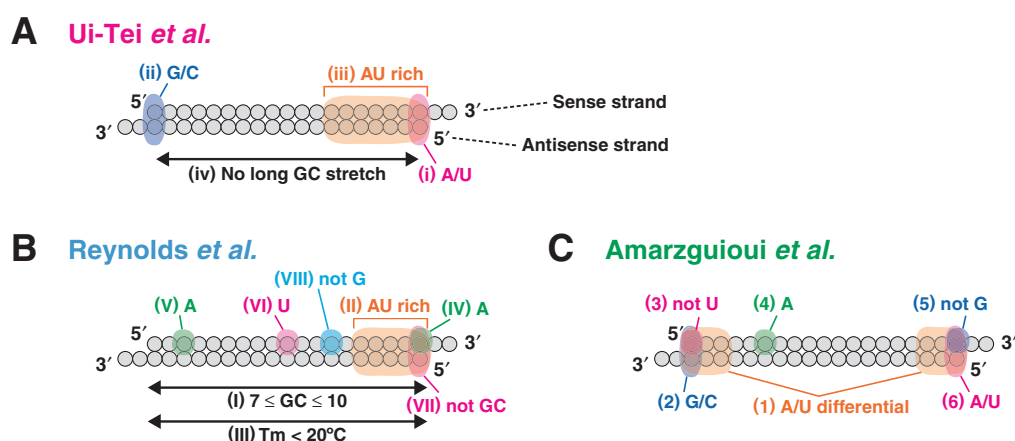


Figure 1-1. 哺乳類細胞で有効な siRNA の設計法. (A) Ui-Tei らの提唱する siRNA 設計ガイドライン (Ui-Tei *et al.*, 2004). (i) ～ (iv) の条件をすべて満たす siRNA を有効と予測している. (B) Reynolds らの siRNA 設計法 (Reynolds *et al.*, 2004). (I) ～ (VIII) の条件をできるだけ満たす siRNA を有効と予測している. (C) Amarzguioui らの siRNA 設計法 (Amarzguioui *et al.*, 2004). (1) ～ (6) の条件をできるだけ満たす siRNA を有効と予測している.

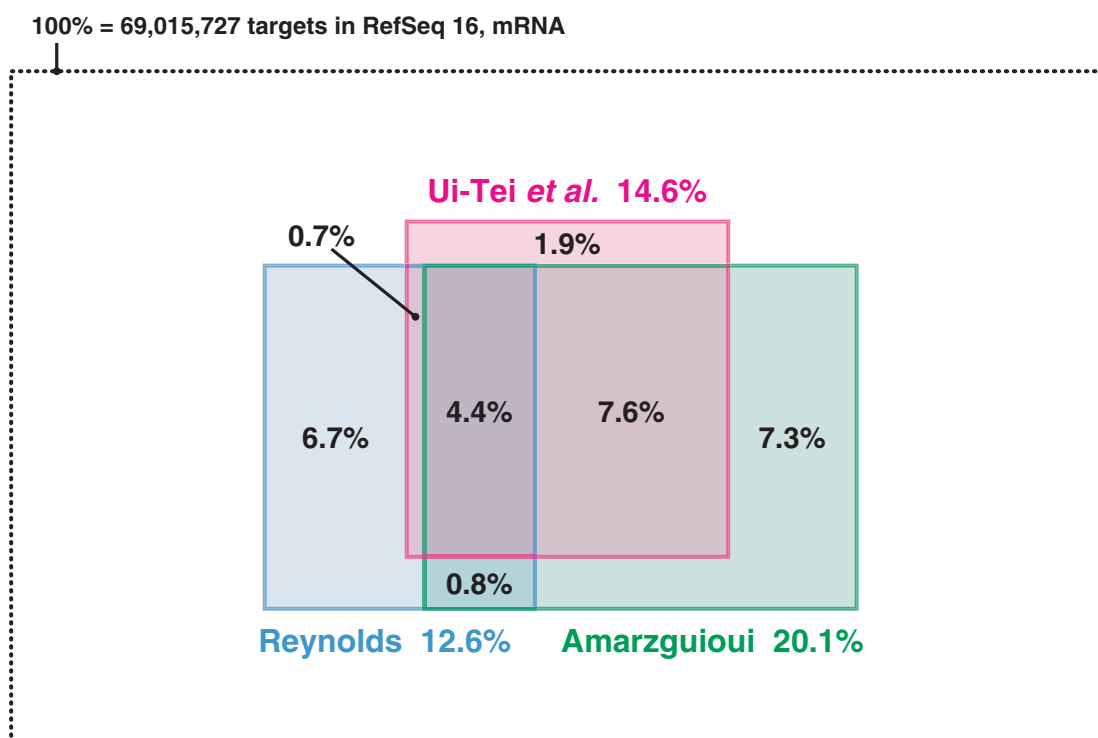


Figure 1-2. ヒトの全遺伝子に対して設計可能な siRNA の全体を 100% として, Ui-Tei ら, Reynolds ら, Amarzguioui らの設計法を満たす siRNA の割合をそれぞれ示した.

かを検討した. RefSeq (Pruitt *et al.*, 2007) release 16 に登録されたヒトの全遺伝子に対して設計可能な 69,015,727 種の siRNA のうち, Ui-Tei ら, Reynolds ら, Amarzguioui らの設計法を満たす siRNA は, それぞれ 14.6%, 12.6%, 20.1% であった (Fig. 1-2). 各 siRNA 設計法は, いくつかの相違点があるものの, siRNA を構成するアンチセンス鎖の 5' 末端が A または U であることが好ましく, アンチセンス鎖の 5' 末端からある程度の長さ (3 ~ 7 塩基) に A または U が多いほうが有効である点で一致している. このため各設計法を満たす siRNA の集合は互いに重なり合い, 3 種の設計法をすべて満たす siRNA も全体の 4.4% 存在した. また特に, Ui-Tei らの設計法および Amarzguioui らの設計法は, センス鎖の 5' 末端が G または C であることが好ましいという点も共通であるため, 互いによく重なり合った.

各 siRNA 設計法の有効性を検証するため, ホタルルシフェラーゼ遺伝子を標的とする 114 種の siRNA (Table S1, 巻末) の RNAi 活性を, ルシフェラーゼアッセイにより評価した. 114 種の siRNA の内訳は下記のとおりである. なお, 詳細は Table S1 (巻末) に示した.

- ・ 3 種の設計法をすべて満たさない siRNA 43 種
- ・ 3 種の設計法のうち 1 種のみを満たす siRNA 30 種
- ・ 3 種の設計法のうち 2 種を満たす siRNA 26 種
- ・ 3 種の設計法をすべて満たす siRNA 15 種

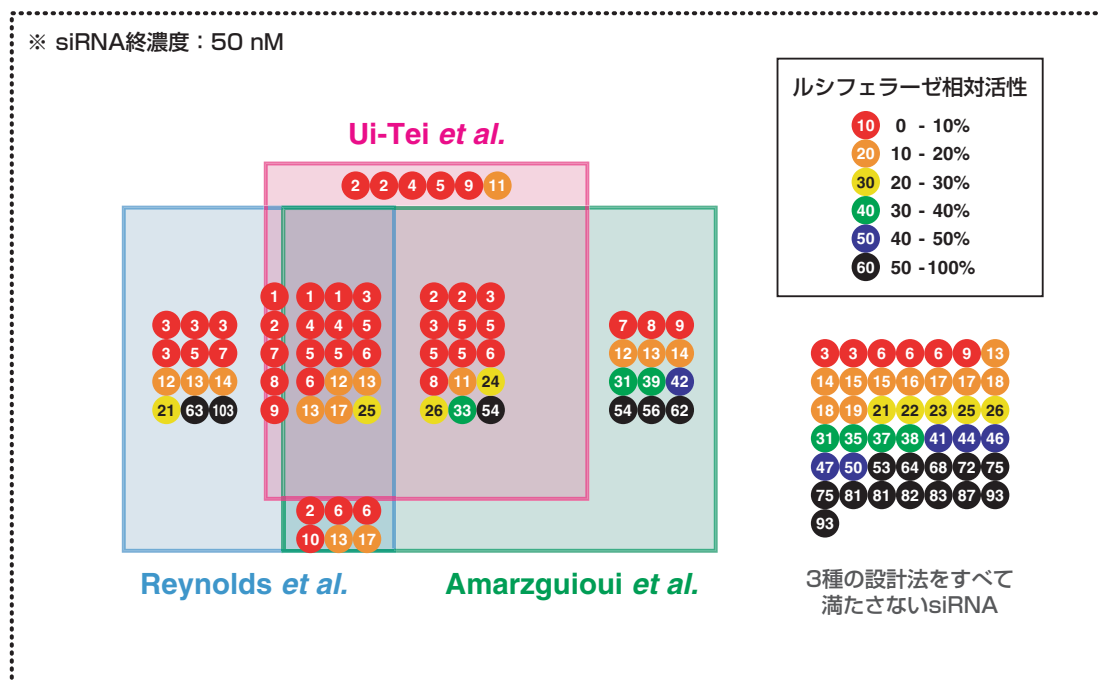


Figure 1-3. ホタルルシフェラーゼ遺伝子を標的とする 114 種の siRNA を、Ui-Tei ら、Reynolds ら、Amarzguioui らの各設計法への適合性に基づき分類し、RNAi 活性とともに示した。ルシフェラーゼ相対活性の値が低いほど、siRNA の RNAi 活性が高いことを示している。

これらの siRNA の RNAi 活性を、ルシフェラーゼアッセイで評価した結果を **Table S1** (巻末) および **Figure 1-3** に示した。Ui-Tei ら、Reynolds ら、Amarzguioui らの各設計法を満たす siRNA のうち、ルシフェラーゼ相対活性を 50% 以下に抑制した siRNA は、それぞれ

- ・ Ui-Tei らの設計法 40 / 41 (97.6%)
- ・ Reynolds らの設計法 36 / 38 (94.7%)
- ・ Amarzguioui らの設計法 44 / 48 (91.7%)

であった。また、ルシフェラーゼ相対活性を 10% 以下に抑制した、高い RNAi 活性を有する siRNA は、

- ・ Ui-Tei らの設計法 30 / 41 (73.2%)
- ・ Reynolds らの設計法 25 / 38 (65.8%)
- ・ Amarzguioui らの設計法 27 / 48 (56.3%)

であった。3 種の設計法とも哺乳類細胞で有効な siRNA を効率よく選択できたが、特に Ui-Tei らの設計法が良好な成績であった。また、Reynolds らの設計法のみを満たす 12 種の siRNA と、Amarzguioui らの設計法のみを満たす 12 種の siRNA とに注目すると、RNAi 活性が低い siRNA が相対的に多いことが判った (**Fig. 1-3**)。一方、Ui-Tei らの設計法のみを満たす 6 種の siRNA は、いずれも有効であった (**Fig. 1-3**)。Ui-Tei らの設計法は、他の設計法よりも判定項目が少

なく単純であるが、すべての判定項目を満たす siRNA のみを有効としている点で Reynolds らや Amarzguioui らの設計法と異なっており、このため有効な siRNA を効率よく選択できたのかも知れない。本研究では以上の結果を考慮し、Ui-Tei らの設計法のみを採用して siRNA 設計をおこなうこととした。

なお、Ui-Tei らの設計法によって有効な siRNA を選択できる理由は以下のように考えられる。RNAi の実行過程において、siRNA は複数のタンパク質と複合体 (RISC) を形成するが、このとき RISC には siRNA を構成する 2 本の RNA 鎖のうち 1 本のみが取り込まれる (Martinez *et al.*, 2002)。siRNA のアンチセンス鎖側が RISC に取り込まれた場合、RISC は標的 mRNA を認識して切断できるが、センス鎖側が RISC に取り込まれた場合は標的 mRNA を認識できず、mRNA は切断されない (Fig. 1-4)。Schwarz らは、ショウジョウバエの *in vitro* の実験系を用いて、siRNA のどちらの RNA 鎖が RISC に取り込まれるかを、siRNA の両末端部分の塩基対形成の強さに注目して解析した。その結果、siRNA のアンチセンス鎖の 5' 末端側が、センス鎖の 5' 末端側よりも塩基対形成が弱い場合に、siRNA のアンチセンス鎖がおもに RISC に取り込まれていることが判明した (Schwarz *et al.*, 2003)。また Khvorova らは、ヒト培養細胞において RNAi 活性が高い siRNA は、アンチセンス鎖の 5' 末端側の塩基対が熱力学的に安定性が低くなっており、このことが RISC にどちらの RNA 鎖が取り込まれるかを規定すると報告している (Khvorova *et al.*,

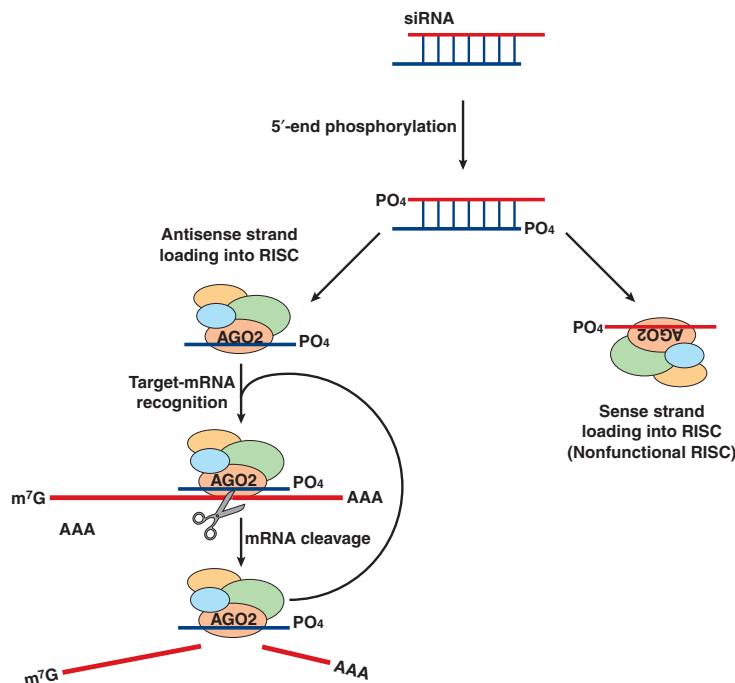


Figure 1-4. RISC には siRNA を構成する 2 本の RNA 鎖のうち 1 本のみが取り込まれる。siRNA のアンチセンス鎖が RISC に取り込まれた場合は標的 mRNA を認識して切断できるが、センス鎖が取り込まれた場合は標的 mRNA を認識できない。

2003). これらの報告は, Ui-Tei らの設計法の, アンチセンス鎖の 5' 末端およびその周辺に A または U が多く, センス鎖の 5' 末端が G または C であるような siRNA が有効であるという条件と対応している. ただし, siFL-1063 (Table S1; 巻末) のように, siRNA のアンチセンス鎖の 5' 末端から 9 塩基が連続して G または C であっても活性の高い siRNA が存在することから, 末端部分の塩基対の安定性以外にも, RISC 形成における RNA 鎖の選択に関わる要因が存在すると考えられる.

1-2. ゲノムワイドな siRNA 設計への適用可能性

Ui-Tei らの設計法では, ヒトの全遺伝子に対して設計可能なすべての siRNA のうち 14.6% が有効と判定された (Fig. 1-2). ただし, この割合は遺伝子の種類によって変動し, Figure 1-5 の赤色のヒストグラムに示すような広い分布を示した. たとえば, *src* 遺伝子 (SRC; NM_005417) では 9.3%, *EGF* 遺伝子 (EGF; NM_001963) では 17.9% の siRNA が有効と判定された (Fig. 1-5). また, RefSeq release 16 に登録されたヒトの全遺伝子に対して siRNA の設計を試みたところ, すべての遺伝子に対して, すくなくとも 1 個の有効な siRNA が設計可能であった.

さらに, Pol III プロモーターをもつベクターから shRNA を発現させる場合,

- ・センス鎖の 5' 末端が G であること. (この G から転写が開始されるため.)
- ・配列中に UUUU もしくはその相補的配列である AAAA を含まないこと. (途中で転写が終結することを防ぐため.)

が必要である (Ui-Tei *et al.*, 2006). この条件を加えると, Ui-Tei らの設計法で有効と判定される siRNA は, ヒトの全遺伝子に対して設計可能なすべての siRNA のうちの 5.7% となった. このとき,

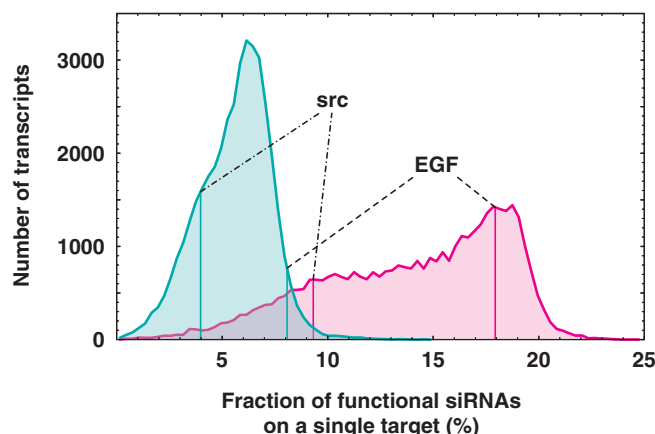


Figure 1-5. ヒトの全遺伝子について, Ui-Tei らの設計法を満たす siRNA の割合を求め, その分布を示した (赤色のヒストグラム). また, Ui-Tei らの設計法を満たす siRNA のうち, Pol III プロモーターから発現させる shRNA としても使用可能な配列の割合を求め, その分布を示した (青色のヒストグラム). ヒストグラムの階級の幅は 0.3% である. *src* 遺伝子および *EGF* 遺伝子における割合を例示した.

個々の遺伝子における割合は Figure 1-5 の青色のヒストグラムのような分布となった。RefSeq release 16 に登録されたヒトの全遺伝子に対して siRNA の設計を試みたところ、99.9% 以上の遺伝子に対して、すくなくとも 1 個の有効な siRNA が設計可能であった。

以上の結果は、Ui-Tei らの siRNA 設計法が、ヒトの全遺伝子をほぼ網羅できるゲノムワイドな siRNA 設計に適用可能であることを示唆した。

1-3. siRNA の標的切断活性を評価するアッセイ系の構築

siRNA が本来の標的とは別の遺伝子の相同部分を認識し、その遺伝子を抑制してしまう現象は、siRNA のオフターゲット効果と呼ばれる (Jackson *et al.*, 2004)。本研究では、siRNA 配列と mRNA 配列との間にどの程度の相同性があるとオフターゲット効果が生じるかを検討するため、任意の標的配列に対して siRNA の標的切断活性を評価できるアッセイ系（以下、**標的切断アッセイ**）を構築した。

標的切断アッセイの概略は、siRNA の標的となる 21 ～ 40 塩基の配列を組み込んだ発現ベクターと siRNA とを同時に培養細胞に導入して、ベクターから発現する標的 mRNA が siRNA によって切断される効率を、リアルタイム RT-PCR で定量するというものである。本アッセイのために構築した標的発現ベクター pTREC (Target RNA Expression Construct の略) の概要を Figure 1-6 に

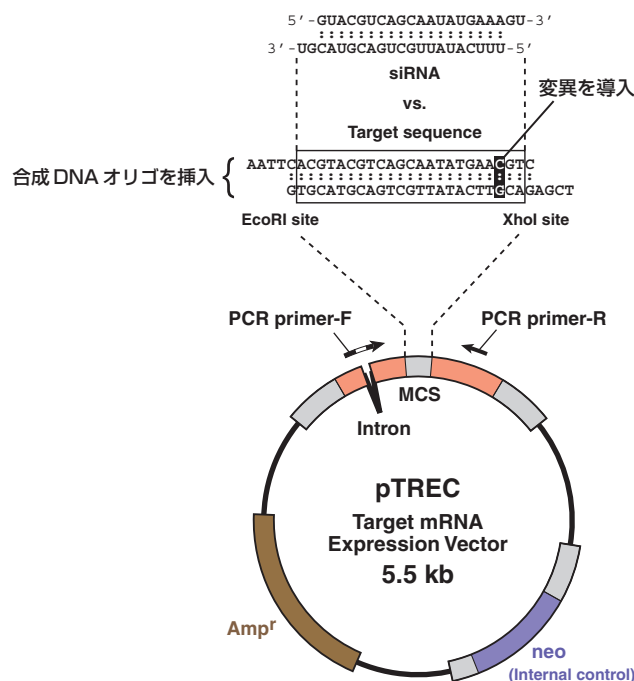


Figure 1-6. 本研究で構築した標的 mRNA 発現ベクター pTREC の構造。RT-PCR に用いる PCR プライマーの一方 (PCR primer-F) の配列は、イントロンによって分断されるように設計した。この工夫により、PCR の際にプラスミドが鋳型となって DNA が増幅することを防ぐことができた。

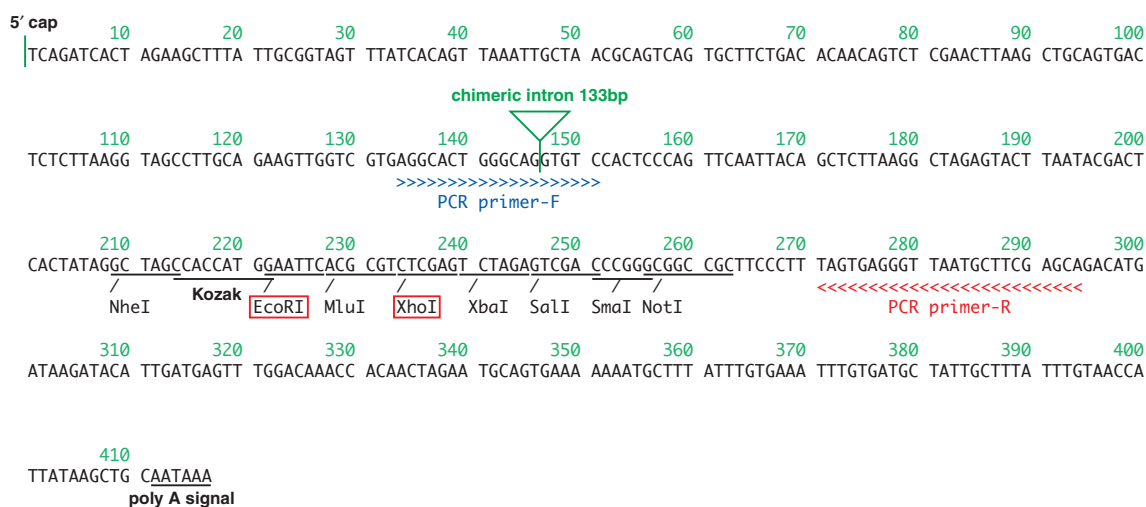


Figure 1-7. pTREC から標的 mRNA として転写される領域の塩基配列. 本研究では, EcoRI 部位と XhoI 部位との間に合成 DNA オリゴで作製した標的配列を挿入した.

示す. 本アッセイでは, 合成 DNA オリゴを pTREC ベクターの MCS に挿入することによって標的配列を自由に設定できるため, ある特定の siRNA に対して標的 mRNA の配列を 1 塩基ずつ変えていくような系統的な解析が可能となる. たとえば **Figure 1-6** の例では, siRNA の標的部分に 1 塩基の置換を導入した場合の切断活性を評価できる. また, pTREC ベクターから発現する mRNA の配列 (**Fig. 1-7**) は標的部分を除き共通であるため, まったく同一の実験条件下で siRNA の標的切断活性を定量的に評価できる. また RT-PCR をおこなう際にも共通の PCR プライマーを使用するため, PCR の条件検討が不要でハイスループットなアッセイが可能である.

まず本アッセイの有効性を検証するために, 本アッセイの結果と, 実際の標的遺伝子に対して siRNA を用いた場合の RNAi 活性とを比較した.

ホタルルシフェラーゼ遺伝子を標的とする 9 種の siRNA について本アッセイにより標的切断活性を評価し, 一方でホタルルシフェラーゼ遺伝子に対する RNAi 効果をルシフェラーゼアッセイで測定した. 本アッセイで得られた結果は, 実際にルシフェラーゼ遺伝子に対して各 siRNA が示した RNAi 活性と高い相関を示した (**Fig. 1-8**). 両者の関係が完全に線形になっていない理由は, 本アッセイが mRNA の量を評価しているのに対し, ルシフェラーゼアッセイは発現したタンパクの活性を評価しているという違いのためであろう.

また, ヒト内在性遺伝子のビメンチン (VIM; NM_003380) を標的とする 2 種の siRNA について本アッセイにより標的切断活性を評価し, 一方で内在性ビメンチンをノックダウンした効果をリアルタイム RT-PCR および抗体染色で評価した. 本アッセイで得られた結果は, 内在性ビメンチンに対して各 siRNA が示した RNAi 活性とよく一致した (**Fig. 1-9**).

以上の結果より, 本アッセイが RNAi 活性を評価する方法として有効であることが確認できた.

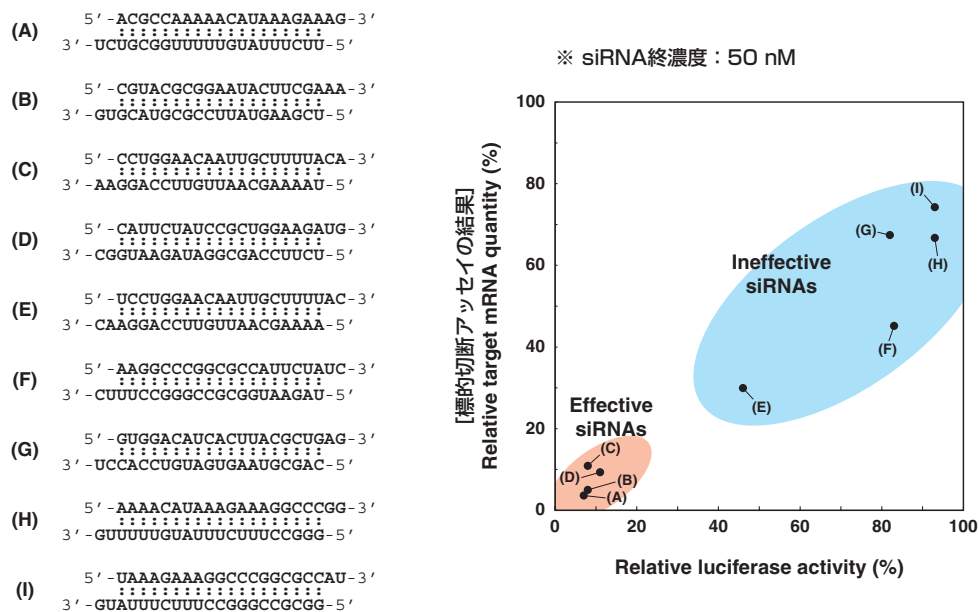


Figure 1-8. ホタルルシフェラーゼ遺伝子を標的とする9種のsiRNAを標的切断アッセイで評価した。標的切断アッセイの結果（縦軸）と、ホタルルシフェラーゼ遺伝子に対するRNAi効果（横軸）の間には高い相関が認められた。ルシフェラーゼ相対活性の値が低いほど、siRNAのRNAi活性が高いことを示している。

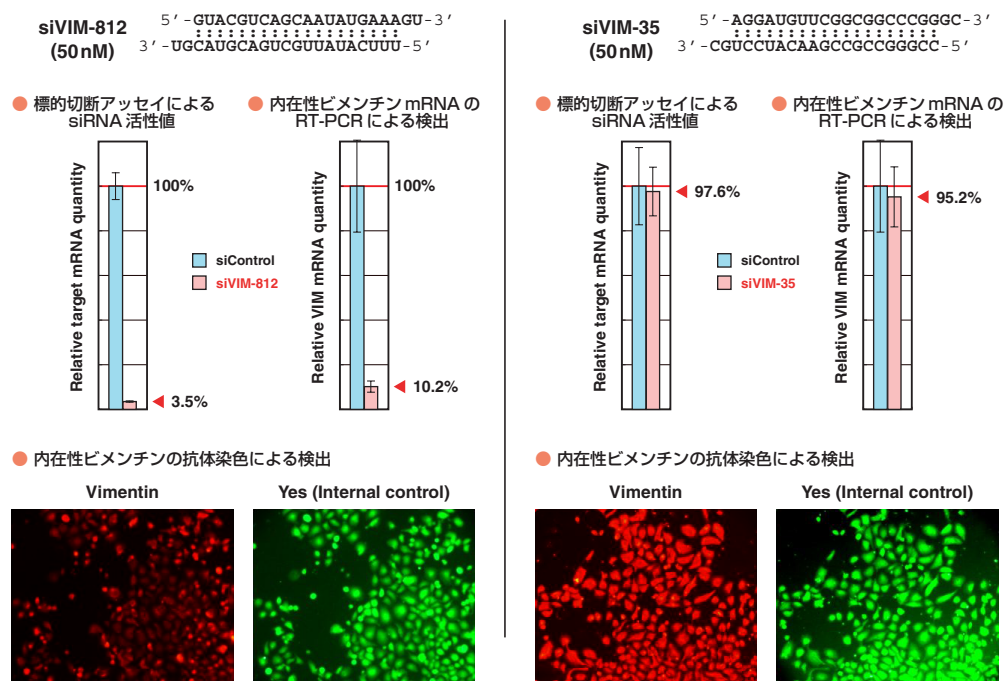


Figure 1-9. ヒト内在性遺伝子のビメンチンを標的とする2種のsiRNAを標的切断アッセイで評価した。標的切断アッセイの結果は、内在性ビメンチンをノックダウンしたときのmRNA定量値および抗体染色の結果と一致した。

さらに、pTREC ベクターから発現させた標的 RNA の siRNA による切断を直接検出するために、プライマーエクステンション法による解析をおこなった。

ホタルルシフェラーゼ遺伝子に対する siRNA (siFL2-153) と相同な 23 塩基を挿入した標的 mRNA を、pTREC から発現させた。それと同時に、siFL2-153 もしくは無関係な配列の siRNA (siControl) を導入した場合について、poly A⁺ RNA を抽出してプライマーエクステンションをおこなった。その結果、いずれの場合も標的 mRNA の 5' 末端に対応するバンドが予想された位置にあらわれた (**Fig. 1-10A**)。また、RNA 抽出の過程で核を分離していないため、イントロンを含む pre-mRNA の 5' 末端に対応するバンドも見られた。

標的 mRNA に siFL2-153 を作用させた場合、核内の pre-mRNA のバンドに大きな変化はなかったが、mRNA の 5' 末端に対応するバンドが薄くなった (**Fig. 1-10A**)。さらに、siRNA の標的配列部分に新たなバンドが生じた (**Fig. 1-10B**)。このバンドは、siRNA によって切断された mRNA のうち 3' 側の RNA 断片の 5' 末端に対応するものと考えられる。すなわち、**Figure 1-10B** のバンドの位置が、siRNA による標的 RNA の切断点であると考えられる。シーケンスラダーを参照すると、この切断点は siRNA のアンチセンス鎖の 5' 末端から 10 番目の塩基と 11 番目の塩基の間の位置に対応することが判明した (**Fig. 1-10C**)。この結果は、ショウジョウバエの *in vitro* での実験における切断点と一致した (Elbashir *et al.*, 2001b,c)。なお、この切断点から 3' 側に 5

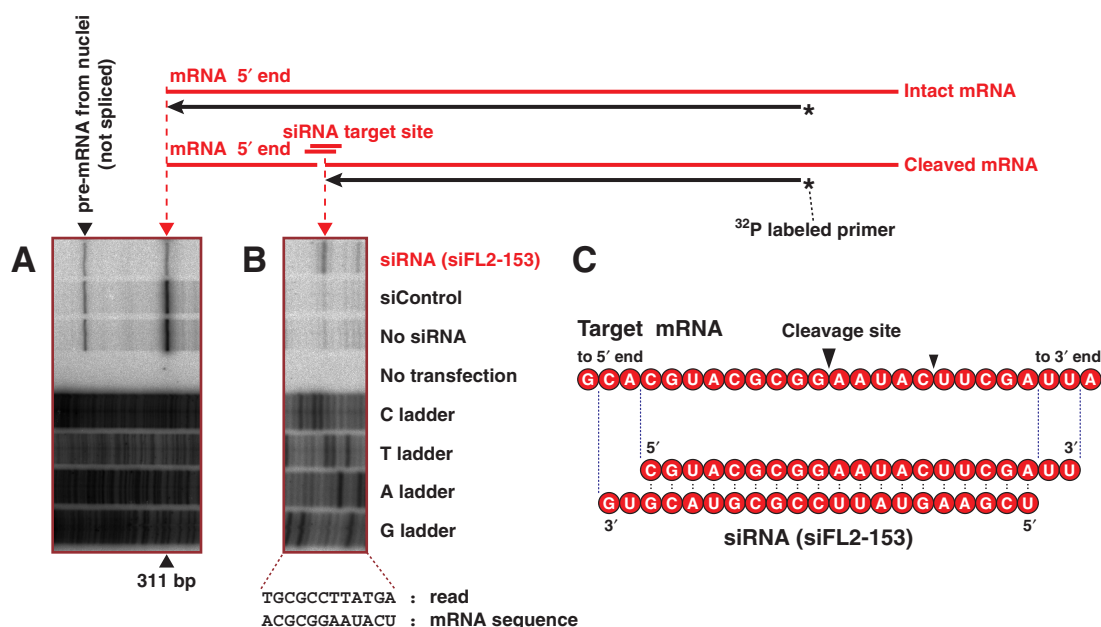


Figure 1-10. プライマーエクステンション法による標的 mRNA の切断の検出. (A) mRNA および pre-mRNA の 5' 末端に対応するバンド. (B) siRNA による切断点に対応するバンド. (C) 本実験のバンドの位置から求めた切断点を図示した.

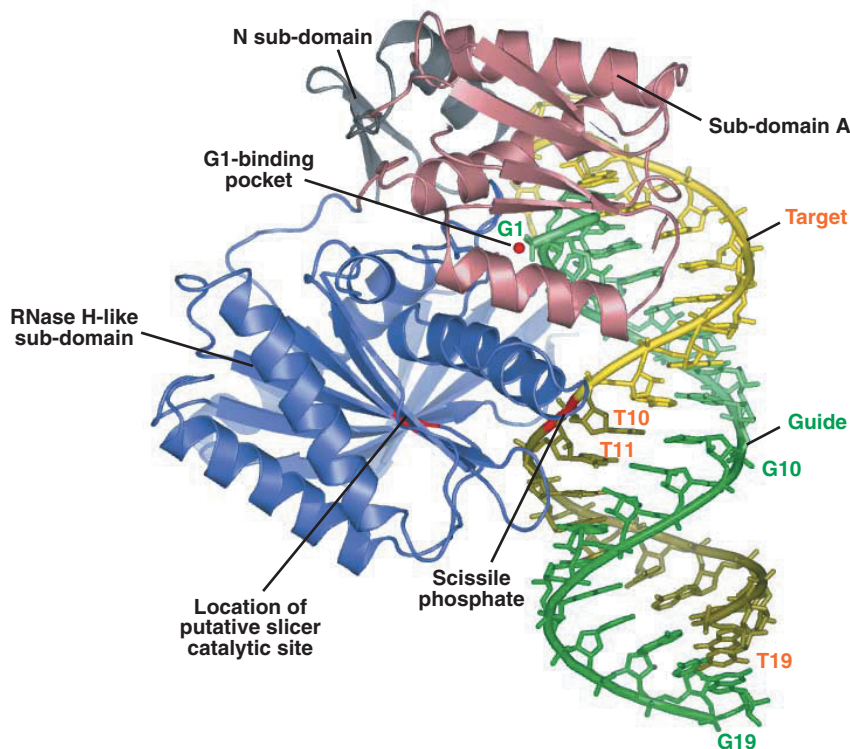


Figure 1-11. 硫酸還元古細菌 *Archaeoglobus fulgidus* の PIWI ドメインと、19 塩基対の 2 本鎖 RNA の複合体の立体構造 (Parker *et al.*, 2005 より作成)。緑色の RNA 鎖が siRNA のアンチセンス鎖、黄色の RNA 鎖が標的 mRNA に相当する。

塩基離れた位置にも薄いバンドが存在し (Fig. 1-10B), この位置でも標的 mRNA の切断が起こっている可能性が示唆された。この第 2 のバンドがアーティファクトである可能性を排除するため、異なる配列の siRNA でも実験をおこなったところ、同様の位置に第 2 のバンドが確認された。また、siRNA のアンチセンス鎖の 5' 末端から 7 塩基を DNA に置換したキメラ型 siRNA で同様の実験をおこなったところ、主バンドはそのままに、第 2 のバンドのみ消失した (データ示さず)。

標的 mRNA の切断点については、RISC を構成するタンパク質である Argonaute がもつ PIWI ドメインの立体構造からも検討できる。Argonaute の PIWI ドメインは RNAi の実行過程において標的 mRNA の切断活性を担っており (Yuan *et al.*, 2005), RNaseH と類似した立体構造を有する (Song *et al.*, 2004)。Figure 1-11 に示すように、Argonaute と結合した siRNA のアンチセンス鎖 (緑色) の 5' 末端 (G1) は PIWI のポケットに収まり、siRNA のアンチセンス鎖と対合した標的 mRNA (黄色) の切断点に相当する部分 (赤色) が、RNaseH の活性中心とされているアミノ酸の近傍に位置している (Ma *et al.*, 2005; Parker *et al.*, 2005)。しかしながら、この切断点から 3' 側に 5 塩基離れたヌクレオチドは PIWI ドメインの反対側に位置しており、どのようなメカニズムでこの部分の切断が起こっているのか説明できず疑問が残った。

1-4. オフターゲット効果が少ない siRNA 設計法の実験的検討 (1)

siRNA 配列と mRNA 配列との間にどの程度の相同性があるとオフターゲット効果が生じるかを明らかにするため、これまでの報告では、標的 mRNA とは少しずつ配列が異なる siRNA を合成しそれらの RNAi 効果を評価する方法がとられてきた。しかしながら、化学合成の siRNA は高価なこともあり、特定の遺伝子の限られた配列しか検討されていなかった。さらに、緒論でも述べたように siRNA の活性はその配列自体によって大きく変化するため、siRNA の配列を変えろという手段ではオフターゲット効果を独立に評価することはできない。オフターゲット効果を正確に評価するためには、siRNA の配列は変えずに、標的 mRNA の配列を変化させる方法が必要である。そこで本研究では、ホタルルシフェラーゼ遺伝子を標的とする siRNA (siFL-36) の配列に対して、様々なミスマッチを含む標的 mRNA を発現する pTREC コンストラクトを系統的に構築した。

まず、siFL-36 の標的配列に 1 塩基の置換を導入したすべての場合を検討した (Fig. 1-12)。siRNA を 50 nM で導入した場合、特に標的配列の中央付近の置換により RNAi 効果が失われた場合があったが (◀で表示)、57 例中 49 例について標的 mRNA が 50% 以下まで分解された。siFL-36 以外の配列でも検討をおこなったところ、やはり標的配列の中央付近の置換で RNAi 効果が失われた場合があった (データ示さず)。ただし、RNAi 効果が失われるような置換の位置や、塩基の種類について規則性は見いだせなかった。標的配列の中央付近と末端付近とでミスマッチの効果が異なる事実は、他の報告 (Du *et al.*, 2005; Birmingham *et al.*, 2006) でも指摘されている。

つぎに、siFL-36 の標的配列に 2 塩基の置換を導入した 23 例を同様に検討した (Fig. 1-13)。標的配列の中央付近に 2 塩基の連続する置換を導入した場合、RNAi 効果がほぼ失われた。ただし、標的配列の末端付近の置換においては、標的 mRNA が 50% 以下まで分解された場合があった (◀で表示)。また、連続しない 2 塩基の置換を導入した場合、それらの位置や種類によってミスマッチの効果は異なっていた (Fig. 1-13)。

さらに、siFL-36 の標的配列に 3 塩基の置換を導入した 12 例を同様に検討した (Fig. 1-14)。その結果、1 例 (◀で表示) を除くすべての例で、RNAi 効果がほぼ失われた。

以上の結果を考慮すると、siRNA 設計においては、標的と無関係なすべての配列に対してかならず 3 塩基以上がミスマッチする配列を選択することが望ましい。後述する情報科学的な解析からは、そのような siRNA は全体の約 10% 存在するが、4 ミスマッチ以上を保証できる siRNA はほとんど存在しない (Naito, Yamada *et al.*, 2004)。したがって、現実的なアプローチとして 3 ミスマッチ以上の siRNA を選択することが妥当と考えられる。

なお、ミスマッチの位置によってその影響が異なる理由は、次のように考えられる。siRNA のアンチセンス鎖と標的 mRNA は、A-form の 2 本鎖 RNA を形成して PIWI ドメインと結合している (Fig. 1-11)。ところが、siRNA のアンチセンス鎖の中央部分が標的 mRNA とミスマッチす

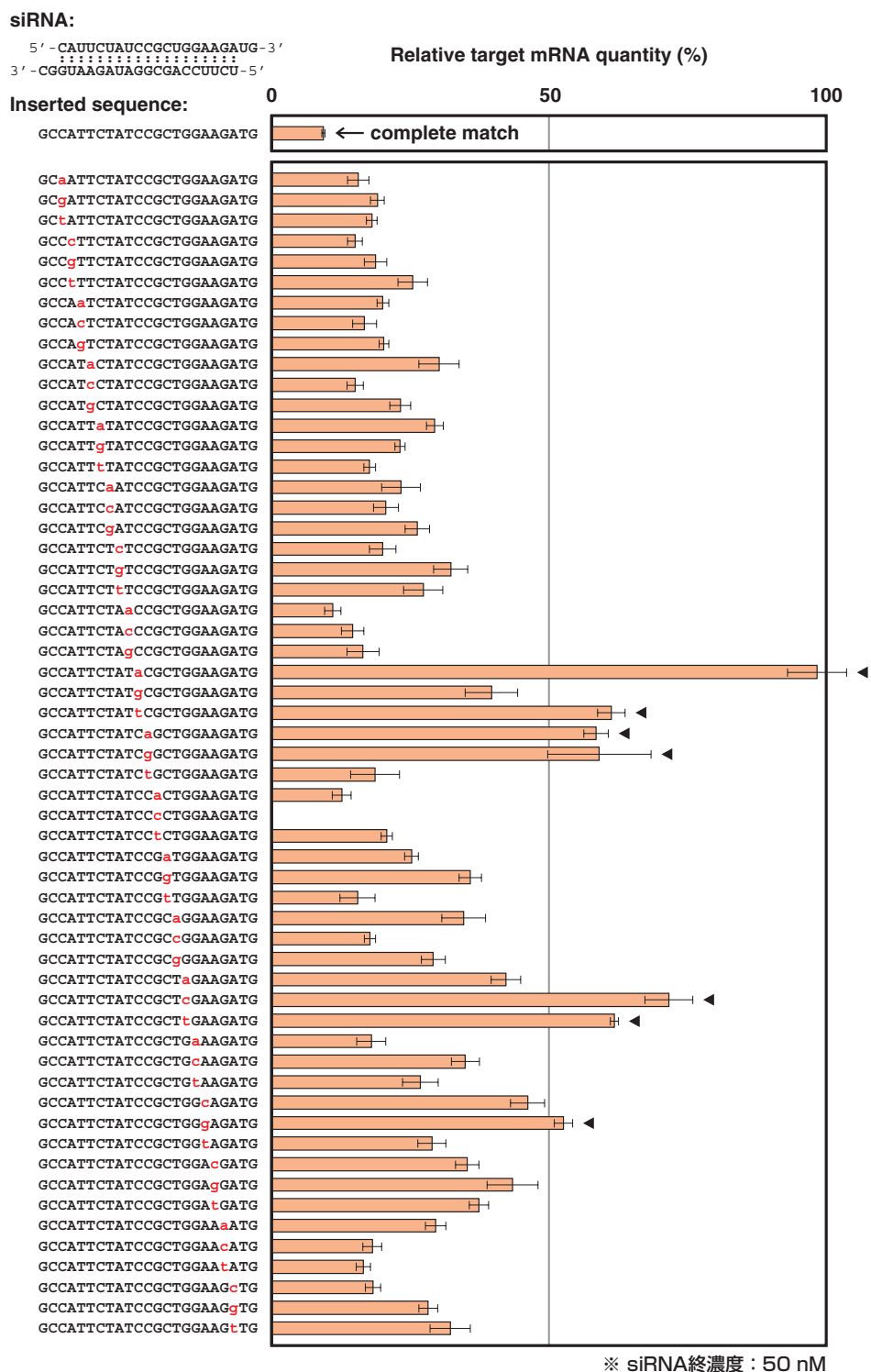


Figure 1-12. siFL-36 の標的配列に 1 塩基の置換を導入した場合の、siFL-36 による RNAi 効果。最上段は、siFL-36 と相同な標的配列に対する RNAi 効果（コントロール）。siFL-36 の配列および、pTREC に挿入した配列を図中に示した。

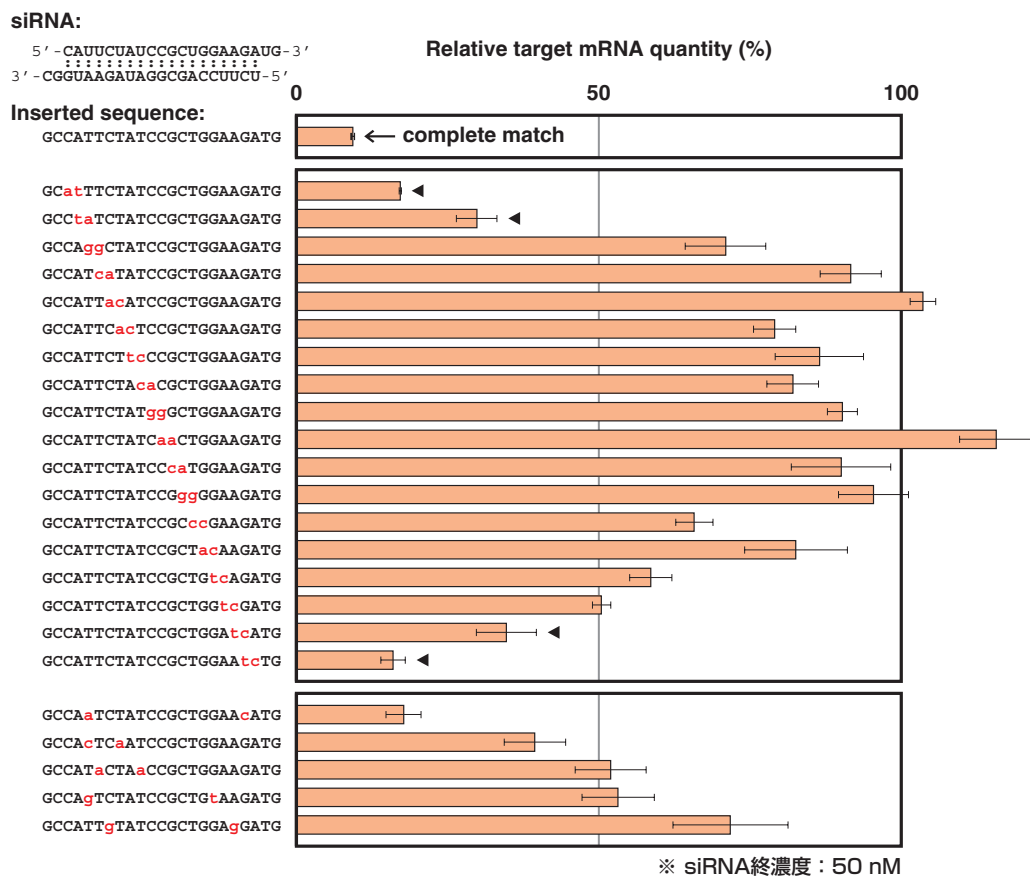


Figure 1-13. siFL-36 の標的配列に 2 塩基の置換を導入した場合の RNAi 効果.

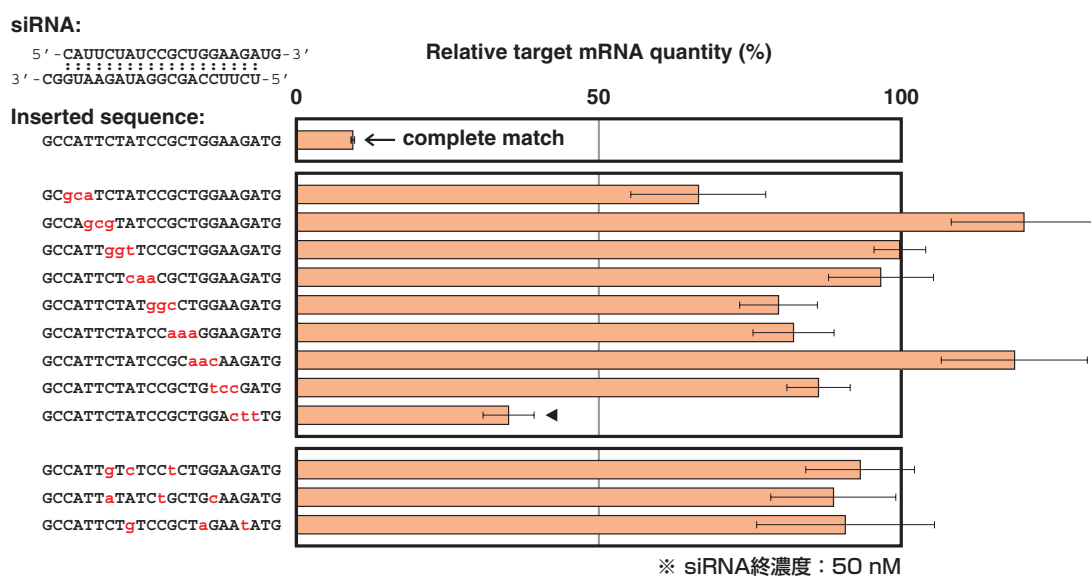


Figure 1-14. siFL-36 の標的配列に 3 塩基の置換を導入した場合の RNAi 効果.

る場合、2本鎖の立体構造がみだれ、標的 mRNA の切断点と PIWI ドメイン活性中心との位置関係が変化すると予想される。その結果、標的 mRNA は切断を受けなくなる可能性が高い。一方、ミスマッチの位置が siRNA の末端部分の場合は、中央付近のミスマッチと比較して立体構造への影響は少なく、このため標的 mRNA の切断活性も保持されるのであろう。

1-5. オフターゲット効果が少ない siRNA 設計法の実験的検討 (2)

近年、siRNA を構成するアンチセンス鎖の 5' 末端から 2 ～ 8 番目の塩基 (以下、**seed** と呼ぶ) が mRNA と完全一致する場合に、翻訳の抑制および mRNA の分解が起こりうることを示された (Lim *et al.*, 2005; Lin *et al.*, 2005; Birmingham *et al.*, 2006)。しかしながら、長さが 7 塩基の seed は計算上は mRNA の約 16 kb に 1 箇所の割合で存在するため、seed の完全一致のみで遺伝子が抑制されるならば、膨大な数の遺伝子が影響を受けることになる。そこで本研究では、seed マッチによるオフターゲット効果を以下の方法で検討した。

ヒト GRK4 遺伝子に対する siRNA (Fig. 1-15A; siGRK4) の配列と完全一致しないにも関わらず、siGRK4 によって抑制される遺伝子として、Hif-1 が報告されている (Lin *et al.*, 2005)。そこで、Hif-1 遺伝子の 3' UTR から、siGRK4 と seed マッチする Box1 および Box2 (Fig. 1-15B) だけを取り出して、ウミシイタケルシフェラーゼ遺伝子の 3' UTR に挿入したコンストラクトを作製した。また同様に、Box1 または Box2 をタンデムに 2 個もしくは 3 個挿入したコンストラクトを作製した。これらのコンストラクトと同時に、siGRK4 を 0.05 ～ 50 nM の濃度で HeLa 細胞に導入したときの、ウミシイタケルシフェラーゼ遺伝子の相対活性を測定した。

その結果、標的配列として Box1 を 1 個挿入した場合、siGRK4 が 5 nM または 50 nM のときにルシフェラーゼ相対活性が 50% 近くまで抑制されたが、Box2 を 1 個挿入した場合ではルシフェ

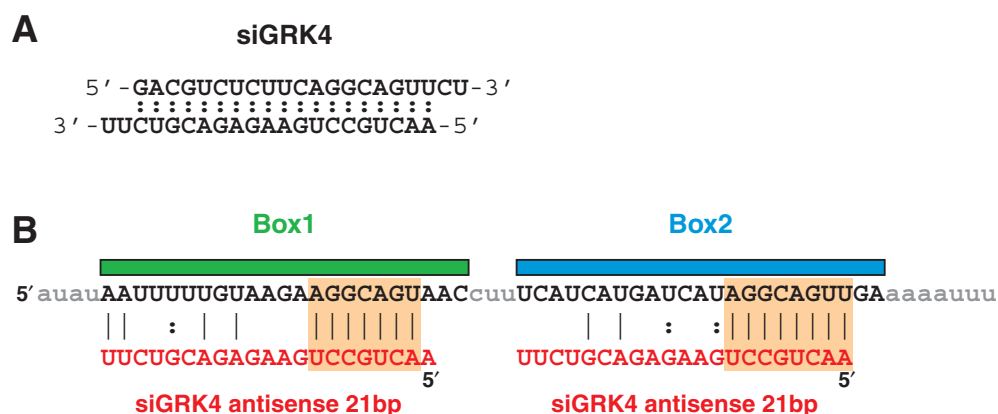


Figure 1-15. (A) siGRK4 の配列. (B) ヒト Hif-1 遺伝子の 3' UTR に存在する、siGRK4 と seed マッチする 2 つの配列. siGRK4 のアンチセンス鎖とのアライメント中、seed マッチしている部分を影付きで示した。

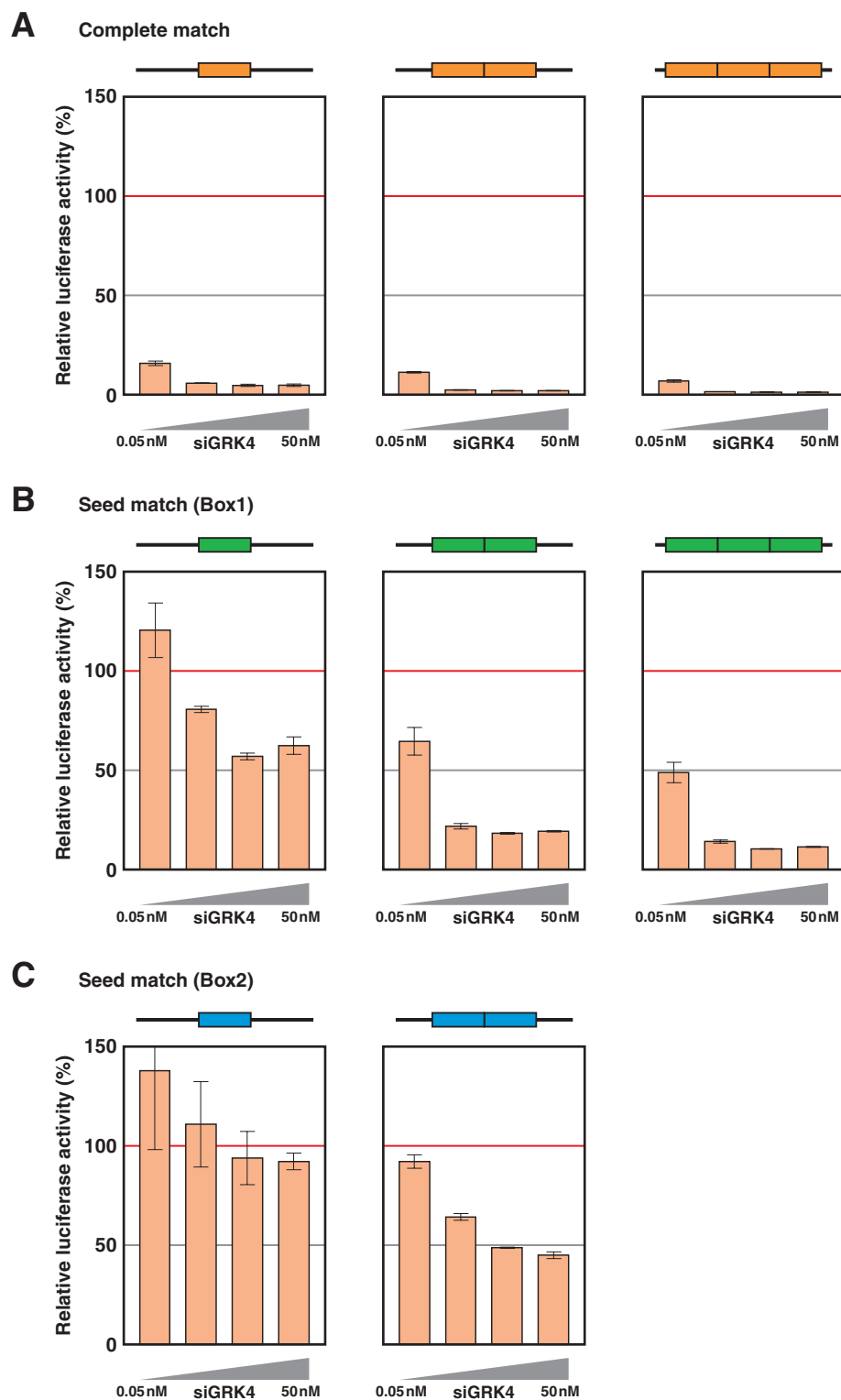


Figure 1-16. (A) siGRK4 と相補的な標的配列（黄色）に対して siGRK4 を作用させた場合の抑制効果. (B) Box1（緑色）を 1 つ, 2 つ, または 3 つ挿入した標的に対して siGRK4 を作用させた場合の抑制効果. (C) Box2（青色）を 1 つまたは 2 つ挿入した標的に対して siGRK4 を作用させた場合の抑制効果. いずれも siGRK4 は 0.05, 0.5, 5, 50 nM の終濃度で導入した.

ラーゼ相対活性の減少は見られなかった。一方、Box1 を 2 個または 3 個挿入した場合、もしくは Box2 を 2 個挿入した場合に、ルシフェラーゼ相対活性が強く抑制された (Fig. 1-16)。

以上の結果は、わずか 7 塩基の seed マッチであっても、同一の mRNA 分子内に複数の seed マッチが存在する場合に大きなオフターゲット効果が起こりうることを示唆した。したがって siRNA の設計においては、標的とは無関係な転写産物のなかに複数の seed マッチが存在するような siRNA 配列を避けることが望ましいと考えられた。ただし、RefSeq release 20 に登録されたヒトの全遺伝子の配列を解析したところ、いかなる遺伝子においても同一の転写産物内に 2 回以上出現しないような seed は、 4^7 種類の seed のうちわずか 5 種類しか存在しなかった。したがって、現実的にはより緩やかな基準を検討する必要がある。そのような方法のひとつとして、ヒトの転写産物において出現頻度の高い seed をなるべく避けるというアプローチが考えられる。

1-6. siRNA 設計システムの構築

これまでに述べた実験的な結果に基づき、哺乳類細胞に有効かつオフターゲット効果の少ない siRNA の設計システムを構築した。哺乳類細胞に有効な siRNA の設計法は、Ui-Tei ら、Reynolds ら、Amarzguioui らの 3 つの設計法を比較検討した結果に基づき、Ui-Tei らの方法を用いた。また、オフターゲット効果の少ない siRNA を選択するために全遺伝子に対する相同性検索をおこない、標的と無関係なすべての配列に対して 3 塩基以上がミスマッチする配列を可能な限り選択することとした。相同性検索には、当初 BLAST (Altschul *et al.*, 1990) を用いたが、

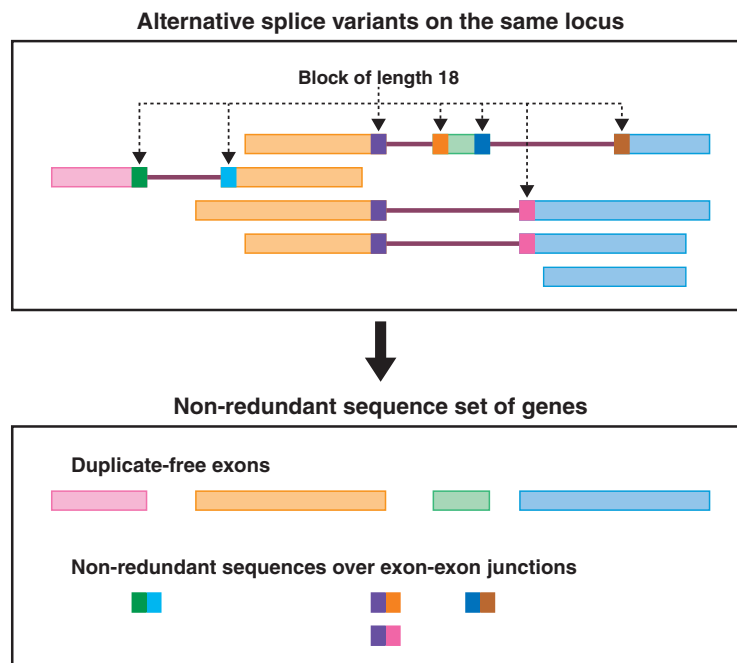


Figure 1-17. NRDB の構築 (Naito, Yamada *et al.*, 2004 より作成)。冗長な配列をゲノムにマッピングすることによって、非冗長なエクソン配列およびエクソンジャンクション配列を得た。

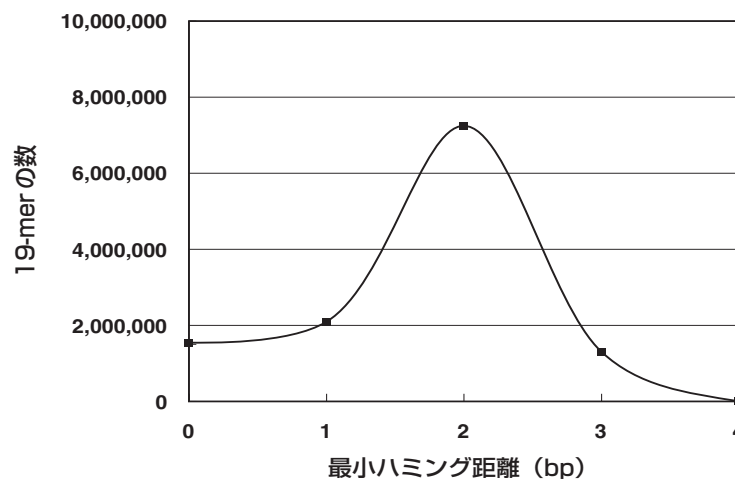


Figure 1-18. ヒトの NRDB を構成する 19-mer の最小ハミング距離の分布 (Naito, Yamada *et al.*, 2004 より作成). 最小ハミング距離が5であるような 19-mer は存在しない.

BLAST は siRNA のような短い配列の検索において見落としが多いという欠点があることが判った (Fig. 4). このため短い配列の検索を見落としなく高速に実行できる Yamada らのアルゴリズム (Yamada *et al.*, 2005) を用いることとした.

全遺伝子に対する相同性検索においては, RefSeq や GenBank など既存のデータベースを用いると, 同一の遺伝子に由来する配列が複数登録されているため siRNA の特異性の評価が難しいという問題があった. たとえば, ヒトの *src* 遺伝子は RefSeq release 20 において NM_005417 および NM_198291 という 2 件の配列として登録されているため, *src* を標的とする siRNA の配列を RefSeq release 20 に対して検索すると, この 2 件の配列がヒットする. このとき, 2 件のヒットが同一の遺伝子に由来するのか, それとも無関係な遺伝子を含むのかを機械的に判断することは困難である. そこで, 非冗長な遺伝子配列データベース (NRDB; Non Redundant DataBase) に対してオフターゲット検索をおこなうこととした (Naito, Yamada *et al.*, 2004). NRDB は, RefSeq および Unique UniGene の配列をゲノムにマッピングして得られた非冗長なエクソン配列と, エクソンジャンクション配列とを統合した配列データベースである (Fig. 1-17).

Yamada および著者らは, ヒトの NRDB を構成する 19-mer の部分配列をすべて生成し, 各配列の NRDB における最小ハミング距離を求めた (Naito, Yamada *et al.*, 2004). 最小ハミング距離とは, NRDB 内でその配列を除きもっとも相同性の高い配列とのミスマッチ数である. たとえばある 19-mer の配列が複数の遺伝子に存在する場合, その 19-mer の最小ハミング距離は 0 である. また, ある 19-mer が他のすべての配列に対して必ず 3 塩基以上のミスマッチを有する場合, すなわち, その 19-mer と 2 ミスマッチ以内で相同な配列が他の遺伝子に存在しない場合, その 19-mer の最小ハミング距離は 3 である. なお, 相補的 (complementary) な配列も相同と見なす.

その結果、ヒトの NRDB を構成する 19-mer のうち最小ハミング距離が 3 以上である配列は 10.7% であったが、4 以上の配列はほとんど存在しなかった (Fig. 1-18)。この結果は、siRNA を設計する際、標的と無関係なすべての配列に対して 3 塩基以上がミスマッチする配列は十分に選択可能であるが、4 ミスマッチ以上を保証できる配列はほとんど選択できないことを示唆した。

本研究で構築した siRNA 設計システムの一部は、siDirect ウェブサーバ (Naito, Yamada *et al.*, 2004; <http://design.RNAi.jp/>) として公開されており、その概略は下記のとおりである。

▶ トップページ (Fig. 1-19A)

基本的にはこの画面上に標的遺伝子の塩基配列またはアクセッション番号を入力するだけで、有効な siRNA の配列が得られる。オフターゲット検索をおこなう生物種 (ヒト、マウス等) を選択できるほか、各種のオプションが用意されている。

▶ siRNA 設計結果の表示 (Fig. 1-19B)

画面上の Effective, both-strand specific siRNA は、Ui-Tei らの設計法を満たし哺乳類細胞で有効と予測される配列で、かつ siRNA のセンス鎖、アンチセンス鎖とも、標的以外のすべての遺伝子に対して 3 塩基以上がミスマッチするような siRNA 配列である。通常はこのリストから siRNA 配列を選択すればよい。これらの siRNA は、画面上の Graphical view of effective siRNA candidates では濃い青色で表示される。


このような siRNA がひとつも設計できない場合は Effective, plus-strand specific siRNA を参照する。これらの siRNA は、少なくとも siRNA のアンチセンス鎖については、標的以外のすべての遺伝子に対して 3 塩基以上がミスマッチするような配列である。この場合、siRNA のセンス鎖側の特異性は保証できないが、もしセンス鎖とハイブリダイズしうるオフターゲット遺伝子が存在しても、Ui-Tei らの設計法を満たす siRNA のセンス鎖側は、RISC に効率よく取り込まれないと考えられるため (Ui-Tei *et al.*, 2004)、その遺伝子に対するオフターゲット効果はおそらく問題にならない。これらの siRNA は、Graphical view of effective siRNA candidates で薄い青色で表示される。

▶ オフターゲット遺伝子の表示 (Fig. 1-19C)

Figure 1-19B のリストで個々の siRNA 配列をクリックすると、その siRNA と 3 ミスマッチ以内で相同なすべての遺伝子が表示され、設計した siRNA の特異性を確認できる (Fig. 1-19C)。siRNA のアンチセンス鎖 (緑) またはセンス鎖 (青) と相同な部分がアライメントとして表示される。

1-7. ヒト全遺伝子に対する siRNA 設計

本研究で構築した siRNA 設計システムにより、RefSeq release 20 に登録されたヒトの全遺伝子に対して siRNA の設計を試みた。各遺伝子について 1 個ずつ、下記の条件を可能なかぎり満た


<http://design.rna.jp/>

Effective, both-strand specific siRNA [map]

Click on each sequence to display siRNA duplex information and its off-target candidates

| target sequence | target position | custom rule range requirement | G/C content longest Cx or longest A's of GC's requirement | G/C content longest Cx or longest A's of GC's requirement | G/C content longest Cx or longest A's of GC's requirement |
|------------------------|-----------------|----------------------------------|--|--|--|
| TGGAGGATGATGCTTTGTGG | 88-110 | >>> | >>> | >>> | >>> |
| AGGGGCTTGGAGGATGGATGC | 144-166 | >>> | >>> | >>> | >>> |
| TGGGGTTCGATGGAATGGATCT | 179-201 | >>> | >>> | >>> | >>> |
| CGGGTTCGATGGAATGGATCT | 180-202 | >>> | >>> | >>> | >>> |
| CGGGTTCGATGGAATGGATCT | 181-203 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 185-207 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 186-208 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 187-209 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 188-210 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 189-211 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 190-212 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 191-213 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 192-214 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 193-215 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 194-216 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 195-217 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 196-218 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 197-219 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 198-220 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 199-221 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 200-222 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 201-223 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 202-224 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 203-225 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 204-226 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 205-227 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 206-228 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 207-229 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 208-230 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 209-231 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 210-232 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 211-233 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 212-234 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 213-235 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 214-236 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 215-237 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 216-238 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 217-239 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 218-240 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 219-241 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 220-242 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 221-243 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 222-244 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 223-245 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 224-246 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 225-247 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 226-248 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 227-249 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 228-250 | >>> | >>> | >>> | >>> |
| TGCAATGCAATGCTTATAGCTC | 229-251 | >>> | >>> | >> | |

[illegible]

Figure 1-19. siDirect ウェブサーバ (<http://design.RNAi.jp/>). (A) トップページ. (B) ヒトの *claudin 17* 遺伝子 (NM_012131) を標的とする siRNA を設計した結果. (C) siRNA 配列と 3 ミスマッチ以内で相同なヒト遺伝子の一覧. 最上段には本来の標的遺伝子である *claudin 17* が 19/19 マッチで表示されている. またその下には, 17/19 マッチで本来の標的遺伝子とは無関係な *claudin 8* が表示されている.

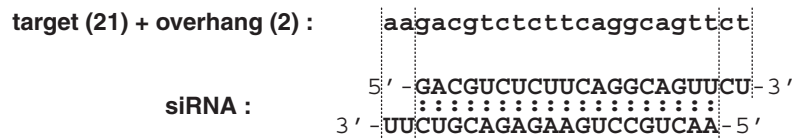


Figure 1-20. 配列表（附録）に記載の塩基配列と、実際の siRNA 配列との対応。

す siRNA の配列を、**配列表**（附録）に示した。ただし、必須と表示されている条件は、かならず満たすこととした。

- ・ Ui-Tei らの siRNA 設計法を満たす、哺乳類で有効と予測される配列（必須）。
- ・ Pol III プロモーターから shRNA を発現させる場合にも使用可能な配列（必須）。
- ・ ヒトの当該遺伝子と対応する、マウスの遺伝子にも共通な配列。なお、共通な配列とは、siRNA のアンチセンス鎖部分の 21 塩基が完全一致していることを意味する。
- ・ 選択的スプライシングによるバリエーションが知られている場合、できるだけ多くのバリエーションに共通な配列。
- ・ 標的と無関係なすべての配列に対して 3 塩基以上がミスマッチする配列。
- ・ 全遺伝子における出現頻度が低い 7 塩基を seed とする配列。

なお、**配列表**（附録）に記載した項目は下記のとおりである。

- ・ spe: Hs（ヒト）または Mm（マウス）。
- ・ Gene ID: Entrez GeneID。
- ・ symbol: 遺伝子名。
- ・ accession: アクセッション番号。
- ・ position: siRNA 配列の最初の塩基の位置。
- ・ target (21) + overhang (2): siRNA のオーバーハングを含む 23 塩基の配列を、a, t, g, c の 4 文字で記載した。G と C の区別を容易にするため、すべて小文字で記載した。詳細は **Figure 1-20** を参照。

1-8. 長い 2 本鎖 RNA から生成される siRNA のオフターゲット効果

化学合成した siRNA を用いて RNAi を誘導する方法以外にも、長い dsRNA を Dicer で処理して得た、さまざまな配列の siRNA の混合物によって RNAi を誘導する方法がある（Kittler *et al.*, 2005）。また、チャイニーズハムスターの CHO-K1 細胞では、長い dsRNA を直接用いて RNAi を誘導できるほか（Ui-Tei *et al.*, 2000）、ショウジョウバエや線虫など哺乳類以外の種では長い dsRNA が一般的に用いられている。長い dsRNA は細胞内で Dicer によって約 21 塩基対の siRNA に分解され、さまざまな配列の siRNA 混合物として標的 mRNA に作用することとなる。このような場合、各 siRNA と相同性の高い配列をもつ遺伝子は、すべてオフターゲット効果を受ける可能性がある。しかしながら、長い dsRNA によるオフターゲット効果を予測する方法としては、BLAST で単純に dsRNA の全長を検索する方法が一般的であった。

本研究では、長い dsRNA によるオフターゲット効果をより正確に予測するため、細胞内における RNAi の実行過程を反映したアルゴリズムを構築した。まず、細胞内で dsRNA が Dicer によって siRNA に分解されることを考慮して、dsRNA 配列を構成する 19 塩基の部分配列（以下、単に 19-mer と呼ぶ）をすべて生成した（Fig. 1-21A）。たとえば、500 塩基の長さの dsRNA 配列からは、482 種の 19-mer が生成できる。このようにして仮想的に生成した各 siRNA について個別にオフターゲット検索をおこない、完全マッチ（19/19 マッチ）、1 ミスマッチ（18/19 マッチ）、2 ミスマッチ（17/19 マッチ）で相同なすべての遺伝子を列挙して集計した（Fig. 1-21B）。実際には dsRNA 配列を構成するすべての部分配列が siRNA として生成されるわけではないが、オフターゲット効果の対象となりうる遺伝子を見落としなく列挙するため、本研究ではすべての部分配列が生成されると仮定して解析をおこなった。なお、部分配列の長さを 21 塩基ではなく 19 塩基とした理由は、siRNA の末端の塩基が標的 mRNA とミスマッチしていても、中央の 19 塩基が完全一致していれば標的の切断が起こりうるためである。

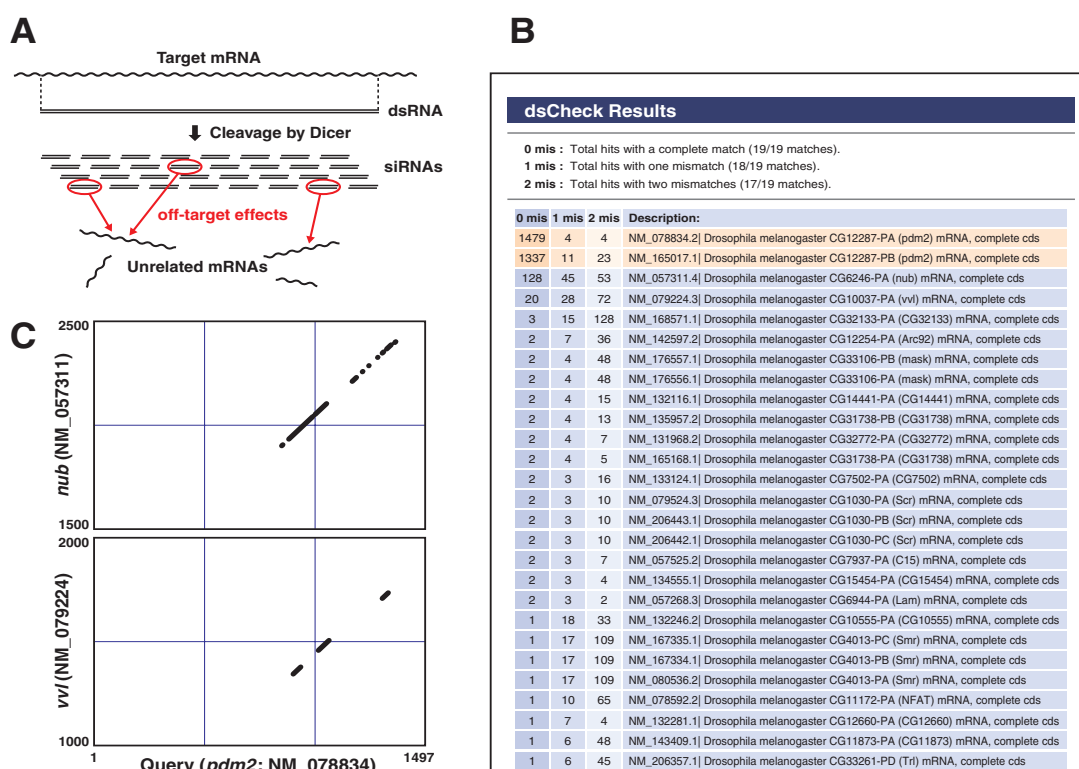


Figure 1-21. Naito, Yamada *et al.*, 2005 より作成。(A) 長い dsRNA は細胞内で Dicer によって約 21 塩基対の siRNA に分解される。(B) ショウジョウバエ *pdm2* 遺伝子の全長を標的とする dsRNA のオフターゲット効果を予測した例。最上段には本来の標的遺伝子である *pdm2* 遺伝子の 2 つのバリエントが表示されている。その下にはオフターゲット遺伝子の候補として、*nub* 遺伝子、*vvl* 遺伝子などが表示されている。(C) *pdm2* 遺伝子、*nub* 遺伝子、*vvl* 遺伝子は、保存された POU ドメインを共有している。17/19 以上のマッチをプロットした。

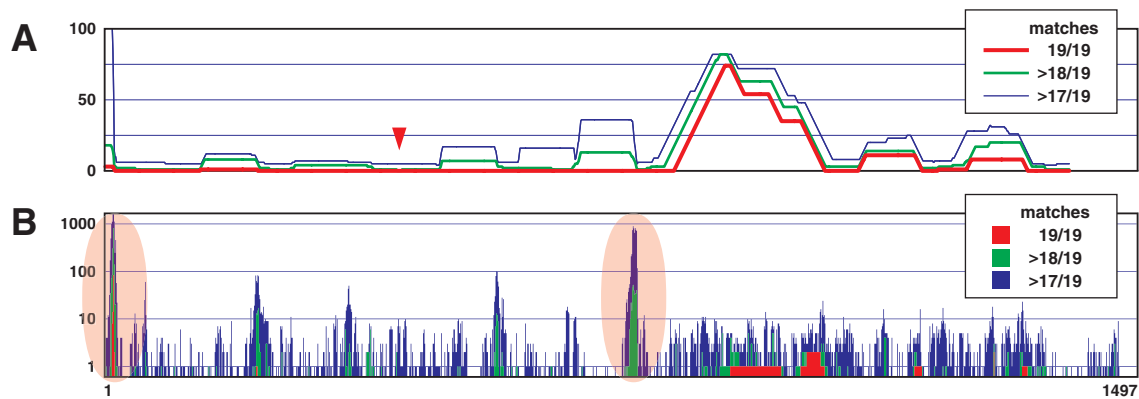


Figure 1-22. *pdm2* 遺伝子に対して特異性の高い dsRNA を設計する方法の検討 (Naito, Yamada *et al.*, 2005 より作成). (A) 100 bp の dsRNA から生成した 82 種の 19-mer による, 協調的ヒットの最大値. 矢印で示した位置が最も協調的ヒットが少なくなる. (B) 各 19-mer と相同性をもつ遺伝子の数. 影付で示した位置の 19-mer は, 多数の遺伝子と相同性を有する.

Figure 1-21B は, ショウジョウバエの POU ドメインタンパクである *pdm2* 遺伝子 (NM_078834) のコード領域の全長 (1497 bp) を標的とする dsRNA について, この方法でオフターゲット効果を予測した結果である. 最上段から順に, *pdm2* 遺伝子の 2 つのバリエント (NM_078834, NM_165017) および, 無関係な遺伝子である, *nub* 遺伝子 (NM_057311), *vvl* 遺伝子 (NM_079224) がヒットしている. **Figure 1-21C** に示すように, *pdm2*, *nub*, *vvl* の各遺伝子は, 高度に保存された POU ドメインを有しており, この領域を標的とする dsRNA は, *nub* 遺伝子や *vvl* 遺伝子に対するオフターゲット効果のリスクが高いと予想される.

このような遺伝子に対して特異性が高い dsRNA を設計するためには, 特定のオフターゲット遺伝子に対して多数の 19-mer がヒットする**協調的ヒット (collaborative off-target hit)**を避けることが望ましい (Naito, Yamada *et al.*, 2005). **Figure 1-22A** は, *pdm2* 遺伝子に対して 100 bp の dsRNA をさまざまな位置に設計した場合に, dsRNA から生成された 82 種の 19-mer による協調的ヒットの最大値を計算したものである. POU ドメインが存在する 1000 bp 付近にピークがあり, この付近では dsRNA から生成した 19-mer の多くが特定のオフターゲット遺伝子 (この場合は *nub* 遺伝子) と相同性を有することを示している. 一方, 500 bp 付近の矢印 (▼) で示した位置は協調的ヒットが少なく, 100 bp の dsRNA を設計するうえで好適と考えられる. また **Figure 1-22B** の影付で示すように, 特定の 19-mer が多数の遺伝子と相同性を有する場合があります, そのような 19-mer を含まないように dsRNA を設計することが望ましいと考えられる.

本研究で構築した, 長い dsRNA によるオフターゲット効果の予測法と, 特異性が高い dsRNA の設計法の一部は, dsCheck ウェブサーバ (<http://dsCheck.RNAi.jp/>) として公開されている (Naito, Yamada *et al.*, 2005).

2. 方 法

2-1. Ui-Tei らによる siRNA 設計法

Ui-Tei *et al.*, 2004 にしたがって、下記 (i) ~ (iv) の条件をすべて満たす siRNA を有効と判定した。Figure 1-1A に (i) ~ (iv) を図示した。以下、siRNA を構成するセンス鎖、アンチセンス鎖を、それぞれ単にセンス鎖、アンチセンス鎖と呼ぶ。

- (i) アンチセンス鎖の 5' 末端が A または U である。
- (ii) センス鎖の 5' 末端が G または C である。
- (iii) アンチセンス鎖の 5' 端の 7 塩基のうち 4 塩基以上が A または U である。
- (iv) 配列中に G/C が 10 塩基以上連続しない。

2-2. Reynolds らによる siRNA 設計法

Reynolds *et al.*, 2004 にしたがって、下記 (I) ~ (VIII) の合計得点が 6 以上の siRNA を有効と判定した。Figure 1-1B に (I) ~ (VIII) を図示した。

- (I) siRNA の 2 本鎖部分における G/C の数が、19 塩基中 7 ~ 10 塩基の範囲であれば +1 点。
- (II) アンチセンス鎖の 5' 端の 5 塩基における A または U の数 (0 ~ 5) を得点とする。
- (III) 2 本鎖部分の Tm が 20℃ 未満。判定不能のため本項目は無視する。
- (IV) センス鎖の 5' 端から 19 番目の塩基が A であれば +1 点。すなわち、アンチセンス鎖の 5' 末端が U であれば +1 点。
- (V) センス鎖の 5' 端から 3 番目の塩基が A であれば +1 点。
- (VI) センス鎖の 5' 端から 10 番目の塩基が U であれば +1 点。
- (VII) アンチセンス鎖の 5' 末端が G または C であれば -1 点。
- (VIII) センス鎖の 5' 端から 13 番目の塩基が G であれば -1 点。

2-3. Amarzguioui らによる siRNA 設計法

Amarzguioui *et al.*, 2004 にしたがって、下記 (1) ~ (6) の合計得点が 3 以上の siRNA を有効と判定した。Figure 1-1C に (1) ~ (6) を図示した。

- (1) $\Delta T3$: アンチセンス鎖の 5' 端の 3 塩基における A または U の数 (0 ~ 3) から、センス鎖の 5' 端の 3 塩基における A または U の数 (0 ~ 3) を引いた値を得点とする。-3 から +3 までの値を取りうる。
- (2) センス鎖の 5' 末端が G または C であれば +1 点。
- (3) センス鎖の 5' 末端が U であれば -1 点。
- (4) センス鎖の 5' 端から 6 番目の塩基が A であれば +1 点。
- (5) センス鎖の 5' 端から 19 番目の塩基が G であれば -1 点。すなわち、アンチセンス鎖の 5' 末端が C であれば -1 点。
- (6) アンチセンス鎖の 5' 末端が A または U であれば +1 点。

2-4. siRNA

Proligo 社より購入したセンス 1 本鎖 RNA とアンチセンス 1 本鎖 RNA をアニーリングして使用した。アニーリングは、終濃度 10 mM Tris-HCl (pH 7.5), 20 mM NaCl の緩衝液中で 95℃

で3分インキュベートし、徐々に室温まで冷却した。もしくは、2本鎖にアニーリングされた siRNA を RNAi 社（東京）より購入した。siRNA の配列は **Table S1**（巻末）、**Figure 1-8, 1-9, 1-10C, 1-12, 1-13, 1-14, 1-15A** にそれぞれ示した。

2-5. 細胞の培養およびトランスフェクション

HeLa 細胞は 10% 非働化ウシ胎児血清（三菱化学）を含有する Dulbecco's modified Eagle medium (DMEM; Gibco 社) に抗生物質（10 U/ml ペニシリン, 50 µg/ml ストレプトマイシン）を添加して 37°C で培養した。HeLa 細胞へのトランスフェクションは、24 ウェルプレートにおいて、Lipofectamine 2000 (Invitrogen 社) を用いて同製品のマニュアルに記載の方法でおこなった。細胞に導入するプラスミドは、Qiagen Plasmid Midi または QIAwell 96 Ultra (Qiagen 社) を用いて抽出した。

2-6. ルシフェラーゼ遺伝子を標的とする RNAi

ホタルルシフェラーゼ遺伝子をコードする pGL3-Control (Promega 社) を 1 µg/well, 内部標準としてウミシイタケルシフェラーゼ遺伝子をコードする pRL-TK (Promega 社) を 0.1 µg/well, siRNA を終濃度 50 nM で HeLa 細胞に導入した。24 時間後に細胞を回収し、Dual-luciferase reporter assay system (Promega 社) でルシフェラーゼ相対活性を測定した。

2-7. 標的 mRNA 発現ベクター pTREC の構築

pCI-neo (Promega 社) の NheI 部位と XbaI 部位との間に、Kozak 配列およびいくつかの制限酵素部位を挿入して pTREC を構築した。挿入した配列は下記のとおりである。

[NheI site]-CACCATGGAATTCACGCGTCTCGAG-[XbaI site]

構築したベクターの構造は **Figure 1-6** に、標的 mRNA として転写される領域の塩基配列は **Figure 1-7** に示した。

2-8. 標的切断アッセイ

pTREC ベクターの EcoRI 部位と XhoI 部位との間に、評価対象の siRNA と相同な 23 塩基の標的配列を **Figure 1-6** の例にしたがって挿入した。ただし、siRNA の配列に対してミスマッチを含む標的配列は、**Figure 1-12, 1-13, 1-14** に示した配列をそれぞれ挿入した。ベクターに挿入した標的配列部分はすべてシーケンスの確認をおこなった。

構築した pTREC コンストラクト (0.5 µg/well) を siRNA (終濃度 50 nM) とともに HeLa 細胞に導入し、24 時間後に細胞を回収後、RNeasy96 (Qiagen 社) で全 RNA を抽出した。得られた RNA を、オリゴ dT をプライマーに SuperScript (Invitrogen 社) で逆転写して cDNA を得た。リアルタイム PCR は、ABI PRISM 7000 (Applied Biosystems 社) を使用して SYBR Green PCR Master Mix (Applied Biosystems 社) でおこなった。プライマーの配列は、

Target-F: 5'-AGGCACTGGGCAGGTGTC
 Target-R: 5'-TGCTCGAAGCATTAACCCTCACTA
 neo-F: 5'-ATCAGGATGATCTGGACGAAG
 neo-R: 5'-CTCTTCAGCAATATCACGGGT

である。ベクターに組み込まれているネオマイシン耐性遺伝子を内部標準として使用し、
 [標的 mRNA 量] / [*neo* mRNA 量] により算出される値を、標的 mRNA 相対量とした。

2-9. ビメンチン遺伝子を標的とする RNAi

抗体染色によるビメンチンタンパク質の検出においては、HeLa 細胞を 24 ウェルプレートにまき、1 日後、2 日後、3 日後にそれぞれ siRNA を終濃度 50 nM で導入した。3 回目の導入操作から 24 時間後に、細胞を 3.7% ホルムアルデヒドで固定し、PBS で洗浄した。ビメンチンの検出は、抗ブタビメンチン抗体 (Oncogene Research Products 社) および Cy3 標識 2 次抗体 (Jackson Immuno Research 社) を用いた。またコントロールとして Yes タンパク質 (YES1; NM_005433) の検出を、抗ヒト Yes 抗体 (Upstate Biotechnology 社) および Cy5 標識 2 次抗体 (Jackson Immuno Research 社) でおこなった。

リアルタイム RT-PCR によるビメンチン mRNA の定量においては、HeLa 細胞に pCAGIPuro-EGFP (ピューロマイシン耐性遺伝子および EGFP をコードするベクター) を 0.5 μ g, siRNA を終濃度 50 nM で導入した。24 時間後より培地をピューロマイシン含有培地に変更して細胞を培養した。導入 3 日後に標的切断アッセイと同様の方法で HeLa 細胞から RNA を抽出し、ビメンチンの mRNA 量および、内部標準として β -アクトニン (ACTB; NM_001101) の mRNA 量を定量した。ビメンチンおよび β -アクトニンの検出に用いたプライマーの配列はそれぞれ、

VIM-F: 5'-GAGCTACGTGACTACGTCCA
 VIM-R: 5'-GTTCTTGAACTCGGTGTTGAT
 ACTB-F: 5'-CACACTGTGCCCATCTACGA
 ACTB-R: 5'-GCCATCTCTTGCTCGAAGTC

である。[VIM mRNA 量] / [ACTB mRNA 量] の値をビメンチン mRNA 相対量とした。

2-10. プライマーエクステンション法

pTREC コンストラクトを 0.5 μ g/well, siRNA を終濃度 5 nM で HeLa 細胞に導入した。24 時間後、1 サンプルあたり 48 ウェルから細胞を回収し、QuickPrep Micro mRNA Purification Kit (Amersham 社) で poly A⁺ RNA を抽出した。抽出した RNA をテンプレートに、³²P で末端標識したプライマー 5'-CTCGAAGCATTAACCCTCACTA を用いて、Primer Extension System (Promega 社) でプライマーエクステンション反応をおこなった。反応産物はシークエンスゲル (尿素 - 6% ポリアクリルアミドゲル) で泳動した。マーカーとして、同じプライマーを用いて pTREC コンストラクトをダイデオキシ法によりシークエンスして得たラダーを泳動した。

| Construct | Inserted sequence |
|-----------|--|
| GRK4.x1 | [XhoI site]-AAGACGTCTCTTCAGGCAGTTCT-[EcoRI site] |
| GRK4.x2 | [XhoI site]-AAGACGTCTCTTCAGGCAGTTCT AAGACGTCTCTTCAGGCAGTTCT-[EcoRI site] |
| GRK4.x3 | [XhoI site]-AAGACGTCTCTTCAGGCAGTTCT AAGACGTCTCTTCAGGCAGTTCT AAGACGTCTCTTCAGGCAGTTCT-[EcoRI site] |
| Box1.x1 | [XhoI site]-AATTTTGTGAAGAAGGCAGTAAC-[EcoRI site] |
| Box1.x2 | [XhoI site]-AATTTTGTGAAGAAGGCAGTAAC AATTTTGTGAAGAAGGCAGTAAC-[EcoRI site] |
| Box1.x3 | [XhoI site]-AATTTTGTGAAGAAGGCAGTAAC AATTTTGTGAAGAAGGCAGTAAC AATTTTGTGAAGAAGGCAGTAAC-[EcoRI site] |
| Box2.x1 | [XhoI site]-TCATCATGATCATAGGCAGTTGA-[EcoRI site] |
| Box2.x2 | [XhoI site]-TCATCATGATCATAGGCAGTTGA TCATCATGATCATAGGCAGTTGA-[EcoRI site] |
| Box2.x3 | [XhoI site]-TCATCATGATCATAGGCAGTTGA TCATCATGATCATAGGCAGTTGA TCATCATGATCATAGGCAGTTGA-[EcoRI site] |

Table 1-1. seed マッチによるオフターゲット効果の検討のため、psiCheck-1 ベクターに挿入した配列の一覧.

2-11. seed マッチによるオフターゲット効果の検討

psiCheck-1 ベクター (Promega 社) の XhoI 部位と EcoRI 部位との間に, **Table 1-1** の配列を挿入したコンストラクトを構築した. 構築したコンストラクトを 0.1 µg/well, pGL3-Control (Promega 社) を 1 µg/well, siRNA を終濃度 0.05 ~ 50 nM で HeLa 細胞に導入した. 24 時間後に細胞を回収し, Dual-luciferase reporter assay system (Promega 社) でルシフェラーゼ相対活性を測定した.

2-12. 共同研究に関する事項

Table S1 (巻末) のうち siRNA の名称に * を表示した箇所は Ui-Tei *et al.*, 2004 によった. また, それ以外の約半数のデータを東京大学の程 久美子博士より供与いただいた. それ以外の箇所は著者が実験をおこない, 全体にわたって著者が主体となって分析および検証をおこなった.

Figure 1-8 のルシフェラーゼ相対活性のデータは, 東京大学の程 久美子博士, 江田 理愛氏より供与いただいた. それに対応する標的切断アッセイは著者がおこない, 全体にわたり著者が主体となって分析および検証をおこなった.

Figure 1-9 の内在性ビメンチンタンパク質の抗体染色のデータは, 東京大学の高橋 史峰博士より供与いただいた. それに対応する標的切断アッセイおよび内在性ビメンチン mRNA の定量は著者がおこない, 全体にわたり著者が主体となって分析および検証をおこなった.

NRDB の構築および siDirect ウェブサーバのプログラム作成は東京大学の山田 智之氏, 森下 真一博士がおこなった. NRDB の概念および siDirect ウェブサーバの仕様は著者が主体となって発案, 構成したほか, siDirect ウェブサーバのプロトタイプを著者が作成した. また, 構築時における評価およびデバッグを著者が主体としておこなった.

dsCheck ウェブサーバのプログラム作成は東京大学の山田 智之氏, 松宮 孝大氏, 森下 真一博士がおこなった. dsCheck ウェブサーバの仕様は著者が主体となって発案, 構成したほか, dsCheck ウェブサーバのプロトタイプを著者が作成した. また, 構築時における評価およびデバッグを著者が主体としておこなった.

第二章

第二章

抗ウイルス siRNA 設計システムの構築および HIV-1 を標的とする RNA 干渉法への応用

1. 結果と考察

1-1. HIV-1 を標的とする siRNA 設計法の検討

HIV-1 や HCV など多くの RNA ウイルスでは、著しく高いゲノム多様性のために siRNA の設計が難しいばかりでなく、1 種類の siRNA を単独で用いた場合、その siRNA に耐性をもつウイルスが容易に出現しうることが報告されている (Boden *et al.*, 2003a; Das *et al.*, 2004). これらの課題を克服するひとつの方法は、ウイルスの配列のなかで高度に保存された領域を標的とすることである。そこで Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov/>) より入手可能な HIV-1 グループ M のゲノム配列 495 種 (Table S2, 巻末) をもとに、siRNA の標的として有力な候補となりうる高度保存領域を同定した。まず、495 種の HIV-1 配列を構成する 21 塩基の部分配列 (以下、単に 21-mer と呼ぶ) をすべて生成し、そのひとつひとつの 21-mer が 495 種のうち何種に共通であるかを調べた。たとえば AATAGCAACAGACATACAAAC という 21-mer が 495 種のうち 300 種に共通であった場合、その 21-mer の **保存度 (conservation)** を $300/495 \times 100 = 60.6 (\%)$ と計算した。ただし、本解析に用いた HIV-1 配列の多くは 5' UTR 部分の配列を欠いているため、5' UTR 領域の保存度を計算する際に分母を 495 とすると、保存度が過小に評価されるという問題があった。そこで、5' UTR 領域のうち LTR 部分については、各々の配列の 3' LTR の配列を用いて補完した。さらに、保存度の計算では分母を一律に 495 とするのではなく、保存度を求めようとしている 21-mer に対応する領域を欠いていない配列の個数をその都度求め、それを分母とした。以上の操作により、5' UTR 領域の 21-mer についても正確に保存度を評価できた。

495 種の HIV-1 配列を構成する 21-mer は合計で 4,417,157 種であり、配列の重複を除くと 1,208,150 種となった。495 種の HIV-1 配列から生成された全 21-mer が、重複を除いた際に 4 分の 1 程度にしかない事実は、HIV-1 ゲノムの多様性が著しく高いことを示唆した。ついで各 21-mer の保存度を siVirus プログラム (Naito *et al.*, 2006) を用いて計算し、青色 (保存度 0%) から赤色 (保存度 100%) の点で示したところ (Fig. 2-1A), それらのほとんどが青色に近く、HIV-1 ゲノムの大部分の領域は 21-mer の単位では保存されていないことが明らかになった。具体的には、保存度が 70% 以上の 21-mer は全体のわずか 1.6% であり、保存度が 50% 以上の 21-mer でも全体の 5.2% に過ぎなかった (Fig. 2-2)。また、すべての HIV-1 配列に共通な、保存度が 100% であるような 21-mer は存在しなかった。これらの結果は、HIV-1 を標的とする siRNA

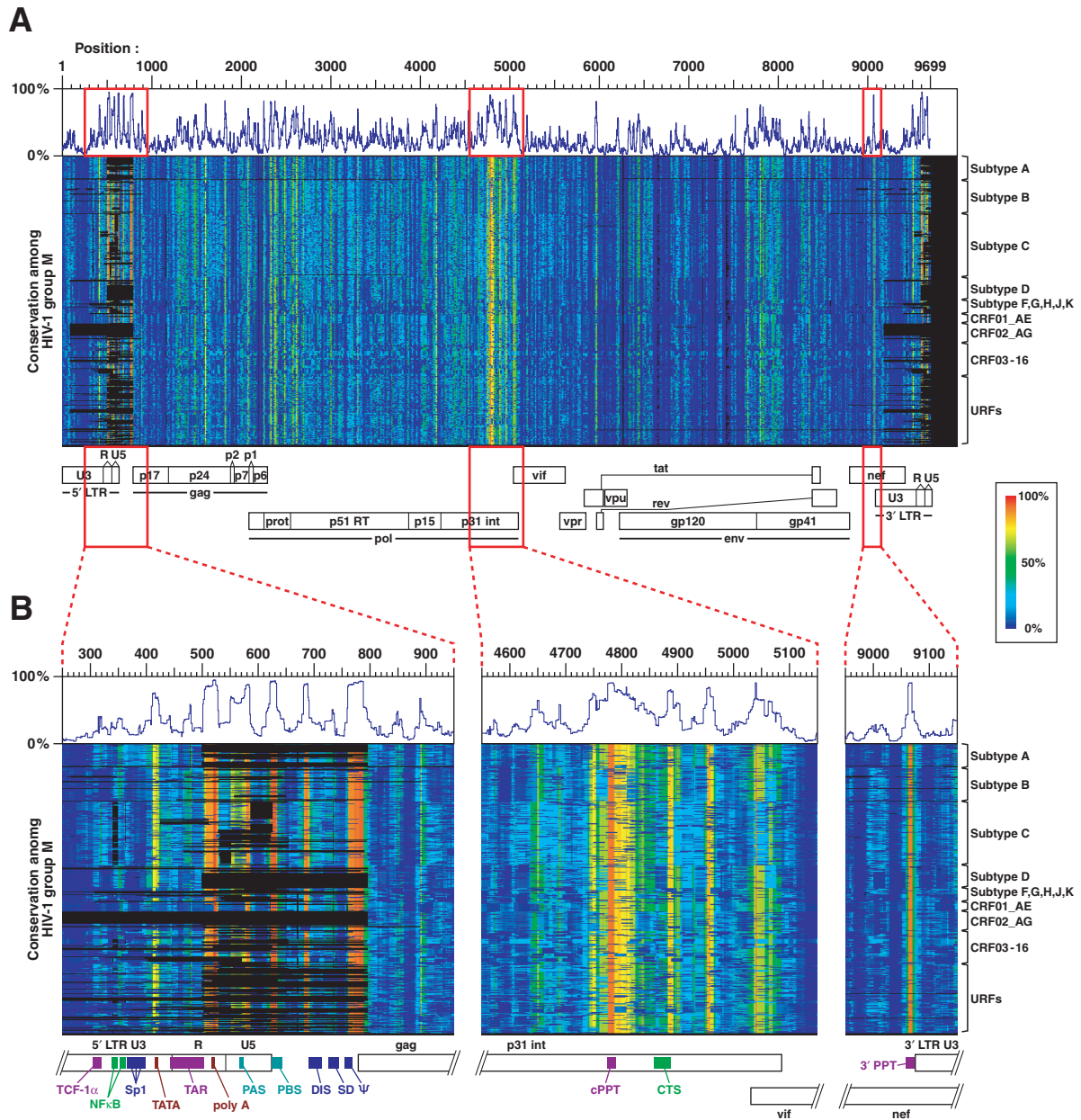


Figure 2-1. HIV-1 ゲノムを構成する 21-mer の保存度. (A) 495 種の HIV-1 配列から生成した 4,417,157 種の 21-mer について保存度を求め、青色（保存度 0%）から赤色（保存度 100%）の点で示した. 折れ線グラフは各位置における保存度の最大値（最頻値と一致）を示す. (B) 5' UTR, インテグラーゼ遺伝子内の cPPT/CTS 領域, および 3' PPT 領域の拡大図.

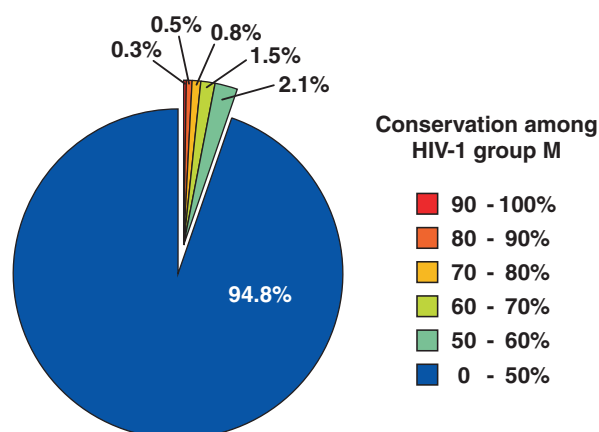
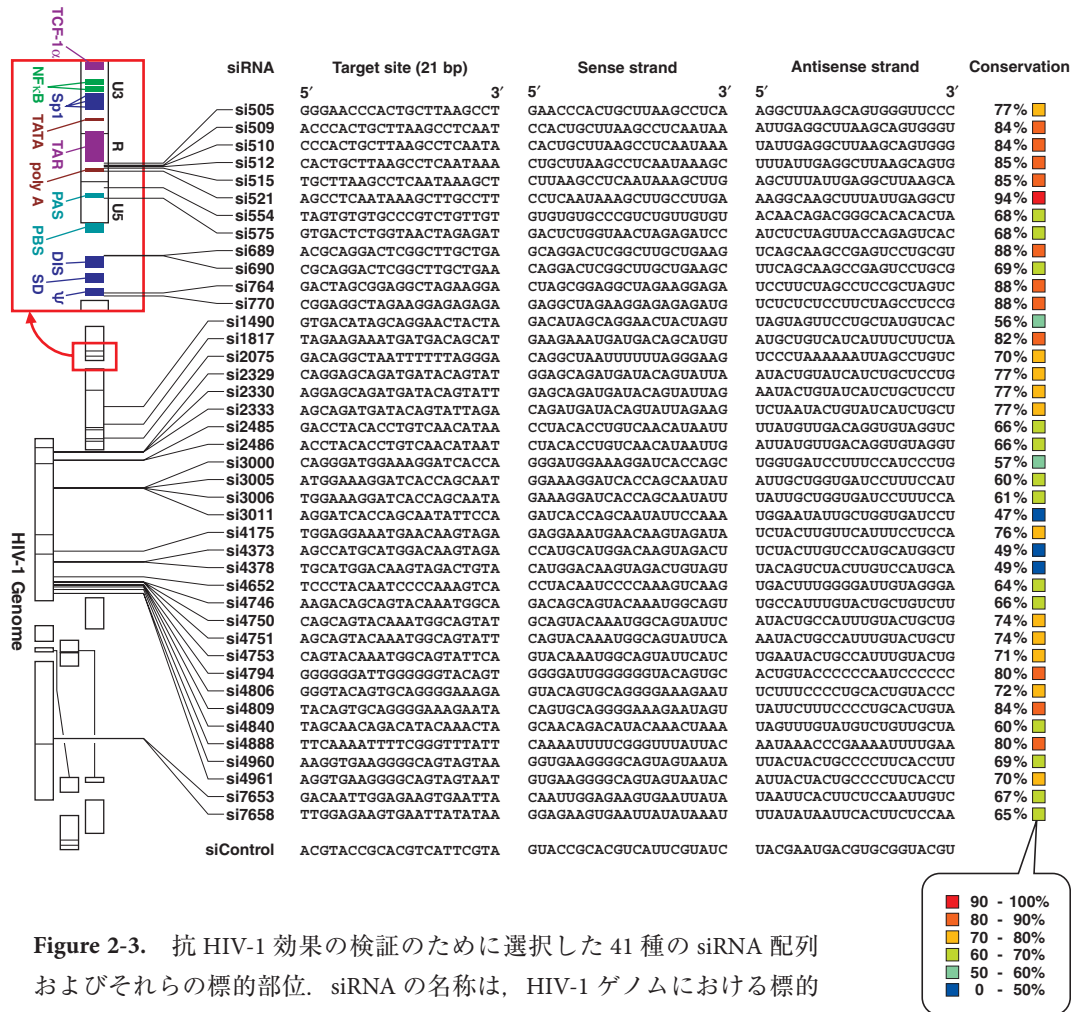


Figure 2-2. 495 種の HIV-1 配列から生成した全 21-mer の保存度の分布.

を無作為に設計しても、高い多様性をもつ HIV-1 を効率よく標的とすることが難しいことを示唆した。なお、HIV-1 を標的とする RNAi 法の試みは既に複数の研究グループより報告されているが、それらの研究で使用されている siRNA の多くは保存度が低かった (Table S3, 巻末)。

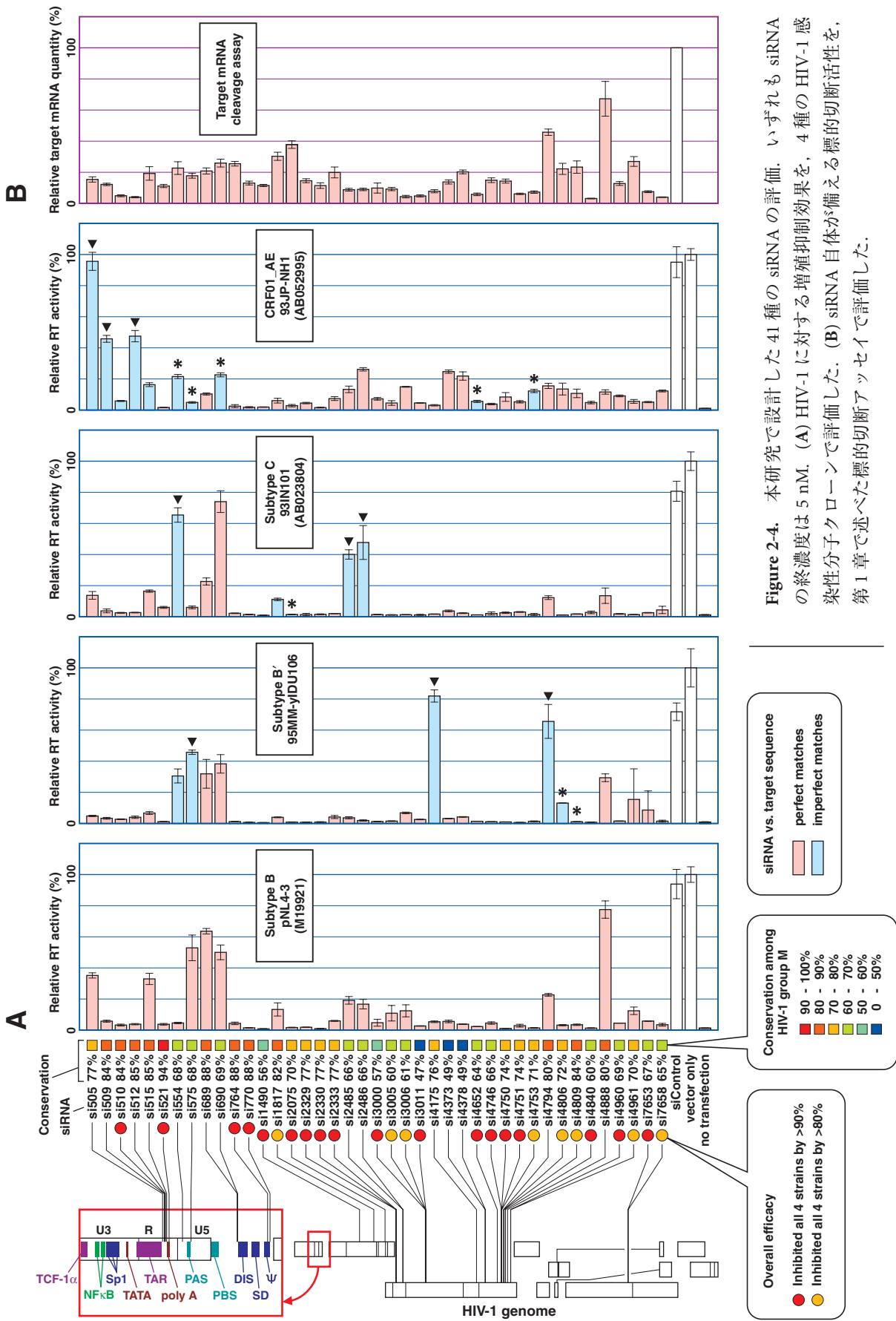
しかしながら、21-mer のうち特定の領域に位置するものは高度に保存されていることが判った (Fig. 2-1B)。これらの領域の多くはウイルスの遺伝子発現や増殖に必須な配列を含んでいた。すなわち、ウイルスの遺伝子発現に必要な TATA 配列やポリ A シグナル (AAUAAA)、逆転写に必要な primer activation signal (PAS) および primer binding site (PBS)、ウイルスのパッケージングに必要な packaging signal (Ψ)、ゲノムの + 鎖 DNA 合成に必要な central polypurine tract (cPPT)、central termination sequence (CTS) ならびに 3' polypurine tract (3' PPT) などが高度に保存されていた。これらはすべて、アミノ酸配列ではなく RNA の配列自体もしくは RNA の 2 次構造を通して機能を発揮するような領域であることは興味ある知見であった。実際、アミノ酸配列のレベルで保存されているような領域の多くは、コドンの 3 番目の塩基の変異を許容するため、塩基配列レベルでは必ずしも保存されていない例が多かった。

以上の解析により同定した、保存度が 70% 以上であるような 21-mer を Table S4 (巻末) に示した。この 216 種の siRNA 配列のうち 36 種が Ui-Tei らの siRNA 設計ガイドラインを満たし、哺乳類細胞で有効と予測された。また、Reynolds らの siRNA 設計法では 30 種、Amarzguioui らの設計法では 44 種がそれぞれ有効と予測された (Table S4, 巻末)。siRNA の抗 HIV-1 効果を検証するため、Table S4 に記載された Ui-Tei らのガイドラインを満たす 36 種の siRNA から、感染性分子クローン pNL4-3 (GenBank: M19921) と配列が完全一致するものを 23 種選択し、また、保存度が中程度で Ui-Tei らのガイドラインを満たし、かつ pNL4-3 と完全一致するものを 18 種選択した。選択した合計 41 種の siRNA の配列を Figure 2-3 に示した。



1-2. siRNA の抗 HIV-1 効果の検証

まず、siRNA 自体が備える標的切断活性を、第 1 章で述べた siRNA の標的切断活性を評価するアッセイ系（以下、**標的切断アッセイ**）で評価した。すなわち、各 siRNA の標的配列を mRNA の一部として発現するコンストラクトを構築し、siRNA とともに HeLa 細胞に導入した。ついで、そのコンストラクトから発現させた標的 mRNA の分解を、リアルタイム RT-PCR 法により定量した。その結果、各 siRNA を 5 nM で用いた場合、41 種の siRNA のうち 39 種が標的 mRNA を 40% 以下に抑制した（Fig. 2-4B）。si4794 と si4888 は有効と予測されたにもかかわらず、5 nM の濃度において十分な抑制を示さなかった。この理由は明らかでないが、si4794 は配列中に連続するグアニン残基（GGGGGG）および連続するシトシン残基（CCCCC）を含むため RNA オリゴの化学合成が難しく、また si4888 は配列中に AAAAUUUU という回文配列が存在し 2 次構造を取りやすいため、いずれも正しく塩基対を形成した 2 本鎖 siRNA の実効濃度が低下していた可能性がある（Ui-Tei *et al.*, 2004; Reynolds *et al.*, 2004）。



ついで、各 siRNA の HIV-1 に対する増殖抑制効果を、サブタイプ B, C, CRF01_AE 属する 4 種の HIV-1 感染性分子クローン (Table 2-2) で評価した。これら 3 種のサブタイプはアジア地域における HIV-1 流行の主要な原因となっており、また塩基配列のうえでは相互に隔たりのあるクラスタを形成している (Fig. 2-5, Takebe *et al.*, 2003)。本アッセイでは、それぞれの感染性分子クローンを siRNA とともに HeLa 細胞に導入し、48 時間後に培地上清を回収して HIV-1 由来の逆転写酵素活性を測定することにより、ウイルスの増殖を評価した。その結果、各 siRNA を 5 nM で導入した場合、41 種の siRNA のうち 26 種がすべての HIV-1 株の増殖を 20% 以下に抑制した (Fig. 2-4A, 赤およびオレンジの丸印で表示)。残りの 15 種の siRNA のうち si4794/4888 を除く 13 種は、標的切断アッセイにおいて活性が認められ (Fig. 2-4B), また si690/4794/4888 以外の 12 種は少なくとも 1 種の HIV-1 株の増殖を 20% 以下に抑制していることから、それらの siRNA 自体は有効であると考えられた。しかし Fig. 2-4A において矢印 (◄) で表示した HIV-1 株では、siRNA 配列と標的配列との間に 1 塩基～数塩基のミスマッチが存在するため (Fig. 2-6), それらの HIV-1 株の増殖を抑制できなかった。一方、塩基の置換が siRNA の末端に近い場合には siRNA の効果が保持される場合があった (Fig. 2-4A, * で表示)。このように末端に近い位置においてミスマッチの影響が少ないという結果は、第 1 章で述べたオフターゲット効果の解析結果 (Fig. 1-10, 1-11, 1-12) とおおむね一致した。また si689/690 については、siRNA 配列と標的配列とが完全に一致しているにもかかわらず HIV-1 の増殖を抑制できなかった (各感染性分子クローンの標的領域をシーケンスして配列が一致することを確認済み)。この領域は、HIV-1 の 5' UTR がとりうる 2 種類の 2 次構造 (Huthoff *et al.*, 2001), すなわち BMH (branched multiple hairpin; Fig. 2-7) と LDI (long distance interaction; Fig. 2-8) の両方において固く閉じており、siRNA が標的配列に接近することを妨げている可能性が示唆された。さらに、si575 の効果が pNL4-3 と 93IN101 とで異なる点が注目された。この理由は明らかでないが、ひとつの可能性として、各感染性分子クロー

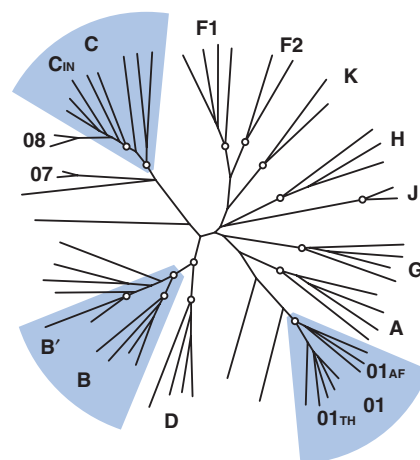


Figure 2-5. HIV-1 グループ M の分子系統樹 (Takebe *et al.*, 2003 より作成)。

| | | | | | |
|---------------|------------------------|---------------|-----------------------|---------------|-------------------------|
| si505 | GGGAACCCACTGCTTAAGCCT | si2075 | GACAGGCTAATTTTTAGGGA | si4746 | AAGACAGCAGTACAAATGGCA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | .G..... | 93IN101 | |
| 93JP-NH1 |AG.C | 93JP-NH1 | | 93JP-NH1 | |
| si509 | ACCCACTGCTTAA-GCCTCAAT | si2329 | CAGGAGCAGATGATACAGTAT | si4750 | CAGCAGTACAAATGGCAGTAT |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 |A..... | 93JP-NH1 | | 93JP-NH1 | |
| si510 | CCCACTGCTTAA-GCCTCAATA | si2330 | AGGAGCAGATGATACAGTATT | si4751 | AGCAGTACAAATGGCAGTATT |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 |A..... | 93JP-NH1 | | 93JP-NH1 | |
| si512 | CACTGCTTAA-GCCTCAATAAA | si2333 | AGCAGATGATACAGTATTAGA | si4753 | CAGTACAAATGGCAGTATTCA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 |A..... | 93JP-NH1 | | 93JP-NH1 |T..... |
| si515 | TGCTTAA-GCCTCAATAAAGCT | si2485 | GACCTACACCTGTCAACATAA | si4794 | GGGGGGATTGGGGGGTACAGT |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 |AC...T.. |
| 93IN101 | | 93IN101 | ...C...A..... | 93IN101 | |
| 93JP-NH1 |A..... | 93JP-NH1 | | 93JP-NH1 | |
| si521 | AGCCTCAATAAAGCTTGCCCT | si2486 | ACCTACACCTGTCAACATAAT | si4806 | GGGTACAGTGCAGGGGAAAGA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 |AC...T..... |
| 93IN101 | | 93IN101 | ...C...A..... | 93IN101 | |
| 93JP-NH1 | | 93JP-NH1 | | 93JP-NH1 | |
| si554 | TAGTGTGTGCCGCTGTGTGT | si3000 | CAGGGATGGAAAGGATCACCA | si4809 | TACAGTGCAGGGGAAAGAATA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | ...A..... | 95MM-yIDU106 | | 95MM-yIDU106 | ...T..... |
| 93IN101 | C.....A..... | 93IN101 | | 93IN101 | |
| 93JP-NH1 | .G.....GT. | 93JP-NH1 | | 93JP-NH1 | |
| si575 | GTGACTCTGGTAACAGAGAT | si3005 | ATGGAAAGGATCACCAGCAAT | si4840 | TAGCAACAGACATACAACTA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 |G..... | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 | AG..... | 93JP-NH1 | | 93JP-NH1 | |
| si689 | ACGCAGGACTCGGCTTGCTGA | si3006 | TGGAAAGGATCACCAGCAATA | si4888 | TTCAAAATTTTCGGGTTTATT |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 | | 93JP-NH1 | | 93JP-NH1 | |
| si690 | CGCAGGACTCGGCTTGCTGAA | si3011 | AGGATCACCAGCAATATCCA | si4960 | AAGGTGAAGGGGCAGTAGTAA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 |G..... | 93JP-NH1 | | 93JP-NH1 | |
| si764 | GACTAGCGGAGGCTAGAAGGA | si4175 | TGGAGGAAATGAACAAGTAGA | si4961 | AGGTGAAGGGGCAGTAGTAAT |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | ...G...A..... | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 | | 93JP-NH1 | | 93JP-NH1 | |
| si770 | CGGAGGCTAGAAGGAGAGAGA | si4373 | AGCCATGCATGGACAAGTAGA | si7653 | GACAAATTGAGAGAAGTGAATTA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 | | 93JP-NH1 | | 93JP-NH1 | |
| si1490 | GTGACATAGCAGGAACACTA | si4378 | TGCATGGACAAGTAGACTGTA | si7658 | TTGGAGAAGTGAATTATATAA |
| pNL4-3 | | pNL4-3 | | pNL4-3 | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | 95MM-yIDU106 | |
| 93IN101 | | 93IN101 | | 93IN101 | |
| 93JP-NH1 | | 93JP-NH1 | | 93JP-NH1 | |
| si1817 | TAGAAGAAATGATGACAGCAT | si4652 | TCCCTACAATCCCAAAGTCA | | |
| pNL4-3 | | pNL4-3 | | | |
| 95MM-yIDU106 | | 95MM-yIDU106 | | | |
| 93IN101 | ...G..... | 93IN101 | | | |
| 93JP-NH1 | | 93JP-NH1 | C..... | | |

Figure 2-6. 本研究で評価した 41 種の siRNA 配列と、各感染性分子クローンの標的配列とのアライメント。siRNA 配列はアンチセンス鎖の 21 塩基に対応するセンス側の配列を示した。いずれも RNA 分子であるが、塩基配列は GenBank に収録された各データにあわせ A, T, G, C の 4 文字を用いて記載した。

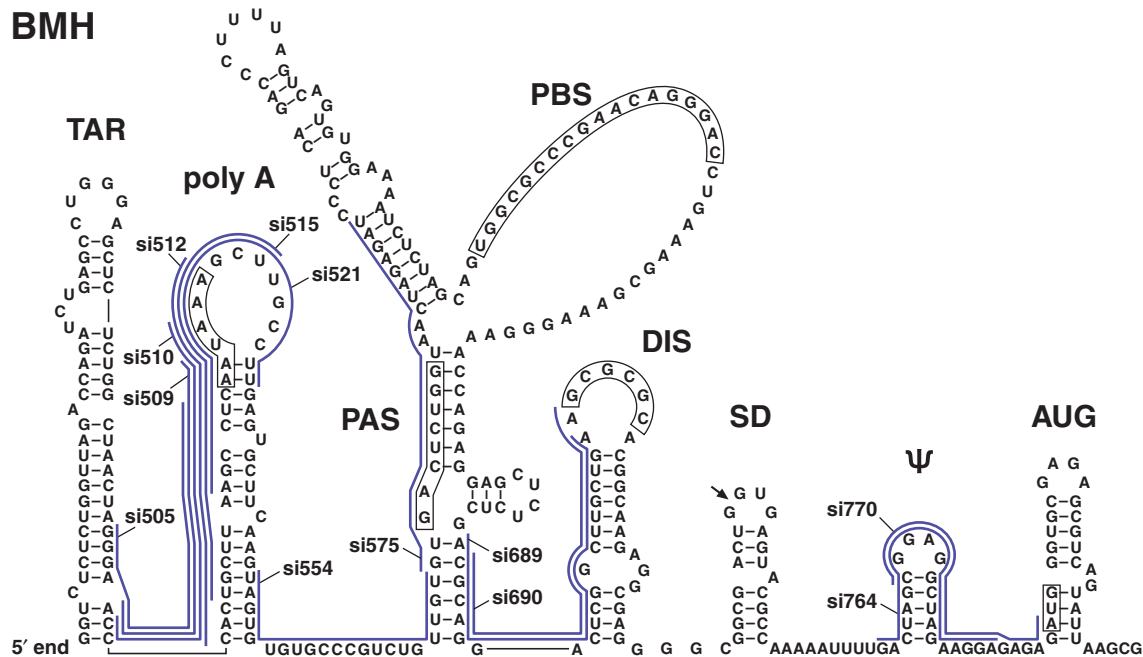


Figure 2-7. BMI (branched multiple hairpin) 型の 2 次構造. 本研究で評価した siRNA の標的部位を示した.

ンで標的部分の 2 次構造に差異があり siRNA の接近のしやすさが異なるのかも知れない. このような 2 次構造の影響を検討するひとつの方法として, 5' UTR 配列の全体を pTREC に挿入したコンストラクトを構築し, 標的切断アッセイをおこなう方法が有効と考えられる.

本研究で構築した抗ウイルス siRNA 設計法により, HIV-1 の高度な多様性に対応しうる siRNA を高い確率で設計できた. 本研究で実際に評価をおこなった siRNA のうち, 保存度が 70% 以上であり, かつウイルス増殖を 10% 以下に抑制できた siRNA は,

- ・ si521 (ポリ A シグナルを標的とする ; 保存度 94%)
- ・ si764/770 (Ψ ; 88%)
- ・ si510 (TAR- ポリ A シグナル ; 84%)
- ・ si2075 (ribosomal slip site; 70%)
- ・ si2329/2330/2333 (プロテアーゼ ; 77%)
- ・ si4750/4751/4753 (インテグラーゼ ; 71-74%)

であった. しかしながら, 1 種の siRNA 単独でデータベースに登録された全 HIV-1 を標的とすることは不可能であった (Fig. 2-9A). これを克服するひとつの方法は, 標的が異なる複数の siRNA を同時に用いることであろう. 本研究で設計し検証をおこなった siRNA は, 2 種を組み合わせると同時に用いることにより 99% 以上の HIV-1 に対応可能である (Fig. 2-9B). また本研究で構築した抗ウイルス siRNA 設計法は, HIV-1 のみならず, 高い多様性をもつ HCV, インフルエンザウイルス, SARS コロナウイルス等にも応用できると考えられる.

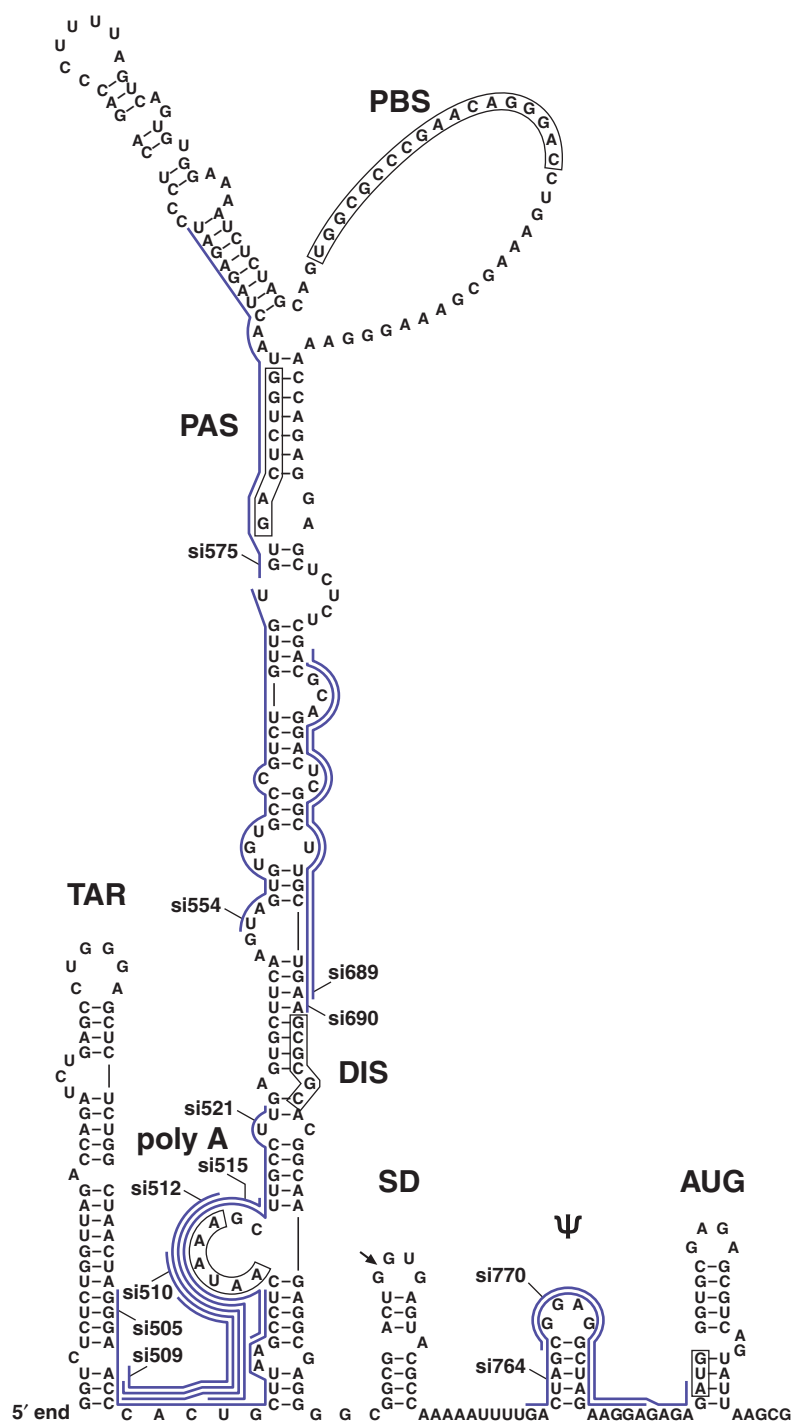


Figure 2-8. LDI (long distance interaction) 型の 2 次構造. 本研究で評価した siRNA の標的部
位を示した.

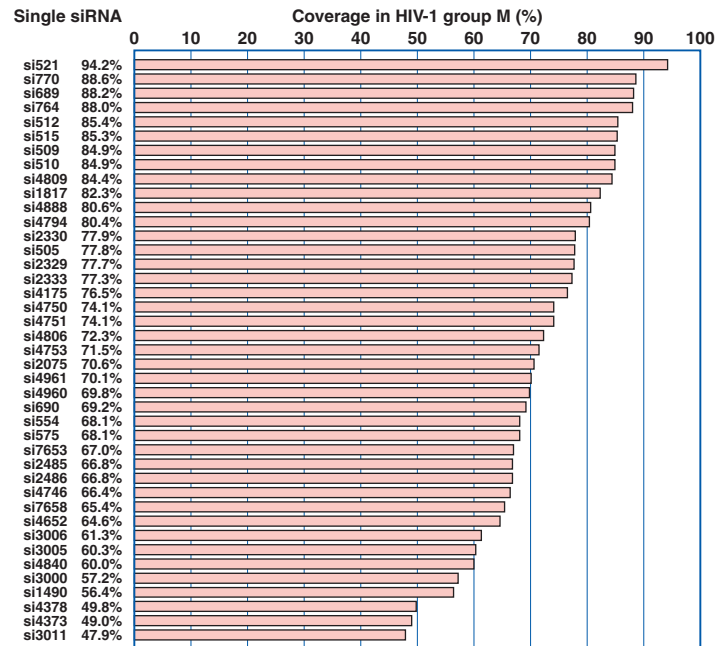
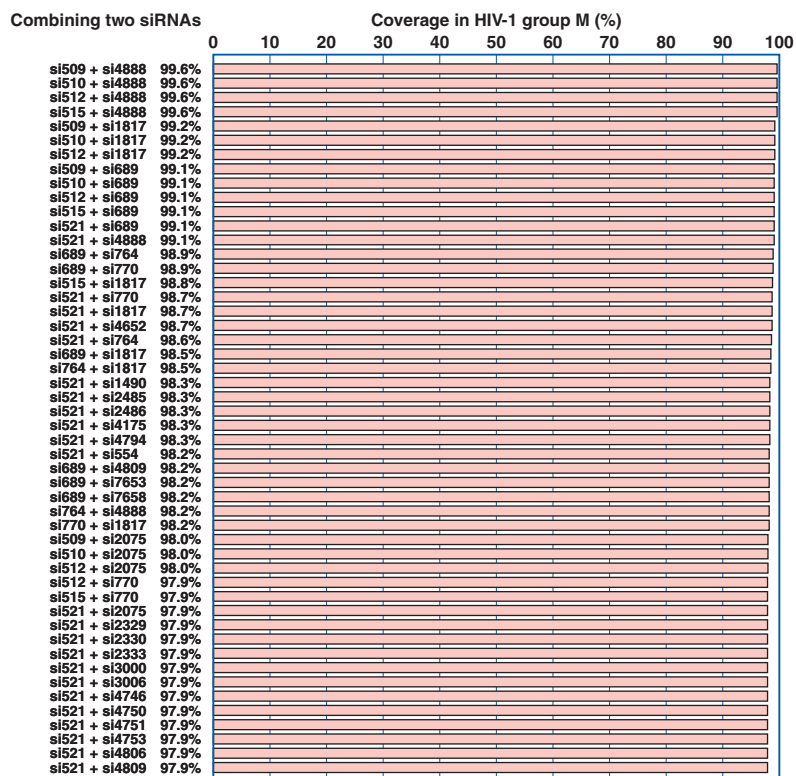
A**B**

Figure 2-9. (A) 本研究で評価した 41 種の siRNA を単独で使った場合に標的とすることが可能な HIV-1 の割合. 各 siRNA 配列の保存度に等しい. (B) 本研究で評価した 41 種の siRNA から 2 つを選択して同時に使った場合に標的とすることが可能な HIV-1 の割合.

1-3. 抗ウイルス siRNA 設計システムの構築

以上に述べた結果に基づき, (1) ウイルス配列の中で保存度が高く, (2) 哺乳類細胞に有効で, (3) オフターゲット効果の少ない siRNA を設計するシステムを構築した. 特に, HIV-1, HCV, A 型インフルエンザウイルス, SARS コロナウイルスを標的とする抗ウイルス siRNA 設計システムを, siVirus ウェブサーバ (<http://siVirus.RNAi.jp/>) として公開した (Naito *et al.*, 2006). siVirus ウェブサーバの概略は下記のとおりである.

▶ トップページ (Fig. 2-10)

HIV-1, HCV, A 型インフルエンザウイルス, SARS コロナウイルスから, siRNA を設計するウイルスを選択する. 各ウイルスの配列はそれぞれ下記のデータベースより入手した.

- ・ Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov/>)
- ・ Los Alamos HCV Sequence Database (Kuiken *et al.*, 2005; <http://hcv.lanl.gov/>)
- ・ NCBI Influenza Virus Sequence Database (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>)
- ・ GenBank (Benson *et al.*, 2007) より SARS コロナウイルスのゲノム 126 種を入手.

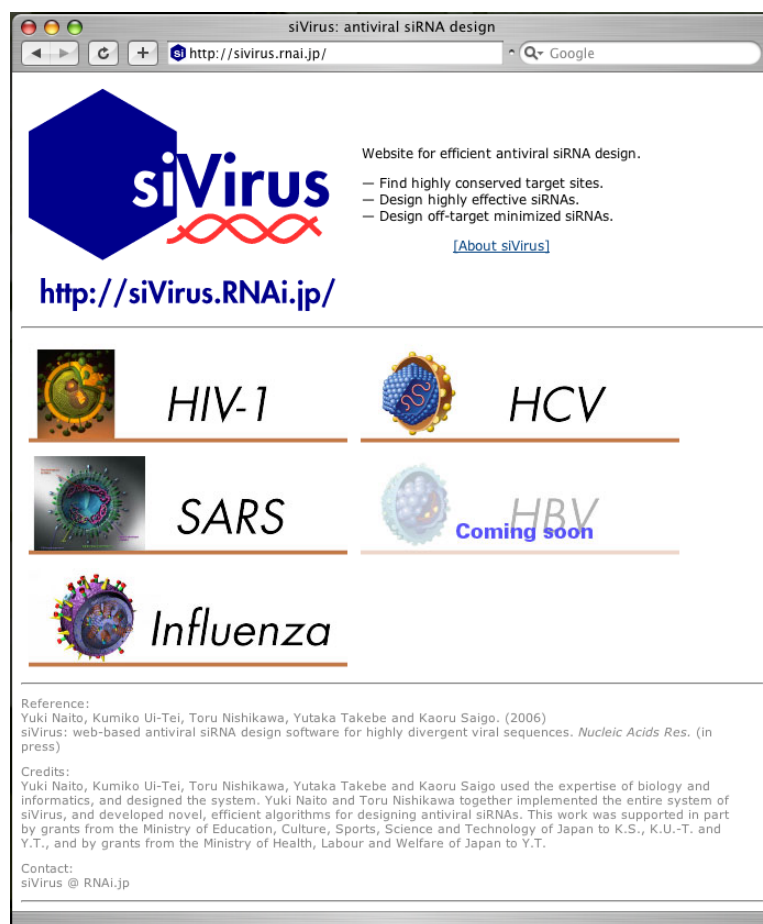


Figure 2-10. siVirus ウェブサーバ (<http://siVirus.RNAi.jp/>) のトップページ.

▶ ウイルス配列の選択画面 (Fig. 2-11)

保存度の計算に用いるウイルス配列の母集団を選択する。特定のサブタイプやサンプリング国に絞って保存度を求めることにより、そのサブタイプや地域でもっとも保存度の高い siRNA を設計できる (Fig. 2-11A)。サブタイプや地域を限定する必要のないときは、単にすべての配列を選択して siRNA 設計をおこなってもよい。ウイルス配列の選択は、チェックボックスによるほか、アクセッション番号のリストをテキスト形式で与えることもできる (Fig. 2-12)。最下部のオプションでは、siRNA の標的遺伝子または標的とする領域の範囲を指定できるほか、ユーザ独自のウイルス配列を FASTA フォーマットで入力できる (Fig. 2-11B)。

A

B

Options:

Target range: or Range: from [begin] to [end] [show landmark]

☒ Filter less effective siRNAs.

Paste in your own sequences no longer than 500000 letters:
(Multiple sequences may be queried if separated by lines starting with '>')

[submit]

Figure 2-11. siVirus による、HIV-1 を標的とする siRNA の設計画面の例。(A) ウイルス配列の選択画面および (B) オプション指定画面。

siVirus: HIV-1

http://sivirus.rnai.jp/HIV/from_text.html

siVirus top → HIV-1 :

- List by subtype
- List by sampling country
- Import list from text

→ Home
siVirus
http://sivirus.RNAi.jp/

HIV-1 - Import list from text

You can download the sequence list in a tab-delimited text, edit this list, and upload the list.

Full list (readonly):

| | | | | | | |
|----------|-------|----|----|----|------|---------|
| AF193275 | HIV-1 | A1 | A1 | BY | 1997 | 97BL006 |
| AF457052 | HIV-1 | A1 | A1 | KE | 2000 | KER2008 |
| AF457053 | HIV-1 | A1 | A1 | KE | 2000 | KER2009 |
| AF457055 | HIV-1 | A1 | A1 | KE | 2000 | KER2012 |
| AF457068 | HIV-1 | A1 | A1 | KE | 2000 | KNH1207 |
| AF457069 | HIV-1 | A1 | A1 | KE | 2000 | KNH1209 |
| AF457070 | HIV-1 | A1 | A1 | KE | 2000 | KNH1211 |

Edit the list, and paste in your list below:

Tips: Only the first word of each line (i.e., an accession number) is processed; anything followed by whitespace or tab is ignored in the calculation.

Options:

Target range: Any or Range: from begin to end [show landmark]

☒ Filter less effective siRNAs.

Paste in your own sequences no longer than 500000 letters:
(Multiple sequences may be queried if separated by lines starting with '>')

Submit

Figure 2-12. siVirus による、HIV-1 を標的とする siRNA の設計画面の例。ウイルス配列の選択は、チェックボックスによるほか、アクセッション番号のリストをテキスト形式で与えることもできる。

▶ siRNA 設計結果の表示 (Fig. 2-13)

ウイルス配列を選択したのち submit ボタンを押下すれば、直ちに結果が表示される (Fig. 2-13)。通常、Ui-Tei, Reynolds, または Amarzguoui の siRNA 設計法にしたがう、哺乳類細胞で有効と予測された siRNA のみが保存度の高い順番に表示される。結果の表示画面には下記の項目が表示される。

- siRNA target site (siRNA の標的配列およびウイルスゲノムにおける位置)
- siRNA efficacy prediction (哺乳類細胞における有効性の予測)
 - Ui-Tei (Ui-Tei *et al.*, 2004 の siRNA 設計法を満たす場合に yes と表示)
 - Reynolds (Reynolds *et al.*, 2004 の siRNA 設計法を満たす場合に yes と表示)
 - Amarzguoui (Amarzguoui *et al.*, 2004 の siRNA 設計法を満たす場合に yes と表示)
- ≤ 2 mismatch off-target hits (2 ミスマッチ以内で相同なヒト遺伝子の数)
- Conservation in the selected sequences (選択した配列における保存度)

なお、≤ 2 mismatch off-target hits の [detail] を押下すると、各 siRNA 配列と 3 ミスマッチ以内で相同なヒト遺伝子の一覧が表示される (Fig. 2-14)。この機能は第 1 章で述べた siDirect ウェブサーバにより実現している。

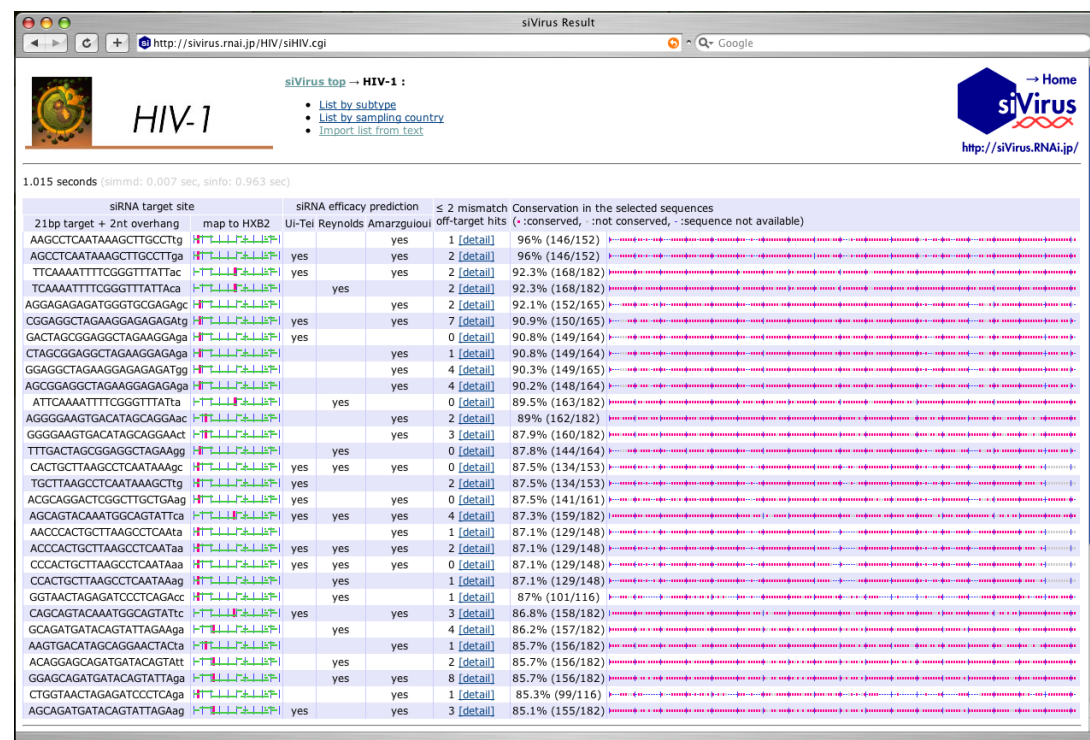


Figure 2-13. siVirus による siRNA の設計結果. HIV-1 のサブタイプ B, C, CRF01_AE の合計 182 種を選択して siRNA 設計をおこなった例.

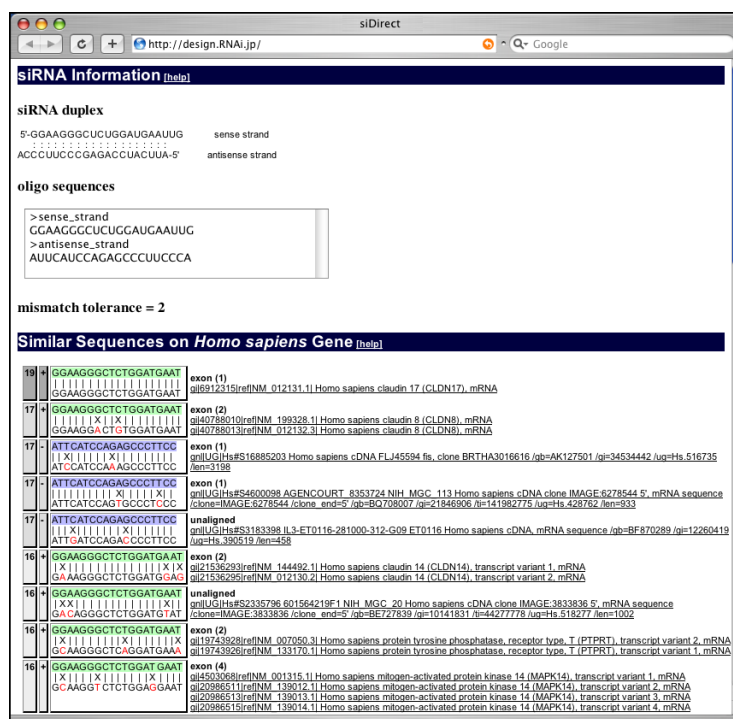


Figure 2-14. 各 siRNA 配列と 3 ミスマッチ以内で相同なヒト遺伝子の一覧.

2. 方 法

2-1. siRNA

化学合成の siRNA は、2 本鎖にアニーリングされたものを RNAi 社（東京）より購入した。siRNA の配列は **Figure 2-3** に示した。各 siRNA の名称は、HIV-1 ゲノム（HXB2; GenBank: K03455）における標的部位の座標に対応している。siControl は、Ui-Tei らのガイドライン（Ui-Tei *et al.*, 2004）を満たし、かつ、ヒトの全転写産物に 3 ミスマッチ以内で相同な配列が存在しないネガティブコントロールである。

2-2. 細胞の培養およびトランスフェクション

細胞の培養およびトランスフェクションは、第 1 章の方法にしたがった。

2-3. 標的切断アッセイ

第 1 章で構築した pTREC ベクターの EcoRI 部位と XhoI 部位との間に、各 siRNA の標的配列を挿入した。ベクターに挿入した標的配列部分はすべてシークエンスの確認をおこなった。各コンストラクトの構築に使用した DNA オリゴヌクレオチドの配列を **Table 2-1** に示した。各コンストラクト（0.5 μ g/well）を siRNA（終濃度 5 nM）とともに HeLa 細胞に導入し、24 時間後に RNeasy 96（Qiagen 社）で細胞を回収した。その他は第 1 章の方法にしたがった。

2-4. HIV-1 逆転写酵素活性の測定

Table 2-2 に示した各 HIV-1 感染性分子クローンを siRNA（終濃度 5 nM）とともに HeLa 細胞に導入し、48 時間後に培地上清を回収した。ついで、HIV-1 に由来する逆転写酵素活性を Willey らの方法（Willey *et al.*, 1988）にしたがって測定した。

2-5. siVirus ウェブサーバの構築

siVirus ウェブサーバは、Mac OS X が稼働する PowerMac G5 Dual 2GHz 主記憶 6GB（Apple 社）1 台の上に構築した。ユーザの要求に対して高速に応答できるよう、あらかじめ全ウイルス配列を構成する 21-mer をすべて生成し、そのひとつひとつの 21-mer がどのウイルス配列と完全一致するかを求め、データベース化した。このデータベースを参照し、ユーザが選択したウイルス配列の組み合わせにおいて保存度が高い 21-mer を短時間で求めるプログラム（siVirus core）を作成した。また、2 ミスマッチ以内で相同なヒト遺伝子の数を、あらかじめ siDirect（Naito, Yamada *et al.*, 2004）で算出してデータベース化した。

2-6. 共同研究に関する事項

Figure 2-4B の初期のデータについて、HeLa 細胞へのトランスフェクションは東京大学の程久美子博士がおこなった。

HIV-1 の増殖実験および逆転写酵素活性の測定 (Fig. 2-4A) は, 国立感染症研究所の武部 豊博士, 納富 香子氏, 小野木 利成氏がおこない, 著者が主体となって分析および検証をおこなった。

siVirus core の仕様は著者が決定し, 実装は東京大学の西川 徹氏がおこなった。なお, ウェブサーバ全体の実装は著者がおこなった。

| pTREC construct | Inserted DNA oligonucleotide duplex | siRNAs assayed |
|-----------------|---|-----------------------|
| pTREC-0505 | 5' -AATTCGGGAACCCACTGCTTAAGCCTCAATAAGCTTGCCTTGA-3' 3' -GCCCTTGGGTGACGAATTTCGGAGTTATTTTCGAACGGAACCTGAGCT-5' | si505/509/510/512/515 |
| pTREC-0521 | 5' -AATTCAGCCTCAATAAGCTTGCCTTGA-3' 3' -GTCGGAGTTATTTTCGAACGGAACCTGAGCT-5' | si521 |
| pTREC-0554 | 5' -AATTCAGTGTGTGCCCGTCTGTGTGTGACTCTGGTAAC-3' 3' -GATCACACACGGGAGACACACACTGAGACCATTTGGAGCT-5' | si554 |
| pTREC-0575 | 5' -AATTCGTGACTCTGGTAAGTACGATCC-3' 3' -GCACTGAGACCATTTGATCTCTAGGGAGCT-5' | si575 |
| pTREC-0689 | 5' -AATTCACGCAGGACTCGGCTTGTGTAAG-3' 3' -GTGCTCTCTGAGCCGAAACGACTTCGGAGCT-5' | si689/690 |
| pTREC-0764 | 5' -AATTCGACTAGCGGAGGCTAGAAGGAGAGAGATG-3' 3' -GCTGATCGCCTCCGATCTTCTCTCTACGAGCT-5' | si764/770 |
| pTREC-1490 | 5' -AATTCGTGACATAGCAGGAACCTACTAGTAC-3' 3' -GCACTGTATCGTCTTGTATGATCATGGAGCT-5' | si1490 |
| pTREC-1817 | 5' -AATTCAGAGAAGATGATGACAGCATGT-3' 3' -GATCTTCTTTACTACTGTCGTACAGAGCT-5' | si1817 |
| pTREC-2075 | 5' -AATTCGACAGGCTAATTTTTCAGGGAAGC-3' 3' -GCTGTCCGATTAATAAATCCCTTCGAGCT-5' | si2075 |
| pTREC-2323 | 5' -AATTCAGATACAGGAGCAGATGATACAGTATTAGAAGC-3' 3' -GATCTATGTCTCTGCTACTATGTCATAATCTTCGAGCT-5' | si2329/2330/2333 |
| pTREC-2482 | 5' -AATTCAGGACCTACACCTGTCAACATAAATGGAAGAAATCT-3' 3' -GATCTGTGATGTGGACAGTTGTATTAACCTTCTTTAGAGAGCT-5' | si2485/2486 |
| pTREC-2999 | 5' -AATTCACAGGGATGGAAGGATCACCAGCAATATCCAAAC-3' 3' -GTGTCCCTACCTTTCCCTAGTGGTCGTTATAAGGTTTGAGCT-5' | si3000/3005/3006/3011 |
| pTREC-4175 | 5' -AATTCAGGAGGAAATGAAACAGTAGATAAATC-3' 3' -GACCTCTTTACTTGTTCATCTATTTAGAGCT-5' | si4175 |
| pTREC-4373 | 5' -AATTCAGCCATGCAATGGACCAAGTAGACTGTAGTC-3' 3' -GTCGGTACGTACCTGTTCATCTGACATCAGAGCT-5' | si4373/4378 |
| pTREC-4652 | 5' -AATTCCTCCCTACAATCCCAAGTCAAGC-3' 3' -GAGGGATGTTAGGGGTTTCAGTTCGAGCT-5' | si4652 |
| pTREC-4746 | 5' -AATTCAGACAGCAGTACAAATGGCAGTATTTCATCA-3' 3' -GTTCTGTGTCATGTTTACCGTCATAAGTAGGAGCT-5' | si4746/4750/4751 |
| pTREC-4746 | 5' -AATTCAGACAGCAGTACAAATGGCAGTATTTCATCA-3' 3' -GTTCTGTGTCATGTTTACCGTCATAAGTAGGAGCT-5' | si4753 |
| pTREC-4794 | 5' -AATTCGGGGGATTTGGGGGTACAGTGCAGGGGAAAGAATAGTC-3' 3' -GCCCCCTTAACCCCAATGTCACGTCCCTTTCTTATCAGAGCT-5' | si4794/4806/4809 |
| pTREC-4840 | 5' -AATTCAGCAACAGACATACAAATCAAC-3' 3' -GATCGTTGTCGTATGTTTGTATTTGAGCT-5' | si4840 |
| pTREC-4888 | 5' -AATTCCTCAAAATTTTCGGGTTTATTAC-3' 3' -GAAGTTTAAAGCCCAATAATGGAGCT-5' | si4888 |
| pTREC-4959 | 5' -AATTCAGAGGTGAAGGGGAGTGTATATCAAGATAATAGTC-3' 3' -GTTTCCACTTCCCGTCATCATTTATTTATTCAGAGCT-5' | si4960/4961 |
| pTREC-7653 | 5' -AATTCGACAATTGGAGAGTGAATTATATAAATATAAAGC-3' 3' -GCTGTTAACCTCTTCACTTAATATATTTATATTTTCGAGCT-5' | si7653 |
| pTREC-7658 | 5' -AATTCCTGGAGAGTGAATTATATAAATATAAAGC-3' 3' -GAACCTCTTCACTTAATATATTTATATTTTCGAGCT-5' | si7658 |

Table 2-1. 抗 HIV-1 siRNA の標的切断アッセイのために構築した pTREC コンストラクトの一覧。挿入した DNA オリゴヌクレオチドの配列中, siRNA の標的部分を影付で示した。

| Molecular clone | Accession | Subtype/CRF | 導入した量 |
|-----------------|---|-------------|-------------|
| pNL4-3 | M19921 | subtype B | 0.2 µg/well |
| 95MM-yIDU106 | Takebe, Y. <i>et al.</i> (unpublished) | subtype B' | 1.0 µg/well |
| 93IN101 | AB023804 | subtype C | 1.0 µg/well |
| 93JP-NH1 | AB052995 | CRF01_AE | 1.0 µg/well |

Table 2-2. siRNA の評価に使用した HIV-1 感染性分子クローンの概要。

文 献

文献

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J.** Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
- Amarzguioui, M. & Prydz, H.** An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun.* **316**, 1050-1058 (2004).
- Banerjea, A., Li, M.-J., Bauer, G., Remling, L., Lee, N.-S., Rossi, J. & Akkina, R.** Inhibition of HIV-1 by lentiviral vector-transduced siRNAs in T lymphocytes differentiated in SCID-hu mice and CD34⁺ progenitor cell-derived macrophages. *Mol. Ther.* **8**, 62-71 (2003).
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L.** GenBank. *Nucleic Acids Res.* **35**, D21-D25 (2007).
- Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J.** Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366 (2001).
- Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., Marshall, W.S. & Khvorova, A.** 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3**, 199-204 (2006).
- Boden, D., Pusch, O., Lee, F., Tucker, L. & Ramratnam, B.** Human immunodeficiency virus type 1 escape from RNA interference. *J. Virol.* **77**, 11531-11535 (2003a).
- Boden, D., Pusch, O., Lee, F., Tucker, L., Shank, P.R. & Ramratnam, B.** Promoter choice affects the potency of HIV-1 specific RNA interference. *Nucleic Acids Res.* **31**, 5033-5038 (2003b).
- Boden, D., Pusch, O., Lee, F., Tucker, L. & Ramratnam, B.** Efficient gene transfer of HIV-1-specific short hairpin RNA into human lymphocytic cells using recombinant adeno-associated virus vectors. *Mol. Ther.* **9**, 396-402 (2004a).
- Boden, D., Pusch, O., Silbermann, R., Lee, F., Tucker, L. & Ramratnam, B.** Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. *Nucleic Acids Res.* **32**, 1154-1158 (2004b).
- Capodici, J., Karikó, K. & Weissman, D.** Inhibition of HIV-1 infection by small interfering RNA-mediated RNA interference. *J. Immunol.* **169**, 5196-5201 (2002).
- Chang, L.-J., Liu, X. & He, J.** Lentiviral siRNAs targeting multiple highly conserved RNA sequences of human immunodeficiency virus type 1. *Gene Ther.* **12**, 1133-1144 (2005).
- Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. & Shiekhattar, R.** TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-744 (2005).

- Coburn, G.A. & Cullen, B.R.** Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference. *J. Virol.* **76**, 9225-9231 (2002).
- Das, A.T., Brummelkamp, T.R., Westerhout, E.M., Vink, M., Madiredjo, M., Bernards, R. & Berkhout, B.** Human immunodeficiency virus type 1 escapes from RNA interference-mediated inhibition. *J. Virol.* **78**, 2601-2605 (2004).
- Dave, R.S. & Pomerantz, R.J.** Antiviral effects of human immunodeficiency virus type 1-specific small interfering RNAs against targets conserved in select neurotropic viral strains. *J. Virol.* **78**, 13687-13696 (2004).
- Doi, N., Zenno, S., Ueda, R., Ohki-Hamazaki, H., Ui-Tei, K. & Saigo, K.** Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr. Biol.* **13**, 41-46 (2003).
- Du, Q., Thonberg, H., Wang, J., Wahlestedt, C. & Liang, Z.** A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites. *Nucleic Acids Res.* **33**, 1671-1677 (2005).
- Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T.** Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498 (2001a).
- Elbashir, S.M., Lendeckel, W. & Tuschl, T.** RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188-200 (2001b).
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T.** Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877-6888 (2001c).
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. & Mello, C.C.** Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811 (1998).
- Grant, C.E., Vasa, M.Z. & Deeley, R.G.** cIRF-3, a new member of the interferon regulatory factor (IRF) family that is rapidly and transiently induced by dsRNA. *Nucleic Acids Res.* **23**, 2137-2146 (1995).
- Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. & Hannon, G.J.** Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146-1150 (2001).
- Han, W., Wind-Rotolo, M., Kirkman, R.L. & Morrow, C.D.** Inhibition of human immunodeficiency virus type 1 replication by siRNA targeted to the highly conserved primer binding site. *Virology* **330**, 221-232 (2004).
- Holen, T., Amarzguoui, M., Wiiger, M.T., Babaie, E. & Prydz, H.** Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor. *Nucleic Acids Res.* **30**, 1757-1766 (2002).
- Hu, W.-Y., Myers, C.P., Kilzer, J.M., Pfaff, S.L. & Bushman, F.D.** Inhibition of retroviral pathogenesis by RNA interference. *Curr. Biol.* **12**, 1301-1311 (2002).
- Huelsmann, P.M., Rauch, P., Allers, K., John, M.J. & Metzner, K.J.** Inhibition of drug-resistant HIV-1 by RNA interference. *Antiviral Res.* **69**, 1-8 (2006).

- Huthoff, H. & Berkhout, B.** Two alternating structures of the HIV-1 leader RNA. *RNA* **7**, 143-157 (2001).
- Jackson, A.L. & Linsley, P.S.** Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* **20**, 521-524 (2004).
- Jacque, J.-M., Triques, K. & Stevenson, M.** Modulation of HIV-1 replication by RNA interference. *Nature* **418**, 435-438 (2002).
- Joshi, P.J., North, T.W. & Prasad, V.R.** Aptamers directed to HIV-1 reverse transcriptase display greater efficacy over small hairpin RNAs targeted to viral RNA in blocking HIV-1 replication. *Mol. Ther.* **11**, 677-686 (2005).
- Kaufman, R.J.** Double-stranded RNA-activated protein kinase mediates virus-induced apoptosis: a new role for an old actor. *Proc. Natl. Acad. Sci. USA* **96**, 11693-11695 (1999).
- Kennerdell, J.R. & Carthew, R.W.** Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* **95**, 1017-1026 (1998).
- Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J. & Plasterk, R.H.A.** Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654-2659 (2001).
- Khvorova, A., Reynolds, A. & Jayasena, S.D.** Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-216 (2003).
- Kittler, R., Heninger, A.-K., Franke, K., Habermann, B. & Buchholz, F.** Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779-784 (2005).
- Kuiken, C., Yusim, K., Boykin, L. & Richardson, R.** The Los Alamos hepatitis C sequence database. *Bioinformatics* **21**, 379-384 (2005).
- Lee, M.-T.M., Coburn, G.A., McClure, M.O. & Cullen, B.R.** Inhibition of human immunodeficiency virus type 1 replication in primary macrophages by using Tat- or CCR5-specific small interfering RNAs expressed from a lentivirus vector. *J. Virol.* **77**, 11964-11972 (2003).
- Lee, N.-S., Dohjima, T., Bauer, G., Li, H., Li, M.-J., Ehsani, A., Salvaterra, P. & Rossi, J.** Expression of small interfering RNAs targeted against HIV-1 *rev* transcripts in human cells. *Nat. Biotechnol.* **20**, 500-505 (2002).
- Lee, S.-K., Dykxhoorn, D.M., Kumar, P., Ranjbar, S., Song, E., Maliszewski, L.E., François-Bongarçon, V., Goldfeld, A., Swamy, N.M., Lieberman, J. & Shankar, P.** Lentiviral delivery of short hairpin RNAs protects CD4 T cells from multiple clades and primary isolates of HIV. *Blood* **106**, 818-826 (2005).
- Lee, Y., Hur, I., Park, S.-Y., Kim, Y.-K., Suh, M.R., Kim, V.N.** The role of PACT in the RNA silencing pathway. *EMBO J.* **25**, 522-532 (2006).
- Leonard, J.N. & Schaffer, D.V.** Antiviral RNAi therapy: emerging approaches for hitting a moving target. *Gene Ther.* **13**, 532-540 (2006).

- Li, M.-J., Bauer, G., Michienzi, A., Yee, J.-K., Lee, N.-S., Kim, J., Li, S., Castanotto, D., Zaia, J. & Rossi, J.J. Inhibition of HIV-1 infection by lentiviral vectors expressing Pol III-promoted anti-HIV RNAs. *Mol. Ther.* **8**, 196-206 (2003).
- Li, M.-J., Kim, J., Li, S., Zaia, J., Yee, J.-K., Anderson, J., Akkina, R. & Rossi, J.J. Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy. *Mol. Ther.* **12**, 900-909 (2005).
- Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S. & Johnson, J.M. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773 (2005).
- Lin, S.-L. & Ying, S.-Y. Combinational therapy for potential HIV-1 eradication and vaccination. *Int. J. Oncol.* **24**, 81-88 (2004).
- Lin, X., Ruan, X., Anderson, M.G., McDowell, J.A., Kroeger, P.E., Fesik, S.W. & Shen, Y. siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.* **33**, 4527-4535 (2005).
- Ma, J.-B., Yuan, Y.-R., Meister, G., Pei, Y., Tuschl, T. & Patel, D.J. Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666-670 (2005).
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563-574 (2002).
- Naito, Y., Yamada, T., Ui-Tei, K., Morishita, S. & Saigo, K. siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.* **32**, W124-W129 (2004).
- Naito, Y., Yamada, T., Matsumiya, T., Ui-Tei, K., Saigo, K. & Morishita, S. dsCheck: highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. *Nucleic Acids Res.* **33**, W589-W591 (2005).
- Naito, Y., Ui-Tei, K., Nishikawa, T., Takebe, Y. & Saigo, K. siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. *Nucleic Acids Res.* **34**, W448-W450 (2006).
- Nishitsuji, H., Ikeda, T., Miyoshi, H., Ohashi, T., Kannagi, M. & Masuda, T. Expression of small hairpin RNA by lentivirus-based vector confers efficient and stable gene-suppression of HIV-1 on human cells including primary non-dividing cells. *Microbes Infect.* **6**, 76-85 (2004).
- Nishitsuji, H., Kohara, M., Kannagi, M. & Masuda, T. Effective suppression of human immunodeficiency virus type 1 through a combination of short- or long-hairpin RNAs targeting essential sequences for retroviral integration. *J. Virol.* **80**, 7658-7666 (2006).
- Novina, C.D., Murray, M.F., Dykxhoorn, D.M., Beresford, P.J., Riess, J., Lee, S.-K., Collman, R.G., Lieberman, J., Shankar, P. & Sharp, P.A. siRNA-directed inhibition of HIV-1 infection. *Nat. Med.* **8**, 681-686 (2002).

- Park, W.-S., Hayafune, M., Miyano-Kurosaki, N. & Takaku, H.** Specific HIV-1 env gene silencing by small interfering RNAs in human peripheral blood mononuclear cells. *Gene Ther.* **10**, 2046-2050 (2003).
- Parker, J.S., Roe, S.M. & Barford, D.** Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663-666 (2005).
- Paul, C.P., Good, P.D., Li, S.X.L., Kleihauer, A., Rossi, J.J. & Engelke, D.R.** Localized expression of small RNA inhibitors in human cells. *Mol. Ther.* **7**, 237-247 (2003).
- Pruitt, K.D., Tatusova, T. & Maglott, D.R.** NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61-D65 (2007).
- Pusch, O., Boden, D., Silberman, R., Lee, F., Tucker, L. & Ramratnam, B.** Nucleotide sequence homology requirements of HIV-1-specific short hairpin RNA. *Nucleic Acids Res.* **31**, 6444-6449 (2003).
- Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S. & Khvorova, A.** Rational siRNA design for RNA interference. *Nat. Biotechnol.* **22**, 326-330 (2004).
- Sabariegos, R., Giménez-Barcons, M., Tàpia, N., Clotet, B. & Martínez, M.A.** Sequence homology required by human immunodeficiency virus type 1 to escape from short interfering RNAs. *J. Virol.* **80**, 571-577 (2006).
- Scherer, L.J., Yildiz, Y., Kim, J., Cagnon, L., Heale, B. & Rossi, J.J.** Rapid assessment of anti-HIV siRNA efficacy using PCR-derived Pol III shRNA cassettes. *Mol. Ther.* **10**, 597-603 (2004).
- Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. & Zamore, P.D.** Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).
- Song, E., Lee, S.-K., Dykxhoorn, D.M., Novina, C., Zhang, D., Crawford, K., Cerny, J., Sharp, P.A., Lieberman, J., Manjunath, N. & Shankar, P.** Sustained small interfering RNA-mediated human immunodeficiency virus type 1 inhibition in primary macrophages. *J. Virol.* **77**, 7174-7181 (2003).
- Song, J.-J., Smith, S.K., Hannon, G.J. & Joshua-Tor, L.** Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434-1437 (2004).
- Surabhi, R.M. & Gaynor, R.B.** RNA interference directed against viral and cellular targets inhibits human immunodeficiency virus type 1 replication. *J. Virol.* **76**, 12963-12973 (2002).
- Suzuki, K., Shijuuku, T., Fukamachi, T., Zaunders, J., Guillemin, G., Cooper, D. & Kelleher, A.** Prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region. *J. RNAi Gene Silencing* **1**, 66-78 (2005).
- Takebe, Y., Motomura, K., Tatsumi, M., Lwin, H.H., Zaw, M. & Kusagawa, S.** High prevalence of diverse forms of HIV-1 intersubtype recombinants in Central Myanmar: geographical hot spot of extensive recombination. *AIDS* **17**, 2077-2087 (2003).
- ter Brake, O., Konstantinova, P., Ceylan, M. & Berkhout, B.** Silencing of HIV-1 with RNA interference: a multiple shRNA approach. *Mol. Ther.* **14**, 883-892 (2006).

- Ui-Tei, K., Zenno, S., Miyata, Y. & Saigo, K. Sensitive assay of RNA interference in *Drosophila* and Chinese hamster cultured cells using firefly luciferase gene as target. *FEBS Lett.* **479**, 79-82 (2000).
- Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R. & Saigo, K. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.* **32**, 936-948 (2004).
- Ui-Tei, K., Naito, Y. & Saigo, K. Guidelines for the selection of effective short-interfering RNA sequences for functional genomics. *Methods Mol. Biol.* **361**, 201-216 (2006).
- Unwalla, H.J., Li, M.-J., Kim, J.D., Li, H.T., Ehsani, A., Alluin, J. & Rossi, J.J. Negative feedback inhibition of HIV-1 by TAT-inducible expression of siRNA. *Nat. Biotechnol.* **22**, 1573-1578 (2004).
- Unwalla, H.J., Li, H.-T., Bahner, I., Li, M.-J., Kohn, D. & Rossi, J.J. Novel Pol II fusion promoter directs human immunodeficiency virus type 1-inducible coexpression of a short hairpin RNA and protein. *J. Virol.* **80**, 1863-1873 (2006).
- Wargelius, A., Ellingsen, S. & Fjose, A. Double-stranded RNA induces specific developmental defects in zebrafish embryos. *Biochem. Biophys. Res. Commun.* **263**, 156-161 (1999).
- Westerhout, E.M., Ooms, M., Vink, M., Das, A.T. & Berkhout, B. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res.* **33**, 796-804 (2005).
- Westerhout, E.M., ter Brake, O. & Berkhout, B. The virion-associated incoming HIV-1 RNA genome is not targeted by RNA interference. *Retrovirology* **3**, 57 (2006a).
- Westerhout, E.M., Vink, M., Haasnoot, P.C.J., Das, A.T. & Berkhout, B. A conditionally replicating HIV-based vector that stably expresses an antiviral shRNA against HIV-1 replication. *Mol. Ther.* **14**, 268-275 (2006b).
- Willey, R.L., Smith, D.H., Lasky, L.A., Theodore, T.S., Earl, P.L., Moss, B., Capon, D.J. & Martin, M.A. In vitro mutagenesis identifies a region within the envelope gene of the human immunodeficiency virus that is critical for infectivity. *J. Virol.* **62**, 139-147 (1988).
- Yamada, T. & Morishita, S. Accelerated off-target search algorithm for siRNA. *Bioinformatics* **21**, 1316-1324 (2005).
- Yamamoto, T., Miyoshi, H., Yamamoto, N., Yamamoto, N., Inoue, J. & Tsunetsugu-Yokota, Y. Lentivirus vectors expressing short hairpin RNAs against the U3-overlapping region of HIV *nef* inhibit HIV replication and infectivity in primary macrophages. *Blood* **108**, 3305-3312 (2006).
- Yoshinari, K., Miyagishi, M. & Taira, K. Effects on RNAi of the tight structure, sequence and position of the targeted region. *Nucleic Acids Res.* **32**, 691-699 (2004).
- Yuan, Y.-R., Pei, Y., Ma, J.-B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.-Y., Dauter, Z., Tuschl, T. & Patel, D.J. Crystal structure of *A. aeolicus* argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell* **19**, 405-419 (2005).

卷 末 表

Table S1

1/2

| siRNA name | Sense strand | Antisense strand | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguoui | | | Relative luc activity (%) |
|------------|------------------------|------------------------|---|-----|-----|------------------------------|
| siFL-5* | GAGGCCAAAAACAUAAGAAA | UCUUUAUGUUUUUGGCGUCUU | yes | yes | yes | 17 |
| siFL-6* | ACGCCAAAAACAUAAGAAA | UUCUUUAUGUUUUUGGCGUCU | | | yes | 7 |
| siFL-12* | AAAACAUAAAGAAAGGCCCGG | GGGCCUUUCUUUAUGUUUUUG | | | | 93 |
| siFL-14* | AACAUAAGAAAGGCCCGGCG | CCGGGCCUUUCUUUAUGUUUU | | | | 81 |
| siFL-18* | UAAAGAAAGGCCCGGCGCAU | GGCGCCGGGCCUUUCUUUAUG | | | | 93 |
| siFL-24* | AAGGCCCGGCGCAUUCUAUC | UAGAAUGGCGCCGGGCCUUUC | | | | 83 |
| siFL-29 | CCGGCGCAUUCUAUCCGCUG | GCGGAUAGAAUGGCGCCGGG | | | | 47 |
| siFL-30* | CGGCGCAUUCUAUCCGCUGG | AGCGGAUAGAAUGGCGCCGGG | | | yes | 54 |
| siFL-31 | GGCGCAUUCUAUCCGCUGGA | CAGCGGAUAGAAUGGCGCCGG | | | | 22 |
| siFL-33 | CGCAUUCUAUCCGCUGGAAG | UCCAGCGGAUAGAAUGGCGCC | | | yes | 14 |
| siFL-34 | GCCAUUCUAUCCGCUGGAAGA | UCCAGCGGAUAGAAUGGCGC | | | yes | 31 |
| siFL-35 | CCAUUCUAUCCGCUGGAAGAU | CUUCCAGCGGAUAGAAUGGCG | | | | 18 |
| siFL-36* | CAUUCUAUCCGCUGGAAGAUG | UCUUCAGCGGAUAGAAUGGC | yes | | | 11 |
| siFL-37 | AUUCUAUCCGCUGGAAGAUGG | AUCUUCAGCGGAUAGAAUGG | | | | 44 |
| siFL-40 | CUAUCCGCUGGAAGAUGGAAC | UCCAUCUCCAGCGGAUAGAA | yes | yes | | 9 |
| siFL-42* | AUCCGCUGGAAGAUGGAACCG | GUUCCAUCUCCAGCGGAUAG | | | | 87 |
| siFL-47 | CUGGAAGAUGGAACCGCUGGA | CAGCGGUUCCAUCUCCAGCG | | | | 17 |
| siFL-48* | UGGAAGAUGGAACCGCUGGAG | CCAGCGGUUCCAUCUCCAGC | | | | 81 |
| siFL-49 | GGAAGAUGGAACCGCUGGAGA | UCCAGCGGUUCCAUCUCCAG | | | yes | 8 |
| siFL-53 | GAUGGAACCGCUGGAGAGCAA | GCUCUCCAGCGGUUCCAUCU | | | | 17 |
| siFL-54* | AUGGAACCGCUGGAGAGCAAC | UGCUCUCCAGCGGUUCCAUCU | | | | 64 |
| siFL-55 | UGGAACCGCUGGAGAGCAACU | UUGCUCUCCAGCGGUUCCAUC | | | | 9 |
| siFL-179 | GUUCGGUUGGCAGAAGCUAUG | UAGCUUCUGCCAACCGAACGG | yes | | | 5 |
| siFL-184* | GUUGGCAGAAGCUAUGAAACG | UUUCAUAGCUUCUGCCAACCG | yes | yes | yes | 6 |
| siFL-257 | CAAUUCUUUAUGCCGUGUUG | ACACCGGCAUAAAGAAUUGAA | | | | 53 |
| siFL-265* | UAUGCCGUGUUGGGCGCGUU | CGGCCCCAACACCGGCAUAAA | | | | 75 |
| siFL-272* | GUGUUGGGCGCGUUAUUUAUC | UAAUAACGCGCCCAACACCG | yes | yes | yes | 13 |
| siFL-295* | AGUUGCAGUUGCGCCCGCGAA | CGCGGGCGCAACUGCAACUCC | | | | 68 |
| siFL-296 | GUUGCAGUUGCGCCCGCGAAC | UCGCGGGCGCAACUGCAACUC | | | | 23 |
| siFL-309* | CCGCGAACGACAUUUUAUAUG | UUUAUAAUGUCGUUCGCGGGC | yes | yes | yes | 6 |
| siFL-329 | GAACGUGAAUUGCUAACAGU | UGUUGAGCAAUUCAGUUCAU | yes | yes | | 7 |
| siFL-370 | CGUGGUGUUCGUUUCAAAAA | UUUGGAAACGAACACACCGGU | yes | | yes | 3 |
| siFL-387 | AAAAGGGGUGUGCAAAAAAUU | AUUUUUUGCAACCCUUUUUG | | yes | | 5 |
| siFL-411* | ACGUGCAAAAAAAGCUCCCAA | GGGAGCUUUUUUUGCACGUUC | | | | 72 |
| siFL-412 | CGUGCAAAAAAAGCUCCCAAU | UGGGAGCUUUUUUUGCACGUU | | | yes | 9 |
| siFL-421 | AAAGCUCCCAUCAUCCAAAA | UUGGAUGAUUGGGAGCUUUU | | yes | | 21 |
| siFL-423 | AGCUCCCAUCAUCCAAAAA | UUUUGGAUGAUUGGGAGCUUU | | yes | yes | 17 |
| siFL-428* | CCAAUCAUCCAAAAAUUAUU | UAAUUUUUUGGAUGAUUGGA | yes | yes | yes | 4 |
| siFL-431 | AUCAUCCAAAAAUUAUUUAUC | UAAUAAUUUUUUGGAUGAUUG | | yes | | 63 |
| siFL-507 | CUCAUCUACCUCGCGUUUA | AAACCGGAGGUAUGAGAU | | | yes | 13 |
| siFL-515* | CCUCCCGGUUUUAUGAAUAC | AUUCAUUAUAAACCGGGAGGUA | yes | yes | yes | 25 |
| siFL-575 | GCACUGAUGAUAACUCCUCU | AGGAGUUAUGAUCAGUGCAA | yes | | | 9 |
| siFL-583 | CAUGAACUCCUCUGGAUCUAC | AGAUCAGAGGAGUUAUGAU | yes | | yes | 26 |
| siFL-630 | CUCAUAGAACUGCCUGCGUGA | ACGCAGGCAGUUCUAUGAGGC | | | yes | 39 |
| siFL-655 | CUCGAUGCCAGAGAUCCUAU | AGGAUCUCUGGCAUGCGAGAA | yes | | yes | 5 |
| siFL-658* | GCAUGCCAGAGAUCCUAUUUU | AAUAGGAUCUCUGGCAUGCGA | yes | yes | yes | 5 |
| siFL-694 | UCCGAUAUCUGCGAUUUUAAG | UAAAAUCGCAGUAUCCGGAAU | | yes | yes | 2 |
| siFL-695* | CCGGAUAUCUGCGAUUUUAAGU | UUAAAAUCGCAGUAUCCGGAA | yes | yes | yes | 5 |
| siFL-721 | UCCAUCCAUCACGGUUUUGG | AAAACCGUGAUGGAAUGGAAC | | yes | | 12 |
| siFL-734* | GGUUUUGGAAUGUUUAUACA | UAGUAAACAUCCAAAACCGU | yes | | yes | 6 |
| siFL-749 | ACUACACUCGGAUAAUUGAUA | UCAAUAUCCGAGUGUAGUAA | | yes | | 3 |
| siFL-750 | CUACACUCGGAUAAUUGAUAU | AUCAAAUAUCCGAGUGUAGUA | yes | yes | | 8 |
| siFL-758 | GGAUAAUUGAUAUGUGGAUUU | AUCCACAUAUCAAUAUCCGA | yes | | yes | 2 |
| siFL-765 | UGAUAUGUGGAUUUCGAGUCG | ACUCGAAAUCCACAUAUCAA | | | | 19 |
| siFL-766 | GAUAUGUGGAUUUCGAGUCGU | GACUCGAAAUCCACAUAUCAA | | | | 50 |
| siFL-767 | AUAUGUGGAUUUCGAGUCGUC | CGACUCGAAAUCCACAUAUCA | | | | 38 |
| siFL-768 | UAUGUGGAUUUCGAGUCGUCU | ACGACUCGAAAUCCACAUAUC | | | | 25 |
| siFL-769 | AUGUGGAUUUCGAGUCGUCUU | GACGACUCGAAAUCCACAUAU | | | | 14 |
| siFL-770 | UGUGGAUUUCGAGUCGUCUUA | AGACGACUCGAAAUCCACAUA | | | | 13 |
| siFL-771 | GUGGAUUUCGAGUCGUCUUA | AAGACGACUCGAAAUCCACAUA | yes | | yes | 2 |

Table S1

2/2

| siRNA name | Sense strand | Antisense strand | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguioui | | | Relative luc activity (%) |
|------------|-----------------------|------------------------|--|-----|-----|---------------------------|
| siFL-772 | UGGAUUUCGAGUCGUCUUAU | UAAGACGACUCGAAAUCCACA | | yes | | 3 |
| siFL-773 | GGAUUUCGAGUCGUCUUAU | UUAAGACGACUCGAAAUCCAC | yes | yes | yes | 3 |
| siFL-774* | GAUUUCGAGUCGUCUUAU | AUUAAGACGACUCGAAAUCCA | yes | yes | yes | 1 |
| siFL-775 | AUUUCGAGUCGUCUUAU | CAUUAAGACGACUCGAAAUCC | | | | 37 |
| siFL-776 | UUUCGAGUCGUCUUAU | ACAUAAGACGACUCGAAAUC | | | | 31 |
| siFL-777 | UUCGAGUCGUCUUAU | UACAUUAAGACGACUCGAAA | | yes | | 7 |
| siFL-778 | UCGAGUCGUCUUAU | AUACAUUAAGACGACUCGAAA | | | | 15 |
| siFL-779 | CGAGUCGUCUUAU | UAUACAUUAAGACGACUCGAA | yes | yes | yes | 1 |
| siFL-780 | GAGUCGUCUUAU | CUAUACAUUAAGACGACUCGA | | | | 3 |
| siFL-781 | AGUCGUCUUAU | UCUAUACAUUAAGACGACUCG | | | | 35 |
| siFL-782 | GUCGUCUUAU | AUCUAUACAUUAAGACGACUC | yes | | yes | 54 |
| siFL-783 | UCGUCUUAU | AAUCUAUACAUUAAGACGACU | | | | 6 |
| siFL-857 | CUGGUGCCAACCCUAUUCUCC | AGAAUAGGGUUGGCACACGCA | yes | | yes | 24 |
| siFL-891* | GCACUCUGAUUGACAAAUACG | UAUUUGUCAAUUCAGAGUGCUU | yes | yes | yes | 5 |
| siFL-903 | ACAAAUACGAUUUAUCUAAU | UUAGAUAUAUCGUAUUUGUCA | | yes | | 14 |
| siFL-904* | CAAAUACGAUUUAUCUAAU | AUUAGAUAAUUCGUAUUUGUC | yes | yes | yes | 13 |
| siFL-998 | CCAGGUUAUCAGGCAAGGAU | AUCCUUGCCUGAUACCCUGGCA | yes | | yes | 5 |
| siFL-1044* | UUCUGAUUACACCCGAGGGGG | CCCUCGGGUGUAUUCAGAAUA | | | | 75 |
| siFL-1048 | GAUUAACCCGAGGGGGGAUGA | AUCCCCUCGGGUGUAUUCAG | | | | 6 |
| siFL-1054 | ACCCGAGGGGGAUGAUAAACC | UUUAUCAUCCCCUCGGGUGU | | yes | yes | 10 |
| siFL-1063 | GGAUUAUAAACCGGGCGCGGU | CGCGCCCGGUUAUCAUCCCC | | | | 6 |
| siFL-1064* | GAUGAUAAACCGGGCGCGGUC | CCGCGCCCGGUUAUCAUCCCC | | | | 46 |
| siFL-1065 | AUGAUAAACCGGGCGCGGUCG | ACCGCGCCCGGUUAUCAUCC | | | | 15 |
| siFL-1074 | CGGGCGCGGUCGGUAAAGUUG | ACUUUACCGACCGCGCCCGGU | yes | | yes | 5 |
| siFL-1075* | GGGCGCGGUCGGUAAAGUUGU | AACUUUACCGACCGCGCCCGG | yes | | yes | 8 |
| siFL-1076 | GGCGCGGUCGGUAAAGUUGUU | CAACUUUACCGACCGCGCCCG | | | | 3 |
| siFL-1081 | GGUCGGUAAAGUUGUCCAUU | UGGAACAACUUUACCGACCGC | yes | | | 2 |
| siFL-1085* | GGUAAAGUUGUCCAUUUUUU | AAAAGGGAACAACUUUACCGA | yes | | yes | 5 |
| siFL-1110 | CGAAGGUUGUGGAUCUGGAUA | UCCAGAUCCACAACCUUCGCU | yes | yes | | 1 |
| siFL-1111 | GAAGGUUGUGGAUCUGGAUAC | AUCCAGAUCCACAACCUUCGC | yes | | | 4 |
| siFL-1120 | GGAUCUGGAUACCGGAAAAC | UUUCCCGGAUCCAGAUCCAC | | yes | yes | 6 |
| siFL-1137 | AAACGUCGGCGUUAUCAA | UGAUUAACGCCAGCGUUUUC | | yes | | 13 |
| siFL-1186* | GAUUAUGUCCGGUUAUGUAAA | UACAUAAACCGGACAUAAUCAU | yes | yes | | 2 |
| siFL-1190 | AUGUCCGGUUAUGUAAACA | UGUUUACAUAAACCGGACAUAA | | yes | | 103 |
| siFL-1228 | CUUGAUUGACAAGGAUGGAUG | UCCAUCCUUGUCAAUCAAGGC | | | | 26 |
| siFL-1234 | UGACAAGGAUGGAUGGCUACA | UAGCCAUCCAUCCUUGUCAAU | | yes | | 3 |
| siFL-1308* | GCCUGAAGUCUCUGAUUAAGU | UUAAUCAGAGACUUCAGGCGG | yes | yes | yes | 4 |
| siFL-1346 | GCUCCCGCUGAAUUGGAAUCC | AUUCCAAUUCAGCGGAGCCA | yes | | yes | 3 |
| siFL-1352 | GCUGAAUUGGAAUCCAUUUUG | AGAUGGAUUCCAAUUCAGCGG | yes | | yes | 11 |
| siFL-1429 | CGGUGAACUCCCGCCGCCGU | GGCGGCGGGAAGUUCACCGGC | | | | 21 |
| siFL-1430* | GGUGAACUCCCGCCGCCGUU | CGGCGGCGGGAAGUUCACCGG | | | | 82 |
| siFL-1431 | GUGAACUCCCGCCGCCGUUG | ACGGCGGCGGGAAGUUCACCG | | | | 16 |
| siFL-1431 | GUGAACUCCCGCCGCCGUUG | ACGGCGGCGGGAAGUUCACCG | | | | 18 |
| siFL-1439* | CCCGCCCGCGUUGUUGUUUUG | AAACAACAACCGCGCGGGAA | | | yes | 62 |
| siFL-1440* | CCGCCCGCGUUGUUGUUUUGG | AAAACAACAACCGCGCGGGGA | yes | | yes | 33 |
| siFL-1441* | CGCCCGCGUUGUUGUUUUGGA | CAAAACAACAACCGCGCGGGG | | | | 41 |
| siFL-1482 | AAAAAGAGAUCGUGGAUUACG | UAAUCCACGAUCUUUUUUCC | | yes | | 3 |
| siFL-1509 | GUCAAGUAACAACCGGAAAA | UUCGCGGUUGUUAUUGACUG | | | yes | 42 |
| siFL-1512 | AAGUAACAACCGCGAAAAAGU | UUUUUCGCGGUUGUUAUUGA | | yes | yes | 12 |
| siFL-1522 | CGCGAAAAAGUUGCGCGGAGG | UCCGCGCAACUUUUUCGCGGU | | | yes | 56 |
| siFL-1539 | GAGGAGUUGUUGUUGGACG | UCCACAAACACAACUCCUCCG | yes | | | 2 |
| siFL-1565 | CCGAAAGGUCUUAACCGAAAA | UUCCGGUAAGACCUUUCGGUA | | | yes | 12 |
| siFL-1586 | CUCGACGCAAGAAAAUACAGA | UGAUUUUUUUCUGCGUGAGUU | yes | yes | yes | 12 |
| siFL-1601 | AUCAGAGAGAUCCUCAUAAAG | UUUAUGAGGAUCUCUCUGAUUU | | yes | yes | 6 |

Table S2

1/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|------------------|---------|---------------------------|------------------|
| AF193275 | 97BL006 | A1 | BY | 1997 |
| AF457052 | KER2008 | A1 | KE | 2000 |
| AF457053 | KER2009 | A1 | KE | 2000 |
| AF457055 | KER2012 | A1 | KE | 2000 |
| AF457068 | KNH1207 | A1 | KE | 2000 |
| AF457069 | KNH1209 | A1 | KE | 2000 |
| AF457070 | KNH1211 | A1 | KE | 2000 |
| AF457077 | KSM4024 | A1 | KE | 2000 |
| AF457079 | KSM4030 | A1 | KE | 2000 |
| AF457080 | MSA4069 | A1 | KE | 2000 |
| AF457081 | MSA4070 | A1 | KE | 2000 |
| AF457083 | MSA4072 | A1 | KE | 2000 |
| AF457084 | MSA4076 | A1 | KE | 2000 |
| AF457086 | MSA4079 | A1 | KE | 2000 |
| AF457089 | NKU3005 | A1 | KE | 2000 |
| AF539405 | ML170_1986 | A1 | KE | 1986 |
| AF004885 | Q23_17 | A1 | KE | 1994 |
| AF457063 | KNH1088 | A1 | KE | 1999 |
| AF457065 | KNH1135 | A1 | KE | 1999 |
| AF457075 | KSM4021 | A1 | KE | 1999 |
| AF069670 | SE7253 | A1 | SE | 1994 |
| AF069671 | SE7535 | A1 | SE | 1994 |
| AF069669 | SE8538 | A1 | SE | 1995 |
| AF069673 | SE8891 | A1 | SE | 1995 |
| AF107771 | UGSE8131 | A1 | SE | 1995 |
| AF361872 | 97TZ02 | A1 | TZ | 1997 |
| AF361873 | 97TZ03 | A1 | TZ | 1997 |
| AF413987 | 98UA0116 | A1 | UA | 2000 |
| AB098333 | UG029 | A1 | UG | - |
| AB098331 | UG031 | A1 | UG | - |
| M62320 | U455 | A1 | UG | 1985 |
| U51190 | 92UG037 | A1 | UG | 1992 |
| AF484507 | 98UG57134 | A1 | UG | 1998 |
| AF484508 | 98UG57135 | A1 | UG | 1998 |
| AF484509 | 98UG57136 | A1 | UG | 1998 |
| AF484512 | 98UG57142 | A1 | UG | 1998 |
| AF484478 | 99UGA07072 | A1 | UG | 1999 |
| AF484493 | 99UGG03379 | A1 | UG | 1999 |
| AF286240 | 97CDKFE4 | A2 | CD | 1997 |
| AF286241 | 97CDKS10 | A2 | CD | 1997 |
| AF286238 | 97CDKTB48 | A2 | CD | 1997 |
| AF286237 | 94CY017_41 | A2 | CY | 1994 |
| U23487 | contaminant_MANC | B | - | - |
| AY037269 | ARMS008 | B | AR | 2000 |
| AY037268 | ARCH054 | B | AR | 1998 |
| AY037282 | ARMA132 | B | AR | 1999 |
| AF538302 | 1181 | B | AU | - |
| AF538303 | 1182 | B | AU | - |
| AF538304 | C24 | B | AU | - |
| AF538305 | C42 | B | AU | - |

Table S2

2/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|----------|---------|---------------------------|------------------|
| AF538306 | C76 | B | AU | - |
| AF538307 | C92 | B | AU | - |
| AF146728 | VH | B | AU | - |
| AF042100 | MBC200 | B | AU | 1986 |
| AF042101 | MBC925 | B | AU | 1987 |
| AF042103 | MBCC54 | B | AU | 1995 |
| AF042104 | MBCC98 | B | AU | 1996 |
| AF042105 | MBCD36 | B | AU | 1996 |
| AY037270 | BOL0122 | B | BO | 1999 |
| AY173956 | BZ167 | B | BR | 1989 |
| AY180905 | CNHN24 | B | CN | - |
| U71182 | RL42 | B | CN | - |
| U43096 | D31 | B | DE | 1986 |
| U43141 | HAN | B | DE | 1986 |
| AY173959 | EC003 | B | EC | 1989 |
| AY173960 | EC102 | B | EC | 1989 |
| AF256210 | S61K15 | B | ES | 1989 |
| K03455 | HXB2 | B | FR | 1983 |
| M26727 | OYI | B | GA | 1988 |
| D10112 | CAM1 | B | GB | 1983 |
| AJ271445 | GB8 | B | GB | 1986 |
| AF224507 | WK | B | KR | 1997 |
| AB097870 | mSTD101 | B | MM | 1999 |
| U34604 | 3202A21 | B | NL | 1986 |
| AY173951 | BK132 | B | TH | 1990 |
| AF086817 | TWCYS | B | TW | 1994 |
| AY268493 | Ac_06 | B | US | 2001 |
| M17451 | RF | B | US | 1983 |
| K02007 | SF2 | B | US | 1983 |
| M17449 | MNCG | B | US | 1984 |
| M38431 | NY5CG | B | US | 1984 |
| AY352275 | SF33 | B | US | 1984 |
| AF004394 | AD87 | B | US | 1986 |
| U63632 | JRFL | B | US | 1986 |
| M93258 | YU2 | B | US | 1986 |
| L02317 | BC | B | US | 1987 |
| AF286365 | WR27 | B | US | 1988 |
| U39362 | P896 | B | US | 1989 |
| AY173952 | US1 | B | US | 1990 |
| AY173953 | US2 | B | US | 1990 |
| AY173954 | US3 | B | US | 1990 |
| AY173955 | US4 | B | US | 1990 |
| U69591 | WCIPR | B | US | 1990 |
| U21135 | WEAU160 | B | US | 1990 |
| AF069140 | DH12_3 | B | US | 1991 |
| AY308761 | 1299_d22 | B | US | 1996 |
| AY308760 | 1333_d2 | B | US | 1996 |
| AB078005 | ARES2 | B | US | 1997 |
| AY247251 | Ac_06 | B | US | 1997 |
| U52953 | 92BR025 | C | BR | 1992 |

Table S2

3/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|-------------|---------|---------------------------|------------------|
| AF286228 | 98BR004 | C | BR | 1998 |
| AF443088 | 00BW07621 | C | BW | 2000 |
| AF443089 | 00BW076820 | C | BW | 2000 |
| AF443090 | 00BW087421 | C | BW | 2000 |
| AF443091 | 00BW147127 | C | BW | 2000 |
| AF443092 | 00BW16162 | C | BW | 2000 |
| AF443093 | 00BW1686 | C | BW | 2000 |
| AF443094 | 00BW17593 | C | BW | 2000 |
| AF443095 | 00BW17732 | C | BW | 2000 |
| AF443096 | 00BW17835 | C | BW | 2000 |
| AF443097 | 00BW17956 | C | BW | 2000 |
| AF443098 | 00BW18113 | C | BW | 2000 |
| AF443099 | 00BW18595 | C | BW | 2000 |
| AF443100 | 00BW18802 | C | BW | 2000 |
| AF443101 | 00BW192113 | C | BW | 2000 |
| AF443102 | 00BW20361 | C | BW | 2000 |
| AF443103 | 00BW20636 | C | BW | 2000 |
| AF443104 | 00BW20872 | C | BW | 2000 |
| AF443105 | 00BW2127214 | C | BW | 2000 |
| AF443106 | 00BW21283 | C | BW | 2000 |
| AF443107 | 00BW22767 | C | BW | 2000 |
| AF443108 | 00BW38193 | C | BW | 2000 |
| AF443109 | 00BW38428 | C | BW | 2000 |
| AF443110 | 00BW38713 | C | BW | 2000 |
| AF443111 | 00BW3876_9 | C | BW | 2000 |
| AF443112 | 00BW3886_8 | C | BW | 2000 |
| AF443113 | 00BW3891_6 | C | BW | 2000 |
| AF443114 | 00BW3970_2 | C | BW | 2000 |
| AF443115 | 00BW5031_1 | C | BW | 2000 |
| AF110959 | 96BW01B03 | C | BW | 1996 |
| AF110962 | 96BW0402 | C | BW | 1996 |
| AF110967 | 96BW0502 | C | BW | 1996 |
| AF290028 | 96BW06 | C | BW | 1996 |
| AF110969 | 96BW1104 | C | BW | 1996 |
| AF110972 | 96BW1210 | C | BW | 1996 |
| AF110974 | 96BW15C02 | C | BW | 1996 |
| AF110976 | 96BW16B01 | C | BW | 1996 |
| AF110980 | 96BW17 | C | BW | 1996 |
| AF443075 | 96BWM032 | C | BW | 1996 |
| AF443074 | 96BWMO1_5 | C | BW | 1996 |
| AF443076 | 98BWMC122 | C | BW | 1998 |
| AF443077 | 98BWMC134 | C | BW | 1998 |
| AF443078 | 98BWMC14A3 | C | BW | 1998 |
| AF443079 | 98BWMO1410 | C | BW | 1998 |
| AF443080 | 98BWMO18D5 | C | BW | 1998 |
| AF443081 | 98BWMO36A5 | C | BW | 1998 |
| AF443082 | 98BWMO37D5 | C | BW | 1998 |
| AF443083 | 99BW393212 | C | BW | 1999 |
| AF443084 | 99BW46424 | C | BW | 1999 |
| AF443085 | 99BW4745 | C | BW | 1999 |

Table S2

4/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|------------|---------|---------------------------|------------------|
| AF443086 | 99BW47547 | C | BW | 1999 |
| AF443087 | 99BWMC168 | C | BW | 1999 |
| U46016 | ETH2220 | C | ET | 1986 |
| AF286233 | 98IS002 | C | IL | 1998 |
| AB023804 | 93IN101 | C | IN | 1993 |
| AF067157 | 93IN904 | C | IN | 1993 |
| AF067158 | 93IN905 | C | IN | 1993 |
| AF067154 | 93IN999 | C | IN | 1993 |
| AF067159 | 94IN11246 | C | IN | 1994 |
| AF286223 | 94IN476 | C | IN | 1994 |
| AF067155 | 95IN21068 | C | IN | 1995 |
| AF286231 | 98IN012 | C | IN | 1998 |
| AF286232 | 98IN022 | C | IN | 1998 |
| AY049708 | 01IN565_10 | C | IN | 1999 |
| AF457054 | KER2010 | C | KE | 2000 |
| AB097871 | mlDU101_3 | C | MM | 1999 |
| AF361874 | 97TZ04 | C | TZ | 1997 |
| AF361875 | 97TZ05 | C | TZ | 1997 |
| AF286234 | 98TZ013 | C | TZ | 1998 |
| AF286235 | 98TZ017 | C | TZ | 1998 |
| AY463228 | J38MA | C | ZA | 2000 |
| AY463223 | J54Ma | C | ZA | 2001 |
| AY118165 | 97ZA003 | C | ZA | 1997 |
| AY118166 | 97ZA009 | C | ZA | 1997 |
| AF286227 | 97ZA012 | C | ZA | 1997 |
| AY158533 | 98ZA445 | C | ZA | 1998 |
| AY158534 | 98ZA502 | C | ZA | 1998 |
| AY158535 | 98ZA528 | C | ZA | 1998 |
| AY162223 | TV001 | C | ZA | 1998 |
| AY162224 | TV002 | C | ZA | 1998 |
| AY162225 | TV012 | C | ZA | 1998 |
| AF411967 | 99ZACM9 | C | ZA | 1999 |
| AY043176 | CTSC2 | C | ZA | 1999 |
| AY043173 | DU151 | C | ZA | 1999 |
| AY043175 | DU422 | C | ZA | 1999 |
| AY043174 | DU179 | C | ZA | 1999 |
| AY463229 | 1119MB | C | ZA | 1999 |
| AY463217 | 1134MB | C | ZA | 2000 |
| AY463218 | 1176MB | C | ZA | 2000 |
| AY463219 | 1192M3M | C | ZA | 2000 |
| AY463220 | 1195MB | C | ZA | 2000 |
| AY463221 | 1210MB | C | ZA | 2000 |
| AY463222 | 1228MB | C | ZA | 2000 |
| AY463224 | 1162MB | C | ZA | 2000 |
| AY463225 | 1170MB | C | ZA | 2000 |
| AY463236 | 1214MB | C | ZA | 2000 |
| AY463226 | 1217MB | C | ZA | 2000 |
| AY463227 | 1225MB | C | ZA | 2000 |
| AY463230 | 1165MB | C | ZA | 2000 |
| AY463231 | 1168MB | C | ZA | 2000 |

Table S2

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|--------------|---------|---------------------------|------------------|
| AY463232 | 1171MB | C | ZA | 2000 |
| AY463233 | 1178MB | C | ZA | 2000 |
| AY463234 | 1197MB | C | ZA | 2000 |
| AY463235 | 2004MB | C | ZA | 2001 |
| AY463237 | 2134MB | C | ZA | 2001 |
| AF411966 | ZASW7 | C | ZA | 1999 |
| AF286224 | 96ZM651 | C | ZM | 1996 |
| AF286225 | 96ZM751 | C | ZM | 1996 |
| K03454 | ELI | D | CD | 1983 |
| M27323 | NDK | D | CD | 1983 |
| U88822 | 84ZR085 | D | CD | 1984 |
| M22639 | Z2Z6 | D | CD | 1985 |
| AY371155 | 01CM_0009BBY | D | CM | 2001 |
| AY371156 | 01CM_0175BA | D | CM | 2001 |
| AY371157 | 01CM_4412HAL | D | CM | 2001 |
| AF133821 | MB2059 | D | KE | - |
| AF457090 | NKU3006 | D | KE | 2001 |
| AJ488926 | MN011 | D | TD | 1999 |
| AJ488927 | MN012 | D | TD | 1999 |
| AJ320484 | 92UG001 | D | UG | 1992 |
| U88824 | 94UG114 | D | UG | 1994 |
| AF484502 | 98UG57128 | D | UG | 1998 |
| AF484504 | 98UG57130 | D | UG | 1998 |
| AF484505 | 98UG57131 | D | UG | 1998 |
| AF484506 | 98UG57132 | D | UG | 1998 |
| AF484511 | 98UG57140 | D | UG | 1998 |
| AF484514 | 98UG57143 | D | UG | 1998 |
| AF484513 | 98UG57146 | D | UG | 1998 |
| AF484518 | 99UGA03349 | D | UG | 1999 |
| AF484477 | 99UGA07412 | D | UG | 1999 |
| AY304496 | 99UGA08483 | D | UG | 1999 |
| AF484480 | 99UGB21875 | D | UG | 1999 |
| AF484481 | 99UGB25647 | D | UG | 1999 |
| AF484483 | 99UGB32394 | D | UG | 1999 |
| AF484485 | 99UGD23550 | D | UG | 1999 |
| AF484486 | 99UGD26830 | D | UG | 1999 |
| AF484487 | 99UGE08364 | D | UG | 1999 |
| AF484515 | 99UGE13613 | D | UG | 1999 |
| AF484489 | 99UGE23438 | D | UG | 1999 |
| AF484519 | 99UGF03726 | D | UG | 1999 |
| AF484490 | 99UGF05734 | D | UG | 1999 |
| AF484494 | 99UGG10555 | D | UG | 1999 |
| AF484495 | 99UGG35093 | D | UG | 1999 |
| AF484497 | 99UGJ27597 | D | UG | 1999 |
| AF484516 | 99UGJ32228 | D | UG | 1998 |
| AF484498 | 99UGK09259 | D | UG | 1999 |
| AF484499 | 99UGK09958 | D | UG | 1999 |
| AF077336 | VI850 | F1 | BE | 1993 |
| AY173957 | BZ126 | F1 | BR | 1989 |
| AY173958 | BZ163 | F1 | BR | 1989 |

Table S2

6/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|--------------|---------|---------------------------|------------------|
| AF005494 | 93BR020_1 | F1 | BR | 1993 |
| AF075703 | FIN9363 | F1 | FI | 1993 |
| AJ249238 | MP411 | F1 | FR | 1996 |
| AY371158 | 02CM_0016BBY | F2 | CM | 2002 |
| AJ249236 | MP255 | F2 | CM | 1995 |
| AJ249237 | MP257 | F2 | CM | 1995 |
| AF377956 | CM53657 | F2 | CM | 1997 |
| AF084936 | DRCBL | G | BE | 1996 |
| AY371121 | 01CM_4049HAN | G | CM | 2001 |
| AF423760 | X558 | G | ES | 2000 |
| AF450098 | X138 | G | ES | 1999 |
| AF061641 | HH8793_12_1 | G | FI | 1993 |
| U88826 | 92NG083 | G | NG | 1992 |
| AF061642 | SE6165 | G | SE | 1993 |
| AF190127 | VI991 | H | BE | 1993 |
| AF190128 | VI997 | H | BE | 1993 |
| AF005496 | 056 | H | CF | 1990 |
| AF082394 | SE7887 | J | SE | 1993 |
| AF082395 | SE7022 | J | SE | 1994 |
| AJ249235 | EQTBI1C | K | CD | 1997 |
| AJ249239 | MP535 | K | CM | 1996 |
| AF197340 | 90CF11697 | 01_AE | CF | 1990 |
| U51188 | 90CF402 | 01_AE | CF | 1990 |
| AF197341 | 90CF4071 | 01_AE | CF | 1990 |
| AY008718 | 97CNGX_11F | 01_AE | CN | 1997 |
| AY008714 | 97CNGX2F | 01_AE | CN | 1997 |
| AB052995 | 93JP_NH1 | 01_AE | JP | 1993 |
| AB070352 | NH25 | 01_AE | JP | 1993 |
| AF259955 | CM235 | 01_AE | TH | 1990 |
| U54771 | CM240 | 01_AE | TH | 1990 |
| AF197338 | 93TH057 | 01_AE | TH | 1993 |
| AF197339 | 93TH065 | 01_AE | TH | 1993 |
| U51189 | 93TH253 | 01_AE | TH | 1993 |
| AF164485 | 93TH9021 | 01_AE | TH | 1993 |
| AB032740 | 95TNIH022 | 01_AE | TH | 1995 |
| AB032741 | 95TNIH047 | 01_AE | TH | 1995 |
| AY125894 | 97TH6_107 | 01_AE | TH | 1997 |
| AY371122 | 01CM_0002BBY | 02_AG | CM | 2001 |
| AY371123 | 01CM_0005BBY | 02_AG | CM | 2001 |
| AY371124 | 01CM_0008BBY | 02_AG | CM | 2001 |
| AY371131 | 01CM_0074NY | 02_AG | CM | 2001 |
| AY371137 | 01CM_0131NY | 02_AG | CM | 2001 |
| AY371132 | 01CM_0158ND | 02_AG | CM | 2001 |
| AY371133 | 01CM_0191ND | 02_AG | CM | 2001 |
| AY371134 | 01CM_0925MO | 02_AG | CM | 2001 |
| AY371136 | 01CM_1237NG | 02_AG | CM | 2001 |
| AY371138 | 01CM_1475MV | 02_AG | CM | 2001 |
| AY371142 | 01CM_4410HAL | 02_AG | CM | 2001 |
| AY371125 | 02CM_0013BBY | 02_AG | CM | 2002 |
| AY371126 | 02CM_0014BBY | 02_AG | CM | 2002 |

Table S2

7/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|--------------|---------|---------------------------|------------------|
| AY371127 | 02CM_0015BBY | 02_AG | CM | 2002 |
| AY371139 | 02CM_1669LE | 02_AG | CM | 2002 |
| AY371140 | 02CM_1677LE | 02_AG | CM | 2002 |
| AY371146 | 02CM_1901LE | 02_AG | CM | 2002 |
| AY371128 | 02CM_1970LE | 02_AG | CM | 2002 |
| AY371129 | 02CM_2162SA | 02_AG | CM | 2002 |
| AY371130 | 02CM_2348SA | 02_AG | CM | 2002 |
| AY371141 | 02CM_4082STN | 02_AG | CM | 2002 |
| AJ286133 | 97CM_MP807 | 02_AG | CM | 1997 |
| AF377954 | CM52885 | 02_AG | CM | 1997 |
| AF377955 | CM53658 | 02_AG | CM | 1997 |
| AY271690 | pBD6_15 | 02_AG | CM | 1999 |
| AY151001 | ECU41 | 02_AG | EC | - |
| AY151002 | ECU42 | 02_AG | EC | - |
| AF063223 | DJ263 | 02_AG | FR | 1991 |
| AF063224 | DJ264 | 02_AG | FR | 1991 |
| AB049811 | 97GHAG1 | 02_AG | GH | 1997 |
| L39106 | IBNG | 02_AG | NG | - |
| AF107770 | SE7812 | 02_AG | SE | 1994 |
| AJ251056 | MP1211 | 02_AG | SN | 1998 |
| AJ251057 | MP1213 | 02_AG | SN | 1998 |
| AF414006 | 98BY10443 | 03_AB | BY | 2000 |
| AF193276 | KAL153 | 03_AB | RU | 1997 |
| AF193277 | RU98001 | 03_AB | RU | 1998 |
| AF049337 | CY032 | 04_cpx | CY | 1994 |
| AF119820 | 97PVCH | 04_cpx | GR | 1991 |
| AF119819 | 97PVMY | 04_cpx | GR | 1997 |
| AF193253 | VI1310 | 05_DF | BE | - |
| AF076998 | VI961 | 05_DF | BE | 1993 |
| AY227107 | X492 | 05_DF | ES | 1999 |
| AF064699 | BFP90 | 06_cpx | AU | 1996 |
| AJ288982 | 95ML127 | 06_cpx | ML | 1995 |
| AJ245481 | 95ML84 | 06_cpx | ML | 1995 |
| AJ288981 | 97SE1078 | 06_cpx | SN | 1997 |
| AF503396 | CNGL179 | 07_BC | CN | - |
| AF286226 | 97CN001 | 07_BC | CN | 1997 |
| AX149771 | CN54 | 07_BC | CN | 1997 |
| AF286230 | 98CN009 | 07_BC | CN | 1998 |
| AY008715 | 97CNGX_6F | 08_BC | CN | 1997 |
| AY008716 | 97CNGX_7F | 08_BC | CN | 1997 |
| AY008717 | 97CNGX_9F | 08_BC | CN | 1997 |
| AF286229 | 98CN006 | 08_BC | CN | 1998 |
| AY093605 | 96GH2911 | 09_cpx | GH | 1996 |
| AY093603 | 95SN1795 | 09_cpx | SN | 1995 |
| AY093604 | 95SN7808 | 09_cpx | SN | 1995 |
| AY093607 | 99DE4057 | 09_cpx | US | 1999 |
| AF289548 | 96TZ_BF061 | 10_CD | TZ | 1996 |
| AF289549 | 96TZ_BF071 | 10_CD | TZ | 1996 |
| AF289550 | 96TZ_BF110 | 10_CD | TZ | 1996 |
| AY371149 | 01CM_0186ND | 11_cpx | CM | 2001 |

Table S2

8/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|--------------|---------|---------------------------|------------------|
| AY371150 | 01CM_4041HAN | 11_cpx | CM | 2001 |
| AY371151 | 02CM_2190SA | 11_cpx | CM | 2002 |
| AY371153 | 02CM_4118STN | 11_cpx | CM | 2002 |
| AF492624 | 1816 | 11_cpx | CM | 1995 |
| AF492623 | 4496 | 11_cpx | CM | 1996 |
| AJ291718 | MP818 | 11_cpx | CM | 1997 |
| AJ291719 | MP1298 | 11_cpx | FR | 1999 |
| AJ291720 | MP1307 | 11_cpx | FR | 1999 |
| AF179368 | GR17 | 11_cpx | GR | - |
| AF408629 | A32879 | 12_BF | AR | 1997 |
| AF408630 | A32989 | 12_BF | AR | 1997 |
| AF385936 | ARMA159 | 12_BF | AR | 1999 |
| AF385934 | URTR23 | 12_BF | UY | 1999 |
| AF385935 | URTR35 | 12_BF | UY | 1999 |
| AY371154 | 02CM_3226MN | 13_cpx | CM | 2002 |
| AF460972 | 1849 | 13_cpx | CM | 1996 |
| AF460974 | 4164 | 13_cpx | CM | 1996 |
| AF423758 | X475 | 14_BG | ES | 2000 |
| AF423759 | X477 | 14_BG | ES | 2000 |
| AF450096 | X605 | 14_BG | ES | 2000 |
| AF450097 | X623 | 14_BG | ES | 2000 |
| AF423756 | X397 | 14_BG | ES | 1999 |
| AF423757 | X421 | 14_BG | ES | 1999 |
| AF529572 | 02TH_OUR1331 | 15_01B | TH | 2002 |
| AF529573 | 02TH_OUR1332 | 15_01B | TH | 2002 |
| AF530576 | 99TH_R2399 | 15_01B | TH | 1999 |
| AF516184 | 99TH_MU2079 | 15_01B | TH | 1999 |
| AF286239 | 97KR004 | 16_A2D | KR | 1997 |
| AF067156 | 95IN21301 | A1C | IN | 1995 |
| AF457061 | KISII5011 | A1C | KE | 2000 |
| AF457087 | MSA4080 | A1C | KE | 2000 |
| AF539406 | ML170_1995 | A1C | KE | 1995 |
| AF457064 | KNH1097 | A1C | KE | 1999 |
| U88823 | 92RW009_06 | A1C | RW | 1992 |
| AF071474 | SE9488 | A1C | SE | 1996 |
| AF361871 | 97TZ01 | A1C | TZ | 1997 |
| AF361876 | 97TZ06 | A1C | TZ | 1997 |
| AF361878 | 97TZ08 | A1C | TZ | 1997 |
| AF361879 | 97TZ09 | A1C | TZ | 1997 |
| AF484491 | 99UGF25926 | A1C | UG | 1999 |
| AF484501 | 99UGK30889 | A1C | UG | 1999 |
| AF457058 | KER2021 | A1D | KE | 2000 |
| AF457078 | KSM4028 | A1D | KE | 2000 |
| AF457082 | MSA4071 | A1D | KE | 2000 |
| AF457059 | KISII5003 | A1D | KE | 1999 |
| AF457073 | KSM4015 | A1D | KE | 1999 |
| AF075701 | SE6954 | A1D | SE | 1993 |
| AF071473 | SE7108 | A1D | SE | 1994 |
| AF442569 | TZBFL0011 | A1D | TZ | 1996 |
| AF442566 | TZBFL0086 | A1D | TZ | 1997 |

Table S2

9/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|-------------|----------|---------------------------|------------------|
| AF442570 | TZBFL0088 | A1D | TZ | 1997 |
| AF484503 | 98UG57129 | A1D | UG | 1998 |
| AF484521 | 98UG57137 | A1D | UG | 1998 |
| AF484510 | 98UG57139 | A1D | UG | 1998 |
| AF484522 | 98UG57147 | A1D | UG | 1998 |
| AF484482 | 99UGB26587 | A1D | UG | 1999 |
| AF484479 | 99UGC06443 | A1D | UG | 1999 |
| AF484517 | 99UGC38442 | A1D | UG | 1999 |
| AF484488 | 99UGE22831 | A1D | UG | 1999 |
| AF484492 | 99UGF27390 | A1D | UG | 1999 |
| AF484496 | 99UGJ21953 | A1D | UG | 1999 |
| AF484520 | 99UGK38855 | A1D | UG | 1999 |
| AY037267 | ARCH003 | BF | AR | 2000 |
| AF408631 | A050 | BF | AR | 1997 |
| AF408632 | A32878 | BF | AR | 1997 |
| AY037266 | ARCH014 | BF | AR | 1998 |
| AF408626 | A025 | BF | AR | 1999 |
| AF332867 | A027 | BF | AR | 1999 |
| AF408627 | A047 | BF | AR | 1999 |
| AF408628 | A063 | BF | AR | 1999 |
| AY037278 | ARMA006 | BF | AR | 1999 |
| AY037283 | ARMA029 | BF | AR | 1999 |
| AY037275 | ARMA036 | BF | AR | 1999 |
| AY037277 | ARMA037 | BF | AR | 1999 |
| AY037281 | ARMA038 | BF | AR | 1999 |
| AY037280 | ARMA097 | BF | AR | 1999 |
| AY037271 | BOL0137 | BF | BO | 1999 |
| AY037272 | URTR17 | BF | UY | 1999 |
| AB097865 | mIDU502 | 01B | MM | 2000 |
| AB097867 | mCSW104 | 01B | MM | 1999 |
| AF490973 | 1269 | 01B | TH | - |
| AY167123 | CM237 | 01B | TH | - |
| AF362994 | NP1623 | 01B | TH | - |
| AY082968 | TH1326 | 01B | TH | - |
| AF468970 | TH283 | 01B | TH | - |
| AF490974 | TH9_95 | 01B | TH | - |
| AF457088 | NKU3004 | A1A2D | KE | 2000 |
| AF457056 | KER2017 | A1CD | KE | 2000 |
| AY371163 | 01CM_1152NG | A1U | CM | 2001 |
| AF411964 | CM4 | A1CDGKU | ZA | 1999 |
| AF075702 | SE8603 | A1CD | SE | 1995 |
| AJ237565 | 97NOGIL3 | A1DHK | NO | 1997 |
| X04415 | MAL | A1DK | CD | 1985 |
| AY371164 | 01CM_1404MV | A1F1 | CM | 2001 |
| AF377959 | CM53379 | A1F1GHJU | CM | 1997 |
| AJ276596 | VI1197 | A1G | BE | 1994 |
| AF377957 | CM53392 | A1G | CM | 1997 |
| AF457062 | KNH1043 | A1G | KE | 1999 |
| U88825 | 92NG003 | A1G | NG | 1992 |
| AF076474 | VI354 | A1GHU | GA | - |

Table S2

10/10

| Accession | Name | Subtype | Country code (ISO3166) | Sampling year |
|-----------|--------------|---------|---------------------------|------------------|
| AF192135 | BW2117 | A1GJ | BW | 1998 |
| U76035 | Z321 | A1GU | CD | 76 |
| AF069672 | SE6594 | A1U | SE | 1993 |
| AF411965 | DU178 | A2C | ZA | 1998 |
| U86780 | ZAM184 | A2C | ZM | 1990 |
| AF457051 | KER2003 | A2D | KE | 1999 |
| AF457072 | KSM4001 | A2D | KE | 1999 |
| AF316544 | 97CDKP58 | A2G | CD | 1997 |
| AY371168 | 01CM_4008HAN | AD | CM | 2001 |
| AY371160 | 02CM_3163MN | AF2 | CM | 2002 |
| AY371166 | 01CM_0989MO | AGU | CM | 2001 |
| AY371169 | 02CM_1918LE | AGU | CM | 2002 |
| AY371161 | 01CM_1278NG | AHJU | CM | 2001 |
| AY371162 | 01CM_1296NG | AHJU | CM | 2001 |
| AP005206 | HH069 | BC | CN | 2000 |
| AP005207 | HH086 | BC | CN | 2000 |
| AB097873 | mIDU103 | BC | MM | 1999 |
| AB097869 | mIDU106 | BC | MM | 1999 |
| AF005495 | 93BR029_4 | BF1 | BR | 1993 |
| AF423755 | X254 | BG | ES | 1999 |
| AY074891 | 00BWMO351 | CD | BW | 2000 |
| AF457085 | MSA4077 | CD | KE | 2000 |
| AF361877 | 97TZ07 | CD | TZ | 1997 |
| AF484500 | 99UGK10192 | CD | UG | 1999 |
| AF076475 | VI1126 | F2KU | BE | 1994 |
| AY371159 | 01CM_0001BBY | 01A | CM | 2001 |
| AY371165 | 02CM_1867LE | 01A | CM | 2002 |
| AY037284 | CM53122 | 01A1 | CM | 1997 |
| AB097872 | mCSW105 | 01A1 | MM | 1999 |
| AY371170 | 01CM_0908MO | 01ADF2 | CM | 2001 |
| AY371167 | 01CM_0130NY | 01AF2U | CM | 2001 |
| AB097866 | mCSW503 | 01BC | MM | 2000 |
| AB097868 | mIDU107 | 01BC | MM | 1999 |
| AY262830 | NP1809 | 01C | TH | - |
| AJ404325 | 97DCKTB49 | 01GHJKU | CD | 1997 |
| AY371145 | 01CM_0190MA | 0102A | CM | 2001 |
| AF184155 | G829 | 02A | GH | - |
| AB052867 | AG2 | 02A1 | GH | 1997 |
| AY371143 | 02CM_2339SA | 02A1U | CM | 2002 |
| AJ276595 | VI1035 | 02C | BE | 1993 |
| AY371147 | 02CM_3228MN | 02G | CM | 2002 |
| AJ293865 | B76 | 06A1 | BJ | - |
| AF286236 | 83CD003 | U | CD | 1983 |
| AF457101 | 90CD121E12 | U | CD | 1990 |
| AY046058 | GR303 | U | GR | 1999 |

Table S3

1/3

| HXB2 coordinate | Target gene | Target site | Conservation (%) | Reference |
|--------------------|--------------------------------------|-----------------------------|---------------------|--------------------|
| 64 | U3, nef | AAGGCTACTTCCCTGATTGGC | 34.5 | Dave (2004) |
| 67 | U3, nef | GCTACTTCCCTGATTGGCAGAACTA | 18.6 | Yamamoto (2006) |
| 72 | U3, nef | TTCCCTGATTGGCAGAACTACACAC | 17.9 | Yamamoto (2006) |
| 130 | U3, nef | GATGGTGCTTCAAGCTAGTAC | 28.3 | Capodici (2002) |
| 184 | U3, nef | AAGGAGAGAACACCAGCTTGT | 0.4 | Chang (2005) |
| 291 | U3, nef | AAGGCCCGAGAGCTGCATCCGG | 0.9 | Suzuki (2005) |
| 321 | U3, nef | AAGACTGCTGACATCGAGCTT | 1.1 | Suzuki (2005) |
| 348 | U3 | AAGGGACTTTCCGCTGGGGAC | 23.5 | Suzuki (2005) |
| 395 | U3 | AACTGGGGAGTGGCGAGCCCT | 0.2 | Suzuki (2005) |
| 456 | R, TAR | GTCTCTCTGGTTAGACCAGAT | 16.2 | Yoshinari (2004) |
| 466 | R, TAR | TTAGACCAGATCTGAGCCTGG | 16.1 | Yoshinari (2004) |
| 468 | R, TAR | AGACCAGATCTGAGCCTGG | 16.1 | Jacque (2002) |
| 483 | R, TAR | CTGGGAGCTCTCTGGCTAACT | 9.5 | Yoshinari (2004) |
| 493 | R, TAR | TCTGGCTAACTAGGGAACCCA | 11.3 | Yoshinari (2004) |
| 515 | R, poly A | TGCTTAAGCCTCAATAAAGCT | 85.3 | Yoshinari (2004) |
| 551 | R, U5 | AAGTAGTGTGTGCCCCGTCTGT | 65.1 | Chang (2005) |
| 626 | U5, PBS | TCTCTAGCAGTGGCGCCCCG | 92.0 | Nishitsuji (2006) |
| 634 | PBS | GGTGGCGCCCCGAACAGGGACCTT | 17.0 | Han (2004) |
| 635 | PBS | GTGGCGCCCCGAACAGGGAC | 90.1 | Das (2004) |
| 636 | PBS | TGGCGCCCCGAACAGGGACT | 54.2 | Das (2004) |
| 769 | U5, Ψ | GCGGAGGCTAGAAGGAGAG | 88.1 | Das (2004) |
| 774 | U5, Ψ, gag (AUG) | GGCTAGAAGGAGAGAGATG | 91.0 | Das (2004) |
| 780 | U5, gag (AUG) | AAGGAGAGAGATGGGTGCGAGAGCGTC | 93.2 | ter Brake (2006) |
| 788 | U5, gag (AUG) | AGATGGGTGCGAGAGCGTC | 92.9 | Westerhout (2006a) |
| 896 | gag (p17) | GGGCAAGCAGGGAGCTAGAAC | 21.5 | Capodici (2002) |
| 896 | gag (p17) | GGGCAAGCAGGGAGCTAGAACTT | 21.5 | Han (2004) |
| 1475 | gag (p24) | GAGAACCAAGGGGAAGTGACA | 44.3 | Capodici (2002) |
| 1600 | gag (p24) | AATAAAATAGTAAGAATGTA | 77.3 | ter Brake (2006) |
| 1605 | gag (p24) | AATAGTAAGAATGTATAGCCC | 75.9 | Chang (2005) |
| 1793 | gag (p24) | AAGCATTGGGACCAGGAGCGA | 0.0 | Hu (2002) |
| 1817 | gag (p24) | TAGAAGAAATGATGACAGCATG | 82.3 | ter Brake (2006) |
| 1820 | gag (p24) | AAGAAATGATGACAGCATGTC | 62.1 | Chang (2005) |
| 2060 | gag (p7) | AAGATTGTACTGAGAGACAGG | 9.3 | Novina (2002) |
| 2060 | gag (p7) | AAGATTGTACTGAGAGACAGGCT | 9.5 | Suzuki (2005) |
| 2062 | gag (p7) | GATTGTACTGAGAGACAGG | 9.3 | Song (2003) |
| 2062 | gag (p7) | GATTGTACTGAGAGACAGG | 9.3 | Lee (2005) |
| 2066 | gag (p7) | GTACTGAGAGACAGGCTAA | 9.1 | Pusch (2003) |
| 2077 | gag (p7,p1), pol, reosomal slip site | CAGGCTAATTTTTTAGGGAA | 70.6 | ter Brake (2006) |
| 2167 | gag (p6), pol | GAGGAGAGCTTCAGGTTTGGG | 4.2 | Lin (2004) |
| 2315 | pol (prot) | AGCTCTATTAGATACAGGA | 44.5 | Paul (2003) |
| 2316 | pol (prot) | GCTCTATTAGATACAGGAGCA | 47.1 | Scherer (2004) |
| 2328 | pol (prot) | ACAGGAGCAGATGATACAGT | 59.5 | ter Brake (2006) |
| 2375 | pol (prot) | ATGGAAACCAAAAATGATAGG | 81.9 | ter Brake (2006) |
| 2559 | pol (RT) | CCTATTGAGACTGTACCAG | 8.0 | Paul (2003) |
| 2586 | pol (RT) | AAGCCAGGAATGGATGGCCCA | 63.5 | Chang (2005) |
| 2748 | pol (RT) | GACAGTACTAAATGGAGAATT | 0.0 | Sabariegos (2006) |
| 2960 | pol (RT) | TGAGACACCAGGGATTAGA | 22.8 | Surabhi (2002) |
| 2961 | pol (RT) | GAGACACCAGGGATTAGAT | 22.8 | Surabhi (2002) |
| 3092 | pol (RT) | TCAATACATGGATGATTTGTA | 26.5 | Huelsmann (2006) |
| 4168 | pol (p15) | AAGGAATTGGAGGAAATGAAC | 47.2 | Chang (2005) |

Table S3

2/3

| HXB2 coordinate | Target gene | Target site | Conservation (%) | Reference |
|--------------------|-----------------|-------------------------------------|---------------------|--------------------|
| 4173 | pol (p15) | ATTGGAGGAAATGAACAAGT | 47.6 | ter Brake (2006) |
| 4285 | pol (int) | GGAGAGCAATGGCTAGTGA | 58.7 | Nishitsuji (2006) |
| 4537 | pol (int) | AATTAGCAGGAAGATGGCCAG | 67.0 | Chang (2005) |
| 4539 | pol (int) | TTAGCAGGAAGATGGCCAGT | 67.4 | ter Brake (2006) |
| 4650 | pol (int) | ATTCCCTACAATCCCCAAAG | 50.7 | ter Brake (2006) |
| 4776 | pol (int), cPPT | CACAATTTTAAAGAAAAGGGGGATTGG GGGG | 90.7 | ter Brake (2006) |
| 4809 | pol (int) | TACAGTGCAGGGGAAAGAATA | 84.4 | ter Brake (2006) |
| 4811 | pol (int) | CAGTGCAGGGGAAAGAATA | 84.4 | Westerhout (2006a) |
| 4884 | pol (int), CTS | AAAATTCAAAATTTTCGGGT | 77.3 | ter Brake (2006) |
| 4888 | pol (int), CTS | TTCAAAATTTTCGGGTTTATT | 80.6 | ter Brake (2006) |
| 4893 | pol (int), CTS | AATTTTCGGGTTTATTACAG | 78.5 | ter Brake (2006) |
| 4953 | pol (int) | CTCTGGAAAGGTGAAGGGGCAGTAGTAAT | 84.0 | ter Brake (2006) |
| 4960 | pol (int) | AAGGTGAAGGGGCAGTAGTAA | 69.8 | Chang (2005) |
| 4966 | pol (int) | AAGGGGCAGTAGTAATAGAAG | 0.0 | Hu (2002) |
| 5040 | pol (int), vif | TATGGAAAACAGATGGCAGGTG | 90.7 | ter Brake (2006) |
| 5138 | vif | GGAAAGCTAAGGACTGGTT | 0.2 | Jacque (2002) |
| 5138 | vif | GGAAAGCTAAGGACTGGTT | 0.2 | Joshi (2005) |
| 5195 | vif | GTTCAAGATACACATCCC | 60.8 | Lee (2005) |
| 5323 | vif | AGCACACAAGTAGACCCTGA | 24.4 | Jacque (2002) |
| 5323 | vif | AGCACACAAGTAGACCCTG | 24.4 | Jacque (2002) |
| 5323 | vif | AGCACACAAGTAGACCCTG | 24.4 | Nishitsuji (2004) |
| 5481 | vif | CTTGGCACTAGCAGCATT | 5.6 | Jacque (2002) |
| 5852 | tat | CTAGAGCCCTGGAAGCATC | 7.8 | Surabhi (2002) |
| 5864 | tat | AAGCATCCAGGAAGTCAGCCT | 10.3 | Dave (2004) |
| 5887 | tat | AACTGCTTGTACCAATTGCTA | 1.8 | Coburn (2002) |
| 5889 | tat | CTGCTTGTACCAATTGCTATT | 5.0 | Boden (2003a,b) |
| 5889 | tat | CTGCTTGTACCAATTGCTA | 3.2 | Boden (2004a) |
| 5889 | tat | CTGCTTGTACCAATTGCTATT | 5.0 | Huelsmann (2006) |
| 5915 | tat | AAGTGTGCTTTTCATTGCCAAG | 5.4 | Lee (2003) |
| 5954 | tat, rev | GCCTTAGGCATCTCCTATG | 1.2 | Das (2004) |
| 5966 | tat, rev | TCCTATGGCAGGAAGAAGCGGAG | 81.5 | ter Brake (2006) |
| 5967 | tat, rev | CCTATGGCAGGAAGAAGCG | 56.8 | Das (2004) |
| 5969 | tat, rev | TATGGCAGGAAGAAGCGGA | 79.9 | Surabhi (2002) |
| 5980 | tat, rev | GAAGCGGAGACAGCGACGA | 45.5 | Surabhi (2002) |
| 5983 | tat, rev | GCGGAGACAGCGACGAAGAGC | 7.4 | Lee (2002) |
| 5983 | tat, rev | GCGGAGACAGCGACGAAGAGC | 7.4 | Li (2003) |
| 5983 | tat, rev | GCGGAGACAGCGACGAAGAGC | 7.4 | Scherer (2004) |
| 5983 | tat, rev | GCGGAGACAGCGACGAAGAGC | 7.4 | Joshi (2005) |
| 5983 | tat, rev | GCGGAGACAGCGACGAAGAGC | 7.4 | Li (2005) |
| 5994 | tat, rev | GACGAAGAGCTCATCAGAACA | 0.2 | Boden (2004b) |
| 6205 | vpu | AAGAGCAGAAGACAGTGGCAA | 53.6 | Chang (2005) |
| 6213 | vpu, env | AAGACAGTGGCAATGAGAGTG | 52.2 | Chang (2005) |
| 7152 | env | AAGAGGGGACCAGGGAGAGCAC | 0.0 | Park (2003) |
| 7368 | env | AAGTTCTACTGTAATTCAACAC | 0.0 | Park (2003) |
| 7384 | env | CAACACAACCTGTTTAATAGTA | 4.0 | Scherer (2004) |
| 7460 | env | CACAATCACACTCCCATGCAG | 1.0 | Scherer (2004) |
| 7464 | env | AAGACACTCCCATGCAGAATAC | 0.0 | Park (2003) |
| 7497 | env | AAGATGTGGCAGGAAGTAGGAC | 0.2 | Park (2003) |
| 7635 | env | GGAGGAGGCGATATGAGGGAC | 0.4 | Scherer (2004) |
| 7798 | env | CAGCAGGAAGCACTATGGGCGC | 75.8 | ter Brake (2006) |

Table S3

| HXB2 coordinate | Target gene | Target site | Conservation (%) | Reference |
|--------------------|---------------|------------------------------|---------------------|----------------------|
| 7858 | env | TGTCTGATATAGTGCAGCAGC | 0.0 | Scherer (2004) |
| 7916 | env | TCTGTTGCAACTCACAGTCTG | 24.1 | Scherer (2004) |
| 7924 | env | AACTCACAGTCTGGGGCATCA | 11.9 | Dave (2004) |
| 8425 | tat, rev, env | AAGAAGAAGGTGGAGAGAGAG | 7.1 | Dave (2004) |
| 8472 | rev, env | AACGGATCCTTGGCACTTATC | 0.4 | Coburn (2002) |
| 8483 | rev, env | GGCACTTATCTGGGACGAT | 0.4 | Lee (2005) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Lee (2002) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Banerjea (2003) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Li (2003) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Scherer (2004) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Unwalla (2004) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Joshi (2005) |
| 8509 | rev, env | GCCTGTGCCTCTTCAGCTACC | 40.6 | Unwalla (2006) |
| 8558 | rev, env | AACGAGGATTGTGGAACCTTCT | 8.7 | Dave (2004) |
| 8759 | env | AAGAATAAGACAGGGCTTGGA | 9.5 | Dave (2004) |
| 8762 | env | AATAAGACAGGGCTTGGAAG | 10.9 | Dave (2004) |
| 8817 | nef | AAGTAGTGTGATTGGATGGCC | 0.4 | Chang (2005) |
| 8844 | nef | AAGGGAAAGAATGAGACGAGC | 6.3 | Dave (2004) |
| 8914 | nef | CAUGGAGCAAUCACAAGUATT | 0.0 | Sabariegos (2006) |
| 8960 | nef | GTGCCTGGCTAGAAGCACA | 11.6 | Jacque (2002) |
| 8960 | nef | GTGCCTGGCTAGAAGCACA | 11.6 | Das (2004) |
| 8960 | nef | GTGCCTGGCTAGAAGCACA | 11.6 | Westerhout (2005) |
| 8960 | nef | GTGCCTGGCTAGAAGCACA | 11.6 | Westerhout (2006a,b) |
| 8999 | nef | TTCCAGTCACACCTCAGGTAC | 1.6 | Scherer (2004) |
| 9063 | nef, U3 | CTTTTTTAAAAGAAAAGGGGGGACTGGA | 90.7 | ter Brake (2006) |
| 9069 | nef, U3 | AAAAGAAAAGGGGGGACTGGA | 90.7 | Dave (2004) |
| 9078 | nef, U3 | GGGGGGACTGGAAGGGCTA | 14.1 | Das (2004) |
| 9081 | nef, U3 | GGGACTGGAAGGGCTAATT | 10.6 | Westerhout (2006a) |
| 9083 | nef, U3 | GACTGGAAGGGCTAATTCA | 2.6 | Nishitsuji (2006) |

* For the target sequence longer than 21 nt, conservation was calculated for each possible 21 nt subsequence, and the highest value is shown. Target sequence shorter than 21 nt was extended to 21 nt for calculating their conservations.

Table S4

1/5

| HXB2 coordinate | Target gene | Target site (21 bp) | Conservation (%) | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguioui | | |
|--------------------|----------------|------------------------|---------------------|--|-----|-----|
| 412 | U3, TATA | CCCTCAGATGCTGCATATAAG | 74.6 | | | |
| 413 | U3, TATA | CCTCAGATGCTGCATATAAGC | 75.1 | | | |
| 414 | U3, TATA | CTCAGATGCTGCATATAAGCA | 74.6 | yes | | |
| 415 | U3, TATA | TCAGATGCTGCATATAAGCAG | 74.6 | | | |
| 416 | U3, TATA | CAGATGCTGCATATAAGCAGC | 75.8 | | | |
| 505 | R, TAR | GGGAACCCACTGCTTAAGCCT | 77.8 | yes | | |
| 506 | R, TAR | GGAACCCACTGCTTAAGCCTC | 81.3 | | | |
| 507 | R, TAR, poly A | GAACCCACTGCTTAAGCCTCA | 84.9 | | | |
| 508 | R, TAR, poly A | AACCCACTGCTTAAGCCTCAA | 85.3 | | | yes |
| 509 | R, TAR, poly A | ACCCACTGCTTAAGCCTCAAT | 84.9 | yes | yes | yes |
| 510 | R, TAR, poly A | CCCACTGCTTAAGCCTCAATA | 84.9 | yes | yes | yes |
| 511 | R, TAR, poly A | CCACTGCTTAAGCCTCAATAA | 85.2 | | yes | |
| 512 | R, poly A | CACTGCTTAAGCCTCAATAAA | 85.4 | yes | yes | yes |
| 513 | R, poly A | ACTGCTTAAGCCTCAATAAAG | 84.7 | | | |
| 514 | R, poly A | CTGCTTAAGCCTCAATAAAGC | 85.3 | | | |
| 515 | R, poly A | TGCTTAAGCCTCAATAAAGCT | 85.3 | yes | | |
| 516 | R, poly A | GCTTAAGCCTCAATAAAGCTT | 85.7 | | | |
| 517 | R, poly A | CTTAAGCCTCAATAAAGCTTG | 85.7 | | | |
| 518 | R, poly A | TTAAGCCTCAATAAAGCTTGC | 86.7 | | | |
| 519 | R, poly A | TAAGCCTCAATAAAGCTTGCC | 86.6 | | | |
| 520 | R, poly A | AAGCCTCAATAAAGCTTGCCCT | 93.9 | | | yes |
| 521 | R, poly A | AGCCTCAATAAAGCTTGCCCTT | 94.2 | yes | | yes |
| 522 | R, poly A | GCCTCAATAAAGCTTGCCCTTG | 95.0 | | | |
| 523 | R, poly A | CCTCAATAAAGCTTGCCCTTGA | 94.2 | | | |
| 524 | R, poly A | CTCAATAAAGCTTGCCCTTGAG | 94.6 | | | |
| 525 | R, poly A | TCAATAAAGCTTGCCCTTGAGT | 87.5 | | | |
| 526 | R, poly A | CAATAAAGCTTGCCCTTGAGTG | 87.5 | | | |
| 527 | R, poly A | AATAAAGCTTGCCCTTGAGTGC | 85.5 | | | |
| 528 | R, poly A | ATAAAGCTTGCCCTTGAGTGCT | 79.5 | | | |
| 552 | U5 | AGTAGTGTGTGCCCCGTCTGTT | 70.3 | | | |
| 576 | U5, PAS | TGACTCTGGTAACTAGAGATC | 71.5 | | | |
| 577 | U5, PAS | GACTCTGGTAACTAGAGATCC | 81.3 | | | |
| 578 | U5, PAS | ACTCTGGTAACTAGAGATCCC | 81.9 | | | |
| 579 | U5, PAS | CTCTGGTAACTAGAGATCCCT | 87.1 | | | yes |
| 580 | U5, PAS | TCTGGTAACTAGAGATCCCTC | 87.1 | | | |
| 581 | U5, PAS | CTGGTAACTAGAGATCCCTCA | 87.7 | | | yes |
| 582 | U5, PAS | TGGTAACTAGAGATCCCTCAG | 88.2 | | | |
| 583 | U5, PAS | GGTAACTAGAGATCCCTCAGA | 88.2 | | yes | |
| 584 | U5, PAS | GTAACCTAGAGATCCCTCAGAC | 87.7 | | | |
| 585 | U5 | TAACTAGAGATCCCTCAGACC | 83.7 | | | |
| 622 | U5, PBS | AAAATCTCTAGCAGTGGCGCC | 90.2 | | | |
| 623 | U5, PBS | AAATCTCTAGCAGTGGCGCCC | 93.1 | | | |
| 624 | U5, PBS | AATCTCTAGCAGTGGCGCCCG | 92.0 | | | |
| 625 | U5, PBS | ATCTCTAGCAGTGGCGCCCGA | 92.1 | | | |
| 626 | U5, PBS | TCTCTAGCAGTGGCGCCCGAA | 93.3 | | | |
| 627 | U5, PBS | CTCTAGCAGTGGCGCCCGAAC | 92.8 | | | |
| 628 | U5, PBS | TCTAGCAGTGGCGCCCGAACA | 92.4 | | | |
| 629 | U5, PBS | CTAGCAGTGGCGCCCGAACAG | 92.4 | | | |
| 630 | U5, PBS | TAGCAGTGGCGCCCGAACAGG | 93.8 | | | |
| 631 | U5, PBS | AGCAGTGGCGCCCGAACAGGG | 93.3 | | | |

Table S4

2/5

| HXB2 coordinate | Target gene | Target site (21 bp) | Conservation (%) | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguioui | | |
|--------------------|---|-----------------------|---------------------|--|-----|-----|
| 632 | U5, PBS | GCAGTGGCGCCCGAACAGGGA | 90.0 | | | |
| 633 | U5, PBS | CAGTGGCGCCCGAACAGGGAC | 90.1 | | | |
| 682 | U5 | TCTCTCGACGCAGGACTCGGC | 85.2 | | | |
| 683 | U5 | CTCTCGACGCAGGACTCGGCT | 88.1 | | | |
| 684 | U5 | TCTCGACGCAGGACTCGGCTT | 88.5 | | | |
| 685 | U5 | CTCGACGCAGGACTCGGCTTG | 88.5 | | | |
| 686 | U5 | TCGACGCAGGACTCGGCTTGC | 88.2 | | | |
| 687 | U5 | CGACGCAGGACTCGGCTTGCT | 88.5 | | | |
| 688 | U5 | GACGCAGGACTCGGCTTGCTG | 90.0 | | | |
| 689 | U5 | ACGCAGGACTCGGCTTGCTGA | 88.2 | yes | | yes |
| 761 | U5, Ψ | TTTGACTAGCGGAGGCTAGAA | 85.2 | | yes | |
| 762 | U5, Ψ | TTGACTAGCGGAGGCTAGAAG | 86.9 | | | |
| 763 | U5, Ψ | TGACTAGCGGAGGCTAGAAGG | 87.6 | | | |
| 764 | U5, Ψ | GACTAGCGGAGGCTAGAAGGA | 88.0 | yes | | |
| 765 | U5, Ψ | ACTAGCGGAGGCTAGAAGGAG | 88.0 | | | |
| 766 | U5, Ψ | CTAGCGGAGGCTAGAAGGAGA | 88.4 | | | yes |
| 767 | U5, Ψ | TAGCGGAGGCTAGAAGGAGAG | 88.1 | | | |
| 768 | U5, Ψ | AGCGGAGGCTAGAAGGAGAGA | 87.8 | | | yes |
| 769 | U5, Ψ | GCGGAGGCTAGAAGGAGAGAG | 88.3 | | | |
| 770 | U5, Ψ, gag (AUG) | CGGAGGCTAGAAGGAGAGAGA | 88.6 | yes | | yes |
| 771 | U5, Ψ, gag (AUG) | GGAGGCTAGAAGGAGAGAGAT | 88.3 | | | yes |
| 772 | U5, Ψ, gag (AUG) | GAGGCTAGAAGGAGAGAGATG | 91.0 | | | |
| 773 | U5, Ψ, gag (AUG) | AGGCTAGAAGGAGAGAGATGG | 91.4 | | | |
| 774 | U5, Ψ, gag (AUG) | GGCTAGAAGGAGAGAGATGGG | 91.4 | | | |
| 775 | U5, Ψ, gag (AUG) | GCTAGAAGGAGAGAGATGGGT | 91.1 | | | |
| 776 | U5, Ψ, gag (AUG) | CTAGAAGGAGAGAGATGGGTG | 90.4 | | | |
| 777 | U5, Ψ, gag (AUG) | TAGAAGGAGAGAGATGGGTGC | 90.4 | | | |
| 778 | U5, Ψ, gag (AUG) | AGAAGGAGAGAGATGGGTGCG | 90.1 | | | |
| 779 | U5, Ψ, gag (AUG) | GAAGGAGAGAGATGGGTGCGA | 90.4 | | | |
| 780 | U5, gag (AUG) | AAGGAGAGAGATGGGTGCGAG | 92.1 | | | |
| 781 | U5, gag (AUG) | AGGAGAGAGATGGGTGCGAGA | 92.1 | | | yes |
| 782 | U5, gag (AUG) | GGAGAGAGATGGGTGCGAGAG | 92.1 | | | |
| 783 | U5, gag (AUG) | GAGAGAGATGGGTGCGAGAGC | 93.2 | | | |
| 784 | U5, gag (AUG) | AGAGAGATGGGTGCGAGAGCG | 92.5 | | | |
| 785 | U5, gag (AUG) | GAGAGATGGGTGCGAGAGCGT | 93.2 | | | |
| 786 | U5, gag (AUG) | AGAGATGGGTGCGAGAGCGTC | 92.9 | | | |
| 787 | U5, gag (AUG) | GAGATGGGTGCGAGAGCGTCA | 89.8 | | | |
| 1597 | gag (p24) | TTAAATAAAATAGTAAGAATG | 72.4 | | | |
| 1598 | gag (p24) | TAAATAAAATAGTAAGAATGT | 77.3 | | | |
| 1599 | gag (p24) | AAATAAAATAGTAAGAATGTA | 76.9 | | yes | |
| 1600 | gag (p24) | AATAAAATAGTAAGAATGTAT | 77.7 | | | |
| 1601 | gag (p24) | ATAAAATAGTAAGAATGTATA | 77.7 | | yes | |
| 1602 | gag (p24) | TAAAATAGTAAGAATGTATAG | 77.5 | | | |
| 1603 | gag (p24) | AAAATAGTAAGAATGTATAGC | 75.9 | | | |
| 1604 | gag (p24) | AAATAGTAAGAATGTATAGCC | 75.9 | | | |
| 1605 | gag (p24) | AATAGTAAGAATGTATAGCCC | 75.9 | | | |
| 1606 | gag (p24) | ATAGTAAGAATGTATAGCCCT | 75.9 | | | |
| 1817 | gag (p24) | TAGAAGAAATGATGACAGCAT | 82.3 | yes | | yes |
| 1818 | gag (p24) | AGAAGAAATGATGACAGCATG | 82.3 | | | |
| 2075 | gag (p7, p1), pol, ribosomal slip site | GACAGGCTAATTTTTTAGGGA | 70.6 | yes | | |

Table S4

3/5

| HXB2 coordinate | Target gene | Target site (21 bp) | Conservation (%) | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguioui | | |
|--------------------|---|-----------------------|---------------------|--|-----|-----|
| 2077 | gag (p7, p1), pol, ribosomal slip site | CAGGCTAATTTTTTAGGGAAA | 70.0 | yes | | yes |
| 2251 | gag (p6), pol (prot) | TCCCTCAAATCACTCTTTGGC | 79.9 | | | |
| 2252 | gag (p6), pol (prot) | CCCTCAAATCACTCTTTGGCA | 81.9 | yes | | yes |
| 2328 | pol (prot) | ACAGGAGCAGATGATACAGTA | 81.3 | | yes | |
| 2329 | pol (prot) | CAGGAGCAGATGATACAGTAT | 77.7 | yes | yes | yes |
| 2330 | pol (prot) | AGGAGCAGATGATACAGTATT | 77.9 | yes | | yes |
| 2331 | pol (prot) | GGAGCAGATGATACAGTATTA | 79.1 | | yes | yes |
| 2332 | pol (prot) | GAGCAGATGATACAGTATTAG | 78.5 | | | |
| 2333 | pol (prot) | AGCAGATGATACAGTATTAGA | 77.3 | yes | | yes |
| 2334 | pol (prot) | GCAGATGATACAGTATTAGAA | 79.7 | | yes | |
| 2335 | pol (prot) | CAGATGATACAGTATTAGAAG | 78.3 | | | |
| 2336 | pol (prot) | AGATGATACAGTATTAGAAGA | 77.7 | | | |
| 2375 | pol (prot) | ATGGAAACCAAAAATGATAGG | 81.9 | | | |
| 2487 | pol (prot) | CCTACACCTGTCAACATAATT | 72.0 | | | |
| 2488 | pol (prot) | CTACACCTGTCAACATAATTG | 72.2 | | | |
| 2489 | pol (prot) | TACACCTGTCAACATAATTGG | 72.2 | | | |
| 2490 | pol (prot) | ACACCTGTCAACATAATTGGA | 73.0 | | | |
| 2611 | pol (RT) | TTAAACAATGGCCATTGACAG | 70.0 | | | |
| 2613 | pol (RT) | AAACAATGGCCATTGACAGAA | 73.0 | | | |
| 2614 | pol (RT) | AACAATGGCCATTGACAGAAG | 73.0 | | | |
| 2615 | pol (RT) | ACAATGGCCATTGACAGAAGA | 73.6 | | yes | |
| 2696 | pol (RT) | AATTGGGCTGAAAATCCATA | 74.8 | | | |
| 3302 | pol (RT) | CTGGACTGTCAATGATATACA | 70.2 | yes | yes | yes |
| 4173 | pol (p15) | ATTGGAGGAAATGAACAAGTA | 76.7 | | yes | |
| 4174 | pol (p15) | TTGGAGGAAATGAACAAGTAG | 76.5 | | | |
| 4175 | pol (p15) | TGGAGGAAATGAACAAGTAGA | 76.5 | yes | | yes |
| 4176 | pol (p15) | GGAGGAAATGAACAAGTAGAT | 71.1 | | | yes |
| 4177 | pol (p15) | GAGGAAATGAACAAGTAGATA | 71.1 | yes | | yes |
| 4178 | pol (p15) | AGGAAATGAACAAGTAGATAA | 70.5 | yes | yes | yes |
| 4179 | pol (p15) | GGAAATGAACAAGTAGATAAA | 70.7 | | yes | |
| 4182 | pol (p15) | AATGAACAAGTAGATAAATTA | 71.7 | | yes | |
| 4183 | pol (p15) | ATGAACAAGTAGATAAATTAG | 71.9 | | | |
| 4184 | pol (p15) | TGAACAAGTAGATAAATTAGT | 71.9 | | | |
| 4749 | pol (int) | ACAGCAGTACAAATGGCAGTA | 74.9 | | | |
| 4750 | pol (int) | CAGCAGTACAAATGGCAGTAT | 74.1 | yes | | yes |
| 4751 | pol (int) | AGCAGTACAAATGGCAGTATT | 74.1 | yes | yes | yes |
| 4752 | pol (int) | GCAGTACAAATGGCAGTATTC | 71.5 | | | |
| 4753 | pol (int) | CAGTACAAATGGCAGTATTCA | 71.5 | yes | yes | yes |
| 4754 | pol (int) | AGTACAAATGGCAGTATTCAT | 71.5 | | | |
| 4776 | pol (int), cPPT | CACAATTTTAAAAGAAAAGGG | 89.8 | | | |
| 4777 | pol (int), cPPT | ACAATTTTAAAAGAAAAGGGG | 89.8 | | | |
| 4778 | pol (int), cPPT | CAATTTTAAAAGAAAAGGGGG | 89.8 | | | |
| 4779 | pol (int), cPPT | AATTTTAAAAGAAAAGGGGGG | 90.5 | | | |
| 4780 | pol (int), cPPT | ATTTTAAAAGAAAAGGGGGGA | 90.5 | | | |
| 4781 | pol (int), cPPT | TTTAAAAGAAAAGGGGGGAT | 90.5 | | | |
| 4782 | pol (int), cPPT | TTTAAAAGAAAAGGGGGGATT | 90.7 | | | |
| 4783 | pol (int), cPPT | TTAAAAGAAAAGGGGGGATTG | 90.7 | | | |
| 4784 | pol (int), cPPT | TAAAAGAAAAGGGGGGATTGG | 90.7 | | | |
| 4785 | pol (int), cPPT | AAAAGAAAAGGGGGGATTGGG | 90.3 | | | |
| 4786 | pol (int), cPPT | AAAGAAAAGGGGGGATTGGGG | 89.4 | | | |

Table S4

4/5

| HXB2 coordinate | Target gene | Target site (21 bp) | Conservation (%) | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguioui | | |
|--------------------|-----------------|-----------------------|---------------------|--|-----|-----|
| 4787 | pol (int), cPPT | AAGAAAAGGGGGGATTGGGGG | 86.6 | | | |
| 4788 | pol (int), cPPT | AGAAAAGGGGGGATTGGGGG | 78.1 | | | |
| 4789 | pol (int), cPPT | GAAAAGGGGGGATTGGGGGGT | 78.1 | | | |
| 4790 | pol (int), cPPT | AAAAGGGGGGATTGGGGGGTA | 78.1 | | | |
| 4791 | pol (int), cPPT | AAAGGGGGGATTGGGGGGTAC | 76.9 | | | |
| 4792 | pol (int), cPPT | AAGGGGGGATTGGGGGGTACA | 76.7 | | | yes |
| 4793 | pol (int), cPPT | AGGGGGGATTGGGGGGTACAG | 80.2 | | | |
| 4794 | pol (int), cPPT | GGGGGGATTGGGGGGTACAGT | 80.4 | yes | | yes |
| 4795 | pol (int), cPPT | GGGGGATTGGGGGGTACAGTG | 80.4 | | | |
| 4796 | pol (int), cPPT | GGGGATTGGGGGGTACAGTGC | 80.0 | | | |
| 4797 | pol (int), cPPT | GGGATTGGGGGGTACAGTGCA | 79.3 | | | |
| 4798 | pol (int), cPPT | GGATTGGGGGGTACAGTGCAG | 78.9 | | | |
| 4799 | pol (int), cPPT | GATTGGGGGGTACAGTGCAGG | 79.1 | | | |
| 4800 | pol (int), cPPT | ATTGGGGGGTACAGTGCAGGG | 76.1 | | | |
| 4801 | pol (int) | TTGGGGGGTACAGTGCAGGGG | 75.9 | | | |
| 4802 | pol (int) | TGGGGGGTACAGTGCAGGGGA | 75.5 | | | yes |
| 4803 | pol (int) | GGGGGGTACAGTGCAGGGGAA | 73.5 | | | yes |
| 4804 | pol (int) | GGGGGTACAGTGCAGGGGAAA | 72.9 | | | yes |
| 4805 | pol (int) | GGGGTACAGTGCAGGGGAAAG | 73.1 | | | |
| 4806 | pol (int) | GGGTACAGTGCAGGGGAAAGA | 72.3 | yes | yes | |
| 4807 | pol (int) | GGTACAGTGCAGGGGAAAGAA | 72.9 | | | |
| 4808 | pol (int) | GTACAGTGCAGGGGAAAGAAT | 74.5 | | yes | |
| 4809 | pol (int) | TACAGTGCAGGGGAAAGAATA | 84.4 | yes | yes | yes |
| 4810 | pol (int) | ACAGTGCAGGGGAAAGAATAA | 71.5 | | yes | yes |
| 4811 | pol (int) | CAGTGCAGGGGAAAGAATAAT | 71.5 | yes | | yes |
| 4812 | pol (int) | AGTGCAGGGGAAAGAATAATA | 73.7 | | yes | |
| 4813 | pol (int) | GTGCAGGGGAAAGAATAATAG | 73.9 | | | |
| 4814 | pol (int) | TGCAGGGGAAAGAATAATAGA | 74.1 | yes | yes | yes |
| 4883 | pol (int), CTS | AAAAATTCAAATTTTCGGGT | 77.3 | | | |
| 4884 | pol (int), CTS | AAAATTCAAATTTTCGGGTT | 76.7 | | | |
| 4885 | pol (int), CTS | AAATTCAAATTTTCGGGTTT | 78.5 | | | |
| 4886 | pol (int), CTS | AATTCAAATTTTCGGGTTTA | 78.7 | | | |
| 4887 | pol (int), CTS | ATTCAAATTTTCGGGTTTAT | 79.3 | | yes | |
| 4888 | pol (int), CTS | TTCAAATTTTCGGGTTTATT | 80.6 | yes | | yes |
| 4889 | pol (int), CTS | TCAAATTTTCGGGTTTATTA | 76.9 | | yes | |
| 4890 | pol (int), CTS | CAAATTTTCGGGTTTATTAC | 77.5 | | | |
| 4891 | pol (int), CTS | AAAATTTTCGGGTTTATTACA | 77.7 | | | |
| 4892 | pol (int), CTS | AAATTTTCGGGTTTATTACAG | 78.5 | | | |
| 4953 | pol (int) | CTCTGGAAGGTGAAGGGGCA | 82.2 | | | yes |
| 4954 | pol (int) | TCTGGAAGGTGAAGGGGCAG | 82.0 | | | |
| 4955 | pol (int) | CTGGAAGGTGAAGGGGCAGT | 82.2 | | | yes |
| 4956 | pol (int) | TGGAAGGTGAAGGGGCAGTA | 84.0 | | | |
| 4957 | pol (int) | GGAAAGGTGAAGGGGCAGTAG | 83.6 | | | |
| 4958 | pol (int) | GAAAGGTGAAGGGGCAGTAGT | 83.4 | | | |
| 4959 | pol (int) | AAAGGTGAAGGGGCAGTAGTA | 72.1 | | yes | yes |
| 4961 | pol (int) | AGGTGAAGGGGCAGTAGTAAT | 70.1 | yes | | yes |
| 5040 | pol (int), vif | TATGGAAAACAGATGGCAGGT | 90.7 | | | |
| 5041 | pol (int), vif | ATGGAAAACAGATGGCAGGTG | 90.1 | | | |
| 5966 | pol (int), vif | TCCTATGGCAGGAAGAAGCGG | 81.5 | | | |
| 5967 | pol (int), vif | CCTATGGCAGGAAGAAGCGGA | 79.9 | | | |

Table S4

5/5

| HXB2 coordinate | Target gene | Target site (21 bp) | Conservation (%) | siRNA efficacy prediction Ui-Tei, Reynolds, Amarzguioui |
|----------------------------|--------------------|----------------------------|-----------------------------|--|
| 5970 | pol (int), vif | ATGGCAGGAAGAAGCGGAGAC | 70.2 | yes |
| 7654 | env (gp120) | ACAATTGGAGAAGTGAATTAT | 70.7 | |
| 7655 | env (gp120) | CAATTGGAGAAGTGAATTATA | 70.5 | |
| 7796 | env (gp41) | AGCAGCAGGAAGCACTATGGG | 71.1 | |
| 7797 | env (gp41) | GCAGCAGGAAGCACTATGGGC | 71.7 | |
| 7798 | env (gp41) | CAGCAGGAAGCACTATGGGCG | 75.4 | |
| 7799 | env (gp41) | AGCAGGAAGCACTATGGGCGC | 75.8 | |
| 7830 | env (gp41) | ACGCTGACGGTACAGGCCAGA | 72.1 | |
| 7831 | env (gp41) | CGCTGACGGTACAGGCCAGAC | 72.9 | yes |
| 9063 | nef, 3'PPT | CTTTTAAAAAGAAAAGGGGGG | 86.1 | |
| 9064 | nef, 3'PPT | TTTTTAAAAAGAAAAGGGGGGA | 89.5 | |
| 9065 | nef, 3'PPT | TTTTTAAAAAGAAAAGGGGGGAC | 90.2 | |
| 9066 | nef, 3'PPT, U3 | TTTAAAAAGAAAAGGGGGGACT | 90.1 | |
| 9067 | nef, 3'PPT, U3 | TTAAAAAGAAAAGGGGGGACTG | 90.3 | |
| 9068 | nef, 3'PPT, U3 | TAAAAAGAAAAGGGGGGACTGG | 90.5 | |
| 9069 | nef, 3'PPT, U3 | AAAAGAAAAGGGGGGACTGGA | 90.7 | |

謝 辞

本研究の機会を恵与され、終始多大なる御指導、御鞭撻を賜りました東京大学教授 西郷 薫博士に謹んで感謝の意を表しますとともに、厚く御礼申し上げます。

本研究全般にわたり直接御指導を賜り、また、本論文の校閲を戴きました東京大学助教授 程 久美子博士ならびに国立感染症研究所室長 武部 豊博士に謹んで謝意を表しますとともに、共同研究を通じて終始御懇篤なる御指導、御討論を賜りました東京大学教授 森下 真一博士ならびに東京大学特任助手 山田 智之氏に深謝いたします。

また、数々の御助言、御討論を戴きました国立遺伝学研究所教授 上田 龍博士、前橋工科大学助教授 善野 修平博士、東京大学助手 高橋 史峰博士に厚く御礼申し上げます。

さらに、実験に御協力、御援助を戴きました江田 理愛氏、田中 愛氏ならびに西郷研究室、程研究室の皆様感謝いたします。

配 列 表