

RNAi 効果の高い sRNA 配列設計方法の構築

池田大樹[†] 権 娟大^{††} 宮崎 智^{††}

本研究では、RNAi 効果を示す既存の sRNA 配列の設計要因について、生体内に存在する sRNA の一次構造を比較・検討し、より高い RNAi 効果を示す配列の設計手法を構築することを目的としている。核酸医薬においてより効果の高い small RNA を作成することにより、現在治療不可能とされている疾患に対する新たな治療法として貢献することができる。DDBJ に登録されている生体内で機能している small RNA を集め、それらの配列情報を解析した。その結果、生体内の sRNA と人工的に作成された sRNA の類似性は非常に低いことが判明した。このような結果にもかかわらず両者が高い効果の RNAi を発揮していることから、small RNA が効果の高い RNAi を発揮するためには、small RNA の配列以外にも他に大きな要因があることが分かった。

In silico design of small RNA sequences with high RNAi effects

Taiki Ikeda[†], Yeondae Kwon^{††}
and Satoru Miyazaki^{††}

In this study, we aim to build the design technique of small RNA sequences that show high RNAi effects by comparing sequences of small RNAs in vivo with those of artificial small RNAs which show RNAi effects. Making a small RNA with higher effects in nucleotide drug contributes a new cure for currently untreatable diseases. We collected small RNA data which function in vivo from DDBJ and analyzed their sequences. As a result, it turns out that the similarity between in vivo small RNAs and artificial small RNAs is low. In spite of this result, since both of small RNAs show high RNAi effects, it is implied that there exist some factors other than the particular sequences of small RNAs so that small RNAs have high RNAi effects.

1. 背景・目的

mRNA のように蛋白質をコードしている RNA を coding RNA と呼び、一方で蛋白質をコードしていない RNA を non coding RNA (以下、ncRNA) と呼ぶ。ヒトゲノム全体の約 98%以上が蛋白質をコードしない領域、ノンコーディング領域であることがわかっている。近年の転写産物の網羅的な解析から、これらノンコーディング領域から産生される転写産物の種類と数は、生物が複雑になるほど増える傾向が見られており、ncRNA によるゲノム情報の発現制御こそが生物の複雑さを生み出しているといえる [1]。このようなことから、種の違いや表現型の変動は蛋白質をコードする配列にあるのではなく、遺伝子発現の制御プログラムの中にある可能性が高い。small RNA (以下、sRNA) はこの ncRNA の一種であり、非常に短い配列でできた直線状の二本鎖 RNA をさす。二本鎖 RNA は一本鎖に分離した後に蛋白質複合体を形成し、自身と相補的な配列を持つ mRNA を認識して結合する。この蛋白質複合体が mRNA を切断もしくは翻訳抑制することによって、mRNA によって翻訳されてできる蛋白質の発現が抑制される。この作用を RNA interference (以下、RNAi) と呼ぶ [2]。

RNAi の性質を利用して人工的に sRNA を設計することにより、理論上あらゆる蛋白質に対して発現抑制が可能になる。そこで本研究では、標的蛋白質に対して高い発現抑制効果を持つ sRNA の配列を設計するプログラムの作製を目的とし、それに必要となる sRNA データベースの構築を行い、sRNA 塩基配列から共通点や特徴がないか解析を行った。

2. 準備

2.1 RNAi の反応機構

sRNA の種類によってその生成機構はそれぞれ異なるが、切断酵素によって二本鎖 sRNA が生成されてから RNAi を起こすまでの反応機構は共通している。二本鎖 sRNA は分離酵素によって片側一本鎖のみが選択され、sRNA-蛋白質複合体 RISC (RNA-induced silencing complex) を形成する (図 1)。この時、選ばれなかったもう一方の sRNA は分解される [3]。RISC は sRNA と相補的な配列を持つ mRNA を認識し結合したのち、mRNA を切断、もしくはそのまま結合し続けることによって蛋白質が翻訳されるのを妨げる。

[†] 東京理科大学大学院 薬学研究科 薬科学専攻
Graduate School of Pharmaceutical Sciences, Tokyo University of Science

^{††} 東京理科大学 薬学部 生命創薬科学科
Faculty of Pharmaceutical Sciences, Tokyo University of Science

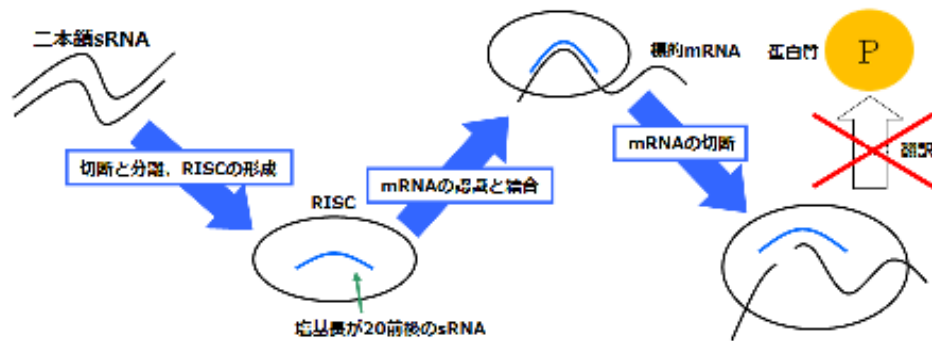


図1 RNAiの反応機構

2.2 従来のsRNA設計方法

標的 mRNA に対して sRNA の配列を設計する際、単純に相補的な配列を作れば抑制効果が高くなるというわけではない。mRNA のどの部分に結合するのか、もしくはその部分が生物において保存されている領域なのか等で RNAi の効果の高さや標的 mRNA も変わってしまう[4][5]。現在利用されている sRNA 設計プログラムの多くはある程度の共通した概念にしたがっており[6]、その代表的な設計プログラムの手順について簡単に説明する。

まず、mRNA の配列上から次のような条件をみたく塩基長 20 前後の部分配列を検索し、それと相補的な配列を sRNA として提供する：(i) 全体の GC 含量が 30~50% の範囲にある (ii) 両末端の GC 含量を比較したとき、3'末端側のほうが高い (iii) G もしくは C が 3 塩基連続で並ばない (iv) 3 塩基周期的に A もしくは U が出現する[8]。

次に、設計プログラムによって次のような独自の条件を付加する：(i) sRNA を特徴づける部分配列を指定する (ii) エネルギー値などをスコア化して基準値と比較を行う[9] (iii) 過去のデータから設計した配列がどれほどの効果を持つか予測する。

最後に、設計した配列が他の mRNA と相補的な配列を持たないか、相同性検索ツール BLAST (Basic Local Alignment Search Tool) [10]等で調べ、結果の良いものを返す[11]。

これらの方法において付加される条件の多くは実験的に決められたものが多い。例えば、まず、ある mRNA に対して網羅的に sRNA を設計し、一つずつ RNAi による標的蛋白質の発現抑制率を調べる。そして実験から得た結果を比較し、RNAi 効果の高いものに見られた sRNA の特徴を解析する。解析から得た sRNA の共通性「10 番目の塩基が A である」等の性質を sRNA の配列設計の条件として付加する。他にも、sRNA が二本鎖から一本鎖に分離する際の性質を利用した条件等がある。

科学的根拠よりも実験的根拠による条件が多いため、設計によって得られた sRNA すべてが高い効果を発揮するというわけではない。また、設計プログラムによっては過去のプログラムから設計された sRNA の実験結果を集め、それらのデータを現在の設計プログラムに生かし改良を加えていく場合もある[12]。

2.3 副作用「オフターゲット効果」

標的 mRNA とは異なる mRNA を切断してしまうことによって生じる作用をオフターゲット効果と呼ぶ。sRNA は短い配列であるため、標的 mRNA とは無関係の mRNA にも相補性を持つ部分配列が存在することがある。そのため sRNA によって関係のない mRNA が翻訳抑制されてしまう場合がある (図 2)。このように間違っ て抑制された mRNA が、本来生体の恒常において不可欠な蛋白質を作り出していた場合、強い毒性を示すことになる[13][14]。

これを避けるためには、設計した sRNA が標的 mRNA 以外の遺伝子に対して相補性が低くなる必要がある。一般的には BLAST を用いて、設計した sRNA の相補配列が標的 mRNA 以外に高い相補性を持つ遺伝子が存在しないかを確認することで回避している。

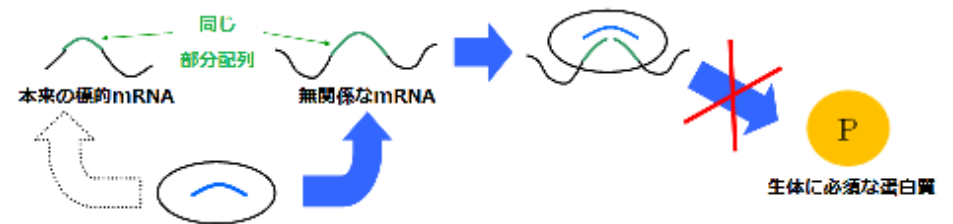


図2 オフターゲット効果

3. 方法と実験

3.1 sRNA の配列情報の取得と分類

国際塩基配列データベース DDBJ (DNA Data Bank of Japan) [15]の web API サービス REST_WEB_API[16]を用いて ncRNA の取得を行った。DDBJ では ncRNA を「A non-protein-coding gene, other than ribosomal RNA and transfer RNA, the functional molecule of which is the RNA transcript. (rRNA と tRNA を除いた、蛋白質をコードしていない機能性の転写産物)」と定義しており、今回探索・収集する sRNA はこれらの ncRNA のうち RNAi を機能として持つ RNA として区別することにした。

まず、DDBJ に登録されている ncRNA の情報をすべて集めた。これらのうち、RNAi の機能を持つ sRNA である siRNA、piRNA、rasiRNA、miRNA に着目することにした。以降、「sRNA」は siRNA、piRNA、rasiRNA、miRNA を指すものとする。さらに収集した sRNA を、sRNA の種類、生物種、配列長毎に分類し、各分類において最も登録件数の多かったグループを各 sRNA から選出した。

また、分類結果を利用し、今後必要な sRNA の配列群をいつでも取得できるよう、独自のデータベースに格納した。

3.2 生体内・人工の sRNA の配列比較

本来生体内の sRNA は特定の mRNA を標的としているため、必ず高い効果で RNAi の機能を発揮している。また、従来の設計方法によって得られた特定条件を満たす sRNA も標的 mRNA に対して高い効果を持っていることから、生体内の sRNA と人工の sRNA の配列間には類似性が見られることが予想される。

本研究では DDBJ に登録されている各 sRNA において登録件数が最も多かった生物種のうち、さらに登録件数が最も多かった配列長をもつ sRNA のグループを選出し、それらの配列群が従来の設計に用いられる代表的な配列条件をどれほど満たしているのか解析を行った。

今回配列に関する解析を目的としたため、2.2 節における代表的な検索条件のうち以下の条件を用いた。①～⑤は配列上の条件であり、これらの条件とは別に生体内に見られる各種 sRNA の配列上の共通点⑥～⑧も解析した。

- ① 「GC」含量が 30～50%の範囲にある。
- ② 両末端の「GC」含量を比較して 5'側が小さい。
- ③ 3 塩基周期性を 4/6 以上みたす。
- ④ 3 塩基以上「G」もしくは「C」が連続しない。

- ⑤ ①～④をすべて満たす配列の割合を計算する。
- ⑥ 共通塩基（各位置における出現率の高い塩基）の解析をする。
- ⑦ 「GC」含量豊富な部分領域の解析をする。
- ⑧ 「AU」含量豊富な部分領域の解析をする。

条件③の 3 塩基周期性とは、5'側から数えて $3n+1$ 番目の塩基(4,7,10,13,16,19)が「A」もしくは「U」であることを意味する。これらの塩基は活性との関係性が高く、「G」や「C」では活性が下がってしまうことが実験的に知られている[8]。

3.3 生体内の sRNA の標的遺伝子候補の探索と収集

人工的に設計する際の sRNA の条件と生体内の sRNA の配列比較の結果から、生体内の sRNA は二本鎖の状態から一本鎖になる過程でどちらが利用されるかわからない状態にあることが分かった。二本鎖のうちどちらを利用するかによって標的となる相補配列が異なるため、別々の mRNA に対して RNAi を起こす可能性がある(図 3)。

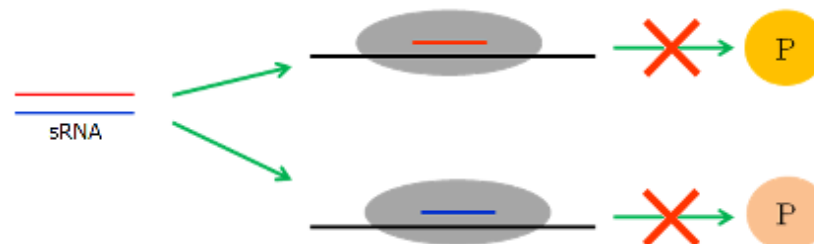


図3 二本鎖 sRNA の選択による標的の違い

場合によっては、その時生体にとって不可欠な蛋白質を作る mRNA に対して RNAi を行い副作用を引き起こす可能性が高い。しかし日常生活においてそのような突発的な副作用が見られないことから、生体内の sRNA には標的に対してのみ確実に RNAi を起こす仕組みがあると考えられる。

そこで本研究では、sRNA の標的遺伝子の解析を行うために、生体内の sRNA データベース miRBase[17][18]と sRNA の標的予測サイト microRNA.ORG[19][20]を利用した。miRBase では sRNA の名称を入力して検索することにより、二本鎖 sRNA それぞれの配列とその標的候補予測サイトへのリンクを得ることができる。これを利用し、これまでに収集してきた配列群のうち、ヒトが持つ二本鎖 sRNA (配列長 22 塩基対) の標的候補を解析した。

3.4 生体内の sRNA の標的解析

生体内の sRNA の配列情報を解析するために、まず、ヒトが持つ二本鎖 sRNA のうち miRBase に登録されている sRNA の標的候補リストから、標的候補としての可能性の高さを示すスコアが高かった mRNA 20 件を選出した[20]。次に、sRNA の相補鎖の標的候補も同様に 20 件選出した。さらに、二本鎖 sRNA 両者の標的候補 20 件を比較し、一致するものがないか調べた。この標的解析を 50 件の二本鎖 sRNA について解析した。

4. 結果

4.1 ncRNA の収集と分類

DDBJ において、ncRNA は全体で 19 種 57178 件登録されており、生物種は 862 種にわたっていた。そのうち今回の解析で用いる sRNA は全体で 9194 件、さらに分類した miRNA は 3413 件、piRNA は 4843 件、rasiRNA は 325 件、siRNA は 613 件登録されていた。各 sRNA で最も登録件数が多い生物種とその配列長のデータ群をまとめた。その結果を表 1 に示す。

また、sRNA の解析基盤を構築するために、収集した sRNA のデータを詳細に分類しデータベースに格納した。このデータベースには miRNA が 3413 件、piRNA が 4843 件、rasiRNA が 325 件、siRNA が 613 件登録されている。

表 1 登録件数の多かった sRNA 群

名称	生物種	配列長	登録件数
miRNA	<i>Homo sapiens</i>	22	122
	<i>Mus musculus</i>	22	164
	<i>Gallus gallus</i>	22	138
siRNA	<i>Gossypium hirsutum</i>	24	361
rasiRNA	<i>Drosophila melanogaster</i>	25	90
piRNA	<i>Mus musculus</i>	26	643
	<i>Drosophila melanogaster</i>	25	290

以下、*Homo sapiens* を Hom、*Mus musculus* を Mus、*Gallus gallus* を Gal、*Gossypium hirsutum* を Gos、*Drosophila melanogaster* を Dro、*Drosophila melanogaster* を Dro と記す。

4.2 生体内 sRNA の配列分析および標的解析

人工的に sRNA の配列を設計する際に用いられる条件と、生体内の sRNA の配列の比較を行った。その結果を表 2 に示す。

表 2 生体内の sRNA 配列と設計条件の比較結果

	22mi Hom*	22mi Mus	22mi Gal	24si Gos	25ra Dro	26pi Mus	25pi Dro
①GC 含量が 30~50%	45% [†]	56%	45%	61%	63%	61%	63%
②両末端の GC 含量の比較で 5'側のほうが小さいもの	47%	30%	30%	52%	41%	45%	59%
③3 塩基周期性を 4/6 以上満たすもの	36%	42%	28%	53%	46%	57%	62%
④3 塩基以上連続で G/C の続かないもの	23%	40%	24%	38%	22%	41%	35%
⑤上記の条件①~④をすべて満たすもの	4%	5%	2%	9%	4%	12%	12%
⑥データ群に見られた共通塩基 (5'からの番号-塩基)	41% (1-U)	43% (1-U)	41% (1-U) 67% (3-S)	64% (1-A) 72% (2-W)	71% (10-A)	81% (1-U)	74% (1-U) 69% (5-W)
⑦データ群に見られた GC 含量豊富な部分領域 (5'からの番号で 8 割以上 GC を含む)	-	-	65% [‡] (2-8)	-	-	-	-
⑧データ群に見られた AU 含量豊富な部分領域 (5'からの番号で 8 割以上 AU を含む)	-	-	-	72% (1-7)	60% (3-12)	60% (1-11)	70% (1-12)

* 配列長が 22 の miRNA で、生物種が *Homo sapiens* であることを意味する。

[†] 22miHom の全データのうち、①を満たす配列が全体で 45%見つかったことを意味する。

[‡] 5'末端側の 2 番目から 8 番目の範囲において GC 含量 80%を超えていた配列が、22miGal 全体の 65%にみられたことを表す。

各データ群において、条件①～④の個々の条件を満たす割合は、6割に到達するものもみられたが、これらの4つの条件（条件⑤）をすべて満たす生体内の sRNA は非常に少ないことが分かった。

人工的に作られる sRNA が 19～22 塩基程度であるのに対して生体内の sRNA の配列長は生物種によってばらつきがあるため、これらの条件がすべての生物間で成り立っているとは一意にいえない。しかしながら、今回解析に用いたデータが登録件数が最多な配列群であったことを踏まえると、少なくともモデル生物種においては、人工的に設計される sRNA と生体内に存在する sRNA の間の配列類似性が非常に低いことが分かった。

また、50 ペアの二本鎖 sRNA の標的比較を行ったが、完全一致したペアは存在しなかった。

5. 考察

本研究における解析の結果、表 2 の条件①～④をすべて満たす生体内の sRNA は非常に少ないことが分かった。したがって、生体内の sRNA が標的とする mRNA が分かっているとき、この mRNA に対して、人工的に sRNA を設計する場合、生体内の sRNA と人工的に設計された sRNA の両者の配列が大きく異なっている可能性が高いといえる（図 5）。

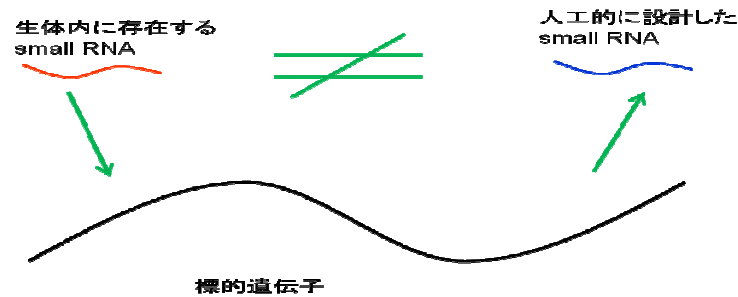


図 5 生体内 sRNA と人工 sRNA の配列の不一致

また、生体内の sRNA の標的解析の結果から、生体内の二本鎖 sRNA が両者同一の標的をもつ可能性が非常に低いことが分かった（図 6）。

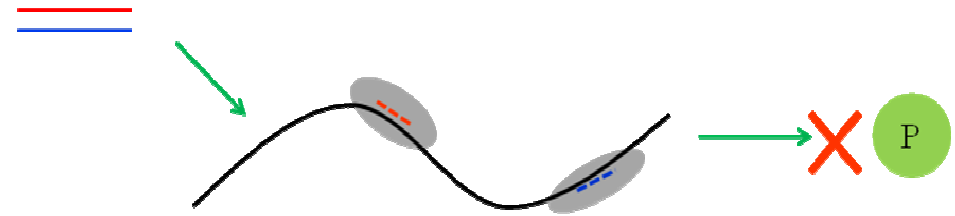


図 6 二本鎖 sRNA が同一の標的 mRNA をもつ場合

今回の解析結果と、人工的に設計された sRNA と生体内の sRNA が共に高い活性をもっていることを踏まえると、RNAi の活性の高さは配列のみに依存するものではない可能性が高い。さらに二本鎖の選択性に関しても、人工的に設計する際とは別の条件に左右される可能性が高く、二本鎖それぞれの標的 mRNA が最終的に同一の mRNA の翻訳抑制につながっているなどの可能性も挙げられる。

今後の展望として、二本鎖 sRNA の選択における他の条件の存在有無、二本鎖 sRNA の生成過程における切断部位、ある標的 mRNA に対して RNAi を行う sRNA 群に見られる共通点・保存配列の有無などの解析を行う予定である。

6. 参考文献

- [1] Haruhiko S.: ノンコーディング RNA 研究の魅力とその現状, 蛋白質 核酸 酵素, Vol.53, No.15, pp. 1926-1931 (2008).
- [2] Schwarz D. S., et al.: Asymmetry in the assembly of the RNAi enzyme complex, Cell, Vol.115, pp. 199-208 (2003).
- [3] DU Z., LEE J., et al.: Reconstitution of human RNA interference in budding yeast, Proc Nat Acad Sci, Vol.105, No.7, pp. 2391-2396 (2008).
- [4] Dykxhoorn D. M., et al.: Short RNAs that silence gene expression, Nat Rev Mol Cell Biol, Vol.4, pp. 457-467(2003).
- [5] Hutvagner G, and Zamore P.: A microRNA in a Multiple-Turnover RNAi Enzyme Complex, Science, Vol.297, pp. 2056-2060 (2002).
- [6] Ui-Tei K, and Takahashi Y.: Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference, Nucleic Acids Res, Vol.32, No.3, pp. 936-948 (2004).
- [7] Aza-Blanc P., et al.: Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening, Cell, Vol.12, pp. 627 (2003).

- [8] Tsutomu S.: RNAi 活性予測アルゴリズムを用いた高効率 siRNA の設計, 細胞工学, Vol.59, No.8, pp. 1710-2715 (2006).
- [9] Ichihara M., Murakumo Y., and Masuda A.: Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities, Nucleic Acid Res, Vol.35, pp. 123 (2007).
- [10] BLAST, <http://blast.ddbj.nig.ac.jp/top-j.html>
- [11] Yamada T., and Morishita S.: Accelerated off-target search algorithm for siRNA, Bioinformatics, Vol.21, pp. 1316-1324 (2005).
- [12] Naito Y., Yamada T., et al.: siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference, Nucleic Acid Res, Vol.32, pp. 124-129 (2004).
- [13] Jackson A., et al.: Prediction profiling reveals off-target gene regulation by RNAi, Nat Biotechnol, Vol.21, pp. 635-637 (2003).
- [14] Saxena S., et al.: Small RNAs with imperfect match to endogenous mRNA repress translation : Implications for off-target activity of small inhibitory RNA in mammalian cells, J Biochem, Vol.278, pp. 44312-44319 (2003).
- [15] DDBJ, <http://www.ddbj.nig.ac.jp/index-j.html>
- [16] REST_WEB_API, http://www.xml.nig.ac.jp/tutorial/rest/index_jp.html
- [17] miRBase, <http://www.mirbase.org/>
- [18] Griffiths J. S., et al.: miRBase: microRNA sequences, targets and gene Nomenclature, NAR, Vol.34, pp. 140-144 (2006).
- [19] microRNA.ORG, <http://www.microrna.org/microrna/home.do>
- [20] Betel D., Wilson M., and Gabow A.: The microRNA.org resource: targets and expression, Nucleic Acids Res, Vol.36, pp. 149-53 (2008).