

siRNA配列設計システム“SISNIPER”の開発

The development of an siRNA design system “SISNIPER”

石川 元一* 野原 祥夫* 佐藤 仁則*
Motokazu Ishikawa, Sachio Nohara, Yoshinori Sato

近年、遺伝子機能解析を行うための方法として、RNAi(RNA interference)が注目を浴びている。RNAiは、細胞や生体内の標的となるmRNAと相補的な20塩基程度の2本鎖RNAである“siRNA(short interfering RNA)”を細胞内に導入することで、mRNAを破壊し遺伝子の機能解析を行う技術である。RNAiに用いるsiRNAには配列設計が重要な役割を果たし、そのための専用ソフトウェアを試薬合成メーカー等が提供しているが、設計精度の点で問題が指摘されていた。そこで我々は高精度なsiRNA配列設計システムの開発を目標として大阪大学、産業技術総合研究所、ジェノメディア(株)と共にコンソーシアムを結成し、その開発成果として、siRNA配列設計システム-SISNIPER^{*1}を完成させた。

SISNIPERは、遺伝子治療のベンチャー企業であるジェノメディア(株)(アンジェスMGグループ)から提供を受けたRNAi実験データの解析結果から生み出された設計アルゴリズムを用いている。このアルゴリズムは、GC含量(GC Score)、RISCへの導入効率(Load Score)、塩基位置(Position Score)、off-target effect(Specificity Score)、活性プロファイル適合度(Profile Score)という5つの基準を用いて各siRNA候補配列をスコア付けすることで活性を予測し、有効なsiRNAを選択するというものである。このうち活性プロファイル適合度の計算は、活性のあるsiRNAの配列の特徴を利用したプロファイルに基づいた予測を行うものであり、既存のソフトウェアには見られない新しい手法である。システムとしての実装に当たっては、コアアルゴリズム以外に、ユーザーごとの設計結果保存機能や、PDF出力機能、プロファイルの自動生成機能等の利便性を高めるための機能を付加した。

Recently, RNAi(RNA interference) has gotten a lot of attention as a tool for gene functional analysis. RNAi is a technology to suppress a function of a specific gene by destructing mRNA of the gene with siRNA(short interfering RNA), a 20 - 30nt RNA duplex with two dT overhangs, which has complementary sequence with the targeted mRNA. Because a sequence design plays a crucial role for making an siRNA, some design softwares for siRNA which is used in RNAi experiment are provided by reagent companies, but the accuracy of these softwares are said to be low. So, we formed a consortium with Osaka University, National Institute of Advanced Industrial Science and Technology and GenomIdea, Inc. with an aim to develop a highly accurate siRNA sequence design software. Finally, we have completed an siRNA sequence design software - SISNIPER - as the result of the consortium.

SISNIPER, an siRNA design system we have developed, is based on an algorithm which was written through the analysis of RNAi experimental data provided by a gene therapy venture - GenomIdea, Inc., a subsidiary of AnGes MG, Inc. In this algorithm, siRNA candidates are graded on five criteria: a GC percentage(GC Score), RISC-loading efficacy(Load Score), a position(Position Score), an off-target effect(Specificity Score) and compatibility to an active profile(Profile Score) and then, active siRNAs will be selected. The active profile is based on the characteristics of active siRNAs sequences, and the calculation with this profile is our original method. Besides a core algorithm, we have implemented some useful functions on the system, such as result-preservation for each user, PDF output and a profile auto-generator.

*1 SISNIPERは、三菱スペース・ソフトウェア(株)の登録商標です。

1. 背景

二本鎖RNAによってその配列と相補的なmRNAが分解され遺伝子の機能が抑制されるRNAi(RNA interference)という現象が、1998年に線虫で最初に報告され、それ以降その他の生物種でも同様の仕組みが次々に報告された。哺乳類でも、siRNA(short interfering RNA)と呼ばれる20数塩基程度の二本鎖RNAがRNAiを引き起こすことが発見された⁽¹⁾。細胞中に導入されたsiRNAは、図1のようにタンパク質複合体であるRISC(RNA-induced silencing complex)に取り込まれ、siRNA配列と相補的な配列を持つmRNAを切断することで、標的遺伝子の機能を抑制する(Knockdownと呼ばれる)。

ある遺伝子の機能を解析したい場合に、その標的遺伝子に対するsiRNAを作成し細胞内に導入してその遺伝子の機能を抑制し、細胞や個体に表れた何らかの変化を観察、またはアッセイを行うことで、その遺伝子の機能を類推することができる。遺伝子機能を抑制する従来の方法として、ノックアウト個体の作製や、アンチセンス法が用いられてきた。ノックアウト個体は遺伝子を完全に破壊できるという効果を持つ反面、その作製には高度な技術と膨大な時間、費用がかかることがネックとなっていた。一方、アンチセンス法は簡便に使えるが、再現性や抑制効果の低さ、生体内での安定性の低さのため普及には至っていない。この2つの技術と比較して、RNAiは簡便、安価に利用でき、非常に高い抑制効果を持つことから、最近登場した技術にも関わらず既に遺伝子機能解析の分野におけるデファクトスタンダードなツールとしての地位を確立している。

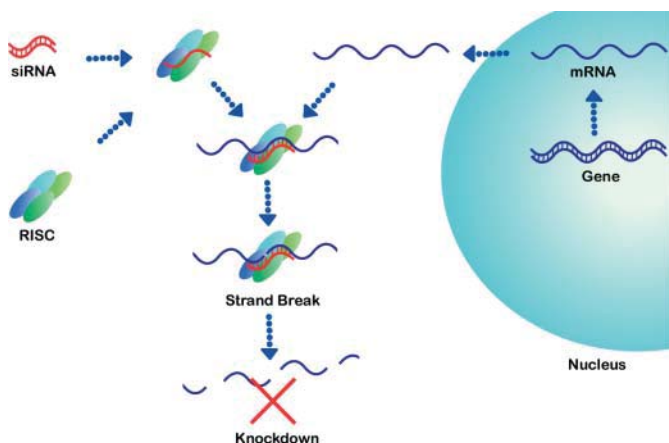


図1 RNAiのメカニズム

2. 現在の市場と技術的な課題

2.1 市場の現状

現在のところ、RNAi関連ビジネスのほとんどをsiRNA試薬の販売やsiRNAを用いた受託解析などの研究開発分野での需要が占めており、現在の市場規模は550億円程度(2004年のデータ)である。RNAiを用いた治療法の開発が進んでいて、実際にC型肝炎、黄斑変性等では臨床段階まで達しているものもあり、これらの売り上げを合わせると、2010年までに1,000億円を超えると予想されている。

siRNA試薬を作製するプロセスは、siRNAに用いる配列の設計、及びその配列を基にした試薬の合成の2つに分けられる。設計の市場について説明すると、まず、siRNA配列設計ソフトウェアの利用形態の主流は、siRNAを販売する試薬合成メーカーが自社のWebサイト内に設計サイトを立ち上げ、ユーザーに利用させるというものであるが、その他にも、配列設計用のパッケージソフトを研究者が購入して使用するケースもある。しかし、ほとんどの研究者がsiRNAを合成メーカーに発注しているため、各メーカーとも設計から発注までを一貫してできるサイトを構築することで顧客の獲得を行っており、いわば試薬を購入する顧客に対するサービスの一環としてsiRNA配列設計を提供している形になっている。

2.2 技術的な課題

実際にsiRNAを実験に使っている研究者から、合成メーカーの設計精度が低いために数種のsiRNAを購入せねばならず、結局合成メーカーとしてはその方が売り上げが上がるという悪弊についても聞いており、市場としては設計精度の高いソフトウェアについてはニーズがあるが、それを満たせるようなものは存在していないことを認識していた。

そこでまず我々は、設計仕様を公開しているソフトウェアについて調査した。まず、EMBOSS(The European Molecular Biology Open Software Suite)が公開したオープンソースソフトウェアであるsirnaは、siRNAを発見したTuschulらの公開した設計プロトコルに基づいて作成されたソフトウェアである。これは、2002年という早い段階でリリースされ、その後の設計ソフトウェアに大きな影響を与えている。

合成メーカーで開発されたソフトウェアは、アルゴリズムの詳細を公開していないところが多いが、数社では論文発表等によりその一部を公開しており^(3,4)、いずれも、特定の位置にある種類の塩基が存在するような配列をsiRNAとして選択することを特色としている。

siRNAが登場して既に数年が経過しているため、特に合成メーカー等は多数の遺伝子についてのsiRNAの実験データを揃えつつあるという状況となっており、設計の精度も徐々に上がりつつはあるが、しかし今だに活性を確実に予見できるような決定的な相関因子は発見されておらず、またRISCによるmRNAの切断の詳細な生化学的様式も報告されていない。したがって、相対的にデータ量は増加しつつあるが、それに対して、Wet系企業が設計ソフトを開発しているため、生物学者が中心となって作製したその解析手法は、何番目にどの塩基があると良い等の恣意的な基準を適用するというレベルに留まっており、統計学的な視点が欠けているために精度に問題があるのだと考えた。

3. WetとDryの融合による、課題の解決

3.1 開発プロジェクトの概要

そこで我々は、Wet系の企業だけでなく、バイオインフォマティクスの分野での経験を持つDry系の企業との共同開発であれば、高精度の設計ソフトを開発可能であると考えた。RNAi分野で共同開発先となるIT企業を探していた大阪大学医学系研究科の金田教授を中心に、ジェノメディア㈱、産業技術総合研究所の湯元先生、及びMSSでコンソーシアムを結成、2003年より共同開発を開始した(平成15、16年度経済産業省の地域新生コンソーシアム事業として予算を獲得)。MSSの役割は、主にジェノメディア㈱によって解析された実験データについて解析を行い、その結果を次回の実験フェーズへとフィードバックしながら、アルゴリズムの開発を行うというものである。

3.2 開発の方針

開発の方針は、他社に見られるような恣意的な基準ではなく、これまでの事業で培ってきた配列解析技術を活用すること、また、最新の知見を取り入れられるような柔軟なアルゴリズム設計とすることの2つとした。実際の解析データとしては以下のようなレポーター遺伝子を用いてsiRNAを導入した際の発現抑制を計測したものをを用いた。

表1 使用したsiRNA

レポーター遺伝子	個数
pGL2	26
MDR1	50
β Gal	10
GL3	13

4. アルゴリズム

4.1 アルゴリズムの概要

標的遺伝子について、21bp長となる全配列を取り出してその各々についてサブスコアを計算し、サブスコアごとに重み付け係数を掛けて合計することでトータルスコアを算出したものを各配列の活性予測値とする。現在は後述する5つのサブスコアを用いているが、将来、新しいデータが発表された場合には、それらを取り入れあるいは、不要なサブスコアを削除することにより、柔軟な設計を行うことができる。また、基本は21bpでの設計を行うが、配列長は20~30bpまでと自由に設定できるようにした。

4.2 アルゴリズムの詳細

以下、各サブスコアについて詳述する。

4.2.1 GC Score

siRNAのGC含量を基にしたスコアである。RNA鎖間の結合力は、A-Uのペアに比べてG-Cのペアの方が高く、GC含量はsiRNAとmRNAとの結合力と相関がある。この結合力はsiRNAとmRNAの結合のしやすさや、RISCによるmRNAの切断の受けやすさ等に影響を及ぼしている可能性が高く、このためGC含量は重要なパラメータの1つとされており、30%~50%で最も活性が高くなる。

4.2.2 Load Score

siRNAの、RISCへの導入効率をスコア化したものである。siRNAの二本鎖のうち片方の鎖がRISCへ取り込まれ、この取り込まれた鎖と相補的な配列を持つmRNAが切断される。そのため、二本鎖のうちどちらの鎖が取り込まれるかによって、標的となるmRNA配列が変わることになる。二本鎖の配列によって、二本鎖のそれぞれが取り込まれる確率に差がでることが実験で示されており^(5,6)、エネルギー的に不安定な端が5'末端となっている方のRNA鎖の方が優先的に取り込まれる(図2)。これは、不安定な端の方が、よりRISCによ

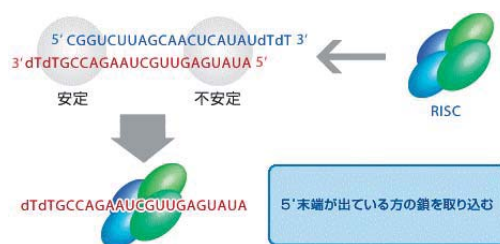


図2 siRNAの内部不安定性とRISCとの関係

二本鎖をほどかれやすく、RISCは二本鎖をほどいた後に5'末端の方の鎖を取り込むためであるという仮説によって説明されている。

siRNAのセンス鎖とアンチセンス鎖のうち、実際に標的の遺伝子配列と相補的なのはアンチセンス鎖のみであるので、アンチセンス鎖がより取り込まれやすいほど、siRNAが実際に標的の切断に使用される確率が高いことになる。そのため、アンチセンス鎖が取り込まれやすいものを優先的に選択するように、二本鎖の両端の自由エネルギーの数値を計算してその差をスコア化し、アンチセンス鎖の5'末端が不安定であるほどスコアを高く計算する。

4.2.3 Position Score

標的遺伝子中でのsiRNAの位置による影響をスコア化したものである。siRNAが標的とするmRNAは細胞中で複雑な二次構造を取るため、mRNA中の箇所によりsiRNAの活性が影響を受ける。Position Scoreはこの位置による効果をスコア化したものであり、5'UTR⁽⁷⁾、転写開始点から50bp⁽¹⁾は活性が悪くスコアが減点される等の基準がある。

4.2.4 Specificity Score

siRNAの標的遺伝子に対する配列特異性をスコア化したものである。一般的に、siRNAの配列特異性は高く、配列が完全一致する標的遺伝子だけを抑制すると言われているが、配列が数塩基異なる非標的遺伝子の発現にも影響を与えてしまう例が報告されており⁽⁸⁾、このsiRNAの非特異的な効果はoff-target effectと呼ばれ、実験データの質の低下を招く原因ともなり、siRNAを設計する際の問題点の一つとなっている。そこで、本アルゴリズムでは、siRNAと4塩基以上のミスマッチがある遺伝子には影響が及ばないと仮定し、siRNAとのミスマッチが3塩基以内となる遺伝子の数によって特異性をスコア化する。ミスマッチが3塩基以内の遺伝子数が少ないほど、そのsiRNAは特異性が高いということなのでスコアは高くなる。

ミスマッチの計算は、NCBIのUniGene⁽⁹⁾をデータベース、全候補siRNAの配列をクエリーとしたBLAST検索によって行っている。BLAST検索の結果を解析し、3つ以内のミスマッチで相同な遺伝子数を求め、その遺伝子数によってSpecificity Scoreを計算する。またBLASTでは、Word-lengthとの関係で、ある位置に非相同な塩基がある場合には相同性を低く見積もってしまう場合があるため、速度はBLASTと比較して遅いが精度の高いSmith-Waterman法を用いることもできる。

4.2.5 Profile Score

活性のあったsiRNAに共通する配列的特徴をプロファイルとして抽出し、これに対する適合度をスコア化したものである。RISCによる切断の際に、siRNAの持つ構造によってその活性は影響を受けるが、RISCによる切断の原子レベルでの反応の詳細が現状で不明であるため、どういった影響があるのかは分からない。したがって、生化学的な原理から演繹的に配列を決定することは困難である。そのため、活性のあった多数の配列から帰納的にこのような配列ならば活性があるはずだというアプローチを採るのが最も有効である。そこで、図3のように、過去の実験データのうち活性のあったものの配列を参考にして、その配列的な特徴を図3下部のような配列プロファイルとして作成しこれを実際の設計に活用するという方法を用いた。

4.3 アルゴリズムの検証

表1のデータを用いて、上記アルゴリズムを用いて計算した予測スコアと実際の活性との相関を見たところ図4のようになった。相関係数は0.65であった。

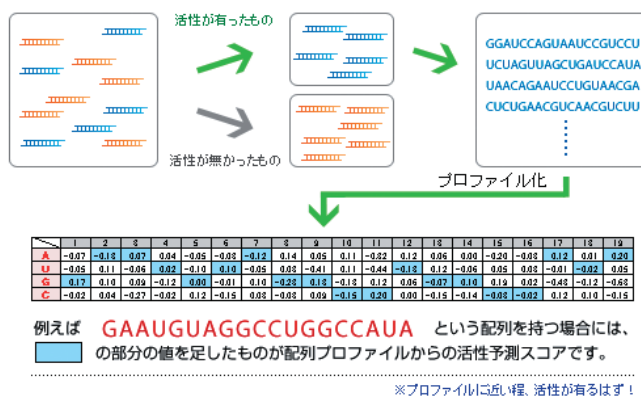


図3 Profile Scoreの概念図

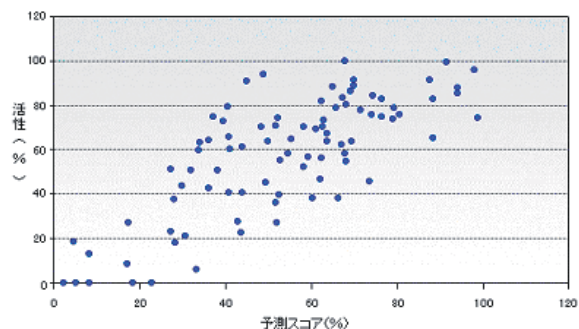


図4 予測スコアと活性の相関図

5. Webシステムとしての実装：SISNIPER

上述したアルゴリズムをシステムとして実装した(製品名：SISNIPER)。実装にあたっては、基本的に合成メーカー向けのパッケージ製品とし、合成メーカー内で想定されるニーズ、使い勝手等を考慮した設計を行った。まず、大まかなユースケースは、受注を受けた合成メーカーの技術者が、受注を受けた遺伝子について合成するための2～3本のsiRNAの配列を決定するというものである。複数の技術者が個人の端末からアクセスすること、各技術者は多数の遺伝子について設計を行うことから、Webベースのシステムとし技術者ごとのアカウントを管理できるようにした。さらに、設計結果はDBに保存し、過去の設計履歴も簡単に閲覧できるようにした。

具体的なデータの流れを図5に示した。BLASTによる処理には数分程度の時間がかかることがあるが、

BLAST以外の設計は1秒以内で終了する。また、BLASTを実行する前に何回か設計を行い、BLAST以外のパラメーターの最適化を行いたいというニーズがあることを考慮して、BLAST以外の処理と、BLAST自体をプロセスとして分離し、設計を開始すると一旦BLAST以外の検索結果を表示し、ユーザーが任意でBLASTを行うことができるような仕様にした。また、BlastMachineを用いてBLAST検索を行うことにより、高速な処理を可能とした。またその他の機能として、ファミリー遺伝子を共通にロックダウンするsiRNAを設計するため、Clustal W⁽¹⁰⁾を用いて複数配列の共通部分に相当する配列のみを設計する機能、データの増加に伴うプロフィールの変更を容易にするためのプロフィール自動作成機能、BLASTまでの自動実行機能等を実装した。図6はSISNIPERの画面例である。

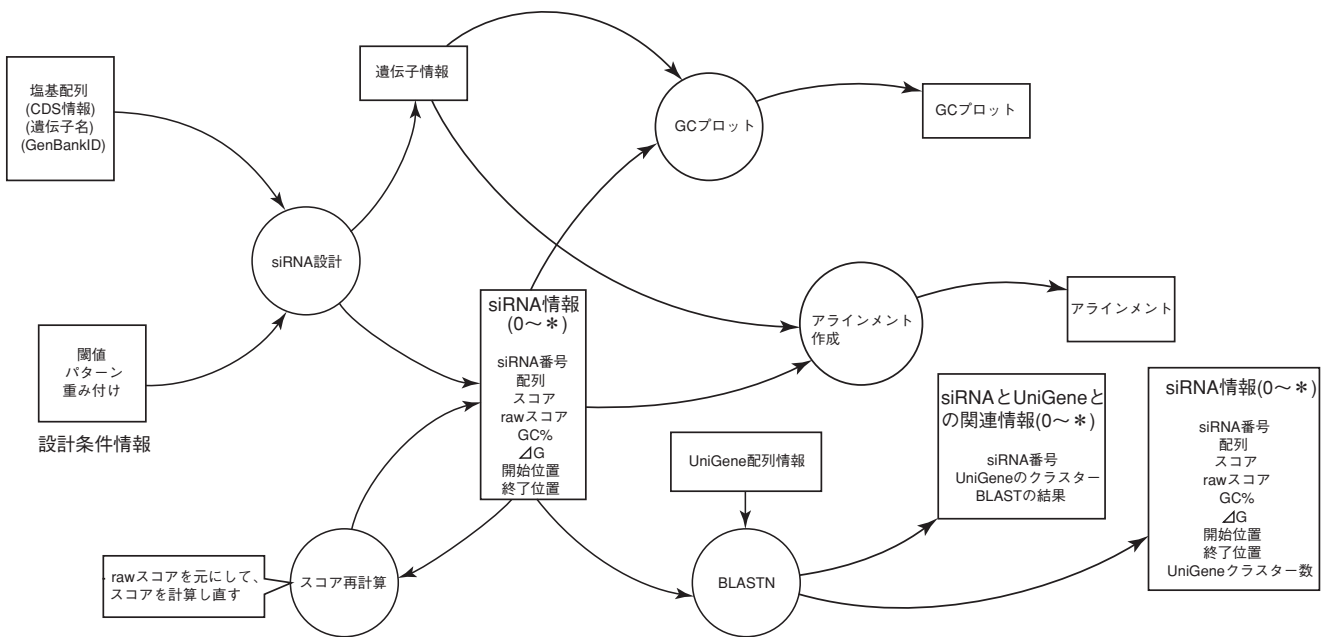


図5 SISNIPERでのデータの流れ



図6 SISNIPERの画面例

6. 今後の展望

現在、合成メーカー2社と共同でビジネスを展開しており、設計サービスを2社に提供している。既に合成、設計市場とも競争が激化しているため販売数は伸び悩んでおり、今後現状を打破するような対策が必要となる。さらに、今までは21bpより長いsiRNAは毒性があると言われていたのが、27bpなどの長さのsiRNAでも毒性がなく、活性もより高いというデータが出される⁽¹¹⁾など、技術の進展は予断を許さない状況である。そこで我々の対策であるが、まず、設計の精度が本システムの要であるため、今後ともジェノメディア株とともに27bpのものも含めたデータの収集に努め、予測精度のさらなる向上を目指すことを第一の目標とする。また現在は、Webベース、結果のDB化等、合成メーカー向けの仕様となっている。今後は、製薬企業、研究機関等への拡販も視野に入れるため、純粋な設計機能だけでなく、何らかの付加価値を付けることで導入をしやすくなるような開発を行う必要がある。例えば、弊社製品であるBioINTEGRA^{※2}、MedRodeo^{※2}と連携させることにより、これらのシステムで探索された興味深い遺伝子について、SISNIPERを使って即座に実験のためのsiRNAを設計、発注できる仕組み等を想定している。また、現在ジェノメディア株と開発を計画しているターゲット遺伝子探索システムについても、SISNIPERと連携させることを考慮した設計を行いたい。

※2 BioINTEGRA、MedRodeoは、三菱スペース・ソフトウェア株の登録商標です。

参考文献

- (1) S.M. Elbashir, et. al.(2001)RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15:188-200
- (2) The European Molecular Open Software Suite (<http://emboss.sourceforge.net/>)
- (3) Reynolds A, et. al(2004)Rational siRNA design for RNA interference. *Nat. Biotechnol.* 22(3) : 326-30
- (4) Naito Y, et. al.(2004)siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.* 32: W124-W129
- (5) Schwarz D.S., et. al.(2003)Asymmetry in the assembly of the RNAi enzyme complex. *Cell.* 155:199-208
- (6) Khvorova A., et. al.(2003)Functional siRNAs and miRNAs exhibit strand bias. *Cell.* 155:209-216
- (7) A.C. Hsieh, et. al.(2004)A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* 32:893-901
- (8) P.C. Scacheri, et. al.(2004)Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 101:1892-1897
- (9) UniGene (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>)
- (10) Clustal W (<http://www.ebi.ac.uk/clustalw/index.html>)
- (11) D.H. Kim, et. al.(2005)Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.* 23:222-226