

学 位 論 文 題 名

RNA interference の効率を決定するパラメータの
解析と siRNA 効果の定量的予測

学位論文内容の要旨

RNA interference (RNAi) は、短い 2 本鎖 RNA (dsRNA) によって遺伝子発現が特異的に抑制 (サイレンシング) される現象である。2 塩基の 3'-オーバーハングを持つ 21 塩基対の siRNA (short interfering RNA) は、標的 mRNA の相補性の高い部位に結合してそれを切断することにより標的遺伝子の特異的にサイレンシングする。一方、ゲノムの遺伝子領域に内在的にコードされている miRNA (micro RNA) も、相補性の高い部位を持つ mRNA を切断する作用を有しており、これらは作用機序の一部を共有していると考えられる。

これまでの研究では、特定の siRNA が特定の標的 mRNA をサイレンシングする程度を定量的に予測することはできなかった。本研究では、これまでに報告された多くの siRNA の効果データを収集し、更に新たな実験データと miRNA の研究成果も付加した上でコンピュータによる幅広い検索と分析を行うことにより、siRNA の効果のより正確で定量的な予測手法を開発することを目的とした。

まず、68 個の mRNA を標的とする合計 679 個の siRNA を新規にデザインし、細胞に導入後リアルタイム RT-PCR で mRNA レベルを定量した。遺伝子サイレンシングの度合いに応じて各 mRNA の F 値 (mRNA レベルでの%抑制) を決定した。次に、これまでに発表された文献や公開情報から siRNA の効果データを集め、827 個の siRNA について F 値を得た。更に、文献から既知の miRNA 492 個の成熟型ガイド鎖についての遺伝子サイレンシング情報を得た。それらの配列データと F 値を使用して siRNA / miRNA の配列情報と F 値との関係を解析した。その結果、高い相関関係を持つ以下の 5 つの主要パラメータを見出した。

1. 'siRNA 両末端の自由エネルギーバランス'

679 個の siRNA の両末端の閉鎖自由エネルギーの組み合わせを解析したところ、siRNA の 2 本鎖の両端の閉鎖自由エネルギーの比である '5'-opening factor (R)' が siRNA 効果と高い相関関係を持つことを見出した。

$$R = R_1^{0.574} R_2^{0.426} \cong \sqrt{R_1 R_2} \quad \text{または} \quad \log R = \log(R_1 R_2)$$

$$\text{ただし, } R_1 = \frac{\Delta G_{18}}{\Delta G_1}, R_2 = \frac{\Delta G_{18} + \Delta G_{17}}{\Delta G_1 + \Delta G_2}, \Delta G_i \text{ は隣接塩基自由エネルギー.}$$

miRNA および効果の高い siRNA はガイド鎖の 5'末端の ΔG が高く、3'末端では低い。これは、5'末端のらせん部分が熱力学的に不安定であればあるほど 2 本鎖が解けやすく RISC (RNA-induced silencing complex) と miRNA / siRNA の結合が起こりやすいことを示唆している。また、センス鎖ではなくガイド鎖が RISC に取り込まれることが重要で、そのために 3'末端は 5'末端より相対的に低いことが求められると推測できる。

2. ‘ヌクレオチド 13~15 塩基対の自由エネルギーの相対的高低差’

ガイド鎖の 3' 側半分のヌクレオチド 13~15 を中心とする部分とその周辺部分との ΔG の相対的な高低差である ‘RISC binding factor (R_{13})’ と siRNA の効果との間に有意な相関関係を見出した。

$$R_{13} = \left(\sum_{i=10}^{12} \Delta G_i + \sum_{i=16}^{18} \Delta G_i \right) / \sum_{i=13}^{15} \Delta G_i \quad \text{または} \quad \log R_{13} = \log(R_{13})$$

これは当該部分が標的認識のためではなく、その後工程の RISC との結合または RITS 複合体 (RNA-induced initiation to transcriptional gene silencing complex) の作用に関係しているためと考えられる。

3. ‘短い回文の出現頻度’

配列モチーフに注目した結果、siRNA 配列中の ‘AU’, ‘CG’ といった短い回文の出現頻度 ‘Palindromic Score (PS)’ は、siRNA の効果と負の相関関係があることを見出した。このような siRNA は自由エネルギーが最小となる構造で安定しておらず、短い回文によって絶えず小さなバルジができては消えるような振動状態が引き起こされることが効果の低下に影響している可能性がある。

4. ‘標的 mRNA の二次構造’

mRNA の二次構造による遮蔽効果について検討するため、標的部位の近傍および離れた部分の閉塞自由エネルギーを計算によりパラメータ化したシールスコア ‘sealScore(SC)’ を求めたところ、siRNA の効果と高い相関関係が認められた。これは siRNA が RISC に取り込まれて RITS 複合体となった後、標的 mRNA へできる限り接近し、近接状態を保たなければ mRNA を切断できないためと推測される。

5. ‘オンターゲットとオフターゲット部位数のバランス’

siRNA が標的としていないオフターゲット mRNA 中にも偶然に標的部位が存在するケースがある。miRNA の実験データの精査により、標的部位たる条件を検索し、オンターゲット部位とオフターゲット部位の出現率のバランスを競合特異度 ‘specificity ratio (Rs)’ として求めたところ、siRNA の効果との高い相関関係を見出した。

$$Rs = TS/NS \quad \text{ただし、} TS \text{ は 1kb 毎のオンターゲット部位出現数、} NS \text{ は 1kb 毎のオフターゲット部位出現数。}$$

これは、siRNA の偏在的な反応により siRNA 分子の競合消費が引き起こされ、本来狙った標的に対して十分な反応を行えないためと考えられる。

また、これまでの研究者によって使用されてきたパラメータの多くは、そのままでは siRNA の効果とわずかな相関しか認められないことが判明したため、パラメータごとに閾値を設定しその超過値を補助パラメータとして採用した。

このような解析の結果得られた主要 5 パラメータと補助パラメータを使用して機械学習を行った。学習アルゴリズムである最小二乗サポートベクターマシン (least squares support vector machine) を使用し、トレーニングデータとして 679 個の siRNA データセットについて主要、補助パラメータおよび対応する F 値を学習させ、残りの 827 個の siRNA データセットの効果を予測し検証を行った結果、良好な回帰結果を得ることができ、配列特異的な siRNA の効果を一定の精度で定量的に予測できることが示された。

siRNA の作用機序の中には 2 本鎖の巻き戻し、RISC との結合、mRNA への近接など、siRNA が細胞内に取り込まれてから最終的に mRNA を切断し RNAi 効果を引き起こすまでの各ステッ

プにおいて実行される一連の作用が含まれている。本研究では、これらの作用を理解して関係要素を駆使することにより、siRNA 効果の高い精度での定量的な予測が可能であることを明らかにした。すなわち、効果が高い siRNA は、それら一連の作用すべてにおいて、それぞれのハードルをうまく乗り越えることができる特徴を併せ持ったものであると考えられた。また、長い進化の過程で完成されてきた miRNA での実験で得られる結果は、siRNA の効果の予測精度を高めるために有用であると考えられた。

学位論文審査の要旨

主 査 教 授 守 内 哲 也

副 査 教 授 畠 山 鎮 次

副 査 教 授 田 中 一 馬

学 位 論 文 題 名

RNA interference の効率を決定するパラメータの 解析と siRNA 効果の定量的予測

RNA interference (RNAi) は、短い2本鎖 RNA (dsRNA) によって遺伝子発現が特異的に抑制 (サイレンシング) される現象である。2 塩基の 3' -オーバーハングを持つ 21 塩基対の siRNA (short interfering RNA) は、標的 mRNA の相補性の高い部位に結合してそれを切断することにより標的遺伝子の特異的にサイレンシングする。これまでの研究では、特定の siRNA が特定の標的 mRNA をサイレンシングする程度を定量的に予測することはできなかった。本研究では、これまでに報告された多くの siRNA の効果データを収集し、更に新たな実験データと miRNA の研究成果も付加した上でコンピュータによる幅広い検索と分析を行うことにより、siRNA の効果のより正確で定量的な予測手法を開発することを目的とした。

本研究では、679 個の siRNA を新規にデザインして細胞に導入後、リアルタイム RT-PCR で mRNA レベルを定量することにより siRNA の効果値を決定した。次に、これまでに発表された文献や公開情報から siRNA の効果データを集め、827 個の siRNA について効果値を得た。更に、文献から既知の miRNA 492 個の成熟型ガイド鎖についての遺伝子サイレンシング情報を得た。それらの配列データと効果値を解析した結果、高い相関関係を持つ以下の 5 つの主要パラメータを見出した。① '5'-opening factor'. siRNA 両末端の自由エネルギーバランスで、二本鎖 RNA 末端のらせん部分の熱力学的不安定性が RISC (RNA-induced silencing complex) との結合、および取り込まれる鎖の選択に関わっていると推測される。② 'RISC binding factor'. ヌクレオチド 13~15 塩基対の自由エネルギーの相対的高低差で、当該部分は標的認識のためではなく、その後工程の RISC との結合作用に関係しているためと考えられる。③ 'Palindromic Score (PS)' (負相関). siRNA 配列中の 'AU', 'CG' 等の短い回文の出現頻度で、siRNA の安定性の低下が効果の低下に影響している可能性がある。④ 'sealScore' (負相関). 標的 mRNA の二次構造による遮蔽効果をパラメータ化したもので、siRNA が RISC に取り込まれた後、標的 mRNA へ接近できなければ mRNA を切断できないためと推測される。⑤ 'specificity ratio' (負相関). オンターゲットとオフターゲットの部位数の比で、siRNA の偏在的な反応により siRNA 分子の競合消費が影響しているためと考えられる。

これらのパラメータと補助パラメータを使用して機械学習を行った結果、良好な回帰結果が得られ、siRNA の効果を一定の精度で定量的に予測できることが示された。

siRNA の作用機序の中には2本鎖の巻き戻し、RISC との結合、mRNA への近接など、最終的に mRNA を切断し RNAi 効果を引き起こすまでの各ステップにおいて実行される一連の作用が含まれている。本研究では、これらの作用を理解して関係要素を駆使することにより、siRNA 効果の高い精度での定量的な予測が可能であることを明らかにした。また、長い進化の過程で完成されてきた miRNA での実験で得られる結果は、siRNA の効果の予測精度を高めるために有用であると考えられた。

公開発表において、副査の畠山教授から、1) タンパクレベルでの発現抑制データの有無、2) 使用データの公開の意思について、3) 実用化に関する考え方について質問があった。続いて副査の田中教授から 1) 予測不適合ケースの原因について、2) 未知のタンパクとの相互作用が関与する可能性について質問があった。最後に主査の守内教授から、1) 最重要パラメータはどれか、2) 開発したシステムの実用例の有無、3) 複数の標的について、4) 本研究の実地応用における意義などについての質問があった。いずれの質問に対しても、申請者は主旨をよく理解し自らの実験データと文献的考察を交えて概ね適切な回答を行った。

この論文は、siRNA の効果を定量的に予測する手法を見出した点で高く評価され、今後この成果は RNAi メカニズムの解明や核酸医薬分野に大いに役立つことが期待される。

審査員一同は、これらの成果を高く評価し、大学院課程における研鑽や取得単位なども併せ申請者が博士（医学）の学位を受けるのに十分な資格を有するものと判定した。