

Title	低分子の結合標的RNAモチーフの新規探索手法の開発 および核酸-低分子相互作用の1分子計測による観測
Author(s)	高島, 裕介
Citation	大阪大学, 2023, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/92169
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

博士論文

低分子の結合標的 RNA モチーフの新規探
索手法の開発および核酸-低分子相互作用
の 1 分子計測による観測

大阪大学大学院
理学研究科 化学専攻

高島 裕介

目次

序章	1
0-1 はじめに	1
0-2 RNA を標的とした創薬	1
0-3 DNA/RNA 結合性低分子	2
0-4 RNA – 低分子データベースについて	3
0-5 ハイスループットシーケンシング	4
第一章 DICER 酵素切断を活用した低分子の結合標的 RNA モチーフの 新規探索手法の開発	6
序論	6
1-1 研究背景	6
結果と考察	7
1-2. Pre-miRNA mutants ライブラリーのシーケンシング解析	7
1-3 NCD を用いた配列解析	13
1-4 NGS で得られた配列の実際の効果の確認	18
1-5 内在性 miRNA への応用	29
1-6 FDA 承認薬とその関連誘導体への適用	31
1-7 大規模データ取得を志向した検討	34
結論	37
EXPERIMENTAL SECTION	38
第二章 水素結合を介した核酸塩基-低分子間相互作用の 1 分子計測による観測	42
序論	42
2-1 研究背景	42
結果と考察	43
2-2 グアニン-リガンド間の単分子カウント解析	43
2-3 グアニン-リガンド相互作用における結合の定量的評価	48
2-4 核酸塩基-リガンド相互作用における選択性の定量的評価	54
2-5 グアニン-リガンド相互作用における水素結合モードの定量的評価	56
2-6 解離定数の比較	57
結論	60
EXPERIMENTAL SECTION	61
総括	64

主論文.....	65
参考論文.....	65
引用文献.....	65
補足資料.....	78
謝辞.....	109

略語表

A	adenine	アデニン
ATP	adenosine triphosphate	アデノシン三リン酸
C	cytosine	シトシン
cDNA	complementary DNA	相補的 DNA
G	guanine	グアニン
miRNA	microRNA	マイクロ RNA
mRNA	messenger RNA	メッセンジャー RNA
ncRNA	non-coding RNA	非翻訳 RNA
NGS	Next generation sequencer	次世代シーケンサー
PCR	polymerase chain reaction	ポリメラーゼ連鎖反応
pre-miRNA	precursor microRNA	マイクロ RNA 前駆体
RNase	ribonuclease	リボヌクレアーゼ
T	thymine	チミン
U	uracil	ウラシル

序章

0-1 はじめに

21 世紀は RNA 研究にとって大躍進の世紀といわれるように、現在において RNA は、生命科学研究のみならず、医学の分野でも研究が盛んに行われるまでになった。筆者が所属する大阪大学産業科学研究所精密制御化学研究室（以下当研究室）も同様に、以前から RNA を標的とした低分子医薬の開発に取り組んできた。しかし、RNA を標的とする低分子設計に明確な指針が見いだされていないことから、RNA を標的とした低分子創薬研究は遅々として進んでいないのが現状である。RNA を標的とした低分子創薬を加速するためには、低分子-RNA 複合体ペアの情報を大規模に収集し、低分子設計の手がかりとなる情報を抽出することが必要であり、RNA の配列・構造の多様性を反映したアッセイ系の確立と、低分子-RNA の相互作用を検出する基盤技術の開発が不可欠である。

このような背景を鑑み、筆者は研究主題として「核酸-低分子相関の観測」を掲げ、これを研究の軸として「低分子の標的となる RNA 配列の網羅的な解析」および「核酸塩基-低分子会合状態の 1 分子レベルでの観測」に挑戦した。いずれの課題も生命科学、創薬分野においてこれまで検討されたことのない「核酸-低分子相関」を観測する新規手法である。これらのテーマに基づき RNA-低分子相関情報の取得を行うことで、将来的に科学的な根拠を背景とした RNA 標的の低分子創薬研究の発展に寄与できると期待した。

本論文ではまず始めに「核酸-低分子相関の観測」に対する重要な概念と研究背景を述べ、実際に行った研究の目的・研究背景・詳細な検討については各章にて記述する。

0-2 RNA を標的とした創薬

従来、RNA はセントラルドグマの流れで遺伝情報を伝達する際の DNA とタンパク質の中間体であると考えられてきたが、2012 年、Encyclopedia of DNA Elements (ENCODE) プロジェクト¹によって、ヒトゲノムの約 30 億塩基対のうち 70%程度は RNA に転写されており、その多くはタンパク質に翻訳されない「非翻訳 RNA (non-coding RNA : ncRNA)」として存在していることが明らかとなった²⁻⁷。さらに、これら ncRNA の中でも特別な機能を有しているものは機能性 ncRNA と呼ばれ、転写や翻訳、エピジェネティクスなどの制御を介して、発生や分化、がん化など、多くの生命現象に関わっていることが分かっており⁸、低分子を用いて特定の ncRNA を標的とし、その機能を調節する戦略は、細胞生物学、医学、バイオテクノロジー、化学など様々な研究分野の研究者の関心を集めている⁹⁻¹³ (Figure 0-1)。

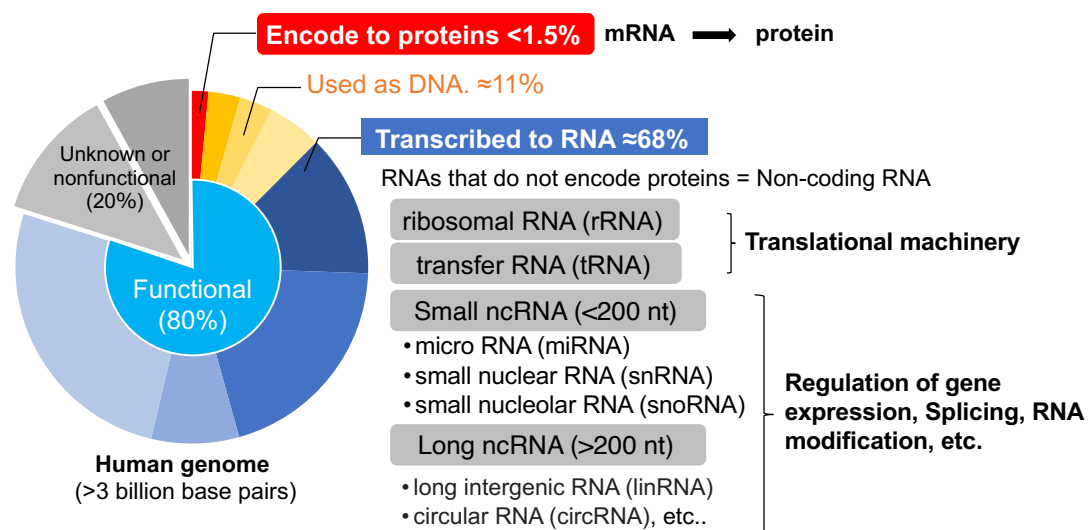


Figure 0-1. Composition of the human genome

RNA を標的とする創薬では RNA 干渉¹⁴ を利用する短鎖二本鎖 RNA (Short interference RNA: siRNA) または長鎖二本鎖 RNA (double-strand RNA: dsRNA) や、アンチセンスオリゴヌクレオチドなどの核酸医薬の研究が盛んであるが、本研究では製造が比較的容易であり、長年にわたって多く使用され、広範囲に渡って研究されてきた低分子医薬品に焦点を当てた。

低分子創薬研究は創薬企業が長年手掛けてきた歴史があり、創薬企業内に多数の化合物ライブラリーや製剤技術などが蓄積されている。つまり、核酸医薬の核酸を低分子に置き換え、核酸標的的低分子創薬が可能となれば、企業内に蓄積された技術を展開でき、より安価かつ高効率な創薬が可能となる。実際に、2020年8月7日に、Roche グループが開発した RNA 結合性の低分子 Risdiplam¹⁵ が SMA 治療薬として FDA に承認されるなど、近年、ncRNA の重要性の認識が進むに連れ、RNA を標的とした低分子の探索、開発研究が急速に進展している。

0-3 DNA/RNA 結合性低分子

当研究室では、以前より遺伝子発現制御の分子ツール等としての応用を目指し、特定の DNA 及び RNA の特異構造に結合する低分子の設計と合成研究に取り組んできた。

NCD (Naphthyridine Carbamate Dimer) は、DNA の G-G ミスマッチ塩基対へ特異的に結合する分子である (Figure 0-2A)。NCD は CGG/CGG モチーフを含む DNA 配列に結合し、その結合にともなう複合体の安定化により、室温では一本鎖である DNA を接着させ二本鎖を形成させる「分子糊」としての機能をもっている

16-18。しかし、NCD を RNA へと適用した際には、NCD の CGG/CGG モチーフへの結合能が、DNA への結合よりも弱いため、RNA の分子糊として機能しなかった¹⁹。NCD のように当研究室で開発された化合物として、DNA の CAG/CAG モチーフ配列に対して結合する NA(0-2B) や²⁰、DNA の C/CC モチーフ配列に対して結合する ANP77(0-3C) などがあるが²¹、どれも限定的な DNA/RNA を対象としてデザイン、合成されており、標的とした DNA/RNA 以外への結合評価はあまり試されていない。しかし近年、筆者らの研究グループは偶然にも NCD が RNA の UGGAA リピート配列に対して高い結合親和性があることを発見した²²。このように、核酸を標的とした低分子には、デザイン対象とした配列以外へも結合し、作用を示す可能性が秘められている。

そこで、筆者は標的としていた配列以外への網羅的なアッセイによって、より強く結合する配列、全く想定されていなかった結合配列など、新規な RNA-低分子結合相関情報が得られるのではないかと考えた。

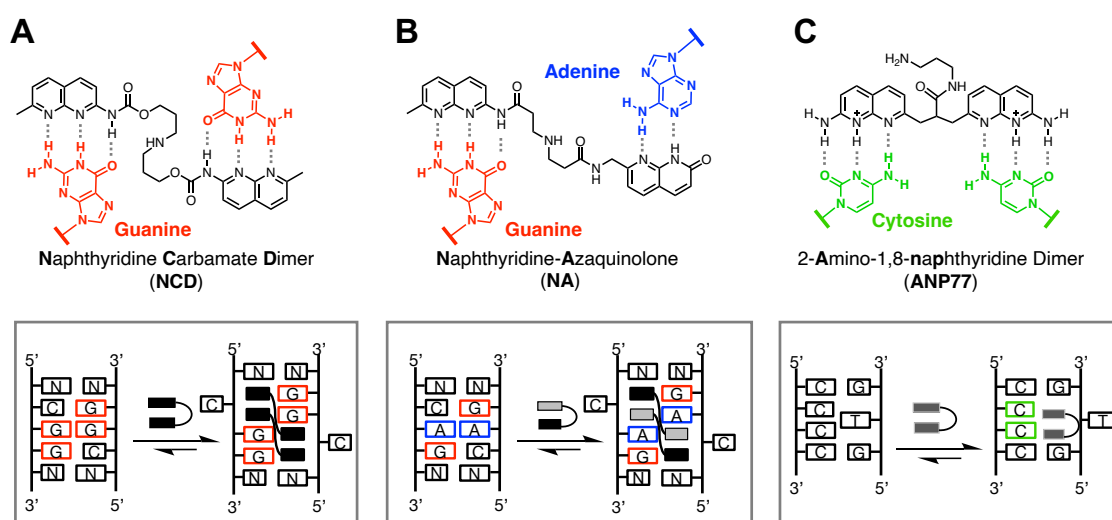


Figure 0-2. DNA binding ligands developed in our laboratory and their putative binding motifs.

0-4 RNA-低分子データベースについて

近年、RNA 結合性の低分子が続々と報告されているが、これらは、化合物ライブラリーの *in vitro*²³⁻²⁵ および *in-silico*²⁶⁻²⁸ スクリーニングにより同定され、天然物の構造改変により取得されたものである²⁹⁻³¹。このように、核酸を標的とした低分子の分子設計研究は活発化してきているが^{26,32-34}、標的 RNA と高い親和性と選択性で結合する分子の同定は依然として課題となっている。これは、低分子-RNA 相互作用に関する理解が限られていることが原因である。

低分子-RNA 相互作用の本質を理解するためには、低分子の RNA 結合ペアを包括的に収集し、分析することが極めて重要となる。そこで、スクリーニングによって得られた情報をまとめ、低分子-RNA 相関情報のデータベース化を図る動きも活発化してきている。例えば Hargrove らの R-BIND³⁵ や Disney らの *informa*³⁶ などが挙げられる。しかしこれらは特定の構造をもつ RNA に対する結合評価に基づく低分子-RNA の相関情報のデータベースであり、RNA の塩基ペアに着目した情報の獲得や、機能評価を元にした情報の獲得は未だ為されていない。そこで本研究では RNA-低分子相関の新たな情報獲得を目指し、第一章では低分子の標的となる RNA 配列を網羅的に獲得する新規機能性スクリーニング法の開発を行った。また第二章ではより詳細な 1 分子レベルにおける低分子と核酸の会合状態に着目し、1 分子スケールで低分子-核酸相関を観測するための新規手法の開発を行った。

0-5 ハイスループットシーケンシング

本研究では次世代シーケンサー (Next Generation Sequencer : NGS)^{37,38} を用いたハイスループットシーケンシング技術を活用した。

シーケンシングとは DNA (RNA) を構成するヌクレオチド (A, G, C, T (U)) の塩基配列を決定することであり、1977 年に開発されたサンガー法によるシーケンシングが一般的である³⁹⁻⁴¹。サンガー法とはヌクレオチドアナログ (ddNTPs) と DNA ポリメラーゼを用いて配列の決定を行う方法である。DNA 鎖は DNA ポリメラーゼによって dNTP が取り込まれ伸長するが、ランダムに取り込まれた ddNTPs は伸長反応を終了させるとともに、得られる DNA 断片を蛍光標識する。この DNA 断片の蛍光色素の種類と DNA 断片の長さをキャピラリー電気泳動装置で検出することで塩基配列を決定する手法がサンガー法である^{42,43}。一方、次世代シーケンサー (NGS) 技術は、ハイスループットであることを特徴とし、数百万の DNA 断片の同時解析が可能であり^{44,45}、ゲノム解析⁴⁶⁻⁴⁸、トランスクリプトームプロファイリング^{49,50}、エピゲノム解析⁵¹ など、さまざまな研究に応用されている。

例えば illumina 社の Genome Analyzer は、フローセルと呼ばれるスライドグラス上に断片化した 1 本鎖の DNA を張り付け、蛍光ラベルされたヌクレオチドを用いて断片の相補鎖を合成し、過程を画像解析することで塩基配列を決定しており、また Thermo Fisher Scientific 社の ion PGMTM シーケンサーはシーケンシング反応中の水素イオン放出を、それに伴う pH の変化により検出する、イオン半導体シーケンスにて塩基配列の決定を行っている。ion PGMTM の解析チップには数百万~数千万の小さなウェルがあり、それぞれのウェルが pH センサーとなっ

ているため、一回のランで大量の配列を決定することができる。断片化した多数の DNA を同時に読んでいく点では illumina 社の Genome Analyzer と同じであるが、核酸が取り込まれる際の pH の変化を電氣的に検出するため、蛍光を光学系、画像解析で読み取るのに比べ、特に連続する同一塩基の読み取りのエラーが少ないという特徴がある。

近年では NGS 技術を用いた合成 RNA ライブラリーのスクリーニングにより、RNA アプタマー^{52,53} や触媒 RNA^{53,54} が同定されている。NGS による RNA シーケンスは、低分子-RNA 相互作用の膨大なデータセットを集めることに応用されており、RNA を標的とする低分子の開発を促進することが期待されており、実際に NGS 等の検出方法とバイオインフォマティクス解析を組み合わせることで、RNA ライブラリーから低分子の標的 RNA モチーフが同定されている⁵⁵。

本論文第一章ではこのハイスループットシーケンシング技術を用いて、低分子と RNA の相互作用に関する情報を大規模に取得し、RNA を標的とする低分子の設計指針の獲得を目指した。また本研究第二章ではシーケンシング技術を応用し、核酸-低分子相互作用を同定する新たな解析手法の開発を行った。

第一章

Dicer 酵素切断を活用した低分子の結合標的 RNA モチーフの 新規探索手法の開発

序論

1-1 研究背景

代表的な ncRNA の 1 つである、マイクロ RNA (miRNA) は、1993 年に Lee らによって、線虫 (*C.elegans*) の発生段階で広範囲に渡り遺伝的制御経路を調節する内在性 RNA 断片として発見され⁵⁶, 2001 年に miRNA という学名が提唱された 21~25 塩基 (nt) の短い RNA である^{57,58}。miRNA の生合成過程を **Figure 1-1** に示した⁵⁹。まず初めに、非コード DNA 領域から転写された primary miRNA (pri-miRNA) が Drosha/DGCR8 複合体により、ヘアピン構造の miRNA 前駆体 (precursor miRNA : pre-miRNA) へとプロセッシングされる。次に、生成した pre-miRNA は核から細胞質へと輸送され、酵素 Dicer によって切断され二本鎖 miRNA (miRNA /miRNA*) が生成する^{60,61}。生成された二本鎖 miRNA は、その一方が RNA 認識タンパク質 Argonaute2 (Ago2) へとロードされ、Ago2 共に RNA-induced silencing complex (RISC) を形成し、特定の標的 mRNA の翻訳抑制調節に関与することが報告されている^{62,63}。

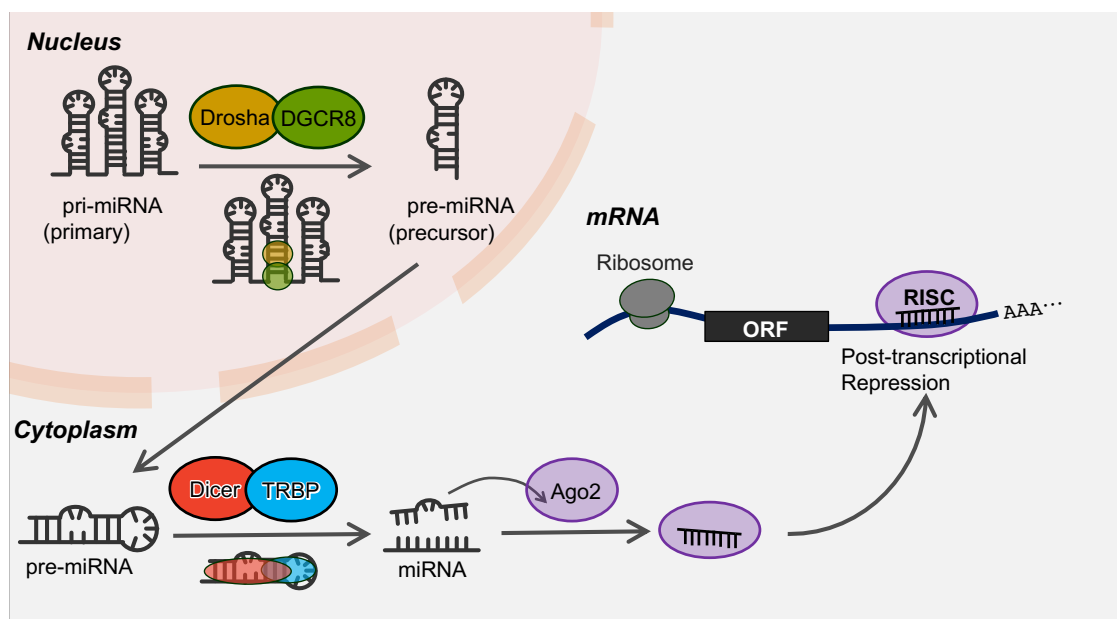


Figure 1-1. Maturation process and function of miRNA

したがって、特定の pre-miRNA の切断反応効率を変化させる低分子は、特定の領域の mRNA の翻訳を調節する薬となる可能性を有しており、また RNA を標的とした創薬におけるリード化合物になることが期待される⁶⁴。

既報において、当研究室では pre-miRNA に結合し miRNA の産生を抑制する化合物 BzDANP や RND を見出している⁶⁵。また Scripps の Disney らは pre-miRNA-525 や pre-miRNA-377 の前駆体に結合し、その産生を抑制する低分子を報告している⁶⁶。しかし、これらの研究は対象とする RNA や化合物が限定的であるため、約 1900 種類ある miRNA の成熟過程を標的とした低分子創薬を加速するためには、多数の標的 RNA と化合物の組み合わせによる、構造-Dicer 切断反応の相関情報の大規模収集が必要である。しかし、現時点においても、特定の RNA を標的とする低分子設計には、明確な指針が未だ見いだされていない。

そこで筆者は、pre-miRNA の Dicer による切断反応を指標することで、当研究室で開発されてきた低分子が結合する RNA 配列を網羅的に探索できないかと考えた。

結果と考察

1-2. Pre-miRNA mutants ライブラリーのシーケンシング解析

本研究を行うにあたり、まず実験のコントロール化合物として当研究室で合成された低分子である、NCD, NA, ANP77 を用いて解析を行った。本章では、その実験方法及び実際の実験結果を示す。

1-2-1 実験手法

本研究では、miR-136 の前駆体である pre-miR-136 を基本骨格とし、pre-miR-136 のループ部分に近い相補鎖部位の 5'末端, 3'末端側をそれぞれ 3 塩基ずつランダム化した、pre-miRNA mutants ライブラリーを設計した。この RNA ライブラリーを Dicer 反応に用いることで A, U, G, および C の 4 種類の塩基と 6 箇所の変異部位の組み合わせで計 4096 配列の混合物となった RNA を同時に解析できる。基本骨格として用いた pre-miRNA136 及び pre-miRNA mutants ライブラリーの推定構造を以下に示す (**Figure 1-2**)。pre-miR-136 は、筆者らの以前の研究で、Dicer により効率よく切断される基質であることが分かっている³⁴。本研究では Dicer の切断部位に近い 5'-GAU-3'/5'-AUC-3' (**Figure1-2** に太字で示す) の 6 塩基を 5'-N₁N₂N₃-3'/5'-N₄N₅N₆-3' としてランダム化して、4096 個の pre-miRNA からなる pre-miRNA mutants ライブラリーを作製した。黒三角は報告されている pre-miR-136 の Dicer による酵素切断部位を示している⁶⁷。

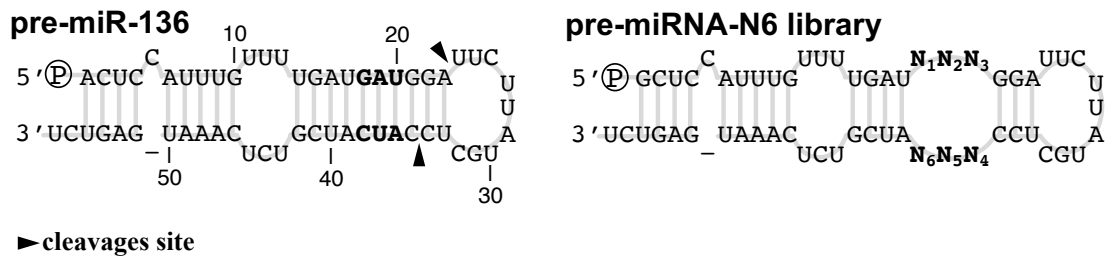


Figure 1-2. Sequences and predicted secondary structures of pre-miR-136 and pre-miRNA-N6 library.

解析手法の概要図を以下に示す(**Figure 1-3**)。解析には NGS を用い、ハイスループットシーケンスを行った。低分子の存在下/非存在下で Dicer 処理した pre-miRNA mutants ライブラリーの未切断基質を逆転写後 PCR によって増幅し、シーケンシング用のアダプター配列、バーコード配列を含む DNA ライブラリーを作製する。作製した DNA ライブラリーをシーケンス解析し、各 pre-miRNA mutants の読み取り数をカウントする。低分子化合物を添加したライブラリーと非添加のライブラリーの各変異体の読み取り数の差を統計学的に解析し、低分子化合物を添加した際の影響を調べた。

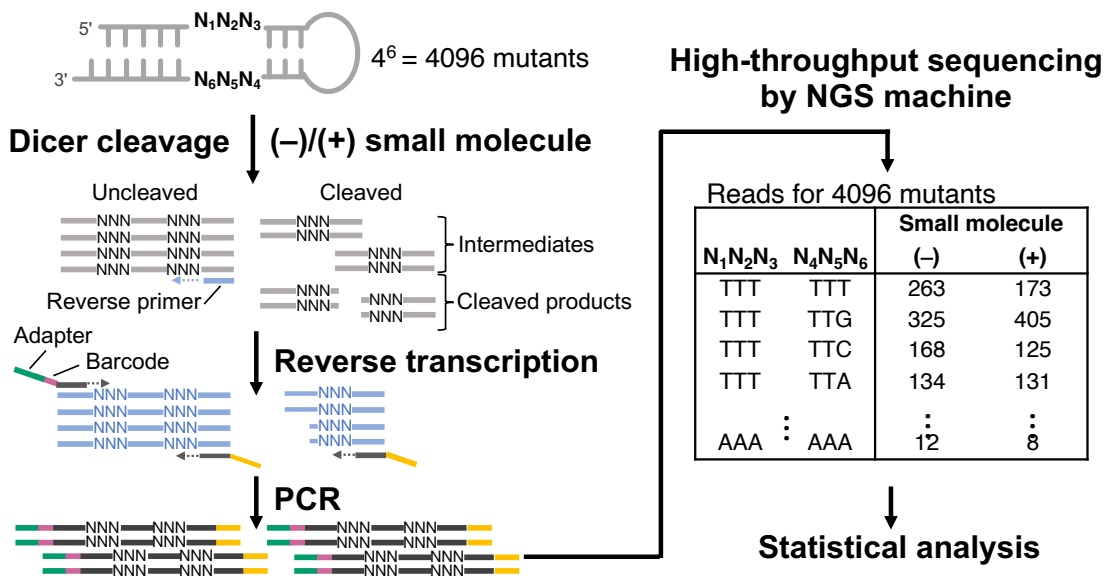


Figure 1-3. Outline of the method in this study.

1-2-2 Pre-miRNA mutants ライブラリーの *in vitro* Dicer 反応

低分子より Dicer 切断効率が増加する配列を探索するにあたり、5'末端がモノリン酸化された pre-miRNA mutants ライブラリーを用いて Dicer 反応を行った。反応はマイクロチューブ下、37 °C で 10 時間行い、反応終了後 Dicer stop solution

(Genlantis) を加え不活性化を行った。以下に反応産物分取し、変性ポリアクリルアミドゲル電気泳動 (以下, 変性 PAGE) にてゲル電気泳動解析を行った結果を示す (Figure 1-4)。NA, NCD, ANP77 のいずれの化合物を添加した際にも Dicer 反応は良好に進行し, 56 塩基長の基質である pre-miRNA mutants ライブラリーの切断と約 20 塩基長の miRNA の産生が確認できた。

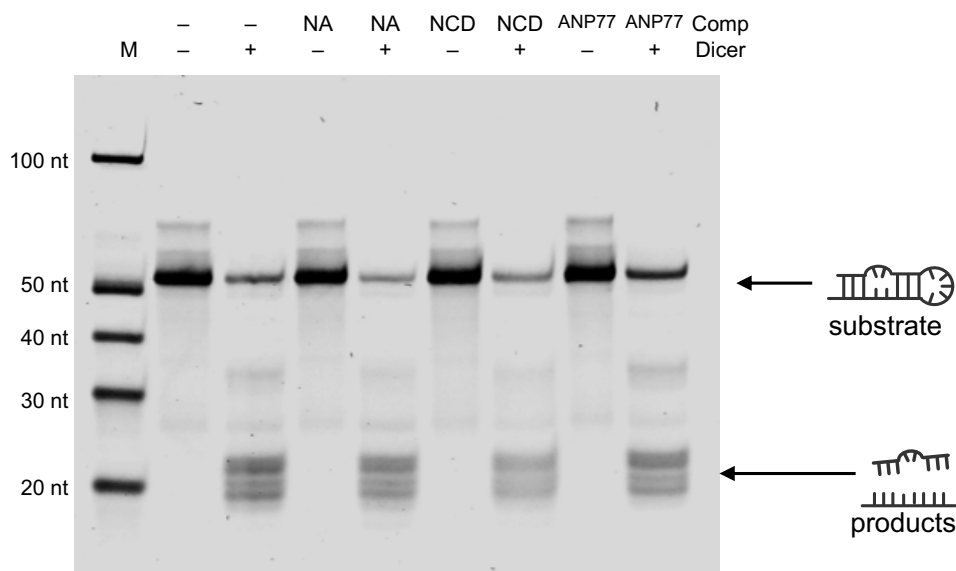


Figure 1-4. Denaturing-PAGE analysis of Dicer cleavage reaction of pre-miR-136 mutants Library.

Conditions : RNA (2 μ M), Compound (100 μ M), with 10x Buffer (Gelantis), $MgCl_2$ (10mM), ATP (1mM), BSA, and Dicer (180 nM) were incubated for 37 $^{\circ}C$, 10 h. Analysis : 15 % denaturing-PAGE, acrylamide: bis-acrylamide = 19: 1, 6 M Urea).

1-2-3 NGS 用 DNA ライブラリーの作製

化合物の存在下/非存在下で *in vitro* Dicer 反応を行ったのち, NGS による配列解析のために DNA ライブラリーの作製を行った (Figure 1-3)。プライマーとして使用したオリゴヌクレオチドを表に示す (Table 1-1)。

Table 1-1. The sequences of DNA oligos used in the study.

#	Oligos	Sequences
1	pr-miR136_F_BC2 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATGCTCCATTTGTTTTGA
2	pr-miR136_F_BC3 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATGCTCCATTTGTTTTGA
3	pr-miR136-F-BC4 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATGCTCCATTTGTTTTGA
4	pr-miR136-F-BC5 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGATGCTCCATTTGTTTTGA
5	pr-miR136-R-trp1 (Rev primer)	CCTCTCTATGGGCAGTCGGTGATAGACTCATTGAGACGA

はじめに, *in vitro* Dicer 反応の反応産物 RNA を, pre-miRNA mutants ライブラリーの 3'-arm 配列に部分的に相補的なプライマーを用いて逆転写酵素によって RNA から cDNA へと変換した。続いて, バーコード配列を含むフォワードプライマー (#1-4) とビーズ固定用の特別な配列を含むリバースプライマー (#5) を当て PCR 反応を行うことで, DNA の増幅とシーケンシング用アダプターの付加を行った。逆転写プライマーは, ライブラリーの 3'-arm 配列を含む未切断 RNA と切断 RNA に結合するが, 5'-arm 配列のみを含む切断 RNA には結合しない。PCR 増幅の際, フォワードプライマーは pre-miRNA mutants ライブラリーの 5'-arm 配列と部分的に相補的な配列を含む cDNA にのみ結合する。この結果, 未切断の RNA にのみに対応する cDNA が選択的に増幅される。PCR 産物は Nano View (*GE Healthcare Life Science*) でおおよその濃度測定したのち, Qubit dsDNA HS Assay Kits (*Thermo Fisher Scientific*) を使用して, ライブラリーの濃度を決定した。測定した濃度を元に, 各サンプル希釈を行い。またゲル電気泳動によってその純度の確認を行った (**Figure 1-5**)。

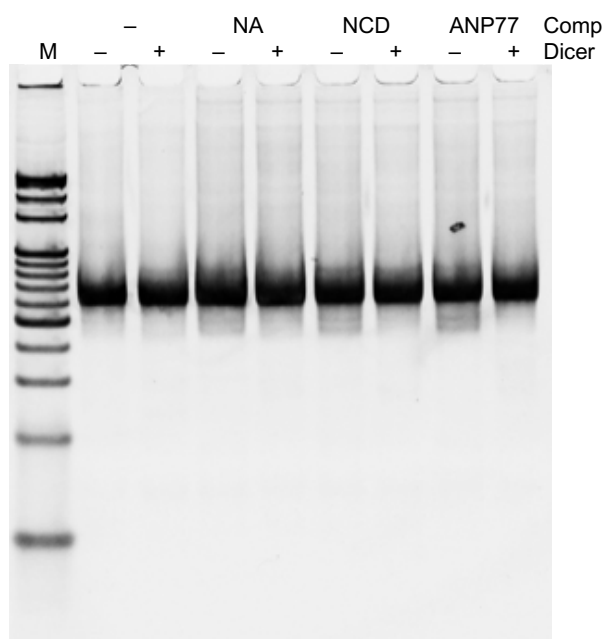


Figure 1-5. Denaturing-PAGE analysis of purity of PCR product.

Analysis : 15 % denaturing-PAGE, acrylamide: bis-acrylamide = 19: 1, 6 M Urea).

1-2-4 配列解析

得られた DNA ライブラリーは, Thermo Fisher Scientific 社の ion PGM で配列決定を行った。それぞれの結果には Dicer の有無, 低分子の有無によって 4 つのサンプルの結果が含まれている。本論文では, この 4 サンプルの内の比較したい 2 サンプルの結果をピックアップし, 各配列の読み取り数を基準に比較した。

今回の実験において低分子が Dicer 反応へ影響を及ぼさないならば, 理論上低分子の有無によって読み取り数に差は生じないはずである。しかし, 実際にはサンプルごとにばらつきが出る。今回は, そのバラツキが偶然起こり得たものであるか, もしくは低分子による有意な差であるのか確認する手法として, ピアソンのカイ二乗検定を用いた (Figure 1-6)。カイ二乗検定とは 2 つデータに関連性があるか否かを確かめる手法であり, 以下のステップで行われる。

- ・各配列の読み取り数をクロス集計表に当てはめる。
- ・式に従って期待値 a~d を求める。
- ・データと期待値との差 e~h (各セルの χ 二乗値) を求める。
- ・e~h を足し合わせ, χ 二乗値を算出する。
- ・ χ 二乗値を p 値に変換する

以上のステップで求めた p 値より, 有意差の検定を行う。p 値とは統計的仮説検定において, 帰無仮説の元で検定統計量はその値となる確率のことである。つまり, P 値が小さいほど, 検定統計量はその値となることはあまり起こりえないことを意味する。一般的に P 値が 5% または 1% 以下の場合に帰無仮説を偽として棄却し, 対立仮説を採択する。つまり, P 値が 0.05 (5%) を下回った場合, その P 値は偶然取る値ではないということの意味し, 「極めて珍しいことが起きた」または「何かしら意味がある事象である (=有意である)」ことを表す。

p 値はデータ数が多いサンプルにおいては, 試行回数が増え, p 値が低い事象が観察されうる。その際, 一般的には p 値を補正した (q 値) を用いて議論を行う。本実験でも, 解析には p 値を補正した値 (q 値) を用いた。補正は Bonferroni 法による多重検定を行った⁶⁸。

今回は pre-miRNA mutants ライブラリーに含まれる 4096 配列それぞれにおいて, カイ二乗検定を行い, 比較したい 2 サンプル間に有意差があるかを ($q < 0.05$) 基準で検定した。

Pearson's chi-squared test

A The number of reads of the sequence α (ex. GGG GGG) in the result of Compound-	B The number of reads of the sequence α in the result of Compound+	A+B The number of reads of the sequence α in the two measurement results
C The number of reads other than the sequence α in the result of Compound-	D The number of reads other than the sequence α in the result of Compound+	C+D The number of reads other than the sequence α in the two measurement results
A+C Total reads of Compound-	A+D Total reads of Compound+	A+B+C+D Total reads of the two measurement results



Expected values for each cell

a $\frac{(A + C) \times (A + B)}{(A + B + C + D)}$	b $\frac{(B + D) \times (A + B)}{(A + B + C + D)}$
c $\frac{(A + C) \times (C + D)}{(A + B + C + D)}$	d $\frac{(B + D) \times (C + D)}{(A + B + C + D)}$

Chi-square values

$$\chi^2 = \frac{(A-a)^2}{a} + \frac{(B-b)^2}{b} + \frac{(C-c)^2}{c} + \frac{(D-d)^2}{d}$$

Figure 1-6. Pearson's chi-squared test.

各配列の読み取り数を低分子の有無で比較した結果をプロットし、以下に示す (**Figure 1-7**)。プロットの際には、比較する各サンプルの総読み取り数が異なることを考慮し、総読み取り数を 10 万となるように規格化し、各配列の読み取り数の割合を存在量とした。この存在量を基底 2 の対数に変換し、グラフにプロットを行った。

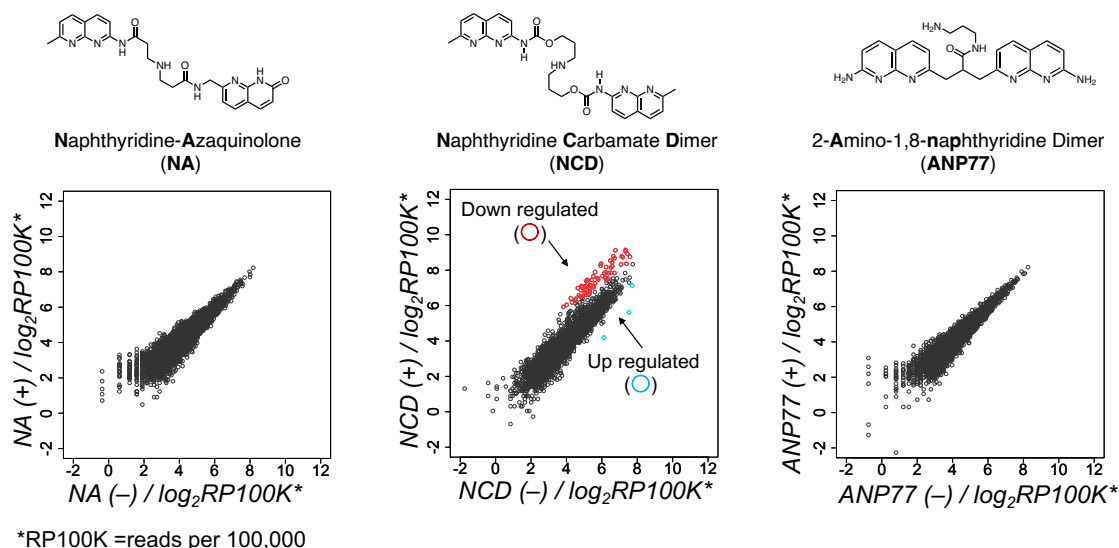


Figure 1-7. Reads number of sequences of detected corresponding to N = 6 in the RNA library. Circle : sequences of $q < 0.05$. Square: sequences of $q > 0.05$.

グラフの横軸は化合物を添加せずに Dicer 反応を行った際の未切断産物の読み取り数, 縦軸は化合物を添加して Dicer 反応を行った際の読み取り数を表している。結果より, NCD を用いた際に有意差のある配列が観測された。しかし一方で, NA と ANP77 を用いた際には有意差のある配列は見られなかった。次節より, 有意差の見られた NCD の結果をより詳細に解析し, 検討を行った結果を示す。

1-3 NCD を用いた配列解析

1-3-1 配列解析結果

本節では 3 つの個別の実験 (以下実験 #1, #2, および #3 とする) を実施し, NCD の追加がライブラリーの各 pre-miRNA 変異体の Dicer 切断にどのように影響するかを確認した。シーケンシングの結果をまとめ, **Table1-2** に示す。近畿大学理工学部・大学院総合理工学研究科飯田慶講師より頂いた perl スクリプト (補足資料) を使用して, 5 側, 3 側, および内部の固定領域に完全に一致する配列を特定した。実験 #1 では, 一致した読み取り数は, NCD (-) サンプルでは 584,738, NCD (+) サンプルでは 484,051 であり, どちらも合計読み取りの約 50%であった。また, 3 回の実験の平均は, NCD (-) サンプルの場合 510,942, NCD (+) サンプルの場合 442,566 であった。一致したリードをすべての 4,096 変異体についてカウントし, NCD (-) サンプルと NCD (+) サンプルで得られた各変異体の頻度の差の統計分析を, ピアソンのカイ二乗検定と Bonferroni 補正を使用して行った。

Table 1-2 Summary statistics for the high-throughput sequencing data.

	Experiment #1		Experiment #2		Experiment #3		Average	
	-	+	-	+	-	+	-	+
NCD								
Matched reads	584738 (52.1%)	484051 (45.6%)	485458 (54.5%)	387013 (49.5%)	462631 (53.0%)	456634 (54.4%)	510942	442566
Non-matched reads	537127 (47.9%)	576637 (54.4%)	405952 (45.5%)	395258 (50.5%)	410202 (47.0%)	383587 (45.6%)	451094	451827
Total reads	1121865	1060688	891410	782271	872833	840221	962036	894393

まずはじめに 4,096 種の変異体すべてについて, マッチしたリードをカウントし, 得られたカウントを log₂RP100K (総読み取り数 10 万個あたりの読み取り数) 値に変換した。各実験の NCD (-) と (+) サンプル間の 4,096 変異体の log-fold-change (logFC) 値の頻度分布を比較するためにヒストグラムを作成した (**Figure**

1-8)。実験#1 と#3 で得られたヒストグラムは, logFC の頻度分布が類似しており, 大多数の変異体が $-1 < \logFC < +1$ の範囲内に存在していた (1-8A, 1-8C, 1-8D)。

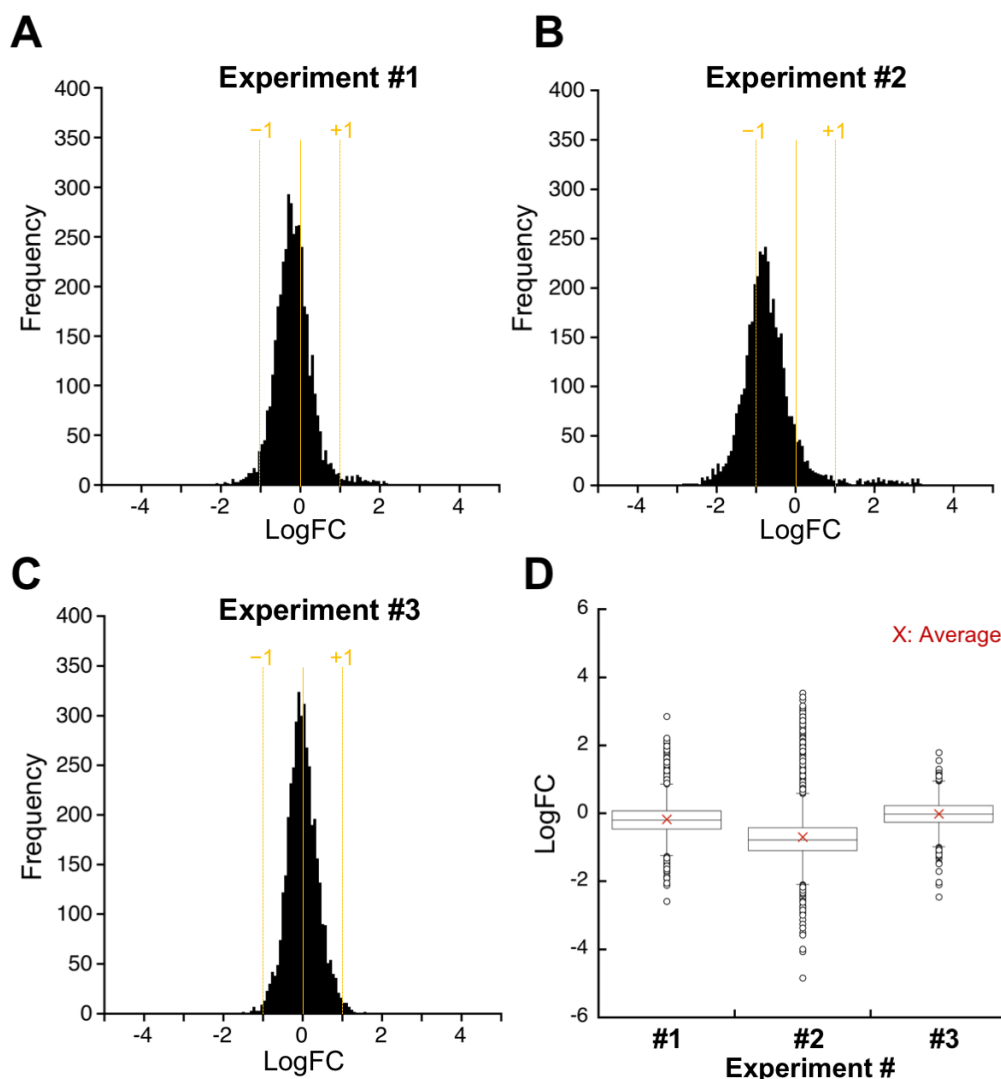


Figure 1-8. Frequency distribution of log-fold-change (logFC) values for 4,096 mutants. The logFC values were calculated by subtracting the log2RP100K values in the NCD (-) sample from the log2RP100K values in the NCD (+) sample in (A) Experiment #1, (B) Experiment #2, and (C) Experiment #3. (D) Box-plot representation of the frequency distribution of logFC values.

一方, 実験#2 の logFC 値の頻度は, 他の実験のそれと比較して, ピークが負に偏った幅広い分布となっていた (1-8B)。また, ほぼ全ての変異体において, NCD 添加時のサンプルの読み取り数が減少しており, logFC 値は負の値となった (1-8B および 1-8D)。3 つの実験間で logFC の頻度分布が異なるため, 同じ統計処理で解析すると誤判定につながってしまう。そこで, logFC 分布のうち, 中央の 50%

のデータを用いて、 $\log_2\text{RP100K}$ 値の線形回帰分析を行い、得られた式によって NCD(-)及び (+)のサンプルの実験データの正規化を行った (Figure 1-9)。

正規化後の NCD (-) 及び (+) のサンプルの $\log_2\text{RP100K}$ 値をプロットし, Figure 1-10A-C に示す。正規化した $\log_2\text{RP100K}$ 値を読み取り数に戻し, NCD (-) と (+) サンプルで得られた各変異体の頻度の差について, 前節と同様にピアソンのカイ二乗検定と Bonferroni 法による多重項検定補正を用いて統計解析を行った。

実験#1, #2, #3 において, NCD (-) と (+) のサンプル間で有意差 ($q < 0.05$) を示した変異体を, それぞれ 70, 110, 14 個同定した。これらの変異体のうち, 67, 105, 11 の変異体 (それぞれ実験#1, #2, #3 から得られた) は, NCD (+) サンプルでより多くの読み取り数を得た。これは, これらの変異体の Dicer による切断が NCD 非存在下よりも存在下で効率よく進行していることを示している。これらの変異体を本論文では阻害配列と呼ぶ (1-10A-C, 赤丸)。一方, 実験#1, #2, #3 からそれぞれ得られた 3, 5, 3 の変異体は, NCD (-) サンプルでより多くの読み取り数が得られ, これらの変異体の Dicer を介した切断が, NCD 非存在下よりも存在下でより効率的に進行していることが示している。これらの変異体を促進配列と呼ぶことにする (1-10A-C, 青丸)。3 回の実験のそれぞれで得られた阻害配列と促進配列の数を表 2 に示す。3 回の実験すべてで同定された阻害配列は 11 個, 促進配列は 2 個であった (Table 1-3)。

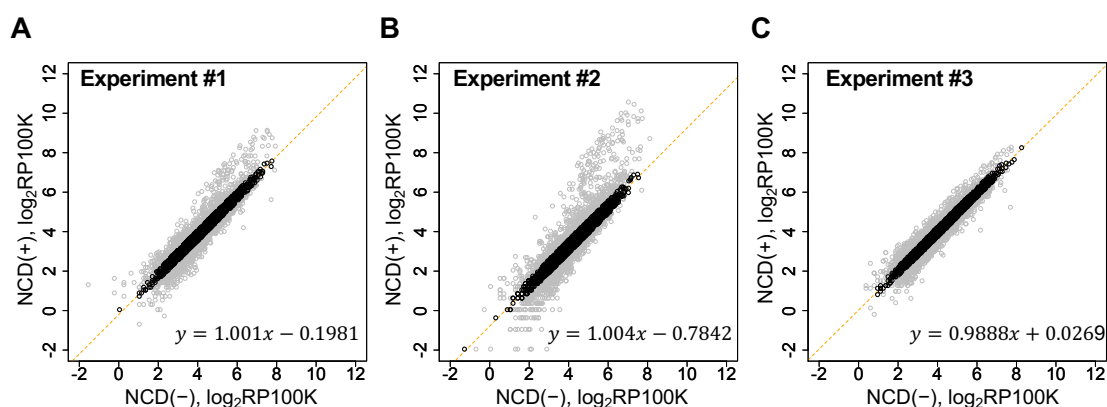


Figure 1-9. Scatter plots depicting \log_2 -transformed normalized read counts in NCD (+) samples (y-axes) compared to NCD (-) samples (x-axes) for (A) Experiment #1, (B) Experiment #2, and (C) Experiment #3. Black dots represent the data for the middle 50% mutants of the $\log\text{FC}$ distribution. Light gray dots represent the data for the other mutants. A linear regression analysis was done on the $\log_2\text{RP100K}$ values between NCD (-) sample and NCD (+) sample using the middle 50% mutants of the $\log\text{FC}$ distribution for each of three experiments. The regression lines (orange dashed lines) and the equations obtained are shown on the plots.

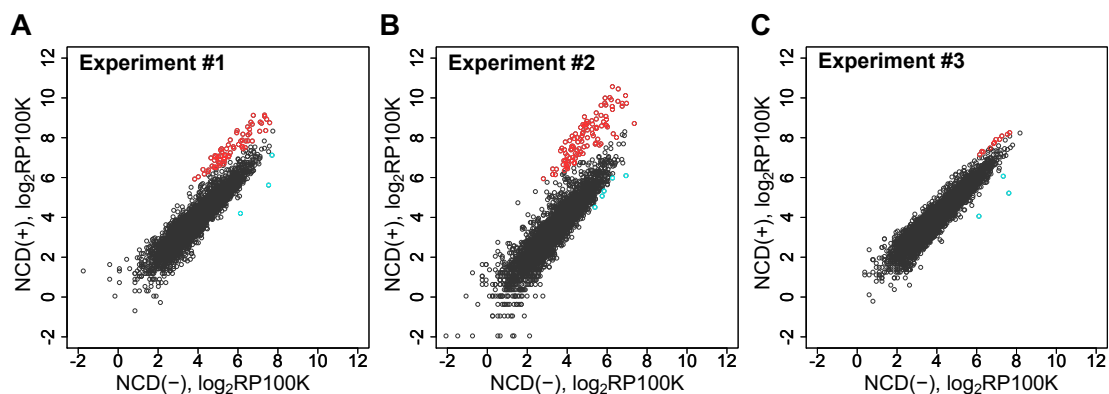


Figure 1-10. Scatter plots representing the normalized log₂-transformed read counts for 4,096 pre-miRNA mutants obtained from the reaction in the absence (x-axis) and presence (y-axis) of NCD. Red dots represent the inhibited mutants, and light blue dots represent the promoted mutants.

Table 1-3. Summary of the number of mutants that showed significant difference in the frequency obtained for NCD (-) sample and NCD (+) sample.

Experiment	Inhibited mutants	Promoted mutants	Total
#1	67	3	70
#2	105	5	110
#3	11	3	14
Identified in #1, #2, and #3	11	2	13

この 11 配列及び 2 配列の傾向をまとめ、ランダム化部位毎の塩基の存在割合をシーケンシングログで示す (**Figure 1-11**)

11 種の阻害配列は、ランダム化領域においてグアニン塩基への明確な選好性を示した (**1-11A**)。ランダム化領域におけるグアニン塩基の含有率は全体で 70%以上であり、他の塩基の含有率よりもかなり高い。N1 (11 変異体中 10 変異体, 91%), N2 (11 変異体中 9 変異体, 82%), N4 (100%) の位置で、有意に高いグアニン頻度が観察された。また、ウラシルは N6 位 (11 変異体中 7 変異体, 64%) に好んで存在した。一方で、ランダム化領域におけるアデニン塩基とシトシン塩基の含有率は、それぞれ 5%と 3%と低いことがわかった。2 種の促進配列はランダム化領域

において同様の配列 (5'-CGG-3'/5'-AGU-3' と 5'-CGG-3'/5'-AGC-3') を共有しており, その違いはN6位の核酸塩基 (UまたはC) だけであることがわかった (1-11B)。

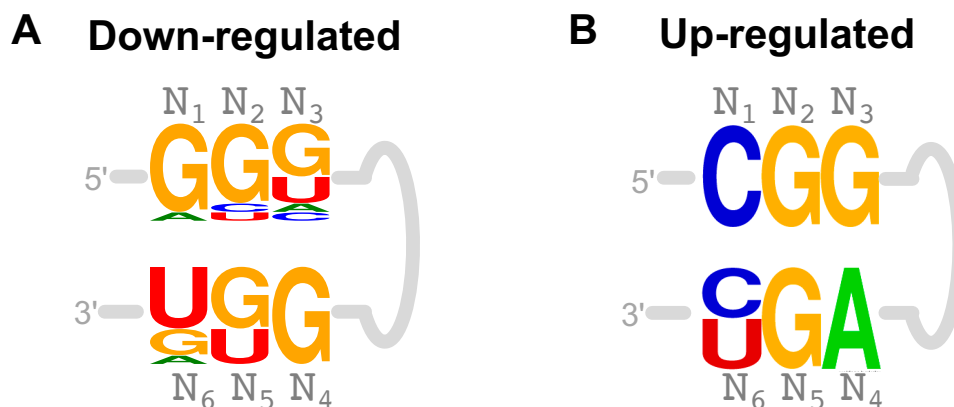


Figure 1-11. Sequence logo of the randomized region for (A) 11 inhibited mutants and (B) two promoted mutants identified in all the three experiments.

続いて, 観察された変異体の頻度の変化が, 本当に Dicer 反応に対する NCD 添加の効果によるものであり, 逆転写やPCRなどの他のステップに対するNCDの効果によるものではないことを確認するために, Dicer 切断にかけずに NCD で処理した pre-miRNA mutants ライブラリーについて配列解析を行った (Table 1-4, Figure 1-12)。実験#3において, NCD (-) と (+) のサンプル間で有意差 ($q < 0.05$) を示した1つの促進配列を確認したが, この変異体は Dicer 切断を行ったサンプルにおいて上記で確認したいずれの変異体とも同一でなかった。

Table 1-4. Summary statistics for the high-throughput sequencing data obtained from three experiments without dicer cleavage reaction.

	Experiment #1		Experiment #2		Experiment #3		Average	
	-	+	-	+	-	+	-	+
NCD								
Matched reads	598838 (57.6%)	557527 (55.5%)	431330 (52.4%)	469051 (52.6%)	472932 (56.6%)	490388 (58.1%)	501033	505655
Non-matched reads	441712 (42.4%)	446665 (44.5%)	391989 (47.6%)	421909 (47.4%)	363200 (43.4%)	353247 (41.9%)	398967	407274
Total reads	1040550	1004192	823319	890960	836132	843635	900000	912929

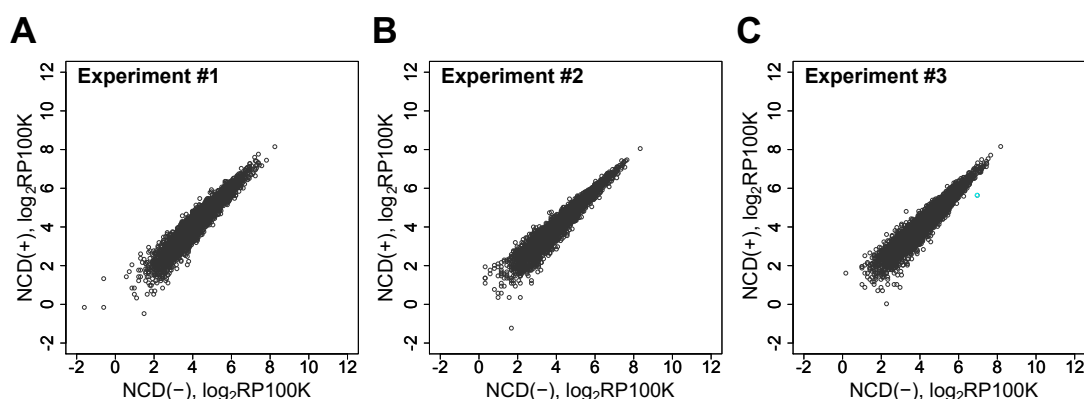


Figure 1-12. The libraries were prepared without subjecting them to Dicer cleavage. Scatter plots depicting \log_2 -transformed normalized read counts in NCD (+) samples (y-axes) compared to NCD (-) samples (x-axes). The light blue dot in the plot for Experiment #3 represents the mutant that yielded more reads in NCD (-) sample with $q < 0.05$.

1-4 NGS で得られた配列の実際の効果の確認

第 1-3 節において, ionPGM を用いた解析により, NCD が特定の配列の pre-miRNA の Dicer 切断反応効率を変化させている可能性が示唆された。本節ではその効果を実際に確かめるべく, 効果の見られた配列を実際に用い, Dicer 反応及び SPR による結合評価を行った。

1-4-1 Dicer 反応による検証

まずはじめに, Dicer 切断によるシーケンスデータの検証を行った。阻害配列と促進配列 (Table 1-5) をそれぞれ個別に準備し, NCD の存在下または非存在下で Dicer による切断反応を行った。また, 3 回の実験すべてで高い q 値を示した (NCD 添加による有意差が見られなかった) 7 つの変異体を同様に準備し, Dicer 切断反応のコントロールとして使用した。

Table 1-5. List of the inhibited mutants and the promoted mutants.

Sequence	Experiment #1			Experiment #2			Experiment #3		
	\log_2 RP100K		LogFC	\log_2 RP100K		LogFC	\log_2 RP100K		LogFC
5'-N ₁ N ₂ N ₃ -3'/5'-N ₄ N ₅ N ₆ -3'	NCD(-)	NCD(+)		NCD(-)	NCD(+)		NCD(-)	NCD(+)	
Inhibited mutants									
GUG GGU	6.71	8.68	1.97	5.93	9.64	3.71	6.87	7.73	0.86
GGU GUG	7.11	8.78	1.67	6.21	9.91	3.70	7.17	7.94	0.77
GGU GUA	6.64	8.78	2.14	6.13	10.02	3.88	6.91	7.78	0.88
GGU GGU	7.33	9.14	1.81	6.56	10.45	3.89	7.56	8.19	0.63
GGG GUU	7.34	9.07	1.73	6.92	10.12	3.20	7.68	8.28	0.60
GGG GUG	6.51	8.14	1.64	5.94	9.07	3.13	6.69	7.50	0.81

GGG GGU	7.18	8.73	1.55	6.54	9.83	3.29	7.28	8.11	0.82
GGG GGG	6.20	7.55	1.35	5.27	8.68	3.42	6.35	7.31	0.96
GGC GGU	6.77	9.13	2.36	6.26	10.57	4.30	6.99	7.93	0.94
GCA GGU	5.93	8.25	2.31	5.46	9.25	3.80	6.17	7.16	1.00
AGG GGU	5.96	8.36	2.40	5.70	9.62	3.93	6.28	7.33	1.06
Promoted mutants									
CGG AGU	7.53	5.62	-1.91	6.95	6.10	-0.85	7.63	5.24	-2.40
CGG AGC	6.12	4.20	-1.92	5.39	4.51	-0.88	6.13	4.08	-2.06

Dicer 反応後、反応生成物を変性ゲルで解析し、SYBR Gold でゲルを染色することで RNA バンドを可視化した (Figure 1-13A-C)。可視化したバンドはデンシトメトリー分析によって定量し、Dicer 反応への NCD 添加が切断生成物と未切断基質の相対量に影響を与えるかどうかを調べた。(1-13D-F)。分析の結果、すべての阻害配列において、NCD の存在下で未切断基質のバンド強度と切断生成物に対する割合が有意に増加した (1-13A, 1-13D)。例えば、5'-GGU-3'/5'-GUG-3'変異体では、NCD 非存在下で未切断基質の割合が 7.2%±3.2%であったものが、NCD 存在下では 88.4%±3.0%と 12.3 倍に増加している。また 5'-GUG-3'/5'-GGU-3'変異体でも同様に 11.8 倍の増加がみられ、未切断基質の割合は NCD 非存在下で 6.6%±2.1%、NCD 存在下で 78.6%±3.0%であった。阻害配列のうち、5'-GGG-3'/5'-GUU-3'、5'-GGG-3'/5'-GUG-3'、5'-GGG-3'/5'-GGG-3'変異体では未切断基質の割合の変化は比較的小さく、それぞれ 1.5 倍、1.4 倍、1.5 倍の増加であった。これは、これらの変異体においては NCD 非存在下においても基質の Dicer による切断が緩やかであったためであると考えられる。以上のように、阻害配列について得られたデータは、これらの変異体に対し、NCD が Dicer による切断に対し阻害効果を有していることを示している。

一方、阻害配列とは対照的に、促進配列の Dicer による切断には NCD 添加の効果は見られなかった (1-13B, E)。5'-CGG-3'/5'-AGU-3'変異体および 5'-CGG-3'/5'-AGC-3'変異体の未切断基質の割合は NCD 存在下で大きく変化せず、それぞれ 1.1 倍および 1.0 倍の変化であることが確認された。

また、コントロールとして使用した配列 (シーケンシング解析で有意差が見られなかった配列) に対する NCD の効果を確認したところ、それらはそれぞれ異なる影響が D 確認された (1-13C, 1-13F)。実験に用いた 7 つの変異体のうち 4 つの変異体では、NCD 添加による未切断基質の割合の有意な変化は見られなかった。これは、NCD がこれらの変異体の Dicer による切断にほとんど影響を及ぼさないことを示唆するものであった。一方、NCD は他の 3 つの変異体の Dicer 切断に対して統計的な有意差をもって影響を及ぼした。デンシトメトリー分析の結果、5'-GGC-3'/5'-GUG-3'変異体および 5'-ACA-3'/5'-GCG-3'変異体は NCD 存

在下で未切断基質の割合がそれぞれ 5.1 倍と 1.9 倍に増加しており, 5'-CAC-3'/5'-GCG-3' 変異体は NCD の添加により, 未切断基質の割合が 0.4 倍減少した。促進配列とコントロールとして使用した配列におけるシーケンシング結果と PAGE の結果の間に観察された不一致については, 本論文 1-4-3 節で考察を示す。

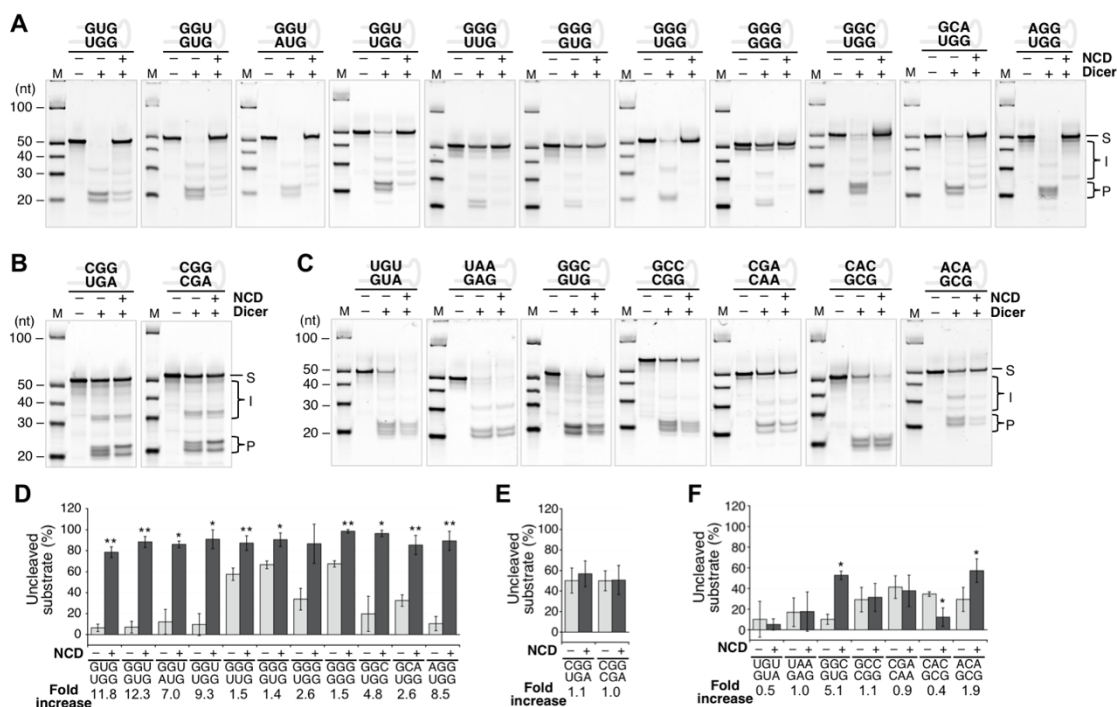


Figure 1-13. Denaturing-PAGE analysis of Dicer-mediated cleavage in the presence or absence of NCD of (A) the inhibited mutants, (B) the promoted mutants, and (C) the nonresponders. Densitometric analyses were performed on SYBRGold-stained gels and the fractions of uncleaved substrate (%) in the absence (light gray bars) and presence (dark gray bars) of NCD are shown as bar graphs in panels (D) the inhibited mutant, (E) the promoted mutants, and (F) the nonresponders. NCD concentration: 100 mM, S: Substrate, I: Intermediate, P: Product. Data are represented as mean \pm SD of three independent experiments. * $p < 0.05$ and ** $p < 0.01$ versus the sample without NCD addition (paired t-test).

また, 5'-GGU-3'/5'-GUG-3' 変異体および 5'-GUG-3'/5'-GGU-3' 変異体, 5'-AGG-3'/5'-GGU-3' 配列に対して, Dicer 切断の NCD の阻害効果は濃度依存的であることを確認した (Figure 1-14)。基質 RNA 2 μ M, Dicer 200 nM を使用し, NCD の濃度依存性の確認を行った結果, NCD の IC₅₀ 値はそれぞれ 4.5 \pm 0.1 μ M および 5.2 \pm 0.4 μ M, 3.8 \pm 0.4 μ M と推定された (1-14B)。

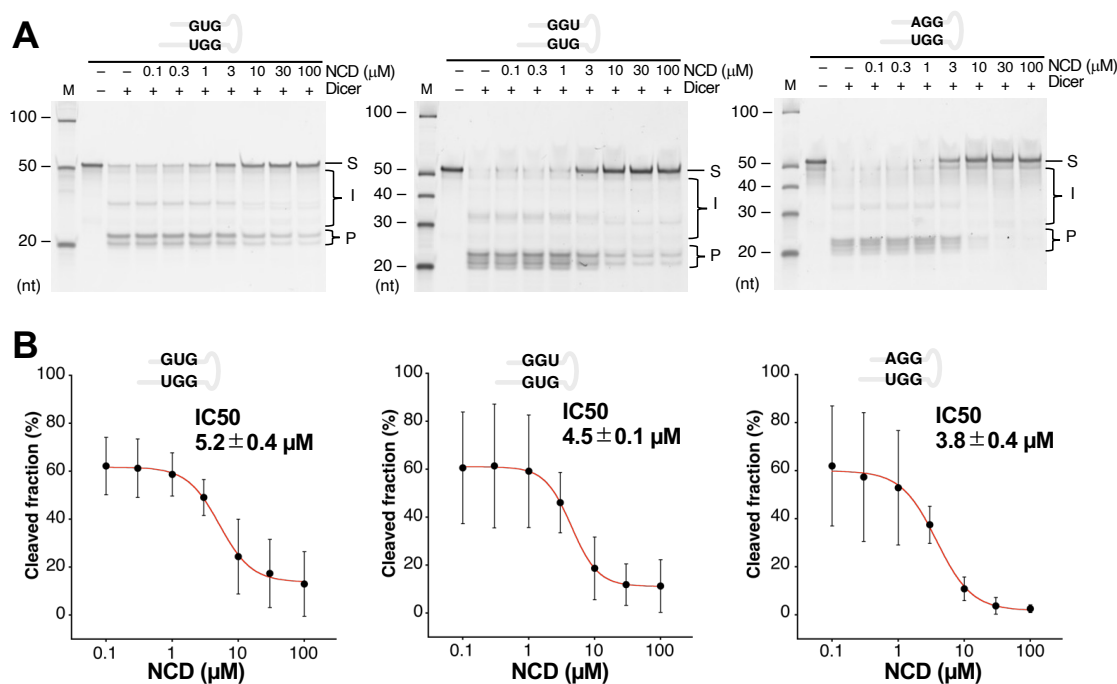


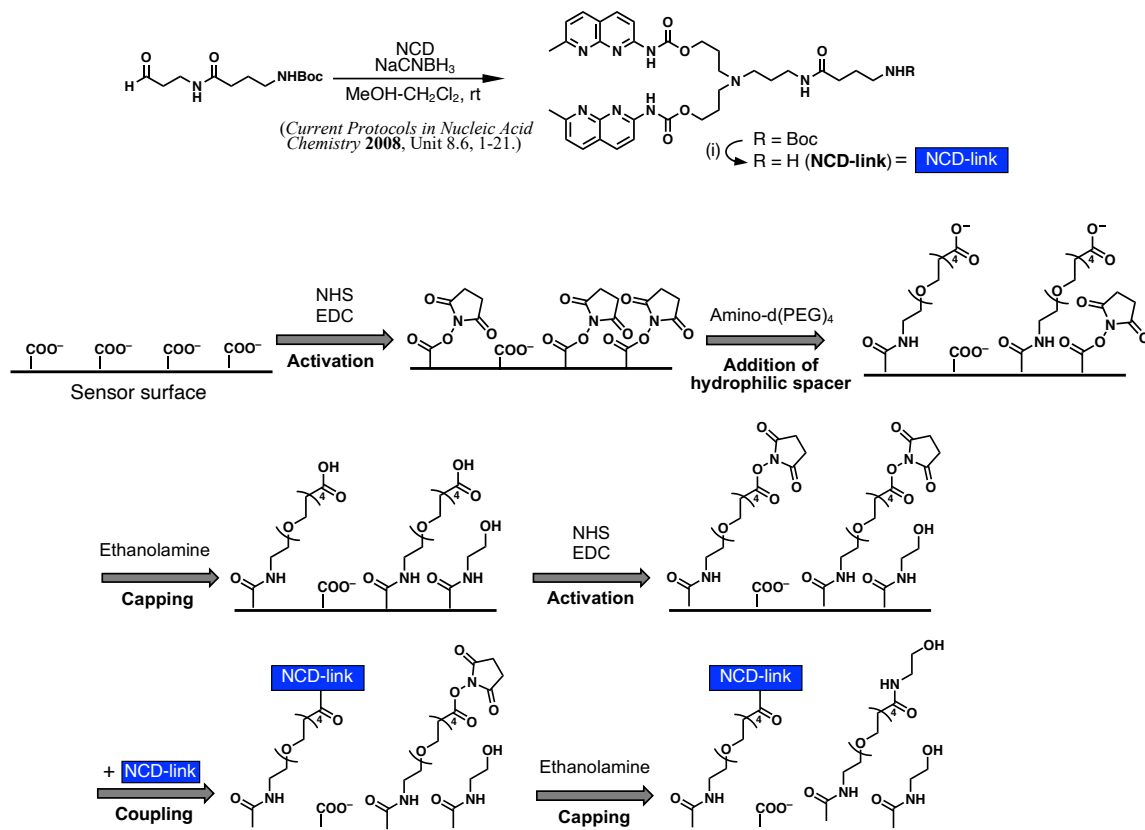
Figure 1-14. Concentration-dependent effect of NCD on the inhibited pre-miRNA mutants. (A) Representative gel images showing the inhibitory effect of NCD at various concentrations. S: Substrate, I: Intermediate, P: Product. (B) Plots showing the fraction of cleaved products (intermediates + products) against the log concentration of NCD. Data are represented as mean \pm SD of three independent experiments.

1-4-2 同定された RNA と NCD の結合評価

次に、NGS で同定されたこれらの変異体に対する NCD の結合評価を表面プラズモン共鳴 (SPR) アッセイによって行った。まず初めに NCD のセンサーチップ表面への固定化を行った (Scheme 1-1)。既報に従い⁶⁹, NCD を 4-((tert-butoxy)carbonylamino)-N-(3-oxopropyl) butanamide で還元アミノ化することにより、アミノアルキルリンカーを NCD に付加させた。Boc 基の脱保護後、得られた化合物 NCD-link を固定化に使用した。

続いて SPR センサーチップへの固定化を行った。まずチップ表面のカルボキシデキストランのカルボキシ基を NHS (N-Hydroxysuccinimide) と EDC (1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide) により活性化して活性エステルにし、アミンカップリングを経て Amino-dPEG4-acid を導入した。次に、未反応の活性エステルをエタノールアミンによりキャッピングした後、水酸化ナトリウム水溶液により、残留エタノールアミンを洗浄した。続いて、PEG のカルボキシ基を NHS と EDC により活性化して活性エステルにし、アミンカップリングを経て NCD-link を導入した。最後に、未反応の活性エステルをエタノールアミンによ

りキャッピングし、水酸化ナトリウム水溶液により残留エタノールアミンを洗浄した。以上の手順で、SPR センサーチップへの NCD の固定化を行った



Scheme 1-1. Immobilization of NCD onto a SPR sensor surface. An aminoalkyl linker was attached to NCD by reductive amination with 4-((*tert*-butoxy)carbonylamino)-*N*-(3-oxopropyl)butanamide, followed by deprotection of the Boc group. (i) 4 M HCl/Ethyl acetate, CHCl_3 . The carboxy groups of the dextran surface of a sensor chip CM5 were activated with NHS/EDC and coupled with Amino-dPEG4. After capping the residual activated ester with ethanolamine, the carboxy groups of Amino-dPEG4 acid were activated with NHS/EDC and coupled with NCD-link to produce the NCD-immobilized sensor surface.

続いて、作成したセンサーチップを用いて $0.01 \sim 1 \mu\text{M}$ の 5 つの濃度でセンサー表面への変異体の結合および解離の速度定数を記録した (**Figure1-15**)。11 種の阻害配列のうち、8 つの変異体は NCD を固定化したセンサー表面に対する SPR シグナルが有意に増加した。一方、3 つの変異体 (5'-GGG-3'/5'-GUU-3', 5'-GGC-3'/5'-GGU-3', 5'-GCA-3'/5'-GGU-3') は他の 8 つと比較すると弱い結合がみられた (**1-15A**)。結合定数を 1:1 と仮定した速度定数から計算した見かけの解離定数

(K_{Dapp}) 値より、これらの阻害配列はナノモル領域 (15.5-110 nM) の NCD への親和性を持つことが明らかになった (Table 1-6)。2つの促進配列は NCD 固定化センサー表面に対して比較的弱い応答を示し (1-15B)、阻害配列よりも高い K_{Dapp} 値を示した (Table 1-6)。コントロールとして使用した配列 (シーケンシング解析で有意差が見られなかった配列) について得られた SPR 応答は、5'-GGC-3'/5'-GUG-3'変異体を除いて、阻害配列で得られた応答より小さかった (1-15C)。またこれらの配列で計算された K_{Dapp} 値は 65.5 nM から 23.1 μ M までと様々であった (Table 1-6)。結果をまとめると、阻害配列で得られた SPR 応答は、促進配列やコントロール配列で得られた SPR 応答より高かった。しかし、NCD の促進配列やコントロール配列への結合評価で得られた K_{Dapp} 値は、阻害配列 5'-GCA-3'/5'-GGU-3'で得られた最高値である 110 nM の 4 倍以内の値であった。しかし、2つのコントロール配列 (5'-GCC-3'/5'-GC-3' と 5'-CAC-3'/5'-GCG-3') が非常に高い K_{Dapp} 値を示していることや、1-4-1 節における低分子の Dicer 切断阻害能と本節における低分子の結合能の間に正の相関が見られないことから、Dicer 切断の制御は、pre-miRNA への結合の強さのみに比例しているのではなく、複数の要因がによって決定される分かった。

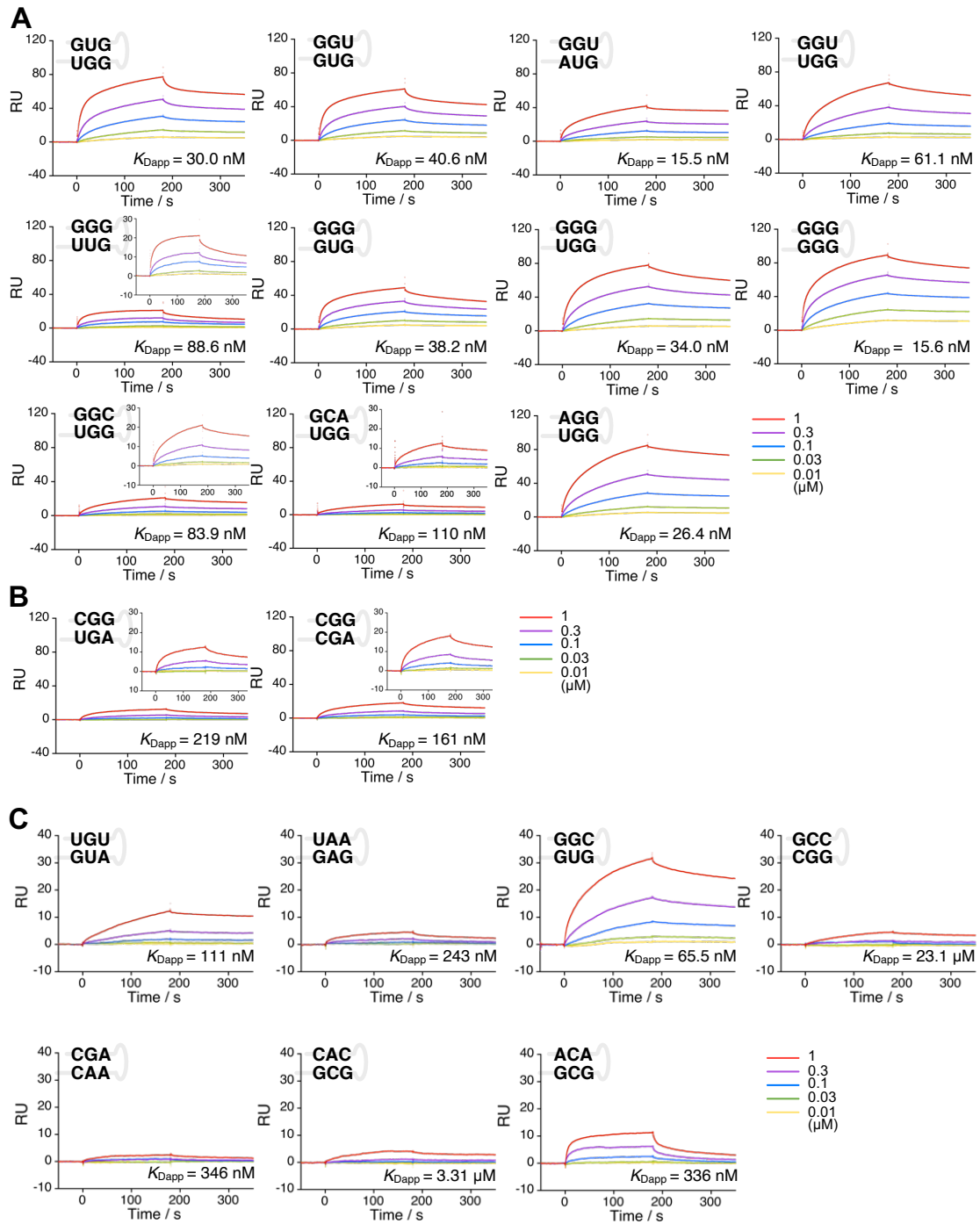


Figure 1-15. SPR analysis of the binding of NCD to (A) the inhibited mutants, (B) the promoted mutants and (C) Non-responders. RNA was added stepwise at 0.01 (yellow), 0.03 (green), 0.1 (blue), 0.3 (purple), and 1.0 μ M (red) to the NCD-immobilized sensor surface. Insets are higher magnification views.

Table 1-6. The kinetic parameters of the interaction between NCD and pre-miRNA mutants were determined by SPR analysis. Eleven inhibited mutants, 2 promoted mutants, and 7 non-responders were evaluated. k_a : association rate constant. k_d : dissociation rate constant. τ : residence time. K_{Dapp} : apparent dissociation constant.

Sequence 5'-N ₁ N ₂ N ₃ -3'/5'-N ₄ N ₅ N ₆ -3'	k_a (1 x 10 ⁴ / Ms)	k_d (1 x 10 ⁻³ / s)	τ (s)	K_{Dapp} (nM)
Inhibited mutants				
GUG GGU	3.16	0.948	1050	30.0
GGU GUG	3.29	1.33	752	40.6
GGU GUA	1.62	0.251	3980	15.5
GGU GGU	1.70	1.04	962	61.1
GGG GUU	3.17	2.81	356	88.6
GGG GUG	3.97	1.52	658	38.2
GGG GGU	3.21	1.09	917	34.0
GGG GGG	4.72	0.736	1360	15.6
GGC GGU	1.38	1.16	862	83.9
GCA GGU	1.03	1.14	877	110
AGG GGU	1.92	0.507	1970	26.4
Promoted mutants				
CGG AGU	1.14	2.49	402	219
CGG AGC	1.09	1.76	568	161
Non-responders				
UGU AUG	0.565	0.626	1600	111
UAA GAG	1.11	2.70	370	243
GGC GUG	1.63	1.07	935	65.5
GCC GGC	0.00744	1.72	581	23100
CGA AAC	0.922	3.19	313	346
CAC GCG	0.0469	1.56	641	3320
ACA GCG	1.99	6.67	150	336

1-4-3 シークエンシング結果と PAGE 及び SPR の結果の不一致について

阻害配列の二次構造モデル (Figure1-16) から, NCD の結合に必要な RNA モチーフは以下のように予測する。11 の阻害配列のうち, 4 つの変異体はランダム化領域に GG ミスマッチを有しており, これは 1 つのコントロール配列 (5'-GGC-3' / 5'-GUG-3') のみに見られ, 他の変異体には見られなかった (Figure 1-17)。

GG ミスマッチの位置は Dicer 切断部位に対して阻害配列間で互いに異なっており、これらの変異体の Dicer を介した切断に対する NCD 結合の影響度合いが異なる可能性があると考ええる。

一方、促進配列の Dicer 切断は、PAGE で解析したところ、NCD の添加によって大きな影響を受けないことが示された。促進配列のシーケンス結果と PAGE 解析の不一致は、NCD (-) と (+) の両方のサンプルにおける不十分な読み取り数によるものではないかと考えられる (Table 1-7)。これらの促進配列は NCD を固定化した SPR センサー表面への結合を示したが、その親和性は阻害配列で得られた親和性より低かった。このことは、PAGE 解析において、これらの促進配列の切断を促進する NCD の明確な効果がなかった理由も説明できるかもしれない。また、実験に使用したコントロール配列のほとんどに、NCD はその Dicer を介した切断に影響を与えなかった。しかし、いくつかの変異体の切断は、NCD の添加によって影響を受け、これは配列決定の結果と矛盾するものであった。また、コントロール配列について得られた読み取り数は NCD の添加に関わらず全て少なく (Table 1-7)、このためシーケンス結果において、これらの変異体の頻度に統計的な有意性は見られなかった。促進配列とコントロール配列のシーケンス結果と Dicer 切断反応の PAGE 解析の間に見られた不一致は、q 値の閾値に加えて、変異体を選択するための読み取り数に関するカットオフ閾値の必要性を示唆するものであった。しかし、全体として、本手法は、低分子化合物の添加により Dicer による切断が阻害される pre-miRNA 変異体を、偽陽性例なく同定することに成功した。

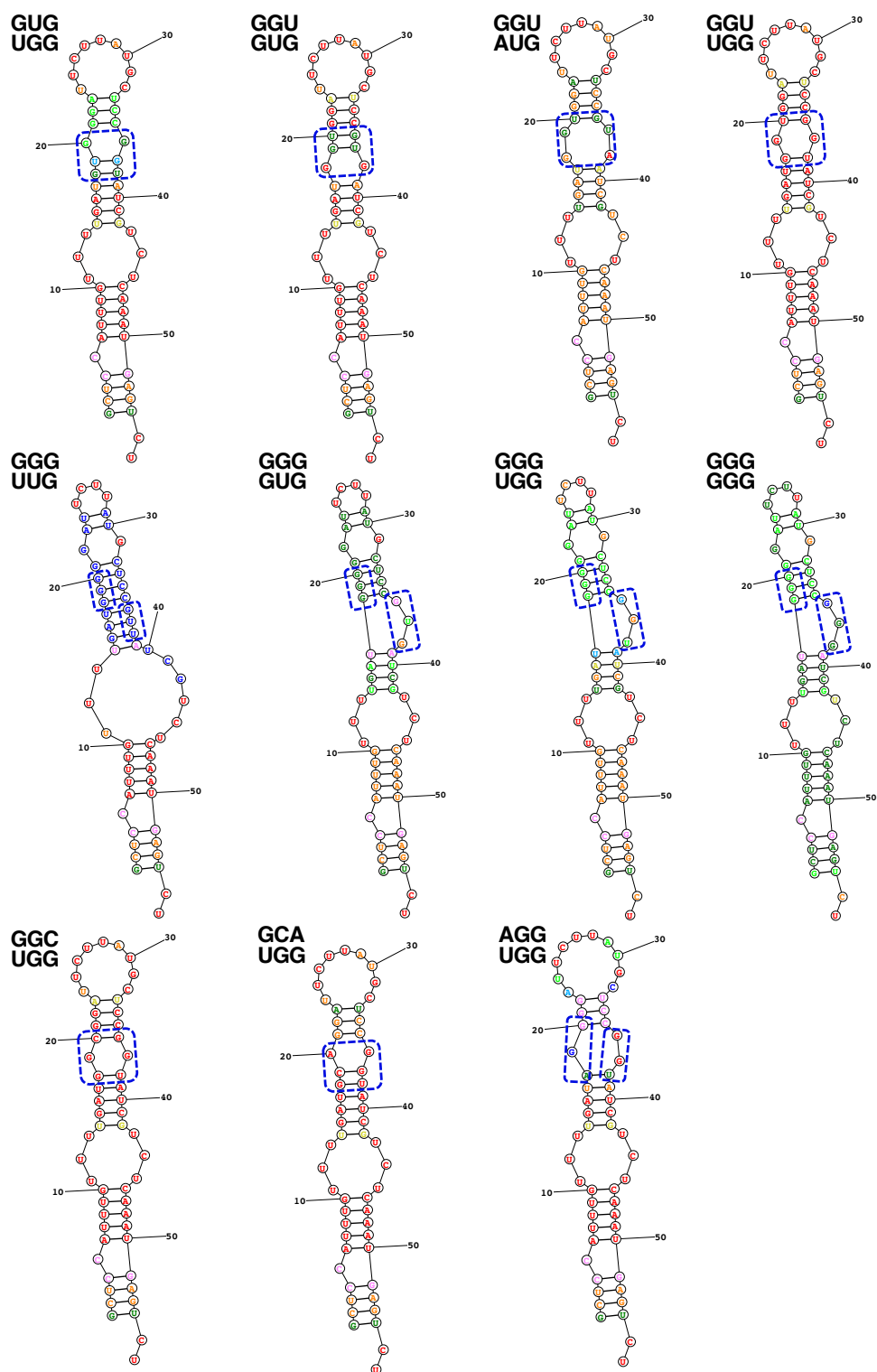


Figure 1-16. Predicted secondary structures of the inhibited mutants.

The structures were predicted by the program RNAstructure Dotted squares indicate the randomized region.

(<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>).

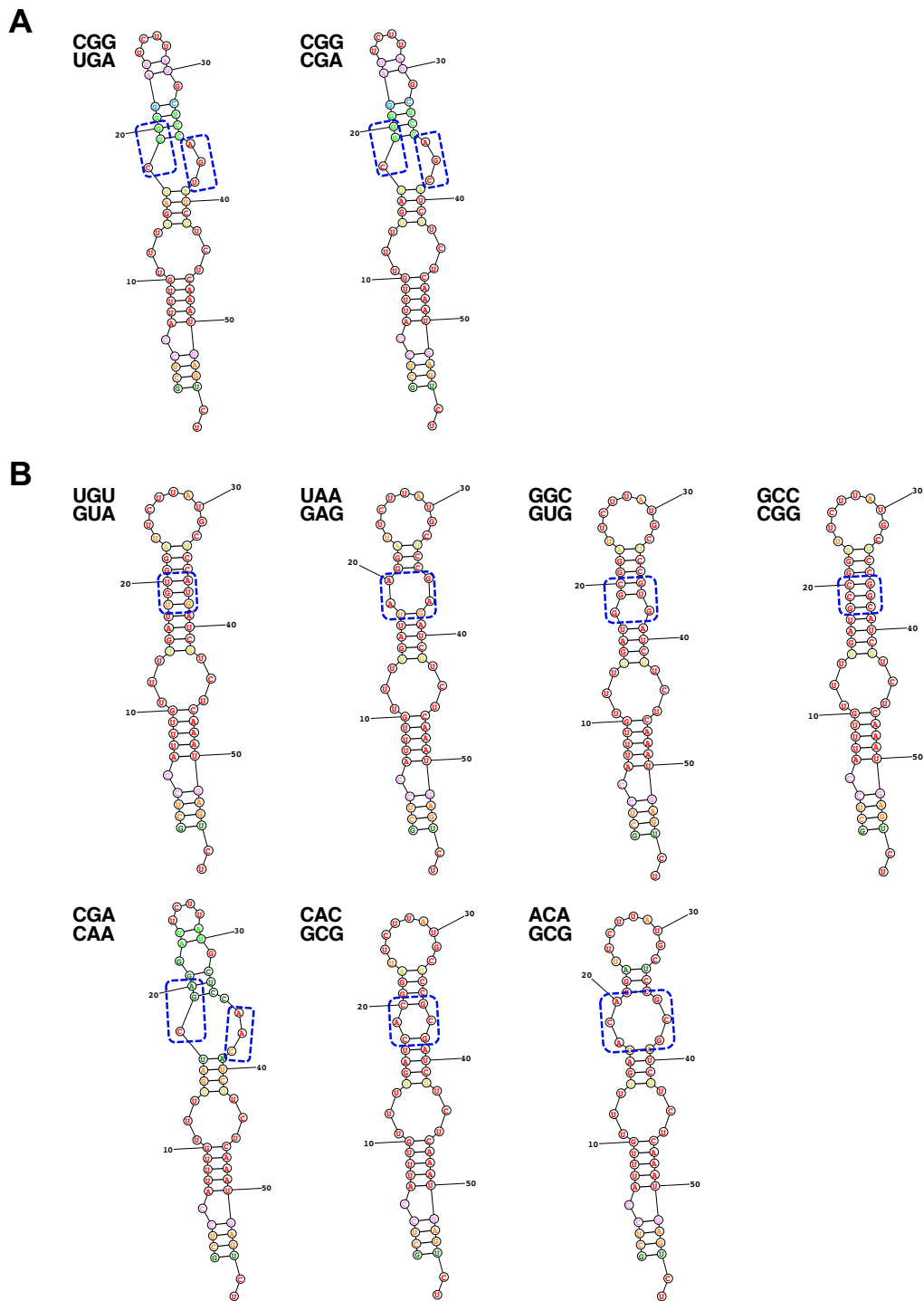


Figure 1-17. Predicted secondary structures of the promoted mutants and the non-responders. The structures were predicted for (A) the promoted mutants and (B) the non-responders by the program RNAstructure Dotted squares indicate the randomized region. (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>).

Table 1-7. Comparison of RP100K values between NCD (-) samples and NCD (+) samples for the selected mutants. The black triangles indicate a negative difference.

Sequence 5'-N ₁ N ₂ N ₃ -3'/5'-N ₄ N ₅ N ₆ -3'	Experiment #1			Experiment #2			Experiment #3		
	NCD(-) RP100K	NCD(+) RP100K	Diference	NCD(-) RP100K	NCD(+) RP100K	Diference	NCD(-) RP100K	NCD(+) RP100K	Diference
Inhibited mutants									
GTG GGT	105.0	410.3	305.3	61.1	800.0	738.9	117.1	212.0	94.9
GGT GTG	138.5	441.1	302.5	73.9	961.7	887.8	143.8	244.8	101.0
GGT GTA	100.1	439.8	339.8	70.2	1036.7	966.4	120.2	220.5	100.4
GGT GGT	160.9	563.6	402.6	94.2	1399.4	1305.2	188.8	291.9	103.1
GGG GTT	162.0	536.9	374.9	121.3	1111.1	989.8	204.6	309.9	105.2
GGG GTG	91.1	283.0	192.0	61.3	538.5	477.2	103.4	181.3	77.9
GGG GGT	145.1	424.7	279.6	93.0	910.0	817.1	155.8	275.7	119.9
GGG GGG	73.5	186.8	113.3	38.5	411.1	372.6	81.5	158.3	76.8
GGC GGT	108.9	558.4	449.5	76.8	1516.2	1439.4	126.7	244.0	117.2
GCA GGT	61.2	304.3	243.1	43.9	609.5	565.7	71.8	143.4	71.7
AGG GGT	62.1	328.3	266.2	51.9	789.1	737.2	77.6	161.4	83.8
Promoted mutants									
CGG AGT	185.1	49.2 ▲	136.0	123.2	68.5 ▲	54.7	198.3	37.7 ▲	160.6
CGG AGC	69.7	18.4 ▲	51.3	41.9	22.7 ▲	19.2	70.1	16.9 ▲	53.2
Non-responders									
TGT ATG	45.1	47.9	2.8	30.2	39.3	9.1	57.4	39.3 ▲	18.2
TAA GAG	8.7	8.9	0.2	5.9	8.8	2.9	10.3	8.8 ▲	1.5
GGC GTG	97.1	109.5	12.4	60.4	101.8	41.5	97.6	101.8	4.2
GCC GGC	2.2	2.1 ▲	0.2	2.3	3.1	0.8	5.2	3.1 ▲	2.1
CGA AAC	7.0	7.0 ▲	0.0	4.1	7.8	3.7	8.2	7.8 ▲	0.4
CAC GCG	12.4	12.8	0.4	6.6	10.1	3.4	11.6	10.1 ▲	1.5
ACA GCG	6.1	7.2	1.1	2.3	3.6	1.3	4.3	3.6 ▲	0.7

1-5 内在性 miRNA への応用

1-5-1 内在性 pre-miRNA ヘアピンにおける同定された RNA モチーフの探索

1-1 節で述べたように miRNA は発がんを含む多くの細胞内プロセスに関与しているため、低分子化合物によって特定の miRNA の生合成を標的とすることは、miRNA 制御不全を伴う疾患の治療戦略として期待できる。

そこで、同定された NCD 結合 RNA モチーフが内在性 pre-miRNA ヘアピンに存在するかどうかを検討した。現在までに、34,675 の pre-miRNA ヘアピンの配列が miRBase データベース (<https://www.mirbase.org>) に登録されている。このデータベースに登録されている pre-miRNA ヘアピンの配列をダウンロードし、ViennaRNA パッケージで二次構造を予測した⁷⁰。11 の阻害配列で同定された RNA モチーフをステム領域で検索し、その RNA モチーフを含むことがわかった pre-miRNA ヘアピンをカウントした。その結果、11 個の RNA モチーフは、合計でデータベース中の pre-miRNA ヘアピンの 738 箇所、ヒト pre-miRNA ヘアピンでは 41 箇所にそれぞれ存在することが判明した (Figure 1-18)。11 の RNA モ

モチーフのうち、5'-GGU-3'/5'-GGU-3'モチーフが、ヒト pre-miRNAヘアピンで最も多く見いだされ、5'-GCA-3'/5'-GGU-3'と5'-GGC-3'/5'-GGU-3'がそれぞれ2番目、3番目に多く観測された。筆者らは阻害配列モチーフに見られる11のモチーフのうち1つ以上を含む pre-miR-216a, pre-miR-3689c, pre-miR-6073 の3つの pre-miRNA を選択し、これらの pre-miRNA の Dicer による切断に対する NCD の影響を調べた (Figure 1-19)。NCD を添加すると、pre-miR-3689c と pre-miR-6073 の切断産物の割合が減少した。一方、pre-miR-216a の反応では、NCD を添加しても、切断された生成物と切断されていない基質の割合は大きく変化しなかった。この NCD の効果の違いは、miRNA 前駆体における NCD 結合部位の位置と関係している可能性があり、過去の研究で観察された結果と一致している⁷¹。

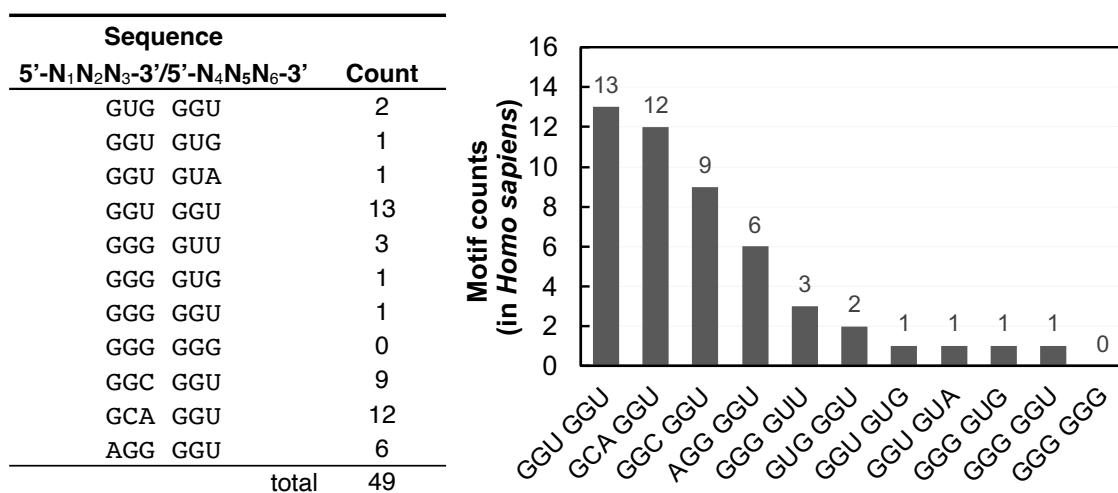


Figure 1-18. Motif search in the human pre-miRNA hairpins. The RNA motifs in the 11 inhibited mutants were searched in the sequences of human pre-miRNA hairpins in the miRbase. The table of counts for each motif that were found in the human pre-miRNA hairpins (left), and the bar-graph representation (right).

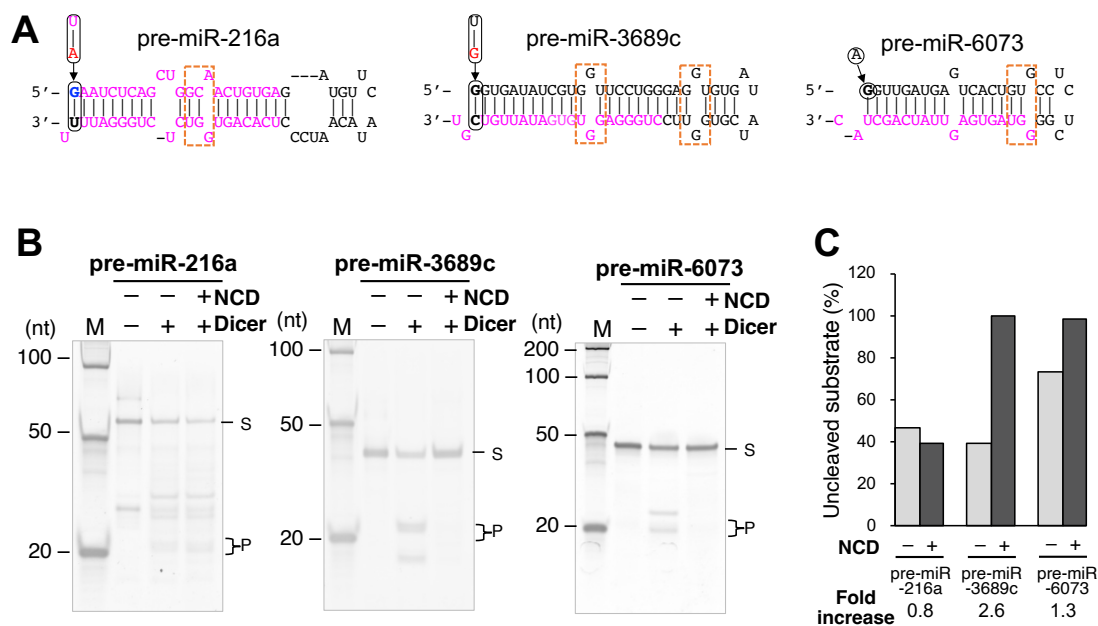


Figure 1-19. The effect of NCD on the Dicer-mediated cleavage of endogenous miRNA precursors. (A) The secondary structures of miRNA precursors used in the Dicer reaction. For efficient initiation of in vitro transcription by T7 RNA polymerase, 5' -end base of the RNAs were replaced with guanine and made the corresponding change near the 3' -end to maintain the predicted secondary structure (shown in circles). Dashed squares represent the one of 11 motifs found in the inhibited mutants. (B) Representative gel images showing the Dicer-mediated cleavage reaction of the selected miRNA precursors in the absence or presence of 100 μ M NCD. S: Substrate, P: Product. (C) Densitometric analysis of uncleaved substrate bands (%) in the absence (light gray bars) and presence (gray bars) of NCD.

1-6 FDA 承認薬とその関連誘導体への適用

これまでに、本手法によって低分子が結合する RNA 配列を網羅的に解析可能であることを証明した。今後の課題は本手法からより実用的なデータを取得すること、データを正しく活用することである。

本節では、pre-mRNA のスプライシングを抑制する効果があると知られている FDA 承認薬である risdiplam および同様の効果が期待されている類似体 branaplam, SMN-C5 へと本手法を適用し、その標的となりうる RNA モチーフの取得を試みた。

1-6-1 配列解析結果

前項の NCD を用いた解析と同様に, 上記 3 つの低分子の有無で pre-miRNA mutants ライブラリーの Dicer による切断反応を行なった。得られた反応産物から NGS 解析用の DNA ライブラリーを作製し, 得られた DNA ライブラリーは, illumina HiSeq 3000 プラットフォームで配列決定した。DNA ライブラリーの作製の際にプライマーとして使用したオリゴヌクレオチドを表に示す(**Table 1-8**)。DNA ライブラリーの調製は 1-2-3 節と同様に, RNA の逆転写反応と生成した cDNA の PCR 増幅によって行なった。

Table 1-8. The sequences of DNA oligos used as primers during DNA library preparation.

#	Oligos	Sequences
1	pre-miR136_F_illumina (Fw primer)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTCCATTTGTTTTGA
2	pre-miR136-R-illumina (Rev primer)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGACTCATTGAGACGA

各配列の読み取り数を低分子の有無で比較した結果をプロットし, 以下に示す(**Figure 1-20**)。図に示すように, risdiplam では 1 つの阻害配列, branaplam では 19 つの促進配列, SMN-C5 では 10 つの促進配列が同定された。

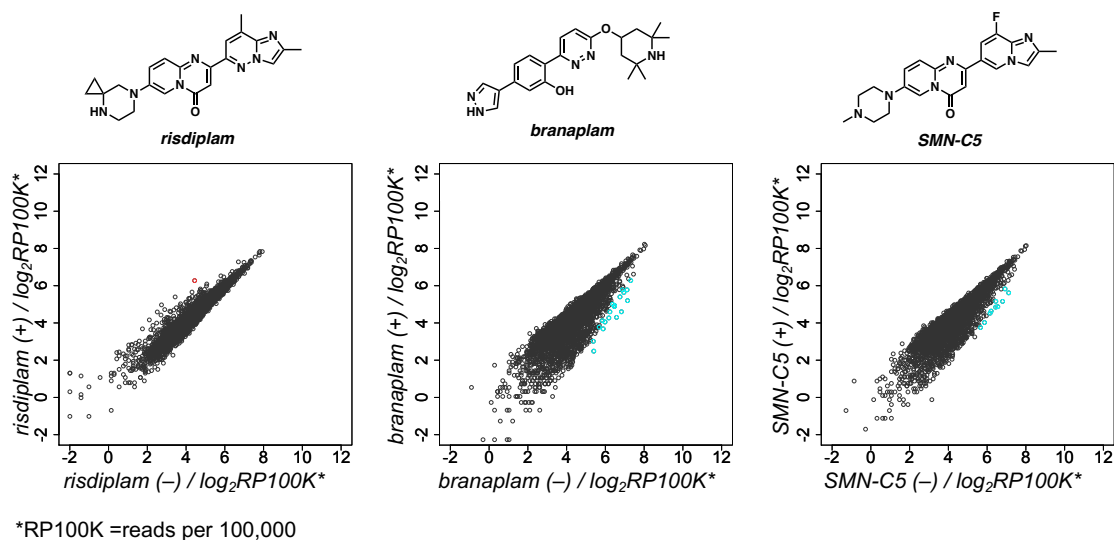


Figure 1-20. Scatter plots representing the normalized log₂-transformed read counts for 4,096 pre-miRNA mutants obtained from the reaction in the absence (x-axis) and presence (y-axis) of FDA-approved drugs and their derivatives. Red dots represent the inhibited mutants, and light blue dots represent the promoted mutants

1-6-2 促進配列の検証

1-4 節では, NCD を用いた際に促進配列が 2 変異体同定された一方で, 個別に合成した基質を用いた際には Dicer 切断反応における促進効果は確認されなかった。そこで, 本節では SMN-C5 を用いた際に同定された促進配列を用い, 実際の効果を変性ゲル解析とデンシトメトリー分析による定量によって確認した (Figure 1-21)。

結果として, SMN-C5 を用いた際にも変異体に対する低分子の Dicer 切断反応の明確な促進効果は確認されなかった。

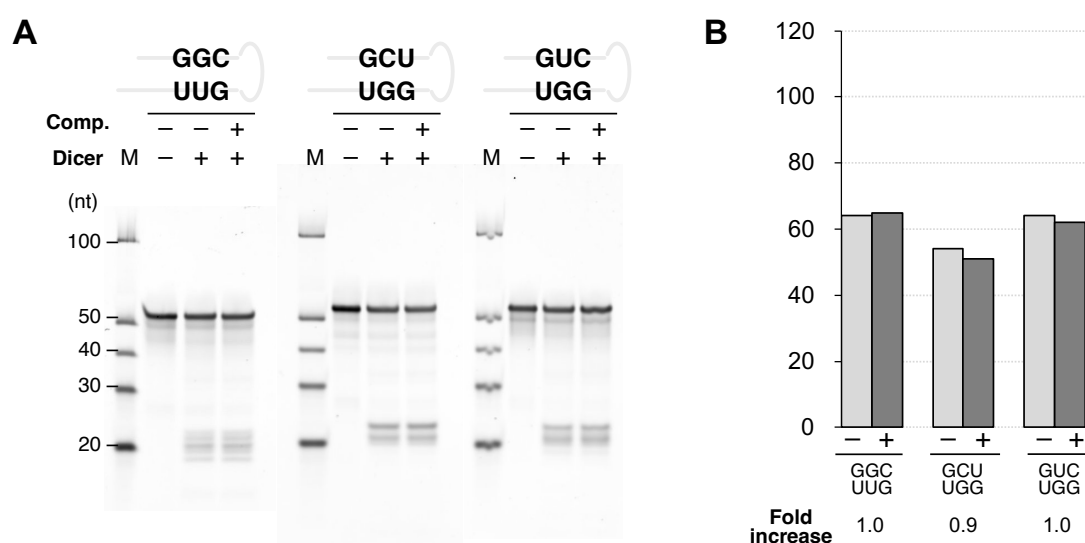


Figure 1-21 Denaturing-PAGE analysis of Dicer-mediated cleavage in the presence or absence of SMN-C5 of the promoted mutants (A). Densitometric analyses were performed on SYBRGold-stained gels and the fractions of uncleaved substrate (%) in the absence (light gray bars) and presence (dark gray bars) of SMN-C5 are shown as bar graphs (B).

SMN-C5 を用いた際にも促進効果が見られなかった原因を調べるために, 筆者は 2 つの調査を行った。

まずはじめに, Dicer 非存在化においても低分子が pre-miRNA に対して影響を及ぼしているのではないかと考え, Dicer を非添加の条件下における低分子化合物の影響を調査した。Dicer 非存在下におけるシーケンシングの結果を Figure 2-22 に示す。図に示すように, Dicer 非存在下では 4096 配列全てにおいて低分子の添加による読み取り数の変化が見られなかった。よって, SMN-C5 は少なくとも Dicer が存在する条件下でのみ pre-RNA へ影響を与えられとされる。

実際の系中における促進配列の挙動の観測は, 今後の課題である。

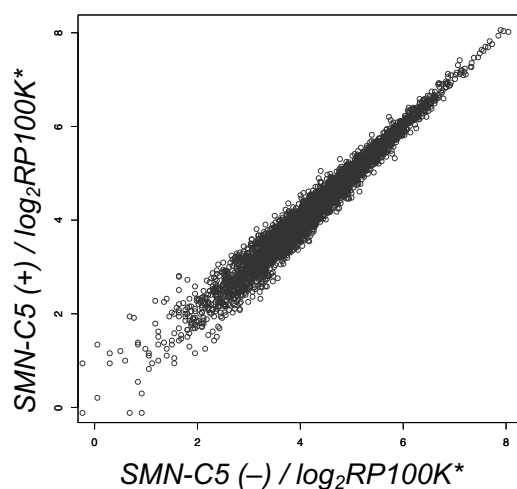


Figure 1-22. Scatter plots representing the normalized log₂-transformed read counts for 4,096 pre-miRNA mutants obtained from the absence of Dicer condition in the absence (x-axis) and presence (y-axis) of SMN-C5..

1-7 大規模データ取得を志向した検討

これまでに、低分子が結合する RNA 配列を網羅的に解析可能であることを実証できた。今後は、より多くの低分子・RNA 配列について本手法を適用して大量のデータを取得する、得られたビッグデータを情報科学的に解析することで、標的 RNA の結合に重要な化学構造を抽出し、新規 RNA 結合低分子を設計する際の指針の獲得を目指す。

本論文ではその大規模データ取得の第一歩として、まず逆転写及び PCR 増幅の条件検討を行った。また、実際に本手法を小規模なライブラリーに適用し、実際の解析が可能であるか、試行した。実験には 96well プレートを使用し、中谷研究室の in-house ライブラリーから 80 化合物 (補足資料) を使用して検討を行った。DNA ライブラリーは, illumina HiSeq 3000 プラットフォームで配列決定した。また解析にあたり、自作の python スクリプト(補足資料)を作成し、全化合物に対するシーケンシングデータを半自動的に一度に解析できるようにした。

実際の検討の結果, 80 化合物中 25 化合物が特定の RNA モチーフに対して Dicer 反応を阻害または促進する効果が観測された (**Figure 1-23**)。25 化合物中, 23 化合物は特定の RNA モチーフに対する阻害効果を有していることが確認され, 7 化合物は促進の効果を有している可能性が示唆された。また, 5 化合物は特定の RNA モチーフに対しては阻害, 別のモチーフには促進の効果を有していることが分かった。

続いて, スクリーニングで得られた低分子の標的となる RNA モチーフのランダム化部位毎の塩基の存在割合をシーケンシングログで示す (**Figure 1-24**)。

阻害配列は、どのランダム化部位においてもグアニン塩基への比較的有意な選好性を示した (1-24A)。一方で、ランダム化部位におけるシトシン塩基の含有率は、他の塩基と比較して極端に低いことがわかった。促進配列は阻害配列と同様に、どのランダム化部位においてもグアニン塩基への比較的有意な選好性を示した (1-24B)。一方で、促進配列においてはランダム化部位におけるアデニン塩基の含有率が、他の塩基と比較して極端に低く、これは阻害配列とは異なるものであった。

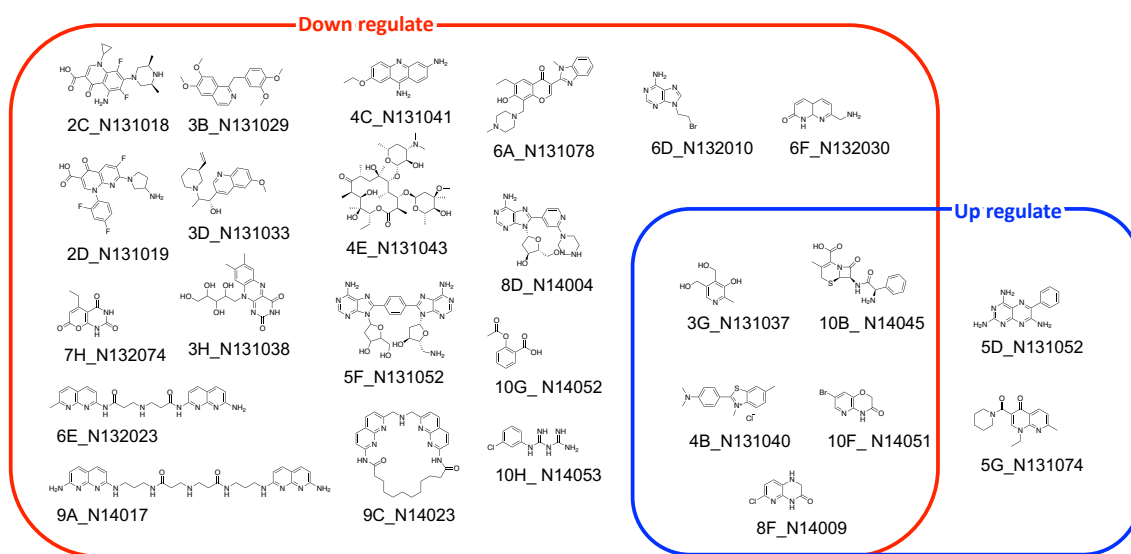


Figure 1-23. Classification of compounds with significant differences

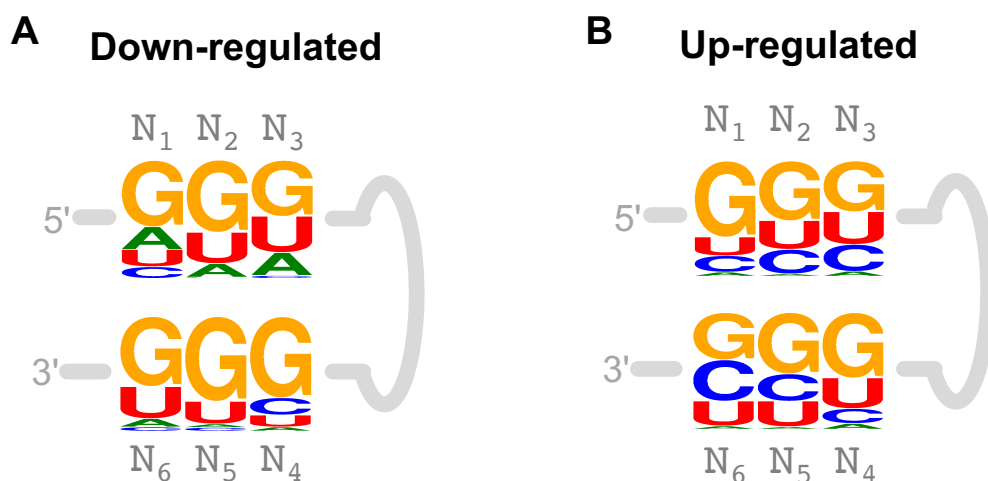


Figure 1-24. Sequence logo of the randomized region for (A) inhibited mutants and (B) promoted mutants identified by analysis of an in-house compound library.

以上より、筆者らが開発した手法は小規模な化合物ライブラリーに対しても適用可能であることが証明された。将来的には本手法を大規模なライブラリーへ適用し、RNA-低分子間の相関情報の網羅的な取得を行い、新規 RNA 結合低分子を設計する際の指針を獲得へと寄与できると期待している。

結論

本研究では、pre-miRNA mutants ライブラリーのハイスループットシーケンシング解析により、Dicer 切断における低分子の効果を評価し、低分子の結合部位となりうる pre-miRNA 変異体の配列モチーフを網羅的に同定した。また配列解析の結果、低分子化合物 NCD の添加によって Dicer による切断が促進または阻害される pre-miRNA 変異体を同定した。これらの同定された pre-miRNA 変異体は、Dicer 切断反応の PAGE とデンストメトリー分析、および SPR による NCD との結合評価によって評価を行った。その結果、シーケンシング解析で同定された 11 種の阻害配列に対して、NCD を添加することで実際に Dicer 切断が阻害されることが証明された。さらに、SPR 解析の結果、これらの阻害配列はナノモル濃度レベルの親和力で NCD と結合することが明らかになった。これらの結果は、Dicer で切断反応において NCD を添加/非添加した pre-miRNA ライブラリーをハイスループットシーケンシング解析することで、NCD 結合により切断が阻害されうる RNA 配列モチーフを特定できたことを示している。また本手法は、RNA や低分子の標識や固定化を必要としないため、異なる標的分子への適用が容易である。今回は 1 種類の低分子を 4096 個の pre-miRNA 変異体に対して検討したが、バーコーディングと全サンプルのプールを利用すれば、より多くの低分子を同時に評価することが可能である。

また、モチーフ探索とその後の解析の結果、同定されたモチーフを持つ miRNA ヘアピンが NCD の標的となり得ることが示唆された。これは我々の手法で同定された RNA モチーフが実際に内在性 RNA 存在すること、そして低分子が作用しうることを確認できたことを示している。

SPR などの結合アッセイを用いると、RNA に対する低分子結合体を同定することができるが、そのような分子のすべてが標的 RNA に対して機能的に活性であるわけではない。我々の手法は、標的 RNA の機能を調節することができる低分子-RNA 相互作用を同定することができる。今後、ハイスループットなシーケンシングデータから標的 pre-miRNA 変異体を同定するための条件をさらに精査する必要があるが、本手法は低分子の新しい RNA 標的の同定に貢献できると考えている。実際に我々は小規模な自社ライブラリーを用いて、化合物×RNA モチーフの大規模スクリーニングが実際に可能であることが証明した。

今後は、得られたデータをもとに、RNA 標的低分子化合物を合成するためのガイドラインをどのように得るかが課題である。

Experimental section

(most parts of experimental section was taken from *ACS Chem. Biol.* **2022**, *17* (10), 2817–2827)

Preparation of the pre-miRNA mutant library.

The double-stranded DNA (dsDNA) library was generated by PCR using a synthesized oligonucleotide containing 6-nt random nucleotides (5'- GCT CCA TTT GTT TTG ATN NNG GAT TCT TAT GCT CCN NNA TCG TCT CAA ATG AGTCT -3', where N represents any nucleotide) as a template with Platinum™ Pfx DNA polymerase (Invitrogen), followed by purification with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel). The primers used for the PCR amplification were: forward primer (5'- TAA TAC GAC TCA CTA TAG CTC CAT TTG TTT TGA -3' [T7 promoter sequence is underlined]) and reverse primer (5' - A*G*A CTC ATT TGA GAC GA -3' [*2'-OMe-modified nucleosides]). The dsDNA library was transcribed *in vitro* with T7 RNA polymerase (MEGAscript™ T7 Transcription Kit, Invitrogen), followed by DNase I treatment. The resulting mixture was separated on an 8% native polyacrylamide gel, and the RNA band similar in size to the wild-type pre-miR-136 was excised and eluted from the gel. The gel-purified RNA was treated with 5' RNA polyphosphatase (Epicentre) to convert triphosphate to monophosphate at the 5' end, and purified by denaturing polyacrylamide electrophoresis.

In vitro Dicer reaction.

RNA was incubated with the in-house-prepared recombinant human Dicer⁷² in the presence or absence of the compound (NCD, 100 μM) at 37 °C. The incubation time varies with different RNA substrates (10 h for the pre-miRNA mutants library and 3 h for the individual mutants). A typical reaction mixture contained 2 μM RNA, 74 nM recombinant human Dicer, 1 mM ATP, 10 mM MgCl₂, 0.5 mg/ml BSA, and 1x Dicer Reaction Buffer (Genlantis) in a 10 μl-reaction. The reaction was stopped by adding Dicer stop solution (Genlantis) and was analyzed by 15% denaturing polyacrylamide gel electrophoresis (acrylamide:bis-acrylamide = 19:1, 6 M urea). The gel was stained with SYBR® Gold and the RNA bands were visualized using a LAS 4010 system and quantified using ImageQuant (GE Healthcare).

Preparation of DNA libraries for sequencing in ion PGM platform.

After *in vitro* Dicer reaction, the reaction mixture was reverse-transcribed with PrimeScript™ II Reverse Transcriptase (TAKARA) at 42 °C for 1 h using the primer (5'- AGA CTC ATT TGA GAC GA -3'). The cDNA was amplified by PCR with the primer

as follows: forward primer (5'- CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG **XXX XXX XXX XGA** TGC TCC ATT TGT TTT GA -3' [**XXXXXXXXXXXX** denotes a barcode sequence for multiplexing]) and reverse primer (5'- CCT CTC TAT GGG CAG TCG GTG ATA GAC TCA TTT GAG ACG A -3'). The PCR products were cleaned with NucleoSpin® Gel and PCR Clean-up kit and quantified by Qubit™ dsDNA HS Assay Kit (Invitrogen). The DNA libraries prepared above were pooled and diluted in nuclease-free water to obtain a final concentration of 8 pM. The library solution was subjected to emulsion PCR with the Ion OneTouch™ 2 system using the Ion PGM™ Hi-Q™ View OT2 200 Kit (Thermo Fisher Scientific). After recovery and enrichment of ion sphere particles, sequencing was performed using The Ion PGM Hi-Q View Sequencing Kit on an Ion 318™ Chip V2 (Thermo Fisher Scientific). The primer sequences used for the library preparation are shown in Table S1-1.

Table S1-1. The sequences of DNA oligos used in the study. The #1~#3 oligos were used for preparation of the pre-miRNA mutants library. The #4~#8 oligos were used for preparation of the libraries for high throughput sequencing. The sequence shown in bold italic in the primer #4~#7 represents the barcode region.

#	Oligos	Sequences
1	pre-miR136_N3x2 (template)	GCTCCATTTGTTTTGATNNNGGATTCTTATGCTCCNNNATCGTCTCAAATGAGTCT
2	pre-miR136_T7-F (Fw primer)	TAATACGACTCACTATAGCTCCATTTGTTTTGA
3	pre-miR136-R_2'Ome (Rev primer)	A*G*ACTCATTGAGACGA (*2'-OMe-modified nucleosides)
4	pre-miR136-F-BC2 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TAAGGAGAAC</i> GATGCTCCATTTGTTTTGA
5	pre-miR136-F-BC3 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>AAGAGGATT</i> CATGCTCCATTTGTTTTGA
6	pre-miR136-F-BC4 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>TACCAAGAT</i> CATGCTCCATTTGTTTTGA
7	pre-miR136-F-BC5 (Fw primer)	CCATCTCATCCCTGCGTGTCTCCGACTCAG <i>CAGAAGGAAC</i> GATGCTCCATTTGTTTTGA
8	pre-miR136-R-trp1 (Rev primer)	CCTCTCTATGGGAGTCCGGTATAGACTCATTTGAGACGA

Preparation of DNA libraries for sequencing in illumina platform.

After *in vitro* Dicer reaction, the reaction mixture was reverse-transcribed with AMV Reverse Transcriptase XL (TOYOBO) at 50 °C for 1 h using the primer (5'- AGA CTC ATT TGA GAC GA -3'). The cDNA was amplified by PCR with the primer as follows: forward primer (5'- TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GCT CCA TTT GTT TTG A -3') and reverse primer (5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GAG ACT CAT TTG AGA CGA -3'). The PCR products were cleaned with Ethanol purification. The resulting DNA libraries were sequenced according to the protocol of the illumina HiSeq 3000 (illumina) platform.

Analysis of sequencing Data.

Sequence reads obtained from each sample were first filtered to remove reads that do not have the 5'-side (GCT CCA TTT GTT TTG AT), 3'-side (ATC GTC TCA AAT GAG TCT), and internal (GGA TTC TTA TGC TCC) fixed sequence. The reads that perfectly matched these fixed sequences were sorted to count the number of reads for each pre-miRNA mutant. Chi-square test of independence was performed to determine whether there is a significant difference in the frequency of each mutant between the library obtained from NCD (-) and NCD (+) samples. The chi-square statistic (χ^2) was calculated for each mutant using the chi-square function in R (version 3.6.3) as follows:

	NCD (-)	NCD (+)	Total
Mutant 1	N_1	N_2	$N_1 + N_2$
Nonmutant 1	n_1	n_2	$n_1 + n_2$
Total	$N_1 + n_1$	$N_2 + n_2$	$N_1 + N_2 + n_1 + n_2$

$\chi^2 = (N_1 + N_2 + n_1 + n_2) (N_1 n_2 - N_2 n_1)^2 / (N_1 + N_2) (N_2 + n_2) (n_2 + n_1) (n_1 + N_1)$, where N_1 and N_2 are the number of reads for mutant 1 obtained from NCD (-) and NCD (+) samples, whereas n_1 and n_2 are the number of reads for nonmutant 1 obtained from NCD (-) and NCD (+) samples. The probability of the chi-square test (p-value) was calculated by applying the R function p-value with one degree of freedom. The resulting p-value was adjusted using Bonferroni's method to obtain the q value by multiplying the p-value by 4096 (the number of pre-miRNA mutant sequences).

Binding analysis by SPR

Interaction between NCD and RNA was evaluated by surface plasmon resonance (SPR) with a Biacore™ T200 SPR system. Chemically synthesized RNAs were diluted to 1 μ M in HBS-EP+ buffer (GE Healthcare), heated at 80°C for 3 min, and then slowly cooled to room temperature for annealing. The resulting RNA solutions were injected in increasing concentrations (0, 0.01, 0.03, 0.1, 0.3, and 1 μ M) to the NCD-immobilized sensor surface with the regeneration step between each concentration. Injections were done at 30 μ l/min with a contact time of 180 sec, followed by dissociation with running buffer for 180 sec. All sensorgrams were corrected by subtraction of blank flow cell response and buffer injection response. The obtained sensorgrams were fitted to a 1:1 binding model to determine the association and dissociation rate constants for RNA-NCD complex using BIAevaluation software (GE Healthcare).

Motif search in endogenous pre-miRNA hairpins.

The sequences of pre-miRNA hairpins (34,675 sequences) were downloaded in FASTA format from the miRBase miRNA database (<https://www.mirbase.org/>). The secondary structure of each input sequence with the lowest free energy was predicted by ViennaRNA (ver.2.4.14). The dot-bracket notation of predicted secondary structure was used to determine the stem region and the terminal loop in each sequence. The RNA motifs identified in the 11 inhibited mutants were searched in the stem region determined above. The hairpin miRNA precursors that were found to contain the RNA motifs were counted and extracted by an in-house script.

第二章

水素結合を介した核酸塩基-低分子間相互作用の

1 分子計測による観測

序論

2-1 研究背景

現在までに、市販されている医薬品の多くはタンパク質を標的としたものであるが、近年では DNA や RNA を標的とする核酸標的分子は、これまでに治療できなかったターゲットへの新しい創薬につながる可能性があるとして、核酸を標的とした医薬品の開発が活発となっている^{73,74}。核酸を標的とする医薬品⁷³⁻⁷⁷としては、核酸医薬の研究が盛んであるが、タンパク質を標的とする医薬品の大半は低分子が占めており、そのうちのいくつかの低分子は核酸を標的とする低分子医薬品となる可能性を秘めている。低分子医薬品のさらなる開発には、塩基分子とリガンドの結合性、選択性、水素結合様式によって決まる、低分子リガンドと塩基分子間の相互作用に関する情報が不可欠である。

ナノポア⁷⁸⁻⁸⁷とナノギャップ⁸⁸⁻⁹³は、電気的計測を利用して単一分子を研究するための強力なナノテクノロジーツールとして発展してきた。ナノポアは、微細なスルーホールを介してイオン電流を変化させることで、単一分子を検出する。実用化されているナノポアの例としては、次世代 DNA シーケンサーがあり、創薬や医学に顕著な影響を与えている⁹⁴⁻⁹⁸。また近年のナノポア開発のトレンドとしては、タンパク質を調べるツールの開発も活発である⁹⁹⁻¹⁰²。

一方、ナノギャップ計測法において 1 分子のナノギャップ間を流れる量子トンネル電流が読み取っているのは、低分子の構造や電子状態である⁸⁸⁻⁹⁰。このナノギャップを用いた電子状態 1 分子計測法の開発は、1 分子計測研究において新たな知見を与えた⁸⁸⁻⁹³。ナノギャップは DNA¹⁰³ や RNA¹⁰⁴ の塩基配列、ペプチドのアミノ酸配列¹⁰⁵ を解読したほか、疾患マーカー⁸⁸ となる化学修飾された塩基分子¹⁰⁶⁻¹⁰⁸ やアミノ酸¹⁰⁵ も解読することができる。また DNA とその認識分子の分子間相互作用、ヌクレオシドやアミノ酸の 1 分子接合も同様に 1 分子計測を用いて研究されている^{109,110}。

現在における単分子計測技術は、AI との融合によってより発展している。例えば、ナノポア次世代 DNA シーケンサーは配列読み取り精度は飛躍的に向上しており¹¹¹、またナノポアと AI の組み合わせは、新規コロナウイルス (SARS-CoV-2) を特異的に高感度かつ迅速に検査する方法を実現させた¹¹²。一方、ナノギャ

ップと AI の組み合わせは、低分子の高精度な同定を実証している^{108,113}。加えて、これら単分子計測技術と AI の組み合わせは、単分子カウント技術の原理である1つの電流時間波形を1つの分子へと帰属させることも可能としており、この分子の種類数をカウントすることができる単分子カウント技術は、単分子計測における定量分析を飛躍的に発展させた^{114,115}。

本研究で筆者らは、核酸標的低分子創薬研究と単分子カウント技術を組み合わせることで、DNA や RNA を用いずに、核酸塩基分子と低分子リガンドとの結合様式、選択性、水素結合様式を1分子レベルで定量的に観測できる技術の開発と、核酸を標的とする低分子創薬における分子デザイン指針を獲得できるツールの開発を目指した。

結果と考察

2-2 グアニン-リガンド間の単分子カウント解析

2-2-1 グアニン認識リガンドの設計

本研究では核酸塩基と低分子リガンド間の水素結合の数に着目し、1分子測定を行った。まずはじめに、グアニンに対し水素結合を形成するリガンドの設計を行った。リガンドの設計は以下のように行った。

グアニンが6つの水素結合形成部位を持つのに対し、7-methylnaphthyridin-2-ylcarbamate (NC) は5つの水素結合形成部位を有している¹¹⁶。したがって、グアニンと NC は1~3つの水素結合をもつ会合体を形成できる (**Figure 2-1a**)。5,7-Methylnaphthyridin-2-ylcarbamate (NCM)¹¹⁷ とナフチリジン (Naph) はそれぞれ3つ、2つの水素結合をもつ会合体形成し (**2-1b**)、キノリン (Quin) は1つ、ナフタレン (NT) は全く形成していない。電子供与性メチル基を持つ NCM は、NC よりも強い水素結合を形成する。以上のように、水素結合の数から、グアニンとリガンドの相互作用の強さが、NCM > NC > Naph > Quin > NT の順となるようにリガンドの設計を行った。

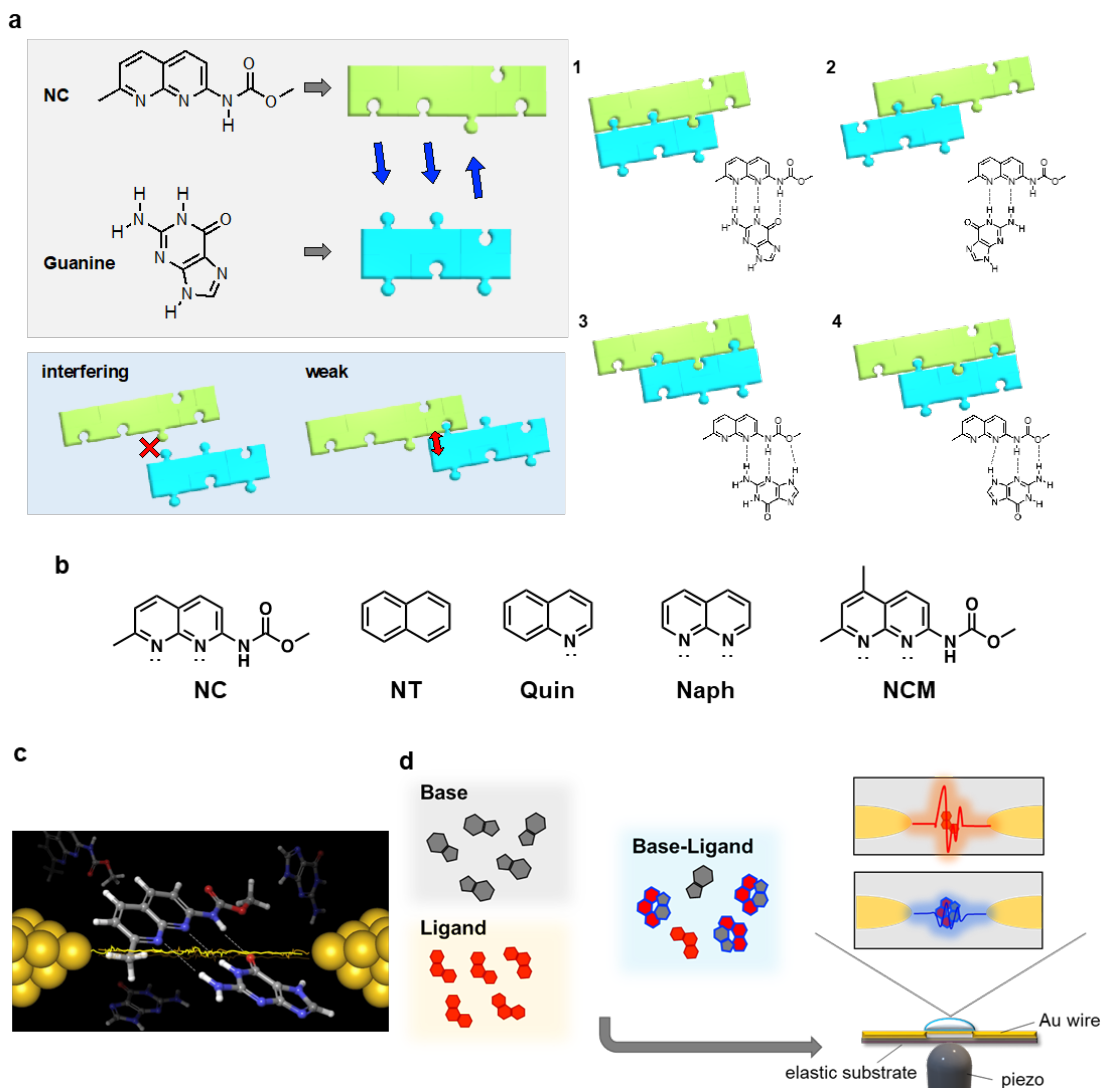


Fig. 2-1 Single-molecule counting of base-ligand aggregates. **a**, Molecular interaction modes between guanine and NC. **b**, Ligand molecules used for aggregate formation. **c**, The principle of the single-molecule counting method. Tunneling current flows between electrodes via single guanine, ligand, and aggregate using nanogap electrodes fabricated by mechanically controllable break junctions (MCBJs). **d**, Solutions of each base molecule, namely, adenine, cytosine, guanine, and thymine, and five ligand molecules were analyzed. Additionally, 1:1 mixed solutions of the base molecules and ligands were also analyzed.

2-2-2 グアニン-リガンド混合溶液の 1 分子測定

続いて、グアニンと低分子リガンド間の 1 分子測定を行った。グアニンとリガンドの混合溶液は、グアニン、リガンド、グアニン-リガンド会合体の 3 種類の分子種を含んでいる。この混合溶液を、電極間距離 0.65 nm の金ナノギャップ上に滴下した。印加電圧 0.1 V でナノギャップ電極間に流れる電流を 10 kHz で測定した (2-1c, d)。さらに、電極間の分子の電子状態を反映するトンネル電流も同時に測定を行った。このトンネル電流時間波形を用いて、グアニン、リガンド、会合体の同定を行った。さらに、グアニンとリガンドの個々のトンネル電流時間波形についても測定を行った。

測定の結果、グアニン及びリガンドの測定ではスパイク状のシグナルが得られた (Figure 2-2a, 2-2b, Figure 2-3a-3i)。これは平均電流 (I_p) と電流時間 (t_d) で特徴付けられる (2-2c, 2-2d)。 I_p は G, NT, Quin, Naph, NC, NCM でそれぞれ 43.4, 24.1, 39.8, 15.2, 40.5, 32.8 pA であった (Table 2-1)。ここで得られた、分子コンダクタンスの平均値 (I_p/V) は透過係数 $4\Gamma_L\Gamma_R/\{(\epsilon_{HOMO} - E_F)^2 + (\Gamma_L + \Gamma_R)^2\}$ に比例していることがわかる¹¹⁸。HOMO と E_F はそれぞれ単分子の最高被占領分子軌道 (HOMO) のエネルギーと金属電極のフェルミエネルギーを表しており、 L, R はそれぞれ、左電極と分子、右電極と分子の間の電子的相互作用を表している。密度汎関数理論 (DFT) に基づく量子化学計算から得られたグアニンと配位子の HOMO エネルギーは、 $G > NT > NC > NCM > Quin > Naph$ の順である (2-2e)¹¹⁹。もし、すべての分子の電極-分子間の電子的相互作用が同じであれば、平均的な単一分子コンダクタンスは HOMO エネルギーと同じオーダーに従い、同じ π 電子系バックボーンを持つリガンドの HOMO エネルギーのオーダーは、平均的な単一分子コンダクタンスのオーダーと同じはずである。しかし、Quin と NT の HOMO エネルギーオーダーは、平均単分子コンダクタンスオーダーと逆になっている。これは、窒素上の非共有電子対により、Quin と Au 電極間の電子的相互作用が、NT と Au 電極間のそれよりも大きくなるためと考えられる⁸⁹。以上のように、ある分子の電子状態を反映する平均単分子コンダクタンスは、グアニンとリガンドを識別するフィンガープリントとなると考えられる。

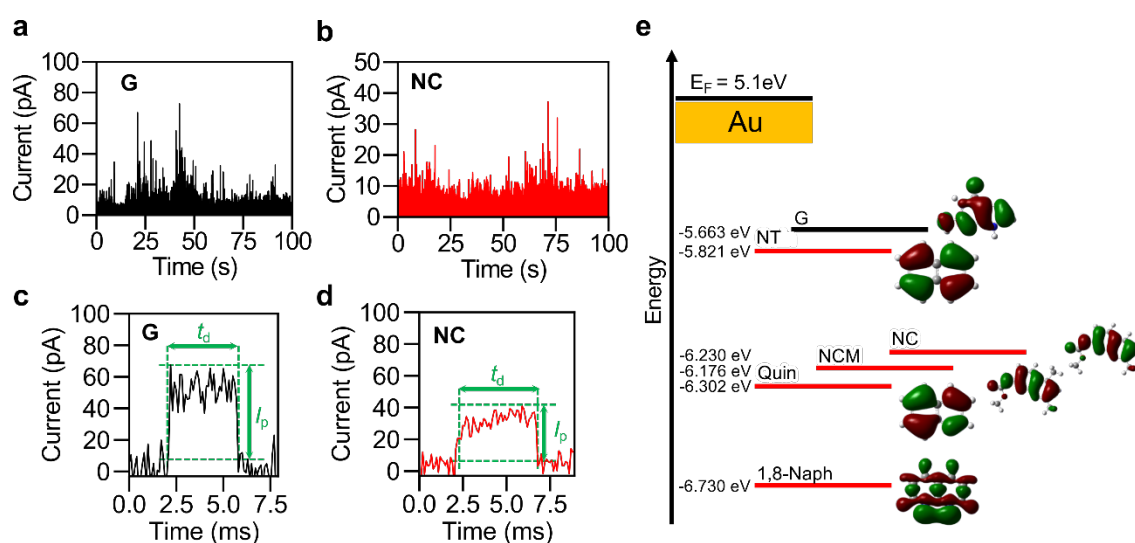


Fig. 2-2 Electrical pulses obtained in single-molecule counting. The current-time profiles obtained when analyzing **a** guanine and **b** NC solutions. **c** and **d** are enlarged views of **a** and **b**. One current-time pulse is characterized by the average current value (I_p) and current duration (t_d). **e**, Energy correlations for guanine, ligand, and gold. The energies for guanine and ligand are the HOMO energies. The gold energy indicates Fermi energy. The phase pattern of the HOMO of each molecule shows isovalue = $0.02\text{ e}^{-\text{\AA}^3}$.

Table 2-1. The average values of each molecule's peak current (I_p) and duration time (t_d).

Molecule	I_p (pA)	t_d (ms)	$I_p V^{-1}$ (pS)
G	43.4	5.1	430
NT	24.1	3.8	241
Quin	39.8	19.7	398
Naph	15.2	17.1	152
NC	40.5	8.0	405
NCM	32.8	4.7	328

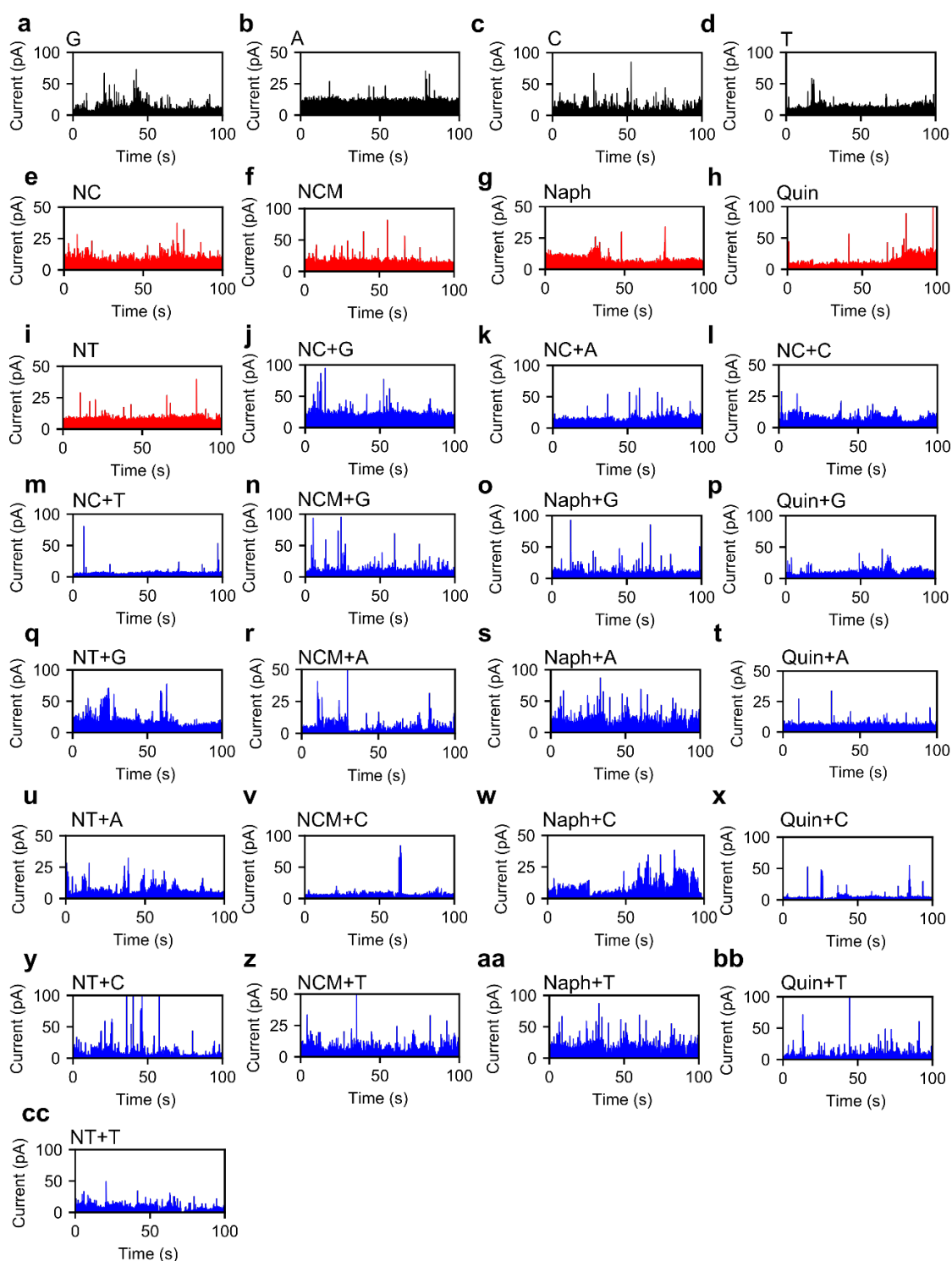


Fig. 2-3 Current-time profiles. **a** Guanine, **b** adenine, **c** cytosine, **d** thymine, **e** NC, **f** NCM, **g** naphthyridine, **h** quinoline, and **i** naphthalene **j** guanine-NC, **k** adenine-NC, **l** cytosine-NC, **m** thymine-NC, **n** guanine-NCM, **o** guanine-Naph, **p** guanine-Quin, and **q** guanine-NT **r** adenine-NCM **s** adenine-Naph **t** adenine-Quin **u** adenine-NT **v** cytosine-NCM **w** cytosine-Naph **x** cytosine-Quin **y** cytosine-NT **z** thymine-NCM **aa** thymine-Naph **bb** thymine-Quin **cc** thymine-NT solutions were analyzed using a single-molecule counting system.

2-3 グアニナーリガンド相互作用における結合の定量的評価

等モル量を混合したグアニナーリガンド溶液を、ナノギャップ電極を用い、個々の分子溶液と同じ測定条件で分析を行った。トンネル電流時間波形における I_p と t_d のヒストグラムを以下に示す (**Figure 2-4a-f, Figure 2-5a-e**)。ヒストグラムの黒線、赤線、青線は、それぞれグアニン溶液、リガンド溶液、グアニナーリガンド溶液で得られた I_p と t_d の値を示している。グアニン溶液の I_p ヒストグラムにおいて、50 pA 付近で見られるショルダーピークは、グアニナーリガンド溶液のヒストグラムでは減少が見られる一方で、グアニナーリガンド溶液の I_p ヒストグラムにおける 20 pA 付近のピークは、それぞれの分子溶液の I_p ヒストグラムと比較して鋭くなっている。この約 20pA の I_p は、グアニナーリガンド会合体の 1 分子コンダクタンスである。しかし、グアニン、リガンド、グアニナーリガンド会合体の I_p ヒストグラムには大きな重なりがあり、この 3 つの分子を識別することは困難であった。

ここで、測定に用いたグアニナーリガンド溶液には、グアニン、リガンド、グアニナーリガンド会合体の 3 種類の分子が存在することに着目する。我々はこの混合溶液に対して、3 種類の分子に対応するスパイクの電気シグナルを得た。このうち、グアニンとリガンドは、それぞれの溶液を測定して得られた電気シグナルと同等である。つまり混合溶液から得られた電気シグナルから、2 つの分子の電気シグナルを差し引くと、グアニナーリガンド会合体の電気シグナルと一致する結果が得られる。

そこで我々はグアニンとリガンドの電気シグナルを正 (positive signal) としてラベル付けし、混合溶液からラベル付けされていない未知の電気シグナル (Unknown signal) の抽出を行った。これを **Positive and Unlabeled Classification (PUC)** 法という (**Figure 2-4g**)¹¹³。

電気シグナルの機械学習には、 I_p と t_d に加えて、電気シグナルの時間依存性のある 10 の特徴量を用いた。(特徴量については **Experimental section** の”**Signal Feature Extraction**”にて後述する。) まず、グアニン、リガンド、グアニナーリガンド溶液の電気シグナルを機械学習した後、PUC 法により電気シグナルを 3 種類の分子に分類した。例えば、100 個の電気シグナルを、グアニン 30 個、リガンド 20 個、グアニナーリガンド会合体 50 個のように分類する。まず初めに、グアニン-NC 混合溶液の 1 分子計測の結果、得られた電気シグナルの 71.8% がグアニン-NC 会合体に起因するものであることが明らかとなった。会合体の I_p 値は、グアニンや NC のものよりも低い値を示している (**Figure 2-4h**)。会合体の電流が小さくなった理由として、水素結合の形成により、窒素原子が金電極と電子的に相互作用することがなくなり、電極との結合が弱くなったこと、トンネル

距離が長くなったことが考えられる。また、水素結合は1分子コンダクタンスを増加させることが既に報告されているが^{120,121}、今回の結果よりトンネル距離の増加は水素結合による電流増加と競合し、電流を減少させることがわかった。また、会合体によるコンフォメーション変化も電流減少の原因となると考えられる。

グアニンと他の4つのリガンドを混合した溶液の1分子計測も同様に行ったところ、得られた電気シグナルの71.7% (NCM), 65.9% (Naph), 54.1% (Quin), 20.1% (NT)が会合体に起因していた (**Figure 2-4i, Table 2-2**)。グアニンとリガンド間の水素結合の数の順序は会合体由来のシグナルの比率の順序と一致した。このことから、得られた会合体由来のシグナルの比率は、リガンド結合の直接的な指標となると考えられる。また、量子化学計算により構造最適化された各集合体の結合エネルギーは-1.169 eV (NCM), -1.157 eV (NC), -0.822 eV (Naph) -0.466 eV (Quin) (**Table 2-3**)で、こちらも会合体由来のシグナルの比率の順序と一致した。

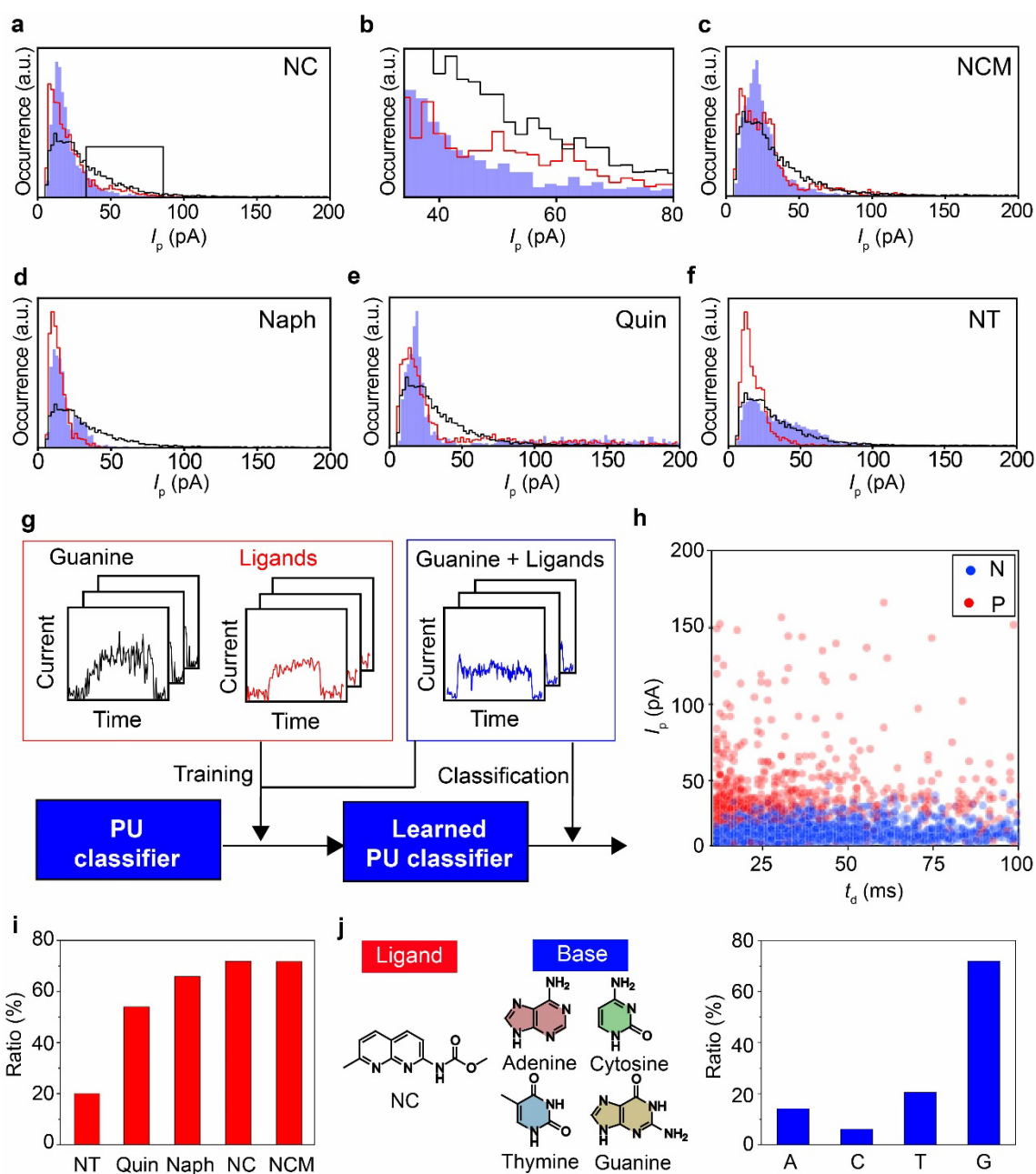


Fig. 2-4 Quantitative analysis of base-ligand binding and selectivity. I_p histograms obtained by analyzing solutions of guanine and **a** NC, **b** close-up view of **a**, **c** NMC, **d** Naph, **e** Quin, and **f** NT aggregates. In the histograms for each ligand, the black, red, and blue lines correspond to solutions of guanine, ligand, and guanine-ligand, respectively. **g**, Machine learning flowchart for extracting the electrical pulses of guanine-ligand aggregates from electrical pulses obtained from guanine-ligand solution analysis. PUC indicates a positive unlabeled classifier. **h**, 2-dimensional maps of I_p and t_d of the electrical signals classified as guanine and NC (red) and guanine-NC aggregates (blue). **i**, Guanine-ligand binding properties. Values on the vertical axis indicate the percentage of aggregates present. **j**, Base molecule selectivity of NC. Again, the vertical axis indicates the percentage of aggregates present.

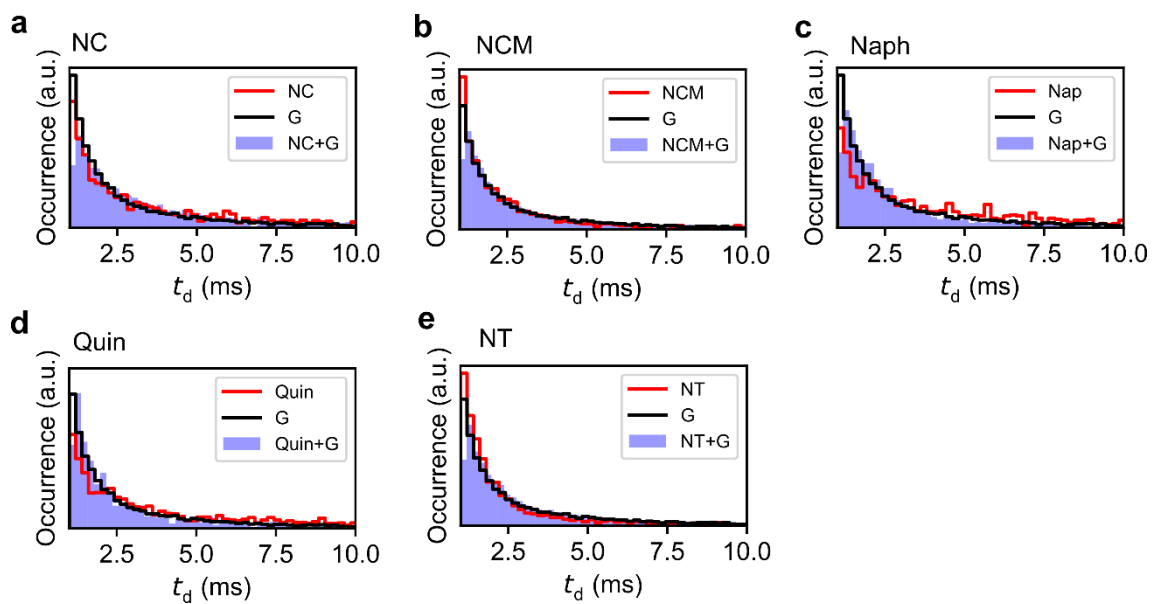


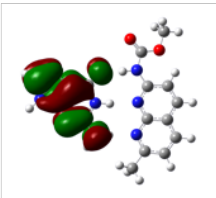
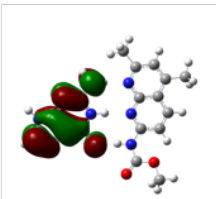
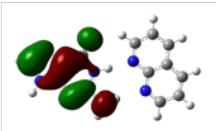
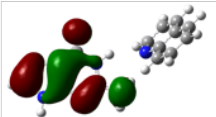
Fig. 2-5 t_d histograms of guanine, ligand, and guanine-ligand equimolar solutions. Histograms of equimolar solutions between guanine and **a** NC, **b** NCM, **c** Naph, **d** Quin, and **e** Naph. In each histogram, red, black, and blue indicate histograms of the ligand, guanine, and equimolar solutions, respectively.

Table 2-2 Ratio of G-ligand signals to the total obtained signals for the base molecules. The analysis was performed ten times under the same conditions, and the average value was calculated. The number of associated signals was determined by the PUC method. N and P indicate the number of signals classified as negative and positive, respectively.

Base	NT			Quin			Naph		
	N	P	N/(N+P)	N	P	N/(N+P)	N	P	N/(N+P)
Number of signals	716	4461	0.26	670	749	0.54	742	300	0.65
	1334	3843	0.27	761	658	0.54	682	360	0.68
	1391	3786	0.26	767	652	0.55	712	330	0.69
	1359	3818	0.12	774	645	0.44	723	319	0.71
	636	4541	0.17	625	794	0.57	735	307	0.59
	874	4303	0.19	803	616	0.60	616	426	0.61
	961	4216	0.16	850	569	0.55	639	403	0.65
	808	4369	0.27	778	641	0.60	679	363	0.62
	1412	3765	0.18	848	571	0.56	646	396	0.66
	906	4271	0.26	792	627	0.54	690	352	0.65
Average	1040	4137	0.20	767	652	0.54	686	356	0.66

Base	NC			NCM		
	N	P	N/(N+P)	N	P	N/(N+P)
Number of signals	2729	822	0.72	3997	1574	0.67
	2572	979	0.75	3754	1817	0.77
	2675	876	0.67	4305	1266	0.59
	2371	1180	0.69	3297	2274	0.76
	2465	1086	0.74	4234	1337	0.63
	2639	912	0.75	3486	2085	0.73
	2665	886	0.67	4077	1494	0.75
	2376	1175	0.67	4186	1385	0.81
	2382	1169	0.74	4494	1077	0.73
	2632	919	0.72	4090	1481	0.67
Average	2551	1000	0.72	3992	1579	0.72

Table 2-3 Intensity of the binding energies for the aggregates. Gaussian09 Package performed all calculations. The experimental results gave the average values of each complex's peak current (I_p) and duration time (t_d). The figures show the HOMO electronic state of the aggregates.

	HOMO	HOMO alignment / eV	Binding Energy / $\text{kJ}\cdot\text{mol}^{-1}$	I_p / pA	t_d / ms
NC		-5.144	-112	30.7	6.8
NCM		-5.113	-113	31.6	5.1
1,8-Naph		-4.896	-79	20.2	4.2
Quin		-5.173	-45	50.5	6.9
NT	—	—	—	43.1	5.2

2-4 核酸塩基-リガンド相互作用における選択性の定量的評価

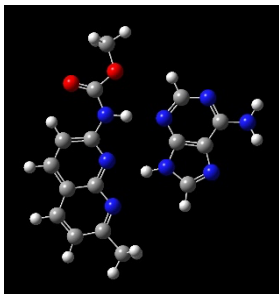
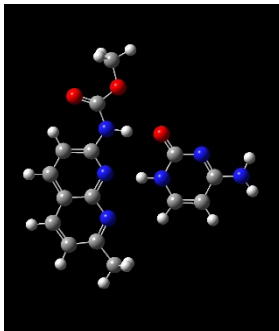
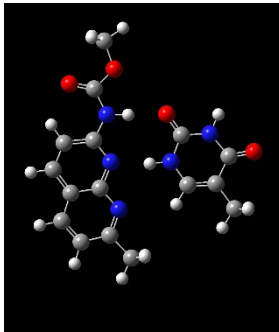
続いて、グアニン-リガンド相互作用と同様に、核酸塩基-リガンド相互作用の1分子測定を行った。NC-核酸塩基分子混合物のトンネル電流時間プロファイルを計測し、得られた電気シグナルをPUC法により機械学習を行った。解析の結果、各NC-塩基性分子混合物の電気信号における塩基-NC会合体の存在比は、71.8% (G), 20.5% (T), 14.1% (A), 6.1% (C) であった (Figure 2-4j, Table 2-4)。また量子化学計算によって構造最適化された各集合体の結合エネルギーは、-1.161 eV (A), -1.223 eV (C), -1.157 eV (G), -1.192 eV (T) であった (Table 2-5)。この結果はグアニン-リガンド会合体とは異なり、結合エネルギーの順序は会合体の比率の順序と一致しなかった。我々の単純な計算では、エンタルピー項のみを計算し、エントロピー項を無視したため、自由エネルギーを評価することができない。また溶媒分子の影響は考慮されておらず、溶媒分子の効果を含めた自由エネルギーを計算するためには、局所的な結合エネルギーだけでなく、分子動力学を用いてアンサンブル構造を計算する必要があり、多大な計算コストがかかる¹²²⁻¹²⁴。

しかし、我々の1分子計測は、溶液中の1分子レベルでの核酸塩基-リガンド相互作用の選択性を定量的に評価することにも成功した。

Table 2-4 Ratios of base-ligand signals to the total obtained signals for A, C, and T. The analysis was performed ten times under the same conditions, and the average value was calculated. The number of associated signals was determined by the PUC method. N and P indicate the number of signals classified as negative and positive, respectively.

Base	A			C			T		
	N	P	N/(N+P)	N	P	N/(N+P)	N	P	N/(N+P)
Number of signals	4	195	0.020	12	971	0.012	7	90	0.072
	22	177	0.11	160	823	0.16	15	82	0.15
	22	177	0.11	12	971	0.012	10	87	0.10
	23	176	0.12	12	971	0.012	23	74	0.24
	22	177	0.11	12	971	0.012	10	87	0.10
	22	177	0.11	27	956	0.27	21	76	0.22
	68	131	0.34	12	971	0.012	54	43	0.56
	30	169	0.15	11	972	0.010	27	70	0.28
	56	143	0.28	22	961	0.022	18	79	0.19
	11	188	0.055	315	668	0.32	14	83	0.14
Average	28	171	0.14	60	924	0.061	20	77	0.21

Table 2-5 Optimized structures and binding energies of NC between A, C, and T. All calculations were performed using the Gaussian09 Package.

Base	Structure	Binding energy (kJ mol ⁻¹)
A		-112
C		-118
T		-115

2-5 グアニン-リガンド相互作用における水素結合モードの定量的評価

核酸塩基-リガンド相互作用の微細構造は、水素結合の組み合わせパターンが複数存在するため、1分子コンダクタンスもそれぞれの構造に対応した複数のパターンが存在する。したがって、会合体の1分子コンダクタンスは、複数の1分子コンダクタンスから構成されていると考えられる。

本研究では、グアニンと NC の混合溶液をモデルとして、ガウス混合モデル (GMM) を用いて電気シグナルの特徴量の抽出とクラスタリングを行った (Figure 2-6a)¹²⁵。最適クラス数はクラス数の増加に伴い誤差が減少しペナルティ項が増加する「ベイズ情報量基準 (BIC)」により決定した。本手法では BIC 値が最小となれば最も妥当なクラス数であると判断できる¹²⁶。

まずはじめに、グアニンおよび NC の非会合状態と予測されたシグナルの混合データを用い、特徴量分類の後、BIC を計算したところ、クラス数は 3 で最小値を示した (Figure 2-6b,c)。3つのクラスの I_p ヒストグラムは、電流値の頻度が低い順に、黒 (低)、赤 (中)、青 (高) で示した (2-6c)。黒色のクラスターでは電流持続時間が短いものの割合が最も多く、これはブランク溶液電流に起因するものであると考えられる。また NC とグアニンの 1 分子コンダクタンスをそれぞれ赤と青のクラスターに示したが、それぞれの個別電気シグナルを用いたため、クラスターは 2 つのみであった。

一方、グアニン-NC 会合体由来の電気シグナルについても特徴量の抽出とクラスタリングを行った。その結果、グアニン-NC 会合体由来の電気シグナルの BIC では、最低 5 つのクラスが示された (Fig.2-6d,e, Table 2-6)。会合体の電気シグナルを抽出するために PUC を用いたため、この 5 クラスにはブランクの電気シグナルは含まれていない。つまり、5 つのクラスターの会合体は、水素結合様式の違いにより、異なる構造を持つ可能性がある。可能性のある会合体構造について、DFT を用いて検討を行った。Watson-Click 構造, Hoogsteen 構造, π -stack 構造を初期構造として計算を行ったところ、5 つの準安定構造が得られた (Figure 2-6f)¹²⁷。5 つの水素結合モードをそれぞれ会合体 1-5 とする。5 つに分解された I_p と t_d は、それぞれ I_p が大きい青と赤、それぞれ 12, 14 pA 付近の緑とオレンジ、 t_d が小さい紫に分類され、紫が会合体 5 に対応することが示された。

同様に, Quin, Naph, NCM とのグアニン会合体における水素結合モードの数は、それぞれ 3, 4, 8 と決定された (Figure 2-7)。水素結合モードの数が少ない順は、NCM > NC > Naph > Quin であり、これは会合体由来電気シグナルの比率の順番と一致していた。

2-6 解離定数の比較

解離定数は塩基-リガンド会合体を評価する上で重要なパラメータである。単分子カウント技術によってカウントされた水素結合数と結合との強い正の相関は、溶液中の解離定数を1分子レベルで計算できる可能性を秘めている。本実験では1分子計測により、グアニン-リガンド会合反応における会合体の比率(r)が実験的に得られた。グアニンとリガンドの初期濃度が同じ a M の場合、平衡時の会合体モル濃度は $2ar/(1+r)$ で与えられ、解離定数 (K_d) は $a(1-r)/2r$ で表される。 $a=1$ M の場合、各会合体の K_d は、1.3 M (NT) > 126 nM (Quin) > 53 nM (Naph) > 36 nM (NCM) > 32 nM (NC) となる。同様に、4種類の塩基分子による NC の K_d は、6.8 M (C) > 2.3 M (A) > 1.3 M (C) > 32 nM (G) であった。マクロな測定による解離定数と比較すると、1分子計測による結果は、表面プラズモン共鳴法による2本鎖DNA中のNCの $K_d=53$ nM よりもわずかに低かった¹²⁸。これはDNA中のNCの水素結合モードの数が、立体障害のために制限されているためであると考えられる。このことは、水素結合モードの数が多いほど結合系は強くなり、溶液中のグアニン-NC会合体の K_d はDNA会合体の K_d より小さくなることを意味している。また、1分子計測の結果は、融解温度 (T_m) 測定と一致した。Gに対するNC二量体の T_m は上昇し、Quin二量体の T_m は上昇しなかった²²。

以上のようにトンネル電流による1分子計測と機械学習を組み合わせた本手法では、溶液中の核酸塩基分子とリガンド間の結合様式、選択性、水素結合様式を定量的に評価しながら、同定した分子数をカウントすることができた。さらに、得られた会合体由来のシグナルの比率から、結合性や選択性を解離定数で評価した。

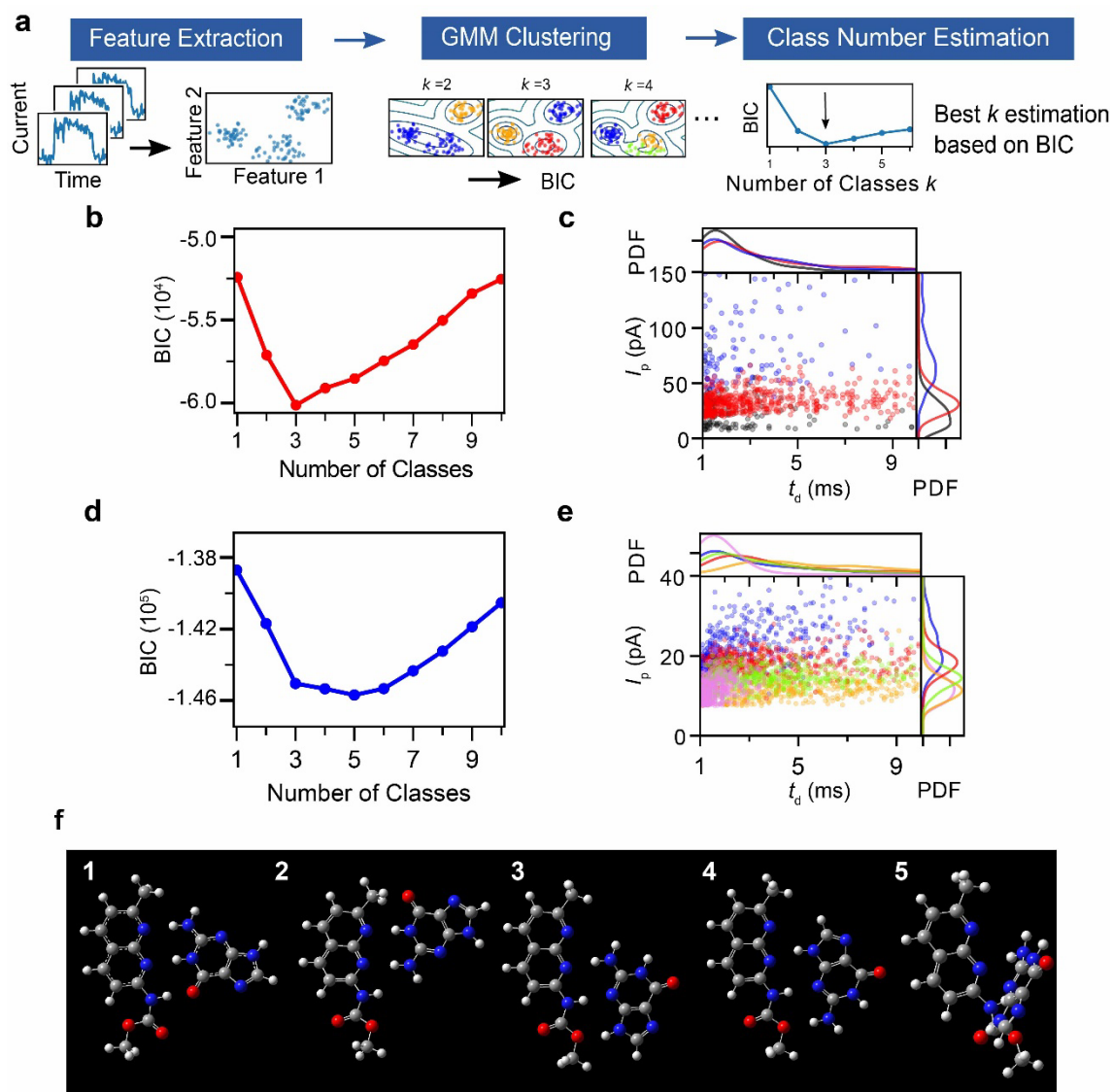


Fig. 2-6 Microscopic hydrogen bonding modes of aggregates. **a**, Estimation protocol for the number of different aggregates. The features extracted from the signals were clustered by Gaussian mixture model (GMM) clustering, assuming the number of clusters (k). Bayesian information criterion (BIC) was calculated from the model. The minimization of BIC determined the most appropriate class number. **b**, BIC with mixed guanine and NC data. **c**, I_p - t_d plots of signals divided into the number of clusters giving minimum BIC. Black, red, and blue correspond to base current, NC, and guanine. **d**, BIC of the classification with each class number for the aggregates between G and NC. **e**, I_p - t_d plot of G-NC aggregates divided into the number of clusters giving minimum BIC. **f**, Five candidate structures of the G-NC aggregate. Density functional theory calculations optimized the five structures.

Table 2-6 The five clusters' average peak current I_p and duration time t_d . Each color code corresponds to the color code in Fig. 4f.

Cluster	I_p (pA)	t_d (ms)
Blue	23	3.2
Red	18	4.1
Green	14	3.5
Orange	12	5.7
Violet	12	1.7

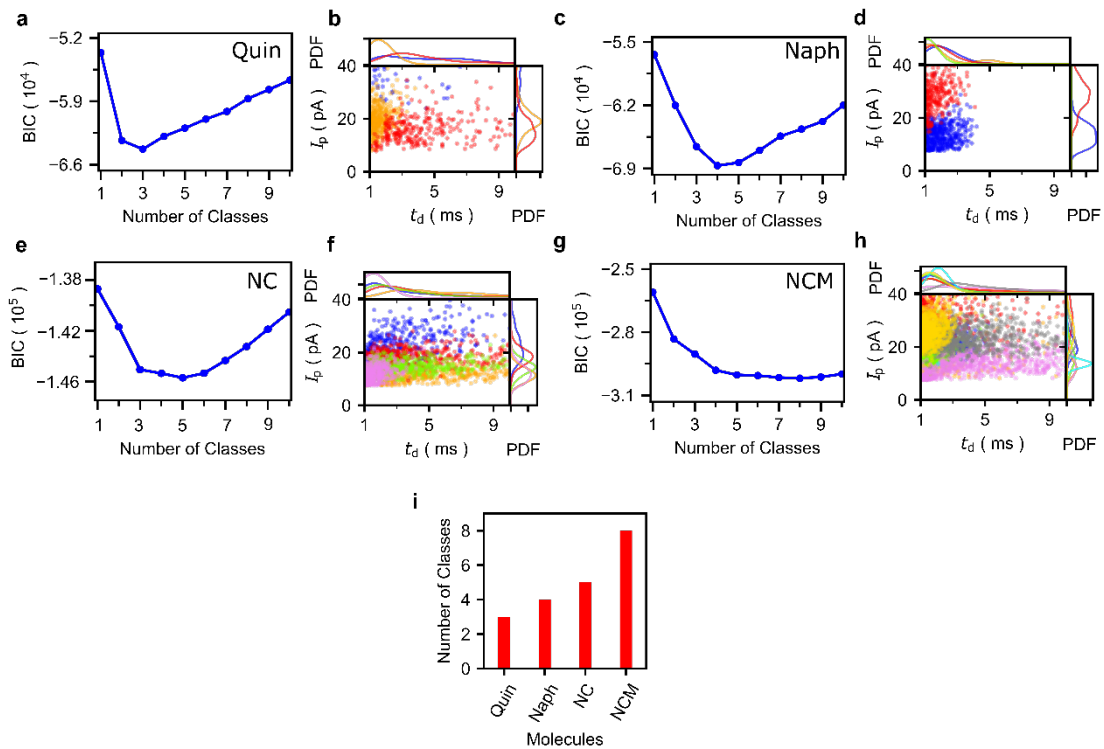


Fig. 2-7 Relation between BIC and the number of classes in GMM clustering. BICs and GMM clusters for **a, b** Quin, **c, d** Naph, **e, f** NC, and **g, h** NCM. **i** The number of classes that provides minimum BIC is 3, 4, 5, and 8 for Quin, Naph, NC, and NCM, respectively.

結論

筆者らはグアニンと低分子リガンドの分子間相互作用を直接評価する方法として、1 分子量子計測法を利用することで、グアニンとリガンド分子の会合状態シグナルの 1 分子スケールでの検出を達成した。その結果、リガンド分子が強く会合しているほど、会合体由来の電気シグナルの全シグナル中に占める割合が有意に大きくなることを見いだした。この結果は、会合状態のシグナルの数が複数の水素結合を介した会合状態の強度と密接に関連していることを示している。さらに、シグナルのクラスタリング法を用いて、会合状態のクラス数を推定した。これは、DFT 計算で推定された準安定状態、および各シグナル統計解析に基づいて、溶液中の各シグナル群の潜在的な会合状態を 1 分子レベルで推定可能であることを示している。これは、各シグナルに対して数種類の会合状態を同定した世界初の例である。統計熱力学を理解するためには、分子の状態を 1 つ 1 つ同定しカウントする必要があるため、これまではマクロな観察により分子状態を積算する方法がとられてきた。しかし、本手法によって、統計熱力学に基づく 1 分子検出の実証の第一歩を実現することができたと考えている。筆者はこの 1 分子計測が、核酸を標的とした低分子化合物の開発を加速させると期待している。

Experimental section

(most parts of experimental section was taken from *J. Am. Chem. Soc.* **2023**, *145* (2), 1310–1318)

Reagents. The reagents and solvents for chemical synthesis obtained from the usual suppliers were used without further purification. Quinoline, naphthalene, and naphthylidene were procured from Sigma-Aldrich. Additionally, 7-methyl-naphthyridin-2-yl-carbamate (NC)¹¹⁶ and 5,7-methyl-naphthyridin-2-yl-carbamate (NCM)¹¹⁷ were synthesized following the procedure reported in a previous study.

Single-molecule electrical measurement. A 0.6 nm electrode gap was prepared in a dilute sample molecule solution in Milli-Q water at 1 μ M concentration. The gap size was set at 0.65 nm for all measurements and tuned finely using the piezoelectric element. Nanofabricated mechanically controllable break-junctions (MCBJs), fabricated using the procedure described elsewhere (S1),¹²⁹ were employed to construct nanogap electrodes. Base molecule solutions were used in single-molecule counting using nanogap electrode devices. The gap distance was maintained via the feedback control of the piezo-actuators during the measurements, which could be estimated using the following direct tunneling current equation: $I \propto \exp(-4\pi\sqrt{2mwl}/h)$, where h , m , w , and l respectively represent the Planck's constant, electron mass, work function of the gold electrode, and gap distance; $m = 9.1 \times 10^{-31}$ kg and $w = 5.1$ eV (work function of Au(111)) were used. A customized current amplifier boosted the current across the electrodes with an NI PXIe-4081 digital multimeter and an NI PXI-5922 high-resolution oscilloscope (National Instruments) at DC bias voltage = 100 mV, measured it at 10 kHz and 100 kHz. The MCBJ sample was replaced with a new one after each hour of measuring I - t . All measurements were performed at 298K.

Signal pickup. Signal pickup was automatically performed using Python. First, the start of the signal from the estimated baseline was considered the signal-rise time, having a value more than six times greater than the threshold noise value. Thus, the signal began when the noise level was exceeded. Next, the signal was considered to have ended with the appearance of a data point below the noise level, as revealed by the estimated baseline post-signal beginning. The baseline can be described as a moving average filtered signal of FFT (Scipy 0.24.2) current data. Signals with a maximum current (I_p) in the range of 7.5–500 pA and a duration (t_d) of 1.00 ms were used for the positive and unlabeled data classification (PUC) analysis.

Data analysis. The discrimination analysis of aggregates constituted the processes of signal pickup, feature extraction, and PUC analysis. First, single-molecule pulse signals were extracted from the current raw data described in the next section, followed by the conversion of each current signal to a 13-dimensional feature vector. The maximum current in the signal (I_p), average current value (I_{ave}), signal duration (t_d), and ten-dimensional shape factors are the elements of the feature vector. The definition of ten-dimensional shape factors is given in “**Signal Feature Extraction**” (Experimental section). A positive and unlabeled classifier was used to classify the feature vectors, as described in a later section. Python 3.9.7 with libraries, NumPy 1.20.3, SciPy 1.7.1, npTDMS 1.4.0, and scikit-learn 0.24.2 were used to perform the analysis.¹²⁵

Signal Feature Extraction

Before signal machine learning analysis, we extracted the signal characters as follows. Each signal was converted into the following feature parameter values, which contain the peak value of the current signal (I_p), average signal value (I_{ave}), duration time (t_d), and ten-dimensional shape factors ($S_n = I_n/I_p$ ($n = 1, 2, \dots, 10$), where the average current value for each time region ($I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9,$ and I_{10}) was calculated.

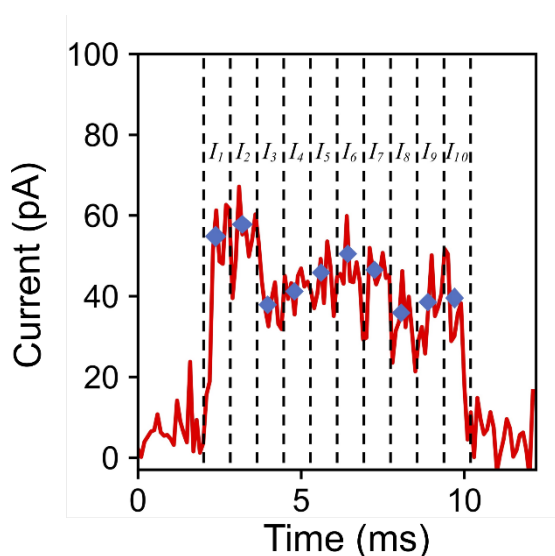


Figure. S2-1 Definitions of features ($S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9,$ and S_{10}) of signals used for machine learning. We used thirteen feature parameters, including the following: the peak value of the current signal (I_p), average signal value (I_{ave}), duration time (t_d), and ten-dimensional shape factors (S_1 – S_{10}). First, the average current value (blue circle) for each time region (I_1 – I_{10}) was calculated. Then, each current value was normalized by the maximum current value of the signal (I_p), and the values in each region were defined as shape factors (S_1 – S_{10}).

Positive and unlabeled data classification for single analysis: PUC analysis was performed to discriminate the associated states, which is suitable for detecting the signals in a single group. Elkan and Noto's method is the basis of the PUC algorithm.¹³⁰ The Gaussian Naïve Bayes was used as the internal classifier in the scikit-learn program (library version 0.24.2). The signals for the base-ligand solutions can be due to the base-ligand aggregates, base, and ligand. While the training signals from the base and ligand solutions were discriminated as positive, the aggregates signals were discriminated as negative. The signals were randomly extracted for training and classification ten times. The existence ratio of aggregates is the average ratio of 10 times PU classification. The number of signals used for PUC analysis is listed in Tables S1 and S3 (SI).

Signal Clustering: After PUC analysis, the signals were discriminated based on the prediction results. The number of clusters k was set from 1 to 10 range. The signals predicted as negative were clustered using the Gaussian mixture model in the scikit-learn program (library version 0.24.2). The same 13-dimensional features utilized in PUC were applied for clustering. After clustering into k clusters, BIC was calculated using the Gaussian Mixture class method. The best class number was determined from the minimum BIC.

総括

以上、本研究においては RNA を標的とする低分子化合物の設計指針の獲得を目指し、核酸-低分子相関の観測を軸とした研究展開を行った。

第一章では、pre-miRNA の Dicer による切断反応を指標として、低分子が結合する RNA 配列を網羅的に探索、評価する手法を提案した。本手法は、特定の低分子化合物と結合する標的 RNA モチーフを網羅的に解析するための簡便かつ効率的な手法である。また本手法は miRNA を対象とした創薬のセレクション技術として用いることが可能であり、これまでに開発されてきた既存薬を用いた、「ドラッグリポジショニング」にも応用可能な技術であると考えられる。つまり、本研究は当研究室で合成された化合物の薬理活性を評価できるだけでなく、得られたデータ、及び本手法そのものが将来的には既存薬や新薬の開発につながると考えている。

第二章では微細加工を用いて作製されたナノギャップ電極間の分子を介したトンネル電流を計測する単分子計測技術と人工知能(AI)による単分子電流シグナル解析を組み合わせた単分子カウント技術を開発し、1分子中における塩基分子と低分子の結合性、選択性、水素結合様式などのパラメータを定量化できることを示した。本手法では低分子リガンドと核酸塩基の混合溶液を分析するのみで、4種類の核酸塩基分子に対する5種類の低分子リガンドそれぞれの結合選択性を、会合体の存在割合によって定量的に評価することが可能である。さらに、1分子計測と量子化学計算により、系中の水素結合の様式と存在割合が低分子リガンドに依存すること分かった。本研究によって筆者らは1分子計測が、DNA や RNA に低分子リガンドを取り込む必要なく、溶液中の会合体の性質について定量的な情報を提供できることを証明した。将来的には、本技術によって得られる核酸塩基-低分子リガンドの相関情報が、新規に低分子リガンドを合成した際の評価基準の一つになると考えている。

以上のように、筆者は2つのアプローチから核酸-低分子相関情報の獲得を志向した新規手法の開発を行った。これらの新規手法は、RNA と低分子の相互作用の理解をさらに深め、低分子と RNA の結合に必要な新要素の発見を促進すると考えている。今後は、より多くの低分子・核酸について本手法を適用し、得られたデータを情報科学的に解析することで、標的 RNA の結合に重要な化学構造を抽出し、新規 RNA 結合低分子を設計する際の指針の獲得に利用できると期待している。

主論文

1. **Method for Identifying Sequence Motifs in Pre-miRNAs for Small-Molecule Binding**

Takashima, Y.; Murata, A.; Iida, K.; Sugai, A.; Hagiwara, M. and Nakatani, K.
ACS Chem. Biol., **2022**, *17* (10), 2817–2827.

2. **Quantitative Microscopic Observation of Base-Ligand Interactions via Hydrogen Bonds by Single-Molecule Counting**

Takashima, Y.; Komoto, Y.; Ohshiro, T. Nakatani, K.; Taniguchi, M.
J. Am. Chem. Soc. **2023**, *145* (2), 1310-1318.

参考論文

1. **Small molecule-induced dimerization of hairpin RNA interfered with the Dicer cleavage reaction**

Murata, A.; Mori, Y.; Di, Y.; Sugai, A.; Das, B.; Takashima, Y.; Nakatani, K.
Biochemistry **2021**, *60*, 245-249.

引用文献

- (1) Pennisi, E. ENCODE Project Writes Eulogy for Junk DNA. *Science (1979)* **2012**, *337* (6099), 1159–1161. <https://doi.org/10.1126/science.337.6099.1159>.
- (2) Kaikkonen, M. U.; Lam, M. T. Y.; Glass, C. K. Non-Coding RNAs as Regulators of Gene Expression and Epigenetics. *Cardiovasc Res* **2011**, *90* (3), 430–440. <https://doi.org/10.1093/cvr/cvr097>.
- (3) Scott, M. S.; Ono, M. From SnoRNA to MiRNA: Dual Function Regulatory Non-Coding RNAs. *Biochimie* **2011**, *93* (11), 1987–1992. <https://doi.org/10.1016/j.biochi.2011.05.026>.
- (4) St.Laurent, G.; Wahlestedt, C.; Kapranov, P. The Landscape of Long Noncoding RNA Classification. *Trends in Genetics* **2015**, *31* (5), 239–251. <https://doi.org/10.1016/j.tig.2015.03.007>.

- (5) Reichetzeder, C. Overweight and Obesity in Pregnancy: Their Impact on Epigenetics. *Eur J Clin Nutr* **2021**, *75* (12), 1710–1722. <https://doi.org/10.1038/s41430-021-00905-6>.
- (6) Gomes, C. P. da C.; Ágg, B.; Andova, A.; Arslan, S.; Baker, A.; Barteková, M.; Beis, D.; Betsou, F.; Bezzina Wettinger, S.; Bugarski, B.; Condorelli, G.; da Silva, G. J. J.; Danilin, S.; de Gonzalo-Calvo, D.; Buil, A.; Carmo-Fonseca, M.; Enguita, F. J.; Felekkis, K.; Ferdinandy, P.; Gyöngyösi, M.; Hackl, M.; Karaduzovic-Hadziabdic, K.; Hellemans, J.; Heymans, S.; Hlavackova, M.; Hoydal, M. A.; Jankovic, A.; Jusic, A.; Kardassis, D.; Kerkelä, R.; Kuster, G. M.; Lakkisto, P.; Leszek, P.; Lustrek, M.; Maegdefessel, L.; Martelli, F.; Novella, S.; O'Brien, T.; Papaneophytou, C.; Pedrazzini, T.; Pinet, F.; Popescu, O.; Potočnjak, I.; Robinson, E.; Sasson, S.; Scholz, M.; Simionescu, M.; Stoll, M.; Varga, Z. v; Vinciguerra, M.; Xuereb, A.; Yilmaz, M. B.; Emanuelli, C.; Devaux, Y. Catalyzing Transcriptomics Research in Cardiovascular Disease: The CardioRNA COST Action CA17129. *Noncoding RNA* **2019**, *5* (2). <https://doi.org/10.3390/ncrna5020031>.
- (7) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat Rev Drug Discov* **2018**, *17* (8), 547–558. <https://doi.org/10.1038/nrd.2018.93>.
- (8) Esteller, M. Non-Coding RNAs in Human Disease. *Nat Rev Genet* **2011**, *12* (12), 861–874. <https://doi.org/10.1038/nrg3074>.
- (9) Tor, Y. Targeting RNA with Small Molecules. *ChemBioChem* **2003**, *4* (10), 998–1007. <https://doi.org/10.1002/cbic.200300680>.
- (10) Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; De Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T. Selective Small-Molecule Inhibition of an RNA Structural Element. *Nature* **2015**, *526* (7575), 672–677. <https://doi.org/10.1038/nature15542>.
- (11) Morgan, B. S.; Forte, J. E.; Culver, R. N.; Zhang, Y.; Hargrove, A. E. Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angewandte Chemie International Edition* **2017**, *56* (43), 13498–13502. <https://doi.org/10.1002/anie.201707641>.

- (12) Donlic, A.; Hargrove, A. E. Targeting RNA in Mammalian Systems with Small Molecules. *Wiley Interdiscip Rev RNA* **2018**, *9* (4), e1477.
<https://doi.org/10.1002/wrna.1477>.
- (13) Sztuba-Solinska, J.; Chavez-Calvillo, G.; Cline, S. E. Unveiling the Druggable RNA Targets and Small Molecule Therapeutics. *Bioorg Med Chem* **2019**, *27* (10), 2149–2165. <https://doi.org/10.1016/j.bmc.2019.03.057>.
- (14) 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* **1998**, *391* (6669), 806–811.
<https://doi.org/10.1038/35888>.
- (15) Ratni, H.; Ebeling, M.; Baird, J.; Bendels, S.; Bylund, J.; Chen, K. S.; Denk, N.; Feng, Z.; Green, L.; Guerard, M.; Jablonski, P.; Jacobsen, B.; Khwaja, O.; Kletzl, H.; Ko, C.-P.; Kustermann, S.; Marquet, A.; Metzger, F.; Mueller, B.; Naryshkin, N. A.; Paushkin, S. v.; Pinard, E.; Poirier, A.; Reutlinger, M.; Weetall, M.; Zeller, A.; Zhao, X.; Mueller, L. Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA). *J Med Chem* **2018**, *61* (15), 6501–6517.
<https://doi.org/10.1021/acs.jmedchem.8b00741>.
- (16) Peng, T.; Dohno, C.; Nakatani, K. Mismatch-Binding Ligands Function as a Molecular Glue for DNA. *Angewandte Chemie - International Edition* **2006**, *45* (34), 5623–5626. <https://doi.org/10.1002/anie.200601190>.
- (17) Dohno, C.; Nakatani, K. Control of DNA Hybridization by Photoswitchable Molecular Glue. *Chem Soc Rev* **2011**, *40* (12), 5718.
<https://doi.org/10.1039/c1cs15062f>.
- (18) Peng, T.; Nakatani, K. Binding of Naphthyridine Carbamate Dimer to the (CGG)_n Repeat Results in the Disruption of the G–C Base Pairing. *Angewandte Chemie International Edition* **2005**, *44* (44), 7280–7283.
<https://doi.org/https://doi.org/10.1002/anie.200502282>.
- (19) Dohno, C.; Kohyama, I.; Kimura, M.; Hagihara, M.; Nakatani, K. A Synthetic Riboswitch That Operates Using a Rationally Designed Ligand-RNA Pair. *Angewandte Chemie - International Edition* **2013**, *52* (38), 9976–9979.
<https://doi.org/10.1002/anie.201303370>.
- (20) Nakatani, K.; Hagihara, S.; Goto, Y.; Kobori, A.; Hagihara, M.; Hayashi, G.; Kyo, M.; Nomura, M.; Mishima, M.; Kojima, C. Small-Molecule Ligand Induces Nucleotide Flipping in (CAG)_n Trinucleotide Repeats. *Nat Chem Biol* **2005**, *1* (1), 39–43. <https://doi.org/10.1038/nchembio708>.

- (21) Das, B.; Nagano, K.; Kawai, G.; Murata, A.; Nakatani, K. 2-Amino-1,8-Naphthyridine Dimer (ANP77), a High-Affinity Binder to the Internal Loops of C/CC and T/CC Sites in Double-Stranded DNA. *J Org Chem* **2022**, *87* (1), 340–350. <https://doi.org/10.1021/acs.joc.1c02383>.
- (22) Shibata, T.; Nagano, K.; Ueyama, M.; Ninomiya, K.; Hirose, T.; Nagai, Y.; Ishikawa, K.; Kawai, G.; Nakatani, K. Small Molecule Targeting r(UGGAA)n Disrupts RNA Foci and Alleviates Disease Phenotype in Drosophila Model. *Nat Commun* **2021**, *12* (1), 236. <https://doi.org/10.1038/s41467-020-20487-4>.
- (23) Murata, A.; Fukuzumi, T.; Umemoto, S.; Nakatani, K. Xanthone Derivatives as Potential Inhibitors of MiRNA Processing by Human Dicer: Targeting Secondary Structures of Pre-MiRNA by Small Molecules. *Bioorg Med Chem Lett* **2013**, *23* (1), 252–255. <https://doi.org/10.1016/j.bmcl.2012.10.108>.
- (24) Diaz, J. P.; Chirayil, R.; Chirayil, S.; Tom, M.; Head, K. J.; Luebke, K. J. Association of a Peptoid Ligand with the Apical Loop of Pri-MiR-21 Inhibits Cleavage by Drosha. *RNA* **2014**, *20* (4), 528–539. <https://doi.org/10.1261/rna.042911.113>.
- (25) Bose, D.; Nahar, S.; Rai, M. K.; Ray, A.; Chakraborty, K.; Maiti, S. Selective Inhibition of MiR-21 by Phage Display Screened Peptide. *Nucleic Acids Res* **2015**, *43* (8), 4342–4352. <https://doi.org/10.1093/nar/gkv185>.
- (26) Velagapudi, S. P.; Gallo, S. M.; Disney, M. D. Sequence-Based Design of Bioactive Small Molecules That Target Precursor MicroRNAs. *Nat Chem Biol* **2014**, *10* (4), 291–297. <https://doi.org/10.1038/nchembio.1452>.
- (27) Periwai, V.; Scaria, V. Machine Learning Approaches toward Building Predictive Models for Small Molecule Modulators of MiRNA and Its Utility in Virtual Screening of Molecular Databases. *Methods in Molecular Biology* **2017**, *1517*, 155–168. https://doi.org/10.1007/978-1-4939-6563-2_11.
- (28) Liu, S.; Yang, Y.; Li, W.; Tian, X.; Cui, H.; Zhang, Q. Identification of Small-Molecule Ligands That Bind to MiR-21 as Potential Therapeutics for Endometriosis by Screening ZINC Database and in-Vitro Assays. *Gene* **2018**, *662*, 46–53. <https://doi.org/10.1016/j.gene.2018.03.094>.
- (29) Maiti, M.; Nauwelaerts, K.; Herdewijn, P. Pre-MicroRNA Binding Aminoglycosides and Antitumor Drugs as Inhibitors of Dicer Catalyzed MicroRNA Processing. *Bioorg Med Chem Lett* **2012**, *22* (4), 1709–1711. <https://doi.org/10.1016/j.bmcl.2011.12.103>.

- (30) Nahar, S.; Ranjan, N.; Ray, A.; Arya, D. P.; Maiti, S. Potent Inhibition of MiR-27a by Neomycin-Bisbenzimidazole Conjugates. *Chem Sci* **2015**, *6* (10), 5837–5846. <https://doi.org/10.1039/c5sc01969a>.
- (31) Vo, D. D.; Tran, T. P. A.; Staedel, C.; Benhida, R.; Darfeuille, F.; Di Giorgio, A.; Duca, M. Oncogenic MicroRNAs Biogenesis as a Drug Target: Structure-Activity Relationship Studies on New Aminoglycoside Conjugates. *Chemistry - A European Journal* **2016**, *22* (15), 5350–5362. <https://doi.org/10.1002/chem.201505094>.
- (32) Childs-Disney, J. L.; Disney, M. D. Small Molecule Targeting of a MicroRNA Associated with Hepatocellular Carcinoma. *ACS Chem Biol* **2016**, *11* (2), 375–380. <https://doi.org/10.1021/acscchembio.5b00615>.
- (33) Murata, A.; Otabe, T.; Zhang, J.; Nakatani, K. BzDANP, a Small-Molecule Modulator of Pre-MiR-29a Maturation by Dicer. *ACS Chem Biol* **2016**, *11* (10), 2790–2796. <https://doi.org/10.1021/acscchembio.6b00214>.
- (34) Otabe, T.; Nagano, K.; Kawai, G.; Murata, A.; Nakatani, K. Inhibition of Pre-MiRNA-136 Processing by Dicer with Small Molecule BzDANP Suggested the Formation of Ternary Complex of Pre-MiR-136–BzDANP–Dicer. *Bioorg Med Chem* **2019**, *27* (10), 2140–2148. <https://doi.org/10.1016/j.bmc.2019.03.031>.
- (35) Morgan, B. S.; Sanaba, B. G.; Donlic, A.; Karloff, D. B.; Forte, J. E.; Zhang, Y.; Hargrove, A. E. R-BIND: An Interactive Database for Exploring and Developing RNA-Targeted Chemical Probes. *ACS Chem Biol* **2019**, *14* (12), 2691–2700. <https://doi.org/10.1021/acscchembio.9b00631>.
- (36) Mardis, E. R. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry* **2013**, *6* (1), 287–303. <https://doi.org/10.1146/annurev-anchem-062012-092628>.
- (37) Beedanagari, S.; John, K. Next Generation Sequencing. In *Encyclopedia of Toxicology (Third Edition)*; Wexler, P., Ed.; Academic Press: Oxford, 2014; pp 501–503. <https://doi.org/https://doi.org/10.1016/B978-0-12-386454-3.00678-3>.
- (38) Sanger, F. SEQUENCES, SEQUENCES, AND SEQUENCES. *Annu Rev Biochem* **1988**, *57* (1), 1–29. <https://doi.org/10.1146/annurev.bi.57.070188.000245>.
- (39) Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, J. C.; Hutchison, C. A.; Slocombe, P. M.; Smith, M. Nucleotide Sequence of Bacteriophage ΦX174 DNA. *Nature* **1977**, *265* (5596), 687–695. <https://doi.org/10.1038/265687a0>.

- (40) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences* **1977**, *74* (12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>.
- (41) Ju, J.; Glazer, A. N.; Mathies, R. A. Energy Transfer Primers: A New Fluorescence Labeling Paradigm for DNA Sequencing and Analysis. *Nat Med* **1996**, *2* (2), 246–249. <https://doi.org/10.1038/nm0296-246>.
- (42) Shendure, J. A.; Porreca, G. J.; Church, G. M.; Gardner, A. F.; Hendrickson, C. L.; Kieleczawa, J.; Slatko, B. E. Overview of DNA Sequencing Strategies. *Curr Protoc Mol Biol* **2011**, *96* (1), 7.1.1-7.1.23. <https://doi.org/https://doi.org/10.1002/0471142727.mb0701s96>.
- (43) Schuster, S. C. Next-Generation Sequencing Transforms Today's Biology. *Nat Methods* **2008**, *5* (1), 16–18. <https://doi.org/10.1038/nmeth1156>.
- (44) Churko, J. M.; Mantalas, G. L.; Snyder, M. P.; Wu, J. C. Overview of High Throughput Sequencing Technologies to Elucidate Molecular Pathways in Cardiovascular Diseases. *Circ Res* **2013**, *112* (12), 1613–1623. <https://doi.org/10.1161/CIRCRESAHA.113.300939>.
- (45) Smith, M. G.; Gianoulis, T. A.; Pukatzki, S.; Mekalanos, J. J.; Ornston, L. N.; Gerstein, M.; Snyder, M. New Insights into *Acinetobacter Baumannii* Pathogenesis Revealed by High-Density Pyrosequencing and Transposon Mutagenesis. *Genes Dev* **2007**, *21* (5), 601–614. <https://doi.org/10.1101/gad.1510307>.
- (46) Nielsen, R.; Paul, J. S.; Albrechtsen, A.; Song, Y. S. Genotype and SNP Calling from Next-Generation Sequencing Data. *Nat Rev Genet* **2011**, *12* (6), 443–451. <https://doi.org/10.1038/nrg2986>.
- (47) Logares, R.; Haverkamp, T. H. A.; Kumar, S.; Lanzén, A.; Nederbragt, A. J.; Quince, C.; Kausserud, H. Environmental Microbiology through the Lens of High-Throughput DNA Sequencing: Synopsis of Current Platforms and Bioinformatics Approaches. *J Microbiol Methods* **2012**, *91* (1), 106–113. <https://doi.org/10.1016/j.mimet.2012.07.017>.
- (48) Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B. B.; Siddiqui, A.; Lao, K.; Surani, M. A. mRNA-Seq Whole-Transcriptome Analysis of a Single Cell. *Nat Methods* **2009**, *6* (5), 377–382. <https://doi.org/10.1038/nmeth.1315>.
- (49) Mutz, K. O.; Heilkenbrinker, A.; Lönne, M.; Walter, J. G.; Stahl, F. Transcriptome Analysis Using Next-Generation Sequencing. *Curr Opin Biotechnol* **2013**, *24* (1), 22–30. <https://doi.org/10.1016/j.copbio.2012.09.004>.

- (50) Leblanc, V. G.; Marra, M. A. Next-Generation Sequencing Approaches in Cancer: Where Have They Brought Us and Wherewill They Take Us? *Cancers (Basel)* **2015**, *7* (3), 1925–1958. <https://doi.org/10.3390/cancers7030869>.
- (51) Jimenez, J. I.; Xulvi-Brunet, R.; Campbell, G. W.; Turk-MacLeod, R.; Chen, I. A. Comprehensive Experimental Fitness Landscape and Evolutionary Network for Small RNA. *Proceedings of the National Academy of Sciences* **2013**, *110* (37), 14984–14989. <https://doi.org/10.1073/pnas.1307604110>.
- (52) Kobori, S.; Takahashi, K.; Yokobayashi, Y. Deep Sequencing Analysis of Aptazyme Variants Based on a Pistol Ribozyme. *ACS Synth Biol* **2017**, *6* (7), 1283–1288. <https://doi.org/10.1021/acssynbio.7b00057>.
- (53) Kobori, S.; Nomura, Y.; Miu, A.; Yokobayashi, Y. High-Throughput Assay and Engineering of Self-Cleaving Ribozymes by Sequencing. *Nucleic Acids Res* **2015**, *43* (13), e85–e85. <https://doi.org/10.1093/nar/gkv265>.
- (54) Velagapudi, S. P.; Seedhouse, S. J.; French, J.; Disney, M. D. Defining the RNA Internal Loops Preferred by Benzimidazole Derivatives via 2D Combinatorial Screening and Computational Analysis. *J Am Chem Soc* **2011**, *133* (26), 10111–10118. <https://doi.org/10.1021/ja200212b>.
- (55) Lee, R. C.; Feinbaum, R. L.; Ambros, V. The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs with Antisense Complementarity to Lin-14. *Cell* **1993**, *75* (5), 843–854. [https://doi.org/10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y).
- (56) Carthew, R. W.; Sontheimer, E. J. Origins and Mechanisms of MiRNAs and SiRNAs. *Cell* **2009**, *136* (4), 642–655. <https://doi.org/10.1016/j.cell.2009.01.035>.
- (57) Cai, Y.; Yu, X.; Hu, S.; Yu, J. A Brief Review on the Mechanisms of MiRNA Regulation. *Genomics Proteomics Bioinformatics* **2009**, *7* (4), 147–154. [https://doi.org/10.1016/S1672-0229\(08\)60044-3](https://doi.org/10.1016/S1672-0229(08)60044-3).
- (58) Ma, C.; Liu, Y.; He, L. MicroRNAs - Powerful Repression Comes from Small RNAs. *Sci China C Life Sci* **2009**, *52* (4), 323–330. <https://doi.org/10.1007/s11427-009-0056-x>.
- (59) Davis-Dusenbery, B. N.; Hata, A. Mechanisms of Control of MicroRNA Biogenesis. *J Biochem* **2010**, *148* (4), 381–392. <https://doi.org/10.1093/jb/mvq096>.
- (60) Ha, M.; Kim, V. N. Regulation of MicroRNA Biogenesis. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group July 16, 2014, pp 509–524. <https://doi.org/10.1038/nrm3838>.

- (61) Guo, R.; Lin, W.; Zhang, J.; Simon, A. E.; Kushner, D. B. Structural Plasticity and Rapid Evolution in a Viral RNA Revealed by in Vivo Genetic Selection. *J Virol* **2009**, *83* (2), 927–939. <https://doi.org/10.1128/JVI.02060-08>.
- (62) Hashimoto, Y.; Akiyama, Y.; Yuasa, Y. Multiple-to-Multiple Relationships between MicroRNAs and Target Genes in Gastric Cancer. *PLoS One* **2013**, *8* (5). <https://doi.org/10.1371/journal.pone.0062589>.
- (63) Zhang, B.; Farwell, M. A. MicroRNAs: A New Emerging Class of Players for Disease Diagnostics and Gene Therapy. *J Cell Mol Med* **2008**, *12* (1), 3–21. <https://doi.org/https://doi.org/10.1111/j.1582-4934.2007.00196.x>.
- (64) Murata, A.; Otabe, T.; Zhang, J.; Nakatani, K. BzDANP, a Small-Molecule Modulator of Pre-MiR-29a Maturation by Dicer. *ACS Chem Biol* **2016**, *11* (10), 2790–2796. <https://doi.org/10.1021/acscchembio.6b00214>.
- (65) Childs-Disney, J. L.; Disney, M. D. Small Molecule Targeting of a MicroRNA Associated with Hepatocellular Carcinoma. *ACS Chem Biol* **2016**, *11* (2), 375–380. <https://doi.org/10.1021/acscchembio.5b00615>.
- (66) Starega-Roslan, J.; Galka-Marciniak, P.; Krzyzosiak, W. J. Nucleotide Sequence of MiRNA Precursor Contributes to Cleavage Site Selection by Dicer. *Nucleic Acids Res* **2015**, *43* (22), 10939–10951. <https://doi.org/10.1093/nar/gkv968>.
- (67) Dunn, O. J. Multiple Comparisons Among Means. *J Am Stat Assoc* **1961**, *56* (293), 52–64. <https://doi.org/10.2307/2282330>.
- (68) Nakatani, K.; He, H.; Uno, S.; Yamamoto, T.; Dohno, C. Synthesis of Dimeric 2-Amino-1,8-Naphthyridine and Related DNA-Binding Molecules. *Curr Protoc Nucleic Acid Chem* **2008**, *32* (1), 8.6.1–8.6.21. <https://doi.org/https://doi.org/10.1002/0471142700.nc0806s32>.
- (69) Lorenz, R.; Bernhart, S. H.; Höner zu Siederdisen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **2011**, *6* (1), 1–14. <https://doi.org/10.1186/1748-7188-6-26/TABLES/2>.
- (70) Murata, A.; Otabe, T.; Zhang, J.; Nakatani, K. BzDANP, a Small-Molecule Modulator of Pre-MiR-29a Maturation by Dicer. *ACS Chem Biol* **2016**, *11* (10), 2790–2796. <https://doi.org/10.1021/acscchembio.6b00214>.
- (71) Murata, A.; Mori, Y.; Di, Y.; Sugai, A.; Das, B.; Takashima, Y.; Nakatani, K. Small Molecule-Induced Dimerization of Hairpin RNA Interfered with the Dicer Cleavage Reaction. *Biochemistry* **2021**, *60* (4), 245–249. <https://doi.org/10.1021/acs.biochem.0c00920>.

- (72) Rupaimoole, R.; Slack, F. J. MicroRNA Therapeutics: Towards a New Era for the Management of Cancer and Other Diseases. *Nat Rev Drug Discov* **2017**, *16* (3), 203–222.
- (73) Zhou, J.; Rossi, J. Aptamers as Targeted Therapeutics: Current Potential and Challenges. *Nat Rev Drug Discov* **2017**, *16* (3), 181–202.
- (74) Falese, J. P.; Donlic, A.; Hargrove, A. E. Targeting RNA with Small Molecules: From Fundamental Principles towards the Clinic. *Chem Soc Rev* **2021**, *50* (4), 2224–2243.
- (75) Ursu, A.; Childs-Disney, J. L.; Andrews, R. J.; O’Leary, C. A.; Meyer, S. M.; Angelbello, A. J.; Moss, W. N.; Disney, M. D. Design of Small Molecules Targeting RNA Structure from Sequence. *Chem Soc Rev* **2020**, *49* (20), 7252–7270.
- (76) Khvorova, A.; Watts, J. K. The Chemical Evolution of Oligonucleotide Therapies of Clinical Utility. *Nat Biotechnol* **2017**, *35* (3), 238–248.
- (77) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proceedings of the National Academy of Sciences* **1996**, *93* (24), 13770–13773.
- (78) Braha, O.; Walker, B.; Cheley, S.; Kasianowicz, J. J.; Song, L.; Gouaux, J. E.; Bayley, H. Designed Protein Pores as Components for Biosensors. *Chem Biol* **1997**, *4* (7), 497–505.
- (79) Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. Rapid Nanopore Discrimination between Single Polynucleotide Molecules. *Proceedings of the National Academy of Sciences* **2000**, *97* (3), 1079–1084.
- (80) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. The Potential and Challenges of Nanopore Sequencing. *Nat Biotechnol* **2008**, *26* (10), 1146–1153.
- (81) Kang, X.; Gu, L.; Cheley, S.; Bayley, H. Single Protein Pores Containing Molecular Adapters at High Temperatures. *Angewandte Chemie International Edition* **2005**, *44* (10), 1495–1499.
- (82) Bayley, H.; Cremer, P. S. Stochastic Sensors Inspired by Biology. *Nature* **2001**, *413* (6852), 226–230.
- (83) Clarke, J.; Wu, H.-C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing. *Nat Nanotechnol* **2009**, *4* (4), 265–270.
- (84) Howorka, S.; Siwy, Z. Nanopore Analytics: Sensing of Single Molecules. *Chem Soc Rev* **2009**, *38* (8), 2360–2384.

- (85) Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A. Ion-Beam Sculpting at Nanometre Length Scales. *Nature* **2001**, *412* (6843), 166–169.
- (86) Dekker, C. Solid-State Nanopores. *Nat Nanotechnol* **2007**, *2* (4), 209–215.
- (87) Di Ventra, M.; Taniguchi, M. Decoding DNA, RNA and Peptides with Quantum Tunnelling. *Nat Nanotechnol* **2016**, *11* (2), 117.
- (88) Su, T. A.; Neupane, M.; Steigerwald, M. L.; Venkataraman, L.; Nuckolls, C. Chemical Principles of Single-Molecule Electronics. *Nat Rev Mater* **2016**, *1* (3), 1–15.
- (89) Aradhya, S. V.; Venkataraman, L. Single-Molecule Junctions beyond Electronic Transport. *Nat Nanotechnol* **2013**, *8* (6), 399–410.
- (90) Xin, N.; Guan, J.; Zhou, C.; Chen, X.; Gu, C.; Li, Y.; Ratner, M. A.; Nitzan, A.; Stoddart, J. F.; Guo, X. Concepts in the Design and Engineering of Single-Molecule Electronic Devices. *Nature Reviews Physics* **2019**, *1* (3), 211–230.
- (91) Xie, X.; Li, P.; Xu, Y.; Zhou, L.; Yan, Y.; Xie, L.; Jia, C.; Guo, X. Single-Molecule Junction: A Reliable Platform for Monitoring Molecular Physical and Chemical Processes. *ACS Nano* **2022**, *16* (3), 3476–3505.
- (92) Komoto, Y.; Fujii, S.; Iwane, M.; Kiguchi, M. Single-Molecule Junctions for Molecular Electronics. *J Mater Chem C Mater* **2016**, *4* (38), 8842–8858.
- (93) Foox, J.; Tighe, S. W.; Nicolet, C. M.; Zook, J. M.; Byrsk-Bishop, M.; Clarke, W. E.; Khayat, M. M.; Mahmoud, M.; Laaguiby, P. K.; Herbert, Z. T. Performance Assessment of DNA Sequencing Platforms in the ABRF Next-Generation Sequencing Study. *Nat Biotechnol* **2021**, *39* (9), 1129–1140.
- (94) Kovaka, S.; Fan, Y.; Ni, B.; Timp, W.; Schatz, M. C. Targeted Nanopore Sequencing by Real-Time Mapping of Raw Electrical Signal with UNCALLED. *Nat Biotechnol* **2021**, *39* (4), 431–441.
- (95) Gilpatrick, T.; Lee, I.; Graham, J. E.; Raimondeau, E.; Bowen, R.; Heron, A.; Downs, B.; Sukumar, S.; Sedlazeck, F. J.; Timp, W. Targeted Nanopore Sequencing with Cas9-Guided Adapter Ligation. *Nat Biotechnol* **2020**, *38* (4), 433–438.
- (96) Jain, M.; Koren, S.; Miga, K. H.; Quick, J.; Rand, A. C.; Sasani, T. A.; Tyson, J. R.; Beggs, A. D.; Dilthey, A. T.; Fiddes, I. T. Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads. *Nat Biotechnol* **2018**, *36* (4), 338–345.
- (97) Ashton, P. M.; Nair, S.; Dallman, T.; Rubino, S.; Rabsch, W.; Mwaigwisya, S.; Wain, J.; O'grady, J. MinION Nanopore Sequencing Identifies the Position and

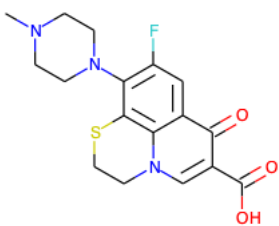
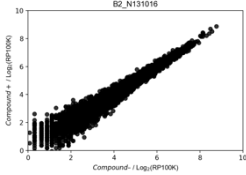
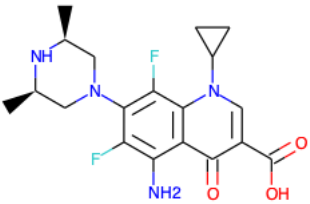
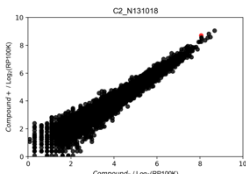
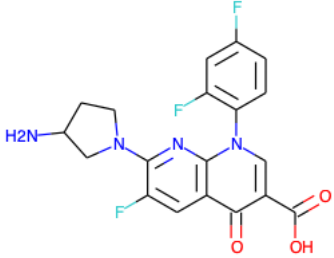
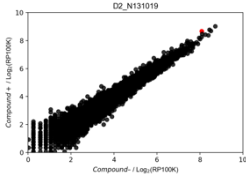
- Structure of a Bacterial Antibiotic Resistance Island. *Nat Biotechnol* **2015**, *33* (3), 296–300.
- (98) Restrepo-Pérez, L.; Joo, C.; Dekker, C. Paving the Way to Single-Molecule Protein Sequencing. *Nat Nanotechnol* **2018**, *13* (9), 786–796.
- (99) Ouldali, H.; Sarthak, K.; Ensslen, T.; Piguet, F.; Manivet, P.; Pelta, J.; Behrends, J. C.; Aksimentiev, A.; Oukhaled, A. Electrical Recognition of the Twenty Proteinogenic Amino Acids Using an Aerolysin Nanopore. *Nat Biotechnol* **2020**, *38* (2), 176–181.
- (100) Alfaro, J. A.; Bohländer, P.; Dai, M.; Filius, M.; Howard, C. J.; van Kooten, X. F.; Ohayon, S.; Pomorski, A.; Schmid, S.; Aksimentiev, A. The Emerging Landscape of Single-Molecule Protein Sequencing Technologies. *Nat Methods* **2021**, *18* (6), 604–617.
- (101) Brinkerhoff, H.; Kang, A. S. W.; Liu, J.; Aksimentiev, A.; Dekker, C. Multiple Rereads of Single Proteins at Single–Amino Acid Resolution Using Nanopores. *Science (1979)* **2021**, *374* (6574), 1509–1513.
- (102) Ohshiro, T.; Tsutsui, M.; Yokota, K.; Taniguchi, M. Quantitative Analysis of DNA with Single-Molecule Sequencing. *Sci Rep* **2018**, *8* (1), 1–8.
- (103) Ohshiro, T.; Matsubara, K.; Tsutsui, M.; Furuhashi, M.; Taniguchi, M.; Kawai, T. Single-Molecule Electrical Random Resequencing of DNA and RNA. *Sci Rep* **2012**, *2*, 501.
- (104) Ohshiro, T.; Tsutsui, M.; Yokota, K.; Furuhashi, M.; Taniguchi, M.; Kawai, T. Detection of Post-Translational Modifications in Single Peptides Using Electron Tunnelling Currents. *Nat Nanotechnol* **2014**, *9* (10), 835–840.
- (105) Tsutsui, M.; Matsubara, K.; Ohshiro, T.; Furuhashi, M.; Taniguchi, M.; Kawai, T. Electrical Detection of Single Methylcytosines in a DNA Oligomer. *J Am Chem Soc* **2011**, *133* (23), 9124–9128.
- (106) Ohshiro, T.; Konno, M.; Asai, A.; Komoto, Y.; Yamagata, A.; Doki, Y.; Eguchi, H.; Ofusa, K.; Taniguchi, M.; Ishii, H. Single-Molecule RNA Sequencing for Simultaneous Detection of M6A and 5mC. **2021**.
- (107) Komoto, Y.; Ohshiro, T.; Taniguchi, M. Detection of Alcohol-Associated Cancer Marker by Single-Molecule Quantum Sequencing. *Chemical Communications* **2020**.
- (108) Zhao, Y.; Ashcroft, B.; Zhang, P.; Liu, H.; Sen, S.; Song, W.; Im, J.; Gyarfas, B.; Manna, S.; Biswas, S.; Borges, C.; Lindsay, S. Single-Molecule Spectroscopy of Amino Acids and Peptides by Recognition Tunnelling. *Nat Nanotechnol* **2014**, *9* (6), 466–473. <https://doi.org/10.1038/nnano.2014.54>.

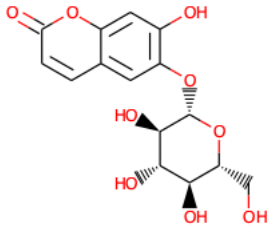
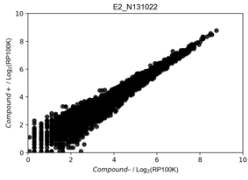
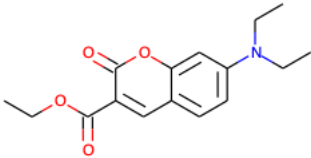
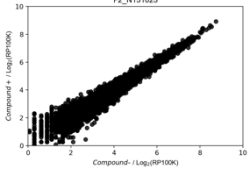
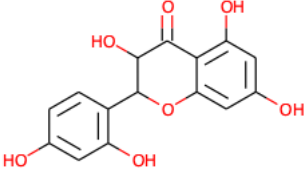
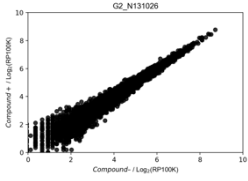
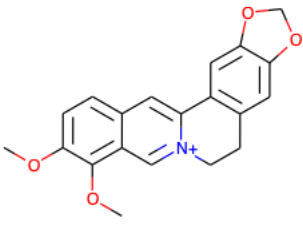
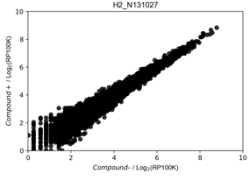
- (109) Chang, S.; Huang, S.; He, J.; Liang, F.; Zhang, P.; Li, S.; Chen, X.; Sankey, O.; Lindsay, S. Electronic Signatures of All Four DNA Nucleosides in a Tunneling Gap. *Nano Lett* **2010**, *10* (3), 1070–1075. <https://doi.org/10.1021/nl1001185>.
- (110) Jain, M.; Fiddes, I. T.; Miga, K. H.; Olsen, H. E.; Paten, B.; Akeson, M. Improved Data Analysis for the MinION Nanopore Sequencer. *Nat Methods* **2015**, *12* (4), 351–356.
- (111) Taniguchi, M.; Minami, S.; Ono, C.; Hamajima, R.; Morimura, A.; Hamaguchi, S.; Akeda, Y.; Kanai, Y.; Kobayashi, T.; Kamitani, W. Combining Machine Learning and Nanopore Construction Creates an Artificial Intelligence Nanopore for Coronavirus Detection. *Nat Commun* **2021**, *12* (1), 1–8.
- (112) Taniguchi, M.; Ohshiro, T.; Komoto, Y.; Takaai, T.; Yoshida, T.; Washio, T. High-Precision Single-Molecule Identification Based on Single-Molecule Information within a Noisy Matrix. *Journal of Physical Chemistry C* **2019**, *123*, 15867–15873. <https://doi.org/10.1021/acs.jpcc.9b03908>.
- (113) Cui, F.; Yue, Y.; Zhang, Y.; Zhang, Z.; Zhou, H. S. Advancing Biosensors with Machine Learning. *ACS Sens* **2020**, *5* (11), 3346–3364.
- (114) Wen, C.; Dematties, D.; Zhang, S.-L. A Guide to Signal Processing Algorithms for Nanopore Sensors. *ACS Sens* **2021**, *6* (10), 3536–3555.
- (115) Nakatani, K.; Natsuhara, N.; Mori, Y.; Mukherjee, S.; Das, B.; Murata, A. Synthesis of Naphthyridine Dimers with Conformational Restriction and Binding to DNA and RNA. *Chemistry—An Asian Journal* **2017**, *12* (23), 3077–3087.
- (116) Murray, T. J.; Zimmerman, S. C. New Triply Hydrogen Bonded Complexes with Highly Variable Stabilities. *J Am Chem Soc* **1992**, *114* (10), 4010–4011.
- (117) Chen, F.; Tao, N. J. Electron Transport in Single Molecules: From Benzene to Graphene. *Acc Chem Res* **2009**, *42* (3), 429–438.
- (118) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. Gaussian 09 Revision A. 1, 2009, Gaussian Inc. *Wallingford CT* **2009**, *66*, 219.
- (119) Nishino, T.; Hayashi, N.; Bui, P. T. Direct Measurement of Electron Transfer through a Hydrogen Bond between Single Molecules. *J Am Chem Soc* **2013**, *135* (12), 4592–4595.
- (120) Zhou, C.; Li, X.; Gong, Z.; Jia, C.; Lin, Y.; Gu, C.; He, G.; Zhong, Y.; Yang, J.; Guo, X. Direct Observation of Single-Molecule Hydrogen-Bond Dynamics with Single-Bond Resolution. *Nat Commun* **2018**, *9* (1), 1–9.
- (121) Hansen, N.; Van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J Chem Theory Comput* **2014**, *10* (7), 2632–2647.

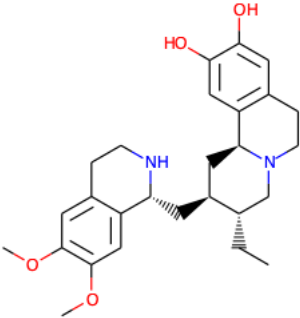
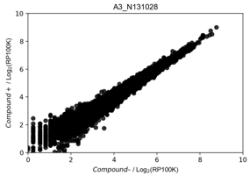
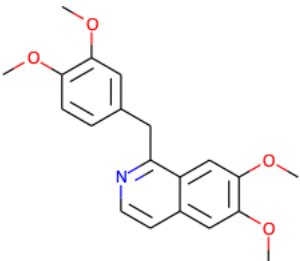
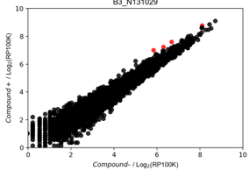
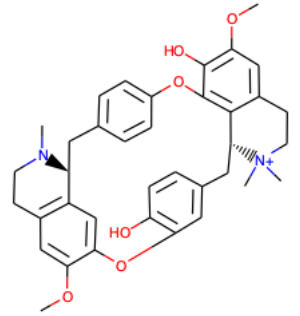
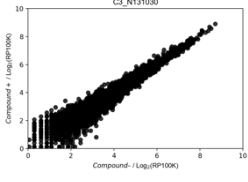
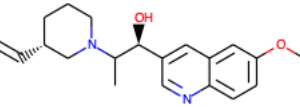
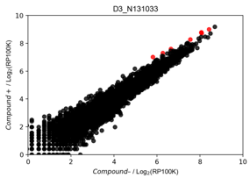
- (122) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J Chem Theory Comput* **2009**, *5* (2), 350–358.
- (123) Paluch, A. S.; Mobley, D. L.; Maginn, E. J. Small Molecule Solvation Free Energy: Enhanced Conformational Sampling Using Expanded Ensemble Molecular Dynamics Simulation. *J Chem Theory Comput* **2011**, *7* (9), 2910–2918.
- (124) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *the Journal of machine Learning research* **2011**, *12*, 2825–2830.
- (125) McLachlan, G. J.; Lee, S. X.; Rathnayake, S. I. Finite Mixture Models. *Annu Rev Stat Appl* **2019**, *6*, 355–378.
- (126) Sessler, J. L.; Lawrence, C. M.; Jayawickramarajah, J. Molecular Recognition via Base-Pairing. *Chem Soc Rev* **2007**, *36* (2), 314–325.
- (127) Nakatani, K.; Sando, S.; Saito, I. Scanning of Guanine–Guanine Mismatches in DNA by Synthetic Ligands Using Surface Plasmon Resonance. *Nat Biotechnol* **2001**, *19* (1), 51–55. <https://doi.org/10.1038/83505>.
- (128) Agrait, N.; Yeyati, A. L.; Van Ruitenbeek, J. M. Quantum Properties of Atomic-Sized Conductors. *Phys Rep* **2003**, *377* (2–3), 81–279.
- (129) Elkan, C.; Noto, K. Learning Classifiers from Only Positive and Unlabeled Data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*; 2008; pp 213–220.

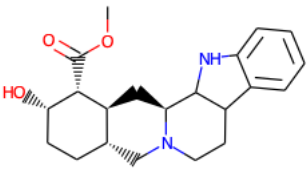
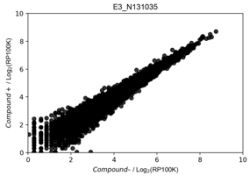
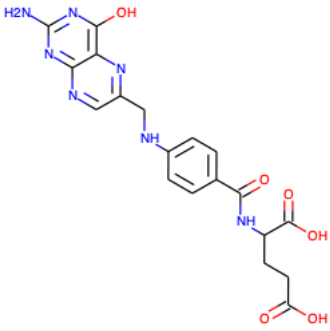
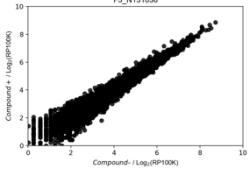
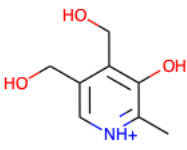
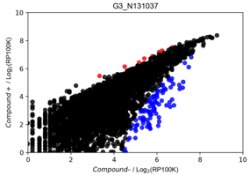
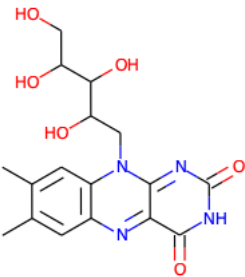
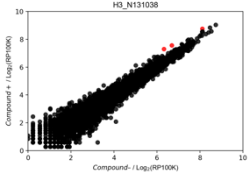
補足資料

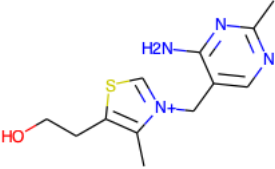
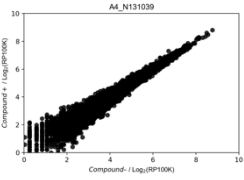
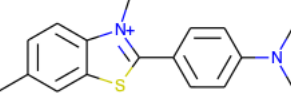
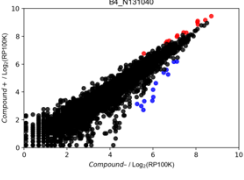
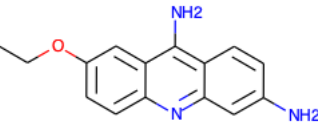
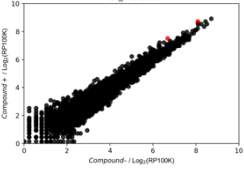
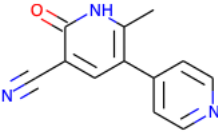
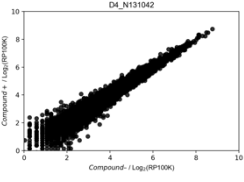
S1. 79 Compound in-house Library

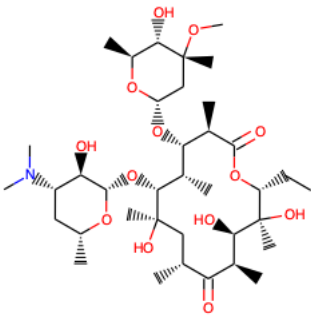
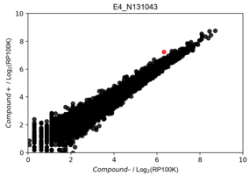
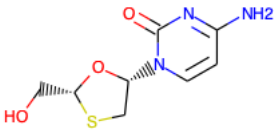
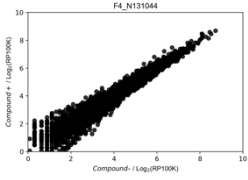
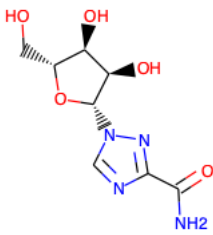
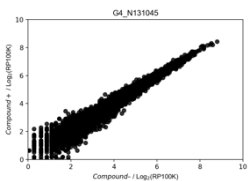
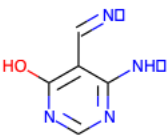
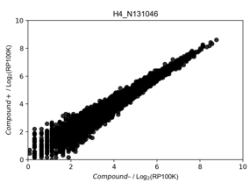
vertical	horizontal	ID	structure	result
A	2	None		
B	2	N131016		
C	2	N131018		
D	2	N131019		

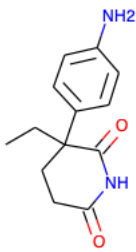
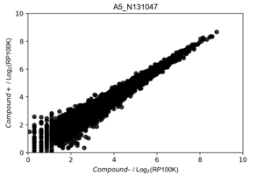
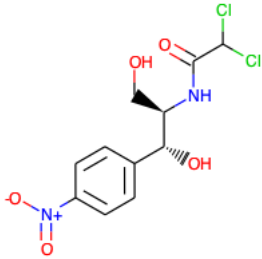
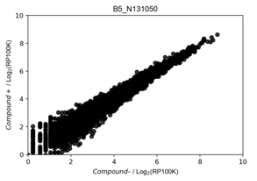
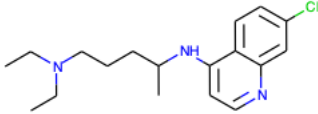
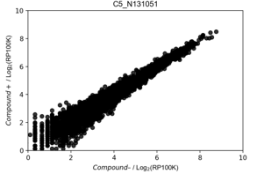
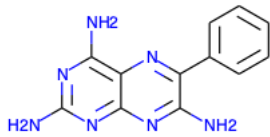
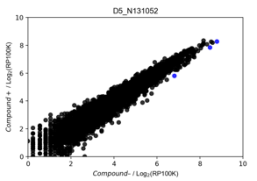
E	2	N131022		
F	2	N131023		
G	2	N131026		
H	2	N131027		

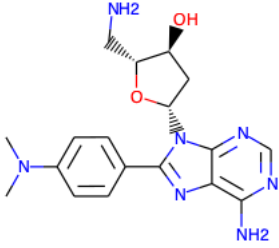
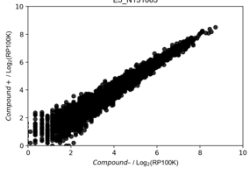
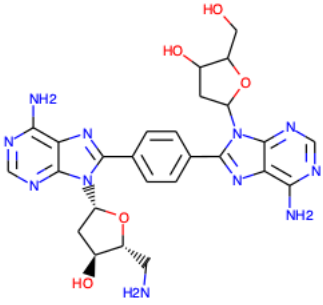
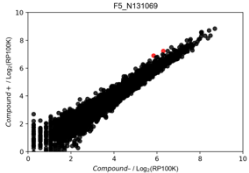
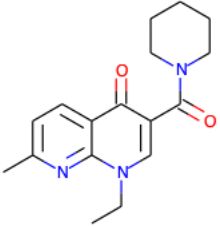
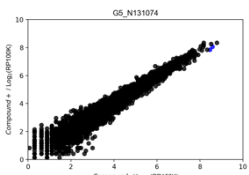
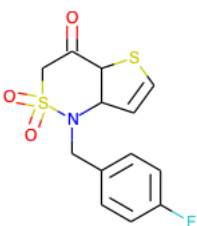
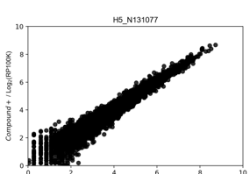
A	3	N131028	 <p>Chemical structure of N131028, a complex molecule featuring a piperazine ring system, a benzene ring with two methoxy groups, and a quinoline ring system with two hydroxyl groups.</p>	 <p>Scatter plot for N131028 showing the relationship between $\text{Log}_2(\text{RP100K})$ (x-axis) and $\text{Log}_2(\text{RP100K})$ (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line. The title of the plot is A3_N131028.</p>
B	3	N131029	 <p>Chemical structure of N131029, a molecule consisting of a quinoline ring system with two methoxy groups and a benzene ring with two methoxy groups.</p>	 <p>Scatter plot for N131029 showing the relationship between $\text{Log}_2(\text{RP100K})$ (x-axis) and $\text{Log}_2(\text{RP100K})$ (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line. The title of the plot is B3_N131029.</p>
C	3	N131030	 <p>Chemical structure of N131030, a complex molecule featuring a piperazine ring system, a benzene ring with two methoxy groups, and a quinoline ring system with two hydroxyl groups.</p>	 <p>Scatter plot for N131030 showing the relationship between $\text{Log}_2(\text{RP100K})$ (x-axis) and $\text{Log}_2(\text{RP100K})$ (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line. The title of the plot is C3_N131030.</p>
D	3	N131033	 <p>Chemical structure of N131033, a molecule consisting of a piperazine ring system, a benzene ring with two methoxy groups, and a quinoline ring system with two hydroxyl groups.</p>	 <p>Scatter plot for N131033 showing the relationship between $\text{Log}_2(\text{RP100K})$ (x-axis) and $\text{Log}_2(\text{RP100K})$ (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line. The title of the plot is D3_N131033.</p>

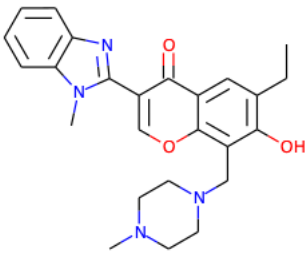
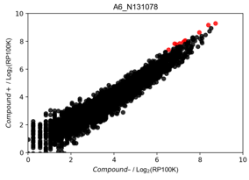
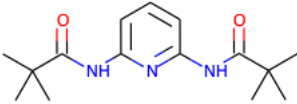
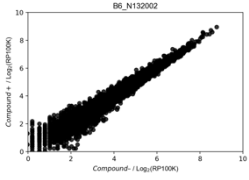
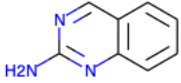
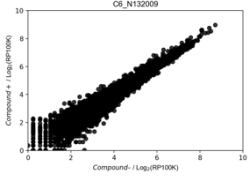
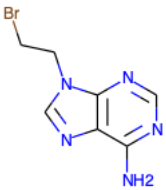
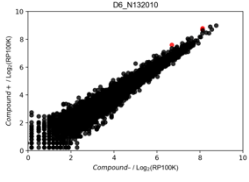
E	3	N131035	 <p>Chemical structure of N131035, a complex bicyclic molecule with a piperidine ring, a benzene ring, and a methyl ester group.</p>	 <p>Scatter plot for N131035 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>
F	3	N131036	 <p>Chemical structure of N131036, a molecule featuring a pyrimidine ring system, a benzene ring, and a carboxylic acid group.</p>	 <p>Scatter plot for N131036 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>
G	3	N131037	 <p>Chemical structure of N131037, a molecule featuring a pyrimidine ring system, a benzene ring, and a hydroxyl group.</p>	 <p>Scatter plot for N131037 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>
H	3	N131038	 <p>Chemical structure of N131038, a molecule featuring a pyrimidine ring system, a benzene ring, and a hydroxyl group.</p>	 <p>Scatter plot for N131038 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>

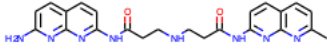
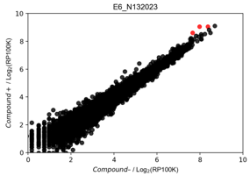
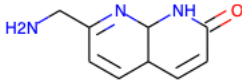
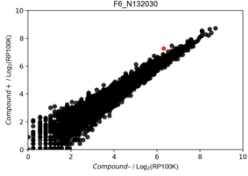
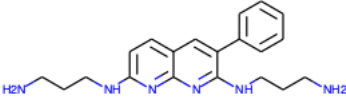
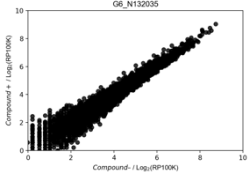
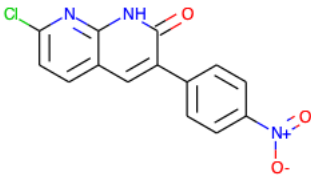
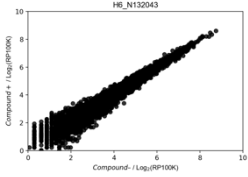
A	4	N131039		
B	4	N131040		
C	4	N131041		
D	4	N131042		

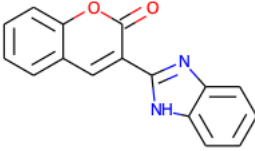
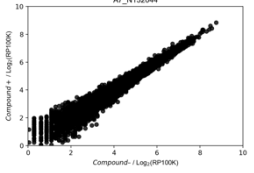
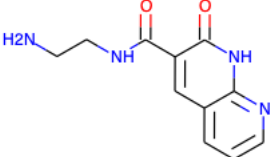
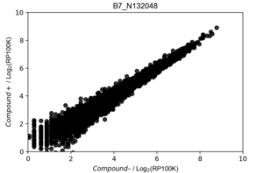
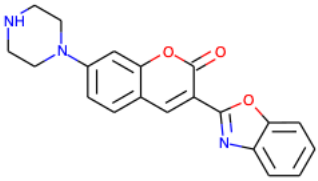
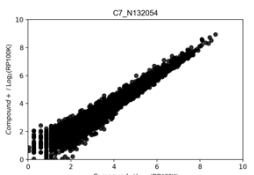
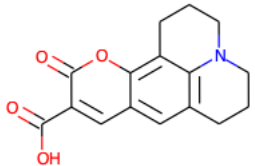
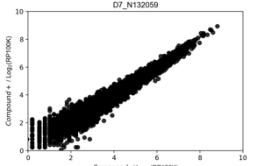
E	4	N131043	 <p>Chemical structure of N131043, a complex polycyclic molecule with multiple hydroxyl groups and a methylamino group.</p>	 <p>Scatter plot for N131043 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>
F	4	N131044	 <p>Chemical structure of N131044, a molecule featuring a thiazolidine ring system and a pyridine ring with an amino group.</p>	 <p>Scatter plot for N131044 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>
G	4	N131045	 <p>Chemical structure of N131045, a molecule featuring a thiazolidine ring system and a pyridine ring with an amino group and a carbonyl group.</p>	 <p>Scatter plot for N131045 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>
H	4	N131046	 <p>Chemical structure of N131046, a molecule featuring a pyridine ring with a hydroxyl group and a nitro group.</p>	 <p>Scatter plot for N131046 showing the relationship between Compound- / Log₁₀(RP100k) (x-axis) and Compound+ / Log₁₀(RP100k) (y-axis). The plot shows a strong positive correlation, with data points clustered around the diagonal line (y=x).</p>

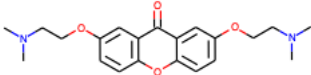
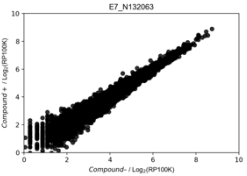
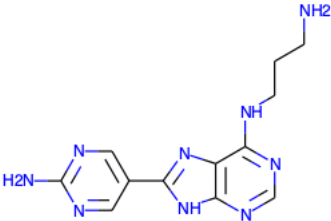
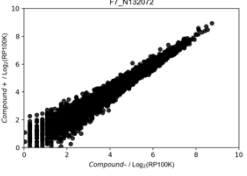
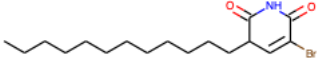
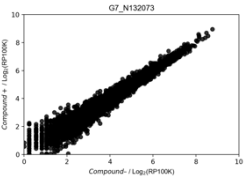
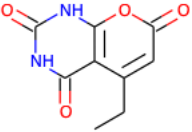
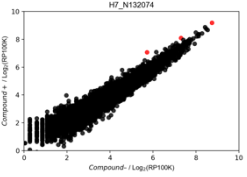
A	5	N131047		
B	5	N131050		
C	5	N131051		
D	5	N131052		

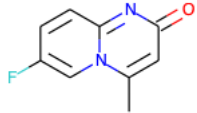
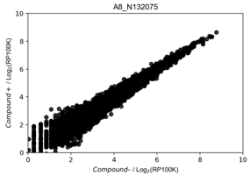
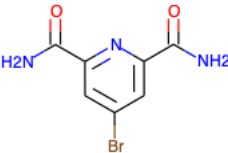
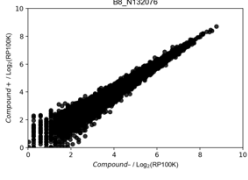
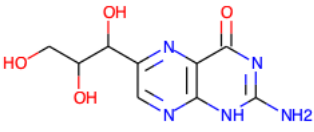
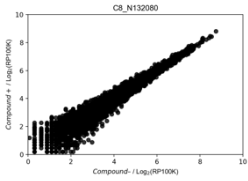
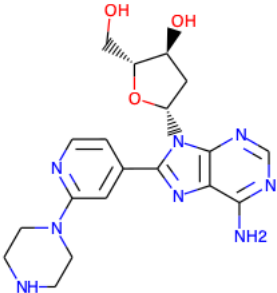
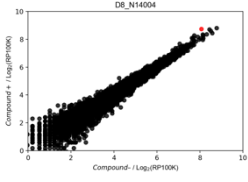
E	5	N131065		
F	5	N131069		
G	5	N131074		
H	5	N131077		

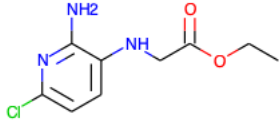
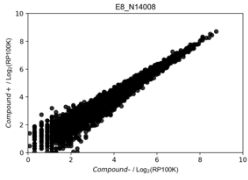
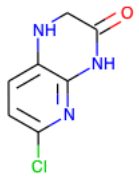
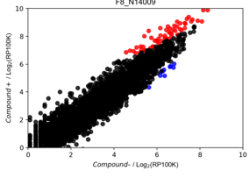
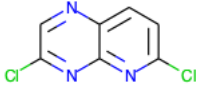
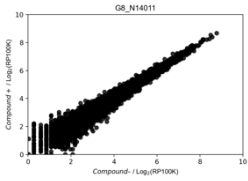
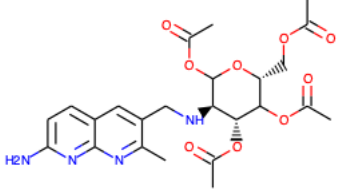
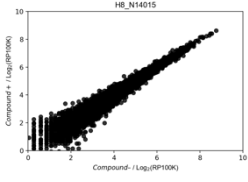
A	6	N131078		
B	6	N132002		
C	6	N132009		
D	6	N132010		

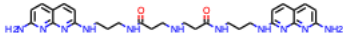
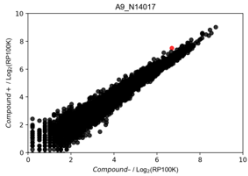
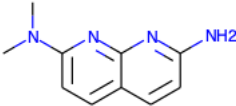
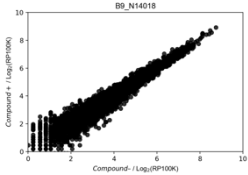
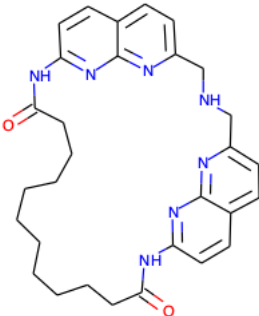
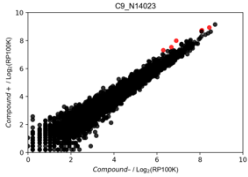
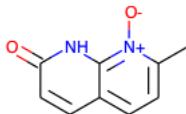
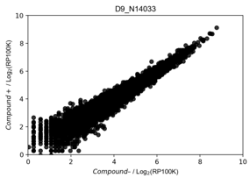
E	6	N132023		
F	6	N132030		
G	6	N132035		
H	6	N132043		

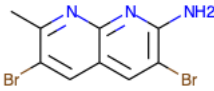
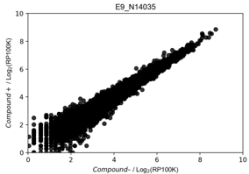
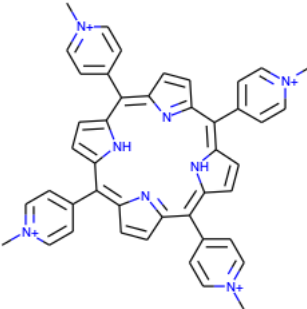
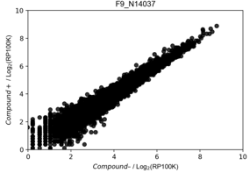
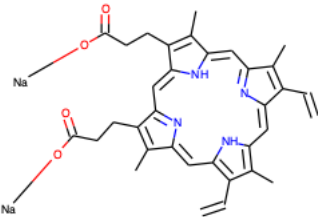
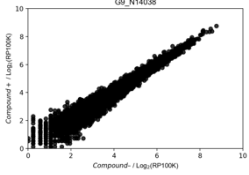
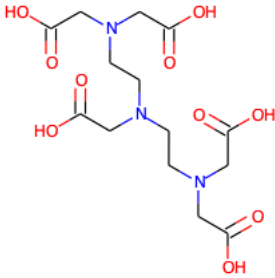
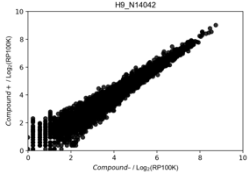
A	7	N132044		
B	7	N132048		
C	7	N132054		
D	7	N132059		

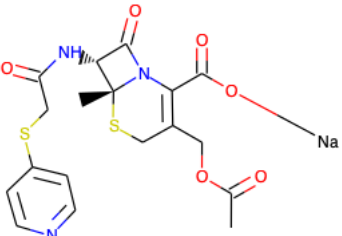
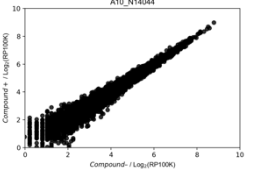
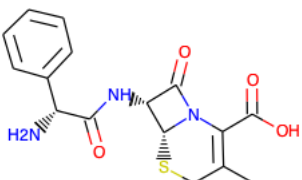
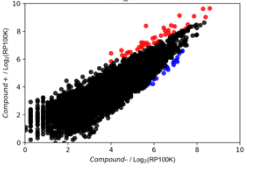
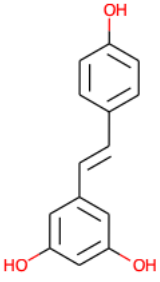
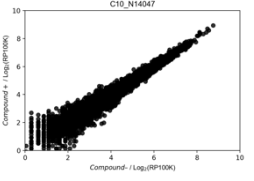
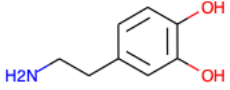
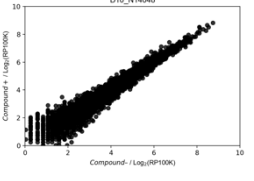
E	7	N132063		
F	7	N132072		
G	7	N132073		
H	7	N132074		

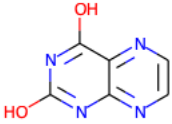
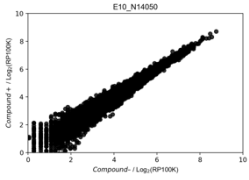
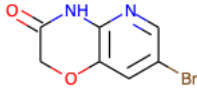
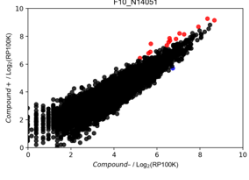
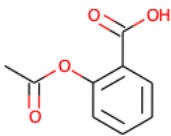
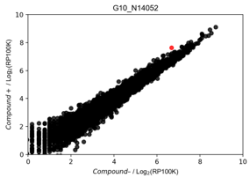
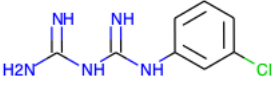
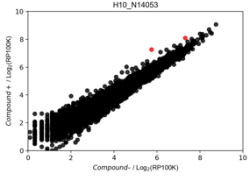
A	8	N132075		
B	8	N132076		
C	8	N132080		
D	8	N14004		

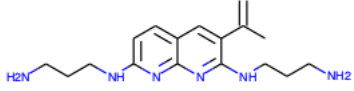
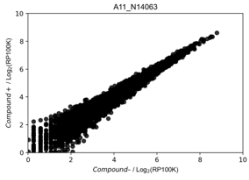
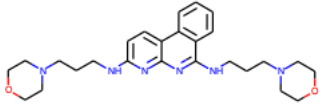
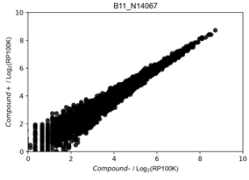
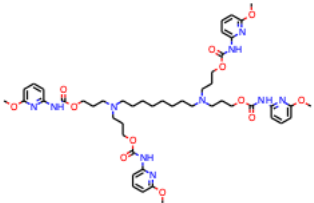
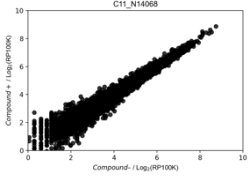
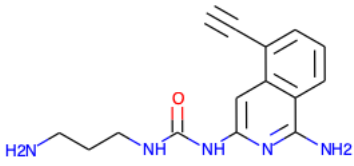
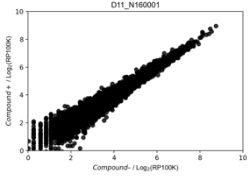
E	8	N14008		
F	8	N14009		
G	8	N14011		
H	8	N14015		

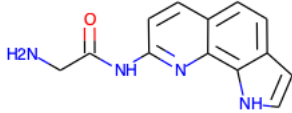
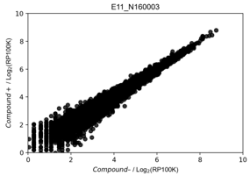
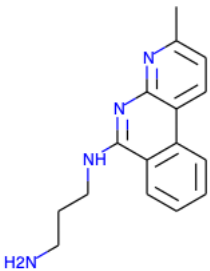
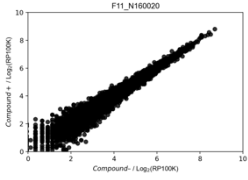
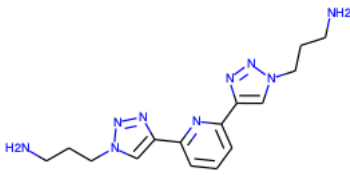
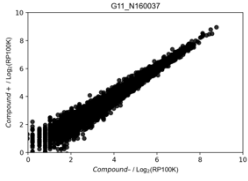
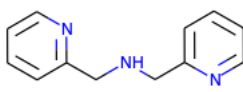
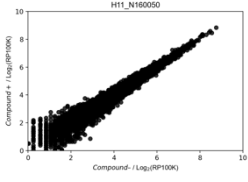
A	9	N14017		
B	9	N14018		
C	9	N14023		
D	9	N14033		

E	9	N14035		
F	9	N14037		
G	9	N14038		
H	9	N14042		

A	10	N14044		
B	10	N14045		
C	10	N14047		
D	10	N14048		

E	10	N14050		
F	10	N14051		
G	10	N14052		
H	10	N14053		

A	11	N14063		
B	11	N14067		
C	11	N14068		
D	11	N160001		

E	11	N160003		
F	11	N160020		
G	11	N160037		
H	11	N160050		

S2. Perl script for sequence Counting

```
#!/usr/local/bin/perl
use strict;

if(@ARGV != 3){
    print "perl $0 in.fastq output1(count_result) output2(alignment_result)\n";
    exit(0);
}

my $in = $ARGV[0];
my $out1 = $ARGV[1];
my $out2 = $ARGV[2];

my $c1 = 0; #for matched reads
my $c2 = 0; #for non-matched reads
my %count= ();
open(F,$in) || die "Cannot open an input file; $in\n";
open(Fo2,"> $out2") || die "Cannot make a output file(2)\n";
print Fo2 "ID\tKeyNucleotids\tStartPos(0based)\tHitSeq\n";

my $fix1 = "GCTCCATTGTTTTGAT";
my $fix2 = "GGATTCTTATGCTCC";
my $fix3 = "ATCGTCTCAAATGAGTCT";

while(<F>){
    chomp($_);
    if($_ =~ /^#@/){ #Seek a header line
        my $key = ""; #for 6base nucleotids
        my $l1 = $_;
        my $id = substr($l1,1); #Remove initial character
        $id =~ s/\s.*//; #Remove after space
        my $l2 = <F>;
        chomp($l2);
        <F>;
        <F>; #Throw away 3rd, 4th lines from the header. They are lines for sequence quality scores.
        if($l2 =~ /$fix1([ATGC]{3})$fix2([ATGC]{3})$fix3/){
            $c1++;
            $key = $1 . $2;
            $count{$key}++;
            my $start = length($);
            my $hit = $fix1 . "[" . $1 . "]" . $fix2 . "[" . $2 . "]" . $fix3;
            printf Fo2 "%s\t%s\t%d\t%s\n",$id,$key,$start,$hit;
        } else {
            $c2++;
        }
    }
}
close(F);
close(F2);

open(Fo1,"> $out1") || die "Cannot make a output file(1)\n";
my @nuc = ("T","G","C","A");

printf Fo1 "KeyNucleotids\t%s\t%s(Frac)\n",$in,$in;
foreach my $n1 (@nuc){
    foreach my $n2 (@nuc){
        foreach my $n3 (@nuc){
            foreach my $n4 (@nuc){
                foreach my $n5 (@nuc){
                    foreach my $n6 (@nuc){
                        my $key = $n1 . $n2 . $n3 . $n4 . $n5 . $n6;
                        printf Fo1 "%s\t%d\t%.4f\n",$key,$count{$key},$count{$key}/$c1;
                    }
                }
            }
        }
    }
}
```

```
}  
}  
}  
}  
}  
printf Fo1 "\n";  
printf Fo1 "Total_Matched\t%d\t%.4f\n", $c1, $c1/($c1+$c2);  
printf Fo1 "Total_Non-Matched\t%d\t%.4f\n", $c2, $c2/($c1+$c2);  
close(Fo1);
```


S3. Python script for sequence Counting

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
from natsort import natsorted
import pandas as pd
import glob
import re
import itertools

folder = #fastaq folder pass name
outputfolder =#output folder pass name

file_names = []
target = '.fastq'
for name in glob.glob(str(folder)+'/*'+target):

    file_names.append(name)
file_names = natsorted(file_names)
#print(file_names)
outputnames = []
for ff in file_names:
    sep = '/'
    t = ff.split(sep)
    r = t[9]
    #target = '.fastq'
    idx = r.find(target)
    q = r[:idx]
    outputnames.append(q)
print(outputnames)
base = 'GCAT'
sixbase = itertools.product(base,repeat=6)
key =[]
for xx in sixbase:
    key.append(xx)
seqlist = []
for yy in key:
```

```

nuc = yy[0]+yy[1]+yy[2]+yy[3]+yy[4]+yy[5]
seqlist.append(nuc)
total = pd.DataFrame(None,columns=['name', 'total_Matched', 'total_nonMatched'])
for file_name, outputname in zip(file_names, outputnames):
    with open(file_name, "r", encoding="utf-8") as df:
        list = df.readlines()
    new_list = []
    for ii in list:
        word = ii.split()
        new_list.append(word)
    index=[]
    data = []
    for number in range(1,len(new_list),4):
        index.append(number)
        data.append(new_list[number])
        print(outputname +' is in progress.')
    match = []
    nonmatch = []
    for df in itertools.chain.from_iterable(data):
        key1 = "GCTCCATTTGTTTTGAT"
        key2 = "GGATTCTTATGCTCC"
        key3 = "ATCGTCTCAAATGAGTCT"
        try:
            s1 = re.search(r'GCTCCATTTGTTTTGAT(.*?)ATCGTCTCAAATGAGTCT',df).group(1)
            try:
                s2 = s1.replace(key2, "")
                if len(s2)==6:
                    match.append(s2)
            else:
                nonmatch.append(df)
        except AttributeError:
            nonmatch.append(df)
        pass
    except:
        nonmatch.append(df)
        pass

```

```
basecount = []
for seq in seqlist:
    basecount.append(match.count(seq))
seqcount = pd.DataFrame(zip(seqlist,basecount),columns=['Keynucleotide','read_number'])
nonmatch = pd.DataFrame(nonmatch)
total = total.append({'name':outputname,'total_Matched' : len(match),'total_nonMatched':
len(nonmatch)},ignore_index=True)
#print(seqcount)
seqcount.to_csv(outputfolder+outputname+'_matched_count.csv')
nonmatch.to_csv(outputfolder+outputname+'_nonmatched_seq.csv')
print('finish')

print(total)
total.to_csv(outputfolder+'result.csv')
```

S4. Python script for normalization

```
import numpy as np
import pandas as pd
def hosei(xx, yy):
    LogFC=[]
    for XX, YY in zip(xx,yy):
        logfc = np.log2(XX) - np.log2(YY)
        LogFC.append(logfc)
    df= pd.DataFrame({'xx':xx,'yy':yy,'LogFC':LogFC})
    df_s= df.sort_values('LogFC')
    df_h=df_s[1024:3172]
    a,b = np.polyfit(np.log2(df_h['xx']),np.log2(df_h['yy']),1)
    log2xx1 = [a * np.log2(df['xx'][nn]) + b if df['xx'][nn] != 0 else 0 for nn in range(len(df['xx'])) ]
    log2yy1 = [np.log2(df['yy'][nn]) if df['yy'][nn] != 0 else 0 for nn in range(len(df['yy']))]
    xx1 = [2**log2xx1[nn] for nn in range(len(df['xx']))]
    yy1 = [2**log2yy1[nn] for nn in range(len(df['yy']))]
    return xx1 , yy1
```

S5. Python script for Analysis and scatter plotting

```
import numpy as np
import pandas as pd
import os
import glob

from natsort import natsorted
import matplotlib.pyplot as plt
import scipy as sp

from scipy import stats
from hosei import hosei

font = {'family': 'Arial',
        'color': 'k',
        }

Comp_minus='A2_None'
folder = '###'
outputfolder = str(folder)+'/analyzed_res'
file_names = []
for name in glob.glob(str(folder)+'/*'+'_matched'+'+*'+'.csv'):
    file_names.append(name)
file_names = natsorted(file_names)
outputnames = []
for ff in file_names:
    q=os.path.basename(ff).split('.', 1)[0]
    target = '_matched'
    idx = q.find(target)
    q = q[:idx]
    outputnames.append(q)
print(outputnames)
df=pd.DataFrame()
for file,outputname in zip(file_names,outputnames):
    df2 = pd.read_csv(file,usecols=['Keynucleotide','read_number'],index_col='Keynucleotide')
    df2 = df2.rename(columns={'read_number': str(outputname)})
    df = pd.concat([df, df2], axis=1)
    #print(df)
print(df)
```

```

xx=df[Comp_minus]
XX = (xx / sum(xx))*100000

Down=[]
Up=[]

for outputname in outputnames:
    if outputname==Comp_minus:
        Down.append(0)
        Up.append(0)
        pass
    else:
        fig, ax = plt.subplots()
        yy=df[outputname]
        YY =(yy / sum(yy))*100000
        xx,yy = hosei(XX,YY)
        xx = pd.DataFrame(xx,columns=[Comp_minus],index=df.index)
        yy = pd.DataFrame(yy,columns=[outputname],index=df.index)
        XX = xx[Comp_minus]
        YY = yy[outputname]
        qvalue=[]
        sigdef=[]
        for nn in range(len(XX)):
            AA = XX[nn]
            BB = YY[nn]
            CC =sum(XX)-AA
            DD =sum(YY)-BB
            df_cross = pd.DataFrame([[AA,BB],[CC,DD]])
            try:
                chi2, pp, dof, ef= sp.stats.chi2_contingency(df_cross, correction=False)
                qq= pp*4096
                if qq<0.05:
                    if AA<BB:
                        sigdef.append('Down')
                    else:
                        sigdef.append('Up')
            else:
                sigdef.append('F')

```

```

        qvalue.append(qq)
    except ValueError:
        qq= 'NA'
        qvalue.append(qq)
        sigdef.append('F')
    pass

qvalue = pd.DataFrame(qvalue,columns=['qvalue'],index=df.index)
sigdef = pd.DataFrame(sigdef,columns=['SigDef'],index=df.index)
df3 = pd.concat([XX,YY,qvalue,sigdef],axis=1,names=[Comp_minus, outputname,'qvalue','SigDef'])
if folder+'qvalue/':
    if not os.path.exists(folder+'qvalue/'):
        os.mkdir(folder+'qvalue/')
df3.to_csv(folder+'qvalue/'+str(outputname)+'_補正後'+'.csv')
print(outputname)
x = df3[Comp_minus].to_numpy()
y = df3[outputname].to_numpy()
x = np.ravel(x)
y = np.ravel(y)
sigdef = sigdef.to_numpy()
sigdef = np.ravel(sigdef)
print('Down regulate')
print(np.count_nonzero(sigdef == 'Down'))
print('Up regulate')
print(np.count_nonzero(sigdef == 'Up'))
Down.append(np.count_nonzero(sigdef == 'Down'))
Up.append(np.count_nonzero(sigdef == 'Up'))
#####
temp = (sigdef == 'Down')
plt.plot( np.log2(x[temp]), np.log2(y[temp]) , 'ro', alpha=0.8)
temp = (sigdef == 'Up')
plt.plot( np.log2(x[temp]), np.log2(y[temp]) , 'bo', alpha=0.8)
temp = (sigdef == 'F')
plt.plot( np.log2(x[temp]), np.log2(y[temp]) , 'ko', alpha=0.8)
plt.xlim(0, 10)
plt.ylim(0, 10)
plt.xlabel('$Y_{it}$ / Log$_2$$(RP100K)$', fontdict=font)

```

```
plt.ylabel('$\gamma_{it}$ {Compound +} $ / \text{Log}_{2}$ (RP100K)', fontdict=font)
plt.title(str(outputname),fontdict=font)
plt.show()
#####
if outputfolder:
    if not os.path.exists(outputfolder+'/'):
        os.mkdir(outputfolder+'/')
fig.savefig(str(outputfolder)+'/'+str(outputname)+".png", dpi=200)
print('-----')
```


S6. Data sets and Python script used for the analysis in Chapter 2

The data sets, analysis results, and computer code used in the studies in Chapter 2 are available on Zenodo by collaborator Yuki Komoto ([10.5281/zenodo.7069764](https://doi.org/10.5281/zenodo.7069764)).

謝辞

本研究を行うにあたり、沢山のご指導、充実した研究環境、そして学会参加機会を与えて下さいました中谷和彦教授に、感謝致します。中谷先生には、研究に関しては勿論のこと、人生の先輩としてのアドバイスを沢山頂きました。中谷先生との出会いが私の人生を変えたと感じています。心より御礼申し上げます。慣れない私に分子生物学の知識や実験手法を一から丁寧に教えて頂き、研究から申請書に至るまで、私の研究生活を支えて下さった村田亜沙子准教授に深く感謝致します。研究室生活の中で、些細な問題から研究に関する的確な助言、アドバイスを頂いた堂野主税准教授に感謝致します。研究に関するだけでなく、日常の問題から将来に関する事にも助言頂き、話を聞いて下さった山田剛史助教に感謝致します。研究において様々な助言や自身の経験などを話して下さった柴田知範助教に感謝致します。疲れている時もいつも笑顔で話し掛けていただき、デスカッションして頂いた、**Bimolendu DAS** 助教に感謝致します。研究室ローテーションから論文化まで、私に新たな研究の道を開いて下さった谷口正輝教授に感謝いたします。大変お世話になりました。MCBJについてご教授いただき、実験でもたくさんのご助力をいただきました、大城敬人特任准教授に感謝致します。研究室ローテーションから長い期間、プログラミングから実験、論文の作成まで大変お世話になりました、小本祐貴助教に感謝致します。NGSのデータ解析に助力いただきました、近畿大学 理工学部・大学院 総合理工学研究科飯田慶先生に厚く御礼申し上げます。

分子生物学実験のテクニックを丁寧に指導して下さい、また日常生活に関するアドバイスも沢山頂きました、須貝亜矢子技術補佐員、原田恭枝技術補佐員、木村真枝技術補佐員に感謝致します。研究に専念できるようにと、研究室生活を支えるお仕事をして頂いただけでなく、日常の悩みや相談にも乗っていただきました、矢口百合子秘書、勝本みほり秘書に感謝致します。

研究生活全般について教えていただき、研究生活を充実したものにして下さった、先輩方に感謝致します。研究室を盛り上げ支えてくれた賢くて愉快的な後輩たちに感謝致します。防衛医科大学校の武井史恵准教授をはじめ、様々な形で研究を支え、盛り上げて下さった方々に感謝致します。最後に、いつも支えて続けてくれた家族に感謝の意を表し謝辞とさせていただきます。

2023年2月

高島