

## キャピラリー電気泳動1ラウンド選抜法による DNA アプタマー配列データのAI解析

齋藤伸吾

### 1 はじめに

分析化学は、分析データを扱う学問でもあり、DX化とは根本的に相性が良い。しかし、ある物質を特異的に認識する分子を創り出し、新たな反応系を用いる分析法を構築するという、方法論確立の観点から見たときは、DX化は利用するにはいまだ遠いものとする研究者も多いだろう。また、DX化を成す手段の一つとしてAI解析（特に機械学習のうちの深層学習や主成分分析など）があるが、これは大規模データを用いて、そのデータ群の持つおよその傾向や特徴を知るため、あるいは系を最適化する場合に最も力を発揮するという印象がある。よって、AI解析を使って一般的な傾向に埋もれた特殊な解（ここでは特異的な反応系）をデータから引き出すには、様々な原理的な工夫が必要と考える。さらに、深層学習などの様に解析結果に理論的バックボーンがない場合には、経験的にその傾向・特徴が強い、という解が現れるに過ぎず、研究のヒントにこそなれ、解析データを学問（真理）に昇華するには相応の工夫が必要であろう。一方で、機械学習は理論的裏付けを与えずとも（合っているとは限らないが）解を与えてくれるという点では非常に便利なものである。

筆者は研究者として、DX化・AI解析を推進したいというより、緻密な計測と洞察に基づきながら学問に貢献したいと考えているが、AI解析を使って学問に貢献する手段はあると考えている。その一つは、大規模なデータ群から、意味のある特殊な多変量データを、理論的なバックボーンに基づいて実験化学的に獲得することであり、その様な高品質データを取り扱うのであれば、最小限の経験論的なDX化でも、何某か学問的に意味のある結果が得られると考える。本稿では、分離化学的に高品質な多変量データを獲得し、得られたデータの機械学習から特殊な分子を獲得する手法について述べたい。具体的には、キャピラリー電気泳動法（CE）を用いて、分子認識素子であるDNAアプタマー配列を含む多量なデータ群を獲得し、それをクラスタリングや深層学習といった機械学習によって解析・判別・比較し、高機能性分子の獲得が可能である、という筆者の研究例を紹介する。

### 2 DNAアプタマーと選抜手法

#### 2.1 DNAアプタマー

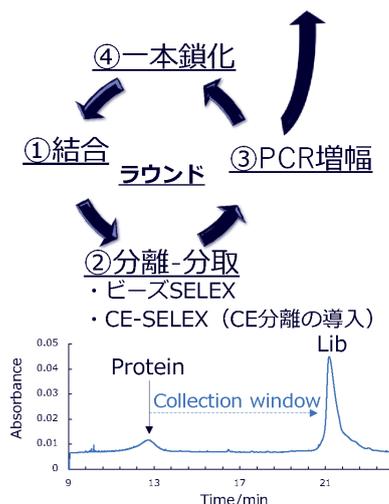
DNAアプタマーとは様々な物質を特異的に認識（結合）する一本鎖DNA（ssDNA）のことである。RNAアプタマーも有用な分子認識材料であるが、ここではより獲得が容易なDNAアプタマーに話を絞る。DNAアプタマーはタンパク質をはじめ、小分子、金属イオン、さらには細菌細胞、動物細胞、ウイルスなど多様な物質を認識するものが報告されている<sup>1)2)</sup>。アプタマー配列の塩基数は数塩基から百塩基近くのもものが報告されているが、一般的には20~40塩基程度のもものが多く、また、その結合親和性は、解離平衡定数 $K_d$ で数nM~数百nMのもものが一般的である。DNAアプタマーは二重鎖形成によるループ、ステム、グアニン四重鎖構造などの高次構造を形成し<sup>3)4)</sup>、それら足場としながら塩基部分などで多点的に標的分子と結合するとされ、抗体の様な多様な認識が可能とされる。

#### 2.2 SELEX法

DNAアプタマーの一般的な獲得方法はSELEX法（systematic evolution of ligands exponential enrichment）である<sup>1)2)</sup>。in vitro選抜とも呼ばれ、多様性の高いランダムDNAライブラリー（Lib）から分子進化学的に選抜する手法である。Libは、PCR増幅に必要な20塩基ほどの二つのプライマー領域が5'および3'末端に位置し、その間にA、G、C、Tの塩基が無作為に並んだランダム領域が設定される。Libは試薬会社から合成物を購入可能であり、理想的には一配列が一分子含まれている。Libの多様性（配列パターン）は理論的には $4^N$ （ $N$ はランダム領域の塩基数）通りあることになり、実験的にはそのうちの $10^{12}$ ~ $10^{15}$ 通り程度が選抜に供されることが多い。

SELEXは以下の工程で行われる（図1参照）。①標的分子（一般には磁気ビーズ等に固定化されたものを用いる）とLibを混合し、Lib中にごくわずかに存在するアプタマーと標的分子を結合させる。②標的-アプタマー複合体を分離する（ビーズを洗浄する等）。③複合体からアプタマーを解離させ、PCR増幅する。これを選抜プールと呼ぶ。④選抜プールの二本鎖DNAを一本鎖化

➤ SELEX法  
⑤配列決定、⑥アプタマーの確認



➤ SR-CE法  
④配列決定・機械学習  
(アプタマー判別・性能予測・プール比較)

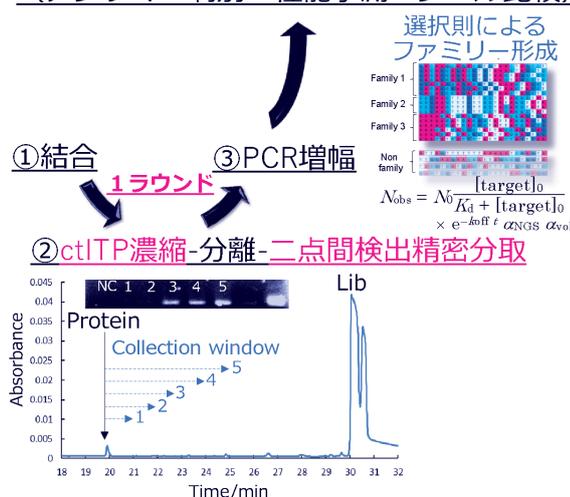


図1 SELEX法およびSR-CE選抜法の概念図

する。この①～④の工程をラウンドと呼び、ラウンドを繰り返すことでDNAアプタマー配列をプール中に濃縮していく。タンパク質には5～10ラウンド程度、細胞等に対しては10～30ラウンド程度が一般的に試行される。その後、⑤十分に濃縮したプール中のDNAの配列を次世代シーケンサー（NGS）等で決定する。ここでよく濃縮された配列（NGSで読みだされた回数（カウント数）の多い配列）がアプタマー配列候補となり、⑥結合実験等で実際にアプタマーであるかの確認とその性能が検証される。

通常の（ビーズ）SELEXを使えば、標的と結合するアプタマーが獲得できることは間違いなく、それゆえに自在に分子認識素子を得られる方法と期待されて多くの研究が発表されてきた。しかし、以下の点がSELEX法で解決できていない点として挙げられる。A) アプタマーが得られない場合、アプタマーが存在しているか自体が不明であり、獲得できるまでSELEXを何度も試行する必要がある。一般に失敗したSELEX結果は論文化されず、成功の可否を決めるファクターが不明である。B) Libの多様性（ $10^{15} \sim 10^{24}$ ）に比して選抜に供する配列数は $10^{12} \sim 10^{13}$ 程度が多く、配列の全体集合のごく一部の部分集合を見ているに過ぎない。よって、より高性能なアプタマーを見逃している可能性も十分にある。これゆえに上記Aの問題が生じる。C) NGS結果のうち、カウント数の多い（あるいはラウンドごとにNGSで観測し、プール中で濃縮傾向のある）配列をアプタマー候補とするが、この結果はPCR増幅によるバイアスを大きく受けることが知られている<sup>5)6)</sup>。つまり、PCR増幅しやすい配列、あるいは何らかの理由（非特異的結合等）で一度増幅されてしまった配列がアプタマー配列候補となってしまう。D) 標的を固定化するこ

とで、自由溶液中とは異なる立体構造や配向の標的分子に対して選抜を行うことになり、思ったような高結合能配列を得られない。実際に得られたアプタマーにおいて、固相化するか自由溶液中で計測するかで解離平衡定数 $K_d$ 、解離速度定数 $k_{off}$ 値が変化することが知られている<sup>7)</sup>、すなわち、アプタマー選抜というものを現実的に定量的、網羅的に扱う研究は筆者の知る限り非常に少ない。

2.3 CE-SELEX法

SELEX法において、上述の工程②を洗練することに成功した代表例にCE-SELEXがある<sup>8)</sup>。これはビーズSELEXと異なり、自由溶液中での高分離が可能なキャピラリー電気泳動法（CE）<sup>9)10)</sup>を分離-分取装置として用いるものである。CEは一般に溶融シリカキャピラリー中に数nLの試料を圧力注入し、その後、数十kVの高電圧を印加することで、電気泳動的にイオンを高分離する手法である。これを用いることでDNA-標的複合体と遊離DNA（Lib）を高分離可能である。数～数十万の理論段数 $N$ および完全分離に相当する1.5以上の分離度 $R_s$ を容易に達成できるCEをアプタマー選抜に利用することは、効率的にアプタマー配列を多く含むプールを獲得するのに合理的である。さらに、上記課題Dに関しては、CEは自由溶液中での泳動分離であるため解決できる。CE-SELEXの開発により通常のラウンド数よりも少ない5ラウンド以下でアプタマー配列が濃縮されたプールを得られるようになった<sup>5)11)</sup>。CE-SELEXでの課題は、導入できる試料体積が小さいため、選抜に供する配列数の少なさ（ $10^{11} \sim 10^{12}$ 程度）であり、これを解決する様々な工夫がなされているものの、根本的なSELEXの課題（上記A～C）は解決できていない。ま

た、近年でも様々な SELEX 自体に関する改良が報告され、配列データに対する機械学習のアプローチもなされているが<sup>1)12)~14)</sup>、ここでは割愛し、課題 A~D に取り組んだ筆者の研究の紹介に留める。

### 3 1 ラウンド CE 選抜と機械学習

#### 3.1 1 ラウンド選抜法

筆者が着想したのは、ラウンドを繰り返さず、1 ラウンドだけで得た選抜結果に対し、溶液反応論を適用すれば、半定量的・半網羅的にアプタマー選抜できるのではないだろうかということである。ただし、これを達成するには仮定と選抜条件を一つずつ設定する必要がある。仮定は「アプタマー配列は分子認識に必要な固定塩基配列以外の数塩基が異なっても同程度の結合能を有すること」、選抜条件は「ランダム領域を短めの 25 塩基程度とすること」、の二つである。25 塩基以下の DNA アプタマーの例は多く知られており<sup>15)~21)</sup>、標的にもよるが多くのアプタマーが存在することが分かっている塩基長である。ここで、分子認識に大きく関与しない 25 塩基中の 4 塩基程度がランダムに配列中に存在することを考える。実際に 25~40 塩基のアプタマーでも類似配列が同様の親和性（解離平衡定数  $K_d$ ）を有している例が報告されている<sup>22)23)</sup>。そうすると、全配列パターンのうちのある任意の配列に対し、4 塩基以下が異なる配列の割合は、簡単な確率計算からおおよそ  $10^{-7}$  % となる。後述するが筆者の CE 選抜では  $10^{12}$ ~ $10^{13}$  配列の導入が可能であったため、ある任意の配列を考えても必ず 1000 分子以上の類似配列が溶液中に存在し、選抜に供されることとなる。すなわち、ある任意のアプタマー配列に対して少なくとも複数の類似配列は必ず選抜に供され、仮想的に網羅的な選抜が可能であると考えた。ここでは一塩基だけの違いで結合能の有無が決まるようなアプタマー配列は獲得できないが、それは通常の SELEX でも同様である。あり得るすべて配列のうちの部分集合に対して淘汰をかけていく SELEX では、一塩基違いも許さない配列を得るには何万回も選抜を試行しなければならない。

ここで筆者が考えたのは、上記の様に Lib 中に 1000 分子以上の同様な配列の集合が存在すると考えるのであれば、これを濃度として近似し、選抜に溶液反応論を導入できるということである。ここで、CE 分離する際に何が起きるかを考える。CE に導入する Lib と標的分子の平衡混合物中で、ある濃度を持つアプタマー分子の錯形成割合は、平衡論から  $[\text{target}]_0 / (K_d + [\text{target}]_0)$  で与えられる（初期標的濃度  $[\text{target}]_0$  が大過剰の場合）。また、Lib と標的分子-アプタマー複合体が CE で完全に分離した際には、複合体ゾーンは非平衡状態に置かれるためキャピラリー中で解離反応が進行する。この反応は一次の速度則に従い、検出までに複合体が解離する割

合は  $\exp(-k_{\text{off}} t)$  で与えられる（ $k_{\text{off}}$  は解離速度定数、 $t$  は分離時間）。したがって、選抜試料中の任意の配列に対して  $N_0$  個の類似配列が含まれていた場合、一度の CE 選抜後に得られる類似配列の数  $N_{\text{obs}}$  は以下の式で与えられる。

$$N_{\text{obs}} = N_0 \frac{[\text{target}]_0}{K_d + [\text{target}]_0} e^{-k_{\text{off}} t} \alpha_{\text{NGS}} \alpha_{\text{vol}} \dots \dots \dots (1)$$

ここで  $\alpha_{\text{NGS}}$  と  $\alpha_{\text{vol}}$  は、それぞれ NGS に供された全配列のうち、実際に読み出された配列の割合および選抜プール溶液体積のうち NGS に導入した割合である（著者らの実験条件では  $\alpha_{\text{NGS}} = 1$  と  $\alpha_{\text{vol}} = 0.2$  である）。この式は 1 ラウンドの選抜における選択則であり、アプタマー配列であれば選抜プール中に類似配列を形成し、 $k_{\text{off}}$  および  $K_d$  が小さい親和性の高い配列ほど大きな  $N_{\text{obs}}$  が観測されることを示している。よって、例えば 1 ラウンドの選抜であっても、アプタマーの特徴を有する類似配列群（ファミリー）が配列解析で見いだせる。つまり、ファミリーによってアプタマー配列を判別すること、および結合能を予測できることになる。また、半網羅性が機能するのであれば、25 塩基のアプタマーの中では最も結合能の高いものを逃さずに選抜できるはずである。加えて、1 ラウンドの選抜であれば一度の選抜操作だけしか行わないため、PCR 増幅のバイアスも最小限で、かつラウンドを重ねる SELEX と比べて高速な選抜法である点も重要である。上記のような思考実験の末、1 ラウンドの選抜によって半網羅的に再現性のある選抜が可能であると着想した。膨大な配列を解析する際に機械学習を用いたが、これについては後述する。本法は SELEX のように選抜プール中の配列の淘汰・濃縮を原理としない。したがって SELEX とは本質的に別種の方法であり、1 ラウンド選抜法（SR 選抜）と呼ぶこととする<sup>24)</sup>。

#### 3.2 SR 選抜における基盤技術

上記の思考実験の SR 選抜を達成するには、一度の分離-分取で出来るだけアプタマー配列を多く含む選抜プールの獲得が必須である。筆者は自由溶液中で高分離が可能な CE を用いることにしたが、さらに CE 選抜にオンキャピラリー濃縮系と精密な CE 分取技術を導入することで初めて SR 選抜を達成できた。以下にそれぞれの基盤技術を説明する。

##### 3.2.1 オンキャピラリー濃縮法

CE 選抜にオンキャピラリー濃縮法を導入した理由は、CE としては大量の試料（500 nL~1  $\mu$ L）を導入することで選抜に供する配列を  $10^{12}$ ~ $10^{13}$  にできることと、濃縮によるゾーンの先鋭化によって  $R_s$  を向上させ、非アプタマー配列の排除を効率化すること、の二点を狙ったためである。オンキャピラリー濃縮法としては過渡的等

速電気泳動法 (tITP, transient isotachopheresis) を用いた<sup>25)</sup>。ここでは ITP の原理の詳細は割愛するが、簡単に述べると、試料イオン ( $S^-$ ) よりも見かけの移動度の大きいリーディングイオン ( $L^-$ ) と小さいターミナルイオン ( $T^-$ ) で試料ゾーンを挟みこみ、それぞれのイオンゾーンが混ざらない状態で泳動する ( $L^- > S^- > T^-$ ) と、 $L^-$  と  $T^-$  の濃度に依存して  $S^-$  が狭いゾーンに濃縮される手法である。さらに ITP 濃縮がなされた後に、各イオンゾーンが混合されるような泳動緩衝液系を使用することで、濃縮から分離モードに自動的に移行させ、ITP 濃縮とキャピラリーゾーン電気泳動 (CZE) 分離を一度の泳動で行うことができる。これが tITP であり、CE としては大容量の試料を注入しつつ (導入体積 500 nL~1  $\mu$ L,  $10^{12}$ ~ $10^{13}$  配列)、高分離が達成できる。Lib と標的タンパク質を分離した典型的泳動図を図 1 中 (右下) に示す。Lib ピークの前端とタンパク質ピークが ITP 濃縮によって先鋭化していることが分かる。ピーク自体は 7~10 倍程度の濃縮がかかっている。Lib ピークの前端を正規分布近似すると、このピーク形状と濃度の計算から  $7\sigma$  以上離れた位置には理論上一分子の Lib 由来の DNA も存在しないということになる ( $\sigma$  は標準偏差)。よって  $7\sigma$  以上離れた位置で選抜プールが獲得できれば、Lib からの混入がなくアプタマーを多く含むプールを得ることができると予想できる。CE-SELEX では、1 ラウンド目で極微量な複合体ピークは検出できないため、タンパク質ピークから Lib ピークの立ち上がりまでを分取する (図 1 左下)。一方、SR-CE 選抜法ではタンパク質ピークから分取範囲を広げて、別個に複数回精密分取を行い、PCR 増幅によって DNA の存在が認められた分画 (図 1 右下のゲル電気泳動図参照) を選抜プールとした。これによって Lib ピークから最も遠く、かつ DNA アプタマーが多く含まれる分画を獲得することを狙った。

また、筆者は生体粒子 (細菌および動物細胞) に対しても独自の CE 濃縮-分離系を開発し、SR 選抜可能であることを示している<sup>26)27)</sup>。細胞などを CE で分離すると通常は細胞の会合体由来と思われる鋭い (スパイク) ピークが多数かつ再現性なく検出され、Lib との効率的な分離ができない。一方、Armstrong らは泳動液にポリエチレンオキシド (PEO) などの水溶性高分子を添加することで細菌のピークが集束 (フォーカシング) することを報告している<sup>28)</sup>。しかし、その再現性は低かった。そこで筆者は細胞を ITP で一つのバンドへと濃縮するとともに PEO のフォーカシング効果によって、細胞を単一のピークとして保持しながら泳動させることを着想した。その結果、細菌細胞 (*B. globigii*, *B. subtilis*, *E. coli*, *S. cerevisiae*)<sup>29)30)</sup>、浮遊がん細胞 (PC-9, HL-60)<sup>27)</sup> に対して単一ピークを再現性良く獲得することに成功し、この分離モードを高分子増強キャピラリー過渡的等

速電気泳動法 polymer-enhanced capillary transient isotachopheresis, PectI) と名付けた<sup>29)31)</sup>。PectI 法をアプタマー選抜に適用することで、高速かつ再現性のある Lib との完全分離 (細胞ピークの理論段数  $N=20$  万段以上、分離度  $R_s=16$  以上) が可能となった。SR-PectI 選抜の際には、単一となった細胞ピークを細胞および細胞-DNA 複合体の混合分画として一度だけ分取した。生体粒子の電気移動度は Ohshima の式などで表されるが<sup>32)33)</sup>、表面に結合する DNA 量が極微量でゼータ電位を大きく変えるほどでなければ、細胞ピークと細胞-DNA 複合体は同時に泳動されると予想されるからである。実際に筆者らは、一度の細胞ピークに分取でアプタマーの獲得に成功している。このように分取すべき複合体ピークを観測しながら分取できるという点は、タンパク質の SR-CE 選抜と大きく異なり、SR-PectI 選抜はより容易な選抜法である。

### 3.2.2 精密分取技術<sup>34)</sup>

SR-CE 選抜を可能とするには、3.2.1 で述べたタンパク質や細胞に対して CE 分離した分画の精密な分取が必須である。分取窓が狭ければ狭いほど非アプタマー配列の混入を防ぐことが可能だからである。しかし、CE 自体その泳動時間の再現性は決して高くなく (RSD  $\pm 3$ ~5 % 程度)、tITP の様な濃縮と分離モードが混在する系では、各ゾーンの移動速度は泳動中に変化してしまうため泳動時間の再現性はさらに落ちる。よって、これまで精密な分取は困難であった。

そこで筆者は、tITP の様に泳動途中でゾーンの移動速度が変化した場合でも精密な区間分取が可能な技術を考案した。tITP では泳動初期は ITP 濃縮モードとなっており、ゾーンの移動速度は明確ではない (電流値も大きく変動する)。しかし、 $L$  および  $T$  イオンゾーンが混合した後の CZE 分離モードでは各ゾーンは一定速度で移動する。そこで市販の CE 装置を改造し、二台の検出器をキャピラリー出口付近に設置することで、各ゾーンが CZE で一定速度になった後の時間と位置をリアルタイムで計測し、キャピラリー出口からゾーンが溶出される正確な時間を算出することとした。この方法により、30 秒程度のピークであればピークの形状通りに精密に分取できる。CE による分取法は様々なものが報告されているが、その中でも最も正確な分取法の一つである。

この方法によって、蛍光色素、蛍光キレート試薬、DNA アプタマーなどの精製にも成功しており、純度 80 % 程度であった試料溶液でも、純度 99 % 以上に精製可能であることが示された。また、この方法は実環境試料からの放射性金属イオンの CE 分離分取などにも応用されている<sup>35)</sup>。筆者らが SR-CE 選抜あるいは DNA アプタマーの結合実験をする際にも、必ずこの手法で DNA を精製することで初めて精密な実験が可能となっ

ており、筆者らのアプタマー研究に欠かせないものとなっている。裏を返せばこのような方法で純度の高い DNA を用いなければ、精密な SR 選抜は不可能である。

### 3.3 SR-CE 選抜における配列解析

上記 3.1 で述べた SR 選抜の原理に従うのであれば、選抜後の配列データ中の類似配列群（ファミリー）はアプタマー配列であり、選択則（式（1））から大きなファミリーであるほど結合能は高い（ $K_d$ ,  $k_{off}$  が小さい）はずである。この概念実証のために標的タンパク質としてゴールドスタンダードであるトロンビン（Thr：血液凝固第 IIa 因子）<sup>15)36)</sup> をモデルとして用い、SR-CE 選抜に供して選抜プール（配列数  $10^5 \sim 10^6$ ）を得た。その後、Thr と選抜プールの CE-LIF 結合実験を行った。その結果、Thr-DNA 複合体ピークを検出し、1 ラウンドで獲得した選抜プール中に 10~30 % 程度のアプタマー配列が含まれていることを確認した。ちなみに、CE-SELEX で分画分取（Thr ピークから DNA ピークの立ち上がりまでの区間、図 1 左下参照）を行ったところ、1 ラウンド後の選抜プールでは複合体ピークは観測されなかった。すなわち、筆者らの dTP 濃縮法と精密分取法の組み合わせであれば、非結合型 DNA を効率的に排除し、1 ラウンドであっても高速に選抜プールが得られることが分かった。

この辺りからようやく DX 化に関連した話題となる。SR-CE 選抜で得た上記選抜プールを NGS 解析に供し、配列決定した。読み取った配列の全カウント数は ~30 万であったが、SR-CE 選抜法では配列の種類の数に着目し、カウント数は考慮しない。特定の配列の濃縮は起こさず、アプタマー配列の特徴のある配列がファミリーを形成することが SR 選抜の原理であるためである。むしろ、カウント数が増加する配列があれば、それは SR 選抜が原理通り働いていないことを意味する。実際に SR-CE 選抜プールの各配列は、カウント数 1~3 がほとんどであり、全配列のうちカウント数 10 未満のものが 90 % 以上と濃縮が全く起きていないことを確認している。カウント数を無視し、配列の種類の違いで数えると ~7 万ほどの配列が得られた。

ここでどのようにしてファミリーを見つけるかが問題である。筆者らは単純なクラスタリング手法を使用することとした（ソフトウェア CD-HIT-EST<sup>37)</sup>）。原理を簡単に述べると、ある配列（親配列）と他の配列（子配列）を比較し、同じ位置に同じ塩基があるかどうかでスコア付ける。ただし、ある塩基がずれた場合やギャップがある場合などでもそれぞれスコア付けし、最終的なスコアに従って類似配列かどうかを判断し、類似配列である子配列を親配列にファミリーとして分類する。これをすべての配列に対し行い、プール中の全ファミリーを探索した。また、ランダム領域の全塩基同士を比較する

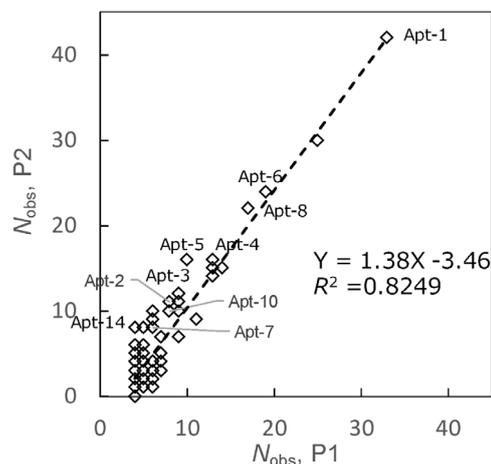


図 2 二つの SR-CE 選抜プール（P1, P2）中のファミリー配列数  $N_{obs}$  の相関  
論文 21 から一部改変して転載。

場合（グローバルアラインメント）と配列中のより短い一部の領域を比較する場合（ローカルアラインメント）とがある。後者はより短い配列モチーフを判別する場合に便利である。ここでは詳細は省くが、両方のアラインメントを行い、どちらの方法でも  $N_{obs}$  の大きなファミリーを抽出している。ここで、構成配列数  $N_{obs} > 4$  をファミリーとすると（Lib の NGS 解析では、同様の解析条件で  $N_{obs} = 3$  のファミリーが最大であったため）、500~1000 種類のファミリーが見つかった。

ここで、SR-CE 選抜実験を独立に二回行い（選抜プール P1 と P2 とする）、二つのプールで同様のファミリーが同じように得られるか、すなわち再現性が得られるかを調査した。ここではプール間比較が可能な CD-HIT-EST-2D プログラム<sup>37)</sup> を用い、P1 のファミリーに対して P2 配列を分類し、P1 と P2 のファミリーの  $N_{obs}$  の相関を取った（図 2）。その結果、良好な直線関係が得られた（ $R_2 = 0.82$ ）。これは、二つの全く独立した実験、すなわち全く異なる（部分集合）配列を含む二つの Lib を使用しても類似の特徴を有する配列群が再現性良く選抜できたことを示している。よって、SR-CE 選抜がアプタマーの特徴を捉えるという点で網羅的であることが判明した。ただし、当然ながら別個の部分集合から獲得した二つのプール間で同じ配列は一つも確認できなかった。よって、ここでは半網羅的な選抜と記載する。この再現性、半網羅性は既存の SELEX にはない SR 選抜だけの大きな特長である。

さて、SR-CE 選抜では、 $N_{obs}$  が大きいほど親和性が高いアプタマーである可能性が高い。これを確認するため、 $N_{obs} = 7 \sim 90$  中の 13 配列を無作為に抽出してアプタマー候補とし、CE-LIF 結合実験に供した。その結果、すべての配列が Thr と結合し、アプタマーであることが分かった。ランダムライブラリーから無作為に選んだ 13 配列がすべてアプタマーである確率はほぼ無いに等

しく、この結果は  $N_{\text{obs}}$  値でアプタマーであるか否かを配列データから判別できることを示す。さらに、結合実験で測定した 13 種のアプタマー配列の  $K_d$  と  $k_{\text{off}}$  を、式 (1) の選択則に代入することで理論的な  $N_{\text{obs}}$  値 ( $N_{\text{cal}}$ ) を計算した。実測値  $N_{\text{obs}}$  と  $N_{\text{cal}}$  の相関を取ったところ、正の相関が得られた ( $R_2=0.76\sim 0.80$ )。完全に理論通りとは言えないものの、 $N_{\text{obs}}$  が高いアプタマーが高性能を示す傾向にあることを示すには十分な結果である。ちなみに、様々なクラスタリング様式で前述の  $N_{\text{obs,P1}}-N_{\text{obs,P2}}$  相関と  $N_{\text{obs}}-N_{\text{cal}}$  相関について解析したところ、前者は 21 塩基長のローカルアラインメントを用いて、21 塩基中 18 塩基が一致する条件で最も相関係数が高く、後者は 20 塩基長のローカルアラインメントを用いて、20 塩基中 19 塩基が一致する条件で最も大きな相関係数が得られた。つまり、SR-CE 選抜では 20 塩基前後のモチーフで Thr と結合するアプタマーが主に獲得できていたことが分かった。

### 3.4 SR-CE 選抜によるアプタマー配列データベース化と新規機能の選抜

#### 3.4.1 DNA アプタマーデータベースとしての可能性

上記の Thr と同様の結果が VEGF-165 タンパク質に対する SR-CE 選抜でも得られている。さらに、Thr 選抜プールの配列とは全く異なる配列が得られていることも確認できている。一方で、トランスフェリン (Tf) を標的にした際には有意なファミリーが得られていない (未発表データ)。すなわち、Tf に対しては、25 塩基アプタマーは存在しないか、あるいは存在しても非常に結合能が低い可能性が高い。実際にこれまで Tf に対するアプタマーの報告はない (酸性条件下で選抜プールを獲得したとの報告はある<sup>38)</sup>が配列決定およびアプタマーの確定には至っていない)。この様にアプタマーが獲得できなかった場合でも、アプタマーが存在するか否かの可能性を議論できる点も SELEX 法にはない SR-CE 選抜の特徴である。

もし、SR-CE 選抜が理想的に働く、すなわちアプタマーの特徴を有する配列を取り逃がさず、再現性良く、半網羅的に選抜できるのであるならば、選抜プールの配列データはアプタマーデータベースとして機能し、以下の利点も生まれるのではないかと。つまり、1) 最も結合能の高いアプタマーを取り逃がすことなく選抜できているはずである。また、2) データベース比較 (プール間比較) から標的選択性のあるアプタマー配列を見つけ出せるはずと考えた。

1) に関しては、SR-CE 選抜で獲得した  $N_{\text{obs}}$  が大きかった Thr および VEGF 結合型アプタマーが既報の高親和性アプタマーの  $K_d$ ,  $k_{\text{off}}$  と比較しても、遜色ない親和性を有することを確かめている。特にランダム領域 25 塩基と若干短めのライブラリーを使用したにもかかわらず、

報告されている中で最も親和性の高いアプタマーと同等以上のものを選抜できたことは意義深い。2) に関しては、標的としてタンパク質だけではなく、細菌細胞 (グラム陽性, グラム陰性, 真菌), 浮遊性がん細胞に対しても SR-PectI 選抜を行い、アプタマーを獲得できているが<sup>26)27)</sup>、これらの選抜プール間比較をしたところ、互いに同じアプタマー候補配列 (ファミリー) はほとんど存在しないことが分かった。これは SR 選抜でそれぞれの標的に特徴的な配列が得られていることを示している。ある標的プールのみの特徴的な配列ということは、その標的への選択性を表していると考えた。そこでこれらの選抜プールを深層学習 (ディープラーニング) に供し、選抜プールの特徴を捉えて配列を分類できるかどうかを試すこととした。深層学習には CNN (convolutional neural networks) と STLM (long short-term memory) を含む比較的単純なアーキテクチャを用いた。深層学習では、それぞれの選抜プールの配列を学習させることで、各配列がどの選抜プールに分類されるかの確率が出力される。未発表データであるため詳細は差し控えるが、 $N_{\text{obs}}$  の大きな配列群は、それら自身の標的の選抜プールに分類され (分類確率が高かった)、他の選抜プールには分類されなかった。また、Thr や VEGF に対し、それぞれ二回の独立の SR-CE 選抜で獲得した選抜プールを深層学習させると、アプタマー候補配列は二つのプール双方にほぼ同じ確率 (0.4~0.6) で分類される。つまり、同じ標的に対しては同様の確率で分類される。すなわち、選抜プールの特徴 (結合選択性) が CNN-STLM といった単純なモデルの実装で明確に捉えられることが明らかとなった。この様に、クラスタリングからは結合能予測 ( $K_d$ ,  $k_{\text{off}}$ ) が可能であり、深層学習からは選択性予測 (配列の特徴量) が可能であると考えている。

#### 3.4.2 データベース比較による機能の選抜<sup>24)</sup>

この様に、SR-CE 選抜法において選抜プールの大規模配列データ群から、アプタマーの判別、結合能の大まかな予測、結合選択性の判別ができることが分かってきた。この SR 選抜の一連の流れは、完全にランダムなデジタルデータの全集合 ( $10^{15}$  配列の Lib) からその一部 ( $10^{12}\sim 10^{13}$  配列) を標本 (入力データ) として SR-CE 選抜操作という演算にかけることで、分布に偏りのある解データ ( $10^5\sim 10^6$  配列) を得ていることに等しい。そこで一つの着想は、入力データを変調することで質の異なる解が出力されないか、すなわち、Lib を変調し、得られた解データを比較することで、ある特殊な機能を選抜できないか、ということである。

ここでは、標的と結合することで立体構造を変化させ、かつ検出シグナルを発生させる機能を有する分子素子の獲得を目指すこととした。これまでも結合時に高

次構造が誘起される DNA アプタマーの報告はある<sup>39)</sup>。ここではそのような「構造誘起アプタマー」を意図的に選抜できる手法を着想した。具体的には、キャピラリー分子ふるい電気泳動 (CSE) を用いて Lib を、標的分子が存在しない状態でも高次構造を形成している DNA 群と形成していない DNA 群とに分離・分取することで、二つのサブライブラリー (構造形成および非形成サブライブラリー: SL1 および SL2) を調製した。これら SL を使って SR-CE 選抜すれば、SL1 からは元から強固な高次構造 (pre-organized 構造) を有する高結合性アプタマーが、SL2 からは標的を分子認識する際に高次構造を誘起させて結合する構造誘起型アプタマーがそれぞれ獲得できるのではないかと着想した。

用いた分子ふるい媒体としては、強い水素結合能がなく、DNA やタンパク質と強く相互作用しないポリエチレンオキシド (PEO) を使用した。もし、高分子と DNA が相互作用してしまうと、その相互作用由来の高次構造が誘起されてしまう可能性が高いからである。さらに、ここでは得られた分画を PCR 増幅に供さないことも重要である。なぜなら、SL 濃度を上げるために PCR 増幅を行ってしまうと、SL 中の配列にバイアスが生じ、一配列一分子の Lib を使うという SR 選抜の前提が崩れ、定量的な解が出力されない、すなわちデータベース比較ができなくなるためである。そこで、tITP 濃縮系を組み込んだ CSE 系を構築し、オンライン濃縮-分離-分取を行った。図 3 に示すように、分取後の SL1 では、構造形成型 DNA が幾つもの鋭いピークの集合として現れ (図 3b)、SL2 ではそのようなピークは現れず、より後方にブロードなピークが観測された (図 3c)。元の Lib の泳動パターンは SL1 と SL2 を合わせた形状で

あり (図 3a)、PCR 増幅による濃度の調整なしで Lib を分割して SL1 と SL2 を獲得できることを確認できた。これらの濃度 (110 nM と 650 nM) は元の Lib 濃度 (3  $\mu$ M) と比べて薄いものの、アプタマー配列が各 SL 中に存在するのであれば、有意に 5~20 配列以上のファミリーを形成する濃度である。この様に、高次構造の違いに基づく Lib の分離挙動の調査、および Lib を構造形成型と非構造形成型の SL に分割して選抜するという試みは、筆者らの研究が初めてである。筆者は、CZE 分離において Lib ピークの幅が単一小分子のピークに比べて広いことや、小さなショルダーピークやフロンティングが観測されることに着目して、Lib 分割の着想を得た。こうした観察と着想は分離化学者の得意とするところであろう。

さて、Thr に対して、SL1 および SL2 を用いて SR-CE 選抜を行い、配列決定後に SL1、SL2 プールのクラスタリングを行った。その結果、それぞれ約 1000 および 200 種類程度のファミリーが見つかった ( $3 \leq N_{\text{obs}} \leq 8$ )。SL1 の方の初期濃度が薄いにもかかわらず選抜プール中のファミリーが多かったことは、多くのアプタマーが高次構造形成型であることを示しており、従来の知見と矛盾しない。得られた二つの選抜プールの配列群をデータベースとして扱い、P1 (分割しない Lib を用いた選抜プール) と比べたところ、多くのファミリー配列は P1 のファミリーに帰属でき、P1 は SL1 と SL2 をほぼ補完していることが分かった。また、SL1 と SL2 プールをデータベース比較したところ、どちらか片方のプールにしか存在しないファミリーは 50 % ほどであった。そこで、SL1 および SL2 プール中から選んだ 15 配列について、Thr との結合能を調査したところ、すべてが Thr 結合性アプタマーであった。そのうち、SL2 プールだけに存在した五つアプタマー配列の  $k_{\text{off}}$  は比較的大きい傾向にあった。これは構造非形成型の SL2 から獲得したアプタマーは速度論的に親和性が低い傾向にあることを示す。また、CSE を用いてアプタマーの回転半径  $r_h$  を測定したところ、SL2 から得られたアプタマー配列は SL1 のものよりも  $r_h$  が 30 % 近く大きいことが分かった。さらに、配列の融解温度  $T_m$  を計測したところ、SL1 だけに存在するアプタマーは高い  $T_m$  を有する一方で、SL2 プールにしか存在しなかったアプタマー 5 配列のうちの 80 % (4 配列) は  $T_m$  の計測ができず、高次構造を形成していないことが分かった。この様に SL1 は構造形成の、SL2 は構造非形成の特徴をそれぞれ有していることが分かった。

ここで、非構造形成型 DNA (SL2 プール) 中のアプタマー配列は、Thr と錯形成する際に高次構造を誘起している可能性がある。そこで、これら五つのアプタマー配列の両端のプライマー部位にフルオレセイン (FAM) と TAMRA 色素を、二重鎖形成を介して導入し、Thr と

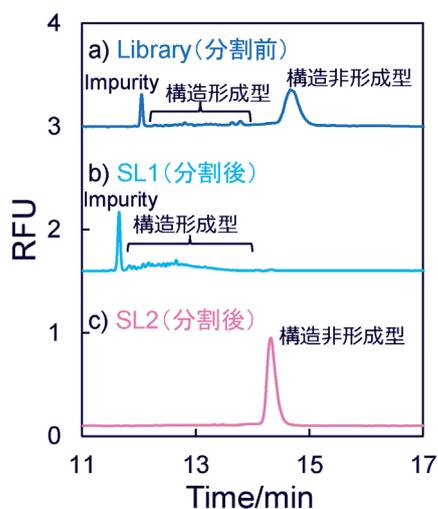


図 3 ランダム DNA ライブラリーの SCE 分離。a) 分割前の Lib。b) SCE 分割後の構造形成型サブライブラリー (SL1)。c) SCE 分割後の構造非形成型サブライブラリー (SL2)

論文 21 から一部改変して転載。

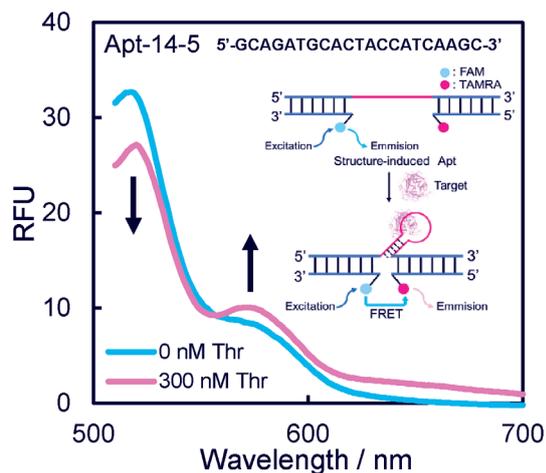


図4 構造誘起型 DNA アプタマーによる Thr の FRET 検出  
論文 21 から一部改変して転載。

の結合前後で蛍光共鳴エネルギー移動 (FRET) 効率が変化するかを測定した。もし、Thr 結合時に初めて高次構造が誘起され、両蛍光色素間距離が近くなれば FRET 効率が上昇するはずである。結果としては 5 配列中の 3 配列で FRET 効率が上昇した。つまり、これら 3 配列は FRET タイプの蛍光センサーとして用いることができる。実際に、これら 3 配列のうちの一つを短縮化した最適配列を用いたところ (図 4)、他のタンパク質の妨害なく、Thr を蛍光レシオメトリーで高感度 (検出限界 2.2 nM) に計測できることが分かった。FRET を起こす Thr センサーの報告は、筆者の知る限り初めてである。

以上の様に、SR 選抜における入力情報 (ライブラリー) を変調し、得られた出力情報 (選抜プール) のデータベース比較により、標的分子を認識し、構造を変化させ、さらに検出シグナルを発生させる機能性分子を意図的に獲得することが可能であることを実証した。

#### 4 おわりに

筆者の一連の研究は、DNA アプタマーの選抜手法を定量的に考察することから始まり、1 ラウンドでの選抜系 (分離-分取系) を構築することによって、ランダムな初期配列分布の変調を再現性良く観測できることを見いだしたものである。これは SR 選抜が半網羅的であり、アプタマーの性能を内包したデータベースになり得ることを示している。このアプタマーデータベースの比較から選択性や機能を引き出すことにも成功した。このような、様々な標的に対し半網羅的に選抜された結果の DNA 配列データを比較する分子認識科学を“Aptaomics”として提案する<sup>24)</sup>。現在は、標的物質の数と種類 (タンパク質、細胞、エクソソーム、ウイルス、藻類等) を増やしている。それらの数千万の配列データを一度に解析することにも成功しており、アプタマーデータベース化の概念実証を推し進めている。

当然ながら課題も多い。第一に挙げられるのは本法のランダム領域の小ささ (25 塩基) である。より長いランダム領域 (30~40 塩基) での SR 選抜を達成するには、現在の NGS よりもハイスループットな DNA 配列決定法が必要となる。また、現在のところ SR 選抜は CE の持つ自由溶液中での高分離能に依存しているため、CE で分離できない標的物質への適用は行っていない。

DX 化という点に目を移すと、筆者の意見では、有意で特殊な高品質データを如何にして取り出すか、また、取り出す原理を理解し、その後、どのようなデータ処理が相応しいかの総合的なデザインが重要であり、それには分離科学および分析化学的思考こそが鍵であると考えられる。データ獲得法の原理・理論がソリッドであれば、獲得できる大規模情報も質が高く、最低限の機械学習で新奇物質を見いだせるというのが筆者のスタンスである。また、DX 化という最近の話題に目を惹かれがちだが、この SR 選抜法の研究においては、新規な生体粒子濃縮法および精密な分取法といった分離分析の実験系の基盤技術が支えている。以上の様に、化学現象とデータ獲得法を深く理解した上での全体論的デザインは、分析化学者にとって得意なところであり、その意味で分析化学者発信の DX 化は大いに期待できると考える。

謝辞 本研究は JSPS 科研費 [22H02104] [23K23372] の助成を受けたものである。ここに謝意を表する。

#### 文 献

- 1) S. Saito : *Anal. Sci.*, **37**, 17 (2021).
- 2) D. J. Chinchilla-Cárdenas, J. S. Cruz-Méndez, J. M. Petano-Duque, R. O. García, L. R. Castro, M. J. Lobo-Castañón, G. O. Cancino-Escalante : *J. Genet. Eng. Biotechnol.*, **22**, 100400 (2024).
- 3) S.-J. Choi, C. Ban : *Sci. Rep.*, **6**, 34998 (2016).
- 4) R. Troisi, V. Napolitano, V. Spiridonova, I. R. Krauss, F. Sica : *Nucleic Acid Res.*, **46**, 12177 (2018).
- 5) M. Jing, M. T. Bowser : *Anal. Chem.*, **85**, 10761 (2013).
- 6) R. Yufa, S. M. Krylova, C. Bruce, E. A. Bagg, C. J. Schofield, S. N. Krylov : *Anal. Chem.*, **87**, 1411 (2015).
- 7) G. K. Uppal, S. Poolsup, E. Zaripov, Y. Gu and M. V. Berezovski : *Anal. Bioanal. Chem.*, **416**, 1697 (2024).
- 8) R. K. Mosing, S. D. Mendonsa, M. T. Bowser : *Anal. Chem.*, **77**, 6107 (2005).
- 9) 齋藤伸吾 : ぶんせき (*Bunseki*), **2023**, 390.
- 10) 日本分析化学会編 : “電気泳動分析”, 初版, (2010), (共立出版).
- 11) M. V. Berezovski, M. U. Musheev, A. P. Drabovich, J. V. Jitkova, S. N. Krylov : *Nat. Protoc.*, **1**, 1359 (2006).
- 12) J. P. Tobia, P.-J. J. Huang, Y. Ding, R. S. Narayan, A. Narayan, J. Liu : *ACS Synth. Biol.*, **12**, 186 (2023).
- 13) N. Iwano, T. Adachi, K. Aoki, Y. Nakamura, M. Hamada : *Nat. Comput. Sci.*, **2**, 378 (2022).
- 14) A. Bashir, Q. Yang, J. Wang, S. Hoyer, W. Chou, C. McLean, G. Davis, Q. Gong, Z. Armstrong, J. Jang, H. Kang, A. Pawlosky, A. Scott, G. E. Dahl, M. Berndt, M. Dimon, B. S. Ferguson : *Nat. Commun.*, **12**, 2366 (2021).

- 15) L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas, J. J. Toole : *Nature*, **355**, 564 (1992).
- 16) B. Wilbanks, J. Smestad, R. M. Heider, A. E. Warrington, M. Rodriguez, L. J. Maher, III : *Nucleic Acid Ther.*, **29**, 126 (2019).
- 17) X. Ren, A. D. Gelinas, I. Carlowitz, N. Janjic, A. M. Pyle : *Nat. Commun.*, **8**, 810 (2017).
- 18) O. A. Alsager, S. Kumar, B. Zhu, J. Travas-Sejdic, K. P. McNatty, J. M. Hodgkiss : *Anal. Chem.*, **87**, 4201 (2015).
- 19) D. J. Scoville, T. K. B. Uhm, J. A. Shallcross, R. J. Whelan : *J. Nucleic Acid*, (2017), DOI : 10.1155/2017/9879135.
- 20) S. M. Krylova, M. Musheev, R. Nutiu, Y. Li, G. Lee, S. N. Krylov : *FEBS Let.*, **579**, 1371 (2005).
- 21) Y. Nonaka, K. Sode, K. Ikebukuro : *Molecules*, **15**, 215 (2010).
- 22) Y. Nonaka, W. Yoshida, K. Abe, S. Ferri, H. Schulze, T. T. Bachmann, K. Ikebukuro : *Anal. Chem.*, **85**, 1132 (2013).
- 23) N. Savory, D. Lednor, K. Tsukakoshi, K. Abe, W. Yoshida, S. Ferri : *Biotechnol. Bioeng.*, **110**, 2573 (2013).
- 24) S. Saito, T. Sakamoto, N. Tanaka, R. Watanabe, T. Kamimura, K. Ota, K. R. Riley, K. Yoshimoto, Y. Tasaki-Handa, M. Shibukawa : *Chem. Eur. J.*, **27**, 10058 (2021).
- 25) M. Urbánek, L. Křivánková, P. Boček : *Electrophoresis*, **24**, 466 (2003).
- 26) S. Saito, K. Hirose, M. Tsuchida, K. Wakui, K. Yoshimoto, Y. Nishiyama, M. Shibukawa : *Chem. Commun.*, **52**, 461 (2016).
- 27) K. Hirose, M. Tsuchida, H. Asakura, K. Wakui, K. Yoshimoto, K. Iida, M. Sato, M. Shibukawa, M. Suganuma, S. Saito : *Analyst*, **142**, 4030 (2017).
- 28) D. W. Armstrong, G. Schulte, J. M. Schneiderheinze, D. J. Westenberg : *Anal. Chem.*, **71**, 5465 (1999).
- 29) S. Saito, T. Massie, T. Maeda, H. Nakazumi, C. L. Colyer : *Anal. Chem.*, **84**, 2454 (2012).
- 30) S. Saito, T. Maeda, H. Nakazumi, C. L. Colyer : *Anal. Sci.*, **29**, 157 (2013).
- 31) 株式会社シノテスト : 齋藤伸吾, 廣瀬和生, 土田真帆, 渋谷雅美, 佐藤誠, 特許 6781883, “アプタマーの選抜方法”, (2020. 10. 21).
- 32) H. Ohshima : *J. Colloid Interface Sci.*, **168**, 269 (1994).
- 33) H. Ohshima : *Colloid Polym. Sci.*, **285**, 1411 (2007).
- 34) T. Haraga, H. Tsujimura, S. Miyauchi, T. Kamimura, M. Shibukawa, S. Saito : *Electrophoresis*, **41**, 1152 (2020).
- 35) K. Ouchi, T. Haraga, K. Hirose, Y. Kurosawa, Y. Sato, M. Shibukawa, S. Saito : *Anal. Chim. Acta*, **1298**, 342399 (2024).
- 36) D. M. Tasset, M. F. Kubik, W. Steiner : *J. Mol. Biol.*, **272**, 688 (1997).
- 37) W. Li, A. Godzik : *Bioinformatics*, **22**, 1658 (2006).
- 38) Q. Li, X. Zhao, H. Liu, F. Qu : *J. Chromatogr. A*, **1364**, 289 (2014).
- 39) H. Qu, A. T. Csordas, J. Wang, S. S. Oh, M. S. Eisenstein, H. T. Soh : *ACS Nano*, **10**, 7558 (2016).



齋藤 伸吾 (SAITO Shingo)

埼玉大学大学院理工学研究科 (〒338-8570 埼玉県さいたま市桜区下大久保 255). 東北大学大学院工学研究科博士後期 3 年の課程応用化学専攻. 工学 (博士). 《現在の研究テーマ》電気泳動を含む分離分析法による新規分子の発見・創出. 《主な著書》“ベーシックマスター 分析化学”, (共著), (オーム社). 《趣味》ラーメン食べ歩き, ギター.

## 原稿募集

話題欄の原稿を募集しています

内容 : 読者に分析化学・分析技術及びその関連分野の話題を提供するもので, 分析に関係ある技術, 化合物, 装置, 公的な基準や標準に関すること, 又それらに関連する提案, 時評的な記事などを分かりやすく述べたもの.

但し, 他誌に未発表のものに限ります.

執筆上の注意 : 1) 広い読者層を対象とするので, 用語, 略語などは分かりやすく記述すること. 2) 啓もう的であること. 3) 図表は適宜用いてもよい. 4) 図表を含めて 4000 字以内 (原則として

図・表は 1 枚 500 字に換算) とする.

なお, 執筆者自身の研究紹介の場とすることのないよう御留意ください.

◇採用の可否は編集委員会にご一任ください. 原稿の送付および問い合わせは下記へお願いします.

〒141-0031 東京都品川区西五反田 1-26-2

五反田サンハイツ 304 号

(公社)日本分析化学会「ぶんせき」編集委員会

[E-mail : bunseki@jsac.or.jp]