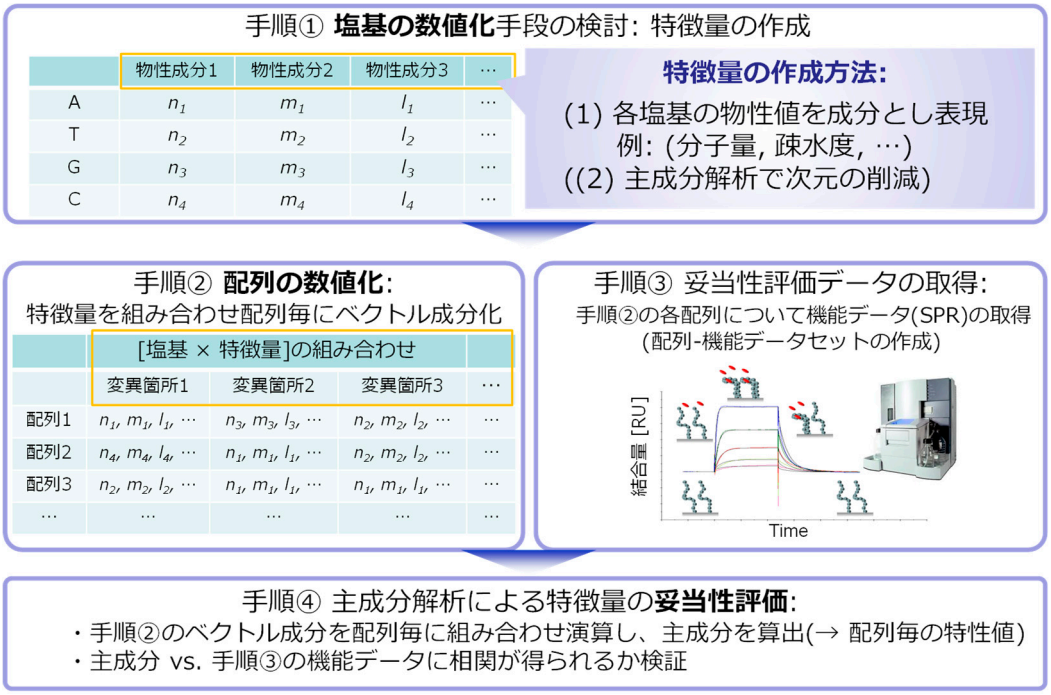
76　機械学習が道先案内する分子標的薬DNAアプタマー開発	梅津　光央

【目的】DNA アプタマーは化学合成が可能であり、有機化合物と同程度のコストで精製できる利点があると共に、抗体分子と同様な特定標的分子へ特異的に結合できる機能を有する。しかし、タンパク質と同様に目的の機能をもつ分子を見つけだす作業はそれなりの資金と時間が必要なため、リスク・リターン・達成速度のアンバランスが参入障壁になっている。我々は、タンパク質で培った機械学習と連携した進化分子工学による配列空間探索法を利用することによって、目的機能をもつ DNA アプタマーを取得できる確率を向上し作業時間を短縮できる手法の開発ができると考え、本研究では機械学習による高結合性 DNA 配列予測や設計を可能とするための塩基配列の数値化手法を検討した。

【方法】数値化用の単位構造としてヌクレオチド単位と塩基単位を抽出し、各々に対して既存ソフトウェアを用いて化学構造から 9 種の物性値を計算し、それら物性成分セットを特徴量とした。また、それら物性値を一度主成分解析し配列機能をより表現可能な特徴量として合成することも試み、それら特徴量と実測測定値との相関について主成分解析を行って検証した。

【結果】トロンビン結合アプタマーおよびその変異配列について、ソフトフェアより計算された 9 種の物性値そのものを特徴量として数値表現に展開した。ヌクレオチド単位と塩基単位の両パターンについて、変異箇所と物成分を組合せた 45 次元の数値で各アプタマーの塩基配列を表現し、配列 - 機能実測値データセットを作成した。そして、主成分解析により各配列が持つ 45 次元の成分を組合せ演算し、各配列の特性値として 16 次元の主成分を得て各主成分と結合量の相関解析を行ったところ、いずれも、異なる結合力を持つ配列が類似の主成分値に偏った結果となり、配列の特性表現性が不足することがわかった。そこで、それらを一度主成分解析し配列機能をより表現可能な特徴量の合成を試みた。塩基の物性成分に対し主成分解析を行い、塩基の特徴づけを強調した 4 成分の特徴量を新たに合成した結果、各配列の主成分値の偏りが減少して結合量との相関係数が向上し、特性表現性が良好な特徴量を構築できた。

DNA アプタマー配列予測・設計を可能とするための塩基配列の数値化



76. 機械学習が道先案内する分子標的薬 DNA アプタマー開発

梅津　光央

東北大学 大学院工学研究科 バイオ工学専攻 生体機能化学講座 タンパク質工学分野

Key words：DNA アプタマー，特徴量，機械学習

緒　言

特定の標的分子だけに優位に結合する分子標的医薬は現在の医薬品開発の主流であり、その構造フォーマットとして、ポリペプチドベースのペプチド・抗体分子や核酸ベースの DNA アプタマーが利用・検討されている。これらポリペプチドと DNA アプタマーの構造的特徴は、アミノ酸や核酸の一元的配列（アミノ酸配列や塩基配列）によって機能が一元的に決定し、生体外で生物の進化を模倣した進化分子工学（2018 年 ノーベル化学賞）を用いることによって、その配列情報を取得できる方法論をもつことである。しかし、それら配列の組み合わせ（配列空間）は膨大で、現在の進化分子工学でも配列空間の中から最適な機能をもつ分子を探索できる確率は高くなく、それ相当の資金と時間が必要である。

その中で我々は、近年、タンパク質において機械学習を組み合わせた進化分子工学を開発し、人工知能である機械学習が小規模な配列集団の情報から進化の方向性を示し、目的機能をもつアミノ酸配列を予測できることを実証し［1, 2］、この成功から、本手法は、タンパク質と同様に特定のモノマー単位の繰り返し構造で機能が決定する DNA アプタマーにも利用可能であると発想した。

DNA アプタマーは、抗体分子と同様に、特定標的分子へ特異的に結合して中和作用を示し、その標的結合性も抗体と匹敵する機能を創出し得る。抗体分子は現在の分子標的薬の主流であるが、その調製には高コスト体質の動物細胞を用いる必要があり、かつ、細胞で生産する生物材であるため精製コストも高い。その一方、DNA は化学合成が可能であり、有機化合物と同程度のコストで精製できる利点があり、世界的に開発意欲は高まっている。しかし、タンパク質と同様に目的の機能をもつ分子を見つけだす作業はそれなりの資金と時間が必要なため、リスク・リターン・達成速度のアンバランスが参入障壁になっている。我々は、タンパク質で培った機械学習と連携した進化分子工学による配列空間探索法を利用することによって、目的機能をもつ DNA アプタマーを取得できる確率を向上し作業時間を短縮できる手法の開発をできると考えた。

この配列空間探索法を DNA アプタマーに応用し、機械学習による高結合性 DNA アプタマーの配列予測・設計を可能とするためには、DNA 塩基配列を何かしらの数値で表現する必要がある。アミノ酸配列においては、各アミノ酸を化学的・物理的特性などに基づいた数値（特徴量）で表現する方法が数多く報告されている。その一方で、DNA アプタマーの領域における機械学習および特徴量に係る研究は発展途上であり、塩基に対する特徴量においては、有用な特徴量化について未だ報告例がないのが現状である。そこで本研究では、機械学習による高結合性 DNA 配列予測や設計を可能とするための塩基配列の数値化手法を検討した。いくつかのパターンで特徴量を作成し、モデル標的およびモデルアプタマーを用いて妥当性検証を行い、最も結果が良好であった特徴量を見出し低分子標的アプタマーへの応用可能性を検証した。

方法および結果

1. 塩基配列の数値化と検証

単位モノマー（ヌクレオチドもしくは塩基側鎖構造）について、化学構造に基づく物性値を計算した。物性値としては、計算機上で化学構造を基に計算可能なものであれば良く、物性値の個数も限定されない方針とした（例: 分子量、疎水度等）。得られた物性値のセットを各単位モノマーの成分とし、特徴量とした。特徴量の成分数（次元数）や物性表現の妥当性等によっては、場合に応じて主成分解析による次元の削減をこの段階で行うが、初期検討については、次元の削減は行わずに進めるものとした。そして、数値化を行いたいDNAアプタマー配列群について、塩基の特徴量を組み合わせ、配列毎に割り当て、ベクトル成分化した。その後、各配列について、実験的にDNAアプタマーの機能データの実測値を取得し、配列 - 機能データセットの作成を行った。データ取得後は、ベクトル成分を配列毎に組み合わせ演算して主成分を算出し、配列毎の特性値として、得られた主成分と機能実測値の相関性の解析し、配列 - 機能データセットとの相関解析を行い、特徴量の妥当性を検証した。

本検討のモデル配列として、既報のトロンビン結合アプタマー（TBA15）および変異配列を使用した。配列は5’-GGT TGG TGT GGT TGG - 3’（15 mer）である。TBA15は形成するG-quadruplex構造を足場としてトロンビンと結合し、その解離定数（KD値）は数nMであることが報告されている。今回、TBA15の塩基配列について、トロンビンとの結合特性に寄与する5箇所に変異を導入し、変異配列を作製した。変異配列を特徴づける情報は、① 変異の場所（5種類）と② 変異箇所に割り当てられた塩基が保持する特徴量（9種類の物性成分）との組み合わせによって表現することができる。従って、各変異配列は、5種類×9種類＝45種類の数値、つまり45次元のベクトル成分として表現した。この手法を用いて、特徴量の構成単位がヌクレオチドの場合と塩基構造の場合の2つのパターンについて変異配列の数値化（ベクトル成分化）を行い、TBA15およびその変異配列の各々を45次元の数値で表現した（図1）。

次に、これら配列の数値化の妥当性を評価するために、構築した単位モノマーの特徴量と配列の機能実測値との相関性を検討した。塩基配列の数値化結果と表面プラズモン共鳴測定で得られる標的結合量との相関性を評価することで作成した特徴量の妥当性を検討するために、塩基配列のベクトル成分を主成分解析し、得られた16次元の主成分について、各主成分と結合量の相関解析を行った。その結果、各主成分と結合量の相関解析を行ったところ、特徴量の作成がヌクレオチドの場合と塩基構造の場合のいずれも、異なる結合力を持つ配列が類似の主成分値に偏った結果となり、配列の特性表現性が不足することがわかった（図2）。

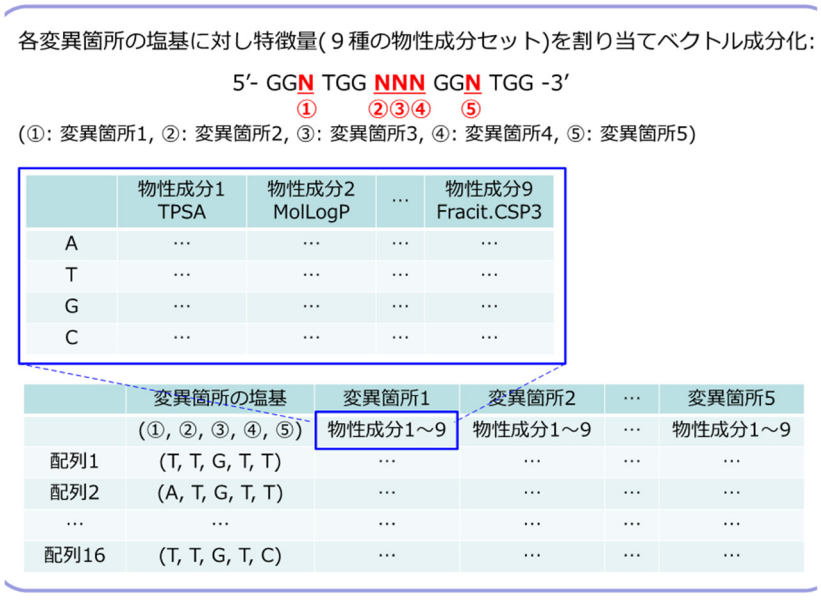


図1. 45次元のベクトル成分による特徴づけのフロー

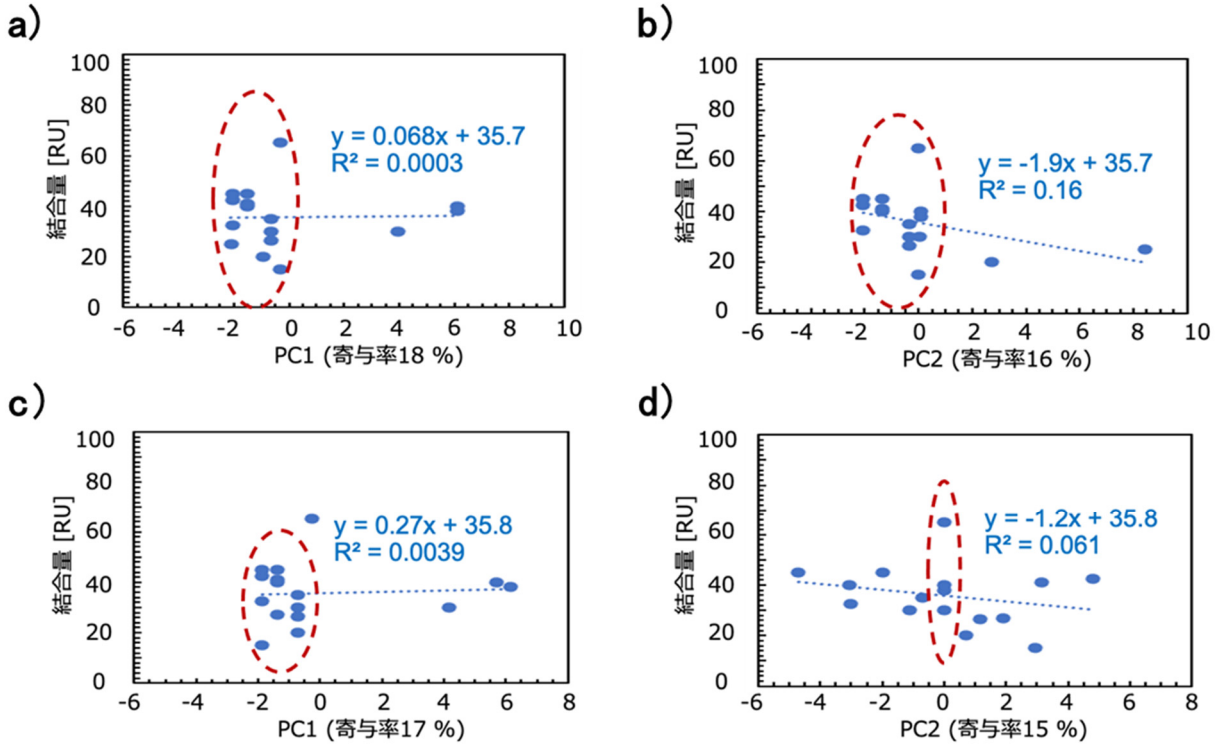


図 2.　主成分と実測機能値との相関

a）　ヌクレオチドの物性成分を得領とした場合の第 1 主成分。

b）　ヌクレオチドの物性成分を得領とした場合の第 2 主成分。

c）　塩基構造の物性成分を得領とした場合の第 1 主成分。

d）　塩基構造の物性成分を得領とした場合の第 2 主成分。

2.　特性表現性を高めるための特徴量の改良

ヌクレオチドもしくは塩基が持つ物性成分のセットを特徴量とした場合、配列の特性表現性が不足していることがわかった。そこで、単位モノマーが持つ物性成分をそのまま各配列の特徴量とするのではなく、単位モノマーの物性成分に対して主成分解析を行い、単位モノマーの特徴づけを強調した特徴量を新たに合成できないか試みた。そこで、単位元モノマーとして塩基構造を対象として、各塩基が持つ 9 種の物性成分について主成分解析を行い、塩基の種類と同数の 4 種類の主成分を得た。これにより、1 塩基を表現する次元を 9 種類から 4 種類に削減した形で集約することができた。次に、この 4 主成分のセットを各塩基構造の特徴量として、16 種類の塩基配列に対し、さらに主成分解析を行った。このようにして、各塩基構造の特徴量から構成される主成分を各配列について取得し、塩基構造の主成分を特徴量とした場合について、各配列の主成分と結合量との相関性が向上するか検証を行った。

2 つの主要な主成分について、それぞれ標的結合量との相関解析を行ったところ、塩基構造の主成分解析結果を特徴量とした場合、各配列の主成分値の偏りが解消し、配列特性の表現性が向上することがわかった（図 3)。さらに、ヌクレオチドの物性成分を特徴量とした場合、塩基構造の物性成分を特徴量とした場合、塩基構造の主成分解析結果を特徴量とした場合について、各配列の主成分 1 と結合量との相関係数を比較したところ、塩基構造の主成分解析結果を特徴量した場合の相関係数は向上しており、トロンビンとの結合特性をうまく表現できている可能性が示された（図 4)。

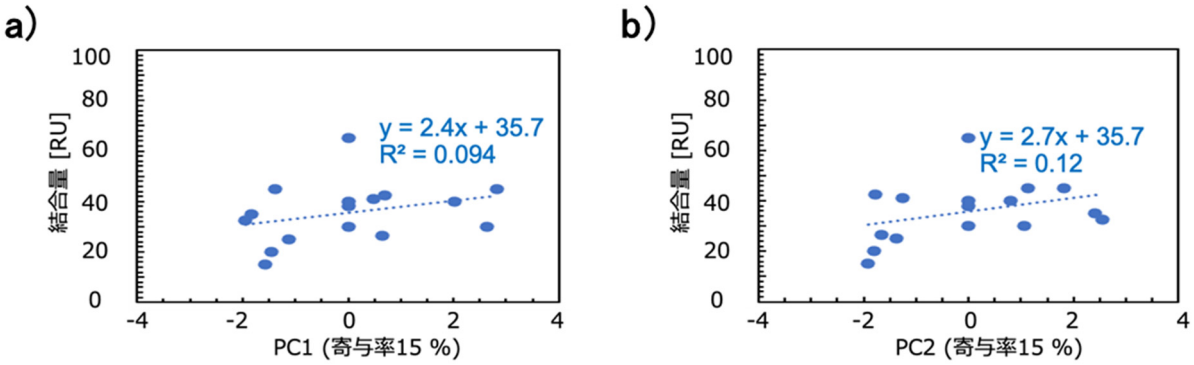


図 3. 塩基構造の物性成分を主成分解析して得られた特徴量をさらに主成分解析して得られた主成分と実測機能値との相関

a) 第 1 主成分。

b) 第 2 主成分。

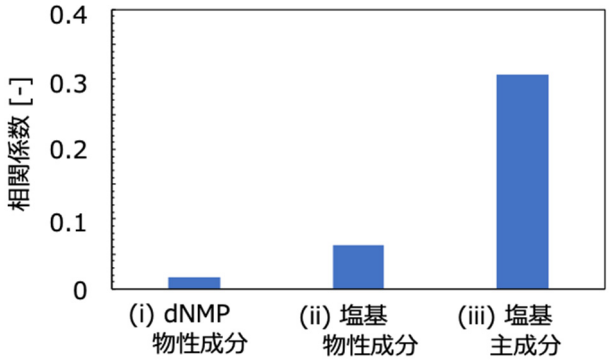


図 4. 特徴量の違いに対する相関係数の比較

i) ヌクレオチドの物性成分を特徴量とした場合。

ii) 塩基構造の物性成分を特徴量とした場合。

iii) 塩基構造の主成分解析結果を特徴量とした場合。

考　察

本研究では、機械学習による高結合性 DNA アプタマーの配列予測・設計のための DNA 塩基配列の数値化手法の構築を目指した。トロンビン結合アプタマーをモデルとし、塩基配列の数値化手法を検討した。その結果、塩基の化学情報に基づく物性成分を主成分解析し、得られた主成分を特徴量とした場合、各配列の主成分と結合量との相関が得られた。本研究で構築した塩基の特徴量および塩基配列の数値化手法を応用展開することで、機械学習により効率的かつ確実な高結合性アプタマーの取得が可能となると考えている。従来は、DNA アプタマーの領域における機械学習および特徴量に係る研究は発展途上であり、塩基に対する特徴量においては、有用な特徴量化について未だ報告例がなく機械学習のインプットデータの作成が困難であった。しかし、今回、塩基配列の機能との相関性を発揮する塩基の特徴量化が可能となったことで、DNA アプタマーの配列を数値表現し機能性との相関をラーニングデータとしてコンピュータに学習させること、その学習データ群に基づいて未知の配列についての機能を予測すること可能性がでてきたと考えている。

共同研究者・謝辞

本研究の共同研究者は、研究代表者が主宰するタンパク質工学分野の大学院生の小野寺真里氏、伊藤智之氏、河田早矢氏である。本研究にご支援賜りました上原記念生命科学財団に心より御礼を申し上げます。

文　献

1） Saito Y, Oikawa M, Nakazawa H, Niide T, Kameda T, Tsuda K, Umetsu M, Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins, ACS Synthetic Biology. Aug 13;7(9):2014–22. PMID: 30103599 DOI: 10.1021/acssynbio.8b00155

2） Saito Y, Oikawa M, Sato T, Nakazawa H, Ito T, Kameda T, Tsuda K, Umetsu M, Machine-learning-guided library design cycle for directed evolution of enzymes: the effects of training data composition on sequence space exploration, ACS Catalysis. 2021 Nov 19;11(23):14615–24. DOI: org/10.1021/acscatal.1c03753.