

【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た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、ソフトウ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

