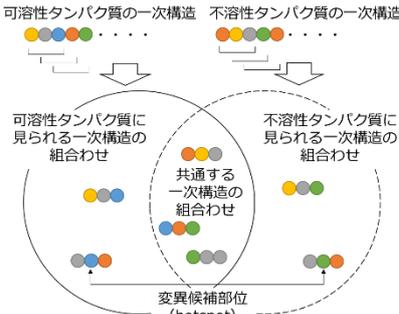


| | | | | |
|---|----------|------------------------------------|-------|--|
| 研究課題 (テーマ) | | タンパク質の一次構造に注目した機械学習による hotspot の探索 | | |
| 研究者 | 所属学科等 | 職 | 氏名 | |
| 代表者 | 生物工学科 | 助教 | 松井 大亮 | |
| 分担者 | 電子・情報工学科 | 准教授 | 榊原 一紀 | |
| 分担者 | 生物工学科 | 教授 | 浅野 泰久 | |
| 研究結果の概要 | | | | |
| <p>酵素の実用化には大量生産技術が必要だが、遺伝子組換えで他の微生物や動植物の酵素を生産する場合、沈殿して目的の酵素を生産できないことが多くある。浅野らは沈殿する植物酵素の一次構造を変えると、機能する可溶性酵素の生産量が大幅に上昇することを発見した。そこで本研究では、酵素化学工学講座 (浅野研) の実験データ (大腸菌生産を検討した 414 種類の高度好熱菌由来のタンパク質のタンパク質の可溶性生産に関する実験データ (可溶性 255 種類、不溶性 159 種類)) を基に榊原研で特徴的な配列 (hotspot) を見出し、一次構造の置換で可溶性酵素として生産する新手法の開発に挑戦した。</p> | | | | |
|  | | | | |
| <p>①形式概念分析 (対象の属性を比較し、その対象間の関連性を調査する分析法) を用いた分析 : 高度好熱菌由来の全てのタンパク質の先頭部分から、連続する 3 つのアミノ酸配列ごとに切り出した配列を比較した結果、可溶性を示す酵素には 726 種類、不溶性を示す酵素に 444 種類含まれることを明らかにした。同様に 4 つの連続するアミノ酸配列で切り出した配列を利用した結果、可溶性を示す酵素には 21846 種類、不溶性を示す酵素に 14509 種類含まれることを明らかにした。</p> | | | | |
| <p>②機械学習 (学習用データセットを使って訓練した後に、未知の例について正確を判断する手法) を用いた分析 : 上記で分類した配列と配列間の距離情報や分子間相互作用に関する情報などを用いて、機械学習 (一部の配列情報で機械に学習させた後に残りの配列から可溶性か不溶性の正答率を算出する) を行ったが、機械学習だけでは良好な結果が得られなかった。</p> | | | | |
| <p>③遺伝的アルゴリズム (適応度の高い解を優先的に選択して交叉・突然変異などの操作を繰り返しながら解を探索する手法) を用いた分析 : 必要なアミノ酸配列のみを抽出し、遺伝的アルゴリズムで解析した結果、最も正答率が高い抽出方法を見出し、その正答率は 0.68 であった。</p> | | | | |
| <p>④機械学習と遺伝的アルゴリズムを用いた分析 : 上記③の遺伝的アルゴリズムを用いた分析時に、上記②の機械学習を取り込むことで、正答率の向上を目指した結果、0.78 まで正答率を上昇させることに成功した。</p> | | | | |
| <p>今後より正答率を上昇させる必要があるが、この手法で一次構造の置換部位と種類を、数分程度で予測することが可能なため、簡便性にも優れ、これまで生産が困難であった多くの酵素の生産が可能となり、その特性の解明につながると考えられる。ひいては、高性能な酵素を見だし、化学工業や医療など産業で広く利用されることに貢献すると期待される。</p> | | | | |
| 今後の展開 | | | | |
| <p>既に見出しているアミノ酸置換で可溶性生産する酵素において、上記①~④の結果が含まれるかを検討する。そのアミノ酸置換の効果やアミノ酸の性質、二次構造情報などをデータ分析に追加し、改良する。さらにその他の沈殿する動植物由来酵素などのタンパク質の配列に、この手法で導き出した可溶性に特徴的な一次構造 (hotspot) に変え、タンパク質の生産量を比較する。</p> | | | | |