

| | | | | | | |
|----------|-----------------------------------|-------|----|----|----|-------------|
| L0021-12 | 単独でフォールドする構造ドメイン同定法の開発及びそのデータベース化 | | | | | |
| | 氏名 | 梅澤 祐貴 | 主査 | 黒田 | 副査 | 朝倉・野口・中村(徳) |

【 背景・目的 】

プロテオミクス研究において、巨大タンパク質を単独で構造を形成する比較的小さな「ドメイン」へと分割することで、迅速な解析を可能にする手法が広く用いられている。そのため、「単独で構造を形成するドメイン」(Independently foldable Structural Domain, ISD) の体系化とデータベース化は、近年のバイオインフォマティクスの重要な研究課題となっている。代表的なドメインデータベースとして、SCOP や CATH が挙げられる。しかし、両者は 1980 年代後半に開発され、ドメインの同定工程の一部を目視で行うという研究者の主観に頼る部分があるうえ、データベースの更新に時間を要する。さらに、両者とも、独立に構造形成することに重きを置いていないためか、その定量的な評価が成されていない。

そこで本研究では「単独で構造形成するドメイン」は、周囲のドメインと接触が少ないものと考え、タンパク質の構造 (原子座標) から ISD を自動的に同定する手法を開発し、計算結果をデータベース化しインターネットで公開した。

【 方法 】

①ドメイン境界決定法 はじめに、タンパク質の Cα原子間距離からドメイン間の Cα 接触密度が少ない残基位置を、ドメイン境界候補とした (図1)。次に、境界候補を用いて定義したドメイン間に形成される水素結合と疎水性クラスタ数が、先行研究で最適化した値より小さい場合、その境界候補を採用し ISD の境界とした。

②Cα間距離の最適化 SCOP/CATH に登録されている多ドメインから成るタンパク質で、個々のドメインが構造を保っていると実験的に検証された Autonomously Foldable Domain (AFD)である 51 タンパク質を「正解」として用いて、Cα間の接触を定義する距離を最適化した。

③ドメイン同定法の評価 ①②で確立した手法を用いて、PDB に存在する全長タンパク質(237,914 タンパク質)に対して構造ドメインの同定を行った。その結果を、他のドメイン同定法及び SCOP や CATH に登録されているドメインと比較した。

【 結果および考察 】

Cα間距離の最適化 接触距離を 13Å に設定した正答率が 49/51 (96.0%)と最大で、ISD 境界と AFD 境界との誤差は平均で 2.3 残基であった。

ドメイン同定法の評価 PDB 上の全タンパク質に対する構造ドメイン同定の結果、101146 ドメインが検出されそのうち 22137 体が AFD であった。この結果は他のドメイン同定ツールと比較しても総数や検出率において優れていた (表 1)。

代表配列比較 本研究により同定された AFD と、SCOP/CATH に登録されている AFD の代表配列を比較した結果、双方のグループの 85%以上と重複があった (図 2)。また、SCOP/CATH では同定されていない独立に構造を形成する構造ドメインが多く存在する点、工程を完全に自動化できる点が本同定法の特長である。

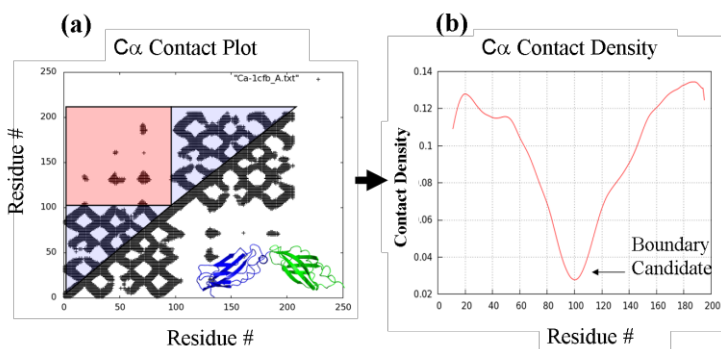


図 1. Cα 接触密度のグラフ (a)青/赤：分割した際に保存/切断される Cα接触 (b)各残基で分割した際に失われる Cα 接触の密度

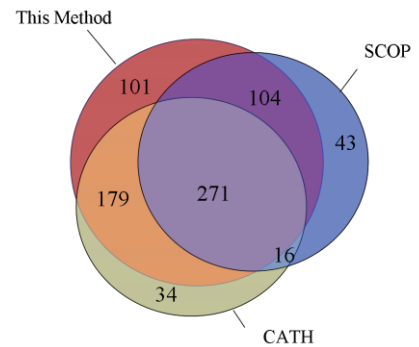


図 2. SCOP/CATH との AFD 配列グループ比較結果

表 1. 手法別の構造ドメイン同定結果

| | This Method | DomainParser | DOMAK |
|-------------------|-------------|--------------|--------|
| Number of domains | 101146 | 128929 | 118193 |
| Number of AFDs | 22137 | 17521 | 12928 |
| Fraction of AFDs | 21.9% | 13.6% | 10.9% |