

構造ドメイン境界予測法に用いるドメインリンカーデータベースの作成		
黒田研究室	03251069	古山 智裕

【背景・目的】

タンパク質のドメイン領域予測法の開発及びそれを支援するデータベース作成が現在注目されている。その背景にタンパク質の立体構造情報蓄積の重要性とそれを担うプロテオミクスの発展がある。プロテオミクス研究では大量のタンパク質を効率良く解析するために分子量の大きなタンパク質を構造的に安定なドメイン（構造ドメインという）単位に分離し解析する事が多く、タンパク質の構造ドメイン領域をより正確に予測出来れば解析効率が高まると期待されている。

そのドメイン領域予測法の1つに構造ドメイン間領域である「ドメインリンカー」を予測する手法がある。予測法の開発には目的に沿ったデータベースが必要となるが、ドメインリンカーデータベースは一般公開されていないため開発者が独自に作成しなければならず手間が掛かっていた。

そこで本研究では、構造ドメイン境界予測法の開発を支援するドメインリンカーデータベースの作成を目的とし、独自の定義に従いドメインリンカー及びドメインを自動で選出するアルゴリズムを開発した。

【方法】

データベース作成手順：

- ①SCOP からマルチドメインタンパクを選出する。
- ② (DSSP) 二次構造予測プログラムを用いドメイン間で水素、S-S 結合をしているものを除外。
- ③疎水性クラスター（本研究にて定義）の有無で疎水性相互作用の有無を計算し、該当を除外。
- ④データベース中の冗長性を破棄するため、配列類似性によるグループ分類を行う。

データベースの評価：

最も影響力の大きい疎水性相互作用計算と本研究での構造ドメイン選出結果の妥当性を調査する。

- ・ 疎水性相互作用計算のパラメータとドメインリンカーの個数の関係を調べる。(図 1)
- ・ BLAST 検索を用いて、「PDB に登録済みの構造的に独立であるタンパク質」と配列類似性が高いドメイン（類似構造ドメインと定義）の個数や割合を調査する。(表 1)

【結果と考察】

②の過程迄で 9037 個のリンカー候補を得た。疎水性相互作用計算にて DISTANCE 値を 5.0Å、要素数を 0 個に設定したところ 2724 個のドメインリンカーを得た。これを nrPDB にて BLAST pvalue 10e⁻⁷ の分類を行ったところ 206 グループのドメインリンカーを含むタンパク質を得た。

(図 1) 疎水性相互作用のパラメータは要素数の影響は小さく、距離によるところが大きい。DISTANCE 値は 5.0Å~7.5Åにてある程度のリンカー数の確保と性質の安定性を実現できると考えられる。(表 1) 本研究のデータベースは元となる SCOP よりも類似構造ドメインの割合が高くなる事を示唆しており本研究のプロトコルが有意であると言える。

表 1 類似構造ドメインを含む割合

距離	類似構造ドメイン数	ドメイン総数	割合
4.663	1814	5668	32.00
5.000	1709	5012	34.10
5.500	1628	4628	35.18
6.000	1573	4414	35.64
6.500	1523	4243	35.89
7.000	1437	3941	36.46
7.500	1291	3531	36.56
8.000	958	2734	35.04
SCOP	3479	23815	14.61

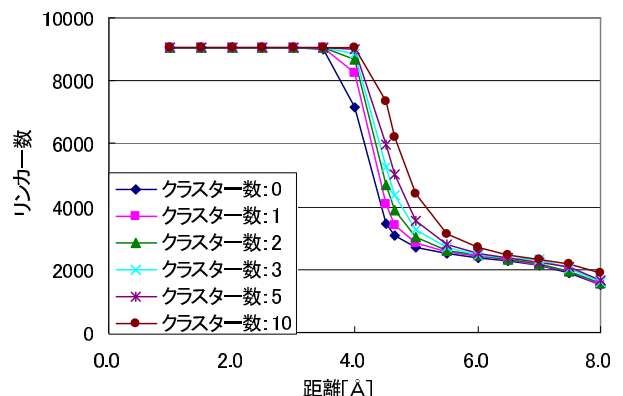


図1 疎水性相互作用計算結果