

**【 背景・目的 】**

生命現象解析の対象は、「遺伝子」から「タンパク質」へと変化しつつある。しかし、タンパク質の中には、分子量が大きく、発現・精製等の実験手法の条件検討が困難なものも存在する。そこで、実験条件の検討が比較的容易なタンパク質の構造単位「構造ドメイン」に分割する手法がとられる。構造ドメインに分割する手法で一般的に利用されるのが限定分解という実験手法であるが、多量のタンパク質が必要な上に、時間や労力も必要される。そこでアミノ酸配列情報から構造ドメイン領域を予測する手法が開発されてきた。

本研究では、より多く独立して構造を取りうるドメインを含んだ「構造ドメイン」データベースの作成を行う。このデータベースでは、マルチドメインタンパク質を SCOP (The Structural Classification Of Protein database) から選出し、そのマルチドメインタンパク質が含むドメイン間に働く相互作用についての定義を検討し、構造ドメインを定義する。

**【 手法 】**

データベース作成手順( fig. 1 ) :

- (I) SCOP (version 1.75) よりマルチドメインタンパク質を取得する。
- (II) 選出されたタンパク質を SCOP の定義でドメインごとに分割する。
- (III) そのドメインの中で PDB (Protein Data Bank) に登録されている結晶化されたタンパク質全長との配列相同性の高いものを独立して構造を取り得るドメインと定義する。
- (IV) 1つのタンパク質内に連続して独立して構造を取り得るドメインが 2 つある場合、ドメインペアを作成する。また1つのタンパク質内に選出されなかったドメインが 2 つある場合もドメインペアを作成する。
- (V) 各ペアのドメイン間相互作用情報として、PDB より疎水性相互作用情報、また PDB と二次構造予測プログラム DSSP を用いて水素結合・ジスルフィド結合情報を取得する。
- (VI) 作成した二つのペアにおいて、各相互作用情報の距離・結合力・数について独立であると定義される AFD と Non AFD の検出感度の差が最大化される場所を構造ドメインと定義した。
- (VII) SCOP より選出したマルチドメインタンパク質全体に対しても、ドメイン間結合情報を取得する。
- (VIII) マルチドメインタンパク質の中から「構造ドメイン」の定義を用いて「構造ドメイン」を選出し、データベース化する。

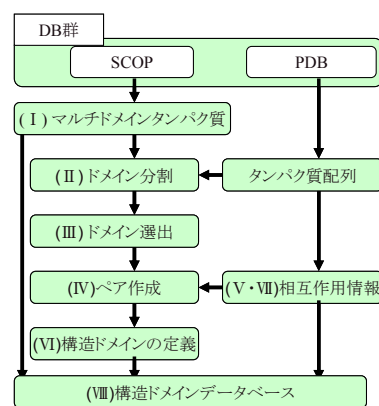


fig.1 データベース作成の手順

**【 結果・考察 】**

SCOP よりマルチドメインタンパク質を 16974 個選出した。得られたタンパク質を SCOP の定義にしたがいドメイン分割し、37051 個のドメインを得た。得られたドメインの中で PDB に登録されたタンパク質全長と配列相同性をもつドメインを 5689 個得た。それを独立して構造を取りうるドメインとし、類似配列ごとにグループ分類することで 107 のペアを作成した。同時に独立でないと定義されたドメインもグループ分類し、1132 のペアを作成した。

そこから以下の定義を決定した。

- ・ -0.5cal の水素結合を 1 つ未満
- ・ 5.0Å の疎水性クラスターを 2 つ未満
- ・ ジスルフィド結合を 1 つ未満

表 1 では、定義を用いて SCOP より得たマルチドメインタンパク質から「構造ドメイン」を選出した。類似配列のグループ分類による結果も含めて、より独立して構造を取りうるドメインを多く含む「構造ドメイン」を検出できた。

表 1 独立して構造を取り得るドメインを含む割合

	独立構造グループ数	グループ総数	割合
SCOP	292	1199	24.35
Our Method for SCOP	109	272	40.07

	独立構造ドメイン数	ドメイン総数	割合
SCOP	5689	37051	15.35
Our Method for SCOP	1925	4928	39.06