

**【背景・目的】**

タンパク質の構造と機能の解明を目的とするプロテオミクスにおいて、解析対象であるタンパク質の多くは巨大であり、発現・精製や実験条件の検討が困難な場合が多い。そこで巨大タンパク質の解析については、比較的サイズが小さく単独での発現・精製が容易なタンパク質の構成単位、「構造ドメイン」を計算的手法により同定し、これらを解析対象とする方法が広く用いられている。この予測手法の開発において、学習データとなる構造ドメインデータは不可欠である。しかし、代表的な構造ドメインデータベースである SCOP や CATH において、同一のタンパク質に対して異なるドメイン定義が行われている場合がある。これは「構造ドメイン」の定義に関して、コンセンサスを得た定義が今もなお存在しない為である。

そこで本研究では、各ドメインが近傍のドメインと持つ相互作用を原子座標から算出し、定量的に評価することで独立して構造を取り得るドメインの定義を検討した。また、この定義を満たすドメインを「構造ドメイン」として、Web 上でデータベースを公開した。

**【方法】**

- (1) SCOP より多ドメインタンパク質を選出し、SCOP の定義によるドメイン配列を得た。
- (2) 得られたドメインの中で、既に単独で結晶構造が登録されているタンパク質と高い配列相同性を持つものを **Autonomously Foldable Domain(AFD)** とし、それ以外を **NonAFD** とした。
- (3) AFD / NonAFD がその近傍残基との間で形成する疎水性クラスタ、水素結合およびジスルフィド結合を検出した。
- (4) 各相互作用の個数の閾値を網羅的に変化させ、Youden's Index(YI, 図 1) 値が最大となる条件を構造ドメインの定義とした。
- (5) 得られたデータベースを Web 上で提供する為のインターフェースを構築した。

**【結果・考察】**

SCOP に登録されている 16,695 個の多ドメインタンパク質より 37,035 個のドメインを得た。そのうち 6,049 ドメインは AFD、30,986 ドメインは NonAFD であった。次に、AFD と NonAFD についてそれぞれのドメインが形成する近傍残基との相互作用数を調べたところ、AFD の形成する水素結合および疎水性クラスタの数は NonAFD と比べ有意に少なかった。そこで、これらの相互作用数を基準とし、構造ドメインの定義を検討した。各相互作用数の閾値を変化させ、YI を最大化する閾値の組み合わせを網羅的に調べた結果、「水素結合（主鎖-主鎖）5 個以下、水素結合（側鎖-主鎖）10 個以下、水素結合(側鎖-側鎖)7 個以下、疎水性クラスタ 8 個以下」の定義を得た。この定義により、SCOP から 14,356 ドメインを「構造ドメイン」として選出し、データベース化した。

我々の構造ドメインの定義は、既に結晶化されている AFD 中 85%に当たる 5129 ドメインがこの定義を満たすことから、独立して構造を取るドメインの性質を良く反映していると考えられる。更に、構造ドメインデータベースとしての妥当性の指標として、データベース中に含まれる AFD の割合を算出したところ、既存のデータベース SCOP と比べ AFD の割合が 2 倍以上に向上していた。(表 1) また本データベースに含まれる AFD 以外のドメインについても、未だ結晶構造が登録されていないが、我々の定義を満たすことから独立して構造を取るドメインであると推察される。

以上、本研究で作成したデータベースを Web 上で公開した。Web 上においては我々の定義のみならず、ユーザーの意図に応じて定義を変更できるようにした。(図 2, URL : <http://domserv.lab.tuat.ac.jp>)

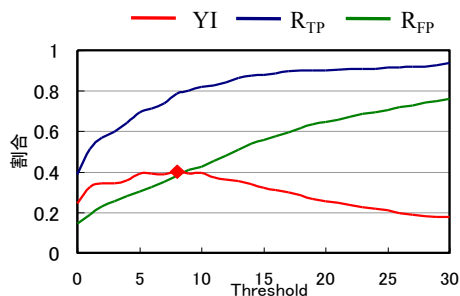


図 1 疎水性クラスタの Youden's Index (YI)  $YI = R_{TP} - R_{FP}$ ,  $R_{TP}/R_{FP}$  はそれぞれ閾値を満たす代表 AFD / NonAFD 数の割合を表す。

◆は YI の最大値を表す。

表 1. 既存データベースとの比較

	AFD数	総数	AFD割合
SCOP	6,049	37,035	16.3%
Our DB	5,129	14,356	35.7%

STRUDOM

Domain Boundary Definition	<input checked="" type="radio"/> SCOP <input type="radio"/> CATH <input type="radio"/> All		Select domain boundary definition.
Hydrogen Bond (main chain)	Strength	# of Hbonds	Strength and number of Hbonds calculated by DSSP (Kabsch, et al.)
	-0.5 (kcal/mol)	6	
Hydrogen Bond (side chain)	Side<->Main	Side<->Side	Number of Side<->Main or Side<->Side bonds. calculated by HBPLUS (L.K.McDonald, et al.)
	5	5	
Hydrophobic Cluster	Distance	# of Clusters	Select distance threshold and number of Clusters calculated from PDB coordinates
	5.0 (A)	10	

図 2 Web 上での提供