

1. はじめに

タンパク質の機能発現部位の検出法については、酵素の凹型形状の検出[1]や既知の活性部位構造との形状比較[2]などのシステムティックな方法が検討されている。我々は、同じ機能を持つタンパク質間において、三次元的に相同なアミノ酸残基群を抽出するという機能発現部位の検出法を提案し、これまでその妥当性や精度向上を検討してきた[3]。

今回、アミノ酸配列の類似性から幾つかのグループに階層に分類されているセリンプロテアーゼを対象に、我々の検出法がどのレベルで有効かを検討し、方法論的な可能性と限界を明らかにしたので報告する。

2. 三次元配置が相同なアミノ酸残基群の検出法

三次元配置の相同なアミノ酸残基群を探る考え方を簡単に説明する。まず、二つのタンパク質に属する同種のアミノ酸残基の対をノードとし、全ての二つのノードにおいて、対応するアミノ酸残基の対間の α 炭素距離が一定の誤差範囲(閾値)である時、そのノード間にエッジを作ることによりグラフを作成する。このグラフの最大完全部分グラフ(全てのノードがエッジで結ばれている部分グラフ)が、三次元配置が重なる最大のアミノ酸残基群を示す。なお、検出システムは高速に最大完全部分グラフを見いだす Carraghan のアルゴリズム[4]を用いた。

3. セリンプロテアーゼ

セリンプロテアーゼは機能等に関する研究が進んでいるタンパク質である。Wallance ら[5]は、Ser-His-Asp を持つタンパク質を集め、そのフォールディングの類似性からまず4つのグループに大きく分類し、さらにアミノ酸配列の類似性に基づきサブグループに分類した。本研究では、これらのサブグループとグループに属するセリンプロテアーゼを対象に、3種類の異なるタイプのペアを取り出し、それらに対して本検出法を適用した。

4. 実験

4.1 同じサブグループのタンパク質

α -chymotrypsin (Bos taurus) と α -chymotrypsin (bobine pancreas) は β -sandwich trypsin-like fold serine proteases group の subgroup 1a に属しているアミノ酸配列の極めて類似したタンパク質である。それらは 245 残基からなり、H-57 と D-102、S-195 が活性中心である。我々は最初に閾値 0.2 Å で検討を始めたが、24 時間かかっても実行が終了しなかったため、閾値を小さくし、0.1 Å、0.05 Å、0.02 Å で比較した。その結果を表 1 に示す。

表 1 同じサブグループのタンパク質を用いた検出結果

Threshold value (Å)	Number of nodes	Calculation time* (s)
0.02	6	68
0.05	10	73
0.10	24	92

*PowerPC G4 1.25GHz

ノード数が 10 個程度の場合を妥当な抽出結果と考え、閾値 0.05 Å によって得られた 10 個のアミノ酸残基を α -chymotrypsin (Bos taurus) の構造の上に表示した(図 2)。これらのアミノ酸残基は活性中心となるクレフトの近傍に位置しているが、機能発現のアミノ酸残基を検出したとは言えない。

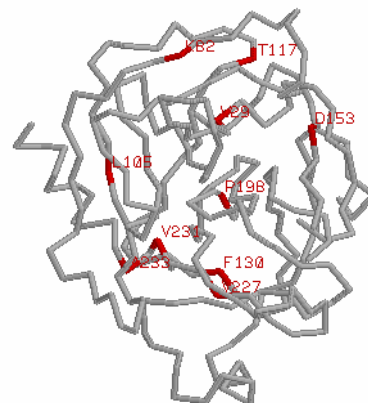


図 2 α -chymotrypsin (Bos taurus) 上の予測される残基

4.2 同じグループで異なるサブグループのタンパク質

α -chymotrypsin (*Bos taurus*) と α -lytic protease (*Lysobacter enzymogenes*) は β -sandwich trypsin-like fold serine protease に属しているが、前者は subgroup 1a、後者は subgroup 1b と異なるサブグループに属している。それらは両方とも 245 残基からなるが、アミノ酸配列の類似性は 30% 以下である。閾値を 1.0 Å、0.5 Å、0.2 Å で検討した結果を表 2 に示す。

表 2 同じグループで異なるサブグループのタンパク質を用いた検出結果

Threshold value (Å)	Number of nodes	Calculation time* (s)
0.2	6	50
0.5	11	72
1.0	17	170

*PowerPC G4 1.25GHz

同様に、閾値 0.5 Å によって得られた 11 個のアミノ酸残基を α -chymotrypsin (*Bos taurus*) の構造の上に表示した (図 3)。

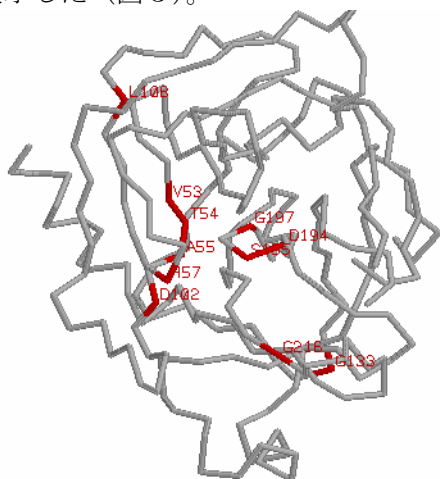


図 3 α -chymotrypsin (*Bos taurus*) 上の予測される残基

これらのアミノ酸残基は、数個のものを除いて活性中心のクレフトに位置しており、H-57 と D-102、S-195 を含んでいるため、検出に成功したといえる。

4.3 異なるグループのタンパク質

α -chymotrypsin (*Bos taurus*) と subtilisin Carlsberg (*B. subtilis*) は前者が β -sandwich trypsin-like fold serine protease に、後者が alternating α - β subtilisin-like fold serine protease にと、フォールディングタイプの異なるグループに属している。前者は 245 残基からなり、後者は 274 残基からなり、アミノ酸配列がかなり異なる。

閾値を 1.0 Å、0.5 Å、0.2 Å で検討した結果を表 3 に示す。

表 3 異なるグループのタンパク質を用いた検出結果

Threshold value (Å)	Number of nodes	Calculation time* (s)
0.2	4	142
0.5	6	229
1.0	9	703

同様に、閾値 1.0 Å によって得られた 9 個のアミノ酸残基を α -chymotrypsin (*Bos taurus*) の構造の上に表示した (図 4)。これらのアミノ酸残基は活性中心となるクレフトとは異なる位置にあり、機能発現のアミノ酸残基を検出したとは言えない。

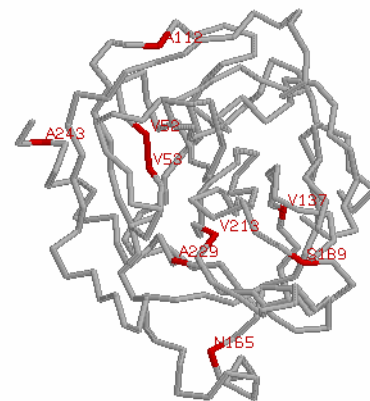


図 4 α -chymotrypsin (*Bos taurus*) 上の予測される残基

5. おわりに

本研究の結果、我々の方法は、(1)対象とする二つのタンパク質が構造的に似すぎている場合その検出がうまくいかないこと、(2)フォールディングのタイプが同じである程度アミノ酸配列が異なっている場合その検出がうまくいくこと、(3)アミノ酸配列が異なってもフォールディングのタイプが違う場合はその検出がうまくいかないことが分かった。

参考文献

- [1] M. Hendlich, F. Rippmann, G. Barnickel, *J. Mol. Graphics Mod.*, 15, 359-363 (1997).
- [2] M. Rosen, S. L. Lin, H. Wolfson, R. Nussinov, *Protein Engng.*, 11(4), 263-277 (1998).
- [3] 中山、矢口、真栄城、第 26 回情化学討論会講演要旨集、129-130 (2003).
- [4] R. Carraghan, P. M. Pardalos, *Operations Res. Lett.*, 9, 375-282 (1990).
- [5] A. C. Wallace, R. A. Laskowski, J. M. Thornton, *Protein Science*, 5, 1001-1013 (1996).