

抗体のアミノ酸配列データベースの統計的解析
 (広島市大院・情報) ○脇田 淳永, 藤原 久志, 石渡 孝
 Statistical analysis of an antibody sequence database
 (Hiroshima City Univ.) ○A.Wakita, H.Fujiwara, and T.Ishiwata

【序】分子認識に関わるアミノ酸配列は特異性が高いため、従来の文字配列解析法（類似性検索に基づく）で取り扱うことが難しい。これに対し、美宅らは、アレルゲンのエピトープ（抗原決定基）の特徴抽出において、統計的解析が有用であることを示した¹⁾。本研究では、こうした統計的解析を抗体分子に適用し、その分子認識部位のアミノ配列の特徴抽出を目指すこととした。

【抗体のアミノ酸配列データの取得】ヒト免疫グロブリン G (Immunoglobulin G, IgG) H 鎖のアミノ酸配列情報ファイル human.ig.hc (Kabat データベース²⁾の一部) を、European Bioinformatics Institute の FTP サーバー³⁾ から取得した。IgG H 鎖のアミノ酸配列には、超可変領域とよばれる多様性の高い比較的短い配列が 3 か所存在する。超可変領域は抗原認識に関与すると考えられ、実際に、相補性決定領域 (CDR) 1、CDR2、CDR3 ともよばれる⁴⁾。本研究では特に高い多様性を持つ H 鎖の CDR3 (CDRH3) を研究対象とした。

【CDRH3 アミノ酸配列の抽出】取得した IgG H 鎖の情報の一例を図 1 に示す。一つのデータの情報は”AA TABLE”から始まり、”//”で終わる。IgG H 鎖のアミノ酸配列は、ファイル内の AA SEQUEN の欄に 3 文字表記で書かれている。ダウンロードしたファイルには、1173 個のデータが含まれていた。

```
AA TABLE : HUMAN HEAVY CHAINS FAMILY I
NUC TABLE : CODONS OF HUMAN HEAVY CHAINS FAMILY I
:
:
KABAT NUM : 0|1|2|3|...|110|111|112|113
AA SEQUEN : ---|GLN|VAL|GLN|...|THR|VAL|SER|SER
:
//
AA TABLE : HUMAN HEAVY CHAINS FAMILY I
:
:
```

図 1 : IgG H 鎖の情報の一例

これらのデータから、欠損のない完全な CDRH3 アミノ酸配列を抽出するために、次の三つの基準を設けた：①CDRH3 近傍の保存度の高いシステインとグリシン（それぞれ Kabat 番号 92 と 104）が存在すること；②これらのシステインからグリシンまでの配列（CDRH3 を含む）に空白がないこと；③TRM（偽遺伝子）がないこと。さらに、冗長性を取り除くために、複数の同一配列は一つの配列として取り扱い、複数の類似配列（同一長でアミノ酸の違いが三つまで）は、相違アミノ酸の重み付け処理を含めつつ、一つの配列としてまとめることとした。

以上の処理を行う Perl スクリプトを実装し、ActivePerl（バージョン 5.16.3.1603）を実行環境として、配列長 3～19 の CDRH3 アミノ酸配列 338 個を抽出した（図

2)。なお、抽出処理が複雑になることを避けるために、配列長 20 以上のデータ 59 個は取り扱わないこととした。図 2 では、配列長 10～14 にピークが見られる。この配列長分布は、先行研究⁵⁾と同様の傾向を示しており、抽出処理の妥当性を支持している。

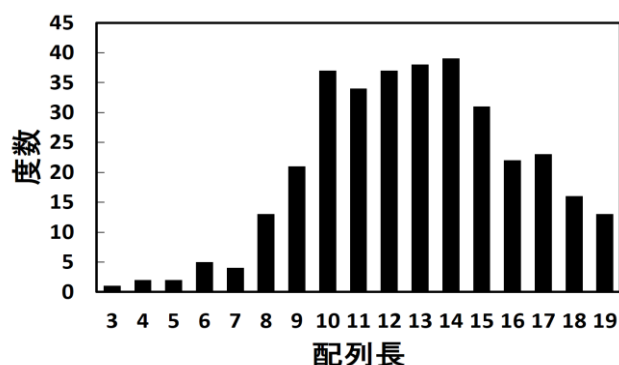


図 2：ヒト CDRH3 の配列長の分布

【CDRH3 アミノ酸配列の統計的解析】CDRH3 はループを形成しており⁶⁾、その中央付近は分子認識に大きく関わると考えられる。そこで、分子認識に関与するアミノ酸を見出すために、配列長 10～19 のヒト CDRH3 アミノ酸配列について、配列全体と中央（三残基）における各アミノ酸出現頻度を比較した（図 3）。なお、出現頻度の計算は、C 言語プログラムを実装・実行して行った。

まず初めに、ヒト CDRH3 では中央も含めて全体的にチロシンの出現頻度が高いことが分かる。これに関連して、セリンとトレオニンでは中央での出現頻度が高くなることが分かる。これらのことより、ヒト CDRH3 の分子認識ではヒドロキシ基が重要であると考えられる。

また中央も含めて全体的にグリシンの出現頻度が高いことは、CDRH3 ループに可塑性を与えて、エピトープとの適切な位置関係を探索することに役立っていると考えている。

さらに種々の統計的解析を行うことで、抗体の分子認識とアミノ酸物性との関係について調べる予定である。

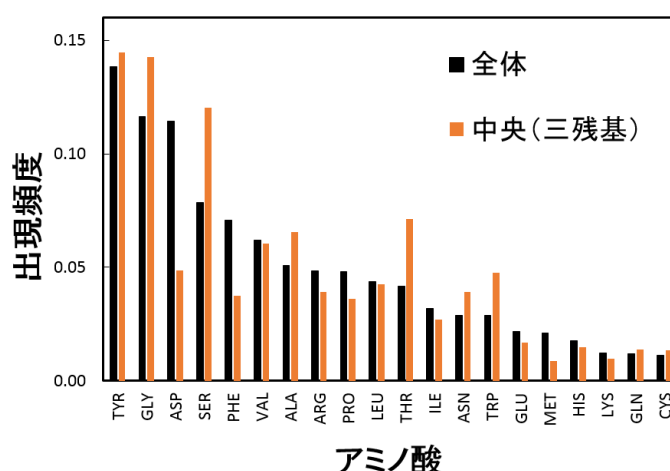


図 3：ヒト CDRH3 における各アミノ酸の出現頻度

【参考文献ほか】

- 1) N. Asakawa, N. Sakiyama, R. Teshima, and S. Mitaku, *J. Biochem.*, **147**, 127 (2010).
- 2) G. Johnson and T. T. Wu, *Nucl. Acids Res.*, **28**, 214 (2000).
- 3) <http://ftp.ebi.ac.uk/pub/databases/kabat>
- 4) Peter Wood （訳：山本一夫），免疫学 （東京化学同人，東京，2010）。
- 5) T. T. Wu, G. Johnson, and E. A. Kabat, *Proteins*, **16**, 1 (1993).
- 6) A. Liljas et al. （訳：田中勲，三木邦夫），構造生物学 （大洋社，東京，2012）。