

情報科学の生命分子科学への応用：

抗体のアミノ酸配列データベース解析方法の探索

(広島市大院・情報) ○脇田 淳永, 藤原 久志, 石渡 孝

Application of information sciences to biomolecular sciences:

Exploration of analysis methods for an antibody sequence database

(Hiroshima City Univ.) ○A. Wakita, H. Fujiwara, and T. Ishiwata

【序】我々は、情報科学の生命分子科学研究への応用を目指して、種々の試みに取り組んでいる。本研究では、統計的解析を抗体分子に適用し、抗原認識に関わるアミノ酸配列の特徴抽出を目指している¹⁾。

【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。本研究では特に高い多様性を持つ H 鎖の相補性決定領域 (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 鎖のアミノ酸配列情報ファイルから、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¹⁾：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った。

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 3~31 の CDRH3 アミノ酸配列 418 個を抽出した (図 1)。図 1 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⁴⁾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。

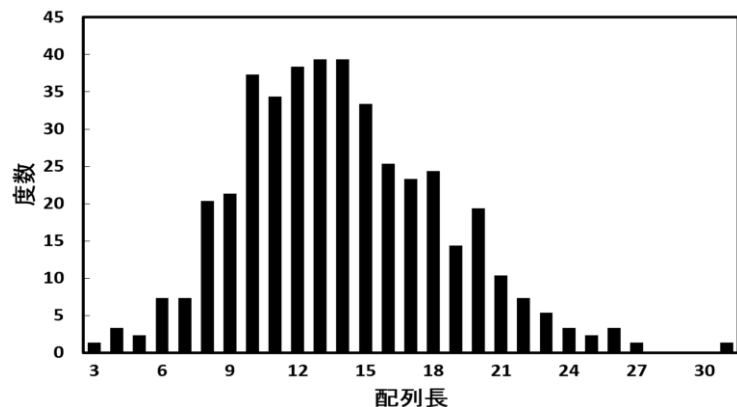


図 1：ヒト CDRH3 の配列長の分布

【Shannon Entropy による解析】配列長 10 以上の CDRH3 アミノ酸配列について、配列長ごとにアミノ酸の位置による多様性を次式のシャノンエントロピー V_i を用いて調べた。ここで i はアミノ酸の位置、 M は位置 i に出現するアミノ酸の数である。

$$V_i = - \sum_{j=1}^M P_{ij} \log_2 P_{ij}$$

図2に CDRH3 アミノ酸配列を解析した例（配列長 11）を示す。図3を含めて、解析した全ての配列長（10から15）において、C末端側三残基のシャノンエントロピーは低く（＝保存度が高い）⁵⁾、それ以外の位置のシャノンエントロピーの値は3.5程度で多様性が高かった。そこで、CDRH3 アミノ酸配列（配列長 10から31）で、分子認識に関わるのはC末端三残基を除いた部分と考え、これを抽出・解析することにした。

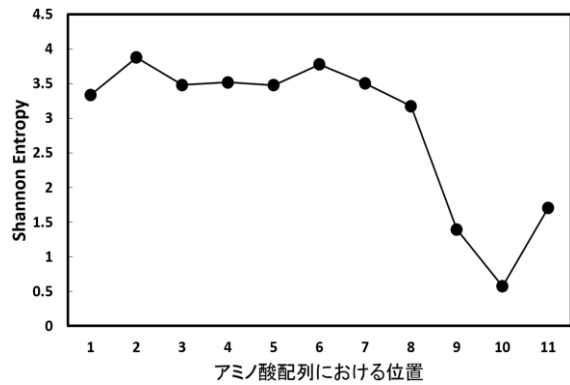


図2：配列長 11 のアミノ酸配列における各位置の Shannon Entropy

【CDRH3におけるアミノ酸出現頻度の解析】CDRH3におけるアミノ酸の出現頻度解析を、配列全体とC末端三残基を除いた配列で行った（図3および4）。

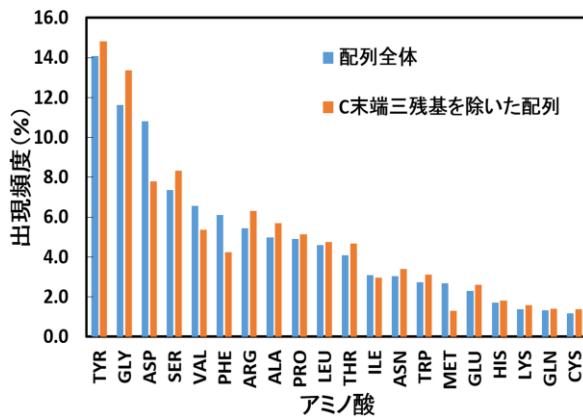


図3：ヒトの CDRH3 における各アミノ酸の出現頻度

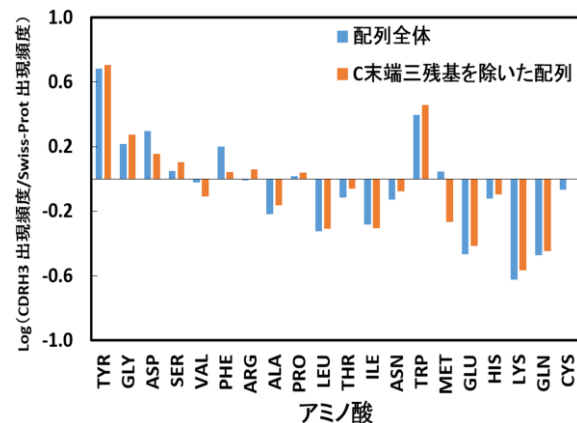


図4：各アミノ酸の出現頻度（図3）の Swiss-Prot における出現頻度との比較

図3および4より、ヒドロキシ基を含む残基としてチロシン、セリンの出現頻度が高い。さらに、これらとトレオニンを含めて「C末端三残基を除いた配列」において、より出現頻度が高くなっている。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と考えられる¹⁾。一方、メチオニンはC末端三残基での出現頻度が高いことが分かる。今後は各アミノ酸配列と他の情報（抗原、掲載論文情報など）とを照合しつつ、抗体認識部位の特徴抽出を目指す予定である。

【参考文献ほか】

- 1) 脇田淳永, 藤原久志, 石渡孝 第8回分子科学討論 3P088 広島 (2014)
- 2) G. Johnson and T. T. Wu, *Nucl. Acids Res.*, **28**, 214 (2000).
- 3) <http://ftp.ebi.ac.uk/pub/databases/kabat>
- 4) T. T. Wu, G. Johnson, and E. A. Kabat, *Proteins*, **16**, 1 (1993).
- 5) L. Wu et al, *J. Immunol.*, **188**, 327 (2012).