3U−2　アミノ酸配列の構造解析

小野　敬規、滝口　伸雄、小谷　善行、西村　恕彦
（東京農工大学　工学部　数理情報工学科）

1．はじめに

タンパクは、２０種類のアミノ酸の１次元結合で構成されている。この２０種類のアミノ酸それぞれを一つの記号で表すと、タンパクの構成を記号列で表現できる。この記号列をアミノ酸配列と呼ぶ。このアミノ酸配列に自然言語におけるような何らかの構造がないかを調べることは、興味のあることである。アミノ酸配列の構造を発見する手法は様々存在するが、本研究では、２文字連接を用いた統計的手法を使った。

2．２文字連接

２文字連接による手順は次の通りである。

①アミノ酸配列の中で、ある二つのアミノ酸がとなりあう頻度を測定する。
②それが特徴のあるものであれば別の記号に置き換える。
③置き換えた記号も含めて、また 、となりあう頻度を測定する。すなわち①から繰り返す。

これをある程度繰り返すことによって、構造が発見できると思われる。
特徴の有無の度合いは、ランダムな配列の場合の期待値と比較する方法と、ただ単純に出現する頻度による方法と二通り試みた。

2．1　ランダムとの比較

アミノ酸配列の中でアミノ酸がランダムに並んでいるかどうかを次の方法で判定する。

アミノ酸配列全体の中で、個々のアミノ酸が出現する頻度と、アミノ酸がとなりあって出現する頻度を求める。

Aというアミノ酸が現れる確率が
　　f（A）
Bというアミノ酸が現れる確率が
　　f（B）
であるとき、アミノ酸配列がランダムならば、ABというとなりあうアミノ酸が現れる確率は
　　f（A）×f（B）×n
となるはずである。ここでnは、アミノ酸の全数である。
実際に、あるとなりあうアミノ酸の出現する確率が、この期待値から大きくずれているなら、そのアミノ酸の組は、特徴があると考えられる。

約１００種類のアミノ酸配列（アミノ酸数約２万）に対して、以上の調査を行った結果、出現する頻度が少ないアミノ酸が他のものに比べ期待値から特に大きくずれてしまうことがわかった。そこで、この方法は、今後多量のデータが入手できたきに行うことにした。

2．2　出現頻度による方法

ランダムとの比較はせずに、ただ出現頻度の多いものを特徴のあるものと考える方法である。この方法は特徴を求めると言うより、同一の配列を抜き出す。

Analysis of Amino Acid Sequence
Takanori ONO, Nobuo TAKIGUCHI, Yoshiyuki KOTANI, Hirohiko NISIMURA
Tokyo University of Agriculture and Technology

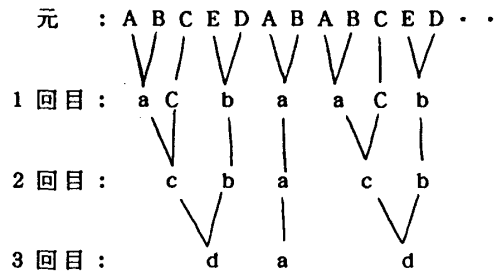
3．出現頻度による構造解析の手順

出現頻度による方法での構造解析について以下に述べる。

1．記号列の中で、ある２つの記号が隣あう回数を数える。
2．出現頻度の多い記号の組（２記号）を、別の記号（１記号）に置き換える置換表を作る。
3．置換表により記号列を出現頻度の多い順に置換してゆく。
4．置き換えた記号を含む新しい記号列で、2．から繰り返す。

これをある２つの記号がとなりあう回数が少なくなるまで（あるいは、２回以上出現しなくなるまで）繰り返す。
手順２．で得られた置換表が、構文規則となる。

（例）



規則：a ::= A B
b ::= E D
c ::= a C
d ::= c d

実際システム上は、記号をコード（１６bitの整数）で表現する。

4．実験例

簡単な実験例を以下に示す。

似た部分の多いアミノ酸配列（長さ１５３）２本と、似た部分の少ないアミノ酸配列１本を用いて解析した。

結果は、８回目で同じアミノ酸配列が出現しなくなり、それまでに作られた規則数は２８０となった。

置換回数	規則数	配列長 α	配列長 β	配列長 γ
0	― ― ―	１５３	１５３	１５３
1	１０９	８８	８５	１１３
2	８４	４７	４６	６９
3	４３	２７	２６	６９
4	２３	１６	１５	６９
5	１２	１０	９	６９
6	６	６	５	６９
7	２	５	４	６９
8	１	４	３	６９

配列αと配列βは同じ配列がかなりの部分を占めているので置換がかなり進んだ。同じ部分の少ない配列γは１回目の以降、置換が行われなかった。

5．まとめ

出現頻度による方法は、結果的には、同一のアミノ酸配列を抜き出した形と似たようなものになる。しかし、長い配列の中に他の短い配列と同じものが含められている場合に、この２文字連接方式による構造解析なら解析木にその情報が残る。この利点を考慮しながらさらに新しい方法を考察する予定である。

本研究の一部は、文部省科研費重点領域研究（ゲノム情報）の支援を受けて行われた。

参考文献

［１］藤本大三郎：タンパク質とは何か、講談社（１９８７）