

蛋白質の二次構造予測支援エキスパート・システム

2L-8

古田 香代里\* 加藤 正明\*\* 椎野 努\*  
\* 沖電気工業(株) \*\* 沖テクノシステムズラボラトリ(株)

1. はじめに

蛋白質の性質は、その立体構造(三次構造)に強く依存することが知られている。従って、希望する性質の蛋白質を人工的に得るためには、希望する三次構造の蛋白質を合成する必要がある。しかし現在、任意のアミノ酸配列(一次構造)を持つ蛋白質がどのような三次構造を持つかを正確に知る方法は存在しない。そこで、限定された、しかも比較的分子数の少ない蛋白質の立体構造解析(X線回折像による解析や磁気共鳴による解析、エネルギー計算法による解析)によって得られたデータを基に、アミノ酸配列と構造との相関関係を統計的な量として捉え、それによって新しいアミノ酸配列の蛋白質の構造を予測する方法がとられる。この際、一次構造と三次構造の間の直接的な相関を求めるのではなく、その中間構造であるアミノ酸配列中の局所的構造(二次構造)を仲介にした相関を求めることが提唱され、一次構造と二次構造の相関を予測する方法がいくつか考案されている。しかしながら、立体構造解析によって現在までに得られているデータは、数も少なく、その試料となる蛋白質の規模、分野共に偏りがあり、統計量として扱うには不十分であることは否めない。従って、これらの予測法は、構造解析によって得られているデータの解釈のしかたによって、それぞれ異なった法則となっている[JMCF85][SCHU79]。現状では、これらの法則から得られた予測結果を基に実験的検証が行われたり、あるいは、より精度の高い予測法の研究が行われている。

本稿では、このような現状から、現在考案されているいくつかの二次構造予測法により、蛋白質の二次構造予測を高速に行うと共に、新しい予測方式の発展にも対応できる支援エキスパート・システムの構成と機能について述べる。

2. 蛋白質の構造予測

1章で述べた予測法は、不完全ではあるが、研究の指標となるという面では非常に有用である[JMCF85]。

現在までに考案されている一次構造から二次構造の予測法の中で、Chou & Fasman法、Robson法、Nagano法の三つの方法が有力とされている。これらの方法は、蛋白質の立体構造解析結果から、アミノ酸が特定の二次構造に多く出現する事実が経験的に知られていることを用いて、この出現頻度を定量的に評価

した結果を傾向性指数と呼ばれる値で示し、それを基準として二次構造を予測する[SCHU79][長野85]。

ここで問題となることは以下の2点である。①個々の予測法を単独に適用するだけでは、その予測法自体の持つ不確かさを判断できない。②立体構造解析結果の増加によって傾向性指数が変化しただけでなく、新しい法則性の発見により、予測手法そのものが変化する可能性がある。これらは、この種の研究支援システムの構築にとつての大きな特徴である。

3. 蛋白質の二次構造予測支援エキスパート・システム

上述の問題に対処するためには、①複数の予測法を同時に適用し、個々の予測法の持つ不確かさを他の予測法と相対的に比較できる。②知識の流動性、すなわち、知識の追加、変更あるいは精密化に対応し得る、という機能が必要である。

そこで、これらの機能を有する蛋白質の二次構造予測支援エキスパート・システムのプロトタイプをAIワークステーションif1000上に構築した。本システムでは、複数の予測法を集結し、さらに知識を独立したモジュールとして知識ベースに実現した[山本86]。本システムは、各々の予測法による不確実性を含む結果を提示し、推論結果が一致する部分と相反する部分を明示し、その論理的根拠および推論過程を説明することにより、研究者の思考を支援する。

3.1 システムの構成と機能

システムの論理構成を図1に示す。

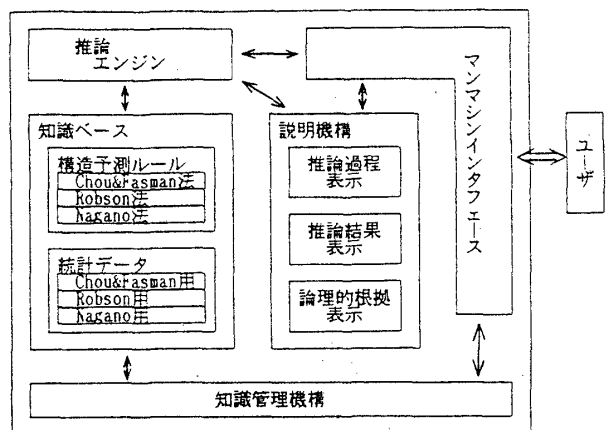


図1. システムの論理構成

本システムは、二次構造予測に必要な知識を格納した知識ベースと、この知識ベースを用いた推論を制御する推論エンジン、推論過程および推論結果とその論理的根拠を出力する説明機構、知識の追加修正を制御する知識管理機構、マンマシンインタフェース部から成る。

知識ベースは、(1) 複数の予測法をルール化した構造予測ルール部と、(2) 構造解析により算出された統計確率的データを記述した統計データ部の、2つの独立したモジュールから構成されている。現在、(1) の構造予測ルール部には2章で述べた3種類の予測法 (Chou & Fasman法、Robson法、Nagano法) をルール化している。

説明機構は、(1) 推論過程説明機能と推論結果の論理的根拠説明機能、(2) 推論結果の画像表示機能を有する。

知識管理機構は、知識ベースに対して、既に格納されている知識の修正や新たな知識の追加等の機能を有する。

マンマシンインタフェース部は、ユーザの入力としてキーボード、マウスを、出力としてディスプレイを制御する。

推論エンジンはif1000のProlog/SLインタプリタを用いる。

### 3.2 システムの出力

本システムの出力例を図2、図3に示す。

図2は、Chou & Fasman法による推論結果を表示したものである。入力された対象アミノ酸配列に対して、推論結果の二次構造 ( $\alpha$ ヘリックス、 $\beta$ シート、折れ曲り構造) の部位を模式的に表現している。

図3は、2章で述べた3種類の予測法による推論結果の対照表示である。入力された対象アミノ酸配列に対する3種類の推論結果から得られた二次構造の部位の一致する部分と相反する部分を明示している。

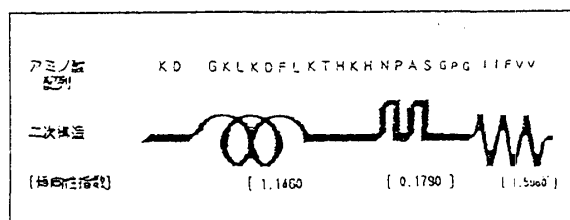


図2. Chou&Fasman法による推論予測結果表示  
( $\circ$ ):  $\alpha$ ヘリックス  $W$ :  $\beta$ シート  $\sim$ : 折れ曲り構造)

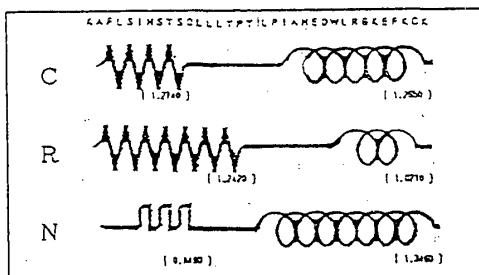


図3. 推論結果の対照表示  
(C: Chou&Fasman法, R: Robson法, N: Nagano法)

## 4. 説明機構の実現形態

エキスパート・システムに要求される説明機能について本システムにおける実現形態を示す。

### 4.1 推論過程の説明

推論に用いた知識ベース中の構造予測ルールを、発火の順に表示する。ユーザは表示されたルール群を見ることで、推論の正当性を検証できる。

### 4.2 推論結果の論理的根拠の説明

推論過程に生じる仮説 (初期状態を含む) を推論結果に至るまで順次に表示する。この時に変化する指数 (不確実性を含む統計確率的な量) の経緯を示す。この表示結果から、各予測法による推論の過程でどのような仮説が生成され、仮説に与えられている指数をもとに、どの仮説が消滅したかの連鎖が論理的に明らかにされる。これによって、最終的に推論結果の導かれた根拠が明確になり、各予測法の論理の違いも明確になる。ユーザは以上の論理的根拠を見ながら、どちらがより確からしいかの判断を行うことになる。

## 5. おわりに

蛋白質の二次構造予測支援エキスパート・システムは、複数の予測法を集めた知識ベースを有し、各予測法の論理的根拠を説明する機能を有する。本システムは、構造の明らかでない蛋白質あるいは未知の蛋白質の構造予測に目安を与え、詳しい構造解析のための初期値設定にも寄与する。今後、現在開発中のプロトタイプ of 知識管理機構や説明機構等の機能強化、特に知識ベースエディタ等のマンマシンインタフェースの拡充を図る予定である。

また、今後のシステムの発展の方向としては、蛋白質工学の発展に伴い、各種データが充実してきた場合、新しい予測法の発案のための支援システム (論理的説明機能等) としても活用でき、それによって発案された新しい予測法は、そのまま本システムにインプリメントすることにより、より正確な構造予測エキスパート・システムとして進化することが期待できる。

## 謝辞

本研究を行うにあたり御指導を頂いた大阪大学富田研一教授、姫路工業大学安岡則武教授に感謝致します。

## 参考文献

- [SCHU79] SCHULZ, et al. "Principles of Protein Structure" Springer-Verlag, 1979.
- [JMCF85] (社)日本機械工業連合会 他「昭和59年度 次世代産業基盤技術調査研究報告書」, 1985.
- [長野85] 長野「プロテインエンジニアリング」pp. 30-52 シーエムシー, 1985.
- [山本86] 山本 他「二次構造予測法における不確実性推論」第33回情報学会全国大会論文集