

(51)Int.Cl. ⁷	識別記号	F I	テ-マコード [*]	(参考)
G06F 17/30	170	G06F 17/30	170	F 5B075

審査請求 未請求 請求項の数 6 O L (全10頁)

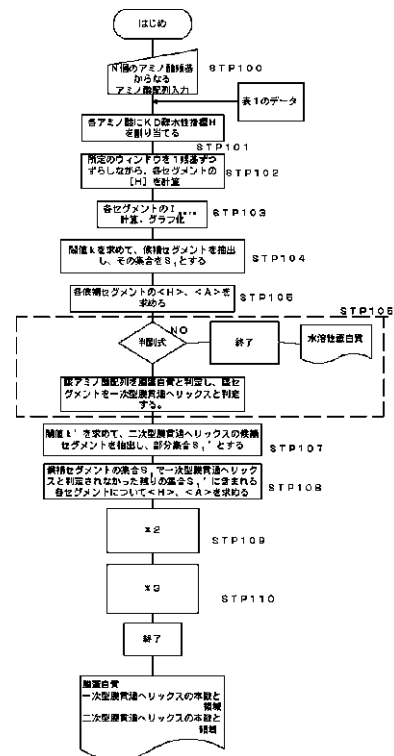
(21)出願番号	特願2001 - 10466(P 2001 - 10466)	(71)出願人	396020800 科学技術振興事業団 埼玉県川口市本町 4 丁目 1 番 8 号
(22)出願日	平成13年 1 月18日(2001.1.18)	(72)発明者	美宅 成樹 東京都国分寺市南町 3 - 21 - 1 ブロード アベニュー国分寺1108
特許法第30条第 1 項適用申請有り		(72)発明者	広川 貴次 千葉県船橋市東船橋 2 - 15 - 1 北辰ビル 308号
		(74)代理人	100087631 弁理士 滝田 清暉 (外 1 名) F タ-ム(参考) 5B075 ND03 ND20 ND22 ND34 NR02 NR12 PP30 PQ02 PQ03 UU19

(54)【発明の名称】膜蛋白質か否かの判別方法、膜蛋白質の機能推定方法、膜貫通ヘリックス領域予測、及びこれらをコンピュータで実施するためのコンピュータプログラム

(57)【要約】 (修正有)

【課題】 アミノ酸の配列が膜蛋白質であるか否かを、高精度で迅速に判別する方法を提供。

【解決手段】 (a) 蛋白質構成要素の 2 0 個のアミノ酸について、疎水性指標 H、両親媒性指標 A、及び非電荷指標 C を割り当て、(b) 5 ~ 9 個のウィンドウ W を用いてアミノ酸配列の疎水性コア指標 I_{core} を算出し、(c) I_{core} 値を縦軸、被判別蛋白質のアミノ酸鎖を横軸にプロットし、(d) 一定の閾値を超えるアミノ酸残基の数が 1 5 以上のセグメントを一次型膜貫通ヘリックスの候補とし、(e) 各候補の二回平均疎水性 $\langle H \rangle$ 及び平均両親媒性 $\langle A \rangle$ を算出して、(f) 一定の判別式に従って一次型膜貫通ヘリックスが存在しない場合には被判別蛋白質を水溶性蛋白質と認定する。



【特許請求の範囲】

【請求項 1】 (a) 蛋白質の構成要素である 2 0 個のアミノ酸それぞれについて、疎水性指標 H、両親媒性指標 A、及び非電荷指標 C を割り当て、(b) 5 ~ 9 個の何れかの数の一連の窓を有するウィンドウ W を用いて被差別蛋白質のアミノ酸配列から抽出される前記ウィンドウ W と対応する全てのアミノ酸配列について、該アミノ酸配列の疎水性コア指標 $I_{c.o.r.e.}$ を、前記疎水性指標を用いて計算した一回平均疎水性 [H] を介して算出し、(c) 得られた $I_{c.o.r.e.}$ 値を縦軸とし、被差別蛋白質のアミノ酸鎖を横軸にしてプロットし、(d) 得られたグラフが一定の閾値を超える領域の一連のアミノ酸残基の数が 1 5 以上となったとき、その領域を構成する一連のアミノ酸残基からなるセグメントを一次型膜貫通ヘリックスの候補とし、(e) 得られた各候補について二回平均疎水性 $\langle H \rangle$ 及び平均両親媒性 $\langle A \rangle$ を算出し、(f) 被差別蛋白質のアミノ酸鎖の長さ(アミノ酸の数)、上記 2 回平均疎水性、及び、平均両親媒性を用いた一定の判別式に従って一次型膜貫通ヘリックスの有無を判別し、(g) 一次型膜貫通ヘリックスが存在しない場合には被差別蛋白質を水溶性蛋白質と認定し、(h) 一次型膜貫通ヘリックスが存在する場合には、少なくとも、前記(d)における閾値を変更すること以外は、前記(b) ~ (d) と全く同様にしてセグメントを抽出し、該抽出したセグメントの集合から既に一次型膜貫通ヘリックスと認定されたセグメントを除去して二次型膜貫通ヘリックスの候補セグメントを得、(i) 次いで前記(e) と同様にして各候補セグメントについて $\langle H \rangle'$ と $\langle A \rangle'$ を算出し、(j) 前記二次型膜貫通ヘリックスの候補セグメントについてその中心が 4 5 残基以下である場合と 4 5 残基以上である場合に分類し、前記 $\langle H \rangle'$ 及び $\langle A \rangle'$ を用いた新たな判別式によって二次型膜貫通ヘリックスの有無を判別すると共に、(k) 全く新しく、前記(b) で抽出された全てのアミノ酸配列の単位に対して非電荷指標 C を割り当て、該非電荷指標 C が 1 であるアミノ酸残基が連続して 1 9 残基以上存在するセグメントを抽出し、該セグメントについて前記 $\langle H \rangle'$ 及び $\langle A \rangle'$ を用いた判別式によって二次型膜貫通ヘリックスの有無を判別する方法であって、前記演算の基本データとして(a) で各アミノ酸に割り当てた各指標及びその値を使用することを特徴とする、膜蛋白質が否かの判別方法。

【請求項 2】 請求項 1 の方法によって得られた膜貫通ヘリックスの全本数を、予め記録された、膜貫通ヘリックスの本数と膜蛋白質の機能との対応表に照し合わせ、判別に係る蛋白質が有する機能を出力することを特徴とする、蛋白質の機能推定方法。

【請求項 3】 請求項 1 で得られた膜貫通ヘリックスのアミノ酸配列について、(1) 前後 7 残基領域の 3 . 6 周期のスペクトル密度を最大エントロピー法で計算する

と共にその領域の平均疎水性値を求め、(2) 膜貫通ヘリックスの中心から N 端及び C 端側に向けて、それぞれ 1 乃至 1 0 迄の数値 j を下記一般式で表される膜環境フィルターに代入しながら、(1) で計算したスペクトル強度を掛け合わせ、(3) 上記(2) で得られた値と平均疎水性値の和の値が連続して設定された閾値を超えた時、その閾値の存在個所によって特定された領域を出力することを特徴とする膜貫通ヘリックスの領域予測方法。

$$f(j) = \frac{1}{1 + \exp((j \pm 10)/2)}$$

【請求項 4】 蛋白質の構成要素である 2 0 個のアミノ酸のそれぞれに対して割り当てられた疎水性指標 H、両親媒性指標 A、及び非電荷指標を用い、選択されたアミノ酸配列について、一回平均疎水性 [H]、疎水性コア指標 $I_{c.o.r.e.}$ 、2 回平均疎水性 $\langle H \rangle$ 、及び平均両親媒性 $\langle A \rangle$ を計算し、これらの値を利用して、アミノ酸配列の決定された蛋白質が膜蛋白質であるか否かを判別するコンピュータプログラムであって、被差別蛋白質の全てのアミノ酸配列を入力し、判別プログラムを実行させると、(a) これらの全てのアミノ酸残基のそれぞれに前記疎水性指標 H を割り当て、(b) 5 ~ 9 から選択された何れかの数の一連の窓を有するウィンドウ W を前記被差別蛋白質の一端から 1 個ずつずらしながら抽出した全てのアミノ酸配列について一回平均疎水性 [H] を算出した後、該 [H] を用いて疎水性コア指標 $I_{c.o.r.e.}$ を算出し、(c) 得られた $I_{c.o.r.e.}$ 値をもとのアミノ酸列の順に対応させて並べた後一定の閾値を超える $I_{c.o.r.e.}$ を抽出し、(d) 該閾値を超える $I_{c.o.r.e.}$ の連続した領域に対応する一連のアミノ酸残基の数が 1 5 以上となる領域を一次型膜貫通ヘリックスの候補セグメントとすると共に、該候補セグメントの集合を S_1 とし、(e) 該 S_1 に属する各候補セグメントについて二回平均疎水性 $\langle H \rangle$ 及び平均両親媒性 $\langle A \rangle$ を算出し、(f) 被差別蛋白質のアミノ酸残基が 1 7 5 以上である場合には、 $\langle A \rangle - 4 \langle H \rangle + 4 . 8$ 及び $\langle A \rangle - 1 \langle H \rangle + 1 . 5$ の二つの判別式を同時に満足するセグメントを検索し、検索されたセグメントを一次型膜貫通ヘリックスと認定し、(g) 被差別蛋白質のアミノ酸残基が 1 7 5 未満である場合には、判別式 $\langle A \rangle - 1 . 2 5 \langle H \rangle + 1 . 6 2 5$ を満足するセグメントを検索し、検索されたセグメントを一次型膜貫通ヘリックスと認定し、(h) 上記(f) 又は(g) の工程で一次型膜貫通ヘリックスが検索されなかった場合には、被差別蛋白質を水溶性蛋白質と認定しその結果を出力してプログラムを終了し、(i) 上記(f) 又は(g) の工程で少なくとも 1 個の一次型膜貫通ヘリックスが検索された場合に、少なくとも、前記(c) で使用した閾値とは異なる閾値を使用すること以外は、(b) ~ (d) と同様

にして抽出したセグメントの集合から、前記一次型膜貫通ヘリックスに対応するセグメントを除いた残りのセグメントの集合 S_i について、その中心が 4 5 残基以下のセグメントについては判別式 $\langle A \rangle - 2 < H \rangle + 1 . 6$ であるセグメントを検索し、その中心が 4 5 残基を超えるセグメントについては、更にその前後 6 5 残基以内に一次型膜貫通ヘリックスが存在することを条件に判別式 $\langle A \rangle - 2 . 7 1 < H \rangle + 2 . 5 7$ を満足するセグメントを検索すると共に、別途、(j) 被判別蛋白質を構成する全アミノ酸残基に非電荷指標を割り当て、非電荷指標 C が 1 であるアミノ酸残基が連続して 1 9 残基以上となるセグメントを抽出し、そのセグメントの中心から前後 5 0 残基以内に前記一次型膜貫通ヘリックス又は、前記 (i) で検索されたセグメントが存在することを条件に、判別式 $\langle A \rangle - 5 < H \rangle + 1 . 9$ を満足するセグメントを検索し、(k) 前記 (i) 及び (j) で検索されたセグメントを二次型膜貫通ヘリックスと認定し、(1) 一次型膜貫通ヘリックスの本数と二次型膜貫通ヘリックスの本数、並びに、必要に応じてその合計本数を出力することを特徴とする、蛋白質が膜蛋白質であるか否かを判別するためのコンピュータソフトウェア；但し、各判別式中の定数は、夫々 $\pm 1 0 \%$ の間で変動することがあるものとする。

【請求項 5】 請求項 4 のコンピュータソフトウェアを用いて得られた膜貫通ヘリックスの全本数を計算し、次いで、予めデータとして用意された膜貫通ヘリックスの本数と膜蛋白質の機能との対応表に照し合わせ、判別に係る蛋白質が有する機能を生じさせることができる、蛋白質の機能を推定するためのコンピュータソフトウェア。

【請求項 6】 請求項 4 のコンピュータソフトウェアによって検索された各膜貫通ヘリックスのアミノ酸配列について、(1) 該アミノ酸配列の前後 7 残基領域の 3 . 6 周期のスペクトル密度を最大エントロピー法で計算すると共にその領域の平均疎水性値を求め、(2) 膜貫通ヘリックスの中心から N 端及び C 端側に向けて、それぞれ 1 乃至 1 0 迄の数値 j を下記一般式で表される膜環境フィルターに代入しながら、前記 (1) で計算したスペクトル強度を掛け合わせ、(3) 前記 (2) で得られた値と平均疎水性値の和の値が連続して設定された閾値を超えた時、その閾値の存在個所によって特定された領域を出力させることができることを特徴とする膜貫通ヘリックスの領域予測のためのコンピュータソフトウェア。

$$f(j) = \frac{1}{1 + \exp((j \pm 10) / 2)}$$

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】 本発明は、アミノ酸配列が決定された蛋白質について、それが膜蛋白質であるか否かを判別する方法に関し、特に、コンピュータを用いて迅

速且つ高い精度で膜蛋白質であるか否かを判別する方法、及び、膜蛋白質である場合には、膜蛋白質であるとの情報を与えるセグメントについての情報を出力すると共に、該膜蛋白質がいかなる機能を発揮するかを推定する方法、及びそれを実施する為のコンピュータソフトウェアに関する。

【 0 0 0 2 】

【従来技術】 膜蛋白質の二次構造の予測は、一般にアミノ酸配列に疎水性指標を割り当てたプロフィールから簡単な計算によって予測することができる [Kyte, J. and Doolittle, R.F., J. Mol. Biol., 157(1982)103-132.]。更に、ニュートラルネットワーク [Rost, B., Casadio, R., Fariseli, P. and Sander, C., Prot. Sci., 4(1995)521-533] や統計的手法 [Jones, D.T., Taylor, W.R. and Thornton, J.M., Biochemistry, 33(1994)3038-3049.] による方法も報告されている。しかしながら、水溶性蛋白質に比べて一見精度よく見える膜蛋白質二次構造予測にも、次のようないくつかの問題点がある。

【 0 0 0 3 】 (1) ある種の蛋白質では、従来の方法では予測が困難な二次構造 (膜貫通ヘリックス) が存在する。

(2) 膜貫通ヘリックスの領域予測が十分でない。

(3) 膜貫通ヘリックスに対して成績の良い方法でも、実際には水溶性蛋白質である多くのものを膜貫通ヘリックスとして誤認予測してしまう。

【 0 0 0 4 】 膜貫通ヘリックスは、基本的には疎水性相互作用によって脂質二重膜の疎水性領域に入り、安定となる。しかしながら、ある疎水性インデックスの値を閾値とし、それを越える一次構造領域を膜貫通ヘリックスとするような従来の単純な方法では、本来単なる水溶性蛋白質であり、膜貫通ヘリックスではないアミノ酸配列の解析からも、多くの膜貫通ヘリックスが誤って予測されるという欠点があった。一方、膜貫通ヘリックスにはしばしば極性残基が含まれており、特に細胞内側には正電荷を持つ残基が多いことなどが知られている [von Heijne, G., J. Mol. Biol., 225(1992)487-494.]。このことは膜貫通ヘリックス形成にも極性相互作用の存在 (極性と疎水性のバランス) を考慮することが必要であることを示している。

【 0 0 0 5 】 一方、膜蛋白質ができるプロセスを考えると、アミノ酸配列にはいくつかの役割があるということが分かる。つまり、あるアミノ酸配列からなるセグメントを想定した場合、①そのセグメント自体が膜貫通ヘリックスを作るべく設計された疎水性の高い配列である場合や、②膜貫通ヘリックスになったときに一定の機能を持つように設計されている場合などに分類することができる。後者の場合、膜蛋白質の機能に本質的に重要なアミノ酸配列は疎水性の低いアミノ酸残基を含む場合が多く、本来膜貫通ヘリックスを作ることに不利に働くが、近傍の疎水性の高いセグメントによって膜内に引き

込まれ、膜貫通ヘリックスになると考えられる。

【0006】つまり、膜貫通ヘリックスには次の2つのタイプがあると考えられる。

1) アミノ酸配列が本来持つ高い疎水性によって、そのセグメントが自発的に膜貫通ヘリックスを形成するタイプ

2) 膜蛋白質の機能的に重要なセグメントなど比較的親水的なセグメントが、上記1)のタイプの膜貫通ヘリックスによって膜内に引き込まれ、膜貫通ヘリックスを形成するタイプ

これらの膜貫通ヘリックスが周囲のヘリックス同士で相互作用することによって強固にパッキングされ、膜蛋白質の立体構造が安定化されるものと考えられる。以上のような膜蛋白質の立体構造形成についての考察から、本発明においては主として上記1)のタイプのものを一次型膜貫通ヘリックス、2)のものを二次型膜貫通ヘリックスと分類する。

【0007】そこで本発明者は、膜貫通ヘリックスをより高い精度で予測することについて鋭意検討した結果、Kyte-Doolittleの疎水性指標(以下、K-D疎水性指標とする)と、新しく定義した両親媒性指標および非電荷指標を用いて膜貫通候補セグメントを抽出・分類し、さらにタンパク質の大きさを考慮して予測した場合には、格段に高精度の予測が可能となることを見出し、本発明に到達した。

【0008】

【発明が解決しようとする課題】従って本発明の第1の目的は、アミノ酸の配列が決定された蛋白質が膜蛋白質であるか否か、及び、膜蛋白質である場合に、その領域並びにその機能がどのようなものであるかを高い精度で判別するための方法を提供することにある。更に本発明の第2の目的は、蛋白質の一次構造情報から、コンピュータを用いてその蛋白質が膜蛋白質であるか否か、膜蛋白質である場合には更にその領域、及び機能を高い精度で判定するための、コンピュータプログラムを提供することにある。

【0009】

【課題を解決するための手段】本発明の上記の諸目的は、アミノ酸配列の決定された蛋白質の判別方法であって、該方法が、(a)蛋白質の構成要素である20個のアミノ酸それぞれについて、疎水性指標H、両親媒性指標A、及び非電荷指標Cを割り当て、(b)5~9個の何れかの数の一連の窓を有するウィンドウWを用いて被判別蛋白質のアミノ酸配列から抽出される前記ウィンドウWと対応する全てのアミノ酸配列について、該アミノ酸配列の疎水性コア指標 $I_{c.o.r.}$ を、前記疎水性指標を用いて計算した一回平均疎水性[H]を介して算出し、(c)得られた $I_{c.o.r.}$ 値を縦軸とし、被判別蛋白質のアミノ酸鎖を横軸にしてプロットし、(d)得られたグラフが一定の閾値を超える領域の一連のアミノ酸

残基の数が15以上となったとき、その領域を構成する一連のアミノ酸残基からなるセグメントを一次型膜貫通ヘリックスの候補とし、(e)得られた各候補について二回平均疎水性<H>及び平均両親媒性<A>を算出し、(f)被判別蛋白質のアミノ酸鎖の長さ(アミノ酸の数)、上記2回平均疎水性、及び、平均両親媒性を用いた一定の判別式に従って一次型膜貫通ヘリックスの有無を判別し、(g)一次型膜貫通ヘリックスが存在しない場合には被判別蛋白質を水溶性蛋白質と認定し、

- 10 (h)一次型膜貫通ヘリックスが存在する場合には、少なくとも、前記(d)における閾値を変更すること以外は、前記(b)~(d)と全く同様にしてセグメントを抽出し、該抽出したセグメントの集合から既に一次型膜貫通ヘリックスと認定されたセグメントを除去して二次型膜貫通ヘリックスの候補セグメントを得、(i)次いで前記(e)と同様にして各候補セグメントについて<H>'と<A>'を算出し、(j)前記二次型膜貫通ヘリックスの候補セグメントについてその中心が45残基以下である場合と45残基以上である場合に分類し、前記<H>'及び<A>'を用いた新たな判別式によって二次型膜貫通ヘリックスの有無を判別すると共に、
- 20 (k)全く新しく、前記(b)で抽出された全てのアミノ酸配列の単位に対して非電荷指標Cを割り当て、該非電荷指標Cが1であるアミノ酸残基が連続して19残基以上存在するセグメントを抽出し、該セグメントについて前記<H>'及び<A>'を用いた判別式によって二次型膜貫通ヘリックスの有無を判別する方法であって、前記演算の基本データとして(a)で各アミノ酸に割り当てた各指標及びその値を使用することを特徴とする、膜蛋白質か否かの判別方法、及び、それを実施する為のコンピュータプログラムによって達成された。

【0010】

【発明の実施の形態】一般的に、一次型膜貫通ヘリックスは、水溶性蛋白質や膜蛋白質の水溶性ドメインには見られないような特徴を持っていて、特に疎水性が高い。しかしながら、疎水性だけではそれらを判断することができない。即ち、脂質からなる二重膜自体が、非常に疎水性の高い炭化水素鎖領域と表面に分布する極性基でできているために、膜貫通ヘリックスも、中心の疎水性部分だけで安定化しているのではなく、両端の極性部位による構造安定化がかなり重要であると考えられる。従って、中心部分に対する疎水性パラメータである疎水性指標Hと共に、もう1つのパラメータとして膜貫通ヘリックスの両端にある極性残基の分布を考慮することが妥当であると考えられる。

- 【0011】また、一般に極性基があると共に炭化水素部分が大きい場合には、分子の両親媒性が高く、ミセルやラメラ構造などの超分子構造を作りやすい。これと同様に、側鎖自体の両親媒性が高いと膜貫通ヘリックスの両端の構造を安定化し易いのではないかと考えられる。

本発明者は、このことを用いて、新たに、下記表 1 に示す両親媒性指標 A_1 , A_2 を作成した。以後これを両親媒性指標 A とする。尚、 C は後述する非電荷指標であり、() 内のアルファベットは、三文字表記のアミノ酸

を一文字表記する場合の記号である。

【 0 0 1 2 】

【表 1】

アミノ酸	H	A		C	アミノ酸	H	A		C
		A_1	A_2				A_1	A_2	
ILE(I)	4.5	0	0	1	TRP(W)	-0.9	0	6.93	1
VAL(V)	4.2	0	0	1	TYR(Y)	-1.3	0	5.06	1
LEU(L)	3.8	0	0	1	PRO(P)	-1.6	0	0	1
PHF(F)	2.8	0	0	1	HIS(H)	-3.2	1.45	0	0
CYS(C)	2.5	0	0	1	ASP(D)	-3.5	0	0	0
MET(M)	1.9	0	0	1	ASN(N)	-3.5	0	0	1
ALA(A)	1.8	0	0	1	GLU(E)	-3.5	1.27	0	0
GLY(G)	-0.4	0	0	1	GLN(Q)	-3.5	1.25	0	1
THR(T)	-0.7	0	0	1	LYS(K)	-3.9	3.87	0	0
SER(S)	-0.8	0	0	1	ARG(R)	-4.5	2.45	0	0

【 0 0 1 3 】以下、本発明を具体的手順に従って説明する。本発明においては、先ず判別しようとする蛋白質を構成する各アミノ酸に、表 1 によって予め設定されている $K - D$ 疎水性指標 H を割り当てる。次に、5 ~ 9 箇から選択されるアミノ酸残基の数、好ましくは 7 箇の連続するアミノ酸残基に当てはめることのできるウィンドウ W を用いて、下記 (1) 式によって一回平均疎水性 $[H]$ を計算する

$$[H (i)] = [\sum_{j=i-W/2}^{i+W/2} H (j)] / W \quad (1)$$

$$I_{c o r e} (i) = P o s i t i v e ([H_{m a x} (i)] + S (i) \cdot [H_{m i n} (i)]) \quad (5)$$

但し、 $[H_{m a x} (i)]$ 、 $[H_{m i n} (i)]$ 及び $S (i)$

但し、 i はウィンドウの中心のアミノ酸の座標である。上記の計算は、例えば 7 残基について行う場合 ($W = 7$ 、 $W / 2 = 3$ とする) には蛋白質を構成するアミノ酸鎖の左端から 7 残基ウィンドウを 1 残基ずつずらしながら、5 残基 ($W = 5$ 、 $W / 2 = 2$ とする) について行う場合には 5 残基ウィンドウを 1 残基ずつずらしながら、全ての単位について計算する。

【 0 0 1 4 】例えば、下記 (5) 式によって、疎水性コア指標である $I_{c o r e} (i)$ を計算する。

$$[H_{m a x} (i)] = M a x ([H (j)] : j = i - n \dots , 0 , \dots , i + n) \quad (6)$$

$$[H_{m i n} (i)] = M i n ([H (j)] : j = i - n , \dots , 0 , \dots , i + n) \quad (7)$$

$$S (i) = S i g n ([H_{m a x} (i)] \cdot [H_{m i n} (i)]) \quad (8)$$

【 0 0 1 5 】上記の $I_{c o r e} (i)$ は正負の間を変動する関数の、正の部分 (Positive) を表す。尚、Positive 関数は、負の値の場合は 0 とする。この正の値が特定の閾値 k より大きくなる部分を抽出し、その部分の連続するアミノ酸残基の数が L より大であるものを、膜蛋白質か否かを判別するためのデータを保有する第一次セグメントの候補 $S (i)$ とし、この候補セグメントの集合を S_1 とする。但し L は、15 ~ 25 から選択された何れかのアミノ酸残基の数、即ち長さであり、好ましくは 18 ~ 20 から選択される整数である。尚、 L を 19 とした場合には、前記 k として、1.15 を設定することが好ましい。また、 $M a x$ 及び $M i n$ は、領域の中から最大値及び最小値を選び出す関数であり、 $S i g n$ は () 内为正のとき $S (i)$ が 1 となり、() 内が負のとき $S (i)$ が - 1 となるように符号を返す関数である。

(i) は下記 6 ~ 8 式で表される。

は前記 (1) 式で求めたものである。但し、 i_c はセグメントの中心、 m はセグメント長の 1 / 2 である。

$$\langle H \rangle = [\sum_{j=i_c-m}^{i_c+m} H (j)] / (2m + 1) \quad (2)$$

$$\bar{A}_1 (i) = [\sum_{j=i-3}^{i+3} A_{E,Q,H,K,R} (j)] / 7 \quad \text{及び} \quad \bar{A}_2 (i) = [\sum_{j=i-3}^{i+3} A_{Y,Z} (j)] / 7 \quad (3)$$

$$\langle A \rangle = [\sum_{j=i_c-m}^{i_c-6} \bar{A}_1 (j) + \sum_{j=i_c+6}^{i_c+m} \bar{A}_1 (j)] / 2(m-6) + [\sum_{j=i_c-m}^{i_c+m} \bar{A}_2 (j)] / (2m + 1) \quad (4)$$

【 0 0 1 6 】上記の如くして抽出された候補 $S (i)$ の各セグメントについて、下記の式により、二回平均疎水性 $\langle H \rangle$ と平均両親媒性 $\langle A \rangle$ を計算する。下記式は、ウィンドウ W を 7 とした場合について示したものである。尚、 A は表 1 に示された両親媒性指標であり、 $[H]$

【 0 0 1 7 】次に、被判別蛋白質の長さ N が約 175 (ここでは 175 と設定) より長い場合と短い場合に分けて、 $S (i)$ の集合 S_1 に含まれる全てのセグメントについて次式の判別式を実行する。但し、通常、 N は 90 ~ 250 の整数である。

①: $N = 175$ の場合: $\langle A \rangle - 4 < H \rangle + 4.8$ 及び $\langle A \rangle - 1 < H \rangle + 1.5$ の両判別式を同時に満足するセグメントは一次型膜貫通ヘリックスである。

②: $N < 175$ の場合: 判別式 $\langle A \rangle - 1.25 < H \rangle + 1.625$ を満足するセグメントは一次型膜貫通ヘ

リックスである。

尚、上記判別式は、一般式 $\langle A \rangle - a \langle H \rangle + b$ で表され、この a 及び b は、既知の多数の蛋白質に対する $\langle A \rangle$ 及び $\langle H \rangle$ を計算し、 $\langle A \rangle$ 及び $\langle H \rangle$ を夫々 X 軸又は Y 軸としてグラフ化した散布図に基づいて選択される。従って、上記判別式における定数 a 、 b は、夫々上記の値 $\pm 10\%$ の範囲で異なった値に設定される場合がある。

【0018】上記判別式を満たすセグメントが前記候補セグメントの中に1つも存在しない場合には、被判別蛋白質は水溶性蛋白質であって、膜蛋白質ではないと判定され、判定操作（プログラム）は終了する。従って、この蛋白質については以降の演算はされず、別の被判別蛋白質について、初めから本判定操作（プログラム）が実行される。一方、上記判別式を満たすセグメントが抽出されたとき、そのセグメントを一次型膜貫通ヘリックスであると判定する。尚、上記の判別式は、既知の蛋白質について、平均両親媒性 $\langle A \rangle$ の二回平均疎水性 $\langle H \rangle$ 依存性を表す散布図を作成することによって求めることができる。

【0019】次に、類似の方法によって二次型膜貫通ヘリックスの存在の有無を判定する。この場合の候補セグメントとして、少なくとも S_1 を求めたときの閾値 k の値を変えること以外は、 S_1 の場合と同様にして候補セグメントを抽出し、この候補セグメントの集合の中から、前記一次型膜貫通ヘリックスを除去した残りの全セグメント（その集合を S_1' とする）、及び、新たに、非電荷指標 C が1であるアミノ酸が、15～25から選択された何れかの長さ L' 以上に連なった全てのセグメント（その集合を S_2 とする）の集合を含める。 S_1' を求めるとき、ウィンドウ W や閾値 k より大きくなる部分の連続するアミノ酸の数 L 等も、 k の値と共に変更しても良い。

【0020】次に、一次型膜貫通ヘリックスの判定の場合と同様にして、次式によって二次型膜貫通ヘリックスの有無を判定する。 S_1' の集合に含まれる各セグメントに対する判別式は下記の通りである。

I c o r e で決まる領域の中心が45残基以下の場合：
 $\langle A \rangle - 2 \langle H \rangle + 1.6$

I c o r e で決まる領域の中心がN端から45残基以上、且つ前記中心から前後65残基以内に一次型膜貫通ヘリックスの中心が存在する場合：
 $\langle A \rangle - 2.71 \langle H \rangle + 2.57$

S_2 の集合に含まれる各セグメントについては、 S_1 の場合と全く同様にして $\langle H \rangle$ 及び $\langle A \rangle$ を求める。次に各候補セグメントについて、中心から前後50残基以内に一次型膜貫通ヘリックス又は上記判別式によって判別されたセグメントが存在することを条件として $\langle A \rangle - 5 \langle H \rangle + 1.9$ の判別式を満たすものを抽出する。但し、これらの判別式における勾配や定数も $\pm 10\%$ の範

囲で変更する場合もある。 S_1' 及び S_2 について上記判別式を満足するセグメントがない場合には二次膜貫通ヘリックスは存在しないことになり、判別操作（プログラム）は終了する。上記判別式を満足するセグメントが存在する場合には、そのセグメントを二次型膜貫通ヘリックスであると判定する。

【0021】以上の判定によって得られた膜貫通ヘリックスの末端の最終的な領域を決定するために、本発明においては、最大エントロピー法によって周期のスペクトル密度を求め、それを下記 $f(j)$ の式で表される膜環境フィルタと組み合わせて膜貫通ヘリックスの末端を決定する。

$f(j) = 1 / (1 + \exp((j \pm 10) / 2))$ 但し、 j はコアの中心から左右へ伸長する値である。膜貫通ヘリックスの末端は、以下の操作によって決定する。

1. 膜貫通ヘリックスのアミノ酸配列について、その前後7残基領域の3.6周期のスペクトル密度を最大エントロピー法で計算し、同時にその領域の平均疎水性値を求める。

20 2. 膜貫通ヘリックスコアの中心からN端、C端側に向けて1, 2, ... 10までの数値 j を膜環境フィルタに代入しながら、(1)で計算したスペクトル強度を掛け合わせる。

3. (2)で得られた値と平均疎水性値を足しあわせてその値が連続してある閾値を超えれば、その領域を膜貫通ヘリックス領域とする。

【0020】被判別蛋白質のアミノ酸配列について以上のデータ入力と演算を行わせ、不等式の判別式を満足する解がない場合には、判別操作（プログラム）はそこで終了する。本発明においては、各判別段階によって得られた全ての膜貫通ヘリックスの本数とそれらの領域についての情報を、モニター及び/又はプリンターによって出力する。勿論、各判別段階における、それらの情報を、個別に出力するようにすることも、全体情報と併せて各個別データを付加的に出力することもできる。

【0021】上記の判別を実施する為のコンピュータプログラムのフローチャートは、図1に示される通りである。本発明のコンピュータプログラムは、C言語等を用いて記載することができる。以下に更に詳述する。

40 【0022】本発明のプログラムを起動し、被判別蛋白質を形成するN個のアミノ酸配列を入力すると、各アミノ酸にN個の連続した整数が割り当てられ(STP100)、入力されたアミノ酸残基数がN個であることが認識される。次に、入力された全てのアミノ酸残基に、予めメモリ部に保存されている前記表1のデータのうち、該当する疎水性指標Hの値を割り当てる(STP101)。

50 【0023】所定のウィンドウ（例えば7つの連続した窓を有するウィンドウ）を、端から1残基ずつずらしながら、各ウィンドウに対応するアミノ酸列を抽出し、抽

出された全てのアミノ酸列に対して、前記(1)式に従って一回平均疎水性[H]を計算する(STP102)。次に前記(5)式に従って、上記の全てのアミノ酸列に対して $I_{c.o.r.}$ を計算しこれを被判定蛋白質のアミノ酸配列の順に対応させて整列させる($I_{c.o.r.}$ のグラフ化: STP103)。

【0024】 $I_{c.o.r.}$ がk(例えば1.15)より大きくなる領域を抽出し、その領域のうちL(例えば19)より大きい数のアミノ酸残基が連続する部分のアミノ酸配列を候補セグメントとして抽出し、その集合を S_1 とする(STP104)。次いで、この各候補セグメントについて、表1のデータを用い、前記した(2)~(4)式に従って二回平均疎水性<H>及び平均両親媒性<A>を計算する(STP105)。

【0025】被判定蛋白質を形成するアミノ酸残基の数Nが175より大きい場合には、STP104で得られた集合 S_1 に含まれる全ての候補セグメントについて、 $\langle A \rangle - 4 \langle H \rangle + 4.8$ 及び $\langle A \rangle - 1 \langle H \rangle + 1.5$ を同時に満たすか否かを判別し、これらの判別式を満たすセグメントが抽出された場合にそれを一次型膜貫通ヘリックスと認定する。一次型膜貫通ヘリックスが存在しない場合には、プログラムは終了し、被判定蛋白質が水溶性蛋白質であるとの結果を出力する(STP106)。

【0026】一次型膜貫通ヘリックスが存在した場合には、次に二次型膜貫通ヘリックスの有無の判断に移る。少なくとも一次型膜貫通ヘリックスの判別操作で使用した閾値kをk'に変更し、必要に応じて、ウィンドウWや、候補セグメントを抽出するための閾値k'を超える領域として設定する一連のアミノ酸残基の数を更に変更すること以外は、 S_1 の場合と同様にして候補セグメントを抽出し、この中から一次型膜貫通ヘリックスに該当するセグメントを除去し、この集合 S_1' を二次型膜貫通ヘリックスの候補セグメントの部分集合と認定する(STP107)。次に、この集合 S_1' に含まれる各セグメントについて、 S_1 の場合と同様にして<H>及び<A>を求める(STP108)。

【0027】セグメントの中心が45残基以下であるものに対しては判別式 $\langle A \rangle - 2 \langle H \rangle + 1.6$ を満足するものを二次型膜貫通ヘリックスと認定し、45残基以上である場合には、更に、該セグメントの前後65残基以内に一次型膜貫通ヘリックスが存在することを条件に、判別式 $\langle A \rangle - 2.71 \langle H \rangle + 2.57$ を満足するものを抽出し、これも二次型膜貫通ヘリックスと認定する(STP109)。尚、これらの判別式における定数は、 $\pm 10\%$ の範囲で変更することもある。

【0028】次に、全く新しく、被蛋白質の全アミノ酸

残基に、表1に基づいてCの値を割り当てる。その結果 $C = 1$ のアミノ酸残基が連続して19残基を超える場合、その領域を二次型膜貫通ヘリックスの候補セグメントとし、その集合を S_2 とする。該候補セグメントのうち、その中心から前後50残基以内に一次型膜貫通ヘリックス又は

【0027】で認定した二次型膜貫通ヘリックスが存在するものに対しては、判別式 $\langle A \rangle - 5 \langle H \rangle + 1.9$ を満足するものを二次型膜貫通ヘリックスと認定する(STP110)。尚、上記判別式の定数は $\pm 10\%$ の範囲で変更することもある。これらの全ての演算が完了すると、一次型膜貫通ヘリックスの本数と領域、及び二次型膜貫通ヘリックスの本数と領域を出力してプログラムは、終了する。

【0029】膜貫通ヘリックスの本数と膜蛋白質の機能とは一定の関係有るので、その対応関係を表1の場合と同様に、予めコンピュータの記憶部に記憶させておくと、被判定蛋白質から抽出された全膜貫通ヘリックスの本数を該データと比較対照して、被判定蛋白質の機能を出力することができる。また、得られた膜貫通ヘリックスのアミノ酸配列について、(1)該アミノ酸配列の前後7残基領域の3.6周期のスペクトル密度を最大エントロピー法で計算すると共にその領域の平均疎水性値を求め、(2)膜貫通ヘリックスの中心からN端及びC端側に向けて、それぞれ1乃至10迄の数値jを下記一般式で表される膜環境フィルターに代入しながら、(1)で計算したスペクトル強度を掛け合わせ、(3)上記(2)で得られた値と平均疎水性値の和の値が連続して設定された閾値を超えた時、その閾値の存在個所によって特定された領域を、膜貫通ヘリックスの領域として出力することができる。

$$f(j) = \frac{1}{1 + \exp((j \pm 10)/2)}$$

【0030】

【発明の効果】本発明によれば、アミノ酸の一次配列が解明された蛋白質について、97%以上の正答率で、迅速に、被判定蛋白質が膜蛋白質であるか、水溶性蛋白質であるかを判定することができる。

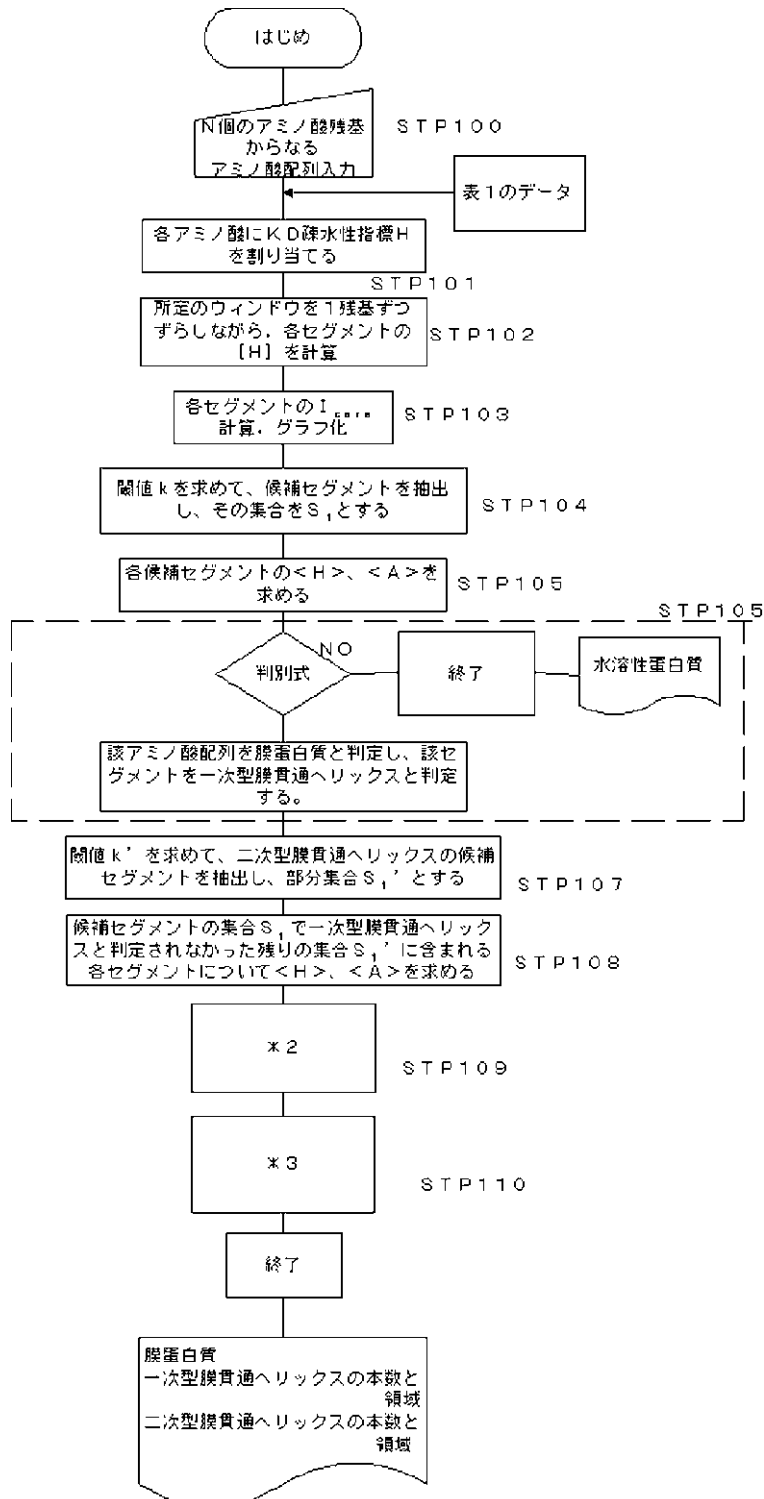
【図面の簡単な説明】

【図1】本発明のコンピュータプログラムのフローチャートの全体図の1例である。

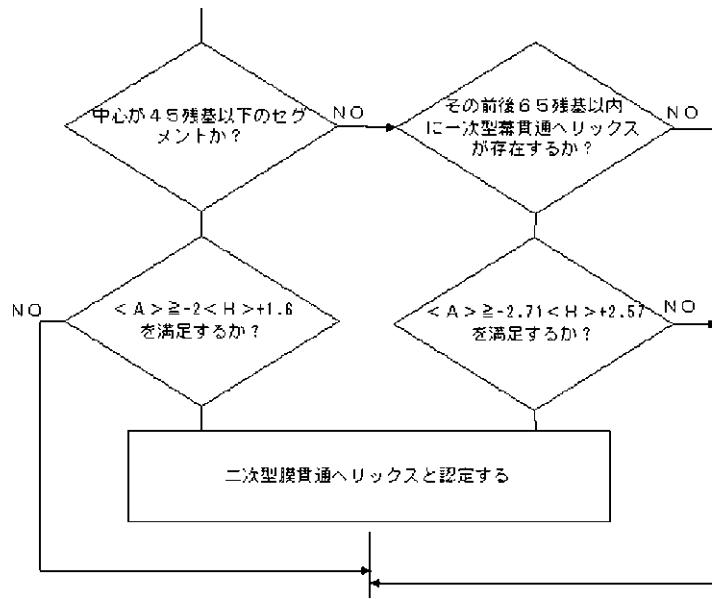
【図2】図1におけるSTP109のフローチャートである。

【図3】図1におけるSTP110のフローチャートである。

【 図 1 】



【図2】



【図3】

