

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002 - 286725

(P 2 0 0 2 - 2 8 6 7 2 5 A)

(43) 公開日 平成14年10月3日 (2002.10.3)

(51) Int.Cl. ⁷	識別記号	F I	テ-マコード [*]	(参考)
G01N 33/68		G01N 33/68		2G045
// G06F 17/30	170	G06F 17/30	170	F 5B075

審査請求 未請求 請求項の数 2 O L (全10頁)

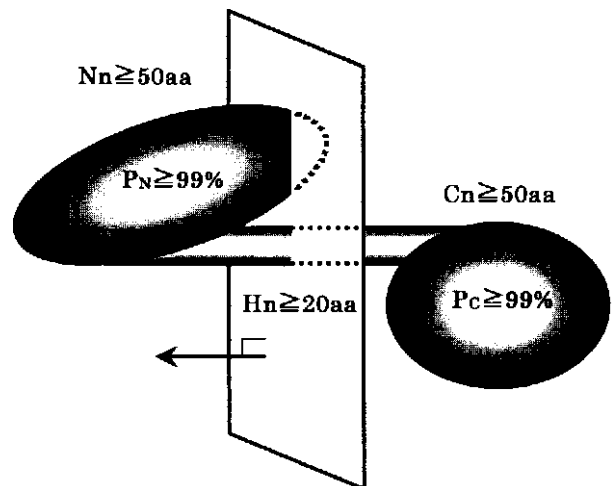
(21) 出願番号	特願2001 - 89619 (P 2001 - 89619)	(71) 出願人	396020800 科学技術振興事業団 埼玉県川口市本町 4 丁目 1 番 8 号
(22) 出願日	平成13年 3 月27日 (2001.3.27)	(72) 発明者	美宅 成樹 東京都小金井市中町 2 丁目24番16号 東京 農工大学工学部 生命工学科 内
		(72) 発明者	高橋 俊哉 東京都小金井市中町 2 丁目24番16号 東京 農工大学工学部 生命工学科 内
		(74) 代理人	100087631 弁理士 滝田 清暉 (外 1 名) F タ-ム(参考) 2G045 AA40 DA36 FB05 JA01 5B075 ND02 UU18

(54) 【発明の名称】 ダンベル型水溶性蛋白質の判別方法、及びその為のコンピュータプログラム

(57) 【要約】 (修正有)

【課題】 アミノ酸配列が決定された水溶液蛋白質について、それがダンベル型であるか否かを判断する方法、及びそのためのコンピュータプログラムの提供。

【解決手段】 蛋白質を構成する各アミノ酸に疎水性指標Hと両親媒性指標Aを割り当て、アミノ酸配列のCore領域を一回平均疎水性 [H] を介して算出し、得られたCore領域について二回平均疎水性 <H> 及び二回平均両親媒性<A> を算出する。次にCore領域の両端残基を除いた領域の [H] の最大値Maxと最小値Minを算出し、Max - Minが一定の閾値以下となる場合に、N端側のアミノ酸残基数N_nとC端側のアミノ酸残基数N_cを算出する。また、候補セグメントの総電荷指標C_Hと候補セグメントを除いたN端側の総電荷指標C_{N D}とC端側の総電荷指標C_{C D}を算出し、これらの総電荷指標が全て正であるか全て負である場合にその蛋白質をダンベル型水溶性蛋白質と判別する。



【特許請求の範囲】

【請求項 1】 アミノ酸配列の決定された水溶性蛋白質が、ダンベル型蛋白質であるか否かを判別する方法であって、該方法が、(a) 蛋白質の構成要素である 20 種のアミノ酸それぞれについて電荷指標 C を割り当てて該蛋白質の総電荷指標の絶対値 $|C_T|$ を算出し、得られた $|C_T|$ が一定の閾値を超える蛋白質については、(b) 蛋白質全体の電荷密度の絶対値 $|D|$ を算出し、該 $|D|$ が一定の閾値を超える蛋白質について、(c) 蛋白質のアミノ酸配列を半分に分割した後、分割されたそれぞれの総電荷指標 C_N 及び C_C を算出し、 $C_N \times C_C$ が一定の閾値以上となる蛋白質をダンベル型蛋白質の候補とし、(d) 該蛋白質を構成する各アミノ酸に疎水性指標 H と両親媒性指標 A を割り当て、(e) 次いで 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 と対応する全てのセグメントについて一回平均疎水性 $[H]$ 及び一回平均両親媒性 $[A]$ を計算し、(f) 前記アミノ酸配列の Core 領域を前記一回平均疎水性 $[H]$ を介して算出し、(g) 得られた Core 領域について二回平均疎水性 $\langle H \rangle$ 及び二回平均両親媒性 $\langle A \rangle$ を算出し、これらがそれぞれ一定の閾値を超える場合に、(h) Core 領域の両端残基を除いた領域の $[H]$ の最大値 Max と最小値 Min を算出し、 $Max - Min$ が一定の閾値以下となる場合に、(i) 中央ヘリックス候補セグメントを、Core 領域外の一平均平均疎水性 $[H]$ を介して算出し、(j) 得られた候補セグメントを除いた N 端側のアミノ酸残基数 N_N (アミノ酸の長さ) と C 端側のアミノ酸残基数 N_C (アミノ酸の長さ) を算出し、該 N_N と N_C が一定の関係を有する場合に、(k) 候補セグメントの総電荷指標 C_H と候補セグメントを除いた N 端側の総電荷指標 C_{N_D} と C 端側の総電荷指標 C_{C_D} を算出し、これらの総電荷指標が全て正であるか全て負である場合にその蛋白質をダンベル型水溶性蛋白質と判別することを特徴とするダンベル型水溶性蛋白質の判別方法。

【請求項 2】 蛋白質の構成要素である 20 種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 、及び電荷指標 C を用い、入力されたアミノ酸配列について、一回平均疎水性 $[H]$ 、二回平均疎水性 $\langle H \rangle$ 、二回平均両親媒性 $\langle A \rangle$ 、総電荷指標 $|C_T|$ 及び電荷密度 $|D|$ 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ダンベル型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、被判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) 予め記憶部に記録された、前記 20 種のアミノ酸に設定された前記疎水性指標 H 、両親媒性指標 A 及び電荷指標 C のデータのうち、電荷指標 C のデータを参照して、入力した全てのアミノ酸残基のそれぞれに前記電荷指標 C を割り当て、該蛋白質の総電荷指標の絶対値 $|C_T|$ を算出し、得られた $|C_T|$ が 20 以上で

あるか否かを判別し、20 未満であればプログラムを終了し、(b) 20 以上となる蛋白質については蛋白質全体の電荷密度の絶対値 $|D|$ を算出し、該 $|D|$ が 0.14 以上であるかを判別する；0.14 未満であればプログラムを終了し、(c) 0.14 以上を満足する蛋白質について、全アミノ酸配列を半分に分割した後、分割されたそれぞれの総電荷指標 C_N 及び C_C を算出し、 $C_N \times C_C$ が 1.00 以上である蛋白質をダンベル型蛋白質の候補とし、該蛋白質を構成する各アミノ酸に疎水性指標 H と両親媒性指標 A を割り当て、(d) 次いで 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 と対応する全てのセグメントについて一回平均疎水性 $[H]$ 及び一回平均両親媒性 $[A]$ を計算する；次に (e) 前記一回平均疎水性 $[H]$ が連続して 10 残基以上 - 0.25 以下となる領域を探索し、該当する領域が無ければプログラムを終了し、該当する領域がある場合には、その領域を Core 領域として定義する；(f) 得られた Core 領域について二回平均疎水性 $\langle H \rangle$ 及び二回平均両親媒性 $\langle A \rangle$ を算出し、 $\langle H \rangle$ の値が - 1.15 以下であると共に、 $\langle A \rangle$ の値が 0.90 以上となるか否かを判別し、この条件が満たされない場合にはプログラムを終了する；(g) 前記 (f) の条件が満たされた場合に、前記 Core 領域の両端 3 残基を除いた領域の一回平均疎水性 $[H]$ の最大値 Max と最小値 Min を算出し、 $Max - Min$ が 2.0 以下となるかを判別し、2.0 を超える場合にはプログラムを終了する；(h) 前記 (g) の条件が満たされた場合、Core 領域外の両端に向けて一回平均疎水性 $[H]$ を探索し、その値が - 0.30 以下になる直前のアミノ酸残基を、それぞれ中央ヘリックス候補セグメントの両末端とする；次に (i) 得られた候補セグメントを除いた N 端側のアミノ酸残基数 N_N (アミノ酸の長さ) と C 端側のアミノ酸残基数 N_C (アミノ酸の長さ) を算出し、 N_C (3/2) N_N 且つ N_C (2/3) N_N の判別条件を実行し、これらの条件が満たされない場合にはプログラムを終了する；(j) 上記判別条件を満たすセグメントが存在した場合、該セグメントを中央ヘリックスの候補セグメントとし、(k) 該候補セグメントの総電荷指標 C_H と候補セグメントを除いた N 端側の総電荷指標 C_{N_D} と C 端側の総電荷指標 C_{C_D} を算出し、これらの総電荷指標が全て正であるか全て負である場合にその蛋白質をダンベル型水溶性蛋白質と判断し、その中央ヘリックスを出力して、全プログラムを終了することを特徴とするダンベル型水溶性蛋白質を判別するためのコンピュータプログラム。

【発明の詳細な説明】

【0001】

【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水溶性蛋白質のコンピュータを用いた判別方法に関し、特に、コンピュータを用いて迅速且つ高い精度

でダンベル型水溶性蛋白質であるか否かを判別する方法、及びダンベル型水溶性蛋白質の中央ヘリックスについての情報を出力する方法、並びにそれを実施する為のコンピュータプログラムに関する。

【 0 0 0 2 】

【従来技術】蛋白質の立体構造の分類方法及び構造予測方法は既に数多く知られているが、水溶性蛋白質の二次構造予測の精度はあまり高くないというのが現状である。その原因の一つとして、二次構造の分類がうまくなされていないということが挙げられる。即ち、統計的・経験的手法による水溶性蛋白質の二次構造予測の正当率は7割を越えていないが、これは、二次構造が同じであれば、それだけで様に分類するということが一因と考えられる。

【 0 0 0 3 】例えば、水溶性蛋白質に含まれるヘリックスでの結合には、水相にむき出しになっているタイプのものであれば、内部コアで形成されているタイプのものや極性と非極性の界面付近でうまく安定しているタイプのものもある。つまり、構造的には同じヘリックスでも環境が異なる場合には蛋白質の立体構造が同一であるとはできないのであるが、統計的・経験的手法ではこれらと同じものとして分類せざるを得ないので予測精度が下がる。従って、このようなヘリックスの環境が異なるものを、異なるものとして分類することが予測精度の向上につながると推定される。

【 0 0 0 4 】本発明が対象としている、水相にむき出しになったヘリックスをもつダンベル型水溶性蛋白質についても状況は同様である。尚、ここでダンベル型蛋白質とは、二つの水溶性ドメインが、一本の長い中央ヘリックスによって結ばれた構造をしている蛋白質である。一般に、水溶性蛋白質の大部分は球状の形態をとることによって安定化しているが、ダンベル型蛋白質の場合には、二つの水溶性ドメインが同程度の大きな電荷を帯びて反発し合うことにより安定化していると考えられる。また、水溶性ドメインを結ぶ中央ヘリックスは、両親媒性残基が多く出現しそれらがヘリックスの安定化に寄与していると考えられる。

【 0 0 0 5 】そこで本発明者は、ダンベル型水溶性蛋白質と中央ヘリックスを高い精度で予測することについて鋭意検討した結果、蛋白質の大きさとその蛋白質を構成する各アミノ酸固有の電荷指標を用いて蛋白質を選別し、さらにKyte-Doolittleの疎水性指標（以下、K-D疎水性指標とする）と、新しく各アミノ酸に対して定義した両親媒性指標を用いて中央ヘリックス領域を予測したところ、極めて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。

【 0 0 0 6 】

【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、アミノ酸配列が決定された水溶性蛋白質がダンベル型水溶性蛋白質であるか否かを迅速且つ高い精度で

判別する方法を提供することにある。本発明の第2の目的は、ダンベル型水溶性蛋白質の中央ヘリックスを、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3の目的は、水溶性蛋白質の一次構造情報を用いて、その蛋白質がダンベル型水溶性蛋白質であるか否か、ダンベル型水溶性蛋白質である場合には更にその中央ヘリックスを、高い精度で判別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。

【 0 0 0 7 】

- 10 【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水溶性蛋白質が、ダンベル型蛋白質であるか否かを判別する方法であって、該方法が、(a) 蛋白質の構成要素である20種のアミノ酸それぞれについて電荷指標 C を割り当てて該蛋白質の総電荷指標の絶対値 $|C_T|$ を算出し、得られた $|C_T|$ が一定の閾値を超える蛋白質について、(b) 蛋白質全体の電荷密度の絶対値 $|D|$ を算出し、該 $|D|$ が一定の閾値を超える蛋白質について、(c) 蛋白質のアミノ酸配列を半分に分割した後、分割されたそれぞれの総電荷指標 C_N 及び C_C を算出し、 $C_N \times C_C$ が一定の閾値以上となる蛋白質をダンベル型蛋白質の候補とし、(d) 該蛋白質を構成する各アミノ酸に疎水性指標 H と両親媒性 A を割り当て、(e) 次いで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 と対応する全てのセグメントについて一回平均疎水性 $[H]$ 及び一回平均両親媒性 $[A]$ を計算し、(f) 前記アミノ酸配列のCore領域を前記一回平均疎水性 $[H]$ を介して算出し、(g) 得られたCore領域について二回平均疎水性 $\langle H \rangle$ 及び二回平均両親媒性 $\langle A \rangle$ を算出し、これらがそれぞれ一定の閾値を超える場合に、(h) Core領域の両端残基を除いた領域の $[H]$ の最大値 Max と最小値 Min を算出し、 $Max - Min$ が一定の閾値以下となる場合に、(i) 中央ヘリックス候補セグメントを、Core領域外の一回平均疎水性 $[H]$ を介して算出し、(j) 得られた候補セグメントを除いたN端側のアミノ酸残基数 N_N （アミノ酸の長さ）とC端側のアミノ酸残基数 N_C （アミノ酸の長さ）を算出し、該 N_N と N_C が一定の関係を有する場合に、(k) 候補セグメントの総電荷指標 C_H と候補セグメントを除いたN端側の総電荷指標 C_{N_D} とC端側の総電荷指標 C_{C_D} を算出し、これらの総電荷指標が全て正であるか全て負である場合にその蛋白質をダンベル型水溶性蛋白質と判別することを特徴とするダンベル型水溶性蛋白質の判別方法、及び、その方法を実施するコンピュータプログラムによって達成された。

【 0 0 0 8 】

- 【発明の実施の形態】一般に、水溶性蛋白質の大部分は球状の形態をとることによって安定化しているが、ダンベル型蛋白質は球状の形態をとっていない。一方、ダンベル型蛋白質は、蛋白質の総電荷指標が大きく、電荷密度が高いという特徴と共に二つのドメインが同程度の大

きな電荷を帯びているという特徴がある。従って、ダンベル型蛋白質の構造安定化には、蛋白質の電荷を考慮することが重要である。

【 0 0 0 9 】また、極性基があると共に炭化水素部分が大きい場合には、一般に分子の両親媒性が高く、これによって極性と非極性の界面が安定化される。これと同様に、側鎖自体の両親媒性が高いと水に露出した中央ヘリ

ックスの構造を安定化し易いと考えられる。本発明者は、このことを用いて、新たに下表表1に示す両親媒性指標Aを作成した。なお、Hは疎水性指標、Cは電荷指標であり、()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 0 】

【表1】

アミノ酸	H	A	αアミノ酸	H	A	C
ILE (I)	4.5	0	TRP (W)	-0.9	0	0
VAL (V)	4.2	0	TYR (Y)	-1.3	0	0
LEU (L)	3.8	0	PRO (P)	-1.6	0	0
PHE (F)	2.8	0	HIS (H)	-3.2	1.45	1
CYS (C)	2.5	0	ASP (D)	-3.5	0	-1
MET (M)	1.9	0	ASN (N)	-3.5	0	0
ALA (A)	1.8	0	GLU (E)	-3.5	1.27	-1
GLY (G)	-0.4	0	GLN (Q)	-3.5	1.25	0
THR (T)	-0.7	0	LYS (K)	-3.9	3.67	1
SER (S)	-0.8	0	ARG (R)	-4.5	2.45	1

【 0 0 1 1 】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、ダンベル型蛋白質を次のように定義する(図1参照)。

(1) PDBに記述されているヘリックス領域のアミノ酸残基[H_N]が20残基以上である。

(2) (1)のヘリックス両端のアミノ酸残基[N_N][C_N]が50残基以上である。

(3) (1)のヘリックスの中央残基を通り、ヘリックス末端2残基のC 原子を結ぶベクトルを法線ベクトルとする平面を考えたとき、ヘリックスN(C)端側の全C 原子がN(C)端側の空間に存在する割合[P_N]([P_C])を計算し、それが99%以上である。

(4) 蛋白質のサイズが大きいとドメインがいくつかあることが考えられるので、総残基数が500残基以内である。

【 0 0 1 2 】従って、本発明においては、蛋白質を構成するアミノ酸残基数(アミノ酸鎖の長さ)Nが100以上500以下の蛋白質を判別の対象とする。尚、100残基以上としたのは、ダンベル型蛋白質の定義が120残基以上であること、及び、100残基未満の場合にはペプチドや蛋白質断片が多く含まれるからである。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電荷指標Cを割り当て、下記(1)式によって総電荷指標の絶対値|C_T|を計算する。

$$C_T = \sum_{i=1}^N C(i)$$

(1)

但し、Nはアミノ酸残基数(アミノ酸鎖の長さ)である。

【 0 0 1 3 】この総電荷指標の絶対値|C_T|が20以上であるか否かを判別し、|C_T|が20以上である場合には下記(2)式によって蛋白質全体の電荷密度の絶対値|D|を計算する。

$$D = C_T / N$$

(2)

尚、閾値を20とした理由は、ダンベル型蛋白質は総電荷指標の絶対値が20以上の大きな電荷を帯びており、総電荷指標の絶対値が20以上の蛋白質の全体に占める割合は1.5%程度である。

【 0 0 1 4 】次に、この蛋白質全体の電荷密度の絶対値|D|が0.14以上であるか否かを判別し、上記|D|が0.14以上である場合には、対象としているアミノ酸配列を半分に分割し、それぞれの総電荷指標C_N、C_Cを、下記(3-1)及び(3-2)式によって計算する。

$$C_N = \sum_{i=1}^{N/2} C(i)$$

(3 - 1)

$$C_c = \sum_{i=N/2+1}^N C(i)$$

(3 - 2)

但し、 $C(i)$ は*i*番目のアミノ酸残基の電荷、 N は全アミノ酸残基の数（アミノ酸鎖の長さ）である。 N が奇数のときは $N/2$ の小数を四捨五入する。尚、既知のダンベル型蛋白質の解析から、負の単位残基当たりの電荷は0.14以上と大きく、0.14以上の蛋白質の全体に占める割合は0.4%程度であることが分かっている。

【0015】上記の C_N 、 C_c より $C_N \times C_c$ を計算し、その値が1.00以上であるか否かを判別し、 $C_N \times C_c$ が1.00以上である場合には、対象としてい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疎水性指標 H 及び両親媒性指標 A を割り当てる。尚、既知のダンベル型蛋白質の解析から、負の単位残基当たりの電荷は1.0以上と大きく、1.0以上の蛋白質の全体に占める割合は0.3%程度であることが分かっている。

【0016】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、下記(4-1)及び(4-2)式によって一回平均疎水性 $[H]$ 及び一回平均両親媒性 $[A]$ を計算する。

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

(4 - 1)

$$[A(i)] = \left[\sum_{j=i-W/2}^{i+W/2} A(j) \right] / W$$

(4 - 2)

但し、 i はウィンドウの中心のアミノ酸の座標である。

【0017】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。次に、連続10残基以上が、一回平均

疎水性 $[H]$ として-0.25以下をとる領域 (Core領域とする) があるか否かを判別し、 $[H]$ が連続10残基以上で-0.25以下であるCore領域があればその領域を抽出する。尚、連続10残基以上とした理由は、中央ヘリックスの疎水性の低い領域が10残基より長いからである。

【0018】上記の如くして抽出されたCore領域に対して、下記(5-1)、(5-2)式によって、二回平均疎水性 $\langle H \rangle$ 及び 二回平均両親媒性 $\langle A \rangle$ を計算する。

$$\langle H \rangle = \left[\sum_{i=S}^E [H(i)] \right] / N_{core}$$

(5 - 1)

但し、 $[H]$ は一回平均疎水性、 S 及び E はCore領域の開始

残基と終始残基である。また、 N_{core} はCore領域の残基数 ($E-S+1$) である。

$$\langle A \rangle = \left[\sum_{i=S}^E [A(i)] \right] / N_{core}$$

(5 - 2)

但し、[A] は一回平均両親媒性、S及びEはCore領域の開始残基と終始残基である。また、 N_{core} はCore領域の残基数 (E-S+1) である。

【 0 0 1 9 】次に、このCore領域の二回平均疎水性 <H>の値が-1.15以下であると共に、二回平均両親媒性 <A>の値が0.90以上であるか否かを判別する。尚、これらの閾値は既知のダンベル型蛋白質の解析に基づいて設定したものである。上記の2条件が満たされた場合には、抽出されたCore領域に対して、両端3残基を除いた領域の一回平均疎水性 [H] に対して最大値Maxと最小値Minを求める。次いで、Max-Minを計算し、その値が2.0以下となるか否かを判断する。

【 0 0 2 0 】Max-Minの値が2.0以下であると判断されたCore領域に対して、その領域外両端に一回平均疎水性 [H] の値を探索していき、-0.30以下になる直前のアミノ酸残基をそれぞれ中央ヘリックス候補セグメントの末端とする。上記の中央ヘリックス候補セグメントに対して、そのセグメントを除いたN端側の総残基数 N_N とC端

側の総残基数 N_C を計算し、次式の判別式を実行する。尚、N端はアミノ基末端、C端はカルボキシル基末端を意味する。

$$10 \quad N_C = (3/2) \times N_N \quad \text{かつ} \quad N_C = (2/3) \times N_N$$

上記判別式を満たすセグメントを抽出する。
【 0 0 2 1 】上記の如くして抽出された中央ヘリックス候補セグメントに対して、下記 (6-1) 式によって総電荷指標 C_H を計算する。また、セグメントを除いたN端側の総電荷指標 C_{ND} とC端側の総電荷指標 C_{CD} を下記 (6-2)、(6-3) 式によって計算する。

$$20 \quad C_H = \sum_{i=S}^E C(i)$$

(6 - 1)

$$C_{ND} = \sum_{i=1}^{S-1} C(i)$$

(6 - 2)

$$C_{CD} = \sum_{i=E+1}^N C(i)$$

(6 - 3)

但し、C(i)はアミノ酸残基iの電荷指標、S及びEは中央ヘリックス候補セグメントの開始残基と終始残基である。Nはアミノ酸鎖の長さ (アミノ酸の長さ) である。

【 0 0 2 2 】次に、次式の判別式を実行し、判別式を満たすセグメントを抽出する。

$$C_H > 0 \quad \text{かつ} \quad C_{ND} > 0 \quad \text{かつ} \quad C_{CD} > 0$$

$$\text{あるいは} \quad C_H < 0 \quad \text{かつ} \quad C_{ND} < 0 \quad \text{かつ} \quad C_{CD} < 0$$

< 0

上記の如くして抽出された中央ヘリックス候補セグメントを中央ヘリックス領域とし、その領域を含む蛋白質をダンベル型水溶性蛋白質と判別する。

【 0 0 2 3 】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2~4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言語などを用いて記載することができる。以下に更に詳述する。

【 0 0 2 4 】本発明のプログラムを起動し、被判断蛋白質を形成するN個のアミノ酸列を入力すると、記憶部に貯えられている前記表1のデータと照らし合わされ、各アミノ酸に、それぞれ固有の電荷指標が割り振られる (S T P 1 0 0)。次いで、前記 (1) 式によって総電荷指標の絶対値 $|C_T|$ が計算される (S T P 1 0 1)。次に、得られた $|C_T|$ が 2 0 以上であるか否かを判別し (S T P 1 0 2)、2 0 未満であればプログラムを終了し、必要に応じて新しい被判断蛋白質について、新たに操作を開始する。

【 0 0 2 5 】 $|C_T|$ が 2 0 以上である場合には、 (2) 式によって蛋白質全体の電荷密度の絶対値 $|D|$ を計算する (S T P 1 0 3)。但し、 (2) 式中のNは、被判断蛋白質の全アミノ酸残基数 (アミノ酸鎖の長さ) である。次に $|D|$ が 0 . 1 4 以上であるか否かを判別する (S T P 1 0 4)。 $|D|$ が 0 . 1 4 未満である場合にはプログラムは終了する。 $|D|$ が 0 . 1 4 以上である場合には、N個からなる全アミノ酸配列を半分に分割し、分割されたそれぞれについて、総電荷指標 C_N (N 端側) 及び C_C (C 端側) を計算する (S T P 1 0 5)。

【 0 0 2 6 】次に、上記 C_N と C_C の積 $C_N \times C_C$ が 1 0 0 以上か否かを判別し (S T P 1 0 6)、1 0 0 未満であればプログラムを終了する。 $C_N \times C_C$ が 1 0 0 以上の場合には、表 1 に照らし合わせて、N個の全アミノ酸にそれぞれ固有の K - D 疎水性指標 H と両親媒性指標 A を割り当てる (S T P 1 0 7)。これらの割り当ては、S T P 1 0 0 で予め行っておいても良い。7 残基ウィンドウを用い、前記式 (4 - 1) によって一回平均疎水性 $[H]$ を計算する (S T P 1 0 8)。

【 0 0 2 7 】次に、 $[H]$ の値が連続 1 0 残基以上 - 0 . 2 5 以下となる C o r e 領域が存在するか否かを判別し (S T P 1 0 9)、存在しなければプログラムを終了する。C o r e 領域が存在する場合には、該 C o r e 領域に対して、前記式 (5 - 1) によって二回平均疎水性 $\langle H \rangle$ を計算すると共に、式 (5 - 2) によって二回平均両親媒性 $\langle A \rangle$ を計算する (S T P 1 1 0)。尚、S T P 1

0 8 における $[A]$ の計算は、S T P 1 0 8 では行わずに S T P 1 1 0 の $\langle A \rangle$ の計算の前に行うようにし、 $[A]$ から $\langle A \rangle$ 迄の計算を一貫して行わせても良い。

【 0 0 2 8 】次に、 $\langle H \rangle - 1 . 1 5$ 且つ $\langle A \rangle - 0 . 9 0$ が満足されるか否かを判別し、満足されない場合にはプログラムを終了する。満足される場合には、C o r e 領域の両端 3 残基を除いた領域について、 $[H]$ の最大値 Max と最小値 Min を求める (S T P 1 1 2)。次いで Max - Min が 2 . 以下であるか否かを判別し (S T P 1 1 3)、2 . 0 を超えればプログラムを終了する。

【 0 0 2 9 】 Max - Min が 2 . 0 以下となる場合には、C o r e 領域外の両端に向って $[H]$ の値を探索していき、 - 0 . 3 0 以下になる直前の残基を候補セグメント末端と決定する。 (S T P 1 1 4)。次に候補セグメントを除いた N 端側総残基数 N_N と C 端側総残基数 N_C を計算し (S T P 1 1 5)、 $N_C - 3 / 2 \times N_N$ かつ $N_C - 2 / 3 \times N_N$ を満足するか否かを判別する (S T P 1 1 6)。上記の条件が満たされない場合にはプログラムを終了する。

【 0 0 3 0 】上記の条件が満足される場合、候補セグメントの総電荷指標 C_H 、候補セグメントを除いた N 端側の総電荷指標 $C_{N D}$ 、及び、C 端側の総電荷指標 $C_{C D}$ を計算する (S T P 1 1 7)。次に、 C_H 、 $C_{N D}$ 及び $C_{C D}$ が何れも正又は負となるか判断する (S T P 1 1 8)。これらの指標の符号が一致して正又は負とならない場合には、プログラムは終了する。上記の条件が満足される場合には、候補セグメントをダンベル型水溶性蛋白質の中央ヘリックスと判断し (S T P 1 1 9)、ダンベル型水溶性蛋白質であるとの判定と中央ヘリックス領域を出力し、プログラムを終了する。尚、本プログラム中で行った判断の各結果を同時に出力するようにできることは当然である。

【 0 0 3 1 】下記表 2 に、1 2 個のダンベル型蛋白質について、本発明のコンピュータプログラムを用いて解析した結果を示す。

【表 2】

PDB ID	Length			Total charges	Net charges				
	N-terminal domain	Center domain	C-terminal domain		N-terminal domain	Central helix	C-terminal domain	N-terminal half	C-terminal half
ICLL	64	28	56	-23	-12	-2	-9	-12	-11
ICLM	64	28	56	-22	-12	-3	-7	-12	-10
INCX	74	31	57	-28	-14	-2	-12	-16	-12
INCY	74	31	57	-28	-14	-2	-12	-16	-12
INCZ	74	31	57	-28	-14	-2	-12	-16	-12
IOSA	64	28	56	-22	-12	-3	-7	-12	-10
1TN4	71	31	57	-28	-14	-3	-11	-16	-12
ITOP	74	31	57	-28	-14	-2	-12	-16	-12
2TN4	71	31	57	-28	-14	-3	-11	-16	-12
3CLN	64	28	56	-22	-12	-2	-8	-12	-10
4CLN	64	28	56	-23	-14	-2	-9	-12	-11
4TNC	74	31	57	-29	-14	-3	-12	-16	-13

【 0 0 3 2 】

【発明の効果】本発明はダンベル型水溶液蛋白質の二次構造の特徴を考慮し、新たに各アミノ酸に対して設定し

た、電化荷指標、K-D疎水性指標、及び両親媒性指標を用いて、特にダンベル型水溶性蛋白質に絞って判別を行うので、従来になく、迅速且つ精度良く判別することが

できる。

【図面の簡単な説明】

【図1】ダンベル型蛋白質の定義を表す図である。

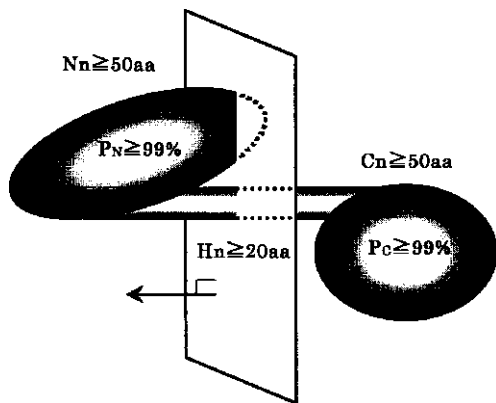
【図2】本発明のコンピュータプログラムのフローチャートの内、ダンベル型蛋白質の候補を抽出するための部分フローチャートである。

【図3】本発明のコンピュータプログラムのフローチャ

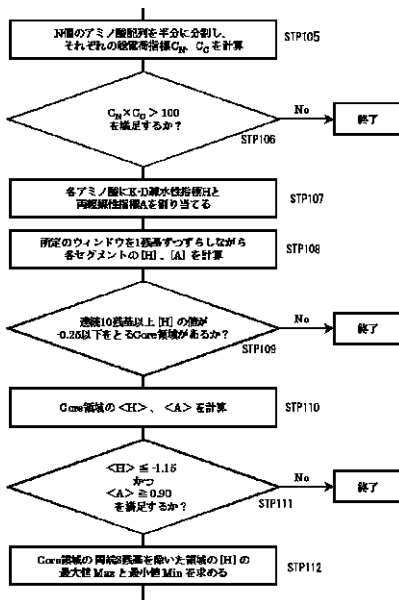
ートの内、ダンベル型蛋白質の候補について、中央ヘリックスの候補となるセグメントを抽出するための部分フローチャートである。

【図4】本発明のコンピュータプログラムのフローチャートの内、中央ヘリックスとダンベル型蛋白質についての最終判断をするための部分フローチャートである。

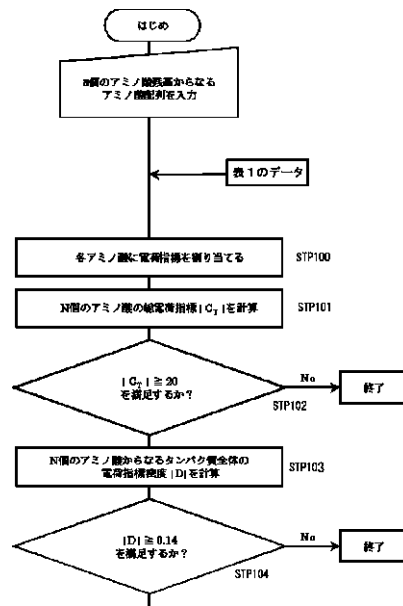
【図1】



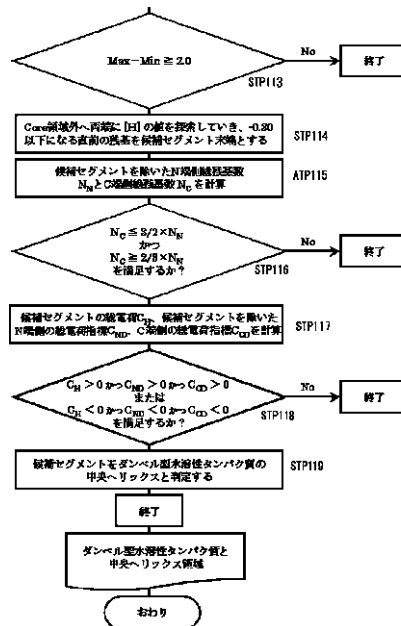
【図3】



【図2】



【図4】



【手続補正書】

【提出日】平成13年3月27日(2001.3.27)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】0031

【補正方法】変更

【補正内容】

【0031】下記表2に、12個のダンベル型蛋白質について、本発明のコンピュータプログラムを用いて解析した結果を示す。

【表2】

PDB ID	アミノ酸長さ			総電荷	正味の電荷				
	N端側ドメイン	中央ヘリックス	C端側ドメイン		N端側ドメイン	中央ヘリックス	C端側ドメイン	N端側の半分	C端側の半分
1CLL	64	28	56	-23	-12	-2	-9	-12	-11
1CLM	64	28	56	-22	-12	-3	-7	-12	-10
1NCX	74	31	57	-28	-14	-2	-12	-16	-12
1NCY	74	31	57	-28	-14	-2	-12	-16	-12
1NCZ	74	31	57	-28	-14	-2	-12	-16	-12
1OSA	64	28	56	-22	-12	-3	-7	-12	-10
1TN4	71	31	57	-28	-14	-3	-11	-16	-12
1TOP	74	31	57	-28	-14	-2	-12	-16	-12
2TN4	71	31	57	-28	-14	-3	-11	-16	-12
3CLN	64	28	56	-22	-12	-2	-8	-12	-10
4CLN	64	28	56	-23	-14	-2	-9	-12	-11
4TNC	74	31	57	-29	-14	-3	-12	-16	-13

【手続補正2】

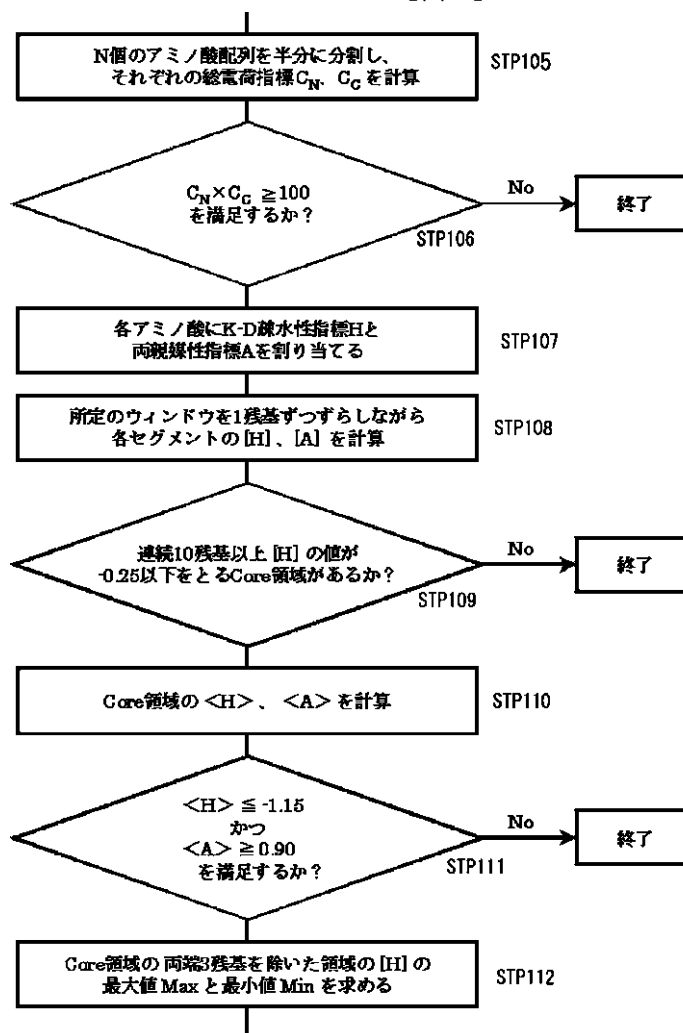
【補正対象書類名】図面

【補正対象項目名】図3

【補正方法】変更

【補正内容】

【図3】



【手続補正3】

【補正対象書類名】図面

【補正対象項目名】図 4
【補正方法】変更

【補正内容】
【図 4】

