

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

特許第4608698号  
(P4608698)

(45) 発行日 平成23年1月12日 (2011. 1. 12)

(24) 登録日 平成22年10月22日 (2010. 10. 22)

(51) Int. Cl. F I  
**G 0 6 F 19/18 (2011. 01)** G O 6 F 19/00 6 1 8  
**G 0 6 F 19/00 (2011. 01)** G O 6 F 19/00 Z N A

請求項の数 15 (全 35 頁)

<p>(21) 出願番号 特願2009-209184 (P2009-209184)                  (22) 出願日 平成21年9月10日 (2009. 9. 10)                  審査請求日 平成22年4月15日 (2010. 4. 15)                   早期審査対象出願</p>	<p>(73) 特許権者 801000027                  学校法人明治大学                  東京都千代田区神田駿河台 1 - 1                  (74) 代理人 100064908                  弁理士 志賀 正武                  (74) 代理人 100106909                  弁理士 棚井 澄雄                  (74) 代理人 100108578                  弁理士 高橋 詔男                  (74) 代理人 100126882                  弁理士 五十嵐 光永                  (72) 発明者 池田 有理                  神奈川県川崎市多摩区東三田 1 - 1 - 1                  学校法人明治大学 生田校舎内</p>
---	--

最終頁に続く

(54) 【発明の名称】 GPI アンカー型タンパク質の判定装置、判定方法及び判定プログラム

(57) 【特許請求の範囲】

【請求項 1】

検査対象タンパク質が GPI アンカー型タンパク質であるか否かを判定する GPI アンカー型タンパク質の判定装置であって、

前記検査対象タンパク質のアミノ酸配列情報を取得する配列取得部と、

前記配列取得部が取得したアミノ酸配列情報における既知の GPI アンカー型タンパク質の N 末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報の N 末端から予め定められた残基数の領域を特定し、当該 N 末端側の高疎水性領域に対応する領域のアミノ酸残基を抽出し、前記 N 末端側の高疎水性領域に対応する領域のアミノ酸残基の疎水性値の平均化に用いる残基数である N 末端側疎水性特性抽出必要数を用いて、連続する当該 N 末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均である N 末端側平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して 1 残基ずつずらしながら複数算出する N 末端側疎水性値算出部と、

前記 N 末端側疎水性値算出部が算出した複数の N 末端側平均疎水性値のうちの最大値が、既知の GPI アンカー型タンパク質における N 末端側平均疎水性値の特性を示す N 末端側疎水性閾値以上であるか否かを判定する N 末端側疎水性判定部と、

前記配列取得部が取得したアミノ酸配列情報のうち前記 N 末端側疎水性値算出部が N 末端側平均疎水性値を算出した範囲以外の領域として、前記アミノ酸配列情報の前記 N 末端から予め定められた残基数のアミノ酸残基以外の領域を特定し、当該 N 末端側平均疎水性値を算出した範囲以外のアミノ酸残基を抽出し、前記 N 末端側平均疎水性値を算出した範

10

20

囲以外のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端外疎水性特性抽出数を用いて、連続する当該N末端外疎水性特性抽出数分のアミノ酸残基の各疎水性指標値の平均であるN末端外平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端外疎水性値算出部と、

前記N末端外疎水性値算出部が算出した複数のN末端外平均疎水性値のうちの最大値が、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示すN末端外疎水性閾値以上であるか否かを判定するN末端外疎水性判定部と、

既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、前記N末端外疎水性値算出部が算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が当該特定した領域内にあるか否かを判定するC末端側最大疎水位置判定部と、

前記配列取得部が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のプロペプチド領域を含む領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、当該プロペプチド領域を含む領域のアミノ酸残基を抽出し、当該抽出したアミノ酸残基のそれぞれに対して、当該プロペプチド領域を含む領域のアミノ酸残基の側鎖サイズの平均化に用いる残基数である側鎖サイズ特性抽出必要数を用いて、連続する当該側鎖サイズ特性抽出必要数分のアミノ酸残基の各側鎖サイズ指標値の平均値である平均側鎖サイズを1残基ずつずらしながら複数算出する側鎖サイズ算出部と、

前記側鎖サイズ算出部が算出した平均側鎖サイズが最小となる位置を基準位置とする、当該基準位置からN末端側及びC末端側に連続する所定の残基数のアミノ酸残基からなる所定の領域におけるアミノ酸残基の部分配列と、既知のGPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列と、の類似度を示すスコアを算出するにあたり、記憶部に記憶されている既知のGPIアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを前記記憶部から取得し、当該位置特異的スコアに基づき、前記部分配列の各アミノ酸残基の位置特異的スコアを特定して数値列を生成し、当該数値列の平均値を算出し、当該算出した平均値を前記スコアとして算出するスコア算出部と、

前記スコア算出部が算出したスコアが、前記検査対象タンパク質と既知のGPIアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定するスコア判定部と、

前記N末端側疎水性判定部が、前記N末端側疎水性値算出部の算出したN末端側平均疎水性値の最大値が前記N末端側疎水性閾値未満であると判定した場合、前記N末端外疎水性判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値の最大値が前記N末端外疎水性閾値未満であると判定した場合、前記C末端側最大疎水位置判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が前記既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内にないと判定した場合、または前記スコア判定部が、前記スコア算出部の算出したスコアが前記スコア判定閾値未満であると判定した場合に、前記検査対象タンパク質がGPIアンカー型タンパク質でないとして判定するGPIアンカー型タンパク質判定部と、

を備えることを特徴とするGPIアンカー型タンパク質の判定装置。

#### 【請求項2】

前記N末端側疎水性閾値は、

予め既知の複数のGPIアンカー型タンパク質に対して前記N末端側平均疎水性値の算出を行い、当該算出されたN末端側平均疎水性値の最大値の集合における最小値であることを特徴とする請求項1に記載のGPIアンカー型タンパク質の判定装置。

#### 【請求項3】

前記N末端外疎水性閾値は、

予め既知の複数のGPIアンカー型タンパク質に対して前記N末端外平均疎水性値の算出を行い、当該算出されたN末端外平均疎水性値の最大値の集合における最小値であることを特徴とする請求項1または請求項2の何れか1項に記載のGPIアンカー型タン

10

20

30

40

50

パク質の判定装置。

【請求項4】

前記既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域は、既知のGPIアンカー型タンパク質において、前記N末端側平均疎水性値が最大となる位置が含まれる領域である、

ことを特徴とする請求項1から請求項3の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項5】

前記既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域は、既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域以外の領域において、前記N末端外平均疎水性値が最大となる位置が含まれる領域である、

ことを特徴とする請求項1から請求項4の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項6】

前記N末端側疎水性特性抽出必要数は、

当該N末端側疎水性特性抽出必要数を用いて、既知の複数のGPIアンカー型タンパク質のN末端側の高疎水性領域のアミノ酸残基のそれぞれに対してN末端側平均疎水性値を算出し、前記既知のGPIアンカー型タンパク質から算出したN末端側平均疎水性値の最大値の集合における最小値を抽出し、前記N末端側疎水性特性抽出必要数を用いて、既知の複数の非GPIアンカー型タンパク質における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域のアミノ酸残基のそれぞれに対してN末端側平均疎水性値を算出した場合に、前記既知の非GPIアンカー型タンパク質から算出したN末端側平均疎水性値の最大値のうち、前記抽出した最小値より値が大きいものの個数が最小となるような値である

ことを特徴とする請求項1から請求項5の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項7】

前記N末端外疎水性特性抽出必要数は、

当該N末端外疎水性特性抽出必要数を用いて、既知の複数のGPIアンカー型タンパク質のN末端側の高疎水性領域以外の領域のアミノ酸残基のそれぞれに対してN末端外平均疎水性値を算出し、前記既知のGPIアンカー型タンパク質から算出したN末端外平均疎水性値の最大値の集合における最小値を抽出し、前記N末端外疎水性特性抽出必要数を用いて、既知の複数の非GPIアンカー型タンパク質における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域以外の領域のアミノ酸残基のそれぞれに対してN末端外平均疎水性値を算出した場合に、前記既知の非GPIアンカー型タンパク質から算出したN末端外平均疎水性値の最大値のうち、前記抽出した最小値より値が大きいものの個数が最小となるような値である

ことを特徴とする請求項1から請求項6の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項8】

前記小側鎖サイズ判定領域は、

既知のGPIアンカー型タンパク質の前記平均側鎖サイズが最小となる位置が含まれる領域である、

ことを特徴とする請求項1から請求項7に何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項9】

前記側鎖サイズ特性抽出必要数は、

当該側鎖サイズ特性抽出必要数を用いて、既知の複数のGPIアンカー型タンパク質の小側鎖サイズ判定領域に対して平均側鎖サイズを算出した場合に、前記GPIアンカー型タンパク質から算出した平均側鎖サイズが最小となるアミノ酸残基のうち、当該アミノ酸

10

20

30

40

50

残基のC末端側に隣接するアミノ酸残基がGPIアンカー修飾部位であるものの個数が最大となるような値である

ことを特徴とする請求項1から請求項8の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項10】

前記位置特異的スコアは、

既知の複数のGPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域内の位置pに存在するアミノ酸残基の種類iの出現頻度を示す $f_{ip}^{positive}$ 、既知の複数の非GPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域内の位置pに存在するアミノ酸残基の種類iの出現頻度を示す $f_{ip}^{negative}$ を用いて、

10

【数1】

$$\ln \frac{f_{ip}^{positive}}{f_{ip}^{negative}}$$

から算出されたものであることを特徴とする請求項1から請求項9の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

20

【請求項11】

前記所定の領域内の位置pに存在するアミノ酸残基の種類iの出現頻度は、

種類iのアミノ酸残基が位置pに存在する既知のGPIアンカー型タンパク質の個数を示す $n_{ip}$ と、当該出現頻度の調整値を示す $\epsilon$ と、アミノ酸残基の種類数sとを用いて、

【数2】

$$\frac{n_{ip} + \epsilon}{\sum_{i=1}^s n_{ip} + \epsilon}$$

30

から算出されたものであることを特徴とする請求項10に記載のGPIアンカー型タンパク質の判定装置。

【請求項12】

前記スコア判定閾値は、

既知の複数のGPIアンカー型タンパク質について算出した前記スコアの最小値と、既知の複数の非GPIアンカー型タンパク質について算出した前記スコアの最大値との間の値である

40

ことを特徴とする請求項1から請求項11の何れか1項に記載のGPIアンカー型タンパク質の判定装置。

【請求項13】

前記スコア判定閾値は、

前記スコアが当該スコア判定閾値以上である既知のGPIアンカー型タンパク質の個数を示すTP、前記スコアが当該スコア判定閾値以上である既知の非GPIアンカー型タンパク質の個数を示すFN、前記スコアが当該スコア判定閾値未満である既知のGPIアンカー型タンパク質の個数を示すFPを用いて、

【数 3】

$$\sqrt{\frac{TP}{TP + FN} \times \frac{TP}{TP + FP}}$$

から算出される値が最大となるような値である

こと特徴とする請求項 1 2 に記載の G P I アンカー型タンパク質の判定装置。

【請求項 1 4】

検査対象タンパク質が G P I アンカー型タンパク質であるか否かを判定する G P I アンカー型タンパク質の判定装置を用いた判定方法であって、

G P I アンカー型タンパク質の判定装置の配列取得部は、前記検査対象タンパク質のアミノ酸配列情報を取得し、

G P I アンカー型タンパク質の判定装置の N 末端側疎水性値算出部は、前記配列取得部が取得したアミノ酸配列情報における既知の G P I アンカー型タンパク質の N 末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報の N 末端から予め定められた残基数の領域を特定し、当該 N 末端側の高疎水性領域に対応する領域のアミノ酸残基を抽出し、前記 N 末端側の高疎水性領域に対応する領域のアミノ酸残基の疎水性値の平均化に用いる残基数である N 末端側疎水性特性抽出必要数を用いて、連続する当該 N 末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均である N 末端側平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して 1 残基ずつずらしながら複数算出し、

G P I アンカー型タンパク質の判定装置の N 末端側疎水性判定部は、前記 N 末端側疎水性値算出部が算出した複数の N 末端側平均疎水性値のうちの最大値が、既知の G P I アンカー型タンパク質における N 末端側平均疎水性値の特性を示す N 末端側疎水性閾値以上であるか否かを判定し、

G P I アンカー型タンパク質の判定装置の N 末端外疎水性値算出部は、前記配列取得部が取得したアミノ酸配列情報のうち前記 N 末端側疎水性値算出部が N 末端側平均疎水性値を算出した範囲以外の領域として、前記アミノ酸配列情報の前記 N 末端から予め定められた残基数のアミノ酸残基以外の領域を特定し、当該 N 末端側平均疎水性値を算出した範囲以外のアミノ酸残基を抽出し、前記 N 末端側平均疎水性値を算出した範囲以外のアミノ酸残基の疎水性値の平均化に用いる残基数である N 末端外疎水性特性抽出数を用いて、連続する当該 N 末端外疎水性特性抽出数分のアミノ酸残基の各疎水性指標値の平均である N 末端外平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して 1 残基ずつずらしながら複数算出し、

G P I アンカー型タンパク質の判定装置の N 末端外疎水性判定部は、前記 N 末端外疎水性値算出部が算出した複数の N 末端外平均疎水性値のうちの最大値が、既知の G P I アンカー型タンパク質における N 末端外平均疎水性値の特性を示す N 末端外疎水性閾値以上であるか否かを判定し、

G P I アンカー型タンパク質の判定装置の C 末端側最大疎水位置判定部は、既知の G P I アンカー型タンパク質の C 末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報の C 末端から予め定められた残基数の領域を特定し、前記 N 末端外疎水性値算出部が算出した N 末端外平均疎水性値が最大となるアミノ酸残基の位置が当該特定した領域内にあるか否かを判定し、

G P I アンカー型タンパク質の判定装置の側鎖サイズ算出部は、前記配列取得部が取得したアミノ酸配列情報における既知の G P I アンカー型タンパク質のプロペプチド領域を含む領域として、前記アミノ酸配列情報の C 末端から予め定められた残基数の領域を特定し、当該プロペプチド領域を含む領域のアミノ酸残基を抽出し、当該抽出したアミノ酸残基のそれぞれに対して、当該プロペプチド領域を含む領域のアミノ酸残基の側鎖サイズの平均化に用いる残基数である側鎖サイズ特性抽出必要数を用いて、連続する当該側鎖サイ

10

20

30

40

50

ズ特性抽出必要数分のアミノ酸残基の各側鎖サイズ指標値の平均値である平均側鎖サイズを1残基ずつずらしながら複数算出し、

G P Iアンカー型タンパク質の判定装置のスコア算出部は、前記側鎖サイズ算出部が算出した平均側鎖サイズが最小となる位置を基準位置とする、当該基準位置からN末端側及びC末端側に連続する所定の残基数のアミノ酸残基からなる所定の領域におけるアミノ酸残基の部分配列と、既知のG P Iアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列との類似度を示すスコアを算出するにあたり、記憶部に記憶されている既知のG P Iアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを前記記憶部から取得し、当該位置特異的スコアに基づき、前記部分配列の各アミノ酸残基の位置特異的スコアを特定して数値列を生成し、当該数値列の平均値を算出し、当該算出した平均値を前記スコアとして算出し、

10

G P Iアンカー型タンパク質の判定装置のスコア判定部は、前記スコア算出部が算出したスコアが、前記検査対象タンパク質と既知のG P Iアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定し、

G P Iアンカー型タンパク質の判定装置のG P Iアンカー型タンパク質判定部は、前記N末端側疎水性判定部が、前記N末端側疎水性値算出部の算出したN末端側平均疎水性値の最大値が前記N末端側疎水性閾値未満であると判定した場合、前記N末端外疎水性判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値の最大値が前記N末端外疎水性閾値未満であると判定した場合、前記C末端側最大疎水位置判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が前記既知のG P Iアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内ないと判定した場合、または前記スコア判定部が、前記スコア算出部の算出したスコアが前記スコア判定閾値未満であると判定した場合に、前記検査対象タンパク質がG P Iアンカー型タンパク質でないと判定する、

20

ことを特徴とする判定方法。

#### 【請求項15】

検査対象タンパク質がG P Iアンカー型タンパク質であるか否かを判定するG P Iアンカー型タンパク質の判定装置を、

前記検査対象タンパク質のアミノ酸配列情報を取得する配列取得部、

30

前記配列取得部が取得したアミノ酸配列情報における既知のG P Iアンカー型タンパク質のN末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のN末端から予め定められた残基数の領域を特定し、当該N末端側の高疎水性領域に対応する領域のアミノ酸残基を抽出し、前記N末端側の高疎水性領域に対応する領域のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端側疎水性特性抽出必要数を用いて、連続する当該N末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均であるN末端側平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端側疎水性値算出部、

前記N末端側疎水性値算出部が算出した複数のN末端側平均疎水性値のうちの最大値が、既知のG P Iアンカー型タンパク質におけるN末端側平均疎水性値の特性を示すN末端側疎水性閾値以上であるか否かを判定するN末端側疎水性判定部、

40

前記配列取得部が取得したアミノ酸配列情報のうち前記N末端側疎水性値算出部がN末端側平均疎水性値を算出した範囲以外の領域として、前記アミノ酸配列情報の前記N末端から予め定められた残基数のアミノ酸残基以外の領域を特定し、当該N末端側平均疎水性値を算出した範囲以外のアミノ酸残基を抽出し、前記N末端側平均疎水性値を算出した範囲以外のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端外疎水性特性抽出数を用いて、連続する当該N末端外疎水性特性抽出数分のアミノ酸残基の各疎水性指標値の平均であるN末端外平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端外疎水性値算出部、

前記N末端外疎水性値算出部が算出した複数のN末端外平均疎水性値のうちの最大値が

50

、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示すN末端外疎水性閾値以上であるか否かを判定するN末端外疎水性判定部、

既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、前記N末端外疎水性値算出部が算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が当該特定した領域内にあるか否かを判定するC末端側最大疎水位置判定部、

前記配列取得部が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のプロペプチド領域を含む領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、当該プロペプチド領域を含む領域のアミノ酸残基を抽出し、当該抽出したアミノ酸残基のそれぞれに対して、当該プロペプチド領域を含む領域のアミノ酸残基の側鎖サイズの平均化に用いる残基数である側鎖サイズ特性抽出必要数を用いて、連続する当該側鎖サイズ特性抽出必要数分のアミノ酸残基の各側鎖サイズ指標値の平均値である平均側鎖サイズを1残基ずつずらしながら複数算出する側鎖サイズ算出部、

前記側鎖サイズ算出部が算出した平均側鎖サイズが最小となる位置を基準位置とする、当該基準位置からN末端側及びC末端側に連続する所定の残基数のアミノ酸残基からなる所定の領域におけるアミノ酸残基の部分配列と、既知のGPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列との類似度を示すスコアを算出するにあたり、記憶部に記憶されている既知のGPIアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを前記記憶部から取得し、当該位置特異的スコアに基づき、前記部分配列の各アミノ酸残基の位置特異的スコアを特定して数値列を生成し、当該数値列の平均値を算出し、当該算出した平均値を前記スコアとして算出するスコア算出部、

前記スコア算出部が算出したスコアが、前記検査対象タンパク質と既知のGPIアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定するスコア判定部、

前記N末端側疎水性判定部が、前記N末端側疎水性値算出部の算出したN末端側平均疎水性値の最大値が前記N末端側疎水性閾値未満であると判定した場合、前記N末端外疎水性判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値の最大値が前記N末端外疎水性閾値未満であると判定した場合、前記C末端側最大疎水位置判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が前記既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内にないと判定した場合、または前記スコア判定部が、前記スコア算出部の算出したスコアが前記スコア判定閾値未満であると判定した場合に、前記検査対象タンパク質がGPIアンカー型タンパク質でないと判定するGPIアンカー型タンパク質判定部、

として機能させるための判定プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検査対象タンパク質がGPI(glycosylphosphatidylinositol)アンカー型タンパク質であるか否かを判定するGPIアンカー型タンパク質の判定装置、判定方法及び判定プログラムに関する。

【背景技術】

【0002】

生体内の多くのタンパク質は、糖鎖、脂質、糖脂質等により翻訳後修飾を受けており、これらの修飾がタンパク質の機能や細胞内局在に影響することが知られている。これらの翻訳後修飾の中でも、脂質と糖鎖とからなる糖脂質であるGPIアンカーによる修飾は、非常に重要な意味を有するとされている。このことは、GPIアンカーが真核生物や古細菌において広く保存されていること、GPIアンカーを欠損した酵母や原虫は生存できず、GPIアンカーを欠損したヒトは造血幹細胞に異常を生じること等からも明らかである。

10

20

30

40

50

GPIにより修飾を受けるタンパク質は、GPIアンカー型タンパク質と呼ばれる。GPIアンカー型タンパク質は、そのアミノ酸配列のN末端に小胞体輸送のシグナルペプチドを有するため、小胞体内に輸送された後に翻訳を完了する。その後、GPIアンカー修飾部位( サイト)のC末端側に存在するプロペプチドが、トランスアミダーゼにより切断及び除去され、GPIアンカー型タンパク質は小胞体内で生合成されたGPIアンカーと結合する。GPIアンカーと結合したGPIアンカー型タンパク質は、ゴルジ体を経て細胞膜表面に輸送され、GPIアンカーにより細胞膜に繋ぎ止められる。

GPIアンカー型タンパク質の特徴としては、N末端のシグナルペプチド及びC末端のプロペプチドの疎水性が高く、 サイトの近隣には残基サイズの小さいアミノ酸が存在することが知られている。

10

#### 【0003】

GPIアンカー型タンパク質としては、CD14、CD16b等の受容体、5'-ヌクレオチダーゼ、アルカリフォスファターゼ等の酵素等の生体反応に極めて重要なタンパク質が多く発見されている。また、狂牛病関連のプリオンタンパク質や、癌関連のヒト癌胎児性抗原(CEA)等、重篤な疾患に関わるタンパク質も見出されている。しかしながら、現在までに真核生物で知られているGPIアンカー型タンパク質は100種類程度であり、未だ発見されていないGPIアンカー型タンパク質が多く存在すると考えられている。そこで、近年では、コンピュータを用いたバイオインフォマティクス手法により、アミノ酸配列からGPIアンカー型タンパク質を新たに見つける試みがなされている。

#### 【0004】

20

例えば、非特許文献1には、真核生物のGPIアンカー型タンパク質を学習のデータセットとして、隠れマルコフモデルとサポートベクターマシン(SVM)とを組み合わせた判定手法を用いて、検査対象タンパク質のアミノ酸配列情報から、検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定する方法が記載されている。

また、非特許文献2には、原核生物及び真核生物のGPIアンカー型タンパク質を学習のデータセットとして、 サイト前後のアミノ酸配列におけるアミノ酸の性質及び出現頻度をスコア化し、GPIアンカー修飾部位を予測し、検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定する方法が記載されている。

さらに、非特許文献3には、ニューラルネットワークの一種であるコホーネン自己組織化マップを用いて、検査対象の真核生物タンパク質がGPIアンカー型タンパク質であるか否かを判定する方法が記載されている。

30

#### 【先行技術文献】

##### 【非特許文献】

#### 【0005】

【非特許文献1】Pierleoniら、「BMC Bioinformatics」、2008年、vol.9、no.392、pp.1-11

【非特許文献2】Eisenhaberら、「Journal of Molecular Biology」、1999年、vol.292、pp.741-758

【非特許文献3】Frankhauserら、「Bioinformatics」、2005年、vol.21、no.9、pp.1846-1852

40

#### 【発明の概要】

##### 【発明が解決しようとする課題】

#### 【0006】

しかしながら、上述したような従来のGPIアンカー型タンパク質判定方法は、新規のGPIアンカー型タンパク質を判定する感度及び選択性が十分ではない。そこで、より高い感度及び選択性で、検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定することへの要求がある。

本発明は、上記事情に鑑みてなされたものであって、高感度且つ高選択的に検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定することが可能なGPIアンカー型タンパク質の判定装置、判定方法及び判定プログラムを提供することを目的とする

50



。【課題を解決するための手段】

【0007】

本発明は上記の課題を解決するためになされたものであり、検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定するGPIアンカー型タンパク質の判定装置であって、前記検査対象タンパク質のアミノ酸配列情報を取得する配列取得部と、前記配列取得部が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のN末端から予め定められた残基数の領域を特定し、当該N末端側の高疎水性領域に対応する領域のアミノ酸残基を抽出し、前記N末端側の高疎水性領域に対応する領域のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端側疎水性特性抽出必要数を用いて、連続する当該N末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均であるN末端側平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端側疎水性値算出部と、前記N末端側疎水性値算出部が算出した複数のN末端側平均疎水性値のうちの最大値が、既知のGPIアンカー型タンパク質におけるN末端側平均疎水性値の特性を示すN末端側疎水性閾値以上であるか否かを判定するN末端側疎水性判定部と、前記配列取得部が取得したアミノ酸配列情報のうち前記N末端側疎水性値算出部がN末端側平均疎水性値を算出した範囲以外の領域として、前記アミノ酸配列情報の前記N末端から予め定められた残基数のアミノ酸残基以外の領域を特定し、当該N末端側平均疎水性値を算出した範囲以外のアミノ酸残基を抽出し、前記N末端側平均疎水性値を算出した範囲以外のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端外疎水性特性抽出数を用いて、連続する当該N末端外疎水性特性抽出数分のアミノ酸残基の各疎水性指標値の平均であるN末端外平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端外疎水性値算出部と、前記N末端外疎水性値算出部が算出した複数のN末端外平均疎水性値のうちの最大値が、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示すN末端外疎水性閾値以上であるか否かを判定するN末端外疎水性判定部と、既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、前記N末端外疎水性値算出部が算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が当該特定した領域内にあるか否かを判定するC末端側最大疎水位置判定部と、前記配列取得部が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のプロペプチド領域を含む領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、当該プロペプチド領域を含む領域のアミノ酸残基を抽出し、当該抽出したアミノ酸残基のそれぞれに対して、当該プロペプチド領域を含む領域のアミノ酸残基の側鎖サイズの平均化に用いる残基数である側鎖サイズ特性抽出必要数を用いて、連続する当該側鎖サイズ特性抽出必要数分のアミノ酸残基の各側鎖サイズ指標値の平均値である平均側鎖サイズを1残基ずつずらしながら複数算出する側鎖サイズ算出部と、前記側鎖サイズ算出部が算出した平均側鎖サイズが最小となる位置を基準位置とする、当該基準位置からN末端側及びC末端側に連続する所定の残基数のアミノ酸残基からなる所定の領域におけるアミノ酸残基の部分配列と、既知のGPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列との類似度を示すスコアを算出するにあたり、記憶部に記憶されている既知のGPIアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを前記記憶部から取得し、当該位置特異的スコアに基づき、前記部分配列の各アミノ酸残基の位置特異的スコアを特定して数値列を生成し、当該数値列の平均値を算出し、当該算出した平均値を前記スコアとして算出するスコア算出部と、前記スコア算出部が算出したスコアが、前記検査対象タンパク質と既知のGPIアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定するスコア判定部と、前記N末端側疎水性判定部が、前記N末端側疎水性値算出部の算出したN末端側平均疎水性値の最大値が前記N末端側疎水性閾値未満で

10

20

30

40

50

あると判定した場合、前記N末端外疎水性判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値の最大値が前記N末端外疎水性閾値未満であると判定した場合、前記C末端側最大疎水位置判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が前記既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内ないと判定した場合、または前記スコア判定部が、前記スコア算出部の算出したスコアが前記スコア判定閾値未満であると判定した場合に、前記検査対象タンパク質がGPIアンカー型タンパク質でないとして判定するGPIアンカー型タンパク質判定部と、を備えることを特徴とする。

【0008】

また、本発明において、前記N末端側疎水性閾値は、予め既知の複数のGPIアンカー型タンパク質に対して前記N末端側平均疎水性値の算出を行い、当該算出されたN末端側平均疎水性値の最大値の集合における最小値であることを特徴とする。

10

【0009】

また、本発明において、前記N末端外疎水性閾値は、予め既知の複数のGPIアンカー型タンパク質に対して前記N末端外平均疎水性値の算出を行い、当該算出されたN末端外平均疎水性値の最大値の集合における最小値であることを特徴とする。

【0010】

また、本発明において、前記既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域は、既知のGPIアンカー型タンパク質において、前記N末端側平均疎水性値が最大となる位置が含まれる領域である、ことを特徴とする。

20

【0011】

また、本発明において、前記既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域は、既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域以外の領域において、前記N末端外平均疎水性値が最大となる位置が含まれる領域である、ことを特徴とする。

【0012】

また、本発明において、前記N末端側疎水性特性抽出必要数は、当該N末端側疎水性特性抽出必要数を用いて、既知の複数のGPIアンカー型タンパク質のN末端側の高疎水性領域のアミノ酸残基のそれぞれに対してN末端側平均疎水性値を算出し、前記既知のGPIアンカー型タンパク質から算出したN末端側平均疎水性値の最大値の集合における最小値を抽出し、前記N末端側疎水性特性抽出必要数を用いて、既知の複数の非GPIアンカー型タンパク質における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域のアミノ酸残基のそれぞれに対してN末端側平均疎水性値を算出した場合に、前記既知の非GPIアンカー型タンパク質から算出したN末端側平均疎水性値の最大値のうち、前記抽出した最小値より値が大きいものの個数が最小となるような値であることを特徴とする。

30

【0013】

また、本発明において、前記N末端外疎水性特性抽出必要数は、当該N末端外疎水性特性抽出必要数を用いて、既知の複数のGPIアンカー型タンパク質のN末端側の高疎水性領域以外の領域のアミノ酸残基のそれぞれに対してN末端外平均疎水性値を算出し、前記既知のGPIアンカー型タンパク質から算出したN末端外平均疎水性値の最大値の集合における最小値を抽出し、前記N末端外疎水性特性抽出必要数を用いて、既知の複数の非GPIアンカー型タンパク質における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域以外の領域のアミノ酸残基のそれぞれに対してN末端外平均疎水性値を算出した場合に、前記既知の非GPIアンカー型タンパク質から算出したN末端外平均疎水性値の最大値のうち、前記抽出した最小値より値が大きいものの個数が最小となるような値であることを特徴とする。

40

【0014】

また、本発明において、前記小側鎖サイズ判定領域は、既知のGPIアンカー型タンパク質の前記平均側鎖サイズが最小となる位置が含まれる領域である、ことを特徴とする。

50

## 【 0 0 1 5 】

また、本発明において、前記側鎖サイズ特性抽出必要数は、当該側鎖サイズ特性抽出必要数を用いて、既知の複数のGPIアンカー型タンパク質の小側鎖サイズ判定領域に対して平均側鎖サイズを算出した場合に、前記GPIアンカー型タンパク質から算出した平均側鎖サイズが最小となるアミノ酸残基のうち、当該アミノ酸残基のC末端側に隣接するアミノ酸残基がGPIアンカー修飾部位であるものの個数が最大となるような値であることを特徴とする。

## 【 0 0 1 7 】

また、本発明において、前記位置特異的スコアは、式(4)から算出されたものであることを特徴とする。

10

## 【 0 0 1 8 】

また、本発明において、前記所定の領域内の位置pに存在するアミノ酸残基の種類iの出現頻度は、式(3)から算出されたものであることを特徴とする。

## 【 0 0 1 9 】

また、本発明において、前記スコア判定閾値は、既知の複数のGPIアンカー型タンパク質について算出した前記スコアの最小値と、既知の複数の非GPIアンカー型タンパク質について算出した前記スコアの最大値との間の値であることを特徴とする。

## 【 0 0 2 0 】

また、本発明において、前記スコア判定閾値は、式(6)から算出される値が最大となるような値であることを特徴とする。

20

## 【 0 0 2 1 】

また、本発明は、検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定するGPIアンカー型タンパク質の判定装置を用いた判定方法であって、GPIアンカー型タンパク質の判定装置の配列取得部は、前記検査対象タンパク質のアミノ酸配列情報を取得し、GPIアンカー型タンパク質の判定装置のN末端側疎水性値算出部は、前記配列取得部が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のN末端から予め定められた残基数の領域を特定し、当該N末端側の高疎水性領域に対応する領域のアミノ酸残基を抽出し、前記N末端側の高疎水性領域に対応する領域のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端側疎水性特性抽出必要数を用いて、連続する当該N末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均であるN末端側平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出し、GPIアンカー型タンパク質の判定装置のN末端側疎水性判定部は、前記N末端側疎水性値算出部が算出した複数のN末端側平均疎水性値のうちの最大値が、既知のGPIアンカー型タンパク質におけるN末端側平均疎水性値の特性を示すN末端側疎水性閾値以上であるか否かを判定し、GPIアンカー型タンパク質の判定装置のN末端外疎水性値算出部は、前記配列取得部が取得したアミノ酸配列情報のうち前記N末端側疎水性値算出部がN末端側平均疎水性値を算出した範囲以外の領域として、前記アミノ酸配列情報の前記N末端から予め定められた残基数のアミノ酸残基以外の領域を特定し、当該N末端側平均疎水性値を算出した範囲以外のアミノ酸残基を抽出し、前記N末端側平均疎水性値を算出した範囲以外のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端外疎水性特性抽出数を用いて、連続する当該N末端外疎水性特性抽出数分のアミノ酸残基の各疎水性指標値の平均であるN末端外平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出し、GPIアンカー型タンパク質の判定装置のN末端外疎水性判定部は、前記N末端外疎水性値算出部が算出した複数のN末端外平均疎水性値のうちの最大値が、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示すN末端外疎水性閾値以上であるか否かを判定し、GPIアンカー型タンパク質の判定装置のC末端側最大疎水位置判定部は、既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、前記N末端外疎水性値算出部が算出したN末端外平均疎

30

40

50

水性値が最大となるアミノ酸残基の位置が当該特定した領域内にあるか否かを判定し、G P Iアンカー型タンパク質の判定装置の側鎖サイズ算出部は、前記配列取得部が取得したアミノ酸配列情報における既知のG P Iアンカー型タンパク質のプロペプチド領域を含む領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、当該プロペプチド領域を含む領域のアミノ酸残基を抽出し、当該抽出したアミノ酸残基のそれぞれに対して、当該プロペプチド領域を含む領域のアミノ酸残基の側鎖サイズの平均化に用いる残基数である側鎖サイズ特性抽出必要数を用いて、連続する当該側鎖サイズ特性抽出必要数分のアミノ酸残基の各側鎖サイズ指標値の平均値である平均側鎖サイズを1残基ずつずらしながら複数算出し、G P Iアンカー型タンパク質の判定装置のスコア算出部は、前記側鎖サイズ算出部が算出した平均側鎖サイズが最小となる位置を基準位置とする、当該基準位置からN末端側及びC末端側に連続する所定の残基数のアミノ酸残基からなる所定の領域におけるアミノ酸残基の部分配列と、既知のG P Iアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列との類似度を示すスコアを算出するにあたり、記憶部に記憶されている既知のG P Iアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを前記記憶部から取得し、当該位置特異的スコアに基づき、前記部分配列の各アミノ酸残基の位置特異的スコアを特定して数値列を生成し、当該数値列の平均値を算出し、当該算出した平均値を前記スコアとして算出し、G P Iアンカー型タンパク質の判定装置のスコア判定部は、前記スコア算出部が算出したスコアが、前記検査対象タンパク質と既知のG P Iアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定し、G P Iアンカー型タンパク質の判定装置のG P Iアンカー型タンパク質判定部は、前記N末端側疎水性判定部が、前記N末端側疎水性値算出部の算出したN末端側平均疎水性値の最大値が前記N末端側疎水性閾値未満であると判定した場合、前記N末端外疎水性判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値の最大値が前記N末端外疎水性閾値未満であると判定した場合、前記C末端側最大疎水位置判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が前記既知のG P Iアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内ないと判定した場合、または前記スコア判定部が、前記スコア算出部の算出したスコアが前記スコア判定閾値未満であると判定した場合に、前記検査対象タンパク質がG P Iアンカー型タンパク質でないと判定する、ことを特徴とする。

#### 【 0 0 2 2 】

また、本発明は、検査対象タンパク質がG P Iアンカー型タンパク質であるか否かを判定するG P Iアンカー型タンパク質の判定装置を、前記検査対象タンパク質のアミノ酸配列情報を取得する配列取得部、前記配列取得部が取得したアミノ酸配列情報における既知のG P Iアンカー型タンパク質のN末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のN末端から予め定められた残基数の領域を特定し、当該N末端側の高疎水性領域に対応する領域のアミノ酸残基を抽出し、前記N末端側の高疎水性領域に対応する領域のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端側疎水性特性抽出必要数を用いて、連続する当該N末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均であるN末端側平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端側疎水性値算出部、前記N末端側疎水性値算出部が算出した複数のN末端側平均疎水性値のうちの最大値が、既知のG P Iアンカー型タンパク質におけるN末端側平均疎水性値の特性を示すN末端側疎水性閾値以上であるか否かを判定するN末端側疎水性判定部、前記配列取得部が取得したアミノ酸配列情報のうち前記N末端側疎水性値算出部がN末端側平均疎水性値を算出した範囲以外の領域として、前記アミノ酸配列情報の前記N末端から予め定められた残基数のアミノ酸残基以外の領域を特定し、当該N末端側平均疎水性値を算出した範囲以外のアミノ酸残基を抽出し、前記N末端側平均疎水性値を算出した範囲以外のアミノ酸残基の疎水性値の平均化に用いる残基数であるN末端外疎水性特性抽出数を用いて、連続する当該N末端外疎水性特

10

20

30

40

50

性抽出数分のアミノ酸残基の各疎水性指標値の平均であるN末端外平均疎水性値を、前記抽出したアミノ酸残基のそれぞれに対して1残基ずつずらしながら複数算出するN末端外疎水性値算出部、前記N末端外疎水性値算出部が算出した複数のN末端外平均疎水性値のうちの最大値が、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示すN末端外疎水性閾値以上であるか否かを判定するN末端外疎水性判定部、既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、前記N末端外疎水性値算出部が算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が当該特定した領域内にあるか否かを判定するC末端側最大疎水位置判定部、前記配列取得部が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のプロペプチド領域を含む領域として、前記アミノ酸配列情報のC末端から予め定められた残基数の領域を特定し、当該プロペプチド領域を含む領域のアミノ酸残基を抽出し、当該抽出したアミノ酸残基のそれぞれに対して、当該プロペプチド領域を含む領域のアミノ酸残基の側鎖サイズの平均化に用いる残基数である側鎖サイズ特性抽出必要数を用いて、連続する当該側鎖サイズ特性抽出必要数分のアミノ酸残基の各側鎖サイズ指標値の平均値である平均側鎖サイズを1残基ずつずらしながら複数算出する側鎖サイズ算出部、前記側鎖サイズ算出部が算出した平均側鎖サイズが最小となる位置を基準位置とする、当該基準位置からN末端側及びC末端側に連続する所定の残基数のアミノ酸残基からなる所定の領域におけるアミノ酸残基の部分配列と、既知のGPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列との類似度を示すスコアを算出するにあたり、記憶部に記憶されている既知のGPIアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを前記記憶部から取得し、当該位置特異的スコアに基づき、前記部分配列の各アミノ酸残基の位置特異的スコアを特定して数値列を生成し、当該数値列の平均値を算出し、当該算出した平均値を前記スコアとして算出するスコア算出部、前記スコア算出部が算出したスコアが、前記検査対象タンパク質と既知のGPIアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定するスコア判定部、前記N末端側疎水性判定部が、前記N末端側疎水性値算出部の算出したN末端側平均疎水性値の最大値が前記N末端側疎水性閾値未満であると判定した場合、前記N末端外疎水性判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値の最大値が前記N末端外疎水性閾値未満であると判定した場合、前記C末端側最大疎水位置判定部が、前記N末端外疎水性値算出部の算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が前記既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内にはないと判定した場合、または前記スコア判定部が、前記スコア算出部の算出したスコアが前記スコア判定閾値未満であると判定した場合に、前記検査対象タンパク質がGPIアンカー型タンパク質でないとして機能させるための判定プログラムである。

【発明の効果】

【0023】

本発明によれば、N末端側の疎水性及びC末端側の疎水性の高低と、新規のPSSM (position specific scoring matrix; 位置特異的スコアリングマトリックス) とを用いているため、高感度且つ高選択的に検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定することができる。

【図面の簡単な説明】

【0024】

【図1】本発明の一実施形態によるGPIアンカー型タンパク質判定装置の構成を示す概略ブロック図である。

【図2】疎水性指標値記憶部が記憶する情報を示す図である。

【図3】側鎖サイズ指標値記憶部が記憶する情報を示す図である。

【図4】PSSM記憶部が記憶するPSSMを示す第1の図である。

10

20

30

40

50

【図5】PSSM記憶部が記憶するPSSMを示す第2の図である。

【図6】GPIアンカー型タンパク質判定装置100の動作を示すフローチャートである。

【図7】GPIアンカー型タンパク質の疎水性プロファイルを示す第1のグラフである。

【図8】N末端側平均疎水性値の算出方法を示す図である。

【図9】既知のGPIアンカー型タンパク質のN末端から30残基以内におけるN末端側平均疎水性値の最大値の分布を示すグラフである。

【図10】GPIアンカー型タンパク質の疎水性プロファイルを示す第2のグラフである。

【図11】既知のGPIアンカー型タンパク質及び既知の非GPIアンカー型タンパク質のN末端外平均疎水性値の最大値を示すグラフである。 10

【図12】GPIアンカー型タンパク質の側鎖サイズのプロファイルを示すグラフである。

【図13】アミノ酸配列の抽出方法を示す図である。

【図14】位置特異的スコアの割り当て方法を示す図である。

【図15】冗長性を排除したGPIアンカー型タンパク質データセットに含まれるエントリーネームを示す表である。

【図16】既知のエントリーに対して算出したスコアの分布を示す図である。

【図17】スコア判定閾値の候補それぞれを用いた場合の感度と選択性とを示す表である。 20

【図18】本実施形態によるGPIアンカー型タンパク質判定装置の判定精度を示す表である。

【図19】N末端側疎水性特性抽出必要数を変更した場合の判定精度を示す表である。

【図20】N末端外疎水性特性抽出必要数を変更した場合の判定精度を示す表である。

【図21】側鎖サイズ特性抽出必要数を変更した場合の判定精度を示す表である。

【図22】基準位置を含む所定の範囲を変更した場合の判定精度を示す表である。

【発明を実施するための形態】

【0025】

以下、図面を参照しながら本発明の実施形態について詳しく説明する。

図1は、本発明の一実施形態によるGPIアンカー型タンパク質判定装置の構成を示す概略ブロック図である。 30

GPIアンカー型タンパク質判定装置100は、配列記憶部101、配列取得部102、疎水性指標値記憶部103、疎水性指標値特定部104、N末端側疎水性値算出部105、N末端側疎水性判定部106、N末端外疎水性値算出部107、N末端外疎水性判定部108、C末端側最大疎水位置判定部109、側鎖サイズ指標値記憶部110、側鎖サイズ指標値特定部111、側鎖サイズ算出部112、PSSM記憶部113、スコア算出部114、スコア判定部115、GPIアンカー型タンパク質判定部116を備える。

【0026】

配列記憶部101は、機能未知の哺乳類のタンパク質の完全長アミノ酸配列情報を記憶する。 40

配列取得部102は、配列記憶部101から検査対象となるタンパク質のアミノ酸配列情報を取得する。

疎水性指標値記憶部103は、アミノ酸残基に対応付けて当該アミノ酸残基の疎水性指標値を記憶する。

疎水性指標値特定部104は、配列取得部102が取得したアミノ酸配列の各アミノ酸残基それぞれの疎水性指標値を疎水性指標値記憶部103が記憶する疎水性指標値から特定し、アミノ酸残基毎の疎水性指標値を示す連続する数値列を生成する。

N末端側疎水性値算出部105は、疎水性指標値特定部104が生成した数値列に基づいて、配列取得部102が取得したアミノ酸配列情報が示すN末端側の連続するアミノ酸残基の平均疎水性値(N末端側平均疎水性値)を算出する。 50

N末端側疎水性判定部106は、N末端側疎水性値算出部105が算出した平均疎水性値の最大値がN末端側疎水性閾値以上であるか否かを判定する。ここで、N末端側疎水性閾値とは、既知のGPIアンカータンパク質におけるN末端側平均疎水性値の特性を示す閾値である。

【0027】

N末端外疎水性値算出部107は、疎水性指標値特定部104が生成した数値列に基づいて、配列取得部102が取得したアミノ酸配列情報のうち、N末端側疎水性値算出部105が平均疎水性値を算出した範囲以外の連続するアミノ酸残基の平均疎水性値(N末端外平均疎水性値)を算出する。

N末端外疎水性判定部108は、N末端外疎水性値算出部107が算出した平均疎水性値の最大値がN末端外疎水性閾値以上であるか否かを判定する。ここで、N末端外疎水性閾値とは、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示す閾値である。

10

C末端側最大疎水位置判定部109は、N末端外疎水性値算出部107が算出した平均疎水性値が最大となるアミノ酸残基の位置が既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内にあるか否かを判定する。

【0028】

側鎖サイズ指標値記憶部110は、アミノ酸残基に対応付けて当該アミノ酸残基の側鎖サイズ指標値を記憶する。

側鎖サイズ指標値特定部111は、配列取得部102が取得したアミノ酸配列の各アミノ酸残基それぞれの側鎖サイズ指標値を、側鎖サイズ指標値記憶部110が記憶する側鎖サイズ指標値から特定し、アミノ酸残基毎の側鎖サイズ指標値を示す連続する数値列を生成する。

20

側鎖サイズ算出部112は、側鎖サイズ指標値特定部111が生成した数値列に基づいて、配列取得部102が取得したアミノ酸配列情報が示すC末端側のアミノ酸残基の平均残基サイズを算出する。

PSSM記憶部113は、GPIアンカー型タンパク質のアミノ酸残基位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを保持するPSSMを記憶する。ここで、位置特異的スコアとは、GPIアンカー型タンパク質である可能性を示す値であり、当該値が大きいほどGPIアンカー型タンパク質である可能性が高いことを表す。

30

スコア算出部114は、PSSM記憶部113が記憶するPSSMに基づいて、側鎖サイズ算出部112が算出した側鎖のサイズの平均が最小となるアミノ酸残基の位置を基準位置とする所定の領域におけるスコアを算出する。ここで算出するスコアは、配列取得部102が取得した検査対象となるタンパク質の所定の領域におけるアミノ酸残基の部分配列のアミノ酸出現傾向と既知のGPIアンカー型タンパク質の所定の領域におけるアミノ酸残基の部分配列のアミノ酸出現傾向との類似度を示すスコアである。

スコア判定部115は、スコア算出部114が算出したスコアが、検査対象タンパク質と既知のGPIアンカー型タンパク質とが類似するか否かを判定するスコア判定閾値以上であるか否かを判定する。

GPIアンカー型タンパク質判定部116は、配列取得部102が取得した検査対象となるタンパク質がGPIアンカー型タンパク質であるか否かを判定する。

40

【0029】

図2は、疎水性指標値記憶部が記憶する情報を示す図である。

疎水性指標値記憶部103は、図2に示すように、アミノ酸残基の各々に対して、当該アミノ酸残基の疎水性を示す指標値を記憶している。なお、本実施形態では、疎水性指標値としてKYTJ820101(Kyte J., Doolittle R., 「Journal of Molecular Biology」、1982年、vol.157、no.1、pp.105-132)で示される疎水性指標値を用いている。図2において、アミノ酸残基の「A」はアラニンを示し、「R」はアルギニンを示し、「N」はアスパラギンを示し、「D」はアスパラギン酸を示し、「C」はシステインを示し、「Q」は

50

グルタミンを示し、「E」はグルタミン酸を示し、「G」はグリシンを示し、「H」はヒスチジンを示し、「I」はイソロイシンを示し、「L」はロイシンを示し、「K」はリシンを示し、「M」はメチオニンを示し、「F」はフェニルアラニンを示し、「P」はプロリンを示し、「S」はセリンを示し、「T」はトレオニンを示し、「W」はトリプトファンを示し、「Y」はチロシンを示し、「V」はバリンを示す。

#### 【0030】

図3は、側鎖サイズ指標値記憶部が記憶する情報を示す図である。

側鎖サイズ指標値記憶部110は、図3に示すように、アミノ酸残基の各々に対して、当該アミノ酸残基の側鎖のサイズを示す指標値を記憶している。なお、本実施形態では、側鎖サイズ指標値としてDAWD720101(Dawson D.M.,「The Biological Genetics of Man」、Academic Press、1972年、pp.1-38)で示される側鎖サイズ指標値を用いている。

10

#### 【0031】

図4及び図5は、PSSM記憶部が記憶するPSSMを示す図である。

PSSM記憶部113は、図4及び図5に示すように、アミノ酸残基の位置におけるアミノ酸残基の種類の出現度合いを示す位置特異的スコアを要素とするPSSMを記憶している。図4及び図5では、アミノ酸残基位置の基準位置を0とし、負数側をN末端側、正数側をC末端側としている。なお、PSSMの作成方法については、後述する。ここで、基準位置とは、GPIアンカー型タンパク質のGPIアンカー修飾部位(サイト)のC末端側に隣接するアミノ酸残基の位置を示す。

20

#### 【0032】

そして、GPIアンカー型タンパク質判定装置100において、まず配列取得部102は、検査対象タンパク質のアミノ酸配列情報を取得する。次に、N末端側疎水性値算出部105は、配列取得部102が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域のアミノ酸残基のそれぞれに対して、連続するN末端側疎水性特性抽出必要数分のアミノ酸残基の各疎水性指標値の平均であるN末端側平均疎水性値を算出する。次に、N末端側疎水性判定部106は、N末端側疎水性値算出部105が算出したN末端側平均疎水性値の最大値が、既知のGPIアンカー型タンパク質におけるN末端側平均疎水性値の特性を示すN末端側疎水性閾値以上であるか否かを判定する。

30

#### 【0033】

また、N末端外疎水性値算出部107は、配列取得部102が取得したアミノ酸配列情報のうちN末端側疎水性値算出部105がN末端側平均疎水性値を算出した範囲以外のアミノ酸残基のそれぞれに対して、連続するN末端外疎水性特性抽出数分のアミノ酸残基の各疎水性指標値の平均であるN末端外平均疎水性値を算出する。次に、N末端外疎水性判定部108は、N末端外疎水性値算出部107が算出したN末端外平均疎水性値の最大値が、既知のGPIアンカー型タンパク質におけるN末端外平均疎水性値の特性を示すN末端外疎水性閾値以上であるか否かを判定する。また、C末端側最大疎水位置判定部109は、N末端外疎水性値算出部107が算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が既知のGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内にあるか否かを判定する。

40

#### 【0034】

他方、側鎖サイズ算出部112は、配列取得部102が取得したアミノ酸配列情報における既知のGPIアンカー型タンパク質のプロペプチド領域に対応する領域のアミノ酸残基のそれぞれに対して、連続する側鎖サイズ特性抽出必要数分のアミノ酸残基の側鎖サイズ指標値の平均値である平均側鎖サイズを算出する。次に、スコア算出部114は、側鎖サイズ算出部112が算出した平均側鎖サイズが最小となる位置を基準位置とする所定の領域におけるアミノ酸残基の部分配列と、既知のGPIアンカー型タンパク質の平均側鎖サイズが最小となる位置を基準位置とする前記所定の領域におけるアミノ酸残基の部分配列との類似度を示すスコアを算出する。次に、スコア判定部115は、スコア算出部11

50



4 が算出したスコアが、検査対象タンパク質と既知の G P I アンカー型タンパク質とが類似するか否かを分類するスコア判定閾値以上であるか否かを判定する。

【 0 0 3 5 】

そして、G P I アンカー型タンパク質判定部 1 1 6 は、N 末端側疎水性判定部 1 0 6、N 末端外疎水性判定部 1 0 8、C 末端側最大疎水位置判定部 1 0 9、及びスコア判定部 1 1 5 の判定結果に基づいて、検査対象タンパク質が G P I アンカー型タンパク質であるか否かを判定する。

これにより、G P I アンカー型タンパク質判定装置 1 0 0 は、高感度且つ高選択的に検査対象タンパク質が G P I アンカー型タンパク質であるか否かを判定する。

【 0 0 3 6 】

次に、G P I アンカー型タンパク質判定装置 1 0 0 の動作を説明する。

図 6 は、G P I アンカー型タンパク質判定装置 1 0 0 の動作を示すフローチャートである。

< ステップ S 1 : 配列を取得 >

まず、使用者による動作開始指示により、G P I アンカー型タンパク質判定装置 1 0 0 が動作を開始すると、配列取得部 1 0 2 は、配列記憶部 1 0 1 から検査対象となるタンパク質のアミノ酸配列情報を取得する。

【 0 0 3 7 】

< ステップ S 2 : 疎水性指標値を特定 >

配列取得部 1 0 2 がアミノ酸配列情報を取得すると、疎水性指標値特定部 1 0 4 は、疎水性指標値記憶部 1 0 3 を参照して、配列取得部 1 0 2 が取得したアミノ酸配列情報の各アミノ酸残基の疎水性指標値を特定し、当該疎水性指標値を示す数値列を生成する。例えば、配列取得部 1 0 2 が取得したアミノ酸配列情報が、「M L L E P G R G C C ...」という配列を示す場合、疎水性指標値特定部 1 0 4 は、疎水性指標値記憶部 1 0 3 が記憶する図 2 に示す指標値より「1.9、3.8、3.8、-3.5、-1.6、-0.4、-4.5、-0.4、2.5、2.5...」という数値列を生成する。

【 0 0 3 8 】

< ステップ S 3 : N 末端側の疎水性指標値を抽出 >

ステップ S 2 で、疎水性指標値特定部 1 0 4 が疎水性指標値を示す数値列を生成すると、N 末端側疎水性値算出部 1 0 5 は、疎水性指標値特定部 1 0 4 が生成した数値列から、G P I アンカー型タンパク質の N 末端側の高疎水性領域に対応する領域のアミノ酸残基を示す部分数値列を抽出する。

本実施形態では、G P I アンカー型タンパク質の N 末端側の高疎水性領域に対応する領域として、N 末端から 3 0 残基以内のアミノ酸残基を用いる。N 末端から 3 0 残基以内のアミノ酸残基の領域は、既知の複数の G P I アンカー型タンパク質のアミノ酸残基のそれぞれに対して、後述するステップ S 4 と同様の処理によって平均疎水性値 ( N 末端側平均疎水性値 ) を算出した場合に、当該算出した平均疎水性値が最大となるアミノ酸残基列の中央に位置するアミノ酸残基が含まれる領域である。

【 0 0 3 9 】

図 7 は、G P I アンカー型タンパク質の疎水性プロファイルを示す第 1 のグラフである。

図 7 は、S W I S S - P R O T v e r 5 4 . 0 の B Y 5 5 \_ H U M A N ( 1 8 1 a a ) エントリーに対して、後述するステップ S 4 と同様の処理によって算出した N 末端側平均疎水性値 ( 1 1 残基平均の場合 ) を示すグラフである。ここで、横軸は、N 末端側疎水性特性抽出必要数の連続するアミノ酸残基列の中央に位置するアミノ酸残基の N 末端からの残基位置を示し、縦軸は N 末端側平均疎水性値の値を示す。

図 7 に示すように、既知の G P I アンカー型タンパク質の N 末端側の領域は疎水性が高く、N 末端から 3 0 残基以内に N 末端側平均疎水性値が最大となる位置が存在する。

【 0 0 4 0 】

< ステップ S 4 : N 末端側平均疎水性値を算出 >

10

20

30

40

50

図 8 は、N 末端側平均疎水性値の算出方法を示す図である。

N 末端側疎水性値算出部 105 は、ステップ S3 で G P I アンカー型タンパク質の N 末端側の高疎水性領域に対応する領域のアミノ酸残基を示す部分数値列を抽出すると、当該部分数値列の連続する N 末端側疎水性特性抽出必要数分の各疎水性指標値の平均である N 末端側平均疎水性値を、図 8 に示すように、1 残基ずつずらしながら算出する。

ここで、N 末端側疎水性特性抽出必要数の連続するアミノ酸残基列の中央のアミノ酸残基の位置が N 末端から r 残基目であるときの N 末端側平均疎水性値は、式 (1) を用いて算出できる。

【 0 0 4 1 】

【 数 1 】

$$\frac{1}{2n+1} \sum_{i=r-n}^{r+n} H(i) \quad \dots (1)$$

【 0 0 4 2 】

但し、n は、平均化に用いる前後の残基数を示す。つまり、2n + 1 は、N 末端側疎水性特性抽出必要数を示す。また、H ( i ) は、N 末端側疎水性特性抽出必要数の連続するアミノ酸残基列の中央のアミノ酸残基の位置が N 末端から i 残基目である場合のアミノ酸残基の疎水性指標値を示す。

つまり、N 末端から r 残基目のアミノ酸残基が中央に位置するアミノ酸残基列の N 末端側平均疎水性値は、N 末端から r - n 残基目のアミノ酸残基から、N 末端から r + n 残基目のアミノ酸残基までの疎水性指標値の平均となる。なお、このとき、N 末端から n 残基以内のアミノ酸残基は、前後 n 残基の平均値を算出できないため、N 末端側平均疎水性値として例えば N U L L 値を代入しておくが良い。

【 0 0 4 3 】

本実施形態では、N 末端側疎水性特性抽出必要数として 11 残基を用いる。つまり、N 末端側平均疎水性値として、N 末端から r 残基目のアミノ酸残基の前後 5 残基のアミノ酸残基の疎水性指標値の平均を算出する。ここで、N 末端側疎水性特性抽出必要数を 11 残基と決定する方法を説明する。

【 0 0 4 4 】

まず、既知の複数の G P I アンカー型タンパク質の N 末端側の高疎水性領域、すなわち N 末端から 30 残基以内のアミノ酸残基から、N 末端側疎水性特性抽出必要数の候補となる範囲の平均疎水性値を、1 残基ずつずらしながら算出する。次に、既知の複数の G P I アンカー型タンパク質のそれぞれの平均疎水性値の最大値を抽出する。そして、抽出した最大値の集合における最小値を抽出する。

次に、既知の複数の非 G P I アンカー型タンパク質における、既知の G P I アンカー型タンパク質の N 末端側の高疎水性領域に対応する領域、すなわち既知の複数の非 G P I アンカー型タンパク質の N 末端から 30 残基以内のアミノ酸残基から、N 末端側疎水性特性抽出必要数の候補となる個数の連続するアミノ酸残基列の平均疎水性値を、1 残基ずつずらしながら算出する。そして、非 G P I アンカー型タンパク質から算出した平均疎水性値の最大値のうち、既知の複数の G P I アンカー型タンパク質のそれぞれの平均疎水性値の最大値の集合から抽出した最小値より値が大きいものの個数を計数する。

この処理を N 末端側疎水性特性抽出必要数の候補となる値を変えて実行し、非 G P I アンカー型タンパク質から算出した平均疎水性値の最大値のうち、既知の複数の G P I アンカー型タンパク質のそれぞれの平均疎水性値の最大値の集合から抽出した最小値より値が大きいものの個数が最小となるような N 末端側疎水性特性抽出必要数の候補を、N 末端側疎水性特性抽出必要数として決定する。

【 0 0 4 5 】

そして、本実施形態では、S W I S S - P R O T v e r 5 4 . 0 より取得した既知の

10

20

30

40

50

哺乳類 G P I アンカー型タンパク質の完全長アミノ酸配列データセット、及び既知の哺乳類非 G P I アンカー型タンパク質の完全長アミノ酸配列データセットを用いて上述した方法を実行した結果、N末端側疎水性特性抽出必要数を 11 残基として決定した。

【 0 0 4 6 】

<ステップ S 5 : N末端側平均疎水性値の最大値の判定>

ステップ S 4 で、N末端側疎水性値算出部 105 が、部分数値列の各疎水性指標値の N末端側平均疎水性値を算出すると、N末端側疎水性判定部 106 は、算出した N末端側平均疎水性値の最大値が N末端側疎水性閾値以上であるか否かを判定する。なお、N末端側疎水性閾値は、G P I アンカー型タンパク質における N末端側平均疎水性値の特性を示す閾値であり、本実施形態では、N末端側疎水性閾値として 1.50 を用いる。1.50 という値は、予め既知の複数の G P I アンカー型タンパク質に対して N末端側平均疎水性値の算出を行い、当該算出された N末端側平均疎水性値の最大値の集合における最小値として算出された値である。

10

【 0 0 4 7 】

図 9 は、既知の G P I アンカー型タンパク質の N末端から 30 残基以内における N末端側平均疎水性値の最大値の分布を示すグラフである。ここで、横軸は N末端側平均疎水性値の最大値を示し、縦軸は G P I アンカー型タンパク質が当該最大値をとる頻度を示す。

図 9 に示すように、既知の G P I アンカー型タンパク質の N末端から 30 残基以内のアミノ酸残基から算出された N末端側平均疎水性値の最大値は、N末端側疎水性閾値である 1.50 以上の値となる。従って、検査対象タンパク質の N末端から 30 残基以内のアミノ酸残基から算出された N末端側平均疎水性値の最大値が 1.50 以上であれば、検査対象タンパク質が G P I アンカー型タンパク質である可能性が高く、当該最大値が 1.50 未満であれば、検査対象タンパク質が G P I アンカー型タンパク質である可能性が低いと判定できる。

20

【 0 0 4 8 】

<ステップ S 6 : N末端外の疎水性指標値を抽出>

ステップ S 5 で N末端側疎水性判定部 106 が、算出した N末端側平均疎水性値の最大値が N末端側疎水性閾値以上であると判定した場合 (ステップ S 5 : Y E S )、N末端外疎水性値算出部 107 は、ステップ S 2 で疎水性指標値特定部 104 が生成した数値列から、ステップ S 3 で N末端側疎水性値算出部 105 が抽出した部分数値列以外の残りの部分数値列を抽出する。すなわち、疎水性指標値特定部 104 が生成した数値列から、N末端から 30 残基以降のアミノ酸残基を示す部分数値列を抽出する。

30

【 0 0 4 9 】

<ステップ S 7 : N末端外平均疎水性値を算出>

次に、N末端外疎水性値算出部 107 は当該部分数値列の連続する N末端外疎水性特性抽出必要数分の各疎水性指標値の平均である N末端外平均疎水性値を、1 残基ずつずらしながら算出する。

ここで、N末端外疎水性特性抽出必要数の連続するアミノ酸残基列の中央のアミノ酸残基の位置が N末端から r 残基目であるときの N末端側平均疎水性値は、N末端側平均疎水性値と同様に、式 (1) を用いて算出できる。なお、このとき、C末端から n 残基以内のアミノ酸残基は、前後 n 残基の平均値を算出できないため、N末端外平均疎水性値として例えば N U L L 値を代入しておくが良い。

40

【 0 0 5 0 】

本実施形態では、N末端外疎水性特性抽出必要数として 17 残基を用いる。つまり、N末端外平均疎水性値として、N末端から r 残基目のアミノ酸残基を中心とする前後 8 残基のアミノ酸残基の疎水性指標値の平均を算出する。ここで、N末端外疎水性特性抽出必要数を 17 残基と決定する方法を説明する。

【 0 0 5 1 】

まず、既知の複数の G P I アンカー型タンパク質の N末端側の高疎水性領域以外の領域、すなわち N末端から 30 残基以降のアミノ酸残基から、N末端外疎水性特性抽出必要数

50

の候補となる個数の連続するアミノ酸残基列の平均疎水性値を、1残基ずつずらしながら算出する。次に、既知の複数のGPIアンカー型タンパク質のそれぞれの平均疎水性値の最大値を抽出する。そして、抽出した最大値の集合における最小値を抽出する。

次に、既知の複数の非GPIアンカー型タンパク質における既知の複数のGPIアンカー型タンパク質のN末端側の高疎水性領域に対応する領域以外の領域、すなわち既知の複数の非GPIアンカー型タンパク質のN末端から30残基以降のアミノ酸残基から、N末端外疎水性特性抽出必要数の候補となる範囲の平均疎水性値を、1残基ずつずらしながら算出する。そして、非GPIアンカー型タンパク質から算出した平均疎水性値の最大値のうち、既知の複数のGPIアンカー型タンパク質のそれぞれの平均疎水性値の最大値の集合から抽出した最小値より値が大きいものの個数を計数する。

10

この処理をN末端外疎水性特性抽出必要数の候補となる値を変えて実行し、非GPIアンカー型タンパク質から算出した平均疎水性値の最大値のうち、既知の複数のGPIアンカー型タンパク質のそれぞれの平均疎水性値の最大値の集合から抽出した最小値より値が大きいものの個数が最小となるN末端外疎水性特性抽出必要数の候補を、N末端外疎水性特性抽出必要数として決定する。

#### 【0052】

図10は、GPIアンカー型タンパク質の疎水性プロファイルを示す第2のグラフである。

図10は、SWISS-PROT ver 54.0のBY55\_HUMAN(181a)エントリーに対して、ステップS7と同様の処理によって算出したN末端外平均疎水性値(17残基平均の場合)を示すグラフである。ここで、横軸は、N末端外疎水性特性抽出必要数の連続するアミノ酸残基列の中央に位置するアミノ酸残基のN末端からの残基位置を示し、縦軸はN末端外平均疎水性値の値を示す。

20

図10に示すように、既知のGPIアンカー型タンパク質のC末端側の領域は、N末端からの30残基に次いで疎水性が高い。

#### 【0053】

そして、本実施形態では、SWISS-PROT ver 54.0より取得した既知の哺乳類GPIアンカー型タンパク質の完全長アミノ酸配列データセット、及び既知の哺乳類非GPIアンカー型タンパク質の完全長アミノ酸配列データセットを用いて上述した方法を実行した結果、N末端外疎水性特性抽出必要数を17残基として決定した。

30

#### 【0054】

<ステップS8：N末端外平均疎水性値の最大値の判定>

ステップS7で、N末端外疎水性値算出部107が、部分数値列の連続するN末端外疎水性特性抽出必要数分の各疎水性指標値の平均であるN末端外平均疎水性値を、1残基ずつずらしながら算出すると、N末端外疎水性判定部108は、算出したN末端外平均疎水性値の最大値がN末端外疎水性閾値以上であるか否かを判定する。なお、N末端外疎水性閾値は、既知のGPIアンカー型タンパク質のN末端外平均疎水性値の特性を示す閾値であり、本実施形態では、N末端外疎水性閾値として1.38を用いている。

1.38という値は、予め既知の複数のGPIアンカー型タンパク質に対してN末端外平均疎水性値の算出を行い、当該算出されたN末端側平均疎水性値の最大値の集合における最小値として算出された値である。

40

#### 【0055】

図11は、既知のGPIアンカー型タンパク質及び既知の非GPIアンカー型タンパク質のN末端外平均疎水性値の最大値を示すグラフである。ここで、横軸は、N末端外疎水性特性抽出必要数の連続するアミノ酸残基列の中央に位置するアミノ酸残基のC末端からの残基位置を示し、縦軸はN末端外平均疎水性値の値を示す。

図11に示すように、既知のGPIアンカー型タンパク質のN末端から30残基以降のアミノ酸残基から算出されたN末端外平均疎水性値の最大値は、N末端外疎水性閾値である1.38以上の値となる。従って、検査対象タンパク質のN末端から30残基以降のアミノ酸残基から算出されたN末端外平均疎水性値の最大値が1.38以上であれば、検査

50

対象タンパク質がGPIアンカー型タンパク質である可能性が高く、当該最大値が1.38未満であれば、検査対象タンパク質がGPIアンカー型タンパク質である可能性が低いと判定できる。

【0056】

<ステップS9：N末端外平均疎水性値が最大となるアミノ酸残基位置の判定>

N末端外疎水性判定部108が、算出したN末端外平均疎水性値の最大値がN末端外疎水性閾値以上であると判定した場合（ステップS8：YES）、C末端側最大疎水位置判定部109は、ステップS7で算出したN末端外平均疎水性値が最大となるアミノ酸残基の位置が、GPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内にあるか否かを判定する。

10

本実施形態では、GPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域として、C末端から14残基以内のアミノ酸残基を用いる。C末端から14残基以内のアミノ酸残基という領域は、既知の複数のGPIアンカー型タンパク質のN末端側の高疎水性領域以外の領域、すなわちN末端から30残基以降のアミノ酸残基のそれぞれに対してN末端外平均疎水性値を算出した場合に、当該算出したN末端外平均疎水性値が最大となる連続するアミノ酸残基列の中央に位置するアミノ酸残基が含まれる領域である。

【0057】

図11に示すように、既知のGPIアンカー型タンパク質のN末端から30残基以降のアミノ酸残基から算出されたN末端外平均疎水性値が最大となるアミノ酸残基列の中央に位置するアミノ酸残基は、C末端側の高疎水性領域内に存在する。従って、検査対象タンパク質のN末端から30残基以降のアミノ酸残基から算出されたN末端外平均疎水性値が最大となるアミノ酸残基列の中央に位置するアミノ酸残基がGPIアンカー型タンパク質のC末端側の高疎水性領域に対応する領域内に存在すれば、検査対象タンパク質がGPIアンカー型タンパク質である可能性が高く、当該領域内に存在しなければ、検査対象タンパク質がGPIアンカー型タンパク質である可能性が低いと判定できる。

20

つまり、図11における網掛け矩形の範囲が、N末端外疎水性閾値及びC末端側の高疎水性領域の条件を満たす範囲を示し、当該範囲内に含まれる非GPIアンカー型タンパク質の個数が最小となるよう、N末端外疎水性閾値及びC末端側の高疎水性領域に対応する領域とを決定している。

【0058】

<ステップS10：小側鎖サイズ判定領域の残基を抽出>

C末端側最大疎水位置判定部109が、N末端外平均疎水性値が最大となるアミノ酸残基の位置がC末端から14残基以内の位置であると判定した場合（ステップS9：YES）側鎖サイズ指標値特定部111は、ステップS1で配列取得部102が取得したアミノ酸配列情報から、小側鎖サイズ判定領域のアミノ酸残基に相当する部分配列を抽出する。ここで、小側鎖サイズ判定領域とは、既知のGPIアンカー型タンパク質のプロペプチド領域を含む領域であり、本実施形態では、C末端から30残基以内のアミノ酸残基を用いる。C末端から30残基以内のアミノ酸残基という領域は、既知のGPIアンカー型タンパク質において、後述するステップS12と同様の処理によって平均側鎖サイズを算出した場合に、当該算出した平均側鎖サイズが最小となるアミノ酸残基列の中央に位置するアミノ酸残基が含まれる領域である。

30

40

【0059】

<ステップS11：側鎖サイズ指標値を特定>

側鎖サイズ指標値特定部111は、ステップS10で小側鎖サイズ判定領域のアミノ酸残基に相当する部分配列を抽出すると、側鎖サイズ指標値記憶部110を参照して、抽出した部分配列が示す各アミノ酸残基に側鎖サイズ指標値を割り当てた数値列を生成する（ステップS11）。例えば、配列取得部102が取得したアミノ酸配列情報が、「MLLEPGRGCC...」という配列を示す場合、側鎖サイズ指標値特定部111は、側鎖サイズ指標値記憶部110が記憶する図3に示す指標値より「6、5.5、5.5、5、5.5、0.5、7.5、0.5、3、3...」という数値列を生成する。

50

## 【 0 0 6 0 】

<ステップ S 1 2 : 平均側鎖サイズを算出>

ステップ S 1 1 で、側鎖サイズ指標値特定部 1 1 1 が側鎖サイズ指標値を示す数値列を生成すると、側鎖サイズ算出部 1 1 2 は、側鎖サイズ指標値特定部 1 1 1 が生成した数値列の連続する側鎖サイズ特性抽出必要数分の各側鎖サイズ指標値の平均である平均側鎖サイズを、1 残基ずつずらしながら算出する。

ここで、平均側鎖サイズ特性抽出必要数の連続するアミノ酸残基列の中央のアミノ酸残基の位置が N 末端から r 残基目であるときの平均側鎖サイズは、式 ( 2 ) を用いて算出できる。

## 【 0 0 6 1 】

【数 2】

$$\frac{1}{2n+1} \sum_{i=r-n}^{r+n} V(i) \quad \cdot \cdot \cdot (2)$$

## 【 0 0 6 2 】

但し、n は、平均化に用いる前後の残基数を示す。つまり、2 n + 1 は、側鎖サイズ特性抽出必要数を示す。また、V ( i ) は N 末端から i 残基目に存在するアミノ酸残基の側鎖サイズ指標値を示す。

つまり、N 末端から r 残基目のアミノ酸残基が中央に位置するアミノ酸残基列の平均側鎖サイズは、N 末端から r - n 残基目のアミノ酸残基から、N 末端から r + n 残基目のアミノ酸残基までの側鎖サイズ指標値の平均となる。なお、このとき、C 末端から n 残基以内のアミノ酸残基は、前後 n 残基の平均値を算出できないため、平均側鎖サイズとして例えば N U L L 値を代入しておくが良い。

## 【 0 0 6 3 】

本実施形態では、側鎖サイズ特性抽出必要数として 3 残基を用いる。つまり、N 末端側平均疎水性値として、N 末端から r 残基目のアミノ酸残基に隣接するアミノ酸残基の疎水性指標値の平均を算出する。ここで、側鎖サイズ特性抽出必要数を 3 残基と決定する方法を説明する。

## 【 0 0 6 4 】

まず、既知の複数の G P I アンカー型タンパク質の小側鎖サイズ判定領域、すなわち C 末端から 3 0 残基以内のアミノ酸残基から、側鎖サイズ特性抽出必要数の候補となる範囲の平均疎水性値を、1 残基ずつずらしながら算出する。次に、既知の複数の G P I アンカー型タンパク質のそれぞれから、平均側鎖サイズが最小となるアミノ酸残基を特定する。そして、当該抽出したアミノ酸残基の C 末端側に隣接するアミノ酸残基が G P I アンカー修飾部位 ( サイト ) であるものの個数を計数する。

この処理を N 末端側疎水性特性抽出必要数の候補となる値を変えて実行し、全 G P I アンカー型タンパク質のうち、平均側鎖サイズが最小となるアミノ酸残基の C 末端側に隣接するアミノ酸残基が G P I アンカー修飾部位であるものの個数が最大となる側鎖サイズ特性抽出必要数の候補を、側鎖サイズ特性抽出必要数として決定する。

## 【 0 0 6 5 】

そして、本実施形態では、S W I S S - P R O T v e r 5 4 . 0 より取得した既知の哺乳類 G P I アンカー型タンパク質の完全長アミノ酸配列データセット、及び既知の哺乳類非 G P I アンカー型タンパク質の完全長アミノ酸配列データセットを用いて上述した方法を実行した結果、N 末端側疎水性特性抽出必要数を 3 残基として決定した。

## 【 0 0 6 6 】

図 1 2 は、G P I アンカー型タンパク質の側鎖サイズのプロファイルを示すグラフである。

図 1 2 は、S W I S S - P R O T v e r 5 4 . 0 の B Y 5 5 \_ H U M A N ( 1 8 1 a

10

20

30

40

50

a) エントリーに対して、ステップS 1 2と同様の処理によって算出した平均側鎖サイズを示すグラフである。ここで、横軸は、平均側鎖サイズのアミノ酸残基列の中央に位置するアミノ酸残基のC末端からの残基位置を示し、縦軸は平均側鎖サイズの値を示す。

図1 2に示すように、既知のG P Iアンカー型タンパク質のG P Iアンカー修飾部位は、平均側鎖サイズが最小となるアミノ酸残基のC末端側に隣接している。

#### 【0067】

<ステップS 1 3：所定の領域のアミノ酸残基を抽出>

図1 3は、アミノ酸配列の抽出方法を示す図である。

ステップS 1 2で、側鎖サイズ算出部1 1 2が平均側鎖サイズを算出すると、スコア算出部1 1 4は、図1 3(1)に示すように、側鎖サイズ算出部1 1 2が算出した平均側鎖サイズが最小となるアミノ酸残基の位置を基準位置として決定する。次に、スコア算出部1 1 4は、図1 3(2)に示すように、当該基準位置を含む所定の領域におけるアミノ酸残基を、ステップS 1で配列取得部1 0 2が取得したアミノ酸配列情報から抽出する。

本実施形態では、当該所定の領域として、基準位置からN末端側に連続する1 0残基のアミノ酸残基とC末端側に連続する1 2残基のアミノ酸残基とを用いる。すなわち、所定の領域とは、既知のG P Iアンカー型タンパク質において、G P Iアンカー修飾部位の前後1 1残基のアミノ酸残基を含む領域である。

#### 【0068】

<ステップS 1 4：位置特異的スコアを割り当てる>

図1 4は、位置特異的スコアの割り当て方法を示す図である。

次に、スコア算出部1 1 4は、P S S M記憶部1 1 3が記憶するP S S Mに基づいて、抽出した所定の範囲の各アミノ酸残基の位置特異的スコアを特定し、当該疎水性指標値を示す数値列を生成する。例えば、抽出した所定の範囲のアミノ酸残基が、図1 4に示すように「C Q N A ... S」という配列を示す場合、スコア算出部1 1 4は、図4及び図5に示すP S S Mを参照して、「1.02、2.44、2.12、1.31、...、1.13」という数値列を生成する。

#### 【0069】

ここで、ステップS 1 4で用いるP S S Mの作成方法を説明する。

まず、既知の哺乳類G P Iアンカー型タンパク質の完全長アミノ酸配列データセット、及び既知の哺乳類非G P Iアンカー型タンパク質の完全長アミノ酸配列データセットを、取得する。本実施形態では、これらのデータセットをS W I S S - P R O T ver 5 4 . 0より取得した。また、G P Iアンカー型タンパク質のデータセットについては、当該アミノ酸配列から翻訳されるG P Iアンカー型タンパク質としての特性が実証されていないもの、明らかに完全長ではないもの等を除外した。その結果、G P Iアンカー型タンパク質のエントリー数は3 9 1であり、非G P Iアンカー型タンパク質のエントリー数は4 8 9 8 3であった。

#### 【0070】

データセットを取得すると、次に、データセットの各エントリーについて、疎水性のスクリーニングを行う。

まず、上述した式(1)及び図2に示す疎水性指標値を用いて、N末端側疎水性特性抽出必要数を1 1残基に設定して(すなわち、式(1)において $n = 5$ に設定して)各エントリーのN末端平均疎水性値を算出し、N末端から3 0残基以内の領域における最大のN末端側平均疎水性値が1 . 5 0以上のものを抽出する。次に、抽出されたデータセット中の各エントリーの平均疎水性値を、前記式(1)及び図2に示す疎水性指標値を用いて、N末端外疎水性特性抽出必要数を1 7残基に設定して(すなわち、式(1)において $n = 8$ に設定して)算出し、N末端から3 0残基を除く全領域における最大のN末端外平均疎水性値が1 . 3 8であり、且つ、該最大のN末端外平均疎水性値を示す残基位置がC末端から1 4残基以内であるものを抽出する。この結果、実際は完全長でないエントリーや、タンパク質としての発現が推定であるエントリーは排除されることとなる。本実施形態では、疎水性スクリーニング後のG P Iアンカー型タンパク質データセットのエントリー数

10

20

30

40

50

は 1 2 1 であり、非 G P I アンカー型タンパク質データセットのエントリー数は 2 1 8 であった。

【 0 0 7 1 】

次いで、疎水性スクリーニングにより抽出されたデータセットに含まれる同一アミノ酸配列を有するエントリーを除き、冗長性を排除する。この結果、本実施形態では、G P I アンカー型タンパク質データセットのエントリー数は 1 1 3 であり、非 G P I アンカー型タンパク質データセットのエントリー数は 2 1 0 であった。冗長性を排除した G P I アンカー型タンパク質データセットに含まれる 1 1 3 の S W I S S - P R O T エントリーネームを図 1 5 に示す。

【 0 0 7 2 】

上記により得られた各データセット中の各エントリーの C 末端から 3 0 アミノ酸残基までの平均側鎖サイズを、上述した式 ( 2 ) 及び図 3 に示す側鎖サイズ指標値を用いて、側鎖サイズ特性抽出必要数を 3 に設定して ( すなわち、式 ( 2 ) において  $n = 1$  に設定して ) 算出する。

そして、データセットのうち G P I アンカー型タンパク質の各エントリーの、平均側鎖サイズが最小となるアミノ酸残基の位置を基準位置とする所定の範囲 ( 基準位置のアミノ酸残基と基準位置から N 末端側に連続する 1 0 残基のアミノ酸残基と C 末端側に連続する 1 2 残基のアミノ酸残基とからなる範囲 ) におけるアミノ酸残基から、式 ( 3 ) を用いて既知の G P I アンカー型タンパク質の所定の領域内の位置  $p$  に存在するアミノ酸残基の種類  $i$  の出現頻度を算出する。

【 0 0 7 3 】

【 数 3 】

$$\frac{n_{ip} + \varepsilon \frac{1}{s}}{\sum_{i=1}^s n_{ip} + \varepsilon} \cdot \cdot \cdot (3)$$

【 0 0 7 4 】

但し、 $n_{ip}$  は、種類  $i$  のアミノ酸残基が位置  $p$  に存在する既知の G P I アンカー型タンパク質の個数を示す。また、 $\varepsilon$  は算出する出現頻度の調整値を示し、本実施形態では 1 を用いている。また、 $s$  は、アミノ酸残基の種類数を示す。

これにより、データセットの全てのエントリーにおいて位置  $p$  に種類  $i$  が存在しない場合にも、ゼロで除算を行うことを防ぐことができる。

同様に、データセットのうち非 G P I アンカー型タンパク質の各エントリーの、平均側鎖サイズが最小となるアミノ酸残基の位置を基準位置とする所定の範囲におけるアミノ酸残基から、式 ( 3 ) を用いて既知の非 G P I アンカー型タンパク質の所定の領域内の位置  $p$  に存在するアミノ酸残基の種類  $i$  の出現頻度を算出する。

【 0 0 7 5 】

既知の G P I アンカー型タンパク質の所定の領域内の位置  $p$  に存在するアミノ酸残基の種類  $i$  の出現頻度、及び既知の非 G P I アンカー型タンパク質の所定の領域内の位置  $p$  に存在するアミノ酸残基の種類  $i$  の出現頻度を算出すると、次に、式 ( 4 ) を用いて、アミノ酸残基の位置  $p$  におけるアミノ酸残基の種類  $i$  の位置特異的スコアを算出する。

【 0 0 7 6 】

10

20

30

40



【数4】

$$\ln \frac{f_{ip}^{positive}}{f_{ip}^{negative}} \quad \dots (4)$$

【0077】

但し、 $f_{ip}^{positive}$  は、既知の GPI アンカー型タンパク質の所定の領域内の位置 p に存在するアミノ酸残基の種類 i の出現頻度を示す。また、 $f_{ip}^{negative}$  は、既知の非 GPI アンカー型タンパク質の所定の領域内の位置 p に存在するアミノ酸残基の種類 i の出現頻度を示す。つまり、位置特異的スコアは、所定の範囲におけるあるアミノ酸残基の位置におけるアミノ酸残基の種類、GPI アンカー型タンパク質における出現度合いを示している。

10

このように算出された位置特異的スコアを要素とする 23 (所定の領域内のアミノ酸残基数) × 20 (アミノ酸残基の種類数) の行列を PSSM として生成し、PSSM 記憶部 113 に格納しておく。これにより、図4及び図5に示す PSSM を生成することができる。

【0078】

<ステップ S15 : スコア算出>

20

スコア算出部 114 は、ステップ S14 で各アミノ酸残基の位置特異的スコアを示す数値列を生成すると、当該生成した数値列の平均値をスコアとして算出する。つまり、スコアは、式(5)を用いて算出できる。

【0079】

【数5】

$$\frac{1}{n+m+1} \sum_{i=n}^m S_{ip} \quad \dots (5)$$

30

【0080】

但し、n は、基準位置から N 末端側に連続するアミノ酸残基の個数を示し、本実施形態では 10 残基を示す。また、m は、基準位置から C 末端側に連続するアミノ酸残基の個数を示し、本実施形態では 12 残基を示す。つまり、n + m + 1 は、所定の領域のアミノ酸残基数を示す。また、 $S_{ip}$  は、基準位置を 0 とした場合の位置 p における種類 i のアミノ酸残基の位置特異的スコアを示す。これにより、スコア算出部 114 は、ステップ S12 で側鎖サイズ算出部 112 が算出した平均側鎖サイズが最小となる位置を基準位置とする所定の領域のアミノ酸残基のそれぞれに対応する位置特異的スコアの平均値を、スコアとして算出する。

なお、検査対象タンパク質の基準位置が C 末端から m 残基以内にある場合は、式(5)の m の代わりに C 末端から基準位置までのアミノ酸残基数を用いる。

40

【0081】

<ステップ S16 : スコアの判定>

ステップ S11 でスコア算出部 114 がスコアを算出すると、スコア判定部 115 は、算出したスコアがスコア判定閾値以上であるか否かを判定する。なお、スコア判定閾値は、検査対象タンパク質と既知の GPI アンカー型タンパク質とが類似するか否かを判定する閾値であり、本実施形態では、スコア判定閾値として -0.102 を用いている。

【0082】

以下に、スコア判定閾値の決定方法を説明する。

まず、上述した既知の GPI アンカー型タンパク質及び既知の非 GPI アンカー型タン

50

パク質のデータセットを用いてPSSMを生成する。なお、ここで生成されるPSSMは、本実施形態で用いたPSSMと同一のものとなる。そして、当該PSSMを用いて、PSSMの生成に用いたデータセットの各エントリーのスコアを算出する。

【0083】

図16は、既知の哺乳類GPIアンカー型タンパク質の完全長アミノ酸配列データセット、及び既知の哺乳類非GPIアンカー型タンパク質の完全長アミノ酸配列データセットのエントリーに対して算出したスコアの分布を示す図である。ここで、横軸はスコアの値を示し、縦軸は当該スコアとなるエントリーの出現度数を示す。

図16に示すように、既知の非GPI型タンパク質のスコアは低く、逆に既知のGPI型タンパク質のスコアは高くなる。

次に、既知のGPI型タンパク質のスコアの最小値と、既知の非GPI型タンパク質のスコアの最大値との間の値をスコア判定閾値の候補として、式(6)を用いて判別の成功率を算出する。

【0084】

【数6】

$$\sqrt{\frac{TP}{TP+FN} \times \frac{TP}{TP+FP}} \quad \dots (6)$$

【0085】

但し、TP(True Positive)は、スコアがスコア判定閾値の候補の値以上である既知のGPIアンカー型タンパク質の個数を示す値であり、GPIアンカー型タンパク質がGPIアンカー型タンパク質であると正しく判定された数を示す。また、FN(False Negative)は、スコアがスコア判定閾値の候補の値未満である既知のGPIアンカー型タンパク質の個数を示す値であり、GPIアンカー型タンパク質が非GPIアンカー型タンパク質であると誤って判定された数を示す。また、FP(False Positive)は、スコアがスコア判定閾値の候補の値以上である既知の非GPIアンカー型タンパク質の個数を示す値であり、非GPIアンカー型タンパク質がGPIアンカー型タンパク質であると誤って判定された数を示す。

すなわち、 $TP / (TP + FN)$ は、判別の感度を示し、 $TP / (TP + FP)$ は、判別の選択性を示す。

上述した式(6)による判別の成功率の計算を、複数のスコア判定閾値の候補(例えば、既知のGPI型タンパク質のスコアの最小値と、既知の非GPI型タンパク質のスコアの最大値との間の値を0.05ずつの間隔で抽出した値)について実行する。

【0086】

図17は、スコア判定閾値の候補それぞれを用いた場合の感度と選択性とを示す表である。

図17に示すように、感度は、スコア判定閾値が低いほど高く、選択性は、スコア判定閾値が高いほど高くなる。

そして、算出された成功率が最大となるスコア判定閾値の候補をスコア判定閾値として決定する。本実施形態では、上述した手法により成功率を算出した結果、スコア判定閾値として-0.102を用いることで、成功率0.969を得ることができた。

【0087】

なお、本実施形態では成功率が最大となるスコア判定閾値の候補をスコア判定閾値として決定する場合を説明したが、これに限られず、GPIアンカー型タンパク質をできるだけ多く検出したい場合、すなわち判別の感度を高めたい場合は、スコア判定閾値をより低い値に設定すると良い。また、非GPIアンカー型タンパク質をできるだけ排除したい場合、すなわち判別の選択性を高めたい場合は、スコア判定閾値をより高い値に設定すると

10

20

30

40

50

良い。この場合、スコア判定閾値は、既知のGPI型タンパク質のスコアの最小値と、既知の非GPI型タンパク質のスコアの最大値との間の値であることが望ましい。

【0088】

<ステップS17：GPIアンカー型タンパク質と判定>

ステップS16でスコア判定部115が、算出したスコアがスコア判定閾値以上であると判定した場合（ステップS16：YES）、GPIアンカー型タンパク質判定部116は、ステップS1で配列取得部102が取得したアミノ酸配列情報が、GPIアンカー型タンパク質のものであると判定する。

【0089】

<ステップS18：非GPIアンカー型タンパク質と判定>

他方、ステップS5でN末端側疎水性判定部106が、算出したN末端側平均疎水性値の最大値がN末端側疎水性閾値未満であると判定した場合（ステップS5：NO）、ステップS8でN末端外疎水性判定部108が、算出したN末端外平均疎水性値の最大値がN末端外疎水性閾値未満であると判定した場合（ステップS8：NO）、ステップS9でC末端側最大疎水位置判定部109が、N末端外平均疎水性値が最大となるアミノ酸残基の位置がC末端側の高疎水性領域に対応する領域内ないと判定した場合（ステップS9：NO）、またはステップS16でスコア判定部115が、算出したスコアがスコア判定閾値未満であると判定した場合（ステップS16：NO）、GPIアンカー型タンパク質判定部116は、ステップS1で配列取得部102が取得したアミノ酸配列情報が、非GPIアンカー型タンパク質のものであると判定する。

【0090】

上述した動作により、GPIアンカー型タンパク質判定装置100は、高感度且つ高選択的に検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定することができる。

図18は、本実施形態によるGPIアンカー型タンパク質判定装置の判定精度を示す表である。ここで、判定精度とは、判定の感度、選択性、及び成功率のことを言う。

図18では、GPIアンカー型タンパク質判定装置100がGPIアンカー型タンパク質であると判定した検査対象タンパク質の判定精度、及び非GPIアンカー型タンパク質であると判定した検査対象タンパク質の判定精度を示している。また、図18では、GPIアンカー型タンパク質及び非GPIアンカー型タンパク質それぞれの判定精度について、Self-consistency（自己無撞着）な手法による判定精度と、4-fold cross validation法（4分割交差検定法）による判定精度とを示している。

【0091】

自己無撞着な手法による判定精度とは、以下の処理により算出した判定精度である。

まず、上述したスコア判定閾値の決定方法と同様に、既知のGPIアンカー型タンパク質及び既知の非GPIアンカー型タンパク質のデータセットを用いてPSSMを生成し、当該PSSMを用いて、PSSMの生成に用いたデータセットの各エントリーのスコアを算出する。そして、当該算出したスコアに基づいてデータセット全体に対する判定精度を算出する。

図18に示すように、本実施形態による、GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が96.5%、選択性が97.3%、成功率が0.969であった。また、非GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が98.6%、選択性が98.1%、成功率が0.983であった。

【0092】

他方、4分割交差検定法による判定精度とは、以下の処理により算出した判定精度である。

まず、既知のGPIアンカー型タンパク質及び既知の非GPIアンカー型タンパク質のデータセットを4等分する。次に、分割したデータセットのうち3つの部分データセットを用いてPSSMを生成し、当該PSSMに基づいて、他の1つの部分データセットの各

10

20

30

40

50

エントリーのスコアを算出する。次に、当該算出したスコアに基づいて、感度、選択性、成功率を算出する。そして、PSSMを生成する部分データセットとスコアを算出する部分データセットとの全ての組み合わせに対して判定精度を算出し、それぞれの平均値をデータセット全体に対する判定精度として算出する。

図18に示すように、本実施形態による、GPIアンカー型タンパク質の4分割交差検定法による判定精度は、感度が86.8%、選択性が96.4%、成功率が0.915であった。また、非GPIアンカー型タンパク質の4分割交差検定法による判定精度は、感度が98.2%、選択性が93.1%、成功率が0.956であった。

#### 【0093】

以下に、N末端側疎水性特性抽出必要数、N末端外疎水性特性抽出必要数、側鎖サイズ特性抽出必要数、及び基準位置を含む所定の範囲をそれぞれ変化させてGPIアンカー型タンパク質の判定を行った場合の判定精度を示す。

図19は、N末端側疎水性特性抽出必要数を11から13に変更した場合の判定精度を示す表である。

図19に示すように、N末端側疎水性特性抽出必要数を13とした場合の、GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が89.9%、選択性が97.2%、成功率が0.935であった。また、4分割交差検定法による判定精度は、感度が88.0%、選択性が83.8%、成功率が0.858であった。

他方、非GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が96.9%、選択性が89.1%、成功率が0.929であった。また、4分割交差検定法による判定精度は、感度が79.4%、選択性が85.0%、成功率が0.820であった。

N末端側疎水性特性抽出必要数を13とした場合の判定精度を、図18に示す本実施形態による判定精度(N末端側疎水性特性抽出必要数を11とした場合の判定精度)とを比較すると、GPIアンカー型タンパク質の4分割交差検定法による感度は、N末端側疎水性特性抽出必要数を13とした場合のほうが高いが、他の判定精度は、本実施形態による判定精度のほうが高いことが分かる。

#### 【0094】

図20は、N末端外疎水性特性抽出必要数を17から13に変更した場合の判定精度を示す表である。

図20に示すように、N末端外疎水性特性抽出必要数を13とした場合の、GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が95.7%、選択性が91.7%、成功率が0.936であった。また、4分割交差検定法による判定精度は、感度が87.3%、選択性が86.4%、成功率が0.868であった。

他方、非GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が90.4%、選択性が95.0%、成功率が0.926であった。また、4分割交差検定法による判定精度は、感度が84.5%、選択性が85.9%、成功率が0.851であった。

N末端外疎水性特性抽出必要数を13とした場合の判定精度を、図18に示す本実施形態による判定精度(N末端外疎水性特性抽出必要数を17とした場合の判定精度)とを比較すると、GPIアンカー型タンパク質の4分割交差検定法による感度は、N末端外疎水性特性抽出必要数を13とした場合のほうが高いが、他の判定精度は、本実施形態による判定精度のほうが高いことが分かる。

#### 【0095】

図21は、側鎖サイズ特性抽出必要数を3から5に変更した場合の判定精度を示す表である。

図21に示すように、側鎖サイズ特性抽出必要数を5とした場合の、GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が92.9%、選択性が98.1%、成功率が0.955であった。また、4分割交差検定法による判定精度は、感度が81.4%、選択性が83.6%、成功率が0.823であった。

他方、非GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が99.1%、選択性が96.3%、成功率が0.977であった。また、4分割交差検定法による判定精度は、感度が90.4%、選択性が90.2%、成功率が0.902であった。

側鎖サイズ特性抽出必要数を5とした場合の判定精度を、図18に示す本実施形態による判定精度(側鎖サイズ特性抽出必要数を3とした場合の判定精度)とを比較すると、GPIアンカー型タンパク質の自己無撞着な手法による選択性、及び非GPI型タンパク質の自己無撞着な手法による感度は、側鎖サイズ特性抽出必要数を5とした場合のほうが高いが、他の判定精度は、本実施形態による判定精度のほうが高いことが分かる。

【0096】

図22は、基準位置を含む所定の範囲を基準位置から(-10残基~+12残基)を(-12残基~+9残基)に変更した場合の判定精度を示す表である。

図22に示すように、所定の範囲を、基準位置からN末端側に12残基、C末端側に9残基の範囲とした場合の、GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が96.5%、選択性が95.6%、成功率が0.960であった。また、4分割交差検定法による判定精度は、感度が80.1%、選択性が92.7%、成功率が0.860であった。

他方、非GPIアンカー型タンパク質の自己無撞着な手法による判定精度は、感度が97.6%、選択性が98.1%、成功率が0.979であった。また、4分割交差検定法による判定精度は、感度が96.4%、選択性が89.9%、成功率が0.931であった。

基準位置を含む所定の範囲を、基準位置からN末端側に12残基、C末端側に9残基の範囲とした場合の判定精度を、図18に示す本実施形態による判定精度(基準位置を含む所定の範囲を、基準位置からN末端側に10残基、C末端側に12残基の範囲とした場合の判定精度)とを比較すると、本実施形態による判定精度のほうが高いことが分かる。

【0097】

このように、本実施形態によれば、N末端側の疎水性及びC末端側の疎水性の高低と、PSSMとを用いているため、高感度且つ高選択的に検査対象タンパク質がGPIアンカー型タンパク質であるか否かを判定することができる。

【0098】

以上、図面を参照してこの発明の一実施形態について詳しく説明してきたが、具体的な構成は上述のものに限られることはなく、この発明の要旨を逸脱しない範囲内において様々な設計変更等を行うことが可能である。

例えば、本実施形態では、タンパク質の完全長アミノ酸配列情報を検査対象として判定を行ったが、これに限られず、完全長塩基配列情報を検査対象として判定を行っても良い。但し、この場合、ステップS1で配列取得部102が完全長塩基配列情報を取得した後、図示しない翻訳処理部が、常法によるイントロ配列の除去処理及びアミノ酸配列情報への翻訳処理を行い、当該アミノ酸配列情報を用いてステップS2以降の処理を行う。

【0099】

上述のGPIアンカー型タンパク質判定装置100は内部に、コンピュータシステムを有している。そして、上述した各処理部の動作は、プログラムの形式でコンピュータ読み取り可能な記録媒体に記憶されており、このプログラムをコンピュータが読み出して実行することによって、上記処理が行われる。ここでコンピュータ読み取り可能な記録媒体とは、磁気ディスク、光磁気ディスク、CD-ROM、DVD-ROM、半導体メモリ等をいう。また、このコンピュータプログラムを通信回線によってコンピュータに配信し、この配信を受けたコンピュータが当該プログラムを実行するようにしても良い。

【0100】

また、上記プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル(差分プログラム)であっても良い。

10

20

30

40

50

【符号の説明】

【0101】

100... G P I アンカー型タンパク質判定装置 101... 配列記憶部 102... 配列取得部 103... 疎水性指標値記憶部 104... 疎水性指標値特定部 105... N末端側疎水性値算出部 106... N末端側疎水性判定部 107... N末端外疎水性値算出部 108... N末端外疎水性判定部 109... C末端側最大疎水位置判定部 110... 側鎖サイズ指標値記憶部 111... 側鎖サイズ指標値特定部 112... 側鎖サイズ算出部 113... P S S M 記憶部 114... スコア算出部 115... スコア判定部 116... G P I アンカー型タンパク質判定部

【要約】

【課題】高感度且つ高選択的に検査対象タンパク質が G P I アンカー型タンパク質であるか否かを判定する。

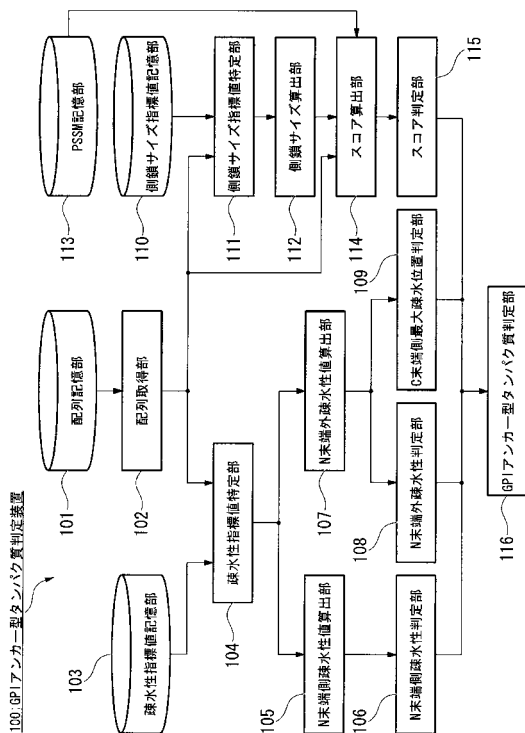
【解決手段】N末端側疎水性判定部106は、検査対象タンパク質と既知の G P I アンカー型タンパク質とのN末端側の疎水性を比較する。次に、N末端外疎水性判定部108は、検査対象タンパク質と既知の G P I アンカー型タンパク質とのN末端側以外の疎水性を比較し、C末端側最大疎水位置判定部109は、検査対象タンパク質と既知の G P I アンカー型タンパク質とのC末端側の高疎水性位置を比較する。次に、スコア判定部115は、P S S M 記憶部113が記憶する P S S M に基づいて算出された検査対象タンパク質のスコアが閾値を超えるか否かを判定する。G P I アンカー型タンパク質判定部116は、上述した比較・判定結果に基づいて、検査対象タンパク質が G P I アンカー型タンパク質

10

20

【選択図】図1

【図1】



【図2】

アミノ酸残基	指標	アミノ酸残基	指標
A	1.8	L	3.8
R	-4.5	K	-3.9
N	-3.5	M	1.9
D	-3.5	F	2.8
C	2.5	P	-1.6
Q	-3.5	S	-0.8
E	-3.5	T	-0.7
G	-0.4	W	-0.9
H	-3.2	Y	-1.3
I	4.5	V	4.2

【図3】

アミノ酸残基	指標	アミノ酸残基	指標
A	2.5	L	5.5
R	7.5	K	7.0
N	5.0	M	6.0
D	2.5	F	6.5
C	3.0	P	5.5
Q	6.0	S	3.0
E	5.0	T	5.0
G	0.5	W	7.0
H	6.0	Y	7.0
I	5.5	V	5.0

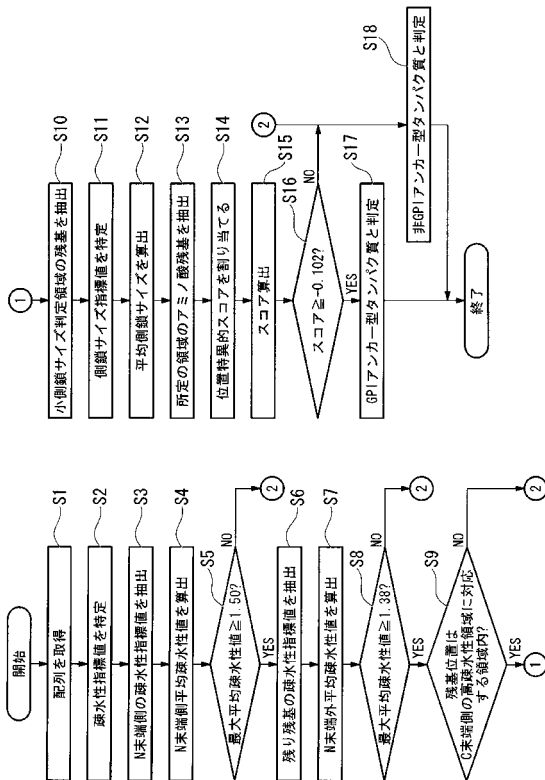
【図4】

	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.84	-0.77	2.46	1.31	-0.86	0.17	-0.07	-0.52	1.11	0.12	0.01
C	1.02	-0.07	1.91	-0.07	0.33	-0.99	1.71	-5.21	-5.50	-5.21	-4.12
D	0.28	-0.59	0.62	0.87	-6.19	-6.06	-5.21	-0.38	-0.08	-2.37	1.24
E	-1.33	0.95	-1.94	-0.08	-0.99	-1.94	0.43	-6.31	-4.12	-6.19	-0.48
F	-0.77	-0.43	-5.50	0.33	-1.17	0.62	0.39	-5.50	-1.68	-4.12	-0.77
G	0.11	-1.57	1.02	-0.36	1.15	0.90	-5.72	0.71	-0.21	0.47	-1.77
H	-1.86	-0.08	-0.08	-0.36	-1.32	-0.48	-1.09	-4.12	1.31	-4.12	-0.08
I	-6.60	-0.99	-0.77	-1.17	-0.08	-1.40	-0.99	-5.90	-1.32	6.66	-0.59
K	-2.21	-0.34	-0.66	-0.56	-0.17	0.03	-0.19	-4.12	-5.72	-6.31	-1.46
L	-1.36	-0.99	-2.08	-0.25	-1.24	-0.48	-1.89	-0.99	-0.48	-4.81	-1.43
M	-0.99	-5.21	-1.94	-5.50	-0.99	2.00	-0.99	1.24	-5.50	1.24	2.41
N	2.69	0.21	2.12	0.21	-0.56	0.03	-0.48	0.70	-5.72	-6.68	-6.88
P	0.16	-0.08	-0.30	-6.68	0.21	-0.68	0.00	-5.21	-0.99	-1.17	-0.18
Q	-0.07	2.44	-0.23	-0.48	-0.08	1.84	-0.68	-4.81	-6.06	-5.50	0.39
R	0.50	-1.33	-1.12	-0.48	-1.33	-0.80	1.02	-5.21	-6.06	-5.21	-0.48
S	1.61	0.00	-1.46	-1.32	0.90	0.11	0.28	0.59	0.40	0.66	0.75
T	0.01	-0.41	-0.72	-0.33	-0.41	0.03	-0.17	-0.59	-1.09	-0.07	0.93
V	-0.84	-1.94	-1.23	-1.86	-0.41	-0.26	-0.99	-7.29	-6.19	-6.19	1.35
W	-4.12	-6.19	-6.19	-4.81	-4.81	-0.99	-1.32	-4.12	-4.12	1.24	-5.72
Y	0.43	1.53	0.71	1.07	4.17	0.06	0.62	1.24	0.62	1.24	-6.41

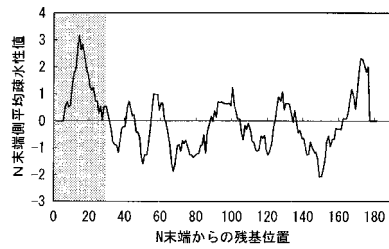
【図5】

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.31	-0.34	-1.33	0.21	-0.54	-0.77	-0.54	-0.70	-1.37	-0.15	0.21	-1.07
C	-1.17	6.66	-0.08	-0.99	-4.81	-6.60	-1.32	-1.17	-1.17	-6.19	-4.81	-6.68
D	5.97	-4.12	-0.07	1.31	-4.12	-4.81	-4.81	-6.19	1.24	-4.81	-4.12	1.24
E	0.33	-1.63	0.62	-0.30	-4.81	-5.90	-5.50	-4.81	-4.12	1.24	1.24	1.24
F	-7.33	0.50	1.91	-4.81	-7.20	-0.54	-0.40	-0.99	0.28	0.39	-6.41	-0.88
G	-0.36	0.06	-2.01	1.13	-0.83	0.37	0.17	-1.52	-0.54	-1.73	-2.59	-7.16
H	0.80	0.62	1.31	-5.50	0.15	-4.81	-0.76	-5.90	-5.72	-4.81	-4.81	1.24
I	0.90	-0.30	1.71	-0.77	-7.00	-1.68	-2.51	-0.19	1.02	-0.48	0.42	1.34
K	-6.41	-2.21	0.15	0.62	1.71	-6.41	-5.72	-6.19	-4.81	-5.21	-5.90	-4.81
L	0.26	1.95	-0.77	-0.55	-2.25	-0.22	-0.58	0.46	0.24	0.82	0.46	-0.13
M	-0.99	-0.07	-5.72	-6.82	-0.48	-5.21	-0.48	0.11	-1.32	-7.16	-0.23	-6.31
N	0.21	-5.90	0.62	0.62	6.66	-5.50	-4.81	-5.50	-4.81	-5.21	-5.72	-4.81
P	-0.08	-0.38	-1.06	-0.73	1.67	1.38	2.07	-1.17	-0.07	0.62	0.90	0.48
Q	-6.51	-0.77	-0.48	-0.77	0.62	-0.48	6.66	-4.81	1.24	5.97	-6.19	-0.99
R	-0.48	-0.88	-1.68	-1.09	-0.63	-0.07	-0.76	-4.12	-4.81	-4.12	-5.21	1.24
S	-0.30	-0.87	1.24	1.13	2.30	0.15	-0.70	-1.12	-0.19	-0.88	-0.36	1.13
T	0.26	-0.70	-0.42	-0.77	-0.39	0.95	0.33	1.09	-0.88	1.13	0.46	1.82
V	0.82	-2.59	-1.28	-1.25	-0.44	-1.12	0.69	1.36	-0.08	-0.43	0.25	-0.83
W	5.97	1.02	-5.21	-0.30	-4.12	0.95	6.66	6.66	-5.72	-4.81	-4.12	-4.81
Y	0.62	1.31	0.62	-5.50	-4.12	-4.81	-5.21	-4.81	1.24	-5.21	-4.12	-4.81

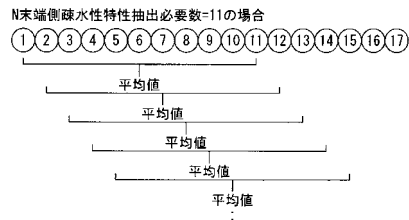
【図6】



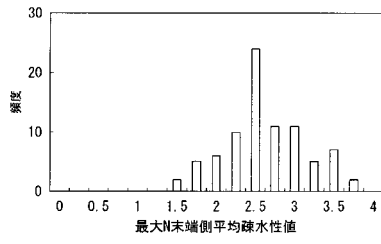
【図7】



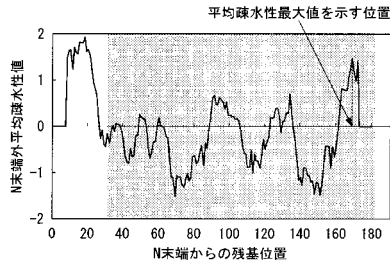
【図8】



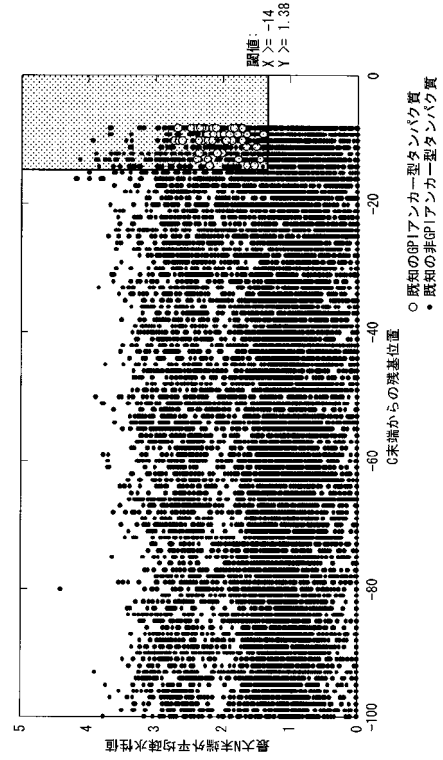
【図9】



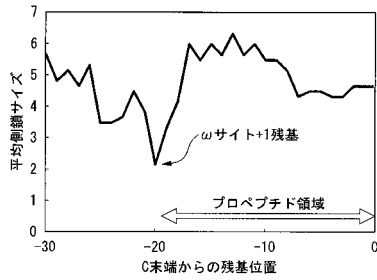
【図10】



【図11】



【図12】

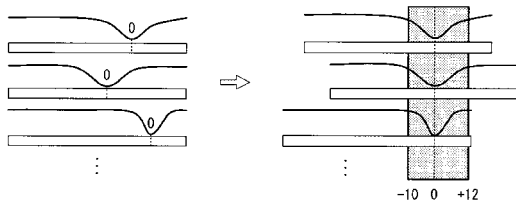


【図14】

$$\begin{array}{cccccccc}
 & -10 & -9 & -8 & -7 & \dots & 12 \\
 C & - & Q & - & N & - & A & - \dots - S \\
 \text{積算スコア} & = & [ & 1.02 & + & 2.44 & + & 2.12 & + & 1.31 & + & \dots & + & (1.13) & ] & / & 23
 \end{array}$$

【図13】

- (1) 平均側鎖サイズが最小値を示す残基位置を基準位置(0)とする。
- (2) 基準位置を含む所定の領域におけるアミノ酸残基の配列を抽出する。

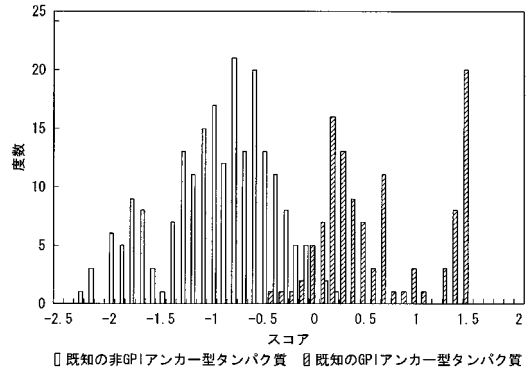




【図15】

5NTD_BOVIN	PPBI_RAT
5NTD_HUMAN	PPBN_HUMAN
5NTD_MOUSE	PPBT_BOVIN
5NTD_RAT	PPBT_FELCA
BY55_HUMAN	PPBT_HUMAN
BY55_MOUSE	PPBT_MOUSE
CAH4_BOVIN	PPBT_RAT
CAH4_HUMAN	PRIO_AILME
CAH4_MOUSE	PRIO_ANTCE
CAH4_RABIT	PRIO_ATEPA
CAH4_RAT	PRIO_BISBI
CBPM_HUMAN	PRIO_BOSTR
CBPM_MOUSE	PRIO_BUSBU
CBPM_PONPY	PRIO_BUDTA
CD48_HUMAN	PRIO_CALJA
CD48_MOUSE	PRIO_CAMDR
CD48_RAT	PRIO_CANFA
CD52_CANFA	PRIO_CAPHI
CD52_HUMAN	PRIO_CEBAP
CD52_MACFA	PRIO_CERAE
CD52_MOUSE	PRIO_CEREL
CD52_RAT	PRIO_COLGU
CD59_AOTTR	PRIO_CRIGR
CD59_CALSQ	PRIO_CRIMI
CD59_CERAE	PRIO_FELCA
CD59_HUMAN	PRIO_GORGO
CD59_PAPSP	PRIO_HUMAN
CD59_PIG	PRIO_HYLLA
CD59_PONPY	PRIO_MACAR
CD59_RABIT	PRIO_MESAU
CD59_RAT	PRIO_MOSCH
CD59_SAISS	PRIO_MOUSE
CD59A_MOUSE	PRIO_MUSPF
CD59B_MOUSE	PRIO_MUSVI
CEAM6_HUMAN	PRIO_ODOHE
CEAM8_HUMAN	PRIO_OVICA
DAF_HUMAN	PRIO_PIG
DAFT_MOUSE	PRIO_PONPY
DPEP1_BOVIN	PRIO_PREFR
DPEP1_HUMAN	PRIO_RABIT
DPEP1_MOUSE	PRIO_RAT
DPEP1_RABIT	PRIO_SAISS
DPEP1_RAT	PRIO_SHEEP
DPEP1_SHEEP	PRIO_SIGHI
FOLR1_HUMAN	PRIO_TRAIN
FOLR1_MOUSE	PRIO_TRIVU
FOLR2_HUMAN	PRIO1_TRAST
FOLR2_MOUSE	PRIO2_TRAST
NAR2A_MOUSE	PSCA_MOUSE
NAR2A_RAT	SACA4_BOVIN
NAR2B_MOUSE	SACA4_MOUSE
NAR2B_RAT	THY1_HUMAN
NAR3_HUMAN	THY1_MOUSE
PPBI_HUMAN	THY1_RAT
PPBI_BOVIN	XPP2_HUMAN
PPBI_HUMAN	XPP2_PIG
PPBI_MOUSE	

【図16】



【図17】

スコア閾値	感度	選択性
-0.6	100%	72%
-0.55	100%	75%
-0.5	99%	78%
-0.45	99%	81%
-0.4	99%	84%
-0.35	99%	88%
-0.3	99%	89%
-0.25	98%	92%
-0.2	97%	93%
-0.15	97%	95%
-0.1	96%	96%
-0.05	94%	97%
0	90%	98%
0.05	87%	98%
0.1	84%	99%
0.15	79%	99%
0.2	72%	100%
0.25	68%	100%

【図18】

	GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	96.5	97.3	0.969
4-fold	86.8	96.4	0.915
	非GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	98.6	98.1	0.983
4-fold	98.2	93.1	0.956

【図20】

N末端疎水性特性抽出必要数=13

	GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	95.7	91.7	0.936
4-fold	87.3	86.4	0.868
	非GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	90.4	95.0	0.926
4-fold	84.5	85.9	0.851

【図19】

N末端疎水性特性抽出必要数=13

	GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	89.9	97.2	0.935
4-fold	88.0	83.8	0.858
	非GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	96.9	89.1	0.929
4-fold	79.4	85.0	0.820

【図21】

側鎖サイズ特性抽出必要数=5

	GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	92.9	98.1	0.955
4-fold	81.4	83.6	0.823
	非GPIアンカー型タンパク質		
	感度 (%)	選択性 (%)	成功率
Self-consistency	99.1	96.3	0.977
4-fold	90.4	90.2	0.902

## 【 図 2 2 】

基準位置を含む所定の範囲:-12~+9

GPIアンカー型タンパク質			
	感度 (%)	選択性 (%)	成功率
Self-consistency	96.5	95.6	0.960
4-fold	80.1	92.7	0.860

非GPIアンカー型タンパク質			
	感度 (%)	選択性 (%)	成功率
Self-consistency	97.6	98.1	0.979
4-fold	96.4	89.9	0.931

---

フロントページの続き

- (72)発明者 大浦 修  
神奈川県川崎市多摩区東三田 1 - 1 - 1 学校法人明治大学 生田校舎内
- (72)発明者 佐々木 貴規  
神奈川県川崎市多摩区東三田 1 - 1 - 1 学校法人明治大学 生田校舎内
- (72)発明者 池田 修己  
神奈川県川崎市多摩区東三田 1 - 1 - 1 学校法人明治大学 生田校舎内

審査官 齋藤 正貴

- (56)参考文献 特開2004 - 125623 (JP, A)  
特開2002 - 215634 (JP, A)  
特開2003 - 014734 (JP, A)

- (58)調査した分野(Int.Cl., DB名)  
G06F 19/00  
JSTPlus / JMEDPlus / JST7580 (JDreamII)  
PubMed