(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年11月24日(24.11.2022)

WIPO | PCT

(10) 国際公開番号
WO 2022/244089 A1

(51) 国際特許分類:
G16B 30/10 (2019.01)

(21) 国際出願番号： PCT/JP2021/018730

(22) 国際出願日： 2021年5月18日(18.05.2021)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(71) 出願人: 富士通株式会社 (**FUJITSU LIMITED**) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 Kanagawa (JP).

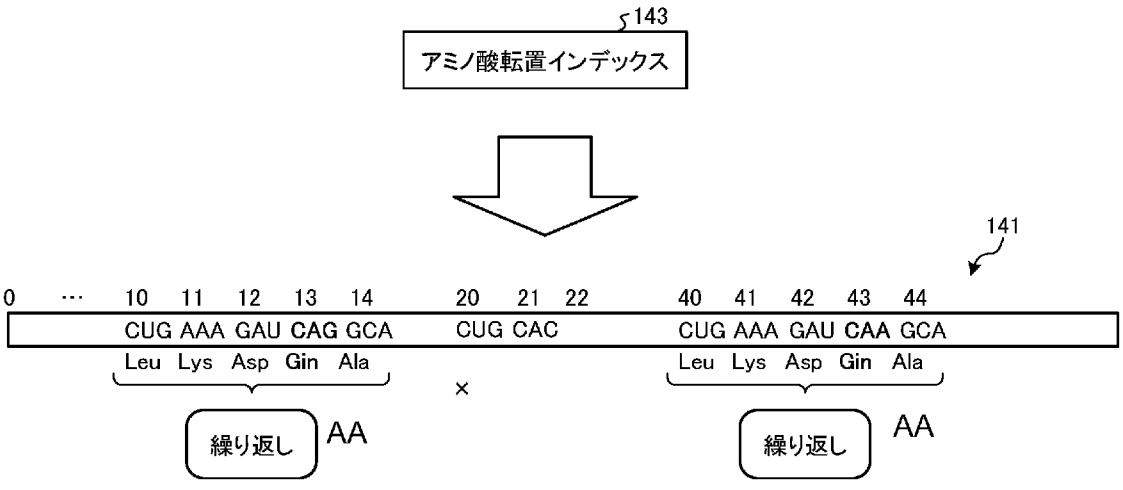
(72) 発明者: 片岡 正弘 (**KATAOKA, Masahiro**); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP). 永浦 良平 (**NAGAURA, Ryohei**); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP). 茂櫛 薫 (**MOGUSHI, Kaoru**); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP).

(74) 代理人: 弁理士法人酒井国際特許事務所 (**SAKAI INTERNATIONAL PATENT OFFICE**); 〒1000013 東京都千代田区霞が関３丁目８番１号 虎の門三井ビルディング Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

(54) **Title:** INFORMATION PROCESSING PROGRAM, INFORMATION PROCESSING METHOD, AND INFORMATION PROCESSING DEVICE

(54) 発明の名称： 情報処理プログラム、情報処理方法および情報処理装置



143 Amino acid transposition index
AA Repeat

(57) **Abstract:** The information processing device calculates a second index that indicates the positions of a single amino acid on a codon file on the basis of a first index that indicates the positions of a plurality of codons indicating the single amino acid and having different nucleotide sequences from each other on the codon file. The information processing device identifies each of the positions of an amino acid sequence appearing repeatedly on the codon file on the basis of the second index. The information processing device identifies codon sequences corresponding to the positions of the amino acid sequence appearing repeatedly on the codon file as codon sequences having homology with each other.

WO 2022/244089 A1

［続葉有］

HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一　国際調査報告（条約第21条(3)）

(57) 要約：情報処理装置は、同一のアミノ酸を示す塩基配列の異なる複数のコドンに関し、コドンファイル上の複数のコドンの位置を示す第１インデックスを基にして、コドンファイル上のアミノ酸の位置を示す第２インデックスを演算する。情報処理装置は、第２インデックスを基にして、コドンファイルに繰り返し発現するアミノ酸配列の位置をそれぞれ特定する。情報処理装置は、コドンファイルに繰り返し発現する各アミノ酸配列の位置に対応する各コドン配列を、それぞれが相同性を有するコドン配列として特定する。

明　細　書

発明の名称 ：

情報処理プログラム、情報処理方法および情報処理装置

技術分野

[0001]　本発明は、情報処理プログラム等に関する。

背景技術

[0002]　ヒトゲノムの塩基配列の研究が進められており、ヒトゲノムを構成するタンパク質の種類が３万種類であることが解明されている。一方で、微生物等のタンパク質の種類は無限であると考えられており、標的となる塩基配列から、繰り返し発現される特異なコドン配列が多数見つかっている。たとえば、繰り返し発現される特異なコドン配列は、ドメインやモチーフ等と呼ばれ、かかる特異なコドン配列を究明することが重要である。

[0003]　ドメインは、タンパク質の配列、構造の一部分で他の部分とは独立に進化し、機能を持った存在である。モチーフは、対称性を持ったコドン配列が特徴である。図１７は、モチーフの一例を示す図である。図１７に示すように、モチーフには、βヘアピン１ａ、グリークキー１ｂ、βバレル１ｃ（ポリンやリポカリン）等がある。フォールドは、タンパク質鎖がその本来の三次元構造、通常は生物学的に機能するコンホメーション(立体構造)を、迅速かつ再現性のある方法で獲得する物理的なプロセスである。

[0004]　たとえば、塩基配列から、モチーフを検索する技術として、ハミング距離を持つ置換塩基配列をキーとしてモチーフを検索する従来技術がある。また、ＤＮＡ（deoxyribonucleic　acid）配列の転写開始点の上流からオーソログ候補の複数の配列断面を抽出し、モチーフ候補を判定する従来技術がある。

先行技術文献

特許文献

[0005] 特許文献1：国際公開第２００５／０９６２０８号

特許文献2：国際公開第２０２０／０４９７４８号

特許文献3：特開２０１４－１１２３０７号公報

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、上述した従来技術では、繰り返し発現するコドン配列の探索を効率よく行うことができないという問題がある。

[0007] ここで、ＤＮＡおよびＲＮＡ（ribonucleic acid）の塩基は４種類で、「Ａ」、「Ｇ」、「Ｃ」、「Ｔ」または「Ｕ」の記号で示される。また、３つの塩基配列がひとかたまりで、２０種類のアミノ酸を決定する。それぞれのアミノ酸は、「Ａ」～「Ｙ」の記号で示される。図１８は、アミノ酸と塩基、およびコドンとの関係を示す図である。３つの塩基配列のかたまりは「コドン」と呼ばれる。各塩基の並びで、コドンが決定され、コドンが決定されるとアミノ酸が決定される。

[0008] 図１８に示すように、一つのアミノ酸に対して、複数種類のコドンが対応付けられる。たとえば、アミノ酸「アラニン（Ala）」は、コドン「ＧＣＵ」、「ＧＣＣ」、「ＧＣＡ」、「ＧＣＧ」に対応付けられ、コドン「ＧＣＵ」、「ＧＣＣ」、「ＧＣＡ」、「ＧＣＧ」は実質的に同一のコドンである。しかしながら、従来技術では、かかるコドンの特性に対応できておらず、繰り返し発現するコドン配列の探索を効率よく行うことができていない。

[0009] １つの側面では、繰り返し発現されるコドン配列の探索を効率よく行うことができる情報処理プログラム、情報処理方法および情報処理装置を提供することを目的とする。

課題を解決するための手段

[0010] 第１の案では、コンピュータに次の処理を実行させる。コンピュータは、同一のアミノ酸を示す塩基配列の異なる複数のコドンに関し、コドンファイル上の複数のコドンの位置を示す第１インデックスを基にして、コドンファイル上のアミノ酸の位置を示す第２インデックスを演算する。コンピュータは、第２インデックスを基にして、コドンファイルに繰り返し発現するアミ

ノ酸配列の位置をそれぞれ特定する。コンピュータは、コドンファイルに繰り返し発現する各アミノ酸配列の位置に対応する各コドン配列を、それぞれが相同性を有するコドン配列として特定する。

発明の効果

[0011] 繰り返し発現するコドン配列の探索を効率良く行うことができる。

図面の簡単な説明

[0012] [図1]図１は、本実施例１に係る情報処理装置の処理を説明するための図（１）である。

[図2]図２は、本実施例１に係る情報処理装置の処理を説明するための図（２）である。

[図3]図３は、本実施例１に係る情報処理装置の構成を示す機能ブロック図である。

[図4]図４は、スコアテーブルのデータ構造の一例を示す図である。

[図5]図５は、コドンファイルのデータ構造の一例を示す図である。

[図6]図６は、コドン転置インデックスのデータ構造の一例を示す図である。

[図7]図７は、アミノ酸転置インデックスのデータ構造の一例を示す図である。

[図8]図８は、特定部の処理を説明するための図（１）である。

[図9]図９は、特定部の処理を説明するための図（２）である。

[図10]図１０は、特定部の処理を説明するための図（３）である。

[図11]図１１は、特定部の処理を説明するための図（４）である。

[図12]図１２は、探索結果情報のデータ構造の一例を示す図である。

[図13]図１３は、本実施例１に係る情報処理装置の処理手順を示すフローチャートである。

[図14]図１４は、本実施例２に係る情報処理装置の処理を説明するための図（１）である。

[図15]図１５は、本実施例２に係る情報処理装置の処理を説明するための図（２）である。

[図16]図１６は、実施例の情報処理装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

[図17]図１７は、モチーフの一例を示す図である。

[図18]図１８は、アミノ酸と塩基、およびコドンとの関係を示す図である。

発明を実施するための形態

[0013] 以下に、本願の開示する情報処理プログラム、情報処理方法および情報処理装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。

実施例 １

[0014] 本実施例１に係る情報処理装置の処理の一例について説明する。図１及び図２は、本実施例１に係る情報処理装置の処理を説明するための図である。

[0015] 図１について説明する。情報処理装置は、塩基配列の情報を含むコドンファイル１４１をコドン単位で走査して、コドン転置インデックス１４２を生成する。

[0016] コドン転置インデックス１４２は、コドンの種類毎のビットマップを有する。コドンの種類は６４種類であるため、コドン転置インデックス１４２には、６４個のビットマップが登録される。コドン転置インデックス１４２の各ビットマップは、コドンの種類と、オフセットと、フラグとが対応付けられる。ビットマップのフラグ「１」が立ったオフセットにおいて、対応する種類のコドンが位置することを示す。ビットマップにおいて、フラグの立っていないオフセットには「０」が対応付けられる。

[0017] たとえば、コドン「ＧＣＵ」に対応するビットマップについて、オフセット「ｎ」にフラグ「１」が対応付けられる場合には、コドンファイル１４１の先頭からｎ＋１番目のコドンが、コドン「ＧＣＵ」であることを示す。なお、本実施例１では、コドンファイル１４１の先頭のコドンのオフセットを「０」とする。

[0018] 情報処理装置は、コドン転置インデックス１４２と、定義テーブルＴ１とを基にして、アミノ酸転置インデックス１４３を生成する。定義テーブルＴ

１は、アミノ酸と、コドンとの対応関係を定義するテーブルである。図１８で説明したように、同一のアミノ酸に対して、複数種類のコドンが対応付けられる場合がある。

[0019] アミノ酸転置インデックス１４３には、各アミノ酸に対応するビットマップが登録される。アミノ酸転置インデックス１４３の各ビットマップは、アミノ酸の種類と、オフセットと、フラグとが対応付けられる。ビットマップのフラグ「１」が立ったオフセットにおいて、対応する種類のアミノ酸が位置することを示す。ビットマップにおいて、フラグの立っていないオフセットには「０」が対応付けられる。

[0020] 情報処理装置が、アミノ酸転置インデックス１４３の各アミノ酸のビットマップのうち、アミノ酸「ＡＩａ」のビットマップを生成する場合について説明する。情報処理装置１００は、定義テーブルＴ１を基にして、アミノ酸「ＡＩａ」に対応するコドンとして「ＧＣＵ」、「ＧＣＣ」、「ＧＣＡ」、「ＧＣＧ」を特定する。

[0021] 情報処理装置は、コドン転置インデックス１４２から、コドン「ＧＣＵ」のビットマップ１４２－１、コドン「ＧＣＣ」のビットマップ１４２－２、コドン「ＧＣＡ」のビットマップ１４２－３、コドン「ＧＣＧ」のビットマップ１４２－４を取得する。情報処理装置は、ビットマップ１４２－１～１４２－４に対してＯＲ演算（論理和）を実行することで、アミノ酸「ＡＩａ」のビットマップ１４３－１を生成する。

[0022] すなわち、情報処理装置は、ビットマップ１４２－１～１４２－１のオフセット「ｎ」のいずれかのフラグが「１」となる場合に、ビットマップ１４３－１のオフセット「ｎ」のフラグを「１」に設定する。一方、情報処理装置は、ビットマップ１４２－１～１４２－１のオフセット「ｎ」のすべてに「０」が設定されている場合、ビットマップ１４３－１のオフセット「ｎ」に「０」を設定する。情報処理装置は、各オフセットにおいて、上記処理を繰り返し実行する。

[0023] 情報処理装置は、他のアミノ酸のビットマップについても、アミノ酸「Ａ

Ｉａ」のビットマップ１４３－１と同様にして生成し、各アミノ酸のビットマップを、アミノ酸転置インデックス１４３に登録する。

[0024] 図２の説明に移行する。情報処理装置は、アミノ酸転置インデックス１４３を基にして、コドンファイル１４１のオフセットと、アミノ酸の種類との関係を特定し、繰り返し発現するアミノ酸配列の位置に対応するコドン配列を、それぞれ相同性を有するコドン配列として特定する。

[0025] たとえば、図２に示す例では、アミノ酸配列「Leu、Lys、Asp、Gln、Ala」が、コドンファイル１４１のオフセット１０～１４，４０～４４等に繰り返し発現している。この場合、情報処理装置は、オフセット１０～１４に含まれるコドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＧ、ＧＣＡ」と、オフセット４０～４４に含まれるコドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＡ、ＧＣＡ」を、相同性を有するコドン配列として特定する。

[0026] コドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＧ、ＧＣＡ」と、コドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＡ、ＧＣＡ」とを比較すると、コドンの粒度では、「ＣＡＧ」と「ＣＡＡ」とが異なる。しかし、「ＣＡＧ」と「ＣＡＡ」とは、同一のアミノ酸「Gln」に対応するため、コドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＧ、ＧＣＡ」と、コドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＡ、ＧＣＡ」とは相同性のコドン配列といえる。

[0027] 上記のように、本実施例１に係る情報処理装置によれば、同一のアミノ酸を示す塩基配列の異なるコドンのビットマップから、アミノ酸単位のビットマップを生成することで、アミノ酸転置インデックス１４３を生成する。情報処理装置は、生成したアミノ酸転置インデックス１４３を用いて、コドンファイル１４１上のアミノ酸の種類との関係を特定し、繰り返し発現するアミノ酸配列の位置に対応するコドン配列を、それぞれ相同性を有するコドン配列として特定する。これによって、繰り返し発現されるコドン配列の探索を効率良く行うことができる。

[0028] 次に、本実施例１に係る情報処理装置の構成の一例について説明する。図３は、本実施例１に係る情報処理装置の構成を示す機能ブロック図である。

図３に示すように、この情報処理装置１００は、通信部１１０、入力部１２０、表示部１３０、記憶部１４０、制御部１５０を有する。

[0029] 通信部１１０は、有線又は無線で外部装置等に接続され、外部装置等との間で情報の送受信を行う。たとえば、通信部１１０は、ＮＩＣ（Network Interface Card）等によって実現される。通信部１１０は、図示しないネットワークに接続されていてもよい。

[0030] 入力部１２０は、各種の情報を、情報処理装置１００に入力する入力装置である。入力部１２０は、キーボードやマウス、タッチパネル等に対応する。

[0031] 表示部１３０は、制御部１５０から出力される情報を表示する表示装置である。表示部１３０は、液晶ディスプレイ、有機ＥＬ（Electro Luminescence）ディスプレイ、タッチパネル等に対応する。

[0032] 記憶部１４０は、定義テーブルＴ１、スコアテーブルＴ２、コドンファイル１４１、コドン転置インデックス１４２、アミノ酸転置インデックス１４３、探索結果情報１４４を有する。記憶部１４０は、たとえば、ＲＡＭ（Random Access Memory)、フラッシュメモリ（Flash Memory）等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。

[0033] 定義テーブルＴ１は、アミノ酸と、コドンとの対応関係を定義するテーブルである。定義テーブルＴ１で定義されるアミノ酸と、コドンとの関係は、図１８で説明したアミノ酸と塩基、およびコドンとの関係と同様である。

[0034] スコアテーブルＴ２は、アミノ酸同士の類似性の程度を定義するテーブルである。図４は、スコアテーブルのデータ構造の一例を示す図である。図４に示すスコアテーブルＴ２の領域Ａ１、Ａ２に示す符号は、図１８で説明したアミノ酸を一意に示す符号である。領域Ａ３の数値は、アミノ酸置換確率を示すスコアであり、スコアが大きいほど、より類似性の程度が大きいことを示す。

[0035] たとえば、図４のスコアテーブルＴ２によれば、アラニン「Ａ（Ala）」と

トレオニン「Ｔ（Thr）」とのスコアが「－４」である。また、アラニン「Ａ（Ala）」とトリプトファン「Ｗ（Trp）」とのスコアが「１」である。このため、アラニンとトレオニンとの組よりも、アラニンとトリプトファンとの組の方が類似性の程度が大きいことを示す。

[0036] コドンファイル１４１は、複数の塩基が並んだ塩基配列の情報を有する。図５は、コドンファイルのデータ構造の一例を示す図である。図５に示すように、コドンファイル１４１は、複数の塩基の記号が並んだ情報である。連続する３つの塩基の組が、一つのコドンに対応する。

[0037] コドン転置インデックス１４２は、コドンファイル１４１の先頭からのオフセットと、コドンの種類とを対応付ける情報である。図６は、コドン転置インデックスのデータ構造の一例を示す図である。コドン転置インデックス１４２の横軸は、オフセットに対応する軸である。コドン転置インデックス１４２の縦軸は、コドンの種類に対応する軸である。

[0038] たとえば、コドンファイル１４１の先頭のコドンのオフセットを「０」とする。コドンファイル１４１の先頭から７番目の位置に、コドン「ＡＵＧ」が含まれる場合、コドン転置インデックス１４２のオフセット「６」の列と、コドン「ＡＵＧ」の行とが交差する位置のビットが「１」となる。

[0039] アミノ酸転置インデックス１４３は、コドンファイル１４１の先頭からのオフセットと、アミノ酸の種類とを対応付ける情報である。図７は、アミノ酸転置インデックスのデータ構造の一例を示す図である。アミノ酸転置インデックス１４３の横軸は、オフセットに対応する軸である。アミノ酸転置インデックス１４３の縦軸は、アミノ酸の種類に対応する軸である。

[0040] たとえば、コドンファイル１４１の先頭のコドン（いずれかのアミノ酸に対応するコドン）のオフセットを「０」とする。コドンファイル１４１の先頭から７番目の位置に、アミノ酸「Ala」に対応するコドン「ＧＣＵ」、「ＧＣＣ」、「ＧＣＡ」、「ＧＣＧ」のいずれかが含まれる場合、アミノ酸転置インデックス１４３のオフセット「６」の列と、アミノ酸「Ala」の行とが交差する位置のビットが「１」となる。

[0041] 探索結果情報１４４は、コドンファイル１４１において繰り返し発現するアミノ酸配列（コドン配列）の情報を有する。たとえば、探索結果情報１４４は、繰り返し発現するアミノ酸配列と、かかるアミノ酸配列の位置との情報を対応付けて保持する。

[0042] 図３の説明に戻る。制御部１５０は、前処理部１５１と、特定部１５２とを有する。制御部１５０は、例えば、ＣＰＵ（Central Processing Unit）やＭＰＵ(Micro Processing Unit)により実現される。また、制御部１５０は、例えばＡＳＩＣ（Application Specific Integrated Circuit）やＦＰＧＡ（Field Programmable Gate Array）等の集積回路により実行されてもよい。

[0043] 前処理部１５１は、コドンファイル１４１および定義テーブルＴ１を基にして、コドン転置インデックス１４２、アミノ酸転置インデックス１４３を生成する。

[0044] 前処理部１５１が、コドン転置インデックス１４２を生成する処理の一例について説明する。前処理部１５１は、定義テーブルＴ１に含まれるコドンの種類から、ターゲットとなるコドンの種類を選択する。前処理部１５１は、コドンファイル１４１の先頭からコドンの粒度（３つの塩基配列をひとかたまりとする粒度）で走査して、選択したコドンの種類が出現するオフセットにフラグ「１」を設定する処理を繰り返し実行し、選択したコドンの種類に対応するビットマップを生成する。

[0045] 前処理部１５１は、他のコドンの種類についても同様にして、ビットマップをそれぞれ生成する。前処理部１５１は、各コドンの種類に対応するビットマップを、コドン転置インデックス１４２に設定することで、コドン転置インデックス１４２を生成する。

[0046] 続いて、前処理部１５１が、アミノ酸転置インデックス１４３を生成する処理の一例について説明する。前処理部１５１は、同一のアミノ酸に対応するコドンの種類を特定し、特定したコドンの種類に対応するビットマップを、コドン転置インデックス１４２から取得する。前処理部１５１は、取得し

た各コドンの種類のビットマップに対して、ＯＲ演算を実行することで、アミノ酸のビットマップを生成する。

[0047] たとえば、前処理部１５１が、アミノ酸転置インデックス１４３の各アミノ酸のビットマップのうち、アミノ酸「Ａｌａ」のビットマップを生成する場合について説明する。図１で説明したように、前処理部１５１は、定義テーブルＴ１を基にして、アミノ酸「Ａｌａ」に対応するコドンとして「ＧＣＵ」、「ＧＣＣ」、「ＧＣＡ」、「ＧＣＧ」を特定する。

[0048] 前処理部１５１は、コドン転置インデックス１４２から、コドン「ＧＣＵ」のビットマップ１４２－１、コドン「ＧＣＣ」のビットマップ１４２－２、コドン「ＧＣＡ」のビットマップ１４２－３、コドン「ＧＣＧ」のビットマップ１４２－４を取得する。前処理部１５１は、ビットマップ１４２－１～１４２－４に対してＯＲ演算（論理和）を実行することで、アミノ酸「Ａｌａ」のビットマップ１４３－１を生成する。

[0049] 前処理部１５１は、他のアミノ酸のビットマップについても、アミノ酸「Ａｌａ」のビットマップ１４３－１と同様にして生成し、各アミノ酸のビットマップを、アミノ酸転置インデックス１４３に設定することで、アミノ酸転置インデックス１４３を生成する。

[0050] 次に、特定部１５２の処理について説明する。特定部１５２は、アミノ酸転置インデックス１４３を基にして、コドンファイル１４１に繰り返し発現するアミノ酸配列の位置（オフセット）をそれぞれ特定する。特定部１５２は、コドンファイル１４１に繰り返し発現するアミノ酸配列の位置（オフセット）に対応する各コドン配列を、それぞれ相同性を有するコドン配列として特定する。

[0051] 特定部１５２は、アミノ酸転置インデックス１４３を基にして、アミノ酸配列の最長一致探索を実行し、最長一致するアミノ酸配列を特定する。特定部１５２は、最長一致するアミノ酸配列の発現回数が、予め設定される発現回数以上である場合に、かかるアミノ酸配列を、「アミノ酸配列候補」として検索する。

[0052] たとえば、図２で説明したように、アミノ酸配列「Leu、Lys、Asp、Gln、Ala」が、コドンファイル１４１のオフセット１０～１４，４０～４４等に繰り返し発現し、発現回数が所定の発現回数以上とする。この場合、特定部１５２は、オフセット１０～１４に含まれるコドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＧ、ＧＣＡ」と、オフセット４０～４４に含まれるコドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ、ＣＡＡ、ＧＣＡ」を、相同性を有するコドン配列として特定する。特定部１５２は、特定した相同性を有するコドン配列の情報を、探索結果情報１４４に登録する。

[0053] ここで、特定部１５２がアミノ酸転置インデックス１４３を基にして、連続するアミノ酸配列を特定する処理の一例について説明する。図８は、特定部の処理を説明するための図（１）である。図８では一例として、アミノ酸配列「Leu、Lys、Asp、Gln」が、コドンファイル１４１に含まれているか否かを特定する場合について説明する。

[0054] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Leu」のビットマップ５０を取得する。ビットマップ５０には、オフセット「１０」、「２０」にフラグ「１」が設定されている。特定部１５２は、ビットマップ５０の左シフトを実行することで、ビットマップ５０ｓを生成する。ビットマップ５０ｓには、オフセット「１１」、「２１」にフラグ「１」が設定されている。

[0055] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Lys」のビットマップ５１を取得する。ビットマップ５１には、オフセット「１１」にフラグ「１」が設定されている。特定部１５２は、ビットマップ５０ｓと、ビットマップ５１とのＡＮＤ演算を実行することで、ビットマップ５２を生成する。

[0056] 図８に示す例では、ビットマップ５２のオフセット「１１」にフラグ「１」が設定されているため、コドンファイル１４１のオフセット「１０～１１」において、「Leu、Lys」が連続して存在していることが特定される。

[0057] 特定部１５２は、ビットマップ５２の左シフトを実行することで、ビット

マップ５２ｓを生成する。ビットマップ５２ｓでは、オフセット「１２」にフラグ「１」が設定される。

[0058] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Asp」のビットマップ５３を取得する。ビットマップ５３には、オフセット「１２」にフラグ「１」が設定されている。特定部１５２は、ビットマップ５２ｓと、ビットマップ５３とのＡＮＤ演算を実行することで、ビットマップ５４を生成する。

[0059] 図８に示す例では、ビットマップ５４のオフセット「１２」にフラグ「１」が設定されているため、コドンファイル１４１のオフセット「１０～１２」において、「Leu、Lys、Asp」が連続して存在していることが特定される。

[0060] 特定部１５２は、ビットマップ５４の左シフトを実行することで、ビットマップ５４ｓを生成する。ビットマップ５４ｓでは、オフセット「１３」にフラグ「１」が設定される。

[0061] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Gln」のビットマップ５５を取得する。ビットマップ５５には、オフセット「１３」にフラグ「１」が設定されている。特定部１５２は、ビットマップ５４ｓと、ビットマップ５５とのＡＮＤ演算を実行することで、ビットマップ５６を生成する。

[0062] 図８に示す例では、ビットマップ５６のオフセット「１３」にフラグ「１」が設定されているため、コドンファイル１４１のオフセット「１０～１３」において、「Leu、Lys、Asp、Gln」が連続して存在していることが特定される。

[0063] 特定部１５２は、各アミノ酸配列について、上記処理を繰り返し実行することで、最長一致のアミノ酸配列を特定し、繰り返し発現するアミノ酸配列を特定する。特定部１５２は、他の技術を用いて、繰り返し発現するアミノ酸配列を特定してもよい。

[0064] 特定部１５２は、上記の処理によって、アミノ酸配列候補を検索した後に、スコアテーブルＴ２を用いて、各アミノ酸配列候補の相同性を評価する。

図９は、特定部の処理を説明するための図（２）である。ここでは、アミノ酸配列候補６０ａ，６０ｂを用いて説明する。アミノ酸配列候補６０ａ，６０ｂは、「Leu、Lys、Asp、Gln、Ala」となる。また、図１８の表（定義テーブルＴ１相当）を基にして、「Leu、Lys、Asp、Gln、Ala」を記号に変換すると「Ｌ（Leu）、Ｋ（Lys）、Ｄ（Asp）、Ｑ（Gln）、Ａ（Ala）」となる。

[0065] 特定部１５２は、スコアテーブルＴ２を基にして、各アミノ酸のスコアを特定し、累計することで、相同性のスコアを算出する。Ｌ（Leu）同士のスコアは、スコアテーブルＴ２に存在しないため「０」となる。Ｋ（Lys）同士のスコアは、スコアテーブルＴ２に基づき「－１」となる。Ｄ（Asp）同士のスコアは、スコアテーブルＴ２に基づき「－１」となる。Ｑ（Gln）同士のスコアは、スコアテーブルＴ２に存在しないため「０」となる。Ａ（Ala）同士のスコアは、スコアテーブルＴ２に基づき「５」となる。このため、特定部１５２は、アミノ酸配列候補６０ａ，６０ｂのスコアを累計値「３」を算出する。

[0066] 特定部１５２は、アミノ酸配列候補のスコアの累計値が閾値以上である場合に、アミノ酸配列候補を、相同性の関係となるアミノ酸配列として特定する。特定部１５２は、特定した結果を、探索結果情報１４４に登録する。閾値は、管理者によって予め設定される。

[0067] ところで、特定部１５２は、相同性の関係となるアミノ酸配列を特定した後に、特定したアミノ酸配列と対称的に発現するアミノ酸配列を更に特定してもよい。図１０は、特定部の処理を説明するための図（３）である。たとえば、特定部１５２は、上記の処理で特定した、アミノ酸配列「Leu、Lys、Asp、Gln、Ala」と、対称的に発現する「Ala、Gln、Asp、Lys、Leu」を、アミノ酸転置インデックス１４３を基にして特定する。図１０に示す例では、特定部１５２は、コドンファイル１４１のオフセット「３０～３４」に存在するアミノ酸配列「Ala、Gln、Asp、Lys、Leu」を特定している。

[0068] ここで、特定部１５２がアミノ酸転置インデックス１４３を基にして、対称性のアミノ酸配列を特定する処理の一例について説明する。図１１は、特

定部の処理を説明するための図（４）である。図１１では一例として、対称的なアミノ酸配列「Ala、Gln、Asp（Lys、Leuについては省略）」が、コドンファイル１４１に含まれているか否かを特定する場合について説明する。

[0069] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Ala」のビットマップ６０を取得する。ビットマップ６０には、オフセット「２４」にフラグ「１」が設定されている。特定部１５２は、ビットマップ６０の右シフトを実行することで、ビットマップ６０ｓを生成する。ビットマップ６０ｓでは、オフセット「２３」にフラグ「１」が設定されている。

[0070] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Gln」のビットマップ６１を取得する。ビットマップ６１には、オフセット「２３」にフラグ「１」が設定されている。特定部１５２は、ビットマップ６０ｓと、ビットマップ６１とのＡＮＤ演算を実行することで、ビットマップ６２を生成する。

[0071] 図１１に示す例では、ビットマップ６２のオフセット「２３」にフラグ「１」が設定されているため、コドンファイル１４１のオフセット「２３～２４」において、「Ala、Gln」が連続して存在していることが特定される。

[0072] 特定部１５２は、ビットマップ６２の右シフトを実行することで、ビットマップ６２ｓを生成する。ビットマップ６２ｓでは、オフセット「２２」にフラグ「１」が設定される。

[0073] 特定部１５２は、アミノ酸転置インデックス１４３から、アミノ酸「Asp」のビットマップ６３を取得する。ビットマップ６３は、オフセット「２２」にフラグ「１」が設定される。特定部１５２は、ビットマップ６２ｓと、ビットマップ６３とのＡＮＤ演算を実行することで、ビットマップ６４を生成する。

[0074] 図１１に示す例では、ビットマップ６４のオフセット「２２」にフラグ「１」が設定されているため、コドンファイル１４１のオフセット「２２～２４」において、「Ala、Gln、Asp」が連続して存在していることが特定される。

[0075] 特定部１５２は、上記処理を実行することで、対称的なアミノ酸配列を特定する。特定部１５２は、特定した結果を、探索結果情報１４４に登録する。特定部１５２は、探索結果情報１４４を、表示部１３０に出力して表示させてもよいし、通信部１１０を介して、外部装置に送信してもよい。

[0076] 図１２は、探索結果情報のデータ構造の一例を示す図である。図１２に示すように、この探索結果情報１４４は、アミノ酸配列、第１オフセット、第２オフセット、累計スコアを対応付ける。アミノ酸配列は、特定部１５２によって特定される相同性のアミノ酸配列である。第１オフセットは、相同性のアミノ酸配列に対応するコドン配列の存在するコドンファイル１４１のオフセットを示す。第２オフセットは、対称性のアミノ酸配列に対応するコドン配列の存在するコドンファイル１４１のオフセットを示す。累計スコアは、図９で説明したスコアの累計値である。

[0077] 図１２では、アミノ酸配列「Leu、Lys、Asp、Gln、Ala」に対応する第１オフセットが、「１０～１４」、「４０～４４」となっている。このため、コドンファイル１４１のオフセット「１０～１４」に対応するコドン配列、オフセット「４０～４４」に対応するコドン配列が、相同性を有するコドン配列となる。

[0078] また、アミノ酸配列「Leu、Lys、Asp、Gln、Ala」に対して対称的なアミノ酸配列「Ala、Gln、Asp、Lys、Leu」の第２オフセットが「３０～３４」となっている。このため、コドンファイル１４１のオフセット「３０～３４」に対応するコドン配列が、対称的なコドン配列となる。

[0079] たとえば、探索結果情報の相同性のアミノ酸配列と、このアミノ酸配列と対称的なアミノ酸配列との間の部分は、モチーフに相当する部分といえる。すなわち、第１オフセット「１０～１４」と、第２オフセット「３０～３４」との間の部分は、モチーフ部分に相当する。

[0080] 次に、本実施例１に係る情報処理装置１００の処理手順の一例について説明する。図１３は、本実施例１に係る情報処理装置の処理手順を示すフローチャートである。図１３に示すように、情報処理装置１００の前処理部１５

１は、コドンファイル１４１と定義テーブルＴ１とを基にして、コドン転置インデックス１４２を生成する（ステップＳ１０１）。

[0081] 前処理部１５１は、定義テーブルＴ１を基にして、同一のアミノ酸に対応する複数のコドンを特定する（ステップＳ１０２）。前処理部１５１は、特定した複数のコドンのビットマップに対してＯＲ演算を実行し、アミノ酸のビットマップを生成して、アミノ酸転置インデックス１４３を生成する（ステップＳ１０３）。

[0082] 情報処理装置１００の特定部１５２は、アミノ酸転置インデックス１４３を基にして、繰り返し発現するアミノ酸配列候補を特定する（ステップＳ１０４）。特定部１５２は、スコアテーブルＴ２を基にして、アミノ酸配列候補のスコアの累積値を算出する（ステップＳ１０５）。

[0083] 特定部１５２は、スコアの累積値を基にして、相同性のアミノ酸配列（相同性を有するコドン配列）を特定する（ステップＳ１０６）。特定部１５２は、相同性のアミノ酸配列と対称的なアミノ酸配列を特定する（ステップＳ１０７）。

[0084] 特定部１５２は、特定した結果を、探索結果情報１４４に登録する（ステップＳ１０８）。特定部１５２は、探索結果情報１４４を出力する（ステップＳ１０９）。

[0085] 次に、本実施例１に係る情報処理装置１００の効果について説明する。情報処理装置１００は、同一のアミノ酸を示す塩基配列の異なるコドンのビットマップから、アミノ酸単位のビットマップを生成することで、アミノ酸転置インデックス１４３を生成する。情報処理装置１００は、生成したアミノ酸転置インデックス１４３を用いて、コドンファイル１４１のアミノ酸の種類との関係を特定し、繰り返し発現するアミノ酸配列の位置に対応するコドン配列を、それぞれ相同性を有するコドン配列として特定する。これによって、繰り返し発現されるコドン配列の探索を効率良く行うことができる。

[0086] 情報処理装置１００は、コドンファイル１４１に繰り返し発現するアミノ酸配列が、相同性を有するアミノ酸である否かを、アミノ酸同士の相同性の

程度を定義したスコアテーブルＴ２を基にして評価する。これによって、単なるアミノ酸の一致だけでなく、アミノ酸配列間の相同性の程度を評価することができる。

[0087] 情報処理装置１００は、複数のコドンに対応するコドン転置インデックス１４２のビットマップの論理和を実行することで、複数のコドンに対応するある一つのアミノ酸のビットマップを演算する。これによって、容易に、複数のコドンに対応するアミノ酸のビットマップを生成して、アミノ酸転置インデックス１４３を生成することができる。

実施例 2

[0088] 実施例１では、アミノ酸の粒度で、相同性を有するアミノ酸配列を特定し、特定したアミノ酸配列のオフセットを基にして、相同性のコドン配列を特定していたが、コドンの粒度で、相同性のコドン配列を特定してもよい。実施例２では、コドンの粒度で、相同性のコドン配列を特定する処理について説明する。

[0089] 図１４は、本実施例２に係る情報処理装置の処理を説明するための図（１）である。情報処理装置は、コドン転置インデックス１４２を基にして、コドンファイル１４１のオフセットと、コドンの種類を特定し、繰り返し発現するコドン配列を特定する。コドン転置インデックス１４２の説明は、実施例１で説明したコドン転置インデックス１４２の説明と同様である。

[0090] たとえば、図１４に示す例では、コドン配列「ＣＵＧ、ＡＡＡ、ＧＡＵ」がコドンファイル１４１のオフセット１０～１２，３０～３２、４０～４２等に繰り返し発現している。この場合、情報処理装置は、オフセット１０～１２，３０～３２、４０～４２のコドン配列を、相同性を有するコドン配列として特定する。情報処理装置は、相同性を有するコドン配列を特定した後に、実施例１で説明したように、アミノ酸の粒度で、相同性を有するアミノ酸配列を特定してもよい。

[0091] 図１５は、本実施例２に係る情報処理装置の処理を説明するための図（２）である。情報処理装置は、相同性のコドン配列を特定した後に、コドンの

粒度で、対称的なコドン配列を特定してもよい。たとえば、相同性を有するコドン配列を「ＣＵＧ、ＡＡＡ、ＧＡＵ」とすると、情報処理装置は、対称的なコドン配列「ＧＡＵ、ＡＡＡ、ＣＵＧ」を、コドンファイル１４１から特定する。図２に示す例では、情報処理装置は、オフセット２３～２５において、対称的なコドン配列「ＧＡＵ、ＡＡＡ、ＣＵＧ」が発現していることを特定する。

[0092] なお、本実施例２に係る情報処理装置が、転置インデックスを用いて、最長一致等のコドン配列を特定する処理は、実施例１で説明したアミノ酸転置インデックス１４３を用いて実行する処理と同様であるため、説明を省略する。

[0093] また、本実施例２に係る情報処理装置の機能ブロック図は、図３に示した情報処理装置１００の機能ブロック図に対応する。なお、図３に示した特定部１５２は、図１４、図１５で説明した処理を追加的に実行するものとする。

[0094] ところで、上述した情報処理装置１００は、相同性のコドン配列、対称性のコドン配列を特定し、モチーフ等に相当する部分を特定していたが、これに限定されるものではなく、多重整列等を特定することができる。多重整列とは、ＤＮＡの塩基配列やタンパク質のアミノ酸配列について、３つ以上の配列間で対応する部分が並ぶように整列したもの、また整列することである。通常、整列する配列群は進化的類縁性を持っていることが仮定される。多重整列の結果に基づいて分子系統樹を推定することができる。

[0095] 次に、上記実施例に示した情報処理装置１００と同様の機能を実現するコンピュータのハードウェア構成の一例について説明する。図１６は、実施例の情報処理装置と同様の機能を実現するコンピュータのハードウェア構成の一例を示す図である。

[0096] 図１６に示すように、コンピュータ３００は、各種演算処理を実行するＣＰＵ３０１と、ユーザからのデータの入力を受け付ける入力装置３０２と、ディスプレイ３０３とを有する。また、コンピュータ３００は、有線または

無線ネットワークを介して、外部装置等との間でデータの授受を行う通信装置３０４と、インタフェース装置３０５とを有する。また、コンピュータ３００は、各種情報を一時記憶するＲＡＭ３０６と、ハードディスク装置３０７とを有する。そして、各装置３０１～３０７は、バス３０８に接続される。

[0097] ハードディスク装置３０７は、前処理プログラム３０７ａ、特定プログラム３０７ｂを有する。また、ＣＰＵ３０１は、各プログラム３０７ａ～３０７ｄを読み出してＲＡＭ３０６に展開する。

[0098] 前処理プログラム３０７ａは、前処理プロセス３０６ａとして機能する。特定プログラム３０７ｂは、特定プロセス３０６ｂとして機能する。

[0099] 前処理プロセス３０６ａの処理は、前処理部１５１の処理に対応する。特定プロセス３０６ｂの処理は、特定部１５２の処理に対応する。

[0100] なお、各プログラム３０７ａ，３０７ｂについては、必ずしも最初からハードディスク装置３０７に記憶させておかなくても良い。例えば、コンピュータ３００に挿入されるフレキシブルディスク（ＦＤ）、ＣＤ－ＲＯＭ、ＤＶＤ、光磁気ディスク、ＩＣカードなどの「可搬用の物理媒体」に各プログラムを記憶させておく。そして、コンピュータ３００が各プログラム３０７ａ，３０７ｂを読み出して実行するようにしてもよい。

符号の説明

[0101]
１００　情報処理装置
１１０　通信部
１２０　入力部
１３０　表示部
１４０　記憶部
１５０　制御部

請求の範囲

[請求項1] 同一のアミノ酸を示す塩基配列の異なる複数のコドンに関し、コドンファイル上の前記複数のコドンの位置を示す第１インデックスを基にして、前記コドンファイル上のアミノ酸の位置を示す第２インデックスを演算し、

前記第２インデックスを基にして、前記コドンファイルに繰り返し発現するアミノ酸配列の位置をそれぞれ特定し、

前記コドンファイルに繰り返し発現する各アミノ酸配列の位置に対応する各コドン配列を、それぞれが相同性を有するコドン配列として特定する

処理をコンピュータに実行させることを特徴とする情報処理プログラム。

[請求項2] 前記コドンファイルに繰り返し発現するアミノ酸配列が、相同性を有するアミノ酸である否かを、アミノ酸同士の相同性の程度を定義したテーブルを基にして評価する処理を更に実行することを特徴とする請求項１に記載の情報処理プログラム。

[請求項3] 前記コドンファイルに繰り返し発現するアミノ酸配列に対して、アミノ酸配列の並び順が逆になる対称的なアミノ酸配列を、前記コドンファイルから特定し、特定した対称的なアミノ酸配列の位置に対応する各コドン配列を特定する処理を更にコンピュータに実行させることを特徴とする請求項１に記載の情報処理プログラム。

[請求項4] 前記演算する処理は、前記複数のコドンに対応する第１インデックスのビットマップの論理和を実行することで、前記複数のコドンに対応するある一つのアミノ酸の第２インデックスのビットマップを演算することを特徴とする請求項１に記載の情報処理プログラム。

[請求項5] コンピュータが実行する情報処理方法であって、

同一のアミノ酸を示す塩基配列の異なる複数のコドンに関し、コドンファイル上の前記複数のコドンの位置を示す第１インデックスを基

にして、前記コドンファイル上のアミノ酸の位置を示す第２インデックスを演算し、

前記第２インデックスを基にして、前記コドンファイルに繰り返し発現するアミノ酸配列の位置をそれぞれ特定し、

前記コドンファイルに繰り返し発現する各アミノ酸配列の位置に対応する各コドン配列を、それぞれが相同性を有するコドン配列として特定する

処理を実行することを特徴とする情報処理方法。

[請求項6] 前記コドンファイルに繰り返し発現するアミノ酸配列が、相同性を有するアミノ酸である否かを、アミノ酸同士の相同性の程度を定義したテーブルを基にして評価する処理を更に実行することを特徴とする請求項５に記載の情報処理方法。

[請求項7] 前記コドンファイルに繰り返し発現するアミノ酸配列に対して、アミノ酸配列の並び順が逆になる対称的なアミノ酸配列を、前記コドンファイルから特定し、特定した対称的なアミノ酸配列の位置に対応する各コドン配列を特定する処理を更にコンピュータに実行させることを特徴とする請求項５に記載の情報処理方法。

[請求項8] 前記演算する処理は、前記複数のコドンに対応する第１インデックスのビットマップの論理和を実行することで、前記複数のコドンに対応するある一つのアミノ酸の第２インデックスのビットマップを演算することを特徴とする請求項５に記載の情報処理方法。

[請求項9] 同一のアミノ酸を示す塩基配列の異なる複数のコドンに関し、コドンファイル上の前記複数のコドンの位置を示す第１インデックスを基にして、前記コドンファイル上のアミノ酸の位置を示す第２インデックスを演算する前処理部と、

前記第２インデックスを基にして、前記コドンファイルに繰り返し発現するアミノ酸配列の位置をそれぞれ特定し、前記コドンファイルに繰り返し発現する各アミノ酸配列の位置に対応する各コドン配列を

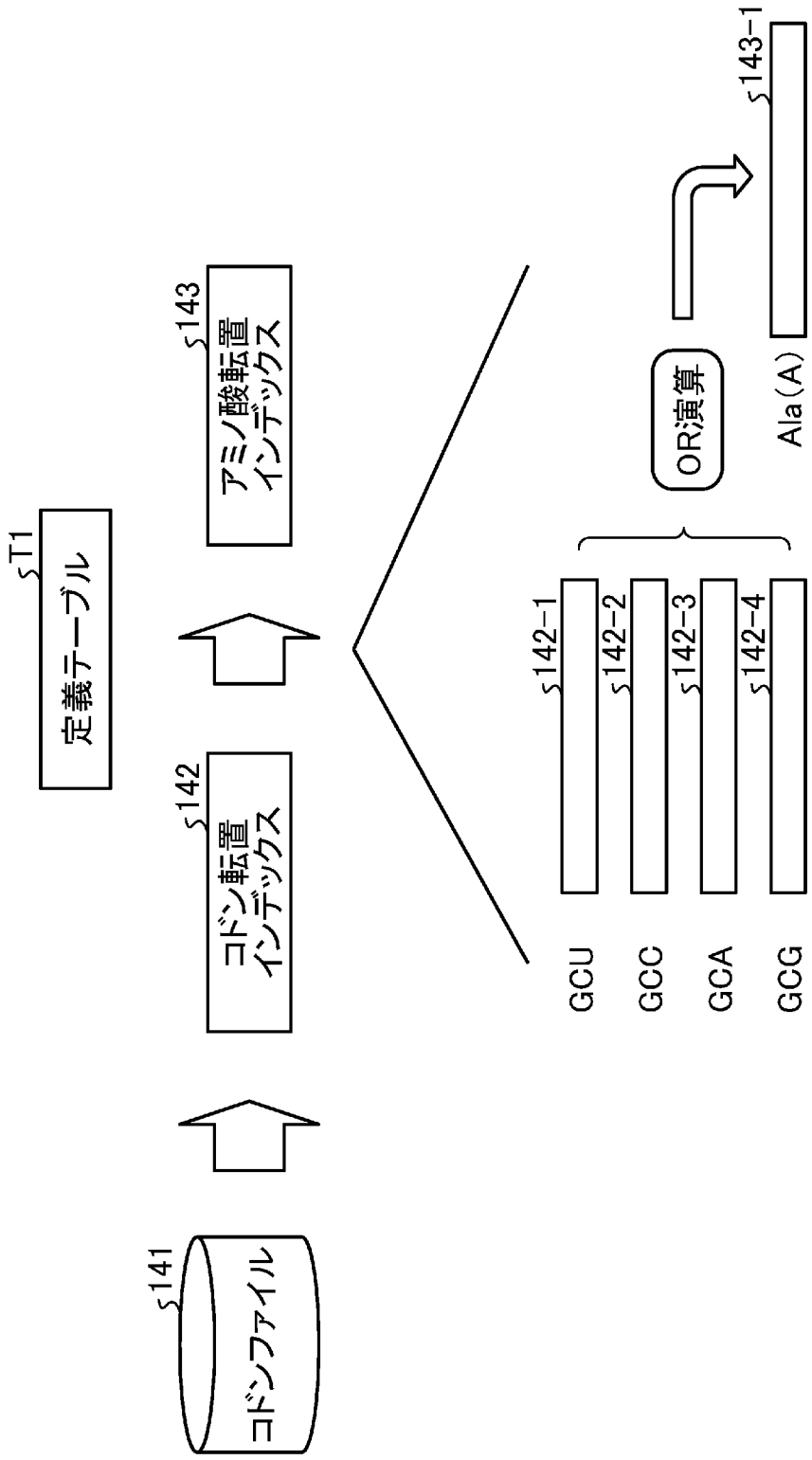
、それぞれが相同性を有するコドン配列として特定する特定部と
を有することを特徴とする情報処理装置。

[請求項10] 前記特定部は、前記コドンファイルに繰り返し発現するアミノ酸配列が、相同性を有するアミノ酸である否かを、アミノ酸同士の相同性の程度を定義したテーブルを基にして評価する処理を更に実行することを特徴とする請求項９に記載の情報処理装置。

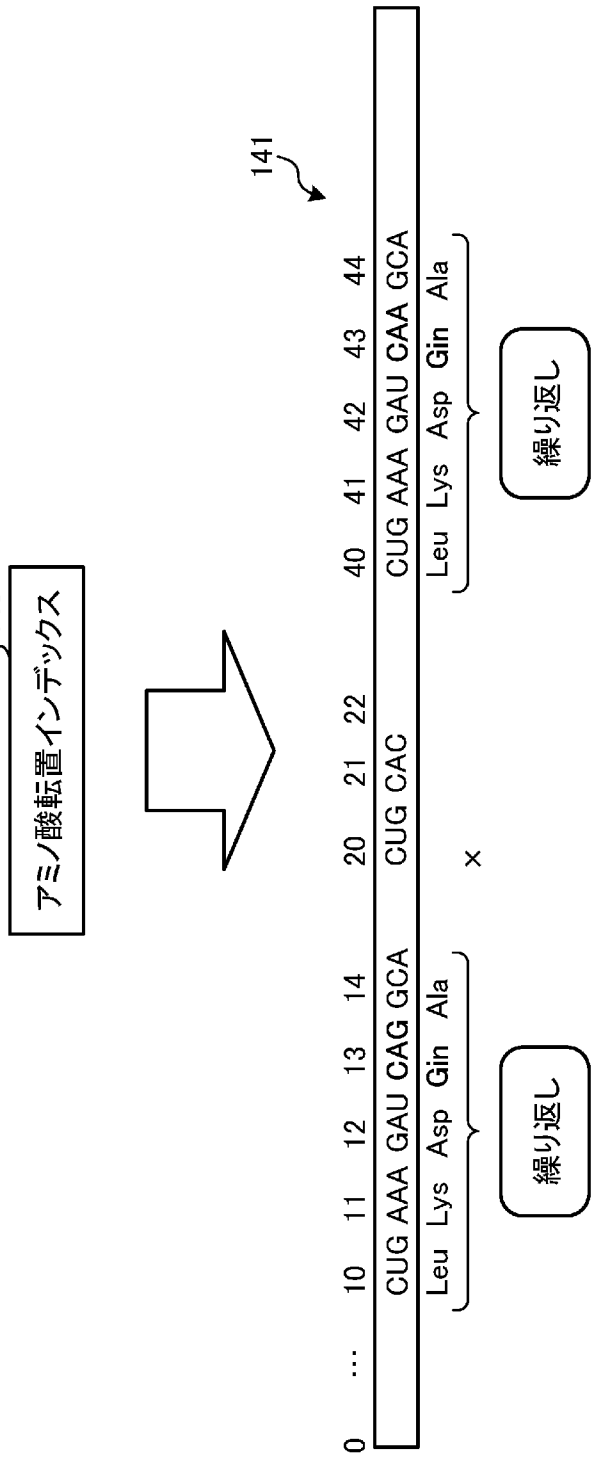
[請求項11] 前記特定部は、前記コドンファイルに繰り返し発現するアミノ酸配列に対して、アミノ酸配列の並び順が逆になる対称的なアミノ酸配列を、前記コドンファイルから特定し、特定した対称的なアミノ酸配列の位置に対応する各コドン配列を特定する処理を更にコンピュータに実行させることを特徴とする請求項９に記載の情報処理装置。

[請求項12] 前記前処理部は、前記複数のコドンに対応する第１インデックスのビットマップの論理和を実行することで、前記複数のコドンに対応するある一つのアミノ酸の第２インデックスのビットマップを演算することを特徴とする請求項９に記載の情報処理装置。

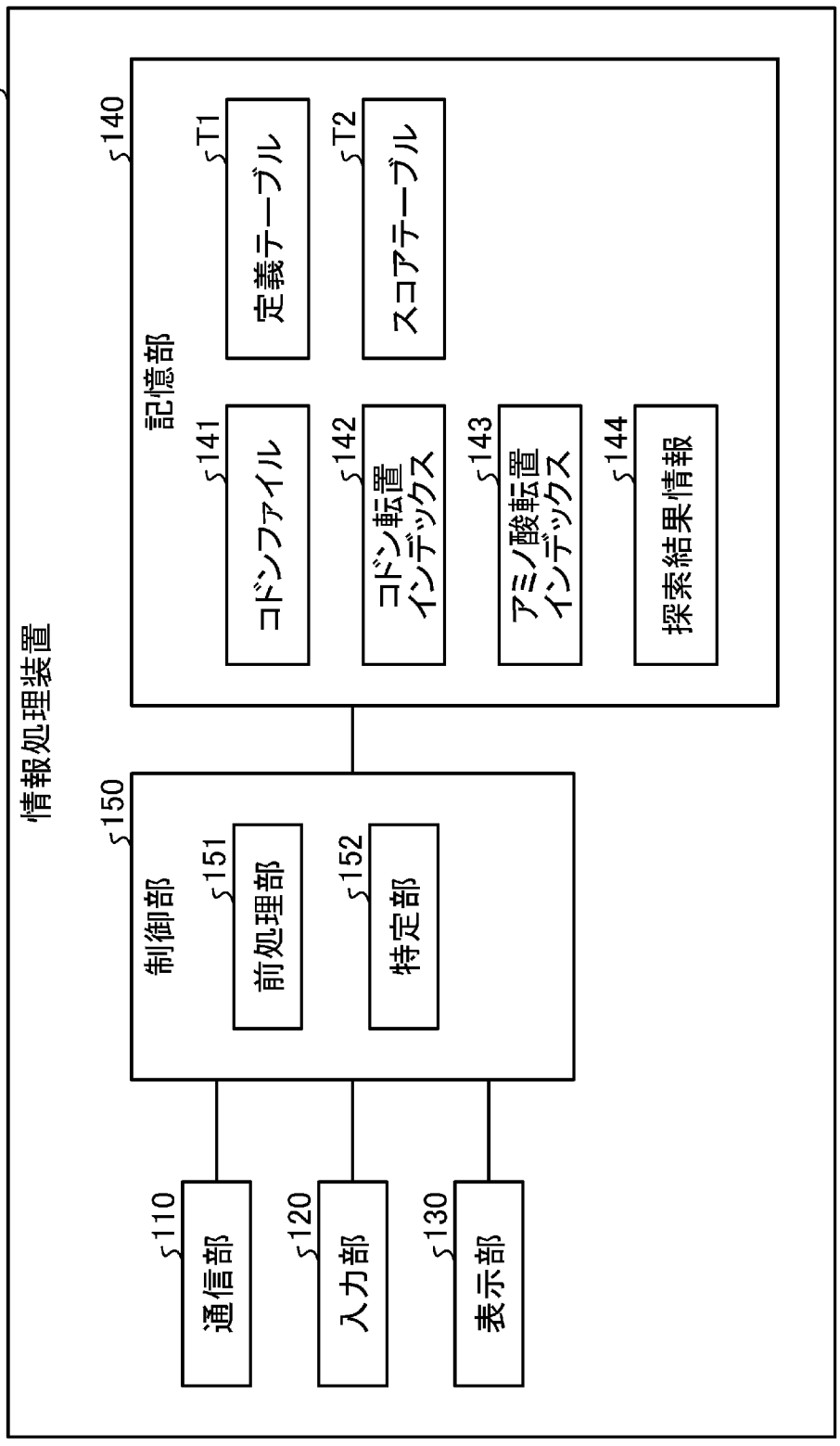
[図1]


141
コドンファイル
142
コドン転置インデックス
143
アミノ酸転置インデックス
T1
定義テーブル
142-1
142-2
142-3
142-4
GCU
GCC
GCA
GCG
OR演算
143-1
Ala (A)

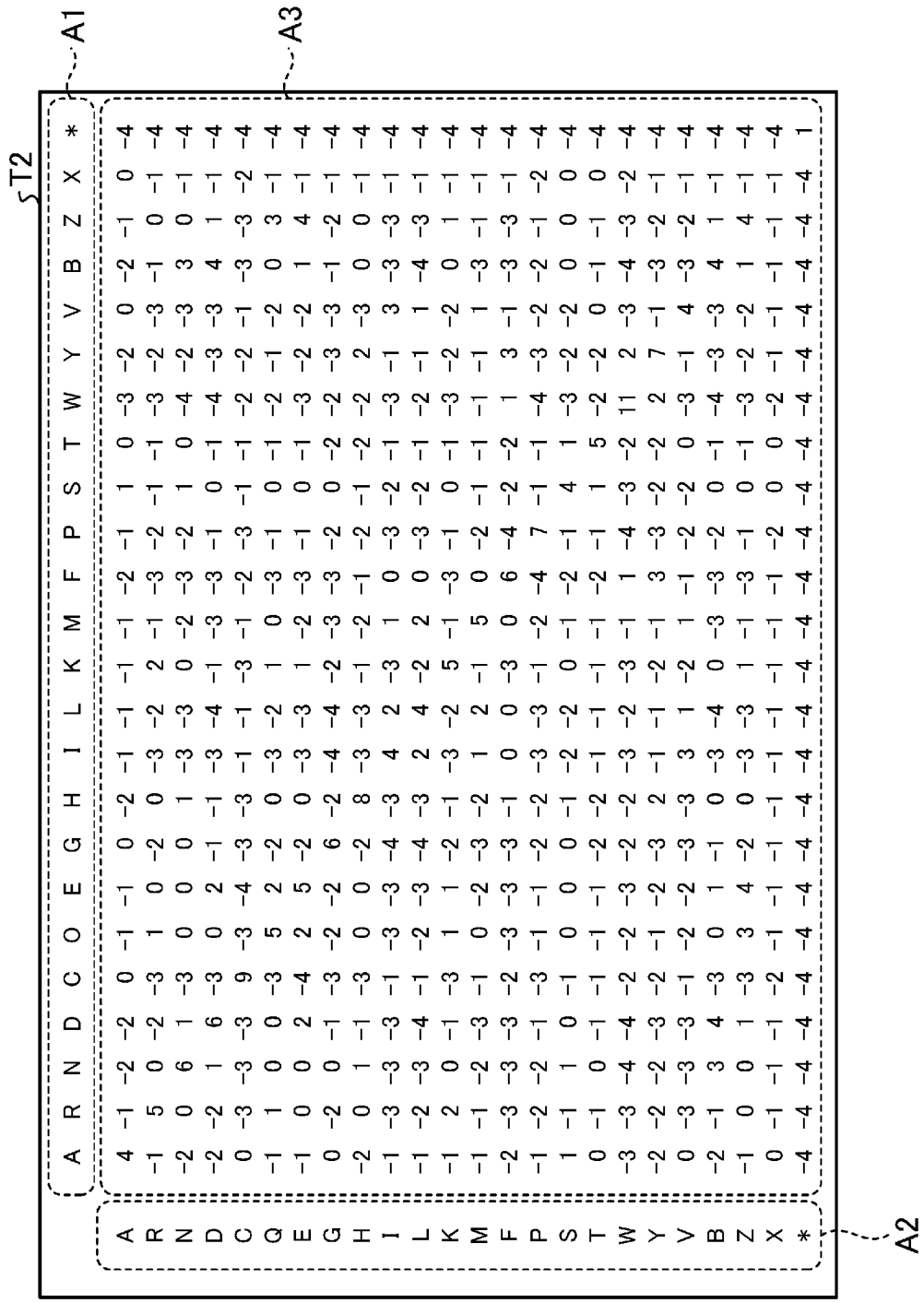
[図2]


141
143
アミノ酸転置インデックス
0 … 10 11 12 13 14 20 21 22 40 41 42 43 44
CUG AAA GAU CAG GCA CUG CAC CUG AAA GAU CAA GCA
Leu Lys Asp Gln Ala
Leu Lys Asp Gln Ala
繰り返し
繰り返し
×

[図3]


100
情報処理装置
140
記憶部
T1
定義テーブル
T2
スコアテーブル
141
コドンファイル
142
コドン転置
インデックス
143
アミノ酸転置
インデックス
144
探索結果情報
150
制御部
151
前処理部
152
特定部
110
通信部
120
入力部
130
表示部

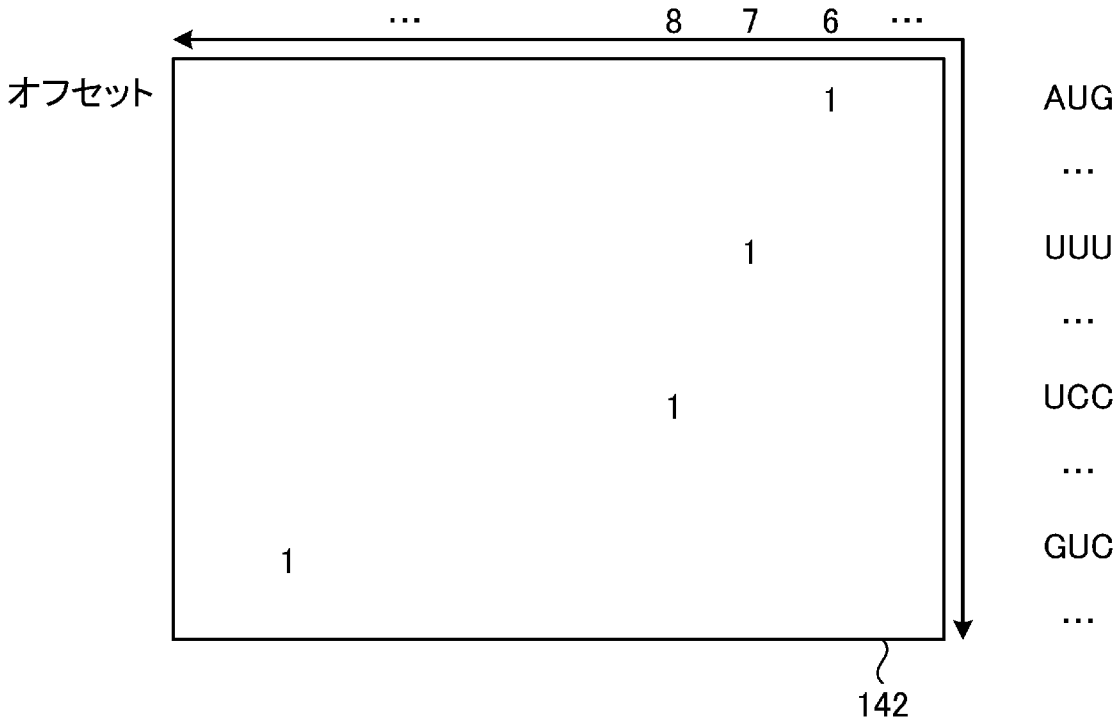
[図4]


T2
A1
A3
A2
A R N D C Q E G H I L K M F P S T W Y V B Z X *
A 4 -1 -2 -2 0 -1 -1 0 -2 -1 -1 -1 -2 -1 1 0 -3 -2 0 -2 -1 0 -4
R -1 5 0 -2 -3 1 0 -2 0 -3 -2 2 -1 -3 -2 -1 -1 -3 -2 -3 -1 0 -1 -4
N -2 0 6 1 -3 0 0 0 1 -3 -3 0 -2 -3 -2 1 0 -4 -2 -3 3 0 -1 -4
D -2 -2 1 6 -3 0 2 -1 -1 -3 -4 -1 -3 -3 -1 0 -1 -4 -3 -3 4 1 -1 -4
C 0 -3 -3 -3 9 -3 -4 -3 -3 -1 -1 -3 -1 -2 -3 -1 0 -2 -1 -3 -2 -1 -4
Q -1 1 0 0 -3 5 2 -2 0 -3 -2 1 0 -3 -1 0 -1 -2 -2 0 3 -1 -4
E -1 0 0 2 -4 2 5 -2 0 -3 -3 1 -2 -3 -1 0 -1 -3 -2 -2 1 4 -1 -4
G 0 -2 0 -1 -3 -2 -2 6 -2 -4 -4 -2 -3 -3 -2 0 -2 -2 -3 -3 -1 -2 -1 -4
H -2 0 1 -1 -3 0 0 -2 8 -3 -3 -1 -2 -1 -2 -1 -2 -2 2 -3 0 0 -1 -4
I -1 -3 -3 -3 -1 -3 -3 -4 -3 4 2 -3 1 0 -3 -2 -1 -3 -1 3 -3 -3 -1 -4
L -1 -2 -3 -4 -1 -2 -3 -4 -3 2 4 -2 2 0 -3 -2 -1 -2 -1 1 -4 -3 -1 -4
K -1 2 0 -1 -3 1 1 -2 -1 -3 -2 5 -1 -3 -1 0 -1 -3 -2 -2 0 1 -1 -4
M -1 -1 -2 -3 -1 0 -2 -3 -2 1 2 -1 5 0 -2 -1 -1 -1 -1 1 -3 -1 -1 -4
F -2 -3 -3 -3 -2 -3 -3 -3 -1 0 0 -3 0 6 -4 -2 -2 1 3 -1 -3 -3 -1 -4
P -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4 7 -1 -1 -4 -3 -2 -2 -1 -2 -4
S 1 -1 1 0 -1 0 0 0 -1 -2 -2 0 -1 -2 -1 4 1 -3 -2 -2 0 0 0 -4
T 0 -1 0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1 1 5 -2 -2 0 -1 -1 0 -4
W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1 1 -4 -3 -2 11 2 -3 -4 -3 -2 -4
Y -2 -2 -2 -3 -2 -1 -2 -3 2 -1 -1 -2 -1 3 -3 -2 -2 2 7 -1 -3 -2 -1 -4
V 0 -3 -3 -3 -1 -2 -2 -3 -3 3 1 -2 1 -1 -2 -2 0 -3 -1 4 -3 -2 -1 -4
B -2 -1 3 4 -3 0 1 -1 0 -3 -4 0 -3 -3 -2 0 -1 -4 -3 -3 4 1 -1 -4
Z -1 0 0 1 -3 3 4 -2 0 -3 -3 1 -1 -3 -1 0 -1 -3 -2 -2 1 4 -1 -4
X 0 -1 -1 -1 -2 -1 -1 -1 -1 -1 -1 -1 -1 -1 -2 0 0 -2 -1 -1 -1 -1 -1 -4
* -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 1

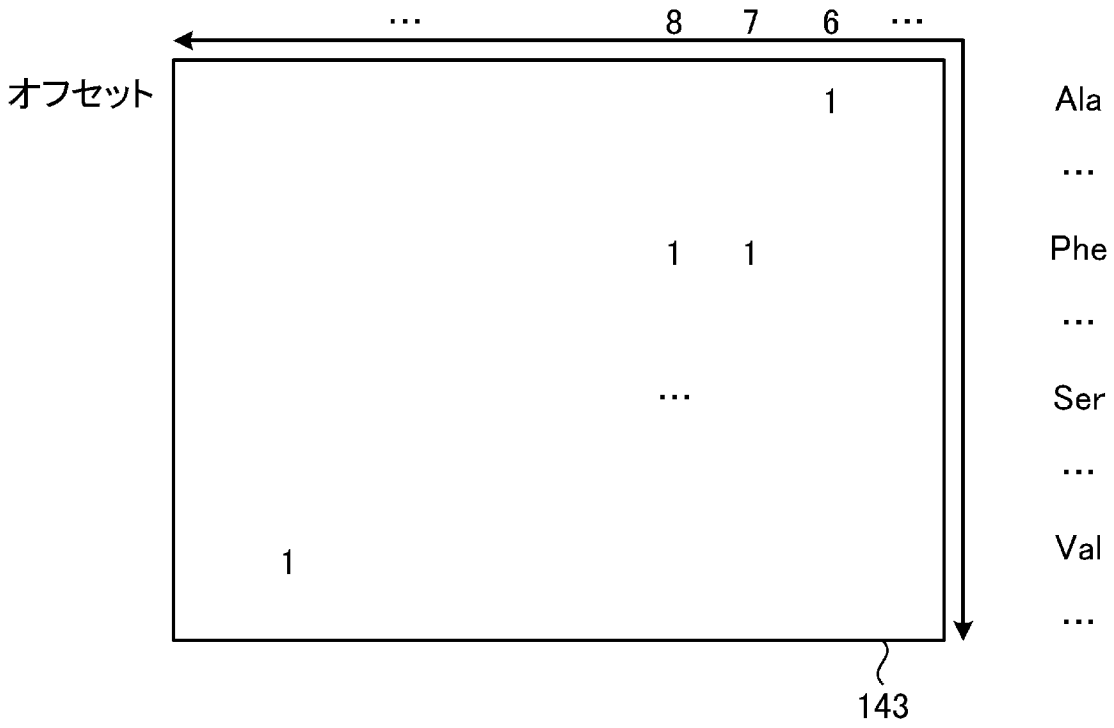
[図5]



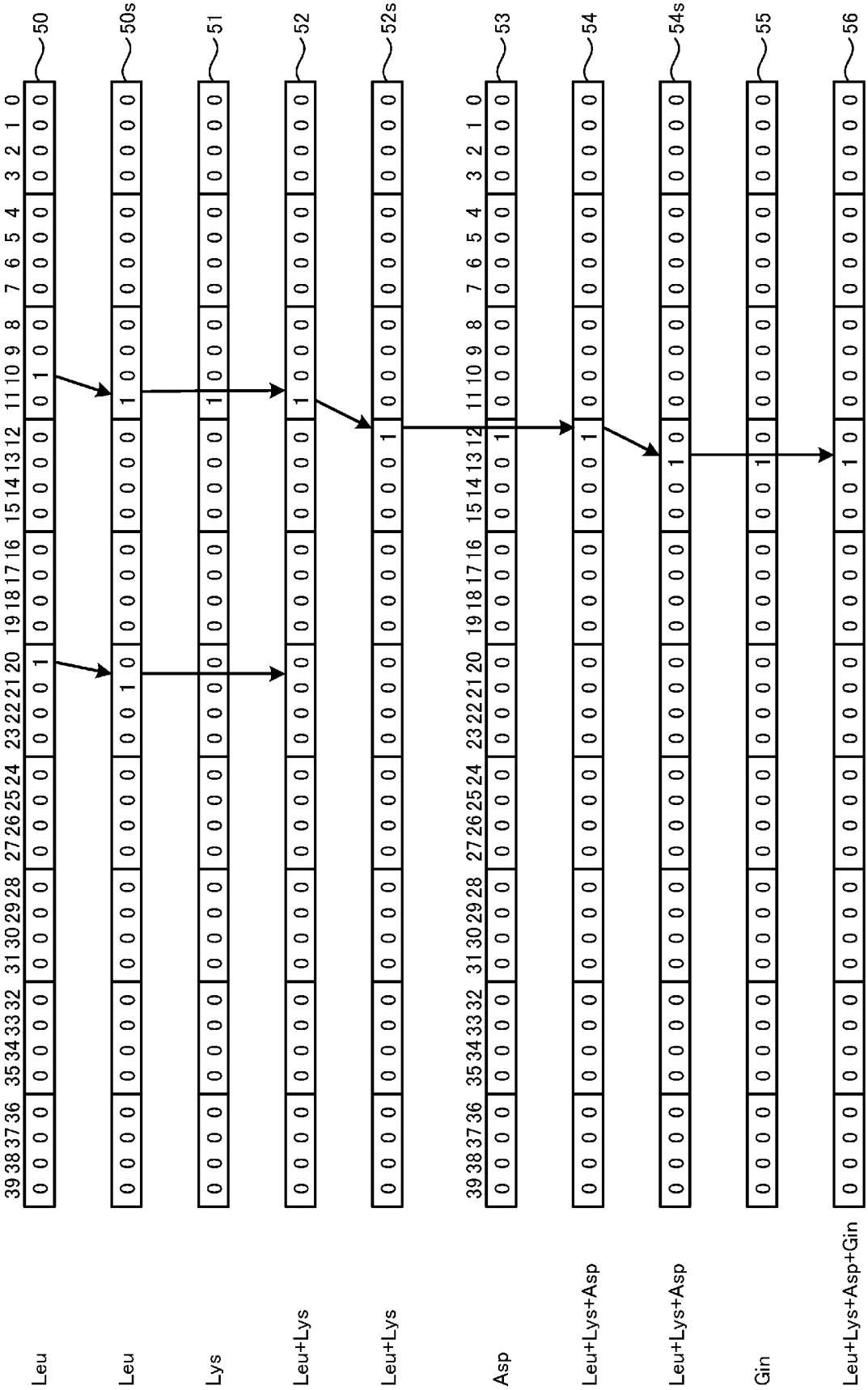
[図6]



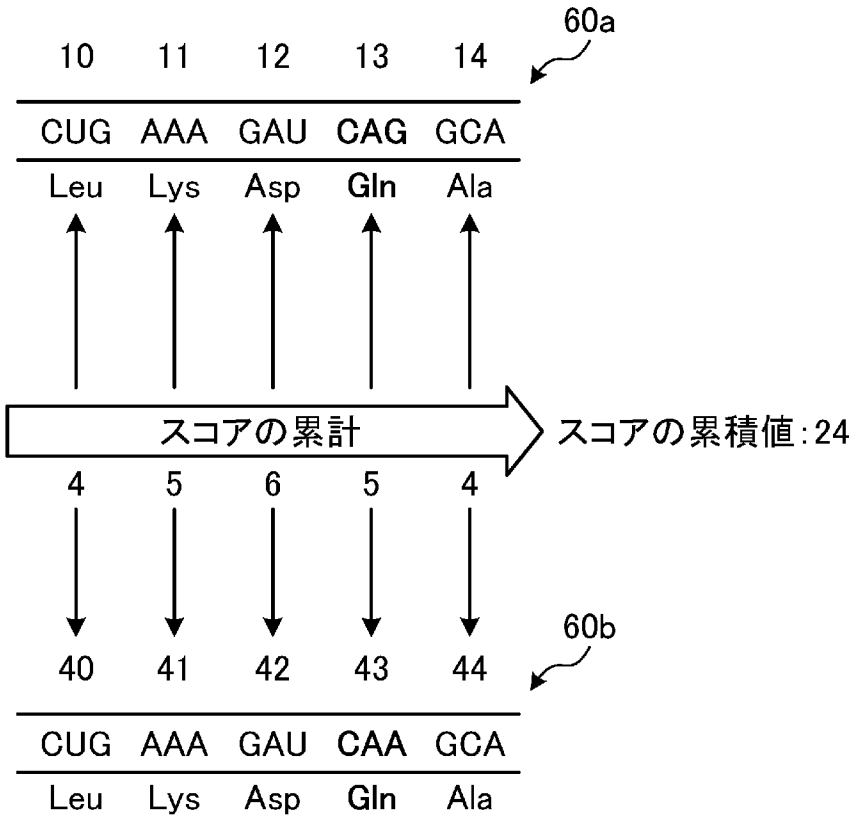
[図7]



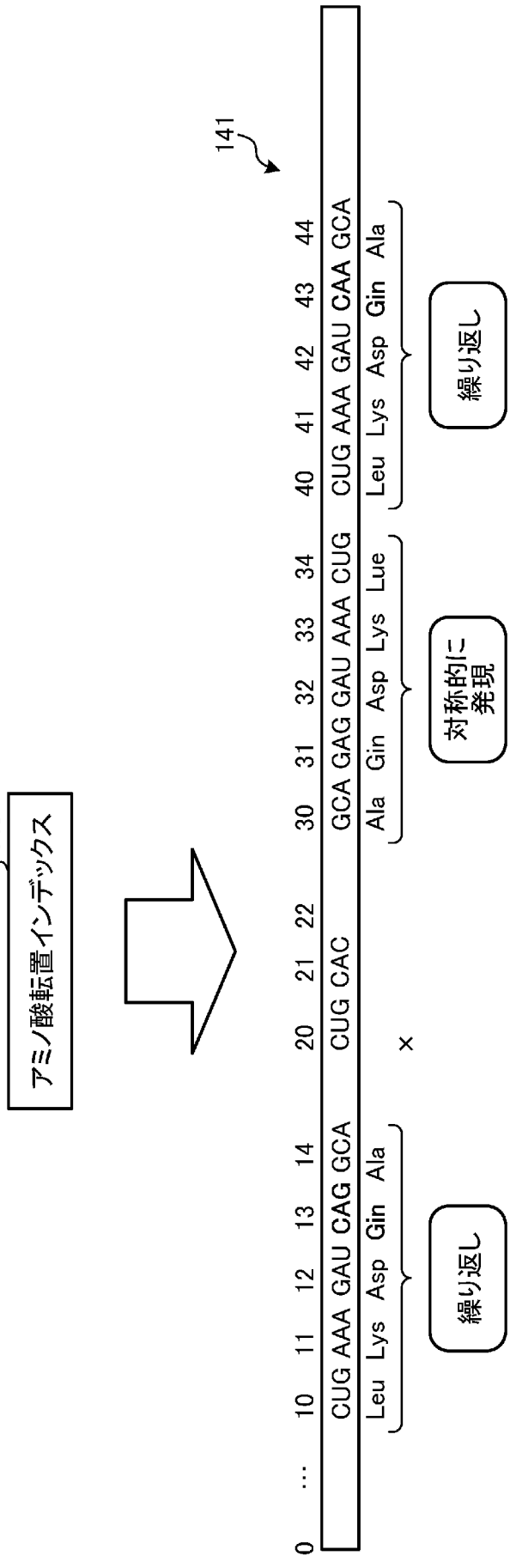
[図8]


39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0
50
50s
51
52
52s
53
54
54s
55
56
Leu
Leu
Lys
Leu+Lys
Leu+Lys
Asp
Leu+Lys+Asp
Leu+Lys+Asp
Gin
Leu+Lys+Asp+Gin

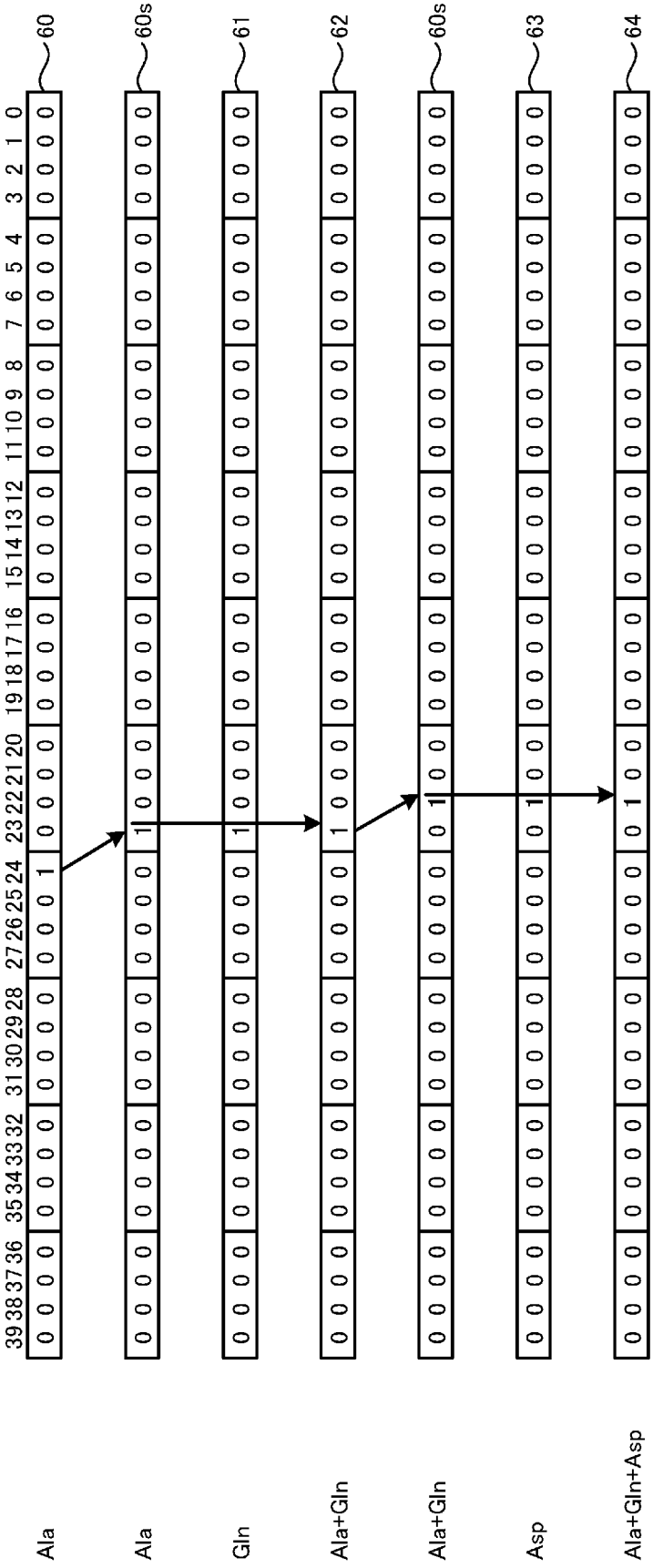
[図9]


60a
10 11 12 13 14
CUG AAA GAU CAG GCA
Leu Lys Asp Gln Ala
スコアの累計
スコアの累積値:24
4 5 6 5 4
60b
40 41 42 43 44
CUG AAA GAU CAA GCA
Leu Lys Asp Gln Ala

[図10]


141
143
アミノ酸転置インデックス
0 … 10 11 12 13 14 20 21 22 30 31 32 33 34 40 41 42 43 44
CUG AAA GAU CAG GCA CUG CAC GCA GAG GAU AAA CUG CUG AAA GAU CAA GCA
Leu Lys Asp Gin Ala
Ala Gin Asp Lys Lue
Leu Lys Asp Gin Ala
繰り返し
対称的に発現
繰り返し
×

[図11]

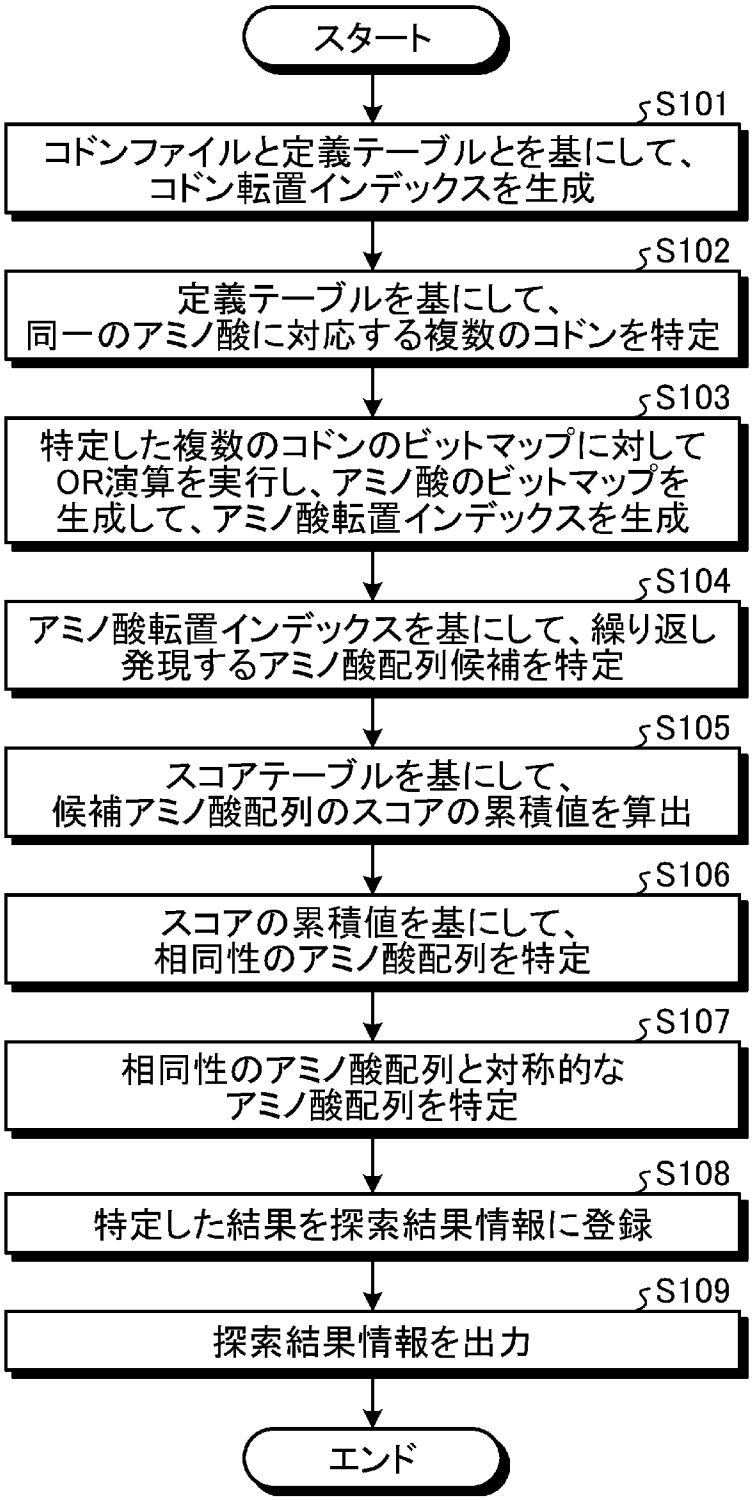

39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0
60
60s
61
62
60s
63
64
Ala
Ala
Gln
Ala+Gln
Ala+Gln
Asp
Ala+Gln+Asp

[図12]

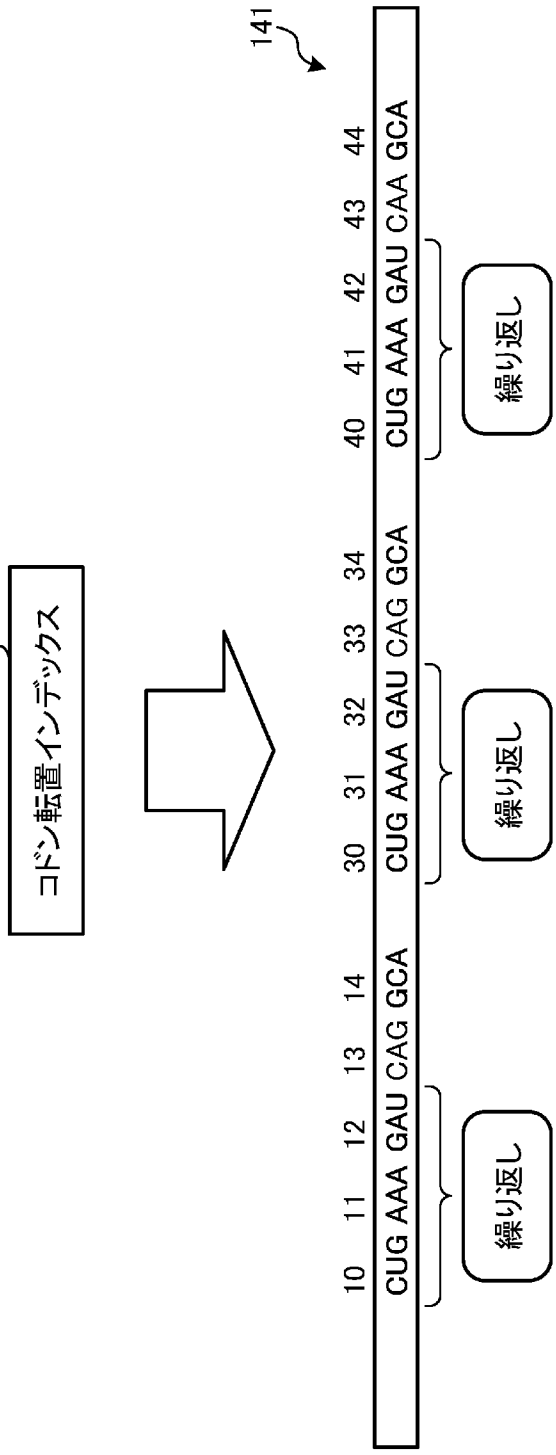
144

アミノ酸配列	第1オフセット	第2オフセット	累計スコア
Leu、Lys、Asp、Gln、Als	10～14、40～44	30～34	24
・・・	・・・	・・・	・・・

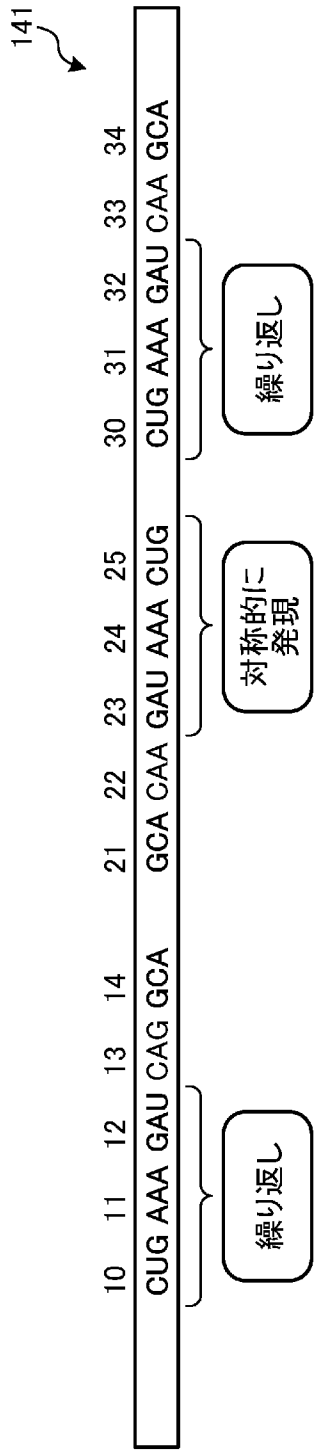
[図13]


スタート
S101
コドンファイルと定義テーブルとを基にして、
コドン転置インデックスを生成
S102
定義テーブルを基にして、
同一のアミノ酸に対応する複数のコドンを特定
S103
特定した複数のコドンのビットマップに対して
OR演算を実行し、アミノ酸のビットマップを
生成して、アミノ酸転置インデックスを生成
S104
アミノ酸転置インデックスを基にして、繰り返し
発現するアミノ酸配列候補を特定
S105
スコアテーブルを基にして、
候補アミノ酸配列のスコアの累積値を算出
S106
スコアの累積値を基にして、
相同性のアミノ酸配列を特定
S107
相同性のアミノ酸配列と対称的な
アミノ酸配列を特定
S108
特定した結果を探索結果情報に登録
S109
探索結果情報を出力
エンド

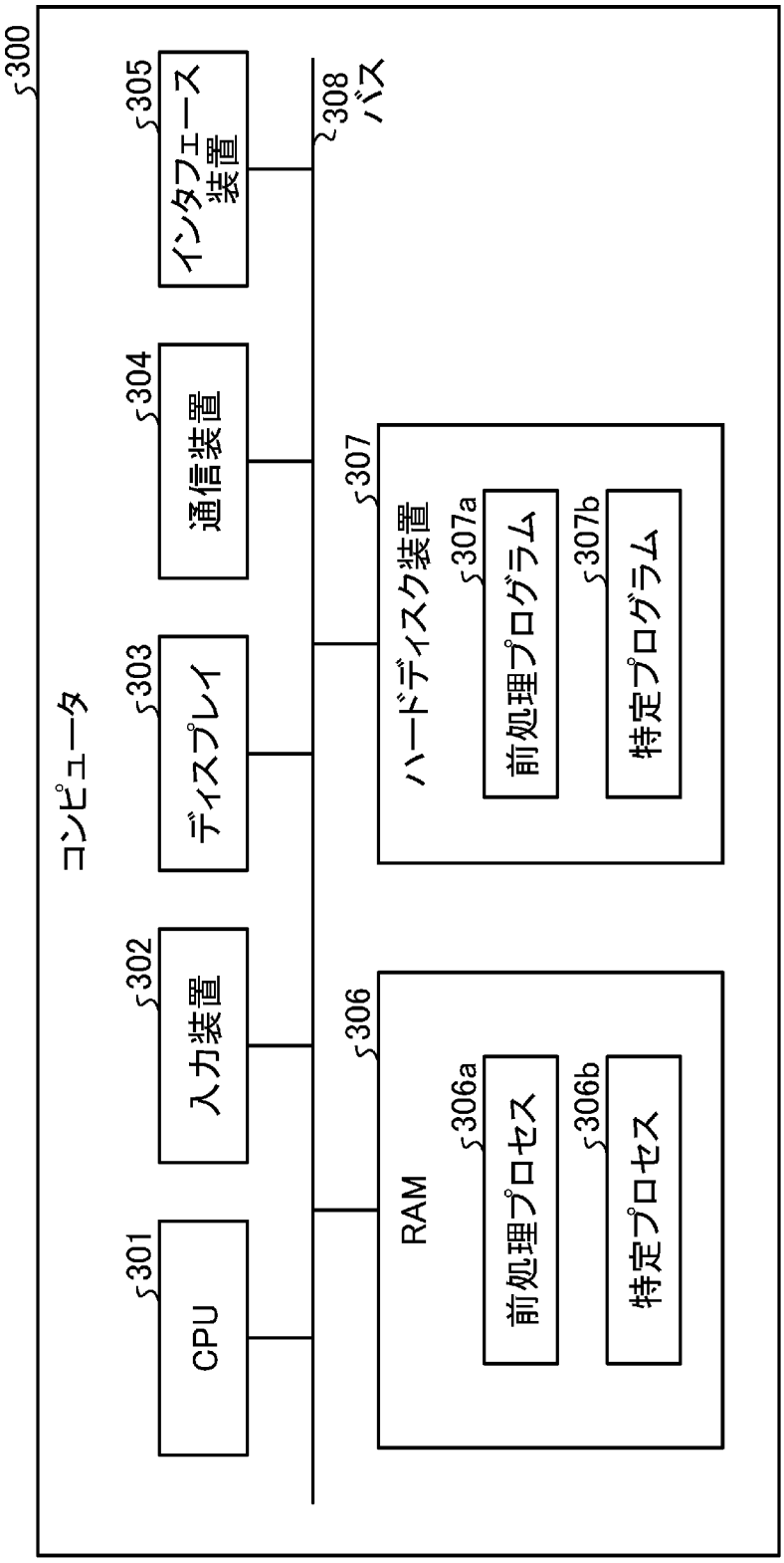
[図14]


141
142
コドン転置インデックス
10 11 12 13 14 30 31 32 33 34 40 41 42 43 44
CUG AAA GAU CAG GCA CUG AAA GAU CAG GCA CUG AAA GAU CAA GCA
繰り返し
繰り返し
繰り返し

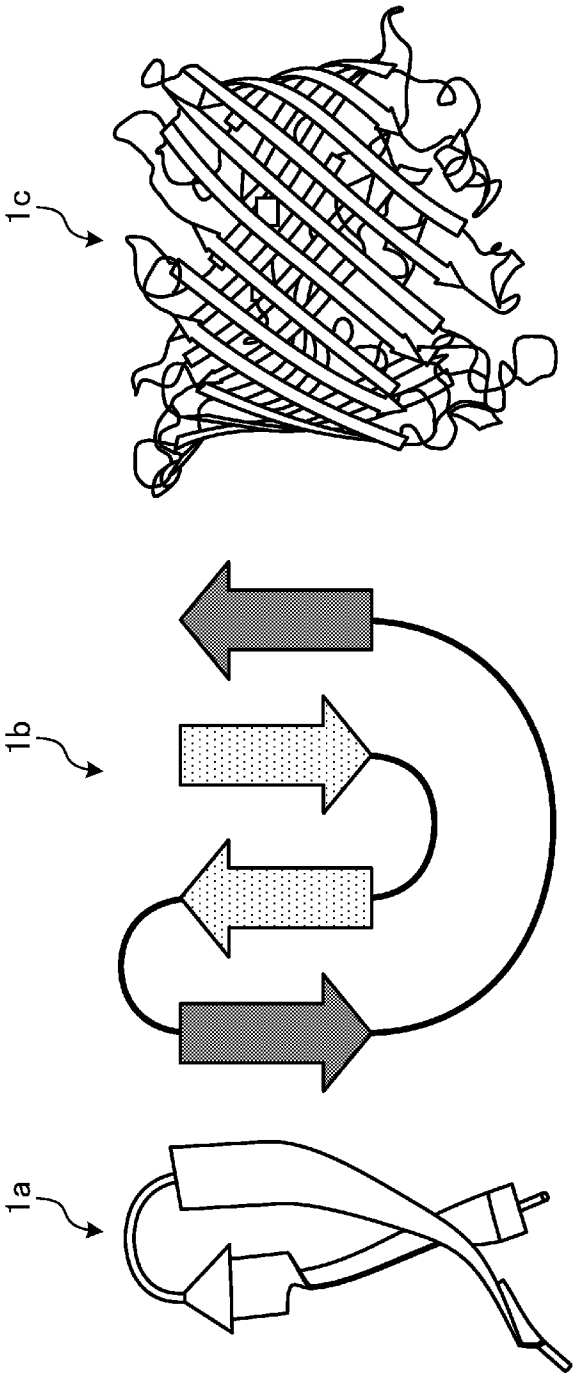
[図15]


141
10 11 12 13 14 21 22 23 24 25 30 31 32 33 34
CUG AAA GAU CAG GCA GCA CAA GAU AAA CUG CUG AAA GAU CAA GCA
繰り返し
対称的に発現
繰り返し

[図16]


300
コンピュータ
301
CPU
302
入力装置
303
ディスプレイ
304
通信装置
305
インタフェース装置
306
RAM
306a
前処理プロセス
306b
特定プロセス
307
ハードディスク装置
307a
前処理プログラム
307b
特定プログラム
308
バス

[図17]



[図18]

アミノ酸			塩基(3塩基配列:コドン)			
名称	略称	記号				
アラニン	Ala	A	GCU	GCC	GCA	GCG
システイン	Cys	C	UGU	UGC		
アスパラギン酸	Asp	D	GAU	GAC		
グルタミン酸	Glu	E			GAA	GAG
フェニルアラニン	Phe	F	UUU	UUC		
グリシン	Gly	G	GGU	GGC	GGA	GGG
ヒスチジン	His	H	CAU	CAC		
イソロイシン	Ile	I	AUU	AUC	AUA	
リシン	Lys	K			AAA	AAG
ロイシン	Leu	L	CUU	CUC	UUA CUA	UUG CUG
メチオニン(開始)	Met	M				AUG
アスパラギン	Asn	N	AAU	AAC		
プロリン	Pro	P	CCU	CCC	CCA	CCG
グルタミン	Gln	Q			CAA	CAG
アルギニン	Arg	R	CGU	CGC	CGA AGA	CGG AGG
セリン	Ser	S	UCU AGU	UCC AGC	UCA	UCG
トレオニン	Thr	T	ACU	ACC	ACA	ACG
バリン	Val	V	GUU	GUC	GUA	GUG
トリプトファン	Trp	W				UGG
チロシン	Tyr	Y	UAU	UAC		
(終止)					UAA	UAG
(終止)					UGA	
決定されてないアミノ酸	Xxx	X				
グルタミン酸か グルタミンのどちらか)	Glx	Z				

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2021/018730

A. CLASSIFICATION OF SUBJECT MATTER
G16B 30/10(2019.01)i
FI: G16B30/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G16B5/00-99/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2021
Registered utility model specifications of Japan	1996-2021
Published registered utility model applications of Japan	1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2020/049748 A1 (FUJITSU LTD) 12 March 2020 (2020-03-12) paragraphs [0129]-[0132], [0187]-[0188], fig. 23, 31	1, 2, 4-6, 8-10, 12
A		3, 7, 11
Y	JP 2015-524658 A (DOW AGROSCIENCES LLC) 27 August 2015 (2015-08-27) paragraph [0071]	1, 2, 4-6, 8-10, 12
A		3, 7, 11
Y	JP 8-77177 A (FUJITSU LTD) 22 March 1996 (1996-03-22) paragraph [0041], fig. 7	2, 6, 10
A	WO 2010/086990 A1 (SPIBER INC) 26 July 2010 (2010-07-26) entire text, all drawings	1-12

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search 12 August 2021 (12.08.2021)	Date of mailing of the international search report 24 August 2021 (24.08.2021)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2021/018730

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2020/049748 A1	12 Mar. 2020	(Family: none)	
JP 2015-524658 A	27 Aug. 2015	US 2015/0175672 A1 paragraph [0084] WO 2014/014950 A1 EP 2872630 A1 AU 2013290374 A CA 2879199 A1 IL 236734 A AR 91774 A1 KR 10-2015-0032334 A CN 104937101 A HK 1211053 A1 RU 2015104984 A ZA 201500638 B	
JP 8-77177 A	22 Mar. 1996	US 5873082 A column 7, lines 45-58, fig. 7	
WO 2010/086990 A1	26 Jul. 2010	US 2011/0281273 A1 whole document EP 2397543 A1 CA 2788452 A	

国際調査報告

国際出願番号
PCT/JP2021/018730

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G16B 30/10(2019.01)i
FI: G16B30/10

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G16B5/00 - 99/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２１年
日本国実用新案登録公報	１９９６－２０２１年
日本国登録実用新案公報	１９９４－２０２１年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
Y	WO 2020/049748 A1（富士通株式会社）12.03.2020（2020-03-12） [0129]-[0132],[0187]-[0188],図23,31	1,2,4-6,8-10,12
A		3,7,11
Y	JP 2015-524658 A（ダウ アグロサイエンシィズ エルエルシー）27.08.2015（2015-08-27） [0071]	1,2,4-6,8-10,12
A		3,7,11
Y	JP 8-77177 A（富士通株式会社）22.03.1996（1996-03-22） [0041],図7	2,6,10
A	WO 2010/086990 A1（スパイバー株式会社）26.07.2010（2010-07-26） 全文,全図	1-12

☐ Ｃ欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

＊ 引用文献のカテゴリー
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
“O” 口頭による開示、使用、展示等に言及する文献
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
“Y” 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
“&” 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
12.08.2021	24.08.2021

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

加舎　理紅子 5R 3054

電話番号 03-3581-1101 内線 3502

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2021/018730

引用文献	公表日	パテントファミリー文献	公表日
WO 2020/049748 A1	12.03.2020	（ファミリーなし）	
JP 2015-524658 A	27.08.2015	US 2015/0175672 A1 [0084] WO 2014/014950 A1 EP 2872630 A1 AU 2013290374 A CA 2879199 A1 IL 236734 A AR 91774 A1 KR 10-2015-0032334 A CN 104937101 A HK 1211053 A1 RU 2015104984 A ZA 201500638 B	
JP 8-77177 A	22.03.1996	US 5873082 A Column 7 Lines 45-58, Fig.7	
WO 2010/086990 A1	26.07.2010	US 2011/0281273 A1 Whole Document EP 2397543 A1 CA 2788452 A	