(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3611601号
(P3611601)
(45)発行日　平成17年1月19日(2005.1.19)　(24)登録日　平成16年10月29日(2004.10.29)

(51)Int.Cl.[7]　F I
G06F 17/30　G06F 17/30　370Z
G06F 17/30　170F

請求項の数 2　(全 27 頁)

(21)出願番号　特願平6-208308
(22)出願日　平成6年9月1日(1994.9.1)
(65)公開番号　特開平8-77177
(43)公開日　平成8年3月22日(1996.3.22)
審査請求日　平成13年8月8日(2001.8.8)

前置審査

(73)特許権者　000005223
富士通株式会社
神奈川県川崎市中原区上小田中４丁目１番１号
(74)代理人　100074099
弁理士　大菅　義之
(74)代理人　100067987
弁理士　久木元　彰
(72)発明者　野口　保
神奈川県川崎市中原区上小田中１０１５番地　富士通株式会社内

審査官　辻本　泰隆

最終頁に続く

(54)【発明の名称】リスト処理システムとその方法

(57)【特許請求の範囲】
【請求項１】
与えられた塩基配列またはアミノ酸配列の配列データに類似する配列データを配列データベースから検索するホモロジー検索により得られたリストを処理するリスト処理システムであって、
リスト格納手段と、
前記与えられた配列データに類似する複数の配列データの識別子を類似度順に含むリストに、前記ホモロジー検索に用いた検索方法名、該ホモロジー検索に用いた配列データベース名、類似度を高めるために連続する２つの塩基またはアミノ酸の間に挿入されたギャップに対するペナルティの値、および検索時に一度に比較される部分配列に含まれる塩基またはアミノ酸の数を示す情報を含むファイル名を付加して、前記リスト格納手段に格納するプロセッサと、
前記与えられた配列データに対する複数回のホモロジー検索の結果として前記プロセッサにより前記リスト格納手段に格納された複数のリストのファイル名を一覧表示し、特定の検索方法または配列データベースが選択されたとき、表示された複数のリストのうち、選択された検索方法または配列データベースを示す情報を含むファイル名が付加されている２つ以上のリストの内容を表示する表示手段とを備え、
前記プロセッサは、前記２つ以上のリストのすべてに含まれている配列データ、前記２つ以上のリストの同じ順位にある同じ配列データ、前記２つ以上のリストの１つに含まれていないが他のリストに含まれている配列データ、または前記２つ以上のリストのすべてに

含まれている配列データとは異なる配列データを特徴として抽出し、前記表示手段は、抽出された特徴を画面表示することを特徴とするリスト処理システム。

【請求項２】

与えられた塩基配列またはアミノ酸配列の配列データに類似する配列データを配列データベースから検索するホモロジー検索により得られたリストを処理するリスト処理方法であって、

プロセッサが、前記与えられた配列データに類似する複数の配列データの識別子を類似度順に含むリストに、前記ホモロジー検索に用いた検索方法名、該ホモロジー検索に用いた配列データベース名、類似度を高めるために連続する２つの塩基またはアミノ酸の間に挿入されたギャップに対するペナルティの値、および検索時に一度に比較される部分配列に含まれる塩基またはアミノ酸の数を示す情報を含むファイル名を付加して、前記リスト格納手段に格納する処理を、該与えられた配列データに対する複数回のホモロジー検索の結果として得られた複数のリストについて繰り返し、

表示手段が、前記リスト格納手段に格納された前記複数のリストのファイル名を一覧表示し、特定の検索方法または配列データベースが選択されたとき、表示された複数のリストのうち、選択された検索方法または配列データベースを示す情報を含むファイル名が付加されている２つ以上のリストの内容を表示し、

前記プロセッサが、前記２つ以上のリストのすべてに含まれている配列データ、前記２つ以上のリストの同じ順位にある同じ配列データ、前記２つ以上のリストの１つに含まれていないが他のリストに含まれている配列データ、または前記２つ以上のリストのすべてに含まれている配列データとは異なる配列データを特徴として抽出し、

前記表示手段が、抽出された特徴を画面表示する

ことを特徴とするリスト処理方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】

本発明は、複数のデータを有する複数のリストを処理して、それらの間の特徴を抽出するリスト処理システムとその方法に関する。

【０００２】

【従来の技術】

近年のバイオテクノロジーの進歩が社会に及ぼす影響は、日増しに増大しつつある。特に、遺伝子の本体であるデオキシリボ核酸（ＤＮＡ）分子に組み換え等の操作を施す遺伝子工学や、既存の蛋白質を基にして新しい蛋白質を作り出す蛋白質工学の発達には著しいものがある。

【０００３】

ＤＮＡは、塩基、糖（デオキシリボース）、リン酸から成るヌクレオチドと呼ばれる構成単位から成る高分子である。このうち、ＤＮＡを構成する塩基には、アデニン（Ａ）、チミン（Ｔ）、シトシン（Ｃ）、グアニン（Ｇ）の４種類がある。それぞれのヌクレオチド間では、デオキシリボースとリン酸同士が結合して鎖状に繋がり、２本の鎖が互いに塩基で結びついた二重螺旋構造を持つ。

【０００４】

塩基同士の結びつきには規則性があり、ＡとＴ、ＣとＧがそれぞれ結合する。これらの塩基のＤＮＡ内における配列順序がＤＮＡの種類、すなわち遺伝子の種類を決めている。

【０００５】

このようにＤＮＡの塩基配列が遺伝情報を記述しているため、遺伝子工学においては、与えられたＤＮＡの複雑な塩基配列（遺伝子配列）を正確かつ迅速に解析する技術が必要不可欠である。

【０００６】

蛋白質は、種類の異なる多数のアミノ酸がペプチド結合により鎖状に連結した高分子化合物である。アミノ酸だけからできているポリペプチドは単純蛋白質と呼ばれ、アミノ酸と

核酸、炭水化物、リン酸等が結合したものは複合蛋白質と呼ばれる。蛋白質の多様な機能は、ポリペプチド鎖を作るアミノ酸の配列順序やポリペプチド鎖の幾何学的配置等により決められている。したがって、蛋白質工学においては、与えられた蛋白質のアミノ酸配列を正確かつ迅速に解析する技術が必要となる。

【０００７】

従来より遺伝子や蛋白質の性質を調べるために、その塩基配列やアミノ酸配列をデータベースに格納された既知の配列データと比較して、与えられた遺伝子等との間にホモロジー（相同性）を有する配列データを求める手法が採られている。このようにデータベース内の類似配列を検索する手法はホモロジー検索と呼ばれている。ホモロジー検索において、一般的には比較する２つの配列データの先頭からそれらの類似部分を検索していく。そして各部分の類似度を計算して配列全体としての類似性を評価する。

【０００８】

しかし、このようなホモロジー検索の手法には確立されたものがあるわけではなく、様々な手法が併用され、比較されながら用いられている。また、検索のために用いられる既存のデータベースにも、複数の異なるものが存在する。同じ配列データを対象にしてホモロジー検索を行っても、その検索結果は用いた手法やデータベース、また、検索の際に与えるパラメータ等によって異なってくる。

【０００９】

従来は、いくつかの手法やパラメータ等を組み合わせて何回かホモロジー検索を行い、それらの結果を比較して検索データの取捨選択を行ったり、最適な検索手法やパラメータを決定したりしていた。

【００１０】

そして、個々のホモロジー検索結果をリスト形式で出力し、手作業でそれらの結果の間の類似点や相違点を調べていた。

【００１１】

【発明が解決しようとする課題】

しかしながら従来の作業方法においては、ホモロジー検索結果が少ない場合は手作業で対処できるが、その数が増えるにしたがって時間もかかるし、間違いも多くなるという問題がある。

【００１２】

また、最近の遺伝子及びアミノ酸配列自動読み取り装置の普及と、遺伝子情報の解明を目的としたヒトゲノムプロジェクトなどのビックプロジェクトの成果として、配列データの件数が飛躍的に増加したのに伴い、個々の検索結果に含まれるデータは既に人間の手作業では対処できない数に達している。

【００１３】

本発明は、複数のデータを有する複数のリストを効率的に処理して、それらの間の特徴を抽出するリスト処理システムとその方法を提供することを目的とする。

【００１４】

さらに詳しくは、バイオテクノロジー等の分野におけるホモロジー検索の結果を項目別に分類管理することにより、それらの結果の類似点や相違点等の特徴を明確に示すことを目的とする。

【００１５】

【課題を解決するための手段】

本発明は、データベースに格納されたデータを処理する情報処理装置におけるリスト処理システムとその方法である。

【００１６】

図１は本発明のリスト処理システムの原理図である。本発明のリスト処理システムは、リスト格納手段１、特徴抽出手段２、および特徴表示手段３を備え、データベース４に格納されたデータを用いて処理を行う。

【００１７】

リスト格納手段１は、データベース４の検索結果であってデータベース４に格納された複数のデータより構成されるリストを複数格納し、データを検索するときに用いた情報を得られたリストのファイル名に付加して格納する。

【００１８】

また、リスト格納手段１は、リストを構成する複数のデータに順位を付加して格納し、データに順位を付加するときに用いた情報をそのリストのファイル名に付加して格納する。

【００１９】

データベース４がホモロジー検索のためのデータベースである場合は、リスト格納手段１は、検索により得られた、与えられたデータに類似したデータの識別子に順位を付加して得られるリストを複数格納する。またリスト格納手段１は、上記与えられたデータの識別名、上記類似したデータの識別子に順位を付加するときに用いた手法の識別名、データベース４の識別名、上記手法のパラメータ等の情報のうちいずれかを前記リストのファイル名に付加して格納し、複数のリストを管理する。

【００２０】

特徴抽出手段２は、リスト格納手段１に格納された複数のリストの間の類似点や相違点等の特徴を抽出して出力する。

特徴表示手段３は、ウィンドウ機能およびグラフィック機能を有し、複数のリストのファイル名を並べて画面表示し、ユーザがいくつかのファイル名を指定すると、指定されたファイル名を持つリストの内容を画面表示する。また特徴表示手段３は、ウィンドウ機能またはグラフィック機能を用いて特徴抽出手段２が出力する上記特徴を画面表示する。

【００２１】

例えば図３に示されるように、本発明のリスト格納手段１は結果ファイル格納部１５であり、特徴抽出手段２はホモロジー検索結果処理プログラム１２を実行する不図示のプロセッサであり、特徴表示手段３はホモロジー検索結果処理プログラム１２を実行するプロセッサと端末１３である。

また、本発明の別のリスト処理システムは、リスト格納手段１、プロセッサ、および表示手段を備え、与えられた塩基配列またはアミノ酸配列の配列データに類似する配列データを配列データベースから検索するホモロジー検索により得られたリストを処理する。プロセッサは、与えられた配列データに類似する複数の配列データの識別子を類似度順に含むリストに、ホモロジー検索に用いた検索方法名、ホモロジー検索に用いた配列データベース名、類似度を高めるために連続する２つの塩基またはアミノ酸の間に挿入されたギャップに対するペナルティの値、および検索時に一度に比較される部分配列に含まれる塩基またはアミノ酸の数を示す情報を含むファイル名を付加して、リスト格納手段１に格納する。表示手段は、与えられた配列データに対する複数回のホモロジー検索の結果としてプロセッサによりリスト格納手段１に格納された複数のリストのファイル名を一覧表示し、特定の検索方法または配列データベースが選択されたとき、表示された複数のリストのうち、選択された検索方法または配列データベースを示す情報を含むファイル名が付加されている２つ以上のリストの内容を表示する。そして、プロセッサは、２つ以上のリストのすべてに含まれている配列データ、２つ以上のリストの同じ順位にある同じ配列データ、２つ以上のリストの１つに含まれていないが他のリストに含まれている配列データ、または２つ以上のリストのすべてに含まれている配列データとは異なる配列データを特徴として抽出し、表示手段は、抽出された特徴を画面表示する。

【００２２】

【作用】

リスト格納手段１が与えられたデータや手法の識別名等の情報をリストのファイル名に付加して管理し、特徴表示手段３がそれらの識別名等を付加されたファイル名を画面に表示するので、ユーザは各リストがどのようにして得られたのかを一目で認識できる。また、表示されたファイル名を持つリストの中から特定の情報を持つリストを容易に選ぶことができる。

【００２３】

例えば、ホモロジー検索により得られた配列データの配列名のリストに、検索対象の配列名や検索手法名等をファイル名として付加すれば、リスト（ファイル）の内容を見なくても、どのような検索結果が格納されたファイルであるかが分かる。

【００２４】

また、特徴抽出手段２が複数のリストの間の類似点や相違点等の特徴を抽出して、特徴表示手段３が抽出された特徴を分かりやすく画面に表示するので、複数のリストの間の特徴を効率よく把握することが出来る。ホモロジー検索結果の処理においては、従来のように検索された配列名を手作業で比較する必要がなく、検索結果に含まれる配列名が膨大な数になる場合や多数の検索結果のリストを比較する場合でも、高速にリストを処理することができる。グラフィック機能を用いて、抽出された特徴をグラフ表示すれば、より明確に特徴を認識できる。

【００２５】

【実施例】

以下、本発明のバイオテクノロジーの分野における実施例について、図面を参照しながら詳細に説明する。

【００２６】

図２は、バイオテクノロジーの分野におけるホモロジー検索結果に影響を与える複数の項目を示している。図２において、配列は、検索の対象となる遺伝子の塩基配列または蛋白質のアミノ酸配列である。

【００２７】

ホモロジー検索手法としては、ＦＡＳＴＡ、ＢＬＡＳＴ（Ｂａｓｉｃ　Ｌｏｃａｌ　Ａｌｉｇｎｍｅｎｔ　Ｓｅａｒｃｈ　Ｔｏｏｌ　）、Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ法等が知られている。ＦＡＳＴＡは、比較する２つの配列データを単位長さ毎に区切って、この比較単位をずらしながら類似度を計算する手法である。ＢＬＡＳＴはＦＡＳＴＡと似ているが、ＦＡＳＴＡより部分的な類似性を高く評価する手法であり、またＦＡＳＴＡよりも高速に結果が得られる。Ｓｍｉｔｈ－Ｗａｔｅｒｍａｎ法は正確な結果が得られるが、複数のアルゴリズムを用いるため時間がかかる。

【００２８】

既知の配列データベースとしては、我が国のＤＤＢＪ（ＤＮＡ　Ｄａｔａ　Ｂａｎｋ　ｏｆ　Ｊａｐａｎ）、アメリカのＧｅｎＢａｎｋ　（Ｇｅｎｅｔｉｃ　Ｓｅｑｕｅｎｃｅ　Ｄａｔａ　Ｂａｎｋ）、ヨーロッパのＥＭＢＬ（Ｅｕｒｏｐｅａｎ　Ｍｏｌｅｃｕｌａｒ　Ｂｉｏｌｏｇｙ　Ｌａｂｏｒａｔｏｒｙ　）のヌクレオチド配列データベース、ＮＢＲＦ（Ｎａｔｉｏｎａｌ　Ｂｉｏｍｅｄｉｃａｌ　Ｒｅｓｅａｒｃｈ　Ｆｏｕｎｄａｔｉｏｎ　）の核酸配列データベース、ＳＷＩＳＳ－ＰＲＯＴ（Ｓｗｉｓｓ　Ｐｒｏｔｅｉｎ　Ｓｅｑｕｅｎｃｅ　Ｄａｔａ　Ｂａｎｋ）等がある。

【００２９】

また、パラメータは、各ホモロジー検索手法において用いられるパラメータであり、その値の取り方によって同じ検索手法を用いても結果が変わってくる。

当然の事ながら、これらの検索結果がユーザ毎に必要になり、それらを必要に応じて表示しなくてはならない。そこで結果ファイル群のデータをファイル名の指定により簡単に読み込むことができるようなシステムが必要になる。

【００３０】

図３は、本発明の実施例のホモロジー検索結果処理システムの構成を示している。図３のシステムは、既知の配列データを格納する配列データベース（配列ＤＢ）１４、ホモロジー検索の結果得られた配列データのリストを格納する結果ファイル格納部１５、入出力に用いられる端末１３、プログラムを格納する不図示のプログラムメモリ、およびプログラムを実行する不図示のプロセッサを有する。ホモロジー検索プログラム１１とホモロジー検索結果処理プログラム１２は、プログラムメモリに格納され、プロセッサにより実行される。

【００３１】

ホモロジー検索はある配列に対して相同性を持った配列を配列データベース１４中より探すもので、相同性の高い順に検索結果を出力する。その際の指標となる数値は用いる手法によって異なるが、一般には同一性（Ｉｄｅｎｔｉｔｙ）と呼ばれる指標で表される。

【００３２】

ホモロジー検索プログラム１１は、複数のホモロジー検索手法に対応する複数の検索プログラムから成り、ホモロジー検索結果処理プログラム１２から起動される。また、配列データベース１４は、複数の異なるデータベースを表している。

【００３３】

検索の対象となる配列は、ユーザによる端末１３の操作により、キー入力、データベース検索、シーケンサ（配列読み取り装置）等の様々な方法で入力される。さらに端末１３よりホモロジー検索手法及びそのパラメータや検索を行うデータベースが指定されると、ホモロジー検索結果処理プログラム１２はその情報を指定されたホモロジー検索プログラム１１に渡し、指定された配列データベース１４内でのホモロジー検索を行わせ、その結果を受け取る。このとき受け取るホモロジー検索結果は、例えば検索された配列データの識別子（配列ＩＤ）やＩｄｅｎｔｉｔｙ等である。

【００３４】

ホモロジー検索結果処理プログラム１２は、ホモロジー検索結果を一旦結果ファイル格納部１５に格納し、次に指定された検索手法やパラメータに従ってホモロジー検索プログラム１１を起動し、同一の配列について検索を行わせる。これを繰り返すことにより、同一の配列を対象とした複数の結果ファイルが得られる。その後ホモロジー検索結果処理プログラム１２は、各結果ファイルの間の類似点や相違点等の特徴を抽出し、端末１３に表示する。抽出された結果ファイル間の特徴は、不図示のプリンタ等を用いて出力することもできる。

【００３５】

本実施例のシステムはＵＮＩＸ環境を前提としており、各ファイルはＵＮＩＸの持つディレクトリ構造の中に作成される。結果ファイルは、図４に示すように、ユーザのホームディレクトリ配下の本システム専用のディレクトリｈｓの配下に作成される。

【００３６】

結果ファイルのファイル名は、図４に示すように、図２の分類項目毎に区切られ、配列名、手法名、データベース（ＤＢ）名の各識別名、およびパラメータから成る。例えば図４のファイル名ＨＩＶ１１＿ＦＡＳＴＡ＿ＳＷ＿５．５．１のうち、ＨＩＶ１１は検索対象となる未知の配列名を表し、ＦＡＳＴＡは検索に用いる手法名を表し、ＳＷ（ＳＷＩＳＳ－ＰＲＯＴ）は検索に用いるデータベース名を表し、５．５．１はＦＡＳＴＡで用いられるパラメータを表す。このような命名規約によりファイル名を管理すれば、どの配列についてどの手法、データベース、パラメータを用いて検索したのかがファイル名から一目瞭然となる。

【００３７】

図５は、ディレクトリｈｓの配下に作成された結果ファイル名のリストの一例である。ＨＩＶ（Ｈｕｍａｎ Ｉｍｍｕｎｏｄｅｆｉｃｉｅｎｃｙ Ｖｉｒｕｓ）、ＬＹＳＯ（Ｌｙｓｏｚｙｍｅ）、ＵＢＩＱ（Ｕｂｉｑｕｉｔｉｎ ）、ＬＥＣＴＩＮ、ＴＲＹＰＳＩＮは検索対象の配列名を表し、ＦＡＳＴＡ１、ＦＡＳＴＡＮ、ＦＡＳＴＡＯは互いに異なる３種のＦＡＳＴＡを表し、ＳＭ－ＷＴはＳｍｉｔｈ－Ｗａｔｅｒｍａｎ法を表す。ＳＷはデータベースＳＷＩＳＳ－ＰＲＯＴを表し、５．２．１等はそれぞれの検索手法において用いたパラメータである。

【００３８】

図５の各結果ファイルのデータ構造の一例が図６に示されている。図６の結果ファイルは、基本的にはそのファイル名に記述された情報の詳細な内容と、検索された配列に関する情報とから成っている。

【００３９】

図６の例では、ファイル名に５．２．１と記されたパラメータが、ギャップペナルティ（

Ｇａｐ　Ｐｅｎａｌｔｙ　）Ｕ＝５、ギャップペナルティＶ＝２、ｋｔｕｐ＝１であることを示しているのがわかる。ｋｔｕｐは、ＦＡＳＴＡにおいて一度に比較する部分配列に含まれる構成単位（塩基、アミノ酸等）の数を表す。例えば、ここでは構成単位を１個ずつ比較していくことを表している。ギャップペナルティＵ、ギャップペナルティＶについては後述する。

【００４０】

また、検索対象（ＴＡＲＧＥＴ）の実際の配列名はＨＩＶ－１　ＰＲＯＴＥＡＳＥであり、検索したデータベースはＳＷＩＳＳ－ＰＲＯＴであることがわかる。

ＬＩＳＴは、検索されたデータベース内の配列のエントリー名をそのスコアとともに、スコアの高い順に並べたリストである。配列のエントリー名は、図２の各データベース毎に決められた配列データの識別子あるいは識別名であり、図６ではＳＷＩＳＳ－ＰＲＯＴのエントリー名として、例えばＨＩＶ＿ＭＡＮＭＡ等が示されている。スコアは配列データのＩｄｅｎｔｉｔｙを表し、その値が大きいほど検索対象の配列に類似していることを示す。例えばＨＩＶ＿ＭＡＮＭＡのスコアは１１３３であり、この検索結果においてはＨＩＶ－１　ＰＲＯＴＥＡＳＥに最も類似していると考えられる。

【００４１】

図７は、ＦＡＳＴＡによるスコア計算に用いられるスコアテーブルの一例を示している。図７のスコアテーブルは、２つのアミノ酸の間の類似度を表すマトリクスであり、その行と列はそれぞれＡからＸまでのアミノ酸の名称を成分として持つ。Ｘはアミノ酸名が具体的に特定できない場合に相当する。各行と各列の交点の数値がそれらのアミノ酸の類似度であり、値が大きいほど類似度が高いことを示す。この数値はそれぞれのアミノ酸の性質等から決められている。このようなスコアテーブルは一般に数種類考案されており、検索の対象となる配列によって使い分けられている。

【００４２】

ＦＡＳＴＡにおいては、検索対象のアミノ酸配列とデータベース中の配列データのうち、一方の配列中のアミノ酸をスコアテーブルの行内で探し、対応するもう一方の配列中のアミノ酸をスコアテーブルの列内で探して、それらの交点の数値をその配列データのスコアに加算する。そして、対象となる全てのアミノ酸について加算が終了した時点で、あるしきい値より大きなスコアを持つ配列データを検索対象のアミノ酸配列の類似配列と考える。

【００４３】

しかしながら、連続する配列のペアを順次比較するだけでは、配列データのスコアは必ずしも大きくならないので、連続する２つのアミノ酸の間にギャップを挿入して類似度を高める手法が用いられる。

【００４４】

図８は、ＦＡＳＴＡにより求められた配列データの一例を示している。図８において、各アルファベットは図７のマトリクスの行に示されるようなアミノ酸の名称を表し、記号「－」はギャップを表す。これらの配列データにおいては、★印を付加した位置のアミノ酸が全て一致している。もしギャップを入れないで配列データを検索すれば、例えば下から３番目の配列Ａ３３８１３の左端には他の配列と異なるアミノ酸Ｇ、Ｓがあるため、他の配列とはかなりずれてしまい、そのスコアは小さくなる。したがって、複数の★印の位置に他の配列と一致するアミノ酸があるにも関わらず、この配列は検索結果には現れない可能性が高い。

【００４５】

このようにギャップを入れることによって、見落とす可能性のあった配列データを検索することができるが、これを多用するとスコアを無制限に大きくすることができるため、類似とはいえないような配列データまで検索される危険性がある。これを防ぐために、ギャップを入れた場合には次式で算出されるペナルティＰをスコアから減じて、スコアの増大を抑えている。

Ｐ＝ＵＬ＋Ｖ　　（１）

（１）式においてＵ、Ｖは図６のギャップペナルティであり、この場合はＵ＝５、Ｖ＝２である。またＬは挿入したギャップの長さ（構成単位数）である。（１）式からわかるように、ペナルティＰはＬの一次関数で表される。ギャップペナルティＵ、Ｖの値は、ｋｔｕｐとともに、ホモロジー検索のパラメータとして与えられる。これらのパラメータのとり方によってスコアが変わるため、検索される配列データも変わってくる。

【００４６】

図９から図１２までは、蛋白質の一種であるＣＹＴＯＣＨＲＯＭＥを対象としてＦＡＳＴＡにより得られたホモロジー検索の結果ファイルの一例を示している。

図９にはデータベース中の検索されたファイル名が示されており、図１０には検索により類似配列として求められた配列データの個数がスコアの範囲とともに示されている。図１０において、左端の列の数値はスコアの範囲を表し、ｉｎｉｔｎ　の列の数値はギャップを入れて計算した場合の該当するスコアを持つ配列データの個数を表し、ｉｎｉｔ１　の列の数値はギャップを入れないで計算した場合の該当するスコアを持つ配列データの個数を表す。

【００４７】

その右側のグラフはｉｎｉｔｎ　およびｉｎｉｔ１　の場合の配列データの個数を示している。「＝」、「－」、「＋」は、それぞれ２個分の配列データを表す。「－」はｉｎｉｔ１　の場合の配列データを表し、「＋」はｉｎｉｔｎ　の場合にギャップを入れることにより新たに増えた配列データを表す。

【００４８】

図１１および図１２は、ｉｎｉｔ１　のスコアが３３を越える配列データの名称とそのスコアのリストである。図１１の１行目には、４２２１５個の配列データ中の１２４１１０７６個のアミノ酸について比較が行われたことが示され、続いてｉｎｉｔｎ　およびｉｎｉｔ１　の場合のスコアの平均値が示されている。また、４行目の５８６４はｉｎｉｔ１　の場合のスコアが３３を越える配列データの個数を表す。

【００４９】

ｉｎｉｔｎ　の列の数値はギャップを入れて計算したスコアを表し、ｉｎｉｔ１　の列の数値はギャップを入れないで計算したスコアを表す。また、ｏｐｔ　の列の数値はｉｎｉｔ１　の場合の結果に対して公知のＮｅｅｄｌｅｍａｎ－Ｗｕｎｓｈ－Ｓｅｌｌｅｒｓのアライメントを行い、スコアを計算し直した値を表す。図１１および図１２においては、ｉｎｉｔｎ　の場合のスコアの大きな順に配列名が並んでいる。検索手法や用いるパラメータが異なれば、一般に結果ファイルにおける配列名の順序が異なってくる。

【００５０】

検索手法やパラメータを変えて得られる多数の結果ファイルが結果ファイル格納部１５に格納されると、ホモロジー検索結果処理プログラム１２は、得られた結果ファイル名のリストを端末１３のディスプレイ上に表示する。

【００５１】

図１３は、図５に示される結果ファイルのうち、ＨＩＶに関するもののファイル名を表示した場合を示している。図１３において、ユーザは任意の複数の結果ファイル名を選択して、選択された結果ファイルの配列データに関する処理をホモロジー検索結果処理プログラム１２に行わせることができる。この選択操作は、端末１３から特定の手法名やデータベース名を入力することにより行われる。

【００５２】

例えば図１３では、配列名ＨＩＶ、データベース名ＳＷ、およびパラメータ５．２．１が指定され、手法名はワイルドカード＊により無指定となっている。その結果該当するＨＩＶ＿ＦＡＳＴＡ１＿ＳＷ＿５．２．１、ＨＩＶ＿ＦＡＳＴＡＮ＿ＳＷ＿５．２．１、ＨＩＶ＿ＦＡＳＴＡＯ＿ＳＷ＿５．２．１の３つのファイル名が網がけ表示される。ホモロジー検索結果処理プログラム１２は、これらの選択された結果ファイルを必要に応じて結果ファイル格納部１５より読み込んで、各検索結果の類似点や相違点等の特徴を抽出し、それを端末１３等に出力する。

【００５３】
次に図１４から３１までを参照しながら、検索結果の特徴の表示例とその抽出方法について説明する。
図１４、１７、２０、２３、２６、および２９は、図１３で選択された３つの結果ファイルについて抽出された様々な特徴の表示例を示している。これらの図において、画面上方にはＴＡＲＧＥＴとして、検索対象の配列名ＨＩＶが表示され、表示領域２１、２２、２３にはそれぞれ結果ファイルＨＩＶ＿ＦＡＳＴＡ１＿ＳＷ＿５．２．１、ＨＩＶ＿ＦＡＳＴＡＮ＿ＳＷ＿５．２．１、ＨＩＶ＿ＦＡＳＴＡＯ＿ＳＷ＿５．２．１に含まれる配列データのエントリー名がスコアの高い順に表示されている。この場合はＬＯＣＵＳ　１　等が１つの配列データに相当するエントリー名である。また、表示領域２１、２２、２３内の上部にはそれぞれの結果ファイルの手法名、パラメータ、データベース名が表示されている。
【００５４】
また、図１５、１６、１８、１９、２１、２２、２４、２５、２７、２８、３０、３１は、ホモロジー検索結果処理プログラム１２による特徴抽出処理のフローチャートである。
【００５５】
図１４は、選択された結果ファイルの全てに含まれている配列の表示例を示している。図１４において、エントリー名ＬＯＣＵＳ　１　、ＬＯＣＵＳ　２　、ＬＯＣＵＳ　３　、ＬＯＣＵＳ　５　は３つの結果ファイルの全てに含まれているので、結果ファイル間の類似点として網がけ表示される。
【００５６】
図１５は、結果ファイルの選択処理および図１４の類似点を抽出する処理のフローチャートである。図１５において処理が開始されると、例えば図１３のような結果ファイル名のリストが表示される（ステップＳ１）。次にユーザにより端末１３から結果ファイル名の指定情報が入力されると（ステップＳ２）、指定情報に該当する結果ファイル名が選択され、網がけ表示される（ステップＳ３）。続いて、選択された各結果ファイルの内容が例えば図１４のように表示される。このとき選択された結果ファイルの数はｊｍａｘとして不図示のメモリに記憶される。次にユーザが、選択された結果ファイルに共通して含まれるエントリー名の検索を指示すると（ステップＳ４）、指示された検索が行われ、その結果が網がけ表示される（ステップＳ５）。
【００５７】
図１６は、ホモロジー検索結果処理プログラム１２による図１５のステップＳ５の処理のフローチャートである。図１６の処理は、図１５の処理から呼び出されるサブルーチン、あるいは図１５の処理とは別のプロセスにより実行される。図１６において、まず各結果ファイル内の配列のエントリー名をサブルーチンあるいは別プロセスに入力し（ステップＳ１１）、ｉ＝１、１番目の結果ファイルのエントリー数をｉｍａｘとおく（ステップＳ１２）。例えば図１４の場合は、１番目の結果ファイルの内容は表示領域２１に表示されており、ｉｍａｘ＝６となる。また２番目、３番目の結果ファイルの内容は、それぞれ表示領域２２、２３に表示されている。
【００５８】
次にｊ＝２とおき（ステップＳ１３）、１番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるかどうかを判定する（ステップＳ１４）。判定結果がＹＥＳの場合は、ｊに１を加算し（ステップＳ１５）、続いてｊとｊｍａｘの値を比較する（ステップＳ１６）。ステップＳ１６でｊがｊｍａｘを越えていなければ（ステップＳ１６、ＹＥＳ）、ステップＳ１４以降の処理を繰り返す。
【００５９】
ステップＳ１６でｊがｊｍａｘを越えると（ステップＳ１６、ＮＯ）、そのエントリー名が全ての結果ファイルに含まれることが分かるので、全ての結果ファイルの該当するエントリー名にフラグ（ＦＬＡＧ）を立てる（ステップＳ１７）。そしてｉに１を加算し（ステップＳ１８）、続いてｉとｉｍａｘの値を比較する（ステップＳ１９）。

【００６０】
ステップＳ１４で判定結果がＮＯの場合は、ステップＳ１５からＳ１７の処理を行わずに、ステップＳ１８の処理に進む。
ステップＳ１９でｉがｉｍａｘを越えていなければ（ステップＳ１９、ＹＥＳ）、ステップＳ１３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ１９、ＮＯ）。その後、図１５のステップＳ５では、フラグの立っている各結果ファイルのエントリー名が網がけ表示される。
【００６１】
例えば図１４の場合は、ｉ＝１、２、３、５にそれぞれ対応するＬＯＣＵＳ　１　、ＬＯＣＵＳ　２　、ＬＯＣＵＳ　３　、ＬＯＣＵＳ　５　の各エントリー名についてステップＳ１７の処理が行われ、表示領域２１、２２、２３内のこれらのエントリー名が網がけ表示される。ｉ＝４、６に対応するエントリー名ＬＯＣＵＳ　４　、ＬＯＣＵＳ　６　については、それぞれｊ＝３、２のときにステップＳ１４の判定結果がＮＯとなり、ステップＳ１７の処理は行われない。したがって、これらのエントリー名は網がけ表示されない。
【００６２】
図１７は、指定した結果ファイルの中の配列とは異なる配列の表示例を示している。図１７においては、ユーザにより表示領域２２の結果ファイルが指定され、他の結果ファイル内のエントリー名のうちＬＯＣＵＳ　６　とＬＯＣＵＳ　８　が指定された結果ファイルに含まれないので、これらのエントリー名が指定された結果ファイルとの相違点として表示領域２１、２２内で網がけ表示される。
【００６３】
図１８は、結果ファイルの選択処理および図１７の相違点を抽出する処理のフローチャートである。図１８のステップＳ２１、Ｓ２２、Ｓ２３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ２３の処理に続いて、ユーザが結果ファイルの指定を行い、その結果ファイルに含まれないエントリー名の検索を指示すると（ステップＳ２４）、指示された検索が行われ、その結果が網がけ表示される（ステップＳ２５）。
【００６４】
図１９は、ホモロジー検索結果処理プログラム１２による図１７のステップＳ２５の処理のフローチャートである。図１９において、まず指定された結果ファイルを１番目として、各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ３１）、ｉ＝１、１番目の結果ファイルのエントリー数をｉｍａｘとおく（ステップＳ３２）。例えば図１７の場合は、１番目の結果ファイルの内容は表示領域２２に表示されており、ｉｍａｘ＝６である。また２番目、３番目の結果ファイルの内容は、それぞれ表示領域２１、２３に表示されている。
【００６５】
次にｊ＝２とおき（ステップＳ３３）、指定された１番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるかどうかを判定する（ステップＳ３４）。判定結果がＹＥＳの場合は、ｊ番目の結果ファイルのそのエントリー名に負のフラグを立て（ステップＳ３５）、ｊに１を加算し（ステップＳ３６）、続いてｊとｊｍａｘの値を比較する（ステップＳ３７）。ステップＳ３４で判定結果がＮＯの場合は、ステップＳ３５の処理を行わずにステップＳ３６の処理に進む。
【００６６】
ステップＳ３７でｊがｊｍａｘを越えていなければ（ステップＳ３７、ＹＥＳ）、ステップＳ３４以降の処理を繰り返す。
ステップＳ３７でｊがｊｍａｘを越えると（ステップＳ３７、ＮＯ）、ｉに１を加算し（ステップＳ３８）、続いてｉとｉｍａｘの値を比較する（ステップＳ３９）。
【００６７】
ステップＳ３９でｉがｉｍａｘを越えていなければ（ステップＳ３９、ＹＥＳ）、ステップＳ３３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ３

９、ＮＯ）。その後、図１８のステップＳ２５では、負のフラグが立っていない、指定されなかった結果ファイルのエントリー名が網がけ表示される。

【００６８】

例えば図１７の場合は、指定された表示領域２２の結果ファイルに含まれるＬＯＣＵＳ　１　、ＬＯＣＵＳ　２　、ＬＯＣＵＳ　４　、ＬＯＣＵＳ　７　、ＬＯＣＵＳ　３　、ＬＯＣＵＳ　５　の各エントリー名についてステップＳ３５の処理が行われ、表示領域２１、２３内のこれらのエントリー名は網がけ表示されない。一方、表示領域２１内のＬＯＣＵＳ　６　と表示領域２３内のＬＯＣＵＳ　８　についてはステップＳ３５の処理は行われないので、これらのエントリー名が網がけ表示される。

【００６９】

図２０は、選択された結果ファイルの全てに含まれている配列とは異なる配列の表示例を示している。選択された結果ファイルの全てに含まれている配列のエントリー名は、図１４に示したとおり、ＬＯＣＵＳ　１　、ＬＯＣＵＳ　２　、ＬＯＣＵＳ　３　、ＬＯＣＵＳ　５　の４つである。図２０においては、それら以外のエントリー名であるＬＯＣＵＳ　４　、ＬＯＣＵＳ　６　、ＬＯＣＵＳ　７　、ＬＯＣＵＳ　８　が、結果ファイル間の相違点として各表示領域内で網がけ表示される。

【００７０】

図２１は、結果ファイルの選択処理および図２０の相違点を抽出する処理のフローチャートである。図２１のステップＳ４１、Ｓ４２、Ｓ４３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ４３の処理に続いて、ユーザが選択された結果ファイルの全てに含まれている配列とは異なる配列の検索を指示すると（ステップＳ４４）、指示された検索が行われ、その結果が網がけ表示される（ステップＳ４５）。

【００７１】

図２２は、ホモロジー検索結果処理プログラム１２による図２１のステップＳ４５の処理のフローチャートである。図２２において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ５１）、ｉ＝１、１番目の結果ファイルのエントリー数をｉｍａｘとおく（ステップＳ５２）。例えば図２０の場合は、１番目の結果ファイルの内容は表示領域２１に表示されており、ｉｍａｘ＝６である。また２番目、３番目の結果ファイルの内容は、それぞれ表示領域２２、２３に表示されている。

【００７２】

次にｊ＝２とおき（ステップＳ５３）、１番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるかどうかを判定する（ステップＳ５４）。判定結果がＹＥＳの場合は、ｊに１を加算し（ステップＳ５５）、続いてｊとｊｍａｘの値を比較する（ステップＳ５６）。ステップＳ５６でｊがｊｍａｘを越えていなければ（ステップＳ５６、ＹＥＳ）、ステップＳ５４以降の処理を繰り返す。

【００７３】

ステップＳ５６でｊがｊｍａｘを越えると（ステップＳ５６、ＮＯ）、そのエントリー名が全ての結果ファイルに含まれることが分かるので、これを網がけ表示させないために全ての結果ファイルの該当するエントリー名に負のフラグを立てる（ステップＳ５７）。そしてｉに１を加算し（ステップＳ５８）、続いてｉとｉｍａｘの値を比較する（ステップＳ５９）。

【００７４】

ステップＳ５４で判定結果がＮＯの場合は、ステップＳ５５からＳ５７の処理を行わずに、ステップＳ５８の処理に進む。

ステップＳ５９でｉがｉｍａｘを越えていなければ（ステップＳ５９、ＹＥＳ）、ステップＳ５３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ５９、ＮＯ）。その後、図２１のステップＳ４５では、負のフラグが立っていない、全ての結果ファイルのエントリー名が網がけ表示される。

【００７５】

例えば図２０の場合は、ｉ＝１、２、３、５にそれぞれ対応するＬＯＣＵＳ　１　、ＬＯ

ＣＵＳ　２　、ＬＯＣＵＳ　３　、ＬＯＣＵＳ　５　の各エントリー名についてステップＳ５７の処理が行われ、表示領域２１、２２、２３内のこれらのエントリー名は網がけ表示されない。ｉ＝４、６に対応するエントリー名ＬＯＣＵＳ　４　、ＬＯＣＵＳ　６　については、それぞれｊ＝３、２のときにステップＳ５４の判定結果がＮＯとなり、ステップＳ５７の処理は行われない。また、１番目の結果ファイルに含まれないエントリー名ＬＯＣＵＳ　７　、ＬＯＣＵＳ　８　についてもステップＳ５７の処理は行われない。したがって、ＬＯＣＵＳ　４　、ＬＯＣＵＳ　６　、ＬＯＣＵＳ　７　、ＬＯＣＵＳ　８　のみが網がけ表示される。

【００７６】

図２３は、指定した配列が結果ファイルの中に含まれているかどうかを表示した例を示している。図２３においては、ユーザによりエントリー名ＬＯＣＵＳ　４　とＬＯＣＵＳ　５　が指定され、それぞれ別の網目を用いて網がけ表示される。このように、ユーザは任意の配列を指定して、それが選択された結果ファイル内に含まれているか否か、また、どの結果ファイル内に含まれているかを認識することができる。

【００７７】

図２４は、結果ファイルの選択処理および図２３に示された指定エントリー名を抽出する処理のフローチャートである。図２４のステップＳ６１、Ｓ６２、Ｓ６３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ６３の処理に続いて、ユーザが特定の配列のエントリー名を指定して、選択された結果ファイル内でそのエントリー名の検索を指示すると（ステップＳ６４）、指示された検索が行われ、その結果が網がけ表示される（ステップＳ６５）。

【００７８】

図２５は、ホモロジー検索結果処理プログラム１２による図２４のステップＳ６５の処理のフローチャートである。図２５において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ７１）、ｊ＝１とおく（ステップＳ７２）。

【００７９】

次にｉ＝１、ｊ番目の結果ファイルのエントリー数をｉｍａｘとおく（ステップＳ７３）。例えば図２３の場合は、表示領域２１、２２、２３の結果ファイルをそれぞれ１、２、３番目の結果ファイルとする。これらの結果ファイルのエントリー数はいずれも６なので、ｉｍａｘは６となる。

【００８０】

次にｊ番目の結果ファイルのｉ番目のエントリー名が指定されたエントリー名かどうかを判定する（ステップＳ７４）。判定結果がＹＥＳの場合は、ｊ番目の結果ファイルのｉ番目のエントリー名にフラグを立てる（ステップＳ７７）。そしてｊに１を加算し（ステップＳ７８）、続いてｊとｊｍａｘの値を比較する（ステップＳ５９）。

【００８１】

ステップＳ７４で判定結果がＮＯの場合は、ｉに１を加算し（ステップＳ７５）、続いてｉとｉｍａｘの値を比較する（ステップＳ７６）。ステップＳ７６でｉがｉｍａｘを越えていなければ（ステップＳ７６、ＹＥＳ）、ステップＳ７４以降の処理を繰り返す。ステップＳ７６でｉがｉｍａｘを越えると（ステップＳ７６、ＮＯ）、ステップＳ７８の処理に進む。

【００８２】

ステップＳ７９でｊがｊｍａｘを越えていなければ（ステップＳ７９、ＹＥＳ）、ステップＳ７３以降の処理を繰り返し、ｊがｊｍａｘを越えると処理を終了する（ステップＳ７９、ＮＯ）。その後、図２４のステップＳ６５では、フラグの立っている各結果ファイルのエントリー名が網がけ表示される。

【００８３】

例えば図２３の場合は、ユーザにより最初にエントリー名ＬＯＣＵＳ　４　が指定され、ｊ＝１、ｉ＝４のときとｊ＝２、ｉ＝３のときにステップＳ７７の処理が行われ、表示領域２１、２２内のエントリー名ＬＯＣＵＳ　４　が網がけ表示される。次にエントリー名

ＬＯＣＵＳ　５　が指定され、ｊ＝１、ｉ＝５のときとｊ＝２、ｉ＝６のとき、およびｊ＝３、ｉ＝４のときにステップＳ７７の処理が行われ、表示領域２１、２２、２３内のエントリー名ＬＯＣＵＳ　５　が、別の網目により網がけ表示される。他のエントリー名については、ステップＳ７４の判定結果がＮＯとなり、ステップＳ７７の処理は行われないので、網がけ表示されない。

【００８４】

図２６は、選択された結果ファイルのデータを重ねて、一致している部分を類似点として強調表示した例を示している。図２６（ａ）においては、エントリー名ＬＯＣＵＳ　１　とＬＯＣＵＳ　２　が全ての表示領域の１番目に表示されており、ＬＯＣＵＳ　３　は表示領域２１と２３の２番目に表示されている。したがって、ユーザが重ね合わせ表示を指示すると、図２６（ｂ）に示すようにＬＯＣＵＳ　１　とＬＯＣＵＳ　２　が同じ網目を用いて網がけ表示され、ＬＯＣＵＳ　３　は別の網目で表示される。他のエントリー名については各結果ファイルの間で重なりが生じていないので、網がけ表示されずに個別に表示される。重なりが多いエントリー名ほどその順位の信頼性が高いので、これにより得られた検索結果の相同性の信頼度が示される。

【００８５】

図２７は、結果ファイルの選択処理および図２６（ｂ）の類似点を抽出する処理のフローチャートである。図２７のステップＳ８１、Ｓ８２、Ｓ８３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ８３の処理に続いて、ユーザが重ね合わせ表示を指示すると（ステップＳ８４）、結果ファイルのリスト内において同順位の同じエントリー名の数が求められ（ステップＳ８５）、その数に応じてエントリー名が網がけ表示される（ステップＳ８６）。

【００８６】

図２８は、ホモロジー検索結果処理プログラム１２による図２７のステップＳ８５の処理のフローチャートである。図２８において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ９１）、ｉ＝１、結果ファイルのエントリー数の最大値をｉｍａｘとおく（ステップＳ９２）。例えば図２６（ａ）の場合は、結果ファイルのエントリー数はいずれも６なので、ｉｍａｘ＝６である。また、表示領域２１、２２、２３に表示された結果ファイルをそれぞれ順に１、２、３番目の結果ファイルとする。

【００８７】

次にｊ＝１、ｎ＝１、ｎｍａｘ＝１、ｋ（１）＝１、ｋ（ｎ）＝０（ｎ＝２，・・・，ｊｍａｘ）とおき（ステップＳ９３）、ｊ番目の結果ファイルのｉ番目のエントリー名をｎ番目の検索対象にする（ステップＳ９４）。次にｊ＝ｊ＋１とおき（ステップＳ９５）、ｎ番目の検索対象のエントリー名がｊ番目の結果ファイルのｉ番目のエントリー名と一致するか否かを判定する（ステップＳ９６）。判定結果がＮＯの場合は、ｎ＝ｎ＋１とおき（ステップＳ９７）、ｎとｎｍａｘの値を比較する（ステップＳ９８）。ｎがｎｍａｘを越えていなければ（ステップＳ９８、ＮＯ）、ステップＳ９６以降の処理を繰り返す。

【００８８】

ステップＳ９８でｎがｎｍａｘを越えれば（ステップＳ９８、ＹＥＳ）、ｊ番目の結果ファイルのｉ番目のエントリー名をｎ番目の検索対象にする（ステップＳ９９）。続いてｎｍａｘに１を加算し（ステップＳ１００）、ｋ（ｎ）に１を加算し（ステップＳ１０１）、ｊに１を加算してｎ＝１とおいて（ステップＳ１０２）、ｊとｊｍａｘの値を比較する（ステップＳ１０３）。

【００８９】

ステップＳ９６で判定結果がＹＥＳの場合は、ステップＳ９７～Ｓ１００の処理を行わずに、ステップＳ１０１以降の処理を行う。

ステップＳ１０３でｊがｊｍａｘを越えていなければ（ステップＳ１０３、ＹＥＳ）、ステップＳ９６以降の処理を繰り返し、ｊがｊｍａｘを越えると（ステップＳ１０３、ＮＯ）、ｉに１を加算して（ステップＳ１０４）、続いてｉとｉｍａｘの値を比較する（ステップＳ１０５）。

【００９０】
ステップＳ１０５でｉがｉｍａｘを越えていなければ（ステップＳ１０５、ＹＥＳ）、ステップＳ９３以降の処理を繰り返し、ｉがｉｍａｘを越えると処理を終了する（ステップＳ１０５、ＮＯ）。その後、図２７のステップＳ８６では、スコアの順位毎にｎ番目の検索対象に指定されたエントリー名が左から順に表示され、ｋ（ｎ）の値に応じて異なる網目を用いて網がけ表示される。どの網目を用いるかは環境変数により決められている。
【００９１】
例えば図２６（ａ）の場合は、ｉ＝１のとき３つの結果ファイルには同じエントリー名ＬＯＣＵＳ　１　が含まれるのでｎ＝１についてのみ検索が行われ、ｊ＝２、３についてステップＳ９６の判定結果がＹＥＳとなる。したがって、ＬＯＣＵＳ　１　に対応するｋ（１）は２回インクリメントされて（ステップＳ１０１）、３になる。ｉ＝２のときも同様である。
【００９２】
ｉ＝３のとき、まず１番目の結果ファイルの３番目のエントリー名ＬＯＣＵＳ　３　が１番目の検索対象となり（ステップＳ９４）、ＬＯＣＵＳ　３　が２番目の結果ファイルの３番目にないので（ステップＳ９６、ＮＯ）、次に２番目の結果ファイルの３番目のエントリー名ＬＯＣＵＳ　４　が２番目の検索対象となる（ステップＳ９９）。そしてＬＯＣＵＳ　４　に対応するｋ（２）がインクリメントされて１になる（ステップＳ１０１）。次にＬＯＣＵＳ　３　が３番目の結果ファイルの３番目にあるので（ステップＳ９６、ＹＥＳ）、ＬＯＣＵＳ　３　に対応するｋ（１）がインクリメントされて２になる（ステップＳ１０１）。ここで、ｊ＝４＞３＝ｊｍａｘとなるため（ステップＳ１０３、ＮＯ）、ｉがインクリメントされる。
【００９３】
ｉ＝４のとき、まず１番目の結果ファイルの４番目のエントリー名ＬＯＣＵＳ　４　が１番目の検索対象となり（ステップＳ９４）、ＬＯＣＵＳ　４　が２番目の結果ファイルの４番目にないので（ステップＳ９６、ＮＯ）、次に２番目の結果ファイルの４番目のエントリー名ＬＯＣＵＳ　７　が２番目の検索対象となる（ステップＳ９９）。そしてＬＯＣＵＳ　７　に対応するｋ（２）がインクリメントされて１になる（ステップＳ１０１）。
【００９４】
ところが、ＬＯＣＵＳ　４　とＬＯＣＵＳ　７　のいずれも３番目の結果ファイルの４番目にないので（ステップＳ９６、ＮＯ）、３番目の結果ファイルの４番目のエントリー名ＬＯＣＵＳ　５　が３番目の検索対象となり（ステップＳ９９）、ＬＯＣＵＳ　５　に対応するｋ（３）がインクリメントされて１になる（ステップＳ１０１）。ここで、ｊ＝４になるため（ステップＳ１０２）、ｉがインクリメントされる。ＬＯＣＵＳ　５　に対応するｋ（１）はインクリメントされず、１のままである（ステップＳ９３）。ｉ＝５、６のときも同様である。
【００９５】
図２７のステップＳ８６では、こうして得られた各ｋ（ｎ）の値毎に異なる網目を用いて、図２６（ｂ）に示すように対応するエントリー名が網がけ表示される。ただし、ｋ（ｎ）＝１のエントリー名については強調する必要がないので、網がけ表示されない。
【００９６】
図２９は、選択された結果ファイル内の配列の数をグラフ表示した例を示している。図２９（ａ）においてユーザがグラフ表示を指示すると、図２９（ｂ）に示すように、全ての結果ファイルに含まれる各エントリー名の個数Ｎがグラフ表示される。結果ファイルに多く含まれるエントリー名の配列は検索対象の配列との類似度が高いと考えられ、グラフ化することにより検索結果の信頼性が示される。例えばＬＯＣＵＳ　４　、ＬＯＣＵＳ　５　、ＬＯＣＵＳ　７　は、図２６（ｂ）の重ね合わせ表示においては強調されていないが、表示領域２１、２２、２３に複数現れていることが明らかになる。
【００９７】
図３０は、結果ファイルの選択処理および図２９（ｂ）のグラフ化処理のフローチャート

である。図３０のステップＳ１１１、Ｓ１１２、Ｓ１１３の処理は、図１５のステップＳ１、Ｓ２、Ｓ３の処理と同様である。ステップＳ１１３の処理に続いて、ユーザがグラフ表示を指示すると（ステップＳ１１４）、結果ファイル内の同じエントリー名の総数が求められ（ステップＳ１１５）、その数がエントリー名毎にグラフ表示される（ステップＳ１１６）。

【００９８】

図３１は、ホモロジー検索結果処理プログラム１２による図３０のステップＳ１１５の処理のフローチャートである。図３１において、まず各結果ファイル内の配列のエントリー名をサブルーチン等に入力し（ステップＳ１２１）、結果ファイルのエントリー数の最大値をｉｍａｘ、ｎ＝１とおく（ステップＳ１２２）。例えば図２９（ａ）の場合はｉｍａｘ＝６である。また、表示領域２１、２２、２３に表示された結果ファイルをそれぞれ順に１、２、３番目の結果ファイルとする。

【００９９】

次にｉ＝１とおき（ステップＳ１２３）、ｎ番目の結果ファイルのｉ番目のエントリー名にフラグが立っているか否かを判定する。判定結果がＮＯであればｊ＝ｎ＋１、ｋ＝１とおき（ステップＳ１２５）、続いてｎ番目の結果ファイルのｉ番目のエントリー名がｊ番目の結果ファイルに含まれるか否かを判定する（ステップＳ１２６）。この判定結果がＹＥＳの場合は、ｋ＝ｋ＋１とおき（ステップＳ１２７）、ｊ番目の結果ファイルのｉ番目のエントリー名にフラグを立てる（ステップＳ１２８）。次にｊに１を加算し（ステップＳ１２９）、ｊとｊｍａｘの値を比較する（ステップＳ１３０）。

【０１００】

ステップＳ１２６で判定結果がＮＯの場合は、ステップＳ１２９以降の処理を行う。

ステップＳ１３０でｊがｊｍａｘを越えていなければ（ステップＳ１３０、ＹＥＳ）、ステップＳ１２６以降の処理を繰り返し、ｊがｊｍａｘを越えると（ステップＳ１３０、ＮＯ）、ｎ番目の結果ファイルのｉ番目のエントリー名をｋの値とともに不図示のメモリに格納する（ステップＳ１３１）。続いてｉに１を加算し（ステップＳ１３２）、ｉとｉｍａｘの値を比較する（ステップＳ１３３）。ステップＳ１２４で判定結果がＹＥＳの場合は、ステップＳ１３２以降の処理を行う。

【０１０１】

ステップＳ１３３でｉがｉｍａｘを越えていなければ（ステップＳ１３３、ＹＥＳ）、ステップＳ１２４以降の処理を繰り返す。ｉがｉｍａｘを越えるとｎに１を加算し（ステップＳ１３４）、ｎとｊｍａｘの値を比較する（ステップＳ１３５）。

【０１０２】

ステップＳ１３５でｎがｊｍａｘを越えていなければ（ステップＳ１３５、ＹＥＳ）、ステップＳ１２３以降の処理を繰り返し、ｎがｊｍａｘを越えると処理を終了する（ステップＳ１３５、ＮＯ）。その後、図３０のステップＳ１１６では、上記メモリに格納された各エントリー名とそのｋの値が順に取り出され、このｋの値がそのエントリー名の個数としてグラフ表示される。

【０１０３】

例えば図２９（ａ）の場合は、ｎ＝１、ｉ＝１のとき、１番目の結果ファイルの１番目のエントリー名ＬＯＣＵＳ　１　が、２番目、３番目の結果ファイルにも含まれているので（ステップＳ１２６、ＹＥＳ）、ｊ＝２、３についてｋがインクリメントされ（ステップＳ１２７）、２番目、３番目の結果ファイルのＬＯＣＵＳ　１　にフラグが立てられる（ステップＳ１２８）。このときｋ＝３がエントリー名ＬＯＣＵＳ　１　の個数として記憶される（ステップＳ１３１）。ｉ＝２、３、５にそれぞれ相当するＬＯＣＵＳ　２　、ＬＯＣＵＳ　３　、ＬＯＣＵＳ　５　についても同様である。

【０１０４】

次にｉ＝４のとき、１番目の結果ファイルの４番目のエントリー名ＬＯＣＵＳ　４　は２番目の結果ファイルには含まれているが（ステップＳ１２６、ＹＥＳ）、３番目の結果ファイルには含まれていないので（ステップＳ１２６、ＮＯ）、ｊ＝２についてのみｋがイ

ンクリメントされ（ステップＳ１２７）、２番目の結果ファイルのＬＯＣＵＳ　４　にフラグが立てられる（ステップＳ１２８）。このときｋ＝２がエントリー名ＬＯＣＵＳ　４　の個数として記憶される（ステップＳ１３１）。

【０１０５】

次にｉ＝６のとき、１番目の結果ファイルの６番目のエントリー名ＬＯＣＵＳ　６　は他の結果ファイルに含まれていないので（ステップＳ１２６、ＮＯ）、ｋはインクリメントされず、ｋ＝１がエントリー名ＬＯＣＵＳ　６　の個数として記憶される（ステップＳ１３１）。

【０１０６】

次にｉをインクリメントすると（ステップＳ１３２）、ｉがｉｍａｘを越えるので（ステップＳ１３３、ＮＯ）、ｎ＝２として（ステップＳ１３４）ｉを初期化し（ステップＳ１２３）、同様の処理を行う。ここでｉ＝１、２、３、５、６のときは、そのエントリー名にフラグが立っているため（ステップＳ１２４、ＹＥＳ）、ｋはインクリメントされずステップＳ１３１の処理も行われない。

【０１０７】

ｉ＝４のとき、２番目の結果ファイルの４番目のエントリー名ＬＯＣＵＳ　７　にはフラグが立っていなくて（ステップＳ１２４、ＮＯ）、しかもＬＯＣＵＳ　７　は３番目の結果ファイルに含まれるので（ステップＳ１２６、ＹＥＳ）、ｊ＝３についてｋがインクリメントされ（ステップＳ１２７）、３番目の結果ファイルのＬＯＣＵＳ　７　にフラグが立てられる（ステップＳ１２８）。このときｋ＝２がエントリー名ＬＯＣＵＳ　７　の個数として記憶される（ステップＳ１３１）。

【０１０８】

次にｎ＝３のときは、ｉ＝６のエントリー名ＬＯＣＵＳ　８　についてのみステップＳ１２６の判定が行われるが、４番目の結果ファイルは存在しないので（ステップＳ１２６、ＮＯ）、ｋはインクリメントされず、ｋ＝１がエントリー名ＬＯＣＵＳ　８　の個数として記憶される（ステップＳ１３１）。

【０１０９】

こうして記憶された各エントリー名の個数が、図２９（ｂ）のようなグラフとして表示される（ステップＳ１１６）。

以上の実施例では、選択された複数の結果ファイルについてのいくつかの特徴抽出の例を示したが、本発明はこれらに限定されることはなく、ホモロジー検索結果処理プログラム１２として他のアプリケーションを用意することにより、任意の他の特徴を抽出する構成とすることもできる。

【０１１０】

また、図１３、１４等において、ファイル名やエントリー名を網がけ表示しているが、多色あるいは多階調のマーカーを用いてエントリー名をマークして表示する構成にしてもよい。本発明においては、結果ファイル内の検索結果をその後の結果の比較やアライメント等の解析のデータとして利用するために、選択により結果ファイル全体を保存したり、マークされた配列の情報のみを保存したり、あるいはマークされなかった配列の情報のみを保存したりすることができる。

【０１１１】

さらに、本発明は、バイオテクノロジー分野におけるホモロジー検索結果の処理に限らず、順位付けられたデータ項目を有する複数のリストの間の任意の特徴を抽出する処理に適用することができる。

【０１１２】

【発明の効果】

本発明によれば、ホモロジー検索結果に見られるような、互いに類似点と相違点を有する複数のリストを効率よく比較することができる。ホモロジー検索結果の場合、膨大な数の配列名を含む多数のリストの比較を迅速に行うことができる。

【０１１３】

複数のリストを比較してそれらの類似点を抽出することにより、信頼性の高いデータが得られる。ホモロジー検索結果においては、多数のリストに共通して含まれる配列名等の良質なデータを効率よく抽出することができる。

【０１１４】

また、抽出された各種の特徴が分かりやすく画面表示されるので、それらの特徴の把握が容易になり、それらを選択して保存することにより、他のシステムや装置へのデータの受け渡しも容易になる。

【０１１５】

さらに、ホモロジー検索結果を格納するファイルを、使用した検索手法等の項目名を用いて管理しているので、多数のファイルの中から特定の項目名やパラメータを持つファイルを選択する操作が容易になる。

【図面の簡単な説明】

【図１】本発明の原理図である。

【図２】ホモロジー検索結果に影響を与える項目を示す図である。

【図３】本発明の実施例の構成図である。

【図４】本発明の実施例における結果ファイルの命名規約を示す図である。

【図５】本発明の実施例における結果ファイル名を示す図である。

【図６】本発明の実施例における結果ファイルのデータ構造を示す図である。

【図７】ホモロジー検索において用いられるスコアテーブルの一例を示す図である。

【図８】検索された配列データを示す図である。

【図９】ＦＡＳＴＡによる結果ファイルの一例を示す図（その１）である。

【図１０】ＦＡＳＴＡによる結果ファイルの一例を示す図（その２）である。

【図１１】ＦＡＳＴＡによる結果ファイルの一例を示す図（その３）である。

【図１２】ＦＡＳＴＡによる結果ファイルの一例を示す図（その４）である。

【図１３】本発明の実施例における結果ファイル名の画面表示を示す図である。

【図１４】本発明の実施例における共通する配列名の画面表示を示す図である。

【図１５】本発明の実施例における共通する配列名の表示処理を示すフローチャートである。

【図１６】本発明の実施例における共通する配列名の抽出処理を示すフローチャートである。

【図１７】本発明の実施例における指定された結果ファイルに含まれない配列名の画面表示を示す図である。

【図１８】本発明の実施例における指定された結果ファイルに含まれない配列名の表示処理を示すフローチャートである。

【図１９】本発明の実施例における指定された結果ファイルに含まれない配列名の抽出処理を示すフローチャートである。

【図２０】本発明の実施例における共通して含まれない配列名の画面表示を示す図である。

【図２１】本発明の実施例における共通して含まれない配列名の表示処理を示すフローチャートである。

【図２２】本発明の実施例における共通して含まれない配列名の抽出処理を示すフローチャートである。

【図２３】本発明の実施例における指定された配列名の画面表示を示す図である。

【図２４】本発明の実施例における指定された配列名の表示処理を示すフローチャートである。

【図２５】本発明の実施例における指定された配列名の抽出処理を示すフローチャートである。

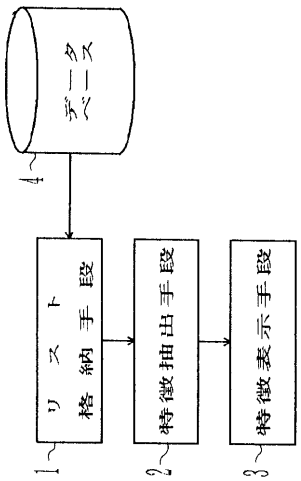
【図２６】本発明の実施例における共通する同順位の配列名の画面表示を示す図である。

【図２７】本発明の実施例における共通する同順位の配列名の表示処理を示すフローチャートである。

【図２８】本発明の実施例における共通する同順位の配列名の抽出処理を示すフローチャートである。
【図２９】本発明の実施例における配列名の数のグラフ表示を示す図である。
【図３０】本発明の実施例における配列名の数のグラフ表示処理を示すフローチャートである。
【図３１】本発明の実施例における配列名の数の計算処理を示すフローチャートである。
【符号の説明】
1　　リスト格納手段
2　　特徴抽出手段
3　　特徴表示手段
4　　データベース

【図１】

本発明の原理図



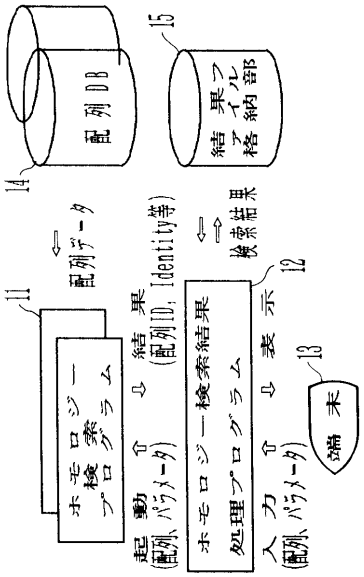

【図２】

ホモロジー検索における分類項目を示す図

| 分類項目 | 内容 |
|---|---|
| 配列 | 検索の対象となる遺伝子またはアミノ酸配列 |
| ホモロジー検索手法 | FASTA, BLAST, Smith-Waterman等の検索手法 |
| データベース | GenBank, EMBL, DDBJ, NRRF, SWISS-PROT等の配列データベース |
| パラメータ | ホモロジー検索手法のパラメータで結果に影響を与える物 |

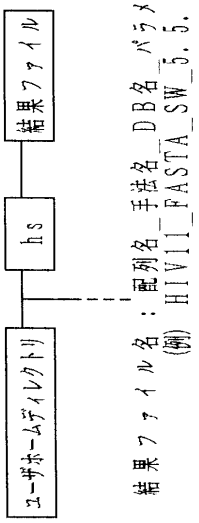
【図３】

実施例の構成図




【図４】

結果ファイルの命名規約を示す図




【図５】

結果ファイル名を示す図

```
HIV_FASTA1_SW_5.2.1      LYSO _FASTA1_SW_5.2.1      UBIQ _FASTA1_SW_5.2.1
HIV_FASTA1_SW_5.5.1      LYSO _FASTAN_SW_5.2.1      UBIQ _FASTAN_SW_5.2.1
HIV_FASTAN_SW_5.2.1      LYSO _FASTA1_SW_5.5.1      UBIQ _FASTA1_SW_5.5.1
HIV_FASTAN_SW_5.5.1      LYSO _FASTAN_SW_5.5.1      UBIQ _FASTAN_SW_5.5.1
HIV_FASTAO_SW_5.2.1                  .                          .
HIV_FASTAO_SW_5.5.1                  .                          .
HIV_SW-WT _SW_5.2.2      LECTIN _FASTA1_SW_5.2.1    TRYPSIN_FASTA1_SW_5.2.1
HIV_SW-WT _SW_5.5.1      LECTIN _FASTAN_SW_5.2.1    TRYPSIN_FASTAN_SW_5.2.1
            .            LECTIN _FASTA1_SW_5.5.1    TRYPSIN_FASTA1_SW_5.5.1
            .            LECTIN _FASTAN_SW_5.5.1    TRYPSIN_FASTAN_SW_5.5.1
```

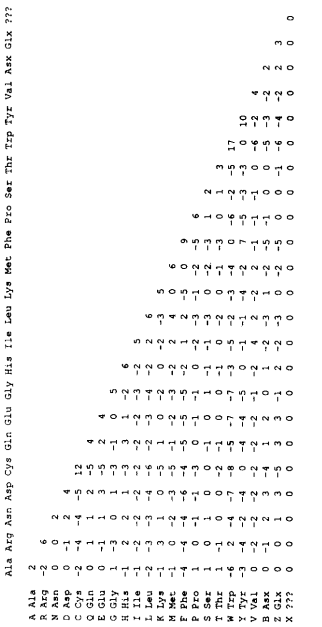
【図６】

結果ファイルのデータ構造を示す図

```
Param：ホモロジー検索の手法で用いられるパラメータの詳細情報
    (例)  Gap Penalty U= 5, Gap Penalty V= 2, ktup=1
Sequence：検索対象(TARGET)配列の名前
    (例) HIV-1 PROTEASE
DB：検索を行うデータベースの名前
    (例) SWISS-PROT
LIST：検索された配列の名前（各データベースごと付けられている配列のエントリー名）
      及びスコア（配列の類似度を示す）
    (例)  HIV _WANWA      1133
          HIV _YEAST      1127
          HIV _THEP3      1018
          HIV _DESVH       999
          HIV _ECOLI       988
          HIV _SCOPO       503
          HIV _ARATH       198
          HIV _PHAAU       174
                .          .
                .          .
```

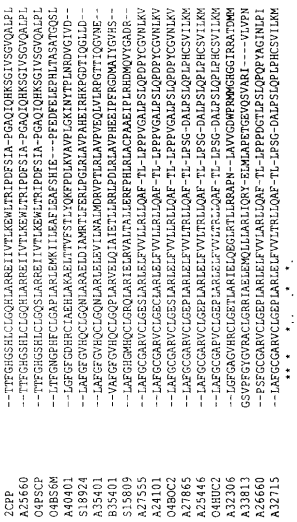
【図７】

スコアテーブルの一例を示す図



【図８】

配列データを示す図



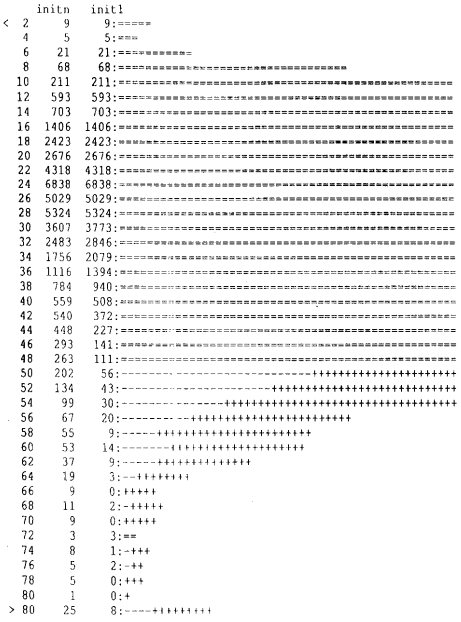
【図９】

FASTAによる結果ファイルを示す図（その1）

```
>2CPP  CYTOCHROME P450CAM (CAMPHOR MONOOXYGENASE) : 405 aa
vs pir library
 search file :
   /rie/homel/pir33/pir1.dat
   /rie/homel/pir33/pir2.dat
   /rie/homel/pir33/pir3.dat
using protein matrix
```

【図１０】

FASTAによる結果ファイルを示す図（その2）



【図１１】

FASTAによる結果ファイルを示す図（その3）

```
12411076 residues in 42215 sequences
  mean initn score:  26.0 (7.30)
  mean init1 score:  25.6 (6.51)
    5864 scores better than 33 saved, ktup: 2, fact: 8  scan time:  0:02:29
The best scores are:                                      initn init1   opt
>A25660  Cytochrome P450 101 - Pseudomonas putida       2050  2050  2050
>O04PSC2 Cytochrome P450 101 - Pseudomonas putida       1999  1801  2003
>O04BSM  Cytochrome P450 106 - Bacillus megaterium       257    99   364
>A40401  Cytochrome P450 55 precursor - Imperfect       214   132   282
>S18924  *Cytochrome P-450soy - Streptomyces grise      172   113   392
>A35401  Cytochrome P450 105A1 - Streptomyces gris      168   125   383
>B35401  Cytochrome P450 105B1 - Streptomyces gris      155    91   367
>S15809  *Cytochrome P450-like protein - Streptomy      143    90   325
>A27555  Steroid 21-monooxygenase cytochrome P450       111    76    78
>A24101  Steroid 21-monooxygenase cytochrome P450       111    74    76
```

【図１２】

FASTAによる結果ファイルを示す図（その4）

```
>O4BOC2   Steroid 21-monooxygenase cytochrome P450  111   76   78
>A27865   Steroid 21-monooxygenase cytochrome P450  111   71   80
>A25446   Steroid 21-monooxygenase cytochrome P450  111   71   80
>O4HUC2   Steroid 21-monooxygenase cytochrome P450  109   71   80
>A32306   Cytochrome P450 103 - Agrobacterium tumef 109   58  248
>A33813   Cytochrome P450 27 precursor, mitochondri 103   62   65
>A26660   Steroid 21-monooxygenase cytochrome P450  102   60   85
>A32715   Steroid 21-monooxygenase cytochrome P450  102   62   62
>JQ1143   Thromboxane-A synthase - Human (EC-number  97   59   68
>A39740   *Sterol 27-hydroxylase precursor - Human   90   55   57
>MNWVRA   Nonstructural polyprotein - Ross River vi  89   60   60
>A29587   Steroid 17alpha-monooxygenase cytochrome   83   55  104
>A40921   *Steroid 17alpha-monooxygenase - Human (E  83   55  104
>A40908   *Steroid 17alpha-monooxygenase (cytochrom  83   55  104
>A26366   Steroid 17alpha-monooxygenase cytochrome   83   55  104
>S12969   *Diacylglycerol kinase - Human (EC-number  80   41   42
>A26289   Steroid 17alpha-monooxygenase cytochrome   78   51   82
>A39607   *Mutation suppressor protein SRP3-1 - Yea  78   35   45
>S04346   Steroid 17alpha-monooxygenase cytochrome   78   51   89
>GNVVTR   Genome polyprotein - Tomato ringspot viru  77   45   46
>QQBE47   DNA-binding protein - Human herpesvirus 4  77   49   55
>TDHULK   Leukocyte antigen-related protein precurs  75   55   73
>B35342   Steroid 11beta-monooxygenase 2 cytochrome  75   50   81
>S09736   *Aldosterone synthase - Rat                75   50   81
>A32693   *Steroid receptor protein svp 1 - Fruit f  75   42   51
>A35342   Steroid 11beta-monooxygenase 1 cytochrome  75   50   81
>MNWVS    Nonstructural polyprotein - Sindbis virus  74   64   64
>VCLJG4   env polyprotein - Simian immunodeficiency  74   41   41
>GNNYBT   Genome polyprotein - Coxsackievirus B3     73   59   59
>GNNYB3   Genome polyprotein - Coxsackievirus B3     73   59   59
>S09156   Diacylglycerol kinase, lymphocyte - Pig    73   35   36
>A35867   Cytochrome P450 71 - Avocado               73   43   49
>A41039   *RNA-directed RNA polymerase - Rabbit hem  73   41   51
>A27124   H+-transporting ATPase - Leishmania donov  73   41   52
>A32525   Steroid 21-monooxygenase cytochrome P450  72   60   84
>MNWVB2   Nonstructural polyprotein - Ockelbo virus  71   61   61
>DJBE28   DNA-directed DNA polymerase - Human herpe  71   41   41
>JX0050   Steroid 11beta-monooxygenase 3 cytochrome  69   41   41
>JX0071   Steroid 11beta-monooxygenase (clone 7-1)   69   41   41
>A38819   *Steroid 11beta-monooxygenase 2 (cytochro  69   41   41
>B26366   Steroid 17alpha-monooxygenase cytochrome   69   51   77
>JX0151   Steroid 11beta-monooxygenase 3 (cytochrom  69   41   41
>A28415   Steroid 11beta-monooxygenase cytochrome P  69   41   41
>S15805   *Cytochrome P450(11beta) - Bovine          69   41   41
>KIECG    GTP pyrophosphokinase - Escherichia coli   69   53   60
>A29943   Toll protein precursor - Fruit fly (Droso  69   43   48
>NCEC7    Exodeoxyribonuclease VII large chain - Es  68   57   66
>C32575   *C-ski protein FB27 - Chicken              68   53   53
>B41378   *Cytochrome c553i precursor - Paracoccus   68   68   79
>TVFVSK   Transforming protein (ski) - Avian erythr  68   53   53
>S18188   *Rat notch protein - Rat                   68   43   43
>A40043   *Notch protein homolog TAN-1 precursor -   68   43   43
>A32575   *C-ski protein FB29 - Chicken              68   53   53
>A22363   Cytochrome P450, phenobarbital-inducible   68   46   51
>S16873   Cytochrome P450 2D4 - Rat                  67   39   87
>A40457   *Replication protein A 70K chain - Human   67   67   76
>B37222   Cytochrome P450 1A2, hepatic - Dog (fragm  67   57   91
>S15435   *Collagen alpha 1(VIII) chain - Human     66   50   58
>A39129   *Catalase HPII - Escherichia coli (EC-num  66   55   58
>A34246   Collagen alpha 1(VIII) chain precursor -   66   50   62
>S05962   Radial spoke protein 3 - Chlamydomonas re  65   39   39
>S13178   *6-Methylsalicylate decarboxylase - Penic  65   41   46
>A40440   *Endothelin 1 and 2 receptor precursor, v  65   43   45
>A25076   Tubulin alpha-1 chain - Yeast (Saccharomy  65   39   40
>S15074   *Calpastatin - Rat                         65   50   51
>A31270   *Radial spoke protein 3 - Chlamydomonas r  65   39   39
```

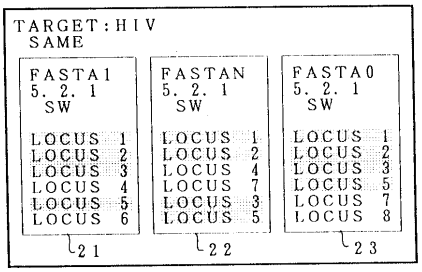
【図１３】

結果ファイル名の画面表示を示す図

```
HIV___*___SW___5. 2. 1
HIV_FASTA1_SW_5. 2. 1
HIV_FASTA1_SW_5. 5. 1
HIV_FASTAN_SW_5. 2. 1
HIV_FASTAN_SW_5. 5. 1
HIV_FASTA0_SW_5. 2. 1
HIV_FASTA0_SW_5. 5. 1
HIV_SM-WT _SW_5. 2. 2
HIV_SM-WT _SW_5. 5. 1
```

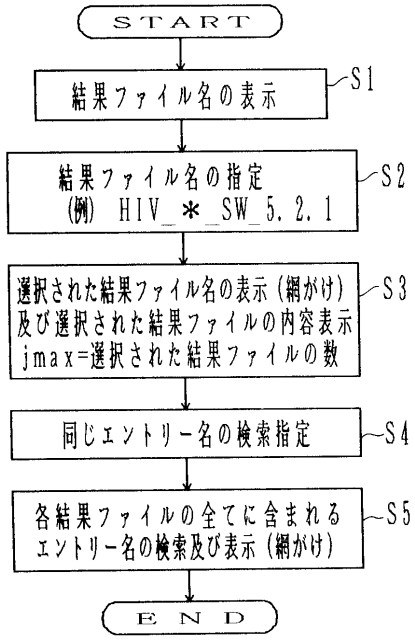
【図１４】

共通する配列名の画面表示を示す図



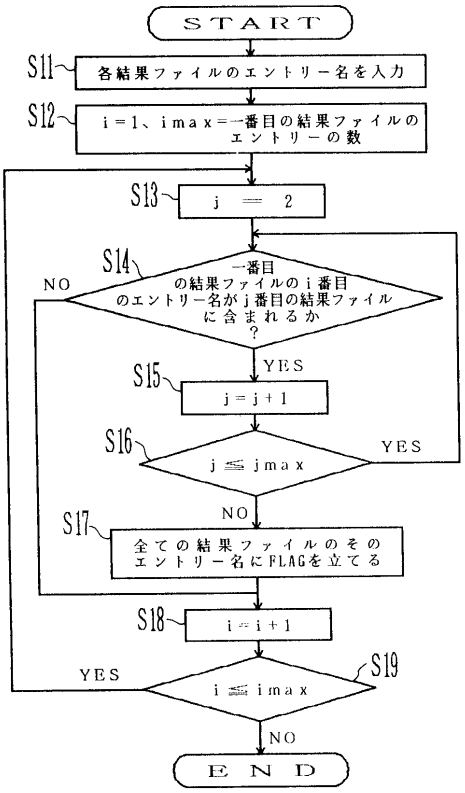
【図１５】

共通する配列名の表示処理のフローチャート



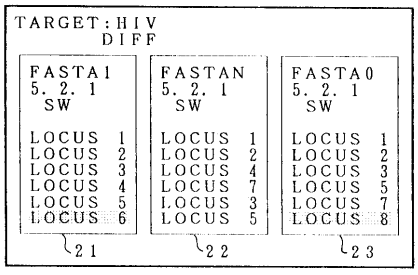

【図１６】

共通する配列名の抽出処理のフローチャート



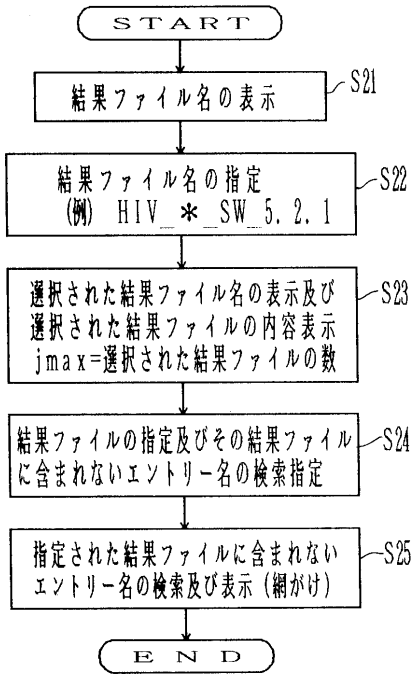

【図１７】

指定された結果ファイルに含まれない配列名の
画面表示を示す図



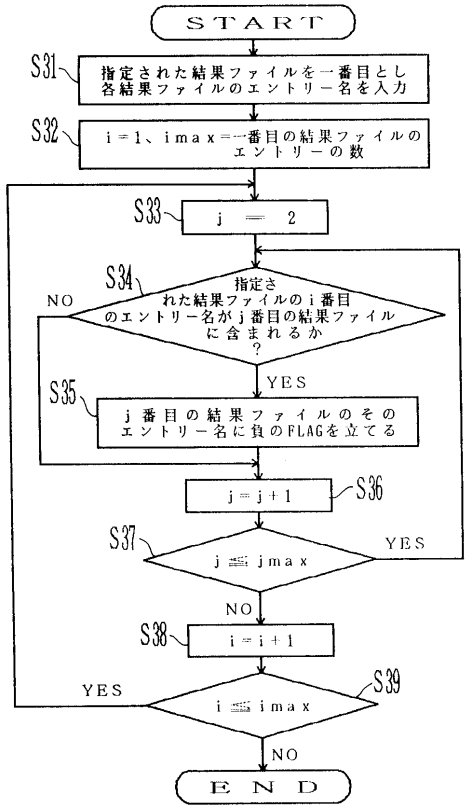

【図１８】

指定された結果ファイルに含まれない配列名の
表示処理のフローチャート



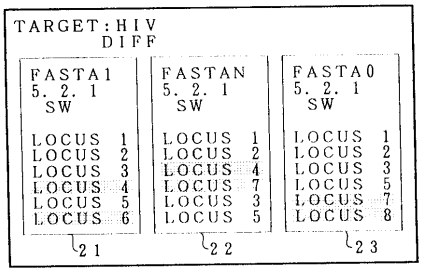
【図１９】

指定された結果ファイルに含まれない配列名の抽出処理のフローチャート



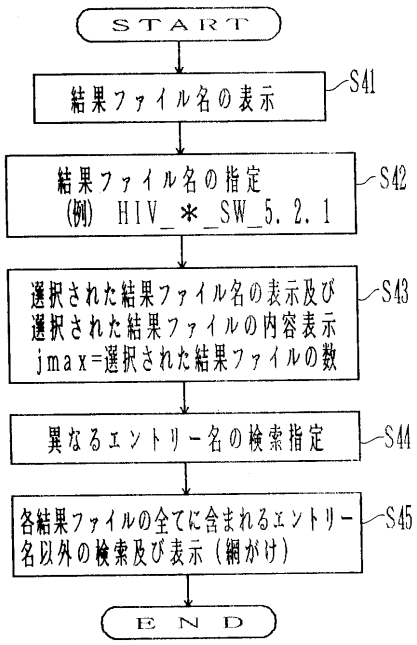

【図２０】

共通して含まれない配列名の画面表示を示す図



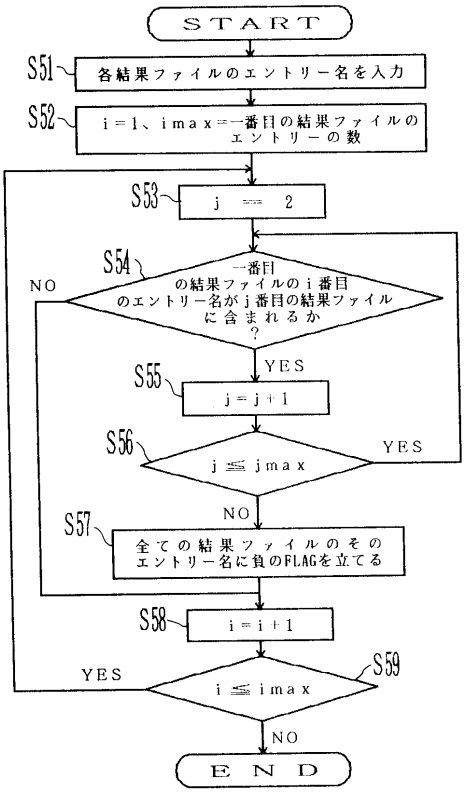

【図２１】

共通して含まれない配列名の表示処理のフローチャート



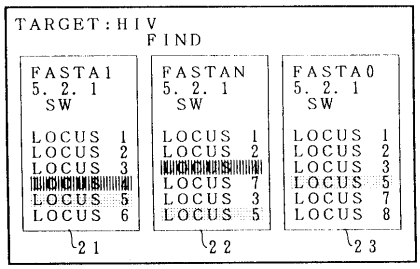

【図２２】

共通して含まれない配列名の抽出処理のフローチャート



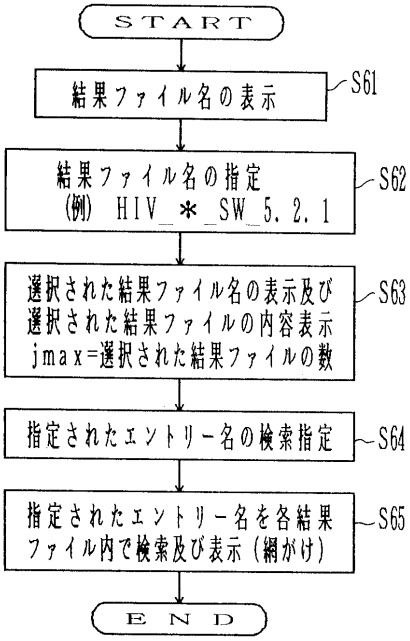
【図２３】

指定された配列名の画面表示を示す図



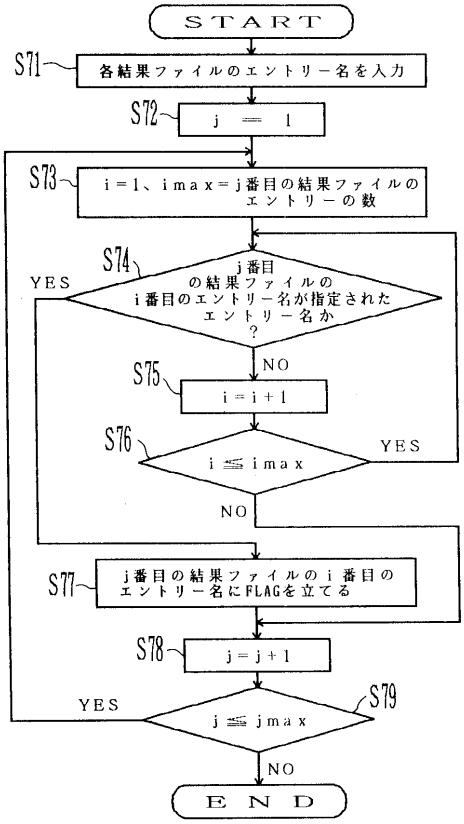

【図２４】

指定された配列名の表示処理のフローチャート



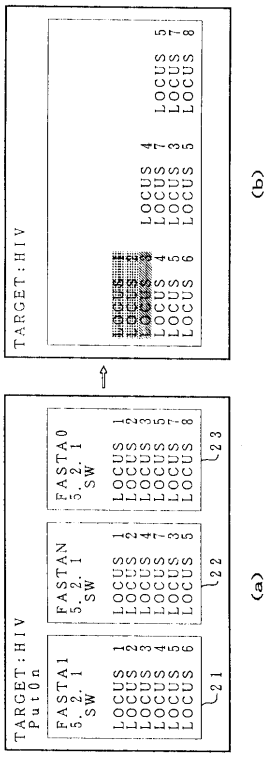

【図２５】

指定された配列名の抽出処理のフローチャート



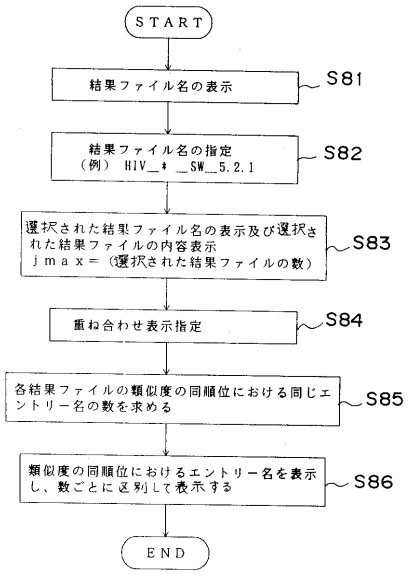

【図２６】

共通する同順位の配列名の
画面表示を示す図



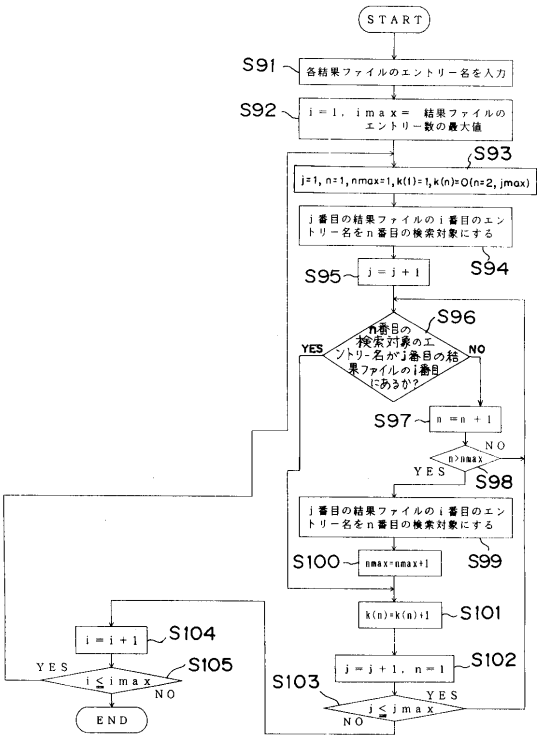
【図２７】

共通する同順位の配列名の表示処理のフローチャート



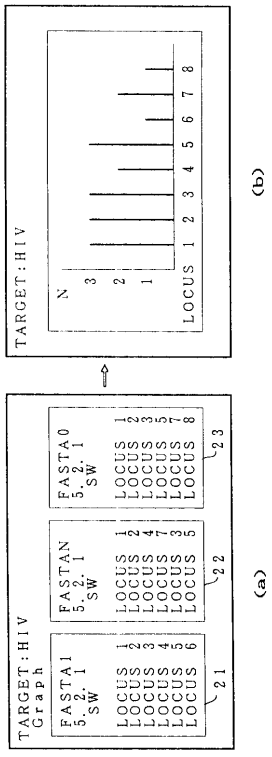

【図２８】

共通する同順位の配列名の抽出処理のフローチャート



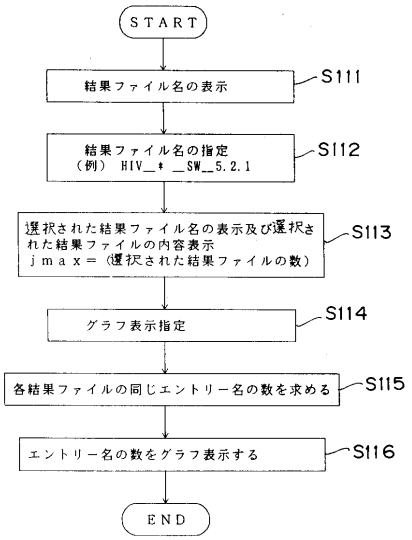

【図２９】

配列名の数のグラフ表示を示す図



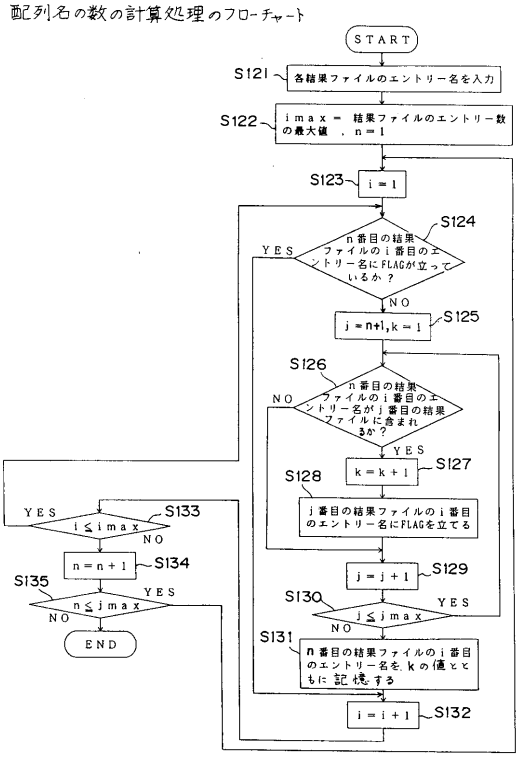

【図３０】

配列名の数のグラフ表示処理のフローチャート



【図３１】

配列名の数の計算処理のフローチャート



フロントページの続き

(56)参考文献 特開平０５－２７４３５８（ＪＰ，Ａ）
国際公開第９４／０１１８３７（WO，Ａ１）
特開平０１－０４４８５４（ＪＰ，Ａ）
特開平０５－１９２１９４（ＪＰ，Ａ）
特開平０５－２２５２３８（ＪＰ，Ａ）
特開平０５－２１６７２６（ＪＰ，Ａ）
土屋勝他３名，お楽しみディスクを２５６倍使う法，ＡＳＣＩＩ，日本，株式会社アスキー，１９９２年 9月 1日，第１６巻第9号，ｐ．２２９－２３６

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G06F 17/30