

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7619443号
(P7619443)

(45)発行日 令和7年1月22日(2025.1.22)

(24)登録日 令和7年1月14日(2025.1.14)

(51)国際特許分類	F I
G 1 6 B 40/20 (2019.01)	G 1 6 B 40/20
G 1 6 B 15/00 (2019.01)	G 1 6 B 15/00
G 1 6 B 50/30 (2019.01)	G 1 6 B 50/30

請求項の数 7 (全29頁)

(21)出願番号	特願2023-515916(P2023-515916)	(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
(86)(22)出願日	令和3年4月20日(2021.4.20)	(74)代理人	110002147 弁理士法人酒井国際特許事務所
(86)国際出願番号	PCT/JP2021/015983	(72)発明者	片岡 正弘 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
(87)国際公開番号	WO2022/224336	(72)発明者	和田 光人 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
(87)国際公開日	令和4年10月27日(2022.10.27)	(72)発明者	松村 量 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
審査請求日	令和5年9月28日(2023.9.28)		最終頁に続く

(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

(57)【特許請求の範囲】

【請求項1】

コンピュータに、

ゲノムに対応するベクトルを入力データとし、前記ゲノムを構成する複数のサブゲノムにそれぞれ対応するベクトルを正解値とする学習データを基にして、前記入力データを学習モデルに入力した際の出力が前記正解値に近づくように前記学習モデルのパラメータを調整する学習を実行し、

分析対象のゲノムを受け付けた場合、前記分析対象のゲノムを前記学習モデルに入力することで、前記分析対象のゲノムに対応する複数のサブゲノムのベクトルを算出する処理を実行させることを特徴とする情報処理プログラム。

10

【請求項2】

前記算出する処理によって、算出された複数のサブゲノムのベクトルと、代替候補となる複数の代替遺伝子ベクターのベクトルとの類似度合いに基づき、前記サブゲノムと代替可能な代替遺伝子ベクターを検索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3】

前記分析対象のゲノムは、タンパク質の二次構造を複数含み、前記分析対象のゲノムに含まれる複数の二次構造のベクトルを積算することで、前記分析対象のゲノムのベクトルを算出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。

20

【請求項 4】

コンピュータに、

ゲノムを製造する際の合成経路に含まれる複数のサブゲノムのベクトルを入力データとし、サブゲノムの構造および遺伝子ベクターの構造のうち共通する構造を示す共通構造のベクトルを正解値とする学習データに基づき、前記入力データを学習モデルに入力した際の出力が前記正解値に近づくように前記学習モデルのパラメータを調整する学習を実行し、分析対象のサブゲノムの入力を受け付けた場合に、前記分析対象のサブゲノムのベクトルを前記学習モデルに入力することで、前記分析対象のサブゲノムに対応する共通構造のベクトルを算出する

処理を実行させることを特徴とする情報処理プログラム。

10

【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、

ゲノムに対応するベクトルを入力データとし、前記ゲノムを構成する複数のサブゲノムにそれぞれ対応するベクトルを正解値とする学習データを基にして、前記入力データを学習モデルに入力した際の出力が前記正解値に近づくように前記学習モデルのパラメータを調整する学習を実行し、

分析対象のゲノムを受け付けた場合、前記分析対象のゲノムを前記学習モデルに入力することで、前記分析対象の遺伝子ベクターに対応する複数のサブベクターのベクトルを算出する

処理を実行することを特徴とする情報処理方法。

20

【請求項 6】

コンピュータが実行する情報処理方法であって、

ゲノムを製造する際の合成経路に含まれる複数のサブゲノムのベクトルを入力データとし、サブゲノムの構造および遺伝子ベクターの構造のうち共通する構造を示す共通構造のベクトルを正解値とする学習データに基づき、前記入力データを学習モデルに入力した際の出力が前記正解値に近づくように前記学習モデルのパラメータを調整する学習を実行し、分析対象のサブゲノムを受け付けた場合、前記分析対象のサブゲノムのベクトルを前記学習モデルに入力することで、前記分析対象のサブゲノムに対応する共通構造のベクトルを算出する

処理を実行することを特徴とする情報処理方法。

30

【請求項 7】

ゲノムに対応するベクトルを入力データとし、前記ゲノムを構成する複数のサブゲノムにそれぞれ対応するベクトルを正解値とする学習データを基にして、前記入力データを学習モデルに入力した際の出力が前記正解値に近づくように前記学習モデルのパラメータを調整する学習を実行する学習部と、

分析対象のゲノムを受け付けた場合、前記分析対象のゲノムを前記学習モデルに入力することで、前記分析対象のゲノムに対応する複数のサブゲノムのベクトルを算出する算出部と

を有することを特徴とする情報処理装置。

【発明の詳細な説明】

40

【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】

遺伝子導入技術の進歩と免疫機構に関する理解の深まりにより、遺伝子ベクターを用いて、遺伝子組み替え操作が行われている。挿入する遺伝子断片の大きさや挿入目的によって、様々な特徴を付加した媒体が遺伝子ベクターとして使い分けられる。それらの操作には、大腸菌や酵母、宿主生物などを由来とする遺伝子ベクターが使用される。

【0003】

50

たとえば、遺伝子改変T細胞を用いたがん免疫療法として、キメラ抗原受容体（Chimeric Antigen Receptor：CAR）導入T細胞治療法が注目されている。CARは、抗原を特異的に認識する抗体由来の部分と、TCR（T Cell Receptor）由来の細胞傷害性機能部分を結合させて人工的に作製された、がん抗原を特異的に認識し、攻撃できる受容体である。

【先行技術文献】

【特許文献】

【0004】

【文献】国際公開第2020/230240号

【文献】国際公開第2007/102578号

10

【発明の概要】

【発明が解決しようとする課題】

【0005】

遺伝子ベクターを用いて、遺伝子治療薬を開発することは非常に有望であるが、多種多様な遺伝子ベクターをそのまま用いて、遺伝子治療薬を合成することは、難しい。

【0006】

そこで、多種多様な遺伝子ベクターを代用して、目的の遺伝子治療薬を合成することが考えられるが、代用可能な遺伝子ベクターをどのように探索し、効率よく遺伝子組み換えすることが、困難であるのが現状である。

【0007】

20

1つの側面では、本発明は、目的とするゲノムに含まれるサブゲノムの代わりとなるゲノムを特定する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】

【0008】

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、ゲノムに対応するベクトルと、ゲノムを構成する複数のサブゲノムにそれぞれ対応するベクトルとの関係を定義した学習データを基にして、学習モデルの学習を実行する。コンピュータは、分析対象のゲノムを受け付けた場合、分析対象のゲノムを学習モデルに入力することで、分析対象のゲノムに対応する複数のサブゲノムのベクトルを算出する。

30

【発明の効果】

【0009】

目的とするゲノムに含まれるサブゲノムの代わりとなるゲノムを特定することができる。

【図面の簡単な説明】

【0010】

【図1】図1は、ゲノムを説明するための図である。

【図2】図2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。

【図3】図3は、タンパク質の一次構造、二次構造、三次構造、高次構造を説明するための図である。

【図4】図4は、遺伝子ベクターの一例を示す図である。

40

【図5】図5は、本実施例に係る情報処理装置の学習フェーズの処理の一例を説明するための図である。

【図6】図6は、本実施例に係る情報処理装置の分析フェーズの処理の一例を説明するための図である。

【図7】図7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8】図8は、塩基ファイルのデータ構造の一例を示す図である。

【図9】図9は、変換テーブルのデータ構造の一例を示す図である。

【図10】図10は、辞書テーブルのデータ構造の一例を示す図である。

【図11】図11は、タンパク質一次構造辞書のデータ構造の一例を示す図である。

【図12】図12は、二次構造辞書のデータ構造の一例を示す図である。

50

【図 1 3】図 1 3 は、三次構造辞書のデータ構造の一例を示す図である。

【図 1 4】図 1 4 は、高次構造辞書のデータ構造の一例を示す図である。

【図 1 5】図 1 5 は、圧縮ファイルテーブルのデータ構造の一例を示す図である。

【図 1 6】図 1 6 は、ベクトルテーブルのデータ構造の一例を示す図である。

【図 1 7】図 1 7 は、タンパク質一次構造ベクトルテーブルのデータ構造の一例を示す図である。

【図 1 8】図 1 8 は、二次構造ベクトルテーブルのデータ構造の一例を示す図である。

【図 1 9】図 1 9 は、三次構造ベクトルテーブルのデータ構造の一例を示す図である。

【図 2 0】図 2 0 は、高次構造ベクトルテーブルのデータ構造の一例を示す図である。

【図 2 1】図 2 1 は、転置インデックステーブルのデータ構造の一例を示す図である。

10

【図 2 2】図 2 2 は、タンパク質一次構造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 2 3】図 2 3 は、二次構造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 2 4】図 2 4 は、三次構造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 2 5】図 2 5 は、高次構造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 2 6】図 2 6 は、ゲノム辞書のデータ構造の一例を示す図である。

【図 2 7】図 2 7 は、本実施例に係る情報処理装置の処理手順を示すフローチャート（1）である。

【図 2 8】図 2 8 は、本実施例に係る情報処理装置の処理手順を示すフローチャート（2）である。

20

【図 2 9】図 2 9 は、本実施例 2 に係る情報処理装置の学習フェーズの処理の一例を説明するための図である。

【図 3 0】図 3 0 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である。

【図 3 1】図 3 1 は、本実施例 2 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 2】図 3 2 は、本実施例 2 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 3 3】図 3 3 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。

【発明を実施するための形態】

【0011】

30

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】

【0012】

本実施例の説明を行う前に、ゲノムについて説明する。図 1 は、ゲノムを説明するための図である。ゲノム 1 は、複数のアミノ酸が連結する順番を規定する遺伝子情報が含まれる。ここで、アミノ酸は、連続する 3 塩基、すなわちコドンによって決定される。また、ゲノム 1 には、タンパク質 1 a の情報が含まれる。タンパク質 1 a は、20 種類のアミノ酸が複数結合し、鎖状に多数が連結したものである。タンパク質 1 a の構造は、タンパク質の一次構造、二次構造、三次構造、高次（四次）構造として捉えることができる。図 1 b は、タンパク質 1 a の高次構造を示している。以下の説明では、適宜、タンパク質の一次構造、タンパク質の二次構造、タンパク質の三次構造、タンパク質の高次構造を、それぞれ、一次構造、二次構造、三次構造、高次構造と表記する。

40

【0013】

DNA および RNA 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 2 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

50

る。

【0014】

図2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。このためコドンが決まると、アミノ酸が決まるが、アミノ酸が決まっても、コドンが一意に特定されるものではない。たとえば、アミノ酸「アラニン(Ala)A」は、コドン「GCU」、「GCC」、「GCA」、または、「GCG」に対応付けられる。

【0015】

また、塩基配列により、タンパク質が一意に決定される。タンパク質の一次構造は、複数のアミノ酸の配列である。二次構造には、ヘリックスやシートなどがあり、局所的に見られる、対称的な副構造である。三次構造には、複数の二次構造が含まれる。また、高次構造には、複数の三次構造が含まれる。図3は、タンパク質の一次構造、二次構造、三次構造、高次構造を説明するための図である。たとえば、図3に示すように、高次構造Z₁には、三次構造Y₁、Y₂、Y₃等が含まれる。三次構造Y₁には、二次構造X₁、X₂、X₃等が含まれる。二次構造X₁には、一次構造W₁、W₂、W₃等が含まれる。一次構造W₁には、アミノ酸A₁、A₂、A₃等が含まれる。

10

【0016】

本実施例で利用する遺伝子ベクターは、外来遺伝物質を別の細胞に人為的に運ぶために利用されるDNAまたはRNA分子である。遺伝子ベクターには、プラスミド、コスミド、ラムダファージ、人口染色体等が存在する。図4は、遺伝子ベクターの一例を示す図である。図4に示す遺伝子ベクターは、pBR322プラスミドであり、クローニングベクターとして広く使用される。遺伝子ベクター自体は、DNAおよびRNAの塩基配列であり、たとえば、図3で説明したタンパク質の高次構造に対応するものとして説明を行う。

20

【0017】

また、遺伝子ベクターは、複数のサブベクターを合成することで生成される。サブベクターは、DNAおよびRNAの塩基配列であり、たとえば、図3で説明したタンパク質の二次構造に対応する。サブベクターには、大腸菌での維持に必要な要素を含む、いわゆる大腸菌ベクターや、酵母や植物、哺乳動物等に由来する細胞系で維持するためのベクターも含まれる。サブベクターは、その他のベクターであってもよい。

【0018】

続いて、本実施例に係る情報処理装置の処理の一例について説明する。

30

【0019】

図5は、本実施例に係る情報処理装置の学習フェーズの処理の一例を説明するための図である。図5に示すように、情報処理装置は、学習データ65を用いて、学習モデル70の機械学習を実行する。学習モデル70は、CNN(Convolutional Neural Network)やRNN(Recurrent Neural Network)等に対応する。

【0020】

学習データ65は、目的ゲノム(治療薬)のベクトルと、この目的ゲノムに含まれる複数のサブゲノムのベクトルとの関係を定義する。たとえば、目的ゲノムのベクトルが、入力データに対応し、複数のサブゲノムが、その出力データの正解値となる。

【0021】

情報処理装置は、目的ゲノムのベクトルを学習モデル70に入力した際の出力が、各サブゲノムのベクトルに近づくように、誤差逆伝播による学習を実行する。情報処理装置は、学習データ65に含まれる目的ゲノムのベクトルと、複数のサブゲノムのベクトルとの関係を基にして、上記処理を繰り返し実行することで、学習モデル70のパラメータを調整する(機械学習を実行する)。

40

【0022】

図6は、本実施例に係る情報処理装置の分析フェーズの処理の一例を説明するための図である。情報処理装置は、分析フェーズにおいて、学習フェーズで学習した学習モデル70を用いて、次の処理を実行する。

【0023】

50

情報処理装置は、目的ゲノム（治療薬）を指定した分析クエリ 80 を受け付けると、分析クエリ 80 の目的ゲノムをベクトル V_{ob80} に変換する。情報処理装置は、ベクトル V_{ob80} を、学習モデル 70 に入力することで、各サブゲノムに対応する複数のベクトル（ V_{sb80-1} 、 V_{sb80-2} 、 V_{sb80-3} 、 \dots 、 V_{sb80-n} ）を算出し、サブゲノムテーブル T_1 に格納する。

【0024】

情報処理装置は、代替遺伝子ベクターテーブル T_2 に格納された、各代替遺伝子ベクターに対応する複数のベクトル（ V_{t1} 、 V_{t2} 、 V_{t3} 、 \dots 、 V_{tn} ）と、複数のベクトル（ V_{sb80-1} 、 V_{sb80-2} 、 V_{sb80-3} 、 \dots 、 V_{sb80-n} ）との類似度を比較して、類似する代替遺伝子ベクターのベクトルを特定する。情報処理装置は、目的ゲノムのベクトルと、サブゲノムのベクトルと、類似する代替遺伝子ベクターのベクトルとを対応付けて、代替管理テーブル 85 に登録する。

10

【0025】

上記のように、本実施例に係る情報処理装置は、目的ゲノムのベクトルと、各サブゲノムのベクトルとの関係を定義した学習データ 65 を基にして、学習モデル 70 の学習を実行しておく。情報処理装置は、学習済みの学習モデル 70 に分析クエリのベクトルを入力することで、分析クエリの目的ゲノムに対応する各サブゲノムのベクトルを算出する。学習モデル 70 から出力される各サブゲノムのベクトルを用いることで、目的ゲノムに含まれるサブゲノムに類似する遺伝子ベクターであって、代替可能な遺伝子ベクターを検出することを容易に実行することができる。

20

【0026】

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 7 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 7 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。

【0027】

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。

【0028】

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。

30

【0029】

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。

【0030】

記憶部 140 は、塩基ファイル 50、変換テーブル 51、辞書テーブル 52、圧縮ファイルテーブル 53、ベクトルテーブル 54、転置インデックステーブル 55 を有する。また、記憶部 140 は、サブゲノムテーブル T_1 、代替遺伝子ベクターテーブル T_2 、ゲノム辞書 D_2 、学習データ 65、学習モデル 70、分析クエリ 80、代替管理テーブル 85 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。

40

【0031】

塩基ファイル 50 は、複数の塩基が配列された情報を保持するファイルである。図 8 は、塩基ファイルのデータ構造の一例を示す図である。図 8 に示すように、塩基ファイル 50 は、4 種類で、「A」、「G」、「C」、「T」または「U」の記号で示される。

【0032】

変換テーブル 51 は、コドンと、コドンの符号とを対応付けるテーブルである。3 つの

50

塩基配列のかたまりは「コドン」と呼ばれる。図9は、変換テーブルのデータ構造の一例を示す図である。図9に示すように、各コドンと、各符号とが対応付けられる。たとえば、コドン「UUU」の符号は「40h(01000000)」となる。「h」は16進数を示すものである。

【0033】

辞書テーブル52は、各種の辞書を保持するテーブルである。図10は、辞書テーブルのデータ構造の一例を示す図である。図10に示すように、この辞書テーブル52は、タンパク質一次構造辞書D1-1、二次構造辞書D1-2、三次構造辞書D1-3、高次構造辞書D1-4を有する。

【0034】

タンパク質一次構造辞書D1-1は、タンパク質の圧縮符号とタンパク質を構成するコドンの配列との関係を定義する辞書データである。図11は、タンパク質一次構造辞書のデータ構造の一例を示す図である。図11に示すように、タンパク質一次構造辞書D1-1は、圧縮符号と、名称と、コドン符号配列とを対応付ける。圧縮符号は、コドンの圧縮符号配列（またはアミノ酸の記号配列）である。名称は、タンパク質の名称である。コドン符号配列は、コドンの圧縮符号の配列である。なお、コドン符号配列の代わりに、アミノ酸の記号の配列を、タンパク質一次構造の圧縮符号と対応付けてもよい。

【0035】

たとえば、タンパク質一次構造「1型コラーゲン」には、圧縮符号「C0008000h」が割り当てられる。圧縮符号「C0008000h」に対応するコドン符号配列は「02h63h78h・・・03h」となる。

【0036】

二次構造辞書D1-2は、タンパク質一次構造の圧縮符号の配列と、二次構造の圧縮符号との関係を定義する辞書データである。図12は、二次構造辞書のデータ構造の一例を示す図である。図12に示すように、二次構造辞書D1-2は、圧縮符号と、名称と、タンパク質一次構造符号配列とを対応付ける。圧縮符号は、タンパク質の二次構造に割り当てられた圧縮符号である。名称は、二次構造の名称である。タンパク質一次構造符号配列は、二次構造に対応するタンパク質一次構造の圧縮符号の配列である。

【0037】

たとえば、二次構造「二次構造」には、圧縮符号「D0000000h」が割り当てられる。圧縮符号「D0000000h」に対応するタンパク質一次構造符号配列は「C0008001hC00・・・」となる。

【0038】

三次構造辞書D1-3は、二次構造の圧縮符号の配列と、三次構造の圧縮符号との関係を定義する辞書データである。図13は、三次構造辞書のデータ構造の一例を示す図である。図13に示すように、三次構造辞書D1-3は、圧縮符号と、名称と、二次構造符号配列とを対応付ける。圧縮符号は、三次構造に割り当てられた圧縮符号である。名称は、三次構造の名称である。二次構造符号配列は、三次構造に対応する二次構造の圧縮符号の配列である。

【0039】

たとえば、三次構造「三次構造」には、圧縮符号「E0000000h」が割り当てられる。圧縮符号「E0000000h」に対応する二次構造符号配列は「D0008031hD00・・・」となる。

【0040】

高次構造辞書D1-4は、三次構造の圧縮符号の配列と、高次構造の圧縮符号との関係を定義する辞書データである。図14は、高次構造辞書のデータ構造の一例を示す図である。図14に示すように、高次構造辞書D1-4は、圧縮符号と、名称と、三次構造符号配列とを対応付ける。圧縮符号は、高次構造に割り当てられた圧縮符号である。名称は、高次構造の名称である。三次構造符号配列は、高次構造に対応する三次構造の圧縮符号の配列である。

10

20

30

40

50

この三次構造の圧縮符号に割り当てられたベクトルとが対応付けられる。三次構造のベクトルは、三次構造に含まれる二次構造のベクトルを積算することにより算出される。

【 0 0 5 2 】

高次構造ベクトルテーブル V T 1 - 4 は、高次構造に対応するベクトルを保持するテーブルである。図 2 0 は、高次構造ベクトルテーブルのデータ構造の一例を示す図である。図 2 0 に示すように、高次構造ベクトルテーブル V T 1 - 4 は、高次構造の圧縮符号と、この高次構造の圧縮符号に割り当てられたベクトルとが対応付けられる。高次構造のベクトルは、高次構造に含まれる三次構造のベクトルを積算することにより算出される。

【 0 0 5 3 】

図 7 の説明に戻る。転置インデックステーブル 5 5 は、各種の転置インデックスを保持するテーブルである。図 2 1 は、転置インデックステーブルのデータ構造の一例を示す図である。図 2 1 に示すように、転置インデックステーブル 5 5 は、タンパク質一次構造転置インデックス I n 1 - 1、二次構造転置インデックス I n 1 - 2、三次構造転置インデックス I n 1 - 3、高次構造転置インデックス I n 1 - 4 を有する。

10

【 0 0 5 4 】

図 2 2 は、タンパク質一次構造転置インデックスのデータ構造の一例を示す図である。タンパク質一次構造転置インデックス I n 1 - 1 の横軸は、オフセットに対応する軸である。タンパク質一次構造転置インデックス I n 1 - 1 の縦軸は、タンパク質一次構造の圧縮符号に対応する軸である。タンパク質一次構造転置インデックス I n 1 - 1 は、「 0 」または「 1 」のビットマップで示され、初期状態では全てのビットマップが「 0 」に設定される。

20

【 0 0 5 5 】

たとえば、タンパク質一次構造圧縮ファイル 5 3 B の先頭のタンパク質一次構造の圧縮符号のオフセットを「 0 」とする。タンパク質一次構造圧縮ファイル 5 3 B の先頭から 8 番目の位置に、タンパク質一次構造の符号「 C 0 0 0 8 0 0 0 h (一型コラーゲン) 」が含まれる場合、タンパク質転置インデックス I n 1 - 1 のオフセット「 7 」の列と、タンパク質の符号「 C 0 0 0 8 0 0 0 h (一型コラーゲン) 」の行とが交差する位置のビットが「 1 」となる。

【 0 0 5 6 】

図 2 3 は、二次構造転置インデックスのデータ構造の一例を示す図である。二次構造転置インデックス I n 1 - 2 の横軸は、オフセットに対応する軸である。二次構造転置インデックス I n 1 - 2 の縦軸は、二次構造の圧縮符号に対応する軸である。二次構造転置インデックス I n 1 - 2 は、「 0 」または「 1 」のビットマップで示され、初期状態では全てのビットマップが「 0 」に設定される。

30

【 0 0 5 7 】

たとえば、二次構造圧縮ファイル 5 3 C の先頭の二次構造の圧縮符号のオフセットを「 0 」とする。二次構造圧縮ファイル 5 3 C の先頭から 8 番目の位置に、二次構造の符号「 D 0 0 0 0 0 0 h (二次構造) 」が含まれる場合、二次構造転置インデックス I n 1 - 2 のオフセット「 7 」の列と、二次構造の圧縮符号「 D 0 0 0 0 0 0 h (二次構造) 」の行とが交差する位置のビットが「 1 」となる。

40

【 0 0 5 8 】

図 2 4 は、三次構造転置インデックスのデータ構造の一例を示す図である。三次構造転置インデックス I n 1 - 3 の横軸は、オフセットに対応する軸である。三次構造転置インデックス I n 1 - 3 の縦軸は、三次構造の圧縮符号に対応する軸である。三次構造転置インデックス I n 1 - 3 は、「 0 」または「 1 」のビットマップで示され、初期状態では全てのビットマップが「 0 」に設定される。

【 0 0 5 9 】

たとえば、三次構造圧縮ファイル 5 3 D の先頭の三次構造の圧縮符号のオフセットを「 0 」とする。三次構造圧縮ファイル 5 3 D の先頭から 1 1 番目の位置に、三次構造の符号「 E 0 0 0 0 0 0 0 h (三次構造) 」が含まれる場合、三次構造転置インデックス I

50

n 1 - 3 のオフセット「1 0」の列と、三次構造の圧縮符号「E 0 0 0 0 0 0 0 h (三次構造)」の行とが交差する位置のビットが「1」となる。

【0 0 6 0】

図 2 5 は、高次構造転置インデックスのデータ構造の一例を示す図である。高次構造転置インデックスのデータ構造の一例を示す図である。高次構造転置インデックス I n 1 - 4 の横軸は、オフセットに対応する軸である。高次構造転置インデックス I n 1 - 4 の縦軸は、高次構造の圧縮符号に対応する軸である。高次構造転置インデックス I n 1 - 4 は、「0」または「1」のビットマップで示され、初期状態では全てのビットマップが「0」に設定される。

【0 0 6 1】

たとえば、高次構造圧縮ファイル 5 3 E の先頭の高次構造の圧縮符号のオフセットを「0」とする。高次構造圧縮ファイル 5 3 E の先頭から 1 1 番目の位置に、高次構造の符号「F 0 0 0 0 0 0 0 h (高次構造)」が含まれる場合、高次構造転置インデックス I n 1 - 4 のオフセット「1 0」の列と、高次構造の圧縮符号「F 0 0 0 0 0 0 0 h (高次構造)」の行とが交差する位置のビットが「1」となる。

【0 0 6 2】

図 7 の説明に戻る。代替遺伝子ベクターテーブル T 2 は、複数の遺伝子ベクターのベクトルを保持する。遺伝子ベクターは、タンパク質の二次構造に対応する。たとえば、代替遺伝子ベクターテーブル T 2 に格納するベクトルは、二次元構造ベクトルテーブル V T 1 - 2 に登録されたベクトルであってもよい。代替遺伝子ベクターテーブル T 2 のデータ構造は、図 6 で説明したように、複数の代替遺伝子ベクターのベクトルが格納される。

【0 0 6 3】

ゲノム辞書 D 2 は、目的ゲノムの名称と、この目的ゲノムに含まれるサブゲノムの名称との関係を定義する。図 2 6 は、ゲノム辞書のデータ構造の一例を示す図である。図 2 6 に示すように、このゲノム辞書 D 2 は、目的ベクターの名称と、複数のサブゲノムの名称とを対応付ける。

【0 0 6 4】

学習データ 6 5 は、目的ゲノムのベクトルと、この目的ゲノムに含まれる複数のサブゲノムのベクトルとの関係を定義する。学習データ 6 5 のデータ構造は、図 5 で説明した学習データのデータ構造に対応する。

【0 0 6 5】

学習モデル 7 0 は、C N N や R N N 等に対応するモデルであり、パラメータが設定される。

【0 0 6 6】

分析クエリ 8 0 は、分析対象となる目的ゲノム（治療薬）の情報を含む。たとえば、目的ゲノムの情報には、高次構造に対応する塩基配列の情報が含まれる。

【0 0 6 7】

代替管理テーブル 8 5 は、目的ゲノムに含まれるサブゲノムのベクトルと、このサブゲノムに類似する遺伝子ベクターであって、代替可能な遺伝子ベクターのベクトルとを対応付けて保持するテーブルである。

【0 0 6 8】

制御部 1 5 0 は、前処理部 1 5 1、学習部 1 5 2、算出部 1 5 3、分析部 1 5 4 を有する。制御部 1 5 0 は、例えば、C P U (Central Processing Unit) や M P U (Micro Processing Unit) により実現される。また、制御部 1 5 0 は、例えば A S I C (Application Specific Integrated Circuit) や F P G A (Field Programmable Gate Array) 等の集積回路により実行されてもよい。

【0 0 6 9】

前処理部 1 5 1 は、下記の各種の処理を実行することで、目的ゲノム（治療薬）に相当する高次構造、または三次構造のベクトルと、サブゲノムに相当する二次構造のベクトル等を算出する。

10

20

30

40

50

【 0 0 7 0 】

まず、前処理部 1 5 1 は、コドン圧縮ファイル 5 3 A を生成する処理、タンパク質一次構造圧縮ファイル 5 3 B を生成する処理、タンパク質一次構造ベクトルテーブル V T 1 - 1、タンパク質一次構造転置インデックス I n 1 - 1 を生成する処理を実行する。

【 0 0 7 1 】

前処理部 1 5 1 は、塩基ファイル 5 0 と、変換テーブル 5 1 とを比較して、塩基ファイル 5 0 の塩基配列を、コドンの単位で圧縮符号を割り当て、コドン圧縮ファイル 5 3 A を生成する。

【 0 0 7 2 】

前処理部 1 5 1 は、コドン圧縮ファイル 5 3 A と、タンパク質一次構造辞書 D 1 - 1 とを比較して、コドン圧縮ファイル 5 3 A に含まれるコドンの圧縮符号の配列を、タンパク質一次構造の単位で圧縮符号を割り当て、タンパク質圧縮ファイル 5 3 B を生成する。

10

【 0 0 7 3 】

前処理部 1 5 1 は、タンパク質一次構造圧縮ファイル 5 3 B を生成すると、タンパク質一次構造の圧縮符号を、ポアンカレ空間に埋め込むことで、タンパク質一次構造（タンパク質一次構造の圧縮符号）のベクトルを算出する。ポアンカレ空間に埋め込んで、ベクトルを算出する処理は、ポアンカレエンベディング（Poincare Embeddings）と呼ばれる技術である。ポアンカレエンベディングは、たとえば、非特許文献「Valentin Khruikov1 et al. 「Hyperbolic Image Embeddings」 Cornell University, 2019 April 3」等に記載された技術を用いればよい。

20

【 0 0 7 4 】

ポアンカレエンベディングでは、ポアンカレ空間に埋め込まれた位置に応じて、ベクトルが割り当てられるものであり、また、類似する情報ほど、近い位置に埋め込まれるという特徴がある。このため、類似の特徴を有する各基は、ポアンカレ空間において、それぞれ近い位置に埋め込まれるため、類似のベクトルが割り当てられる。図示を省略するが、前処理部 1 5 1 は、類似するタンパク質一次構造同士を定義したタンパク質一次構造類似テーブルを参照して、各タンパク質一次構造の圧縮符号をポアンカレ空間に埋め込み、各タンパク質一次構造の圧縮符号のベクトルを算出する。前処理部 1 5 1 は、タンパク質辞書一次構造 D 1 - 1 に定義された各タンパク質一次構造の圧縮符号に対して、ポアンカレエンベディングを事前に実行しておいてもよい。

30

【 0 0 7 5 】

前処理部 1 5 1 は、タンパク質一次構造（タンパク質一次構造の圧縮符号）と、タンパク質一次構造のベクトルとを対応付けることで、タンパク質一次構造ベクトルテーブル V T 1 - 1 を生成する。前処理部 1 5 1 は、タンパク質一次構造のベクトルと、タンパク質一次構造圧縮ファイル 5 3 B におけるタンパク質一次構造（タンパク質一次構造の圧縮符号）の位置との関係を基にして、タンパク質一次構造転置インデックス I n 1 - 1 を生成する。

【 0 0 7 6 】

続いて、前処理部 1 5 1 は、二次構造圧縮ファイル 5 3 C を生成する処理、二次構造ベクトルテーブル V T 1 - 2、二次構造転置インデックス I n 1 - 2 を生成する処理を実行する。

40

【 0 0 7 7 】

前処理部 1 5 1 は、タンパク質一次構造圧縮ファイル 5 3 B と、二次構造辞書 D 1 - 2 とを比較して、タンパク質一次構造圧縮ファイル 5 3 B に含まれるタンパク質一次構造の圧縮符号の配列を、二次構造の単位で圧縮符号を割り当て、二次構造圧縮ファイル 5 3 C を生成する。

【 0 0 7 8 】

前処理部 1 5 1 は、二次構造辞書 D 1 - 2 を参照して、二次構造の圧縮符号に対応するタンパク質一次構造符号配列（タンパク質一次構造の圧縮符号の配列）を特定する。前処理部 1 5 1 は、特定した各タンパク質一次構造の圧縮符号のベクトルを、タンパク質一次

50

構造ベクトルテーブルV T 1 - 1 から取得し、取得した各ベクトルを加算することで、二次構造の圧縮符号のベクトルを算出する。前処理部 1 5 1 は、上記処理を繰り返し実行することで、各二次構造のベクトルを算出する。

【 0 0 7 9 】

前処理部 1 5 1 は、二次構造（二次構造の圧縮符号）と、二次構造のベクトルとを対応付けることで、二次構造ベクトルテーブルV T 1 - 2 を生成する。前処理部 1 5 1 は、二次構造のベクトルと、二次構造圧縮ファイル 5 3 C における二次構造（二次構造の圧縮符号）の位置との関係を基にして、二次構造転置インデックス I n 1 - 2 を生成する。

【 0 0 8 0 】

続いて、前処理部 1 5 1 は、三次構造圧縮ファイル 5 3 D を生成する処理、三次構造ベクトルテーブルV T 1 - 3、三次構造転置インデックス I n 1 - 3 を生成する処理を実行する。

10

【 0 0 8 1 】

前処理部 1 5 1 は、二次構造圧縮ファイル 5 3 C と、三次構造辞書 D 1 - 3 とを比較して、二次構造圧縮ファイル 5 3 C に含まれる二次構造の圧縮符号の配列を、三次構造の単位で圧縮符号を割り当て、三次構造圧縮ファイル 5 3 D を生成する。

【 0 0 8 2 】

前処理部 1 5 1 は、三次構造辞書 D 1 - 3 を参照して、三次構造の圧縮符号に対応する二次構造符号配列（二次構造の圧縮符号の配列）を特定する。前処理部 1 5 1 は、特定した各二次構造の圧縮符号のベクトルを、二次構造ベクトルテーブルV T 1 - 2 から取得し、取得した各ベクトルを加算することで、三次構造の圧縮符号のベクトルを算出する。前処理部 1 5 1 は、上記処理を繰り返し実行することで、各三次構造のベクトルを算出する。

20

【 0 0 8 3 】

前処理部 1 5 1 は、三次構造（三次構造の圧縮符号）と、三次構造のベクトルとを対応付けることで、三次構造ベクトルテーブルV T 1 - 3 を生成する。前処理部 1 5 1 は、三次構造のベクトルと、三次構造圧縮ファイル 5 3 D における三次構造（三次構造の圧縮符号）の位置との関係を基にして、三次構造転置インデックス I n 1 - 3 を生成する。

【 0 0 8 4 】

続いて、前処理部 1 5 1 は、高次構造圧縮ファイル 5 3 E を生成する処理、高次構造ベクトルテーブルV T 1 - 4、高次構造転置インデックス I n 1 - 4 を生成する処理を実行する。

30

【 0 0 8 5 】

前処理部 1 5 1 は、三次構造圧縮ファイル 5 3 D と、高次構造辞書 D 1 - 4 とを比較して、三次構造圧縮ファイル 5 3 D に含まれる三次構造の圧縮符号の配列を、高次構造の単位で圧縮符号を割り当て、高次構造圧縮ファイル 5 3 E を生成する。

【 0 0 8 6 】

前処理部 1 5 1 は、高次構造辞書 D 1 - 4 を参照して、高次構造の圧縮符号に対応する三次構造符号配列（三次構造の圧縮符号の配列）を特定する。前処理部 1 5 1 は、特定した各三次構造の圧縮符号のベクトルを、三次構造ベクトルテーブルV T 1 - 3 から取得し、取得した各ベクトルを加算することで、高次構造の圧縮符号のベクトルを算出する。前処理部 1 5 1 は、上記処理を繰り返し実行することで、各高次構造のベクトルを算出する。

40

【 0 0 8 7 】

前処理部 1 5 1 は、高次構造（高次構造の圧縮符号）と、高次構造のベクトルとを対応付けることで、高次構造ベクトルテーブルV T 1 - 4 を生成する。前処理部 1 5 1 は、高次構造のベクトルと、高次構造圧縮ファイル 5 3 E における高次構造（高次構造の圧縮符号）の位置との関係を基にして、高次構造転置インデックス I n 1 - 4 を生成する。

【 0 0 8 8 】

前処理部 1 5 1 が、代替遺伝子ベクターテーブルT 2 を生成する処理の一例について説明する。たとえば、前処理部 1 5 1 は、二次構造ベクトルテーブルV T 1 - 2 に含まれる三次構造のベクトルを、そのまま、代替遺伝子ベクターテーブルT 2 に設定する。なお、

50

前処理部 151 は、入力部 120 を介してベクトルの指定を受け付けた場合には、指定されたベクトルを、代替遺伝子ゲノムテーブル T2 に設定してもよい。

【0089】

前処理部 151 が、学習データ 65 を生成する処理の一例について説明する。前処理部 151 は、ゲノム辞書 D2 を基にして、目的ゲノムの名称と、サブゲノムの名称との関係特定する。前処理部 151 は、高次構造辞書 D1 - 4 と高次構造ベクトルテーブル VT1 - 4、または、三次構造辞書 D1 - 4 と三次構造ベクトルテーブル VT1 - 3、および目的ゲノムの名称とを基にして、目的ゲノムのベクトルを特定する。前処理部 151 は、二次構造辞書 D1 - 2 および二次構造ベクトルテーブル VT1 - 2 と、サブゲノムの名称とを基にして、サブゲノムのベクトルを特定する。前処理部 151 は、かかる処理によつて、目的ゲノムと、サブゲノムとの関係を特定し、学習データ 65 に登録する。

10

【0090】

前処理部 151 は、上記処理を繰り返し実行することで、学習データ 65 を生成する。なお、情報処理装置 100 は、外部装置等から、作成済みの学習データ 65 を取得して利用してもよい。

【0091】

図 7 の説明に戻る。学習部 152 は、学習データ 65 を用いて、学習モデル 70 の学習を実行する。学習部 152 の処理は、図 5 で説明した処理に対応する。学習部 152 は、学習データ 65 から、目的ゲノム（治療薬）のベクトルと、この目的ゲノムのベクトルに対応する各サブゲノムのベクトルとの組を取得する。学習部 152 は、目的ゲノムのベクトルを、学習モデル 70 に入力した場合の、学習モデル 70 の出力の値が、各サブゲノムのベクトルの値に近づくように、誤差逆伝播による学習を実行することで、学習モデル 70 のパラメータを調整する。

20

【0092】

学習部 152 は、学習データ 65 の目的ゲノムのベクトルと、各サブゲノムのベクトルとの組について、上記処理を繰り返し実行することで、学習モデル 70 の学習を実行する。

【0093】

算出部 153 は、分析クエリ 80 の指定を受け付けた場合に、学習済みの学習モデル 70 を用いて、分析クエリ 80 の目的ゲノムに含まれる各サブゲノムのベクトルを算出する。算出部 153 の処理は、図 6 で説明した処理に対応する。算出部 153 は、分析クエリ 80 を、入力部 120 から受け付けてもよいし、通信部 110 を介して、外部装置から受け付けてもよい。

30

【0094】

算出部 153 は、分析クエリ 80 に含まれる目的ゲノムの塩基配列を取得する。算出部 153 は、目的ゲノムの塩基配列と、変換テーブル 51 とを比較して、目的ゲノムの塩基配列に含まれるコドン特定し、コドンの単位で、目的ゲノムの塩基配列をそれぞれ圧縮符号に変換する。また、算出部 153 は、コドンの単位で圧縮したコドン符号配列と、タンパク質一次構造辞書 D1 - 1 とを比較して、タンパク質一次構造の単位で、コドン符号配列をそれぞれ圧縮符号に変換する。

【0095】

算出部 153 は、変換した各タンパク質一次構造の圧縮符号と、タンパク質一次構造ベクトルテーブル VT1 - 1 とを比較して、各タンパク質一次構造の圧縮符号のベクトルを特定する。算出部 153 は、特定した各タンパク質一次構造の圧縮符号のベクトルを積算することで、分析クエリ 80 に含まれる目的ゲノムに対応するベクトル Vob80 を算出する。

40

【0096】

なお、目的ゲノムが、複数のサブゲノムの二次構造によって指定されている場合には、算出部 153 は、次の処理を実行する。算出部 153 は、目的ゲノムのサブゲノムの各二次構造と、二次構造辞書 D1 - 2 および二次構造ベクトルテーブル VT1 - 2 とを比較して、目的ゲノムに含まれるサブゲノムの二次構造のベクトルを特定する。算出部 153 は

50

、特定したサブゲノムの二次構造のベクトルを積算することで、目的ゲノムのベクトルを算出する。

【0097】

算出部153は、ベクトルVob80を、学習モデル70に入力することで、各サブゲノムに対応する複数のベクトルを算出する。算出部153は、算出した各サブゲノムのベクトルを、分析部154に出力する。以下の説明では、算出部153が算出した各サブゲノムのベクトルを、それぞれ「分析ベクトル」と表記する。算出部153は、各サブゲノムのベクトル（分析ベクトル）は、サブゲノムテーブルT1に格納する。

【0098】

分析部154は、分析ベクトルを基にして、分析ベクトルに類似するベクトルを有する代替遺伝子ベクターの情報を検索する。分析部154は、検索結果を基にして、目的ゲノムに含まれる各サブゲノムのベクトルと類似した各代替遺伝子ベクターのベクトル（以下に示す類似ベクトル）とを対応付けて、代替管理テーブル85に登録する。

10

【0099】

たとえば、分析部154は、分析ベクトルと、代替遺伝子ベクターテーブルT2に含まれる各ベクトルとの距離をそれぞれ算出し、分析ベクトルとの距離が閾値未満となるベクトルを特定する。代替遺伝子ベクターテーブルT2に含まれるベクトルであって、分析ベクトルとの距離が閾値未満となるベクトルが「類似ベクトル」となる。この類似ベクトルに対応する遺伝子ベクターが、代替可能な遺伝子ベクターとなる。

【0100】

分析部154は、二次構造ベクトルテーブルVT1-2を基にして、類似ベクトルに対応する遺伝子ベクターの圧縮符号を特定し、特定した遺伝子ベクターの圧縮符号、二次構造辞書D1-2、タンパク質一次構造辞書D1-1を基にして遺伝子ベクターに含まれるタンパク質一次構造を特定してもよい。分析部154は、かかる処理を実行することで、類似ベクトルに対応する代替可能な遺伝子ベクターの特徴を検索し、代替管理テーブル85に登録する。代替可能な遺伝子ベクターの特徴は、遺伝子ベクターに含まれるタンパク質、タンパク質の一次構造となる。

20

【0101】

分析部154は、各分析ベクトルについて、上記処理を繰り返し実行することで、分析ベクトル毎に、類似ベクトルに対応する遺伝子ベクターの特徴を検索し、代替管理テーブル85に登録してもよい。分析部154は、代替管理テーブル85を、表示部130に出力して表示させてもよいし、ネットワークに接続された外部装置に送信してもよい。

30

【0102】

次に、本実施例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27は、本実施例に係る情報処理装置の処理手順を示すフローチャート(1)である。図27に示すように、情報処理装置100の前処理部151は、ポアンカレエンベディングを実行することで、各タンパク質の圧縮符号のベクトルを算出する(ステップS101)。

【0103】

前処理部151は、塩基ファイル50、変換テーブル51、辞書テーブル52を基にして、圧縮ファイルテーブル53、ベクトルテーブル54、転置インデックステーブル55を生成する(ステップS102)。

40

【0104】

前処理部151は、学習データ65を生成する(ステップS103)。情報処理装置100の学習部152は、学習データ65を基にして、学習モデル70の学習を実行する(ステップS104)。

【0105】

図28は、本実施例に係る情報処理装置の処理手順を示すフローチャート(2)である。情報処理装置100の算出部153は、分析クエリ80を受け付ける(ステップS201)。算出部153は、分析クエリ80(目的ゲノム)のベクトルを算出する(ステップS202)。

50

【0106】

算出部153は、算出した分析クエリ80のベクトルを、学習済みの学習モデル70に入力することで、各サブゲノムのベクトルを算出する(ステップS203)。情報処理装置100の分析部154は、各サブゲノムのベクトルと代替遺伝子ベクターテーブルT2のベクトルとを比較する(ステップS204)。

【0107】

分析部154は、各サブゲノムに対応する代替可能な遺伝子ベクターを検索する(ステップS205)。分析部154は、検索結果を、代替管理テーブル85に登録する(ステップS206)。

【0108】

次に、本実施例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、学習フェーズで、目的ゲノム(治療薬)のベクトルと、サブゲノムのベクトルとの関係を定義した学習データ65を基にして、学習モデル70の学習を実行しておく。情報処理装置100は、分析フェーズで、学習済みの学習モデル70に分析クエリのベクトルを入力することで、分析クエリ(目的ゲノム)に対応する各サブゲノムのベクトルを算出する。学習モデル70から出力される各サブゲノムのベクトルを用いることで、目的ゲノムに含まれるサブゲノムに類似する代替可能な遺伝子ベクターを検出することを容易に実行することができる。

【0109】

たとえば、目的ゲノムに含まれるサブゲノムが、稀少なサブゲノムである場合には、情報処理装置100の処理を実行することで、かかるサブゲノムの代わりとなる安価な遺伝子ベクターを容易に探索することができる。

【0110】

ところで、上述した実施例では、サブゲノム(二次構造)の粒度で比較を行い、代替可能な遺伝子ベクターを探索していたが、これに限定されるものではない。たとえば、情報処理装置100は、サブゲノムを構成する複数の一次構造の粒度で比較を行い、代替可能な一次構造を探索してもよい。

【実施例2】

【0111】

続いて、実施例2について説明する。図29は、本実施例2に係る情報処理装置の学習フェーズの処理の一例を説明するための図である。図29に示すように、情報処理装置は、学習データ90を用いて、学習モデル91の学習を実行する。学習モデル91は、CNNやRNN等に対応する。

【0112】

学習データ90は、目的ゲノム(治療薬)を合成する複数のサブゲノムのベクトルと、遺伝子ベクターを基に遺伝子組み換えで保持される共通構造のベクトルとの関係を定義する。たとえば、サブゲノムのベクトルが、入力データに対応し、複数の共通構造のベクトルが正解値となる。

【0113】

情報処理装置は、サブゲノムのベクトルを学習モデル91に入力した際の出力が、各共通構造のベクトルに近づくように、誤差逆伝播による学習を実行する。情報処理装置は、学習データ90に含まれるサブゲノムのベクトルと、共通構造のベクトルとの関係を基にして、上記処理を繰り返し実行することで、学習モデル91のパラメータを調整する(機械学習を実行する)。

【0114】

図30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である。実施例2に係る情報処理装置は、実施例1の情報処理装置100と同様にして、学習モデル90を学習しておいてもよい。また、情報処理装置は、図29で説明したように、学習モデル70とは別の、学習モデル91を学習する。学習モデル91は、分析クエリ(サブゲノム)92のベクトルが入力された場合に、共通構造のベクトルを出力する。

10

20

30

40

50

【 0 1 1 5 】

情報処理装置は、サブゲノムを指定した分析クエリ 9 2 を受け付けると、サブゲノムベクトルテーブル T 1 を用いて、分析クエリ 9 2 のサブゲノムをベクトル V_{sb92-1} に変換する。情報処理装置は、サブゲノムのベクトル V_{sb92-1} を、学習モデル 9 1 に入力することで、共通構造に対応するベクトル V_{cm92-1} を算出する。

【 0 1 1 6 】

ここで、情報処理装置は、サブゲノムのベクトル V_{sb92-1} と、代替遺伝子ベクターテーブル T 2 に含まれる複数の遺伝子ベクターのベクトルとを比較する。代替遺伝子ベクターベクトルテーブル T 2 は、実施例 1 で説明した代替遺伝子ベクターベクトルテーブル T 2 に対応する。

10

【 0 1 1 7 】

情報処理装置は、サブゲノムのベクトル V_{sb92-1} について、類似する遺伝子ベクターのベクトルを特定する。たとえば、サブゲノムのベクトル V_{sb92-1} と類似する遺伝子ベクターのベクトルを V_{t92-1} とする。そうすると、ベクトル V_{sb92-1} のサブゲノムと、ベクトル V_{t92-1} の遺伝子ベクターとで共通する共通構造のベクトルが、学習モデル 9 1 から出力されるベクトル V_{cm92-1} となることわかる。また、遺伝子ベクターのベクトル V_{t92-1} から、共通構造のベクトル V_{cm92-1} を減算した結果が、類似する遺伝子ベクターとサブゲノムとで相違する「遺伝子組換構造」のベクトルとなる。

【 0 1 1 8 】

情報処理装置は、共通構造のベクトルと、遺伝子組換構造とのベクトルとの関係を、共通構造・遺伝子組換構造テーブル 9 3 に登録する。情報処理装置は、各サブゲノムのベクトルについて、上記処理を繰り返し実行することで、共通構造・遺伝子組換構造テーブル 9 3 を生成する。

20

【 0 1 1 9 】

上記のように、本実施例 2 に係る情報処理装置は、学習済みの学習モデル 9 1 に分析クエリ 9 2 のベクトルを入力し、分析クエリのサブゲノムに対応する各共通構造のベクトルを算出する。また、サブゲノムに類似する遺伝子ベクターの各ベクトルから、共通構造のベクトルを減算することで、類似するサブゲノムと、遺伝子ベクターとで相違する遺伝子組換構造のベクトルを算出する。上記の共通構造のベクトルや、遺伝子組換構造のベクトルを用いることで、目的ゲノムの合成や製造に使用可能なより良い遺伝子ベクターを容易に分析することができる。

30

【 0 1 2 0 】

次に、本実施例 2 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 1 は、本実施例 2 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 1 に示すように、この情報処理装置 2 0 0 は、通信部 2 1 0、入力部 2 2 0、表示部 2 3 0、記憶部 2 4 0、制御部 2 5 0 を有する。

【 0 1 2 1 】

通信部 2 1 0、入力部 2 2 0、表示部 2 3 0 に関する説明は、実施例 1 で説明した通信部 1 1 0、入力部 1 2 0、表示部 1 3 0 に関する説明と同様である。

40

【 0 1 2 2 】

記憶部 2 4 0 は、塩基ファイル 5 0、変換テーブル 5 1、辞書テーブル 5 2、圧縮ファイルテーブル 5 3、ベクトルテーブル 5 4、転置インデックステーブル 5 5 を有する。また、記憶部 2 4 0 は、サブゲノムテーブル T 1、代替遺伝子ベクターテーブル T 2、ゲノム辞書 D 2、学習データ 9 0、学習モデル 9 1、分析クエリ 9 2、共通構造・遺伝子組換構造テーブル 9 3 を有する。記憶部 2 4 0 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。

【 0 1 2 3 】

塩基ファイル 5 0、変換テーブル 5 1、辞書テーブル 5 2、圧縮ファイルテーブル 5 3

50

、ベクトルテーブル54、転置インデックステーブル55、サブゲノムテーブルT1、代替遺伝子ベクターテーブルT2、ゲノム辞書D2に関する説明は、実施例1で説明した内容と同様である。学習データ90は、図29で説明した内容と同様である。学習モデル91、分析クエリ92に関する説明は、図30で説明した内容と同様である。

【0124】

共通構造・遺伝子組換構造テーブル93は、図30で説明したように、共通構造ベクトルに類似した遺伝子ベクターからサブゲノムに遺伝子組み換えするための遺伝子組換構造ベクトルの情報が含まれる。図30では、たとえば、共通構造・遺伝子組換構造テーブル93には、Vcm92-1に対応した遺伝子組換構造ベクトルが含まれる。共通構造のベクトルと、遺伝子組換構造のベクトルとを積算したベクトルが、遺伝子ベクターのベクトル

10

【0125】

図31に説明に戻る。制御部250は、前処理部251、学習部252、算出部253、分析部254を有する。制御部250は、例えば、CPUやMPUにより実現される。また、制御部250は、例えばASICやFPGA等の集積回路により実行されてもよい。

【0126】

前処理部251に関する説明は、実施例1で説明した前処理部151に関する処理の説明と同様である。前処理部251によって、塩基ファイル50、変換テーブル51、辞書テーブル52、圧縮ファイルテーブル53、ベクトルテーブル54、転置インデックステーブル55、サブゲノムテーブルT1、代替遺伝子ベクターテーブルT2が生成される。前処理部251は、学習データ90を、外部装置から取得してもよいし、前処理部251が生成してもよい。

20

【0127】

算出部253は、分析クエリ92の指定を受け付けた場合に、学習済みの学習モデル91を用いて、分析クエリ92のサブゲノムの合成経路で遺伝子組み換えさせる各共通構造のベクトルを算出する。算出部253は、算出した各共通構造のベクトルを、分析部254に出力する。

【0128】

以下の説明では、算出部253が算出した各共通構造のベクトルを、それぞれ「共通構造ベクトル」と表記する。

30

【0129】

分析部254は、分析クエリ92のサブゲノムのベクトルと、共通構造ベクトルと、遺伝子ベクターベクトルテーブルT2とを基にして、共通構造・遺伝子組換機構テーブル93を生成する。以下において、分析部254の処理の一例について説明する。

【0130】

分析部254は、サブゲノムのベクトルと、代替遺伝子ベクターテーブルT2に含まれる各ベクトルとの距離をそれぞれ算出し、サブゲノムのベクトルとの距離が閾値未満となるベクトルを特定する。代替遺伝子ベクターテーブルT2に含まれるベクトルであって、サブゲノムのベクトルとの距離が閾値未満となるベクトルを「類似ベクトル」と表記する。

【0131】

分析部254は、類似ベクトルから、共通構造ベクトルを減算することで、遺伝子組換構造のベクトルを算出し、共通構造ベクトルと、遺伝子組換構造のベクトルとの対応関係を特定する。分析部254は、共通構造ベクトルと、遺伝子組換構造のベクトルとを共通構造・遺伝子組換構造テーブル93に登録する。分析部245は、上記処理を繰り返し実行することで、共通構造・遺伝子組換構造テーブル93を生成する。分析部245は、共通構造・遺伝子組換構造テーブル93を、表示部230に出力して表示させてもよいし、ネットワークに接続された外部装置に送信してもよい。

40

【0132】

次に、本実施例2に係る情報処理装置200の処理手順の一例について説明する。図32は、本実施例2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。情報処理

50

装置 200 の算出部 253 は、分析クエリ 92 を受け付ける (ステップ S301)。

【0133】

算出部 253 は、サブゲノムテーブル T1 を基にして、分析クエリ 92 のサブゲノムをベクトルに変換する (ステップ S302)。

【0134】

算出部 253 は、サブゲノムのベクトルを、学習済みの学習モデル 91 に入力することで、共通構造のベクトルを算出する (ステップ S303)。情報処理装置 200 の分析部 254 は、共通構造のベクトルと、代替遺伝子ベクターテーブル T2 の各ベクトルとの距離を基にして、類似ベクトルを特定する (ステップ S304)。

【0135】

分析部 254 は、サブゲノムと類似する遺伝子ベクターの各ベクトルから共通構造のベクトルを減算することで、遺伝子組換構造のベクトルを算出する (ステップ S305)。分析部 254 は、共通構造のベクトルと、遺伝子組換構造のベクトルとの関係を、共通構造・遺伝子組換構造テーブル 93 に登録する (ステップ S306)。分析部 254 は、共通構造・遺伝子組換構造テーブルの情報を出力する (ステップ S307)。

【0136】

次に、本実施例 2 に係る情報処理装置 200 の効果について説明する。情報処理装置 100 は、学習済みの学習モデル 91 に分析クエリ 92 のベクトルを入力し、分析クエリのサブゲノムに対応する各共通構造のベクトルを算出する。また、サブゲノムに類似する遺伝子ベクターのベクトルから各、共通構造のベクトルを減算することで、類似するサブゲノムと、遺伝子ベクターとで相違する遺伝子組換構造のベクトルを算出する。上記の共通構造のベクトルや、遺伝子組換構造のベクトルを用いることで、目的ゲノムへの遺伝子組換や再合成や製造に活用可能な、より良い遺伝子ベクターを容易に分析することができる。

【0137】

サブゲノムと遺伝子ベクターは、複数のタンパク質一次構造で構成される二次構造である。また、タンパク質一次構造の分散ベクトルを用いることで、あるタンパク質一次構造に隣接するタンパク質一次構造を推定することができ、各タンパク質一次構造の結合度や安定性の評価に応用することができる。実績のあるサブゲノムへ遺伝子ベクターからの遺伝子組み換えに関し、サブゲノムや遺伝子ベクターの二次構造を構成する複数のタンパク質一次構造の分散ベクトルをもとに、機械学習を行うことで、遺伝子ベクターからの流用と、遺伝子組み換えと、再合成の分析精度を向上することができる。

【0138】

次に、上記実施例に示した情報処理装置 100 (200) 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 33 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。

【0139】

図 9 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 CPU 301 と、ユーザからのデータの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 RAM 306 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。

【0140】

ハードディスク装置 307 は、前処理プログラム 307a、学習プログラム 307b、算出プログラム 307c、分析プログラム 307d を有する。また、CPU 301 は、各プログラム 307a ~ 307d を読み出して RAM 306 に展開する。

【0141】

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。学習プログラム 307b は、学習プロセス 306b として機能する。算出プログラム 307c は、算出プロセス 306c として機能する。分析プログラム 307d は、分析プロセス 306d と

10

20

30

40

50

して機能する。

【 0 1 4 2 】

前処理プロセス 3 0 6 a の処理は、前処理部 1 5 1 , 2 5 1 の処理に対応する。学習プロセス 3 0 6 b の処理は、学習部 1 5 2 , 2 5 2 の処理に対応する。算出プロセス 3 0 6 c の処理は、算出部 1 5 3 , 2 5 3 の処理に対応する。分析プロセス 3 0 6 d の処理は、分析部 1 5 4 の処理に対応する。

【 0 1 4 3 】

なお、各プログラム 3 0 7 a ~ 3 0 7 d については、必ずしも最初からハードディスク装置 3 0 7 に記憶させておかなくても良い。例えば、コンピュータ 3 0 0 に挿入されるフレキシブルディスク (F D)、C D - R O M、D V D、光磁気ディスク、I C 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 3 0 0 が各プログラム 3 0 7 a ~ 3 0 7 d を読み出して実行するようにしてもよい。

10

【符号の説明】

【 0 1 4 4 】

1 1 0 , 2 1 0	通信部
1 2 0 , 2 2 0	入力部
1 3 0 , 2 3 0	表示部
1 4 0 , 2 4 0	記憶部
1 5 0 , 2 5 0	制御部
1 5 1 , 2 5 1	前処理部
1 5 2 , 2 5 2	学習部
1 5 3 , 2 5 3	算出部
1 5 4 , 2 5 4	分析部

20

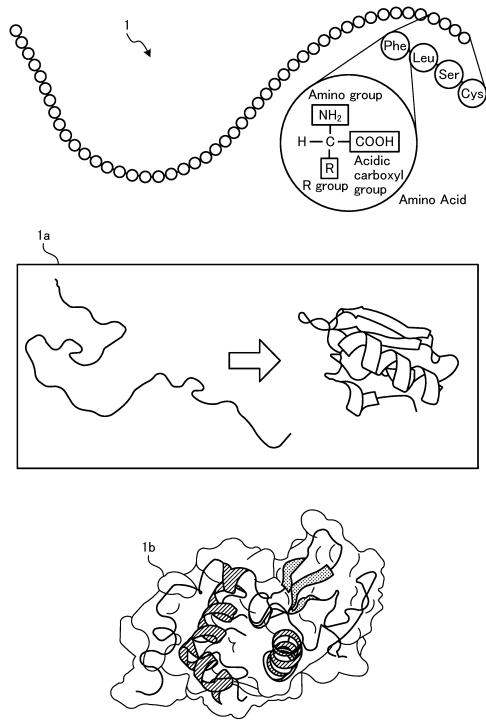
30

40

50

【 図 面 】

【 図 1 】



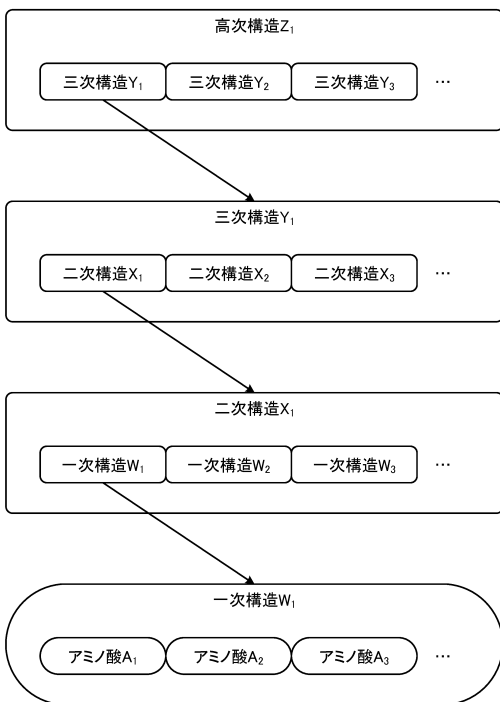
【 図 2 】

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	CUA	CUG
メチオニン(開始)	Met	M			UUA	UUG
アスパラギン	Asn	N	AAU	AAC		AUG
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	CGG
					AGA	AGG
セリン	Ser	S	UCU	UCC	UCA	UCG
			AGU	AGC		
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	
決定されていないアミノ酸	Xxx	X				
グルタミン酸かグルタミンのどちらか	Glx	Z				

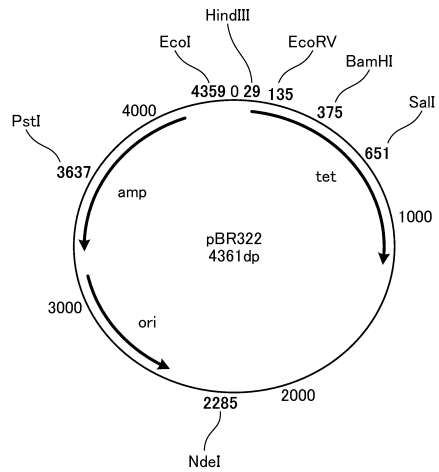
10

20

【 図 3 】



【 図 4 】

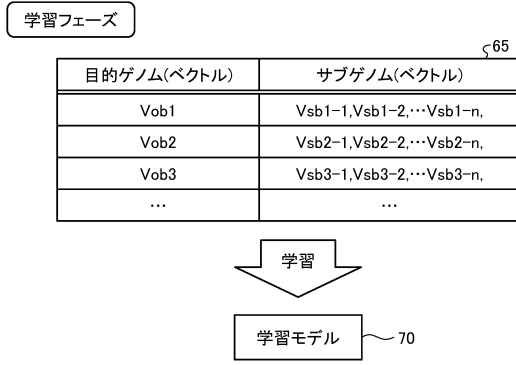


30

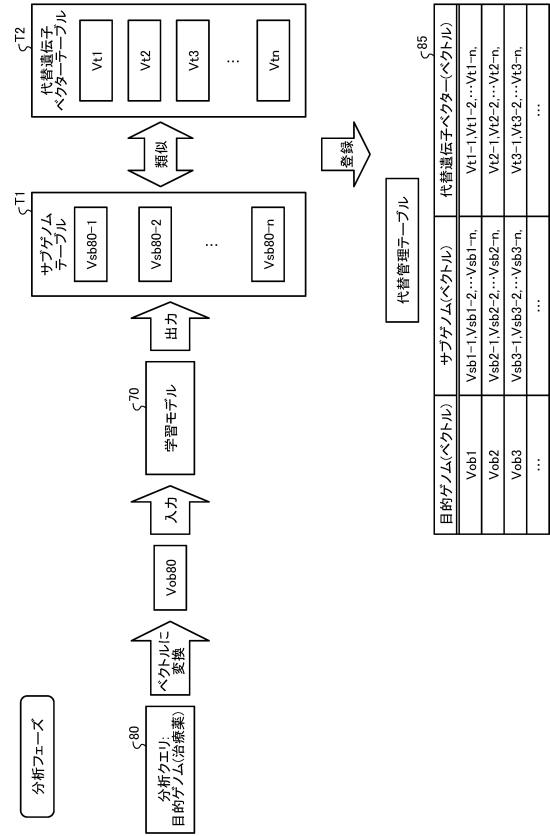
40

50

【図5】



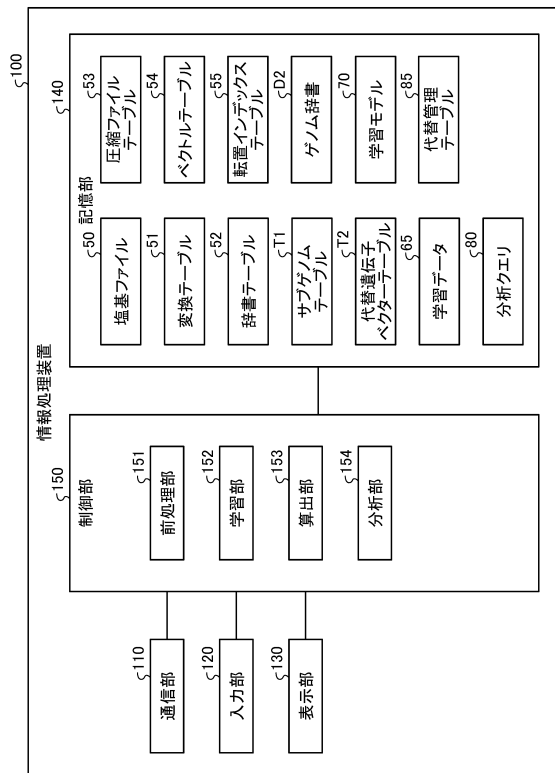
【図6】



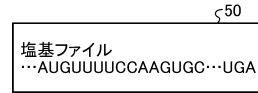
10

20

【図7】



【図8】



30

40

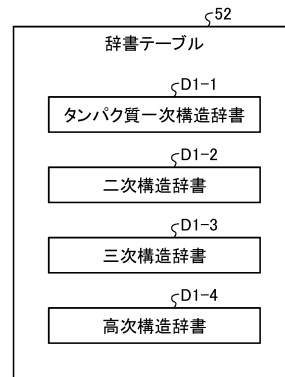
50

【 図 9 】

ζ51

	4(0100)	5(0101)	6(0110)	7(0111)
0(0000)	UUU	CUU	AUU	GUU
1(0001)	UUC	CUC	AUC	GUC
2(0010)	UUA	CUA	AUA	GUA
3(0011)	UUG	CUG	AUG	GUG
4(0100)	UCU	CCU	ACU	GCU
5(0101)	UCC	CCC	ACC	GCC
6(0110)	UCA	CCA	ACA	GCA
7(0111)	UCG	CCG	ACG	GCG
8(1000)	UAU	CAU	AAU	GAU
9(1001)	UAC	CAC	AAC	GAC
A(1010)	UAA	CAA	AAA	GAA
B(1011)	UAG	CAG	AAG	GAG
C(1100)	UGU	CGU	AGU	GGU
D(1101)	UGC	CGC	AGC	GGC
E(1110)	UGA	CGA	AGA	GGA
F(1111)	UGG	CGG	AGG	GGG

【 図 1 0 】



10

【 図 1 1 】

ζD1-1

圧縮符号	名称	コドン符号配列
C0008000h	I 型コラーゲン	02h63h78h.....03h
C0008001h	II 型コラーゲン	02h52h79h.....03h
C0008002h	III 型コラーゲン	02h66h.....03h
C0008003h	IV 型コラーゲン	...
...

【 図 1 2 】

ζD1-2

圧縮符号	名称	タンパク質一次構造符号配列
D0000000h	α 二次構造	C0008001hC00...
D0000001h	β 二次構造	C0008103hC00...
D0000002h	γ 二次構造	C0001258hC00...
...

20

30

40

50

【 図 1 3 】

ζD1-3

圧縮符号	名称	二次構造符号配列
E000000h	α α 三次構造	D0008031hD00...
E000001h	α β 三次構造	D0008003hD00...
E000002h	γ β 三次構造	D0331258hD00...
...

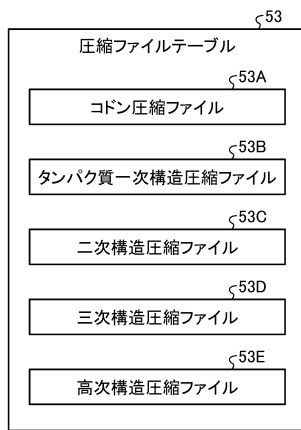
【 図 1 4 】

ζD1-4

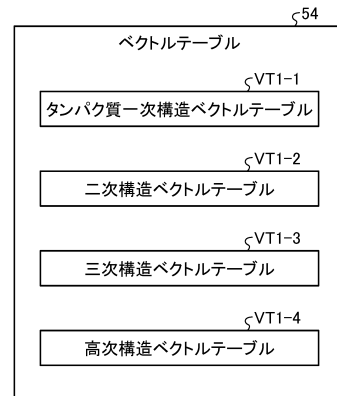
圧縮符号	名称	三次構造符号配列
F000000h	α α α 高次構造	E0000031hE00...
F000001h	α α β 高次構造	E0000831hE00...
F000002h	α β β 高次構造	E0033031hE00...
...

10

【 図 1 5 】



【 図 1 6 】



20

30

40

50

【図 17】

ζVT1-1

圧縮符号(タンパク質一次構造)	ベクトル
C0008000h	V1-1
C0008001h	V1-2
C0008002h	V1-3
C0008003h	V1-4
...	...

【図 18】

ζVT1-2

圧縮符号(二次構造)	ベクトル
D0000000h	V2-1
D0000001h	V2-2
D0000002h	V2-3
...	...

10

【図 19】

ζVT1-3

圧縮符号(三次構造)	ベクトル
E0000000h	V3-1
E0000001h	V3-2
E0000002h	V3-3
...	...

【図 20】

ζVT1-4

圧縮符号(高次構造)	ベクトル
F0000000h	V4-1
F0000001h	V4-2
F0000002h	V4-3
...	...

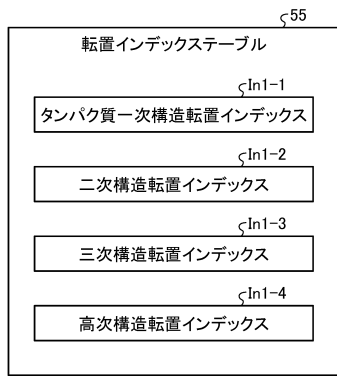
20

30

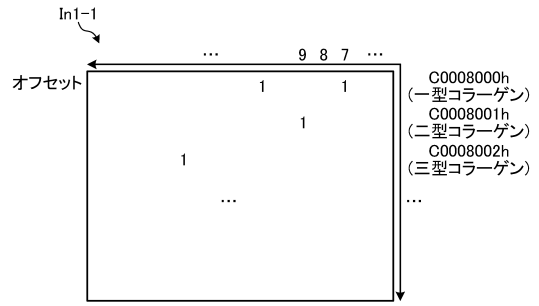
40

50

【 図 2 1 】

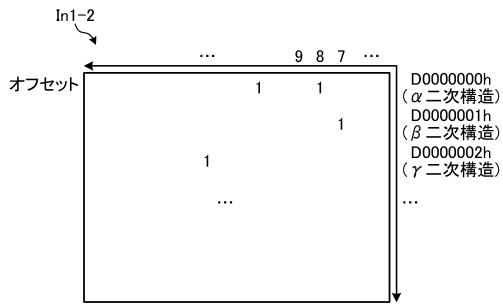


【 図 2 2 】

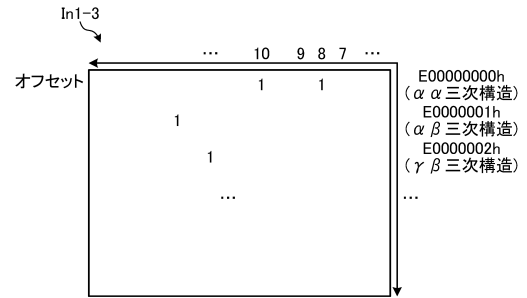


10

【 図 2 3 】



【 図 2 4 】



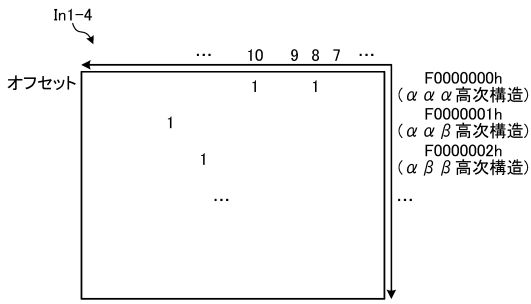
20

30

40

50

【図 2 5】

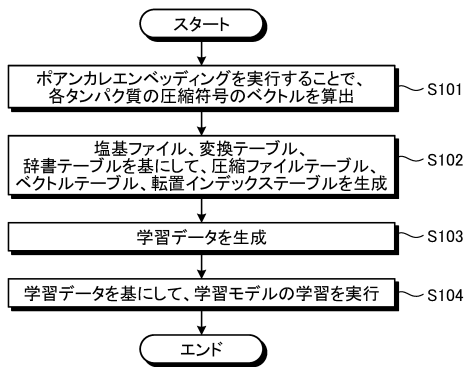


【図 2 6】

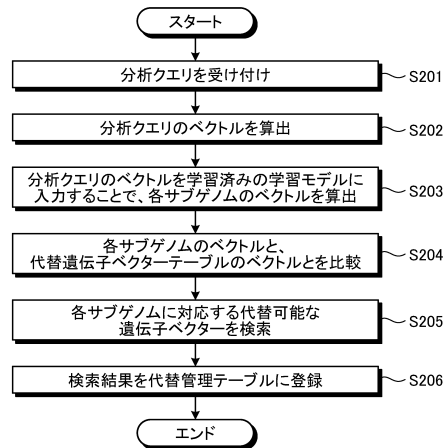
ζ_{D2}	
目的ゲノム(名称)	サブゲノム(名称)
目的ゲノムA	サブゲノムA1,サブゲノムA2;...
目的ゲノムB	サブゲノムB1,サブゲノムB2;...
目的ゲノムC	サブゲノムC1,サブゲノムC2;...
...	...

10

【図 2 7】



【図 2 8】



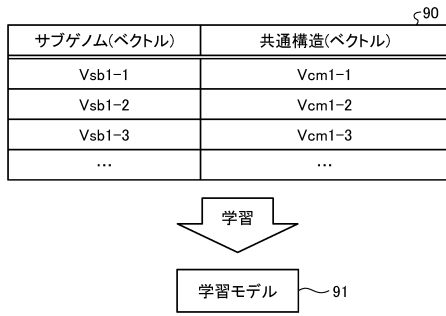
20

30

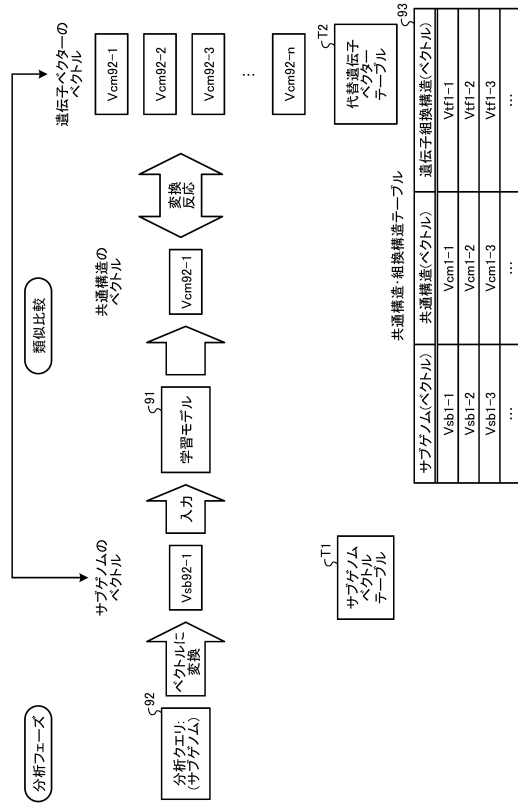
40

50

【図 29】



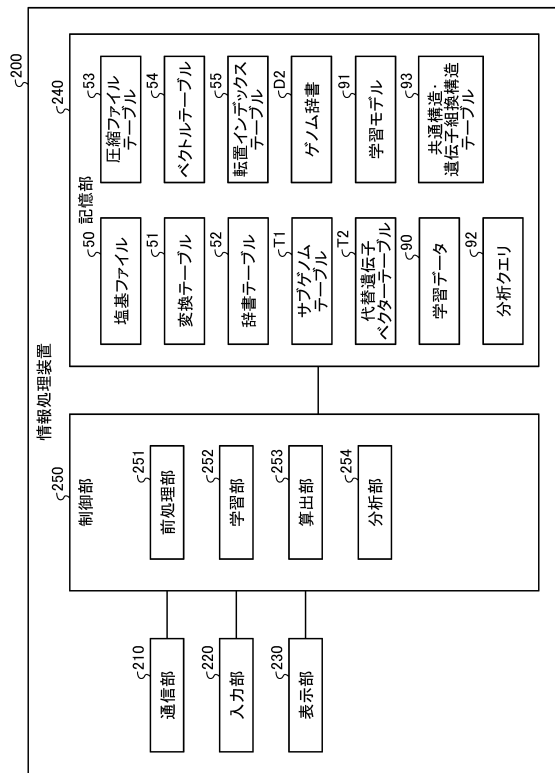
【図 30】



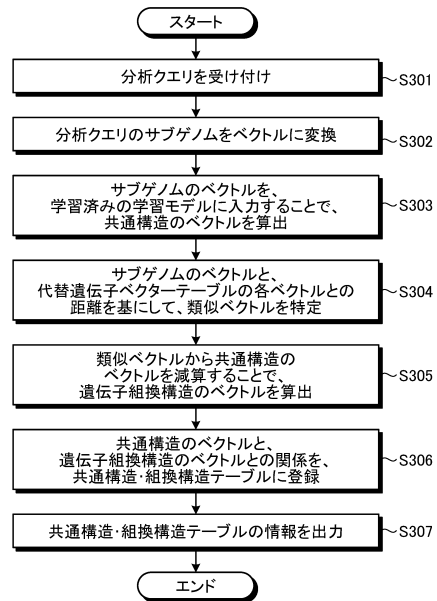
10

20

【図 31】



【図 32】

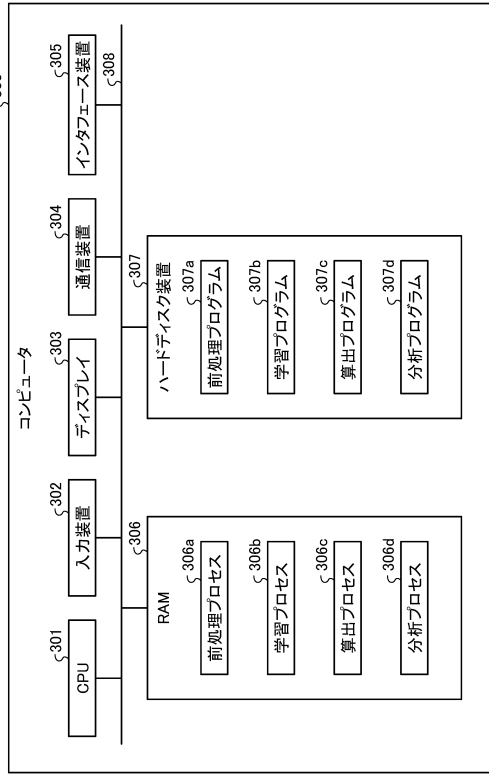


30

40

50

【 図 33 】



10

20

30

40

50

フロントページの続き

審査官 渡邊 加寿磨

- (56)参考文献 特表2020-530918(JP,A)
特表2017-504913(JP,A)
特開2020-154442(JP,A)
特表2000-507940(JP,A)
中国特許出願公開第103189550(CN,A)
米国特許出願公開第2011/0131027(US,A1)
- (58)調査した分野 (Int.Cl., DB名)
G06Q 10/00-99/00
G16B 5/00-99/00
G16C 10/00-99/00
G16Z 99/00