

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】令和4年6月30日(2022.6.30)

【国際公開番号】WO2020/058176

【公表番号】特表2022-501694(P2022-501694A)

【公表日】令和4年1月6日(2022.1.6)

【出願番号】特願2021-509152(P2021-509152)

【国際特許分類】

G 1 6 B 1 5 / 2 0 (2 0 1 9 . 0 1)

10

【 F I 】

G 1 6 B 1 5 / 2 0

【誤訳訂正書】

【提出日】令和4年6月22日(2022.6.22)

【誤訳訂正1】

【訂正対象書類名】特許請求の範囲

【訂正対象項目名】全文

【訂正方法】変更

【訂正の内容】

20

【特許請求の範囲】

【請求項1】

所与のタンパク質に対する距離マップを生成するためのコンピュータで実施される方法であって、前記所与のタンパク質が、ある構造に並べられるアミノ酸残基の配列によって定義され、前記距離マップが、前記構造の中の前記アミノ酸残基間の推定距離を特徴付け、前記方法が、

複数の距離マップクロップを生成するステップであって、各距離マップクロップが、前記タンパク質の前記構造の中の(i)前記配列において1つまたは複数のそれぞれの第1の位置の各々にあるアミノ酸残基と、(ii)前記配列において1つまたは複数のそれぞれの第2の位置の各々にあるアミノ酸残基との間の推定距離を特徴付け、各距離マップクロップが、前記距離マップの真部分集合であり、距離マップクロップを生成するステップが、前記配列における1つまたは複数の第1の位置および前記配列における1つまたは複数の第2の位置を特定するステップであって、前記第1の位置が前記配列の真部分集合である、ステップと、

30

前記配列における前記第1の位置にある前記アミノ酸残基および前記配列における前記第2の位置にある前記アミノ酸残基からネットワーク入力を決定するステップと、

前記ネットワーク入力を距離予測ニューラルネットワークに提供するステップであって、前記距離予測ニューラルネットワークが、距離予測ニューラルネットワーク重みの現在値に従って前記ネットワーク入力を処理して前記距離マップクロップを備えるネットワーク出力を生成するように構成される、ステップと

40

を備える、ステップと、

前記複数の距離マップクロップを使用して、前記所与のタンパク質に対する前記距離マップを生成するステップとを備える、コンピュータで実施される方法。

【請求項2】

前記配列における前記第1の位置にある前記アミノ酸残基および前記配列における前記第2の位置にある前記アミノ酸残基からネットワーク入力を決定するステップが、

前記配列における前記第1の位置にある前記アミノ酸残基および前記配列における前記第2の位置にある前記アミノ酸残基に対応する、(i)アミノ酸残基の前記配列の表現、および(ii)アミノ酸残基の前記配列を含む多重配列整列から導かれる整列特徴量の成分を抽出するステップを備える、請求項1に記載の方法。

50

【請求項 3】

前記配列における1つまたは複数の第1の位置および前記配列における1つまたは複数の第2の位置を特定するステップが、
第1の所定の長さの連続する位置の第1の配列として前記第1の位置を確率的にサンプリングするステップと、
第2の所定の長さの連続する位置の第2の配列として前記第2の位置を確率的にサンプリングするステップとを備える、請求項1または2に記載の方法。

【請求項 4】

前記複数の距離マップクロープを使用して前記所与のタンパク質に対する前記距離マップを生成するステップが、
前記複数の距離マップクロープを平均化するステップを備える、請求項1から3のいずれか一項に記載の方法。

10

【請求項 5】

前記複数の距離マップクロープを使用して前記所与のタンパク質に対する前記距離マップを生成するステップが、
融合ニューラルネットワークパラメータの現在値に従って融合ニューラルネットワークを使用して前記複数の距離マップクロープを処理し、前記距離マップを備える出力を生成するステップを備える、請求項1から3のいずれか一項に記載の方法。

【請求項 6】

各距離マップクロープが、第1の位置にある第1のアミノ酸残基および第2の位置にある第2のアミノ酸残基を含むアミノ酸残基の各ペアに対して、アミノ酸残基の前記ペア間の距離範囲の所定のセットにわたるそれぞれの確率分布を定義する、請求項1から5のいずれか一項に記載の方法。

20

【請求項 7】

各距離マップクロープが、第1の位置にある第1のアミノ酸残基および第2の位置にある第2のアミノ酸残基を含むアミノ酸残基の各ペアに対して、アミノ酸残基の前記ペア間の距離が所定の閾値未満であるかどうかを示す二進変数を定義する、請求項1から5のいずれか一項に記載の方法。

【請求項 8】

各距離マップクロープが、第1の位置にある第1のアミノ酸残基および第2の位置にある第2のアミノ酸残基を含むアミノ酸残基の各ペアに対して、アミノ酸残基の前記ペア間の距離を定義する連続値の数を定義する、請求項1から5のいずれか一項に記載の方法。

30

【請求項 9】

前記距離予測ニューラルネットワークが、複数の訓練例を備える訓練データのセットについて訓練され、各訓練例が、(i)訓練タンパク質の中のアミノ酸残基の配列から決定される訓練ネットワーク入力と、(ii)目標距離マップとを備え、前記目標距離マップが、前記訓練タンパク質の中の前記アミノ酸残基間の実際の距離を特徴付ける、請求項1から8のいずれか一項に記載の方法。

【請求項 10】

前記距離予測ニューラルネットワークが、複数の訓練例を備える訓練データのセットについて訓練され、各訓練例が、(i)訓練タンパク質の中のアミノ酸残基の配列から決定される訓練ネットワーク入力と、(ii)目標距離マップとを備え、前記目標距離マップが、前記訓練タンパク質の中の前記アミノ酸残基間の推定距離を特徴付け、前記訓練ネットワーク入力を処理することによって教師ニューラルネットワークの出力として生成される、請求項1から8のいずれか一項に記載の方法。

40

【請求項 11】

複数の前記訓練例に対して、ランダムノイズが前記目標距離マップに追加される、請求項9または10に記載の方法。

【請求項 12】

前記距離予測ニューラルネットワークが、1つまたは複数の拡張畳み込みニューラルネッ

50

トワーク層、1つまたは複数の残差ブロック、1つまたは複数の注意層、またはこれらの組合せを備える、請求項1から11のいずれか一項に記載の方法。

【請求項13】

前記距離予測ニューラルネットワークが、前記配列の中の前記第1の位置における前記アミノ酸残基および前記配列の中の前記第2の位置における前記アミノ酸残基の二次構造を特徴付ける補助出力を生成するように訓練される、請求項11または12に記載の方法。

【請求項14】

前記距離予測ニューラルネットワークが、前記配列の中の前記第1の位置におけるアミノ酸残基と前記配列の中の前記第2の位置におけるアミノ酸残基との間のねじれ角を特徴付ける補助出力を生成するように訓練される、請求項1から13のいずれか一項に記載の方法。

10

【請求項15】

前記所与のタンパク質の予測構造を決定するための1つまたは複数のデータ処理装置によって実行され、前記所与のタンパク質の前記予測構造が複数の構造パラメータの値によって定義され、

前記予測構造を定義する前記複数の構造パラメータの初期値を取得するステップと、前記複数の構造パラメータの前記初期値を更新するステップであって、複数の更新反復の各々において、

前記距離マップを使用して前記構造パラメータの現在値によって定義される前記予測構造の品質を特徴付ける品質スコアを決定するステップと、

20

前記複数の構造パラメータの1つまたは複数に対して、

前記構造パラメータの前記現在値を調整することによって前記品質スコアを最適化するステップと

を備える、ステップと、

前記複数の更新反復の最後の更新反復の後の前記複数の構造パラメータの前記現在値によって定義されるものとして、前記所与のタンパク質の前記予測構造を決定するステップとを備える、請求項1から14のいずれか一項に記載の方法。

【請求項16】

リガンドを取得する方法であって、前記リガンドが薬品または産業用酵素のリガンドであり、前記方法が、

30

標的アミノ酸配列を取得するステップであって、前記標的アミノ酸配列が標的タンパク質のアミノ酸配列である、ステップと、

アミノ酸またはアミノ酸残基の前記配列として前記標的アミノ酸配列を使用して請求項15に記載の方法を実行して、前記標的タンパク質の構造を決定するステップと、

前記標的タンパク質の前記構造との1つまたは複数のリガンド候補の相互作用を評価するステップと、

前記評価の結果に応じて、前記リガンド候補のうちの1つまたは複数を選択するステップとを備える、コンピュータで実施される方法。

【請求項17】

前記標的タンパク質が受容体または酵素を備え、前記リガンドが前記受容体または酵素のアゴニストまたはアンタゴニストである、請求項16に記載の方法。

40

【請求項18】

ポリペプチドリガンドを取得する方法であって、前記リガンドが、薬品または産業用酵素のリガンドであり、前記方法が、

1つまたは複数のポリペプチドリガンド候補のアミノ酸配列を取得するステップと、

前記1つまたは複数のポリペプチドリガンド候補の各々に対して、アミノ酸またはアミノ酸残基の前記配列として前記ポリペプチドリガンド候補の前記アミノ酸配列を使用して請求項15に記載の方法を実行して、前記ポリペプチドリガンド候補の構造を決定するステップと、

標的タンパク質の標的タンパク質構造を取得するステップと、

50

前記1つまたは複数のポリペプチドリガンド候補の各々の前記構造と前記標的タンパク質構造との間の相互作用を評価するステップと、
前記評価の結果に応じて、前記ポリペプチドリガンド候補の前記1つまたは複数のうちの1つを前記ポリペプチドリガンドとして選択するステップとを備える、コンピュータで実施される方法。

【請求項19】

前記リガンド候補のうちの1つとの前記相互作用を評価する前記ステップが、前記リガンド候補に対する相互作用スコアを決定するステップを備え、前記相互作用スコアが、前記リガンド候補と前記標的タンパク質との間の相互作用の尺度を備える、請求項16から18のいずれか一項に記載の方法。

10

【請求項20】

前記リガンドを合成するステップをさらに備える、請求項16から19のいずれか一項に記載の方法。

【請求項21】

in vitroおよびin vivoで前記リガンドの生物学的な活性を試験するステップをさらに備える、請求項20に記載の方法。

【請求項22】

タンパク質ミスフォールディング疾患の存在を特定する方法であって、
タンパク質のアミノ酸配列を取得するステップと、

アミノ酸またはアミノ酸残基の前記配列として前記タンパク質のアミノ酸配列を使用して
請求項15に記載の方法を実行して、前記タンパク質の構造を決定するステップと、
人体または動物の体から得られた前記タンパク質のあるバージョンの構造を取得するステップと、

20

人体または動物の体から得られた前記タンパク質のあるバージョンの前記構造と前記タンパク質の前記構造を比較するステップと、
前記比較の結果に応じてタンパク質ミスフォールディング疾患の存在を特定するステップとを備える、コンピュータで実施される方法。

【請求項23】

1つまたは複数のコンピュータと、前記1つまたは複数のコンピュータによって実行されると、前記1つまたは複数のコンピュータに請求項1から19および22のいずれか一項に記載の方法の動作を実行させる命令を記憶した1つまたは複数の記憶デバイスとを備える、システム。

30

【請求項24】

1つまたは複数のコンピュータによって実行されると、前記1つまたは複数のコンピュータに、請求項1から19および22のいずれか一項に記載の方法の動作を実行させる命令を記憶した、1つまたは複数のコンピュータ記憶媒体。

【誤訳訂正2】

【訂正対象書類名】明細書

【訂正対象項目名】0025

【訂正方法】変更

40

【訂正の内容】

【0025】

方法は、複数の距離マップクロップを生成するステップを備えてもよく、距離マップクロップの各々が、タンパク質の構造の中の(i)配列において1つまたは複数のそれぞれの第1の位置の各々にあるアミノ酸残基と、(ii)配列において1つまたは複数のそれぞれの第2の位置の各々にあるアミノ酸残基との間の推定距離を特徴付ける。距離マップクロップを生成するステップは、配列における1つまたは複数の第1の位置および配列における1つまたは複数の第2の位置を特定するステップを備えてもよく、第1の位置は配列の真部分集合であってもよい。距離マップクロップを生成するステップはさらに、配列における第1の位置にあるアミノ酸残基および配列における第2の位置にあるアミノ酸残基からネッ

50

トワーク入力を決定するステップを備え得る。距離マップクロープを生成するステップはさらに、距離予測ニューラルネットワーク重みの現在値に従ってネットワーク入力を処理して距離マップクロープを備えるネットワーク出力を生成するように構成される、距離予測ニューラルネットワークにネットワーク入力を提供するステップを備え得る。そして、所与のタンパク質に対する距離マップが、複数の距離マップクロープを使用して生成され得る。

【誤訳訂正 3】

【訂正対象書類名】明細書

【訂正対象項目名】0140

【訂正方法】変更

【訂正の内容】

【0140】

予測構造の品質スコアは一般に、予測構造の品質を特徴付ける。たとえば、図5を参照してより詳しく説明されるように、予測構造の品質スコアは、予測構造とタンパク質104の実際の構造との間の推定される類似性、予測構造の中のアミノ酸の主鎖原子間の距離に基づく予測構造の尤度、または両方を特徴付け得る。一般に、予測構造の品質スコアは、アミノ酸配列102の中のアミノ酸の原子の真部分集合に基づいて(たとえば、主鎖原子だけに基つき、いずれの他の原子には基づかず)、またはアミノ酸配列102の中のアミノ酸の各々の原子のすべてに基づいて、決定され得る。アミノ酸の中の主鎖原子だけに基づく予測構造の品質スコアは主鎖原子スコアと呼ばれることがあり、一方、アミノ酸の各々の原子のすべてに関して決定される品質スコアは全原子スコアと呼ばれることがある。

【誤訳訂正 4】

【訂正対象書類名】明細書

【訂正対象項目名】0169

【訂正方法】変更

【訂正の内容】

【0169】

いくつかの場合、ネットワーク入力410は、アミノ酸配列102全体を特徴付けるデータを含む(たとえば、ネットワーク入力410は、アミノ酸配列102の中の各アミノ酸の表現を含む)。いくつかの他の場合には、ネットワーク入力410は、アミノ酸配列102の中のアミノ酸の真部分集合(すなわち、アミノ酸配列102の中の各アミノ酸ではなく)だけを特徴付けるデータを含む。

【誤訳訂正 5】

【訂正対象書類名】明細書

【訂正対象項目名】0170

【訂正方法】変更

【訂正の内容】

【0170】

断片生成システム310は、生成ニューラルネットワーク重みの現在値に従ってネットワーク入力410を処理して構造パラメータ404を生成するように構成される、生成ニューラルネットワーク406にネットワーク入力410を提供する。構造パラメータは、たとえば、アミノ酸配列102のアミノ酸のセットの主鎖原子に対する、主鎖原子ねじれ角または主鎖原子座標であり得る。いくつかの実装形態では、構造パラメータ404はアミノ酸配列102全体の構造を定義し、断片生成システム310は構造パラメータ404を断片化することによって構造断片を生成する。いくつかの実装形態では、構造パラメータ404は、アミノ酸配列102の断片(すなわち、真部分集合)の構造を定義する。

【誤訳訂正 6】

【訂正対象書類名】明細書

【訂正対象項目名】0218

【訂正方法】変更

10

20

30

40

50

【訂正の内容】

【0218】

距離マップ608を生成するために、システム528は距離マップクロップ612のセットを生成し得る。各距離マップクロップ612は、全距離マップ608の真部分集合の推定である。より具体的には、各距離マップクロップ612は、(i)アミノ酸配列102の中の1つまたは複数の第1の位置の各々におけるアミノ酸残基と、(ii)アミノ酸配列102の中の1つまたは複数の第2の位置の各々におけるアミノ酸残基との間の推定距離を特徴付け、第1の位置、第2の位置、または両方が、アミノ酸配列102におけるすべての位置のうちの真部分集合である。第1の位置は、距離マップ行列608のそれぞれの行を特定するものとして理解されてもよく、第2の位置は、距離マップ行列608のそれぞれの列を特定するものとして理解されてもよい。以下でより詳しく説明されるように、システム528は、融合エンジン614を使用して距離マップクロップ612を組み合わせて、全距離マップ608を生成する。

10

【誤訳訂正7】

【訂正対象書類名】明細書

【訂正対象項目名】0326

【訂正方法】変更

【訂正の内容】

【0326】

前に言及されたように、スコアリングシステム916は、距離マップ全体1018の真部分集合である距離マップクロップのセットを生成し、次いで距離マップクロップを融合する(たとえば、重複する距離マップクロップを平均化する)ことによって、図6に関連して説明されるような距離マップ1018を決定することができる。

20

【誤訳訂正8】

【訂正対象書類名】明細書

【訂正対象項目名】0358

【訂正方法】変更

【訂正の内容】

【0358】

システム1200は、ドメインセグメンテーションエンジン1216を使用してドメインセグメンテーション候補1230を生成する。いくつかの実装形態では、ドメインセグメンテーションエンジン1216は、タンパク質1204の1つ1つのあり得るドメインセグメンテーションに対応するドメインセグメンテーション候補1230を生成する。いくつかの他の実装形態では、ドメインセグメンテーションエンジン1216は、タンパク質1204のあり得るドメインセグメンテーションの真部分集合に対応するドメインセグメンテーション候補1230を(たとえば、タンパク質1204の所定の数のあり得るドメインセグメンテーションをランダムにサンプリングすることによって)生成する。

30

【誤訳訂正9】

【訂正対象書類名】明細書

【訂正対象項目名】0370

【訂正方法】変更

【訂正の内容】

【0370】

システムは、タンパク質の複数のドメインセグメンテーション候補を取得する(1302)。各ドメインセグメンテーション候補は、複数のそれぞれのドメイン候補へのタンパク質のアミノ酸配列の区分を定義する。いくつかの実装形態では、システムは、タンパク質の1つ1つのあり得るドメインセグメンテーションに対応するドメインセグメンテーション候補を生成する。いくつかの他の実装形態では、システムは、タンパク質のあり得るドメインセグメンテーションの真部分集合に対応するドメインセグメンテーション候補を(たとえば、タンパク質の所定の数のあり得るドメインセグメンテーションをランダムにサンブ

40

50

リングすることによって)生成する。

10

20

30

40

50