

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2024-537793

(P2024-537793A)

(43)公表日 令和6年10月16日(2024.10.16)

(51)国際特許分類

F I

G 1 6 B 15/30 (2019.01)

G 1 6 B 15/30

G 1 6 B 40/00 (2019.01)

G 1 6 B 40/00

審査請求 未請求 予備審査請求 未請求 (全83頁)

(21)出願番号	特願2024-519522(P2024-519522)	(71)出願人	516331306
(86)(22)出願日	令和4年9月29日(2022.9.29)		アトムワイズ, インコーポレイテッド
(85)翻訳文提出日	令和6年5月15日(2024.5.15)		アメリカ合衆国, カリフォルニア州 9
(86)国際出願番号	PCT/US2022/045250		4 1 0 3 サンフランシスコ マーケット
(87)国際公開番号	WO2023/055949		ストリート 7 1 7 スイート 8 0 0
(87)国際公開日	令和5年4月6日(2023.4.6)	(74)代理人	100079108
(31)優先権主張番号	63/251,142		弁理士 稲葉 良幸
(32)優先日	令和3年10月1日(2021.10.1)	(74)代理人	100109346
(33)優先権主張国・地域又は機関			弁理士 大貫 敏史
	米国(US)	(74)代理人	100117189
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA		弁理士 江口 昭彦
	,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(	(74)代理人	100134120
	AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A		弁理士 内藤 和彦
	T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR	(72)発明者	グニエウェク, パーヴェル
	,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,		アメリカ合衆国, カリフォルニア州 9
	最終頁に続く		最終頁に続く

(54)【発明の名称】 負のポーズデータ及びモデルコンディショニングを使用した化合物とポリマーとの間の相互作用の特徴付け

(57)【要約】

試験化合物とポリマーとの間の相互作用を特徴付けるためのシステム及び方法は、ポリマーの配位、及び化合物の訓練データセットを使用する。各化合物は、正の相互作用スコアを有する標的ポリマー配位に関する正のポーズと、標的ポリマー配位及び負の相互作用スコアに関する化合物の負のポーズと、を有する。モデルは、各化合物に対して、少なくとも (i) 化合物の正の相互作用スコアに対する、モデルへの入力としての正のポーズの正のスコアと、(ii) 化合物の負の相互作用スコアに対する、モデルへの入力としての負のポーズの負のスコアと、を適用することによって訓練され、それによって、モデルのパラメータを調整する。次に、モデルの出力は、少なくとも部分的に、試験化合物とポリマーとの間の相互作用を特徴付けるために使用される。

【選択図】 なし

【特許請求の範囲】

【請求項 1】

試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのコンピュータシステムであって、前記コンピュータシステムが、

1つ以上のプロセッサと、

前記1つ以上のプロセッサによってアドレス指定可能なメモリと、を備え、前記メモリが、前記1つ以上のプロセッサによる実行のための少なくとも1つのプログラムを記憶し、前記少なくとも1つのプログラムが、

(A) 前記標的ポリマーの複数の原子座標を取得することであって、前記複数の原子座標が、少なくとも400個の原子の原子座標を含む、複数の原子座標を取得することと、

(B) 複数の訓練化合物における各訓練化合物のそれぞれの電子記述を含む訓練データセットを取得することであって、前記複数の訓練化合物が、少なくとも100個の化合物を含み、各それぞれの電子記述が、

(i) 対応する第1の正の相互作用スコアと結合された複数の原子空間座標に関する対応する前記訓練化合物の対応する正のポーズと、

(i i) 対応する第1の負の相互作用スコアと結合された前記複数の原子空間座標に関する前記対応する訓練化合物の対応する負のポーズと、を含む、訓練データセットを取得することと、

(C) 少なくとも第1のモデルを訓練することであって、前記第1のモデルが、第1の複数のパラメータを有し、前記第1の複数のパラメータが、400個よりも多いパラメータを含み、前記訓練が、前記複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも

(i) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第1の正の相互作用スコアに対する、前記第1のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの対応する正のスコアと、

(i i) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第1の負の相互作用スコアに対する、前記第1のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの対応する負のスコアと、を使用し、それによって、前記第1の複数のパラメータを調整し、少なくとも前記第1のモデルの出力が、少なくとも部分的に、前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けを提供するために使用される、訓練することと、を行うための命令を含む、コンピュータシステム。

【請求項 2】

前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコアが、

前記対応する正のポーズにおける前記標的ポリマーに関する前記対応する訓練化合物の対応する正のボクセルマップを取得することと、

前記対応する正のボクセルマップを、対応する正のベクトルに展開することと、

前記対応する正のベクトルを、重畳型ニューラルネットワークに入力することと、を行うことによって取得され、それによって、前記対応する正のポーズの前記対応する正のスコアを取得し、前記重畳型ニューラルネットワークが、500個よりも多いパラメータを含み、

前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコアが、

前記対応する負のポーズにおける前記標的ポリマーに関する前記対応する訓練化合物の対応する負のボクセルマップを取得することと、

前記対応する負のボクセルマップを、対応する負のベクトルに展開することと、

前記対応する負のベクトルを、前記重畳型ニューラルネットワークに入力し、それによって、前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコアを取得することと、によって取得される、請求項1に記載のコ

10

20

30

40

50

ンピュータシステム。

【請求項 3】

前記対応する正のベクトルが、第 1 の一次元ベクトルであり、

前記対応する負のベクトルが、第 2 の一次元ベクトルである、請求項 2 に記載のコンピュータシステム。

【請求項 4】

前記第 1 のモデルが、第 1 の完全接続ニューラルネットワークである、請求項 1 ～ 3 のいずれか一項に記載のコンピュータシステム。

【請求項 5】

前記訓練が、前記第 1 の複数のパラメータが、関連付けられた損失関数を通じた逆伝播によって調整される、回帰タスクであり、

前記対応する第 1 の正の相互作用スコアが、以下の式によって、前記対応する第 1 の負の相互作用スコアに関係付けられており、

$$B = N \times A、$$

式中、

A は、対応する前記正の相互作用スコアであり、

B は、対応する前記負の相互作用スコアであり、

N は、ゼロ超かつ 1 未満の実数である、請求項 1 ～ 4 のいずれか一項に記載のコンピュータシステム。

【請求項 6】

前記関連付けられた損失関数が、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber 損失関数、Log-Cosh 損失関数、又は分位損失関数である、請求項 5 に記載のコンピュータシステム。

【請求項 7】

前記対応する第 1 の正の相互作用スコア及び前記対応する第 1 の負の相互作用スコアが、各々、結合係数を表し、

前記対応する第 1 の正の相互作用スコアが、前記標的ポリマーに対する前記対応する訓練化合物の前記結合係数のインビトロ測定値である、請求項 5 に記載のコンピュータシステム。

【請求項 8】

前記第 1 の正の相互作用スコアが、前記標的ポリマーに関するそれぞれの前記訓練化合物の IC₅₀、EC₅₀、K_d、K_I、又は pK_I である、請求項 7 に記載のコンピュータシステム。

【請求項 9】

前記対応する第 1 の正の相互作用スコア及び前記対応する第 1 の負の相互作用スコアが、各々、前記標的ポリマーに対する前記対応する訓練化合物の結合係数又はインシリコポーズ品質スコアを表す、請求項 1 ～ 4 のいずれか一項に記載のコンピュータシステム。

【請求項 10】

前記訓練データセット内の各それぞれの電子記述が、前記対応する訓練化合物の前記対応する正のポーズの対応する正の活性スコアと、前記対応する訓練化合物の前記対応する負のポーズの対応する負の活性スコアと、を更に含み、

少なくとも前記第 1 のモデルを前記訓練すること (C) が、第 2 のモデルを前記第 1 のモデルと共同で訓練することを更に含み、前記第 2 のモデルが、第 2 の複数のパラメータを有し、前記訓練が、前記複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも

(i i i) 前記対応する訓練化合物の前記対応する正の活性スコアに対する、前記第 2 のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコアと、

(i v) 前記対応する訓練化合物の前記対応する負の活性スコアに対する、前記第 2 のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応

10

20

30

40

50

する負のポーズの前記対応する負のスコアと、を更に使用し、それによって、前記第 2 の複数のパラメータを調整し、前記第 2 のモデルが、少なくとも部分的に、前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けを提供するために、前記第 1 のモデルの前記出力とともに使用される前記試験化合物と前記標的ポリマーとの間の前記相互作用の活性を提供する、請求項 1 ～ 9 のいずれか一項に記載のコンピュータシステム。

【請求項 1 1】

前記第 2 のモデルが、第 2 の完全接続ニューラルネットワークである、請求項 1 0 に記載のコンピュータシステム。

【請求項 1 2】

前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けが、バイナリ活性スコアである、請求項 1 ～ 1 1 のいずれか一項に記載のコンピュータシステム。

10

【請求項 1 3】

前記訓練データセット内の各それぞれの電子記述が、前記対応する訓練化合物の前記対応する正のポーズの対応する正の活性スコアと、前記対応する訓練化合物の前記対応する負のポーズの対応する負の活性スコアと、を更に含み、

少なくとも前記第 1 のモデルを前記訓練すること (C) が、第 2 のモデルを前記第 1 のモデルと共同で訓練することを更に含み、前記第 2 のモデルが、第 2 の複数のパラメータを有し、前記訓練が、前記複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも

(i i i) 前記対応する訓練化合物の前記対応する正の活性スコアに対する、前記第 2 のモデルへの結合入力としての前記標的ポリマー及び前記対応する第 1 の正の相互作用スコアに関する前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコアと、

20

(i v) 前記対応する訓練化合物の前記対応する負の活性スコアに対する、前記第 2 のモデルへの結合入力としての前記標的ポリマー及び前記対応する第 1 の負の相互作用スコアに関する前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコアと、を更に使用し、それによって、前記第 2 の複数のパラメータを調整し、前記第 2 のモデルが、前記第 1 のモデルの前記出力とともに、少なくとも部分的に、前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けを提供するために使用される、請求項 1 ～ 1 2 のいずれか一項に記載のコンピュータシステム。

30

【請求項 1 4】

前記対応する正の活性スコアが、第 1 のバイナリ活性スコアであり、前記対応する負の活性スコアが、第 2 のバイナリ活性スコアである、請求項 1 3 に記載のコンピュータシステム。

【請求項 1 5】

対応する前記第 1 のバイナリ活性スコアが、前記標的ポリマーに対する前記対応する化合物の測定された活性に基づいて、1 の値を割り当てられており、

対応する前記第 2 のバイナリ活性スコアが、0 の値を割り当てられている、請求項 1 4 に記載のコンピュータシステム。

【請求項 1 6】

40

前記第 1 のモデルの前記訓練が、前記第 1 の複数のパラメータが、第 1 の関連付けられた損失関数を通した逆伝播によって調整されている、回帰タスクであり、

前記第 2 のモデルの前記訓練が、前記第 2 の複数のパラメータが、第 2 の関連付けられた損失関数を通した逆伝播によって調整されている、分類タスクである、請求項 1 3 に記載のコンピュータシステム。

【請求項 1 7】

前記対応する第 1 の正の相互作用スコア及び前記対応する第 1 の負の相互作用スコアが、各々、前記標的ポリマーに対する前記対応する訓練化合物の結合係数又はインシリコポーズ品質スコアを表し、

前記対応する正の活性スコアが、第 1 のバイナリ活性スコアであり、前記対応する負の

50

活性スコアが、第 2 のバイナリ活性スコアである、請求項 16 に記載のコンピュータシステム。

【請求項 18】

前記第 1 の関連付けられた損失関数が、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber 損失関数、Log-Cosh 損失関数、又は分位損失関数であり、

前記第 2 の関連付けられた損失関数が、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数である、請求項 16 又は 17 に記載のコンピュータシステム。

【請求項 19】

前記第 2 のモデルが、第 2 の完全接続ニューラルネットワークである、請求項 13 ～ 18 のいずれか一項に記載のコンピュータシステム。

10

【請求項 20】

前記訓練データセット内の各それぞれの電子記述が、前記対応する訓練化合物の前記対応する正のポーズの対応する第 2 の正の相互作用スコアと、前記対応する訓練化合物の前記対応する負のポーズの対応する第 2 の負の相互作用スコアと、を更に含み、

前記訓練データセット内の各それぞれの電子記述が、前記対応する訓練化合物の前記対応する正のポーズの対応する正の活性スコアと、前記対応する訓練化合物の前記対応する負のポーズの対応する負の活性スコアと、を更に含み、

少なくとも前記第 1 のモデルを前記訓練すること (C) が、第 2 のモデル及び第 3 のモデルを前記第 1 のモデルと共同で訓練することを更に含み、前記第 2 のモデルが、第 2 の複数のパラメータを有し、前記第 3 のモデルが、第 3 の複数のパラメータを有し、前記訓練が、前記複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも

20

(i i i) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第 2 の正の相互作用スコアに対する、前記第 2 のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコアと、

(i v) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第 2 の負の相互作用スコアに対する、前記第 2 のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコアであって、それによって、前記第 2 の複数のパラメータを調整する、対応する負のスコアと、

(v) 前記対応する訓練化合物の前記対応する正の活性スコアに対する、前記第 3 のモデルへの結合入力としての、前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコア、前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコアの入力時の前記第 1 のモデルの前記出力及び前記第 2 のモデルの前記出力と、

30

(v i) 前記対応する訓練化合物の前記対応する負の活性スコアに対する、前記第 3 のモデルへの結合入力としての、前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコア、並びに前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコアの入力時の前記第 1 のモデルの前記出力及び前記第 2 のモデルの前記出力と、を更に使用し、それによって、前記第 3 のモデルの前記第 3 の複数のパラメータを調整し、前記第 3 のモデルの出力が、前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けを提供する、請求項 1 に記載のコンピュータシステム。

40

【請求項 21】

前記第 2 のモデルが、第 2 の完全接続ニューラルネットワークであり、

前記第 3 のモデルが、第 3 の完全接続ニューラルネットワークである、請求項 20 に記載のコンピュータシステム。

【請求項 22】

前記対応する正の活性スコアが、第 1 のバイナリ活性スコアであり、前記対応する負の活性スコアが、第 2 のバイナリ活性スコアである、請求項 20 又は 21 に記載のコンピュータシステム。

50

【請求項 23】

前記対応する第1のバイナリ活性スコアが、前記標的ポリマーに対する前記対応する化合物の測定された活性に基づいて、1の値を割り当てられており、

前記対応する第2のバイナリ活性スコアが、0の値を割り当てられている、請求項22に記載のコンピュータシステム。

【請求項 24】

前記第1のモデルの前記訓練が、前記第1の複数のパラメータが、第1の関連付けられた損失関数を通した逆伝播によって調整されている、第1の回帰タスクであり、

前記第2のモデルの前記訓練が、前記第2の複数のパラメータが、第2の関連付けられた損失関数を通した逆伝播によって調整されている、第2の回帰タスクであり、

前記第3のモデルの前記訓練が、前記第3の複数のパラメータが、第3の関連付けられた損失関数を通した逆伝播によって調整されている、分類タスクである、請求項20に記載のコンピュータシステム。

10

【請求項 25】

前記対応する第1の正の相互作用スコア及び前記対応する第1の負の相互作用スコアが、各々、前記標的ポリマーに対する前記対応する訓練化合物のインシリコポーズ品質スコアを表し、

前記対応する第2の正の相互作用スコア及び前記対応する第2の負の相互作用スコアが、各々、前記標的ポリマーに対する前記対応する訓練化合物の結合係数を表し、

前記対応する正の活性スコアが、第1のバイナリ活性スコアであり、前記対応する負の活性スコアが、第2のバイナリ活性スコアである、請求項24に記載のコンピュータシステム。

20

【請求項 26】

前記第1の関連付けられた損失関数が、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数であり、

前記第2の関連付けられた損失関数が、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数であり、

前記第3の関連付けられた損失関数が、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数である、請求項25に記載のコンピュータシステム。

【請求項 27】

前記ポリマーが、タンパク質、ポリペプチド、ポリ核酸、ポリリボ核酸、多糖、又はそれらの任意の組み合わせのアセンブリである、請求項1～26のいずれか一項に記載のコンピュータシステム。

30

【請求項 28】

前記複数の原子座標が、分解能2.5 Å以上又は分解能3.3 Å以上で分解された前記標的ポリマーの結晶構造の三次元座標のセット $\{x_1, \dots, x_N\}$ である、請求項1～27のいずれか一項に記載のコンピュータシステム。

【請求項 29】

前記標的ポリマーの前記複数の原子座標が、核磁気共鳴、中性子回折、又は低温電子顕微鏡によって判定された前記標的ポリマーの三次元座標の集合体を含む、請求項1～27のいずれか一項に記載のコンピュータシステム。

40

【請求項 30】

前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けが、バイナリスコアであり、

前記バイナリスコアの第1の値が、第1の閾値を上回る前記標的ポリマーに関する前記試験化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iを表し、

前記バイナリスコアの第2の値が、前記第1の閾値を下回る前記標的ポリマーに関する前記試験化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iを表す、請求項1に記載のコンピュータシステム。

【請求項 31】

50

前記訓練データセット内の各訓練化合物が、リピンスキーのルールオブファイブ：(i) 5個以下の水素結合ドナー、(ii) 10個以下の水素結合アクセプター、(iii) 500ダルトン未満の分子量、及び(iv) 5未満のLogP、のうちの2つ以上のルール、3つ以上のルール、又は4つのルール全てを満たす、請求項1～30のいずれか一項に記載のコンピュータシステム。

【請求項32】

前記訓練データセット内の各訓練化合物が、500ダルトン未満、1000ダルトン未満、2000ダルトン未満、4000ダルトン未満、6000ダルトン未満、8000ダルトン未満、10000ダルトン未満、又は20000ダルトン未満の分子量を有する有機化合物である、請求項1～30のいずれか一項に記載のコンピュータシステム。

10

【請求項33】

前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの前記対応する正のスコアが、前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズを前記重畳型ニューラルネットワークに入力すると、重畳型ニューラルネットワークから取得され、

前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの前記対応する負のスコアが、前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズを前記重畳型ニューラルネットワークに入力すると、前記重畳型ニューラルネットワークから取得される、請求項1に記載のコンピュータシステム。

【請求項34】

20

前記重畳型ニューラルネットワークが、グラフ重畳型ニューラルネットワーク、等変量ニューラルネットワーク、又はメッセージパッシングニューラルネットワークである、請求項33に記載の方法コンピュータシステム。

【請求項35】

試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法であって、前記方法が、

メモリを備えるコンピュータシステムにおいて、

(A) 前記標的ポリマーの複数の原子座標を取得することであって、前記複数の原子座標が、少なくとも400個の原子の原子座標を含む、複数の原子座標を取得することと、

(B) 複数の訓練化合物における各訓練化合物のそれぞれの電子記述を含む訓練データセットを取得することであって、前記複数の訓練化合物が、少なくとも100個の化合物を含み、各それぞれの電子記述が、

30

(i) 対応する第1の正の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する正のポーズと、

(ii) 対応する第1の負の相互作用スコアと結合された前記複数の原子空間座標に関する前記対応する訓練化合物の対応する負のポーズと、を含む、訓練データセットを取得することと、

(C) 少なくとも第1のモデルを訓練することであって、前記第1のモデルが、第1の複数のパラメータを有し、前記第1の複数のパラメータが、400個よりも多いパラメータを含み、前記訓練が、前記複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも

40

(i) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第1の正の相互作用スコアに対する、前記第1のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの対応する正のスコアと、

(ii) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第1の負の相互作用スコアに対する、前記第1のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの対応する負のスコアと、を使用し、それによって、前記第1の複数のパラメータを調整し、少なくとも前記第1のモデルの出力が、少なくとも部分的に、前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けを提供するために使用される、訓練することと、を含む、方法。

50

【請求項 36】

非一時的なコンピュータ可読記憶媒体であって、前記非一時的なコンピュータ可読記憶媒体が、コンピュータシステムによって実行されるときに、前記コンピュータシステムに、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を実行させる命令を記憶し、前記方法が、

(A) 前記標的ポリマーの複数の原子座標を取得することであって、前記複数の原子座標が、少なくとも400個の原子の原子座標を含む、複数の原子座標を取得することと、

(B) 複数の訓練化合物における各訓練化合物のそれぞれの電子記述を含む訓練データセットを取得することであって、前記複数の訓練化合物が、少なくとも100個の化合物を含み、各それぞれの電子記述が、

(i) 対応する第1の正の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する正のポーズと、

(ii) 対応する第1の負の相互作用スコアと結合された前記複数の原子空間座標に関する前記対応する訓練化合物の対応する負のポーズと、を含む、訓練データセットを取得することと、

(C) 少なくとも第1のモデルを訓練することであって、前記第1のモデルが、第1の複数のパラメータを有し、前記第1の複数のパラメータが、400個よりも多いパラメータを含み、前記訓練が、前記複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも

(i) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第1の正の相互作用スコアに対する、前記第1のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する正のポーズの対応する正のスコアと、

(ii) 前記標的ポリマーに関する前記対応する訓練化合物の前記対応する第1の負の相互作用スコアに対する、前記第1のモデルへの入力としての前記標的ポリマーに関する前記対応する訓練化合物の前記対応する負のポーズの対応する負のスコアと、を使用し、それによって、前記第1の複数のパラメータを調整し、少なくとも前記第1のモデルの出力が、少なくとも部分的に、前記試験化合物と前記標的ポリマーとの間の前記相互作用の前記特徴付けを提供するために使用される、訓練することと、を含む、非一時的なコンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、参照により本明細書に組み込まれる、2021年10月1日に出願の「CHARACTERIZATION OF INTERACTIONS BETWEEN COMPOUNDS AND POLYMERS USING NEGATIVE POSITIVE DATA AND MODEL CONDITIONING」と題された米国仮特許出願第63/251,142号に対する優先権を主張する。

【0002】

本出願は、モデルを使用して、試験化合物と標的ポリマーとの間の相互作用を特徴付けることを対象とする。

【背景技術】

【0003】

基本的に、生物学的システムは、標的ポリマーを有する化合物などの分子の物理的相互作用を通して機能する。構造ベースの仮想高スループットスクリーニング(vHTS)機械学習方法が、機械学習アプローチを通して、候補(試験)化合物と標的ポリマーとの間の相互作用を特徴付けるために使用されている。そのような特徴付けは、例えば、連続的又は分類的な活性標識、PKa、又は任意の他の好適なメトリックを報告して、候補化合物と標的ポリマーとの間の相互作用を特徴付けることができる。

【0004】

v H T S 機械学習方法の1つの欠点は、そのような方法で呼び出された機械学習モデルが、化合物と結合部位との間のポーズを解釈する方法である。モデルは、化合物及びポリマーについての構造情報が提供されているにも関わらず、これら2つを別々に表す。そのため、ポリマー及び化合物の識別を可能にする任意の提供されたポーズは、同じスコアを与える。モデルは、ポーズに対し非感受性である。図19に示す。これは、重畳型ニューラルネットワークなどの機械学習モデルが、全ての正しい構成要素を持っているが、全体としては根本的に誤っているポーズを誤って選ぶ可能性がある、ピカソ問題に例示されている。図18に示す。左側のポーズ及び右側のポーズは、どちらも同じ部分、2つの目、2つの眉毛、鼻、唇、及び頭の全体形状を有する。したがって、左側のポーズが正しいことを重畳型ニューラルネットワークに教えることは困難であることが分かる。このため、従来のv H T S 機械学習方法には、固有のポーズ非感受性が存在する。このようなポーズ非感受性は、試験化合物と標的ポリマーとの間の相互作用の誤った又は不正確な特徴付けにつながる可能性がある。例えば、このようなポーズ非感受性は、スクリーニングライブラリ内の各化合物の分類的な活性標識を提供するv H T S 機械学習アプローチに、スクリーニングライブラリ内の化合物の特定割合を誤って標識する可能性がある。

10

【0005】

上記の背景を考えると、当該技術分野で必要とされるのは、v H T S 機械学習方法にポーズ感受性を与えるための方法である。

【発明の概要】

【0006】

20

本開示は、v H T S 機械学習モデルがポーズ感受性であるように、それらをコンディショニングすることによって、背景技術で識別された問題に対処する。そのようなモデルは、それぞれの訓練化合物と標的ポリマーとの間の相互作用の特徴付けが既知である訓練化合物について訓練される。しかしながら、各そのような訓練化合物に対して、v H T S 機械学習モデルは、訓練化合物の正のポーズ及び訓練化合物の負のポーズの両方について訓練され、そのような正のポーズ及び負のポーズは、独立したポーズ生成プロセスを使用して選択される。このようにして、v H T S 機械学習モデルは、ポーズ感受性であるように訓練される。

【0007】

したがって、本開示の一態様は、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するためのコンピュータシステムである。コンピュータシステムは、1つ以上のプロセッサと、1つ以上のプロセッサによってアドレス指定可能なメモリと、を備える。メモリは、1つ以上のプロセッサによる実行のための少なくとも1つのプログラムを記憶する。いくつかの実施形態では、試験化合物と標的ポリマーとの間の相互作用の特徴付けは、バイナリ活性スコアである。いくつかの実施形態では、標的ポリマーは、タンパク質、ポリペプチド、ポリ核酸、ポリリボ核酸、多糖、又はそれらの任意の組み合わせのアセンブリである。本開示によれば、標的ポリマーの複数の原子座標が取得される。いくつかの実施形態では、複数の原子座標は、少なくとも400個の原子の原子座標を含む。

30

【0008】

いくつかの実施形態では、複数の原子座標は、分解能2.5 Å以上又は分解能3.3 Å以上で分解された標的ポリマーの結晶構造の三次元座標のセット $\{x_1, \dots, x_N\}$ である。

40

【0009】

いくつかの実施形態では、標的ポリマーの複数の原子座標は、核磁気共鳴、中性子回折、又は低温電子顕微鏡によって判定された標的ポリマーの三次元座標の集合体を含む。

【0010】

いくつかの実施形態では、試験化合物と標的ポリマーとの間の相互作用の特徴付けは、バイナリスコアであり、バイナリスコアの第1の値は、第1の閾値を上回る標的ポリマーに関する試験化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iを表し、バイナリスコアの第2の値は、第1の閾値を下回る標的ポリマーに関する試験化合物のIC₅₀、E

50

C₅₀、K_d、K_I、又はpK_Iを表す。

【0011】

本開示によれば、複数の訓練化合物における各訓練化合物のそれぞれの電子記述を含む訓練データセットが取得される。いくつかの実施形態では、複数の訓練化合物は、少なくとも100個の化合物を含む。各それぞれの電子記述は、(i) 対応する第1の正の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する正のポーズと、(ii) 対応する第1の負の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する負のポーズと、を含む。

【0012】

いくつかの実施形態では、標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアは、対応する正のポーズにおける標的ポリマーに関する対応する訓練化合物の対応する正のボクセルマップを検索し、対応する正のボクセルマップを、対応する正のベクトルに展開し、及び対応する正のベクトルを、ニューラルネットワークに入力することによって取得され、それによって、対応する正のポーズの対応する正のスコアを取得する。いくつかの実施形態では、ニューラルネットワークは、500個よりも多いパラメータを含む。いくつかのそのような実施形態では、対応する正のベクトルは、第1の一次元ベクトルである。

10

【0013】

いくつかの実施形態では、標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアは、対応する負のポーズにおける標的ポリマーに関する対応する訓練化合物の対応する負のボクセルマップを検索し、対応する負のボクセルマップを、対応する負のベクトルに展開し、及び対応する負のベクトルを、ニューラルネットワークに入力することによって取得され、それによって、標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアを取得する。

20

【0014】

いくつかのそのような実施形態では、対応する負のベクトルは、第2の一次元ベクトルである。

【0015】

いくつかの実施形態では、対応する第1の正の相互作用スコア及び対応する第1の負の相互作用スコアは、各々、標的ポリマーに対する対応する訓練化合物の結合係数又はインシリコポーズ品質スコアを表す。

30

【0016】

いくつかの実施形態では、訓練データセット内の各訓練化合物は、リピンスキーのルールオブファイブ：(i) 5個以下の水素結合ドナー、(ii) 10個以下の水素結合アクセプター、(iii) 500ダルトン未満の分子量、及び(iv) 5未満のLogP、のうちの2つ以上のルール、3つ以上のルール、又は4つのルール全てを満たす。

【0017】

いくつかの実施形態では、訓練データセット内の各訓練化合物は、500ダルトン未満、1000ダルトン未満、2000ダルトン未満、4000ダルトン未満、6000ダルトン未満、8000ダルトン未満、10000ダルトン未満、又は20000ダルトン未満の分子量を有する有機化合物である。

40

【0018】

本開示によれば、少なくとも第1のモデルが訓練される。第1のモデルは、第1の複数のパラメータを有する。いくつかの実施形態では、第1の複数のパラメータは、400個よりも多いパラメータを含む。訓練は、複数の訓練化合物中の各対応する訓練化合物46に対して、少なくとも(i) 標的ポリマーに関する対応する訓練化合物の対応する第1の正の相互作用スコアに対する、第1のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアと、(ii) 標的ポリマーに関する対応する訓練化合物の対応する第1の負の相互作用スコアに対する、第1のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応

50

する負のスコアと、を使用し、それによって、第1の複数のパラメータを調整し、少なくとも第1のモデルの出力は、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用される。

【0019】

いくつかの実施形態では、第1のモデルは、第1の完全接続ニューラルネットワークである。

【0020】

いくつかの実施形態では、訓練は、第1の複数のパラメータが、関連付けられた損失関数を通じた逆伝播によって調整される、回帰タスクである。そのような実施形態では、対応する第1の正の相互作用スコアは、式

10

$$B = N \times A$$

によって、対応する第1の負の相互作用スコアに関係付けられており、式中、Aは、対応する正の相互作用スコアであり、Bは、対応する負の相互作用スコアであり、Nは、ゼロ超かつ1未満の実数である（例えば、0.90）。

【0021】

いくつかのそのような実施形態では、関連付けられた損失関数は、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数である。

【0022】

いくつかのそのような実施形態では、対応する第1の正の相互作用スコア及び対応する第1の負の相互作用スコアは、各々、結合係数を表し、対応する第1の正の相互作用スコアは、標的ポリマーに対する対応する訓練化合物の結合係数のインビトロ測定値である。

20

【0023】

いくつかのそのような実施形態では、第1の正の相互作用スコアは、標的ポリマーに関するそれぞれの訓練化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iである。

【0024】

いくつかの実施形態では、訓練データセット内の各それぞれの電子記述は、対応する訓練化合物の対応する正のポーズの対応する正の活性スコアと、対応する訓練化合物の対応する負のポーズの対応する負の活性スコアと、を更に含む。そのような実施形態では、少なくとも第1のモデルを訓練することは、第2のモデルを第1のモデルと共同で訓練することを更に含む。第2のモデルは、第2の複数のパラメータを有する。そのような訓練は、複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも(iii)対応する訓練化合物の対応する正の活性スコアに対する、第2のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアと、(iv)対応する訓練化合物の対応する負の活性スコアに対する、第2のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアと、を更に使用する。このようにして、第2の複数のパラメータは、第2のモデルが、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために、第1のモデルの出力とともに使用される、試験化合物と標的ポリマーとの間の相互作用の活性を提供するように、調整される。

30

【0025】

いくつかのそのような実施形態では、第2のモデルは、第2の完全接続ニューラルネットワークである。

【0026】

いくつかの実施形態では、訓練データセット内の各それぞれの電子記述は、対応する訓練化合物の対応する正のポーズの対応する正の活性スコアと、対応する訓練化合物の対応する負のポーズの対応する負の活性スコアと、を更に含む。そのような実施形態では、少なくとも第1のモデルを訓練することは、第2のモデルを第1のモデルと共同で訓練することを更に含み、第2のモデルは、第2の複数のパラメータを有する。そのような実施形態における訓練は、複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも（

40

50

i i i) 対応する訓練化合物の対応する正の活性スコアに対する、第2のモデルへの結合入力としての標的ポリマー及び対応する第1の正の相互作用スコアに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアと、(i v) 対応する訓練化合物の対応する負の活性スコアに対する、第2のモデルへの結合入力としての標的ポリマー及び対応する第1の負の相互作用スコアに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアと、を更に使用する。このようにして、第2の複数のパラメータは、第2のモデルが、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために、第1のモデルの出力とともに使用され得るように、調整される。

【0027】

いくつかのそのような実施形態では、対応する正の活性スコアは、第1のバイナリ活性スコアであり、対応する負の活性スコアは、第2のバイナリ活性スコアである。いくつかの実施形態では、対応する第1のバイナリ活性スコアは、標的ポリマーに対する対応する化合物の測定された活性に基づいて、1の値を割り当てられており、対応する第2のバイナリ活性スコアは、0の値を割り当てられている。いくつかのそのような実施形態では、第1のモデルの訓練は、第1の複数のパラメータが、第1の関連付けられた損失関数を通じた逆伝播によって調整されている、回帰タスクであり、第2のモデルの訓練は、第2の複数のパラメータが、第2の関連付けられた損失関数を通じた逆伝播によって調整されている、分類タスクである。いくつかのそのような実施形態では、対応する第1の正の相互作用スコア及び対応する第1の負の相互作用スコアは、各々、標的ポリマーに対する対応する訓練化合物の結合係数又はインシリコポーズ品質スコアを表し、対応する正の活性スコアは、第1のバイナリ活性スコアであり、対応する負の活性スコアは、第2のバイナリ活性スコアである。いくつかのそのような実施形態では、第1の関連付けられた損失関数は、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数であり、第2の関連付けられた損失関数は、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数である。いくつかのそのような実施形態では、第2のモデルは、第2の完全接続ニューラルネットワークである。

【0028】

いくつかの実施形態では、訓練データセット内の各それぞれの電子記述は、対応する訓練化合物の対応する正のポーズの対応する第2の正の相互作用スコアと、対応する訓練化合物の対応する負のポーズの対応する第2の負の相互作用スコアと、を更に含む。そのような実施形態では、訓練データセット内のそれぞれの電子記述は、対応する訓練化合物の対応する正のポーズの対応する正の活性スコアと、対応する訓練化合物の対応する負のポーズの対応する負の活性スコアと、を更に含む。そのような実施形態では、少なくとも第1のモデルを訓練することは、第2のモデル及び第3のモデルを第1のモデルと共同で訓練することを更に含む。第2のモデルは、第2の複数のパラメータを有し、第3のモデルは、第3の複数のパラメータを有する。そのような実施形態では、複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも(i i i) 標的ポリマーに関する対応する訓練化合物の対応する第2の正の相互作用スコアに対する、第2のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアと、(i v) 標的ポリマーに関する対応する訓練化合物の対応する第2の負の相互作用スコアに対する、第2のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアであって、それによって、第2の複数のパラメータを調整する、対応する負のスコアと、(v) 対応する訓練化合物の対応する正の活性スコアに対する、第3のモデルへの結合入力としての、標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコア、対応する訓練化合物の対応する正のポーズの対応する正のスコアの入力時の第1のモデル及び第2のモデルの出力と、(v i) 対応する訓練化合物の対応する負の活性スコアに対する、第3のモデルへの結合入力としての、標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコア、並びに対応する訓練化合物の対応する負のポーズの対応する負のスコアの入力

10

20

30

40

50

時の、第1のモデル及び第2のモデルの出力と、を更に使用し、それによって、第3のモデルの第3の複数のパラメータを調整する。そのような実施形態では、第3のモデルの出力は、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供する。いくつかのそのような実施形態では、第2のモデルは、第2の完全接続ニューラルネットワークであり、第3のモデルは、第3の完全接続ニューラルネットワークである。いくつかのそのような実施形態では、対応する正の活性スコアは、第1のバイナリ活性スコアであり、対応する負の活性スコアは、第2のバイナリ活性スコアである。いくつかの実施形態では、対応する第1のバイナリ活性スコアは、標的ポリマーに対する対応する化合物の測定された活性に基づいて、1の値を割り当てられており、対応する第2のバイナリ活性スコアは、0の値を割り当てられている。いくつかのそのような実施形態では、第1のモデルの訓練は、第1の複数のパラメータが、第1の関連付けられた損失関数を通した逆伝播によって調整されている、第1の回帰タスクであり、第2のモデルの訓練は、第2の複数のパラメータが、第2の関連付けられた損失関数を通した逆伝播によって調整されている、第2の回帰タスクであり、第3のモデルの訓練は、第3の複数のパラメータが、第3の関連付けられた損失関数を通した逆伝播によって調整されている、分類タスクである。いくつかのそのような実施形態では、対応する第1の正の相互作用スコア及び対応する第1の負の相互作用スコアは、各々、標的ポリマーに対する対応する訓練化合物のインシリコポーズ品質スコアを表し、対応する第2の正の相互作用スコア及び対応する第2の負の相互作用スコアは、各々、標的ポリマーに対する対応する訓練化合物の結合係数を表し、対応する正の活性スコアは、第1のバイナリ活性スコアであり、対応する負の活性スコアは、第2のバイナリ活性スコアである。いくつかのそのような実施形態では、第1の関連付けられた損失関数は、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数であり、第2の関連付けられた損失関数は、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数であり、第3の関連付けられた損失関数は、バイナリクロスエントロピー損失関数、ヒンジ損失関数、二乗ヒンジ損失関数、又は第1の若しくは第2の関連付けられた損失関数として使用されるような本明細書に記載の任意の他の損失関数である。

10

20

【0029】

本開示の別の態様は、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を提供し、方法は、メモリを備えるコンピュータシステムにおいて、標的ポリマーの複数の原子座標を取得することを含む。いくつかの実施形態では、複数の原子座標は、少なくとも400個の原子の原子座標を含む。訓練データセットが取得される。複数の訓練化合物における各訓練化合物のそれぞれの電子記述を含む訓練データセット。いくつかの実施形態では、複数の訓練化合物は、少なくとも100個の化合物を含む。各それぞれの電子記述は、(i) 対応する第1の正の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する正のポーズと、(ii) 対応する第1の負の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する負のポーズと、を含む。少なくとも第1のモデルが訓練され、第1のモデルは、第1の複数のパラメータを有する。いくつかの実施形態では、第1の複数のパラメータは、400個よりも多いパラメータを含む。訓練は、複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも(i) 標的ポリマーに関する対応する訓練化合物の対応する第1の正の相互作用スコアに対する、第1のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアと、(ii) 標的ポリマーに関する対応する訓練化合物の対応する第1の負の相互作用スコアに対する、第1のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアと、を使用する。このようにして、第1の複数のパラメータが調整される。訓練後、少なくとも第1のモデルの出力は、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用される。

30

40

【0030】

50

本開示の別の態様は、非一時的なコンピュータ可読記憶媒体を提供する。非一時的なコンピュータ可読記憶媒体は、コンピュータシステムによって実行されるときに、コンピュータシステムに、方法に従って、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を実行させる命令を記憶する。方法は、標的ポリマーの複数の原子座標を取得することを含む。いくつかの実施形態では、複数の原子座標は、少なくとも400個の原子の原子座標を含む。複数の訓練化合物における各訓練化合物のそれぞれの電子記述を含む訓練データセットが取得される。複数の訓練化合物は、少なくとも100個の化合物を含む。各それぞれの電子記述は、(i) 対応する第1の正の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する正のポーズと、(ii) 対応する第1の負の相互作用スコアと結合された複数の原子空間座標に関する対応する訓練化合物の対応する負のポーズと、を含む。方法では、少なくとも第1のモデルが訓練される。第1のモデルは、第1の複数のパラメータを有する。いくつかの実施形態では、第1の複数のパラメータは、400個よりも多いパラメータを含む。訓練は、複数の訓練化合物中の各対応する訓練化合物に対して、少なくとも(i) 標的ポリマーに関する対応する訓練化合物の対応する第1の正の相互作用スコアに対する、第1のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する正のポーズの対応する正のスコアと、(ii) 標的ポリマーに関する対応する訓練化合物の対応する第1の負の相互作用スコアに対する、第1のモデルへの入力としての標的ポリマーに関する対応する訓練化合物の対応する負のポーズの対応する負のスコアと、を使用し、それによって、第1の複数のパラメータを調整する。少なくとも第1のモデルの出力は、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用される。

10

20

【0031】

図面では、本開示のシステム及び方法の実施形態が例として示されている。説明及び図面は、単に例示を目的とし、理解を助けるものであり、本開示のシステム及び方法の限定の定義として意図されないことを明確に理解されたい。

【図面の簡単な説明】

【0032】

【図1】 本開示のいくつかの実施形態による、コンピュータシステムを示す。

【図2A】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

30

【図2B】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図2C】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図2D】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図2E】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図2F】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

40

【図2G】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図2H】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図2I】 本開示のいくつかの実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるための方法を示す。

【図3】 本開示のいくつかの実施形態による、標的ポリマーに対するポーズにおける例示的な訓練化合物の概略図である。

【図4】 本開示のいくつかの実施形態による、ボクセルの三次元グリッドの形態である入力特徴の幾何学的表現の概略図である。

50

【図 5】本開示のいくつかの実施形態による、ボクセルの二次元グリッド上に符号化された化合物の図である。

【図 6】本開示のいくつかの実施形態による、ボクセルの二次元グリッド上に符号化された化合物の図である。

【図 7】本開示のいくつかの実施形態による、ボクセルに番号が付けられた図 6 の視覚化の図である。

【図 8】本開示のいくつかの実施形態による、原子中心の座標位置の形態である入力特徴の幾何学的表現の概略図である。

【図 9 A】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、化合物結合モードスコアであり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

10

【図 9 B】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、活性及び化合物結合モードスコアであり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

【図 9 C】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練され、最終出力モデルは、2つの異なるポーズ品質モデルに対してコンディショニングされる。

【図 10】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、(i) バイナリ離散活性及び (ii) pK_i であり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

20

【図 11】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは pK_i であり、 pK_i は、部分的に、活性に対してコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

【図 12】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、活性は、部分的に、 pK_i 及びポーズ品質スコアの両方に対してコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

30

【図 13】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、活性は、部分的に、 pK_i 及び化合物結合モードスコアの両方に対してコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

【図 14】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、活性及び2つの異なる化合物結合モードスコアであり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

【図 15】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、活性、2つの異なる化合物結合モードスコア及び pK_i であり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

40

【図 16 A】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、活性は、部分的に、 pK_i 及び結合モードスコアに対してコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。

【図 16 B】本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、活性は、部分的に、 pK_i 及び2つの異なる結合モードスコアに対してコンディショニングされ、システムは、化合物

50

を訓練するための結合された正及び負のポーズを使用して訓練される。

【図 17】本開示のいくつかの実施形態による、複数の関数計算素子 (g_1 、 g_2 、...) をボクセル入力 (x_1 、 x_2 、...、 x_{100}) に適用することと、 $g()$ を使用して、関数計算素子出力と一緒に構成することと、の描写である。

【図 18】従来技術による、標的ポリマーに関する化合物のポーズを特徴付ける際に、機械学習モデルが直面する非感受性を示す。

【図 19】化合物-ポリマーポーズの品質に対する従来の機械学習モデルの非感受性を示し、図に示すように、考え得る最良のポーズは、機械学習モデルによる、悪いポーズと同じスコアを受け取り、現実にはあり得ないポーズは、機械学習モデルによる、考え得る最良のポーズと同じスコアを受け取る。

【図 20】注釈付きの ATP 結合部位 (灰色)、アロステリック部位 (赤色)、及び SH2 ドメインにおける対照結合部位 (青色) を有する、ヒト ZAP70 タンパク質を示す。使用 PDB ID: 2ozo。

【図 21】本開示の実施形態による、受信者動作曲線 AUC 性能の様々なベンチマークを示す。

【図 22】本開示の実施形態による、約 300 個のキナーゼ阻害剤 (1 と標識された、結合剤) と混合された 10^5 個の多様な化合物 (0 と標識された、非結合剤) が、以下の 3 つの結合部位とドッキングしてスコアを付けた、ピカソ問題の実験を示す: i) ATP 結合部位、ii) アロステリック結合部位、及び iii) SH2 ドメインにおける結合部位。

【図 23】本開示の実施形態による、良いポーズと、悪いポーズ (左側のパネル) 又は現実にはあり得ないポーズ (右側のパネル) との間のメジアン確率低下を示す。

【図 24】本開示の実施形態による、PoseRanker 及び Vina スコアに対してコンディショニングされた活性タスクを示す。

【0033】

同様の参照番号は、図面全体を通して対応する部分を指す。

【発明を実施するための形態】

【0034】

ここで、添付の図面にその例が示される実施形態を詳細に参照する。以下の詳細な説明には、本開示の完全な理解を提供するために、多くの特定の詳細が示されている。しかしながら、本開示は、これらの特定の詳細なしで実施され得ることが当業者に明らかであろう。他の場合では、実施形態の態様を不必要に曖昧にしないように、周知の方法、手順、構成要素、回路、及びネットワークは、詳細には説明されていない。

【0035】

また、第 1 の、第 2 のなどの用語は、様々な要素を説明するために本明細書で使用され得るが、これらの要素はこれらの用語によって制限されるべきではないことも理解されたい。これらの用語は、ある要素を別の要素と区別するためにのみ使用される。例えば、第 1 の対象は、本発明の範囲から逸脱することなく、第 2 の対象と称され得、同様に、第 2 の対象は、第 1 の対象と称され得る。第 1 の対象と第 2 の対象はどちらも対象であるが、同じ対象ではない。

【0036】

本開示に使用される用語は、特定の実施形態を記載する目的のためだけであり、本発明を制限することを意図しない。本発明の説明及び添付の特許請求の範囲の説明で 사용되는場合、単数形の「a」、「an」及び「the」は、文脈により明らかにそうではないと指示されない限り、複数形態も同様に含むことが意図される。「及び／又は (and/or)」という用語は、本明細書で 사용되는場合、関連する列挙した品目のうちの 1 つ以上の任意の及び全ての可能な組み合わせを指し、包含することも理解されるであろう。「含む (comprises)」及び／又は「含む (comprising)」という用語は、本明細書で 사용되는場合、記載された特徴、整数、ステップ、動作、要素、及び／又は構成要素の存在を指定するが、1 つ以上の他の特徴、整数、ステップ、動作、要素

10

20

30

40

50

、構成要素、及び／又はそれらの群の存在又は追加を排除しないことが更に理解されるであろう。

【0037】

本明細書で使用される場合、用語「～する場合 (i f) 」は、文脈に応じて、「～するとき (w h e n) 」又は「～した後 (u p o n) 」又は記載された先行する条件が真であることの「判定に応じて」又は「判定に従って」又は「文脈に応じて」を意味すると解釈され得る。同様に、「判定された場合」又は「[記載された条件又は事象]が検出された場合」という句は、文脈に応じて、「判定時」又は「判定に応答して」又は「[記載された条件又は事象]の検出時」、又は「[記載された条件又は事象]を検出したとき」を意味すると解釈され得る。

10

【0038】

本開示は、ポリマーの配位、及び化合物の訓練データセットを使用して、試験化合物とポリマーとの間の相互作用を特徴付けるためのシステム及び方法を提供する。各それぞれの訓練化合物は、正の相互作用スコアを有する標的ポリマー配位に関する正のポーズを有する。化合物の訓練データセット内のそれぞれの訓練化合物のうちの少なくともいくつかはまた、標的ポリマー配位に関するそれぞれの訓練化合物の負のポーズ、及び負の相互作用スコアも有する。モデルは、訓練セット内の各それぞれの化合物に対して、少なくとも (i) 化合物の正の相互作用スコアに対する、モデルへの入力としての正のポーズの正のスコアと、 (i i) ある場合は、化合物の負の相互作用スコアに対する、モデルへの入力としての負のポーズの負のスコアと、を適用することによって訓練され、それによって、モデルのパラメータを調整し、訓練セット内の化合物のうちの少なくともいくつかは、正のポーズ及び負のポーズの両方を有する。いくつかの実施形態では、訓練セット内の化合物の少なくとも5%、10%、20%、50%、又は70%は、正のポーズ及び負のポーズの両方を有し、一方、訓練セット内の残りの化合物は、正のポーズのみを有する。いくつかの実施形態では、訓練セット内の化合物の全ては、正のポーズ及び負のポーズの両方を有する。

20

【0039】

いくつかの実施形態では、正のポーズの正のスコアは、ポリマーに関するそれぞれの正のポーズにおいて、それぞれの訓練化合物の対応する正のボクセルマップを形成することによって取得される。いくつかの実施形態では、対応する正のボクセルマップは、ベクトル化され、ニューラルネットワークに供給される。いくつかの実施形態では、ボクセルマップは、ベクトル化なしでニューラルネットワークに入力される。

30

【0040】

いくつかの実施形態では、ニューラルネットワークは、重畳型ニューラルネットワークである。いくつかのそのような実施形態では、重畳型ニューラルネットワークは、入力層、複数の個別に重み付けされた畳み込み層、及び出力スコアラを含む。畳み込み層は、初期層及び最終層を含む。入力に応答して、入力層は、初期畳み込み層に値を供給する。最終畳み込み層以外の各それぞれの畳み込み層は、それぞれの畳み込み層の重みの関数としての中間値と、それぞれの畳み込み層の入力値と、を畳み込み層のうちの他の一つに供給する。最終畳み込み層は、最終層の重み及び入力値の関数として、値をスコアラに供給する。このようにして、スコアラは、それぞれの化合物の正のポーズのスコアを付け、それぞれの化合物の正のポーズの正のスコアに達する。

40

【0041】

いくつかの実施形態では、負のポーズの負のスコアは、ポリマーに関するそれぞれの負のポーズにおいて、それぞれの訓練化合物の対応する負のボクセルマップを形成することによって取得される。いくつかの実施形態では、対応する負のボクセルマップは、ベクトル化され、上で説明されるニューラルネットワーク (例えば、重畳型ニューラルネットワーク) に供給される。いくつかの実施形態では、ボクセルマップは、ベクトル化なしでニューラルネットワークに入力される。このようにして、スコアラは、それぞれの化合物の負のポーズのスコアを付け、それぞれの化合物の負のポーズの負のスコアに達する。

50

【0042】

モデルが訓練化合物に対して訓練されると、モデルを使用して、試験化合物とポリマーとの間の相互作用を特徴付けることができる。いくつかの実施形態では、試験化合物及び標的ポリマーの正のポーズがニューラルネットワークに入力されることに応答して、正のポーズのスコアが、ニューラルネットワーク及び第2の（又は第3の、第4の、…第xの）モデルによって提供される。埋設層を介したコンディショニング時、ニューラルネットワークによって提供された正のポーズのスコアは、訓練されたモデルへの入力として機能し、これにより、試験化合物とポリマーとの間の相互作用の特徴付けが提供される。

【0043】

図1は、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのコンピュータシステム100を示す。それは、例えば、1つ以上の試験化合物の標的ポリマーとの結合親和性に関する正確な予測を生成するための結合親和性予測システムとして使用され得る。

10

【0044】

図1を参照すると、典型的な実施形態では、コンピュータシステム100は、1つ以上のコンピュータを備える。図1における例示の目的のために、コンピュータシステム100は、開示されたコンピュータシステム100の機能の全てを含む単一のコンピュータとして表される。しかしながら、本開示は、そのように限定されない。コンピュータシステム100の機能は、任意の数のネットワーク化されたコンピュータに分散され得、並びに／又はいくつかのネットワーク化されたコンピュータ及び／若しくは仮想マシンの各々に存在し得る。当業者は、多様な異なるコンピュータトポロジがコンピュータシステム100にとって可能であり、そのようなトポロジは全て本開示の範囲内であることを理解するであろう。

20

【0045】

上記を念頭に置いて図1を参照すると、コンピュータシステム100は、1つ以上の処理ユニット（CPU）59と、ネットワーク又は他の通信インターフェース84と、ユーザインターフェース78（例えば、ディスプレイ82及び任意選択的なキーボード80又は他の形態の入力デバイスを含む）と、メモリ92（例えば、ランダムアクセスメモリ）と、1つ以上のコントローラ88によって任意選択的にアクセスされる1つ以上の磁気ディスクストレージ及び／又は持続性デバイス90と、前述の構成要素を相互接続するための1つ以上の通信バス12と、前述の構成要素に電力を供給するための電源79と、を備える。メモリ92内のデータは、キャッシングなどの周知のコンピューティング技術を使用して、不揮発性メモリ90とシームレスに共有され得る。メモリ92及び／又はメモリ90は、中央処理装置（複数可）59に対して遠隔に位置している大容量ストレージを含むことができる。言い換えると、メモリ92及び／又はメモリ90に格納されたいくつかのデータは、実際には、コンピュータシステム100の外部にあるが、ネットワークインターフェース84を使用して、インターネット、イントラネット、又は他の形態のネットワーク若しくは電子ケーブルを経由してコンピュータシステム100によって電子的にアクセスすることができるコンピュータ上にホストされ得る。いくつかの実施形態では、コンピュータシステム100は、システムの速度及び性能を改善するために、1つ以上の図形処理ユニット50に関連付けられたメモリ52から実行されるニューラルネットワークを活用する。いくつかの代替の実施形態では、コンピュータシステム100は、図形処理ユニット50に関連付けられたメモリではなく、メモリ92から実行されるニューラルネットワークを活用する。

30

40

【0046】

コンピュータシステム100のメモリ92、及び／又は任意選択的にメモリ52は、

●様々な基本システムサービスを処理するための手順を含む、任意選択的なオペレーティングシステム34と、

●試験化合物と標的ポリマーとの間の相互作用を特徴付けるための空間データ評価モジュール36と、

50

●構造データ（例えば、標的ポリマーの複数の原子空間座標 40）及び／又は、任意選択的に、標的ポリマーの活性部位情報 42 を含む、標的ポリマー用のデータ 38 と、

●複数の訓練化合物における各訓練化合物のそれぞれの電子記述 46 を含む訓練データセット 44 であって、訓練データセット 44 の少なくともサブセットにおける各それぞれの電子記述は、（i）対応する第 1 の正の相互作用スコア 50 と結合された複数の原子空間座標 40 に関する対応する訓練化合物の対応する正のポーズ 48、及び（ii）対応する第 1 の負の相互作用スコア 62 と結合された複数の原子空間座標 40 に関する対応する訓練化合物の対応する負のポーズ 60 を含む、訓練データセット 44 と、

●第 1 の複数のパラメータ 73 を含む第 1 のモデル 72 であって、第 1 のモデルの出力は、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用される、第 1 のモデル 72 と、

●空間データにニューラルネットワーク 24 を適用するための（例えば、標的ポリマーにドッキングされた試験又は訓練化合物を試験にニューラルネットワークを適用するための）査定モジュール 20 と、

●ボクセルマップの 1 つ以上の（任意選択的に）ベクトル化された表現 54 / 66 と、

●任意選択的に入力層 26 を含み、任意選択的に 1 つ以上の畳み込み層 28 を含み、端末スコアラ 30 を含む、ニューラルネットワーク 24 と、

●第 2 の複数のパラメータ 75 を含む第 2 のモデル 74 であって、第 2 のモデルの出力は、少なくとも部分的に、（i）試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用され、及び／又は（ii）第 1 のモデルをコンディショニングするために使用される、第 2 のモデル 74 と、

●任意選択的に、第 3 の複数のパラメータ 77 を含む第 3 のモデル 76 であって、第 3 のモデルの出力は、少なくとも部分的に、（i）試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用され、及び／又は（ii）第 1 のモデル及び／又は第 2 のモデルをコンディショニングするために使用される、第 3 のモデル 76 と、

●任意選択的に、任意の数の追加の第 x のモデルであって、各々のそのような追加の第 x のモデルは、対応する複数のパラメータを含み、追加の第 x のモデルの出力は、少なくとも部分的に、（i）試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用され、及び／又は（ii）任意の他の単一モデル及び／又はモデルのグループをコンディショニングするために使用される、第 x のモデルと、を格納する。

【0047】

いくつかの実施態様では、コンピュータシステム 100 の上記の識別されたデータ要素又はモジュールのうちの 1 つ以上は、前述のメモリデバイスのうちの 1 つ以上に格納され、上で説明される機能を実行するための命令のセットに対応する。上記の識別されたデータ、モジュール又はプログラム（例えば、命令のセット）は、別個のソフトウェアプログラム、手順、又はモジュールとして実装される必要はなく、したがって、これらのモジュールの様々なサブセットは、様々な実装で組み合わせられ得るか、又は別様に再配置され得る。いくつかの実施態様では、メモリ 92 及び／又は 90（及び任意選択的に 52）は、任意選択的に、上記で識別されたモジュール及びデータ構造のサブセットを格納する。更に、いくつかの実施形態では、メモリ 92 及び／又は 90（及び任意選択的に 52）は、

【0048】

ここで、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムが開示されており、そのような特徴付けを実行するための方法を、図 2 を参照して詳述し、以下で考察する。

【0049】

ブロック 200。図 2 A のブロック 200 を参照すると、試験化合物と標的ポリマー 38 との間の相互作用の特徴付けを提供するコンピュータシステム 100 が開示されている。図 1 と併せて上記で考察されるように、コンピュータシステムは、1 つ以上のプロセッサ 74 と、1 つ以上のプロセッサによってアドレス指定可能なメモリ 90 / 92 と、を備

える。メモリは、1つ以上のプロセッサによる実行のための少なくとも1つのプログラムを記憶する。図2の残りは、コンピュータシステムの訓練と、訓練されたコンピュータシステムの使用と、を含む、少なくとも1つのプログラムの特徴を詳述する。

【0050】

ブロック202～ブロック204。図2Aのブロック202を参照すると、いくつかの実施形態では、基準化合物に対して訓練されると、空間データ評価モジュール36は、試験化合物と標的ポリマー38との間の相互作用を特徴付けることができる。いくつかのそのような実施形態では、この特徴付けは、離散（例えば、離散バイナリ）活性スコアである。言い換えると、特徴付けは分類的である。例えば、いくつかの実施形態では、特徴付けは離散バイナリであり、コンピュータシステムは、試験化合物が、空間データ評価モジュール36に実装され、以下で更に詳細に考察されるインシリコ法によって、標的ポリマーに対して活性であると判定されたとき、1つの値、例えば、「1」を提供し、試験化合物が標的ポリマーに対して活性でないと判定されたとき、別の値、例えば、「0」を提供する。

10

【0051】

いくつかの実施形態では、特徴付けは、バイナリ以外の離散スケールである。例えば、いくつかの実施形態では、特徴付けは、試験化合物が、空間データ評価モジュール36に実装され、以下で更に詳細に考察されるインシリコ法によって、第1の閾値を下回った活性を有すると判定されると、第1の値、例えば、「0」を提供し、試験化合物が、第1の閾値と第2の閾値との間の活性を有すると判定されると、第2の値、例えば、「1」を提供し、試験化合物が、第2の閾値を上回る活性を有すると判定されると、第3の値、例えば、「2」を提供する。そのような実施形態では、第1及び第2の閾値は所定のものであり、特定の実験（例えば、特定の標的ポリマーに対する試験化合物の特定のデータベース、セット、又は集合の特定の評価）に対して一定であり、試験ポリマーに対する活性のための試験化合物の特定のデータベース、セット、又は集合から好適な試験化合物を識別するのに有用であることが証明する値を有するように選択される。例えば、いくつかの実施形態では、本明細書に開示される閾値のいずれも、試験化合物のデータベースのうちの0.1パーセント以下、0.5パーセント以下、1パーセント以下、2パーセント以下、5パーセント以下、10パーセント以下、20パーセント以下、又は50パーセント以下を、標的ポリマーに対して活性であるとして識別するように設計されており、試験化合物のデータベースは、100個以上の化合物、1000個以上の化合物、10,000個以上の化合物、100,000個以上の化合物、 1×10^6 個の化合物、 10×10^6 個の化合物、又はそれ以上を含む。

20

30

【0052】

代替的な実施形態では、基準化合物に対して訓練されると、空間データ評価モジュール36は、試験化合物と標的ポリマー38との間の相互作用を連続スケールでの活性として特徴付けることができる。すなわち、空間データ評価モジュール36は、標的ポリマーに対する試験化合物の活性を示す連続スケールの数値を提供する。連続スケールの活性値は、例えば、訓練された空間データ評価モジュール36によって割り当てられた標的ポリマーに対して、試験化合物のデータベース内の各試験化合物の活性を比較するのに有用である。

40

【0053】

ブロック204を参照すると、開示されたシステム及び方法は、試験化合物と標的ポリマー38との間の相互作用を連続スケール又は離散スケールの活性として特徴付けることに限定されない。代替的な実施形態では、空間データ評価モジュール36は、実際には、基準化合物に対して訓練されると、試験化合物と標的ポリマーとの間の相互作用を、連続スケール又は離散（分類的な）スケールの標的ポリマーに対する試験化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iとして特徴付けることができる。

【0054】

3つの可能な結果を有するバイナリ離散スケール及び離散スケールが識別されているが

50

、本開示は、試験化合物と標的ポリマー 38 との間の相互作用の特徴付けのための離散スケールのこれら 2 つの例に限定されない。実際には、任意の離散スケールを、非限定的な例として、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、又は 20 個の異なる結果を有する離散スケールを含む、試験化合物と標的ポリマー 38 との間の相互作用の特徴付けに使用することができる。

【0055】

ブロック 206。図 2A のブロック 204 を参照すると、いくつかの実施形態では、標的ポリマー 38 は、タンパク質、ポリペプチド、ポリ核酸、ポリリボ核酸、多糖、金属タンパク質、又はそれらの任意の組み合わせのアセンブリである。いくつかの実施形態では、標的ポリマー 38 は、反復残基からなる巨大分子である。いくつかの実施形態では、標的ポリマー 38 は、天然材料である。いくつかの実施形態では、標的ポリマー 38 は、合成材料である。いくつかの実施形態では、標的ポリマー 38 は、エラストマー、シェラック、こはく、天然又は合成ゴム、セルロース、ペークライト、ナイロン、ポリスチレン、ポリエチレン、ポリプロピレン、ポリアクリロニトリル、ポリエチレングリコール、又は多糖である。

【0056】

いくつかの実施形態では、標的ポリマー 38 は、ヘテロポリマー（コポリマー）である。コポリマーは、1 つだけのモノマーが使用されるホモポリマーとは対照的に、2 つ（又はそれ以上）のモノマー種に由来するポリマーである。共重合は、コポリマーを化学的に合成するために使用される方法を指す。コポリマーの例としては、ABS プラスチック、SBR、ニトリルゴム、スチレンアクリロニトリル、スチレン・イソプレン・スチレン（SIS）、及びエチレン酢酸ビニルが挙げられるが、これらに限定されない。コポリマーは、少なくとも 2 つのタイプの構成単位を（構造単位、又は粒子も同様に）含むため、コポリマーは、これらの単位の鎖に沿ってどのように配置されるかに基づいて分類することができる。これらは、A 単位及び B 単位が規則的に交互になっている、交互コポリマーを含む。例えば、Jenkins, 1996, "Glossary of Basic Terms in Polymer Science," Pure Appl. Chem. 68 (12): 2287-2311 を参照されたく、これは参照によりその全体が本明細書に組み込まれる。コポリマーの追加の例は、反復配列（例えば、 $(A-B-A-B-B-A-A-A-A-B-B-B)_n$ ）に配置された A 単位及び B 単位を有する周期的コポリマーである。コポリマーの追加の例は、コポリマー中のモノマー残基の配列が統計的法則に従う、統計的コポリマーである。例えば、Painter, 1997, Fundamentals of Polymer Science, CRC Press, 1997, p 14 を参照されたく、これは参照によりその全体が本明細書に組み込まれる。開示されたシステム及び方法を使用して評価され得るコポリマーの更に他の例は、共有結合によって結び付いている 2 つ以上のホモポリマーサブユニットを含むブロックコポリマーである。ホモポリマーサブユニットの結合は、ジャンクションブロックとして知られている、中間非反復サブユニットを必要とし得る。2 つ又は 3 つの別個のブロックを有するブロックコポリマーは、それぞれ、ジブロックコポリマー及びトリブロックコポリマーと呼ばれる。

【0057】

いくつかの実施形態では、標的ポリマー 38 は、50 個以上、100 個以上、150 個以上、200 個以上、300 個以上、400 個以上、500 個以上、600 個以上、700 個以上、800 個以上、900 個以上、又は 1000 個以上の原子を含む。

【0058】

いくつかの実施形態では、標的ポリマー 38 は、実際には、複数のポリマー（例えば、2 個以上、3 個以上、10 個以上、100 個以上、1000 個以上、又は 5000 個以上のポリマー）であり、複数のポリマー中のそれぞれのポリマーは、全て同じ分子量を有するわけではない。いくつかのそのような実施形態では、複数のポリマー中の標的ポリマー 38 は、少なくとも 50%、少なくとも 60%、少なくとも 70%、少なくとも 80%、

又は少なくとも90%の配列同一性を共有し、対応する鎖長の分布を有する重量範囲に入る。いくつかの実施形態では、標的ポリマー38は、1つ以上の置換基の側鎖又は分岐を有する主鎖を含む分岐ポリマー分子である。分岐ポリマーのタイプとしては、星型ポリマー、形ポリマー、ブラシポリマー、デンドロン化ポリマー、ラダー、及びデンドリマーが挙げられるが、これらに限定されない。例えば、Rubinstein et al., 2003, Polymer physics, Oxford; New York: Oxford University Press. p. 6を参照されたく、これは参照によりその全体が本明細書に組み込まれる。

【0059】

いくつかの実施形態では、標的ポリマーは、ポリペプチドである。本明細書で 사용되는場合、「ポリペプチド」という用語は、ペプチド結合によって結び付いている2つ以上のアミノ酸又は残基を意味する。「ポリペプチド」及び「タンパク質」という用語は、本明細書では同義に使用され、オリゴペプチド及びペプチドを含む。「アミノ酸」、「残基」又は「ペプチド」は、プロリン及びヒドロキシプロリンなどのイミノ酸を含む、当技術分野で周知のように、タンパク質の20個の標準的な構造単位のうちのいずれかを指す。アミノ酸異性体の名称には、D、L、R、及びSが含まれ得る。アミノ酸の定義には、非天然アミノ酸が含まれる。したがって、セレノシステイン、ピロリジン、ランチオニン、2-アミノイソ酪酸、ガンマーアミノ酪酸、デヒドロアラニン、オルニチン、シトルリン、及びホモシステインは、非限定的な例として、全てアミノ酸とみなされる。アミノ酸の他の変異体又は類似体は、当技術分野で周知である。したがって、ポリペプチドは、ペプトイドなどの合成ペプチド類似構造を含み得る。Simon et al., 1992, Proceedings of the National Academy of Sciences USA, 89, 9367を参照されたく、これは参照によりその全体が本明細書に組み込まれる。Chin et al., 2003, Science 301, 964及びChin et al., 2003, Chemistry & Biology 10, 511もまた参照されたく、これらの各々は参照によりその全体が本明細書に組み込まれる。

【0060】

開示されたシステム及び方法のいくつかの実施形態に従って評価された標的ポリマー38はまた、任意の数の翻訳後修飾も有し得る。したがって、標的ポリマー38は、アシル化、アルキル化、アミド化、ビオチン化、ホルミル化、 γ -カルボキシル化、グルタミル化、グリコシル化、グリシル化、ヒドロキシル化、ヨウ素化、イソプレニル化、リポイル化、（例えば、ヘム、フラビン、金属などの）補因子付加、ヌクレオシド及びそれらの誘導体の付加、酸化、還元、peg化、ホスファチジルイノシトール付加、ホスホパンテテイル化、リン酸化、ピログルタミン酸形成、ラセミ化、tRNAによるアミノ酸の付加（例えば、アルギニル化）、硫酸化、セレノイル化、ISG化、SUMO化、ユビキチン化、化学修飾（例えば、シトルリン化及びアミド分解）、並びに他の酵素（例えば、プロテアーゼ、ホスファターゼ及びキナーゼ）による処理によって修飾されているそれらのポリマーを含む。他のタイプの翻訳後修飾は、当技術分野で周知であり、本開示の標的ポリマー38の範囲内である。

【0061】

いくつかの実施形態では、標的ポリマー38は、界面活性剤である。界面活性剤は、液体の表面張力、2つの液体間の界面張力、又は液体と固体との間の界面張力を低下させる化合物である。界面活性剤は、洗剤、湿潤剤、乳化剤、発泡剤、及び分散剤として機能し得る。界面活性剤は、通常、両親媒性である有機化合物であり、疎水性基（それらの尾部）及び親水性基（それらの頭部）の両方を含有することを意味する。したがって、界面活性剤分子は、不水溶性（又は油溶性）成分及び水溶性成分の両方を含有する。界面活性剤分子は、水中に拡散し、水が油と混合される場合、空気と水との間の界面又は油と水との間の界面に吸着する。不溶性疎水性基は、バルク水相から、空气中又は油相中に延在し得、一方、水溶性頭部基は、水相中に留まる。表面での界面活性剤分子のこのようなアライ

メントは、水／空気又は水／油界面での水の表面特性を変更する。

【0062】

イオン性界面活性剤の例としては、アニオン性、カチオン性、又は両性イオン性（アンポテリック）界面活性剤などのイオン性界面活性剤が挙げられる。いくつかの実施形態では、標的物体58は、逆ミセル又はリポソームである。

【0063】

いくつかの実施形態では、標的ポリマー38は、フラーレンである。フラーレンは、中空球、楕円体又は管の形態で、完全に炭素からなる任意の分子である。球状フラーレンは、バッキーボールとも呼ばれ、サッカーで使用されるボールに似ている。円筒形のものは、カーボンナノチューブ又はバキータンチューブと呼ばれる。フラーレンは、構造がグラファイトに類似しており、グラファイトは、結び付いている六角形環の積層グラフェンシートからなるが、五角形（又は時には七角形）環もまた含有し得る。

10

【0064】

ブロック208～ブロック212。図2Aのブロック208を参照すると、標的ポリマー38の複数の原子座標40が取得される。いくつかの実施形態では、複数の原子座標は、標的ポリマーの少なくとも400個の原子の原子座標を含む。いくつかの実施形態では、複数の原子座標は、標的ポリマーの少なくとも25個の原子、少なくとも50個の原子、少なくとも100個の原子、少なくとも200個の原子、少なくとも300個の原子、少なくとも400個の原子、少なくとも1000個の原子、少なくとも2000個の原子、又は少なくとも5000個の原子の原子座標を含む。いくつかの実施形態では、配位子が標的ポリマーを結合させると見込まれる標的ポリマー38の活性部位の座標のみが取得される。ブロック210を参照すると、いくつかの実施形態では、複数の原子座標は、分解能2.5 Å以上又は分解能3.3 Å以上で分解された標的ポリマーの結晶構造の三次元座標のセット $\{x_1, \dots, x_N\}$ である。ブロック212を参照すると、いくつかの実施形態では、標的ポリマーの複数の原子座標は、核磁気共鳴、中性子回折、又は低温電子顕微鏡によって判定された標的ポリマーの三次元座標の集合体を含む。

20

【0065】

いくつかの実施形態では、複数の原子座標は、（例えば、X線結晶学的技術によって）3.3 Å以上、3.2 Å以上、3.1 Å以上、3.0 Å以上、2.5 Å以上、2.2 Å以上、2.0 Å以上、1.9 Å以上、1.85 Å以上、1.80 Å以上、1.75 Å以上、又は1.70 Å以上の分解能で分解された標的ポリマー38の結晶構造の三次元座標のセット $\{x_1, \dots, x_N\}$ である。

30

【0066】

いくつかの実施形態では、標的ポリマー38の複数の原子座標は、核磁気共鳴によって判定された標的ポリマーの10個以上、20個以上、30個以上又はそれ以上の三次元座標の集合体であり、集合体は、1.0 Å以上、0.9 Å以上、0.8 Å以上、0.7 Å以上、0.6 Å以上、0.5 Å以上、0.4 Å以上、0.3 Å以上、又は0.2 Å以上の骨格の平均二乗偏差（RMSD）を有する。いくつかの実施形態では、複数の原子座標は、中性子回折又は低温電子顕微鏡によって判定される。

【0067】

いくつかの実施形態では、標的ポリマー38は、ポリペプチドに結合した核酸などの、2つの異なるタイプのポリマーを含む。いくつかの実施形態では、天然の標的ポリマーは、互いに結合した2つのポリペプチドを含む。いくつかの実施形態では、研究中の天然の標的ポリマーは、1つ以上の金属イオン（例えば、1つ以上の亜鉛原子を有するメタロプロテインナーゼ）を含む。そのような場合では、金属イオン及び／又は有機小分子は、標的ポリマーの原子座標40に含まれ得る。

40

【0068】

いくつかの実施形態では、標的ポリマー38はポリマーであり、標的ポリマー中には、10個以上、20個以上、30個以上、50個以上、100個以上、100個～1000個、又は500個未満の残基が存在する。

50

【0069】

いくつかの実施形態では、標的ポリマー38の原子座標は、アブイニシオ法、密度関数法、半経験的及び経験的方法、分子力学、化学力学、又は分子力学などのモデリング方法を使用して判定される。

【0070】

いくつかの実施形態では、原子座標40は、標的ポリマー38を含む原子の中心のデカルト座標によって表される。いくつかの代替の実施形態では、標的ポリマー38の空間座標40は、例えば、X線結晶学によって測定される標的ポリマーの電子密度によって表される。例えば、いくつかの実施形態では、原子座標40は、標的ポリマー38の計算された原子座標を使用して算出された $2F_{\text{observed}} - F_{\text{calculated}}$ 電子密度マップを含み、 F_{observed} は、標的ポリマーの観察された構造因子の振幅であり、 F_c は、標的ポリマー38の計算された原子座標から計算された構造因子の振幅である。

10

【0071】

様々な他の実施形態では、標的ポリマー38の原子座標40は、ブロック206に従って、溶液NMRによって生成された構造集合体、X線結晶学、中性子回折、低温電子顕微鏡、コンピュータによるシミュレーションからのサンプリング、ホモロジーモデリング、回転異性体ライブラリのサンプリング、又はそれらの任意の組み合わせから解釈される共複合体を含むが、これらに限定されない多種多様な供給源から取得される。

【0072】

ブロック214。図2Bのブロック214を参照すると、複数の訓練化合物における各訓練化合物46のそれぞれの電子記述を含む訓練データセット44が取得される。いくつかの実施形態では、複数の訓練化合物は、少なくとも50個、100個、200個、1000個、5000個、10,000個、50,000個、100,000個、 1×10^6 個、 1×10^7 個、又は 1×10^8 個の訓練化合物を含む。訓練データセットの少なくともサブセットにおける各訓練化合物46のそれぞれの電子記述は、(i) 対応する第1の正の相互作用スコア50と結合された複数の原子空間座標に関する対応する訓練化合物46の対応する正のポーズ48、及び(ii) 対応する第1の負の相互作用スコア62と結合された複数の原子空間座標に関する対応する訓練化合物の対応する負のポーズ60を含む。図3は、標的ポリマー38の活性部位における訓練化合物46の正のポーズ48を示す。いくつかの実施形態では、訓練化合物46のいくつかは、負のポーズ60を有さず、対応する第1の負の相互作用スコア62を有しない。いくつかの実施形態では、訓練化合物46のいくつかは、正のポーズ48を有さず、対応する第1の正の相互作用スコア50を有しない。いくつかの実施形態では、訓練化合物46の全ては、正のポーズ及び負のポーズの両方、並びに対応する第1の正の相互作用スコア及び第1の負の相互作用スコアの両方を有する。

20

30

【0073】

いくつかの実施形態では、標的ポリマー38は、活性部位を有するポリマーであり、正及び負のポーズは、訓練化合物をポリマーの活性部位にドッキングすることによって取得される。いくつかの実施形態では、訓練化合物は、複数のポーズを形成するために、標的ポリマー38に複数回ドッキングされる。いくつかの実施形態では、各訓練化合物は、標的化合物38に、2回、3回、4回、5回以上、10回以上、50回以上、100回以上、又は1000回以上ドッキングされる。各々のそのようなドッキングは、標的ポリマー38にドッキングされた訓練化合物の異なるポーズを表す。いくつかの実施形態では、標的ポリマー38は、活性部位を有するポリマーであり、各訓練化合物は、複数の異なる方法の各々で活性部位にドッキングされており、各々のそのような方法は、異なるポーズを表す。これらのポーズの多くは正しくないと予想され、そのようなポーズは、訓練化合物と、自然に生じる標的ポリマーとの間の真の相互作用を表していないことを意味する。

40

【0074】

いくつかの実施形態では、訓練化合物の各ポーズは、AutoDock Vinaによって判定される。Trott and Olson, "AutoDock Vina: i

50

improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading,” *Journal of Computational Chemistry* 31 (2010) 455–461を参照されたい。そのような実施形態では、それぞれの訓練化合物について、AutoDock Vinaによる最良のスコアを受け取ったポーズには、正のポーズ48が割り当てられ、AutoDock Vinaによる最悪のスコアを受け取ったポーズには、負のポーズ60が割り当てられる。いくつかの実施形態では、異なるドッキングプログラムを使用して、それぞれの訓練化合物の正のポーズ48及び負のポーズ60を判定する。例えば、いくつかの実施形態では、Quick Vina 2 (Alhossary et al., 2015, “Fast, accurate, and reliable molecular docking with QuickVina,” *Bioinformatics* 31:13, pp. 2214–2216)、VinaLC (Zhang et al., 2013, “Message Passing Interface and Multithreading Hybrid for Parallel Molecular Docking of Large Databases on Petascale High Performance Computing Machines,” *J. Comput. Chem.* DOI:10.1002/jcc.23214)、Smina (Koes et al., 2013, “Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise,” *Journal of chemical information and modeling* 53:8, pp. 1893–1904)、又はCuina (Morrison et al., “Efficient GPU Implementation of AutoDock Vina,” COMP poster 3432389) が使用される。

10

20

30

40

50

【0075】

いくつかの実施形態では、正のポーズ48は、ポーズの正の集合体であり、負のポーズ60は、ポーズの負の集合体である。例えば、いくつかの実施形態では、正のポーズ48は、2～500個の構造的に類似したポーズからなる対応する第1の集合体であり、負のポーズ48は、2～500個の構造的に類似したポーズからなる対応する第2の集合体であり、対応する第1の集合体は、対応する第2の集合体よりも良い全体的なドッキングスコアを有する。そのような集合体を取得するための方法は、Stafford et al., 2019, “Modeling protein flexibility with conformational sampling improves ligand pose and bioactivity prediction,” *Abstracts of Papers of the American Chemical Society*, Volume 258に開示されており、これは参照により本明細書に組み込まれる。いくつかの実施形態では、各対応する第1の集合体（まとめて正のポーズ48を表す）は、2～30個、2～20個、2～10個、100個よりも多い、2～1000個の構造的に類似したポーズである。いくつかの実施形態では、各対応する第2の集合体（まとめて負のポーズ48を表す）は、2～30個、2～20個、2～10個、100個よりも多い、2～1000個の構造的に類似したポーズである。

【0076】

いくつかの実施形態では、（例えば、ポーズの集合体における）各ポーズは、標的タンパク質のいくつか（例えば、2～100個）の異なる立体構造に対してスコアリングされる。いくつかの実施形態では、（例えば、ポーズの集合体における）各ポーズは、標的タンパク質の固定された立体構造に対してスコアリングされる。

【0077】

いくつかの実施形態では、訓練化合物は、ランダムポーズ生成技術、又はバイアスポーズ生成のいずれかによって、標的ポリマー38にドッキングされる。いくつかの実施形態

では、訓練化合物は、マルコフ連鎖モンテカルロサンプリングによって、標的ポリマー 38 にドッキングされる。いくつかの実施形態では、そのようなサンプリングは、ドッキング計算における訓練化合物の十分な柔軟性、及び訓練化合物と標的ポリマー 38 との間の相互作用エネルギーと訓練（又はテスト）物体の立体構造エネルギーの和であるスコアリング関数を可能にする。例えば、Liu and Wang, 1999, "MCDOCK: A Monte Carlo simulation approach to the molecular docking problem," *Journal of Computer-Aided Molecular Design* 13, 435-451 を参照されたく、これは参照により本明細書に組み込まれる。そのような実施形態では、所与の訓練化合物について、最良のドッキングスコアを受け取ったポーズには、正のポーズ 48 が割り当てられ、最悪のドッキングスコアを受け取ったポーズには、正のポーズが割り当てられる。

10

【0078】

いくつかの実施形態では、DOCK（各々が参照により本明細書に組み込まれる、Shoichet, Bodian, and Kuntz, 1992, "Molecular docking using shape descriptors," *Journal of Computational Chemistry* 13 (3), pp. 380-397、及び Knegtel, Kuntz, and Oshiro, 1997 "Molecular docking to ensembles of protein structures," *Journal of Molecular Biology* 266, pp. 424-440）などのアルゴリズムを使用して、標的ポリマー 38 に対する訓練化合物の各々についての複数のポーズを見出す。そのようなアルゴリズムは、標的ポリマー 38 及び訓練化合物を剛体としてモデル化する。ドッキングされた立体構造を、相補的な表面を使用して検索し、ポーズを見出す。

20

【0079】

いくつかの実施形態では、AutoDOCK（各々が参照により本明細書に組み込まれる、Morris et al., 2009, "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility," *J. Comput. Chem.* 30 (16), pp. 2785-2791; Sotriffer et al., 2000, "Automated docking of ligands to antibodies: methods and applications," *Methods: A Companion to Methods in Enzymology* 20, pp. 280-291 及び "Morris et al., 1998, "Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function," *Journal of Computational Chemistry* 19: pp. 1639-1662）を使用して、標的ポリマー 38 に対する訓練化合物の各々についての複数のポーズを見出す。AutoDOCK は、配位子の動力学モデルを使用し、モンテカルロ、模擬アニーリング、ラマルク遺伝的アルゴリズム、及び遺伝的アルゴリズムをサポートしている。したがって、いくつかの実施形態では、（所与の訓練化合物についての）複数の異なるポーズは、ドッキングスコアリング関数を使用して、マルコフ連鎖モンテカルロサンプリング、模擬アニーリング、ラマルク遺伝的アルゴリズム、又は遺伝的アルゴリズムによって取得される。

30

40

【0080】

いくつかの実施形態では、FlexX（参照により本明細書に組み込まれる、Rarey et al., 1996, "A Fast Flexible Docking Method Using an Incremental Construction Algorithm," *Journal of Molecular Biology* 261, pp. 470-489）などのアルゴリズムを使用して、標的ポリマーに対する訓練

50

化合物ごとの複数のポーズを見出す。Flex Xは、欲張りアルゴリズムを使用して、標的ポリマー38の活性部位で訓練化合物の逐次構築を行う。したがって、いくつかの実施形態では、（所与の標的化合物についての）複数の異なるポーズは、欲張りアルゴリズムによって取得される。

【0081】

いくつかの実施形態では、GOLD（参照により本明細書に組み込まれる、Jones et al., 1997, "Development and Validation of a Genetic Algorithm for flexible Docking," Journal Molecular Biology 267, pp. 727-748）などのアルゴリズムを使用して、標的ポリマー38に対する訓練化合物の各々についての複数のポーズを見出す。GOLDは、Genetic Optimization for Ligand Docking（配位子ドッキングの遺伝的最適化）の略である。GOLDは、訓練化合物と標的ポリマー38との間に遺伝的に最適化された水素結合ネットワークを構築する。

10

【0082】

いくつかの実施形態では、分子力学は、標的ポリマー（又は、標的ポリマーの活性部位などの、その一部分）及び各それぞれの訓練化合物に対して実行され、各それぞれの訓練化合物について正のポーズ48及び負のポーズ60を識別する。分子力学の実行中に、標的ポリマー及び訓練化合物の原子が一定期間相互作用することを可能にし、システムの力学的進化のビューを示す。標的ポリマー及び訓練化合物内の原子の軌道は、相互作用粒子系のニュートンの運動方程式を数値的に解くことによって判定され、粒子とそれぞれのポテンシャルエネルギーとの間の力は、原子間ポテンシャル又は分子力学力場を使用して計算される。Alder and Wainwright, 1959, "Studies in Molecular Dynamics. I. General Method," J. Chem. Phys. 31 (2): 459、及びBibcode, 1959, J. Ch. Ph. 31, 459A, doi:10.1063/1.1730376を参照されたく、これらは各々、参照によりその全体が本明細書に組み込まれる。したがって、このようにして、分子力学の実行により、経時的に標的ポリマー及びそれぞれの訓練化合物の軌道を生成する。この軌道は、標的ポリマー及び訓練化合物内の原子の軌道を含む。いくつかの実施形態では、複数の異なるポーズのサブセットは、一定の期間にわたってこの軌道のスナップ写真を撮ることによって取得される。いくつかの実施形態では、ポーズは、いくつかの異なる軌道のスナップ写真から取得され、各軌道は、訓練化合物と相互作用する標的ポリマーの異なる分子力学の実行を含む。いくつかの実施形態では、分子力学の実行の前に、訓練化合物はまず、ドッキング技術を使用して、標的ポリマーの活性部位にドッキングされる。

20

30

【0083】

いくつかの実施形態では、標的ポリマーに対するそれぞれの訓練化合物の複数のポーズの中からの任意の対（ポーズ対の一方のポーズが、対の他方のポーズよりも良いドッキングスコアを有する）は、それぞれの訓練化合物の正のポーズ48及び負のポーズ60としてそれぞれ機能することができる。

40

【0084】

ブロック216。ポーズを見出し、そのようなポーズのインシリコポーズ品質スコアを判定するためのいくつかの異なる非限定的な方法及びプログラムが、図2Bのブロック214と併せて上記に開示されている。いくつかの実施形態では、正のポーズ48の第1の正の相互作用スコア50は、これらの非限定的な方法及びプログラムのいずれか、又はそれらの任意の組み合わせ、又は任意の同等若しくは類似のプログラムによって、標的ポリマー38に関する正のポーズ48について算出された、インシリコポーズ品質スコアである。いくつかの実施形態では、正のポーズ48は、ブロック214において上記で考察されるように、ポーズの集合体であり、正のポーズ48の第1の正の相互作用スコア50は、これらの非限定的な方法及びプログラムのいずれかによって、標的ポリマー38に関す

50

る正のポーズ48について算出された、インシリコポーズ品質スコアである。それに応じて、いくつかの実施形態では、負のポーズ60の第1の負の相互作用スコア62は、これらの非限定的な方法及びプログラムのいずれかによって、標的ポリマー38に関する負のポーズ60について算出された、インシリコポーズ品質スコアである。いくつかの実施形態では、負のポーズ60は、ブロック214において上記で考察されるように、ポーズの集合体であり、負のポーズ60の第1の負の相互作用スコア62は、これらの非限定的な方法及びプログラムのいずれかによって、標的ポリマー38に関する負のポーズ60について算出された、インシリコポーズ品質スコアである。

【0085】

いくつかの実施形態では、訓練化合物のインシリコポーズ品質スコアを使用するのではなく、第1の正の相互作用スコアは、実験手段によって判定された、標的ポリマー38に対する対応する訓練化合物46の測定された結合係数、 IC_{50} 、 EC_{50} 、 K_d 、 K_I 、又は pK_I である。 IC_{50} 、 EC_{50} 、 K_d 、 K_I 、及び pK_I などの測定された結合係数は、一般に、Huser ed., 2006, High-Throughput Screening in Drug Discovery, Methods and Principles in Medicinal Chemistry 35及びChen ed., 2019, A Practical Guide to Assay Development and High-Throughput Screening in Drug Discoveryに記載されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

【0086】

ブロック218。図2Bのブロック218を参照すると、いくつかの実施形態では、訓練データセット内の各訓練化合物は、リピンスキーのルールオブファイブ：(i) 5個以下の水素結合ドナー、(ii) 10個以下の水素結合アクセプター、(iii) 500ダルトン未満の分子量、及び(iv) 5未満の $\log P$ 、のうちの2つ以上のルール、3つ以上のルール、又は4つのルール全てを満たす。Lipinski, 1997, Adv. Drug Del. Rev. 23, 3を参照されたく、これは参照によりその全体が本明細書に組み込まれる。

【0087】

いくつかの実施形態では、訓練化合物は、リピンスキーのルールオブファイブに加えて、1つの基準、又は2つ以上の基準を満たす。例えば、いくつかの実施形態では、訓練化合物は、5個以下の芳香族環、4個以下の芳香族環、3個以下の芳香族環、又は2個以下の芳香族環を有する。いくつかの実施形態では、訓練化合物は、2000ダルトン未満、4000ダルトン未満、6000ダルトン未満、8000ダルトン未満、10000ダルトン未満、又は20000ダルトン未満の分子量を有する任意の有機化合物である。

【0088】

しかしながら、開示されたシステム及び方法のいくつかの実施形態は、訓練化合物のサイズの限定を有しない。例えば、いくつかの実施形態では、そのような訓練化合物は、抗体などの大きいポリマーである。

【0089】

ブロック220を参照すると、いくつかの実施形態では、訓練データセット内の各訓練化合物は、500ダルトン未満、1000ダルトン未満、2000ダルトン未満、4000ダルトン未満、6000ダルトン未満、8000ダルトン未満、10000ダルトン未満、又は20000ダルトン未満の分子量を有する有機化合物である。

【0090】

ブロック224～ブロック226。ブロック224を参照すると、方法では、少なくとも第1のモデル72が訓練される。訓練は、複数の訓練化合物の少なくとも第1のサブセットにおける各対応する訓練化合物46に対して、少なくとも(i) 標的ポリマーに関する対応する訓練化合物の対応する第1の正の相互作用スコア50に対する、第1のモデル72への入力としての標的ポリマー38に関する対応する訓練化合物46の対応する正の

ポーズ 48 の対応する正のスコアと、(i i) 標的ポリマーに関する対応する訓練化合物の対応する第 1 の負の相互作用スコア 62 に対する、第 1 のモデル 72 への入力としての標的ポリマーに関する対応する訓練化合物の対応する負のポーズ 60 の対応する負のスコアと、を使用し、それによって、第 1 の複数のパラメータ 73 を調整し、少なくとも第 1 のモデルの出力は、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用される。いくつかのそのような実施形態では、訓練は、複数の訓練化合物の第 2 のサブセットにおける各対応する訓練化合物 46 に対して、少なくとも、標的ポリマーに関する対応する訓練化合物の対応する第 1 の正の相互作用スコア 50 に対する、第 1 のモデル 72 への入力としての標的ポリマー 38 に関する対応する訓練化合物 46 の対応する正のポーズ 48 の対応する正のスコアを更に使用する。いくつかの実施形態では、訓練化合物の全ては、正のポーズ及び負のポーズの両方を有する。いくつかの実施形態では、複数の訓練化合物中の訓練化合物の一部のみが、正のポーズ及び負のポーズの両方を有するが、複数の訓練化合物中の他の訓練化合物は、正のポーズを有しているが、負のポーズを有していない。いくつかの実施形態では、複数の訓練化合物中の訓練化合物の一部のみが、正のポーズ及び負のポーズの両方を有するが、複数の訓練化合物中の他の訓練化合物は、(i) 1 つ以上の正のポーズを有するが負のポーズを有していないか、又は (i i) 1 つ以上の負のポーズを有するが正のポーズを有していないかのいずれかである。

10

【0091】

ブロック 234 を参照すると、いくつかの実施形態では、第 1 のモデル 72 は、第 1 の完全接続ニューラルネットワークである。

20

【0092】

図 9A では、第 1 のモデル 72 は、化合物のポーズ品質の推定値を提供する。モデル 72 を訓練するために、複数の訓練化合物中の各対応する訓練化合物 46 に対して、訓練セット 44 内のデータが使用される。各訓練化合物に対して、対応する訓練化合物 46 の対応する正のポーズ 48 の対応する正のスコアが、第 1 のモデル 72 への入力として、標的ポリマー 38 に関して取得される。

【0093】

図 9A の実施形態によれば、対応する正のポーズ 48 の対応する正のスコアは、以下のブロック 228 でより詳細に考察されるように、正のポーズ 48 をニューラルネットワーク 24 に入力したときの、ニューラルネットワーク 24 の出力である。図 9A に示されるように、典型的な実施形態では、正のスコアは、埋設層 96 からの埋設の形態であり、これは、少なくとも、第 1 のモデルへの入力として機能するのに必要な寸法に合わせて正のスコアを特定の大きさにする目的を果たす。第 1 のモデル 72 の出力は、対応する正のスコアをニューラルネットワーク 24 から入力すると、標的ポリマー 38 に関する対応する訓練化合物の対応する第 1 の正の相互作用スコア 50 に対して比較される。第 1 のモデル 72 の出力と対応する第 1 の正の相互作用スコア 50 との差は、72 逆伝播技術を通して第 1 のモデルの重みを調整するために、損失関数によって評価される。

30

【0094】

更に、図 9A の実施形態によれば、対応する負のポーズ 60 の対応する負のスコアは、負のポーズを有する訓練セット内のそれらの化合物について、以下のブロック 232 でより詳細に考察されるように、負のポーズ 60 をニューラルネットワーク 24 に入力したときの、ニューラルネットワーク 24 の出力である。図 9A に示されるように、典型的な実施形態では、負のスコアは、埋設層 96 からの埋設の形態であり、これは、少なくとも、第 1 のモデルへの入力として機能するのに必要な寸法に合わせて負のスコアを特定の大きさにする目的を果たす。第 1 のモデル 72 の出力は、対応する負のスコアをニューラルネットワーク 24 から入力すると、標的ポリマー 38 に関する対応する訓練化合物の対応する第 1 の負の相互作用スコア 62 に対して比較される。第 1 のモデル 72 の出力と対応する第 1 の負の相互作用スコア 62 との差はまた、逆伝播技術を通して第 1 のモデルの重みを調整するために、損失関数によって評価される。

40

50

【0095】

第1のモデル72は、第1の複数のパラメータ73を有する。いくつかの実施形態では、第1の複数のパラメータは、100個、200個、300個、400個、500個、600個、700個、800個、900個、1000個、10,000個、50,000個、100,000個又は 1×10^6 個よりも多いパラメータを含む。

【0096】

ブロック226を参照すると、いくつかの実施形態では、第1のモデル72は、多層パーセプトロン（MLP）としても知られる、完全接続ニューラルネットワークである。いくつかの実施形態では、MLPは、少なくとも3つの層：入力層、隠れ層及び出力層のノードを含む、フィードフォワード人工ニューラルネットワーク（ANN）の一種である。そのような実施形態では、入力ノードを除いて、各ノードは、非線形活性化機能を使用するニューロンである。いくつかの実施形態では、第1のモデル72として機能する好適なMLPに関する更なる開示は、参照により本明細書に組み込まれる、Vang-mata et al., 2020, Multilayer Perceptrons: Theory and Applications, Nova Science Publishers, Hauppauge, New Yorkにおいて見られる。

10

【0097】

ブロック228～ブロック230。図2Bのブロック228を参照すると、いくつかの実施形態では、標的ポリマー38に関する対応する訓練化合物46の対応する正のポーズ48の対応する正のスコアは、対応する正のポーズ48における標的ポリマー38に関する対応する訓練化合物46の対応する正のボクセルマップ52を検索し、対応する正のボクセルマップ52を、対応する正のベクトル54に展開し、及び対応する正のベクトル54を、ニューラルネットワーク（例えば、重畳型ニューラルネットワーク、グラフニューラルネットワークなど）の形態で、ニューラルネットワーク24に入力することによって取得される。グラフニューラルネットワーク又は重畳型ニューラルネットワーク24は、次に、出力時に、対応する正のポーズ48の対応する正のスコアを提供する。

20

【0098】

いくつかの実施形態では、ニューラルネットワーク24は、ボクセルマップを使用するかどうかに関わらず、500個よりも多いパラメータ、1000個よりも多いパラメータ、2000個よりも多いパラメータ、5000個よりも多いパラメータ、10,000個よりも多いパラメータ、100,000個よりも多いパラメータ、又は 1×10^6 個よりも多いパラメータを含む。

30

【0099】

いくつかのそのような実施形態では、ブロック230を参照すると、上で参照される対応する正のベクトル54は、第1の一次元ベクトルである。いくつかの実施形態では、対応する正のベクトル54は、10個以上の要素、20個以上の要素、100個以上の要素、500個以上の要素、1000個以上の要素、又は10,000個以上の要素を含む。

【0100】

いくつかの実施形態では、ニューラルネットワーク24は、Wallach et al., 2015, "AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery," arXiv:1510.02855v1、又は米国特許第11,080,570号、同第10,546,237号、同第10,482,355号、同第10,002,312号、若しくは同第9,373,059号に開示される重畳型ニューラルネットワーク24のいずれかであり、これらは各々、参照によりその全体が本明細書に組み込まれる。重畳型ニューラルネットワークを使用して、標的ポリマー38に関する対応する訓練化合物46の対応する正のポーズ48の対応する正のスコアを取得することに関する更なる詳細は、「重畳型ニューラルネットワークを使用して、ポーズのスコアを取得する」と題された節で以下に開示されている。

40

50

【0101】

いくつかの実施形態では、ニューラルネットワーク24は、等変量ニューラルネットワークである。等変量重畳型ニューラルネットワークの非限定的な例は、Thomas et al., 2018, "Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds," arXiv:1802.08219、Anderson et al., 2019, "Cormorant: Covariant Molecular Neural Networks," Neural Information Processing Systems、Johannes et al., 2020, "Directional Message Passing For Molecular Graphs," International Conference on Learning Representations、Townshend et al., 2021, "ATOM3D: Tasks On Molecules in Three Dimensions," International Conference on Learning Representations、Jing et al., 2009, "Learning from Protein Structure with Geometric Vector Perceptrons," arXiv:2009.01411、及びSatorras et al., 2021, "E(n) Equivariant Graph Neural Networks," arXiv:2102.09844に開示されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

10

20

【0102】

いくつかの実施形態では、ニューラルネットワーク24は、グラフニューラルネットワーク（例えば、グラフ重畳型ニューラルネットワーク）である。グラフ重畳型ニューラルネットワークの非限定的な例は、Behler Parrinello, 2007, "Generalized Neural-Network Representation of High Dimensional Potential-Energy Surfaces," Physical Review Letters 98, 146401、Chmiela et al., 2017, "Machine learning of accurate energy-conserving molecular force fields," Science Advances 3 (5): e1603015、Schuett et al., 2017, "SchNet: A continuous-filter convolutional neural network for modeling quantum interactions," Advances in Neural Information Processing Systems 30, pp. 992-1002、Feinberg et al., 2018, "PotentialNet for Molecular Property Prediction," ACS Cent. Sci. 4, 11, 1520-1530、及びStafford et al., "AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High Throughput Screens," <https://chemrxiv.org/engage/chemrxiv/article-details/614b905e39ef6a1c36268003>に開示されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

30

40

【0103】

いくつかの実施形態では、ニューラルネットワーク24は、参照により本明細書に組み込まれる、2022年5月10日に出願の"Characterization of Interactions Between Compounds and Polymers Using Pose Ensembles"と題された米国仮特許出願第63/336,841号に開示されたグラフニューラルネットワークのいずれかである。

50

【0104】

ブロック232～ブロック234。図2Dのブロック232を参照すると、いくつかの実施形態では、標的ポリマー38に関する対応する訓練化合物46の対応する負のポーズ60の対応する負のスコアは、標的ポリマー38に関する対応する訓練化合物46の対応する正のポーズ48の対応する正のスコアが得られたのと同じ様式で得られる。例えば、いくつかの実施形態では、それは、対応する負のポーズ60における標的ポリマーに関する対応する訓練化合物の対応する負のボクセルマップを検索し、対応する負のボクセルマップを、対応する負のベクトル66に展開し、及び対応する負のベクトルを、ニューラルネットワーク24に入力することによって取得され、それによって、標的ポリマー38に関する対応する訓練化合物46の対応する負のポーズ60の対応する負のスコアを取得する。いくつかのそのような実施形態では、図2Dのブロック234を参照すると、対応する負のベクトル66は、第2の一次元ベクトルである。

10

【0105】

ブロック236～ブロック244。図2Dのブロック236を参照すると、いくつかの実施形態では、モデル72の訓練は、第1のモデル72の第1の複数のパラメータ73が、関連付けられた損失関数を通した逆伝播によって調整されている、回帰タスクであり、対応する第1の正の相互作用スコア50は、式

$$B = N \times A$$

によって対応する第1の負の相互作用スコア62に関係付けられており、式中、Aは、対応する正の相互作用スコアであり、Bは、対応する負の相互作用スコアであり、Nは、ゼロ超かつ1未満の実数である。回帰タスクとしてのモデル72の訓練は、第1の正の相互作用スコアが、ウェットラボ（例えば、インビボ又はインビトロの）アッセイからのそれぞれの訓練化合物の測定された特性である場合に好適である。それぞれの訓練化合物のそのような測定された特性の例としては、標的ポリマーに関するそれぞれの訓練化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iが挙げられるが、これらに限定されない。そのような実施形態では、第1の正の相互作用スコア50に、訓練化合物の測定された特性を割り当てるのが妥当である。次いで、訓練目的のために、訓練化合物の測定された特性を考えると、訓練化合物の第1の負の相互作用スコア62を何に割り当てるかという質問になる。図2Dのブロック236によれば、いくつかの実施形態では、負の相互作用スコア62には、測定された特性の固定割引価値Nが割り当てられる。固定とは、対応する第1の負の相互作用スコア62の値を計算するために、各それぞれの訓練化合物について各第1の正の相互作用スコア50に、同じ値Nが適用されることを意味する。したがって、Nの値が0.90である場合、各それぞれの訓練化合物について、対応する第1の負の相互作用スコア62は、対応する第1の正の相互作用スコア50の0.90である値を有する。いくつかの実施形態では、Nは、0.10～0.99の値である。いくつかの実施形態では、Nは、0.20～0.95の値である。いくつかの実施形態では、Nは、0.30～0.90の値である。いくつかの実施形態では、Nは、0.25～0.85の値である。いくつかの実施形態では、Nは、0.60～0.95の値である。いくつかの代替の実施形態では、負の相互作用スコア62には、測定された特性の対数が割り当てられる。したがって、そのような実施形態では、各それぞれの訓練化合物について、対応する第1の負の相互作用スコアは、対応する第1の正の相互作用スコア50の対数である。対数は、自然対数、底数10などの任意の底数であり得る。

20

30

40

【0106】

ブロック238を参照すると、いくつかの実施形態では、ブロック232に関して上で説明される関連付けられた損失関数は、任意の好適な回帰タスク損失関数である。そのような損失関数の例としては、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数が挙げられるが、これらに限定されない。Wang et al., 2020, "A Comprehensive Survey of Loss Functions in Machine Learning," Annals of Data Science, <https://doi.org/10.1007/s401060190086>

50

rg/10.1007/s40745-020-00253-5, last accessed September 15, 2021を参照されたく、それらの各々は参照により本明細書に組み込まれる。

【0107】

図2Dのブロック240を参照すると、いくつかの特定の実施形態では、対応する第1の正の相互作用スコア50及び対応する第1の負の相互作用スコア62は、各々、結合係数を表し、対応する第1の正の相互作用スコアは、標的ポリマー38に対する対応する訓練化合物46の結合係数のインビボ又はインビトロ測定値である。

【0108】

図2Eのブロック244を参照すると、いくつかの実施形態では、第1の正の相互作用スコアは、標的ポリマーに関するそれぞれの訓練化合物のIC₅₀、EC₅₀、K_d、K_I、又はpK_Iである。測定された結合係数は、一般に、Huser ed., 2006, High-Throughput-Screening in Drug Discovery, Methods and Principles in Medicinal Chemistry 35及びChen ed., 2019, A Practical Guide to Assay Development and High-Throughput Screening in Drug Discoveryに記載されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

10

【0109】

ブロック246～ブロック248。図2Eのブロック246を参照すると、いくつかの実施形態では、訓練データセット44内の電子記述46の少なくとも1つのサブセットにおける各それぞれの電子記述46は、対応する訓練化合物46の対応する正のポーズ48の対応する正のスコア56と、対応する訓練化合物の対応する負のポーズ60の対応する負の活性スコア58と、を更に含む。いくつかの実施形態では、訓練化合物のうちの少なくともいくつかは、負の活性スコア58を有しない。図9Bを参照すると、いくつかの実施形態では、少なくとも第1のモデル72を訓練することは、第2のモデル74を第1のモデルと共同で訓練することを更に含む。

20

【0110】

第1のモデル72と同様に、第2のモデル74は、複数のパラメータ75（第2の複数のパラメータ）を有する。いくつかの実施形態では、第2の複数のパラメータは、100個、200個、300個、400個、500個、600個、700個、800個、900個、1000個、10,000個、50,000個、100,000個又は 1×10^6 個よりも多いパラメータを含む。

30

【0111】

図9Bの実施形態では、第2のモデル74は、化合物のポーズ品質の推定値を提供する。第2のモデル74を訓練するために、複数の訓練化合物中の各対応する訓練化合物46に対して、訓練セット44内のデータが使用される。各訓練化合物に対して、対応する訓練化合物46の対応する正のポーズ48の対応する正のスコアが、第2のモデル74への入力として、標的ポリマー38に関して取得される。

【0112】

図9Bの実施形態によれば、対応する正のポーズ48の対応する正のスコアは、正のポーズ48をニューラルネットワーク24に入力したときの、ニューラルネットワーク24の出力である。図9Bに示されるように、典型的な実施形態では、正のスコアは、埋設層96からの埋設の形態であり、これは、少なくとも、第2のモデルへの入力として機能するのに必要な寸法に合わせて正のスコアを特定の大きさにする目的を果たす。第2のモデル74の出力は、エッジ920によって示されるように、対応する正のスコアをニューラルネットワーク24から第2のモデル74に入力すると、標的ポリマー38に関する対応する訓練化合物の対応する第1の正の相互作用スコア50に対して比較される。第2のモデル74の出力と対応する第1の正の相互作用スコア50との差は、74逆伝播技術を通して第2のモデルの重みを調整するために、損失関数によって評価される。

40

50

【0113】

更に、図9Bの実施形態によれば、対応する負のポーズ60の対応する負のスコアは、負のポーズを有する訓練セット内のそれらの化合物について、負のポーズ60をニューラルネットワーク24に入力したときの、ニューラルネットワーク24の出力である。図9Bに示されるように、典型的な実施形態では、負のスコアは、埋設層96からの埋設の形態であり、これは、少なくとも、第2のモデルへの入力として機能するのに必要な寸法に合わせて負のスコアを特定の大きさにする目的を果たす。第2のモデル74の出力は、エッジ920によって示されるように、対応する負のスコアをニューラルネットワーク24から第2のモデル74に入力すると、標的ポリマー38に関する対応する訓練化合物の対応する第1の負の相互作用スコア62に対して比較される。第2のモデル74の出力と対応する第1の負の相互作用スコア62との差はまた、逆伝播技術を通して第2のモデルの複数のパラメータ75を調整するために、損失関数によって評価される。

10

【0114】

更に、図9Bに示される実施形態では、複数の訓練化合物中の各対応する訓練化合物に対して、ブロック224の訓練は、少なくとも、訓練化合物の少なくともサブセットについて、(iii)対応する訓練化合物の対応する正の活性スコア56に対する、(図9Bのエッジ930によって示した)第1のモデル72への入力としての標的ポリマー38に関する対応する訓練化合物46の対応する正のポーズ48の対応する正のスコアと、(iv)対応する訓練化合物46の対応する負の活性スコア68に対する、(繰り返しになるが、図9Bのエッジ930によって示した)第1のモデル72への入力としての標的ポリマー38に関する対応する訓練化合物46の対応する負のポーズ60の対応する負のスコアと、を更に使用する。このようにして、第1のモデルの第1の複数のパラメータ73は、訓練中に調整される。

20

【0115】

したがって、図9Bの実施形態では、第2のモデル74は、それぞれの第1の正の相互作用スコア50及び第1の負の相互作用スコア62に対して訓練され、一方、第1のモデル72は、正の活性スコア56及び負の活性スコア68に対して訓練される。いくつかのそのような実施形態では、第1の正の相互作用スコア50及び第1の負の相互作用スコア62は、ドッキングスコアであり、正の活性スコア及び負の活性スコアは、バイナリ離散活性値である。例えば、バイナリ離散活性値の2つの可能な値のうちの一方は、対応する訓練が標的ポリマーの活性を阻害することを示し、一方、バイナリ離散活性値の2つの可能な値のうちの他方は、対応する訓練が標的ポリマーのその活性を阻害しないことを示すであろう。

30

【0116】

図9Bに示されるように、訓練されると、試験化合物のポーズがニューラルネットワーク24に伝わり、標的ポリマーに対する試験化合物のポーズのスコアを生じる。標的ポリマーに関する試験化合物のポーズのこのスコアは、(試験化合物と標的ポリマーとの間の相互作用の特徴付けを、ポーズ品質スコアの形態で提供するための)第2のモデル74と、(試験化合物と標的ポリマーとの間の相互作用の特徴付けを、試験化合物と標的ポリマー38との間の相互作用の活性の形態で提供するための)第1のモデル72と、の両方に入力される。したがって、図9Bの実施形態では、試験化合物と標的ポリマーとの間の相互作用の特徴付けは、活性スコア(例えば、離散バイナリスコア又はスカラースコア)及びポーズ品質スコアの両方である。

40

【0117】

ブロック248を参照すると、いくつかのそのような実施形態では、第1のモデル72及び第2のモデル74は、各々、多層パーセプトロン(MLP)としても知られる、完全接続ニューラルネットワークである。いくつかの実施形態では、MLPは、少なくとも3つの層：入力層、隠れ層及び出力層のノードを含む、フィードフォワード人工ニューラルネットワーク(ANN)の一種である。そのような実施形態では、入力ノードを除いて、各ノードは、非線形活性化機能を使用するニューロンである。いくつかの実施形態では、

50

第1のモデル72として機能する好適なMLPに関する更なる開示は、参照により本明細書に組み込まれる、Vang-mata ed., 2020, Multilayer Perceptrons: Theory and Applications, Nova Science Publishers, Hauppauge, New Yorkにおいて見られる。

【0118】

ブロック252～ブロック256。図2Fのブロック252を参照すると、図9Cに示されるように、いくつかの実施形態では、訓練データセット44の少なくともサブセットにおける各それぞれの電子記述46は、対応する訓練化合物46の対応する正のポーズ48の対応する正の活性スコア56と、対応する訓練化合物の対応する負のポーズ60の対応する負の活性スコア58と、を更に含む。そのような実施形態では、ブロック224で上述した訓練（少なくとも第1のモデル72の訓練）は、第2のモデル74を第1のモデル72と共同で訓練することを更に含む。第2のモデル74は、第2の複数のパラメータ75を有する。

10

【0119】

図9Cの実施形態では、第2のモデル74は、化合物のポーズ品質の推定値を提供する。第2のモデル74を訓練するために、複数の訓練化合物中の各対応する訓練化合物46に対して、訓練セット44内のデータが使用される。各訓練化合物に対して、対応する訓練化合物46の対応する正のポーズ48の対応する正のスコアが、第2のモデル74への入力として、標的ポリマー38に関して取得される。

20

【0120】

図9Cの実施形態によれば、対応する正のポーズ48の対応する正のスコアは、正のポーズ48をニューラルネットワーク24に入力したときの、ニューラルネットワーク24の出力である。図9Cに示されるように、典型的な実施形態では、正のスコアは、埋設層96からの埋設の形態であり、これは、少なくとも、第1のモデル及び第2のモデルへの入力として機能するのに必要な寸法に合わせて正のスコアを特定の大きさにする目的を果たす。第2のモデル74の出力は、エッジ940によって示されるように、対応する正のスコアをニューラルネットワーク24から第2のモデル74に入力すると、標的ポリマー38に関する対応する訓練化合物の対応する第1の正の相互作用スコア50に対して比較される。第2のモデル74の出力と対応する第1の正の相互作用スコア50との差は、74逆伝播技術を通して第2のモデルの重みを調整するために、損失関数によって評価される。

30

【0121】

更に、図9Cの実施形態によれば、対応する負のポーズ60の対応する負のスコアは、負のポーズを有する訓練セット内のそれらの化合物について、負のポーズ60をニューラルネットワーク24に入力したときの、ニューラルネットワーク24の出力である。図9Cに示されるように、典型的な実施形態では、負のスコアは、埋設層96からの埋設の形態であり、これは、少なくとも、第1のモデル及び第2のモデルの両方への入力として機能するのに必要な寸法に合わせて負のスコアを特定の大きさにする目的を果たす。第2のモデル74の出力は、エッジ940によって示されるように、対応する負のスコアをニューラルネットワーク24から第2のモデル74に入力すると、標的ポリマー38に関する対応する訓練化合物の対応する第1の負の相互作用スコア62に対して比較される。第2のモデル74の出力と対応する第1の負の相互作用スコア62との差はまた、逆伝播技術を通して第2のモデルの複数のパラメータ75を調整するために、損失関数によって評価される。

40

【0122】

図9Cによる訓練は、複数の訓練化合物の少なくともサブセットにおける各対応する訓練化合物46に対して、少なくとも対応する訓練化合物の対応する正の活性スコア56に対する、第1のモデル72への結合入力としてのモデル24（エッジ950を通して）及び第2のモデル74（エッジ930を通して）の両方によって提供された標的ポリマー3

50

8に関する対応する訓練化合物の対応する正のポーズ48の対応する正のスコアと、対応する訓練化合物の対応する負の活性スコア68に対する、モデル24（繰り返しになるが、エッジ950を通して）及び第2のモデル74（繰り返しになるが、エッジ930を通して）の両方によって提供された標的ポリマー38に関する対応する訓練化合物46の対応する負のポーズ60の対応する負のスコアと、を更に使用する。このようにして、第1のモデル72の第1の複数のパラメータ73が（例えば、損失関数を使用する逆伝播法を通して）調整される。

【0123】

第2のモデル74は、少なくとも部分的に、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために、第1のモデル72の出力とともに使用される。例えば、図9Cに示されるように、訓練されると、試験化合物のポーズがニューラルネットワーク24に伝わり、標的ポリマー38に対する試験化合物のポーズのスコアを生じる。標的ポリマーに関するこのスコアは、第1のモデル72（エッジ950を通して）及び第2のモデル74（エッジ940を通して）の両方に入力される。更に、試験化合物の第2のモデル74の出力（ポーズ品質スコア、pK_Aなどの、相互作用スコアの計算である）は、エッジ930を通して第1のモデル72に入力される。したがって、第1のモデル72は、試験化合物のポーズのモデル24への入力に応答して、第2のモデルの出力及びモデル24の出力の両方を受け取る。第1のモデル72は、これらの入力の両方を使用して、試験化合物と標的ポリマーとの間の相互作用の特徴付けを判定する。いくつかの実施形態では、この特徴付けは、試験化合物の活性スコアである。いくつかの実施形態では、この活性スコアは、例えば、「1」が、試験化合物が標的ポリマーに対して活性であることを示し、「0」が、試験化合物が標的ポリマーに対して不活性であることを示す、離散バイナリスコアである。いくつかの実施形態では、第1のモデル72によって提供された活性スコアは、スカラーである。モデル24及び第2のモデル74の出力の両方に対する第1のモデル72の離散バイナリ活性スコアのコンディショニングは、試験化合物を特徴付ける際の第1のモデルの性能を改善するのに役立つ。

【0124】

図2Fのブロック254を参照すると、いくつかのそのような実施形態では、対応する正の活性スコア56は、第1のバイナリ活性スコアであり、対応する負の活性スコア68は、第2のバイナリ活性スコアである。いくつかの実施形態では、活性基準を満たすことに基づいて、対応する第1のバイナリ活性スコアは、標的ポリマーに対する対応する化合物の測定された活性に基づいて、1の値を割り当てられており、対応する第2のバイナリ活性スコアは、活性基準を満たさないことに基づいて0の値を割り当てられている。いくつかの実施形態では、訓練化合物のこれらの活性値は、インビボ又はインビトロのアッセイによって取得される。そのようなアッセイは、一般に、Huser ed., 2006, High-Throughput Screening in Drug Discovery, Methods and Principles in Medicinal Chemistry 35及びChen ed., 2019, A Practical Guide to Assay Development and High-Throughput Screening in Drug Discoveryに記載されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

【0125】

図2Fのブロック256を参照すると、いくつかの実施形態では、第2のモデル74の訓練は、第2の複数のパラメータ75が、第2の関連付けられた損失関数を通じた逆伝播によって調整されている、回帰タスクである。回帰タスクに好適な損失関数の非限定的な例としては、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数が挙げられるが、これらに限定されない。Wang et al., 2020, "A Comprehensive Survey of Loss Functions in Machine Learning," Annals of Data Science, <https://doi.org/10.100>

7 / s 4 0 7 4 5 - 0 2 0 - 0 0 2 5 3 - 5 , l a s t a c c e s s e d S e p t e m b e r 1 5 , 2 0 2 1 を参照されたく、これは参照によりその全体が本明細書に組み込まれる。更に、いくつかの実施形態では、第 1 のモデル 7 2 の訓練は、第 1 の複数のパラメータ 7 3 が、第 1 の関連付けられた損失関数を通した逆伝播によって調整されている、分類タスクである。分類タスクに好適な損失関数の非限定的な例としては、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数が挙げられるが、これらに限定されない。

【0126】

いくつかの実施形態では、第 1 のモデルの出力は、バイナリ以外の離散値である。例えば、(図 9 C に示される構成の分類子 2 4 にポーズを入力することに応答した) 第 2 のモデルの第 1 の出力値は、標的ポリマーに対する試験化合物の不十分な活性を示し、第 2 の出力値は、標的ポリマーに対する試験化合物の中間活性を示し、第 3 の出力値は、標的ポリマーに対する試験化合物の良好な活性を示す。いくつかのそのような実施形態では、第 1 の分類子を訓練するために使用される損失関数は、マルチクラスクロスエントロピー損失関数、スパースマルチクラスクロスエントロピー損失関数、又はカルバックライブラーダイバージェンス損失関数などのマルチクラス分類損失関数であり得る。

10

【0127】

ブロック 2 6 0。図 2 G のブロック 2 6 0 を参照すると、いくつかの実施形態では、対応する第 1 の正の相互作用スコア 5 0 及び対応する第 1 の負の相互作用スコア 6 2 は、各々、標的ポリマーに対する対応する訓練化合物の結合係数又はインシリコポーズ品質スコアを表し、対応する正の活性スコア 5 6 は、第 1 のバイナリ活性スコアであり、対応する負の活性スコア 6 8 は、第 2 のバイナリ活性スコアである。

20

【0128】

ブロック 2 6 2。図 2 G のブロック 2 6 2 を参照すると、いくつかの実施形態では、第 1 の関連付けられた損失関数は、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber 損失関数、Log-Cosh 損失関数、又は分位損失関数であり、第 2 の関連付けられた損失関数は、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数である。

【0129】

ブロック 2 6 4。図 2 G のブロック 2 6 4 を参照すると、いくつかの実施形態では、第 2 のモデル 7 4 は、多層パーセプトロン (MLP) としても知られる、第 2 の完全接続ニューラルネットワークである。いくつかの実施形態では、MLP は、少なくとも 3 つの層：入力層、隠れ層及び出力層のノードを含む、フィードフォワード人工ニューラルネットワーク (ANN) の一種である。そのような実施形態では、入力ノードを除いて、各ノードは、非線形活性化機能を使用するニューロンである。いくつかの実施形態では、第 1 のモデル 7 2 として機能する好適な MLP に関する更なる開示は、参照により本明細書に組み込まれる、Vang-mata ed., 2020, Multilayer Perceptrons: Theory and Applications, Nova Science Publishers, Hauppauge, New York において見られる。

30

40

【0130】

ブロック 2 6 8 ~ ブロック 2 7 6。図 2 H のブロック 2 6 8 を参照すると、図 1 6 A に示されるように、いくつかの実施形態では、訓練データセット内の各それぞれの電子記述は、対応する訓練化合物 4 6 の対応する正のポーズ 4 8 の対応する第 2 の正の相互作用スコアと、対応する訓練化合物の対応する負のポーズ 6 0 の対応する第 2 の負の相互作用スコアと、を更に含む。更に、訓練データセット内の各それぞれの電子記述はまた、対応する訓練化合物 4 6 の対応する正のポーズ 4 8 の対応する正の活性スコア 5 6 と、対応する訓練化合物の対応する負のポーズ 6 0 の対応する負の活性スコア 6 8 と、を含む。

【0131】

そのような実施形態では、少なくとも第 1 のモデル 7 2、第 2 のモデル 7 4、及び第 3

50

のモデル 7 6 の訓練は、共同で訓練される。

【0132】

第 2 のモデル 7 4 は、第 2 の複数のパラメータ 7 5 を有する。いくつかの実施形態では、第 2 の複数のパラメータは、100 個、200 個、300 個、400 個、500 個、600 個、700 個、800 個、900 個、1000 個、10,000 個、50,000 個、100,000 個又は 1×10^6 個よりも多いパラメータを含む。

【0133】

第 3 のモデル 7 6 は、第 3 の複数のパラメータ 7 7 を有する。いくつかの実施形態では、第 3 の複数のパラメータは、100 個、200 個、300 個、400 個、500 個、600 個、700 個、800 個、900 個、1000 個、10,000 個、50,000 個、100,000 個又は 1×10^6 個よりも多いパラメータを含む。

10

【0134】

モデル共訓練は、複数の訓練化合物の少なくともサブセットにおける各対応する訓練化合物 4 6 に対して、少なくとも、(i) 標的ポリマー 3 8 に関する対応する訓練化合物の対応する第 1 の正の相互作用スコア 5 0 に対する、第 2 のモデル 7 4 への結合入力としてのモデル 2 4 (エッジ 1 6 1 0 を通して) によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する正のポーズ 4 8 の対応する正のスコアと、(ii) 標的ポリマー 3 8 に関する対応する訓練化合物 4 6 の対応する第 1 の負の相互作用スコア 6 2 に対する、第 2 のモデル 7 4 への入力としてのモデル 2 4 (繰り返しになるが、エッジ 1 6 1 0 を通して) によって提供された標的ポリマー 3 8 に関する対応する訓練化合物 4 6 の対応する負のポーズ 6 0 の対応する負のスコアと、を使用し、それによって、第 2 のモデルの第 2 の複数のパラメータを調整する。

20

【0135】

モデル共訓練は、複数の訓練化合物の少なくともサブセットにおける各対応する訓練化合物 4 6 に対して、少なくとも、標的ポリマー 3 8 に関する対応する訓練化合物の対応する第 2 の正の相互作用スコア 5 8 に対する、第 3 のモデル 7 6 への入力としてのモデル 2 4 (エッジ 1 6 2 0 を通して) によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する正のポーズ 4 8 の対応する正のスコアと、標的ポリマー 3 8 に関する対応する訓練化合物 4 6 の対応する第 2 の負の相互作用スコア 7 0 に対する、第 3 のモデル 7 6 への入力としてのモデル 2 4 (繰り返しになるが、エッジ 1 6 2 0 を通して) によって提供された標的ポリマー 3 8 に関する対応する訓練化合物 4 6 の対応する負のポーズ 6 0 の対応する負のスコアと、を更に使用し、それによって、第 3 のモデル 7 6 の第 3 の複数のパラメータ 7 7 を調整する。

30

【0136】

モデル共訓練は、複数の訓練化合物の少なくともサブセットにおける各対応する訓練化合物 4 6 に対して、少なくとも、(i) モデル 2 4 (エッジ 1 6 3 0 を通して) によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する正のポーズ 4 8 の対応する正のスコアと、(ii) モデル 2 4 によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する正のポーズ 4 8 の対応する正のスコアの第 2 のモデル 7 4 への入力時の、エッジ 1 6 4 0 を通した第 2 のモデル 7 4 の出力と、(iii) 標的ポリマー 3 8 に関する対応する訓練化合物の対応する正の活性スコアに対する、第 1 のモデル 7 2 への一括入力としての、モデル 2 4 によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する正のポーズ 4 8 の対応する正のスコアの第 3 のモデル 7 6 への入力時の、エッジ 1 6 5 0 を通した第 3 のモデル 7 6 の出力と、かつ、少なくとも、(i) モデル 2 4 (エッジ 1 6 3 0 を通して) によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する負のポーズの対応する負のスコアと、(ii) モデル 2 4 によって提供された標的ポリマー 3 8 に関する対応する訓練化合物の対応する負のポーズの対応する負のスコアの第 2 のモデル 7 4 への入力時の、エッジ 1 6 4 0 を通した第 2 のモデル 7 4 の出力と、(iii) 標的ポリマー 3 8 に関する対応する訓練化合物の対応する負の活性スコアに対する、第 1 のモデル 7 2 への一括入力としての、モデル 2 4 によ

40

50

って提供された標的ポリマー 38 に関する対応する訓練化合物の対応する負のポーズの対応する負のスコアの第 3 のモデル 76 への入力時の、エッジ 1650 を通した第 3 のモデル 76 の出力と、を更に使用し、それによって、第 1 のモデルの第 1 の複数のパラメータを調整する。

【0137】

第 1 のモデル 74 は、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供するために使用される。例えば、図 16A に示されるように、訓練されると、試験化合物のポーズがニューラルネットワーク 24 に伝わり、標的ポリマー 38 に対する試験化合物のポーズのスコアを生じる。標的ポリマーに関するこのスコアは、第 1 のモデル 72 (エッジ 1630 を通して)、第 2 のモデル 74 (エッジ 1610 を通して)、及び第 3 のモデル (エッジ 1620 を通して) に入力される。更に、試験化合物の第 2 のモデル 74 の出力 (ポーズ品質スコアなどの、相互作用スコアの計算である) は、エッジ 1640 を通して第 1 のモデル 72 に入力される。更に、試験化合物の第 3 のモデル 76 の出力 (pK_A などの、相互作用スコアの計算である) は、エッジ 1650 を通して第 1 のモデル 72 に入力される。したがって、第 3 のモデルは、試験化合物のポーズのモデル 24 への入力に
10 応答して、第 1 のモデル、第 2 のモデル、及びモデル 24 の出力を受け取る。第 1 のモデル 72 は、これらの入力の各々を使用して、試験化合物と標的ポリマーとの間の相互作用の特徴付けをまとめて判定する。いくつかの実施形態では、この特徴付けは、試験化合物の活性スコアである。いくつかの実施形態では、この活性スコアは、例えば、「1」が、
20 試験化合物が標的ポリマーに対して活性であることを示し、「0」が、試験化合物が標的ポリマーに対して不活性であることを示す、離散バイナリスコアである。いくつかの実施形態では、第 3 のモデル 74 によって提供された活性スコアは、スカラーである。モデル 24、第 2 のモデル 74、及び第 3 のモデル 76 の出力に対する第 1 のモデル 72 の離散バイナリ) 活性スコアのコンディショニングは、この第 1 のモデルに、活性を算出するときに結合モードを強制的に考慮させることによって、試験化合物を特徴付ける際の第 1 のモデルの性能を改善することに役立ち、したがって、機械学習で生じるピカソ問題に対処する。したがって、第 1 のモデルの出力は、試験化合物と標的ポリマーとの間の相互作用の特徴付けを提供する。

【0138】

いくつかの実施形態では、図 16A を参照すると、ニューラルネットワーク 24 によって生成された埋設 96 は、以下の 3 つの出力を予測するために使用される: (第 1 のモデル 72 を通した) 活性、(第 2 のモデル 74 を通した) C U i n a ポーズ品質スコア、及び (第 3 のモデル 76 を通した) p K i スコア。これは、図 16A に示される実施形態では、2 段階で実行される。第 1 に、C U i n a 及び p K i スコア予測は、標的ポリマー 38 に対する試験化合物のポーズのスコアを、ニューラルネットワーク 24 から第 2 のモデル 74 及び第 3 のモデル 76 を通して (埋設 96 として) 伝えることによって、算出される。第 2 に、コンディショニングされた埋設 1690 は、(i) 入力埋設 96 (ニューラルネットワークスコアからの標的ポリマー 38 に対する試験化合物のポーズのスコア) と、
30 (i i) 第 1 の段階から結果として生じる第 2 のモデル 74 のスコア予測と、(i i i) 第 1 の段階からの第 3 のモデル 76 のスコア予測と、を連結させることによって、形成
40 される。次いで、この埋設 1690 は、多層パーセプトロンの形態である第 1 のモデル 72 に伝えられ、試験化合物の活性予測を算出する。いくつかの実施形態では、埋設 1690 は、(i) 入力埋設 96 (ニューラルネットワークスコアからの標的ポリマー 38 に対する試験化合物のポーズのスコア) と、(i i) 第 1 の段階から結果として生じる第 2 のモデル 74 のスコア予測と、(i i i) 第 1 の段階からの第 3 のモデル 76 のスコア予測と、を単に連結させるのではなく、これらの 3 つのソースを互いに乗算し、乗算の積を埋設 1690 として第 3 のモデルに入力する。いくつかの実施形態では、図 16A の埋設 1690 は、(i) 入力埋設 96 (ニューラルネットワークスコアからの標的ポリマー 38 に対する試験化合物のポーズのスコア) と、(i i) 第 1 の段階から結果として生じる第 2 のモデル 74 のスコア予測と、(i i i) 第 1 の段階からの第 3 のモデル 76 のスコア
50

予測と、を単に連結させるのではなく、これらの3つのソースを互いに乗算し、乗算の積を埋設1690として第3のモデルに入力する。いくつかの実施形態では、埋設1690は、連結させるではなく、埋設1690において3つのソースの各々を変換し、この変換は、第1のモデル72への入力として機能する。より一般的には、埋設1690は、第1のモデル72に伝えられるコンディショニング埋設を形成するために、乗算、連結、線形又は非線形変換を含むが、これに限定されない埋設1690への入力のいずれかの全て又は任意の部分に対して任意の数学的関数を実行することができる。

【0139】

図16Bを参照すると、第1のモデル72を追加のモデルに対しても同様にコンディショニングすることは可能である。したがって、図16Bでは、第1のモデル72は、ネットワーク24の出力に加えて、例えば、訓練化合物のC U i n aスコアについて訓練された第2のモデル74、例えば、訓練化合物のp K iスコアについて訓練された第3のモデル76、及び、例えば、訓練化合物のP o s e N e tスコアについて訓練された第4のモデル990の出力に対してコンディショニングされる。

【0140】

図2Iのブロック272を参照すると、いくつかのそのような実施形態では、第1のモデル、第2のモデル74、第3のモデル76、及び第4のモデル990は、各々、完全接続ニューラルネットワークである。そのような完全接続ニューラルネットワークは、多層パーセプトロン(MLP)としても知られている。いくつかの実施形態では、MLPは、少なくとも3つの層：入力層、隠れ層及び出力層のノードを含む、フィードフォワード人工ニューラルネットワーク(ANN)の一種である。そのような実施形態では、入力ノードを除いて、各ノードは、非線形活性化機能を使用するニューロンである。いくつかの実施形態では、第1のモデル72として機能する好適なMLPに関する更なる開示は、参照により本明細書に組み込まれる、V a n g - m a t a e d . , 2020, M u l t i l a y e r P e r c e p t r o n s : T h e o r y a n d A p p l i c a t i o n s , N o v a S c i e n c e P u b l i s h e r s , H a u p p a u g e , N e w Y o r kにおいて見られる。

【0141】

図2Iのブロック274を参照すると、いくつかの実施形態では、第1のモデル72によって提供された対応する正の活性スコアは、第1のバイナリ活性スコアであり、第1のモデル72によって提供された対応する負の活性スコアは、第2のバイナリ活性スコアである。いくつかの実施形態では、対応する第1のバイナリ活性スコアは、標的ポリマーに対する対応する訓練化合物の測定された活性に基づいて、「1」の値を割り当てられており、対応する第2のバイナリ活性スコアは、「0」の値を割り当てられている。

【0142】

図2Iのブロック276を参照すると、いくつかの実施形態では、第2のモデル74の訓練は、第2のモデルに関連付けられた第2の複数のパラメータが、第2の関連付けられた損失関数を通した逆伝播によって調整されている、回帰タスクである。更に、いくつかの実施形態では、第3のモデル76の訓練は、第3のモデルに関連付けられた第3の複数のパラメータが、第3の関連付けられた損失関数を通した逆伝播によって調整されている、回帰タスクである。更に、いくつかの実施形態では、第4のモデル990の訓練は、第4のモデル990に関連付けられた第4の複数のパラメータが、第4の関連付けられた損失関数を通した逆伝播によって調整されている、回帰タスクである。これらの回帰タスクに好適な損失関数の非限定的な例としては、平均二乗誤差損失関数、平均絶対誤差損失関数、H u b e r損失関数、L o g - C o s h損失関数、又は分位損失関数が挙げられるが、これらに限定されない。W a n g e t a l . , 2020, " A C o m p r e h e n s i v e S u r v e y o f L o s s F u n c t i o n s i n M a c h i n e L e a r n i n g , " A n n a l s o f D a t a S c i e n c e , <https://doi.org/10.1007/s40745-020-00253-5>, last accessed September 15, 2021を参照されたく、これは参照に

よりその全体が本明細書に組み込まれる。更に、いくつかの実施形態では、第1のモデル72の訓練は、第1のモデル72に関連付けられた第1の複数のパラメータが、第1の関連付けられた損失関数を通した逆伝播によって調整されている、分類タスクである。分類タスクに好適な損失関数の非限定的な例としては、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数が挙げられるが、これらに限定されない。

【0143】

いくつかのそのような実施形態では、対応する第1の正の相互作用スコア及び対応する第1の負の相互作用スコアは、各々、標的ポリマーに対する対応する訓練化合物のインシリコポーズ品質スコアを表し、対応する第2の正の相互作用スコア及び対応する第2の負の相互作用スコアは、各々、標的ポリマーに対する対応する訓練化合物の結合係数を表し、対応する正の活性スコアは、第1のバイナリ活性スコアであり、対応する負の活性スコアは、第2のバイナリ活性スコアである。いくつかのそのような実施形態では、第2、第3、及び第4の関連付けられた損失関数は、各々独立して、平均二乗誤差損失関数、平均絶対誤差損失関数、Huber損失関数、Log-Cosh損失関数、又は分位損失関数であり、一方、第1の関連付けられた損失関数は、バイナリクロスエントロピー損失関数、ヒンジ損失関数、又は二乗ヒンジ損失関数である。

【0144】

いくつかの実施形態では、図16Bを参照すると、ニューラルネットワーク24によって生成された埋設96は、以下の4つの出力を予測するために使用される：（第1のモデル72を通した）活性、（第2のモデル74を通した）C U i n a ポーズ品質スコア、（第3のモデル76を通した）p K i スコア、及び（第4のモデル990を通した）P o s e N e t スコア。これは、図16Bに示される実施形態では、2段階で実行される。第1に、C U i n a、p K i、及びP o s e N e t スコア予測は、標的ポリマー38に対する試験化合物のポーズのスコアを、ニューラルネットワーク24から第2のモデル74、第3のモデル76、及び第4のモデル990を通して（埋設96として）伝えることによって、算出される。第2に、コンディショニングされた埋設1690は、（i）入力埋設96（ニューラルネットワークスコアからの標的ポリマー38に対する試験化合物のポーズのスコア）と、（i i）第1の段階から結果として生じる第2のモデル74のスコア予測と、（i i i）第1の段階からの第3のモデル76のスコア予測と、を連結させることによって、形成される。次いで、この埋設1690は、第4のモデルの出力とともに、多層パーセプトロンの形態である第1のモデル72に伝えられ、試験化合物の活性予測を算出する。いくつかの実施形態では、図16Bの埋設1690は、（i）入力埋設96（ニューラルネットワークスコアからの標的ポリマー38に対する試験化合物のポーズのスコア）と、（i i）第1の段階から結果として生じる第2のモデル74のスコア予測と、（i i i）第1の段階からの第3のモデル76のスコア予測と、を単に連結させるのではなく、これらの3つのソースを互いに乗算し、乗算の積を埋設1690として第3のモデルに入力する。いくつかの実施形態では、埋設1690は、連結させるではなく、埋設1690において3つのソースの各々を変換し、この変換は、第1のモデル72への入力として機能する。より一般的には、埋設1690は、第1のモデル72に伝えられるコンディショニング埋設を形成するために、乗算、連結、線形又は非線形変換を含むが、これに限定されない埋設1690への入力のいずれかの全て又は任意の部分に対して任意の数学的関数を実行することができる。

【0145】

図10は、本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、（i）バイナリ離散活性及び（i i）p K i であり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。図10には示されていないが、共有埋設層は、化合物のボクセレート（v o x e l a t e d）されたポーズのニューラルネットワーク24への入力時に、ニューラルネットワーク24から出力を受け取る。図10のシステムでは、p K i モデル及び活性モデルは、互いに独立している。いくつかの実施形態では、p K i モデルは、平均二乗誤

差などの損失関数を使用して、回帰タスクとして訓練されるが、一方で、活性モデルは、バイナリコストエントロピーなどの損失関数を使用して、分類タスクとして訓練される。

【0146】

図11は、本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは pK_i であり、 pK_i は、部分的に、活性がコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。図11には示されていないが、共有埋設層は、化合物のボクセレート (voxelated) されたポーズのニューラルネットワーク24への入力時に、ニューラルネットワーク24から出力を受け取る。図11のシステムでは、 pK_i モデルは、活性モデルに対してコンディショニングされる。いくつかの実施形態では、 pK_i モデルは、平均二乗誤差などの損失関数を使用して、回帰タスクとして訓練されるが、一方で、活性モデルは、バイナリコストエントロピーなどの損失関数を使用して、分類タスクとして訓練される。

10

【0147】

図12は、本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、活性は、部分的に、 pK_i 及びポーズ品質スコアの両方に対してコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。図12には示されていないが、共有埋設層は、化合物のボクセレート (voxelated) されたポーズのニューラルネットワーク24への入力時に、ニューラルネットワーク24から出力を受け取る。図12のシステムでは、活性モデルは、 pK_i モデルに対してコンディショニングされる。いくつかの実施形態では、 pK_i モデルは、平均二乗誤差などの損失関数を使用して、回帰タスクとして訓練されるが、一方で、活性モデルは、バイナリコストエントロピーなどの損失関数を使用して、分類タスクとして訓練される。

20

【0148】

図13は、本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは活性であり、活性は、部分的に、 pK_i 及び化合物結合モードスコアの両方に対してコンディショニングされ、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。図13には示されていないが、共有埋設層は、化合物のボクセレート (voxelated) されたポーズのニューラルネットワーク24への入力時に、ニューラルネットワーク24から出力を受け取る。図13のシステムでは、活性モデルは、 pK_i モデル及びpose netモデルの両方に対してコンディショニングされる。いくつかの実施形態では、 pK_i モデル及びpose netモデルは、平均二乗誤差などの損失関数を使用して、回帰タスクとして訓練されるが、一方で、活性モデルは、バイナリコストエントロピーなどの損失関数を使用して、分類タスクとして訓練される。

30

【0149】

図14は、本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、活性及び2つの異なる化合物結合モードスコアであり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。図14のシステムでは、活性モデルは、ポーズ品質スコアモデルに対してコンディショニングされる。いくつかの実施形態では、ポーズ品質モデルは、平均二乗誤差などの損失関数を使用して、回帰タスクとして訓練されるが、一方で、活性モデルは、バイナリコストエントロピーなどの損失関数を使用して、分類タスクとして訓練される。

40

【0150】

図15は、本開示の一実施形態による、試験化合物と標的ポリマーとの間の相互作用を特徴付けるためのシステムであり、特徴付けは、活性、2つの異なる化合物結合モードスコア及び pK_i であり、システムは、化合物を訓練するための結合された正及び負のポーズを使用して訓練される。図15のシステムでは、活性モデルは、ポーズ品質スコアモデ

50

ルに対してコンディショニングされる。いくつかの実施形態では、ポーズ品質モデルは、平均二乗誤差などの損失関数を使用して、回帰タスクとして訓練されるが、一方で、活性モデルは、バイナリコストエントロピーなどの損失関数を使用して、分類タスクとして訓練される。

【0151】

代表的な試験化合物及び訓練化合物。試験化合物と訓練化合物との有意な差は、訓練化合物は（例えば、ウェットラボ結合アッセイから取得された相補的結合データなどで）標識されており、そのような標識が、ニューラルネットワーク24及び本開示の他のモデルを訓練するために使用されるのに対し、試験化合物は標識されず、ニューラルネットワーク24及び本開示の他のモデルが、試験化合物を分類するために使用されることである。言い換えると、訓練化合物は、標識によって既に分類されており、そのような分類は、ニューラルネットワーク24及び本開示の他のモデルを訓練するために使用され、その結果、次いで、本開示のモデルは、試験化合物を分類し得る。試験化合物は、典型的には、ニューラルネットワーク24及び本開示の他のモデルの適用前には分類されない。典型的な実施形態では、訓練化合物に関連付けられた分類は、ウェットラボ結合アッセイによって取得された標的ポリマー38に対する結合データである。

10

【0152】

予測モデルを訓練すること。ディープニューラルネットワーク（例えば、ニューラルネットワーク24）が実装されている、いくつかの実施形態では、ネットワーク24は、幾何学的データ入力を受け取り、所与の試験化合物が標的ポリマーに結合するかどうかの予測（確率）を出力するように訓練される。例えば、いくつかの実施形態では、（それぞれの関連付けられた結合データのために）標的ポリマーに対する周知の結合データを有する訓練化合物は、図2との関連において上記で考察された技術を使用して、本開示のニューラルネットワーク24及びモデルを順次通過し、ニューラルネットワーク24は、各それぞれの訓練化合物に対して、単一値を提供する。

20

【0153】

そのようないくつかの実施形態では、本開示のシステムは、所与の標的化合物に対する各訓練物体に対して、2つの可能な活性クラスのうちの1つを出力する。例えば、本開示のシステムによって各それぞれの訓練化合物に対して提供された単一値は、それが所定の閾値を下回る場合には、第1の活性クラス（例えば、結合剤）にあり、その数が所定の閾値を上回る場合には、第2の活性クラス（例えば、非結合剤）にある。本開示のシステムによって割り当てられた活性クラスは、訓練化合物結合データによって表されるように、実際の活性クラスと比較される。典型的な非限定的な実施形態では、そのような訓練化合物結合データは、独立したウェットラボ結合アッセイからのものである。本開示のシステムによってなされる活性クラス割り当ての誤差は、結合データに対して検証されるように、次いで、システムを訓練するために、本開示のシステムの各モデル（例えば、24、72、74など）の重みを通して逆伝播される。例えば、ネットワークの任意選択的な畳み込み層28内のそれぞれのフィルタのフィルタ重みは、そのような逆伝播で調整される。例示的な実施形態では、ニューラルネットワーク24は、ネットワークにより行われるクラス割り当てにおける誤差に対して、結合データを考慮して、AdaDelta適応学習方法での確率的勾配降下（参照により本明細書に組み込まれる、Zeiler, 2012 “ADADelta: an adaptive learning rate method,” ‘CoRR, vol. abs/1212.5701）と、参照により本明細書に組み込まれる、Rumelhart et al., 1988, “Neurocomputing: Foundations of research,” ch. Learning Representations by Back-propagating Errors, pp. 696-699, Cambridge, MA, USA: MIT Pressに示されている逆伝播アルゴリズムと、によって訓練される。そのようないくつかの実施形態では、2つの可能な活性クラスは、それぞれ、所与の閾値量よりも大きい結合定数（例えば、1ナノモル、10ナノモル、100ナノモル、1マイクロモル、10マイクロモ

30

40

50

ル、100マイクロモル、又は1ミリモルよりも大きい標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKI)及び所与の閾値量を下回る結合定数(例えば、1ナノモル、10ナノモル、100ナノモル、1マイクロモル、10マイクロモル、100マイクロモル、又は1ミリモルよりも小さい標的化合物に関する訓練化合物のIC₅₀、EC₅₀、又はKI)である。

【0154】

いくつかの実施形態では、本開示のシステムは、所与の標的ポリマーに対する各訓練化合物に対して、複数の可能な活性クラス(例えば、3つ以上の活性クラス、4つ以上の活性クラス、5つ以上の活性クラス)のうちの1つを出力する。例えば、本開示のシステムによって各それぞれの訓練化合物に対して提供された単一値は、その数が第1の範囲に入る場合は、第1の活性クラスにあり、その数が第2の範囲に入る場合は、第2の活性クラスにあり、その数が第3の範囲に入る場合は、第3の活性クラスにある、などである。本開示のシステムによって割り当てられた活性クラスは、他の形態の訓練データの訓練化合物結合データによって表されるように、実際の活性クラスと比較される。本開示のシステムによってなされる活性クラス割り当ての誤差は、結合データ(又は他の形態の測定又は独立して計算されたデータ)に対して検証されるように、上記で考察された技術を使用して、本開示のシステムを訓練するために使用される。いくつかの実施形態では、複数の分類における各それぞれ分類は、標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、pK_A、又はKI範囲である。

【0155】

いくつかの実施形態では、本開示のシステムによる複数の訓練化合物の分類は、非パラメトリック技術を使用して、訓練データ(例えば、訓練化合物についての結合データ又は他の独立して測定されたデータ)に対して比較される。例えば、本開示のシステムは、所与の特性(例えば、所与の標的ポリマーに対する結合)に関して複数の訓練化合物をランク付けするために使用され、このランク順序は、複数の訓練化合物のウェットラボ結合アッセイによって取得される訓練データによって提供されるランク順序に対して比較される。これは、上記で考察されたシステム誤差修正技術を使用して、計算されたランク順序の誤差に対して本開示のシステムを訓練する能力を生じさせる。いくつかの実施形態では、本開示のシステムによる訓練化合物によるランク付けと、結合データ(又は訓練化合物についての他の独立して測定されたデータ)によって判定された訓練化合物のランク付けとの間の誤差(差)は、Wilcoxon Mann Whitney関数(Wilcoxon符号順位検定)又は他の非パラメトリック検定を使用して算出され、このような誤差は、上記で考察された誤差修正技術を使用してシステムを更に訓練するために、本開示のシステム(例えば、モデル72、モデル74、モデル24など)を通して逆伝播される。

【0156】

深層学習技術が上で説明されるようにニューラルネットワーク24を利用する実施形態では、ネットワーク24を含むシステムの訓練は、その予測の精度を改善するために、任意選択的な畳み込み層28内のフィルタ内の重み、及びネットワーク層内のバイアスを修正することを含み得る。重み及びバイアスは、L1、L2、重み減衰、及びドロップアウトなどの正規化の様々な形態で、更に制約され得る。

【0157】

ある実施形態では、ニューラルネットワーク24又は本明細書に開示されるモデルのいずれかは、訓練データが(例えば、結合データで)標識される場合、任意選択的に、それぞれのパラメータ(例えば、重み)を調節され(システムの予測された結合親和性及び／又はカテゴリ化と、訓練データの報告された結合親和性及び／又はカテゴリ化との間の誤差を潜在的に最小限に抑えるように調整され)得る。様々な方法は、勾配降下法などの誤差関数を最小限に抑えるために使用され得、これには、対数損失、平方和誤差、ヒンジ損失方法が含まれ得るが、これらに限定されない。これらの方法は、運動量、ヘシアンフリー推定、ネステロフの加速勾配、アダグラッドなどの二次法又は近似を含み得る。標識されていない発生事前訓練及び標識された識別訓練は、組み合わせられ得る。

【0158】

入力された幾何学的データは、訓練例にグループ化され得る。例えば、分子、補因子、及びタンパク質の単一セットは、複数の幾何学的測定値を有する場合が多く、各「スナップ写真」は、標的ポリマー及び訓練化合物が採用し得る代替的な立体構造及びポーズを描写する。同様に、標的ポリマーがタンパク質である場合、タンパク質側鎖、補因子、及び訓練化合物の異なる互変異性体もまたサンプリングされ得る。これらの状態は全て、生物学的システムの挙動に寄与するため、ボルツマン分布に従って、結合親和性を予測するシステムは、（例えば、これらのサンプリングの加重平均を取ることによって）これらの状態をまとめて考慮するように構成され得る。任意選択的に、これらの訓練例は、結合情報で標識され得る。定量的結合情報（例えば、結合データ）が利用可能である場合、そのような標識は、数値的結合親和性であり得る。代替的に、訓練例は、2つ以上の順序付けられたカテゴリ（例えば、結合剤及び非結合剤の2つのカテゴリ、又は1モル未満、1ミリモル未満、100マイクロモル未満、10マイクロモル未満、1マイクロモル未満、100ナノモル未満、10ナノモル未満、1ナノモル未満の力価の結合剤として配位子を記述するいくつかの重複する可能性のあるカテゴリ）のセットから標識を割り当てられ得る。訓練結合データは、実験的測定値、算出された推定値、専門家の洞察、又は推測（例えば、分子及びタンパク質の無作為な対は、結合する可能性が非常に低い）などの、様々なソースから導出又は受け取られ得る。

10

【0159】

ニューラルネットワーク24を使用して、ポーズのスコアを取得する。ニューラルネットワーク24を使用してポーズをスコアリングするために、いくつかの実施形態では、ポーズについてボクセルマップが作成される（例えば、正のポーズの場合は正のボクセルマップ52、負のポーズ60の場合は負のボクセルマップ64）。いくつかの実施形態では、ボクセルマップは、（i）訓練化合物を、正のポーズ48（又はその集合体）又は負のポーズ（又はその集合体）のいずれかでサンプリングすることと、標的ポリマー38を三次元グリッドベースでサンプリングすることによって、対応する複数の空間充填（三次元）多面体セルを含む対応する三次元均一空間充填ハニカムを形成することと、（ii）対応する複数の三次元セル内の各それぞれの三次元多面体セルに対して、それぞれの三次元多面体セルの特性（例えば、化学的特性）に基づいて、それぞれのボクセルマップにボクセル（一定間隔の多面体セルの離散的セット）を投入することと、によって作成される。したがって、特定の訓練化合物に対して、2つのボクセルマップ、正のボクセルマップ52及び負のボクセルマップ65が作成される。空間充填ハニカムの例としては、平行六面体セルを有する立方体ハニカム、六角柱セルを有する六角柱ハニカム、ひし形十二面体セルを有するひし形十二面体、細長い十二面体セルを有する細長い十二面体、及び切頂八面体セルを有する切頂八面体が挙げられる。

20

30

【0160】

いくつかの実施形態では、空間充填ハニカムは、立方体セルを有する立方体ハニカムであり、そのようなボクセルの寸法は、それぞれの分解能を決定する。例えば、各ボクセルが、そのような実施形態では、1 Å寸法（例えば、それぞれのセルのそれぞれの高さ、幅、及び深さにおいて1 Å×1 Å×1 Å）を有する幾何学的データの対応する立方体を表すことを意味する、分解能1 Åが選択され得る。しかしながら、いくつかの実施形態では、より細かいグリッド間隔（例えば、0.1 Å、又は更には0.01 Å）又はより粗いグリッド間隔（例えば、4 Å）が使用され、その間隔は、入力された幾何学的データを網羅するための整数個のボクセルを生じる。いくつかの実施形態では、サンプリングは、0.1 Å～10 Åの分解能で発生する。例示として、40 Å入力立方体の場合、分解能1 Åで、そのような配置は、 $40 \times 40 \times 40 = 64,000$ 入力ボクセルを生じるであろう。

40

【0161】

いくつかの実施形態では、サンプリング（i）で発生した原子の特徴は、それぞれのボクセルマップの単一ボクセルに配列され、複数のボクセルの各ボクセルは、最大1つの原子の特徴を表す。いくつかの実施形態では、原子の特徴は、原子タイプの列挙からなる。

50

一例として、開示されたシステム及び方法のいくつかの実施形態は、ボクセルマップ 40 の所与のボクセル内の全ての原子の存在を、そのエントリの異なる数として表すように構成されており、例えば、炭素がボクセル内にある場合、炭素の原子番号が 6 であるため、6 の値がそのボクセルに割り当てられる。しかしながら、そのような符号化は、近い原子番号を有する原子は、同様に挙動することを暗に意味し得るが、これは、用途によっては、特に有用ではない場合がある。更に、元素の挙動は、族内（周期表の列）でより類似し得、したがって、そのような符号化は、ニューラルネットワーク 24 が復号するための追加の作業を課す。

【0162】

いくつかの実施形態では、原子の特徴は、バイナリカテゴリ変数としてボクセルで符号化される。そのような実施形態では、原子タイプは、「ワンホット」符号化と呼ばれるもので符号化され、全ての原子タイプは、別個のチャンネルを有する。したがって、そのような実施形態では、各ボクセルは、複数のチャンネルを有し、複数のチャンネルの少なくともサブセットは、原子タイプを表す。例えば、各ボクセル内の 1 つのチャンネルは、炭素を表し得、一方で、各ボクセル内の別のチャンネルは、酸素を表し得る。所与のボクセルに対応する三次元グリッド要素に所与の原子タイプが含まれている場合、所与のボクセル内のその原子タイプのチャンネルは、「1」などのバイナリカテゴリ変数の第 1 の値を割り当てられ、所与のボクセルに対応する三次元グリッド要素にその原子タイプが含まれていない場合、その原子タイプのチャンネルは、所与のボクセル内の「0」などのバイナリカテゴリ変数の第 2 の値を割り当てられる。

【0163】

100 個以上の元素があるが、大部分は生物学で遭遇しない。しかしながら、最も一般的な生物元素（例えば、H、C、N、O、F、P、S、Cl、Br、I、Li、Na、Mg、K、Ca、Mn、Fe、Co、Zn）を表すことでさえ、ボクセル当たり 18 チャンネル、すなわち $10,483 \times 18 = 188,694$ 個の入力を受容体フィールドに生じ得る。したがって、いくつかの実施形態では、ボクセルマップ内の各それぞれのボクセルは、複数のチャンネルを含み、複数のチャンネル内の各々のチャンネルは、それぞれのボクセルに対応する三次元空間充填多面体セルで生じ得る異なる特性を表す。所与のボクセルに対して可能なチャンネルの数は、原子の追加の特徴（例えば、部分電荷、タンパク質標的に対する配位子の存在、電気陰性度、又は SYBYL 原子タイプ）が、各ボクセルに対して独立したチャンネルとして更に提示される実施形態では更に高く、そうでなければ同等な原子を区別するためにより多くの入力チャンネルが必要となる。

【0164】

いくつかの実施形態では、各ボクセルは、5 個以上の入力チャンネルを有する。いくつかの実施形態では、各ボクセルは、15 個以上の入力チャンネルを有する。いくつかの実施形態では、各ボクセルは、20 個以上の入力チャンネル、25 個以上の入力チャンネル、30 個以上の入力チャンネル、50 個以上の入力チャンネル、又は 100 個以上の入力チャンネルを有する。いくつかの実施形態では、各ボクセルは、以下の表 1 に含まれている記述子から選択された 5 個以上の入力チャンネルを有する。例えば、いくつかの実施形態では、各ボクセルは、5 個以上のチャンネルを有し、各々、バイナリカテゴリ変数として符号化され、各々のそのようなチャンネルは、以下の表 1 から選択された SYBYL 原子タイプを表す。例えば、いくつかの実施形態では、ボクセルマップ内の各それぞれのボクセルは、それぞれのボクセルによって表される所与の試験物体—標的物体（又は訓練物体—標的物体）複合体の空間内のグリッドが sp3 炭素を包含する場合、チャンネルは第 1 の値（例えば、「1」）を採用し、そうでなければ第 2 の値（例えば、「0」）であることを意味する、C.3（sp3 炭素）原子タイプのチャンネルを含む。

【0165】

10

20

30

40

50

【表 1－1】

表 1－SYBYL 原子タイプ

SYBYL 原子タイプ	記述
C.3	sp3 炭素
C.2	sp2 炭素
C.ar	芳香族炭素
C.1	sp 炭素
N.3	sp3 窒素
N.2	sp2 窒素
N.1	sp 窒素
O.3	sp3 酸素
O.2	sp2 酸素
S.3	sp3 硫黄
N.ar	芳香族窒素
P.3	sp3 リン
H	水素
Br	臭素
Cl	塩素
F	フッ素
I	ヨウ素
S.2	sp2 硫黄
N.pl3	pl3 平面三角形窒素
LP	孤立電子対
Na	ナトリウム
K	カリウム
Ca	カルシウム
Li	リチウム
Al アルミニウム	アルミニウム
Si	シリコン
N.am	アミド窒素
S.o	スルホキシド硫黄
S.o2	スルホン硫黄
N.4	正に荷電した窒素
O.co2	カルボキシレート基又はリン酸塩基中の酸素
C.cat	グアジニウム基でのみ使用されるカルボカチオン
H.spc	SPC 水モデルにおける水素
O.spc	SPC 水モデルにおける酸素

【 0 1 6 6 】

10

20

30

40

50

【表 1－2】

SYBYL 原子タイプ	記述
H.t3p	TIP3P 水モデルにおける水素
O.t3p	TIP3P 水モデル中の酸素
ANY	任意の原子
HEV	重(非 H)原子
HET	ヘテロ原子(N、O、S、P)
HAL	ハロゲン
Mg	マグネシウム
Cr.oh	ヒドロキシクロム
Cr.th	クロム
Se	セレン
Fe	鉄
Cu	銅
Zn	亜鉛
Sn	tin
Mo	モリブデン
Mn	マンガン
Co.oh	ヒドロキシコバルト

【0167】

いくつかの実施形態では、各ボクセルは、上記の表 1 に含まれている記述子から選択された 10 個以上の入力チャネル、15 個以上の入力チャネル、又は 20 個以上の入力チャネルを含む。いくつかの実施形態では、各ボクセルは、ハロゲン用のチャネルを含む。

【0168】

いくつかの実施形態では、第 1 の構造タンパク質配位子相互作用フィンガープリント (SPLIF) スコアは、それぞれの訓練化合物の正のポーズ 48 に対して生成され、第 2 の SPLIF は、訓練化合物の負のポーズ 60 に対して生成される。そのような実施形態では、これらの SPLIF スコアは、基本的なニューラルネットワークへの追加の入力として使用されるか、又はボクセルマップで個別に符号化される。SPLIF の説明については、Da and Kireev, 2014, J. Chem. Inf. Model. 54, pp. 2555–2561, “Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure–Based Virtual Screening: Method and Benchmark Study” を参照されたく、これは参照によりその全体が本明細書に組み込まれる。SPLIF は、訓練化合物の相互作用するフラグメントと標的ポリマー 38 との間に発生し得る全ての可能な相互作用タイプ (例えば、 π – π 、CH– π など) を暗黙で符号化する。第 1 のステップでは、訓練化合物–標的ポリマー 38 の分子間接触を検査する。2 つの原子間の距離が、特定の閾値内 (例えば、4.5 Å 内) であ

る場合、それらは、接触しているとみなされる。そのような各分子間原子対に対して、それぞれの訓練原子及び標的ポリマー原子は、円形のフラグメント、例えば、問題の原子及びそれぞれの連続した近傍を特定の距離まで含むフラグメントに拡張される。各タイプの円形のフラグメントは、識別子を割り当てられる。いくつかの実施形態では、そのような識別子は、それぞれのボクセル内の個々のチャンネルで符号化される。いくつかの実施形態では、パイプラインパイロットソフトウェアで定義されるように、第1の最も近い隣接物（ECFP2）までの拡張接続フィンガープリントを使用することができる。Pipeline Pilot, ver. 8.5, Accelrys Software Inc., 2009を参照されたく、これは参照によりその全体が本明細書に組み込まれる。ECFPは、全ての原子／結合タイプに関する情報を保持し、1つの一意の整数識別子を使用して、1つの下部構造（例えば、円形のフラグメント）を表す。SPLIFフィンガープリントは、含まれる全ての円形のフラグメント識別子を符号化する。いくつかの実施形態では、SPLIFフィンガープリントは、個々のボクセルを符号化されないが、以下で考察されるニューラルネットワーク24における別個の独立した入力として機能する。

10

【0169】

いくつかの実施形態では、SPLIFではなく、又はそれに加えて、構造相互作用フィンガープリント（SIFT）は、標的ポリマーに対する所与の訓練化合物の各ポーズ（正のポーズ48及び負のポーズ60）に対して算出され、かつ、ニューラルネットワーク24への入力として独立して提供されるか、又はボクセルマップで符号化される。SIFTの計算については、Deng et al., 2003, “Structural Interaction Fingerprint (SIFT): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions” J. Med. Chem. 47 (2), pp. 337-344を参照されたく、これは参照によりその全体が本明細書に組み込まれる。

20

【0170】

いくつかの実施形態では、SPLIF及びSIFTではなく、又はそれらに加えて、原子対ベース相互作用フラグメント（APIF）は、標的ポリマー38に対する所与の訓練化合物の各ポーズ（正のポーズ48及び負のポーズ60）に対して算出され、かつ、ニューラルネットワーク24への入力として独立して提供されるか、又はボクセルマップで個別に符号化される。APIFの計算については、Perez-Nueno et al., 2009, “APIF: a new interaction fingerprint based on atom pairs and its application to virtual screening” J. Chem. Inf. Model. 49 (5), pp. 1245-1260を参照されたく、これは参照によりその全体が本明細書に組み込まれる。

30

【0171】

データ表現は、例えば、分子／タンパク質に関連付けられた様々な構造的関係の式を可能にする方法で符号化され得る。幾何学的表現は、様々な実施形態によれば、様々な方法及びトポグラフィで実装され得る。幾何学的表現は、データの視覚化及び分析のために使用される。例えば、ある実施形態では、ジオメトリは、2-D、3-Dデカルト／ユークリッド空間、3-D非ユークリッド空間、マニホールドなどの様々なトポグラフィ上にレイアウトされたボクセルを使用して表され得る。例えば、図4は、ある実施形態による、一連のサブコンテナを含むサンプルの三次元グリッド構造400を示す。各サブコンテナ402は、ボクセルに対応し得る。座標系は、各サブコンテナが識別子を有するように、グリッドのために定義され得る。開示されたシステム及び方法のいくつかの実施形態では、座標系は、3-D空間におけるデカルト系であるが、システムの他の実施形態では、座標系は、とりわけ、偏球、円筒座標系又は球面座標系、極座標系、様々なマニホールド及びベクトル空間用に設計された他の座標系などの任意の他のタイプの座標系であり得る。いくつかの実施形態では、ボクセルは、それぞれに関連付けられた特定の値を有し得、こ

40

50

れは、例えば、とりわけ、標識を適用すること、及び／又はそれぞれの位置決めを判定することによって表され得る。

【0172】

ニューラルネットワークは固定の入力サイズを必要とするため、開示されたシステム及び方法のいくつかの実施形態は、幾何学的データ（標的試験又は標的訓練物体複合体）を、適切な境界ボックス内に収まるようにトリミングする。例えば、側面に対する25～40 Åの立方体が使用され得る。標的及び／又は試験物体が標的物体58の活性部位にドッキングされているいくつかの実施形態では、活性部位の中心は、立方体の中心として機能する。

【0173】

いくつかの実施形態では、標的ポリマー38の活性部位を中心とする固定寸法の四角い立方体が、空間をボクセルグリッドに分割するために使用されるが、開示されたシステムは、そのように限定されない。いくつかの実施形態では、様々な形状のいずれかが、空間をボクセルグリッドに分割するために使用される。いくつかの実施形態では、直角プリズム、多面体形状などの多面体を使用して空間を分割する。

【0174】

ある実施形態では、グリッド構造は、ボクセルの配置に類似するように構成され得る。例えば、各サブ構造は、分析される各原子に対してチャンネルに関連付けられ得る。また、各原子を数値的に表すための符号化方法が提供され得る。

【0175】

いくつかの実施形態では、ボクセルマップは、（例えば、訓練化合物ポーズ及び標的ポリマーの分子力学の実行に沿った）時間の因子を考慮し、したがって、四次元（X、Y、Z、及び時間）であり得る。

【0176】

いくつかの実施形態では、画素、点、多角形、多面体、又は複数の次元の任意の他のタイプの形状（例えば、3D、4Dなどの形状）などの他の実施態様は、ボクセルの代わりに使用され得る。

【0177】

いくつかの実施形態では、幾何学的データは、空洞フラッディングアルゴリズムによって判定されるように、標的ポリマー38の結合部位の質量中心であるように、X座標、Y座標及びZ座標の原点を選択することによって正規化される。そのようなアルゴリズムの代表的な詳細については、Ho and Marshall, 1990, "Cavity search: An algorithm for the isolation and display of cavity-like binding regions" Journal of Computer-Aided Molecular Design 4, pp. 337-354及びHendlich et al., 1997, "Ligsite: automatic and efficient detection of potential small molecule-binding sites in proteins" J. Mol. Graph. Model 15:6を参照されたく、これらは各々、参照によりその全体が本明細書に組み込まれる。代替的に、いくつかの実施形態では、ボクセルマップの原点は、（標的ポリマーに結合したそれぞれのポーズ—正のポーズ48又は負のポーズ60—でドッキングされた訓練化合物の）共複合体全体の質量中心を中心とする。いくつかの実施形態では、ボクセルマップの原点は、訓練化合物の質量中心を中心とする。いくつかの実施形態では、ボクセルマップの原点は、標的ポリマー38の質量中心を中心とする。基底ベクトルは、任意選択的に、共複合体全体、標的ポリマーのみ、又は訓練化合物のみの主慣性モーメントであるように選択され得る。いくつかの実施形態では、標的ポリマー38は、活性部位を有し、サンプリングは、正のポーズ48及び負のポーズ60の両方で訓練化合物をサンプリングし、活性部位の質量中心が、原点、及びサンプリング用の対応する三次元の均一なハニカムとしてとられる三次元グリッドベースでの活性部位は、質量中心を中心とするポリマー及び訓練化合

10

20

30

40

50

物の一部を表す。いくつかの実施形態では、均一なハニカムは、規則的な立方体のハニカムであり、ポリマー及び試験物体の一部分は、所定の固定寸法の立方体である。そのような実施形態では、所定の固定寸法の立方体の使用により、幾何学的データの関連部分を使用され、各ボクセルマップが同じサイズであることを確実にする。いくつかの実施形態では、立方体の所定の固定寸法は、 $N \text{ \AA} \times N \text{ \AA} \times N \text{ \AA}$ であり、式中、 N は、 $5 \sim 100$ の整数若しくは実数、 $8 \sim 50$ の整数、又は $15 \sim 40$ の整数である。いくつかの実施形態では、均一なハニカムは、直角プリズムハニカムであり、ポリマー及び訓練化合物の部分は、直角プリズム所定の固定寸法 $Q \text{ \AA} \times R \text{ \AA} \times S \text{ \AA}$ であり、式中、 Q は、 $5 \sim 100$ の第1の整数であり、 R は、 $5 \sim 100$ の第2の整数であり、 S は、 $5 \sim 100$ の第3の整数又は実数であり、セット $\{Q, R, S\}$ 内の少なくとも1つの数は、セット $\{Q, R, S\}$ 内の別の値に等しくない。

10

【0178】

ある実施形態では、全てのボクセルは、1つ以上の入力チャンネルを有し、入力チャンネルは、それぞれに関連付けられた様々な値を有し得、単純な実装では、オン／オフであり得、あるタイプの原子に対して符号化するように構成され得る。原子タイプは、原子の元素を表示し得るか、又は原子タイプは、他の原子の特徴を区別するために更に精緻化され得る。次いで、存在している原子は、各ボクセルで符号化され得る。様々なタイプの符号化は、様々な技術及び／又は方法論を使用して利用され得る。例示的な符号化方法として、原子の原子番号が利用され得、水素の1からウンウンオクチウム（又は任意の他の元素）の118までのボクセルごとに1つの値を生じる。

20

【0179】

しかしながら、上記で考察されるように、「ワンホット符号化」などの他の符号化方法を利用し得、各ボクセルは、それぞれオン又はオフのいずれかであり、あるタイプの原子に対して符号化する、多くの並列入力チャンネルを有する。原子タイプは、原子の元素を表示し得るか、又は原子タイプは、他の原子の特徴を区別するために更に精緻化され得る。例えば、SYBYL原子タイプは、単一結合炭素を、二重結合炭素、三重結合炭素、又は芳香族炭素と区別する。SYBYL原子タイプについては、Clark et al., 1989, "Validation of the General Purpose Tripos Force Field, 1989, J. Comput. Chem. 10, pp. 982-1012を参照されたく、これは参照により本明細書に組み込まれる。

30

【0180】

いくつかの実施形態では、各ボクセルは、標的ポリマー38の一部である原子、又は訓練化合物の一部に対する補因子を区別するための1つ以上のチャンネルを更に含む。例えば、一実施形態では、各ボクセルは、標的ポリマー38用の第1のチャンネル及び訓練化合物用の第2のチャンネルを更に含む。ボクセルによって表される空間の部分における原子が、標的ポリマー38からである場合、第1のチャンネルは、「1」などの値に設定され、そうでなければゼロである（例えば、ボクセルによって表される空間の部分が、原子を含まないか、又は訓練化合物からの1つ以上の原子を含むため）。更に、ボクセルによって表される空間の部分における原子が、訓練化合物からである場合、第2のチャンネルは、「1」などの値に設定され、そうでなければゼロである（例えば、ボクセルによって表される空間の部分が、原子を含まないか、又は標的ポリマー38からの1つ以上の原子を含むため）。同様に、他のチャンネルは、追加的に（又は代替的に）部分電荷、分極率、電気陰性度、溶媒接触可能空間、及び電子密度などの更なる情報を指定し得る。例えば、いくつかの実施形態では、標的物体の電子密度マップは、三次元座標のセットを覆い、ボクセルマップの作成は、電子密度マップを更にサンプリングする。好適な電子密度マップの例としては、複数の同形の置換マップ、異常信号マップを有する単一の同形置換、単一波長異常分散マップ、多波長異常分散マップ、及び $2F_o - F_c$ マップが挙げられるが、これらに限定されない（260）。McRee, 1993, Practical Protein Crystallography, Academic Pressを参照されたく、これは参照により本明細書に組み込まれる。

40

50

【0181】

いくつかの実施形態では、開示されたシステム及び方法によるボクセル符号化は、追加の任意の符号化改良を含み得る。以下の2つが、例として提供される。

【0182】

第1の符号化改良では、必要なメモリは、大部分の元素が生物学的システムにめったに発生しないことに基づいて、ボクセルによって表される原子のセットを低減することによって（例えば、ボクセルによって表されるチャンネルの数を低減することによって）低減され得る。原子は、希原子を組み合わせることによるか（したがって、システムの性能にほとんど影響を与えない場合がある）、又は類似した特性を有する原子を組み合わせることによるか（したがって、組み合わせからの不正確さを最小限に抑え得る）のいずれかにより、ボクセル内の同じチャンネルを共有するようにマッピングされ得る。いくつかの実施形態では、2個、3個、4個、5個、6個、7個、8個、9個、又は10個の異なる原子が、ボクセル内の同じチャンネルを共有する。

10

【0183】

符号化の改良は、隣接ボクセルを部分的に活性化することによって、ボクセルが原子の位置を表すようにすることである。これは、後続のニューラルネットワークにおける隣接ニューロンの部分的な活性化をもたらし、ワンホット符号化から「複数ウォーム（several-warm）」符号化に移行する。例えば、 $1\text{ \AA}^3$ グリッドが配列されるときに、ファンデルワールス径が $3.5\text{ \AA}$ 、したがって体積が $22.4\text{ \AA}^3$ である塩素原子を考慮することが例示的であり得、塩素原子内のボクセルは、完全に充填され、原子の端縁でのボクセルは、部分的に充填されるだけである。したがって、部分的に充填されたボクセル内の塩素を表すチャンネルは、そのようなボクセルが塩素原子内に収まる量に比例してオンになる。例えば、ボクセル体積の50%が塩素原子内に収まる場合、塩素を表すボクセル内のチャンネルは、50%活性化される。これは、離散ワンホット符号化に対して「平滑化された」より正確な表現をもたらし得る。したがって、いくつかの実施形態では、サンプリングで発生した原子の特徴は、ボクセルマップ内のボクセルのサブセットに分散され、ボクセルのこのサブセットは、2個以上のボクセル、3個以上のボクセル、5個以上のボクセル、10個以上のボクセル、又は25個以上のボクセルを含む。いくつかの実施形態では、原子の特徴は、原子タイプ（例えば、SYBYL原子タイプのうちの1つ）の列挙からなる。

20

30

【0184】

したがって、符号化された幾何学的データのボクセル化（ラスタ化）（試験又は訓練物体の標的物体へのドッキング）は、入力データに適用された様々なルールに基づいている。

【0185】

図5及び図6は、いくつかの実施形態による、ボクセルの二次元グリッド500上に符号化された2つの分子502の図を提供する。図5は、二次元グリッド上に重畳された2つの分子を提供する。図6は、異なる斜線パターンを使用して、酸素、窒素、炭素、及び空きスペースの存在をそれぞれ符号化する、ワンホット符号化を提供する。上で述べられるように、そのような符号化は、「ワンホット」符号化と称され得る。図6は、分子502を省略した図5のグリッド500を示す。図7は、ボクセルに番号が付けられた図6のボクセルの二次元グリッドの図を提供する。

40

【0186】

いくつかの実施形態では、特徴形状は、ボクセル以外の形態で表される。図8は、特徴（例えば、原子中心）が0-D点（表現802）、1-D点（表現804）、2-D点（表現806）、又は3-D点（表現808）として表現されている様々な表現の図を提供する。最初に、点間の間隔は、無作為に選択され得る。しかしながら、予測モデルが訓練されるにつれて、点は、互いに近づき得るか、又はより遠くに離れ得る。

【0187】

いくつかの実施形態では、入力された表現は、三次元座標を含むが、これらに限定され

50

ない、特徴の1Dアレイの形態であり得る。

【0188】

いくつかの実施形態では、ニューラルネットワーク24は、グラフ重畳型ニューラルネットワークである。グラフ重畳型ニューラルネットワークの非限定的な例は、Behler Parrinello, 2007, "Generalized Neural-Network Representation of High Dimensional Potential-Energy Surfaces," *Physical Review Letters* 98, 146401、Chmiela et al., 2017, "Machine learning of accurate energy-conserving molecular force fields," *Science Advances* 3 (5): e1603015、Schuett et al., 2017, "SchNet: A continuous-filter convolutional neural network for modeling quantum interactions," *Advances in Neural Information Processing Systems* 30, pp. 992-1002、Feinberg et al., 2018, "PotentialNet for Molecular Property Prediction," *ACS Cent. Sci.* 4, 11, 1520-1530、及びStafford et al., "AtomNet PoseRanker: Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High Throughput Screens," <https://chemrxiv.org/engage/chemrxiv/article-details/614b905e39ef6a1c36268003>に開示されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

【0189】

いくつかの実施形態では、ニューラルネットワークは、等変量ニューラルネットワークである。等変量重畳型ニューラルネットワークの非限定的な例は、Thomas et al., 2018, "Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds," *arXiv:1802.08219*、Anderson et al., 2019, "Cormorant: Covariant Molecular Neural Networks," *Neural Information Processing Systems*、Johannes et al., 2020, "Directional Message Passing For Molecular Graphs," *International Conference on Learning Representations*、Townshend et al., 2021, "ATOM3D: Tasks On Molecules in Three Dimensions," *International Conference on Learning Representations*、Jing et al., 2009, "Learning from Protein Structure with Geometric Vector Perceptrons," *arXiv:2009.01411*、及びSatorras et al., 2021, "E(n) Equivariant Graph Neural Networks," *arXiv:2102.09844*に開示されており、これらは各々、参照によりその全体が本明細書に組み込まれる。

【0190】

いくつかの実施形態では、ニューラルネットワーク24は、参照により本明細書に組み込まれる、2022年5月10日出願の"Characterization of Interactions Between Compounds and Polymers Using Pose Ensembles"と題された米国仮特許出願第63

／ 336, 841号に開示されたグラフニューラルネットワークのいずれかである。

【0191】

ボクセルマップを、対応するベクトルに展開する。各ボクセルマップ（例えば、正のボクセルマップ52及び負のボクセルマップ64）は、任意選択的に、対応するベクトル（例えば、訓練データセット40の各訓練化合物に対する正のベクトル54及び負のベクトル66）に展開される。いくつかの実施形態では、そのような各ベクトルは、一次元ベクトルである。例えば、いくつかの実施形態では、各側面の20 Åの立方体は、標的ポリマー38の活性部位を中心とし、1 Åの三次元固定グリッド間隔でサンプリングされて、上記で考察されるように、原子タイプ、及び、任意選択的に、より複雑な訓練化合物－標的ポリマー記述子などのボクセル構造特徴の基本的なそれぞれのチャンネルに適用される、ボクセルマップの対応するボクセルを形成する。いくつかの実施形態では、この三次元ボクセルマップのボクセルは、一次元浮動小数点ベクトルに展開される。

10

【0192】

いくつかの実施形態では、ボクセルマップ（例えば、訓練データセット40の各訓練化合物に対する正のベクトル54及び負のベクトル66）のベクトル化された表現は、ニューラルネットワーク24に供される。いくつかの実施形態では、図1Bに示されるように、ボクセルマップのベクトル化された表現は、査定モジュール20、及びニューラルネットワーク24とともにGPUメモリ52に格納される。これは、ニューラルネットワーク24を通してより速い速度でボクセルマップのベクトル化された表現を処理することの利点を提供する。しかしながら、他の実施形態では、ボクセルマップのベクトル化された表現（例えば、訓練データセット40の各訓練化合物に対する正のベクトル54及び負のベクトル66）、査定モジュール20、及びニューラルネットワーク24のいずれか又は全ては、システム100のメモリ92内にあるか、又は単にネットワーク全域でシステム92によってアドレス指定可能である。いくつかの実施形態では、ボクセルマップのベクトル化された表現、査定モジュール20、及びニューラルネットワーク24のいずれか又は全ては、クラウドコンピューティング環境にある。

20

【0193】

いくつかの実施形態では、ベクトル（例えば、訓練データセット40の各訓練化合物に対する正のベクトル54及び負のベクトル66）は、図形処理ユニットメモリ52に提供され、図形処理ユニットメモリは、複数のベクトルを順次受け取るための入力層26、任意選択的に、複数の畳み込み層28、及びスコアラ30を含むニューラルネットワーク24を含む、ネットワークアーキテクチャを含む。いくつかの実施形態では、任意の複数の畳み込み層は、初期畳み込み層及び最終畳み込み層を含む。いくつかの実施形態では、ニューラルネットワーク24は、GPUメモリ内にはないが、システム100の汎用メモリ内にある。いくつかの実施形態では、ボクセルマップは、ネットワーク24に入力される前にベクトル化されない。

30

【0194】

畳み込み層28のユーザを作るいくつかの実施形態では、複数の畳み込み層の畳み込み層28は、学習可能なフィルタ（カーネルとも呼ばれる）のセットを備える。各フィルタは、畳み込み層の入力体積の深さ、高さ、及び幅にわたって畳み込まれる（所定のステッププレートでステップされる）固定された三次元サイズを有し、フィルタのエントリ（重み、又はより一般的にはパラメータ）と入力との間のドット積（又は他の関数）を算出し、それによって、そのフィルタの多次元活性化マップを作成する。いくつかの実施形態では、フィルタのステッププレートは、入力空間の1個の要素、2個の要素、3個の要素、4個の要素、5個の要素、6個の要素、7個の要素、8個の要素、9個の要素、10個の要素、又は10個よりも多い要素である。したがって、フィルタのサイズが 5^3 の場合を考慮する。いくつかの実施形態では、このフィルタは、ボクセルチャンネル当たり125個の入力空間の値の総数に対し、5個の要素からなる深さ、5個の要素からなる幅、及び5個の要素からなる高さを有する入力空間の連続した立方体の間のドット積（又は他の数学的関数）を算出する。

40

50

【0195】

初期畳み込み層（例えば、入力層26からの出力）への入力空間は、ボクセルマップ、又はボクセルマップのベクトル化された表現（例えば、訓練データセット40の各訓練化合物に対する正のベクトル54及び負のベクトル66）のいずれかから形成される。いくつかの実施形態では、ボクセルマップのベクトル化された表現は、初期畳み込み層への入力空間として機能するボクセルマップの一次元ベクトル化された表現である。それにもかかわらず、フィルタが、その入力空間を畳み込み、入力空間が、ボクセルマップの一次元ベクトル化された表現である場合、フィルタは、それでも一次元ベクトル化された表現から、標的ポリマー38－訓練化合物複合体の固定空間の対応する連続した立方体を表す要素を取得する。いくつかの実施形態では、フィルタは、簿記技術を使用して、標的ポリマー38－訓練化合物複合体の固定空間の対応する連続した立方体を形成する一次元ベクトル化された表現の中から、それらの要素を選択する。したがって、いくつかの例では、これは、標的ポリマー38－訓練化合物複合体の固定空間の対応する連続した立方体の要素値を取得するために、一次元ベクトル化された表現における要素の非連続的なサブセットを取ることを必然的に伴う。

10

【0196】

いくつかの実施形態では、フィルタは、フィルタに対応する活性化層の第1の単一値（又は値のセット）を算出するために、125個の入力空間値のドット積（又は図14に開示される関数などの、他の何らかの形態の数学的演算）を取るための（入力チャンネル当たり）125個の対応する重みを有するように（例えば、ガウスノイズに）初期化されるか、訓練される。いくつかの実施形態では、フィルタによって算出された値は、合計され、重み付けされ、及び／又はバイアスされる。フィルタに対応する活性化層の加算値を算出するために、フィルタは、次いで、フィルタに関連付けられたステップレート（ストライド）によって入力体積の三次元のうちの1つでステップ（畳み込み）され、そこで、フィルタ重みと（チャンネル当たり）125個の入力空間値のドット積（又は図17に開示される数学的関数などの、他の何らかの形態の数学的演算）が、入力体積内の新しい場所で取られる。このステップすること（畳み込み）は、フィルタがステップレートに従って入力空間全体をサンプリングするまで繰り返される。いくつかの実施形態では、入力空間の辺縁は、畳み込み層によって生成された出力空間の空間体積を制御するために、ゼロパディングされる。典型的な実施形態では、畳み込み層のフィルタの各々は、このようにして三次元入力体積全体をキャンバスで覆い、それによって、対応する活性化マップを形成する。畳み込み層のフィルタからの活性化マップの集合は、1つの畳み込み層の三次元出力体積をまとめて形成し、それによって、後続の畳み込み層の三次元（3つの空間次元）入力として機能する。したがって、出力体積内の全てのエントリはまた、畳み込み層への入力空間内の小さい領域を見て、同じ活性化マップ内のニューロンとパラメータを共有する単一のニューロン（又はニューロンのセット）の出力として解釈することもできる。したがって、いくつかの実施形態では、複数の畳み込み層の畳み込み層は、複数のフィルタを有し、複数のフィルタの各フィルタは、（3つの空間次元で）ストライドYを有する N^3 の立体入力空間を畳み込み、ここで、Nは、2つ以上（例えば、2、3、4、5、6、7、8、9、10、又は10超）の整数であり、Yは、正の整数（例えば、1、2、3、4、5、6、7、8、9、10、又は10超）である。

20

30

40

【0197】

複数の畳み込み層の各層は、異なる重みのセット、又はより一般的には異なるパラメータのセットに関連付けられている。より詳細には、複数の畳み込み層の各層は、複数のフィルタを含み、各フィルタは、独立した複数のパラメータ（例えば、重み）を含む。いくつかの実施形態では、畳み込み層は、次元 5^3 の128個のフィルタを有し、したがって、畳み込み層は、ボクセルマップ内のチャンネル当たり $128 \times 5 \times 5 \times 5$ すなわち16,000個のパラメータ（例えば、重み）を有する。したがって、ボクセルマップに5つのチャンネルがある場合、畳み込み層は、 $16,000 \times 5$ 個のパラメータ（例えば、重み）、すなわち80,000個パラメータ（例えば、重み）を有する。いくつかの実施形態で

50

は、所与の畳み込み層内の全てのフィルタの一部又は全部のそのようなパラメータ（及び、任意選択的に、バイアス）は、ともに結び付けられ、例えば、同一であるように制約され得る。

【0198】

それぞれのベクトル（例えば、正のベクトル54又は負のベクトル66）の入力にตอบสนองして、入力層26は、それぞれのベクトル内の値の第1の関数として、第1の複数の値を初期畳み込み層に供給し、第1の関数は、任意選択的に、図形処理ユニット50を使用して算出される。いくつかの実施形態では、コンピュータシステム100は、2つ以上の図形処理ユニット50を有する。

【0199】

最終畳み込み層以外の各それぞれの畳み込み層28は、(i)それぞれの畳み込み層に関連付けられたパラメータ（例えば、重み）の異なるセット及び(ii)それぞれの畳み込み層によって受け取られた入力値のそれぞれの第2の関数として、中間値を複数の畳み込み層の別の畳み込み層に供給する。いくつかの実施形態では、第2の関数は、図形処理ユニット50を使用して算出される。例えば、いくつかの実施形態では、それぞれの畳み込み層28の各それぞれのフィルタは、畳み込み層の特徴的な三次元ストライドに従って、畳み込み層に対して入力体積を（3つの空間次元で）キャンバスで覆い、各それぞれのフィルタ位置において、それぞれのフィルタのフィルタパラメータ（例えば、重み）と、それぞれのフィルタ位置における入力体積の値（総入力空間のサブセットである連続した立方体）のドット積（又は他の何らかの数学的関数）を取り、それによって、それぞれのフィルタ位置に対応する活性化層上の計算されたポイント（又はポイントのセット）を生成する。それぞれの畳み込み層のフィルタの活性化層は、それぞれの畳み込み層の中間値をまとめて表す。

【0200】

最終畳み込み層は、(i)最終畳み込み層に関連付けられた異なるパラメータのセット（例えば、重み）及び(ii)任意選択的に、図形処理ユニット50を使用して算出される最終畳み込み層によって受け取られた入力値の第3の関数として、最終値をスコアラに供給する。例えば、最終畳み込み層28の各それぞれのフィルタは、畳み込み層の特徴的な三次元ストライドに従って、最終畳み込み層に対して入力体積を（3つの空間次元で）キャンバスで覆い、各それぞれのフィルタ位置において、フィルタのフィルタ重みと、それぞれのフィルタ位置における入力体積の値のドット積（又は他の何らかの数学的関数）を取り、それによって、それぞれのフィルタ位置に対応する活性化層上の計算されたポイント（又はポイントのセット）を計算する。最終畳み込み層のフィルタの活性化層は、スコアラ30に供給される最終値をまとめて表す。

【0201】

いくつかの実施形態では、重畳型ニューラルネットワークは、1つ以上の活性化層を有する。いくつかの実施形態では、活性化層は、非飽和活性化関数 $f(x) = \max(0, x)$ を適用するニューロンの層である。それは、畳み込み層の受容野に影響を与えずに、決定関数及び全体ネットワークの非線形特性を増加させる。他の実施形態では、活性化層は、非線形性を増加させるための他の関数、例えば、飽和双曲線正接関数 $f(x) = \tanh(x)$ 、 $f(x) = |\tanh(x)|$ 、及びシグモイド関数、及びシグモイド関数 $f(x) = (1 + e^{-x})^{-1}$ を有する。ニューラルネットワークのためのいくつかの実施形態では、他の活性化層に含まれている他の活性化関数の非限定的な例としては、ロジスティック（又はシグモイド）、ソフトマックス、ガウス、ボルツマン重み付け平均化、絶対値、線形、整流線形、有界整流線形、ソフト整流線形、パラメータ化整流線形、平均、最大、最小、何らかのベクトルノルム L_p ($p = 1, 2, 3, \dots, \infty$ の場合)、符号、平方、平方根、複二次、逆二次、逆複二次、多高調波スプライン、及び薄板スプラインが挙げられ得るが、これらに限定されない。

【0202】

ネットワーク24は、入力内の何らかの空間位置において、何らかの特定のタイプの特

10

20

30

40

50

徴を見たときに活性化する畳み込み層 28 内のフィルタを学習する。いくつかの実施形態では、畳み込み層内の各フィルタの初期パラメータ（例えば、重み）は、化合物訓練ライブラリに対して重畳型ニューラルネットワークを訓練することによって、取得される。したがって、重畳型ニューラルネットワーク 24 の動作により、結合親和性予測を実施するために歴史的に使用された特徴よりも複雑な特徴を生じ得る。例えば、水素結合検出器として機能するネットワーク 24 の所与の畳み込み層内のフィルタは、水素結合ドナー及びアクセプターが、所与の距離及び角度にあることを認識するだけでなく、ドナー及びアクセプターの周りの生化学的環境が、結合を強める又は弱めることも認識することが可能であり得る。加えて、ネットワーク 24 内のフィルタは、基本的なデータ内の結合剤と非結合剤を効果的に区別するように訓練され得る。

10

【0203】

上で説明されるように、いくつかの実施形態では、ニューラルネットワーク 24 は、三次元畳み込み層を形成するように構成される。最下位レベルの畳み込み層 28 への入力領域は、受容野からのボクセルチャネルの立方体（又は他の連続した領域）であり得る。より高位の畳み込み層 28 は、より低位の畳み込み層からの出力を評価するが、それでもそれぞれの出力を、（3-D ユークリッド距離で）互いに近いボクセルの有界領域の関数とする。

【0204】

ある実施形態では、ネットワーク 24 は、正規化技術を適用して、モデルの訓練データを過剰適合する傾向を低減するように構成されている。

20

【0205】

ネットワーク 24 内のネットワーク層のうちのゼロ個以上は、プーリング層からなり得る。畳み込み層と同様に、プーリング層は、異なる空間局所的な入力のパッチに同じ関数を適用する関数計算のセットである。プーリング層の場合、出力は、プーリング演算子、例えば、いくつかのボクセルにわたる、 $p = 1, 2, 3, \dots, \infty$ の場合の何らかのベクトルノルム L_p によって与えられる。プーリングは、典型的には、チャンネルにわたってではなく、チャンネルごとに行われる。プーリングは、入力空間を三次元ボックスのセットに分割し、そのような各サブ領域に対して、最大値を出力する。プーリング演算は、並進不変の一形態を提供する。プーリング層の機能は、ネットワーク内のパラメータ及び計算の量を低減するために、よってまた、過剰適合も制御するために、表現の空間サイズを漸次的に低減することである。いくつかの実施形態では、プーリング層は、ネットワーク 24 内の連続した畳み込み 28 層の間に挿入される。そのようなプーリング層は、入力の深さスライスごとに独立して作用し、サイズを空間的に変更する。最大プーリングに加えて、プーリングユニットはまた、平均プーリング又は L_2 -ノルムプーリングなどの他の機能も実行することができる。

30

【0206】

ネットワーク 24 内のゼロ個以上の層は、局所反応正規化又は局所コントラスト正規化などの正規化層からなり得、これらは、同じ位置でチャンネルにわたるか、又はいくつかの位置にわたる特定のチャンネルに対して適用され得る。これらの正規化層は、同じ入力に対するいくつかの関数計算の応答の多様性を促進し得る。

40

【0207】

いくつかの実施形態では、スコアラ 30 は、複数の完全接続層と、複数の完全接続層の完全接続層が評価層に供給する評価層と、を含む。規則正しいニューラルネットワークに見られるように、完全接続層のニューロンは、前の層の全ての活性化に対する完全接続を有する。よって、その活性化は、バイアスオフセットが続く行列乗算で算出することができる。いくつかの実施形態では、各完全接続層は、512 個の隠れユニット、1024 個の隠れユニット、又は 2048 個の隠れユニットを有する。いくつかの実施形態では、スコアラには、完全接続層はない、1 個の完全接続層、2 個の完全接続層、3 個の完全接続層、4 個の完全接続層、5 個の完全接続層、6 個以上の完全接続層、又は 10 個以上の完全接続層がある。

50

【0208】

いくつかの実施形態では、評価層は、複数の活性クラスを区別する。いくつかの実施形態では、評価層は、2つの活性クラス、3つの活性クラス、4つの活性クラス、5つの活性クラス、又は6つ以上の活性クラスにわたるロジスティック回帰コスト層を含む。

【0209】

いくつかの実施形態では、評価層は、複数の活性クラスにわたるロジスティック回帰コスト層を含む。いくつかの実施形態では、評価層は、2つの活性クラス、3つの活性クラス、4つの活性クラス、5つの活性クラス、又は6つ以上の活性クラスにわたるロジスティック回帰コスト層を含む。

【0210】

いくつかの実施形態では、評価層は、2つの活性クラスを区別し、第1の活性クラス（第1の分類）は、第1の結合値を上回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀又はKIを表し、第2の活性クラス（第2の分類）は、第1の結合値を下回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKIである。いくつかの実施形態では、第1の結合値は、1ナノモル、10ナノモル、100ナノモル、1マイクロモル、10マイクロモル、100マイクロモル、又は1ミリモルである。

【0211】

いくつかの実施形態では、評価層は、2つの活性クラスにわたるロジスティック回帰コスト層を含み、第1の活性クラス（第1の分類）は、第1の結合値を上回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀又はKIを表し、第2の活性クラス（第2の分類）は、第1の結合値を下回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKIである。いくつかの実施形態では、第1の結合値は、1ナノモル、10ナノモル、100ナノモル、1マイクロモル、10マイクロモル、100マイクロモル、又はミリモルである。

【0212】

いくつかの実施形態では、評価層は、3つの活性クラスを区別し、第1の活性クラス（第1の分類）は、第1の結合値を上回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀又はKIを表し、第2の活性クラス（第2の分類）は、第1の結合値と第2の結合値との間にある標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKIであり、第3の活性クラス（第3の分類）は、第2の結合値を下回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKIであり、第1の結合値は、第2の結合値以外である。

【0213】

いくつかの実施形態では、評価層は、3つの活性クラスにわたるロジスティック回帰コスト層を含み、第1の活性クラス（第1の分類）は、第1の結合値を上回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀又はKIを表し、第2の活性クラス（第2の分類）は、第1の結合値と第2の結合値との間にある標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKIであり、第3の活性クラス（第3の分類）は、第2の結合値を下回る標的ポリマーに関する訓練化合物のIC₅₀、EC₅₀、又はKIであり、第1の結合値は、第2の結合値以外である。

【0214】

いくつかの実施形態では、スコアラ30は、完全接続単層又は多層パーセプトロンを備える。いくつかの実施形態では、スコアラは、サポートベクターマシン、ランダムフォレスト、最も近い隣接物を含む。いくつかの実施形態では、スコアラ30は、入力を様々な出力カテゴリに分類する強度（又は確実性又は確率）を示す数値スコアを割り当てる。いくつかの例では、カテゴリは、結合剤及び非結合剤、又は代替的に、効力レベル（例えば、1モル未満、1ミリモル未満、100マイクロモル未満、10マイクロモル未満、1マイクロモル未満、100ナノモル未満、10ナノモル未満、1ナノモル未満のIC₅₀、EC₅₀又はKI効力）である。

【0215】

使用事例。

以下は、本開示のいくつかの実施形態のいくつかの用途を説明する、例示的な目的のためにのみ提供されるサンプル使用事例である。他の使用が考慮され得、以下に提供される実施例は、非限定的であり、変形、省略を受けることがあるか、又は追加の要素を含み得る。

【0216】

ヒット発見。製薬会社は、新規の有望な薬物リードを発見するために、化合物のスクリーニングに数百万ドルを費やしている。大規模な化合物コレクションを試験して、関心対象の疾患標的との任意の相互作用を有するわずかな化合物を発見する。残念ながら、ウェットラボスクリーニングは実験誤差を被り、アッセイ実験を実行するためのコスト及び時間に加えて、大規模なスクリーニングコレクションの収集は、保管の制約、棚安定性、又は化学薬品コストを通じた重要な課題を課す。最大の製薬会社でさえ、数十万から数百万の化合物を有するにすぎないのに対し、数千万の市販の分子及び数億のシミュレート可能な分子がある。

10

【0217】

物理的実験に対する潜在的により効率的な代替案は、仮想高スループットスクリーニングである。物理学シミュレーションが、モデルが物理的に試験される前に、航空宇宙エンジニアが考えられる翼の設計を評価するのに役立つ場合があるのと同じように、コンピュータによる分子のスクリーニングは、見込みが大いにある分子の小さいサブセットに実験的試験を集中させることができる。これにより、スクリーニングのコスト及び時間を低減し得、偽陰性を低減し、成功率を改善し、及び／又は化学的環境のより広範なスワスを網羅する。

20

【0218】

本出願では、タンパク質標的が、システムへの入力として提供され得る。分子の大規模なセットもまた提供され得る。各分子に対して、タンパク質標的に対する結合親和性が予測される。結果として生じるスコアは、分子をランク付けするために使用され得、最高得点の分子は、標的タンパク質を結合させる可能性が最も高い。任意選択的に、ランク付けされた分子リストは、類似した分子のクラスタについて分析され得るか、大きいクラスタは、分子結合のより強力な予測として使用され得るか、又は分子は、確認実験において多様性を確実にするために、クラスタにわたって選択され得る。

30

【0219】

オフ標的副作用予測。多くの薬物は副作用を有することが発覚することがある。これらの副作用は、薬物の治療効果を担う経路以外の生物学的経路との相互作用に起因する場合が多い。これらのオフ標的副作用は、不快又は危険であり得、薬物の使用が安全である患者集団を制限する場合がある。したがって、オフ標的副作用は、どの薬物候補を更に開発するかを評価するための重要な基準である。多くの代替の生物学的標的との薬物の相互作用を特徴付けることが重要であるが、そのような試験は、開発及び実行するのに高価かつ時間がかかる場合がある。コンピュータによる予測は、このようなプロセスをより効率的にすることができる。

【0220】

本発明の実施形態を適用する際に、重要な生物学的応答及び／又は副作用に関連付けられている生物学的標的のパネルが構築され得る。次いで、システムは、パネル内の各タンパク質に対する結合を順番に予測するように構成され得る。特定の標的に対する強い活性（すなわち、オフ標的タンパク質を活性化することで知られている化合物と同じくらい強力な活性）は、オフ標的効果による副作用に分子を関与させ得る。

40

【0221】

毒性予測。毒性予測は、オフ標的副作用予測の特に重要な特別な事例である。後期臨床試験において薬物候補の約半数は、許容不能な毒性のために不合格になる。新薬承認プロセスの一部として（及び薬物候補をヒトで試験できる前に）、FDAは、（その抑制が、薬物－薬物相互作用からの毒性につながる場合がある）シトクロムP450肝臓酵素又は

50

(その結合が、心室性不整脈及び他の心臓有害影響につながるQT延長につながる場合がある) hERGチャネルを含む標的のセットに対する毒性試験データを要求する。

【0222】

毒性予測では、システムは、オフ標的タンパク質を主要な抗標的(例えば、CYP450、hERG、又は5-HT_{2B}受容体)であるように制約するように構成され得る。次いで、薬物候補に対する結合親和性は、これらのタンパク質に対して予測され得る。任意選択的に、分子は、代謝物のセット(元の分子の代謝/分解中に身体によって生成された後続の分子)を予測するために分析され得、これはまた、抗標的に対する結合について分析され得る。問題がある分子は、毒性を回避するために識別及び修正され得るか、又は分子系列に関する開発は、追加の資源の浪費を回避するために中止され得る。

10

【0223】

効力の最適化。薬物候補の主要な要件のうちの1つは、その疾患標的に対する強力な結合である。スクリーニングが臨床上効果的であるのに十分に強力に結合する化合物を見つけることはまれである。したがって、初期化合物は、医薬品化学者が分子構造を繰り返し修正して、標的結合の強度を増加させた新しい分子を提案する、最適化の長いプロセスを施す。各新しい分子は、変化が無事に結合を改善したかどうかを判定するために、合成及び試験される。システムは、物理試験をコンピュータによる予測で置き換えることによって、このプロセスを容易にするように構成され得る。

【0224】

本出願では、疾患標的、及びリード分子のセットが、システムに入力され得る。システムは、リードのセット結合親和性予測を生成するように構成され得る。任意選択的に、システムは、結合親和性における予測される差の理由を知らせるのに役立ち得る候補分子間の差を強調し得る。医薬化学者ユーザは、この情報を使用して、願わくは標的に対する活性が改善された新しい分子のセットを提案することができる。これらの新しい代替分子は、同じ様式で分析され得る。

20

【0225】

選択性の最適化。上記で考察されるように、分子は、様々な強度で多くのタンパク質を結合する傾向がある。例えば、(よく知られている化学療法標的である)タンパク質キナーゼの結合ポケットは、非常に類似しており、大部分のキナーゼ阻害剤は、多くの異なるキナーゼに影響を与える。これは、様々な生物学的経路が同時に修正され、「汚れた」医薬プロファイル及び多くの副作用を生じることを意味する。したがって、多くの薬物の設計における重要な課題は、活性自体ではなく、特異性：おそらく密接に関連するタンパク質のセットから1つのタンパク質(又はタンパク質のサブセット)を選択的に標的とする能力である。

30

【0226】

システムは、候補薬物の選択性を最適化するための時間及びコストを低減することができる。本出願では、ユーザは、2つのセットのタンパク質を入力して得る。一方のセットは、化合物が活性であるべきタンパク質を記述し、他方のセットは、化合物が不活性であるべきタンパク質を記述する。システムは、両方のセット内のタンパク質の全てに対して分子の予測を行い、相互作用強度のプロファイルを確立するように構成され得る。任意選択的に、これらのプロファイル进行分析して、タンパク質における説明パターンを示唆することができる。ユーザは、システムによって生成された情報を使用して、異なるタンパク質セットへの相対的な結合を改善する分子に対する構造的変化を考慮し、より優れた特異性を有する新しい候補分子を設計することができる。任意選択的に、システムは、選択性における予測される差の理由を知らせるのに役立ち得る候補分子間の差を強調するように構成され得る。提案された候補は、それぞれの活性プロファイルの特異性を更に精緻化するために、繰り返し分析され得る。

40

【0227】

自動分子設計のための適合度関数：前述の最適化を実行するための自動化ツールは貴重である。成功した分子は、効力、選択性、及び毒性の間の最適化並びにバランスを必要と

50

する。「足場ホッピング」(リード化合物の活性が保たれるが、化学構造が大幅に変更されるとき)は、改善された薬物動態、薬力学、毒性、又は知的財産のプロファイルを生じさせることができる。分子の無作為な生成、所与の結合部位を充填するための分子フラグメントの成長、分子の集団を「突然変異」及び「異種交配」させるための遺伝的アルゴリズム、並びに分子の部分のバイオイソステリック置換との交換などのアルゴリズムは、新しい分子を繰り返し示唆するために存在する。これらの方法の各々によって生成された薬物候補は、上で説明される複数の目的(効力、選択性、毒性)に対して評価されなければならない。技術が前述の手動設定(結合予測、選択性、副作用、及び毒性予測)の各々について有益であり得るのと同じように、それは、自動分子設計システムに組み込まれ得る。

【0228】

10

薬物の別の目的での使用。全ての薬物は副作用を有し、ときおり、これらの副作用が有益である。最もよく知られている例は、一般的に頭痛治療として使用されるが、心臓血管の健康のためにも使用されるアスピリンであり得る。薬物リポジショニングは、薬物がヒトで安全であることが既に示されており、患者における迅速な吸収及び好ましい安定性について最適化されているため、創薬のコスト、時間、及びリスクを大幅に低減することができる。残念ながら、薬物リポジショニングは、大部分は偶然である。例えば、シルデナフィル(バイアグラ)は、血圧降下剤として開発され、勃起不全に対する効果的な治療であることが思いがけず観察された。オフ標的効果のコンピュータによる予測は、代替疾患を治療するために使用され得る化合物を識別するための、薬物の別の目的での使用の文脈で使用され得る。

20

【0229】

本出願では、オフ標的副作用予測と同様に、ユーザは、考えられる標的タンパク質のセットを組み立て得、各タンパク質は、疾患に結び付けられる。すなわち、各タンパク質の抑制は、(おそらく異なる)疾患を治療するであろう。例えば、シクロオキシゲナーゼ-2の阻害剤は、炎症を緩和することができるが、一方で、Xa因子の阻害剤は、抗凝固剤として使用することができる。これらのタンパク質は、存在する場合、承認された薬物の結合親和性で注釈を付けられる。次いで、分子のセットを組み立て、分子のセットをヒトでの使用のために承認又は調査された分子に限定する。最後に、タンパク質及び分子の各対に対して、ユーザは、システムを使用して、結合親和性を予測し得る。薬物の別の目的での使用のための候補は、分子の予測される結合親和性が、タンパク質にとって効果的な薬物の結合親和性に近い場合に識別され得る。

30

【0230】

薬物耐性予測。薬物耐性は、病原体集団を急速に分裂させ突然変異させることに対して選択圧をかける、薬剤使用の不可避の結果である。薬物耐性は、ウイルス(HIV)、外因性微生物(MRSA)、及び調節不全宿主細胞(癌)などの多様な病原体で見られる。経時的に、薬が抗生物質であるか化学療法であるかに関わらず、所与の薬は効果がなくなる。その点で、介入は、願わくは、まだ強力な異なる薬に移行することができる。HIVでは、患者が治療されている間に、ウイルスが蓄積する変異によって定義される、周知の疾患の進行経路がある。

【0231】

40

病原体が医学的介入にどのように適応するかを予測することに相当な関心が寄せられている。1つのアプローチは、治療中にどの変異が病原体で発生するのかを特徴付けることである。具体的には、薬のタンパク質標的は、その天然基質を結合し続けながら、同時に薬物を結合することを回避するように変異する必要がある。

【0232】

本出願では、標的タンパク質の考えられる変異のセットが提案され得る。各変異に対して、結果として生じるタンパク質の形状を予測し得る。これらの変異タンパク質形態のそれぞれに対して、システムは、天然基質及び薬物の両方に対する結合親和性を予測するように構成され得る。タンパク質にもはや薬物に結合させないが、天然基質に対する結合を続行させる変異は、薬物耐性を付与するための候補である。これらの変異したタンパク質

50

は、例えば、これらのタンパク質をこれらの他の予測使用事例のうちの1つへの入力として使用することによって、薬物を設計するための標的として使用され得る。

【0233】

オーダーメイド医療。効果のない薬は投与されるべきではない。コストと手間に加えて、全ての薬は副作用を有する。道徳的及び経済的な考慮により、利益がこれらの害を上回るときにだけ薬を与えることが不可欠になる。薬がいつ有用であるかを予測できることが重要な場合がある。人は、少数の変異によって互いに異なる。しかしながら、小さい変異は、計り知れない影響を及ぼす可能性がある。これらの変異が疾患標的の活性（オルソステリック）又は調節（アロステリック）部位で発生すると、それらは薬物が結合するのを防止し、したがって、薬の活性を阻害する場合がある。特定の人のタンパク質構造が周知である（又は予測されている）とき、システムは、薬物が効果的であるかどうかを予測するように構成され得、又はシステムは、薬物がいつ機能しないかを予測するように構成され得る。

10

【0234】

本出願では、システムは、薬物の化学構造及び特定の患者の特定の発現タンパク質を入力として受け取るように構成され得る。システムは、薬物とタンパク質との間の結合を予測するように構成され得、薬物の予測された結合親和性が、特定の患者のタンパク質構造が弱すぎて臨床上効果的ではない場合、臨床医又は開業医は、その薬物が患者に無益に処方されるのを防止し得る。

【0235】

治験設計。本出願は、上記のオーダーメイド医療の使用事例を患者集団の事例に一般化する。システムが、薬物が特定の患者の表現型に効果的であるかどうかを予測できるとき、この情報を使用して、臨床試験を設計するのに役立てることができる。特定の疾患標的が薬物によって十分に影響を受けることがない患者を除外することによって、臨床試験は、より少ない患者を使用して、統計的検出力を達成することができる。より少ない患者は、臨床試験のコスト及び複雑さを直接的に低減する。

20

【0236】

本出願では、ユーザは、（例えば、変異又はイソ型に起因する）異なるタンパク質の発現によって特徴付けられる亜集団に、考えられる患者集団を分割し得る。システムは、異なるタンパク質タイプに対する薬物候補の結合強度を予測するように構成され得る。特定のタンパク質タイプに対する予測される結合強度が、（例えば、試験管、動物モデル、又は健常ボランティアにおける物理的特徴付けに基づくように）臨床上達成可能な患者内濃度を下回る必要な薬物濃度を示す場合、薬物候補は、そのタンパク質亜集団に対して不合格になると予測される。次いで、そのタンパク質を有する患者は、治験から除外され得る。

30

【0237】

農薬設計。医薬用途に加えて、農薬産業は、新しい殺虫剤の設計で結合予測を使用する。例えば、殺虫剤のための1つの必要なものは、任意の他の種に悪影響を与えることなく、関心対象の単種を止めることである。環境安全性のために、人は、マルハナバチを殺さずにゾウムシを殺すことを望むだろう。

40

【0238】

本出願では、ユーザは、検討中の異なる種からタンパク質構造のセットをシステムに入力することができる。タンパク質のサブセットは、それに対して活性であるタンパク質として指定することができる。一方、残りは、それに対して分子が不活性であるべきタンパク質として指定することができる。前の使用事例と同様に、（既存のデータベース内か、新たに生成されたかに関わりなく）分子の何らかのセットが各標的に対して考慮され、システムは、タンパク質の第2のグループを回避しながら、第1のグループに対して最大の効果を有する分子を返すだろう。

【0239】

材料科学。新しい材料の挙動及び特性を予測するために、分子相互作用を分析すること

50

が有用であり得る。例えば、溶媒化を研究するために、ユーザは、所与の小分子の反復結晶構造を入力し得、結晶の表面上の小分子の別の例の結合親和性を評価し得る。ポリマー強度を研究するために、ポリマーストランドのセットは、タンパク質標的構造に類似して入力され得、ポリマーのオリゴマーは、小分子として入力され得る。したがって、ポリマーストランド間の結合親和性は、システムによって予測され得る。

【0240】

シミュレーション。シミュレータは、タンパク質の領域に留まる分子の傾向がタンパク質の結合親和性と相関するため、タンパク質に対する分子の結合親和性を測定することが多い。結合を支配する特徴の正確な記述を使用して、特に高い又は低い結合エネルギーを有する領域及びポーズを識別することができる。エネルギーの記述は、分子の運動及びタンパク質結合領域の占有を説明するために、モンテカルロシミュレーションに折り畳むことができる。同様に、システム生物学を研究し、モデル化するための確率論的シミュレータは、分子濃度のわずかな変化が生物学的ネットワークにどのように影響を与えるかの正確な予測から利益を得ることができる。

【実施例】

【0241】

AtomNet（登録商標）Carbon：物理学及び幾何学を学習することにより、構造ベースの仮想高スループットスクリーニングアーキテクチャにポーズ感受性が付与される。

【0242】

分子生物活性は、集合体の特性であり、受容体化合物複合体形成のエンタルピー及びエントロピー成分によって判定される。構造ベースの深層学習方法は、活性予測には成功しているが、ドッキングされたポーズに対し非感受性であり得、ヒット検出の信頼性を低下させる。更に、構造ベースの深層学習方法は、自由エネルギーの変化へのエントロピー寄与を無視することが多い。集合体がポーズに対して感受性がある場合、集合体アプローチは成功する。この実施例では、ドッキングされたポーズに対する感受性を増加させた深層学習マルチタスクアーキテクチャについて説明する。

【0243】

1 序文

ENAMINE又はMoleculeのような膨大なオンデマンド化学ライブラリは、薬剤、構造ベースの仮想高スループットスクリーニング（vHTS）キャンペーンの規模を変革した[1]。候補分子のライブラリから「ヒット」を識別するために、構造ベースの仮想スクリーニング方法は、タンパク質と配位子との間の結合親和性を、それぞれのドッキングされた結合複合体から予測し、それによって、実験的に観察された親和性が、タンパク質-配位子相互作用と相関すると仮定する。従来の方法は、複合体形成の結合自由エネルギーを計算しようとする、経験的な物理学ベースのアプローチを使用する。それに反して、機械学習（ML）及び深層学習（DL）アプローチは、活性を予測するために明示的（ML）又は暗黙的（DL）特徴及び標識を使用して、大規模なデータセット上で訓練される。これらの統計モデルは、一般的に、活性を予測するための遡及試験において物理学ベースのアプローチよりも優れている。

【0244】

重畳型ニューラルネットワーク（CNN）を中心としたvHTSの初期の構造ベースのDL方法は、3Dグリッドによってタンパク質-配位子構造を表して活動を予測する[2～5]。概して効果的であるが[6]、CNNの欠点は、それらが回転不変ではなく、代替表現よりも多くのパラメータを必要とすることである。その結果として、グラフ重畳型ネットワーク[7]、又はより一般的には、メッセージパッシングニューラルネットワーク[8～10]は人気を得ている。最近の研究では、構造ベースの機械学習方法の性能は、タンパク質化学的な特徴によって部分的に促進されることが示唆されている[11、12、5]。配位子と結合部位との間の特異性相互作用に応答するのではなく、モデルは、一般的な配位子-タンパク質の特徴的な性質を学習する。この欠陥は、モデルが、同じタ

ンパク質上のこれまでに見られなかった結合部位に直面したとき、特に、その部位が正規の部位と部分的に重複するときに、予測性能の低下によって現れる。例えば、モデルは、キナーゼ上のアロステリック部位のATP競合結合剤を上位にランク付けし得る。このような限定は、新しい化学物質の発見、又はタンパク質上の新規部位を標的とする能力を大きく阻害する。

【0245】

配位子のポーズの品質及び親和性に関する同時訓練は、ポーズ感受性を改善することができる[13]。ここでは、その観察に基づいて、生物活性、及びポーズの物理学ベースのvina[14]スコアを同時に評価する、生物活性予測のためのマルチタスクアーキテクチャを構築及び提示する。更に、その生体活性タスクをポーズの品質にコンディショニングする。最後に、生体活性分子の生体活性標識を無効にしながら、モデルを悪いポーズに曝露し、したがって、負の例として、真の結合剤の悪いポーズを提示する。当社のアーキテクチャが、いくつかの厳密なベンチマークでポーズ感受性を改善することを実証する。

10

【0246】

2.1 ニューラルネットワークアーキテクチャ

この実施例のシステムは、位置に依存するエッジを有するグラフニューラルネットワークベースのアーキテクチャである。これは、本開示の重畳型ニューラルネットワーク24の実施例である。この実施例では、任意の配位子原子の7Å以内の受容体原子のみを考慮する。(配位子及び受容体)原子が、互いに4Å以内にある場合に隣接する、2つのグラフ畳み込み層を使用する。次いで、配位子のみの特徴を抽出し、更に2つの配位子のみの層を更に付け加える。この配位子のみの層は、合計プーリング層を使用してプールされる。次いで、プールされた特徴は、ネットワークの最上部にあるマルチタスク多層パーセプトロン(第1のモデル72、第2のモデル74など)の埋設として使用される。グラフニューラルネットワークによって生成された埋設は、この実施例では、活性、PoseRankerポーズ品質スコア、及びVinaドッキングスコアの3つの出力を予測するために使用される。これは、2段階で実行される。最初に、PoseRanker及びVinaスコア予測は、埋設を2つの独立した多層パーセプトロンを通過させることによって算出される。次いで、コンディショニングされた埋設は、入力埋設をPoseRankerスコア予測と連結することによって形成され、活性予測を算出するために第3の多層パーセプトロンに渡される[15]。4.3節は、モデル訓練パラメータの詳細を提供した。

20

30

【0247】

2.2 データ

訓練データセットは、ChEMBL又はPubChemのような一般提供されている供給源と、Reaxys又はLigandExptのような商用データベースと、から収集された結合親和性の測定値からなっていた。この実施例では、pKi2(0;11)での定量的測定値のみを考慮した。測定したpKi(又はIC50)が10µM未満である場合、化合物は、活性として標識され、そうでない場合、化合物は不活性として標識された。測定された活性化合物の数は不活性のものよりも多いため、活性化合物の各々を、別の異なるタンパク質標的のためのデコイとして無作為に割り当てることによって、訓練データセットを増強した。加えて、いくつかのモデルでは、ポーズネガティブ(ポーズが悪く、不活性として標識された活性化合物)を使用した。詳細については、4.2節を参照されたい。訓練から12個の多様なタンパク質(D12)のセットを除外し、それは、ホールドアウト試験セットとして機能した。加えて、訓練セットから、D12セットのタンパク質の全ての近接相同体を除外した(95%未満の配列と同様に)。訓練セットは、3800個以上の多様なタンパク質を網羅し、ポーズネガティブなしで4.8M(ある場合は5.8M)のデータポイントを考慮する。ホールドアウトセットは、12個のタンパク質にわたって分布する、約33000個の化合物を考慮する全ての化合物は、開示アーキテクチャであるC vina[16]とドッキングされ、(PoseRankerモデル[10]によってランク付けされたように)利用可能な最良のポーズが、DLモデルとのスコアリ

40

50

ングに使用された。

【0248】

2.3 数値実験

当社のモデルのポーズ感受性を研究するために、D12における活性標的ー化合物対の各々を、以下を用いて3回スコアリングした：i) トップポーズ、ii) 悪いポーズ、iii) 配位子の質量中心の周りの配位子の（4回繰り返された）無作為な回転によってトップポーズから取得された、物理的に非現実的なポーズ。良いポーズは、PoseRankerによる最高ランクのポーズだった（ポーズはCulinaで生成された）。悪いポーズは、PoseRankerによる最悪ランクのポーズだった。非現実的なポーズは、質量中心の周りの配位子の無作為な回転によって、良いポーズから取得された。全てのポーズをスコアリングし、次いで、悪いポーズ及び非現実的なポーズのスコアを、良いポーズのスコアから減算した。ポーズ感受性の尺度は、良いポーズと悪い／非現実的なポーズとの間の活性スコアの低下の中央値である。重畳型ニューラルネットワークは、入力データの知覚フィールドにおける特徴を検出することができる。そのフィールドが十分に大きく複雑である場合、モデルは、保存されている結合部位、例えば、タンパク質キナーゼにおけるATP結合部位に特徴的である原子の集まりを検出することができる。しかしながら、知覚フィールドの範囲を制限することは、例えば、プールすることによって、検出された特徴間の空間情報を省略する。結果として、モデルは、入力データに提供された化学的に無関係な特徴、いわゆるピカソ問題を検出することによって、バイアスされ得る。隣接結合部位が、モデルの推論にどのように干渉するかを監視するために、周知のキナーゼ阻害剤の多様なセット（活性として標識された約300個の多様な化合物）を選択し、それらを、利用可能なスクリーニングライブラリ（MCULE、2017／18／10時点、不活性として標識）から無作為に選択された10⁵個の化合物と混合した。各化合物を、ATP結合部位、及びATP部位から6～10 Å離れたアロステリック部位にドッキングさせた。モデルの潜在的なバイアスを監視するために、全ての化合物もまた、（ATP結合部位から50 Å超離れた）遠隔のSH2ドメイン上に位置する暫定的な結合部位にドッキングさせた（図20）。期待されたのは、良好な性能を有するモデルは、ATP結合部位にドッキングしたときに、キナーゼ阻害剤をバックグラウンドの無作為分子から適切に区別することができるということであった（0.5よりもはるかに高いROC AUCを期待する）。その一方で、ポーズ感受性モデルは、化合物がアロステリック部位（0.5に近いROC AUC）にドッキングされるとき、隣接するATP部位によってバイアスされるべきではない（図20）。訓練セット内の任意の可能なバイアスを説明するために、SH2ドメインに位置する空間的に遠い結合部位のROC AUCを計算し、図20の青色の点を示した。

10

20

30

40

50

【0249】

3 結果

図21の結果は、この実施例で研究されたモデルが、ホールドアウトセットで良好な性能を有し、GCNは、CNNよりもわずかに優れていることを示している。しかしながら、単一タスクモデルの両方を、ヒトZAP70タンパク質のアロステリック部位の仮想スクリーニングで使った場合、それらの両方は、周知のATP部位キナーゼ阻害剤を改良するであろう。これは、モデルが、配位子ー受容体相互作用の特徴を学習せず、代わりに配位子及び受容体の独立した表現を学習するためである。次いで、これらの学習された表現／埋設は、モデルの推論に使用される。ATP結合部位は、これら2つのネットワークの知覚フィールドにあるため、GCN及びCNNモデルは、高度に保存されているATP結合部位の特徴を識別し得（図20、図22）、モデルが、あまり一般的ではないアロステリック部位の代わりに、ATP部位について尋ねられたかのように予測を行うことができる（図22）。この結果は、原発部位（ATP部位）から空間的に離れている結合部位のスクリーニングが、キナーゼ阻害剤（SH2部位、図20、図22）の改良をもたらさなかったため、バイアス訓練セットによって説明することができない。これは、これら2つのモデル、CNN及びGCNが、タンパク質化学的性質（配位子及び受容体表現が使用

されるが、一方が他方から独立している）であることを示唆する。これは、結合部位における配位子の誤配列（悪いポーズ）だけでなく（図23の左パネル）、配位子-受容体界面の破壊（図23の右パネル）にも非感受性であることによって、更に裏付けられる。この奇妙な挙動は、3DグリッドベースのCNN[4, 13]に関する以前の研究でも観察されているが、一般に適用可能な解決策は提案されていない。PCMモデルの主な欠点は、推論に使用されるポーズに対する生来の非感受性である。したがって、ピカソ問題の解決策は、モデルがポーズ感受性であることを確認することである。この実施例では、ポーズ感受性とみなされるモデルの最小要件は、i) 物理的に非現実的なポーズ（例えば、複数の原子同士の重複）が物理的に非現実的な特徴のないポーズと比較して、ペナルティを課されること、ii) 結合ポケットの外側にある配位子を有するポーズは、結合部位にある配位子を有するポーズよりもペナルティを課されるべきであること、及びiii) 標的部位の近傍にある結合部位は、予測に干渉してはならないことである。

10

【0250】

最初、構造データで訓練された単一タスク（活性）モデルが、配位子-受容体相互作用に関するその構造情報を使用しないことは直感的ではない。しかしながら、訓練中、主な目的は、指定された損失関数を最小化することであり、配位子-受容体相互作用の使用が、このタスクにおいてモデルにエッジを与えることができるという仮定であるため、これは事実であり得る。実際には、インシリコ生成ポーズは、誤差及び不確実性の影響を受け、それらに過度に依存すると、モデルの性能が損なわれる可能性がある。モデルは、配位子-受容体相互作用の構造的な特徴を学習する動機を持たないため、モデルはそれらを見ることが多い。したがって、追加のタスクが、構造感受性である埋設を必要とするマルチタスクモデルを訓練することは、理論的には、問題を軽減するはずである。これは、図20及び図22のMTモデルについて見るように、実際にそうである。別のタスク（モデルMT-1）としてドッキングスコア回帰を追加することは、明らかに不適切なポーズ（非現実的なポーズ）にペナルティを課すモデルに既につながり、hZAP70タンパク質のアロステリック部位のトップヒットで改良されたキナーゼ阻害剤の量を減少させることが分かる（図22）。悪いポーズは誤配列されているが、原子衝突はないため、MT-1モデルは、依然としてスクリーニングで使用されている良いポーズと悪いポーズとを区別することができない（図23の左パネル）。興味深いことに、この問題は、ポーズ品質回帰を第3のタスクとして追加することによって、又は活性タスクをポーズ品質、モデルMT-2及びモデルMT-3をコンディショニングすることによってだけでは解決することができない（図23）。これは、モデルが良いポーズのみを見せられ、悪いポーズがどのように見えるかのイメージを学習することができないからである。

20

30

【0251】

欠けている情報を補うために、ポーズネガティブと呼ばれるデータ増大技術が使用される。ポーズネガティブは、もともと正のデータポイントとして標識され、利用可能な最良のポーズとともに使用された例である。しかしながら、（この場合はPoseRankerスコアである任意のメトリックに従って）利用可能な最悪のポーズを選択し、標識が変更されたモデルにネガティブな例として提示することができる。このアプローチで、モデル（MT-4a及びMT-4b）が、物理的に非現実的なポーズと悪いポーズの両方にペナルティを課することができることが観察された（図23）。更に、同じモデルはまた、ピカソ問題も緩和した。しかしながら、この場合、ポーズ品質に対する活性のコンディショニングの欠如が、ピカソ問題をより起こしやすいモデルにつながることを観察された（図22）。

40

【0252】

4 結論

マルチタスクアーキテクチャは、化合物の生物学的活性を予測することができ、また、推論のために提供された構造データを十分に利用することができるモデルをもたらす。モデルに直交タスクを強制的に学習させ、最終モデルを正規化する。提案された解決策は、3Dグリッドベースのモデル及びグラフベースのモデルの両方に、一般的に適用可能であ

50

る（データ図示せず）。このアプローチは、深層学習及び構造ベースの創薬の分野を、新規の結合部位及び以前は治療不可能なタンパク質に開放する。

【0253】

本業績は、初期段階の創薬に関連付けられたコスト及び開発時間を低減するための取り組みの文脈で開発された。この分野での成功は、長期的には、医薬品へのアクセスを改善し得、医療コストを低減し得る。ここに記載されている訓練データセットは、一般提供されているデータからなるため、必然的に様々な疾患及び健康状態への研究資金の配分に偏りを反映することを認識する必要がある。タンパク質結合部位にわたるモデルの一般化可能性を改善するための努力が、訓練データにおけるこのような制限を軽減するのに役立つことを期待する。

10

【0254】

4.1 条件付きマルチタスクアーキテクチャ

実際には、回帰タスク（MSE、MAE、Huber、Log-Coshなど）及び分類タスク（BCE、ヒンジ損失、二乗ヒンジ、局所損失など）に使用できる損失関数に制限はなかった。補助タスクは、i）共有された埋設 x_{em} をタスクの出力 s_i と変換すること、

【数1】

$$i: x'_{em} = f(x_{em}; s_i)$$

20

i i) 出力スコア s_j を入力埋設と連結すること、及び i i i) i) と i i)

【数2】

$$x_{em}: x'_{em} = (x_{em}; s_j)$$

30

の組み合わせによって、活性タスクへの入力をコンディショニングすることができる。図24は、入力埋設が、最初に、PoseRankerスコア(i)によってコンディショニングされ、次に、Vinaスコアが埋設(i i i)と連結されたアーキテクチャを示す。

【0255】

4.2 ポーズネガティブでのデータ増大

C U i n a ドッキングを使用して、配位子-標的対ごとに64個のポーズが生成された。次に、PoseRankerを使用してポーズをそれぞれの品質に従って並べ替え[10]、上位16個のポーズを選択した。最高ランクのポーズは、良いポーズとして訓練及びスコアリングで使用されたが、最後の(16番目の)ポーズは、ポーズネガティブとして使用され、不活性(非結合剤)とみなされた。

40

【0256】

4.3 訓練

各モデルは、10個のエポックに対して訓練された。全てのニューラルネットワークアーキテクチャについて、6個のモデルが訓練され、各々が、5/6番目のデータを訓練セットとして使用し、1/6番目のがクロスフォールド検証のために除外された。各データのクロスフォールドには、配列の類似性の70%超を共有するタンパク質のクラスタが含まれる。モデルを、学習率が $lr = 0.001$ のADAMオプティマイザを使用して訓練し、標的を、その標的に関連付けられた活性化化合物の数に比例して、置換でサンプリングした(測定された活性化化合物のない標的を訓練セットから取り除いた)。

50

【0257】

参考文献

[1] Irwin and Shoichet, 2016, "Docking Screens for Novel Ligands Conferring New Biology: Miniperspective. Journal of Medicinal Chemistry," 59 (9) : 4103-4120, May 2016. ISSN 167 0022-2623, 1520-4804. doi:0.1021/acs.jmedchem.5b02008. URL <https://pubs.acs.org/doi/10.1021/acs.jmedchem.5b02008>.

10

【0258】

[2] Wallach, Dzamba, and Heifets, 2015, "AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery," arXiv:1510.02855 [cs, q-bio, stat].

【0259】

[3] Ragoza et al, 2017, "Protein-Ligand Scoring with Convolutional Neural Networks, Journal of Chemical Information and Modeling 57 (4), pp. 942-957.

20

【0260】

[4] Stepniowska-Dziubinska et al., 2018, "Development and evaluation of a deep learning model for protein-ligand binding affinity prediction," Bioinformatics, 34 (21), pp. 3666-3674.

【0261】

[5] Boyles et al., 2019, "Learning from the ligand: using ligand-based features to improve binding affinity prediction," Bioinformatics, page btz665.

30

【0262】

[6] Hsieh et al., 2019, "Miro1 Marks Parkinson's Disease Subset and Miro1 Reducer Rescues Neuron Loss in Parkinson's Models. Cell Metabolism," 30 (6), pp. 1131-1140.

【0263】

[7] Kipf and Welling, 2017, "Semi-Supervised Classification with Graph Convolutional Networks," arXiv:1609.02907 [cs, stat], February 2017. URL <http://arxiv.org/abs/1609.02907>. arXiv:1609.02907.

40

【0264】

[8] Feinberg et al., 2018, "PotentialNet for Molecular Property Prediction," ACS Central Science 4 (11), pp. 1520-1530.

【0265】

[9] Lim et al., 2019, "Predicting Drug-Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Gr

50

aph Representation,” Journal of Chemical Information and Modeling, 59 (9), pp. 3981–3988.

[0266]

[10] Stafford et al., 2021, “Enriching Ligand Pose Quality for Dynamic Proteins in Virtual High Throughput Screens,” doi:10.33774/chemrxiv-2021-t6xkj. URL <https://chemrxiv.org/engage/chemrxiv/article-details/614b905e39ef6a1c36268003>.

10

[0267]

[11] Siegel et al., 2019, “In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening,” Journal of Chemical Information and Modeling 59 (3), pp. 947–961.

[0268]

[12] Chen et al., 2019, “Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening,” PLOS ONE 14 (8): e0220113.

20

[0269]

[13] Francoeur et al., 2020, “Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design,” Journal of Chemical Information and Modeling 60 (9), pp. 4200–4215.

[0270]

[14] Trott and Olson, 2010, “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,” Journal of Computational Chemistry 31 (2) pp. 455–461.

30

[0271]

[15] Long et al., 2018, “Conditional Adversarial Domain Adaptation,” arXiv:1705.10667 [cs], December 2018.

[0272]

[16] Morrison et al., 2020, “CUina: An Efficient GPU Implementation of AutoDock Vina,” August 2020. URL <https://blog.atomwise.com/efficient-gpu-implementation-of-autodock-vina>.

40

[0273]

結論

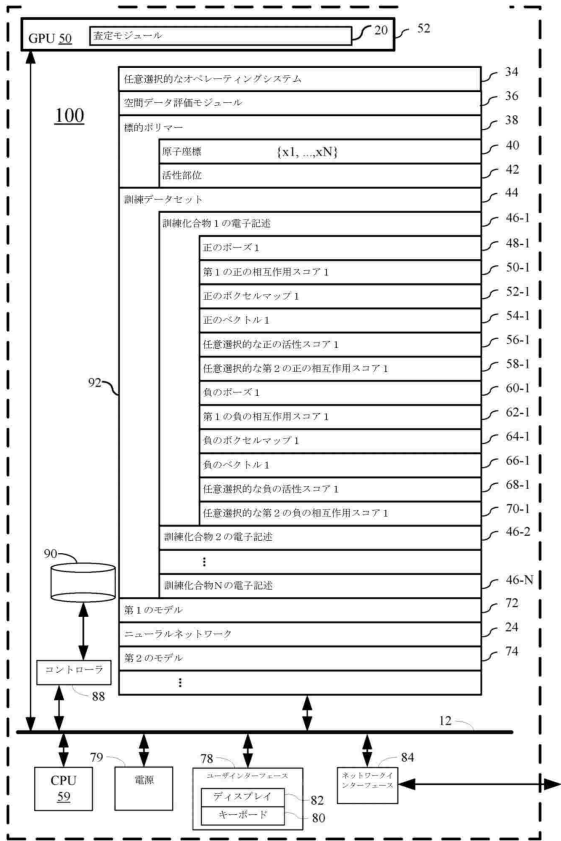
説明の目的において、前述の説明は、特定の実装を参照して説明されている。しかしながら、以上の例示的な考察は、網羅的であることを意図するものではなく、実施を開示される正確な形態に限定することを意図するものでもない。上記の教示を考慮して、多くの

50

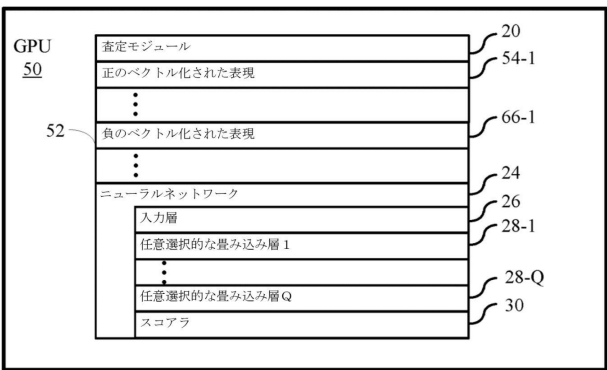
修正及び変形が可能である。実装は、原理及びそれらの実際の用途を最もよく説明するために選択及び説明され、それによって、当業者が、企図される特定の使用に適した実装及び様々な修正を伴う様々な実装を最もよく利用できるようにした。

【図面】

【図 1 A】



【図 1 B】



10

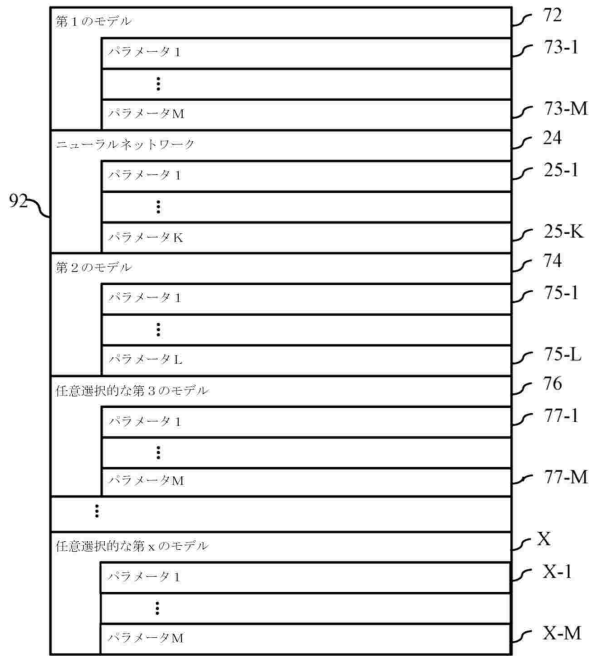
20

30

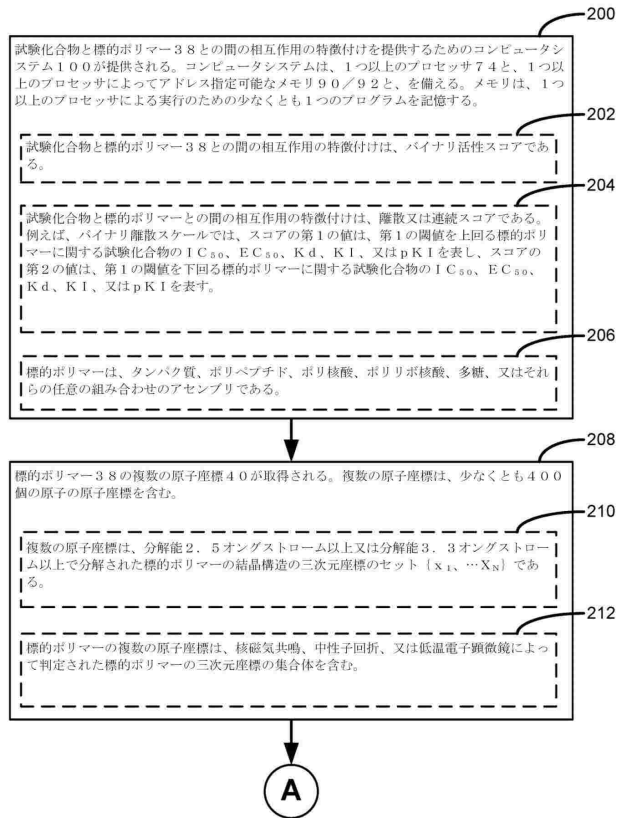
40

50

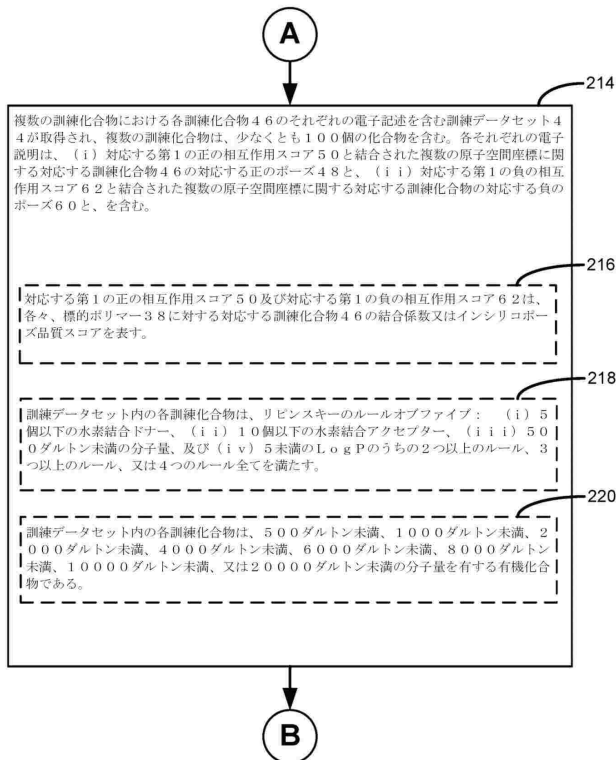
【図 1 C】



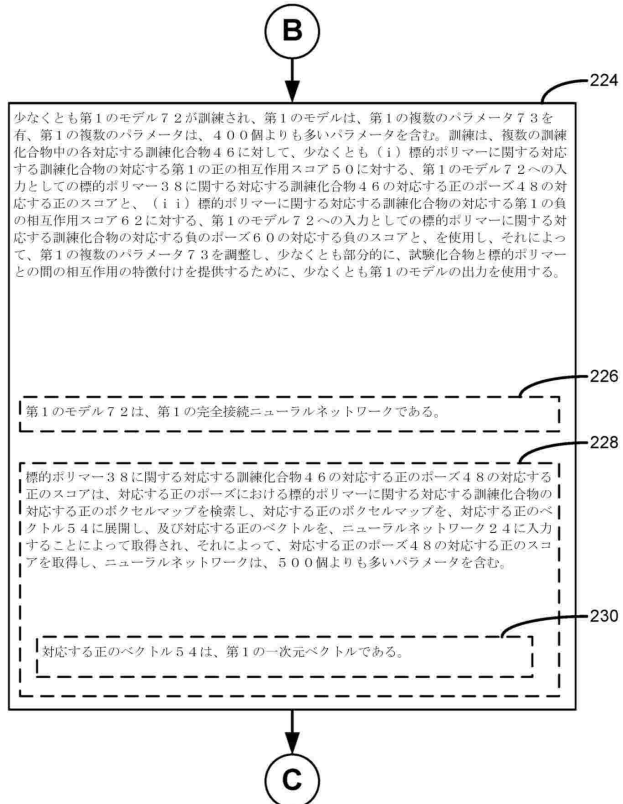
【図 2 A】



【図 2 B】



【図 2 C】



10

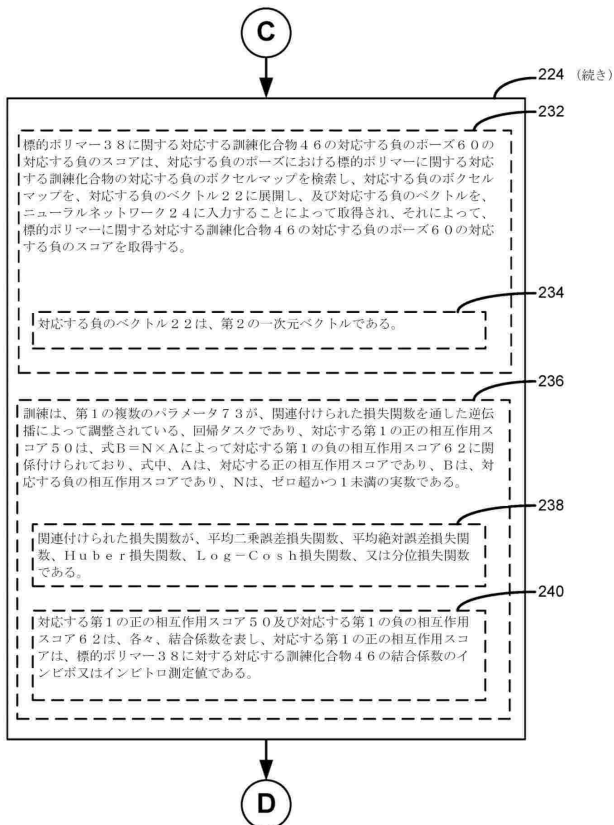
20

30

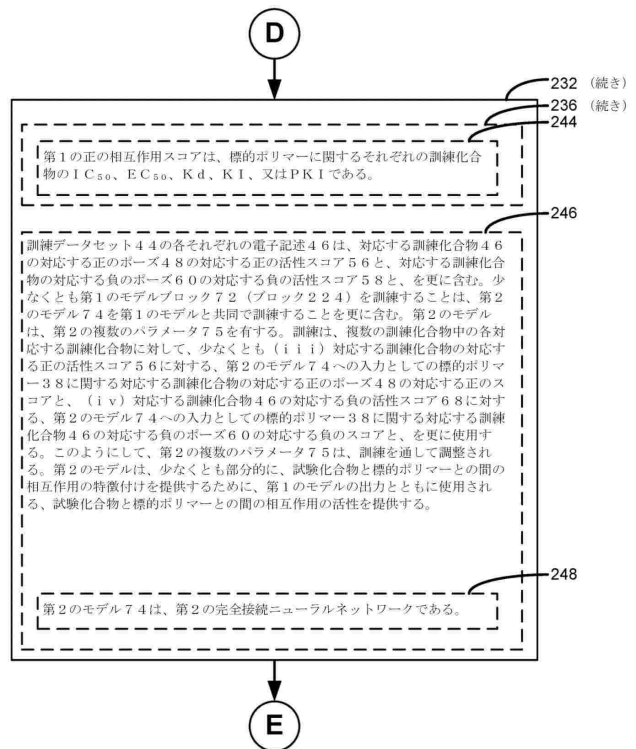
40

50

【図 2 D】



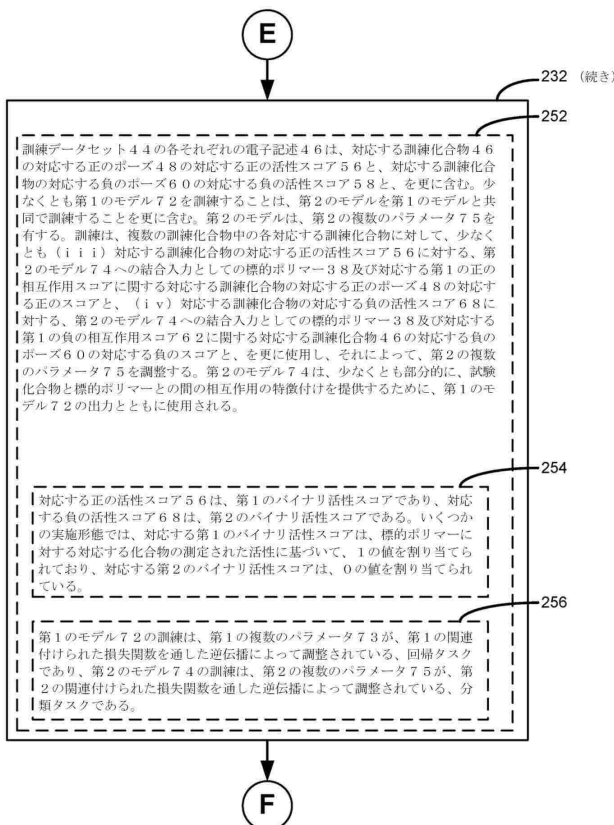
【図 2 E】



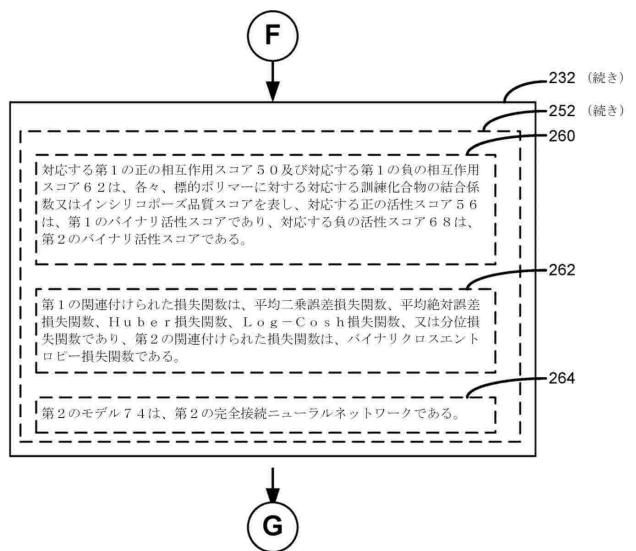
10

20

【図 2 F】



【図 2 G】

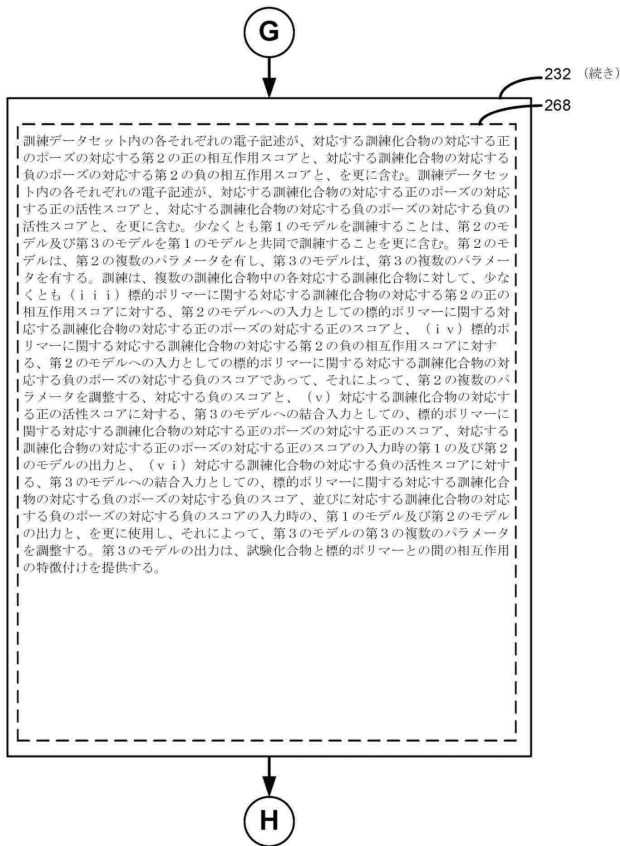


30

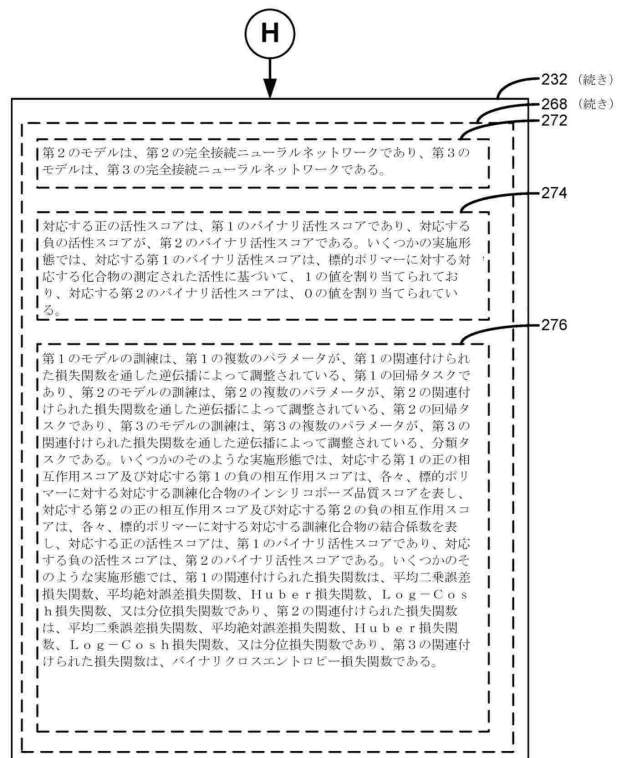
40

50

【図 2 H】



【図 2 I】



【図 3】

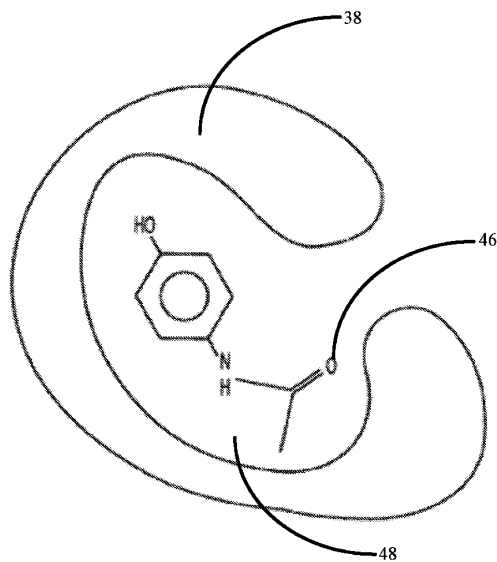


Fig. 3

【図 4】

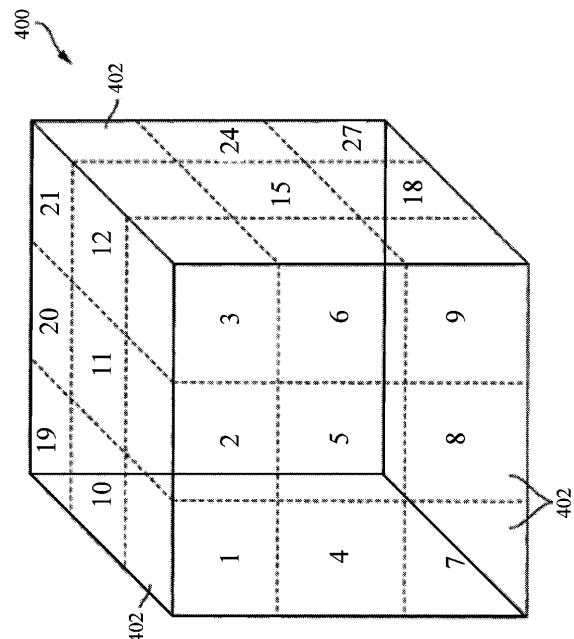


Fig. 4

10

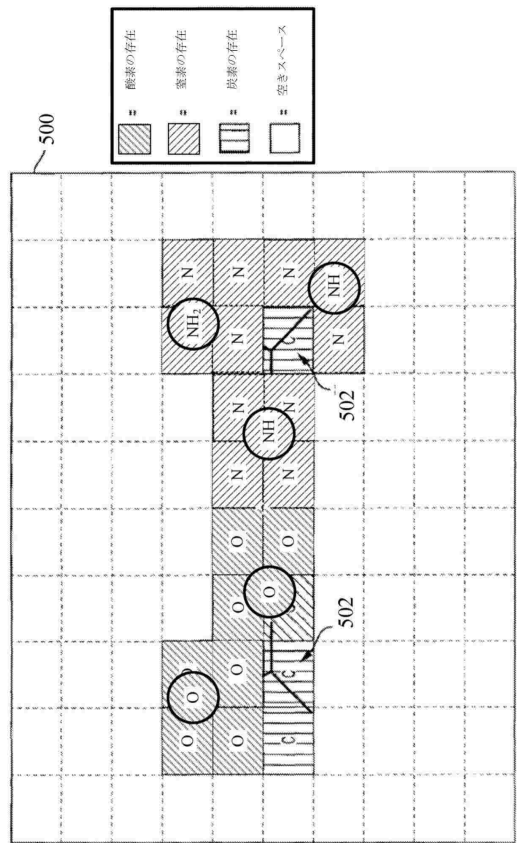
20

30

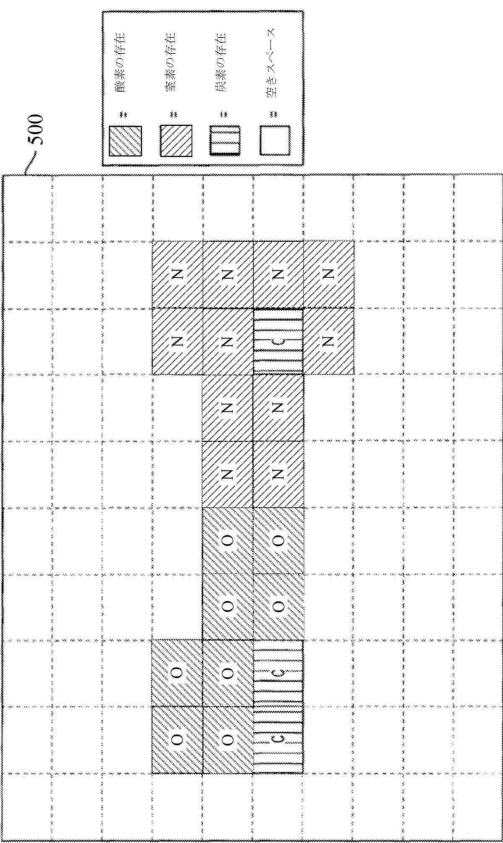
40

50

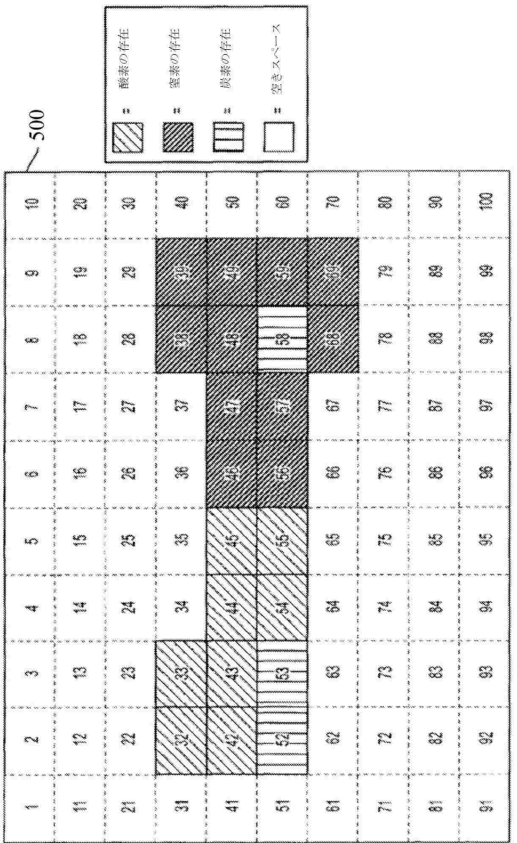
【図 5】



【図 6】



【図 7】



【図 8】

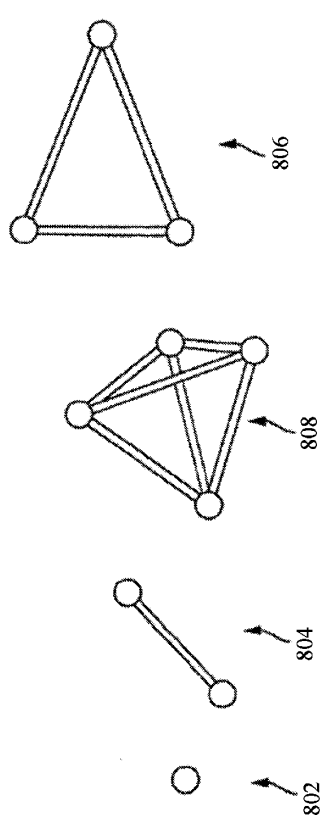


Fig. 8

10

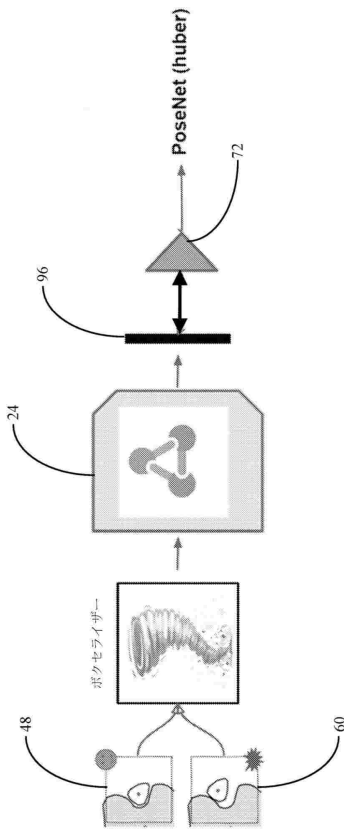
20

30

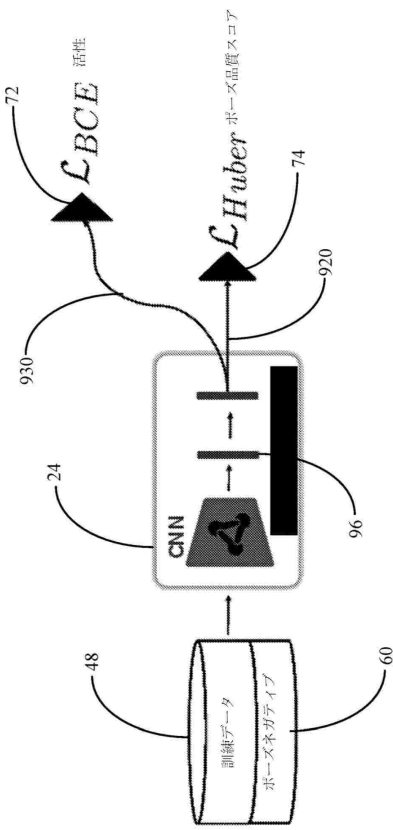
40

50

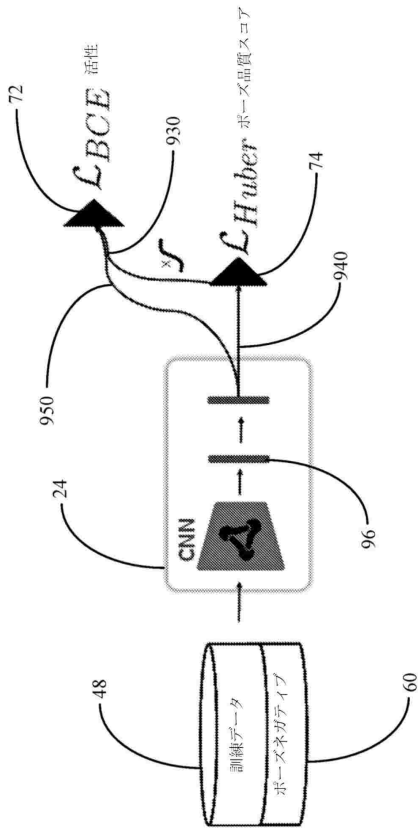
【図 9 A】



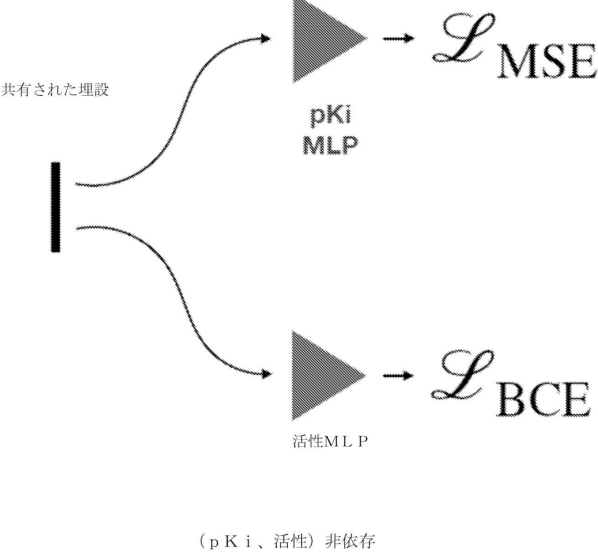
【図 9 B】



【図 9 C】



【図 10】



10

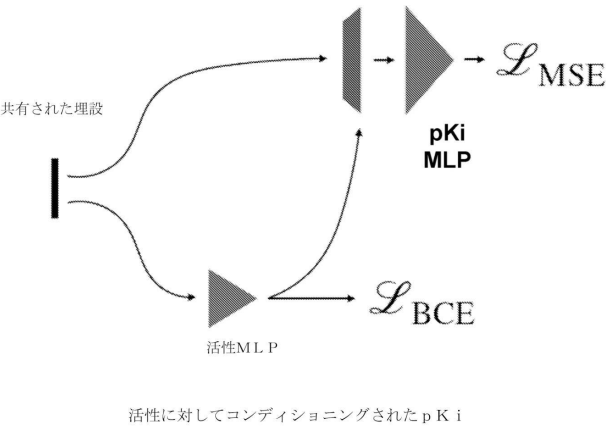
20

30

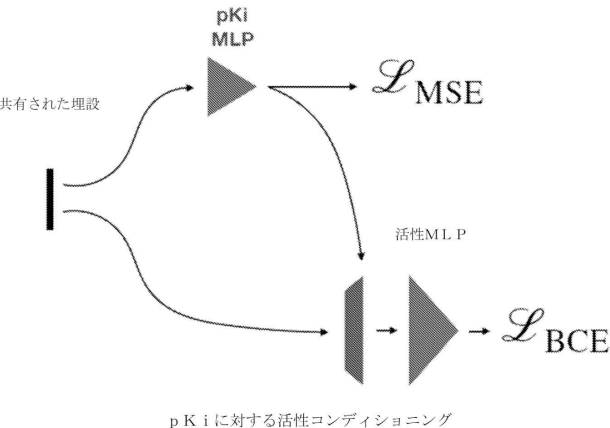
40

50

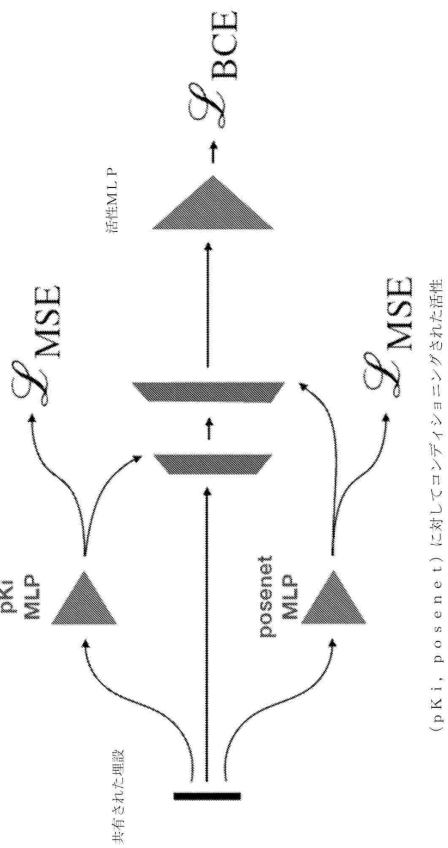
【図 1 1】



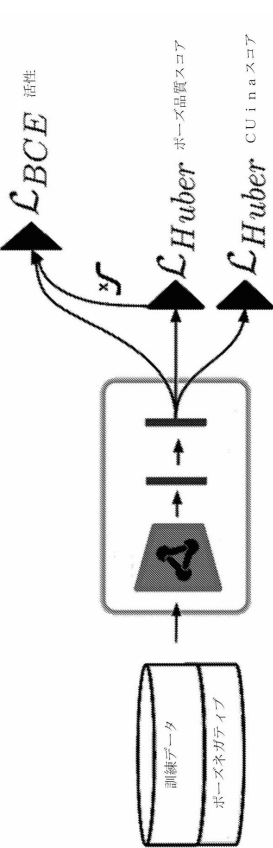
【図 1 2】



【図 1 3】



【図 1 4】



10

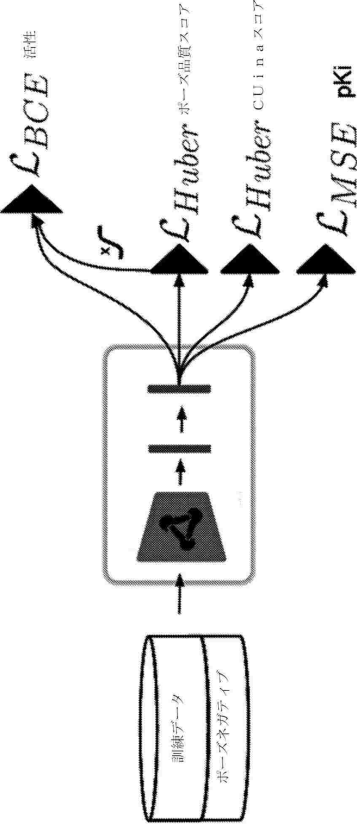
20

30

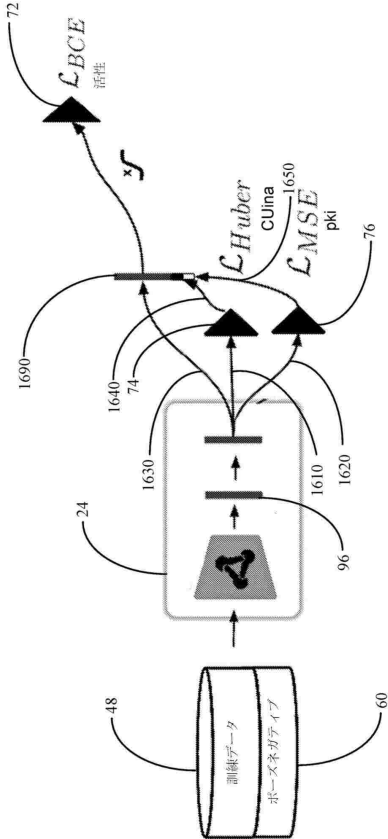
40

50

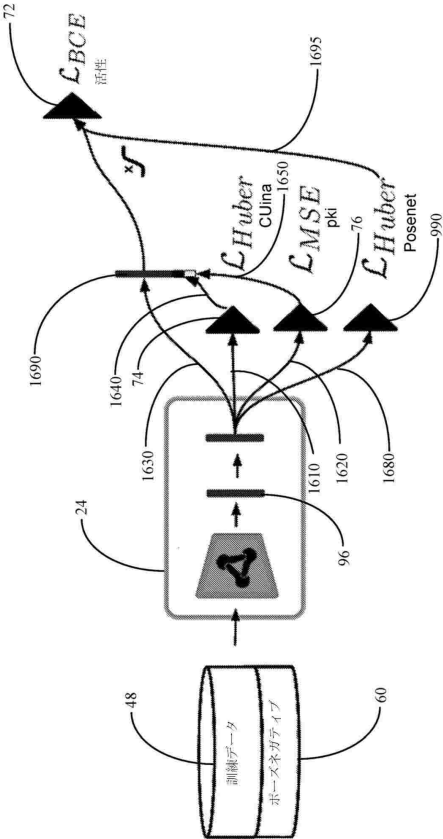
【図 15】



【図 16 A】



【図 16 B】



【図 17】

$$f(x_1, x_2, \dots, x_{100}) = g(g_0(x_1, x_2, \dots, x_{100}), g_1(x_1, x_2, \dots, x_{100}), \dots, g_n(x_1, x_2, \dots, x_{100}))$$
$$x_1 = (x_1^1, x_1^2, \dots, x_1^{40})$$
$$x_2 = (x_2^1, x_2^2, \dots, x_2^{40})$$

Fig. 17

10

20

30

40

50

【図 18】

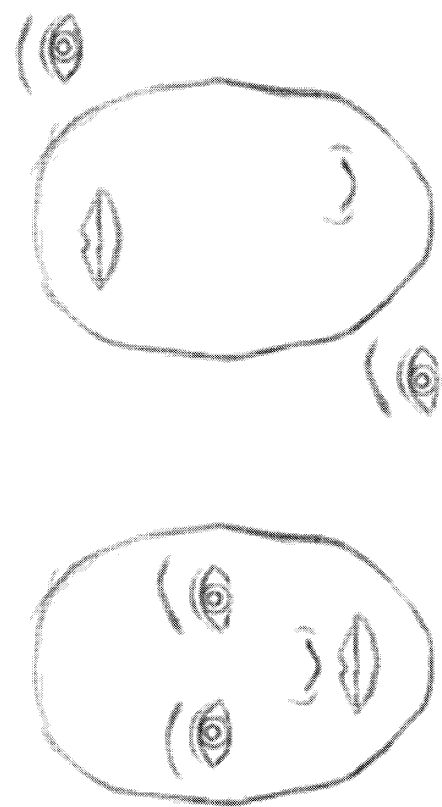
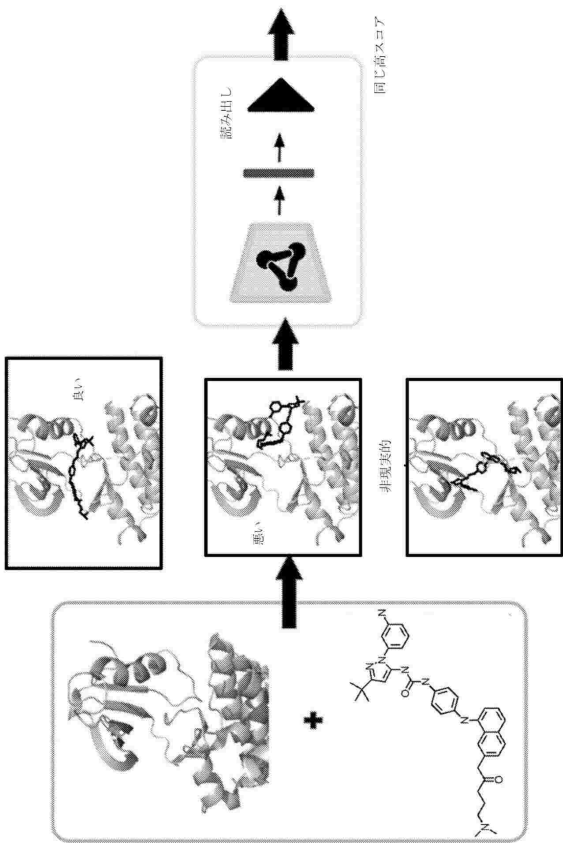


Fig. 18
(Prior Art)

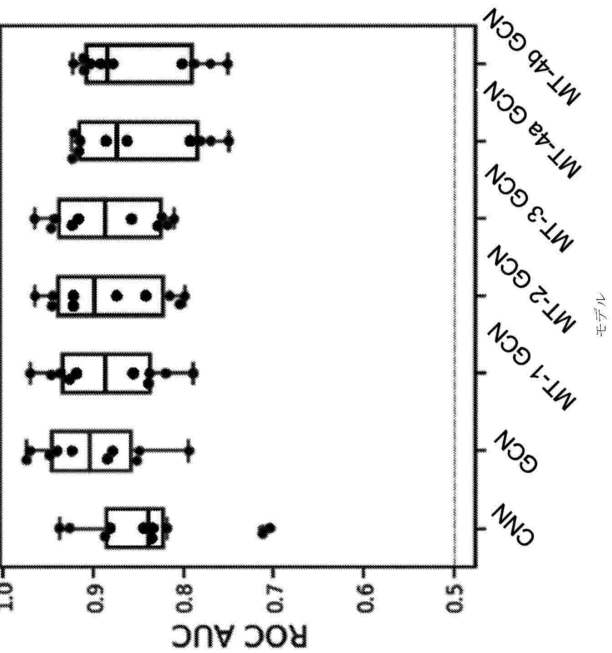
【図 19】



【図 20】



【図 21】



10

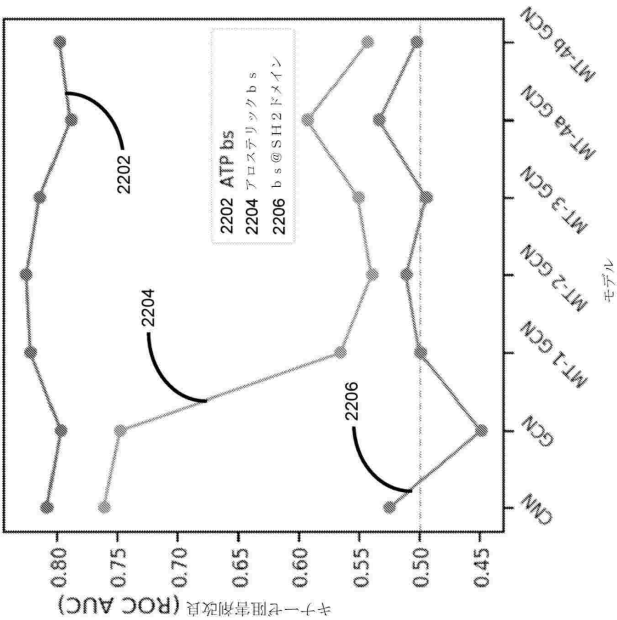
20

30

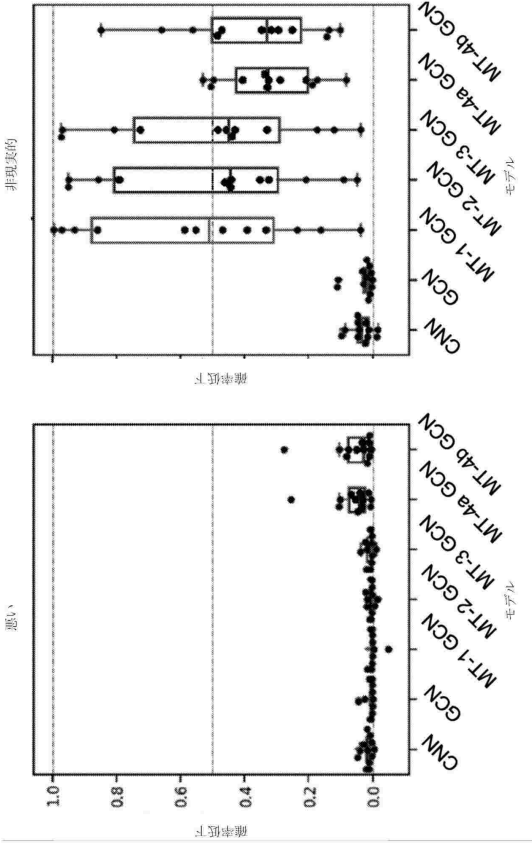
40

50

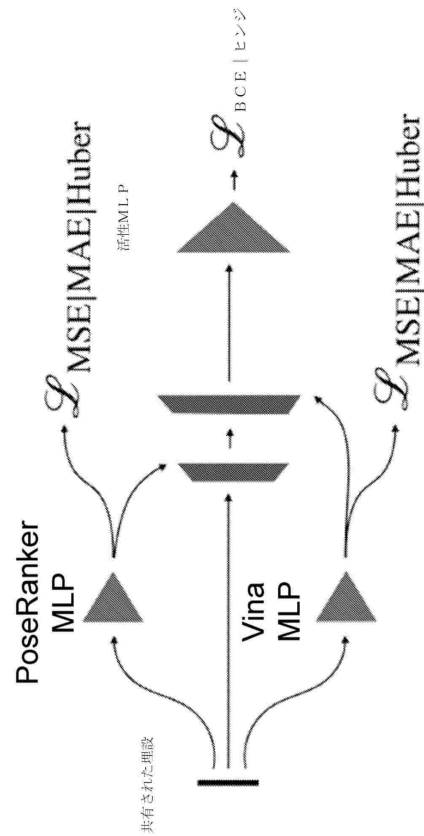
【図 2 2】



【図 2 3】



【図 2 4】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/045250

A. CLASSIFICATION OF SUBJECT MATTER INV. G16C20/50 G16B15/30 ADD.		
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) G16C G06F G16B		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 3 356 999 B1 (ATOMWISE INC [US]) 27 November 2019 (2019-11-27)	1-36
Y	paragraphs [0006] - [0007], [0017], [0042], [0063], [0064]; claims 1-16 -----	1-36
X	MATTHEW RAGOZA ET AL: "Protein-Ligand Scoring with Convolutional Neural Networks", ARKIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 8 December 2016 (2016-12-08), XP080737867, DOI: 10.1021/ACS.JCIM.6B00740	1-36
Y	abstract page 4, paragraph 3 page 5, line 2 - page 6, line 17 ----- -/--	1-36
☒ 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 5 January 2023		Date of mailing of the international search report 13/01/2023
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Nurmi, Jussi

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/045250

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2018/183263 A2 (ATOMWISE INC [US]) 4 October 2018 (2018-10-04)	1-36
Y	the whole document	1-36

X,P	PAWEL GNIEWEK ET AL: "Learning physics confers pose-sensitivity in structure-based virtual screening", ARKIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 2 December 2021 (2021-12-02), XP091110774, the whole document	1-36

10

20

30

40

50

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2022/045250

Patent document cited in search report		Publication date		Patent family member(s)	Publication date
EP 3356999	B1	27-11-2019	CN	108140131 A	08-06-2018
			DK	3356999 T3	03-02-2020
			EP	3356999 A1	08-08-2018
			EP	3680820 A1	15-07-2020
			ES	2772687 T3	08-07-2020
			HU	E047807 T2	28-05-2020
			JP	6975140 B2	01-12-2021
			JP	2019501433 A	17-01-2019
			SG	11201802759Y A	30-05-2018
			WO	2017062382 A1	13-04-2017

WO 2018183263	A2	04-10-2018	CN	109964278 A	02-07-2019
			EP	3433780 A2	30-01-2019
			JP	7121725 B2	18-08-2022
			JP	2020515922 A	28-05-2020
			SG	11201809343R A	29-11-2018
			US	2018285731 A1	04-10-2018
			US	2020334528 A1	22-10-2020
			WO	2018183263 A2	04-10-2018

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,
CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,IT,J
M,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,T
H,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

4 1 0 3 サンフランシスコ マーケット ストリート 7 1 7 スイート 8 0 0 アトムワイズ, イ
ンコーポレイテッド内

(72)発明者 ウォーリー, ブラッド

アメリカ合衆国, カリフォルニア州 9 4 1 0 3 サンフランシスコ マーケット ストリート 7 1
7 スイート 8 0 0 アトムワイズ, インコーポレイテッド内

(72)発明者 アンダーソン, ブランドン

アメリカ合衆国, カリフォルニア州 9 4 1 0 3 サンフランシスコ マーケット ストリート 7 1
7 スイート 8 0 0 アトムワイズ, インコーポレイテッド内

(72)発明者 スタッフォード, ケイト

アメリカ合衆国, カリフォルニア州 9 4 1 0 3 サンフランシスコ マーケット ストリート 7 1
7 スイート 8 0 0 アトムワイズ, インコーポレイテッド内

(72)発明者 ミシンガー, マイケル

アメリカ合衆国, カリフォルニア州 9 4 1 0 3 サンフランシスコ マーケット ストリート 7 1
7 スイート 8 0 0 アトムワイズ, インコーポレイテッド内