

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2025年2月13日(13.02.2025)



(10) 国際公開番号

WO 2025/033296 A1

- (51) 国際特許分類:
G16C 20/50 (2019.01)
- (21) 国際出願番号: PCT/JP2024/027412
- (22) 国際出願日: 2024年7月31日(31.07.2024)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2023-131649 2023年8月10日(10.08.2023) JP
- (71) 出願人: 中外製薬株式会社(CHUGAI SEIYAKU KABUSHIKI KAISHA) [JP/JP]; 〒1158543 東京都北区浮間5丁目5番1号 Tokyo (JP).
- (72) 発明者: 滝沢 秀季 (TAKIZAWA Shuuki); 〒2448602 神奈川県横浜市戸塚区戸塚町2 1 6 番地 中外製薬株式会社内 Kanagawa (JP). 森 啓太 (MORI Keita); 〒2448602 神奈川県横浜市戸塚区戸塚町2 1 6 番地 中外製薬株式

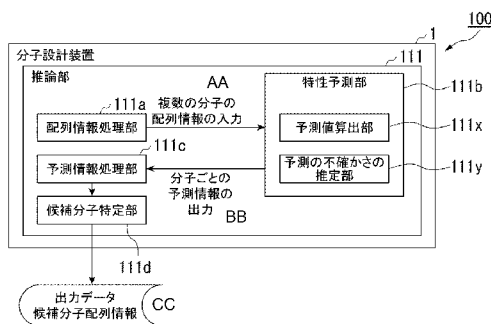
会社内 Kanagawa (JP). 吉村 大(YOSHIMURA Dai); 〒2448602 神奈川県横浜市戸塚区戸塚町2 1 6 番地 中外製薬株式会社内 Kanagawa (JP). 寺本 礼仁(TERAMOTO Reiji); 〒2448602 神奈川県横浜市戸塚区戸塚町2 1 6 番地 中外製薬株式会社内 Kanagawa (JP).

(74) 代理人: 長谷川 芳樹, 外(HASEGAWA Yoshiki et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 9階 創英国際特許法律事務所 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, 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, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,

(54) Title: INFORMATION PROCESSING SYSTEM, INFORMATION PROCESSING METHOD, PROGRAM, AND METHOD FOR MANUFACTURING MOLECULAR COMPOUND

(54) 発明の名称: 情報処理システム、情報処理方法、プログラム、および分子化合物を製造する方法



- 1 Molecule design device
111 Inference unit
111a Sequence information processing unit
111b Characteristic prediction unit
111c Prediction information processing unit
111d Candidate molecule identification unit
111x Prediction value calculation unit
111y Prediction uncertainty estimation unit
AA Input of sequence information for plurality of molecules
BB Output of prediction information for each molecule
CC Output data, candidate molecule sequence information

(57) Abstract: This information processing system for identifying molecules that are suitable for a drug candidate comprises: a characteristic prediction unit that, for molecules constituting elements of a sequence information set that is a set of sequence information for a plurality of different molecules, calculates a characteristic prediction value for each molecule using a prediction model for predicting the characteristics of the molecules from the sequence information for the molecules, and estimates the uncertainty of the prediction; and a candidate molecule identification unit that searches for candidates for molecules having desired characteristics on the basis of the characteristic prediction value and an estimated value of the uncertainty of the prediction.

(57) 要約: 薬物候補に適した分子を特定する情報処理システムは、複数の異なる分子の配列情報の集合である配列情報集合の要素をなす分子について、分子の配列情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する特性予測部と、特性予測値と予測の不確かさの推定値とに基づいて、所望の特性を有する分子の候補を探索する候補分子特定部とを備える。

LY, MA, MD, MG, MK, MN, MU, MW, MX, MY,
MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,
PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,
SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

規則4.17に規定する申立て:

- 出願し及び特許を与えられる出願人の資格に関する申立て (規則4.17(ii))

添付公開書類:

- 国際調査報告 (条約第21条(3))

明 細 書

発明の名称：

情報処理システム、情報処理方法、プログラム、および分子化合物を製造する方法

技術分野

[0001] 本開示の一側面は、情報処理システム、分子設計装置、情報処理方法、プログラム、および分子化合物を製造する方法に関する。

背景技術

[0002] 近年、医薬分野において、創薬の負担の軽減のため、機械学習の情報処理技術が活用されている（特許文献1）。機械学習による予測を用いて、薬として求められる性質を備えた分子を効率よく発見しようとする試みが行われている。

先行技術文献

特許文献

[0003] 特許文献1：国際公開第2018／132752号

発明の概要

発明が解決しようとする課題

[0004] 機械学習による予測は不確実性を伴うため、これを考慮せずに機械学習による予測を活用することは、期待しない結果を招くことがある。また、創薬では解析すべき対象が多く、機械学習による予測を実施可能とする全ての対象に適用することは困難である。

[0005] 上記事情に鑑み、本開示の一側面は、創薬に要する負担をより軽減する技術を提供することを目的としている。

課題を解決するための手段

[0006] 本開示の一側面では以下の態様を提供する：

[A1]

薬物候補に適した分子を特定する情報処理システムであって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する特性予測部と、

前記特性予測値と前記予測の不確かさの推定値とに基づいて、所望の特性を有する分子の候補を探索する候補分子特定部と、

を備える、情報処理システム。

[A 2]

前記ビルディングブロックコンビネーション情報は、前記分子配列情報である、

A 1 に記載の情報処理システム。

[A 3]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

A 1 または A 2 に記載の情報処理システム。

[A 4]

前記分子は、タンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

A 1 から A 3 のいずれかに記載の情報処理システム。

[A 5]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

A 1 から A 4 のいずれかに記載の情報処理システム。

[A 6]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子

に対する結合能である、

A 1 から A 5 のいずれかに記載の情報処理システム。

[A 7]

さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出する予測情報処理部を備え、

前記候補分子特定部は、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合の要素から少なくとも一つの候補分子を特定する、

A 1 から A 6 のいずれかに記載の情報処理システム。

[A 8]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

A 1 から A 7 のいずれかに記載の情報処理システム。

[A 9]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

A 8 に記載の情報処理システム。

[A 1 0]

前記予測情報処理部は、前記特性品質値を与える目的関数として平均分散を算出する、

A 7 から A 9 のいずれかに記載の情報処理システム。

[A 1 1]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、

A 1 から A 1 0 のいずれかに記載の情報処理システム。

[A 1 2]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて前記ビルディングブロックコンビネーション情報集合は組合せ最適化アルゴリズムを用いて取得され、前記分子の特性予

測値あるいは前記特性品質値に基づいて前記組合せ最適化アルゴリズムの抽出パラメータは更新される、

A 1 から A 1 1 のいずれかに記載の情報処理システム。

[A 1 3]

前記組合せ最適化アルゴリズムは、木構造パルツェン推定器である、

A 1 2 に記載の情報処理システム。

[A 1 4]

前記予測モデルは、訓練用の複数の分子のビルディングブロックコンビネーション情報および該分子の特性評価の結果に基づく学習により生成された予測モデルである、

A 1 から A 1 3 のいずれかに記載の情報処理システム。

[0007] [B 1]

所定の予測モデルを用いて分子のビルディングブロックコンビネーション情報から該分子の特性を推論する制御部を備える分子設計装置であって、

前記制御部は、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を取得し、

前記予測モデルにより、前記ビルディングブロックコンビネーション情報集合の要素をなす分子について、該分子のビルディングブロックコンビネーション情報から該分子の特性予測値を算出し、該予測の不確かさを推定し、

前記特性予測値と前記予測の不確かさの推定値とに基づいて、所望の特性を有する分子の候補を探索する、

分子設計装置。

[B 2]

前記ビルディングブロックコンビネーション情報は、前記分子の配列情報である、 B 1 に記載の分子設計装置。

[B 3]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および低

分子化合物のうち少なくとも一つである、

B 1 または B 2 に記載の分子設計装置。

[B 4]

前記分子は、タンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、B 1 から B 3 のいずれかに記載の分子設計装置。

[B 5]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

B 1 から B 4 のいずれかに記載の分子設計装置。

[B 6]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

B 1 から B 5 のいずれかに記載の分子設計装置。

[B 7]

前記制御部は、さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出し、

前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合の要素から少なくとも一つの候補分子を特定する、

B 1 から B 6 のいずれかに記載の分子設計装置。

[B 8]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

B 7 のいずれかに記載の分子設計装置。

[B 9]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

B 8 に記載の分子設計装置。

[B 1 0]

前記制御部は、前記特性品質値を与える目的関数として平均分散を算出する、

B 7 から B 9 のいずれかに記載の分子設計装置。

[B 1 1]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、

B 1 から B 1 0 のいずれかに記載の分子設計装置。

[B 1 2]

前記制御部は、組合せ最適化アルゴリズムを用いて前記ビルディングブロックコンビネーション情報集合を取得し、前記特性予測値あるいは前記特性品質値に基づいて前記組合せ最適化アルゴリズムの抽出パラメータを更新する、

B 1 から B 1 1 のいずれかに記載の分子設計装置。

[B 1 3]

前記組合せ最適化アルゴリズムは、木構造パルツェン推定器である、
B 1 2 に記載の分子設計装置。

[B 1 4]

前記候補分子のビルディングブロックコンビネーション情報を出力する出力部をさらに備える、

B 1 から B 1 3 のいずれかに記載の分子設計装置。

[B 1 5]

前記予測モデルは、訓練用の複数の分子のビルディングブロックコンビネーション情報および該分子の特性評価の結果に基づく学習により生成された予測モデルである、

B 1 から B 1 4 のいずれかに記載の情報処理システム。

[0008] [C 1]

薬物候補に適した分子を特定する情報処理システムであって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合で

あるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得するビルディングブロックコンビネーション情報処理部と、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子について、該分子の特性予測値を算出する特性予測部と、

前記特性予測値に基づいて、所望の特性を有する分子の候補を探索する候補分子特定部と、を備え、

前記ビルディングブロックコンビネーション情報処理部は、

前記分子の特性予測値に基づいて、より所望の特性を有する分子が含まれるように、前記組合せ最適化アルゴリズムの抽出パラメータを更新する情報処理システム。

[C 2]

前記候補分子特定部は、前記抽出パラメータが更新される前に取得されたビルディングブロックコンビネーション情報集合である第1ビルディングブロックコンビネーション情報集合の要素をなす分子について算出された第1特性予測値と、前記抽出パラメータが更新された後に取得されたビルディングブロックコンビネーション情報集合である第2ビルディングブロックコンビネーション情報集合の要素をなす分子について算出された第2特性予測値と、に基づいて、所望の特性を有する分子の候補を探索する、

C 1 の情報処理システム。

[C 3]

前記第1ビルディングブロックコンビネーション情報集合と前記第2ビルディングブロックコンビネーション情報集合とは互いに異なる分子のビルディングブロックコンビネーション情報を含む集合である、

C 2 に記載の情報処理システム。

[C 4]

前記ビルディングブロックコンビネーション情報は、前記分子の配列情報である、

C 1 から C 3 のいずれかに記載の情報処理システム。

[C 5]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

C 1 から C 4 のいずれかに記載の情報処理システム。

[C 6]

前記分子は、タンパク質、抗体、ペプチド、または環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

C 1 から C 5 のいずれかに記載の情報処理システム。

[C 7]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

C 1 から C 6 のいずれかに記載の情報処理システム。

[C 8]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

C 1 から C 7 のいずれかに記載の情報処理システム。

[C 9]

前記特性予測部は、前記特性予測値を算出し、前記予測の不確かさの推定値を算出する、

C 1 から C 8 のいずれかに記載の情報処理システム。

[C 10]

前記ビルディングブロックコンビネーション情報処理部は、

前記分子の特性予測値および前記予測の不確かさの推定値に基づいて、より所望の特性を有する分子が含まれるように組合せ最適化アルゴリズムの抽

出パラメータを更新する、

C 1 から C 9 のいずれかに記載の情報処理システム。

[C 1 1]

前記特性予測部より出力された、前記特性予測値および／または前記予測の不確かさの推定値を含む予測情報の処理を行う予測情報処理部をさらに備え、

前記予測情報処理部は、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出し、

前記候補分子特定部は、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合からの少なくとも一つの分子を特定する、

C 9 または C 1 0 に記載の情報処理システム。

[C 1 2]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

C 1 1 に記載の情報処理システム。

[C 1 3]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

C 1 2 に記載の情報処理システム。

[C 1 4]

前記予測情報処理部は、前記特性品質値を与える目的関数として平均分散を算出する、

C 1 1 から C 1 3 のいずれかに記載の情報処理システム。

[C 1 5]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、

C 9 から C 1 4 のいずれかに記載の情報処理システム。

[C 1 6]

前記組合せ最適化アルゴリズムとして、木構造パルツェン推定器を用いる

、

C 1 から C 1 5 に記載の情報処理システム。

[0009] [D 1]

所定の予測モデルを用いて分子のビルディングブロックコンビネーション情報から該分子の特性を推論する制御部を備える分子設計装置であって、

前記制御部は、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得し、

前記予測モデルを用いた前記ビルディングブロックコンビネーション情報集合の要素をなす分子についての前記特性予測値を算出し、

前記予測モデルによる分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新し、

前記特性予測値に基づいて、所望の特性を有する分子の候補を探索する、分子設計装置。

[D 2]

前記制御部は、

前記抽出パラメータの更新前に複数の異なる分子のビルディングブロックコンビネーション情報の集合である第 1 ビルディングブロックコンビネーション情報集合を取得し、

前記第 1 ビルディングブロックコンビネーション情報集合の要素をなす分子について、前記予測モデルを用いて第 1 特性予測値を算出し、

前記抽出パラメータの更新後に、複数の異なる分子のビルディングブロックコンビネーション情報の集合である第 2 ビルディングブロックコンビネーション情報集合をさらに取得し、

前記第 2 ビルディングブロックコンビネーション情報集合の要素をなす分子について、前記予測モデルを用いて第 2 特性予測値を算出し、

前記第 1 特性予測値と前記第 2 特性予測値とに基づいて、所望の特性を有する分子の候補を探索する、

D 1 に記載の分子設計装置。

[D 3]

前記第 1 ビルディングブロックコンビネーション情報集合と前記第 2 ビルディングブロックコンビネーション情報集合とは互いに異なる分子の配列情報を含む集合である、

D 2 に記載の分子設計装置。

[D 4]

前記ビルディングブロックコンビネーション情報は、前記分子の配列情報である、

D 1 から D 3 のいずれかに記載の分子設計装置。

[D 5]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

D 1 から D 5 のいずれかに記載の分子設計装置。

[D 6]

前記分子は、タンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

D 1 から D 5 のいずれかに記載の分子設計装置。

[D 7]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

D 1 から D 6 のいずれかに記載の分子設計装置。

[D 8]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

D 1 から D 7 のいずれかに記載の分子設計装置。

[D 9]

前記制御部は、前記特性予測値を算出し、さらに予測の不確かさの推定値を算出する、

D 1 から D 8 のいずれかに記載の分子設計装置。

[D 10]

前記制御部は、前記分子の特性の予測値および前記予測の不確かさの推定値に基づいて、より所望の特性を有する分子が含まれるように前記ビルディングブロックコンビネーション情報集合を取得する、

D 1 から D 9 のいずれかに記載の情報処理システム。

[D 11]

前記制御部は、さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出し、

前記候補分子特定部は、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合から少なくとも一つの分子を特定する、

D 10 に記載の分子設計装置。

[D 12]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

D 11 に記載の分子設計装置。

[D 13]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

D 12 に記載の分子設計装置。

[D 14]

前記制御部は、前記特性品質値を与える目的関数として平均分散を算出する、

D 11 から D 13 のいずれかに記載の分子設計装置。

[D 1 5]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、
D 9 から D 1 4 のいずれかに記載の分子設計装置。

[D 1 6]

前記組合せ最適化アルゴリズムとして、木構造パルツェン推定器を用いる
、
D 1 から D 1 5 のいずれかに記載の分子設計装置。

[D 1 7]

前記予測モデルは、分子ごとに異なるビルディングブロックコンビネーション情報および該分子の特性を示す訓練データに基づいて学習されたモデルである、

D 1 から D 1 6 のいずれかに記載の分子設計装置。

[0010] [E 1]

薬物候補に適した分子を特定する情報処理システムにおける情報処理方法であって、複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定するステップと、

前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索するステップと、を含む、
情報処理方法。

[E 2]

前記ビルディングブロックコンビネーション情報は、前記分子の配列の情報である、

E 1 に記載の情報処理方法。

[E 3]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および

、低分子化合物のうち少なくとも一つである、

E 1 または E 2 に記載の情報処理方法。

[E 4]

前記分子は、タンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報は、アミノ酸配列の情報である、

E 1 から E 3 のいずれかに記載の情報処理方法。

[E 5]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

E 1 から E 4 のいずれかに記載の情報処理方法。

[E 6]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

E 1 から E 5 のいずれかに記載の情報処理方法。

[E 7]

さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出するステップを含み、

前記探索するステップは、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合から少なくとも一つの分子を特定するステップを含む、

E 1 から E 6 のいずれかに記載の情報処理方法。

[E 8]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

E 7 に記載の情報処理方法。

[E 9]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変

数を入力とする任意の関数の出力値である、

E 8 に記載の情報処理方法。

[E 1 0]

前記特性品質値を算出するステップは、前記特性品質値を与える目的関数として平均分散を算出する、

E 7 から E 9 のいずれかに記載の情報処理方法。

[E 1 1]

前記予測の不確かさの推定値は、前記予測値の標準偏差である、

E 1 から E 1 0 のいずれかに記載の情報処理方法。

[E 1 2]

特定した前記候分子のビルディングブロックコンビネーション情報を出力するステップを含む、

E 1 から E 1 1 のいずれかに記載の情報処理方法。

[E 1 3]

特定した前記分子の特性予測値の情報をさらに出力する、

E 1 2 に記載の情報処理方法。

[E 1 4]

前記探索するステップは、前記ビルディングブロックコンビネーション情報集合から前記特性品質値が所定値以上の分子を特定するステップを含む、

E 7 から E 1 3 のいずれかに記載の情報処理方法。

[E 1 5]

前記探索するステップは、分子ごとに前記特性品質値の順位を特定し、所定の順位内の分子を特定するステップを含む、

E 7 から E 1 3 のいずれかに記載の情報処理方法。

[E 1 6]

前記探索するステップは、前記ビルディングブロックコンビネーション情報集合から、前記特性予測値と前記予測の不確かさがそれぞれ所定の条件を満たす分子を、少なくとも一つ選択するステップを含む

E 1 から E 1 5 のいずれかに記載の情報処理方法。

[E 1 7]

組合せ最適化アルゴリズムを用いて、前記ビルディングブロックコンビネーション情報集合を取得し、前記特性予測値あるいは前記特性品質値に基づいて前記組合せ最適化アルゴリズムの抽出パラメータを更新する、

E 1 から E 1 6 のいずれかに記載の情報処理方法。

[E 1 8]

前記組合せ最適化アルゴリズムは、木構造パルツェン推定器である、

E 1 7 に記載の情報処理方法。

[E 1 9]

前記予測モデルは、分子ごとに異なるビルディングブロックコンビネーション情報および該分子の特性を示す訓練データに基づいて学習されたモデルである、

E 1 から E 1 8 のいずれかに記載の情報処理方法。

[0011] [F 1]

コンピュータに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する手順と、

前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索する手順と、

を実行させるためのプログラム。

[F 2]

前記ビルディングブロックコンビネーション情報は、前記分子の配列情報である、

F 1 に記載のプログラム。

[F 3]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

F 1 または F 2 に記載のプログラム。

[F 4]

前記分子は、タンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報は、アミノ酸配列の情報である、

F 1 から F 3 のいずれかに記載のプログラム。

[F 5]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

F 1 から F 4 のいずれかに記載のプログラム。

[F 6]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

F 1 から F 5 のいずれかに記載のプログラム。

[F 7]

さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出する手順を有し、

前記探索する手順は、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合から少なくとも一つの分子を特定する手順を含む、

F 1 から F 6 のいずれかに記載のプログラム。

[F 8]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

F 7 に記載のプログラム。

[F 9]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

F 8 に記載のプログラム。

[F 1 0]

前記特性品質値を算出する手順は、前記特性品質値を与える目的関数として平均分散を算出する、

F 7 から F 9 のいずれかに記載のプログラム。

[F 1 1]

前記不確かさの推定値は、前記予測値の標準偏差である、

F 1 から F 1 0 のいずれかに記載のプログラム。

[F 1 2]

より所望の特性を有する分子の候補を探索し特定した候補分子のビルディングブロックコンビネーション情報を出力する手順を含む、

F 1 から F 1 1 のいずれかに記載のプログラム。

[F 1 3]

より所望の特性を有する分子の候補を探索し特定した候補分子の特性予測値の情報をさらに出力する、

F 1 2 に記載のプログラム。

[F 1 4]

前記探索する手順は、前記ビルディングブロックコンビネーション情報集合から、前記特性品質値が所定値以上の分子を特定する手順を含む、

F 7 から F 1 3 のいずれかに記載のプログラム。

[F 1 5]

前記探索する手順は、分子ごとに前記特性品質値の順位を特定し、所定の順位内の分子を特定する手順を含む、

F 7 から F 1 3 のいずれかに記載のプログラム。

[F 1 6]

前記探索するステップは、前記ビルディングブロックコンビネーション情報集合から、前記特性予測値と前記予測の不確かさの推定値がそれぞれ所定の条件を満たす分子を、少なくとも一つ選択する手順を含む

F 1 から F 1 5 いずれかに記載のプログラム。

[F 1 7]

組合せ最適化アルゴリズムを用いて、前記分子の特性予測値あるいは前記特性品質値に基づいて前記ビルディングブロックコンビネーション情報集合取得の抽出パラメータを更新する手順をさらに含む、

F 1 から F 1 6 のいずれかに記載のプログラム。

[F 1 8]

前記組合せ最適化アルゴリズムは、木構造パルツェン推定器である、

F 1 7 に記載のプログラム。

[F 1 9]

前記予測モデルは、分子ごとに異なるビルディングブロックコンビネーション情報および該分子の特性を示す訓練データに基づいて学習されたモデルである、

F 1 から F 1 8 のいずれかに記載のプログラム。

[0012] [G 1]

分子化合物を製造する方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合にアクセスし、

前記ビルディングブロックコンビネーション情報集合を予測モデルに入力する入力ステップと、

前記予測モデルより出力される、前記ビルディングブロックコンビネーション情報集合に含まれる分子ごとの特性予測値および予測の不確かさの推定値に基づき、前記ビルディングブロックコンビネーション情報集合からより所望の特性を有する分子を探索し、候補分子として特定する推論ステップと、

当該候補分子にかかるビルディングブロックコンビネーション情報を出力する出力ステップと、

前記ビルディングブロックコンビネーション情報に示される分子配列を有する前記分子化合物を生成する生成ステップと、
を有する方法。

[G 2]

前記候補分子は生物学的配列を有し、前記ビルディングブロックコンビネーション情報は配列情報である、

G 1 に記載の方法。

[G 3]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

G 1 または G 2 に記載の方法。

[G 4]

前記分子はタンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

G 1 から G 3 のいずれかに記載の方法。

[G 5]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

G 1 から G 4 のいずれかに記載の方法。

[G 6]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

G 1 から G 5 のいずれかに記載の方法。

[G 7]

さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性

品質値を算出し、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合から少なくとも一つの候補分子を特定するステップを含む、

G 1 から G 6 のいずれかに記載の方法。

[G 8]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

G 7 に記載の方法。

[G 9]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

G 8 に記載の方法。

[G 1 0]

前記特性品質値を算出するステップは、前記特性品質値を与える目的関数として平均分散を算出する、

G 7 から G 9 のいずれかに記載の方法。

[G 1 1]

前記不確かさの推定値は、前記予測値の標準偏差である、

G 1 から G 1 0 のいずれかに記載の方法。

[G 1 2]

前記候補分子にかかる情報は、前記候補分子のビルディングブロックコンビネーション情報を含む、

G 1 から G 1 1 のいずれかに記載の方法。

[G 1 3]

前記候補分子にかかる情報は、前記候補分子の特性予測値を含む、

G 1 から G 1 2 のいずれかに記載の方法。

[G 1 4]

前記候補分子を探索するステップは、前記ビルディングブロックコンビネ

ーション情報集合から、前記特性品質値が所定値以上の分子を、前記候補分子として特定するステップを含む、

G 7 から G 1 3 のいずれかに記載の方法。

[G 1 5]

前記候補分子を探索するステップは、分子ごとに前記特性品質値の順位を特定し、所定の順位内の分子を、前記候補分子として特定するステップを含む、

G 7 から G 1 3 のいずれかに記載の方法。

[G 1 6]

前記候補分子を探索するステップは、前記ビルディングブロックコンビネーション情報集合から、前記特性予測値と前記予測の不確かさの推定値がそれぞれ所定の条件を満たす分子を、少なくとも一つ選択するステップを含む、

G 1 から G 1 5 のいずれかに記載の方法。

[G 1 7]

前記ビルディングブロックコンビネーション情報集合は組合せ最適化アルゴリズムを用いて取得され、

前記組合せ最適アルゴリズムの抽出パラメータを前記分子の特性予測値あるいは前記特性品質値に基づいて更新する手順をさらに含む、

G 1 から G 1 6 のいずれかに記載の方法。

[G 1 8]

前記組合せ最適化アルゴリズムは、木構造パルツェン推定器である、
G 1 7 に記載の方法。

[G 1 9]

複数の候補分子それぞれの特性の実験値のうち、最良の実験値を有する候補分子を選択する選択ステップを有する

G 1 から G 1 8 のいずれかに記載の方法。

[0013] [H 1]

薬物候補に適した分子を特定する情報処理システムにおける情報処理方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得するステップと、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子について、該分子の特性予測値を算出するステップと、

前記特性予測値に基づいて、より所望の特性を有する分子を探索するステップと、

前記分子の特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新するステップを含む、
情報処理方法。

[H 2]

前記抽出パラメータが更新される前に取得されたビルディングブロックコンビネーション情報集合である第1ビルディングブロックコンビネーション情報集合の要素をなす分子について算出された第1特性予測値と、前記抽出パラメータが更新された後に取得されたビルディングブロックコンビネーション情報集合である第2ビルディングブロックコンビネーション情報集合の要素をなす分子について算出された第2特性予測値と、に基づいて、所望の特性を有する分子の候補を探索するステップを有する

H 1 に記載の情報処理方法。

[H 3]

前記第1ビルディングブロックコンビネーション情報集合と前記第2ビルディングブロックコンビネーション情報集合とは互いに異なる分子のビルディングブロックコンビネーション情報を含む集合である、

H 2 に記載の情報処理方法。

[H 4]

前記ビルディングブロックコンビネーション情報は、前記分子の配列の情報である、

H 1 から H 3 のいずれかに記載の情報処理方法。

[H 5]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

H 1 から H 4 のいずれかに記載の情報処理システム。

[H 6]

前記分子は、タンパク質、抗体、ペプチド、または環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

H 1 から H 5 のいずれかに記載の情報処理方法。

[H 7]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

H 6 のいずれかに記載の情報処理方法。

[H 8]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

H 1 から H 7 のいずれかに記載の情報処理方法。

[H 9]

前記特性予測値と、前記予測の不確かさの推定値を算出するステップを有する、

H 1 から H 8 のいずれかに記載の情報処理方法。

[H 1 0]

前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を

算出するステップを有し、

前記探索するステップは、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合からの少なくとも一つの分子を特定するステップを含む、

H 9 に記載の情報処理方法。

[H 1 1]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

H 1 0 に記載の情報処理方法。

[H 1 2]

前記特性品質値は、前記特性予測値と前記予測の不確かさの二変数を入力とする所定の関数の出力値である、

H 1 1 に記載の情報処理方法。

[H 1 3]

前記特性品質値を算出するステップは、前記特性品質値を与える目的関数として平均分散を算出するステップを含む、

H 1 0 から H 1 2 のいずれかに記載の情報処理方法。

[H 1 4]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、

H 9 から H 1 3 のいずれかに記載の情報処理方法。

[H 1 5]

前記抽出パラメータの更新に係る組合せ最適化アルゴリズムとして、木構造パルツェン推定器を用いる、

H 1 から H 1 4 のいずれかに記載の情報処理方法。

[0014] [I 1]

コンピュータに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴ

リズムに従って取得する取得手順と、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子の特性予測値を算出する予測手順と、

分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新する更新手順と、

前記特性予測値に基づいて、所望の特性を有する分子を探索する探索手順と、

を実行させるためのプログラム。

[1 2]

前記取得手順は、

前記抽出パラメータの更新前に複数の異なる分子のビルディングブロックコンビネーション情報の集合である第1ビルディングブロックコンビネーション情報集合を取得する手順と、

前記抽出パラメータの更新後に、複数の異なる分子のビルディングブロックコンビネーション情報の集合である第2ビルディングブロックコンビネーション情報集合をさらに取得する手順と、を有し、

前記予測手順は、

前記第1ビルディングブロックコンビネーション情報集合の要素をなす分子について、前記予測モデルを用いて第1特性予測値を算出する手順と、

前記第2ビルディングブロックコンビネーション情報集合の要素をなす分子について、前記予測モデルを用いて第2特性予測値を算出する手順と、を有し、

前記探索手順は、

前記第1のビルディングブロックコンビネーション情報集合と前記第2のビルディングブロックコンビネーション情報集合とから、前記第1特性予測値と前記第2特性予測値とに基づいて、所望の特性を有する分子の候補を探

索する探索手順を有する

Ⅰ 1 に記載のプログラム。

[Ⅰ 3]

前記第 1 ビルディングブロックコンビネーション情報集合と前記第 2 ビルディングブロックコンビネーション情報集合とは互いに異なる分子のビルディングブロックコンビネーション情報を含む集合である、

Ⅰ 2 に記載のプログラム。

[Ⅰ 4]

前記ビルディングブロックコンビネーション情報は、前記分子の配列の情報である、

Ⅰ 1 から Ⅰ 3 のいずれかに記載のプログラム。

[Ⅰ 5]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

Ⅰ 1 から Ⅰ 4 のいずれかに記載のプログラム。

[Ⅰ 6]

前記分子は、タンパク質、抗体、ペプチド、または、環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

Ⅰ 1 から Ⅰ 5 のいずれかに記載のプログラム。

[Ⅰ 7]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

Ⅰ 1 から Ⅰ 6 のいずれかに記載のプログラム。

[Ⅰ 8]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

Ⅰ 1 から Ⅰ 7 のいずれかに記載のプログラム。

[1 9]

前記特性予測値を算出し、前記予測の不確かさの推定値を算出する手順を有する、

1 1 から 1 8 のいずれかに記載のプログラム。

[1 1 0]

さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出する手順を有し、

前記探索手順は、

前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合から少なくとも一つの分子を特定する手順を有する、

1 9 に記載のプログラム。

[1 1 1]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

1 1 0 に記載のプログラム。

[1 1 2]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

1 1 1 に記載のプログラム。

[1 1 3]

前記特性品質値を算出する手順は、前記特性品質値を与える目的関数として平均分散を算出する、

1 1 0 から 1 1 2 のいずれかに記載のプログラム。

[1 1 4]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、

1 9 から 1 1 3 のいずれかに記載のプログラム。

[1 1 5]

前記ビルディングブロックコンビネーション情報集合の更新に係る組合せ

最適化アルゴリズムとして、木構造パルツェン推定器を用いる、

11から114のいずれかに記載のプログラム。

[116]

前記予測モデルは、分子ごとに異なるビルディングブロックコンビネーション情報および該分子の特性を示す訓練データに基づいて学習されたモデルである、

11から115のいずれかに記載のプログラム。

[0015] [J1]

分子化合物を製造する方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得する取得ステップと、

前記ビルディングブロックコンビネーション情報集合を予測モデルに入力する入力ステップと、

前記予測モデルより出力される、前記ビルディングブロックコンビネーション情報集合に含まれる分子ごとの特性予測値に基づき、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新する更新ステップと、

前記特性予測値に基づき、前記ビルディングブロックコンビネーション情報集合からより所望の特性を有する分子を探索し、候補分子として特定するステップと、

当該候補分子にかかるビルディングブロックコンビネーション情報を出力する出力ステップと、

前記ビルディングブロックコンビネーション情報に示される分子配列を有する前記分子化合物を生成する生成ステップと、

を有する方法。

[J2]

前記抽出パラメータが更新される前に取得されたビルディングブロックコ

ンビネーション情報集合である第1ビルディングブロックコンビネーション情報集合の要素をなす分子について算出された第1特性予測値と、前記抽出パラメータが更新された後に取得されたビルディングブロックコンビネーション情報集合である第2ビルディングブロックコンビネーション情報集合の要素をなす分子について算出された第2特性予測値と、に基づいて、前記候補分子を探索するステップを有する

J 1 に記載の方法。

[J 3]

前記第1ビルディングブロックコンビネーション情報集合と前記第2ビルディングブロックコンビネーション情報集合とは互いに異なる分子のビルディングブロックコンビネーション情報を含む集合である、

J 2 に記載の方法。

[J 4]

前記候補分子は、生物学的配列を有する、

J 1 から J 3 のいずれかに記載の方法。

[J 5]

前記ビルディングブロックコンビネーション情報は、前記分子の配列情報である、

J 1 から J 4 のいずれかに記載の方法。

[J 6]

前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、

J 1 から J 5 のいずれかに記載の方法。

[J 7]

前記分子は、タンパク質、抗体、ペプチド、または環状ペプチドであり、前記ビルディングブロックコンビネーション情報はアミノ酸配列の情報である、

J 1 から J 6 のいずれかに記載の方法。

[J 8]

前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、

J 1 から J 7 のいずれかに記載の方法。

[J 9]

前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、

J 1 から J 8 のいずれかに記載の方法。

[J 1 0]

前記特性予測値と、前記予測の不確かさの推定値を算出するステップを有する、

J 1 から J 9 のいずれかに記載の方法。

[J 1 1]

前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出するステップを有し、

前記候補分子を探索する推論ステップは、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合からの少なくとも一つの分子を前記候補分子として特定するステップを含む、

J 1 0 に記載の方法。

[J 1 2]

前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、

J 1 1 に記載の方法。

[J 1 3]

前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、

J 1 2 に記載の方法。

[J 1 4]

前記特性品質値を算出するステップは、前記特性品質値を与える目的関数として平均分散を算出するステップを含む、

J 1 1 から J 1 3 のいずれかに記載の方法。

[J 1 5]

前記予測の不確かさの推定値は、前記特性予測値の標準偏差である、

J 1 0 から J 1 4 のいずれかに記載の方法。

[J 1 6]

前記ビルディングブロックコンビネーション情報集合の更新に係る組合せ最適化アルゴリズムとして、木構造パルツェン推定器を用いる、

J 1 から J 1 5 のいずれかに記載の方法。

[J 1 7]

複数の候補分子それぞれの特性の実験値のうち、最良の実験値を有する候補分子を選択するステップを有する

J 1 から J 1 6 のいずれかに記載の方法。

[0016] [K]

プロセッサとメモリを備えるコンピュータシステムであって、前記メモリは 1 以上の指令を記憶するように構成され、

当該指令は、前記プロセッサに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出させ、該予測の不確かさを推定させ、

前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索させる、

コンピュータシステム。

[0017] [L]

1 以上の指令を記憶する非一時的コンピュータ読み取り可能記憶媒体であ

って、

当該指令は、コンピュータに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出させ、該予測の不確かさを推定させ、

前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を候補分子として探索させる、

非一時的コンピュータ読み取り可能記憶媒体。

[0018] [M]

プロセッサとメモリを備えるコンピュータシステムであって、前記メモリは1以上の指令を記憶するように構成され、

当該指令は、前記プロセッサに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得させ、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子の特性予測値を算出させ、

分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新させ、

前記特性予測値に基づいて、所望の特性を有する分子を探索させる、

コンピュータシステム。

[0019] [N]

1以上の指令を記憶する非一時的コンピュータ読み取り可能記憶媒体であって、

当該指令は、コンピュータに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得させ、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子の特性予測値を算出させ、

分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新させ、

前記特性予測値に基づいて、所望の特性を有する分子を探索させる、
非一時的コンピュータ読み取り可能記憶媒体。

発明の効果

[0020] 本開示の一側面によれば創薬に要する負担をより軽減することができる。

図面の簡単な説明

[0021] [図1]第一実施形態の分子設計装置を含む創薬システムの一例を示す図。

[図2]第二実施形態の分子設計装置を含む創薬システムの一例を示す図。

[図3]本願の実施形態における複数の分子配列情報の一例を説明する説明図。

[図4]本願の実施形態における分子設計装置のハードウェア構成の一例を示す図。

[図5]本願の実施形態における制御部の構成の一例を示す図。

[図6]第一実施形態における分子設計装置が実行する処理の流れの一例を示すフローチャート。

[図7]第二実施形態における分子設計装置が実行する処理の流れの一例を示すフローチャート。

[図8]第二実施形態における組合せ最適化処理の一例を示すフローチャート。

[図9]第二実施形態における組合せ最適化処理の他の例を示すフローチャート。

[図10]本願の実施形態における適用例に係る最適化課題設定の概略図。

[図11]第一検証例に係るTPEでの最適化による目標関数値の算出例を示す

図。

[図12]第一検証例に係るTPEによるサンプリングで得られた配列の予測平均値と予測標準偏差値との関係を示す図。

[図13]第一検証例に係るTPEによるサンプリングで得られた配列の編集距離の密度分布を示す図。

[図14]第一検証例に係る提案配列に対する疑似正解モデルスコアの分布を示す図。

[図15]第一検証例に係る提案配列の編集距離を示す図。

[図16]第一検証例に係る提案配列の予測標準偏差値を示す図。

[図17]第一検証例に係る探索空間における変異候補箇所ごとのアミノ酸の候補の例を示す表。

[図18]第一検証例に係るTPEのパラメータ設定例を示す表。

[図19]第二検証例に係るTPEによるサンプリングで得られた配列の予測平均値と予測標準偏差値の分布を例示する図。

[図20]第二検証例に係るTPEのサンプリングによる配列を例示するt-SNE可視化図。

[図21]第二検証例に係る提案配列の予測平均値を例示する図。

[図22]第二検証例に係る提案配列の予測分散値を例示する図。

[図23]第二検証例に係る提案配列の発現量の分布を例示する図。

[図24]第二検証例に係る配列ごとのオクテット値の分布を例示する図。

発明を実施するための形態

[0022] 1. 定義

本明細書で用語「および／または」は、「および／または」の前後に記載される各対象またはそれらの任意の組合せを示すために用いられる。例えば、「A、Bおよび／またはC」は、「A」、「B」、「C」の各対象のほか、組合せ「AおよびB」、「AおよびC」、「BおよびC」、ならびに「AおよびBおよびC」も含む。

[0023] ・アミノ酸

本明細書において、天然アミノ酸と非天然アミノ酸とを含んでもよい。天然アミノ酸の場合、例えば、Ala/A、Leu/L、Arg/R、Lys/K、Asn/N、Met/M、Asp/D、Phe/F、Cys/C、Pro/P、Gln/Q、Ser/S、Glu/E、Thr/T、Gly/G、Trp/W、His/H、Tyr/Y、Ile/I、Val/Vと表されるように、アミノ酸は1文字コードまたは3文字コード、またはその両方で表記されている。

[0024] ・アミノ酸の改変

抗原結合分子のアミノ酸配列中のアミノ酸の改変のためには、部位特異的変異誘発法（Kunkelら（Proc. Natl. Acad. Sci. USA（1985）82，488-492））、Overlap extension PCR等の公知の方法が適宜採用され得る。また、天然のアミノ酸以外のアミノ酸に置換するアミノ酸の改変方法として、複数の公知の方法もまた採用され得る（Annu. Rev. Biophys. Biomol. Struct.（2006）35，225-249、Proc. Natl. Acad. Sci. U.S.A.（2003）100（11），6353-6357）。例えば、終止コドンの1つであるUAGコドン（アンバーコドン）の相補的アンバーサプレッサーtRNAに非天然アミノ酸が結合されたtRNAが含まれる無細胞翻訳系システム（Clover Direct（Protein Express））等も用いられてよい。

[0025] ・抗原

本明細書において「抗原」は抗原結合ドメインが結合するエピトープを含む限りその構造は特定の構造に限定されない。ある態様において、抗原は4アミノ酸以上のペプチド、またはポリペプチド、またはタンパク質である。

上記の抗原の例示には、細胞膜に発現する膜型分子、および細胞から細胞外に分泌される可溶性分子が含まれる。

[0026] ・抗原結合ドメイン

本明細書において、「抗原結合ドメイン」は目的とする抗原に結合するかぎりどのような構造のドメインも使用され得る。そのようなドメインの例として、例えば、抗体の重鎖および軽鎖の可変領域、生体内に存在する細胞膜タンパクであるAvimerに含まれる35アミノ酸程度のAドメインと呼ばれるモジュール（国際公開W02004/044011、W02005/040229）、細胞膜に発現する糖タ

ンパク質であるfibronectin中のタンパク質に結合するドメインである10Fn3ドメインを含むAdnectin（国際公開W02002/032925）、ProteinAの58アミノ酸からなる3つのヘリックスの束（bundle）を構成するIgG結合ドメインをscaffoldとするAffibody（国際公開W01995/001937）、33アミノ酸残基を含むターンと2つの逆並行ヘリックスおよびループのサブユニットが繰り返し積み重なった構造を有するアンキリン反復（ankyrin repeat：AR）の分子表面に露出する領域であるDARPs（Designed Ankyrin Repeat proteins）（国際公開W02002/020565）、好中球ゲラチナーゼ結合リポカリン（neutrophil gelatinase associated lipocalin（NGAL））等のリポカリン分子において高度に保存された8つの逆並行ストランドが中央方向にねじれたバレル構造の片側を支える4つのループ領域であるAnticalin等（国際公開W02003/029462）、ヤツメウナギ、ヌタウナギなど無顎類の獲得免疫システムとしてイムノグロブリンの構造を有さない可変性リンパ球受容体（variable lymphocyte receptor（VLR））のロイシン残基に富んだりリピート（leucine-rich-repeat（LRR））モジュールが繰り返し積み重なった馬てい形の構造の内部の並行型シート構造のくぼんだ領域（国際公開W02008/016854）が挙げられる。

本開示の抗原結合ドメインの例として、抗体の重鎖および軽鎖の可変領域を含む抗原結合ドメインが挙げられる。こうした抗原結合ドメインの例としては、「scFv（single chain Fv）」、「単鎖抗体（single chain antibody）」、「Fv」、「scFv2（single chain Fv 2）」、「Fab」または「F(ab')₂」等が挙げられる。

[0027] ・ 抗原結合分子

本開示において、抗原結合ドメインを含む抗原結合分子は最も広義な意味として使用されており、具体的には、それらが抗原結合ドメインを含む限り、様々な分子型が含まれる。抗原結合分子は、抗原結合ドメインのみからなる分子であってもよく、抗原結合ドメインおよび他のドメインを含む分子であってもよい。例えば、抗原結合分子が抗原結合ドメインとFc領域が結合した分子で有る場合、例として、完全抗体および抗体断片が挙げられる。抗体

には、単一のモノクローナル抗体（アゴニストおよびアンタゴニスト抗体を含む）、ヒト抗体、ヒト化抗体、キメラ抗体等が含まれ得る。既存の安定な α/β バレルタンパク質構造等の立体構造がscaffold（土台）として用いられ、その一部分の構造のみが抗原結合ドメインの構築のためにライブラリ化されたスキャフォールド分子も、本開示の抗原結合分子に含まれ得る。

[0028] 本明細書で用語「抗体」は、最も広い意味で使用され、所望の抗原結合活性を示す限りは、これらに限定されるものではないが、モノクローナル抗体、ポリクローナル抗体、多重特異性抗体（例えば、二重特異性抗体）および抗体断片を含む、種々の抗体構造を包含する。

抗体はそれが天然に存在する血漿、血清等の天然資源、または抗体を産生するハイブリドーマ細胞の培養上清から単離され得るし、または遺伝子組換え等の手法を用いることによって部分的にもしくは完全に合成され得る。抗体の例としては免疫グロブリンのアイソタイプおよびそれらのアイソタイプのサブクラスが挙げられる。ヒトの免疫グロブリンとして、IgG1、IgG2、IgG3、IgG4、IgA1、IgA2、IgD、IgE、IgMの9種類のクラス（アイソタイプ）が知られている。本開示の抗体には、これらのアイソタイプのうちIgG1、IgG2、IgG3、IgG4が含まれ得る。ヒトIgG1、ヒトIgG2、ヒトIgG3、ヒトIgG4定常領域としては、遺伝子多型による複数のアロタイプ配列がSequences of proteins of immunological interest, NIH Publication No.91-3242に記載されているが、本開示においてはそのいずれであってもよい。特にヒトIgG1の配列としては、EUナンバリングで表される356-358位のアミノ酸配列がDELであってもEEMであってもよい。また、ヒトIg κ (Kappa)定常領域とヒトIg λ (Lambda)定常領域としては、遺伝子多型による複数のアロタイプ配列がSequences of proteins of immunological interest, NIH Publication No.91-3242に記載されているが、本開示においてはそのいずれであってもよい。

[0029] 「抗体断片」は、完全抗体が結合する抗原に結合する当該完全抗体の一部を含む、当該完全抗体以外の分子のことをいう。抗体断片の例は、これら

に限定されるものではないが、Fv、Fab、Fab'、Fab' -SH、F(ab')₂；ダイアボディ；線状抗体；単鎖抗体分子（例えば、scFv）；および、抗体断片から形成された多重特異性抗体を含む。

[0030] 用語「全長抗体」、「完全抗体」、および「全部抗体」は、本明細書では相互に交換可能に用いられ、天然型抗体構造に実質的に類似した構造を有する、または本明細書で定義するFc領域を含む重鎖を有する抗体のことをいう。

[0031] 用語「可変領域」または「可変ドメイン」は、抗体を抗原へと結合させることに関与する、抗体の重鎖または軽鎖のドメインのことをいう。天然型抗体の重鎖および軽鎖の可変ドメイン（それぞれVHおよびVL）は、通常、各ドメインが4つの保存されたフレームワーク領域（FR）および3つの超可変領域（HVR）を含む、類似の構造を有する。（例えば、Kindt et al., Kuby Immunology, 6th ed., W.H. Freeman and Co., page 91 (2007) 参照。）1つのVHまたはVLドメインで、抗原結合特異性を与えるに充分であろう。さらに、ある特定の抗原に結合する抗体は、当該抗原に結合する抗体からのVHまたはVLドメインを使ってそれぞれVLまたはVHドメインの相補的ライブラリをスクリーニングして、単離されてもよい。例えば、Portolano et al., J. Immunol. 150:880-887 (1993)；Clarkson et al., Nature 352: 624-628 (1991) 参照。

[0032] 本明細書において「分子量」とは、化合物分子を構成する原子の原子量の総和（単位：「g/mol」）を意味し、分子構造式に含まれている原子の原子量の総和を算出することで得られる。本明細書においては分子量の単位を省略することがある。なお、分子量は、例えば、液体クロマトグラフィー質量分析（LC/MS）により測定することができる。

[0033] 本明細書において中分子化合物、または中分子とは、分子量が500 g/mol以上かつ30000 g/mol未満の化合物である。

中分子化合物としては、例えば分子量が500 g/mol以上かつ6000 g/mol未満の化合物であり、分子量が500 g/mol以上かつ40

00 g/mol未満の化合物でもよく、分子量が600 g/mol以上かつ4000 g/mol未満の化合物でもよく、分子量が700 g/mol以上かつ3000 g/mol未満の化合物でもよい。中分子化合物としては、例えばペプチド鎖を含むペプチド化合物、核酸、または糖鎖であり、ペプチド化合物でもよく、5～30残基のアミノ酸を含むペプチド化合物でもよく、7～25残基のアミノ酸を含むペプチド化合物でもよく、9～20残基のアミノ酸を含むペプチド化合物でもよい。中分子化合物としては、例えば分子量が500 g/mol以上かつ6000 g/mol未満のペプチド化合物であり、分子量が500 g/mol以上かつ4000 g/mol未満のペプチド化合物でもよく、分子量が600 g/mol以上かつ4000 g/mol未満のペプチド化合物でもよく、分子量が700 g/mol以上かつ3000 g/mol未満のペプチド化合物でもよい。

[0034] 本明細書において低分子化合物とは、分子量が500 g/mol未満の化合物である。また、本明細書において、高分子化合物とは、分子量が30000 g/mol以上の化合物である。

[0035] (実験手法の定義)

[特性評価]

本開示の一態様として、分子の配列情報、および、当該分子の特性評価の評価結果情報に基づいて、機械学習を行うことで予測モデルが生成される。分子の特性評価の非限定な一態様として、分子の結合能評価、薬理活性評価、物性評価、動態評価、安全性評価が例示されるがこれらの評価に限定されるものではない。

[0036] ・結合能評価

標的分子結合分子の標的分子への結合能評価の手法は特に限定されないが、標的分子結合分子の標的分子に対する結合を定量評価することで可能である。標的分子は例えば標的タンパクである。標的分子結合分子は例えば抗原結合分子であり、標的分子は例えば抗原である。例えば、標的分子が抗原である場合、抗原結合分子と抗原の結合活性を測定することで評価することが

できる。結合活性 (binding activity) は、分子 (例えば、抗体) の 1 個またはそれ以上の結合部位と、分子の結合パートナー (例えば、抗原) との間の、非共有結合的な相互作用の合計の強度のことをいう。ここで、「結合活性 (binding activity)」は、ある結合対のメンバー (例えば、抗体と抗原) の間の 1 : 1 相互作用に厳密に限定されない。例えば、結合対のメンバーが 1 価での 1 : 1 相互作用を反映する場合、結合活性は固有の結合アフィニティ (単に「アフィニティ」 (affinity) と呼ばれることもある) のことをいう。結合対のメンバーが、1 価での結合および多価での結合の両方が可能である場合、結合活性は、これらの結合力の総和となる。分子 X のそのパートナー Y に対する結合活性は、一般的に、解離定数 (KD) または「単位リガンド量当たりのアナライト結合量」により表すことができる。結合活性は、本明細書に記載のものを含む、当該技術分野において知られた通常の方法によって測定され得る。標的組織特異的な化合物の濃度以外の条件については当業者が適宜決定することが可能である。特定の態様において、本明細書で提供される抗原結合分子は抗体であり、抗体の結合活性 (binding activity) は、 $\leq 1\mu\text{M}$ 、 $\leq 100\text{nM}$ 、 $\leq 10\text{nM}$ 、 $\leq 1\text{nM}$ 、 $\leq 0.1\text{nM}$ 、 $\leq 0.01\text{nM}$ または $\leq 0.001\text{nM}$ (例えば、 10^{-8}M 以下、例えば $10^{-8}\text{M} \sim 10^{-13}\text{M}$ 、例えば $10^{-9}\text{M} \sim 10^{-13}\text{M}$) の解離定数 (KD) である。

[0037] 一態様において、抗体の結合活性 (binding activity) は表面プラズモン共鳴分析法を測定原理とする、例えば、BIAcore (商標登録) T200 または BIAcore (商標登録) 4000 (GE Healthcare, Uppsala, Sweden) を用いたリガンド捕捉法が用いられる。機器操作には、例えば、BIAcore (商標登録) Control Software が用いられる。一態様においてアミンカップリングキット (GE Healthcare, Uppsala, Sweden) を供給元の指示に従って使用し、カルボキシメチルデキストランをコーティングしたセンサーチップ (GE Healthcare, Uppsala, Sweden) にリガンド捕捉用分子、例えば、抗タグ抗体、抗 IgG 抗体、プロテイン A など、を固相化する。リガンド捕捉分子は、適切な pH の 10 mM 酢酸ナトリウム溶液を用いて希釈され、適切な流速および注入時間で注入

される。結合活性測定は0.05%ポリソルベート20（その他の名称としてTween（商標登録）20）含有緩衝液を測定用緩衝液として使用し、流速は10-30 $\mu\text{L}/\text{分}$ 、測定温度は例えば25°Cまたは37°Cで測定される。リガンド捕捉用分子に抗体をリガンドとして捕捉させて測定を実施する場合は、抗体を注入して目的量を捕捉させたのち、測定用緩衝液を用いて調製された抗原およびまたはFc受容体の段階希釈物（アナライト）が注入される。リガンド捕捉用分子に抗原およびまたはFc受容体をリガンドとして捕捉させて測定を実施する場合は、抗原およびまたはFc受容体を注入して目的量を捕捉させたのち、測定用緩衝液を用いて調製された抗体の段階希釈物（アナライト）が注入される。

[0038] 一態様において、測定結果はBIAcore（登録商標）Evaluation Softwareを用いて解析される。速度論的パラメータ（kinetics parameter）算出は1:1 Bindingのモデルを用いて、結合および解離のセンサグラムを同時にフィッティングすることによって実施され、結合速度（ k_{on} もしくは k_a ）、解離速度（ k_{off} もしくは k_d ）、平衡解離定数（KD）が計算され得る。結合活性が弱い、特に解離が早く速度論的パラメータ算出が困難な場合は、Steady stateモデルを用いて平衡解離定数（KD）を計算してもよい。結合活性の他のパラメータとしては、特定の濃度のアナライトの結合量（RU）をリガンドの捕捉量（RU）で除して「単位リガンド量当たりのアナライト結合量」も算出され得る。

[0039] 抗原結合活性の値として、抗原が可溶性分子の場合はKD（解離速度定数）を用いることが可能であり、抗原が膜型分子の場合は見かけの k_d （Apparent dissociation rate constant：見かけの解離速度定数）を用いることが可能である。 k_d （解離速度定数）、および、見かけのKD（見かけの解離速度定数）は、当業者公知の方法で測定することが可能であり、例えば、Biacore（GE healthcare）、フローサイトメーター等を用いることが可能である。

[0040] 特性評価の異なる一態様として、ディスプレイライブラリによる抗原結合分子の選択手法が挙げられる。一態様において、ファージディスプレイを用いたパニングが挙げられる。親和性評価を例に挙げると、複数の異なる抗原

結合分子が提示されたファージライブラリを調製し、標的抗原と調製したファージを接触させた後に、未結合のファージの洗浄操作をすることで、標的抗原と相互作用する抗原結合分子を提示したファージを濃縮することが可能である。濃縮後のファージに含まれる抗原結合分子をコードする核酸配列を解析することで、標的抗原に親和性のある配列を同定することが可能である。また、一態様において、哺乳細胞ディスプレイを用いたパニングが挙げられる。当該ディスプレイシステムを用いた薬理活性評価を例に挙げると、複数の異なる抗原結合分子を含むライブラリを標的とする哺乳細胞に発現させ、それが同じ細胞に示す作用に応じてレポーター活性等を変化させることで、望む薬理活性を有する抗原結合分子遺伝子を有する細胞をフローサイトメーター等で単離することが可能である。また、当該ディスプレイシステムを用いた物性評価を例に挙げると、複数の異なる抗原結合分子を含むライブラリを標的とする哺乳細胞に発現させ、その発現量を抗原結合分子に対して特異的な抗体で染色することで、安定的に高発現できる抗原結合分子遺伝子を有する細胞をフローサイトメーター等で単離することが可能である。パニングによる抗原結合分子の特性評価には、前記ファージまたは哺乳細胞を用いる手法に限られるものではなく、抗原結合分子を提示できる限り様々な手法を利用することが可能であり、例えばリボソームに提示させる手法、mRNAに提示させる手法、ファージ以外のウイルスに提示させる手法、大腸菌等の細菌に提示させる手法等が例示させるがこれらに限定されるものではない。

[0041] 特性評価の異なる一態様として、個体に由来する免疫細胞から抗体遺伝子配列を取得する方法、または血清から抗体蛋白質配列を取得する方法が挙げられる。免疫細胞から抗体遺伝子配列を抽出する親和性評価を例に挙げると、標的抗原蛋白質を個体に投与することで免疫感作を誘導し、標的抗原に対して結合する抗体遺伝子を有する免疫細胞から遺伝子を抽出することで、標的抗原に親和性のある配列を同定することが可能である。

免疫感作を引き起こす抗原には、前記蛋白質を用いる手法に限られるものではなく、当該蛋白質をコードする遺伝子または当該蛋白質を発現する細胞

を利用することが可能である。また、対象とする個体は、ヒト、マウス、ラット、ハムスター、ウサギ、サル、ニワトリ、ラクダ、ラマ、アルパカが挙げられるが、これらに限定されるものではない。

また、前記核酸配列または出現頻度を解析する手法として、各抗原結合分子の核酸配列を有する遺伝子組み換え生物をクローニングし、キャピラリー電気泳動を利用したサンガー法で解析する手法、および次世代シーケンサーを用いて解析する手法が挙げられるが、これらに限定させるものではない。

前記核酸配列を解析する場合において、出現頻度に基づき特性の強弱を判断することも可能である。例えば濃縮後の核酸配列の解析により出現頻度の高い配列がコードする抗原結合分子は当該特性が高く、濃縮後に出現頻度の低い配列がコードする抗原結合分子は、出現頻度が高い配列がコードする抗原結合分子よりも特性が低いと推定することが可能である。

また、前記ディスプレイライブラリまたは個体に由来する抗原結合分子情報を取得する手法は様々な特性評価に応用可能であり、前記に限定されるものではない。

[0042] ・薬理活性評価

分子の薬理活性評価の手法は特に限定されないが、例えば分子が示す中和活性、アゴニスト活性、または細胞傷害活性を測定することで評価することができる。薬理活性評価として細胞傷害活性評価を例に挙げた場合、抗体依存性細胞介在性細胞傷害 (antibody-dependent cell-mediated cytotoxicity : ADCC) 活性、補体依存性細胞傷害 (complement-dependent cytotoxicity : CDC) 活性、T細胞による細胞傷害 (T-cell-dependent cytotoxicity : TDCC) 活性および抗体依存性細胞貪食 (Antibody-Dependent Cellular Phagocytosis : ADCP) 活性等が例示される。CDC活性とは補体系による細胞傷害活性を意味する。またADCC活性とは、標的細胞の細胞膜に発現された膜型分子に結合する抗原結合ドメインを含む抗原結合分子のFc領域に、免疫細胞等が当該免疫細胞に発現したFcγレセプターを介して結合し、当該免疫細胞が標

的細胞に傷害を与える活性を意味する。またT D C C活性とは、標的細胞の細胞膜に発現された膜型分子に結合する抗原結合ドメイン、およびT細胞上のT細胞レセプター（TCR）複合体の構成サブユニットのいずれかに対する抗原結合ドメイン、特にCD3 epsilon鎖に結合する抗原結合ドメインを含むbi-specific抗体を用いることで標的細胞とT細胞を接近させることにより、T細胞が標的細胞に障害を与える活性を意味する。目的の抗原結合分子がADCC活性、C D C活性、TDCC活性またはADCP活性を有するか否かは公知の方法により測定され得る。

また中和活性とは、ウイルス、毒素など、細胞に対して生物学的活性を有するリガンドの当該生物学的活性を阻害する活性をいう。即ち、中和活性を有する物質とは、当該リガンドまたは当該リガンドが結合するレセプターに結合し、当該リガンドとレセプターの結合を阻害する物質をさす。中和活性によりリガンドとの結合を阻止されたレセプターは、当該レセプターを通じた生物学的活性を発揮することができなくなる。抗原結合分子が抗体である場合、このような中和活性を有する抗体は一般に中和抗体と呼ばれ、前記リガンドとレセプターの結合の阻害活性を測定することで中和活性を測定することが可能である。細胞に対して生物学的活性を有するリガンドはウイルスおよび毒素などに限定されず、サイトカイン、ケモカインなどの内在性リガンドがレセプターに結合することで惹起する生理作用を阻害活性も中和活性として理解される。また、中和活性はリガンドとレセプターの結合を阻害する場合に限らず、生物学的活性を有するタンパク質の機能を阻害する活性も中和活性として理解され、前記タンパク質の機能として酵素活性が例示され得る。

[0043] ・物性評価

分子の物性評価の手法は特に限定されないが、物性としては例えば熱安定性、化学安定性、溶解性、粘性、光安定性、長期保存安定性、非特異的吸着性、脂溶性、膜透過性が例示され、例示した種々の物性評価は当業者公知の方法により測定され得る。評価方法は特に限定されないが、例えば熱安定性

、化学安定性、光安定性、機械刺激に対する安定性、長期保存安定性等の安定性評価は、安定性評価の目的とする熱処理、低pH環境への暴露、光暴露、機械による攪拌、長期保存等の処理前と処理後において、当該分子の分解または化学的修飾、会合化を測定することにより評価することができる。そのような安定性評価を実施する測定法の非限定な一態様として、イオン交換クロマトグラフィー法、サイズ排除クロマトグラフィーなどのクロマトグラフィーを用いる手法、質量分析法、電気泳動法が挙げられるがこれらに限定されず、当業者公知の様々な手法により測定することができる。

また、上記以外の物性評価として、ポリエチレングリコール沈殿法によるタンパク質の溶解度の評価、小角X線散乱法による粘性の評価、Extra Cellular Matrix(ECM)への結合評価に基づく非特異的結合評価等が例示されるが、これらに限定されるものではない。

また、物性評価として、タンパク質発現量評価、精製用樹脂または精製用リガンドへの結合評価、表面電荷の評価などについても、当業者公知の手法により測定できる限り評価することが可能である。

[0044] ・動態評価

分子の動態評価の手法は特に限定されないが、マウス、ラット、サル、イヌなどの動物に投与し、投与後の血中の分子の量を経時的に測定することで評価が可能であり、当業者においてPharmacokinetics (PK) 評価として広く公知の手法により評価可能である。PKを直接評価する手法以外に、分子の表面電荷、等電点等をソフトウェア上で計算することにより分子のアミノ酸配列から動態の挙動を予測することも可能である。

[0045] ・安全性評価

分子の安全性評価の手法は特に限定されないが、ISPRI Web-Based Immunogenicity Screening (EpiVax)等の免疫原性予測ツール、抗原結合分子の断片ペプチドのHLA結合評価、MAPPs (MHC-Associated Peptide Proteomics) またはT細胞増殖評価などを用いたT細胞エピトープの検出および免疫原性の評価が挙げられる。また、リウマトイド因子(RF)との結合、PBMCまたはwho

le bloodを用いた免疫反応の評価、血小板凝集評価など当業者公知の手法により測定できる限り評価することが可能である。

[0046] (機械学習に用いられている用語、手法の定義)

・ M B O (Model Based Optimization)

M B O (Model Based Optimization、モデルベース最適化)とは、ある特性の特性値に関して、それ自身を直接の対象として最適化を行うのではなく、何らかのモデルによって推定される、ある特性の特性値の予測値を対象として実施される最適化のことである。

[0047] ・ T P E (Tree-structured Parzen Estimator)

T P E (Tree-structured Parzen Estimator、木構造パルツェン推定器)とは、ベイズ最適化の一種である。T P Eは、最適化しようとする関数について、入力値に対する出力値による条件付き確率と、出力値に対する確率に基づいて、ある入力値に対する出力値の期待改善量を計算する過程を有する。即ち、T P Eは、このようにして計算される期待改善量を、最大にするような入力値を用いて関数の最適化を行う手法のことである。

[0048] (対象物の製造方法)

本開示の一側面では、取得される薬物候補に適した対象物の製造方法は、当業者に周知の方法で実施することができる。

対象物が抗体である場合、例えば、米国特許第4,816,567号に記載されたとおり、抗体は組み換えの方法または構成を用いて製造することができる。本明細書に記載の候補分子化合物である抗体の発現に好適な条件下で、当該抗体をコードする核酸を含む宿主細胞を培養すること、および任意で、当該抗体を宿主細胞（または宿主細胞培養培地）から回収することを含む、抗体を作製する方法が一態様として挙げられる。抗体をコードする、単離された核酸は、抗体のVLを含むアミノ酸配列および／またはVHを含むアミノ酸配列（例えば、抗体の軽鎖および／または重鎖）をコードしてもよい。このような核酸を含む宿主細胞は、(1)抗体のVLを含むアミノ酸配列および抗体のVHを含むアミノ酸配列をコードする核酸を含むベクター、または、(2)抗体のVLを含

むアミノ酸配列をコードする核酸を含む第一のベクターと抗体のVHを含むアミノ酸配列をコードする核酸を含む第二のベクターを含む（例えば、形質転換されている）。一態様において、宿主細胞は、真核性である（例えば、チャイニーズハムスター卵巣（CHO）細胞）またはリンパ系の細胞（例えば、Y0、NS0、Sp2/0細胞））。抗体をコードするベクターのクローニングまたは発現に好適な宿主細胞は、原核細胞または真核細胞を含む。例えば、抗体は、特にグリコシル化およびFcエフェクター機能が必要とされない場合は、細菌で製造してもよい。細菌での抗体断片およびポリペプチドの発現に関して、例えば、米国特許第5,648,237号、第5,789,199号、および第5,840,523号を参照のこと。（加えて、大腸菌における抗体断片の発現について記載したCharlton, *Methods in Molecular Biology*, Vol. 248 (B.K.C. Lo, ed., Humana Press, Totowa, NJ, 2003), pp.245-254も参照のこと。）発現後、抗体は細菌細胞ペーストから可溶性フラクション中に単離されてもよく、またさらに精製することができる。

[0049] 対象物がペプチド化合物、または環状ペプチド化合物である場合、液相合成法、Fmoc合成、Boc合成等を用いた固相合成法、およびこれらの組み合わせ等により製造することができる。液相合成法および固相合成法は当業者に周知の方法で実施することができる。固相合成法とは、化合物を固体に結合させ、その固体の樹脂上で前記化合物と試薬を化学反応させ目的の化合物を合成する方法である。ペプチドの固相合成法は、固体の樹脂に所望のアミノ酸またはペプチドを結合させ、固体の樹脂に結合したアミノ酸またはペプチドに対して更に所望のアミノ酸またはペプチドを順次連結することでペプチド鎖を伸長し、ペプチドを合成する方法である。この固体の樹脂に結合されたペプチドを固体の樹脂から切り離すことで、目的のペプチドを得ることができる。

[0050] （第一実施形態）

図1は、第一実施形態の分子設計装置1を含む創薬システム100の一例を示すブロック図である。

創薬システム１００は、薬物候補に適した新しい対象物を創出するシステムである。本システムにより、特定の生理活性（例えば、特定のタンパク質との結合）のような所定の特性を有する新しい対象物を生成するための方法が提供される。薬物としては、これらに限定されるものではないが、低分子医薬品、中分子医薬品、生物薬剤、細胞、核酸医薬品、バイオ医薬品、または他の活性剤のような潜在的活性剤などが含まれる。対象物は、所望のまたは定義された生物活性（例えば、他のタンパク質よりも優先的に特定のタンパク質と結合する）を有する、分子構造などが含まれる。薬物候補となる分子には生体分子、化合物が含まれ、核酸、ペプチド、環状ペプチド、タンパク質、抗体、標的分子結合分子、高分子化合物、中分子化合物、低分子化合物等の種々の分子を包含する。

本明細書には記載はないが、創薬システム１００は、薬剤のターゲットと相互作用する分子の選定装置、リード分子の創出装置などを含んでもよい。創薬システム１００は、例えば、W02020/246617の開示を含んで構成される情報処理システムであってもよい。

[0051] 創薬システム１００は、分子設計装置１を含んで構成される。分子設計装置１は、所望の特性を有する分子の候補を探索し、特定した候補の情報を出力する。出力される情報は、候補分子のビルディングブロックコンビネーション情報である。言い換えると、分子設計装置１は、候補分子を特定し、特定した候補分子のビルディングブロックコンビネーション情報を出力する。ここで候補分子とは、所望の特性を有すると期待される分子のことである。候補分子ビルディングブロックコンビネーション情報は、候補分子についての情報であって、候補分子の一部または全部のビルディングブロックコンビネーション情報である。出力される候補分子ビルディングブロックコンビネーション情報は、一つの候補分子を示す情報を含んでもよく、複数の候補分子を示す情報を含んでもよい。

[0052] 創薬システム１００は、分子設計装置１より出力された候補分子ビルディングブロックコンビネーション情報を用いて、薬物候補に適した新しい対象

物を選択する。例えば、創薬システム100は、分子設計装置1より出力された候補分子ビルディングブロックコンビネーション情報に基づき候補分子を生成し、実験的に候補分子の特性評価を行い、特性評価の結果に基づき、所望の特性を有する分子を薬物候補に適した新しい対象物として選択する。即ち、創薬システム100は分子設計装置1より出力された候補分子ビルディングブロックコンビネーション情報を用いて、実際に候補分子を生成して行われた特性評価の結果に基づき、薬物候補に適した新しい対象物を創出することができる。このような場合には、候補分子とは、医薬品の主要成分の候補を絞り込むための検証対象となりうる分子のことであるといえる。

[0053] 分子のビルディングブロックコンビネーション情報とは、分子の一部または全部のビルディングブロックの組合せの情報である。ビルディングブロックコンビネーション情報が分子の一部のビルディングブロックの組合せ情報である場合、その配列の範囲は任意に設定可能としてもよい。

ビルディングブロックとは、分子を構成する単位である。分子のビルディングブロックコンビネーション情報は、分子を構成するビルディングブロックの組合せに関する。本願では、個々の構成要素を含んでなる配列を「組合せ」と呼ぶことがある。また、ビルディングブロックコンビネーションの一例として配列という用語を用いることがある。

[0054] 設計される分子は、例えば、タンパク質である。分子がタンパク質である場合、ビルディングブロックはアミノ酸であり、分子のビルディングブロックコンビネーション情報は、例えば、タンパク質のアミノ酸配列の情報である。

設計される分子は、例えば、核酸である。分子が核酸である場合、ビルディングブロックはヌクレオチドである。分子の配列は、例えば、核酸のヌクレオチド配列である。分子ビルディングブロックコンビネーション情報は、ヌクレオチド配列に関する情報である。

より具体的には、設計される分子が抗体である場合、分子配列はアミノ酸配列であり、ビルディングブロックはアミノ酸である。分子のビルディング

ブロックコンビネーション情報は、例えば抗体全長配列であり、例えば、VHもしくはVLのアミノ酸配列であり、またはCDR、FRなどの抗体の一部分の配列である。

また、例えば設計される分子が環状ペプチドである場合、分子配列は非天然アミノ酸を含むアミノ酸配列であり、ビルディングブロックは天然アミノ酸と非天然アミノ酸である。分子のビルディングブロックコンビネーション情報は、非天然アミノ酸を含むアミノ酸配列の情報である。

設計される分子が低分子である場合、分子のビルディングブロックコンビネーションはフラグメントの組合せであり、ビルディングブロックはフラグメント（低分子を構成する断片の分子）である。

また、例えば、設計される分子が核酸である場合、分子配列は塩基配列であり、ビルディングブロックは塩基である。

[0055] 所望の特性とは、薬物候補に適した新しい対象物に求められる特性であり、任意に設定可能である。特性の非限定な一態様として、所定の生体内ターゲットへの結合能、結合能、薬理活性、物性、動態、安全性が例示されるが、これらに限定されるものではない。例えば、設計される分子が抗体であれば、特性は、例えば、所定の抗原に対する薬剤の結合能である。例えば、設計される分子がmRNA（messenger-RNA）であれば、特性は、例えば、タンパク質の翻訳力である。

[0056] 分子設計装置1は、後述する予測情報から、所望の特性を有する分子を推論する推論装置の一例である。所望の特性とは、例えば、標的分子への結合能があること、効能があること、膜透過性を含む薬らしさがあることなどが挙げられる。以下、予測情報から所望の特性を有する分子を推論することを、候補分子を特定すると言い換えてもよい。

本実施形態において、所望の特性を有すると期待される分子とは、所望の特性に対して予測モデルによる良い予測値と低い予測不確かさを示す分子のことである。

[0057] 以下、ビルディングブロックコンビネーション情報として配列情報を処理

する場合について説明する。

分子設計装置 1 は、例えば、配列情報処理部 1 1 1 a、特性予測部 1 1 1 b、予測情報処理部 1 1 1 c および候補分子特定部 1 1 1 d を備える推論部 1 1 1 を有する。

配列情報処理部 1 1 1 a は配列情報集合を用意し、用意した配列情報集合を特性予測部 1 1 1 b に出力する。配列情報集合とは、複数の分子の配列情報の集合である。配列情報処理部 1 1 1 a は、自律的に分子ごとの配列情報を生成してもよいし、他機器から分子ごとの配列情報を入力してもよい。なお、配列情報集合はビルディングブロックコンビネーション情報集合と言い換えてもよく、配列情報処理部 1 1 1 a はビルディングブロックコンビネーション情報処理部と言い換えてもよい。

特性予測部 1 1 1 b は、配列情報処理部 1 1 1 a から入力される配列情報集合の各要素である分子の配列情報について、その分子の特性を予測し、予測した特性に係る予測情報を予測情報処理部 1 1 1 c に出力する。

予測情報処理部 1 1 1 c は、特性予測部 1 1 1 b より分子ごとの予測情報を取得する。予測情報処理部 1 1 1 c は、取得した予測情報を候補分子特定部 1 1 1 d に出力する。

候補分子特定部 1 1 1 d は、予測情報に基づき、少なくとも一つの候補分子の配列情報を特定する。候補分子特定部 1 1 1 d は、特定した配列情報を示す出力データを他の機器に出力してもよいし、記憶部 1 4（後述）に記憶してもよい。

[0058] 本実施形態において、配列情報集合は機械学習により生成された仮想的な分子配列（「仮想配列」と呼ばれることがある）を示す配列情報の集合であってもよく、現実の分子配列（「実配列」と呼ばれることがある）の配列情報の集合であってもよく、仮想配列を示す配列情報と実配列を示す配列情報のどちらをも含む集合であってもよい。例えば、配列情報集合は、仮想配列生成モデルにより生成された仮想配列の配列情報を有してもよい。また、配列情報集合は、例えば、既存のデータベースや実験結果として得られた実配

列を示す配列情報を有してもよい。後述する組合せ最適化により、ビルディングブロックの組合せの全候補から抽出された配列情報であってもよい。

配列情報処理部 111a は、配列情報集合を用意する処理を行う。配列情報処理部 111a は、機械学習モデルを用いて、ある入力情報に対する出力となる配列情報を取得するための取得プロセスを実行する配列情報集合取得部（図示せず）をさらに備えてもよい。この場合、具体的には、配列情報処理部 111a は、配列情報集合を取得し、取得した配列情報集合を特性予測部 111b に出力する。

[0059] 本実施形態において、特性予測部 111b には配列情報処理部 111a から複数の分子の配列情報が入力される。

特性予測部 111b は、入力された複数の分子の配列情報ごとに、特性予測値および予測の不確かさ（uncertainty）の推定値を演算する。

本実施形態において、特性予測部 111b は、予測値算出部 111x と、予測の不確かさの推定部 111y とを備える。予測値算出部 111x は、分子の配列情報を入力し、予測モデルを用いて特性予測値を算出する。特性予測値は予測モデルを用いて予測された特性値である。予測の不確かさの推定部 111y は、予測値算出部 111x により算出された特性予測値の不確かさを推定する。推定された不確かさを示す値を、予測の不確かさの推定値と呼ぶ。

予測モデルは、例えば、個々の分子の配列情報および該分子についての特性評価を複数セット有する訓練データに基づく学習により生成される。

予測値算出部 111x は、複数の特性についての特性予測値を算出してもよい。複数の特性について特性値が予測される場合、予測の不確かさの推定部 111y は、それぞれの特性の特性予測値について予測の不確かさを推定してもよく、いずれか一つの特性予測値について予測の不確かさを推定してもよい。

特性予測部 111b は、分子ごとの予測情報として、特性予測値と、該特性予測値についての予測の不確かさの推定値を予測情報処理部 111c に出

力する。

[0060] 候補分子特定部 1 1 1 d は、予測情報に基づいて所望の特性を有する分子を推論する処理を行い、候補分子を特定する。

本実施形態において、候補分子特定部 1 1 1 d は、特性予測部 1 1 1 b より出力された各々の分子の配列情報についての特性予測値と予測の不確かさの推定値に基づき、所望の特性を有する分子を推論する。候補分子特定部 1 1 1 d は、予測情報処理部 1 1 1 c から得られた予測情報に基づいて、配列情報処理部 1 1 1 a からの配列情報集合に配列情報が含まれる複数の分子から、少なくとも一つの候補分子を特定する。候補分子特定部 1 1 1 d は、候補分子の特定において、少なくとも一つの特性についての特性予測値を用いてもよいし、複数の特性についての特性予測値を用いてもよい。候補分子特定部 1 1 1 d は、候補分子の特定において、予測情報として、複数の特性予測値と、少なくとも一つの特性予測値についての予測の不確かさの推定値と、を用いてもよい。すなわち、候補分子の特定において、第一特性値（特性予測値）の最適化を図るにあたり、第二特性値（特性予測値についての予測の不確かさ）の制約を考慮してもよい。また、候補分子特定部 1 1 1 d は、後述する特性品質値に基づいて所望の特性を有する分子を候補分子として特定してもよい。

候補分子特定部 1 1 1 d は、例えば、特性予測値と予測の不確かさとがそれぞれ所定の条件を満たす少なくとも一つの分子を候補分子として選択してもよい。候補分子特定部 1 1 1 d は、複数の特性予測値と、少なくとも一つの特性予測値についての予測の不確かさの推定値とがそれぞれ所定の条件を満たす少なくとも一つの分子を候補分子として選択してもよい。

[0061] 候補分子特定部 1 1 1 d は、特性予測値と予測の不確かさの推定値とに基づき算出された特性品質値に基づいて、少なくとも一つの候補分子を特定してもよい。この場合、特性品質値は、例えば、予測情報処理部 1 1 1 c において算出されてもよいし、特性予測部 1 1 1 b において算出されてもよい。

特性品質値は、特性予測値と予測の不確かさの推定値に基づく値である。

特性品質値は、特性予測値と予測の不確かさの推定値とを説明変数とし所定の関数を用いて算出される目的変数とみなすこともできる。

例えば、特性品質値は、特性予測の不確かさの推定値が大きいほど小さな値を与え、特性予測値が大きいほど大きな値を与える指標値であってもよい。言い換えれば、特性品質値は、特性予測値の増加に応じて増加し、予測の不確かさの推定値の増加に応じて減少するものであればよい。一例として、特性品質値は、線形領域における特性予測値と予測の不確かさの推定値の所定の係数倍との差、あるいは、特性予測値の所定の係数倍と予測の不確かさの推定値との差である。所定の係数は、正の定数であり、特性品質値に対する特性値または予測の不確かさの推定値の寄与の度合いを示す。

したがって、所望の特性を有すると予測されるほど、もしくは、その特性の度合いが高いほど、大きい値を与える特性予測値が定義される場合、特性品質値は、特性予測値の値が上がり、予測の不確かさの推定値が小さくなるほど大きな値を示す指標である。この場合、特性品質値が大きいほど、実際に生成される分子が期待される特性を示す可能性が高いことを意味する。即ち、分子設計において特性品質値を用いた推定処理はリスク回避型の推定処理を実現するといえる。

[0062] なお、特性品質値は、線形領域における特性予測値と予測の不確かさの推定値の所定の係数倍との差分である必要はない。特性品質値は、特性予測値および予測の不確かさの推定値のいずれか一方を所定の回数累乗し、他方の値との差、あるいは、商を算出した値であってもよい。また、所望の特性を有する可能性もしくは、その特性の程度が高いほど小さい値を与える予測値が定義される場合、特性予測値は、特性予測値の予測の不確かさの推定値による除算もしくは、これらの対数領域での演算などの他種の演算で定められてもよい。特性品質値の算出には、予測値とその不確かさの両者を用いて導出されればよく、その関数、手順、などについては限定されない。

[0063] 候補分子特定部 111d は、算出された特性品質値に基づいて、所望の特性を有する分子を特定する。一例として、候補分子特定部 111d は、特性

予測部 1 1 1 b において特性予測値が得られた分子の中で、特性品質値の良い値を与える分子ほど、所望の特性を有する分子として特定される可能性が高くなる。ここで、「特性品質値が良い」とは、所望の特性を有すると予測される可能性もしくは、その特性の程度が高いことを示す特性品質値が得られることを意味する。

他の例として、候補分子特定部 1 1 1 d は、特性予測部 1 1 1 b において特性予測値が算出された分子の中で、特性品質値が所定の条件を満たす分子を所望の特性を有する分子として特定してもよい。さらに他の例として、候補分子特定部 1 1 1 d は、特性予測部 1 1 1 b において特性予測値が算出された複数の分子を、特性品質値を用いて順位づけして並び替え、並び替えた分子から所定の範囲内の分子を所望の特性を有する分子として特定してもよい。ここで、候補分子特定部 1 1 1 d は、より良好な特性品質値を有する分子ほど優先されるように一定数の複数の分子を所望の特性を有する分子として特定してもよい。また、候補分子特定部 1 1 1 d は、特性予測値と予測の不確かさの推定値がそれぞれ所定の条件を満たす分子を所望の特性を有する分子として特定してもよい。

[0064] 特性予測部 1 1 1 b により算出される特性予測値の精度が分子の特定に影響する。特性予測モデルの訓練データによっては、配列情報に基づく特性予測の精度に差が生じる。ここで、予測の精度は、予測の信頼性、または予測の確かさと言い換えられてもよい。予測の精度が低いとは、予測の信頼性が低い、予測の確かさが低い、また予測の不確かさが高い、と言い換えられる。また、予測の精度が高いとは、予測の信頼性が高い、予測の確かさが高い、または、予測の不確かさが低いと言い換えられる。

本実施形態の分子設計装置 1 では、所望の特性を有する分子の推論処理に、特性予測値だけではなく予測の不確かさの推定値をも指標としている。即ち、本実施形態の分子設計装置 1 では、候補分子を特定する際に、候補分子特定部が使用する対象として、特性値を予測する数理モデルではなく、特性値の予測と予測の不確かさの推定を行う数理モデルを用いる。このことによ

り、特性予測値が高くても、その予測の確からしさが低い分子を推論処理により候補分子として特定することを防ぐことができる。即ち、実際に分子を生成した際に、期待されるほどの特性を有しない可能性が高い分子を特定することが防止される。このことにより、本実施形態により、創薬に要する負担をより軽減することができる。

[0065] (第二実施形態)

図2は、第二実施形態の分子設計装置1を含む創薬システム100の一例を示すブロック図である。以下の説明では、第一実施形態との差異点を主とする。特に断らない限り、第一実施形態との共通点については、その説明を援用する。

[0066] 本実施形態に係る分子設計装置1は、後述する予測情報から、所望の特性を有すると期待される分子を推論する推論装置の一例である。

分子設計装置1は、例えば、配列情報処理部111a、特性予測部111b、予測情報処理部111cおよび候補分子特定部111dを備える推論部111を有する。

[0067] 本実施形態に係る推論部111は、組合せ最適化アルゴリズム（本願では、単に「組合せ最適化」と呼ぶことがある）を実行し、所望の特性を有すると期待される分子を推論する。

配列情報処理部111aは、組合せ最適化を実行することにより、予め設定された探索空間から、任意の数の分子の配列情報を抽出する。即ち、配列情報処理部111aによれば、組合せ最適化により、探索空間におけるビルディングブロックの配列の候補から所望の特性を有すると期待される一部の分子の配列情報を、更新された抽出パラメータを用いて抽出する処理を繰り返す。探索空間は、任意に設定されうる。探索空間は、例えば、特性予測に用いられる予測モデルの訓練に用いられる訓練データに基づいて設定されてもよい。配列情報取得回数、即ち、配列情報取得の繰り返し回数は、1以上の任意の自然数であればよい。推論部111には、配列情報取得回数が予め設定されてもよい。推論部111は、配列情報処理部111aが現実配列

情報を抽出した回数を繰り返し回数として計数してもよい。本実施形態において、探索空間に含まれる分子の配列情報は、配列情報集合の要素になりえることを意味する。

[0068] 配列情報処理部 1 1 1 a は、抽出した複数の分子の配列情報を特性予測部 1 1 1 b に出力する。

本実施形態に係る特性予測部 1 1 1 b は、予測モデルを用い、配列情報処理部 1 1 1 a から入力される分子ごとの配列情報に基づいて、その分子の特性を示す予測情報を生成し、分子ごとに生成した予測情報を予測情報処理部 1 1 1 c に出力する。予測モデルは、例えば、分子ごとの配列情報および該分子についての特性評価結果を複数セット含んで構成される訓練データに基づく学習により生成される。

分子ごとの予測情報には、例えば、分子ごとの特性予測値が含まれる。そこで、特性予測部 1 1 1 b は、分子ごとの配列情報から特性予測値を算出する予測値算出部 1 1 1 x を備えてもよい。

分子ごとの予測情報には、例えば、分子ごとの特性予測値および該特性予測値の不確かさの推定値が含まれてもよい。特性予測部 1 1 1 b は、特性予測値の不確かさの推定値を算出する予測の不確かさの推定部 1 1 1 y を含んで構成されてもよい。よって、特性予測部 1 1 1 b より出力される分子ごとの予測情報には、分子ごとの特性予測値の他、該特性予測値の不確かさの推定値がいずれも含まれうる。

[0069] 予測情報処理部 1 1 1 c は、特性予測部 1 1 1 b から分子ごとの予測情報を取得する。予測情報が特性予測値および特性予測値の不確かさの評価値を含む場合、予測情報処理部 1 1 1 c は、前記配列情報集合に含まれる各々の分子の配列情報について分子ごとの特性予測値と特性予測値の不確かさとに基づいて特性品質値を算出してもよい。本実施形態では、配列情報処理部 1 1 1 a は、予測情報処理部 1 1 1 c により算出された特性品質値を予測情報として抽出パラメータの更新に用いる。抽出パラメータの具体例については後述する。

計数された繰り返し回数が予め設定された配列情報取得回数に満たない場合、配列情報処理部 1 1 1 a は、取得した予測情報に基づいて抽出パラメータを更新する。従って、組合せ最適化において、配列情報処理部 1 1 1 a における更新後の抽出パラメータに基づく配列情報の抽出、特性予測部 1 1 1 b による抽出された配列情報に基づく予測情報の取得、および、配列情報処理部 1 1 1 a による抽出パラメータの更新を含む一連の処理が繰り返される。

[0070] 計数された繰り返し回数が予め設定された配列情報取得回数に達した場合には、配列情報処理部 1 1 1 a は、組合せ最適化アルゴリズムの処理を停止し、予測情報処理部 1 1 1 c は、特性予測部 1 1 1 b から入力された分子ごとの予測情報を候補分子特定部 1 1 1 d に出力する。候補分子特定部 1 1 1 d は、予測情報処理部 1 1 1 c から入力される分子ごとの予測情報に基づき、各回における配列情報集合から少なくとも一つの候補分子を特定する。配列情報取得回数が N 回である場合には、予測情報処理部 1 1 1 c は、N 回分の抽出された分子の配列情報から、少なくとも一つの候補分子の配列情報を特定する。また、候補分子特定部 1 1 1 d は、N 回分の抽出された分子の配列情報集合から、少なくとも一つの候補分子の配列情報を特定してもよい。なお、予測情報に特性品質値が含まれる場合には、候補分子特定部 1 1 1 d は、分子ごとの特性品質値に基づいて候補分子を特定することができる。

候補分子特定部 1 1 1 d は、特定された候補分子の配列情報を出力する。

[0071] 本実施形態では、組合せ最適化において、これに限定されないが、ベイズ最適化を適用することができる。ベイズ最適化の手法として、例えば、TPE (Tree-Structured Parzen Estimator)、進化的アルゴリズムの手法として、例えば、NSGA II (Elitist Non-dominated Sorting Genetic Algorithm) 等のいずれのアルゴリズムが用いられてもよい。

[0072] 本実施形態においては、組合せ最適化を用いることで、探索空間に含まれる全ての分子の配列情報のそれぞれについて特性値を予測することなく、所望の特性を有する分子を推論することができる。仮に膨大な分子をいずれも

処理対象として扱う場合、莫大な計算時間がかかるが、本実施形態のように探索空間に対して組合せ最適化を実施することで、すべての分子を扱うことなく、所望の特性を有する分子を推論することができる。このことにより、創薬に要する負担をより軽減することができる。

[0073] 本実施形態において、分子設計装置 1 は、組合せ最適化の目的関数として、配列情報から特性値を推定するための数理モデルを用いてもよく、特性品質値を推定するための数理モデルを用いてもよい。そして、その数理モデルの出力が組合せ最適化により最適化される。特性品質値を推定する数理モデルを用いる場合には、特性予測値で指示される特性が好適であってもその予測の確からしさが低い分子を、候補分子として特定することを防ぐことができる。即ち、実際に分子を生成した際に期待されるほどの特性を有しない可能性が高い分子を、候補分子として特定する可能性を低減することができる。このことにより、創薬に要する負担をより軽減することができる。

[0074] 図 3 は、第一実施形態および第二実施形態における配列情報集合の一例を説明する説明図である。図 3 の“配列 ID”は、配列情報集合に含まれる各分子の配列情報を識別する識別子である。図 3 は、配列情報集合に含まれる各分子について配列情報を示す。各分子の配列情報は、複数のビルディングブロックとビルディングブロックごとの配置を示す。個々のビルディングブロックの配置は、情報 D 1 0 1 から情報 D 1 0 2 までの間に示される。例えば、情報 D 1 0 1 は、各分子の配列のポジション H 1 に係るビルディングブロックの情報である。V H L 0 0 0 1、V H L 0 0 0 2、V H L 0 0 0 3、V H L 0 0 0 4、などのいずれも、ポジション H 1 におけるビルディングブロックが M である分子を示す。

[0075] 図 4 は、第一実施形態および第二実施形態に係る分子設計装置 1 のハードウェア構成の一例を示す図である。分子設計装置 1 は、バスで相互に接続された CPU 等のプロセッサ 9 1 とメモリ 9 2 とを備える制御部 1 1 を備え、所定のプログラムを実行するコンピュータシステムを有する。コンピュータシステムは、プログラムの実行によって制御部 1 1、入力部 1 2、通信部 1

3、記憶部14および出力部15を備える分子設計装置1として機能するとみなすこともできる。本願では、「プログラムの実行」もしくは「プログラムを実行する」とは、プログラムに記述された1以上の指令のそれぞれにより指示される処理を実行するとの意味を含む。

[0076] より具体的には、プロセッサ91が記憶部14に記憶されているプログラムを読み出し、読み出したプログラムをメモリ92に記憶させる。プロセッサ91が、メモリ92に記憶させたプログラムを実行することによって、制御部11、入力部12、通信部13、記憶部14および出力部15を備える分子設計装置1として機能する。プログラムは、例えば記憶部14に予め記憶されている。

[0077] 制御部11は、分子設計装置1が備える各種機能部の動作を制御する。制御部11は、例えば、所望の特性を有する分子を推論する推論処理を実行する推論部111の機能を備える。

[0078] 入力部12は、マウス、キーボード、タッチパネル等の入力装置を含んで構成される。入力部12は、これらの入力装置を分子設計装置1に接続するインタフェースとして構成されてもよい。入力部12は、分子設計装置1に対する各種情報の入力を受け付ける。入力部12には、例えば、予測モデルの訓練データが入力される。

[0079] 通信部13は、分子設計装置1を外部装置に接続するための通信インタフェースを含んで構成される。通信部13は、有線または無線を介して外部装置と通信する。外部装置は例えば、予測モデルの訓練データの送信元の装置である。外部装置は、候補分子配列情報の送信先とする装置であってもよい。なお、分子設計装置1は、インターネットに接続され、1個または複数個の端末装置のそれぞれと相互に通信可能なサーバ装置として実現されてもよい。

[0080] 記憶部14は、磁気ハードディスク装置、半導体記憶装置などのコンピュータ読み出し可能な記憶媒体装置 (non-transitory computer-readable recording medium) を用

いて構成される。記憶部 1 4 は分子設計装置 1 に関する各種情報を記憶する。記憶部 1 4 は、例えば入力部 1 2 または通信部 1 3 を介して入力された情報を記憶する。記憶部 1 4 は、例えば制御部 1 1 の実行する処理で生じた各種情報を記憶する。記憶部 1 4 は、例えば、予測モデル、配列情報、組合せ最適化に用いる各種のパラメータ（上記の抽出パラメータが含まれる）などを記憶する。記憶部 1 4 は、例えば、上述したプログラムを記憶する。

[0081] 出力部 1 5 は、各種情報を出力する。出力部 1 5 は、例えば C R T (C a t h o d e R a y T u b e) ディスプレイ、液晶ディスプレイ、有機 E L (E l e c t r o - L u m i n e s c e n c e) ディスプレイ等の表示装置を含んで構成される。出力部 1 5 は、これらの表示装置を分子設計装置 1 に接続するインタフェースとして構成されてもよい。出力部 1 5 は、例えば入力部 1 2 または通信部 1 3 に入力された情報を出力する。出力部 1 5 は、例えば制御部 1 1 の実行する処理で生じた各種情報を出力してもよい。

[0082] 図 5 は、第一実施形態および第二実施形態のそれぞれに係る制御部 1 1 の構成の一例を示す図である。制御部 1 1 は、推論部 1 1 1、入力制御部 1 1 2、通信制御部 1 1 3、記憶制御部 1 1 4 および出力制御部 1 1 5 を備える。推論部 1 1 1 は、上記の推論処理を実行する。入力制御部 1 1 2 は、入力部 1 2 の動作を制御する。通信制御部 1 1 3 は、通信部 1 3 の動作を制御する。記憶制御部 1 1 4 は、記憶部 1 4 の動作を制御する。出力制御部 1 1 5 は、出力部 1 5 の動作を制御する。

[0083] 次に、各実施形態に係る推論処理の例について説明する。図 6 は、第一実施形態における分子設計装置 1 が実行する処理の流れの一例を示すフローチャートである。

配列情報処理部 1 1 1 a は、複数の分子の配列情報を含む配列情報集合を取得する（ステップ S 1 0 1）。

特性予測部 1 1 1 b は、配列情報集合をなす各分子について、特性予測値を算出し、特性値の不確かさの推定値を算出する（ステップ S 1 0 2）。

候補分子特定部 1 1 1 d は、分子ごとに算出した特性予測値および予測の

不確かさの推定値に基づき、候補分子を特定する（ステップS 1 0 3）。特性予測部 1 1 1 b が予測の不確かさの推定値は不確かさの定量化(Uncertainty Quantification)することができれば、その手法は限定されない。例えば、予測の不確かさの推定値として、特性予測値の標準偏差が算出されてもよい。また、不確かさの定量化において、コンフォーマル予測 (Conformal Prediction) が用いられてもよい。特性予測部 1 1 1 b は、学習データを複数の部分に分割して、部分ごとに予測に対するエラー（予測誤差）を計算し、そのエラーの分布に基づいて不確かさを定量化してもよい。

候補分子特定部 1 1 1 d は、特定した候補分子の配列情報を候補分子配列情報として出力部 1 5 を用いて出力する（ステップS 1 0 4）。その後、制御部 1 1 は、図 6 の処理を終了する。

[0084] なお、予測情報処理部 1 1 1 c は、ステップS 1 0 3 に先行して、ステップS 1 0 2 において得られた特性予測値と予測の不確かさの推定値とに基づいて特性品質値を算出してもよい。予測情報処理部 1 1 1 c は、特性品質値の一例として平均分散 (MV : Mean Variance) を算出してもよい。MV は、リスク耐性パラメータ ρ と所定の予測モデルによる特性予測値 $f'(x)$ との積からペナルティ関数 $g(x)$ を差し引いて得られる ($MV = \rho f'(x) - g(x)$)。リスク耐性パラメータは、特性予測値 $f'(x)$ の不確かさ $g(x)$ に対する耐性、つまり、予測値 $f'(x)$ の信頼度を示す正の実数値である。予測情報処理部 1 1 1 c は、不確かさの推定値 $g(x)$ の一例として標準偏差を算出する。なお、MV は、金融工学においてポートフォリオ最適化に応用されることがある。金融工学では、MV は、平均とする期待報酬とリスクとしての報酬の分散を総合した指標として用いられることがある。MV は、最適化アルゴリズムに関わらず種々の分野、目的で利用可能な指標である（例えば、Q. Zhu and V. Y. F. Tan: Thompson Sampling Algorithms for Mean-Variance Bandits (2020 ICML), S. Takeuchi: Distributionally-Aware Kernelized Bandit Problems for Risk Aversion (2022 ICML) 参照）。

[0085] 予測情報処理部 111c は、特性品質値の一例として不確かさ $g(x)$ を制約条件として用いて、計算を実行してもよい。例えば、予測情報処理部 111c は或る不確かさ $g(x)$ の閾値 τ を設け、不確かさ $g(x)$ が τ 以下である場合には特性品質値を特性予測値 $f'(x)$ に設定し、不確かさ $g(x)$ が τ より大きい場合には特性品質値を十分小さい値（例えば $f'(x)$ が取りうる最小値）に設定してもよい。このような設定は式（１）により表される。

[数1]

$$\text{特性品質値} = \begin{cases} f'(x) & \text{if } g(x) \leq \tau \\ \min f'(x) & \text{if } g(x) > \tau \end{cases} \quad \dots(1)$$

[0086] 予測情報処理部 111c は複数の物性値に基づいて特性品質値を算出してもよい。例えば、予測情報処理部 111c は第一物性値 $g_1(x)$ の閾値 τ_1 と、第二物性値 $g_2(x)$ の閾値 τ_2 とを設ける。第一物性値 $g_1(x)$ が τ_1 より小さく、かつ、第二物性値 $g_2(x)$ が τ_2 より小さい場合には、予測情報処理部 111c は特性品質値を特性予測値 $f'(x)$ に設定してもよい。一方、第一物性値 $g_1(x)$ が τ_1 以上、または、第二物性値 $g_2(x)$ が τ_2 以上の場合には、予測情報処理部 111c は特性品質値を十分小さい値（例えば $f'(x)$ が取りうる最小値）に設定してもよい。

[0087] あるいは、予測情報処理部 111c は、制約を満たす確率の予測値 $Pr(x \in F | x)$ に基づいて特性品質値を算出してもよい。すなわち、予測情報処理部 111c は、予測モデルによる特性予測値 $f'(x)$ に、制約を満たす確率の予測値を乗じて、制約を満たす確率の予測が高い配列情報のみをサンプリングしてもよい。この例では、制約を満たす確率が低い場合には特性品質値は 0 に近づき、制約を満たす確率が高いほど（１に近いほど）特性品質値は目的指標の値そのものに近づく。このような特性品質値は式（２）により表される。

[数2]

$$\text{特性品質値} = f'(x) \cdot Pr(x \in F | x) \quad \dots(2)$$

[0088] なお、制約の個数は1でもよいし2以上でもよい。例えば、複数の物性値に基づく特性品質値を算出する場合には、予測情報処理部111cは、各物性値についての制約条件を設定し、すべての条件を満たす確率を、予測モデルによる特性予測値 $f'(x)$ に掛けてもよい。この場合に、すべての条件を満たす確率は、各物性値についての制約条件を満たす確率の積であってもよい。

[0089] 図7は、第二実施形態における分子設計装置1が実行する処理の流れの一例を示すフローチャートである。

配列情報処理部111aは、入力値 x に対する測定値 y の組であるデータペアを複数個含むデータセット D を訓練データとして取得する（ステップS201）。例えば、記憶部14には、入力部12または通信部13から入力されたデータセット D を予め記憶しておく。配列情報処理部111aは、記憶部14からデータセット D を読み出す。

配列情報処理部111aは、取得した訓練データを用いて、例えば、ガウス過程回帰を実行して入力値に対する測定値を予測するための予測モデルを学習する（ステップS202）。

推論部111は、組合せ最適化を実行し、候補分子を特定する（ステップS203）。候補分子特定部111dは、特定した候補分子の配列情報を、出力部15を用いて出力する（ステップS204）。その後、図7の処理を終了する。

[0090] <MBOへの適用例>

次に、MBOへの適用例について説明する。

本適用例においては、in-silico薬剤設計の方法の一種であるオフラインMB0とビルディングブロックベースの分子設計とが組合せられて適用される。

オフラインMB0は、獲得されたデータから生成された「代理」予測モデル内で最適な分子を探索する方法である。当該方法は、ブラックボックスとして代表モデルを扱うブラックボックス最適化に係り、逆解析としても知られる。ここでは、MBOは、予め蓄積された実験データを訓練データとして用い

て予測モデルを学習しておき、学習により得られた予測モデルを用いて得られる特性値に基づいて薬剤候補分子を評価し、評価結果に基づいて予測モデルを更新する手法である。下記の説明では、予測モデルを代表モデル (proxy model) と呼ぶことがある。

[0091] 本適用例において離散入力値がMBOの最適化対象になりうる。MBOでは、離散入力値 $x \in X$ に対する測定値 y が未知の関数 $f(x)$ として与えられ、予め離散入力値 x に対する測定値 y の組を n 個 (n は、2以上の整数) 有するデータセット D が与えられ、さらに追加の入力値 x に対する測定値 y を取得できないことを前提とする。離散入力値 x に対する測定値 y は、関数 $f(x)$ の正解 (オラクル (oracle) とも呼ばれる) を与える。一般的なMBOは、離散値空間 X において関数 $f(x)$ が最大となる入力値 x を求めることを目的とする。上記の手法は、関数 $f(x)$ に代え、特性品質値 $f(x) - \lambda g(x)$ が最大化する入力値 x を求める問題と捉えることができる。ここで、入力値 x として配列が適用され、関数 $f(x)$ は、特性値に相当する実数値を予測するための関数に相当する。ペナルティ関数 $g(x)$ は、関数 $f(x)$ の値の不確かさに相当する実数値を与える関数である。 λ は、上記の係数 α に相当する。入力値 x は、ベクトル表現を用いて表される。

[0092] オフラインMBOでは、配列情報処理部111aは、離散入力値 x に対する測定値 y の組を複数個有するデータセット D を訓練データとして、予測モデルを予め学習させておく。配列情報処理部111aは、予測モデルの学習において、離散入力値 x に対する目的関数の関数値を予測値 y' とし、予測値 y' とその離散入力値 x に対する測定値 y との差分の大きさを表す指標値を最小化するように、予測モデルのパラメータを定める。本適用例では、以下に説明するように、推論部111は、離散入力値 x をサンプリングし、サンプルごとに予測モデルを用いて得られる予測結果に基づいて組合せ最適化のパラメータを更新する。

[0093] 推論部111は、組合せ最適化において、例えば、上記のTPEを用いることができる。TPEは、ベイズ最適化にパルツェン窓密度推定を組合せて

適用したブラックボックス最適化アルゴリズムである。TPEは、カテゴリパラメータを取り扱えるため、ビルディングブロックベースの分子設計に応用することを可能とする。

[0094] 本適用例では、推論部111は、TPEの機能を実装し、分子設計において次のように応用される。TPEは、目的関数の期待改善量（EI: Expected Improvement）の最大化を志向する手法である。まず、配列情報処理部111aは、薬剤候補分子をなすビルディングブロックコンビネーションを探索空間に適用し、予測モデルからの出力を探索空間からサンプリングされる入力分子集合Xに対応する目的スコア集合Yに割り当てる。例えば、配列情報処理部111aは、予測モデルにより得られる特性予測値をTPEの目的スコアとして設定する。配列情報処理部111aは、探索空間から入力分子集合Xをサンプリングする、すなわちビルディングブロックコンビネーションにおける改変位置（modification position）ごとにビルディングブロックの候補を割り当てる。例えば、薬剤候補分子がタンパク質であった場合には、タンパク質の配列における改変位置ごとに候補アミノ酸が割り当てられる。一例として、TPEサンプラを実行するためにオプチュナ（Optuna）実装を用いることができる。

[0095] 次に、特性予測部111bは、サンプリングにより定めた入力値に対して予測モデルを用いて目的関数の推定値を算出する（評価）。

予測情報処理部111cは、選択した入力値と算出した推定値を配列情報処理部111aに出力し、配列情報処理部111aは、予測情報処理部111cから入力された入力値と推定値を用いて、最適化アルゴリズムのパラメータを更新する（更新）。ここで、配列情報処理部111aは、選択した入力値と算出した推定値を、それぞれ探索空間からサンプリングされる入力分子集合X、目的スコアの集合Yに割り当てる。推論部111は、データ分割、サンプリング、評価、および、更新の処理を繰り返すことで、サンプリングにより得られた入力値（サンプル値）の評価結果が予測モデルに反映され、目的関数の推定値がより大きいビルディングブロックの組合せが確率的に

探索される。

[0096] 本適用例に係る分子設計装置 1 が実行する組合せ最適化処理の流れの一例を図 8 に示す。

推論部 111 は、配列情報取得回数の初期値 0 を設定する。ループ R 20 は、ステップ S 203 a ~ S 203 c の処理を含む。制御部 11 は、繰り返し回数が所定のサンプリング回数 N 以下である場合を、ループ R 20 の実行条件とする。サンプリング回数は、上記の配列情報取得回数に相当する。

[0097] 配列情報処理部 111 a は、組合せ最適化のアルゴリズムに従って、複数の分子の配列情報を取得する（ステップ S 203 a）。配列情報処理部 111 a は、例えば、組合せ最適化のアルゴリズムとして T P E を用いる場合、各入力値に対して予測モデルを用いて得られる推定値（上記の特性予測値に相当）と所定の閾値 γ とに基づいて、入力値の集合を 2 つのセットに分割する。一方のセット（「第 1 セット」と呼ぶ）は、推定値が閾値 γ 以上となる入力値を含んで構成される。他方のセット（「第 2 セット」と呼ぶ）は、推定値が閾値 γ 未満となる入力値を含んで構成される。配列情報処理部 111 a は、目的関数の期待改善量を最大化する入力値をサンプリングする。期待改善量は、更新前後の目的関数の期待値の増加量に相当し、入力値 x に対して第 1 セットに属する $p(x | y_1) / p(x | y_2)$ に比例することが知られている。 $p(x | y_1)$ は、第 1 セットに対する入力値 x の密度分布を示す。 $p(x | y_2)$ は、第 2 セットに対する入力値 x の密度分布を示す。即ち、E 1 が抽出パラメータの一例となる。配列情報処理部 111 a は、算出した E 1 を最大化する入力値をサンプリングする。

特性予測部 111 b は、予測モデルにより配列評価を行う。ここで、特性予測部 111 b は、予測モデルを用いて、サンプリングした入力値に対する目的関数の推定値を算出する（ステップ S 203 b）。

配列情報処理部 111 a は、選択した入力値と算出した推定値を探索空間からサンプリングされる入力分子集合 X、目的スコアの集合 Y に割り当てる。従って、新たな入力値が目的関数と対応付けて探索空間からサンプリング

される入力分子集合Xに加わることで、配列情報集合の抽出に係る組合せ最適化アルゴリズムの抽出パラメータが更新される（ステップS203c）。

[0098] 推論部111は、ステップS203a～S203cの処理が終了する都度、1を加算することにより繰り返し回数を更新する（インクリメント）。推論部111は、繰り返し回数がサンプリング回数N以下のとき、ステップS203a～S203cの処理を繰り返す。

推論部111は、繰り返し回数がサンプリング回数Nを超えると、ループR20の処理を終了する。候補分子特定部111dは、目的関数の推定値の最も高いものから降順に所定数のビルディングブロックコンビネーションを、候補分子配列情報として出力部15を用いて出力してもよい。その後、図8の処理を終了する。

[0099] 本適用例に係る分子設計装置1が実行する組合せ最適化処理の流れの他の例を図9に示す。図9の処理は、ステップS203a～S203cの処理を繰り返す点で図8の処理と共通するが、ループR20に代え、ステップS203dとステップS203eを備える点で図8の処理と異なる。ステップS203dにおいて、推論部111は、サンプリング回数Nと繰り返し回数の初期値0を設定する。ステップS203bの処理の後、ステップS203eの処理に進む。ステップS203eにおいて、推論部111は、その時点における繰り返し回数に1を加算し、繰り返し回数がN回に達したか否かを判定する。達したと判定されるとき（ステップS203e YES）、図9の処理を終了する。達していないと判定されるとき（ステップS203e NO）、ステップS203cの処理に進む。ステップS203cの処理の後、ステップS203aの処理に進む。よって、図9の処理は、推論部111が、繰り返し回数がサンプリング回数Nを超えると、抽出パラメータの更新を行わない点で図8の処理と異なる。

[0100] なお、図8および図9の例では、終了条件を繰り返し回数が配列情報取得の回数を超える時として設定する場合を仮定したが、必ずしもこれに限られない。終了条件は、例えば、目的関数の推定値の目標値を予め設定しておく

、目的関数の推定値が、その目標値に到達することであってもよい。

[0101] なお、本願の検証例では、目的関数としてガウス過程回帰の予測平均値と予測分散値から算出されるMVを用いるTPEを $MV - TPE$ と呼ぶことがある。目的関数としてガウス過程回帰の予測平均 (Mean) を用いるTPEを $Mean - TPE$ と呼ぶことがある。本願では、目的関数としてのガウス過程回帰の予測平均は、特性値自体を指す。

[0102] <第一検証例>

次に、本実施形態の検証例について説明する。第一検証例では、予測モデルの訓練データとしてGFPデータセットを用いて本適用例の安全性を検証した。GFPデータセットは、56086個のGFP (Green Fluorescent Protein、緑色蛍光蛋白質) 配列とその特性としての輝度の情報を有する。GFPは、医学ならびに生物学研究において広く用いられる試料である。明るいGFPを生成することが研究目標とし、モデルベースの最適化性能のベンチマークとして用いられることがある。個々のGFP配列を示す入力値は、756次元のベクトル表現を用いて表される。本検証では、ベクトル表現を取得する際、蛋白質埋め込みモデルとして、タスク評価蛋白質埋め込み法 (TAPPE: Tasks Assessing Protein Embeddings method) を用いた。

[0103] 訓練データとして、親配列avGFP (aequorea victoria [オワンクラゲ] GFP) からの変異数が2以下となるGFP配列に基づいて、ガウス過程 (GP: Gaussian Process) を用いて予測モデルを学習した。以下の説明では、この学習により得られる予測モデルを代表モデル (proxy model) と呼ぶことがある。また、親配列avGFPを単に親配列またはテンプレートGFPと呼ぶことがある。この手順により、親配列からの残基の置換が少ない配列が検証されるので、実用的な分子最適化過程が実現される。本検証では、図10に例示されるように、2以下の残基置換 (編集距離 (edit distance)) を伴うGFP配列を採用した。疑似正解モデルとして、GFPデータセットの全データを用いて、Light GBM (Light Gradient Boosting Machine) を学習した。Light GBMは、決定木アルゴリズムに基づく、

ランク付け、分類、その他のタスクに用いられることがある。かかる設定により、代表モデルで親配列周りのG F P配列が網羅され、疑似正解モデルで、より広範なG F P空間が探索空間として網羅される。

[0104] 探索空間を訓練データの上位100配列における変異として定義した。探索空間には、図17に示されるように、37か所の変位候補箇所のそれぞれについて2～5個のアミノ酸の候補が含まれる。T P Eのパラメータとして、図18に示すように、サンプリング回数3000、多変量「なし」（即ち、目的関数が1変数）と設定した。訓練データの上位10配列を、M V - T P E、M e a n - T P Eのそれぞれについて、ウォームスタート初期化に用いた。

[0105] 通例、薬剤発見では、1回のバッチにおいて、10～100配列が評価される。本検証では、図10に例示されるように、疑似正解モデルを用いてスコア、即ち目的関数の推定値が高い上位96配列を提案配列として評価対象とした。

[0106] 検証では、目的関数としてM Vと平均をそれぞれ用いて図7、8の処理を実行した。平均を、比較例とした。M Vと平均には、それぞれ特性値が含まれるが、その特性値としてG F Pの輝度を用いた。図11は、T P Eによる最適化軌跡として、目的関数の推定値としての平均を上段に示し、M Vを下段に示す。いずれも10回分の最適化処理の結果を示す。破線は、各回のサンプルごとの推定値を示す。

図12は、各回におけるサンプル値の平均（G P Mean）と標準偏差（G P Std）との関係を示す。目的関数としてM Vを用いることで、平均を用いる場合よりも標準偏差が小さくなる。このことは、M Vを用いることで推定値の不確かさが低減することを意味する。

図13は、テンプレートG F Pを基準とする各サンプルの編集距離の密度分布を示す。目的関数としてM Vを用いることで、平均を用いる場合よりも全体として編集距離が小さくなる。このことは、M Vの利用により、テンプレートG F Pからの変異が少ないG F Pがサンプルとして選択されることを

意味する。

[0107] 図14は、疑似正解モデルを用いて得られた提案配列のスコアの分布を示す。スコアは、GFPにより発される蛍光の輝度の推定値を示す。目的関数としてMVを用いることで平均を用いる場合よりも輝度が高くなることを示す。

図15は、テンプレートGFPを基準とする提案配列の編集距離を示す。目的関数としてMVを用いることで平均を用いる場合よりも編集距離が小さくなる。このことは、MVを用いることで変異が少ない配列がサンプルされること、即ち安全な最適化を実現できることを示す。

図16は、提案配列の標準偏差 (GP Std) を示す。目的関数としてMVを用いることで平均を用いる場合よりも標準偏差が小さくなる。このことは、MVを用いることで特性値とする輝度の不確かさが少ない配列がサンプルされることで、安全な最適化を実現できることを裏付ける。

[0108] <第二検証例>

次に、第二検証例について説明する。第二検証例では、訓練データとして二重特異性抗体に関するデータを用い、ガウス過程を実行して予測モデルを学習した。本検証例において用いられた訓練データは、MarvelD3とCD3とを抗原とする二重特異性抗体（バイスペシフィック抗体）の配列と結合能の特性値とを示すデータである。結合能の特性値としてオクテット値 (Octet Value) を用いた。予測モデルに対する入力値としてTAPEを用いて得られるベクトル表現を用いた。このベクトル表現により、試料とする二重特異性抗体の蛋白質の配列が表現される。

[0109] MarvelD3は、4回膜貫通構造を有する密着結合蛋白質である。本検証例では、MarvelD3を抗がん薬の標的候補として設定した。がん抗原とT cell上の抗原を架橋する二重特異性抗体の開発は、がん治療への適用が期待される。本検証では、学習された予測モデルを、リード抗体からより優れた特性を持つ抗MarvelD3配列の候補を定める際に用いた。

[0110] 本検証では、抗体配列のオクテット値の測定を複数回行い、抗体配列と測

定により得られた抗体配列のオクテット値とを予測モデルへの入力として用いた。バッチ測定により、1回の測定で、典型的には100個以内の抗体配列のオクテット値が取得される。

ここで、測定に係る抗体を次の手順で取得した。まず、予め設計された重鎖または軽鎖をエンコードするプラスミドを準備し、Expi293F細胞を用いて組み換え抗体を一過性に発現させた。培養上清からプロテインAを用いて抗体を捕獲し、緩衝液に溶出させた。溶出した緩衝液を還元条件で混合することでMarvelD3/CD3二重特異性抗体を調製した。ここでは、同じ重鎖間で電荷斥力を印加して選択的重鎖ヘテロ二量化を実行した。280 nm吸光度により緩衝液における抗体の濃度を決定した。その後、二重特異性抗体を含む緩衝液に対してイオン交換クロマトグラフィーを行って、目的とするMarvelD3/CD3抗体が準備できたことを確認した。

[0111] オクテット値の測定に際しオクテットHTXシステムを用いた。CD81タンパク質とヒトMarvelD3タンパク質を表面に持つExtracellular vesiclesを、抗CD81抗体を用いてセンサーチップ上に捕獲した。0.1%BSAを含むD-PBS(-)溶液における600秒間のベースラインステップの後で、抗体20 nMを含む同じ緩衝液での会合応答と解離応答をそれぞれ900秒間および1500秒間測定した。抗体の結合能は、ベースラインステップと会合段階の終期との間での波長の変異として表れる。測定は、ベースラインステップ、会合、解離の各段階において、温度30°Cとし、毎分1000回の振動速度で振動させて行った。

[0112] 予測モデルの最適化において、第一検証例と同様にMV-TPEとMean-TPEをそれぞれ実行した。本検証では、サンプリングにより得られた候補配列のうち、目的関数の推定値について上位48配列を提案配列として評価した。

[0113] 図19は、各回におけるサンプリングにより得られた推定値の平均値と標準偏差の分布をMV-TPEとMean-TPEのそれぞれについて示す。標準偏差は不確実性を示し、推定値の平均値は高いほど良好なオクテット値を示す。図19によれば、目的関数として平均(Mean)を用いる場合に

は、標準偏差の最頻値が1.0となるのに対し、目的関数としてMVを用いる場合には、0.3となる。また、目的関数として平均を用いる場合には、平均の最頻値は2.3となるのに対し、MVを用いる場合には1.5となる。これらの結果は、目的関数としてMVを用いる場合よりも平均を用いる場合の方が、特性予測値としてのオクテット値の不確かさが高くなることを示す。

[0114] 図20は、サンプリングにより得られた配列のt-SNE (t-distributed Stochastic Neighbor Embedding) 可視化図である。t-SNEは、高次元データを低次元に圧縮して可視化する手法である。図20は、個々の配列を表すTAP Eによるベクトル表現の分布を、訓練データ (Train)、平均 (Mean)、MVのそれぞれについて二次元平面上に表す。図20によれば、目的関数として平均を用いる場合にサンプリングされた配列は訓練データからかけ離れているのに対し、MVを用いる場合には、訓練データにより近い。このことは、MV-TPEにより、Mean-TPEよりも不確かさを抑制しながら、より安全な領域を探索できることを示すとともに、病的なサンプルを回避できることが示される。

[0115] 図21は、評価対象の48配列のオクテット値の推定値の平均値 (GP Mean) をTPEの目的関数を平均 (Mean) とした場合と、MVとした場合のそれぞれについて示す。図21によれば、目的関数として平均を用いて得られるオクテット値の平均値よりもMVを用いて得られるオクテット値の平均値のほうが低くなった。

図22は、当該推定値の標準偏差 (GP Std) を平均 (Mean)、MVのそれぞれについて示す。図22によれば、目的関数として平均を用いて得られるオクテット値の標準偏差よりもMVを用いて得られるオクテット値の標準偏差のほうが低くなった。この結果から、MVによりサンプリングされる配列では、不確かさが低いことから、安定した発現および結合能が失われる可能性が低いことが推認される。

[0116] 図23は、評価対象の48配列の発現量の分布を平均 (Mean)、MV

のそれぞれについて示す。図23によれば、目的関数として平均を用いる場合、上位48配列の発現量が極めて低く抑えられ、オクテット値の測定に十分な試料が得られなかった。これに対し、目的関数としてMVを用いる場合、上位48配列のほぼ全てが有意に発現する。

図24は、配列ごとのオクテット値の分布をMV、訓練データのそれぞれについて示す。オクテット値の分布傾向は、MV、訓練データ間で類似する。これにより、目的関数としてMVを用いることで訓練データ、ひいては、親配列と同様の結合能を有する配列を取得できることが示される。さらに、MVによれば、訓練セットでの最大値よりもオクテット値が高い配列が数個得られた。

[0117] <総括>

上記の実証例によれば、本適用例に係るオフラインのモデルベースの最適化(MBO)を、薬剤設計における常套手段となるビルディングブロックベースの分子設計に応用し、in-silico実装(即ち、計算機による実装)を実現できることが示された。ここで、ベイズ最適化の一手法である木構造パルツェン推定器(TPE)がカテゴリ変数のコンビネーションの最適化に係るビルディングブロックベースの分子設計に適用可能とする。また、金融工学で用いられるMVがオフラインのMBOを適用することで、病的な挙動を回避し、安全な配列の探索に寄与することが示された。即ち、分子設計のMBOにあたって、risk averse predictionを組み込むことにより、候補分子の特性予測における外挿領域の過大評価を防止し、予測信頼度の低い分子の提案を防ぐことができた。また、上記の実証例により、本実施形態が治療抗体設計において安全な配列を探索するうえで有用であることが示された。

[0118] このように構成された、実施形態の分子設計装置1は、薬剤候補分子配列情報に基づき特性品質値を得るモデルを目的関数とするベイズ最適化を行う。そのため、創薬に要する負担をより軽減することができる。

[0119] (変形例)

なお、分子設計装置1は、ネットワークを介して通信可能に接続された複

数台の情報処理装置を用いて実装されてもよい。この場合、分子設計装置 1 が備える各機能部は、複数の情報処理装置に分散して実装されてもよい。また、創薬システム 100 は、分子設計装置 1 で推定された最適分子を生成するプラント（図示せず）を備えてもよい。分子設計装置 1 は、上記の手法により推定された最適分子が有する生物学的配列を示す出力情報をプラントに出力する。プラントは、分子設計装置 1 から入力される出力情報で指示される生物学的配列を有する分子化合物を生成する工程を実行する。

[0120] 分子設計装置 1 の各機能の全てまたは一部は、ASIC (Application Specific integrated Circuit)、PLD (Programmable Logic Device)、FPGA (Field Programmable Gate Array) 等のハードウェアを用いて実現されてもよい。プログラムは、コンピュータ読み取り可能な記録媒体に記録されてもよい。コンピュータ読み取り可能な記録媒体とは、例えばフレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置である。プログラムは、電気通信回線を介して送信されてもよい。

[0121] なお、分子設計装置 1 は推定装置の一例である。上記の機能は、分子設計を主機能としない装置、汎用のコンピュータシステムを備える情報機器など、他種の機器によって実現されてもよい。

[0122] 例えば、本実施形態は、プロセッサとメモリを備えるシステムとして実現されてもよい。当該システムでは、メモリが 1 以上の指令を記憶するように構成され、当該指令は、プロセッサに、複数の異なる分子の配列情報の集合である配列情報集合の要素をなす分子について、分子の配列情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する手順と、前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索する手順と、を実行させるための指令であってもよい。

また、当該指令は、プロセッサに、複数の異なる分子の配列情報の集合である配列情報集合を組合せ最適化アルゴリズムに従って取得する取得手順と

、分子の配列情報から該分子の特性を予測するための予測モデルを用いて、前記配列情報集合の要素をなす分子の特性予測値を算出する予測手順と、分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新する更新手順と、前記特性予測値に基づいて、所望の特性を有する分子を探索する探索手順と、を実行させるための指令であってもよい。

[0123] 本実施形態は、1以上の指令を記憶する非一時的コンピュータ読み取り可能媒体として実現されてもよい。当該指令は、コンピュータに、複数の異なる分子の配列情報の集合である配列情報集合の要素をなす分子について、分子の配列情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する手順と、前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索する手順と、を実行させるための指令であってもよい。また、当該指令は、コンピュータに、複数の異なる分子の配列情報の集合である配列情報集合を組合せ最適化アルゴリズムに従って取得する取得手順と、分子の配列情報から該分子の特性を予測するための予測モデルを用いて、前記配列情報集合の要素をなす分子の特性予測値を算出する予測手順と、分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新する更新手順と、前記特性予測値に基づいて、所望の特性を有する分子を探索する探索手順と、を実行させるための指令であってもよい。

[0124] 本実施形態は、人工知能エンジンを用いるコンピュータ実装方法として実現されてもよい。当該コンピュータ実装方法は、薬剤候補分子を構成する単位の配列の情報である薬剤候補分子配列情報、を要素とする集合である候補集合に対する処理であって、候補集合の各要素が示す各薬剤候補分子のうち、所望の特性の最適値を与える薬剤候補分子である最適分子を、組合せ最適化を実行することで推定する制御ステップと、当該最適分子にかかる情報を出力するステップと、を有し、組合せ最適化の目的関数は、薬剤候補分子配

列情報に基づき所望の特性の特性値の推定と前記推定の結果の不確かさの取得とを行う数理モデルの実行と、特性値と前記不確かさの所定に基づき算出された特性品質値の取得と、を行うモデルである。

- [0125] 上記では、いくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態およびその変形は、発明の範囲および要旨に含まれると同様に、特許請求の範囲に記載された発明とその均等の範囲に含まれるものである。本明細書において引用された文献は、参照により本明細書に組み入れられる。

符号の説明

- [0126] 1…分子設計装置、11…制御部、12…入力部、13…通信部、14…記憶部、15…出力部、100…創薬システム、111…推論部、112…入力制御部、113…通信制御部、114…記憶制御部、115…出力制御部、91…プロセッサ、92…メモリ。

請求の範囲

- [請求項1] 薬物候補に適した分子を特定する情報処理システムであって、
複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する特性予測部と、
前記特性予測値と前記予測の不確かさの推定値とに基づいて、所望の特性を有する分子の候補を探索する候補分子特定部と、
を備える情報処理システム。
- [請求項2] さらに、前記特性予測値と前記予測の不確かさの推定値とに基づいて特性品質値を算出する予測情報処理部を備え、
前記候補分子特定部は、前記特性品質値に基づき、前記ビルディングブロックコンビネーション情報集合の要素から少なくとも一つの候補分子を特定する、
請求項1に記載の情報処理システム。
- [請求項3] 前記特性品質値は、前記特性予測値と前記予測の不確かさの推定値の二変数を入力とする所定の関数の出力値である、
請求項2に記載の情報処理システム。
- [請求項4] 前記予測情報処理部は、前記特性品質値を与える目的関数として平均分散を算出する、
請求項2または3に記載の情報処理システム。
- [請求項5] 前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて減少する、
請求項2～4のいずれか一項に記載の情報処理システム。
- [請求項6] 組合せ最適化アルゴリズムを用いて、前記分子の特性予測値あるいは前記特性品質値に基づいて前記ビルディングブロックコンビネーション情報集合を取得するビルディングブロックコンビネーション情報

処理部を備える請求項2～5のいずれか一項に記載の情報処理システム。

[請求項7] 前記特性品質値は、前記特性予測値の増加に応じて増加し、前記予測の不確かさの推定値の減少に応じて前記ビルディングブロックコンビネーション情報集合は組合せ最適化アルゴリズムを用いて取得され、前記分子の特性予測値あるいは前記特性品質値に基づいて前記組合せ最適化アルゴリズムの抽出パラメータは更新される、請求項2～6のいずれか一項に記載の情報処理システム。

[請求項8] 前記組合せ最適化アルゴリズムは、木構造パルツェン推定器である、請求項7に記載の情報処理システム。

[請求項9] 前記予測モデルは、訓練用の複数の分子のビルディングブロックコンビネーション情報および該分子の特性評価の結果に基づく学習により生成された予測モデルである、請求項1～8のいずれか一項に記載の情報処理システム。

[請求項10] 前記ビルディングブロックコンビネーション情報は、前記分子の配列情報である、請求項1～9のいずれか一項に記載の情報処理システム。

[請求項11] 前記分子は、核酸、ペプチド、環状ペプチド、タンパク質、抗体、および、低分子化合物のうち少なくとも一つである、請求項1～10のいずれか一項に記載の情報処理システム。

[請求項12] 前記特性は、結合能、薬理活性、物性、動態、および、安全性のうち少なくとも一つの特性である、請求項1～11のいずれか一項に記載の情報処理システム。

[請求項13] 前記分子は、標的分子に結合する分子であり、前記特性は、前記標的分子に対する結合能である、請求項1～12のいずれか一項に記載の情報処理システム。

[請求項14] 薬物候補に適した分子を特定する情報処理システムにおける情報処

理方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定するステップと、

前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索するステップと、
を含む情報処理方法。

[請求項15]

コンピュータに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合の要素をなす分子について、分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、該分子の特性予測値を算出し、該予測の不確かさを推定する手順と、

前記特性予測値と前記予測の不確かさとに基づいて、より所望の特性を有する分子の候補を探索する手順と、
を実行させるためのプログラム。

[請求項16]

分子化合物を製造する方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合にアクセスし、

前記ビルディングブロックコンビネーション情報集合を予測モデルに入力する入力ステップと、

前記予測モデルより出力される、前記ビルディングブロックコンビネーション情報集合に含まれる分子ごとの特性予測値および予測の不確かさの推定値に基づき、前記ビルディングブロックコンビネーション情報集合からより所望の特性を有する分子を探索し、候補分子とし

て特定する推論ステップと、

当該候補分子にかかるビルディングブロックコンビネーション情報を出力する出力ステップと、

前記ビルディングブロックコンビネーション情報に示される分子配列を有する前記分子化合物を生成する生成ステップと、

を有する方法。

[請求項17]

薬物候補に適した分子を特定する情報処理システムであって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得するビルディングブロックコンビネーション情報処理部と、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子について、該分子の特性予測値を算出する特性予測部と、

前記特性予測値に基づいて、所望の特性を有する分子の候補を探索する候補分子特定部と、を備え、

前記ビルディングブロックコンビネーション情報処理部は、

前記分子の特性予測値に基づいて、より所望の特性を有する分子が含まれるように、前記組合せ最適化アルゴリズムの抽出パラメータを更新する

情報処理システム。

[請求項18]

薬物候補に適した分子を特定する情報処理システムにおける情報処理方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得するステップと、

分子のビルディングブロックコンビネーション情報から該分子の特

性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子について、該分子の特性予測値を算出するステップと、

前記特性予測値に基づいて、より所望の特性を有する分子を探索するステップと、

前記分子の特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新するステップを含む、

情報処理方法。

[請求項19]

コンピュータに、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得する取得手順と、

分子のビルディングブロックコンビネーション情報から該分子の特性を予測するための予測モデルを用いて、前記ビルディングブロックコンビネーション情報集合の要素をなす分子の特性予測値を算出する予測手順と、

分子ごとの特性予測値に基づいて、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新する更新手順と、

前記特性予測値に基づいて、所望の特性を有する分子を探索する探索手順と、

を実行させるためのプログラム。

[請求項20]

分子化合物を製造する方法であって、

複数の異なる分子のビルディングブロックコンビネーション情報の集合であるビルディングブロックコンビネーション情報集合を組合せ最適化アルゴリズムに従って取得する取得ステップと、

前記ビルディングブロックコンビネーション情報集合を予測モデル

に入力する入力ステップと、

前記予測モデルより出力される、前記ビルディングブロックコンビネーション情報集合に含まれる分子ごとの特性予測値に基づき、より所望の特性を有する分子が含まれるように前記組合せ最適化アルゴリズムの抽出パラメータを更新する更新ステップと、

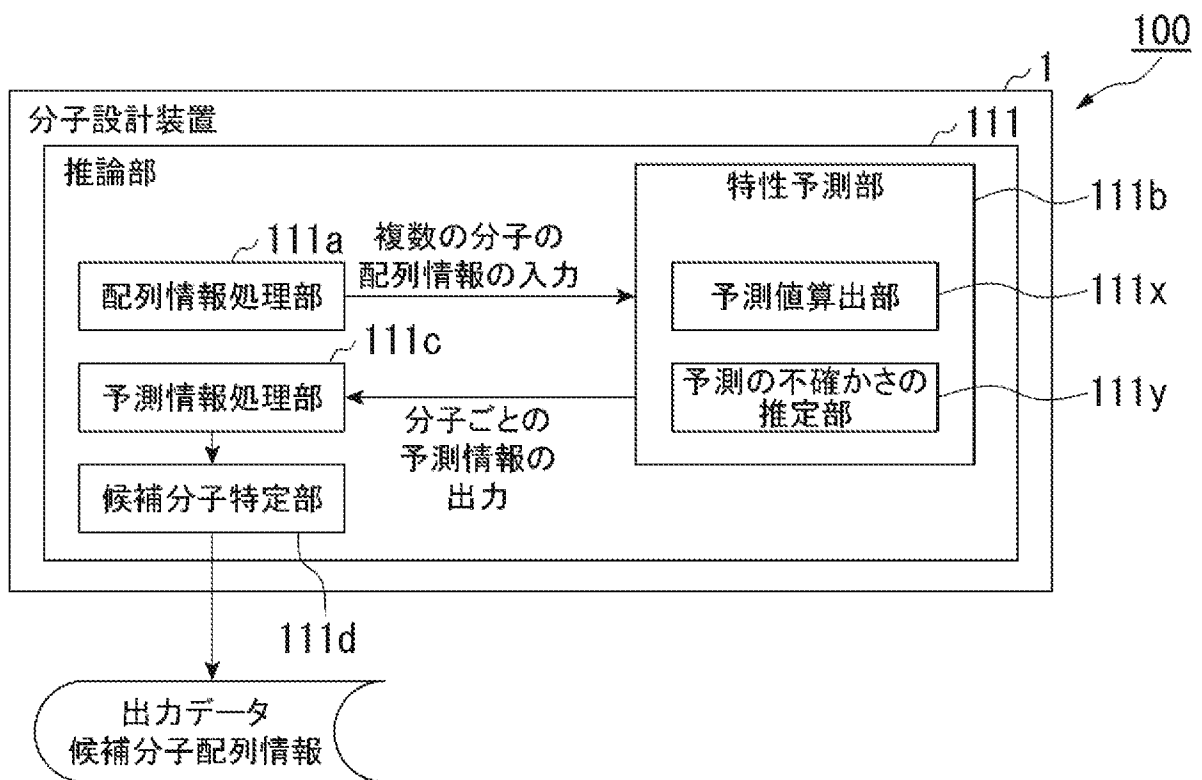
前記特性予測値に基づき、前記ビルディングブロックコンビネーション情報集合からより所望の特性を有する分子を探索し、候補分子として特定するステップと、

当該候補分子にかかるビルディングブロックコンビネーション情報を出力する出力ステップと、

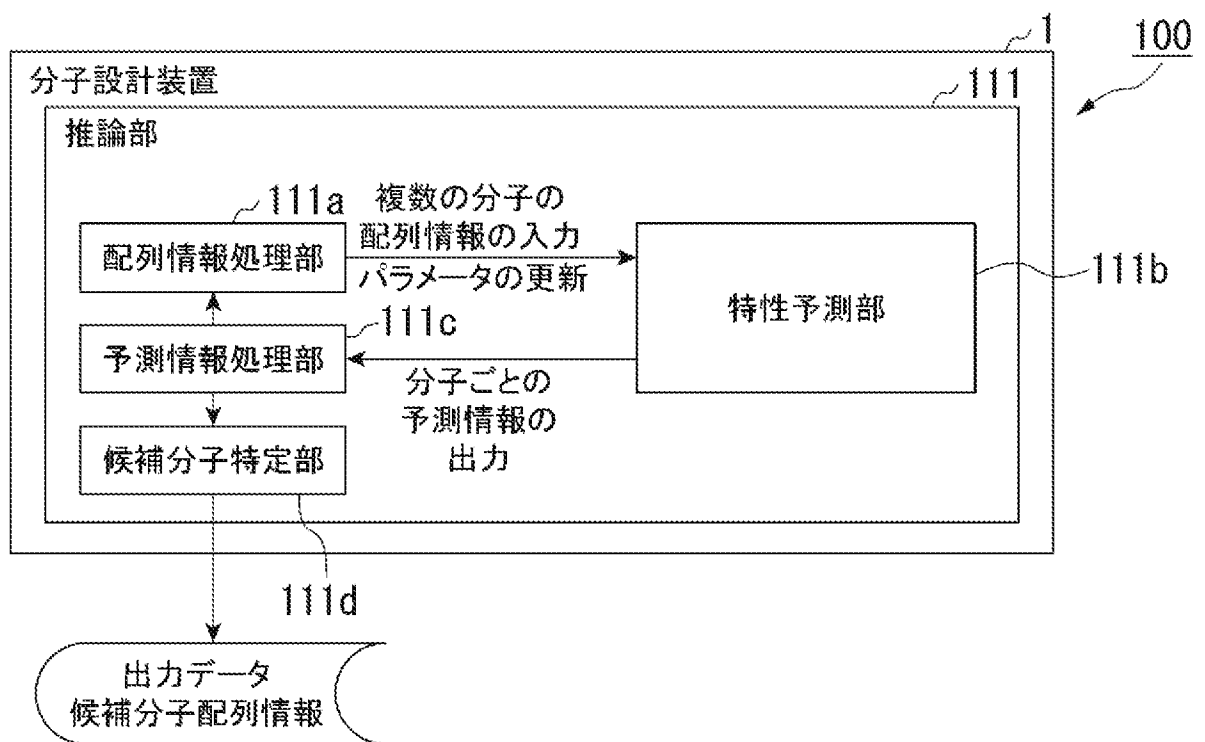
前記ビルディングブロックコンビネーション情報に示される分子配列を有する前記分子化合物を生成する生成ステップと、

を有する方法。

[図1]



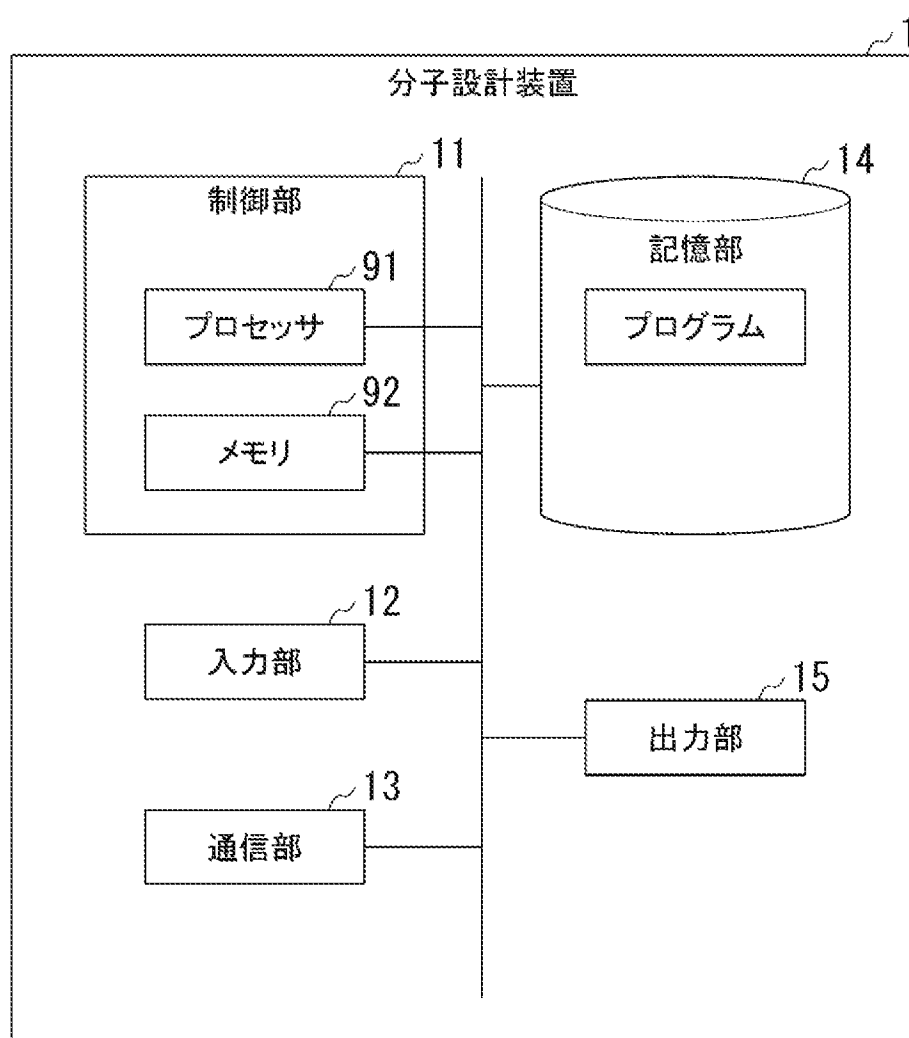
[図2]



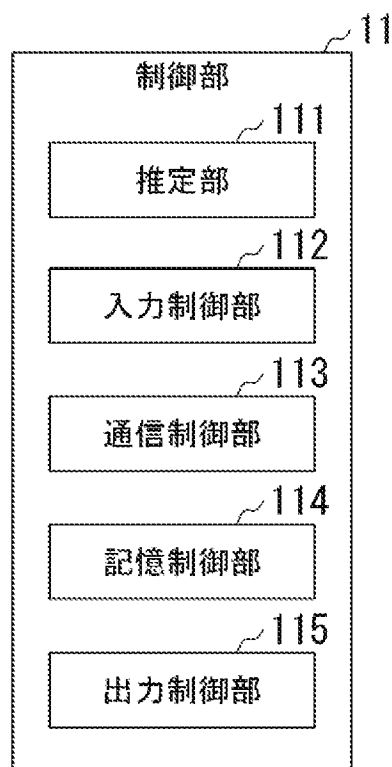
[図3]

[illegible]

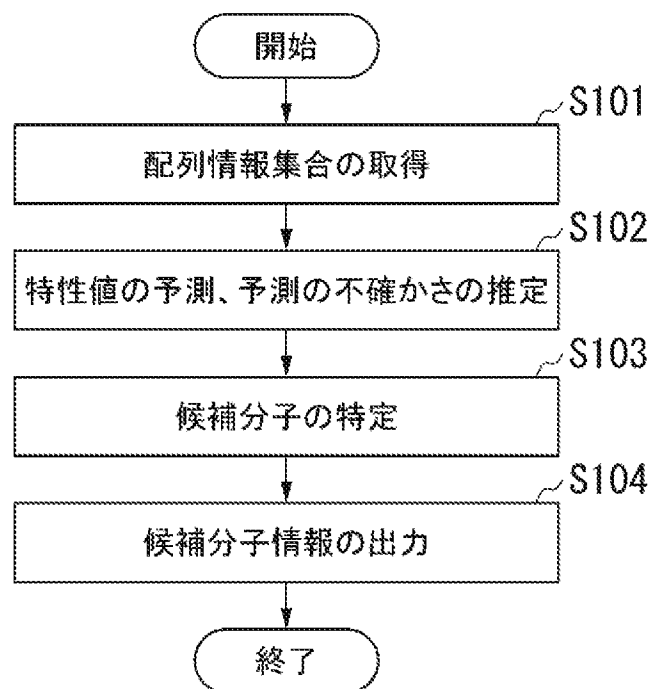
[図4]



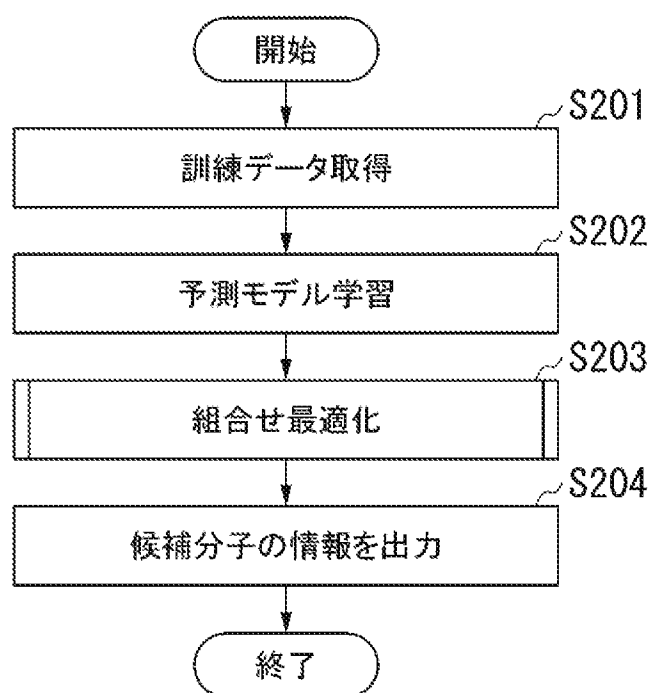
[図5]



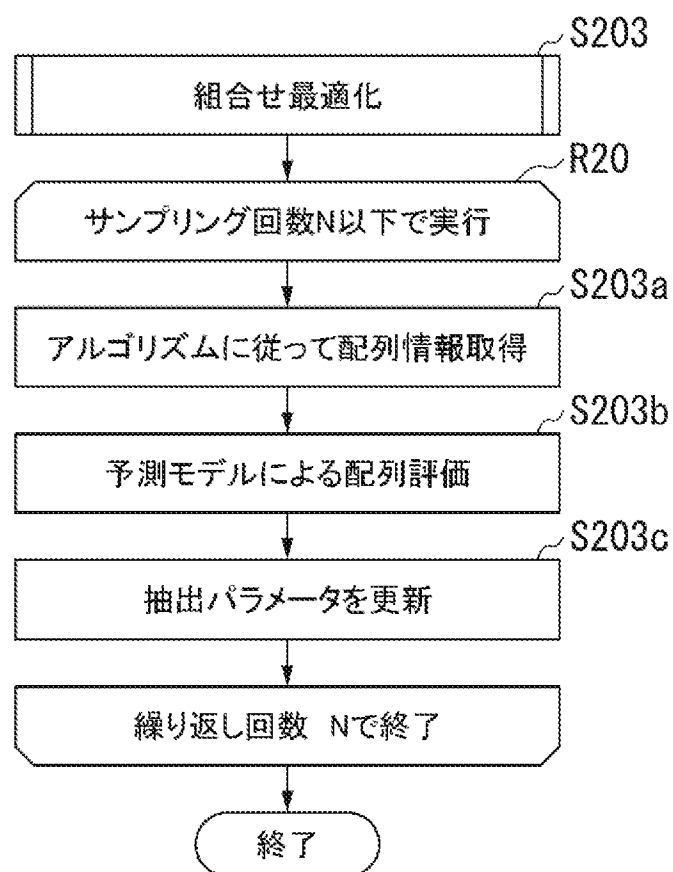
[図6]



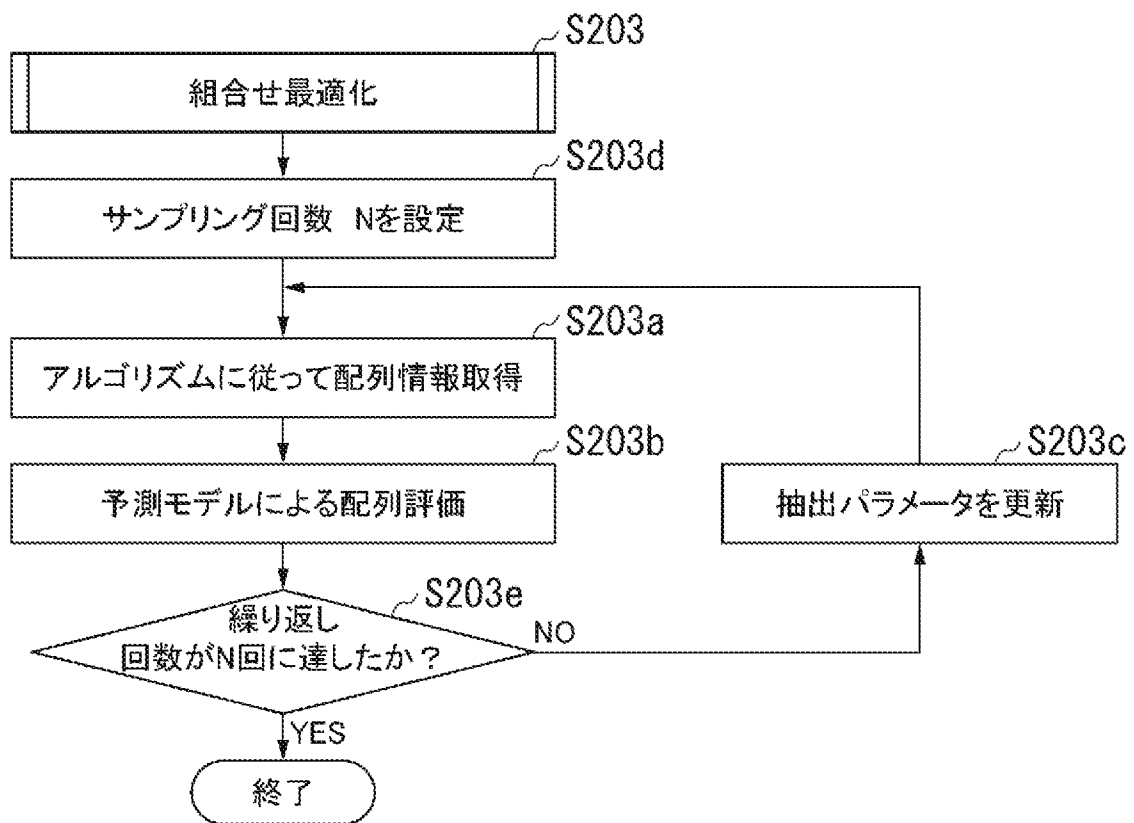
[図7]



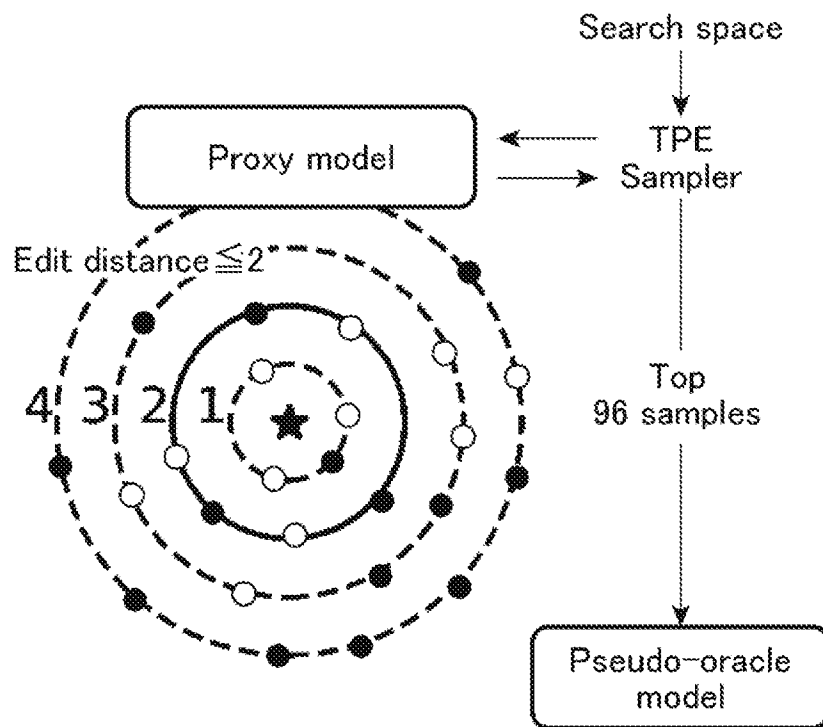
[図8]



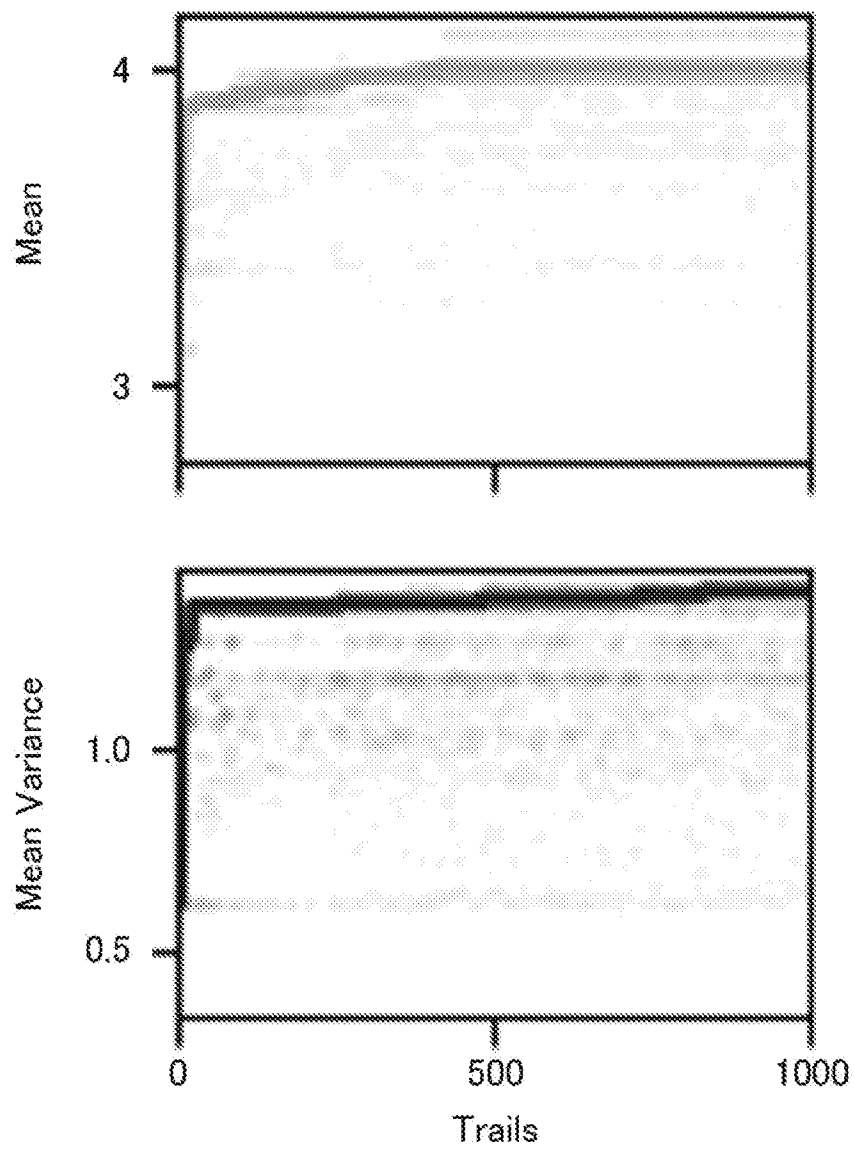
[図9]



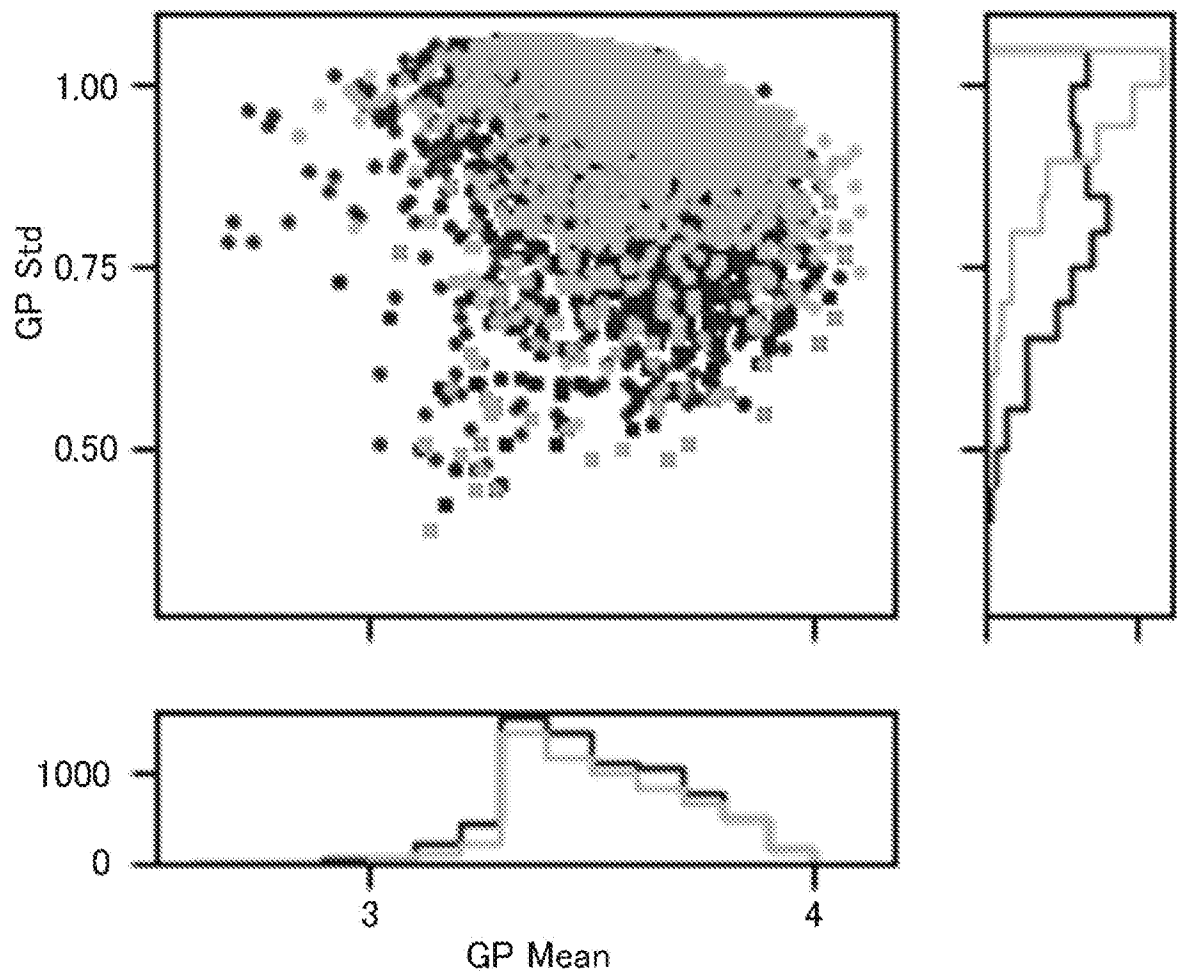
[図10]



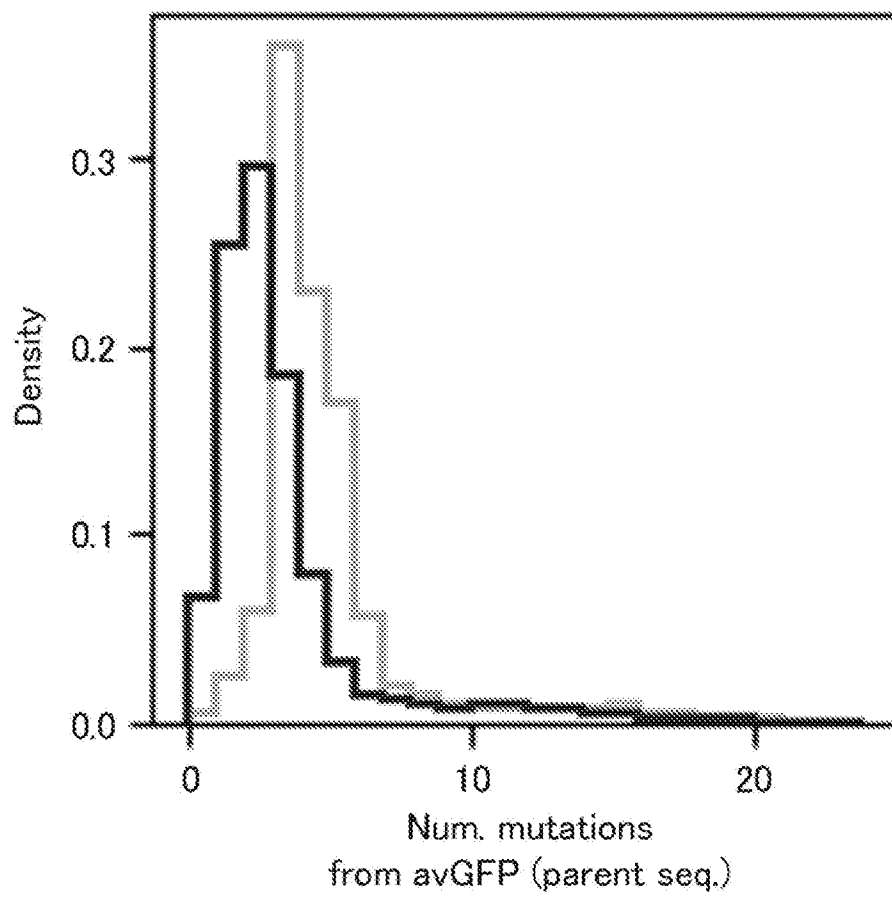
[図11]



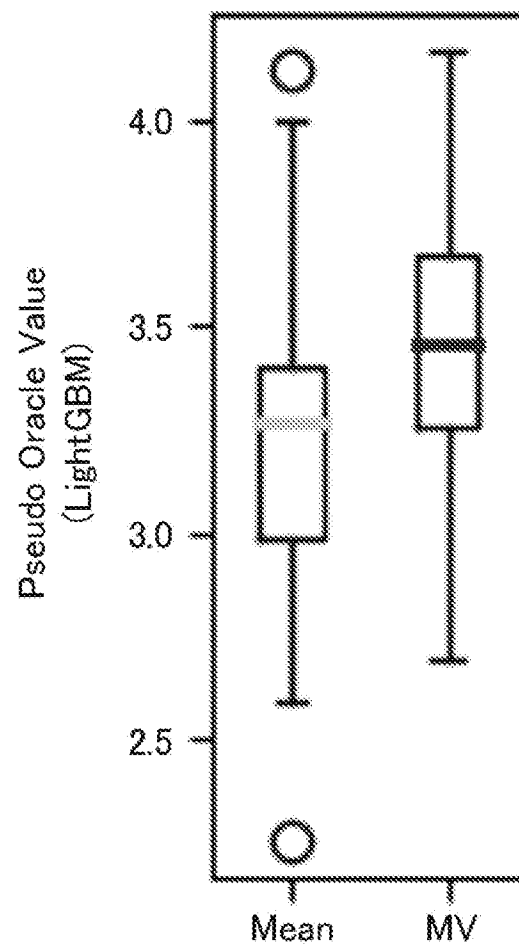
[図12]



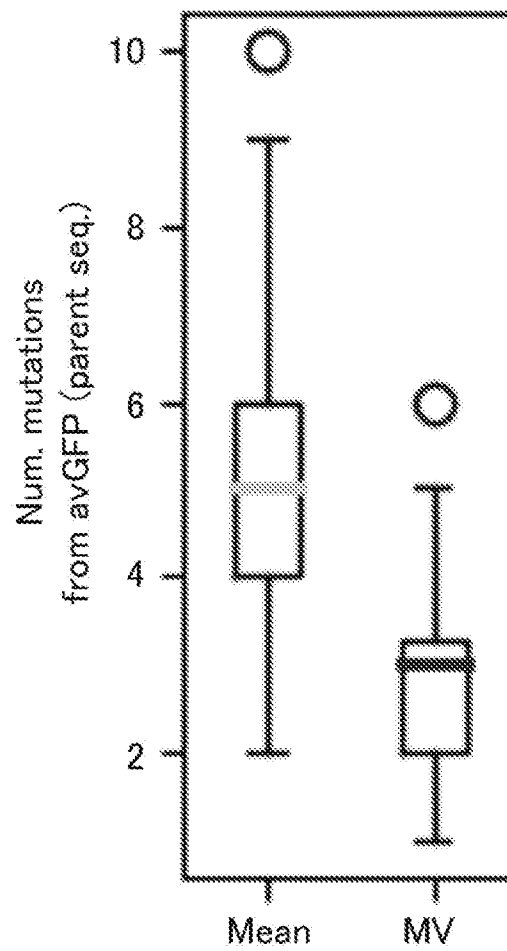
[図13]



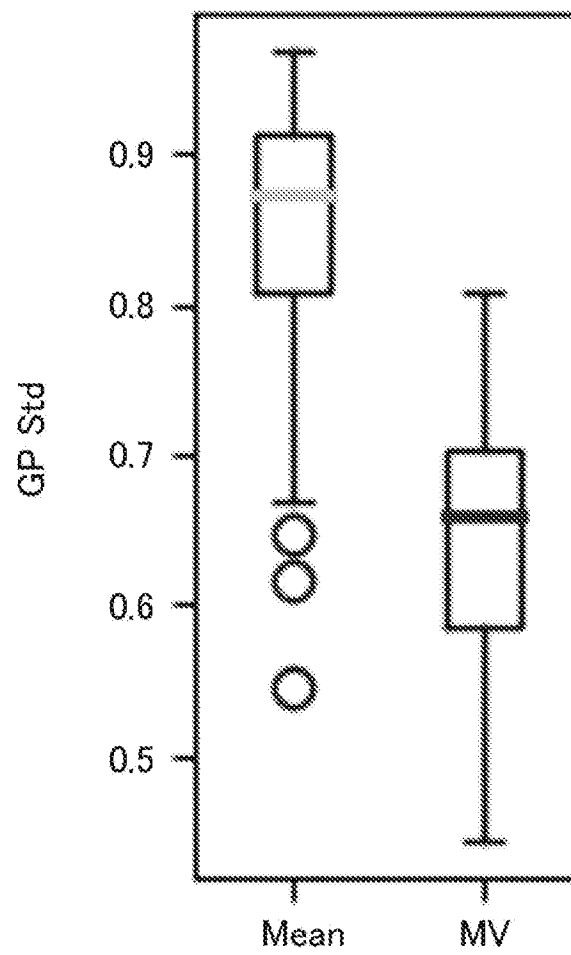
[図14]



[図15]



[図16]



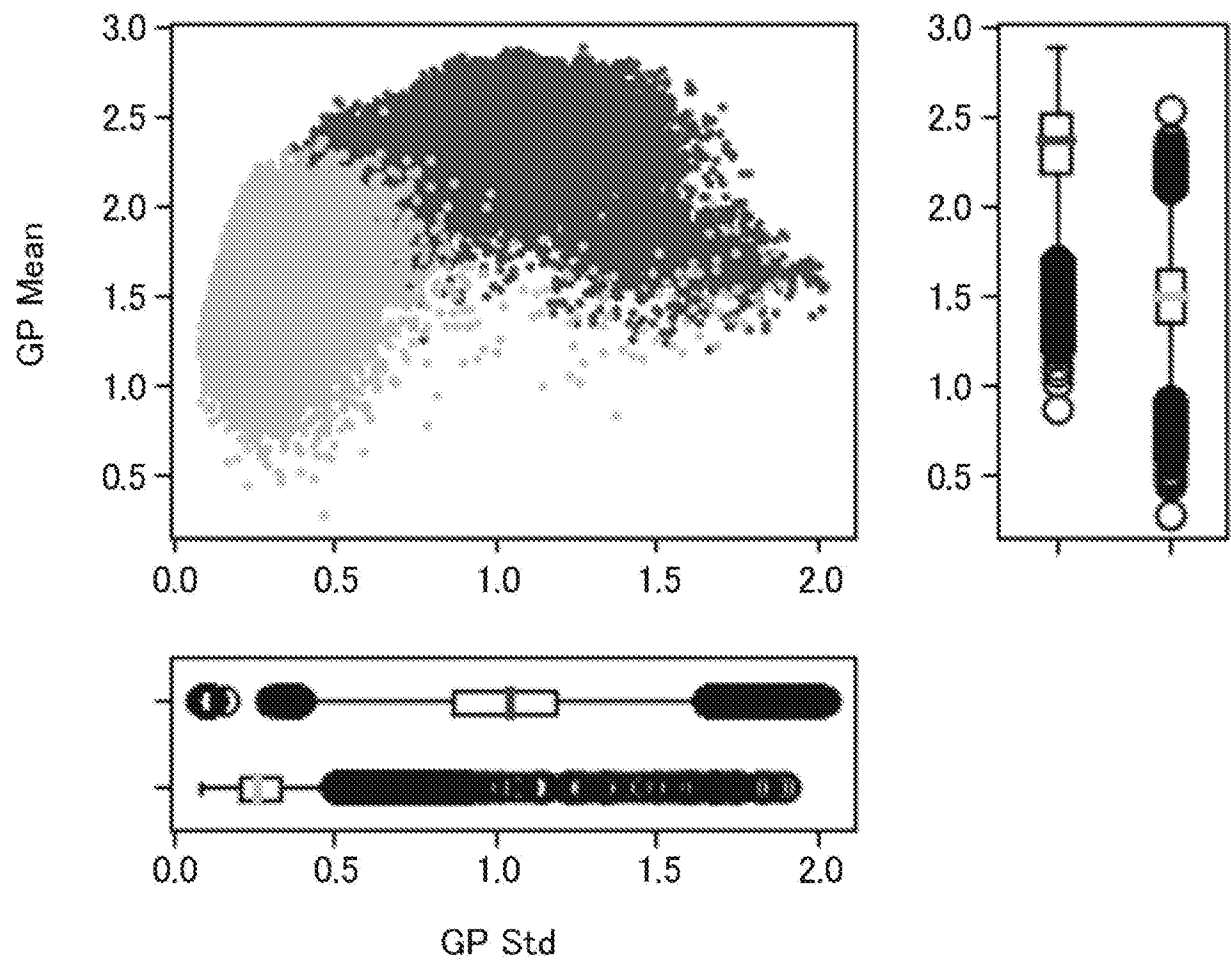
[図17]

位置	候補	位置	候補	位置	候補	位置	候補	位置	候補
3	AG	72	RL	136	LM	174	ST	220	LV
18	DE	74	HPT	143	NG	177	ILV	230	HY
24	QH	104	NCHIY	144	LY	188	CG	231	AG
29	CST	105	FY	157	GKV	190	GV	233	DY
38	NHSY	109	AS	162	AV	197	ND	236	NY
41	LM	110	QE	163	NT	202	IT		
43	LM	127	AILT	167	RGH	203	QH		
61	PST	128	DG	170	IL	205	AV		

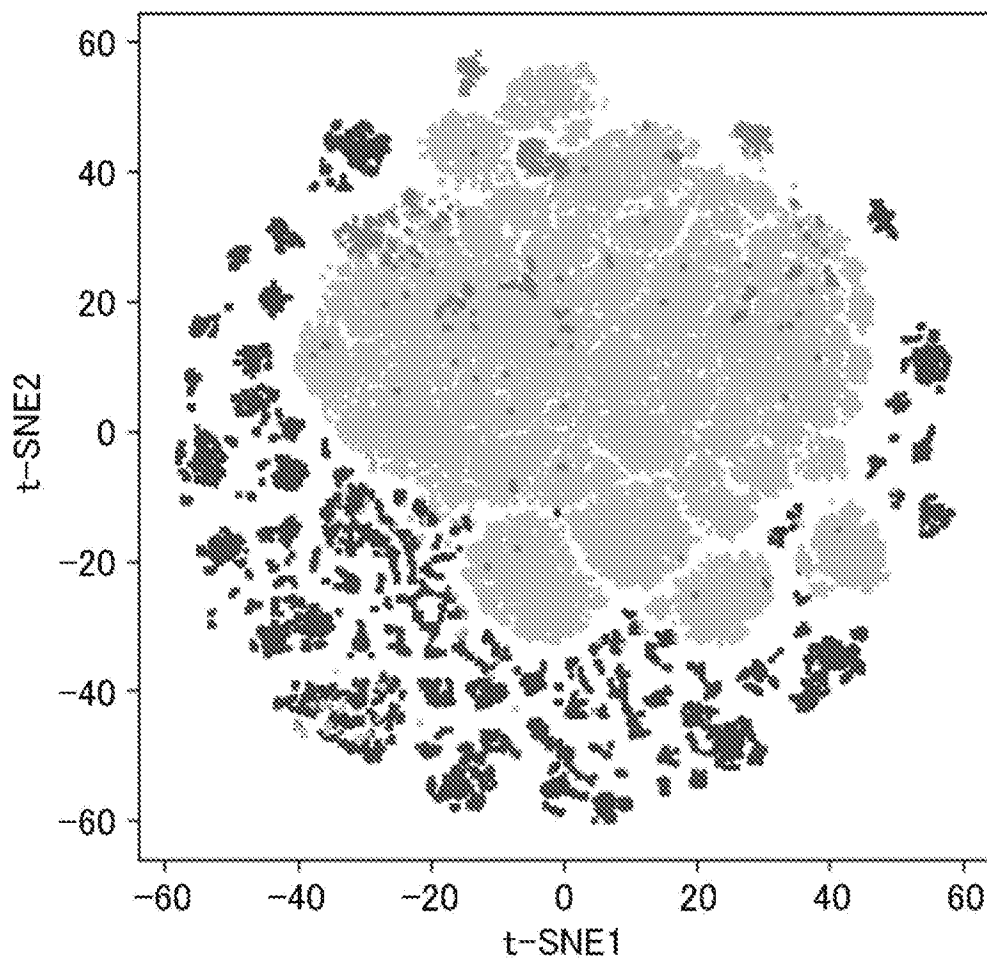
[図18]

名称	値
サンプリング回数	3000
多変量	なし

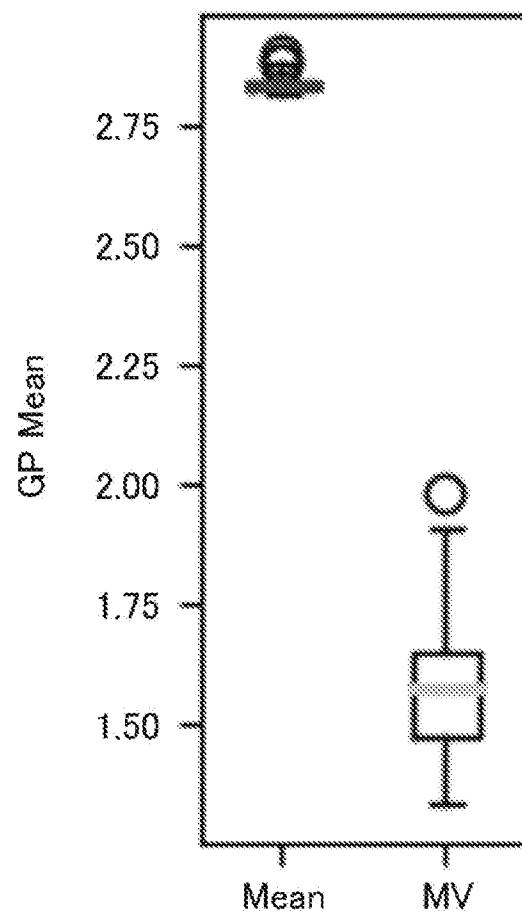
[図19]



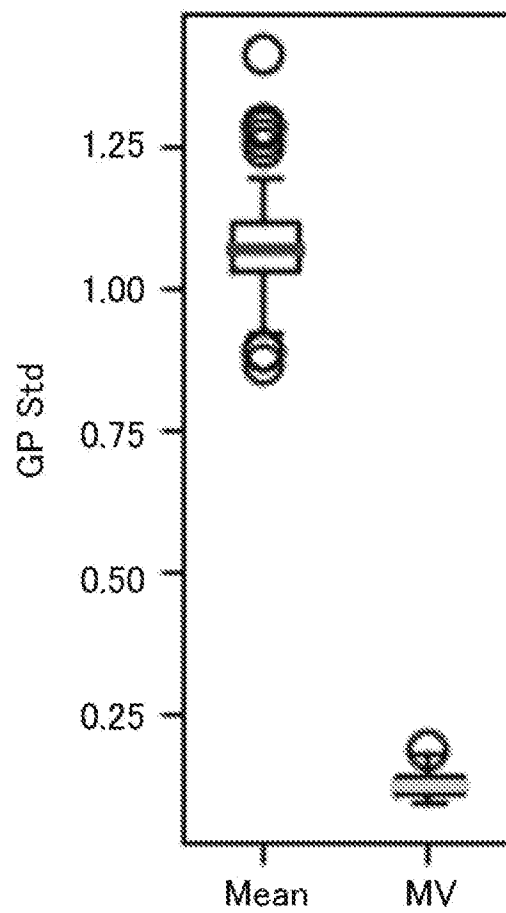
[図20]



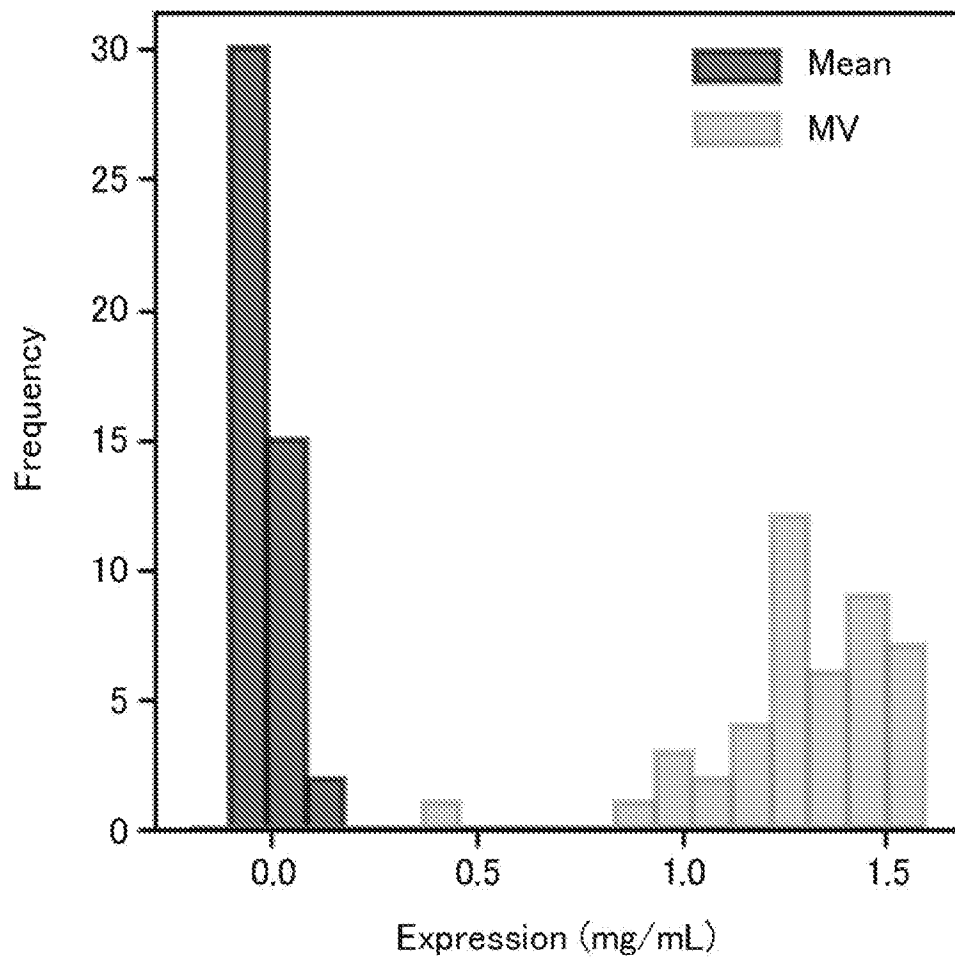
[図21]



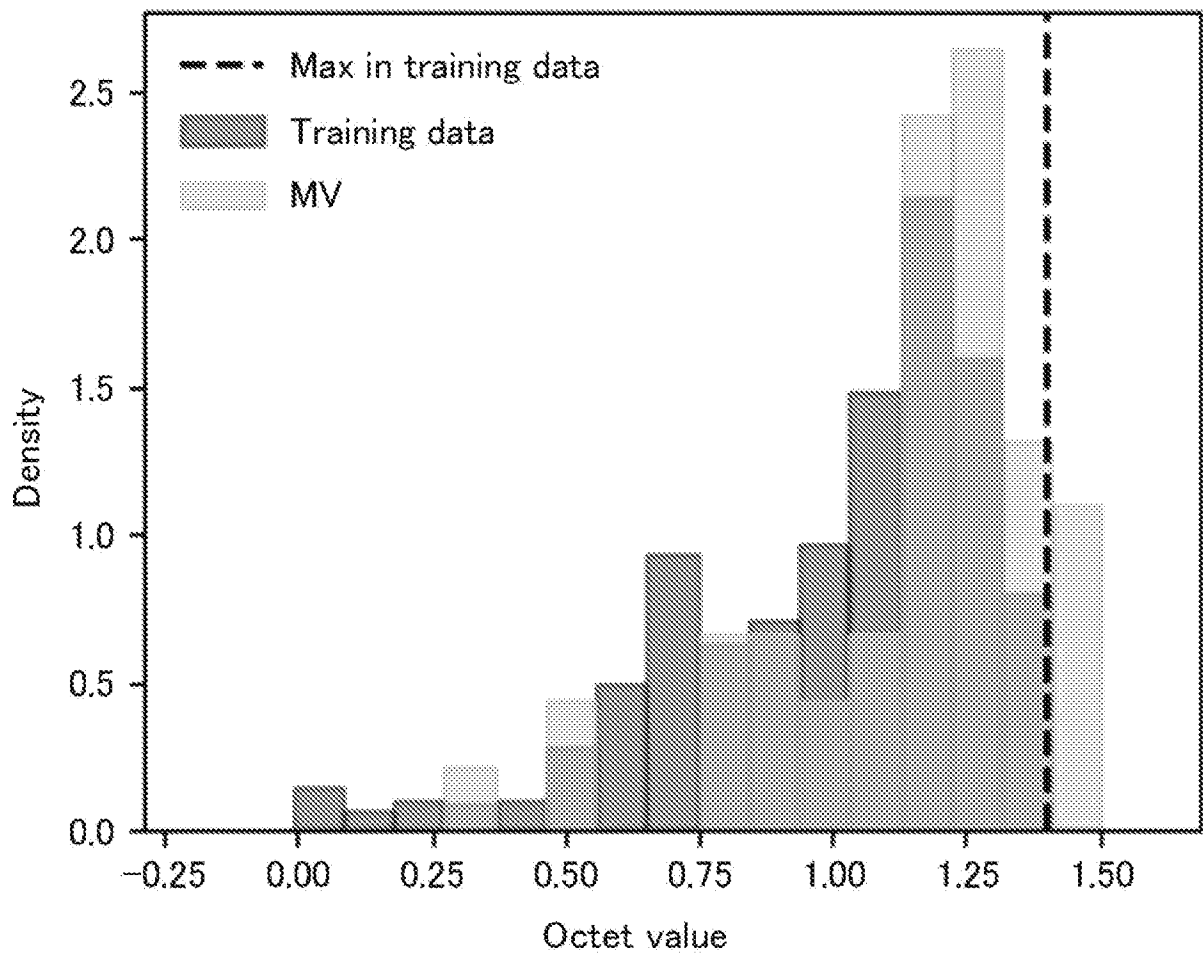
[図22]



[図23]



[図24]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2024/027412

A. CLASSIFICATION OF SUBJECT MATTER <i>G16C 20/50</i> (2019.01)i FI: G16C20/50 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) G16C20/50 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2024 Registered utility model specifications of Japan 1996-2024 Published registered utility model applications of Japan 1994-2024 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2008-174503 A (NEC CORPORATION) 31 July 2008 (2008-07-31) entire text, all drawings	1-20
A	JP 2022-150078 A (FUJITSU LIMITED) 07 October 2022 (2022-10-07) entire text, all drawings	1-20
☐ 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 “D” document cited by the applicant in the international application “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 15 August 2024		Date of mailing of the international search report 03 September 2024
Name and mailing address of the ISA/JP Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2024/027412

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
JP	2008-174503	A	31 July 2008	(Family: none)			
JP	2022-150078	A	07 October 2022	US	2022/0310211	A1	
				EP	4071764	A1	
				CN	115132292	A	

A. 発明の属する分野の分類（国際特許分類（IPC））

G16C 20/50(2019.01)i

FI: G16C20/50

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

G16C20/50

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報

1922 - 1996年

日本国公開実用新案公報

1971 - 2024年

日本国実用新案登録公報

1996 - 2024年

日本国登録実用新案公報

1994 - 2024年

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

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2008-174503 A（日本電気株式会社）31.07.2008（2008 - 07 - 31） 全文、全図	1-20
A	JP 2022-150078 A（富士通株式会社）07.10.2022（2022 - 10 - 07） 全文、全図	1-20

☐ C欄の続きにも文献が列挙されている。

☒ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

“A” 特に関連のある文献ではなく、一般的技術水準を示すもの

“D” 国際出願で出願人が先行技術文献として記載した文献

“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

“O” 口頭による開示、使用、展示等に言及する文献

“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献

“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの

“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

“&” 同一パテントファミリー文献

国際調査を完了した日

15.08.2024

国際調査報告の発送日

03.09.2024

名称及びあて先

日本国特許庁(ISA/JP)

〒100-8915

日本国

東京都千代田区霞が関三丁目4番3号

権限のある職員（特許庁審査官）

塩田 徳彦 5R 4533

電話番号 03-3581-1101 内線 3560

様式 PCT/ISA/210（第2ページ）（2022年7月）

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2024/027412

引用文献			公表日	パテントファミリー文献			公表日
JP	2008-174503	A	31.07.2008	(ファミリーなし)			
JP	2022-150078	A	07.10.2022	US	2022/0310211	A1	
				EP	4071764	A1	
				CN	115132292	A	