

明 細 書

発明の名称：

特徴量算出方法、特徴量算出プログラム、特徴量算出装置、スクリーニング方法、スクリーニングプログラム、及び化合物創出方法

技術分野

[0001] 本発明は特徴量の算出、化合物のスクリーニング、及び化合物の立体構造の創出を行う方法、プログラム、及び装置に関し、特に医薬候補化合物を探索する技術に関する。

背景技術

[0002] 従来、コンピュータを用いた創薬研究では、既存の化合物を数万から十萬程度集めたライブラリを用意し化合物の構造式を与えて1つの標的タンパク質に対する結合力を調べることで、医薬候補化合物（以下、「ヒット」という）を探索してきた。例えば、下記特許文献1では、化合物の構造式を与えて結合力を予測している。また特許文献2においても、構造式の発生と結合力の予測とを繰り返して、望ましい結合力を有する化合物を少しずつ探索する（試行錯誤する）ことが記載されている。

[0003] また、特許文献3には「化合物指紋（フィンガープリント）」と呼ばれる記述子を用いて探索を行うことが記載されている。「記述子」とは化合物の構造式から得られる情報であり、「化合物指紋」は様々な官能基の有無などの情報を示す。このような記述子は「化合物の記述子が類似ならば、化合物の骨格が類似する」のが特徴である。

先行技術文献

特許文献

[0004] 特許文献1：米国特許第9373059号公報

特許文献2：特許第5946045号公報

特許文献3：特許第4564097号公報

発明の概要

発明が解決しようとする課題

[0005] 最近ニーズの高い標的タンパク質が複雑で高難度になり、単なるライブラリのスクリーニングではヒットを見つけづらい。他方、化合物の理論上の数は分子量500以下の低分子に限っても（10の60乗）個はあり、分子量1,000前後の中分子まで拡大するとその数はさらに増え、有史以来合成された化合物が（10の9乗）個程度であることを考えると、まだまだヒット発見の可能性がある。しかしながら、このような天文学的な数の化合物すべてに対して結合力を調べることは、実験はもちろん、シミュレーションでもほとんど不可能である。一部の化合物に対して結合力を調べる場合でも、上述した特許文献1,2のような試行錯誤の繰り返しでは効率が低い。また、特許文献3に記載されたFingerprintのような従来の記述子（特徴量）の場合、同じ薬効を示す化合物であっても特徴量が類似しているとは限らず、特徴量が対象構造体の化学的性質を的確に示していなかったため、特徴量を用いた探索の効率が低かった。

[0006] このように、従来の技術では特徴量が対象構造体の化学的性質を的確に示しておらず、このため特徴量を用いたスクリーニング、立体構造創出の効率が低かった。

[0007] 本発明はこのような事情に鑑みてなされたもので、対象構造体の化学的性質を的確に示す特徴量を算出できる特徴量算出方法、特徴量算出プログラム、及び特徴量算出装置を提供することを目的とする。また、本発明は特徴量を用いて医薬候補化合物のスクリーニングを効率よく行うことができるスクリーニング方法及びスクリーニングプログラムを提供することを目的とする。また、本発明は特徴量を用いて医薬候補化合物の立体構造を効率よく創出することができる化合物創出方法を提供することを目的とする。

課題を解決するための手段

[0008] 上述した目的を達成するため、本発明の第1の態様に係る特徴量算出方法は、化学的性質を有する複数の単位構造体から構成される対象構造体を指定する対象構造体指定工程と、対象構造体の立体構造の周辺における1種類以

上のプローブの集積度合いを3次元空間において定量化した特徴量を算出する特徴量算出工程であって、機械学習により構成された生成器を用いて対象構造体から特徴量を算出する特徴量算出工程と、を有し、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である。

[0009] 対象構造体の化学的性質は対象構造体とその周辺における1種類以上のプローブとの相互作用の結果として表されるので、対象構造体どうしでプローブの集積度合いが類似であることはそれら対象構造体の化学的性質が類似していることを示す。すなわち、第1の態様により算出される特徴量が類似な対象構造体は類似の化学的性質を示す。したがって第1の態様により対象構造体の化学的性質を的確に示す特徴量を算出することができる。

[0010] 第1の態様において、特徴量の算出に用いる「生成器」（予測モデル、学習済みモデル）は、例えば「『複数の化合物についての、化合物の立体構造を示す情報（3次元画像等）を教師データとし、特徴量を説明変数とした学習用データセット』を生成器に入力して機械学習により学習させる学習処理工程を有する生成器の構築方法」により構築することができる。階層的ニューラルネットワークを用いて生成器を構築してもよい。

[0011] 第1の態様において、プローブを構成する「点」は数学的な「点」とは異なり大きさがあってもよい（実際の原子、イオン等）。また実際の原子等ではなく仮想的な点電荷（「点」の一態様）やダイポール等をプローブとしてもよい。プローブは、ターゲット（対象構造体）となる化合物に応じて選択することができる。

[0012] なお、第1の態様に係る特徴量算出方法は、CPU（Central Processing Unit）等の各種プロセッサを1つ以上用いて実現されるコンピュータ、あるいはそのようなコンピュータを備える特徴量算出装置、特徴量算出システム等により実行することができる。以下の各態様に係る特徴量算出方法、スクリーニング方法、及び化合物創出方法についても同様である。

[0013] 第2の態様に係る特徴量算出方法は第1の態様において、対象構造体指定

工程では対象構造体として化合物を指定し、特徴量算出工程では、生成器を用いて、化合物の立体構造の周辺における、プローブとしてのアミノ酸の集積度合いを3次元空間において定量化した特徴量である第1の特徴量を算出する。第2の態様は、第1の態様における「プローブ」、「対象構造体」、「複数の単位構造体」をそれぞれアミノ酸、化合物、複数の原子としたものである。集積度合いを定量化するアミノ酸は1種類に限らず、2種類以上のアミノ酸が結合したペプチドでもよい。

[0014] 第1の態様と同様に、化合物の薬効（例えば、標的タンパク質への結合力）は局所的には化合物と各アミノ酸（プローブの一態様）との相互作用の結果として表されるので、化合物間でアミノ酸の集積度合いが類似であれば、それら化合物は（標的タンパク質との）結合力が類似していることを示す。すなわち、第2の態様に係る特徴量（第1の特徴量）が類似な化合物は類似の薬効を示す。したがって第2の態様により化合物の化学的性質を的確に示す特徴量を算出することができる。なお第2の態様において、生体リガンド等、立体構造及び標的タンパク質への結合が知られている化合物を対象構造体として指定することができる。

[0015] 第3の態様に係る特徴量算出方法は第1の態様において、対象構造体指定工程では、標的タンパク質の活性部位であるポケットに結合するポケット構造体を対象構造体として指定し、特徴量算出工程では、生成器を用いて、複数の仮想的球体によるポケット構造体の立体構造の周辺における、プローブとしてのアミノ酸の集積度合いを3次元空間において定量化した特徴量である第2の特徴量を算出する。第3の態様は、第1の態様における「プローブ」、「対象構造体」、「単位構造体」をそれぞれアミノ酸、ポケット構造体、複数の仮想的球体としたものである。標的タンパク質の「活性部位」とはポケット構造体が結合することにより標的タンパク質の活性が促進または抑制される部位を意味し、「仮想的球体」はファンデルワールス半径、電荷等の化学的性質を有すると考えることができる。

[0016] 第2の態様について上述したのと同様に第3の態様に係る特徴量（第2の

特徴量) が類似なポケット構造体は類似の化学的性質を示すので、第3の態様によりポケット構造体の化学的性質を的確に示す特徴量を算出することができる。なお、ポケット構造体は標的タンパク質のポケットと結合する化合物に対応する。また第3の態様において、標的タンパク質の立体構造の実測結果、ポケットの位置情報等に基づくシミュレーションを第2の特徴量の算出に用いることができる。なお、標的タンパク質の立体構造は、アミノ酸の各残基が識別できる分解能の立体構造であれば、測定技術(X線結晶構造、NMR構造(NMR: Nuclear Magnetic Resonance)、クライオTEM構造(TEM: Transmission Electron Microscopy)、など)は限定されない。

[0017] 第4の態様に係る特徴量算出方法は第1の態様において、対象構造体指定工程では対象構造体として化合物を指定し、特徴量算出工程では、生成器を用いて、化合物の立体構造の周辺におけるプローブの集積度合いであって、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子、水、複数の原子で構成される1種類以上のイオンのうち1つ以上をプローブとした集積度合いを3次元空間において定量化した特徴量である第3の特徴量を算出する。第4の態様は、第1の態様における「プローブ」、「対象構造体」、「複数の単位構造体」をそれぞれ1種類以上の核酸塩基等(任意の種類、数、組合せでよい)、化合物、複数の原子としたものである。

[0018] 本発明では、医薬のターゲットとしてタンパク質以外の生体高分子(化合物)であるDNA(Deoxyribonucleic Acid)、RNA(Ribonucleic Acid)、細胞膜、多糖を扱うことができる。第4の態様はこれらのターゲット化合物についての特徴量の算出方法を規定するもので、プローブはアミノ酸ではなく別の物質(各ターゲットのビルディングブロック)にする。具体的には、ターゲットがDNA、RNA、細胞膜、多糖の場合、プローブはそれぞれ1種類以上の核酸塩基、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子とする。また、これらをプローブとして集積度合いを定量化する際に、水、複数の原子で構成される1種類以上のイオンを考慮してもよい。第1から第3の態様と同様に、化合物の薬効(DNA等のターゲッ

トへの結合力)は局所的には化合物と核酸塩基等(プローブ)との相互作用の結果として表されるので、化合物間で核酸塩基等の集積度合いが類似であれば、それら化合物はターゲットとの結合力が類似していることを示す。すなわち第4の態様に係る特徴量(第3の特徴量)が類似な化合物は類似の薬効を示す。したがって、第4の態様により化合物の化学的性質を的確に示す特徴量を算出することができる。

[0019] 第1から第4の態様において、特徴量を化合物の回転及び並進について不変量化する不変量化工程を設けてもよい。このような不変量化により、特徴量を扱いやすくまたデータ容量を小さくすることができる。特徴量の不変量化は、フーリエ変換、相関関数の角度積分等により行うことができる。

[0020] 上述した目的を達成するため、本発明の第5の態様に係る特徴量算出プログラムは第1から第4の態様のいずれか1つに係る特徴量算出方法をコンピュータに実行させる。第5の態様における「コンピュータ」は、CPU (Central Processing Unit) 等の各種プロセッサを1つ以上用いて実現することができる。この「コンピュータ」は特徴量算出装置、特徴量算出システムに備えられたものでよい。なお、第5の態様に係る特徴量算出プログラムのコンピュータ読み取り可能なコードが記録された非一時的記録媒体も、本発明の態様として挙げることができる。

[0021] 上述した目的を達成するため、本発明の第6の態様に係る特徴量算出装置は化学的性質を有する複数の単位構造体から構成される対象構造体を指定する対象構造体指定部と、対象構造体の立体構造の周辺における1種類以上のプローブの集積度合いを3次元空間において定量化した特徴量を算出する特徴量算出部であって、機械学習により構成された生成器を用いて対象構造体から特徴量を算出する特徴量算出部と、を備え、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である。第6の態様によれば、第1の態様と同様に対象構造体の化学的性質を的確に示す特徴量を算出することができる。なお、第6の態様に係る特徴量算出装置は、CPU等の各種プロセッサにより本発明に係る特徴量算出プ

ログラムのコードを読み取って本発明に係る特徴量算出方法を実行してもよい。

[0022] 上述した目的を達成するため、本発明の第7の態様に係るスクリーニング方法は複数の化合物から標的タンパク質に結合する標的化合物を抽出するスクリーニング方法であって、複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、化合物の立体構造について第2の態様に係る特徴量算出方法を用いて算出した第1の特徴量と、を関連付けて記憶する記憶工程と、標的タンパク質との結合が確認されている化合物であるリガンドについて第2の態様に係る特徴量算出方法を用いて第1の特徴量を算出する特徴量算出工程と、複数の化合物についての第1の特徴量とリガンドについての第1の特徴量との類似度を算出する類似度算出工程と、複数の化合物から類似度に基づいて標的化合物を抽出する化合物抽出工程と、を有する。第2の態様について上述したように、リガンドと対象化合物とで第1の特徴量（3次元で定量化した特徴量）が類似ならば、両者の薬効が類似している。したがって第7の態様によれば、第1の特徴量に基づいてリガンドと薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。なお、化合物抽出工程では類似度が目標値を満たす（例えば、しきい値以上である）化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0023] 上述した目的を達成するため、本発明の第8の態様に係るスクリーニング方法は複数の化合物から標的タンパク質に結合する標的化合物を抽出するスクリーニング方法であって、複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、第2の態様に係る特徴量算出方法を用いて算出した第1の特徴量と、を関連付けて記憶する記憶工程と、標的タンパク質のポケット構造体について、第3の態様に係る特徴量算出方法を用いて第2の特徴量を算出する特徴量算出工程と、複数の化合物についての第1の特徴量とポケット構造体についての第2の特徴量との類似度を算出する類似度算出工程と、複数の化合物から類似度に基づいて標的化合物を抽出する化合物抽

出工程と、を有する。

- [0024] 第3の態様について上述したように、ポケット構造体と対象化合物とで第2の特徴量が類似ならば、両者の化学的性質が類似している。したがって第8の態様によれば、ポケット構造体と化学的性質が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。なおポケット構造体は標的タンパク質と結合する化合物に対応するので、ポケット構造体についての特徴量（第2の特徴量）と化合物についての特徴量（第1の特徴量）とは比較対照が可能であり類似度を算出する。
- [0025] 上述した目的を達成するため、本発明の第9の態様に係るスクリーニング方法は複数の化合物からタンパク質以外の標的生体高分子に結合する標的化合物を抽出するスクリーニング方法であって、複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、化合物の立体構造について第4の態様に係る特徴量算出方法を用いて算出した第3の特徴量と、を関連付けて記憶する記憶工程と、タンパク質以外の標的生体高分子に結合することが確認されている化合物である結合化合物について第4の態様に係る特徴量算出方法を用いて第3の特徴量を算出する特徴量算出工程と、複数の化合物についての第3の特徴量と結合化合物についての第3の特徴量との類似度を算出する類似度算出工程と、複数の化合物から類似度に基づいて標的化合物を抽出する化合物抽出工程と、を有する。第4の態様について上述したように、本発明はタンパク質以外の標的生体高分子であるDNA等を扱うことができ、標的生体高分子に結合する結合化合物と対象化合物とで第3の特徴量が類似ならば、両者の薬効が類似している。したがって第9の態様によれば、第3の特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。
- [0026] 上述した目的を達成するため、本発明の第10の態様に係るスクリーニング方法は複数の化合物について立体構造を示す情報を入力する入力工程と、複数の化合物のそれぞれについて、情報に基づいて第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する特徴量算出工程と

、複数の化合物のそれぞれについて、情報に基づいて、物性を示す物性値及び／または毒性を示す毒性値を算出する特性値算出工程と、複数の化合物から、特徴量の類似度が目標値を満たし、かつ物性値及び／または毒性値が目標値を満たす化合物を抽出する抽出工程と、を有する。

[0027] 本発明に係る特徴量（第1から第4の態様に係る特徴量算出方法により算出される特徴量）によれば、既存の化合物に物性や毒性の面で問題がある場合でも、標的タンパク質等との結合力が同程度の（すなわち、特徴量が同程度の）新規化合物に作り変えることで、骨格が大きく変わり物性や毒性の問題を克服する可能性を高めることができる。そして、このような「結合力維持」と「物性や毒性の問題点回避」の両方を達成するには、本発明に係る特徴量と物性や毒性を示す指標とを組み合わせることが有効である。第10の態様はそのような着想に基づくスクリーニング方法の一態様を規定するもので、第1から第4の態様のいずれか1つに係る特徴量算出方法を用いて、複数の化合物から、特徴量の類似度が目標値を満たし、かつ物性値及び／または毒性値が目標値を満たす化合物を抽出するので、医薬候補化合物のスクリーニングを効率よく行うことができる。

[0028] なお第10の態様において、類似度を判断する化合物間では同一の態様により特徴量を算出する（特徴量算出方法を揃える）ことが好ましい。後述する第11～第13の態様においても同様である。

[0029] 上述した目的を達成するため、本発明の第11の態様に係るスクリーニング方法は複数の化合物について立体構造を示す情報を入力する入力工程と、複数の化合物のそれぞれについて、情報に基づいて第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する第1の特徴量算出工程と、複数の化合物のそれぞれについて、情報に基づいて化合物の構造を示す構造特徴量を算出する第2の特徴量算出工程と、複数の化合物のうちから、特徴量の類似度が目標値を満たし、かつ構造特徴量の類似度がしきい値以下である化合物を抽出する抽出工程と、を有する。

[0030] 探索により得られた新規化合物が探索の起点となった元の（既存の）化合

物と別骨格であることを確実にするためには、化合物の探索時に「元の化合物と（本発明に係る）特徴量は同程度であるが、構造式では違う」という条件を課すことが有効である。第11の態様はそのような着想に基づくスクリーニング方法の一態様を規定するもので、第1から第4の態様のいずれか1つに係る特徴量算出方法を用いて、複数の化合物のうちから、特徴量の類似度が目標値を満たし、かつ構造特徴量の類似度がしきい値以下である化合物を抽出するので、医薬候補化合物のスクリーニングを効率よく行うことができる。なお、「構造式では違う」ことを示すには構造式での類似性の指標を用いることが有効であるが、多数提案されている公知の指標をそのような指標として用いることができる。

[0031] 上述した目的を達成するため、本発明の第12の態様に係るスクリーニング方法は複数の化合物について立体構造を示す情報を入力する入力工程と、複数の化合物のそれぞれについて、情報に基づいて第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する特徴量算出工程と、複数の化合物のそれぞれについて、特徴量に基づいて他の化合物との結合力を出力する予測モデルを用いて結合力を予測する予測工程と、複数の化合物のうちから、結合力が目標値を満たす化合物を抽出する抽出工程と、を有する。第12の態様に係るスクリーニング方法は、探索の起点となる情報を一つの化合物の構造式だけから「本発明の特徴量に基づく結合力の予測モデル」及び「結合力の目標値（現状の値より高く設定することができる）」に変更することにより、結合力が目標値を満たす化合物を効率的に抽出する（医薬候補化合物のスクリーニングを効率よく行う）ことができる。

[0032] 上述した目的を達成するため、本発明の第13の態様に係るスクリーニング方法は複数の化合物について立体構造を示す情報を入力する入力工程と、他の化合物との結合力の目標値を設定する設定工程と、第1から第4の態様のいずれか1つに係る特徴量算出方法により算出された特徴量に基づいて結合力を出力する予測モデルを用いて、目標値に対応する特徴量を推定する推定工程と、複数の化合物のそれぞれについて、情報に基づいて第1から第4

の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する特徴量算出工程と、算出の結果に基づいて、複数の化合物のうちから、推定した特徴量との類似度がしきい値以上である特徴量を有する化合物を抽出する抽出工程と、を有する。上述した第12の態様では「予測モデル」を用いて結合力を予測し、「結合力の目標値」を満たす化合物を抽出している。これに対し第13の態様に係るスクリーニング方法では、「予測モデル」について第12の態様と逆の使い方をする。具体的には、「予測モデル」を用いて「結合力の目標値」から特徴量（第1から第4の態様のいずれか1つに係る特徴量）を推定し、推定した特徴量を持つ化合物を探索する。このような手法によっても、第12の態様と同様に医薬候補化合物のスクリーニングを効率よく行うことができる。

[0033] なお、第12、13の態様において、「結合力の予測モデル」は例えば複数の結合化合物の構造式とその結合力の実験値のセットから生成することができる。

[0034] 上述した目的を達成するため、本発明の第14の態様に係るスクリーニングプログラムは第7から第13の態様のいずれか1つに係るスクリーニング方法をコンピュータに実行させる。第14の態様における「コンピュータ」は、CPU等の各種プロセッサを1つ以上用いて実現することができる。この「コンピュータ」はスクリーニング装置、スクリーニングシステムに備えられたものでもよい。なお、第14の態様に係るスクリーニングプログラムのコンピュータ読み取り可能なコードが記録された非一時的記録媒体も、本発明の態様として挙げることができる。

[0035] 上述した目的を達成するため、本発明の第15の態様に係る化合物創出方法は複数の化合物から標的タンパク質に結合する標的化合物の立体構造を創出する化合物創出方法であって、複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、第2の態様に係る特徴量算出方法を用いて算出した第1の特徴量と、を関連付けて記憶する記憶工程と、標的タンパク質との結合が確認されている化合物であるリガンドについて、第2の態様に

係る特徴量算出方法を用いて第1の特徴量を算出する特徴量算出工程と、複数の化合物の立体構造を教師データとし、第1の特徴量を説明変数とした機械学習により生成器を構築する生成器構築工程と、生成器を用いて、リガンドの第1の特徴量から標的化合物の立体構造を生成する化合物立体構造生成工程と、を有する。

[0036] 上述した第7から第13の態様に係るスクリーニング方法では、構造式が既に決定されている（書き下されている）複数の化合物の中から、リガンドあるいは標的タンパク質に適合する化合物を見出している。このため、化合物の特徴量を算出した上で、別途算出したリガンドあるいは標的タンパク質のポケット構造体の特徴量との類似度に基づいて化合物を抽出する方策、つまり検索の方策を採る。したがって化合物の構造式と特徴量との対応関係を記録しておけば、類似度が高い（あるいはしきい値以上の）構造式を見出すことができる。これに対し第15の態様では、検索をせずに、リガンドの特徴量（第1の特徴量）に特徴量が類似した（したがって薬効が類似した）化合物の構造式を生成する。

[0037] 特徴量が与えられた場合の構造式の生成は、機械学習により構築した生成器を用いて行うことができる。具体的には、第15の態様では化合物の立体構造を教師データとし、第1の特徴量を説明変数とした機械学習（学習手法は特に限定されない）により生成器を構築し、この生成器を用いて、リガンドの第1の特徴量から標的化合物の立体構造を生成する。第15の態様では検索を行わないので、「スクリーニングによる検索の結果、解なし」となる場合でも化合物の立体構造を生成することができ、したがって医薬候補化合物の立体構造を効率よく創出することができる。なお、第15の態様において生成される立体構造は教師データとして与える化合物の特徴に影響される。したがって、教師データとして与える化合物の特徴を選択することで、特徴が異なる立体構造を有する化合物を生成できる。例えば、合成が容易な化合物を教師データとして与えることで、合成が容易な立体構造を有する化合物を生成することができる。

[0038] 上述した目的を達成するため、本発明の第16の態様に係る化合物創出方法は複数の化合物から標的タンパク質に結合する標的化合物の立体構造を創出する化合物創出方法であって、複数の化合物のそれぞれについて、第2の態様に係る特徴量算出方法を用いて算出した第1の特徴量と、を関連付けて記憶する記憶工程と、標的タンパク質のポケット構造体について、第3の態様に係る特徴量算出方法を用いて第2の特徴量を算出する特徴量算出工程と、複数の化合物の立体構造を教師データとし、第1の特徴量を説明変数とした機械学習により生成器を構築する生成器構築工程と、生成器を用いて、ポケット構造体の第2の特徴量から標的化合物の立体構造を生成する化合物立体構造生成工程と、を有する。第16の態様によれば、第15の態様と同様に、検索をせずに、ポケット構造体の特徴量（第2の特徴量）に特徴量が類似した（したがって薬効が類似した）化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。なお第15の態様と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成できる。

[0039] 上述した目的を達成するため、本発明の第17の態様に係る化合物創出方法は複数の化合物からタンパク質以外の標的生体高分子に結合する標的化合物の立体構造を創出する化合物創出方法であって、複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、第3の特徴量と、を関連付けて記憶する記憶工程と、タンパク質以外の標的生体高分子との結合が確認されている化合物である結合化合物について、第4の態様に係る特徴量算出方法を用いて第3の特徴量を算出する特徴量算出工程と、複数の化合物の立体構造を教師データとし、第3の特徴量を説明変数とした機械学習により生成器を構築する生成器構築工程と、生成器を用いて、結合化合物の第3の特徴量から標的化合物の立体構造を生成する化合物立体構造生成工程と、を有する。

[0040] 第17の態様によれば、第15、第16の態様と同様に、検索をせずに、結合化合物の特徴量（第3の特徴量）に特徴量が類似した（したがって薬効

が類似した)化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。なお第15、第16の態様と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

発明の効果

[0041] 以上説明したように、本発明の特徴量算出方法、特徴量算出プログラム、及び特徴量算出装置によれば、対象構造体の化学的性質を的確に示す特徴量を算出することができる。また、本発明のスクリーニング方法及びスクリーニングプログラムによれば医薬候補化合物のスクリーニングを効率よく行うことができる。また、本発明の化合物創出方法によれば医薬候補化合物の立体構造を効率よく創出することができる。

図面の簡単な説明

- [0042] [図1]図1は、第1の実施形態に係るスクリーニング装置の構成を示すブロック図である。
- [図2]図2は、処理部の構成を示すブロック図である。
- [図3]図3は、記憶部に記憶される情報を示す図である。
- [図4]図4は、化合物の構造情報と特徴量とを関連付けて記憶する様子を示す図である。
- [図5]図5は、化合物の3次元AAM特徴量の算出手順を示すフローチャートである。
- [図6]図6は、構造式の3次元化の例を示す図である。
- [図7]図7は、3次元AAM特徴量の例を示す図である。
- [図8]図8は、3次元AAM特徴量の例を示す表である。
- [図9]図9は、ポケット構造体についての3次元AAM記述子の手順を示すフローチャートである。
- [図10]図10は、ポケット構造体についての3次元AAM記述子の様子を示す概念図である。
- [図11]図11は、不変量化AAM記述子の例を示す図である。

[図12]図12は、不変量化AAM記述子が類似である化合物の例を示す図である。

[図13]図13は、不変量化AAM記述子を用いた場合のヒットのを見つけやすさを示す図である。

[図14]図14は、不変量化AAM記述子を用いた場合のヒットの探索時間を示す図である。

[図15]図15は、不変量化AAM記述子を用いた場合のヒットの探索時間を示す他の図である。

[図16]図16は、記述子の類似度に基づく標的化合物の抽出手順を示す図である。

[図17]図17は、記述子の類似度に基づく標的化合物の抽出結果の例を示す表である。

[図18]図18は、記述子の類似度に基づく標的化合物の抽出手順を示す他の図である。

[図19]図19は、記述子の類似度に基づく標的化合物の抽出結果の例を示す他の表である。

[図20]図20は、第2の実施形態に係る化合物創出装置の構成を示すブロック図である。

[図21]図21は、処理部の構成を示す図である。

[図22]図22は、記憶部に記憶される情報を示す図である。

[図23]図23は、リガンド入力の場合の立体構造生成手順を示すフローチャートである。

[図24]図24は、機械学習の結果を用いた立体構造生成の様子を示す図である。

[図25]図25は、中間層の数とcos類似度との関係を示す図である。

[図26]図26は、立体構造の生成例を示す図である。

[図27]図27は、標的タンパク質入力の場合の立体構造生成手順を示す図である。

[図28]図28は、第3の実施形態に係る医薬候補化合物探索装置の構成を示すブロック図である。

[図29]図29は、処理部の構成を示す図である。

[図30]図30は、記憶部に記憶される情報を示す図である。

[図31]図31は、ヒットの見つけやすさの比較結果を示す図である。

[図32]図32は、階層型ネットワークの構成例を示す図である。

[図33]図33は、本発明に係る特徴量と他の特徴量とを併用した場合のスクリーニングの処理を示すフローチャートである。

[図34]図34は、処理部の構成の他の例を示す図である。

[図35]図35は、本発明に係る特徴量と物性値との併用によるスクリーニングの様子を示す図である。

[図36]図36は、本発明に係る特徴量と物性値との併用によるスクリーニングの様子を示す他の図である。

[図37]図37は、本発明に係る特徴量と物性値との併用によるスクリーニングの様子を示すさらに他の図である。

[図38]図38は、本発明に係る特徴量と他の特徴量とを併用した場合のスクリーニングの処理を示すフローチャートである。

[図39]図39は、処理部の構成の他の例を示す図である。

[図40]図40は、本発明に係る特徴量と構造特徴量との併用によるスクリーニングの様子を示す図である。

[図41]図41は、本発明に係る特徴量と構造特徴量との併用によるスクリーニングの様子を示す他の図である。

[図42]図42は、本発明に係る特徴量と構造特徴量との併用によるスクリーニングの様子を示すさらに他の図である。

[図43]図43は、本発明に係る特徴量と他の特徴量とを併用した場合の出力例を示す図である。

[図44]図44は、本発明に係る特徴量と他の特徴量とを併用した場合の出力例を示す他の図である。

[図45]図45は、本発明に係る特徴量及び予測モデルを用いた場合のスクリーニングの処理を示す他のフローチャートである。

[図46]図46は、処理部の構成のさらに他の例を示す図である。

[図47]図47は、本発明に係る特徴量及び予測モデルを用いた場合のスクリーニングの処理を示すさらに他のフローチャートである。

[図48]図48は、本発明に係る特徴量と他の特徴量とを併用した場合の化合物創出の処理を示すフローチャートである。

[図49]図49は、本発明に係る特徴量と他の特徴量とを併用した場合の化合物創出の処理を示す他のフローチャートである。

[図50]図50は、本発明に係る特徴量と他の特徴量とを併用した場合の化合物創出の処理を示すさらに他のフローチャートである。

[図51]図51は、本発明に係る特徴量と他の特徴量とを併用した場合の化合物創出の処理を示すさらに他のフローチャートである。

[図52]図52は、本発明に係る特徴量と構造特徴量とを併用した場合の化合物創出の処理を示すフローチャートである。

[図53]図53は、本発明に係る特徴量と構造特徴量とを併用した場合の化合物創出の処理を示す他のフローチャートである。

[図54]図54は、本発明に係る特徴量と構造特徴量とを併用した場合の化合物創出の処理を示すさらに他のフローチャートである。

[図55]図55は、本発明に係る特徴量と構造特徴量とを併用した場合の化合物創出の処理を示すさらに他のフローチャートである。

[図56]図56は、本発明に係る特徴量及び予測モデルを用いた場合の化合物創出の処理を示すフローチャートである。

[図57]図57は、本発明に係る特徴量及び予測モデルを用いた場合の化合物創出の処理を示す他のフローチャートである。

[図58]図58は、本発明に係る特徴量及び予測モデルを用いた場合の化合物創出の処理を示すさらに他のフローチャートである。

[図59]図59は、本発明に係るスクリーニングの処理を示すフローチャート

である。

[図60]図60は、本発明に係る化合物創出の処理を示すフローチャートである。

[図61]図61は、本発明に係るスクリーニングの処理を示す他のフローチャートである。

[図62]図62は、本発明に係る化合物創出の処理を示す他のフローチャートである。

[図63]図63は、本発明に係るスクリーニングの処理を示すさらに他のフローチャートである。

[図64]図64は、本発明に係る化合物創出の処理を示すさらに他のフローチャートである。

[図65]図65は、本発明に係る化合物情報処理を示すさらに他のフローチャートである。

[図66]図66は、アミノ酸の集積度合いの出力例を示す図である。

[図67]図67は、アミノ酸の集積度合いの出力例を示す他の図である。

[図68]図68は、アミノ酸の集積度合いの2次元表示の例を示す図である。

[図69]図69は、アミノ酸の集積度合いの2次元表示の例を示す他の図である。

[図70]図70は、アミノ酸の集積度合いの2次元表示の例を示すさらに他の図である。

[図71]図71は、アミノ酸の集積度合いのグラフ表示の例を示す図である。

[図72]図72は、アミノ酸の集積度合いのグラフ表示の例を示す他の図である。

[図73]図73は、特徴量の疑似的3次元表示の例を示す図である。

[図74]図74は、アミノ酸の集積度合いの差分を類似度として表現した様子を示す図である。

[図75]図75は、結合力の予測モデルの性能の出力例を示す図である。

[図76]図76は、本発明に係る化合物探索システムの構成例を示す図である。

。

[図77]図77は、化合物探索サーバの機能構成の例を示す図である。

[図78]図78は、クライアント装置の機能構成の例を示す図である。

[図79]図79は、化合物創出サーバの機能構成の例を示す図である。

[図80]図80は、化合物探索サーバの機能構成の他の例を示す図である。

発明を実施するための形態

[0043] 以下、添付図面を参照しつつ、本発明の特徴量算出方法、特徴量算出プログラム、特徴量算出装置、スクリーニング方法、スクリーニングプログラム、及び化合物創出装置の実施形態について、詳細に説明する。

[0044] <第1の実施形態>

図1は第1の実施形態に係るスクリーニング装置10（特徴量算出装置、スクリーニング装置）の構成を示すブロック図である。スクリーニング装置10は化合物（対象構造体）及び／またはポケット構造体（対象構造体）についての特徴量の算出、及び標的化合物の抽出（スクリーニング）を行う装置であり、コンピュータを用いて実現することができる。図1に示すように、スクリーニング装置10は処理部100、記憶部200、表示部300、及び操作部400を備え、互いに接続されて必要な情報が送受信される。これらの構成要素については各種の設置形態を採用することができ、各構成要素が1箇所（1筐体内、1室内等）に設置されていてもよいし、離れた場所に設置されネットワークを介して接続されていてもよい。また、スクリーニング装置10はインターネット等のネットワークNWを介して外部サーバ500、及びPDB（Protein Data Bank）等の外部データベース510に接続し、必要に応じて化合物の構造式、タンパク質の結晶構造等の情報を取得することができる。

[0045] <処理部の構成>

図2は処理部100の構成を示す図である。処理部100は情報入力部110、特徴量算出部120、類似度算出部130、化合物抽出部140、表示制御部150、CPU160（CPU：Central Processing Unit）、RO

M 1 7 0 (R O M : Read Only Memory)、及びR A M 1 8 0 (R A M : Random Access Memory)を備える。

- [0046] 情報入力部 1 1 0 は、不図示のDVDドライブ、半導体メモリ用端子等の記録媒体インタフェース及び／またはネットワークNWを介して化合物の構造、標的タンパク質のX結晶構造及びポケット位置等の情報を入力する。特徴量算出部 1 2 0 (対象構造体指定部、立体構造発生部、特徴量算出部、不変量化部)は、本発明に係る特徴量を算出する。類似度算出部 1 3 0 (類似度算出部)は、算出した特徴量どうしの類似度を算出する。化合物抽出部 1 4 0 (化合物抽出部)は、複数の化合物から類似度に基づいて標的化合物を抽出する。表示制御部 1 5 0 は、入力した情報及び処理結果のモニタ 3 1 0 への表示を制御する。処理部 1 0 0 のこれらの機能を用いた特徴量算出及び標的化合物のスクリーニングの処理については、詳細を後述する。なお、これらの機能による処理はC P U 1 6 0 の制御の下で行われる。
- [0047] 上述した処理部 1 0 0 の各部の機能は、各種のプロセッサ (processor) を用いて実現できる。各種のプロセッサには、例えばソフトウェア (プログラム) を実行して各種の機能を実現する汎用的なプロセッサであるC P U が含まれる。また、上述した各種のプロセッサには、画像処理に特化したプロセッサであるG P U (Graphics Processing Unit)、F P G A (Field Programmable Gate Array) などの製造後に回路構成を変更可能なプロセッサであるプログラマブルロジックデバイス (Programmable Logic Device : P L D) も含まれる。さらに、A S I C (Application Specific Integrated Circuit) などの特定の処理を実行させるために専用に設計された回路構成を有するプロセッサである専用電気回路なども上述した各種のプロセッサに含まれる。
- [0048] 各部の機能は1つのプロセッサにより実現されてもよいし、同種または異種の複数のプロセッサ (例えば、複数のF P G A、あるいはC P U とF P G A の組み合わせ、またはC P U とG P U の組み合わせ) で実現されてもよい。また、複数の機能を1つのプロセッサで実現してもよい。複数の機能を1つのプロセッサで構成する例としては、第1に、クライアント、サーバなど

のコンピュータに代表されるように、1つ以上のCPUとソフトウェアの組合せで1つのプロセッサを構成し、このプロセッサが複数の機能として実現する形態がある。第2に、システムオンチップ (System On Chip: SoC) などに代表されるように、システム全体の機能を1つのIC (Integrated Circuit) チップで実現するプロセッサを使用する形態がある。このように、各種の機能は、ハードウェア的な構造として、上述した各種のプロセッサを1つ以上用いて構成される。さらに、これらの各種のプロセッサのハードウェア的な構造は、より具体的には、半導体素子などの回路素子を組み合わせた電気回路 (circuitry) である。

[0049] 上述したプロセッサあるいは電気回路がソフトウェア (プログラム) を実行する際は、実行するソフトウェアのコンピュータ (例えば、処理部100を構成する各種のプロセッサや電気回路、及び／またはそれらの組み合わせ) で読み取り可能なコードをROM170 (図2を参照) 等の非一時的記録媒体に記憶しておき、プロセッサがそのソフトウェアを参照する。非一時的記録媒体に記憶しておくソフトウェアは、本発明に係る特徴量算出方法及び標的化合物の抽出処理を実行するためのプログラム (特徴量算出プログラム及びスクリーニングプログラム) を含む。ROM170ではなく各種光磁気記録装置、半導体メモリ等の非一時的記録媒体にコードを記録してもよい。ソフトウェアを用いた処理の際には例えばRAM180が一時的記憶領域として用いられ、また例えば不図示のEEPROM (Electrically Erasable and Programmable Read Only Memory) に記憶されたデータを参照することもできる。

[0050] <記憶部の構成>

記憶部200はDVD (Digital Versatile Disk)、ハードディスク (Hard Disk)、各種半導体メモリ等の非一時的記録媒体及びその制御部により構成され、図3に示す画像及び情報が記憶される。構造情報210は化合物の構造式、標的タンパク質の立体構造及びポケット位置を含む。立体構造情報220は、構造情報210から発生させた化合物及び／またはポケット構造

体の立体構造の情報である。3次元AAM記述子230は、化合物またはポケット構造体の立体構造の周辺における1種類以上のアミノ酸の集積度合いを3次元空間において定量化した特徴量であり、後述する特徴量算出方法により算出される。なお、「AAM」は「アミノ酸マッピング (Amino Acid Mapping)」を意味する。不変量化AAM記述子240は、3次元AAM記述子230を化合物またはポケット構造体の回転及び並進について不変量化した特徴量である。類似度情報250は特徴量どうしの類似度を示す情報であり、化合物抽出結果260は類似度に基づいて抽出した標的化合物を示す情報である。

[0051] 図4は、N個 (Nは2以上の整数) の化合物について、構造情報210、立体構造情報220、3次元AAM記述子230、及び不変量化AAM記述子240が関連付けられて記憶部200に記憶される様子を示す図である。図4において、例えば構造式を構造情報210とし、立体化した構造式 (後述) を立体構造情報220とすることができる。また図4では、各化合物に対し、20種類のアミノ酸のそれぞれについて、3次元AAM記述子230 ($g_a(r)$) のように記載; aはアミノ酸の種類を表す添字) と、その3次元AAM記述子230に対応する不変量化AAM記述子240 ($F_{ab}(s)$) のように記載; a, bはアミノ酸の種類を表す添字) を関連付けて記憶している。3次元AAM記述子230及び不変量化AAM記述子240は、20種類のアミノ酸すべてについてではなく、スクリーニングに用いる記述子の数に応じて一部のアミノ酸について記憶してもよい。

[0052] 記憶部200において、図4に示すような情報のセット (ライブラリ) を複数記憶してもよい。なお、図4では化合物についての情報の記憶の様子を示したが、標的タンパク質についても同様の構成で情報を記憶することができる。また、このような構造情報及び立体構造情報を用いた3次元AAM記述子及び/または不変量化AAM記述子の算出方法は後述する。

[0053] <表示部及び操作部の構成>

表示部300はモニタ310 (表示装置) を備えており、入力した画像、

記憶部 200 に記憶された画像及び情報、処理部 100 による処理の結果等を表示することができる。操作部 400 は入力デバイス及び／またはポインティングデバイスとしてのキーボード 410 及びマウス 420 を含んでおり、ユーザはこれらのデバイス及びモニタ 310 の画面を介して、本発明に係る特徴量算出方法の実行及び標的化合物の抽出に必要な操作を行うことができる（後述）。ユーザが実行できる操作には、例えば処理モード、算出する記述子の種類、スクリーニングに用いる記述子、類似度に対するしきい値の指定等が含まれる。

[0054] <スクリーニング装置における処理>

上述した構成のスクリーニング装置 10 では、操作部 400 を介したユーザの指示に応じて、特徴量（記述子）の算出及び／または標的化合物の抽出を行うことができる。以下、各処理の詳細を説明する。

[0055] <特徴量の算出>

スクリーニング装置 10 は、操作部 400 を介したユーザの指示に応じて、3次元 AAM 記述子及び／または不変量化 AAM 記述子を算出することができる。

[0056] <化合物に対する 3 次元 AAM 記述子の算出>

図 5 は、化合物（対象構造体）についての 3 次元 AAM 記述子の算出手順を示すフローチャートである。なおリガンドは標的タンパク質との結合が確認されている化合物であり、図 5 の手順で 3 次元 AAM 記述子を算出することができる。ステップ S100 で、情報入力部 110 はユーザの操作に応じて化合物の構造式を入力する。これにより、入力した化学式で表される化合物が対象構造体として指定される（対象構造体指定工程）。

[0057] 特徴量算出部 120 は、入力した構造式を 3 次元化して、複数の原子（化学的性質を有する複数の単位構造体）による化合物の立体構造を発生させる（ステップ S102：立体構造発生工程）。構造式の 3 次元化は種々の手法が知られており、ステップ S102 で用いる手法は特に限定されない。図 6 は構造式の立体化の例、（a）部分は入力した構造式、（b）部分は 3 次元

化した構造式を示す。

[0058] 特徴量算出部 120 は、アミノ酸 “a” (a はアミノ酸の種類を表す数字；1 から 20) の各原子 “ μ ” が感じる自由エネルギーの空間分布 $\Delta G_{a\mu}(r)$ を算出する (ステップ S104；特徴量算出工程)。 $\Delta G_{a\mu}(r)$ の算出方法としては分子動力学法 (MD: Molecular Dynamics) を採用することができるが、これに限定されない。特徴量を算出するアミノ酸はあらかじめ決められた種類でもよいし、ユーザの指示に応じて決定してもよい (1 種類以上であればよく、複数種類でもよい)。

[0059] 特徴量算出部 120 は、 $\Delta G_{a\mu}(r)$ からアミノ酸 “a” の各原子 “ μ ” の分布関数 $g_{a\mu}(r)$ を算出する (ステップ S106；特徴量算出工程)。 $g_{a\mu}(r)$ は、T を室温、 K_B をボルツマン定数とすると、以下の式 (1) により表される。

[0060] [数1]

$$g_{a\mu}(\mathbf{r}) = \exp(-\Delta G_{a\mu}(\mathbf{r}) / K_B T) \quad \dots (1)$$

[0061] 特徴量算出部 120 は、分布関数 $g_{a\mu}(r)$ からアミノ酸の重心の分布関数 $g_a(r)$ を算出する (ステップ S108；特徴量算出工程)。算出のためには、 $g_{a\mu}(r)$ を各原子 “ μ ” について相乗平均する。この分布関数 $g_a(r)$ が、化合物の立体構造の周辺における 1 種類以上のアミノ酸 “a” の集積度合いを 3 次元空間において定量化した 3 次元 AAM 記述子である。特徴量算出部 120 は、算出した 3 次元 AAM 記述子を、3 次元 AAM 記述子 230 として化合物の構造情報 (構造情報 210)、立体構造情報 (立体構造情報 220) と関連付けて記憶部 200 に記憶する (図 4 参照)。

[0062] 図 7 は、図 6 に示す化合物についての 3 次元 AAM 記述子の例である。(a) 部分はアラニンについての 3 次元 AAM 記述子を示し、(b) 部分はバリリンについての 3 次元 AAM 記述子を示す。図 7 において色の濃い領域は、アミノ酸の集積度合い (存在確率) が高い領域である。図 8 は図 6 に示す化合物についての 3 次元 AAM 記述子の他の例を示す表であり、それぞれ異なる

る方向1、方向2、方向3（それぞれ表の1段目、2段目、3段目）についての3次元AAM記述子を示す。表の左欄は3次元AAM記述子（しきい値に対する等高面）を示し、右欄は3次元AAM記述子（しきい値に対する等高面）及び化合物の立体構造を示す。

[0063] <ポケット構造体に対する3次元AAM記述子の算出>

スクリーニング装置10では、化合物ではなく標的タンパク質に結合するポケット構造体を対象構造体として指定し、このポケット構造体に対する特徴量（3次元AAM記述子）を算出することができる。ポケット構造体は標的タンパク質の活性部位であるポケットに結合する対象構造体であり、「活性部位」とはポケット構造体が結合することにより標的タンパク質の活性が促進または抑制される部位を意味する。図9はポケット構造体に対する3次元AAM記述子の算出手順を示すフローチャートである。また、図10はポケット構造体に対する3次元AAM記述子の様子を示す概念図である。

[0064] 図9のフローチャートでは、情報入力部110が標的タンパク質の立体構造の実測及びポケットの位置情報を入力する（ステップS200：対象構造体指定工程）。図10の（a）部分は、標的タンパク質TPにおけるポケットPOを示す。ステップS200の処理によりポケット構造体が対象構造として指定される。

[0065] 特徴量算出部120は、標的タンパク質のポケットに複数の仮想的球体（化学的性質を有する複数の単位構造体）を詰める（ステップS202：対象構造体指定工程、立体構造発生工程）。「仮想的球体」はファンデルワールス半径、電荷等の化学的性質を有すると考えることができ、「仮想的球体を詰める」のはシミュレーション（例えば分子動力学法）により行うことができる。ステップS202により、詰められた仮想的球体の集まり（立体構造）をポケット構造体（対象構造体）の立体構造として得ることができる（ステップS204：立体構造発生工程）。図10の（b）部分に、標的タンパク質TPに対するポケット構造体PSの例を示す。

[0066] 特徴量算出部120は、標的タンパク質の立体構造の実測を用いて、ポケ

ット構造体の周辺への１種類以上のアミノ酸の集積度合いを３次元で定量化する（ステップＳ２０６：特徴量算出工程）。実際に、ポケット構造体の周辺にどの種類のアミノ酸が集積しているか読み出すことが可能である。図１０の（ｃ）部分は、ポケット構造体ＰＳの周辺に３種類のアミノ酸Ａ１，Ａ２，Ａ３が集積している様子を示す。なお集積度合いを定量化するアミノ酸は１種類以上であればよい（複数種類でもよい）。またあらかじめ決められた種類のアミノ酸について定量化してもよいし、ユーザの操作に応じて設定したアミノ酸について定量化してもよい。特徴量算出部１２０は、算出した３次元ＡＡＭ記述子を、３次元ＡＡＭ記述子２３０として化合物の構造情報（構造情報２１０）、立体構造情報（立体構造情報２２０）と関連付けて記憶部２００に記憶する（図３，４参照；記憶工程）。後述する不変量化ＡＡＭ記述子が算出されている場合は、特徴量算出部１２０は３次元ＡＡＭ記述子と不変量化ＡＡＭ記述子との関連付けを行う。

[0067] ＜３次元ＡＡＭ記述子の不変量化＞

上述した３次元ＡＡＭ記述子はアミノ酸の３次元的な集積度合いを示しているが、化合物が同じでも重心移動、回転等が起きると値が変化し、また３次元情報であるのでデータ容量が大きい。そこで第１の実施形態に係るスクリーニング装置１０では、３次元ＡＡＭ記述子に加え、またはこれに代えて「３次元ＡＡＭ記述子を化合物の回転及び並進に対して不変量化した不変量化ＡＡＭ記述子」（不変量化特徴量）を算出することができる。なお、化合物の場合もポケット構造体の場合も、同じ手順で不変量化を行うことができる。化合物についての３次元ＡＡＭ記述子を用いた場合は化合物についての不変量化ＡＡＭ記述子が得られ、ポケット構造体についての３次元ＡＡＭ記述子を用いた場合はポケット構造体についての不変量化ＡＡＭ記述子が得られる。

[0068] 特徴量算出部１２０（不変量化部）は、以下の式（２）に示すようにフーリエ変換を用いて $f_a(k)$ を算出する（不変量化工程）。上述したように、“ a ”はアミノ酸の種類を示す添字（１～２０）である。また、“ i ”は虚数

単位である。

[0069] [数2]

$$f_a(\mathbf{k}) = \int d^3\mathbf{r} (g_a(\mathbf{r}) - 1) e^{i\mathbf{k}\cdot\mathbf{r}} \quad \dots (2)$$

[0070] 特徴量算出部120は、この $f_a(\mathbf{k})$ を用いて以下の式(3)により不変量化AAM記述子である $F_{ab}(s)$ (不変量化特徴量)を算出することができる(不変量化工程)。式(3)では、2種類の異なるアミノ酸(“a”, “b”で示す)についての3次元AAM記述子($g_a(r)$, $g_b(r)$)を用いて、相関関数の角度積分により不変量化AAM記述子を算出する。なお、20種類のアミノ酸のうちで不変量化AAM記述子の算出に用いる2種類のアミノ酸の組合せは特に限定されない。

[0071] [数3]

$$F_{ab}(s) = \int d^3\mathbf{k} f_a(-\mathbf{k}) f_b(+\mathbf{k}) \delta(\mathbf{k}^2 - s) \quad \dots (3)$$

[0072] 式(3)では不変量化の際にデルタ関数を用いているが、以下の式(4)に示すように、任意の関数($h(\mathbf{k}^2 - s)$)を用いて不変量化を行うことができる。

[0073] [数4]

$$F_{ab}(s) = \int d^3\mathbf{k} f_a(-\mathbf{k}) f_b(+\mathbf{k}) h(\mathbf{k}^2 - s) \quad \dots (4)$$

[0074] このようにして算出した不変量化AAM記述子の例を図11に示す。図11の(a)部分是不変量化AAM記述子である $F_{12}(s)$ (アミノ酸1, 2についての不変量化AAM記述子)の実部であり、(b)部分は虚部である。このように2種類の異なるアミノ酸についての3次元AAM記述子を用いて不変量化を行うことで、アミノ酸の相互作用の情報を維持しつつ不変量化を行うことができ、特徴量(不変量化特徴量)に基づく化合物の比較(薬効判定)を正確に行うことができる。

[0075] 特徴量算出部120は、算出した不変量化AAM記述子を、不変量化AAM記述子240として化合物の構造情報(構造情報210)、立体構造情報

(立体構造情報 220)、及び元の 3 次元 AAM 記述子 230 と関連付けて記憶部 200 に記憶する (図 3, 4 参照; 記憶工程)。なお、第 1 の実施形態では 2 種類の異なるアミノ酸についての 3 次元 AAM 記述子を用いて不変量化 AAM 記述子を算出するので、3 次元 AAM 記述子と不変量化 AAM 記述子との関連付けも複数ありうる。

[0076] <不変量化 AAM 記述子の有効性評価>

上述の処理により算出した不変量化 AAM 記述子の有効性を説明する。

[0077] <類似の不変量化 AAM 記述子を有する化合物の活性の例>

図 12 の (a) 部分は、タンパク質 ABL1 (標的タンパク質の一例) に対するリガンドの構造式を示す。このリガンドの結合力は、IC50 (50 % 阻害濃度) で 1 μ M レベルである。これに対し、図 12 の (b) 部分はリガンドとほぼ同じ不変量化 AAM 記述子を持つ化合物の構造式である。この化合物の活性を実測すると、リガンドと同じレベルであった。すなわち、図 11 は不変量化 AAM 記述子が類似の化合物が類似の薬効を有することを示す例である。このように、第 1 の実施形態によれば対象構造体の化学的性質を的確に示す特徴量 (不変量化 AAM 記述子) を得ることができる。なお、結合力の強さを表す指標や単位は、二通りに分類することができる。一つは結合エネルギーで、単位としては kJ/mol や kcal/mol が使われる。もう一つは上述した IC50 あるいは Kd (結合解離定数) であり、単位は濃度の M (=mol/L; モーラー) が用いられる。

[0078] <ヒットの見つけやすさ>

以下の手順 1 ~ 5 により、不変量化 AAM 記述子によるヒットの見つけやすさを評価した。

(手順 1) あるタンパク質 (標的タンパク質) に対して、ヒット化合物 X 個とヒットでない化合物 Y 個とを混ぜる。

(手順 2) (X + Y) 個の化合物すべてについて、不変量化 AAM 記述子を計算する。

(手順 3) 記述子ごとの類似度を算出する。

(手順4) (X+Y) 個の化合物を、不変量化 AAM 記述子の類似度に基づいてチーム分けする。

(手順5) ヒットが集まったチームが機械的に発生するかどうかチェックする。

[0079] タンパク質 ABL1 (キナーゼ) に対し 183 個のヒットを含む 10,933 個の化合物 (ヒット含有率 1.6%) を上述の手順によりチーム分けしたところ、221 チームに分かれた。あるチームは 16 個のヒットとその他の化合物を 14 個含んでおり、ヒット含有率は 53.3% であった。またこのチームは図 12 の (a) 部分に示す化合物及び (b) 部分に示す化合物を含んでいたが、従来の記述子である Fingerprint を用いた場合、これら化合物の類似度は 25% であり、本来ならばヒットであるにも関わらず別物と認識されていた。このように、上述のチームにおいて本発明の第 1 の実施形態に係る不変量化 AAM 記述子を用いた場合、Fingerprint ではまともになかったヒットでも同一チームに属することが分かる。

[0080] 上述した 221 チームについて、チームごとのヒットの見つけやすさ (= 期待値; ヒット数 × ヒット含有率) を求めた結果を図 13 に示す。比較のため、ランダムにチーム分けした場合、及び Fingerprint を用いてチーム分けした場合の結果を示す。この結果より、上述の化合物群に対し不変量化 AAM 記述子を用いた場合、ランダムなチーム分けや Fingerprint を用いたチーム分けよりも多くのヒットを含むチームが発生することが分かる。なお、図 13 においてチーム番号はチーム分け方法 (ランダム、不変量化 AAM 記述子、Fingerprint) によって異なるので、チーム分けの優劣は、同一のチーム番号での期待値を比較するのではなく「期待値の高い (より多くのヒットを含む) チームが含まれるかどうか」により判断している。

[0081] < ヒット探索時間 (その 1) >

図 14 は、上述した化合物群についての Importance Sampling (重点サンプリング) のシミュレーション結果を示すグラフである。第 1 の実施形態に係る不変量化 AAM 記述子を用いた場合、ランダムなチーム分けの場合と比較

したヒット探索時間（同じヒット数を発見するための薬効評価の回数）は、50%探索の場合で約2分の1、25%探索の場合約4分の1に短縮された。一方、Fingerprintを用いたチーム分けの場合、ヒット探索時間は短縮しなかった。なお、ここでのImportance Samplingのシミュレーションとは、チームごとに優先度を表す変数を与え、測定ごとに、ヒットが出たチームの優先度を上げ、ヒットが出なかったチームの優先度を下げないように、変数を更新しながら、より少ない測定回数でより多くのヒットを集める方法を意味する。図14では優先度の制御に機械学習手法の一種であるNB法（NB: Naive Bayes）を用いた例を示したが、これに限定する必要はない。

[0082] <ヒット探索時間（その2）>

タンパク質AA2ARに対する32,464個の化合物（483個のヒットを含む）について、（その1）と同様にImportance Sampling（重点サンプリング）のシミュレーション結果を図15に示す。タンパク質AA2ARはX線結晶構造を得るのが困難な膜タンパクであるが、そのような膜タンパクの場合でも、ヒット探索時間は50%探索で約2分の1まで短縮された。

[0083] 上述した不変量化AAM記述子によれば、記述子が類似の化合物は類似した薬効（標的タンパク質への結合）を示すので、対象構造体（化合物、ポケット構造体）の化学的性質を的確に示す。3次元AAM記述子を変量化した不変量化AAM記述子によれば、2種類の異なるアミノ酸についての3次元AAM記述子を用いて不変量化を行うことで、記述子に基づく化合物の比較（薬効判定）を正確に行いつつ、扱いやすくデータ容量を削減することができる。さらに不変量化AAM記述子によればヒットを見つけやすく、探索を高速化することができる。

[0084] <特徴量算出方法及び特徴量算出プログラムの効果>

以上説明したように、第1の実施形態に係るスクリーニング装置10では、本発明に係る特徴量算出方法及び特徴量算出プログラムを用いて、対象構造体の化学的性質を的確に示す特徴量（3次元AAM記述子、不変量化AAM記述子）を算出することができる。

[0085] <標的化合物の抽出（スクリーニング）>

上述した３次元ＡＡＭ記述子、不変量化ＡＡＭ記述子を用いた、複数の化合物からの標的化合物（医薬候補化合物）の抽出について説明する。標的化合物の抽出はリガンドの記述子（３次元ＡＡＭ記述子、不変量化ＡＡＭ記述子）に基づいて行うモード（第１のモード）と標的タンパク質のポケット構造体の記述子（３次元ＡＡＭ記述子、不変量化ＡＡＭ記述子）に基づいて行うモード（第２のモード）がある。いずれのモードにより抽出を行うかは、操作部４００を介したユーザの操作に応じて選択することができる。

[0086] <リガンド入力のスクリーニング>

図１６はリガンドの３次元ＡＡＭ記述子を用いたスクリーニングの手順を示すフローチャートである。処理が開始すると、特徴量算出部１２０はリガンドの３次元ＡＡＭ記述子を算出する（ステップＳ３００：特徴量算出工程）。なおリガンドは標的タンパク質との結合が確認されている化合物なので、ステップＳ３００における３次元ＡＡＭ記述子の算出は図５のフローチャートに示す手順により算出することができる。

[0087] 図４について上述したように、スクリーニング装置１０では、複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、この立体構造に対応する３次元ＡＡＭ記述子とが関連付けて記憶部２００に記憶されている。類似度算出部１３０は、化合物についての３次元ＡＡＭ記述子と、ステップＳ３００で算出したリガンドの３次元ＡＡＭ記述子との類似度を算出する（ステップＳ３０２：類似度算出工程）。類似度を算出したら、化合物抽出部１４０は類似度に基づいて標的化合物を抽出する（ステップＳ３０４：標的化合物抽出工程）。上述のように３次元ＡＡＭ記述子が類似であれば類似の薬効（標的タンパク質への結合）を示すので、３次元ＡＡＭ記述子の類似度を用いることにより、リガンドと類似の薬効を有する化合物（すなわち、医薬候補である標的化合物）を抽出することができる。なお類似度に基づく標的化合物の抽出（ステップＳ３０４）は、具体的には「類似度がしきい値以上の化合物を抽出する」、「類似度が高い順に化合物を抽出する」等

により行うことができる。

[0088] 図16では3次元AAM記述子を用いたスクリーニングの手順について説明しているが、不変量化AAM記述子を用いたスクリーニングも同様の手順で行うことができる。具体的には、特徴量算出部120が図5の手順及び上述した式(2)、(3)によりリガンドの不変量化AAM記述子(不変量化特徴量)を算出し、類似度算出部130が、記憶部200に記憶されている化合物の不変量化AAM記述子との類似度を算出する。類似度を算出したら、化合物抽出部140が類似度に基づいて標的化合物を抽出する。類似度に基づく標的化合物抽出の具体的態様は、3次元AAM記述子と同様に行うことができる。

[0089] 図17はリガンド入力のスクリーニング結果の例を示す表である。図17の(a)部分は3次元AAM記述子を用い「類似度がしきい値以上の化合物を抽出する」とした場合の結果を示し、(b)部分は不変量化AAM記述子を用い「類似度が高い順に化合物を抽出する」とした場合の結果を示す。なお、図17の(a)部分ではアミノ酸1についての3次元AAM記述子($g_1(r)$)に基づいて化合物を抽出しているが、他のアミノ酸(アミノ酸2~20)についての3次元AAM記述子(例えば、 $g_2(r)$)に基づいて化合物を抽出してもよい。また、異なるアミノ酸についての複数の3次元AAM記述子(例えば、 $g_1(r)$ と $g_2(r)$)の類似度($g_1(r)$ どうしの類似度と $g_2(r)$ どうしの類似度)をそれぞれ算出し、これに基づいて化合物を抽出してもよい。化合物の抽出に用いる3次元AAM記述子は1種類でよいが、複数種類の3次元AAM記述子を用いることにより、類似度に基づく化合物の抽出を正確に行うことができる。なお、複数種類の3次元AAM記述子を用いる場合、それら記述子の間でのアミノ酸の組み合わせは特に限定されない(例えば、 $g_1(r)$ と $g_2(r)$ でもよいし、 $g_3(r)$ と $g_4(r)$ でもよい)。

[0090] 同様に、図17の(b)部分ではアミノ酸1, 2についての不変量化AAM記述子($F_{12}(s)$)に基づいて化合物を抽出しているが、不変量化AAM記述子の算出を行うアミノ酸は他の組合せ(例えば、アミノ酸3, 4による

$F_{34}(s)$))であってもよい。また、アミノ酸の組合せが異なる複数の不変量化AAM記述子(例えば、 $F_{12}(s)$ と $F_{34}(s)$))に基づいて化合物の抽出を行ってもよい(例えば、 $F_{12}(s)$ どうしの類似度と $F_{34}(s)$ どうしの類似度を用いる)。化合物の抽出に用いる不変量化AAM記述子は1種類でよいが、複数種類の不変量化AAM記述子を用いることにより、類似度に基づく化合物の抽出を正確に行うことができる。なお、複数種類の不変量化AAM記述子を用いる場合、それら記述子の間でのアミノ酸の組み合わせは特に限定されない(例えば、 $F_{12}(s)$ と $F_{34}(s)$ でもよいし、 $F_{12}(s)$ と $F_{13}(s)$ でもよい)。いずれのアミノ酸について記述子及び類似度を算出するかは、操作部400を介したユーザの指示に応じて処理部100(特徴量算出部120、類似度算出部130、化合物抽出部140)が決定してもよいし、ユーザの指示によらず処理部100が決定してもよい。

[0091] なお、(a)部分では類似度のしきい値を80%とし(b)部分では抽出個数を100個としているが、これらの値は例示であり、スクリーニングの精度等の条件に応じてしきい値及び抽出個数を設定することができる。設定は、操作部400を介したユーザの入力に応じて行うことができる。また、図17と逆に3次元AAM記述子を用いた場合に「類似度が高い順に化合物を抽出する」とし、不変量化AAM記述子を用いた場合に「類似度がしきい値以上の化合物を抽出する」としてもよい。化合物抽出部140は、図17に示すような抽出結果を、化合物抽出結果260として記憶部200に記憶させる(図3参照)。

[0092] <標的タンパク質入力のスクリーニング>

図18は標的タンパク質のポケット構造体についての3次元AAM記述子を用いたスクリーニングの手順を示すフローチャートである。処理が開始すると、特徴量算出部120は標的タンパク質のポケット構造体についての3次元AAM記述子を算出する(ステップS400:特徴量算出工程)。ステップS400における3次元AAM記述子の算出は図9のフローチャートに示す手順により算出することができる。類似度算出部130は、化合物につ

いての3次元AAM記述子と、ステップS400で算出したポケット構造体についての3次元AAM記述子との類似度を算出する（ステップS402：類似度算出工程）。類似度を算出したら、化合物抽出部140は類似度に基づいて標的化合物を抽出する（ステップS404：標的化合物抽出工程）。上述したリガンド入力の場合と同様に、類似度に基づく標的化合物の抽出（ステップS404）は、具体的には「類似度がしきい値以上の化合物を抽出する」、「類似度が高い順に化合物を抽出する」等により行うことができる。

[0093] 不変量化AAM記述子を用いる場合も、図18のフローチャートと同様の手順で標的化合物を抽出することができる。

[0094] 図19は標的タンパク質入力のスクリーニング結果の例を示す表である。図19の（a）部分は3次元AAM記述子を用い「類似度がしきい値以上の化合物を抽出する」とした場合の結果を示し、（b）部分は不変量化AAM記述子を用い「類似度が高い順に化合物を抽出する」とした場合の結果を示す。類似度のしきい値及び抽出個数は、スクリーニングの精度等の条件に応じて設定することができる。設定は、操作部400を介したユーザの入力に応じて行うことができる。また、図19と逆に3次元AAM記述子を用いた場合に「類似度が高い順に化合物を抽出する」とし、不変量化AAM記述子を用いた場合に「類似度がしきい値以上の化合物を抽出する」としてもよい。

[0095] 標的タンパク質入力のスクリーニングの場合も、リガンド入力のスクリーニングの場合（図17及びその説明を参照）と同様にアミノ酸の種類を変えてもよいし、異なるアミノ酸についての複数の記述子（3次元AAM記述子、不変量化AAM記述子）を用いてもよい。化合物の抽出に用いる記述子は1種類でよいが、複数種類の記述子を用いることにより、類似度に基づく化合物の抽出を正確に行うことができる。なお、複数種類の記述子を用いる場合、それら記述子の間でのアミノ酸の組み合わせは特に限定されない。いずれのアミノ酸について記述子及び類似度を算出するかは、操作部400を介

したユーザの指示に応じて処理部１００（特徴量算出部１２０、類似度算出部１３０、化合物抽出部１４０）が決定してもよいし、ユーザの指示によらず処理部１００が決定してもよい。

[0096] 化合物抽出部１４０は、図１９に示すような抽出結果を、化合物抽出結果２６０として記憶部２００に記憶させる（図３参照）。

[0097] <スクリーニング装置の効果>

以上説明したように、第１の実施形態に係るスクリーニング装置１０では、本発明に係る特徴量算出方法及び特徴量算出プログラムにより算出した特徴量（３次元ＡＡＭ記述子、不変量化ＡＡＭ記述子）を用いて、本発明に係るスクリーニング方法及びスクリーニングプログラムにより医薬候補化合物のスクリーニングを効率よく行うことができる。

[0098] <第２の実施形態>

本発明の第２の実施形態に係る化合物創出装置について説明する。図２０は化合物創出装置２０（特徴量算出装置、化合物創出装置）の構成を示すブロック図である。なお、第１の実施形態と同様の要素には同一の参照符号を付し、詳細な説明を省略する。

[0099] 化合物創出装置２０は処理部１０１を含む。処理部１０１は図２１のように構成され、情報入力部１１０、特徴量算出部１２０（特徴量算出部）、生成器構築部１３２（生成器構築部）、化合物立体構造生成部１４２（化合物立体構造生成部）、表示制御部１５０を含む。情報入力部１１０、特徴量算出部１２０、表示制御部１５０の機能は上述したスクリーニング装置１０における情報入力部１１０、特徴量算出部１２０、表示制御部１５０とそれぞれ同様である。これら各部の機能は、スクリーニング装置１０について上述したのと同様に、各種のプロセッサ（processor）を用いて実現することができる。

[0100] 図２２は記憶部２０１に記憶される情報を示す図である。記憶部２０１には、スクリーニング装置１０における化合物抽出結果２６０に代えて立体構造生成結果２７０が記憶される。記憶部２０１に記憶される情報は、図４に

ついて上述したのと同様に、互いに関連付けて記憶される。

[0101] <標的化合物の立体構造生成>

上述した3次元AAM記述子、不変量化AAM記述子を用いた、標的化合物（医薬候補化合物）の立体構造生成について説明する。化合物創出装置200による標的化合物の立体構造生成では、検索を行わないので「スクリーニングによる検索の結果、解なし」となる場合でも化合物の立体構造を生成することができ、したがって医薬候補化合物の立体構造を効率よく創出することができる。立体構造の生成は、リガンドの記述子（3次元AAM記述子、不変量化AAM記述子）に基づいて行うモードと、標的タンパク質のポケット構造体の記述子（3次元AAM記述子、不変量化AAM記述子）に基づいて行うモードとがある。いずれのモードにより立体構造の生成を行うかは、操作部400を介したユーザの操作に応じて選択することができる。

[0102] <リガンドが入力の立体構造生成>

図23はリガンド入力の場合の立体構造生成手順を示すフローチャートである。処理が開始すると、特徴量算出部120はリガンドの記述子（3次元AAM記述子）を算出する（ステップS500：対象構造体指定工程、立体構造発生工程、特徴量算出工程）。ステップS500の処理は、第1の実施形態と同様に本発明に係る特徴量算出方法及び特徴量算出プログラムを用いて行うことができる（図5～8及びそれらの図についての説明を参照）。

[0103] ステップS502では、生成器構築部132は機械学習により生成器を構築する（生成器構築工程）。以下、図24を参照してステップS502の処理を説明する。（Step1）特徴量算出部120が複数の化合物について3次元AAM記述子を算出し、構造式（立体化された構造式）と3次元AAM記述子とのペア（3次元データどうし）を作る。（Step2）生成器構築部132が、化合物の立体構造を教師データとし3次元AAM記述子を説明変数とした機械学習（深層学習）により生成器を構築する。深層学習の手法は特定の手法に限定されず、例えば単純な全結合のニューラルネットでもよいし、畳み込みニューラルネット（CNN：Convolutional Neural Networ

k) でもよい。ただし、立体構造の生成精度は用いる学習手法に依存するので、立体構造の生成条件、要求精度等の条件に応じて学習手法を選択することが好ましい。生成器構築方法として、後述する付記 14, 15 に記載の方法を用いてもよい。

[0104] 上述した Step 1, 2 の処理が終わると図 23 のフローチャートに戻る。化合物立体構造生成部 142 は、構築した生成器を用いて、リガンドの 3 次元 AAM 記述子から標的化合物（ヒット）の立体構造（立体化された構造式）を生成する（ステップ S504：化合物立体構造生成工程）。これにより、リガンドと類似の薬効（標的タンパク質との結合）を有する化合物、すなわち医薬候補化合物の立体構造を得ることができる。なお、同一の 3 次元 AAM 記述子を与える立体構造は複数存在しうる。化合物立体構造生成部 142 は、生成した立体構造を、立体構造生成結果 270 として 3 次元 AAM 記述子（3 次元 AAM 記述子 230）と関連付けて記憶部 201 に記憶させる（図 22 参照）。操作部 400 を介したユーザの指示に応じて、表示制御部 150 が、生成した立体構造をモニタ 310 に表示してもよい。

[0105] なお、上述した手順において、機械学習のための 3 次元 AAM 記述子を算出するアミノ酸は 1 種類でもよいし複数種類でもよい。ただし、複数種類のアミノ酸について 3 次元 AAM 記述子を算出し学習に供することにより、生成される立体構造の精度を向上させることができる。なお、アミノ酸の種類が異なる複数の 3 次元 AAM 記述子を用いる場合、それら記述子の間でのアミノ酸の組み合わせは特に限定されない。いずれのアミノ酸について 3 次元 AAM 記述子を算出し学習に供するかは、操作部 400 を介したユーザの指示に応じて処理部 100（特徴量算出部 120、類似度算出部 130、化合物抽出部 140）が決定してもよいし、ユーザの指示によらず処理部 100 が決定してもよい。

[0106] <立体構造の生成例>

機械学習により構築した生成器を用いて生成した立体構造の例について説明する。この例では単純な全結合のニューラルネットを用いてライブラリ化

化合物のうち1,800個を上述の手法により学習し、残り200個の化合物の立体構造がどの程度再現できるかを調べた。その結果を図25に示す。ニューラルネットの中間層の数を増やすと、平均cos類似度は59%になった。図26はそのような類似度を示す例についての、3次元AAM記述子から生成した立体構造（構造式）と正解構造式（それぞれ図26の（a）部分、（b）部分）を示す図である。

[0107] <教師データの特徴と生成される立体構造との関係>

上述した手順により生成される立体構造は、教師データとして与える化合物の特徴に影響される。したがって、教師データとして与える化合物の特徴を選択することで、特徴が異なる立体構造を有する化合物を生成できる。例えば、合成が容易な立体構造を有する化合物の3次元AAM記述子を教師データとして与えることで、リガンドと類似の薬効を有し、かつ合成が容易な立体構造を有する化合物を生成することができる。どのような化合物についての3次元AAM記述子を教師データとして与えるかは、生成したい化合物の特徴に合わせて選択することができる。

[0108] <不変量化AAM記述子を用いた立体構造の生成>

図23～26では3次元AAM記述子を用いた立体構造の生成について説明した。これに対し不変量化AAM記述子（不変量化特徴量）を用いた場合も、3次元AAM記述子を用いる場合と同様に、不変量化AAM記述子を教師データとし立体構造（立体化した構造式）を説明変数とした機械学習（深層学習）により標的化合物の立体構造を生成することができる。

[0109] <標的タンパク質が入力の立体構造生成>

化合物創出装置20では、上述したリガンド入力による立体構造生成に加えて、標的タンパク質を入力として標的化合物の立体構造を生成することができる。この場合も、リガンド入力の場合と同様に、3次元AAM記述子を用いた立体構造生成と不変量化AAM記述子を用いた立体構造生成とを行うことができる。

[0110] 図27は、標的タンパク質が入力の場合（3次元AAM記述子を用いるも

のとしている)の立体構造生成手順を示すフローチャートである。処理が開始すると、特徴量算出部120は標的タンパク質のポケット構造体の3次元AAM記述子を算出する(ステップS600:対象構造体指定工程、立体構造発生工程、特徴量算出工程)。ステップS600の処理は、第1の実施形態と同様に本発明に係る特徴量算出方法を用いて行うことができる(図9, 10及びそれらの図についての説明を参照)。

[0111] ステップS602では、生成器構築部132は、リガンド入力の場合と同様に機械学習(深層学習)により生成器を構築する(生成器構築工程)。生成器の構築は上述したStep1, 2と同様にして行うことができる。化合物立体構造生成部142は、構築した生成器を用いて、ポケット構造体の3次元AAM記述子から標的化合物(ヒット)の立体構造(立体化された構造式)を生成する(ステップS604:化合物立体構造生成工程)。これにより、ポケット構造体と類似の薬効(標的タンパク質との結合)を有する化合物、すなわち医薬候補化合物の立体構造を得ることができる。なお、同一の3次元AAM記述子を与える立体構造は複数存在しうる。化合物立体構造生成部142は、生成した立体構造を、立体構造生成結果270として3次元AAM記述子(3次元AAM記述子230)と関連付けて記憶部201に記憶させる(図22参照)。操作部400を介したユーザの指示に応じて、表示制御部150が、生成した立体構造をモニタ310に表示してもよい。

[0112] <化合物創出装置の効果>

以上説明したように、第2の実施形態に係る化合物創出装置20では、本発明に係る特徴量算出方法及び特徴量算出プログラムにより算出した特徴量(3次元AAM記述子、不変量化AAM記述子)を用いて、本発明に係る化合物創出方法及び化合物創出プログラムにより医薬候補化合物の立体構造を効率よく創出することができる。

[0113] <第3の実施形態>

上述した第1の実施形態は特徴量の算出及びこれに基づくスクリーニングを行う態様であり、第2の実施形態は特徴量の算出及びこれに基づく標的化

化合物の立体構造創出を行う態様であるが、特徴量の算出に加えてスクリーニングと標的化合物の立体構造創出の両方を行ってもよい。そのため第3の実施形態に係る医薬候補化合物探索装置30（特徴量算出装置、スクリーニング装置、化合物創出装置；図28参照）では、図1に示すスクリーニング装置10の処理部100、あるいは図20に示す化合物創出装置20の処理部101に代えて図28に示す処理部102を有する。図29に示すように、処理部102は特徴量算出部120（特徴量算出部、不変量化部）、類似度算出部130（類似度算出部）、生成器構築部132（生成器構築部）、化合物抽出部140（化合物抽出部）、化合物立体構造生成部142（化合物立体構造生成部）を有し、特徴量の算出、スクリーニング、及び化合物の立体構造創出を行うことができる。また、医薬候補化合物探索装置30はこれに合わせた情報を記憶部202に記憶する。具体的には、図30に示すように、記憶部200及び記憶部201に記憶される情報（図3，22参照）が合わせて記憶部202に記憶される。

[0114] その他の要素は図1に示すスクリーニング装置10、図20に示す化合物創出装置20と同様であるので同一の参照符号を付し詳細な説明を省略する。

[0115] 上述した構成により、第3の実施形態に係る医薬候補化合物探索装置30においても、スクリーニング装置10、化合物創出装置20と同様に、対象構造体の化学的性質を的確に示す特徴量を算出し、医薬候補化合物のスクリーニングを効率よく行い、医薬候補化合物の立体構造を効率よく創出することができる。

[0116] 以上で本発明の実施形態に関して説明してきたが、本発明は上述した態様に限定されず、以下に例示するように本発明の精神を逸脱しない範囲で種々の変形が可能である。

[0117] <扱える医薬のターゲット>

本発明では、医薬のターゲットとしてタンパク質以外に、DNA（Deoxyribonucleic Acid）、RNA（Ribonucleic Acid）、細胞膜、多糖を扱うこと

ができる。ただし、第1～第3の実施形態において、アミノ酸を別のものに変更する必要がある。具体的には、DNAの場合はアミノ酸を核酸塩基へ変更し、RNAの場合はアミノ酸を核酸塩基へ変更し、細胞膜の場合はアミノ酸を脂質分子へ変更し、多糖の場合はアミノ酸を単糖分子へ変更する。以下では、この変更で、本発明でDNA、RNA、細胞膜、多糖も扱える理由を説明する。タンパク質、DNA、RNA、細胞膜、多糖は纏めて生体高分子と呼ばれ、固有のビルディングブロックから成り立っている。具体的には、タンパク質のビルディングブロックはアミノ酸、DNAのビルディングブロックは核酸塩基、RNAのビルディングブロックは同様に核酸塩基、細胞膜のビルディングブロックは脂質分子、多糖のビルディングブロックは単糖分子である。タンパク質以外の生体高分子であるDNA、RNA、細胞膜、多糖にも、タンパク質と同様に、活性部位であるポケットがあるため、本発明は、医薬のターゲット（標的生体高分子）がDNA、RNA、細胞膜、多糖の場合にも、タンパク質の場合に示した第1～第3の実施形態において、アミノ酸をターゲットのビルディングブロックへ変更することで、対応できる。なお、化合物あるいはポケット構造体の周辺におけるアミノ酸、核酸塩基、脂質分子、単糖分子の集積度合いの定量化時に、水を考慮することもできる。

[0118] <扱える活性>

本発明では、「化合物による標的生体分子単独での活性」という通常の活性以外にも、「化合物による、標的生体分子に加えてその他の生体分子からなる複合体である細胞の活性」についても扱うことができる。

[0119] <（変形例1）アミノ酸以外の生体高分子についての特徴量及びその利用>

<ターゲット及びプローブ>

医薬のターゲット（標的生体高分子）としてタンパク質以外の生体高分子（化合物）であるDNA、RNA、細胞膜、多糖を扱う場合、特徴量の算出においてプローブはアミノ酸ではなく別の物質（各ターゲットのビルディン

グブロック)にする。具体的には、ターゲットが「DNA、RNA、細胞膜、多糖」の場合、プローブはそれぞれ「1種類以上の核酸塩基、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子」とする。また、これらをプローブとして集積度合いを定量化する際に、水、1種類以上のイオンを考慮してもよい。また、ターゲットが「DNA、RNA、細胞膜、多糖」のうち複数種類の生体高分子により構成される場合、プローブもターゲットの構成に応じて「1種類以上の核酸塩基、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子、水、1種類以上のイオン」のうち1つ以上（ターゲットの構成に応じた種類、数、及びそれらの組合せでよい）とすることができる。プローブを構成するイオンは、単原子イオンでもよいし多原子イオンでもよい。なお、いずれものプローブも、ファンデルワールス力を発生させることを前提とする。

[0120] <特徴量の算出及びスクリーニング>

特徴量（変形例1に係る特徴量）の算出及びスクリーニングを行う装置（特徴量算出装置、スクリーニング装置）の構成は、第1の実施形態と同様である（図1～3を参照）。ただし、図3における3次元AAM記述子230に代えて変形例1に係る記述子を算出及び記憶し、不変量化AAM記述子240に代えて変形例1に係る不変量化特徴量を算出及び記憶する。特徴量の算出及びスクリーニング手順は第1の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、スクリーニング方法、及びスクリーニングプログラムを用いることができる。具体的には、第1の実施形態に係る特徴量の算出（図5参照）におけるプローブとしての「アミノ酸」を「1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子、水、1種類以上のイオンのうち1つ以上（任意の種類、数、及び組合せで良い）」として分布関数を算出して（式（1）参照）この分布関数から変形例1に係る特徴量を算出し、変形例1に係る特徴量をフーリエ変換して変形例1に係る不変量化特徴量を算出する（式（2）参照）。また、2種類の異なるプローブ（1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分

子、水、及び1種類以上のイオンのうち1つ以上で構成される第1のプロープと、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子、水、及び1種類以上のイオンのうち1つ以上で構成される第2のプロープであって第1のプロープとは異なる第2のプロープ) についての変形例1に係る特徴量を用いて、相関関数の角度積分により変形例1に係る不変量特徴量を算出してもよい(式(3)、(4)参照)。

[0121] また、第1の実施形態における3次元AAM記述子に代えて変形例1に係る特徴量を用いて、複数の化合物についての変形例1に係る特徴量と、結合化合物についての変形例1に係る特徴量の類似度に基づいて標的化合物を抽出することができる。類似度がしきい値以上の化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0122] <特徴量の算出及び化合物の創出>

特徴量(変形例1に係る特徴量)の算出及び化合物の創出を行う装置(特徴量算出装置、化合物創出装置)の構成は、第2の実施形態と同様である(図20~22を参照)。ただし、図22における3次元AAM記述子230に代えて変形例1に係る記述子(変形例1に係る特徴量)を算出及び記憶し、不変量AAM記述子240に代えて変形例1に係る不変量特徴量を算出及び記憶する。特徴量の算出及び化合物の創出手順は第2の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、化合物創出方法、及び化合物創出プログラムを用いることができる。変形例1においては、化合物の立体構造を教師データとし、変形例1に係る特徴量を説明変数とした機械学習(深層学習)により生成器を構築し、構築した生成器を用いて、標的生体高分子との結合が確認されている化合物である結合化合物についての変形例1に係る特徴量から標的生体高分子の立体構造を生成することができる。なお第1から第3の実施形態と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

[0123] 第1から第3の実施形態について上述したのと同様に、化合物の薬効(D

N A等のターゲットへの結合力)は局所的には化合物と核酸塩基等(プローブ)との相互作用の結果として表されるので、化合物間で核酸塩基等の集積度合いが類似であれば、それら化合物はターゲットとの結合力が類似している。すなわち変形例1に係る特徴量が類似な化合物は類似の薬効を示す。したがって変形例1では、変形例1に係る特徴量により化合物の化学的性質を的確に示すことができる。また、標的生体高分子に結合する結合化合物と対象化合物とで変形例1に係る特徴量が類似ならば、両者の薬効が類似している。したがって変形例1によれば、変形例1に係る特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。また変形例1によれば、上述した実施形態と同様に、検索をせずに結合化合物の特徴量(変形例1に係る特徴量)に特徴量が類似した(したがって薬効が類似した)化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。

[0124] なお、特徴量算出、スクリーニング、及び化合物創出を行う場合(第3の実施形態に対応)は、図28~30と同様の構成により行うことができる。

[0125] <(変形例2) アミノ酸及びアミノ酸以外についての特徴量及びその利用>

<ターゲット及びプローブ>

変形例2では、「タンパク質とタンパク質以外の生体高分子(DNA、RNA、細胞膜、多糖)との複合体」をターゲットとする。また、「1種類以上のアミノ酸」(第1のプローブ)と「1種類以上の核酸塩基、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子、水、1種類以上のイオンのうち1つ以上」(第2のプローブ;任意の種類、数、組合せでよい)とをプローブとする。第1、第2のプローブの構成(種類、数、及びそれらの組合せ)はターゲットの構成に応じて設定することができる。プローブを構成するイオンは、単原子イオンでもよいし多原子イオンでもよい。なお、いずれものプローブもファンデルワールス力を発生させることを前提とする。

[0126] <特徴量の算出及びスクリーニング>

特徴量（変形例 2 に係る特徴量）の算出及びスクリーニングを行う装置（特徴量算出装置、スクリーニング装置）の構成は、第 1 の実施形態と同様である（図 1 ～ 3 を参照）。ただし、図 3 における 3 次元 AAM 記述子 230 に代えて変形例 2 に係る記述子（変形例 2 に係る特徴量）を算出及び記憶し、不変量化 AAM 記述子 240 に代えて変形例 2 に係る不変量化特徴量を算出及び記憶する。特徴量の算出及びスクリーニング手順は第 1 の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、スクリーニング方法、及びスクリーニングプログラムを用いることができる。具体的には、変形例 2 に係る特徴量の算出（図 5 参照）におけるプローブとしての「アミノ酸」を「1 種類以上のアミノ酸」（第 1 のプローブ）と「1 種類以上の核酸塩基、1 種類以上の脂質分子、水、1 種類以上の単糖分子、1 種類以上のイオンのうち 1 つ以上」（第 2 のプローブ；任意の種類、数、組合せでよい）として分布関数を算出して（式（1）参照）この分布関数から変形例 2 に係る特徴量を算出し、変形例 2 に係る特徴量をフーリエ変換して変形例 2 に係る不変量化特徴量を算出する（式（2）参照）。また、「1 種類以上のアミノ酸」（第 1 のプローブ）と「1 種類以上の核酸塩基、1 種類以上の脂質分子、1 種類以上の単糖分子、水、1 種類以上のイオンのうち 1 つ以上」（第 2 のプローブ；任意の種類、数、組合せでよい）とのうち少なくとも一方が異なる 2 種類のプローブについての変形例 2 に係る特徴量を用いて、相関関数の角度積分により変形例 2 に係る不変量化特徴量を算出してもよい（式（3）、（4）参照）。

[0127] また、第 1 の実施形態における 3 次元 AAM 記述子に代えて変形例 2 に係る特徴量を用い、複数の化合物についての変形例 2 に係る特徴量と、結合化合物についての変形例 2 に係る特徴量の類似度に基づいて標的化合物を抽出することができる。類似度がしきい値以上の化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0128] <特徴量の算出及び化合物の創出>

特徴量（変形例 2 に係る特徴量）の算出及び化合物の創出を行う装置（特徴量算出装置、化合物創出装置）の構成は、第 2 の実施形態と同様である（図 20～22 を参照）。ただし、図 22 における 3 次元 AAM 記述子 230 に代えて変形例 2 に係る記述子（変形例 2 に係る特徴量）を算出及び記憶し、不変量化 AAM 記述子 240 に代えて変形例 2 に係る不変量化特徴量を算出及び記憶する。特徴量の算出及び化合物の創出手順は第 2 の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、化合物創出方法、及び化合物創出プログラムを用いることができる。変形例 2 においては、化合物の立体構造を教師データとし、変形例 2 に係る特徴量を説明変数とした機械学習（深層学習）により生成器を構築し、構築した生成器を用いて、標的生体高分子との結合が確認されている化合物である結合化合物についての変形例 2 に係るから標的生体高分子の立体構造を生成することができる。なお第 1 から第 3 の実施形態及び変形例 1 と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

[0129] 第 1 から第 3 の実施形態及び変形例 1 について上述したのと同様に、化合物の薬効（ターゲットへの結合力）は局所的には化合物とプローブとの相互作用の結果として表されるので、化合物間でプローブの集積度合いが類似であれば、それら化合物はターゲットとの結合力が類似している。すなわち変形例 2 に係る特徴量が類似な化合物は類似の薬効を示す。したがって変形例 2 では、変形例 2 に係る特徴量により化合物の化学的性質を的確に示すことができる。また、標的生体高分子に結合する結合化合物と対象化合物とで変形例 2 に係る特徴量が類似ならば、両者の薬効が類似している。したがって変形例 2 によれば、変形例 2 に係る特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。また変形例 2 によれば、上述した実施形態及び変形例 1 と同様に、検索をせずに、結合化合物の特徴量（変形例 2 に係る特徴量）に特徴量が類似した（したがって薬効が類似した）化合物の構造式を生成するの

で、医薬候補化合物の立体構造を効率よく創出することができる。

[0130] なお、特徴量算出、スクリーニング、及び化合物創出を行う場合（第3の実施形態に対応）は、図28～30と同様の構成により行うことができる。

[0131] <（変形例3）仮想的な点電荷等についての特徴量及びその利用>

<ターゲット及びプローブ>

変形例3では生体高分子（化合物）をターゲットとし、「電荷が+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0.1である第3の点電荷、電荷が-0.1である第4の点電荷、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及びそれらの組合せでよい）」（実数電荷を持ちファンデルワールス力を発生させる仮想的な点電荷）をプローブとする。これらの点電荷は「単一の点」の一例である。

[0132] <特徴量の算出及びスクリーニング>

特徴量（変形例3に係る特徴量；第2の特徴量、第2の不変量化特徴量）の算出及びスクリーニングを行う装置（特徴量算出装置、スクリーニング装置）の構成は、第1の実施形態と同様である（図1～3を参照）。ただし、図3における3次元AAM記述子230に代えて第2の記述子（第2の特徴量）を算出及び記憶し、不変量化AAM記述子240に代えて第2の不変量化特徴量を算出及び記憶する。特徴量の算出及びスクリーニング手順は第1の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、スクリーニング方法、及びスクリーニングプログラムを用いることができる。具体的には、第1の実施形態に係る特徴量の算出（図5参照）におけるプローブとしての「アミノ酸」を「電荷が+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0.1である第3の点電荷、電荷が-0.1である第4の点電荷、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及びそれらの組合せでよい）」として分布関数を算出して（式（1）参照）この分布関数から第2の特徴量を算出し、第2の特徴量をフーリエ変換して第2の不変量化特徴量を算出する（式（2）参照）。また、2種類の異なるプローブ（第1の点電荷、第2の点電荷、第3の点電荷、

第4の点電荷、第5の点電荷のうち1つ以上で構成される第1のプローブと、第1の点電荷、第2の点電荷、第3の点電荷、第4の点電荷、第5の点電荷のうち1つ以上で構成される第2のプローブであって第1のプローブとは異なる第2のプローブ)についての第2の特徴量を用いて、相関関数の角度積分により第2の不変量特徴量を算出してもよい(式(3)、(4)参照)。

[0133] また、第1の実施形態における3次元AAM記述子に代えて第2の特徴量を用い、複数の化合物についての第2の特徴量と、結合化合物についての第2の特徴量の類似度に基づいて標的化合物を抽出することができる。類似度がしきい値以上の化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0134] <特徴量の算出及び化合物の創出>

特徴量(第2の特徴量)の算出及び化合物の創出を行う装置(特徴量算出装置、化合物創出装置)の構成は、第2の実施形態と同様である(図20～22を参照)。ただし、図22における3次元AAM記述子230に代えて第2の記述子(第2の特徴量)を算出及び記憶し、不変量AAM記述子240に代えて第2の不変量特徴量を算出及び記憶する。特徴量の算出及び化合物の創出手順は第2の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、化合物創出方法、及び化合物創出プログラムを用いることができる。変形例3においては、化合物の立体構造を教師データとし、第2の特徴量を説明変数とした機械学習(深層学習)により生成器を構築し、構築した生成器を用いて、標的生体高分子との結合が確認されている化合物である結合化合物についての第2の特徴量から標的生体高分子の立体構造を生成することができる。なお第1から第3の実施形態及び変形例1、2と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

[0135] 第1から第3の実施形態及び変形例1、2について上述したのと同様に、化合物の薬効(ターゲットへの結合力)は局所的には化合物とプローブとの

相互作用の結果として表されるので、化合物間でプローブの集積度合いが類似であれば、それら化合物はターゲットとの結合力が類似している。すなわち第2の特徴量が類似な化合物は類似の薬効を示す。したがって変形例3では、第2の特徴量により化合物の化学的性質を的確に示すことができる。また、標的生体高分子に結合する結合化合物と対象化合物とで第2の特徴量が類似ならば、両者の薬効が類似している。したがって変形例3によれば、第2の特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。また変形例3によれば、第1から第3の実施形態及び変形例1、2と同様に、検索をせずに、結合化合物の特徴量（第2の特徴量）に特徴量が類似した（したがって薬効が類似した）化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。

[0136] なお、特徴量算出、スクリーニング、及び化合物創出を行う場合（第3の実施形態に対応）は、図28～30と同様の構成により行うことができる。

[0137] <（変形例4）アミノ酸及び仮想的な点電荷等についての特徴量及びその利用>

<ターゲット及びプローブ>

変形例4では生体高分子（化合物）をターゲットとし、「1種類以上のアミノ酸である第1のプローブと、電荷が+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0.1である第3の点電荷、電荷が-0.1である第4の点電荷、第1の点電荷と第2の点電荷とが離間して配置されたダイポール、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及び組合せで良い）である第2のプローブ」とをプローブとする。第2のプローブはダイポールを除く点電荷（第1～第5の点電荷のうち1つ以上）でもよい。第1から第5の点電荷は、「単一の点」の一例である。

[0138] <特徴量の算出及びスクリーニング>

特徴量（変形例4に係る特徴量）の算出及びスクリーニングを行う装置（特徴量算出装置、スクリーニング装置）の構成は、第1の実施形態と同様で

ある（図1～3を参照）。ただし、図3における3次元AAM記述子230に代えて変形例4に係る記述子（変形例4に係る特徴量）を算出及び記憶し、不変量化AAM記述子240に代えて変形例4に係る不変量化特徴量を算出及び記憶する。特徴量の算出及びスクリーニング手順は第1の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、スクリーニング方法、及びスクリーニングプログラムを用いることができる。具体的には、第1の実施形態に係る特徴量の算出（図5参照）におけるプローブとしての「アミノ酸」を「1種類以上のアミノ酸である第1のプローブと、電荷が+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0.1である第3の点電荷、電荷が-0.1である第4の点電荷、第1の点電荷と第2の点電荷とが離間して配置されたダイポール、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及び組合せで良い）である第2のプローブ」として分布関数を算出して（式（1）参照）この分布関数から変形例4に係る特徴量を算出し、変形例4に係る特徴量をフーリエ変換して変形例4に係る不変量化特徴量を算出する（式（2）参照）。第2のプローブはダイポールを除く点電荷（第1～第5の点電荷のうち1つ以上）でもよい。また、第1のプローブと第2のプローブとのうち少なくとも一方が異なる2種類のプローブについての変形例4に係る特徴量を用いて、相関関数の角度積分により変形例4に係る不変量化特徴量を算出してもよい（式（3）、（4）参照）。

[0139] また、第1の実施形態における3次元AAM記述子に代えて変形例4に係る特徴量を用いて、複数の化合物についての変形例4に係る特徴量と、結合化合物についての変形例4に係る特徴量の類似度に基づいて標的化合物を抽出することができる。類似度がしきい値以上の化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0140] <特徴量の算出及び化合物の創出>

特徴量（変形例4に係る特徴量）の算出及び化合物の創出を行う装置（特徴量算出装置、化合物創出装置）の構成は、第2の実施形態と同様である（

図20～22を参照)。ただし、図22における3次元AAM記述子230に代えて変形例4に係る記述子(変形例4に係る特徴量)を算出及び記憶し、不変量化AAM記述子240に代えて変形例4に係る不変量化特徴量を算出及び記憶する。特徴量の算出及び化合物の創出手順は第2の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、化合物創出方法、及び化合物創出プログラムを用いることができる。変形例4においては、化合物の立体構造を教師データとし、変形例4に係る特徴量を説明変数とした機械学習(深層学習)により生成器を構築し、構築した生成器を用いて、標的生体高分子との結合が確認されている化合物である結合化合物についての変形例4に係る特徴量から標的生体高分子の立体構造を生成することができる。なお第1から第3の実施形態と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

[0141] 第1から第3の実施形態について上述したのと同様に、化合物の薬効(ターゲットへの結合力)は局所的には化合物とプローブとの相互作用の結果として表されるので、化合物間でプローブの集積度合いが類似であれば、それら化合物はターゲットとの結合力が類似している。すなわち変形例4に係る特徴量が類似な化合物は類似の薬効を示す。したがって変形例4では、変形例4に係る特徴量により化合物の化学的性質を的確に示すことができる。また、標的生体高分子に結合する結合化合物と対象化合物とで変形例4に係る特徴量が類似ならば、両者の薬効が類似している。したがって変形例4によれば、変形例4に係る特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。また変形例4によれば、上述した実施形態と同様に、検索をせずに結合化合物の特徴量(変形例4に係る特徴量)に特徴量が類似した(したがって薬効が類似した)化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。

[0142] なお、特徴量算出、スクリーニング、及び化合物創出を行う場合(第3の

実施形態に対応)は、図28～30と同様の構成により行うことができる。

[0143] <(変形例5) 核酸塩基等及び仮想的な点電荷等についての特徴量及びその利用>

<ターゲット及びプローブ>

変形例5では生体高分子(化合物)をターゲットとし、「1種類以上の核酸塩基、1種類以上の脂質分子、水、1種類以上の単糖分子、1種類以上のイオンのうち1つ以上(任意の種類、数、及び組合せでよい)」である第1のプローブと、「電荷が+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0.1である第3の点電荷、電荷が-0.1である第4の点電荷、第1の点電荷と第2の点電荷とが離間して配置されたダイポール、電荷がゼロである第5の点電荷のうち1つ以上(任意の種類、数、及び組合せでよい)」である第2のプローブとをプローブとする。第1のプローブを1種類以上の単原子イオンとし、第2のプローブを点電荷(第1～第5の点電荷のうち1つ以上)としてもよい。「単原子イオン」及び「点電荷」は「単一の点」の一例である。

[0144] <特徴量の算出及びスクリーニング>

特徴量(変形例5に係る特徴量;第3の特徴量、第3の不変量化特徴量)の算出及びスクリーニングを行う装置(特徴量算出装置、スクリーニング装置)の構成は、第1の実施形態と同様である(図1～3を参照)。ただし、図3における3次元AAM記述子230に代えて第3の記述子(第3の特徴量)を算出及び記憶し、不変量化AAM記述子240に代えて第3の不変量化特徴量を算出及び記憶する。特徴量の算出及びスクリーニング手順は第1の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、スクリーニング方法、及びスクリーニングプログラムを用いることができる。具体的には、第1の実施形態に係る特徴量の算出(図5参照)におけるプローブとしての「アミノ酸」を「1種類以上の核酸塩基、1種類以上の脂質分子、水、1種類以上の単糖分子、1種類以上のイオンのうち1つ以上(任意の種類、数、及び組合せでよい)」である第1のプローブと、「電荷が

+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0.1である第3の点電荷、電荷が-0.1である第4の点電荷、第1の点電荷と第2の点電荷とが離間して配置されたダイポール、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及び組合せでよい）」である第2のプローブとして分布関数を算出して（式（1）参照）この分布関数から第3の特徴量を算出し、第3の特徴量をフーリエ変換して第3の不変量化特徴量を算出する（式（2）参照）。また、第1のプローブと第2のプローブとのうち少なくとも一方が異なる2種類のプローブについての第3の特徴量を用いて、相関関数の角度積分により第3の不変量化特徴量を算出してもよい（式（3）、（4）参照）。上述のように、第1のプローブを1種類以上の単原子イオンとし、第2のプローブを点電荷（第1～第5の点電荷のうち1つ以上）としてもよい。「単原子イオン」及び「点電荷」は「単一の点」の一例である。

[0145] また、第1の実施形態における3次元AAM記述子に代えて第3の特徴量を用いて、複数の化合物についての第3の特徴量と、結合化合物についての第3の特徴量の類似度に基づいて標的化合物を抽出することができる。類似度がしきい値以上の化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0146] <特徴量の算出及び化合物の創出>

特徴量（第3の特徴量）の算出及び化合物の創出を行う装置（特徴量算出装置、化合物創出装置）の構成は、第2の実施形態と同様である（図20～22を参照）。ただし、図22における3次元AAM記述子230に代えて第3の記述子（第3の特徴量）を算出及び記憶し、不変量化AAM記述子240に代えて第3の不変量化特徴量を算出及び記憶する。特徴量の算出及び化合物の創出手順は第2の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、化合物創出方法、及び化合物創出プログラムを用いることができる。変形例5においては、化合物の立体構造を教師データとし、第3の特徴量を説明変数とした機械学習（深層学習）により生成器を構

築し、構築した生成器を用いて、標的生体高分子との結合が確認されている化合物である結合化合物についての第3の特徴量から標的生体高分子の立体構造を生成することができる。なお第1から第3の実施形態と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

[0147] 第1から第3の実施形態について上述したのと同様に、化合物の薬効（ターゲットへの結合力）は局所的には化合物とプローブとの相互作用の結果として表されるので、化合物間でプローブの集積度合いが類似であれば、それら化合物はターゲットとの結合力が類似している。すなわち第3の特徴量が類似な化合物は類似の薬効を示す。したがって変形例5では、第3の特徴量により化合物の化学的性質を的確に示すことができる。また、標的生体高分子に結合する結合化合物と対象化合物とで第3の特徴量が類似ならば、両者の薬効が類似している。したがって変形例5によれば、第3の特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。また変形例5によれば、上述した実施形態と同様に、検索をせずに結合化合物の特徴量（第3の特徴量）に特徴量が類似した（したがって薬効が類似した）化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。

[0148] なお、特徴量算出、スクリーニング、及び化合物創出を行う場合（第3の実施形態に対応）は、図28～30と同様の構成により行うことができる。

[0149] <（変形例6）アミノ酸、核酸塩基等、及び仮想的な点電荷等についての特徴量及びその利用>

<ターゲット及びプローブ>

変形例6では生体高分子（化合物）をターゲットとし、「1種類以上のアミノ酸」である第1のプローブと、「1種類以上の核酸塩基、1種類以上の脂質分子、水、1種類以上の単糖分子、1種類以上のイオンのうち1つ以上（任意の種類、数、及び組合せで良い）」である第2のプローブと、「電荷が+1である第1の点電荷、電荷が-1である第2の点電荷、電荷が+0、

1である第3の点電荷、電荷が -0.1 である第4の点電荷、第1の点電荷と第2の点電荷とが離間して配置されたダイポール、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及び組合せで良い）」である第3のプローブと、をプローブとする。第2のプローブを1種類以上の単原子イオンとし、第3のプローブを第1～第5の点電荷のうち1つ以上としてもよい。「単原子イオン」及び「点電荷」は「単一の点」の一例である。

[0150] <特徴量の算出及びスクリーニング>

特徴量（変形例6に係る特徴量）の算出及びスクリーニングを行う装置（特徴量算出装置、スクリーニング装置）の構成は、第1の実施形態と同様である（図1～3を参照）。ただし、図3における3次元AAM記述子230に代えて変形例6に係る記述子（変形例6に係る特徴量）を算出及び記憶し、不変量化AAM記述子240に代えて変形例6に係る不変量化特徴量を算出及び記憶する。特徴量の算出及びスクリーニング手順は第1の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、スクリーニング方法、及びスクリーニングプログラムを用いることができる。具体的には、第1の実施形態に係る特徴量の算出（図5参照）におけるプローブとしての「アミノ酸」を「1種類以上のアミノ酸」である第1のプローブと、「1種類以上の核酸塩基、1種類以上の脂質分子、水、1種類以上の単糖分子、1種類以上のイオンのうち1つ以上（任意の種類、数、及び組合せで良い）」である第2のプローブと、「電荷が $+1$ である第1の点電荷、電荷が -1 である第2の点電荷、電荷が $+0.1$ である第3の点電荷、電荷が -0.1 である第4の点電荷、第1の点電荷と第2の点電荷とが離間して配置されたダイポール、電荷がゼロである第5の点電荷のうち1つ以上（任意の種類、数、及び組合せで良い）」である第3のプローブと、として分布関数を算出して（式（1）参照）この分布関数から変形例6に係る特徴量を算出し、変形例6に係る特徴量をフーリエ変換して変形例6に係る不変量化特徴量を算出する（式（2）参照）。また、第1のプローブと、第2のプローブと、第3のプローブとのうち少なくとも1つが異なる2種類のプローブについて

の変形例 6 に係る特徴量を用いて、相関関数の角度積分により変形例 6 に係る不変量化特徴量を算出してもよい（式（3）、（4）参照）。上述のように、第 2 のプローブを 1 種類以上の単原子イオンとし、第 3 のプローブを第 1 ～第 5 の点電荷のうち 1 つ以上としてもよい。

[0151] また、第 1 の実施形態における 3 次元 AAM 記述子に代えて変形例 6 に係る特徴量を用いて、複数の化合物についての変形例 6 に係る特徴量と、結合化合物についての変形例 6 に係る特徴量の類似度に基づいて標的化合物を抽出することができる。類似度がしきい値以上の化合物を抽出してもよいし、類似度が高い順に化合物を抽出してもよい。

[0152] <特徴量の算出及び化合物の創出>

特徴量（変形例 6 に係る特徴量）の算出及び化合物の創出を行う装置（特徴量算出装置、化合物創出装置）の構成は、第 2 の実施形態と同様である（図 20 ～ 22 を参照）。ただし、図 22 における 3 次元 AAM 記述子 230 に代えて変形例 6 に係る記述子（変形例 6 に係る特徴量）を算出及び記憶し、不変量化 AAM 記述子 240 に代えて変形例 6 に係る不変量化特徴量を算出及び記憶する。特徴量の算出及び化合物の創出手順は第 2 の実施形態と同様であり、本発明の特徴量算出方法、特徴量算出プログラム、化合物創出方法、及び化合物創出プログラムを用いることができる。変形例 6 においては、化合物の立体構造を教師データとし、変形例 6 に係る特徴量を説明変数とした機械学習（深層学習）により生成器を構築し、構築した生成器を用いて、標的生体高分子との結合が確認されている化合物である結合化合物についての変形例 6 に係る特徴量から標的生体高分子の立体構造を生成することができる。なお第 1 から第 3 の実施形態と同様に、教師データとして与える化合物の特徴を選択することにより、特徴が異なる立体構造を有する化合物を生成することができる。

[0153] 第 1 から第 3 の実施形態について上述したのと同様に、化合物の薬効（ターゲットへの結合力）は局所的には化合物とプローブとの相互作用の結果として表されるので、化合物間でプローブの集積度合いが類似であれば、それ

ら化合物はターゲットとの結合力が類似している。すなわち変形例 6 に係る特徴量が類似な化合物は類似の薬効を示す。したがって変形例 6 では、変形例 6 に係る特徴量により化合物の化学的性質を的確に示すことができる。また、標的生体高分子に結合する結合化合物と対象化合物とで変形例 6 に係る特徴量が類似ならば、両者の薬効が類似している。したがって変形例 6 によれば、変形例 6 に係る特徴量に基づいて結合化合物と薬効が類似した標的化合物を抽出し、医薬候補化合物のスクリーニングを効率よく行うことができる。また変形例 6 によれば、上述した実施形態と同様に、検索をせずに結合化合物の特徴量（変形例 6 に係る特徴量）に特徴量が類似した（したがって薬効が類似した）化合物の構造式を生成するので、医薬候補化合物の立体構造を効率よく創出することができる。

[0154] なお、特徴量算出、スクリーニング、及び化合物創出を行う場合（第 3 の実施形態に対応）は、図 28～30 と同様の構成により行うことができる。

[0155] <各特徴量によるヒット数の比較>

図 31 は、上述した図 13 と同様の系（タンパク質 ABL1）について、化合物を対象構造体とする不変量特徴量（アミノ酸をプローブとする不変量特徴量、及びアミノ酸以外をプローブとする不変量特徴量）によるヒットのを見つけやすさを比較評価した結果の一例を示す図である。図 31 によれば、記述子（特徴量）の種類によって効果（ヒット数の期待値）に差があるものの、ランダムの場合（図 13 参照）と比較して期待値が向上することが分かる。なお、図 31 は（チーム数＝183）でのクラスタリングの結果なので、「AAM」についてのヒット数が図 13 の場合（チーム数＝221 でのクラスタリングの結果）と異なっている。このように、アミノ酸以外をプローブとする不変量特徴量を用いる場合においても、医薬候補化合物のスクリーニングを効率よく行うことができる。

[0156] <生成器を用いた特徴量の算出>

上述した第 1～第 3 の実施形態及び変形例 1～7 では「対象構造体について、（構造式等から）複数の単位構造体による立体構造を発生させる立体構

造発生工程」を有する特徴量算出方法により特徴量を算出しているが、本発明では「機械学習により構成された、立体構造から特徴量を算出する生成器」を用いて特徴量（特徴量、第1～第3の特徴量）を算出してもよい。生成器は、例えば「『複数の化合物についての、化合物の立体構造を示す情報（3次元画像等）を教師データとし、特徴量（例えば、第1～第4の態様に係る特徴量のいずれか1つ）を説明変数とした学習用データセット』を生成器に入力して機械学習により学習させる学習処理工程を有する生成器の構築方法」により構築することができる。

[0157] 上述した生成器（予測モデル、学習済みモデル）は化合物の立体構造を示す情報として3次元画像を受け付けてもよく、また階層型ネットワークを備えていてもよい。階層型ネットワークは、3次元フィルタを用いた畳み込み演算を行って特徴マップ（入力層から得られる情報）を縮小する畳み込み層と、3次元フィルタを用いた逆畳み込み演算を行って特徴マップを拡大する逆畳み込み層と、を含んでいてもよい。階層型ネットワークは、学習処理工程において、化合物を構成する原子の種類を色の相違（例えば、赤、緑、青を重み付けした組み合わせを用いる）で表現した3次元画像を受け付けてもよい。3次元画像をそれぞれ異なる色に対応した複数のチャンネル（例えば、赤、緑、青の3チャンネル）に分けて受け付けてもよい。また、階層型ネットワークは、学習処理工程において、3次元画像をそれぞれ異なる種類の原子に対応した複数のチャンネルに分けて受け付けてもよい。階層型ネットワークは畳み込みニューラルネットワークでもよい。

[0158] 図32は階層型ネットワークを用いて構成した「立体構造から特徴量を算出する生成器」の例を示す図である。図32の（a）部分は、階層型ネットワーク600は入力層610と、中間層620と、出力層630とを含む畳み込みニューラルネットワークを示す。中間層620は、第1の中間層622と、第2の中間層624とを備える。第1の中間層622は、畳み込み演算を行う畳み込み層と、プーリング処理を行うプーリング層とを含み（図32ではこれら2つの層をまとめて層625としている）、入力側に近い層で

は低次の特徴抽出が行われ、出力側に近づくにつれて高次の特徴抽出が行われる。学習時は、入力層 6 1 0 から出力層 6 3 0 に向けて情報（各層の重みパラメータの値等）が更新されてゆく。

[0159] 第 2 の中間層 6 2 4 は、3 次元フィルタを用いた逆畳み込み演算を行う逆畳み込み層と、逆プーリング処理を行う逆プーリング層とを含む（図 3 2 の（a）部分ではこれら 2 つの層をまとめて層 6 2 7 としている）。第 1 の中間層 6 2 2 では畳み込み演算及びプーリング処理により特徴マップがダウンスケーリングされ、第 2 の中間層 6 2 4 では逆畳み込み演算及び逆プーリング処理により特徴マップがアップスケーリングされる。3 次元フィルタのサイズを変更すること等によりダウンスケーリング及びアップスケーリングの度合いを調節することができ、これにより入力した 3 次元画像と同じサイズの 3 次元情報（対象構造体の周囲の 3 次元空間におけるプローブの集積度合いを定量化した特徴量）を得ることができる。なお、第 1 の中間層 6 2 2 及び第 2 の中間層 6 2 4 における層（畳み込み層、プーリング層、逆畳み込み層、逆プーリング層）の数及び組み合わせは特に限定されない（図 3 2 中に点線で記載した矢印は、層が省略されていることを示す）。

[0160] 上述した構成の階層型ネットワーク 6 0 0 を用いる場合、学習の過程において、出力層 6 3 0 が出力する結果と画像セットに対する認識（プローブの集積度合いを示す特徴量）の正解とを比較して損失（誤差）を計算し、損失が小さくなるように中間層 6 2 0 における重みパラメータを出力側の層から入力側の層に向けて更新していく処理（誤差逆伝播）を行うことが好ましい。なお、中間層 6 2 0 は畳み込み層及びプーリング層（、並びに逆畳み込み層及び逆プーリング層）の他にバッチノーマライゼーションを行う層を含んでもよい。バッチノーマライゼーション処理は学習を行う際のミニバッチを単位としてデータの分布を正規化する処理であり、学習を速く進行させる、初期値への依存性を下げる、過学習を抑制する等の役割を担う。

[0161] 上述の階層型ネットワーク 6 0 0 の学習が終了すると、化合物の 3 次元画像（立体構造を示す情報）を入力して本発明に係る特徴量（特徴量、第 1 ～

第3の特徴量)を算出することができる。この場合、学習データとしての立体構造情報に代えて実データとしての立体構造情報を入力層610に入力すると、入力層610から出力層630に向けて(図32の(a)部分における矢印の方向に)各層の重みパラメータ等の情報が更新されてゆき、出力層630から特徴量が出力される。

- [0162] なお、上述の階層型ネットワーク600において、入力と出力を逆にする(出力層630に特徴量を入力して入力層610から3次元画像を出力させる)ことで、特徴量に基づいて化合物の立体構造を創出することができる。この場合、特徴量を出力層630に入力すると、出力層630から入力層610に向けて(図32の(b)部分における矢印の方向に)情報が更新されてゆき、入力層610から化合物の立体構造を示す3次元画像が出力される。

- [0163] <他の特徴量との併用によるスクリーニング>

<物性値及び／または毒性値との併用>

図33は、本発明に係る特徴量と物性を示す物性値及び／または毒性を示す毒性値とに基づくスクリーニングの処理を示すフローチャートである。この処理は、第1の実施形態に係るスクリーニング装置10(特徴量算出装置、スクリーニング装置)、及び第3の実施形態に係る医薬候補化合物探索装置30(特徴量算出装置、スクリーニング装置、化合物創出装置)と同様の装置により実行することができる。これらの装置において、上述した生成器を用いて特徴量を算出してもよい。例えば、スクリーニング装置10において、処理部100に代えて図34に示す構成の処理部103を用いてスクリーニングを行うことができる。

- [0164] 情報入力部110は複数の化合物について立体構造を示す情報(例えば、構造式、3次元画像等)を入力し(ステップS310:入力工程)、特徴量算出部122(特徴量算出部、生成器)は複数の化合物のそれぞれについて、入力された情報に基づいて特徴量を算出する(ステップS312:特徴量算出工程)。図34に示す処理部103において、特徴量算出部122は上

述した階層型ネットワーク600等の生成器を備え、第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量（特徴量、第1～第3の特徴量）を算出することができる。また、特性値算出部134（特性値算出部）は、複数の化合物のそれぞれについて、入力された情報に基づいて、物性を示す物性値及び／または毒性を示す毒性値を算出する（ステップS314：特性値算出工程）。特性値算出部134は、物性値として例えば簡単な物性（分子量、log P、極性表面積、分極率）や、複雑な物性（溶解性、膜透過性）を算出することができる。これらの物性は公知の技術（一般に市販されているソフトウェア）により算出することができる。log P等の実験で得られる物性値について、Clog P等の予測値を用いてもよい。また、特性値算出部134は、物性値として「ローテータブルボンドの数（the number of rotatable bonds; RBN）」を算出することができる。「ローテータブルボンドの数」は、例えば“http://www.taletе.mi.it/help/dproperties_help/index.html?constitutional_descriptors.htm”に記載されているように、「一重結合であり、かつ、環には含まれず、終端の重い元素にも結合していないものの数」と定義することができる。

- [0165] また、特性値算出部134は、毒性として例えばlog BCF（化合物の魚類生物濃縮性）、Ames（変異原性）、あるいは皮膚感作性等を算出することができる。これらの毒性については、例えば「改正化審法における濃縮性予測手法活用，住友化学（株） 生物環境科学研究所，2006」（https://www.sumitomo-chem.co.jp/rd/report/files/docs/20060106_vpv.pdf）、「Mortelmans K & Zeiger E (2000): The Ames Salmonella/microsome mutagenicity assay. Mutat. Res. 455(1-2), 29-60.」あるいは<http://www.oyc-bio.jp/pages/reagent/ames/index>、<https://staff.aist.go.jp/yamane.t/RS/index.files/42RS-Kanazawa.pdf>に記載されている。また、これらの毒性は公知のソフトウェア“DEREK”（http://ls.ctc-g.co.jp/products/lhasa/derek_nexus.html）や“OECD QSAR ToolBox”（<https://www.nite.go.jp/data/000084383.pdf>）で計算することができる。

[0166] 複数の化合物全てについて特徴量、物性値及び／または毒性値を算出した
ら（ステップS 3 1 6でYES）、情報入力部1 1 0は特徴量、物性値及び
／または毒性値の目標値を入力する（ステップS 3 1 8：目標値入力工程）。
化合物抽出部1 4 0は特徴量の類似度（基準となる化合物の特徴量と上述
した複数の化合物のそれぞれについての特徴量との間の類似度）が目標値を
満たすかどうか、物性値及び／または毒性値が目標値を満たすか否かを判断
し（ステップS 3 2 0、S 3 2 2：抽出工程）、複数の化合物から、特徴量
の類似度が目標値を満たし、かつ物性値及び／または毒性値が目標値を満た
す化合物を抽出する（ステップS 3 2 4：抽出工程）。化合物抽出部1 4 0
は、複数の化合物全てについて判断及び抽出が終了したら、探索を終了する
（ステップS 3 2 6）。このような処理により、結合力（特徴量と対応）を
維持しつつ物性及び／または毒性が改善された化合物を抽出することができる。

[0167] <物性値との併用によるスクリーニングの実施例>

本発明に係る特徴量と物性を示す物性値とに基づくスクリーニングの実施
例について説明する。ここでは、元の化合物に対し、複数の化合物（ライブ
ラリ化合物）の中から「本発明に係る特徴量の類似度が基準値以下で、物性
値が目標値を満たす化合物」を抽出する場合について説明する。上述したよ
うに、図3 4に示す構成の処理部1 0 3を有するスクリーニング装置1 0が
、このようなスクリーニングの処理を実行することができる。スクリーニン
グ装置1 0は図3 3のフローチャートに示す手順で処理を行ってもよいし、
以下に説明するようにこの手順を本発明の第1 0の態様の範囲内で変更した
手順（例えば、処理の順番を入れ替えてもよいし、抽出工程を個々の化合物
について繰り返す行うのではなく複数の化合物について一括して行ってもよ
い）で処理を行ってもよい。

[0168] 情報入力部1 1 0は、図3 5の（a）部分に示す元の化合物C 0、及び同
図の（b）部分に示す複数の化合物C 1～C 3（ライブラリ化合物）につい
て立体構造を示す情報（構造式や3次元画像等）を入力し（入力工程）、特

微量算出部 1 2 2（特徴量算出部、生成器）は、これら化合物について、入力された情報に基づいて本発明に係る特徴量を算出する（特徴量算出工程）。特徴量算出部 1 2 2 は、本実施例ではプローブを電荷中性のアミノ酸とし、第 2 の態様に係る特徴量算出方法を用いて特徴量（第 2 の特徴量）を算出する。そして、化合物抽出部 1 4 0 は特徴量の類似度（元の化合物の特徴量とライブラリ化合物のそれぞれについての特徴量との間の類似度）を算出する（図 3 6 の（a）部分及び（b）部分；抽出工程）。本実施例のケースでは、化合物 C 0 と化合物 C 1 ～C 3 との間での特徴量の類似度は図 3 6 の（b）部分に示す通りである。なお、図 3 6 における「A A M 類似度」は、元の化合物 C 0 とライブラリの化合物 C 1 ～C 3 との間での、本発明に係る特徴量の類似度（類似性距離）を示す。

[0169] 化合物抽出部 1 4 0 は、ライブラリ化合物のそれぞれについて、特徴量の類似度が目標値を満たすかどうかを判断する（抽出工程）。ここで類似度（類似性距離）に対するしきい値を 1 0 0. 0 とした場合、化合物 C 1 ～C 3 のうち化合物 C 2, C 3 について判断が肯定される。また、特性値算出部 1 3 4（特性値算出部）は、化合物 C 2, C 3 について C l o g P（物性値の一例）を算出する（図 3 6 の（c）部分及び（d）部分；特性値算出工程）。算出結果は図 3 6 の（d）部分に示す通りである。化合物抽出部 1 4 0 は化合物 C 2, C 3 について C l o g P が目標値（ここでは、しきい値を 1. 0 以下とする）を満たすか否かを判断し（抽出工程）、目標値を満たす化合物 C 2 を抽出する（図 3 6 の（e）部分）。

[0170] このようにして、ライブラリにおける化合物 C 1 ～C 3 の探索で、元の化合物 C 0 と特徴量が類似しており（類似度が 1 0 0. 0 以下）、かつ C l o g P が低い（1. 0 以下）化合物 C 2 に骨格改変することができた（図 3 7 参照）。なお、C l o g P が低いことは、化合物の親水性が高く経口吸収性が優れていることを示す。

[0171] <構造特徴量との併用>

上述した物性値及び／または毒性値との併用の場合と同様に、図 3 8 のフ

ローチャートを参照して以下で説明するように、本発明に係る特徴量と化合物の構造を示す特徴量（構造特徴量）とを用いてスクリーニングを行うこともできる。この処理は、例えばスクリーニング装置10において処理部100に代えて図39に示す構成の処理部104を用いてスクリーニングを行うことができる。なお、図38のフローチャートに関し、図33と同様の処理には同一のステップ番号を付し、詳細な説明を省略する。

[0172] 入力された化合物について、特徴量算出部122は第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出し（ステップS312：第1の特徴量算出工程）、構造特徴量算出部136（構造特徴量算出部）は複数の化合物のそれぞれについて、入力された情報に基づいて、化合物の構造を示す構造特徴量を算出する（ステップS315：第2の特徴量算出工程）。構造特徴量算出部136は、構造特徴量として例えばFingerprint記述子（官能基や部分構造の有無を0と1で表現した記述子）を算出することができる。情報入力部110は特徴量の目標値、及び構造特徴量に対するしきい値を入力し（ステップS319：目標値入力工程）、化合物抽出部140は特徴量の類似度（基準となる化合物の特徴量と上述した複数の化合物のそれぞれについての特徴量との間の類似度）が目標値を満たし、かつ構造特徴量の類似度（基準となる化合物の構造特徴量と上述した複数の化合物のそれぞれについての構造特徴量との間の類似度）がしきい値以下である（ステップS321及びステップS323でYES）化合物を抽出する（ステップS324：抽出工程）。構造特徴量の類似度としては、Fingerprint記述子に対するTanimoto係数を用いることができる。Tanimoto係数は、比較したい2化合物のFingerprint記述子を与えると0から1までの実数として算出され、1に近くなるほど類似度が高くなるように定義されている。このような処理により、結合力（特徴量と対応）を維持しつつ骨格が異なる化合物を抽出することができる。

[0173] <構造特徴量との併用によるスクリーニングの実施例>

本発明に係る特徴量と構造特徴量とに基づくスクリーニングの実施例につ

いて説明する。ここでは、図40の(a)部分に示す元の化合物C0に対し、複数の化合物C4～C6(ライブラリ化合物)の中から「本発明に係る特徴量の類似度が基準値以下で、元の化合物との構造特徴量の類似度が目標値を満たす化合物」を抽出する場合について説明する。上述したように、図39に示す構成の処理部104を有するスクリーニング装置10がスクリーニングの処理を実行することができる。スクリーニング装置10は図38のフローチャートに示す手順で処理を行ってもよいし、以下で説明するようにこの手順を本発明の第11の態様の範囲内で変更した手順(例えば、処理の順番を入れ替えてもよいし、抽出工程を個々の化合物について繰り返し行うのではなく複数の化合物について一括して行ってもよい)で処理を行ってもよい。

[0174] 構造特徴量算出部136(構造特徴量算出部)は化合物C4～C6のそれぞれについて、入力された情報に基づいて、化合物の構造を示す構造特徴量を算出する(第2の特徴量算出工程)。本実施例では、構造特徴量算出部136は構造特徴量としてFingerprint記述子を算出するものとする。そして、化合物抽出部140は、元の化合物C0と化合物C4～C6(ライブラリ化合物)との間で、構造特徴量の類似度を示すTanimoto係数を算出し(図41の(a)部分及び(b)部分;抽出工程)、Tanimoto係数が0.2(しきい値)以下となる化合物を抽出する(抽出工程)。本実施例では、図41の(c)部分に示すように、化合物C5、C6が抽出される。

[0175] 特徴量算出部122は、Tanimoto係数で抽出した化合物C5、C6について、本発明に係る特徴量を算出する(特徴量算出工程)。特徴量算出部122は、本実施例ではプローブを電荷中性のアミノ酸とし、第2の態様に係る特徴量算出方法を用いて特徴量(第2の特徴量)を算出する。化合物抽出部140は元の化合物C0と化合物C5、C6との間で特徴量の類似度を算出し(図41の(c)部分及び(d)部分)、類似度(類似性距離)が目標値を満たす(しきい値:100.0以下)化合物C6を抽出する(図41の(e))。なお、図41における「AAM類似度」は、元の化合物C0とライ

ブラリの化合物C 4～C 6との間での、本発明に係る特徴量の類似度（類似性距離）を示す。

[0176] このようにして、ライブラリの化合物C 4～C 6の探索で、元の化合物C 0と特徴量が類似しており（類似度が100.0以下）、かつ構造が異なる（Tanimoto係数が0.2以下）化合物C 6に骨格改変することができた（図42参照）。

[0177] <他の特徴量を併用した場合の出力例>

図43は本発明に係る特徴量と他の特徴量（物性値、毒性値等）を用いた場合の出力例を示す図である。図43の（a）部分はケース1～3について2つの化合物（化合物A，B）の構造を示す図であり、同図の（b）部分は各ケースについて特徴量の類似性距離及びClogP（logPの予測値；物性値の一例）を示す。図中の「類似性距離（AAM記述子）」とは、20種類のアミノ酸をプローブとした場合の、本発明に係る特徴量（回転、並進に対し不変量化している）の類似度を示す指標であり、この値が100以下であることを「特徴量が類似しているか否か」の基準とすることができる。すなわち、図43に示すケース1～3においては、「化合物A，Bの特徴量は類似している（したがって、結合力が同程度である）」ということができる。また、図43の（b）部分に示すClogPについては、「値が1違うと、物性としての差が大きい」（したがって、ケース1～3の化合物は物性の差が大きい）ということができる。以上より、本発明に係る特徴量と他の特徴量を併用することにより、結合力を維持しつつ物性及び／または毒性が改善された化合物を抽出することができることが分かる。

[0178] なお、図44は類似性距離が大きい化合物についての物性値の例を示す。

図44の（a）部分は化合物の構造を示し、同図の（b）部分は特徴量の類似性距離及びClogPを示す。

[0179] <結合力が向上する化合物のスクリーニング（例1）>

図33，38について上述した態様（フローチャート）では、基準とする化合物に対し結合力（特徴量と対応）が維持され、かつ物性値等が目標値を

満たす化合物を抽出している。これに対し図45のフローチャートに示す処理によれば、結合力が向上する化合物（基準とする化合物よりも結合力が高い化合物）のスクリーニングを行うこともできる。このような処理は、処理部100に代えて例えば図46に示す処理部105を備えるスクリーニング装置10（図1を参照）により行うことができる。なお、図45のフローチャートに関し、図33、38と同様の処理には同一のステップ番号を付し、詳細な説明を省略する。

[0180] 結合力予測部133は、「特徴量に基づいて他の化合物との結合力を出力する予測モデル」（後述する付記19を参照）を用いて、化合物の結合力を予測する（ステップS313：予測工程）。付記19に係る予測モデルは、付記18に記載の「予測モデル生成方法」により生成することができる。全化合物について特徴量の算出及び結合力の予測が終了したら（ステップS316でYES）、情報入力部110は操作部400（図1を参照）を介したユーザの操作等により結合力の目標値を入力する（ステップS317：抽出工程）。化合物抽出部140は、複数の化合物のうちから、結合力が目標値を満たす化合物を抽出する（ステップS317-2でYESの場合、ステップS324で化合物を抽出：抽出工程）。このような処理により、結合力が向上する化合物のスクリーニングを行うことができる。

[0181] <結合力が向上する化合物のスクリーニング（例2）>

図47は、結合力が向上する化合物のスクリーニングの他の例を示すフローチャートである。これらの処理は、図46と同様の処理部105を備えるスクリーニング装置10により行うことができる。なお、図47において、図33、35、39のフローチャートと同様の処理については同一のステップ番号を付し、詳細な説明を省略する。

[0182] 結合力予測部133はユーザの操作等に基づいて結合力の目標値を設定する（ステップS310-2：設定工程）。この際、結合力予測部133は、基準とする化合物の結合力よりも高い目標値を設定することができる。そして、結合力予測部133は、上述した予測モデルを用いて、結合力の目標値

に対応する特徴量を推定して、推定した特徴量を特徴量の目標値とする（ステップS 3 1 0－3：推定工程）。情報入力部 1 1 0 は、ユーザの操作等に基づいて特徴量の目標値を入力する（ステップS 3 1 0－4）。特徴量算出部 1 2 2 は第 1 から第 4 の態様のいずれか 1 つに係る特徴量算出方法により特徴量を算出し（ステップS 3 1 2：特徴量算出工程）、化合物抽出部 1 4 0 はステップS 3 1 0－3 で推定した特徴量（すなわち、特徴量の目標値）との類似度がしきい値以上である特徴量を有する化合物を抽出する（ステップS 3 2 4：抽出工程）。このような処理により、結合力の目標値に対応する特徴量を有する化合物を抽出することができ、結合力が向上する化合物のスクリーニングを行うことができる。

[0183] 以上説明したように、本発明の特徴量算出方法、特徴量算出プログラム、及び特徴量算出装置よれば、対象構造体の化学的性質を的確に示す特徴量を算出することができる。また、本発明のスクリーニング方法及びスクリーニングプログラムによれば、特徴量を用いて医薬候補化合物のスクリーニングを効率よく行うことができる。また、本発明に係る化合物創出方法によれば、特徴量を用いて医薬候補化合物の立体構造を効率よく創出することができる。

[0184] （付記）

上述した態様に加えて、以下に記載した内容も本発明の範囲に含まれる。なお、以下に記載の方法、装置、プログラム等は、第 1 ～第 3 の実施形態と同様の構成により実現することができる。

[0185] （付記 1）

付記 1 に係る化合物創出方法は、化合物について、立体構造を示す情報と、第 1 から第 4 の態様のいずれか 1 つに係る特徴量についての第 1 の目標値と、物性を示す物性特性値についての第 2 の目標値及び／または毒性を示す毒性特性値についての第 3 の目標値と、を入力する入力工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構造について第 1 から第 4 の態様のいずれか 1 つに係る特徴量算出方法により特徴量を算出する特

微量算出工程と、候補構造について物性特性値及び／または毒性特性値を算出する特性値算出工程と、候補構造を採用または棄却する候補構造採用工程であって、算出した特徴量と、算出した物性特性値及び／または毒性特性値と、が第1の目標値ないし第3の目標値を満たしている場合は候補構造を採用し、第1の目標値ないし第3の目標値を満たしていない場合は候補構造を棄却する候補構造採用工程と、特徴量と、算出した物性特性値及び／または毒性特性値と、が第1の目標値ないし第3の目標値を満たすまで候補構造取得工程、特徴量算出工程、特性値算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、を有する。

[0186] 図48は付記1に係る化合物創出方法の処理を示すフローチャートである。第10の態様について上述したように、本発明に係る特徴量と、物性値及び／または毒性値と、を用いて化合物のスクリーニングを行うことができるが、図48に示す処理により、スクリーニングの場合と同様に、本発明に係る特徴量が目標値（第1の目標値）を満たし、かつ物性値及び／または毒性値が目標値（第2、第3の目標値）を満たす化合物を創出することができる。なお、物性値及び／または毒性値として、「他の特徴量との併用によるスクリーニング」で記載したものと同様の指標を用いることができる。また、図48に示す処理は、処理部101を備える化合物創出装置20（特徴量算出装置、化合物創出装置；図20、21参照）や、処理部102を備える医薬候補化合物探索装置30（特徴量算出装置、スクリーニング装置、化合物創出装置；図28参照）、あるいはこれらと同様の構成を有する装置、システムにより行うことができる（付記2から7に示す態様についても同様である）。

[0187] （付記2）

付記2に係る化合物創出方法は、化合物について、立体構造を示す情報と、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、物性を示す物性特性値についての第2の目標値及び／または毒性を示す毒性特性値についての第3の目標値と、を入力する入力工程と、立体構造を

変化させて候補構造を得る候補構造取得工程と、候補構造について第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する特徴量算出工程と、候補構造を採用または棄却する候補構造採用工程であって、算出した特徴量が第1の目標値に近づいている場合は候補構造を採用し、算出した特徴量が第1の目標値に近づいていない場合は候補構造を棄却する候補構造採用工程と、特徴量が第1の目標値を満たすまで候補構造取得工程、特徴量算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、特徴量が第1の目標値を満たす候補構造について、物性特性値及び／または毒性特性値を算出する特性値算出工程と、物性特性値及び／または毒性特性値を算出した候補構造に対応する化合物のうちから、物性特性値及び／または毒性特性値が第2の目標値及び／または第3の目標値を満たす化合物を抽出する抽出工程と、を有する。

[0188] 図49は、付記2に係る化合物創出方法の処理を示すフローチャートである。図49に示す処理により、付記1と同様に本発明に係る特徴量が目標値（第1の目標値）を満たし、かつ物性値及び／または毒性値が目標値（第2、第3の目標値）を満たす化合物を創出することができる。

[0189] （付記3）

付記3に係る化合物創出方法は、化合物について、立体構造を示す情報と、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、物性を示す物性特性値についての第2の目標値及び／または毒性を示す毒性特性値についての第3の目標値と、を入力する入力工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構造について物性特性値及び／または毒性特性値を算出する特性値算出工程と、候補構造を採用または棄却する候補構造採用工程であって算出した物性特性値及び／または毒性特性値が第2の目標値及び／または第3の目標値に近づいている場合は候補構造を採用し、算出した物性特性値及び／または毒性特性値が第2の目標値及び／または第3の目標値に近づいていない場合は候補構造を棄却する候補構造採用工程と、算出した物性特性値及び／または毒性特性値が第2の目

標値及び／または第3の目標値を満たすまで候補構造取得工程、特性値算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、算出した物性特性値及び／または毒性特性値が第2の目標値及び／または第3の目標値を満たす候補構造に対応する化合物について、第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する特徴量算出工程と、特徴量を算出した化合物のうちから、特徴量が第1の目標値を満たす化合物を抽出する抽出工程と、を有する。

[0190] 図50は、付記3に係る化合物創出方法の処理を示すフローチャートである。図50に示す処理により、付記1、2と同様に本発明に係る特徴量が目標値（第1の目標値）を満たし、かつ物性値及び／または毒性値が目標値（第2，第3の目標値）を満たす化合物を創出することができる。

[0191] （付記4）

付記4に係る化合物創出方法は、化合物について、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、物性を示す物性特性値についての第2の目標値及び／または毒性を示す毒性特性値についての第3の目標値と、を入力する入力工程と、第1の目標値に基づいて、第15から第17の態様のいずれか1つに係る化合物創出方法により複数の化合物を創出する創出工程と、複数の化合物について、物性特性値及び／または毒性特性値を算出する特性値算出工程と、複数の化合物のうちから、物性特性値及び／または毒性特性値が第2の目標値及び／または第3の目標値を満たす化合物を抽出する抽出工程と、を有する。

[0192] 図51は、付記4に係る化合物創出方法の処理を示すフローチャートである。付記4においては、候補構造の改変により化合物を創出する付記1～3と異なり、化合物の創出において第15から第17の態様のいずれか1つに係る化合物創出方法を用いている。図51に示す処理によっても、付記1～3と同様に本発明に係る特徴量が目標値（第1の目標値）を満たし、かつ物性値及び／または毒性値が目標値（第2，第3の目標値）を満たす化合物を創出することができる。

[0193] (付記5)

付記5に係る化合物創出方法は、付記1から3のいずれか1つに係る化合物創出方法において、入力された情報に基づいて第1から第4の態様に係る特徴量算出方法のうちいずれか1つを選択する第1の選択工程をさらに有し、特徴量算出工程では選択した特徴量算出方法により特徴量を算出する。処理部101を備える化合物創出装置20や、処理部102を備える医薬候補化合物探索装置30、あるいはこれらと同様の構成を有する装置、システムは、ユーザの操作に応じて特徴量算出方法を選択することができる。

[0194] (付記6)

付記6に係る化合物創出方法は、付記1から5のいずれか1つに係る化合物創出方法において、物性特性値の算出方法及び／または毒性特性値の算出方法を選択する第2の選択工程をさらに有し、特性値算出工程では選択した算出方法により物性特性値及び／または毒性特性値を算出する。処理部101を備える化合物創出装置20や、処理部102を備える医薬候補化合物探索装置30、あるいはこれらと同様の構成を有する装置、システムは、ユーザの指示に応じて物性特性値の算出方法及び／または毒性特性値の算出方法を選択することができる。

[0195] (付記7)

付記7に係る化合物創出方法は、付記1から6のいずれか1つに係る化合物創出方法において、物性特性値は分子量、ClogP、極性表面積、分極率、及びローテータブルボンドの数(RBN)のうち1つ以上である。RBNの定義は「物性値及び／または毒性値との併用」の項で上述した通りである。

[0196] (付記8)

付記8に係る化合物創出方法は、化合物について、立体構造を示す情報と、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、立体構造を示す構造特徴量についての第2の目標値と、を入力する入力工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構

造について、第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する第1の特徴量算出工程と、候補構造について構造特徴量を算出する第2の特徴量算出工程と、候補構造を採用または棄却する候補構造採用工程であって、算出した特徴量及び算出した構造特徴量が第1の目標値及び第2の目標値を満たしている場合は候補構造を採用し、第1の目標値及び第2の目標値を満たしていない場合は候補構造を棄却する候補構造採用工程と、特徴量及び構造特徴量が第1の目標値及び第2の目標値を満たすまで候補構造取得工程、第1の特徴量算出工程、第2の特徴量算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、を有する。

[0197] 図52は付記8に係る化合物創出方法の処理を示すフローチャートである。第11の態様について上述したように、本発明に係る特徴量と化合物の構造を示す特徴量（構造特徴量）とを用いてスクリーニングを行うことができるが、スクリーニングの場合と同様に本発明に係る特徴量と化合物の構造を示す特徴量（構造特徴量）とを用いて化合物の創出を行うことができる。図52に示す処理によれば、結合力（特徴量と対応）を維持しつつ骨格が異なる化合物を創出することができる。なお、図52に示す処理は、処理部101を備える化合物創出装置20（特徴量算出装置、化合物創出装置；図20、21参照）や、処理部102を備える医薬候補化合物探索装置30（特徴量算出装置、スクリーニング装置、化合物創出装置；図28参照）、あるいはこれらと同様の構成を有する装置、システムにより行うことができる（付記9から14に示す態様についても同様である）。

[0198] （付記9）

付記9に係る化合物創出方法は、化合物について、立体構造を示す情報と、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、立体構造を示す構造特徴量についての第2の目標値と、を入力する入力工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構造について、第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する第1の特徴量算出工程と、候補構造を採用または棄却す

る候補構造採用工程であって、算出した特徴量が第1の目標値を満たしている場合は候補構造を採用し、算出した特徴量が第1の目標値を満たしていない場合は候補構造を棄却する候補構造採用工程と、特徴量が第1の目標値を満たすまで候補構造取得工程、第1の特徴量算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、特徴量が第1の目標値を満たす候補構造について構造特徴量を算出する構造特徴量算出工程と、構造特徴量を算出した候補構造に対応する化合物から、第2の目標値との類似度がしきい値以下である構造特徴量を有する化合物を抽出する抽出工程と、を有する。

[0199] 図53は、付記9に係る化合物創出方法の処理を示すフローチャートである。図53に示す処理により、付記8と同様に結合力（特徴量と対応）を維持しつつ骨格が異なる化合物を創出することができる。

[0200] （付記10）

付記10に係る化合物創出方法は、化合物について、立体構造を示す情報と、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、立体構造を示す構造特徴量についての第2の目標値と、を入力する入力工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構造について構造特徴量を算出する第2の特徴量算出工程と、候補構造を採用または棄却する候補構造採用工程であって、算出した構造特徴量が第2の目標値を満たしている場合は候補構造を採用し、算出した構造特徴量が第2の目標値を満たしていない場合は候補構造を棄却する候補構造採用工程と、構造特徴量が第2の目標値を満たすまで候補構造取得工程、第2の特徴量算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、構造特徴量が第2の目標値を満たす候補構造について特徴量を算出する第1の特徴量算出工程と、特徴量を算出した候補構造に対応する化合物から、第1の目標値との類似度がしきい値以下である構造特徴量を有する化合物を抽出する抽出工程と、を有する。

[0201] 図54は、付記10に係る化合物創出方法の処理を示すフローチャートで

ある。図54に示す処理により、付記8、9と同様に結合力（特徴量と対応）を維持しつつ骨格が異なる化合物を創出することができる。

[0202] （付記11）

付記11に係る化合物創出方法は、化合物について、立体構造を示す情報と、第1から第4の態様のいずれか1つに係る特徴量についての第1の目標値と、立体構造を示す構造特徴量についての第2の目標値と、を入力する入力工程と、第15から第17の態様のいずれか1つに係る化合物創出方法により、特徴量が第1の目標値を満たす複数の候補構造を生成する生成工程と、複数の候補構造について構造特徴量を算出する算出工程と、複数の候補構造から、構造特徴量がしきい値以下である化合物を抽出する抽出工程と、を有する。

[0203] 図55は、付記11に係る化合物創出方法の処理を示すフローチャートである。候補構造の改変により化合物を創出する付記8～10と異なり、化合物の創出において第15から第17の態様のいずれか1つに係る化合物創出方法を用いている。図55に示す処理によっても、付記8から10と同様に結合力（特徴量と対応）を維持しつつ骨格が異なる化合物を創出することができる。

[0204] （付記12）

付記12に係る化合物創出方法は、付記8から10のいずれか1つに係る化合物創出方法において、第1から第4の態様に係る特徴量算出方法のうちいずれか1つを情報に基づいて選択する第1の選択工程をさらに有し、第1の特徴量算出工程では、選択した特徴量算出方法により特徴量を算出する。特徴量算出方法の選択は、ユーザの指示に基づいて行うことができる。

[0205] （付記13）

付記13に係る化合物創出方法は付記8から12のいずれか1つに係る化合物創出方法において、構造特徴量は化合物についてのFingerprint（Fingerprint記述子）である。Fingerprint記述子は、特定の官能基や部分構造の有無を0と1で表現した記述子である。

[0206] (付記 1 4)

付記 1 4 に係る化合物創出方法は付記 1 2 または 1 3 に係る化合物創出方法において、構造特徴量の類似度を示す指標としてTanimoto係数を用いる。Tanimoto係数は、比較したい 2 化合物のFingerprint記述子を与えると 0 から 1 までの実数として算出され、1 に近くなるほど類似度が高くなるように定義されている。

[0207] (付記 1 5)

付記 1 5 に係る化合物創出方法は、化合物について、立体構造を示す情報と、結合力の目標値と、を入力する入力工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構造について、第 1 から第 4 の態様のいずれか 1 つに係る特徴量算出方法により特徴量を算出する特徴量算出工程と、第 1 から第 4 の態様のいずれか 1 つに係る特徴量に基づいて結合力を出力する予測モデルを用いて、算出した特徴量に対応する候補構造の結合力を推定する推定工程と、候補構造を採用または棄却する候補構造採用工程であって、推定した結合力が目標値を満たしている場合は候補構造を採用し、推定した結合力が目標値を満たしていない場合は候補構造を棄却する候補構造採用工程と、結合力が目標値を満たすまで候補構造取得工程、特徴量算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、を有する。

[0208] 図 5 6 は、付記 1 5 に係る化合物創出方法の処理を示すフローチャートである。図 5 6 に示す処理によれば、上述した「結合力が向上する化合物のスクリーニング（例 1、2）」の項で説明した「予測モデルを用いたスクリーニング」と同様に、予測モデルを用いて「結合力が向上する化合物の創出」を行うことができる。付記 1 5 に係る化合物創出方法は、化合物創出装置 20、医薬候補化合物探索装置 30 等により実行することができる。

[0209] (付記 1 6)

付記 1 6 に係る化合物創出方法は、化合物について、結合力の目標値を入力する入力工程と、第 1 から第 4 の態様のいずれか 1 つに係る特徴量に基づ

いて結合力を出力する予測モデルを用いて、目標値を満たす特徴量を推定する推定工程と、立体構造を変化させて候補構造を得る候補構造取得工程と、候補構造について、第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する第1の特徴量算出工程と、候補構造を採用または棄却する候補構造採用工程であって、算出した特徴量が推定した特徴量を満たしている場合は候補構造を採用し、算出した特徴量が推定した特徴量を満たしていない場合は候補構造を棄却する候補構造採用工程と、算出した特徴量が推定した特徴量を満たすまで候補構造取得工程、第1の特徴量算出工程、及び候補構造採用工程における処理を繰り返させる制御工程と、を有する。

[0210] 図57は、付記16に係る化合物創出方法の処理を示すフローチャートである。図57に示す処理により、付記15と同様に予測モデルを用いて「結合力が向上する化合物の創出」を行うことができる。

[0211] (付記17)

付記17に係る化合物創出方法は、化合物について、結合力の目標値を入力する入力工程と、第1から第4の態様のいずれか1つに係る特徴量に基づいて結合力を出力する予測モデルを用いて、目標値を満たす特徴量を推定する推定工程と、推定した特徴量に基づいて、第15から第17の態様のいずれか1つに係る化合物創出方法により、目標値に対応する化合物の立体構造を生成する生成工程と、を有する。

[0212] 図58は、付記17に係る化合物創出方法の処理を示すフローチャートである。図58に示す処理により、付記15、16と同様に予測モデルを用いて「結合力が向上する化合物の創出」を行うことができる。

[0213] (付記18)

付記18に係る予測モデル生成方法は、複数の化合物について立体構造を示す情報を入力する入力工程と、複数の化合物のそれぞれについて、第1から第4の態様のいずれか1つに係る特徴量算出方法を用いて特徴量を算出する特徴量算出工程と、特徴量を説明変数とし、特徴量に対応する結合力を教

師データとした機械学習により化合物の結合力の予測モデルを生成する工程と、を有する。

[0214] (付記 19)

付記 19に係る予測モデルは、コンピュータに、化合物の特徴量に基づいて結合力を出力させる予測モデルであって、特徴量を説明変数とし結合力を教師データとした機械学習により構成されたニューラルネットワークを備え、入力された特徴量に対し学習済みの重み付け係数に基づく演算を行い、結合力を出力させる。付記 18に係る予測モデル生成方法、及び付記 19に係る予測モデルを用いて、「結合力が向上する化合物のスクリーニング（例 1、2）」の項で説明したスクリーニングや付記 15～17で説明した化合物の創出を行うことができる。なお、図 32の（b）部分について上述したのと同様に、付記 19に係る予測モデルのニューラルネットワークに対し出力側の層に結合力を与えて出力層から入力層に向けて情報を更新することで、入力層から化合物の特徴量（例えば、第 1 から第 4 の態様のいずれか 1 つに係る特徴量；学習のさせ方に依存する）が出力される。これにより、例えば付記 16、17において、結合力の目標値から特徴量を推定することができる。

[0215] <予測モデルの毒性予測への応用>

上述した予測モデルの生成方法、及び生成された予測モデルは毒性予測に応用することができる。ここで「毒性」として、例えば hERG 阻害（hERG : Human Ether-a-go-go Related Gene）、CYP 阻害（CYP : Cytochrome P450）を挙げることができる。hERG、CYP は生体高分子なので、生体高分子をターゲットとした特徴量を用いることにより、hERG 阻害、CYP 阻害を予測することができる。具体的には、hERG 阻害による毒性の回避は、hERG との結合力を「下げる」（基準となる化合物の結合力よりも下げる、結合力を目標値よりも下げる等）ことで実現できる。CYP 阻害による毒性の回避についても同様である。このような毒性の予測は、毒性の低い化合物のスクリーニングあるいは創出において行うことができる。

[0216] (付記 20)

付記 20 に係る化合物創出プログラムは、第 15 から第 17 の態様、付記 1 から 17 のいずれか 1 つに係る化合物創出方法をコンピュータに実行させる。付記 20 に係る化合物創出プログラムのコンピュータ読み取り可能なコードを記録した非一時的記録媒体も、本発明の一態様として挙げることができる。

[0217] (付記 21)

付記 21 に係る特徴量算出方法は、化学的性質を有する複数の単位構造体から構成される対象構造体を指定する対象構造体指定工程と、対象構造体について複数の単位構造体による立体構造を発生させる立体構造発生工程と、立体構造の周辺における 1 種類以上のプローブの集積度合いを 3 次元空間において定量化した特徴量を算出する特徴量算出工程と、を有し、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である。なお、以下に説明する付記 22 から 40 においては、付記 21 に係る特徴量算出方法を用いてもよいし、上述した第 1 ～第 4 の態様に係る特徴量算出方法を用いてもよい。

[0218] (付記 22)

付記 22 に係るスクリーニング方法は、付記 21 に係る特徴量算出方法を用いたスクリーニング方法であって、対象構造体指定工程では対象構造体として結合化合物の構造式と、複数の探索対象化合物の構造式と、を指定し、立体構造発生工程では結合化合物の構造式から結合化合物の立体構造を発生させる方法と、複数の探索対象化合物の構造式から複数の探索対象化合物の立体構造を発生させる方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化して特徴量を算出する方法と、を指定する。付記 22 に係るスクリーニング方法はさらに、結合化合物の特徴量と複数の探索対象化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、結合化合物の立体構造に基づいて、複数の化合物

から、特徴量が結合化合物についての特徴量としきい値以上の類似度である化合物を抽出する工程と、を有する。

[0219] 図59は付記22に係るスクリーニング方法の処理を示すフローチャートであり、これらの処理により、例えば一つの結合化合物の構造式から、その結合化合物と同じ結合力を持つ化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。図59に示す処理は、第1、第3の実施形態に係る装置（スクリーニング装置10、第3の実施形態に係る医薬候補化合物探索装置30）、あるいはこれらの装置と同様の構成を有する装置、システムにより実行することができる。これらの装置やシステムは、装置やシステムにスクリーニング方法を実行させるプログラムを用いることができ、実行に際してそのプログラムのコンピュータ読み取り可能なコードが記録された非一時的記録媒体を参照することができる。後述する付記に係るスクリーニング方法についても同様である。なお、これらの装置やシステムは、ユーザの選択操作等に応じて立体構造を発生させる方法等を指定することができる（後述する付記についても同様である）。

[0220] なお、付記22は結合化合物（タンパク質以外の標的生体高分子に結合することが確認されている化合物）の情報を用いたスクリーニング方法であるが、リガンド（標的タンパク質との結合が確認されている化合物）の情報を用いたスクリーニングも同様に行うことができる。具体的には、付記22において「結合化合物の構造式」、「結合化合物の立体構造」、「結合化合物（について）の特徴量」をそれぞれ「リガンドの構造式」、「リガンドの立体構造」、「リガンド（について）の特徴量」に置き換えることで、リガンドの情報を用いたスクリーニングを行うことができる（一つのリガンドの構造式から、そのリガンドと同じ結合力を持つ化合物を、既存の化合物群からスクリーニングする）。同様に、以下の各「付記」における他のスクリーニング方法や化合物創出方法も、結合化合物だけでなくリガンドの情報を用いて行うことができる。

[0221] （付記23）

付記 23 に係る化合物創出方法は、付記 21 に係る特徴量算出方法を用いた化合物創出方法であって、対象構造体指定工程では対象構造体として結合化合物の構造式を指定し、立体構造発生工程では結合化合物の構造式から結合化合物の立体構造を発生させる方法と、創出する化合物の立体構造の生成方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化して結合化合物の特徴量を算出する方法と、を指定する。付記 23 に係る化合物創出方法はさらに、結合化合物の特徴量と他の化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、結合化合物の特徴量との類似度がしきい値以上である特徴量を有する化合物の立体構造を発生させる工程と、を有する。

[0222] 図 60 は付記 23 に係る化合物創出方法の処理を示すフローチャートであり、これらの処理により、例えば一つの結合化合物の構造式から、その結合化合物と同じ結合力を持つ化合物を、逆問題を解くことにより新たに創出することができる。図 60 に示す処理は、第 2、第 3 の実施形態に係る装置（化合物創出装置 20、医薬候補化合物探索装置 30）、あるいはこれらの装置と同様の構成を有する装置、システムにより実行することができる。これらの装置やシステムは、装置やシステムにスクリーニング方法を実行させるプログラムを用いることができ、実行に際してそのプログラムのコンピュータ読み取り可能なコードが記録された非一時的記録媒体を参照することができる。後述する付記に係る化合物創出方法についても同様である。なお、「化合物の立体構造を与えて特徴量（例えば、第 1 から第 4 の態様のいずれか 1 つに係る特徴量）を求める」ことを「順問題」ということができ、「与えられた特徴量を有する立体構造を求める」ことを「逆問題」ということができる。

[0223] （付記 24）

付記 24 に係るスクリーニング方法は、付記 21 に係る特徴量算出方法を

用いたスクリーニング方法であって、対象構造体指定工程では生体高分子の活性部位の立体構造と、複数の探索対象化合物の構造式と、を指定し、立体構造発生工程では活性部位に結合するポケット構造体の立体構造を発生させる方法と、複数の探索対象化合物の構造式から複数の探索対象化合物の立体構造を発生させる方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、ポケット構造体に対するプローブの集積度合いを定量化する方法と、集積度合いに基づいて特徴量を算出する方法と、を指定する。付記 2 4 に係るスクリーニング方法はさらに、ポケット構造体の特徴量と複数の探索対象化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、生体高分子の活性部位の立体構造に基づいて、複数の探索対象化合物から、特徴量が生体高分子についての特徴量としきい値以上の類似度である化合物を抽出する工程と、を有する。図 6 1 は付記 2 4 に係るスクリーニング方法の処理を示すフローチャートであり、これらの処理により、例えば生体高分子の活性部位の立体構造から、その活性部位に適合する化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。

[0224] (付記 2 5)

付記 2 5 に係る化合物創出方法は、付記 2 1 に係る特徴量算出方法を用いた化合物創出方法であって、対象構造体指定工程では対象構造体として生体高分子の活性部位の構造式を指定し、立体構造発生工程では活性部位に結合するポケット構造体の立体構造を発生させる方法と、創出する化合物の立体構造の生成方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、ポケット構造体に対するプローブの集積度合いを定量化する方法と、プローブの集積度合いを定量化してポケット構造体の特徴量を算出する方法と、を指定する。付記 2 5 に係る化合物創出方法はさらに、ポケット構造体の特徴量と化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定す

る工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、ポケット構造体の特徴量との類似度がしきい値以上である特徴量を有する化合物の立体構造を発生させる工程と、を有する。図62は付記25に係る化合物創出方法の処理を示すフローチャートであり、これらの処理により、例えば生体高分子の活性部位の立体構造から、その活性部位に適合する化合物を、逆問題を解くことにより新たに創出することができる。

[0225] (付記26)

付記26に係るスクリーニング方法は、付記21に係る特徴量算出方法を用いたスクリーニング方法であって、対象構造体指定工程では複数の結合化合物の構造式と、複数の探索対象化合物の構造式と、を指定し、立体構造発生工程では複数の結合化合物の立体構造を発生させる方法と、複数の探索対象化合物の立体構造を発生させる方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、複数の結合化合物のそれぞれに対するプローブの集積度合いを定量化する方法と、集積度合いに基づいて特徴量を算出する方法と、を指定する。付記26に係るスクリーニング方法はさらに、複数の結合化合物のそれぞれについての結合力の値を入力する工程と、複数の結合化合物についての構造式と結合力の値との組から化合物の結合力の予測モデルを生成する方法を指定する工程と、結合力の目標値を指定する工程と、複数の探索対象化合物から、予測モデルを用いて目標値の結合力を有する化合物を抽出する工程と、を有する。図63は付記26に係るスクリーニング方法の処理を示すフローチャートであり、これらの処理により、例えば複数の化合物の構造式と結合力の値のセット(組)を用いて、目標の結合力を有する化合物を、既存の化合物群(複数の探索対象化合物)からスクリーニングすることができる。なお、付記26において、付記18、19に示す予測モデル生成方法及び予測モデルを用いることができる。

[0226] (付記27)

付記 27 に係る化合物創出方法は、付記 21 に係る特徴量算出方法を用いた化合物創出方法であって、対象構造体指定工程では複数の結合化合物の構造式を指定し、立体構造発生工程では複数の結合化合物の立体構造を発生させる方法と、創出する化合物の立体構造の生成方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化する方法と、集積度合いに基づいて特徴量を算出する方法と、を指定する。付記 27 に係る化合物創出方法はさらに、複数の結合化合物のそれぞれについての結合力の値を指定する工程と、複数の結合化合物についての構造式と結合力の値との組から化合物の結合力の予測モデルを生成する方法を指定する工程と、結合力の目標値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、結合力が目標値以上である化合物の立体構造を発生させる工程と、を有する。図 64 は付記 27 に係る化合物創出方法の処理を示すフローチャートであり、これらの処理により、例えば複数の化合物の構造式と結合力の値のセット（組）を用いて、目標の結合力を有する化合物を、逆問題を解くことにより新たに創出することができる。なお、付記 27 において、付記 18, 19 に示す予測モデル生成方法及び予測モデルを用いることができる。

[0227] (付記 28)

付記 28 に係るスクリーニング方法は、付記 21 に係る特徴量算出方法を用いたスクリーニング方法であり、対象構造体指定工程では複数の探索対象化合物の構造式を指定し、立体構造発生工程では複数の探索対象化合物の構造式から複数の探索対象化合物の立体構造を発生させる方法を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化して特徴量を算出する方法と、を指定する。付記 28 に係るスクリーニング方法はさらに、結合化合物の立体構造を示す情報を入力する工程と、結合化合物の特徴量と複数の探索対象化合物の特徴量との類似度の算出方法を指定する工程と、類似

度のしきい値を指定する工程と、結合化合物の立体構造に基づいて、複数の化合物から、特徴量が結合化合物についての特徴量としきい値以上の類似度である化合物を抽出する工程と、を有する。付記 28 に係るスクリーニング方法は付記 22 に係るスクリーニング方法を改変したものであり、結合化合物の構造式に代えて結合化合物の立体構造を入力する。このため、付記 21 に係る特徴量算出方法の工程のうち、結合化合物については立体構造を発生させるための工程（対象構造体指定工程と立体構造発生工程）は不要である。付記 28 によっても、付記 22 と同様に一つの結合化合物の立体構造から、その結合化合物と同じ結合力を持つ化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。

[0228] （付記 29）

付記 29 に係る化合物創出方法は、特徴量算出方法を用いた化合物創出方法である。付記 29 で用いる特徴量算出方法は、対象構造体について複数の単位構造体による立体構造を発生させる立体構造発生工程と、立体構造の周辺における 1 種類以上のプローブの集積度合いを 3 次元空間において定量化した特徴量を算出する特徴量算出工程と、を有し、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である。付記 29 に係る化合物創出方法は、結合化合物の立体構造を示す情報を入力する工程をさらに有し、立体構造発生工程では創出する化合物の立体構造の生成方法を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化して結合化合物の特徴量を算出する方法と、を指定する。付記 29 に係る化合物創出方法はさらに、結合化合物の特徴量と他の化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、結合化合物の特徴量との類似度がしきい値以上である特徴量を有する化合物の立体構造を発生させる工程と、を有する。付記 29 に係る化合物創出方法は付記 23 に係る化合物創出方法

を改変したものであり、結合化合物の構造式に代えて結合化合物の立体構造を入力する。このため、結合化合物については、立体構造を発生させるための工程（対象構造体指定工程と立体構造発生工程）は不要である。付記 2 9 によっても、付記 2 3 と同様に一つの結合化合物の立体構造から、その結合化合物と同じ結合力を持つ化合物を、逆問題を解くことにより新たに創出することができる。

[0229] (付記 3 0)

付記 3 0 に係るスクリーニング方法は、付記 2 1 に係る特徴量算出方法を用いたスクリーニング方法であって、生体高分子の立体構造を示す情報を入力する工程と、情報に基づいて生体高分子の活性部位を特定する工程と、を有し、対象構造体指定工程では複数の探索対象化合物の構造式を指定し、立体構造発生工程では活性部位に結合するポケット構造体の立体構造を発生させる方法と、複数の探索対象化合物の構造式から複数の探索対象化合物の立体構造を発生させる方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、ポケット構造体に対するプローブの集積度合いを定量化する方法と、集積度合いに基づいて特徴量を算出する方法と、を指定する。付記 3 0 に係るスクリーニング方法はさらに、ポケット構造体の特徴量と複数の探索対象化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、生体高分子の活性部位の立体構造に基づいて、複数の探索対象化合物から、特徴量が生体高分子についての特徴量としきい値以上の類似度である化合物を抽出する工程と、を有する。付記 3 0 に係るスクリーニング方法は付記 2 4 に係るスクリーニング方法を改変したものであり、生体高分子の活性部位の立体構造に代えて生体高分子の立体構造を入力する。このため、付記 3 0 に係るスクリーニング方法は「生体高分子の立体構造を示す情報に基づいて、生体高分子の活性部位を特定する工程」を有する。付記 3 0 によっても、付記 2 4 と同様に生体高分子の立体構造から、その活性部位に適合する化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニン

グすることができる。

[0230] (付記 3 1)

付記 3 1 に係る化合物創出方法は、付記 2 1 に係る特徴量算出方法を用いた化合物創出方法であり、生体高分子の立体構造を示す情報を入力する工程と、情報に基づいて生体高分子の活性部位を特定する工程と、を有し、対象構造体指定工程では生体高分子の活性部位の立体構造を指定し、立体構造発生工程では活性部位に結合するポケット構造体の立体構造を発生させる方法と、創出する化合物の立体構造の生成方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、ポケット構造体に対するプローブの集積度合いを定量化する方法と、プローブの集積度合いを定量化してポケット構造体の特徴量を算出する方法と、を指定する。付記 3 1 に係る化合物創出方法はさらに、ポケット構造体の特徴量と化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、ポケット構造体の特徴量との類似度がしきい値以上である特徴量を有する化合物の立体構造を発生させる工程と、を有する。付記 3 1 に係る化合物創出方法は付記 2 5 に係る化合物創出方法を改変したものであり、生体高分子の活性部位の立体構造に代えて生体高分子の立体構造を入力する。このため、付記 3 1 に係る化合物創出方法は「生体高分子の立体構造を示す情報に基づいて、生体高分子の活性部位を特定する工程」を有する。付記 3 1 によっても、付記 2 5 と同様に生体高分子の立体構造から、その活性部位に適合する化合物を、逆問題を解くことにより新たに創出することができる。

[0231] (付記 3 2)

付記 3 2 に係るスクリーニング方法は、付記 2 1 に係る特徴量算出方法を用いたスクリーニング方法であって、生体高分子の一次配列を示す情報を入力する工程と、情報に基づいて生体高分子の立体構造を発生させる工程と、立体構造に基づいて生体高分子の活性部位を特定する工程と、を有し、対象

構造体指定工程では複数の探索対象化合物の構造式を指定し、立体構造発生工程では活性部位に結合するポケット構造体の立体構造を発生させる方法と、複数の探索対象化合物の構造式から複数の探索対象化合物の立体構造を発生させる方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、ポケット構造体に対するプローブの集積度合いを定量化する方法と、集積度合いに基づいて特徴量を算出する方法と、を指定する。付記 3 2 に係るスクリーニング方法はさらに、ポケット構造体の特徴量と複数の探索対象化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、生体高分子の活性部位の立体構造に基づいて、複数の探索対象化合物から、特徴量が生体高分子についての特徴量としきい値以上の類似度である化合物を抽出する工程と、を有する。付記 3 2 に係るスクリーニング方法は付記 2 4 に係るスクリーニング方法を改変したものであり、生体高分子の活性部位の立体構造に代えて生体高分子の一次配列を入力する。このため、付記 3 2 に係るスクリーニング方法は「生体高分子の一次配列を示す情報に基づいて、生体高分子の立体構造を発生させる工程」と「生体高分子の立体構造に基づいて生体高分子の活性部位を特定する工程」とを有する。付記 3 2 によっても、付記 2 4 と同様に生体高分子の一次配列情報から、その活性部位に適合する化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。

[0232] (付記 3 3)

付記 3 3 に係る化合物創出方法では、対象構造体について立体構造を発生させる立体構造発生工程と、立体構造の周辺における 1 種類以上のプローブの集積度合いを 3 次元空間において定量化した特徴量を算出する特徴量算出工程と、を有し、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である特徴量算出方法を用いる。また、付記 3 3 に係る化合物創出方法は、生体高分子の一次配列を示す情報を入力する工程と、情報に基づいて生体高分子の立体構造を発生させる工

程と、立体構造に基づいて生体高分子の活性部位を特定する工程と、を有し、立体構造発生工程では活性部位に結合するポケット構造体の立体構造を発生させる方法と、創出する化合物の立体構造の生成方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、ポケット構造体に対するプローブの集積度合いを定量化する方法と、プローブの集積度合いを定量化してポケット構造体の特徴量を算出する方法と、を指定する。付記 33 に係る化合物創出方法はさらに、ポケット構造体の特徴量と化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、ポケット構造体の特徴量との類似度がしきい値以上である特徴量を有する化合物の立体構造を発生させる工程と、を有する。

[0233] 付記 33 に係る化合物創出方法は付記 25 に係る化合物創出方法を改変したものであり、生体高分子の活性部位の立体構造に代えて生体高分子の一次配列を入力する。このため、付記 33 に係る化合物創出方法は「生体高分子の一次配列を示す情報に基づいて生体高分子の立体構造を発生させる工程」と、「生体高分子の立体構造に基づいて生体高分子の活性部位を特定する工程」と、を有する。付記 33 によっても、付記 25 と同様に生体高分子の一次配列から、その活性部位に適合する化合物を、逆問題を解くことにより新たに創出することができる。

[0234] (付記 34)

付記 34 に係るスクリーニング方法は、付記 21 に係る特徴量算出方法を用いたスクリーニング方法であって、複数の結合化合物について立体構造を入力する工程を有し、対象構造体指定工程では複数の探索対象化合物の構造式と、を指定し、立体構造発生工程では複数の探索対象化合物の立体構造を発生させる方法と、を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、複数の結合化合物のそれぞれに対するプローブの集積度合いを定量化する方法と、集積度

合いに基づいて特徴量を算出する方法と、を指定する。付記 3 4 に係るスクリーニング方法はさらに、複数の結合化合物のそれぞれについての結合力の値を指定する工程と、複数の結合化合物についての立体構造と結合力の値との組から化合物の結合力の予測モデルを生成する方法を指定する工程と、結合力の目標値を指定する工程と、複数の探索対象化合物から、予測モデルを用いて目標値の結合力を有する化合物を抽出する工程と、を有する。

[0235] 付記 3 4 に係るスクリーニング方法は付記 2 6 に係るスクリーニング方法を改変したもので、結合化合物の構造式ではなく、結合化合物の立体構造を入力する。このため、付記 2 1 に係る特徴量算出方法の工程のうち、結合化合物については立体構造を発生させるための工程（対象構造体指定工程と立体構造発生工程）は不要である。付記 3 4 に係るスクリーニング方法によっても、付記 2 6 と同様に複数の化合物の立体構造と結合力の値のセット（組）に基づいて、目標の結合力を有する化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。なお、付記 3 4 においても、付記 1 8, 1 9 に示す予測モデル生成方法及び予測モデルを用いることができる。

[0236] （付記 3 5）

付記 3 5 に係る化合物創出方法では、対象構造体について立体構造を発生させる立体構造発生工程と、立体構造の周辺における 1 種類以上のプローブの集積度合いを 3 次元空間において定量化した特徴量を算出する特徴量算出工程と、を有し、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である特徴量算出方法を用いる。また、付記 3 5 に係る化合物創出方法は、複数の結合化合物のそれぞれについて立体構造を示す情報を入力する工程を有し、立体構造発生工程では創出する化合物の立体構造の生成方法を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化する方法と、集積度合いに基づいて特徴量を算出する方法と、を指定する。付記 3 5 に係る化合物創出方法はさらに、

複数の結合化合物のそれぞれについての結合力の値を指定する工程と、複数の結合化合物についての立体構造と結合力の値との組から化合物の結合力の予測モデルを生成する方法を指定する工程と、結合力の目標値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、結合力が目標値以上である化合物の立体構造を発生させる工程と、を有する。

[0237] 付記 35 に係る化合物創出方法は付記 27 に係る化合物創出方法を改変したもので、結合化合物の構造式ではなく、結合化合物の立体構造を入力する。このため、結合化合物については立体構造を発生させるための工程（対象構造体指定工程と立体構造発生工程）は不要である。付記 35 に係る化合物創出方法によっても、付記 27 と同様に複数の化合物の立体構造と結合力の値のセット（組）を用いて、目標の結合力を有する化合物を、逆問題を解くことにより新たに創出することができる。なお、付記 35 においても、付記 18, 19 に示す予測モデル生成方法及び予測モデルを用いることができる。

[0238] （付記 36）

付記 36 に係るスクリーニング方法は、付記 21 に係る特徴量算出方法を用いたスクリーニング方法であって、結合化合物について特徴量を入力する工程を有し、対象構造体指定工程では複数の探索対象化合物の構造式を指定し、立体構造発生工程では複数の探索対象化合物の構造式から複数の探索対象化合物の立体構造を発生させる方法を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化して特徴量を算出する方法と、を指定する。付記 36 に係るスクリーニング方法はさらに、結合化合物の特徴量と複数の探索対象化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、複数の化合物から、特徴量が化合物についての特徴量としきい値以上の類似度である化合物を抽出する工程と、を有する。付記 36 に係るスクリーニング方法は付記 22 に係るスクリーニング方

法を改変したもので、結合化合物の構造式ではなく特徴量を入力する。このため、結合化合物については、付記 2 1 に係る特徴量算出方法の各工程（対象構造体指定工程、立体構造発生工程、及び特徴量算出工程）は不要である。付記 3 6 に係るスクリーニング方法によっても、付記 2 2 と同様に一つの結合化合物の特徴量から、その結合化合物と同じ結合力を持つ化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。

[0239] （付記 3 7）

付記 3 7 に係る化合物創出方法では、対象構造体について立体構造を発生させる立体構造発生工程と、立体構造の周辺における 1 種類以上のプローブの集積度合いを 3 次元空間において定量化した特徴量を算出する特徴量算出工程と、を有し、プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である特徴量算出方法を用いる。また、付記 3 7 に係る化合物創出方法は、結合化合物について特徴量を入力する工程を有し、立体構造発生工程では創出する化合物の立体構造の生成方法を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、プローブの集積度合いを定量化して結合化合物の特徴量を算出する方法と、を指定する。付記 3 7 に係る化合物創出方法はさらに、結合化合物の特徴量と他の化合物の特徴量との類似度の算出方法を指定する工程と、類似度のしきい値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、結合化合物の特徴量との類似度がしきい値以上である特徴量を有する化合物の立体構造を発生させる工程と、を有する。

[0240] 付記 3 7 に係る化合物創出方法は付記 2 3 に係る化合物創出方法を改変したもので、結合化合物の構造式ではなく特徴量を入力する。このため、結合化合物については特徴量算出方法の各工程（対象構造体指定工程、立体構造発生工程、及び特徴量算出工程）は不要である。付記 3 7 に係る化合物創出方法によっても、付記 2 3 と同様に一つの結合化合物の特徴量から、その結

合化合物と同じ結合力を持つ化合物を、逆問題を解くことにより新たに創出することができる。

[0241] (付記 38)

付記 38に係るスクリーニング方法は、付記 21に係る特徴量算出方法を用いたスクリーニング方法であって、結合化合物についての特徴量を入力する工程を有し、対象構造体指定工程では複数の探索対象化合物の構造式を指定し、立体構造発生工程では複数の探索対象化合物の立体構造を発生させる方法を指定し、特徴量算出工程ではプローブの構造式と、プローブの構造式からプローブの立体構造を発生させる方法と、を指定する。付記 38に係るスクリーニング方法はさらに、複数の結合化合物のそれぞれについての結合力の値を指定する工程と、複数の結合化合物についての特徴量と結合力の値との組から化合物の結合力の予測モデルを生成する方法を指定する工程と、結合力の目標値を指定する工程と、複数の探索対象化合物から、予測モデルを用いて目標値の結合力を有する化合物を抽出する工程と、を有する。

[0242] 付記 38に係るスクリーニング方法は付記 26に係るスクリーニング方法を改変したもので、結合化合物の構造式ではなく特徴量を入力する。このため、結合化合物については、付記 21に係る特徴量算出方法の各工程（対象構造体指定工程、立体構造発生工程、及び特徴量算出工程）は不要である。付記 38に係るスクリーニング方法によっても、付記 26と同様に複数の化合物の特徴量と結合力の値のセット（組）を用いて、目標の結合力を有する化合物を、既存の化合物群（複数の探索対象化合物）からスクリーニングすることができる。なお、付記 38においても、付記 18, 19に示す予測モデル生成方法及び予測モデルを用いることができる。

[0243] (付記 39)

付記 39に係る化合物創出方法は、付記 21に係る特徴量算出方法を用いた化合物創出方法であって、複数の結合化合物についての特徴量を入力する工程を有し、立体構造発生工程では創出する化合物の立体構造の生成方法を指定する。付記 39に係る化合物創出方法はさらに、複数の結合化合物のそ

れぞれについて特徴量及び結合力の値を指定する工程と、複数の結合化合物についての特徴量と結合力の値との組から化合物の結合力の予測モデルを生成する方法を指定する工程と、結合力の目標値を指定する工程と、特徴量に基づいて化合物の立体構造を発生させる際の条件を指定する工程と、生成方法及び条件の下で、結合力が目標値以上である化合物の立体構造を発生させる工程と、を有する。付記 39 に係る化合物創出方法は付記 27 に係る化合物創出方法を改変したもので、結合化合物の構造式ではなく特徴量を入力する。このため、結合化合物については、付記 21 に係る特徴量算出方法の各工程（対象構造体指定工程、立体構造発生工程、及び特徴量算出工程）は不要である。付記 39 に係る化合物創出方法によっても、付記 27 と同様に複数の化合物の特徴量と結合力の値のセット（組）を用いて、目標の結合力を有する化合物を、逆問題を解くことにより新たに創出することができる。なお、付記 39 においても、付記 18, 19 に示す予測モデル生成方法及び予測モデルを用いることができる。

[0244] （付記 40）

付記 40 に係る化合物情報処理方法は、付記 21 に係る特徴量算出方法を用いた化合物情報処理方法であって、対象構造体指定工程では対象構造体として化合物を指定し、立体構造発生工程では複数の原子による化合物の立体構造を発生させ、特徴量算出工程では、立体構造発生工程で発生させた化合物の立体構造の周辺における、プローブとしてのアミノ酸の集積度合いを 3 次元空間において定量化した特徴量である第 1 の特徴量を算出する。付記 40 に係る化合物情報処理方法はさらに、化合物に対するアミノ酸の集積度合いを、複数のタンパク質についての立体構造を示す情報が格納されたデータベースと照合する工程と、照合に基づいて複数のタンパク質から化合物が結合する標的タンパク質の候補を抽出する工程であって、化合物についてのアミノ酸の集積度合いとタンパク質のポケットを含む領域におけるアミノ酸の分布パターンとの一致度に基づいて標的タンパク質の候補を抽出する工程と、を有する。タンパク質のデータベースとしては例えば P D B（Protein Dat

a Bank) を用いることができるが、この例に限定されるものではない。

[0245] 創薬研究においては、「新規薬剤の候補となる化合物がどのタンパク質（標的タンパク質）に作用しているかを特定したい」というニーズが存在する。付記40に係る化合物情報処理方法はそのようなニーズに対応するもので、図65のフローチャートに示す処理により標的タンパク質の候補を抽出する。抽出した候補の名称や立体構造を示す情報を出力（表示装置に表示する、等）してもよい。このような処理は、例えば第1～第3の実施形態に係る装置（スクリーニング装置10、化合物創出装置20、医薬候補化合物探索装置30）あるいはこれらの装置と同様の構成を有する装置、システムにより行うことができる。なお、これらの装置やシステムに図65のフローチャートに示す処理を行わせるプログラム、及び斯かるプログラムのコンピュータ読み取り可能なコードを記録した非一時的記録媒体も本発明の態様として挙げることができる。

[0246] <付記22～40に関する出力例>

(1) 1つの化合物に対する情報の出力例

付記22から27に係るスクリーニング方法及び化合物創出方法において、化合物の構造式や他の記述子（特徴量）と併せて、各アミノ酸の集積度合いを図や数値として示すことができる。例えば、図66に示すように、化合物の名前、構造式、及び立体構造と対応させて、本発明に係る特徴量算出方法により算出した各アミノ酸の集積度合いを3次元的に示すことができる。また、図67に示すように、各アミノ酸の集積度合いを選択原子群（図67の右端列において丸印及び矢印で示した部分）の持ち分について表示することができる。

[0247] 集積度合いの表示は、図66、61に例示したように3次元のかつ写実的でもよいし、2次元のかつ／または簡易的でもよい。例えば、図68に示すように、対象とする化合物を中心とした仮想的な円（図68中、点線で表示）の円周上に、プローブ（アミノ酸等）の集積度合いを太さ及び色の異なる円弧により示すことができる。この態様において、例えば集積度合いが高

い範囲を太い赤線で表示し、集積度合いが低い範囲を細い青線で表示することができる。色は複色色用いてもよいし、単色の濃淡（例えば、白色、灰色、黒色等）で表示してもよい。また、集積度合いが非常に低い（あるいはゼロ）の範囲は線なしとしてもよい。なお、図68に示す円弧の太さ及び色と集積度合いとの関係は例示である。

[0248] 図68の態様において、集積度合いの全体における各アミノ酸の持ち分を示した例が図69である。図69の（a）部分が、あるアミノ酸の集積度合いを示し、同図の（b）部分は別のアミノ酸の集積度合いを示す。また、図70に示すように、集積度合いにおける選択部分（同図中、実線の丸印及び矢印で示した部分）の持ち分を表示してもよい。

[0249] 図71はプローブの集積度合いを2次元グラフで表示した様子を示す図である。同図の横軸は指定された方向を基準とした角度を示し、縦軸は集積度合いを示す。図71はプローブ全体（例えば20種類のアミノ酸）の集積度合いを示す図であり、これを各プローブ（例えば、個々のアミノ酸）に分解して示したのが図72である。

[0250] 図73は特徴量の疑似的3次元表示の例を示す図である。図73に示す例では、対象化合物を中心とする球面上でのプローブの集積度合いを、世界地図の作成において用いられるメルカトル図法により2次元に射影して示している。図73の（a）部分は球面上の領域を緯度、経度に対応する線で示し、同図の（b）部分は、図68～64と同様に、集積度合いの高低を色に対応させて示している。例えば、集積度合いが高い領域（点線で囲んだ領域等）を赤色で表示し、集積度合いが低い領域を青色で表示することができる。なお、図68～67における例は表示態様を概念的に示すものであり、実際の化合物に対する集積度合いを表すものではない。

[0251] 図66～67に例示する態様により、ユーザはプローブの集積度合いを視覚的に容易に把握することができる。これらの図における集積度合い等は、例えば第1～第3の態様に係る装置（スクリーニング装置10、化合物創出装置20、医薬候補化合物探索装置30）の表示制御部150が、ユーザの

操作に応じてモニタ 310（表示装置）に表示させることができる（変形例や付記においても同様である）。

[0252] （２）結合力維持でのスクリーニング、創出に関する情報の出力例

結合化合物や生体高分子の活性部位、及びスクリーニングや創出の対象となる化合物の差を見やすくするための出力態様として以下の例が考えられる。例えば、「各アミノ酸の集積度合いの差分を類似度として表現して提示する」、あるいは「各アミノ酸の集積度合いの差分を図や数値として提示する」、といった態様が考えられる。また、「各アミノ酸の集積度合いの差分をとり、一致している集積度合いの部分を検出し、原子ごとの持ち分の集積度合いと照合することで、一致している集積度合いの部分に対応する原子を特定し、その原子を構造式上に提示する」、といった態様も考えられる。また、逆問題を解くことによる立体構造の創出において、化合物の構造式の他、各アミノ酸の集積度合いが徐々に結合化合物や生体高分子の活性部位の集積度合いに近づいていく様子を提示することが考えられる。図 74 は、目標化合物（図中の“lig001”）に対し、スクリーニング化合物（“lig002”等）の各アミノ酸の集積度合いの差分を類似度として表現した例を示す図である。

[0253] （３）結合力向上のためのスクリーニング、創出に関する情報の出力例

複数の結合化合物と結合力の値のセットからの、結合力向上のためのスクリーニング、創出に関する情報の出力について説明する。例えば、図 75 に示すように、結合力の予測モデルの性能（実測と予測の対応、 R^2 等）を提示することが考えられる。また、結合力にとって重要だった要因を、各アミノ酸の集積度合いや構造式の部分として提示することが考えられる。また、化合物の結合力の予測値を出力することも考えられる。

[0254] （４）アミノ酸をグループ分けした出力

上述の（１）～（３）の態様により、ユーザはアミノ酸（プローブ）の集積度合いやこれに関連する情報を視覚により容易に把握することができるが、出力の際に、アミノ酸の全情報を出力するのではなく、各アミノ酸を親水

性か疎水性か、あるいは中性か塩基性か酸性か、といった素性ごとにグループ分けし、それぞれで平均化した結果を出力することが考えられる。

[0255] (付記 4 1)

付記 4 1 に係る化合物探索サーバは、第 1 の化合物の立体構造を示す第 1 の情報と、探索対象となる複数の化合物を特定する第 2 の情報と、をクライアント装置から受信する受信部と、第 1 の化合物について第 1 から第 4 の態様のいずれか 1 つに係る特徴量算出方法で算出した特徴量と、複数の化合物についての特徴量と、に基づいて、複数の化合物から、第 1 の化合物についての特徴量との類似度が指定された範囲である目標特徴量を有する標的化合物を抽出する抽出部と、標的化合物の立体構造を示す第 3 の情報をクライアント装置に送信する送信部と、を備える。

[0256] (付記 4 2)

付記 4 2 に係るクライアント装置は、第 1 の情報及び第 2 の情報を付記 4 1 に係る化合物探索サーバに送信する送信部と、第 3 の情報を化合物探索サーバから受信する受信部と、第 3 の情報を出力する出力部と、を備える。また、第 1 の情報及び第 2 の情報を付記 4 1 に係る化合物探索サーバに送信する送信機能と、第 3 の情報を化合物探索サーバから受信する受信機能と、第 3 の情報を出力する出力機能と、をコンピュータに実行させるプログラムや、そのプログラムのコンピュータ読み取り可能なコードが記録された非一時的記録媒体も、付記 4 2 の変形例として挙げることができる。

[0257] (付記 4 3)

付記 4 3 に係る化合物探索システムは、付記 4 1 に係る化合物探索サーバと、付記 4 2 に係るクライアント装置と、を含む。

[0258] <付記 4 1 ～ 4 3 に係る化合物探索システムの構成例>

図 7 6 は付記 4 1 ～ 4 3 に係る化合物探索システム 7 0 0 (化合物探索システム) の構成を示す図である。化合物探索システム 7 0 0 は、化合物探索サーバ 7 1 0 (化合物探索サーバ) 及びクライアント装置 7 3 0 (クライアント装置) を含み、これらがネットワーク NW で接続されて構成されている。

。化合物探索サーバ710は、タンパク質やその他化合物の立体構造、特性（例えば特徴量、結合力、物性値、特性値）等が格納されたデータベース720を有する。化合物探索サーバ710は、PDB等の既存のデータベースにアクセスしてもよい。また、化合物探索サーバ710は、図77の機能ブロック図に示すように、受信部712（受信部）と、抽出部714（抽出部）と、送信部716（送信部）と、を備える。一方、図76、72に示すように、クライアント装置730の処理部740は送信部742（送信部）と、受信部744（受信部）と、出力部（746）と、を備える。これらの機能は、第1の実施形態について上述したようにCPU等のプロセッサを1つ以上用いて実現することができ、そのプロセッサが付記41に係る処理を実行する。クライアント装置730や、後述する他の付記についても同様である。

[0259] 化合物探索システム700はこれらの構成要素を一事業者が所有及び／または使用してもよい。また、例えば化合物探索サービスの提供者が化合物探索サーバ710及びデータベース720を所有し、ユーザがクライアント装置730を所有及び／または使用してもよい。この場合、クライアント装置730のユーザは、操作部760を操作すること等により1つ以上の化合物の構造式（「第1の情報」の一態様；3次元画像でもよい）及び探索対象となる複数の化合物を特定する情報（「第2の情報」の一態様）を指定し、クライアント装置730は処理部740の送信部742を介してこれらの情報を化合物探索サーバ710に送信する。ユーザは「探索対象となる複数の化合物を特定する情報」として各化合物の構造式や立体構造を指定してもよいし、化合物の立体構造等が格納されたデータベースの名称（例えば、PDB（Protein Data Bank）等）やネットワーク上の所在（URL（Uniform Resource Locator）等）を指定してもよい。

[0260] クライアント装置730から情報（第1の情報及び第2の情報）を受信した化合物探索サーバ710は、付記41について上述した処理により標的化合物を抽出し、標的化合物の立体構造を示す情報（構造式や3次元画像；第

3の情報)をクライアント装置730に送信する。クライアント装置730は、化合物探索サーバ710から受信した情報(第3の情報)を出力(モニタ772に表示、等)する。これによりクライアント装置730のユーザは、自らが指定した「探索対象となる複数の化合物」から抽出された、「1つ以上の化合物」と同じ結合力を持つ別の化合物の構造式や、その結合力の値を受け取ることができる。化合物探索サーバ710の所有者等のサービス提供者が、構造式等を提供する化合物の現物(実際の化合物)を合成してユーザに提供してもよい。

[0261] (付記44)

付記44に係る化合物創出サーバは、第1の化合物の立体構造を示す第1の情報をクライアント装置から受信する受信部と、第1の情報に基づいて、第1の化合物について第1から第4の態様のいずれか1つに係る特徴量算出方法を用いて特徴量を算出する算出部と、特徴量と付記15から17, 23, 25, 27, 29, 31, 33, 35, 37, 39のうちいずれか1つに係る化合物創出方法とに基づいて、目標特徴量を有する第2の化合物であって、目標特徴量が特徴量に対し指定された範囲である第2の化合物の立体構造を示す第2の情報を生成する生成部と、第2の情報をクライアント装置に送信する送信部と、を備える。

[0262] (付記45)

付記45に係るクライアント装置は、第1の化合物の立体構造を示す第1の情報を付記44に係る化合物創出サーバに送信する送信部と、化合物創出サーバから第2の情報を受信する受信部と、第2の情報を出力する出力部と、を備える。

[0263] (付記46)

付記46に係る化合物創出システムは、付記44に係る化合物創出サーバと、付記45に係るクライアント装置と、を含む。

[0264] <付記44～46に係る化合物創出システムの構成例>

付記44～46に係る化合物創出システムは付記41～43に係る化合物

探索システム 700（化合物探索システム）と同様の構成（サーバとクライアント装置がネットワークを介して接続された構成；図 76 を参照）を採用することができる。ここで、図 79 の機能ブロック図に示すように、付記 44 に係る化合物創出サーバ 710A（化合物創出サーバ）は、受信部 712（受信部）と、送信部 716（送信部）と、算出部 717（算出部）と、生成部 718（生成部）と、を有する。化合物創出システムにおけるクライアント装置は、図 76 に示すクライアント装置 730 と同様の構成を用いることができる。

[0265] このような構成の下で、クライアント装置 730 のユーザは操作部 760 を操作すること等により 1 つ以上の化合物の構造式（「第 1 の情報」の一態様；3 次元画像でもよい）を指定し、クライアント装置 730 は処理部 740 の送信部 742 を介してこの情報を化合物創出サーバ 710A に送信する。クライアント装置 730 は目標特微量及び特微量の範囲を合わせて送信してもよい。クライアント装置 730 から情報（第 1 の情報）を受信した化合物創出サーバ 710A は、付記 44 について上述した処理により特微量が目標特微量に対し指定された範囲である化合物（第 2 の化合物）の立体構造を示す情報（構造式や 3 次元画像；第 2 の情報）を生成（化合物を創出）し、その情報をクライアント装置 730 に送信する。化合物創出サーバ 710A は第 2 の化合物について結合力を算出し、合わせて送信してもよい。クライアント装置 730 は、化合物創出サーバ 710A から受信した情報（第 2 の情報）を出力（モニタ 772 に表示、等）する。これによりクライアント装置 730 のユーザは、自らが指定した「1 つ以上の化合物」と同じ結合力を持つ（特微量が目標特微量に対し指定された範囲である）既存の、あるいは新規な化合物の構造式や、その結合力の値を受け取ることができる。化合物創出サーバ 710A の所有者等のサービス提供者が、構造式等を提供する化合物の現物（実際の化合物）を合成してユーザに提供してもよい。

[0266] （付記 47）

付記 47 に係る化合物探索サーバは、複数の化合物についての立体構造を

示す第1の情報及び結合力を示す第2の情報と、複数の探索対象化合物の立体構造を特定する第3の情報と、をクライアント装置から受信する受信部と、複数の化合物についての第1の情報及び第2の情報から、第1から第4の態様のいずれか1つに係る特徴量算出方法に係る特徴量に基づいて化合物の結合力を出力する予測モデルを生成する生成部と、複数の探索対象化合物のそれぞれについて、第3の情報に基づいて、第1から第4の態様のいずれか1つに係る特徴量算出方法により特徴量を算出する算出部と、複数の探索対象化合物のそれぞれについて、算出した特徴量を予測モデルに適用して結合力を推定する推定部と、複数の探索対象化合物から、複数の化合物についての結合力より高い結合力を有する化合物を抽出する抽出部と、抽出した化合物の立体構造を示す情報及び抽出した化合物の結合力を示す情報をクライアント装置に送信する送信部と、を有する。

[0267] (付記48)

付記48に係るクライアント装置は、複数の化合物についての立体構造を示す第1の情報及び結合力を示す第2の情報と、複数の探索対象化合物の立体構造を特定する第3の情報と、を付記47に係る化合物探索サーバに送信する送信部と、化合物探索サーバが抽出した、化合物の立体構造を示す情報及び抽出した化合物の結合力を示す情報を化合物探索サーバから受信する受信部と、化合物の立体構造を示す情報及び抽出した化合物の結合力を示す情報を出力する出力部と、を備える。

[0268] (付記49)

付記49に係る化合物探索システムは、付記47に係る化合物探索サーバと、付記48に係るクライアント装置と、を含む。

[0269] <付記47～49に係る化合物探索システムの構成例>

付記47～49に係る化合物探索システムは付記41～43に係る化合物探索システム700（化合物探索システム）と同様の構成（サーバとクライアント装置がネットワークを介して接続された構成；図76を参照）を採用することができる。ここで、図80の機能ブロック図に示すように、付記4

7に係る化合物探索サーバ710B（化合物探索サーバ）は、受信部712（受信部）と、抽出部714（抽出部）と、送信部716（送信部）と、算出部717（算出部）と、生成部718（生成部）と、推定部719（推定部）と、を有する。化合物探索システムにおけるクライアント装置は、図76に示すクライアント装置730と同様の構成を用いることができる。

[0270] このような構成の下で、クライアント装置730のユーザは操作部760を操作すること等により複数の化合物についての立体構造を示す第1の情報（構造式、3次元画像等）及び結合力を示す第2の情報と、複数の探索対象化合物の立体構造を特定する第3の情報と、を指定し、クライアント装置730は処理部740の送信部742を介してこの情報を化合物探索サーバ710Bに送信する。ユーザは「複数の化合物の立体構造を示す情報」として各化合物の構造式や立体構造を指定してもよいし、化合物の立体構造等が格納されたデータベースの名称（例えば、PDB等）やネットワーク上の所在（URL等）を指定してもよい。クライアント装置730から情報（第1～第3の情報）を受信した化合物探索サーバ710Bは、付記47について上述した処理により化合物の抽出を行い、その化合物の立体構造を示す情報及び結合力を示す情報をクライアント装置730に送信する。クライアント装置730は、化合物探索サーバ710Bから受信した情報を出力（モニター772に表示、等）する。これによりクライアント装置730のユーザは、自らが指定した化合物より高い結合力を有する新規な化合物の構造式や、その結合力の値を受け取ることができる。化合物探索サーバ710の所有者等のサービス提供者が、構造式等を提供する化合物の現物（実際の化合物）を合成してユーザに提供してもよい。

[0271] （付記50）

付記50に係る化合物創出サーバは、複数の化合物についての立体構造を示す第1の情報及び結合力を示す第2の情報をクライアント装置から受信する受信部と、複数の化合物についての第1の情報及び第2の情報から、第1から第4の態様のいずれか1つに係る特徴量算出方法に係る特徴量に基づい

て化合物の結合力を出力する予測モデルを生成する生成部と、予測モデルと付記 15 から 17, 23, 25, 27, 29, 31, 33, 35, 37, 39 のうちいずれか 1 つに係る化合物創出方法とに基づいて、複数の化合物についての結合力より高い結合力を有する化合物の立体構造を示す情報を生成する生成部と、生成した情報及び生成した情報に対応する化合物の結合力を示す情報をクライアント装置に送信する送信部と、を有する。

[0272] (付記 5 1)

付記 5 1 に係るクライアント装置は、複数の化合物についての立体構造を示す第 1 の情報及び結合力を示す第 2 の情報を付記 5 0 に係る化合物創出サーバに送信する送信部と、化合物創出サーバが生成した化合物の立体構造を示す情報及び生成した情報に対応する化合物の結合力を示す情報を化合物創出サーバから受信する受信部と、化合物創出サーバから受信した立体構造を示す情報及び結合力を示す情報を出力する出力部と、を備える。

[0273] (付記 5 2)

付記 5 2 に係る化合物創出システムは、付記 5 0 に係る化合物創出サーバと、付記 5 1 に係るクライアント装置と、を含む。

[0274] <付記 5 0 ～ 5 2 に係る化合物創出システムの構成例>

上述した構成により、付記 5 0 ～ 5 2 に係る化合物創出システムのクライアント装置のユーザは、自らが指定した化合物より高い結合力を有する既存あるいは新規な化合物の構造式や、その結合力の値を受け取ることができる。化合物探索サーバ 710 の所有者等のサービス提供者が、構造式等を提供する化合物の現物（実際の化合物）を合成してユーザに提供してもよい。

符号の説明

- [0275] 10 スクリーニング装置
20 化合物創出装置
30 医薬候補化合物探索装置
100 処理部
101 処理部

1 0 2 処理部
1 0 3 処理部
1 0 4 処理部
1 0 5 処理部
1 1 0 情報入力部
1 2 0 特徴量算出部
1 2 2 特徴量算出部
1 3 0 類似度算出部
1 3 2 生成器構築部
1 3 3 結合力予測部
1 3 4 特性値算出部
1 3 6 構造特徴量算出部
1 4 0 化合物抽出部
1 4 2 化合物立体構造生成部
1 5 0 表示制御部
1 6 0 C P U
1 7 0 R O M
1 8 0 R A M
2 0 0 記憶部
2 0 1 記憶部
2 0 2 記憶部
2 1 0 構造情報
2 2 0 立体構造情報
2 3 0 3次元A A M記述子
2 4 0 不変量化A A M記述子
2 5 0 類似度情報
2 6 0 化合物抽出結果
2 7 0 立体構造生成結果

- 3 0 0 表示部
- 3 1 0 モニタ
- 4 0 0 操作部
- 4 1 0 キーボード
- 4 2 0 マウス
- 5 0 0 外部サーバ
- 5 1 0 外部データベース
- 6 0 0 階層型ネットワーク
- 6 1 0 入力層
- 6 2 0 中間層
- 6 2 2 第 1 の中間層
- 6 2 4 第 2 の中間層
- 6 2 5 層
- 6 2 7 層
- 6 3 0 出力層
- 7 0 0 化合物探索システム
- 7 1 0 化合物探索サーバ
- 7 1 0 A 化合物創出サーバ
- 7 1 0 B 化合物探索サーバ
- 7 1 2 受信部
- 7 1 4 抽出部
- 7 1 6 送信部
- 7 1 7 算出部
- 7 1 8 生成部
- 7 1 9 推定部
- 7 2 0 データベース
- 7 3 0 クライアント装置
- 7 4 0 処理部

7 4 2 送信部
7 4 4 受信部
7 5 0 記憶部
7 6 0 操作部
7 6 2 キーボード
7 6 4 マウス
7 7 0 表示部
7 7 2 モニタ
A 1 アミノ酸
A 2 アミノ酸
A 3 アミノ酸
A A 2 A R タンパク質
A B L 1 タンパク質
C 0 化合物
C 1 化合物
C 2 化合物
C 3 化合物
C 4 化合物
C 5 化合物
C 6 化合物
N W ネットワーク
P O ポケット
P S ポケット構造体
S 1 0 0 ~ S 1 0 8 特徴量算出方法の各ステップ
S 2 0 0 ~ S 2 0 6 特徴量算出方法の各ステップ
S 3 0 0 ~ S 3 0 4 標的化合物抽出方法の各ステップ
S 3 1 0 ~ S 3 2 6 スクリーニング方法の各ステップ
S 4 0 0 ~ S 4 0 4 標的化合物抽出方法の各ステップ

S 5 0 0 ～ S 5 0 4 立体構造創出方法の各ステップ

S 6 0 0 ～ S 6 0 4 立体構造創出方法の各ステップ

T P 標的タンパク質

請求の範囲

- [請求項1] 化学的性質を有する複数の単位構造体から構成される対象構造体を指定する対象構造体指定工程と、
- 前記対象構造体の立体構造の周辺における1種類以上のプローブの集積度合いを3次元空間において定量化した特徴量を算出する特徴量算出工程であって、機械学習により構成された生成器を用いて前記対象構造体から前記特徴量を算出する特徴量算出工程と、
- を有し、
- 前記プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である特徴量算出方法。
- [請求項2] 前記対象構造体指定工程では前記対象構造体として化合物を指定し、
- 前記特徴量算出工程では、前記生成器を用いて、前記化合物の前記立体構造の周辺における、前記プローブとしてのアミノ酸の集積度合いを前記3次元空間において定量化した特徴量である第1の特徴量を算出する請求項1に記載の特徴量算出方法。
- [請求項3] 前記対象構造体指定工程では、標的タンパク質の活性部位であるポケットに結合するポケット構造体を前記対象構造体として指定し、
- 前記特徴量算出工程では、前記生成器を用いて、複数の仮想的球体による前記ポケット構造体の立体構造の周辺における、前記プローブとしてのアミノ酸の集積度合いを前記3次元空間において定量化した特徴量である第2の特徴量を算出する請求項1に記載の特徴量算出方法。
- [請求項4] 前記対象構造体指定工程では前記対象構造体として化合物を指定し、
- 前記特徴量算出工程では、前記生成器を用いて、前記化合物の前記立体構造の周辺における前記プローブの集積度合いであって、1種類以上の核酸塩基、1種類以上の脂質分子、1種類以上の単糖分子、水

、複数の原子で構成される１種類以上のイオンのうち１つ以上を前記プローブとした集積度合いを前記３次元空間において定量化した特徴量である第３の特徴量を算出する請求項１に記載の特徴量算出方法。

[請求項５] 請求項１から４のいずれか１項に記載の特徴量算出方法をコンピュータに実行させる特徴量算出プログラム。

[請求項６] 非一時的かつコンピュータ読取可能な記録媒体であって、前記記録媒体に格納された指令がコンピュータによって読み取られた場合に請求項５に記載の特徴量算出プログラムをコンピュータに実行させる記録媒体。

[請求項７] 化学的性質を有する複数の単位構造体から構成される対象構造体を指定する対象構造体指定部と、

前記対象構造体の立体構造の周辺における１種類以上のプローブの集積度合いを３次元空間において定量化した特徴量を算出する特徴量算出部であって、機械学習により構成された生成器を用いて前記対象構造体から前記特徴量を算出する特徴量算出部と、

を備え、

前記プローブは、実数電荷を持ちファンデルワールス力を発生させる複数の点が離間して配置された構造体である特徴量算出装置。

[請求項８] 複数の化合物から標的タンパク質に結合する標的化合物を抽出するスクリーニング方法であって、

前記複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、前記化合物の前記立体構造について請求項２に記載の特徴量算出方法を用いて算出した前記第１の特徴量と、を関連付けて記憶する記憶工程と、

前記標的タンパク質との結合が確認されている化合物であるリガンドについて請求項２に記載の特徴量算出方法を用いて前記第１の特徴量を算出する特徴量算出工程と、

前記複数の化合物についての前記第１の特徴量と前記リガンドにつ

いての前記第1の特徴量との類似度を算出する類似度算出工程と、
前記複数の化合物から前記類似度に基づいて前記標的化合物を抽出する化合物抽出工程と、
を有するスクリーニング方法。

[請求項9]

複数の化合物から標的タンパク質に結合する標的化合物を抽出するスクリーニング方法であって、

前記複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、請求項2に記載の特徴量算出方法を用いて算出した前記第1の特徴量と、を関連付けて記憶する記憶工程と、

前記標的タンパク質の前記ポケット構造体について、請求項3に記載の特徴量算出方法を用いて前記第2の特徴量を算出する特徴量算出工程と、

前記複数の化合物についての前記第1の特徴量と前記ポケット構造体についての前記第2の特徴量との類似度を算出する類似度算出工程と、

前記複数の化合物から前記類似度に基づいて前記標的化合物を抽出する化合物抽出工程と、

を有するスクリーニング方法。

[請求項10]

複数の化合物からタンパク質以外の標的生体高分子に結合する標的化合物を抽出するスクリーニング方法であって、

前記複数の化合物のそれぞれについて、複数の原子による化合物の立体構造と、前記化合物の前記立体構造について請求項4に記載の特徴量算出方法を用いて算出した前記第3の特徴量と、を関連付けて記憶する記憶工程と、

前記タンパク質以外の前記標的生体高分子に結合することが確認されている化合物である結合化合物について請求項4に記載の特徴量算出方法を用いて前記第3の特徴量を算出する特徴量算出工程と、

前記複数の化合物についての前記第3の特徴量と前記結合化合物に

ついでの前記第3の特徴量との類似度を算出する類似度算出工程と、
前記複数の化合物から前記類似度に基づいて前記標的化合物を抽出する化合物抽出工程と、
を有するスクリーニング方法。

[請求項11] 複数の化合物について立体構造を示す情報を入力する入力工程と、
前記複数の化合物のそれぞれについて、前記情報に基づいて請求項1から4のいずれか1項に記載の特徴量算出方法により前記特徴量を算出する特徴量算出工程と、
前記複数の化合物のそれぞれについて、前記情報に基づいて、物性を示す物性値及び／または毒性を示す毒性値を算出する特性値算出工程と、
前記複数の化合物から、前記特徴量の類似度が目標値を満たし、かつ前記物性値及び／または前記毒性値が目標値を満たす化合物を抽出する抽出工程と、

を有するスクリーニング方法。

[請求項12] 複数の化合物について立体構造を示す情報を入力する入力工程と、
前記複数の化合物のそれぞれについて、前記情報に基づいて請求項1から4のいずれか1項に記載の特徴量算出方法により前記特徴量を算出する第1の特徴量算出工程と、
前記複数の化合物のそれぞれについて、前記情報に基づいて化合物の構造を示す構造特徴量を算出する第2の特徴量算出工程と、
前記複数の化合物のうちから、前記特徴量の類似度が目標値を満たし、かつ前記構造特徴量の類似度がしきい値以下である化合物を抽出する抽出工程と、

を有するスクリーニング方法。

[請求項13] 複数の化合物について立体構造を示す情報を入力する入力工程と、
前記複数の化合物のそれぞれについて、前記情報に基づいて請求項1から4のいずれか1項に記載の特徴量算出方法により前記特徴量を

算出する特徴量算出工程と、

前記複数の化合物のそれぞれについて、前記特徴量に基づいて他の化合物との結合力を出力する予測モデルを用いて前記結合力を予測する予測工程と、

前記複数の化合物のうちから、前記結合力が目標値を満たす化合物を抽出する抽出工程と、

を有するスクリーニング方法。

[請求項14]

複数の化合物について立体構造を示す情報を入力する入力工程と、他の化合物との結合力の目標値を設定する設定工程と、

請求項1から4のいずれか1項に記載の特徴量算出方法により算出された特徴量に基づいて前記結合力を出力する予測モデルを用いて、前記目標値に対応する前記特徴量を推定する推定工程と、

前記複数の化合物のそれぞれについて、前記情報に基づいて請求項1から4のいずれか1項に記載の特徴量算出方法により前記特徴量を算出する特徴量算出工程と、

前記算出の結果に基づいて、前記複数の化合物のうちから、前記推定した前記特徴量との類似度がしきい値以上である前記特徴量を有する化合物を抽出する抽出工程と、

を有するスクリーニング方法。

[請求項15]

請求項8から14のいずれか1項に記載のスクリーニング方法をコンピュータに実行させるスクリーニングプログラム。

[請求項16]

非一時的かつコンピュータ読取可能な記録媒体であって、前記記録媒体に格納された指令がコンピュータによって読み取られた場合に請求項15に記載のスクリーニングプログラムをコンピュータに実行させる記録媒体。

[請求項17]

複数の化合物から標的タンパク質に結合する標的化合物の立体構造を創出する化合物創出方法であって、

複数の化合物のそれぞれについて、複数の原子による化合物の立体

構造と、請求項2に記載の特徴量算出方法を用いて算出した前記第1の特徴量と、を関連付けて記憶する記憶工程と、

前記標的タンパク質との結合が確認されている化合物であるリガンドについて、請求項2に記載の特徴量算出方法を用いて前記第1の特徴量を算出する特徴量算出工程と、

前記複数の化合物の前記立体構造を教師データとし、前記第1の特徴量を説明変数とした機械学習により生成器を構築する生成器構築工程と、

前記生成器を用いて、前記リガンドの前記第1の特徴量から前記標的化合物の立体構造を生成する化合物立体構造生成工程と、

を有する化合物創出方法。

[請求項18]

複数の化合物から標的タンパク質に結合する標的化合物の立体構造を創出する化合物創出方法であって、

前記複数の化合物のそれぞれについて、請求項2に記載の特徴量算出方法を用いて算出した前記第1の特徴量と、を関連付けて記憶する記憶工程と、

前記標的タンパク質の前記ポケット構造体について、請求項3に記載の特徴量算出方法を用いて前記第2の特徴量を算出する特徴量算出工程と、

前記複数の化合物の立体構造を教師データとし、前記第1の特徴量を説明変数とした機械学習により生成器を構築する生成器構築工程と、

前記生成器を用いて、前記ポケット構造体の前記第2の特徴量から前記標的化合物の立体構造を生成する化合物立体構造生成工程と、

を有する化合物創出方法。

[請求項19]

複数の化合物からタンパク質以外の標的生体高分子に結合する標的化合物の立体構造を創出する化合物創出方法であって、

複数の化合物のそれぞれについて、複数の原子による化合物の立体

構造と、前記第3の特徴量と、を関連付けて記憶する記憶工程と、

前記タンパク質以外の前記標的生体高分子との結合が確認されている化合物である結合化合物について、請求項4に記載の特徴量算出方法を用いて前記第3の特徴量を算出する特徴量算出工程と、

前記複数の化合物の前記立体構造を教師データとし、前記第3の特徴量を説明変数とした機械学習により生成器を構築する生成器構築工程と、

前記生成器を用いて、前記結合化合物の前記第3の特徴量から前記標的化合物の立体構造を生成する化合物立体構造生成工程と、

を有する化合物創出方法。