

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2011年8月25日(25.08.2011)

PCT

(10) 国際公開番号
WO 2011/102384 A1

- (51) 国際特許分類:
C40B 30/02 (2006.01) G06F 17/50 (2006.01)
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2011/053280
- (22) 国際出願日: 2011年2月16日(16.02.2011)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2010-031526 2010年2月16日(16.02.2010) JP
- (71) 出願人 (米国を除く全ての指定国について): 学校法人関西文理総合学園(EDUCATIONAL CORPORATION KANSAI BUNRI SOUGOUAKUEN) [JP/JP]; 〒5260829 滋賀県長浜市田村町1266番地 Shiga (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 白井 剛 (SHIRAI, Tsuyoshi) [JP/JP]; 〒5260829 滋賀県長浜市田村町1266番地 学校法人関西文理総合

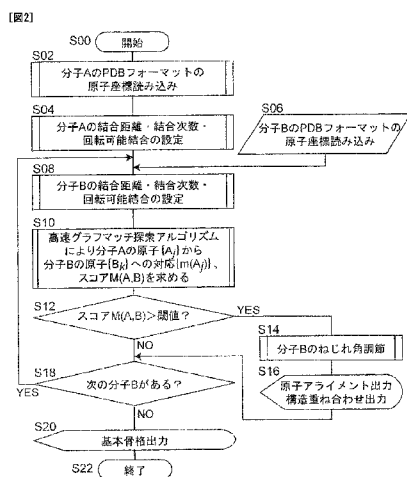
合学園 長浜バイオ大学 バイオサイエンス
学部内 Shiga (JP).

- (74) 代理人: 田中 光雄, 外(TANAKA, Mitsuo et al.); 〒5400001 大阪府大阪市中央区域見1丁目3番7号IMPビル青山特許事務所 Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT,

[続葉有]

(54) Title: DEVICE AND METHOD FOR HIGH SPEED GRAPH MATCHING SEARCH TO DETERMINE SIMILARITY BETWEEN MOLECULES

(54) 発明の名称: 分子間の類似度を評価するための高速グラフマッチ検索装置及び方法



(57) Abstract: Disclosed is a high speed molecular graph matching algorithm for use with a method for determining the matching between atoms of two molecules and superimposing the molecules accordingly. Specifically disclosed is a method for determining the optimum atomic matching between a first molecule A and a second molecule B by determining the degree of matching between each atom of the first molecule A and each atom of the second molecule B (represented by the expression $m(A_i) = B_k$) and superimposing the molecules, said method comprising: a step of calculating $S1(A_i, B_k)$, $S2(A_i, B_k)$, and $S3(A_i, B_k)$ for all sets $\{i, k\}$; a step of determining the overall graph matching score $M(A, B)$ by first associating the pair of atoms (A_i, B_k) having the largest calculated value for $S3(A_i, B_k)$ with that having the second largest value for $S3(A_j, B_l)$, then associating the pair of atoms having the second largest value for $S3(A_j, B_l)$ with that having the third largest value, and repeating this procedure until all pairs of atoms have been associated with one another; and a step of, if the determined $M(A, B)$ exceeds a threshold value, outputting superimposition of the molecules A and B.

(57) 要約:

[続葉有]

- S00 START
- S02 READ ATOMIC COORDINATES OF MOLECULE A IN PDB FORMAT
- S04 DETERMINE BOND DISTANCES, BOND ORDERS, AND ROTATABLE BONDS FOR MOLECULE A
- S06 READ ATOMIC COORDINATES OF MOLECULE B IN PDB FORMAT
- S08 DETERMINE BOND DISTANCES, BOND ORDERS, AND ROTATABLE BONDS FOR MOLECULE B
- S10 DETERMINE DEGREE OF MATCHING ($m(A_i)$) BETWEEN ATOM $\{A_j\}$ OF MOLECULE A AND ATOM $\{B_k\}$ OF MOLECULE B AND SCORE $M(A, B)$ IN ACCORDANCE WITH HIGH SPEED GRAPH MATCHING SEARCH ALGORITHM
- S12 SCORE $M(A, B) >$ THRESHOLD VALUE?
- S14 ADJUST ANGLE OF TWIST IN MOLECULE B
- S16 OUTPUT ATOMIC ALIGNMENT DATA, AND OUTPUT STRUCTURAL SUPERIMPOSITION
- S18 NEXT MOLECULE B?
- S20 OUTPUT BASIC STRUCTURE DATA
- S22 END



WO 2011/102384 A1



NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI 添付公開書類:
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, — 国際調査報告 (条約第 21 条(3))
NE, SN, TD, TG).

分子グラフに関して、2分子間の原子対応を求め該対応に基づいて2分子を重ね合わせする方法を実現する高速グラフマッチアルゴリズムを提供する。第1の分子Aの夫々の原子と第2の分子Bの夫々の原子を対応付け ($m(A_i) = B_k$) で重ね合わせ、ABの間の最適な原子間対応を求める方法であつて、 $S_1(A_i, B_k)$ 、 $S_2(A_i, B_k)$ 及び $S_3(A_i, B_k)$ の各々を全ての $\{i, k\}$ の組について求める工程と、その工程にて最大の $S_3(A_i, B_k)$ を算出した原子 (A_i, B_k) の対応から開始して、未対応の原子の対の中で最大の $S_3(A_j, B_l)$ を持つものを対応させることを、対応可能原子の組が無くなるまで続けたときの全体の対応におけるグラフマッチスコア $M(A, B)$ を求める工程と、その工程における $M(A, B)$ が閾値より大きいならば、AとBにつき重ね合わせを出力する工程を含む方法を提供する。

明 細 書

発明の名称：

分子間の類似度を評価するための高速グラフマッチ検索装置及び方法

技術分野

[0001] 本発明は、高速グラフマッチ検索アルゴリズムを利用して、2分子間の原子対応を求め対応に基づいて2分子を仮想的に重ね合わせ、2分子間の類似度を求めて評価する、高速グラフマッチ検索装置及び方法に関する。

背景技術

[0002] 医薬や農薬の分子設計において、2つの分子に係る分子構造を仮想空間にて重ね合わせることが頻繁に行われる。図13は、そのような、2つの分子（Cholic acid [CHD] と Corticosteron [COR]）を仮想空間にて重ね合わせすることを模式的に示す図である。しかしながら、2つの分子についての最適な重ね合わせを探索し決定することは非常に困難な問題である。

[0003] 例えば、分子Aと分子Bとの重ね合わせの問題について、片方の分子Aを『CMP』とした場合に、それに基づき重ね合わせにて探索可能な重ね合わせの対象の分子Bを求める場合を検討する。ここで「探索可能な」というのは、全探査を8時間労働・週休2日の労働時間で50年程度行って、探索が解決され得ると想定される、という程の意味である。例えば、人手による計算による場合では、分子BがCysteineである場合、 1.3×10^7 通り程度の重ね合わせの計算を行い、最適な重ね合わせを求めることが可能となる（図14（a））。同様に、デスクトップコンピュータによる場合では、分子BがDiaminopimelateである場合、 1.5×10^{15} 通り程度の重ね合わせの計算を行い、最適な重ね合わせを求めることが可能となる（図14（b））。更に同様に、超高速電子計算機による場合でも、分子BがAMPである場合、 8.3×10^{21} 通りの重ね合わせの計算を行い、最適な重ね合わせを求めることが可能となる（図14（c））。このように

、分子Aと分子Bの最適な重ね合わせを全探索に拠ることは、膨大な時間が掛かるため、必ずしも現実的な方法ではない。

[0004] よって、2分子間の原子対応を求め該対応に基づいて2分子の最適な重ね合わせを実現するグラフマッチにおいて、多少の間違いを許容しつつも発見的に高速に行うことが求められている。

[0005] なお、化合物検索のアルゴリズムに関する先行技術文献として、以下のよ
うな6件が挙げられる。

先行技術文献

特許文献

- [0006] 特許文献1：特許第4001657号
特許文献2：特許第3928000号
特許文献3：国際出願01/097094号
特許文献4：国際出願02/41184号
特許文献5：国際出願2007/004643号

非特許文献

- [0007] 非特許文献1：J. Computer-Aided Molecular Design, 13:499-512, 1999 Esti
mation of active conformations of drugs by a new molecular superposin
g procedure

発明の概要

発明が解決しようとする課題

- [0008] 本発明は、原子をノード、化学結合をエッジとして表現した分子グラフに
関して、2分子間の原子対応を求め該対応に基づいて2分子を重ね合わせす
る方法を高速に実現する、グラフマッチ検索装置及び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

- [0009] 本発明は、上記の目的を達成するために為されたものである。本発明に係
る請求項1に記載の、分子間の類似度を評価するための高速グラフマッチ検

索装置は、

第1の分子Aを構成する原子 (A_i, A_j, \dots) の各々に係る座標データと第2の分子Bを構成する原子 (B_k, B_l, \dots) の各々に係る座標データを記憶部から入力し、演算部にロードされるコンピュータプログラムに従って、演算部及び記憶部に構築される仮想メモリ空間において第1の分子Aの夫々の原子 (A_i, A_j, \dots) と第2の分子Bの夫々の原子 (B_k, B_l, \dots) との対応付け ($m(A_i) = B_k$) を求めて重ね合わせを行い (i, j, k, l はいずれも自然数)、第1の分子Aと第2の分子Bの間の最適な原子間対応、及び第1の分子Aと第2の分子Bの類似度に係るデータを出力部に出力する、第1の分子Aと第2の分子Bとの類似度を評価するための高速グラフマッチ検索装置において、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i, B_k の対の各原子からみて、周囲の環境が相互にどれだけ似ているかを示す第1の類似指標 $S_1(A_i, B_k)$ を求める第1の算出手段と、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i, B_k の対の各原子からみて、等しい結合距離にある周囲の原子 A_j, B_l の全ての組につき、第1の類似指標 $S_1(A_j, B_l)$ を積算して算出する第2の類似指標 $S_2(A_i, B_k)$ を求める算出手段であって、その原子 A_i, B_k の対の各原子から等しい結合距離にある周囲の原子 A_j, B_l が同じ元素であれば、更に第1の類似指標 $S_1(A_j, B_l)$ に係数を掛けた上で積算する、第2の類似指標 $S_2(A_i, B_k)$ を求める第2の算出手段と、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i, B_k の対を始点とし、第1の分子Aの原子と第2の分子Bの原子とを順次対応付けして全体の対応を作成し、そのときに算出されるグラフマッチスコア $M(A, B)$ を値とする第3の類似指標 $S_3(A_i, B_k)$ を求める算出手段であって、

対応付け作成時には、既に対応付け済みの原子に直接結合する原子を次に選択すること、及び第2の類似指標 S_2 が高い対を選択することを優先することを、条件とする、第3の類似指標 $S_3(A_i, B_k)$ を求める第3の算出手段と、

第3の算出手段にて最大の $S_3(A_i, B_k)$ を算出した際の、始点の原子 (A_i, B_k) の対から開始して、未対応の原子の対の中で最大の $S_3(A_j, B_l)$ を持つものを対応させることを、対応可能原子の組が無くなるまで続けたときの、全体の対応におけるグラフマッチスコア $M(A, B)$ を求める第4の算出手段と、

第4の算出手段におけるグラフマッチスコア $M(A, B)$ が閾値より大きいならば、第1の分子Aと第2の分子Bにつき第4の算出手段で算出した原子間対応及びグラフマッチスコア $M(A, B)$ を出力する第5の出力手段とを含むことを特徴とする。

発明の効果

[0010] 本発明により、原子をノード、化学結合をエッジとして表現した分子グラフに関して、2分子間の原子を対応させ該対応に基づいて2分子を重ね合わせするにあたり、最適な重ね合わせを高速に且つ精度よく求めることができる。

図面の簡単な説明

[0011] [図1]本発明の実施形態に係るグラフマッチによる分子構造の高速アルゴリズムを実現するコンピュータシステムの構成の例を示す図である。

[図2]本発明の実施形態に係る高速グラフマッチ探索アルゴリズムによる分子構造の重ね合わせ及びその表示のためのプログラムのフローチャートである。

[図3]分子Aと分子Bの、原子(ノード)及び結合(エッジ)を模式的に示す図(図3(1))と、図3(1)に示す分子Aと分子Bに基づいて算出された分子グラフマッチスコアの例(図3(2))である。

[図4]図2に示すステップS10において、分子Aと分子Bの間の原子対応関

係 $\{m(A_i)\}$ とグラフマッチスコア $M(A, B)$ を求める高速グラフマッチ探索アルゴリズムのフローチャートである。

[図5] 分子Aの一部及び分子Bの一部を示す図であって、原子 A_i と原子 B_k に関する指標 S_1 の算出を説明するための図である。

[図6] 分子Aの一部及び分子Bの一部を示す図であって、原子 A_i と原子 B_k に関する指標 S_2 の算出を説明するための図である。

[図7] 分子Aの一部及び分子Bの一部を示す図であって、原子 A_i と原子 B_k に関する指標 S_3 の算出を説明するための図である。

[図8] 分子Aの一部及び分子Bの一部を示す図であって、図4のステップ S_{1008} におけるグラフマッチスコア M_0 の算出を説明するための図である。

[図9] 分子Aの一部及び分子Bの一部を示す図であって、図4のステップ S_{1012} における微調整を説明するための図である。

[図10] ねじれ角を調整して、分子Aと分子Bをより重ね合わせて表示することを模式的に示す図である。

[図11] クエリの原子 $\{A_i\}$ に対応した $\{m(A_i)\}$ の組で、重ね合わせを行い原子座標を出力した図 (図11(a)) と、クエリ構造から共通骨格にあたる原子座標を出力した図 (図11(b)) である。

[図12] 総組み合わせ数 10^{12} 以下の問題に対して、全探査により最大スコアを求め、本発明の実施形態に係る高速グラフマッチ探索アルゴリズムによる解と比較を行った際の、全探査組み合わせ数に対する計算時間をグラフ化したもの (図12(1)) と、全探査組み合わせ数に対する正解率をグラフ化したもの (図12(2)) と、正解スコア差と累積正解率の関係をグラフ化したもの (図12(3)) である。

[図13] 2つの分子 (Cholic acid [CHD] と Corticosteron [COR]) を仮想空間にて重ね合わせすることを模式的に示す図である。

[図14] 分子Aと分子Bとの重ね合わせの問題について、片方の分子Aを『CMP』とした場合に、それに基づき重ね合わせにて探索可能な重ね合わせの

対象の分子Bの例を示した図である。

発明を実施するための形態

- [0012] 以下、図面を参照して本発明に係る好適な実施の形態を説明する。
- [0013] 本実施形態に係るグラフマッチによる分子構造の高速アルゴリズムは、コンピュータを用いて行われるものであり、C言語などの適切なプログラム言語によって記述されたプログラムをコンピュータで実行し、（後で説明する）様々な分子を構成する原子に関する座標データをコンピュータ上で構築される仮想メモリ空間に展開することにより、実現されるものである。
- [0014] 図1は、本実施形態に係るグラフマッチによる分子構造の高速アルゴリズムを実現するコンピュータシステム2の構成の例を示す図である。コンピュータシステム2は、ディスプレイ等の出力部12、キーボード16やマウス18などの入力部、並びに、演算部、記憶部及び通信制御部等を含む中央処理部14から構成される。中央処理部14は、インターネット4等の外部ネットワークを介して、外部サーバ8や外部データベース10と接続しそれら外部サーバ8や外部データベース10とデータを送受信することができるように、構成されている。
- [0015] 本実施形態で利用される、様々な分子についての原子座標に係るデータは、PDB（プロテインデータバンク；蛋白質構造データバンク）フォーマットのデータであり、通常、外部の商用及び公開データベース10等から提供される。例えば、PDBフォーマットの様々な分子についての原子座標に係るデータは、外部ネットワーク4を介して外部の商用及び公開データベース10からダウンロードされ、コンピュータシステム2に付属する記憶部に格納される。これらのデータは、図2に示すフローチャートに係る処理を実行する際、記憶部から読み出されて利用される。
- [0016] 1. 高速グラフマッチ探索アルゴリズムによる分子構造の重ね合わせ処理
- 図2は、本実施形態に係る高速グラフマッチ探索アルゴリズムによる分子構造の重ね合わせ及びその処理のフローチャートである。図2を参照して本実施形態に係る分子構造の重ね合わせ処理を説明する。まず、重ね合わせの

一方の分子（分子Aとする）についてのPDBフォーマットの原子座標を読み込む（ステップS02）。読み込んだPDBフォーマットの原子座標に基づいて、分子Aの結合距離・結合次数・回転可能結合の設定を行う（ステップS04）。分子の結合距離・結合次数・回転可能結合の設定については後で説明する。

[0017] ステップS02及びステップS04と並行して、重ね合わせのもう一方の分子（分子Bとする）についてのPDBフォーマットの原子座標を読み込む（ステップS06）。なお、分子Bは複数であることがある。次に、分子Bの一つについて結合距離・結合次数・回転可能結合の設定を行う（ステップS08）。

[0018] 続いて、高速グラフマッチ探索アルゴリズムを行い、分子Aの原子（ A_i ）から分子B（ B_k ）への対応関係 $\{m(A_i)\}$ 及びそのときのグラフマッチスコア $M(A, B)$ を求める（ i, k はいずれも自然数）（ステップS10）。ここで、グラフマッチスコア $M(A, B)$ とは、2分子間の原子対応を求め該対応に基づいて2分子の最適な重ね合わせを実現するグラフマッチにおいて、最適さの程度を示す指標である。なお、グラフマッチスコア $M(A, B)$ 、対応関係 $\{m(A_i)\}$ 、及び高速グラフマッチ探索アルゴリズムの、夫々の詳細については、後で説明する。

[0019] グラフマッチスコア $M(A, B)$ が閾値より大きいかどうか確認される（ステップS12）。閾値より大きいということは、そのグラフマッチスコア $M(A, B)$ を実現する重ね合わせのための対応関係（ $m(A_i)$ ）が十分に適切であることを意味する（ステップS12のYes）。このとき、分子Aに対する分子Bのねじれ角が調節され（ステップS14）、分子Aと分子Bにつき原子アラインメント及び構造重ね合わせが出力される（ステップS16）。ねじれ角の調節、並びに、原子アラインメント及び構造重ね合わせの出力についても、後述する。

[0020] 更に、次の分子Bがあるかどうか判断される（ステップS18）。次の分子Bがあれば（ステップS18のYes）、次の分子Bについての結合距離

・結合次数・回転可能結合の設定（ステップS08）以降の処理が繰り返される。

[0021] 分子Bが無くなれば（ステップS18・No）、出力部12に基本骨格（又は共通骨格）を出力して（ステップS20）処理を終了する。

[0022] 2. 結合距離・結合次数・回転可能結合の設定

図2のステップS04及びS08で行われる「結合距離・結合次数・回転可能結合の設定」について説明する。

[0023] (2. 1) 結合距離

PDBフォーマットに係るデータが示す分子構造では、原子間の結合が定義されていないことがある。そこで本実施形態では、一つの分子において、原子*i*と原子*j*の間の原子間距離が2.00 Åより短い場合は化学結合が存在するものとしてデータ上、化学結合を設定する（*i*、*j*はいずれも自然数）。この「原子間距離」は、PDBから読み込まれる原子座標に基づいて計算される。更に、一つの分子において二つの原子を取り上げたとき、それら2原子を繋ぐ化学結合の数を「結合距離」とする。それら2原子を繋ぐ経路が複数存在するときは最小のものを取る。結合を一つずつ延長することで、一つの分子内の全ての原子間に結合距離が設定される。

[0024] (2. 2) 結合次数

PDBフォーマットに係るデータが示す分子構造では、原子間の結合次数が定義されておらず、且つ、一般に水素原子を含んでいない。そこで、以下の表1の示すルールに従い、原子間距離に基づき結合次数を求める。

[0025]

[表1]

結合の種類	ルール (条件)	結合次数
炭素 (C) - 炭素 (C)	{原子間距離} > 1.44 Å	単結合
	1.44 Å ≥ {原子間距離} > 1.25 Å	二重結合
	1.25 Å ≥ {原子間距離}	三重結合
炭素 (C) - 窒素 (N)	{原子間距離} > 1.32 Å	単結合
	1.32 Å ≥ {原子間距離} > 1.21 Å	二重結合
	1.21 Å ≥ {原子間距離}	三重結合
炭素 (C) - 酸素 (O)	{原子間距離} > 1.25 Å	単結合
	1.25 Å ≥ {原子間距離}	二重結合
窒素 (N) - 窒素 (N)	{原子間距離} > 1.30 Å	単結合
	1.30 Å ≥ {原子間距離}	二重結合
酸素 (O) - 酸素 (O)	{原子間距離} > 1.25 Å	単結合
	1.25 Å ≥ {原子間距離}	二重結合
その他の結合		単結合

[0026] (2. 3) 回転可能結合

直接結合する原子の対 (原子 i と原子 j) の全てについて、上記「(2. 1) 結合距離」の定義プロセスを、原子の対間の結合が存在しないものとして実行する。その結果、原子の対 (原子 i と原子 j) 間に結合距離が設定されず、且つ、原子 i と原子 j の間の結合が単結合である場合は、原子 i と原子 j の対の間の結合は「回転可能結合」と設定する。

[0027] 3. 分子グラフマッチスコア定義

本実施形態に係るグラフマッチによる分子構造の高速アルゴリズムでは、分子グラフマッチスコア $M(A, B)$ を定義している。なお $\{M(A, B)\}$ は、分子 A と分子 B との間の分子グラフマッチスコアであることを示す。図 3 (1) は、分子 A と分子 B の、原子 (ノード) 及び結合 (エッジ) を模式的に示す図である。

以下に、本実施形態で利用する分子グラフマッチスコア $M(A, B)$ の定義 (定義 1)、(定義 2)、(定義 3) 及び (定義 4) について説明する。

[0028] (定義 1) ; 「 A_i 」は、分子 A の i 番目の原子であることを示す。「 $A_i - A_j$ 」は、 A_i と A_j の結合を示す。

[0029] (定義2) ; 分子Aの原子i (A_i) が、分子Bの原子k (B_k) に対応することを、「 $m(A_i) = B_k$ 」と表すものとする。即ち、 $m(A_i) = B_k$ とは、分子Aの原子iが対応する分子Bの原子kを示す。

[0030] (定義3)

分子グラフマッチスコアM(A, B)は以下の式(数1)で定義される

[0031] [数1]

$$M(A,B) = \sum_{i,j} E(A_i, A_j) + \sum_i N(A_i)$$

数1の各項は、以下の通り定義される。なおE(A_i, A_j)は、実行modeにより異なる値を持つ。この実行modeは、図1に示す入力部等を介して事後的に外部から設定され得るものである。

[0032] [表2]

項	値	条件	
E(A _i , A _j)	1	A _i -A _j と、m(A _i)-m(A _j)とが、それぞれ直接結合する。	実行mode = 0の場合
	1	A _i -A _j と、m(A _i)-m(A _j)とが、それぞれ直接結合し、A _i とm(A _i)、A _j とm(A _j)はそれぞれ同一元素である。	実行mode = 1の場合
	1	A _i -A _j と、m(A _i)-m(A _j)とが、それぞれ直接結合し、A _i -A _j 、m(A _i)-m(A _j)は結合次数が同じである。	実行mode = 2の場合
	1	A _i -A _j と、m(A _i)-m(A _j)とが、それぞれ直接結合し、A _i とm(A _i)、A _j とm(A _j)はそれぞれ同一元素であり、A _i -A _j 、m(A _i)-m(A _j)は結合次数が同じである。	実行mode = 3の場合
	0	A _i -A _j と、m(A _i)-m(A _j)とが、直接共有結合しないか、上述のいずれも満たさない。	
N(A _i)	1	A _i とm(A _i)が、同一元素である。	
	0	A _i とm(A _i)が、同一元素でない。	

[0033] 図3(2)は、上述の定義に従い、図3(1)に示す分子Aと分子Bに基づいて算出された分子グラフマッチスコアの例である。模様が同じであれば同じ元素であり、エッジは全て単結合であるとしているので、実行modeに関わり無く、図3(2)に示す値(特に、M(A, B) = 14)となる。

[0034] 4. 高速グラフマッチ探索アルゴリズム

図4は、図2に示すステップS10において、分子Aと分子Bの間の、原子の対応関係 $\{m(A_i)\}$ とグラフマッチスコア $M(A, B)$ を求める高速グラフマッチ探索アルゴリズムのフローチャートである。以下、このフローチャートを参照し、高速グラフマッチ探索アルゴリズムを具体的に説明する。

[0035] [ステップS1002] ; まず、分子Aを構成する原子と、分子Bを構成する原子との全ての組み合わせ (A_i, B_k) について、以下の数2及び表3で定義される「 $S_1(A_i, B_k)$ 」を求める。

[0036] [数2]

$$S_1(A_i, B_k) = \sum_j \sum_l s_1(A_j, B_l)$$

[0037] [表3]

	値	条件
$s_1(A_j, B_l)$	1	A_i-A_j と、 B_k-B_l とは同じ結合距離にあり、 A_j と B_l とは同一元素である。
	0	上記条件が成立しない。

[0038] $S_1(A_i, B_k)$ は、原子 A_i と原子 B_k の対において、周囲の環境（同じ結合距離に同じ種類の原子があるか）がどれだけ似ているかを示す指標である。

[0039] 例えば、図5に示される分子Aの一部、及び分子Bの一部において、原子 A_i から2の結合距離にある原子 A_j と、原子 A_i に対応する原子 B_k から2の結合距離にある原子 B_l とが同一元素であれば、 $s_1(A_j, B_l)$ の値は“1”になる。原子 A_i と原子 B_k の対を中心として、同じ結合距離にある、分子Aの原子と分子Bの原子が同じかどうか、全 $\{j, l\}$ の組について確認し、“1”又は“0”を設定して積算する。上記の $S_1(A_i, B_k)$ は、対応する原子 A_i, B_k からみて、同じ結合距離の位置に同じ元素がある、という場合が多い程、大きくなる。

[0040] [ステップS1004] ; 次に、以下の数3及び表4で定義される「 S_2 （

$A_i, B_k)$ 」を、全ての $\{i, k\}$ の組について求める。

[0041] [数3]

$$S_2(A_i, B_k) = \sum_j \sum_l s_2(A_j, B_l)$$

[0042] [表4]

	値	条件
$s_2(A_j, B_l)$	$S_1(A_j, B_l) \times 12$	$A_i - A_j$ と、 $B_k - B_l$ とは同じ結合距離にあり、 A_j と B_l とは同一元素である。
	$S_1(A_j, B_l)$	$A_i - A_j$ と、 $B_k - B_l$ とは同じ結合距離にあるが、 A_j と B_l とは同一元素ではない。
	0	上記条件が成立しない。

[0043] $S_2(A_i, B_k)$ は、対応する原子 A_i, B_k の対の夫々において、その対の各原子から等しい結合距離にある周囲の原子 A_j, B_l の全ての組について、上記の、周囲の環境がどれだけ似ているかを示す指標である $S_1(A_j, B_l)$ を積算する指標であるが、その対の各原子から等しい結合距離にある周囲の原子 A_j, B_l が同じものであれば、更に $S_1(A_j, B_l)$ に係数（上記表では12）を掛けて積算される。従って、対応する原子 A_i, B_k の対の各原子について、周囲の環境が類似し、更に周囲の環境のその周囲の環境が類似すれば、大きくなる指標である。

[0044] 例えば、図6に示される、原子 A_i を含む分子Aの一部、及び原子 B_k を含む分子Bの一部において、 $S_2(A_i, B_k)$ を検討する。原子 A_i からある結合距離（図6では2）にある原子 A_j と、原子 B_k からそれと等しい結合距離にある原子 B_l との全ての対につき、 $s_2(A_j, B_l)$ 、即ち $S_1(A_j, B_l)$ 、又は $S_1(A_j, B_l) \times 12$ を積算する。特に、 A_j と B_l が同じ元素であれば、 $S_1(A_j, B_l)$ は所定数倍（ここでは12倍）されて積算されて、 S_2 が求められる。原子 A_i 及び原子 B_k からの結合距離は、1から最大値（即ち、原子 A_i 又は原子 B_k から最も遠い原子までの結合距離）まで変動することが想定される。上述のとおり、 $S_1(A_j, B_l)$ は、原子 A_j, B_l の対において、（図6の A_m, B_n などの）周囲の環境がどれだけ似ているかを示す指標である。

[0045] 上記の $S_2(A_i, B_k)$ では、対応する2つの原子 A_i, B_k に関して、同じ結合距離の位置の原子の対 (A_j, B_l) の $S_1(A_j, B_l)$ が積算されるが、 (A_j, B_l) が同じ元素であれば、 $S_1(A_j, B_l)$ が所定数倍（1.2倍）されて積算されるから、周囲の原子の構成が近似するように対応付けされていると、やはり $S_2(A_i, B_k)$ は大きくなる。なお、係数「1.2」は別の数値であってもよい。

[0046] [ステップS1006]；次に、以下の数4及び表5で定義される「 $S_3(A_i, A_k)$ 」を、全ての $\{i, k\}$ の組について求める。

[0047] [数4]

$$S_3(A_i, B_k) = m_{0i,k}(A, B)$$

[0048] [表5]

	値条件
$m_{0i,k}(A, B)$	原子 (A_i, B_k) の対応から開始して、既に対応した原子に直接結合し、未対応で最大の $S_2(A_i, B_k)$ を持つ原子の組に対応させることを、順次、対応可能原子が無くなるまで続けたときの、全体の対応におけるグラフマッチスコア $M(A, B)$

[0049] $S_3(A_i, B_k)$ は、原子 A_i, B_k の対を始点とし、次々に分子Aの原子と分子Bの原子を対応付けして全体の対応を作成し、そのときのグラフマッチスコア $M(A, B)$ を値とする指標である。ここで、対応付け作成時には、既に対応付け済みの原子に直接結合する原子を次に選択すること、及び指標 S_2 が高い対を選択するのを優先することを、条件としている。

[0050] 例えば、図7に示される、原子 A_i を含む分子Aの一部、及び原子 B_k を含む分子Bの一部において、 $S_3(A_i, B_k)$ を検討する。始点は、原子 A_i, B_k の対である。原子 A_i には、原子 A_j, A_p, A_r が直接結合する。原子 B_k には、原子 B_l, B_q, B_s が直接結合する。 $\{A_j, A_p, A_r\}$ と $\{B_l, B_q, B_s\}$ とから形成され得る原子同士の（ $3 \times 3 = 9$ 通りの）対のうちから、 (A_j, B_l) の対の S_2 が最大であるとする、原子 A_j と原子 B_l を対応付けすることになる。

[0051] 次に、分子Aにおいて対応付けが済んだ $A_i - A_j$ には、原子 A_p, A_r

、 A_t 、 A_v が直接結合する。分子Bにおいて対応付けが済んだ $B_k - B_l$ には、原子 B_q 、 B_s 、 B_u 、 B_w が直接結合する。 $\{A_p, A_r, A_t, A_v\}$ と $\{B_q, B_s, B_u, B_w\}$ とから形成され得る原子同士の($4 \times 4 = 16$ 通りの)対のうちから、(A_p, B_q)の対の S_2 が最大であるとすると、原子 A_p と原子 B_q を対応付けすることになる。これにより、分子Aにおいては、原子 A_i 、 A_j 、 A_p の対応付けが完了し、分子Bにおいては、原子 B_k 、 B_l 、 B_q の対応付けが完了する。

[0052] このような対応付けを、対応可能原子の対が無くなるまで、順次繰り返して行う。対応付けが終われば、その対応付けの下でのグラフマッチスコア M を求める。このような対応付け及びグラフマッチスコア M 算出が、全ての $\{i, k\}$ の組について行われる。

[0053] 上記の $S_3(A_i, B_k)$ では、全ての $\{A_i, B_k\}$ の組み合わせの各々において、原子の対の始点 $\{A_i, B_k\}$ の周囲から徐々に、 S_2 (対応する原子 A_i 、 B_k の対について、周囲の環境が類似し、更に周囲の環境のその周囲の環境が類似すれば、大きくなる指標)の大きさに着目して、分子Aの原子と分子Bの原子とが対応付けされ、グラフマッチスコアが計算されることになる。

[0054] [ステップS1008]; 次に、ステップS1006にて最大の $S_3(A_i, B_k)$ を算出した際の、始点の原子(A_i, B_k)の対応から開始して、未対応の原子の対の中で最大の $S_3(A_j, B_l)$ を持つものを対応させることを、対応可能原子の対が無くなるまで続け、全体の対応におけるグラフマッチスコア $M_0(A, B)$ を求める。このとき、途中、原子の対応の対と、次の原子の対応の対とにおいて、分子Aの原子は直接結合していなくてもよく、同様に、分子Bの原子も直接結合していなくてもよい。

[0055] 例えば、図8に示される、原子 A_i を含む分子Aの一部、及び原子 B_k を含む分子Bの一部において、ステップS1008で行われる原子の対の対応付けを検討する。まず、分子Aを構成する(例えば、 a 個の)全ての原子と、分子Bを構成する(例えば、 b 個の)全ての原子とから形成され得る原子

同士の ($a \times b$ 通りの) 対のうち、原子 A_i 、 B_k の対において、(ステップ S1006で求めた) S_3 が、他のどの対よりも大きい、即ち最大であるとする。そうすると、まず原子 A_i 、 B_k の対が対応付けされる。

次に、 A_i を除いた分子Aを構成する ($a-1$)個の原子と、 B_k を除いた分子Bを構成する ($b-1$)個の原子とから、形成され得る原子同士の ($a-1$) \times ($b-1$)通りの対のうち、原子 A_j 、 B_l の対において、 S_3 が、他のどの対よりも大きいとする。そうするとそこで原子 A_j 、 B_l の対が対応付けされる。このとき、 A_j は A_i と直接結合しているとは限らず、 B_l は B_j と直接結合しているとは限らない(このことは以下、同様である)。

[0056] 次に、 A_i と A_j を除いた分子Aを構成する ($a-2$)個の原子と、 B_k と B_l を除いた分子Bを構成する ($b-2$)個の原子とから、形成され得る原子同士の ($a-2$) \times ($b-2$)通りの対のうち、原子 A_{j2} 、 B_{l2} の対において、 S_3 が、他のどの対よりも大きいとする。そうするとそこで原子 A_{j2} 、 B_{l2} の対が対応付けされる。

更に次に、 A_i と A_j と A_{j2} を除いた分子Aを構成する ($a-3$)個の原子と、 B_k と B_l と B_{l2} を除いた分子Bを構成する ($b-3$)個の原子とから、形成され得る原子同士の ($a-3$) \times ($b-3$)通りの対のうち、原子 A_{j3} 、 B_{l3} の対において、 S_3 が、他のどの対よりも大きいとする。そうするとそこで原子 A_{j3} 、 B_{l3} の対が対応付けされる。

更に次に、 A_i と A_j と A_{j2} と A_{j3} を除いた分子Aを構成する ($a-4$)個の原子と、 B_k と B_l と B_{l2} と B_{l3} を除いた分子Bを構成する ($b-4$)個の原子とから、形成され得る原子同士の ($a-4$) \times ($b-4$)通りの対のうち、原子 A_{j4} 、 B_{l4} の対において、 S_3 が、他のどの対よりも大きいとする。そうするとそこで原子 A_{j4} 、 B_{l4} の対が対応付けされる。

[0057] このような対応付けを、対応可能原子の対が無くなるまで、順次繰り返して行う。対応付けが終われば、その対応付けの下でのグラフマッチスコアM

0を求める。

- [0058] ステップS1008では、ステップS1006で求めた多数の（例えば、 $a \times b$ 通りの）S3、即ちM(A, B)に基づいて、最終候補となり得る対応付け、及びその対応付けの下でのグラフマッチスコアM0の算出が行われる。
- [0059] [ステップS1010]；算出したM0(A, B)が、想定され得る最大値であるか否かが確認される。具体的には、M0(A, B)が、M(A, A)又はM(B, B)に等しいかどうか、確認される。図3(2)に示すように、M(A, A)（又はM(B, B)）は、最大値であると考えられるから、このステップS1010はこれ以上、グラフマッチスコアを算出する必要がないかどうかを確認するために行われる。
- [0060] 等しければ（ステップS1010・Yes）、ステップS1016にて原子対応 {m(A_i)} とグラフマッチスコアM0(A, B)を出力して終了する。等しくなければ（ステップS1010・No）、ステップS1012に移行する。
- [0061] [ステップS1012]；ステップS1012では、最終候補となり得る対応付けの微調整が行われる。
- [0062] 分子Aにおけるひとつの結合した {A_i, A_j} の組に対応する、分子Bの {B_k, B_l} において、一方を他の原子B_nと入れ換え、入れ換えた原子についてのみ、分子Aにおける原子の対応を変更して、グラフマッチスコアM1(A, B)を求める。なお、原子B_nが分子Aにおいて対応する原子を持たない場合であってもよい。
- [0063] 例えば、図9に示される、原子A_i、A_j、A_mを含む分子Aの一部、及び、原子B_k、B_l、B_nを含む分子Bの一部において、ステップS1012で行われる原子の対の対応付けの変更の例を、説明する。A_iとB_k、A_jとB_l、及び、A_mとB_nが、対応付けられており、A_iとA_jが結合しているとする。ここで、{B_k, B_l} のうち的一方であるB_lと、B_nとを入れ換え、A_jとB_nを対応付け、同時に、A_mとB_lを対応付ける。即

ち、 $m(A_j) = B_l$ であったものを $m(A_j) = B_n$ とし、 $m(A_m) = B_n$ であったものを $m(A_m) = B_l$ とする。その他の原子に係る対応付けは動かされない。この一部のみ変更された対応付けに基づいて、グラフマッチスコア $M_1(A, B)$ を求める。

[0064] 図9の例における原子 B_k を(図示しない) B_p と入れ替える、というような対応付けの変更であってもよい。

[0065] [ステップS1014]; 算出した $M_1(A, B)$ が、 $M_0(A, B)$ より大きいかどうか、確認される。即ち、ステップS1012にて、微調整を施した原子対応付けから算出されるグラフマッチスコア $M_1(A, B)$ の変動が確認される。算出した $M_1(A, B)$ が、 $M_0(A, B)$ より大きければ(ステップS1014・Yes)、 $M_1(A, B)$ の値が $M_0(A, B)$ に上書きされ(ステップS1015)、S1012にて更に微調整が施された原子対応付けから算出されるグラフマッチスコア $M_1(A, B)$ が求められる。

[0066] [ステップS1016]; 算出した $M_1(A, B)$ が、 $M_0(A, B)$ より大きくなければ(ステップS1014・No)、原子対応付けとグラフマッチスコア $M_0(A, B)$ を出力して終了する。

[0067] 5. 分子の構造重ね合わせ

図4及び図2に示すフローチャートにより求めた原子対応に基づく、構造重ね合わせの表示について説明する。分子Aと分子Bの分子構造の重ね合わせにおいて、分子Aの原子 $\{A_i\}$ に対応した原子 $\{m(A_i)\}$ は適宜、重ね合わせられて表示される。このとき、Kabschの方法(McLachlan, AD. Gene duplications in the structural evolution of chymotrypsin. Journal of Molecular Biology, 128, 49-79, 1979. Kabsch, W. A solution for the best rotation to relate two sets of vectors. Acta Crystallographica, 32A, 922-923, 1976.)が用いられてもよい。

[0068] このとき、2分子間で対応するねじれ角は、以下の方法でそろえられる。

(1) グラフマッチにより結合した分子Aの原子 $\{A_i, A_j, A_k, A$

l} が、同様に結合した分子Bの原子 {m (A i), m (A j), m (A k), m (A l)} に対応し、かつ、結合 A j - A k と、m (A j) - m (A k) がいずれも回転可能結合であれば、分子Bのねじれ角 {m (A i), m (A j), m (A k), m (A l)} を、分子Aの対応するねじれ角 {A i, A j, A k, A l} と同値にする。

[0069] 上記 (1) の様子を模式的に表現したのが、図6である。「回転可能結合」であるか否かの判断においては、図2におけるステップS04及びS08や、「(2.3) 回転可能結合」にて設定されるデータが利用される。

[0070] 6. 高速グラフマッチ探索アルゴリズムによる分子構造の重ね合わせ及びその処理の実施例。

図2に示される高速グラフマッチ探索アルゴリズムによる分子構造の重ね合わせ及びその処理のためのフローチャートを実現するプログラムを実装し、クエリ (分子A) をG39 (タフミル) とし、探索ターゲットデータベースをPDBの全リガンド (9445種) として、計算を行った。動作周波数 2.4 GHz のデスクトップコンピュータを利用した。計算時間は、8分56秒であった。

[0071] 以下の表6で、上述の計算によりクエリの原子 {A i} に対応した {m (A i)} を、下に並べて表示している。左端カラムには、クエリ (分子A) 及び探索対象分子 (分子B) の例を示している。左端から2番目のカラムには、原子数を示している。左端から3番目のカラムには、分子A (G39 (タフミル)) と分子Bとのグラフマッチスコアを示している。左端から4番目のカラムには、自己 (自己の分子) とのグラフマッチスコアを示している。そして、右部には原子アラインメントを示している。各カラム (縦) に並んだ原子種が一致する場合は共通骨格に当たる。ここで、各カラム (縦) に並んだ原子種が90%以上一致すれば最下行に “**” を、50%以上一致すれば最下行に “+” を、示している。

[0072]

12 (1) ~ (3) に示すグラフ及び数値から、本実施形態に係る高速グラフマッチ探索アルゴリズムは、高い性能を持つと考えられる。

[0077] (7. 2 アルゴリズム性能評価 (2))

ブロンのクリーク探索アルゴリズム (Bron G. & Kerbosch J. Algorithm 457: Finding all cliques of an undirected graph. Communications of the Association for Computing Machinery, 16, 575-577, 1973) を用いて発見的にグラフマッチを行う方法である `simcomp` の方法 (Hattori, M., Okuno, Y., Goto, S. & Kanehisa, M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. Journal of American Chemical Society, 125, 11853-11865, 2003) と、成績比較を行った。

[0078] ランダムに選んだ同じ 50 種の分子集合に対し総当たりグラフマッチを行い、全比較 1225 例 (`all`)、及びいずれかの方法が部分グラフ (確実に正解である) を発見した 136 例 (`partial`) について、本実施形態の定義によるスコアと実行時間 (グラフマッチに要した実時間) を比較した。`simcomp` の方法における最大試行回数 (`Rmax`) を、 1.5×10^4 (デフォルト値) ~ 10^8 で変化させた。

[0079] その結果 (以下、表 7 参照)、本実施形態に係る高速グラフマッチ探索アルゴリズムは、136 の部分グラフ (`partial`) をすべて発見したのに対し、`simcomp` の方法は 10 例 (7%) で失敗した。実行時間は一例を除いて本法が高速で、平均 48 ミリ秒高速であった。`Rmax` を増大させても発見できる部分グラフに逆転はなく、`simcomp` の方法の実行時間が増大するだけであった。また全比較 (`all`) においても、`Rmax` = 1.5×10^4 で本実施形態に係る高速グラフマッチ探索アルゴリズムが 96 ミリ秒遅い (但し、発見したグラフマッチのスコアは高い) 以外は、どの `Rmax` においても、より高速により高スコアのグラフマッチを発見した。これらの数値から、本実施形態に係る高速グラフマッチ探索アルゴリズムは高い性能を持つと考えられる。

[0080] [表7]

実行時間差とスコア差 { Δ (本実施形態 - simcomp) } の平均				
R m a x	$\langle \Delta t_{sec} \rangle_{all}$	$\langle \Delta S \rangle_{all}$	$\langle \Delta t_{sec} \rangle_{partial}$	$\langle \Delta S \rangle_{partial}$
1.5×10^4	0.096	2.43	-0.048	0.16
1.5×10^5	-0.127	2.13	-0.505	0.04
1.5×10^6	-1.612	2.00	-4.779	0.01
1.5×10^7	-12.912	1.98	-23.020	0.01
1.5×10^8	-95.416	1.94	-95.723	0.01

符号の説明

[0081] 2・・・コンピュータシステム、4・・・インターネット、8・・・外部サーバ、10・・・外部データベース、12・・・出力部、14・・・中央処理部、16・・・キーボード、18・・・マウス。

請求の範囲

[請求項1]

第1の分子Aを構成する原子 (A_i, A_j, \dots) の各々に係る座標データと第2の分子Bを構成する原子 (B_k, B_l, \dots) の各々に係る座標データを記憶部から入力し、演算部にロードされるコンピュータプログラムに従って、演算部及び記憶部に構築される仮想メモリ空間において第1の分子Aの夫々の原子 (A_i, A_j, \dots) と第2の分子Bの夫々の原子 (B_k, B_l, \dots) との対応付け ($m(A_i) = B_k$) を求めて重ね合わせを行い (i, j, k, l はいずれも自然数)、第1の分子Aと第2の分子Bの間の最適な原子間対応、及び第1の分子Aと第2の分子Bの類似度に係るデータを出力部に出力する、第1の分子Aと第2の分子Bとの類似度を評価するための高速グラフマッチ検索装置において、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対の各原子からみて、周囲の環境が相互にどれだけ似ているかを示す第1の類似指標 $S_1(A_i, B_k)$ を求める第1の算出手段と、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対の各原子からみて、等しい結合距離にある周囲の原子 A_j 、 B_l の全ての組につき、第1の類似指標 $S_1(A_j, B_l)$ を積算して算出する第2の類似指標 $S_2(A_i, B_k)$ を求める算出手段であって、その原子 A_i 、 B_k の対の各原子から等しい結合距離にある周囲の原子 A_j 、 B_l が同じ元素であれば、更に第1の類似指標 $S_1(A_j, B_l)$ に係数を掛けた上で積算する、第2の類似指標 $S_2(A_i, B_k)$ を求める第2の算出手段と、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、

B_kの対を始点とし、第1の分子Aの原子と第2の分子Bの原子とを順次対応付けして全体の対応を作成し、そのときに算出されるグラフマッチスコアM(A, B)を値とする第3の類似指標S₃(A_i, B_k)を求める算出手段であって、対応付け作成時には、既に対応付け済みの原子に直接結合する原子を次に選択すること、及び第2の類似指標S₂が高い対を選択することを優先することを、条件とする、第3の類似指標S₃(A_i, B_k)を求める第3の算出手段と、

第3の算出手段にて最大のS₃(A_i, B_k)を算出した際の、始点の原子(A_i, B_k)の対から開始して、未対応の原子の対の中で最大のS₃(A_j, B_l)を持つものを対応させることを、対応可能原子の組が無くなるまで続けたときの、全体の対応におけるグラフマッチスコアM(A, B)を求める第4の算出手段と、

第4の算出手段におけるグラフマッチスコアM(A, B)が閾値より大きいならば、第1の分子Aと第2の分子Bにつき第4の算出手段で算出した原子間対応及びグラフマッチスコアM(A, B)を出力する第5の出力手段と

を含む、分子間の類似度を評価するための高速グラフマッチ検索装置。

[請求項2]

更に、

上記第4の算出手段によりグラフマッチスコアM(A, B)を算出した後に、第1の分子Aにおけるひとつの結合した{A_i, A_j}の組に対応する、第2の分子Bの{B_k, B_l}において、一方を他の原子B_nと入れ換え、入れ換えた原子についてのみ、第1の分子Aにおける原子の対応を変更して、微調整されたグラフマッチスコアM(A, B)を求める第6の算出手段を含み、

上記第6の算出手段にて算出された、微調整されたグラフマッチスコアM(A, B)が、上記第3の算出手段にて算出されたグラフマッチスコアM(A, B)より大きければ、上記第5の出力手段が、微調

整されたグラフマッチスコアM (A , B) をグラフマッチスコアM (A , B) に上書きして出力を行う

ことを特徴とする請求項 1 に記載の高速グラフマッチ検索装置。

[請求項3]

グラフマッチスコアM (A , B) が、下記の数 1 で定義され、下記数 1 の各項は、下記の表 1 で定義され、下記表 1 の実行modeの値は入力手段を介して外部から設定されることを特徴とする

請求項 1 又は 2 に記載の高速グラフマッチ検索装置。

[数1]

$$M(A,B) = \sum_{i,j} E(A_i, A_j) + \sum_i N(A_i)$$

[表1]

項	値	条件	
E (A i , A j)	1	Ai-Ajと、m(Ai)-m(Aj)とが、それぞれ直接共有結合する。	実行mode= 0 の場合
	1	Ai-Ajと、m(Ai)-m(Aj)とが、それぞれ直接結合し、Aiとm(Ai)、Ajとm(Aj)はそれぞれ同一元素である。	実行mode= 1 の場合
	1	Ai-Ajと、m(Ai)-m(Aj)とが、それぞれ直接結合し、Ai-Aj、m(Ai)-m(Aj)は結合次数が同じである。	実行mode= 2 の場合
	1	Ai-Ajと、m(Ai)-m(Aj)とが、それぞれ直接結合し、Aiとm(Ai)、Ajとm(Aj)はそれぞれ同一元素であり、Ai-Aj、m(Ai)-m(Aj)は結合次数が同じである。	実行mode= 3 の場合
N (A i)	0	Ai-Ajと、m(Ai)-m(Aj)とが、直接結合しないか、上述のいずれも満たさない。	
	1	Aiとm(Ai)が、同一元素である。	
	0	Aiとm(Ai)が、同一元素でない。	

[請求項4]

上記第 1 の算出手段における第 1 の類似指標「S 1 (A i , B k)」

が、下記の数 2 及び表 2 で定義され、

上記第 2 の算出手段における第 2 の類似指標「S 2 (A i , B k)」

が、下記の数 3 及び表 3 で定義され、

上記第 3 の算出手段における第 3 の類似指標「S 3 (A i , B k)」

が、下記の数 4 及び表 4 で定義される

ことを特徴とする請求項 1 乃至 3 のうちのいずれか一に記載の高速グラフマッチ検索装置。

[数2]

$$S_1(A_i, B_k) = \sum_j \sum_l s_1(A_j, B_l)$$

[表2]

	値	条件
$s_1(A_j, B_l)$	1	$A_i - A_j$ と、 $B_k - B_l$ とは同じ結合距離にあり、 A_j と B_l とは同一元素である。
	0	上記条件が成立しない。

[数3]

$$S_2(A_i, B_k) = \sum_j \sum_l s_2(A_j, B_l)$$

[表3]

	値	条件
$s_2(A_j, B_l)$	$S_1(A_j, B_l) \times$ 係数	$A_i - A_j$ と、 $B_k - B_l$ とは同じ結合距離にあり、 A_j と B_l とは同一元素である。
	$S_1(A_j, B_l)$	$A_i - A_j$ と、 $B_k - B_l$ とは同じ結合距離にあるが、 A_j と B_l とは同一元素ではない。
	0	上記条件が成立しない。

[数4]

$$S_3(A_i, B_k) = m_{0i,k}(A, B)$$

[表4]

	値	条件
$m_{0i,k}(A, B)$		原子(A_i, B_k)の対応から開始して、既に対応した原子に直接結合し、未対応で最大の $S_2(A_i, B_k)$ を持つ原子の組を対応させることを、順次、対応可能原子が無くなるまで続けたときの、全体の対応におけるグラフマッチスコア $M(A, B)$

[請求項5] 上記表 3 における係数が、1 2であることを特徴とする請求項 4 に記載の高速グラフマッチ検索装置。

[請求項6] 記憶部に格納される第 1 の分子 A を構成する原子 (A_i, A_j, \dots) の各々に係る座標データと第 2 の分子 B を構成する原子 (B_k, B_l, \dots) の各々に係る座標データを入力し、演算部にロード

される所与のコンピュータプログラムに従って、コンピュータ上に構築される仮想メモリ空間において第1の分子Aの夫々の原子 (A_i , A_j , ...) と第2の分子Bの夫々の原子 (B_k , B_l , ...) との対応付け ($m(A_i) = B_k$) を求めて重ね合わせを行い (i , j , k , l はいずれも自然数)、第1の分子Aと第2の分子Bの間の最適な原子間対応、及び第1の分子Aと第2の分子Bの類似度を出力部に出力する、コンピュータを用いて第1の分子Aと第2の分子Bとの類似度を評価するための高速グラフマッチ検索方法において、

第1の分子Aの全ての原子 A_i と第1の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対の各原子からみて、周囲の環境が相互にどれだけ似ているかを示す第1の類似指標 $S_1(A_i, B_k)$ を求める第1の工程と、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対の各原子からみて、等しい結合距離にある周囲の原子 A_j 、 B_l の全ての組につき、第1の類似指標 $S_1(A_j, B_l)$ を積算する第2の類似指標 $S_2(A_i, B_k)$ を求める工程であって、その A_i 、 B_k の対の各原子から等しい結合距離にある周囲の原子 A_j 、 B_l が同じ元素であれば、更に第1の類似指標 $S_1(A_j, B_l)$ に係数を掛けた上で積算する、第2の類似指標 $S_2(A_i, B_k)$ を求める第2の工程と、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対を始点とし、第1の分子Aの原子と第2の分子Bの原子とを順次対応付けして全体の対応を作成し、そのときに算出されるグラフマッチスコア $M(A, B)$ を値とする第3の類似指標 $S_3(A_i, B_k)$ を求める工程であって、対応付け作成時には、既に対応付け済みの原子に直接結合する原子を次に選択すること、及び第2の類似指標

S₂が高い対を選択することを優先することを、条件とする、第3の類似指標S₃(A_i, B_k)を求める第3の工程と、

第3の工程にて最大のS₃(A_i, B_k)を算出した際の、始点の原子(A_i, B_k)の対から開始して、未対応の原子の対の中で最大のS₃(A_j, B_l)を持つものを対応させることを、対応可能原子の組が無くなるまで続けたときの、全体の対応におけるグラフマッチスコアM(A, B)を求める第4の工程と、

第4の工程におけるグラフマッチスコアM(A, B)が閾値より大きいならば、第1の分子Aと第2の分子Bにつき第4の工程で算出した原子間対応及びグラフマッチスコアM(A, B)を出力する第5の工程と

を含む、分子間の類似度を評価するための高速グラフマッチ検索方法。

[請求項7]

記憶部に格納される第1の分子Aを構成する原子(A_i, A_j, ...)の各々に係る座標データと第2の分子Bを構成する原子(B_k, B_l, ...)の各々に係る座標データを入力し、コンピュータ上に構築される仮想メモリ空間において、第1の分子Aの夫々の原子(A_i, A_j, ...)と第2の分子Bの夫々の原子(B_k, B_l, ...)との対応付け(m(A_i) = B_k)を求めて重ね合わせを行い(i, j, k, lはいずれも自然数)、第1の分子Aと第2の分子Bの間の最適な原子間対応、及び第1の分子Aと第2の分子Bとの類似度を評価する処理を、コンピュータに実行させるコンピュータプログラムにおいて、

第1の分子Aの全ての原子A_iと第2の分子Bの全ての原子B_kとで形成される、原子A_iと原子B_kの組の全てに関して、原子A_i, B_kの対の各原子からみて、周囲の環境が相互にどれだけ似ているかを示す第1の類似指標S₁(A_i, B_k)を求める第1の算出ステップと、

第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対の各原子からみて、等しい結合距離にある周囲の原子 A_j 、 B_l の全ての組につき、第1の類似指標 $S_1(A_j, B_l)$ を積算する第2の類似指標 $S_2(A_i, B_k)$ を求める算出ステップであって、その A_i 、 B_k の対の各原子から等しい結合距離にある周囲の原子 A_j 、 B_l が同じ元素であれば、更に第1の類似指標 $S_1(A_j, B_l)$ に係数を掛けた上で積算する、第2の類似指標 $S_2(A_i, B_k)$ を求める第2の算出ステップと、

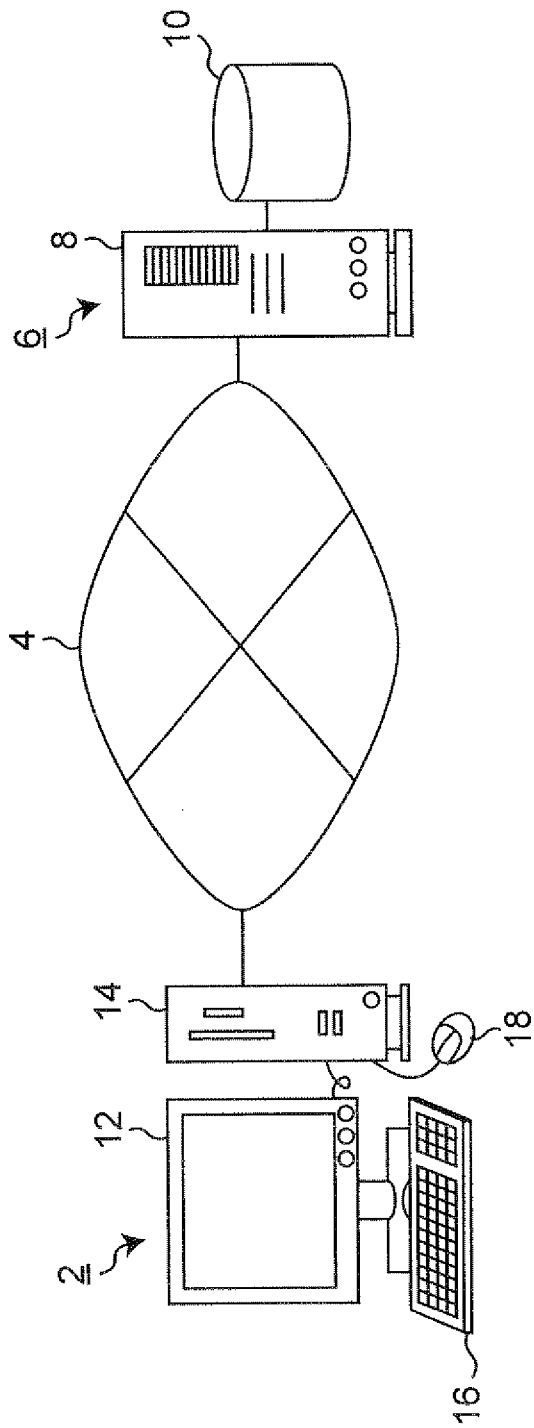
第1の分子Aの全ての原子 A_i と第2の分子Bの全ての原子 B_k とで形成される、原子 A_i と原子 B_k の組の全てに関して、原子 A_i 、 B_k の対を始点とし、第1の分子Aの原子と第2の分子Bの原子とを順次対応付けして全体の対応を作成し、そのときに算出されるグラフマッチスコア $M(A, B)$ を値とする第3の類似指標 $S_3(A_i, B_k)$ を求める算出ステップであって、対応付け作成時には、既に対応付け済みの原子に直接結合する原子を次に選択すること、及び第2の類似指標 S_2 が高い対を選択するのを優先することを、条件とする、第3の類似指標 $S_3(A_i, B_k)$ を求める第3の算出ステップと、

第3の算出ステップにて最大の $S_3(A_i, B_k)$ を算出した際の、始点の原子(A_i, B_k)の対から開始して、未対応の原子の対の中で最大の $S_3(A_j, B_l)$ を持つものを対応させることを、対応可能原子の組が無くなるまで続けたときの、全体の対応におけるグラフマッチスコア $M(A, B)$ を求める第4の算出ステップと、

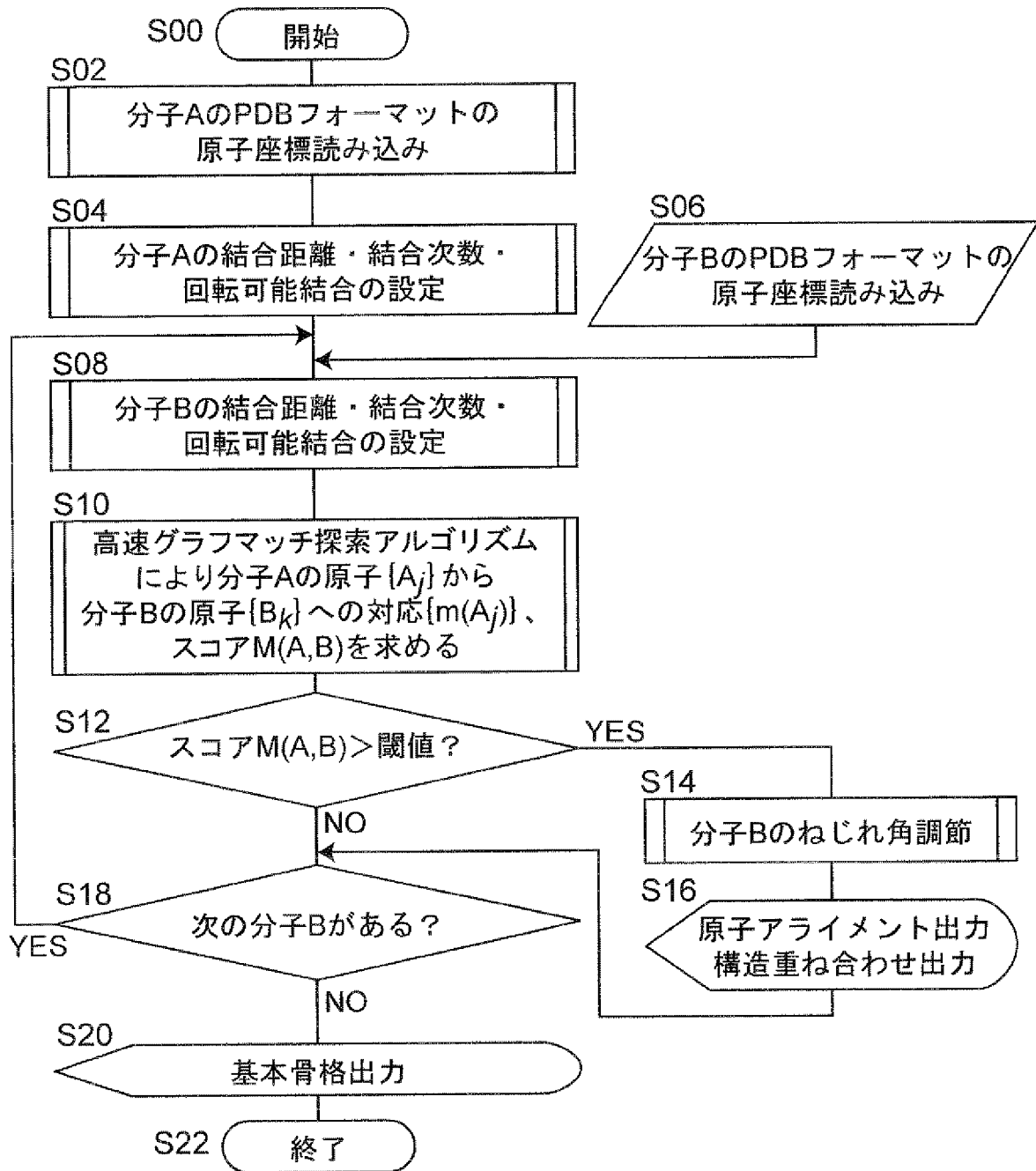
第4の算出ステップにおけるグラフマッチスコア $M(A, B)$ が閾値より大きいならば、第1の分子Aと第2の分子Bにつき第4の工程で算出した原子間対応及びグラフマッチスコア $M(A, B)$ を出力する第5の出力ステップとを

コンピュータに実行させるコンピュータプログラム。

[図1]

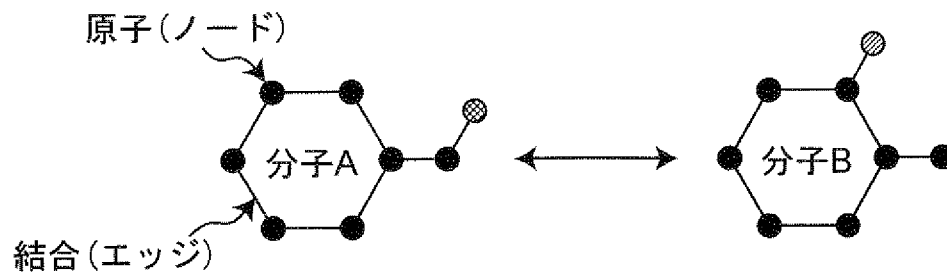


[図2]



[図3]

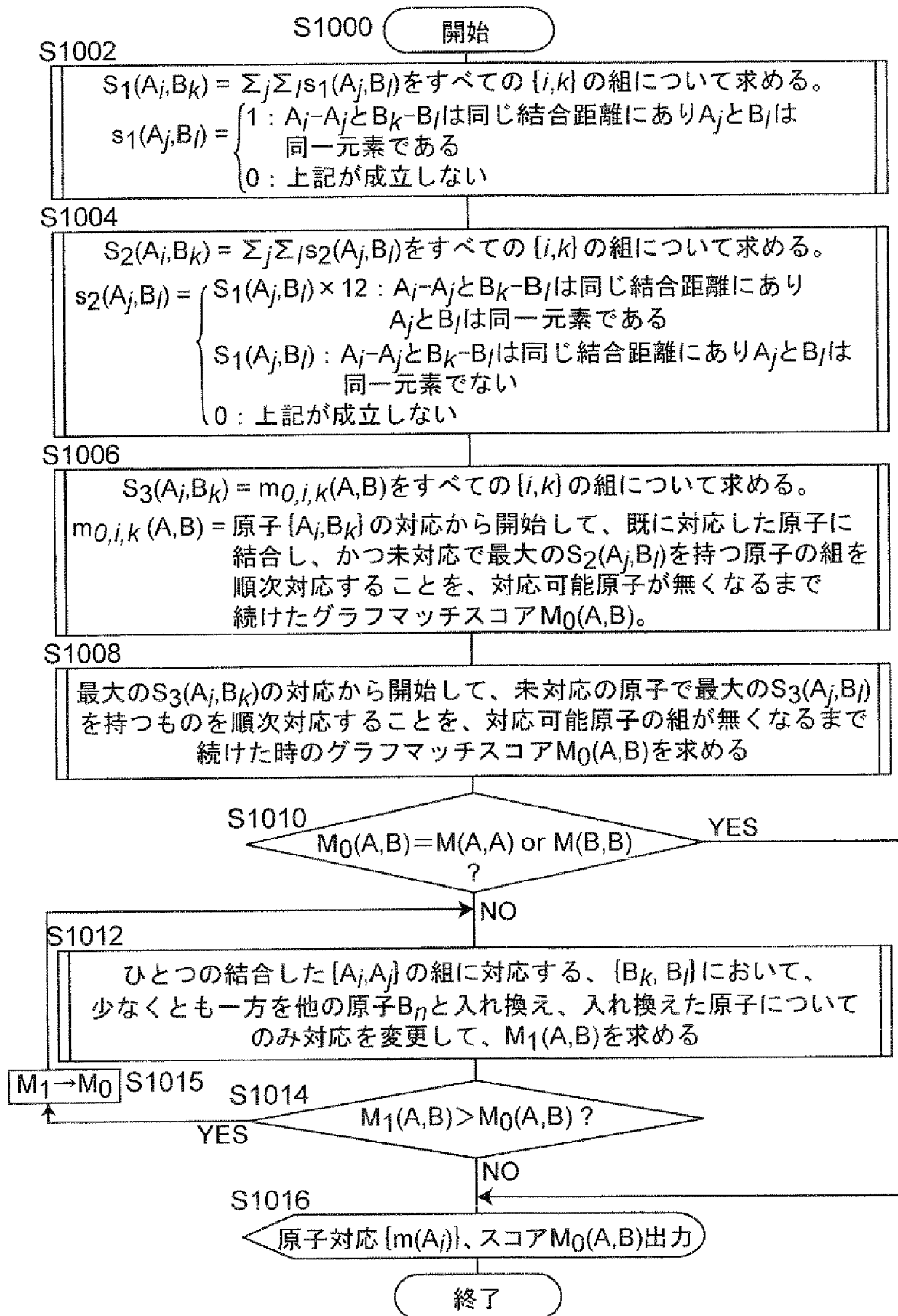
(1)



(2)

$$M(A,A) = 16,$$
$$M(B,B) = 16,$$
$$M(A,B) = 14$$

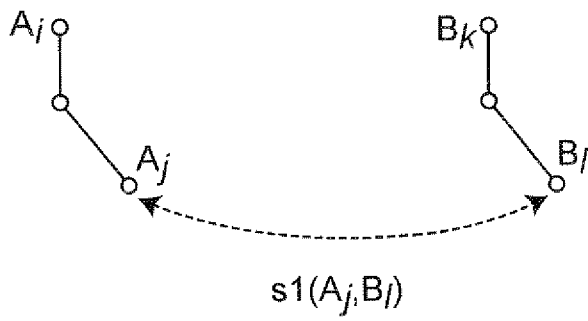
[図4]



[图5]

[分子A]

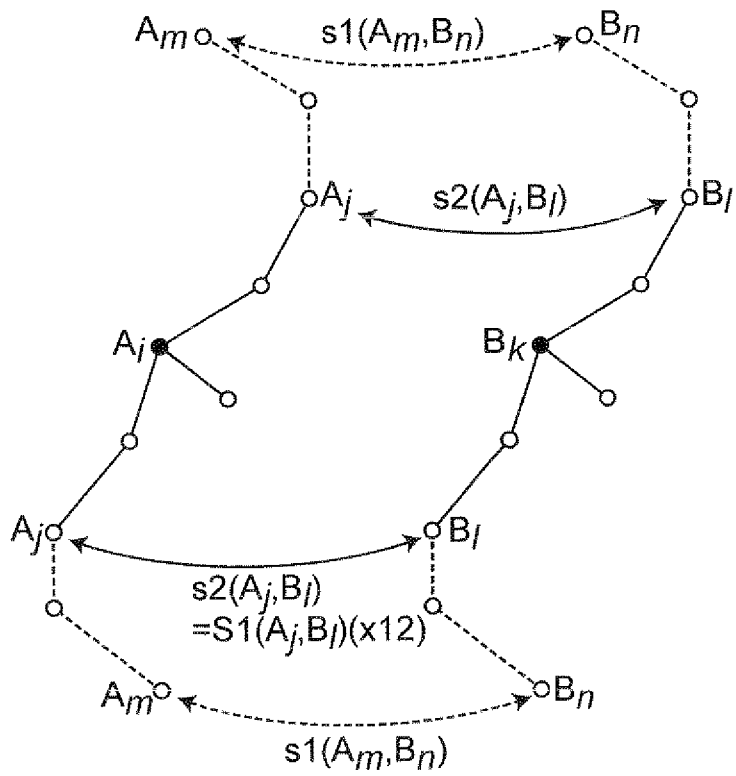
[分子B]



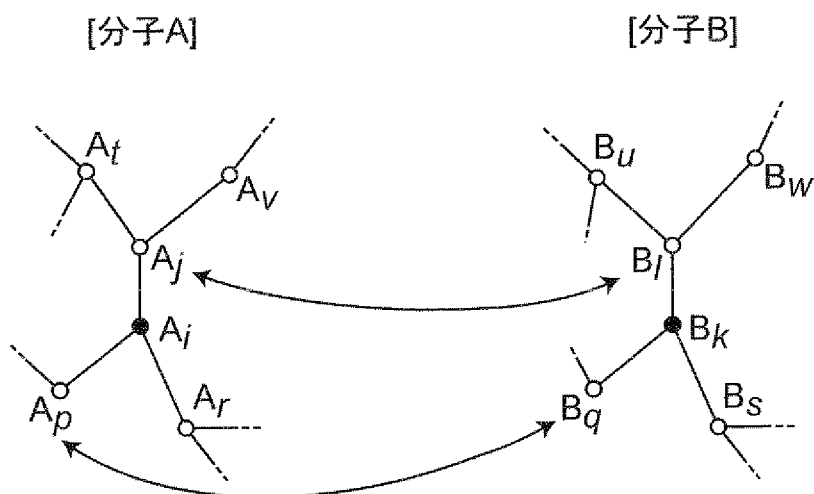
[图6]

[分子A]

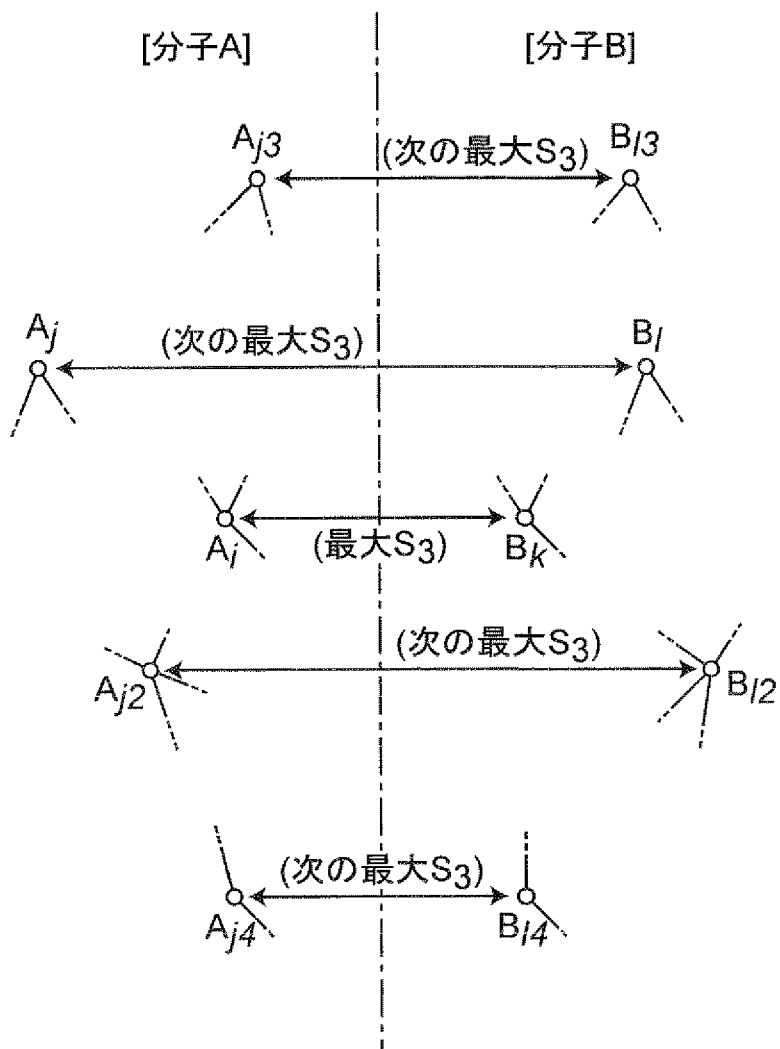
[分子B]



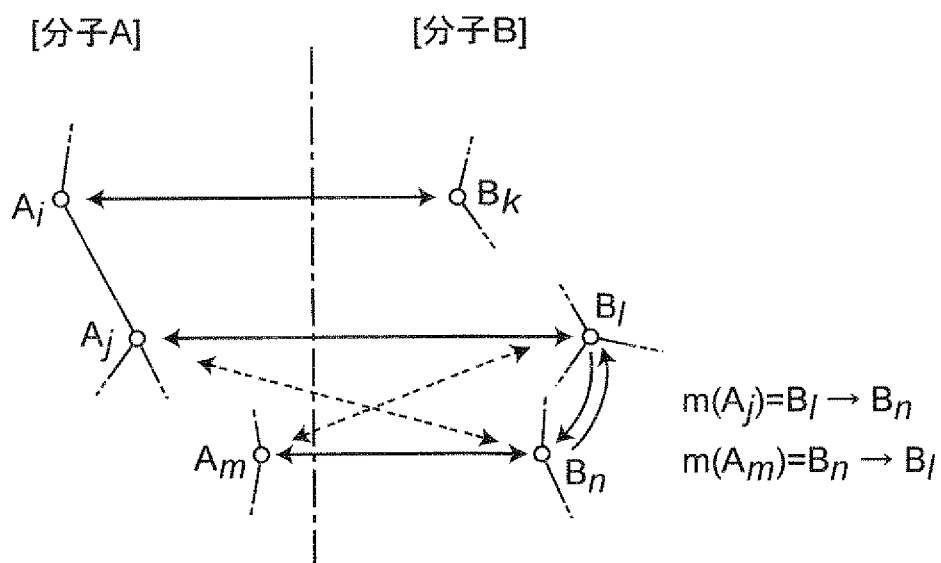
[図7]



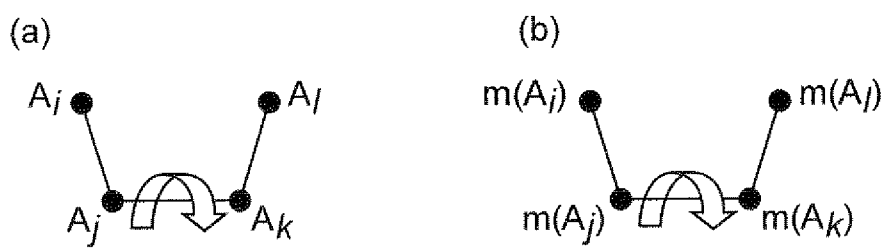
[図8]



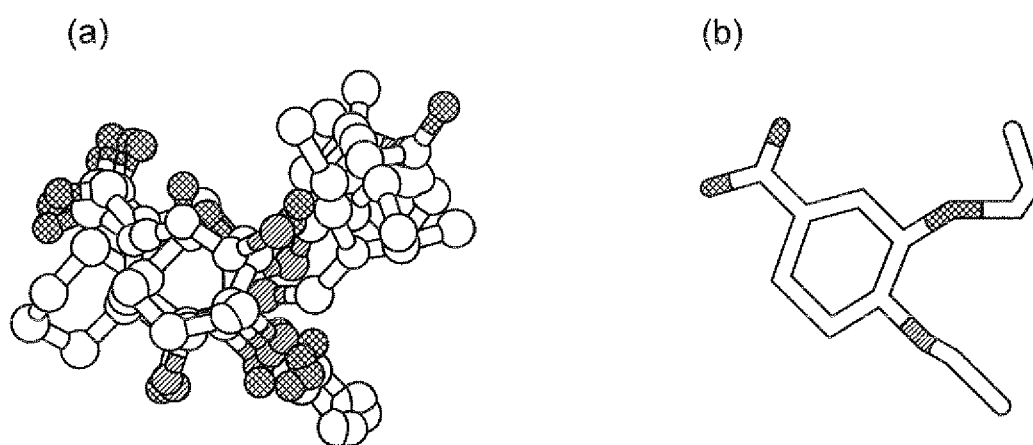
[図9]



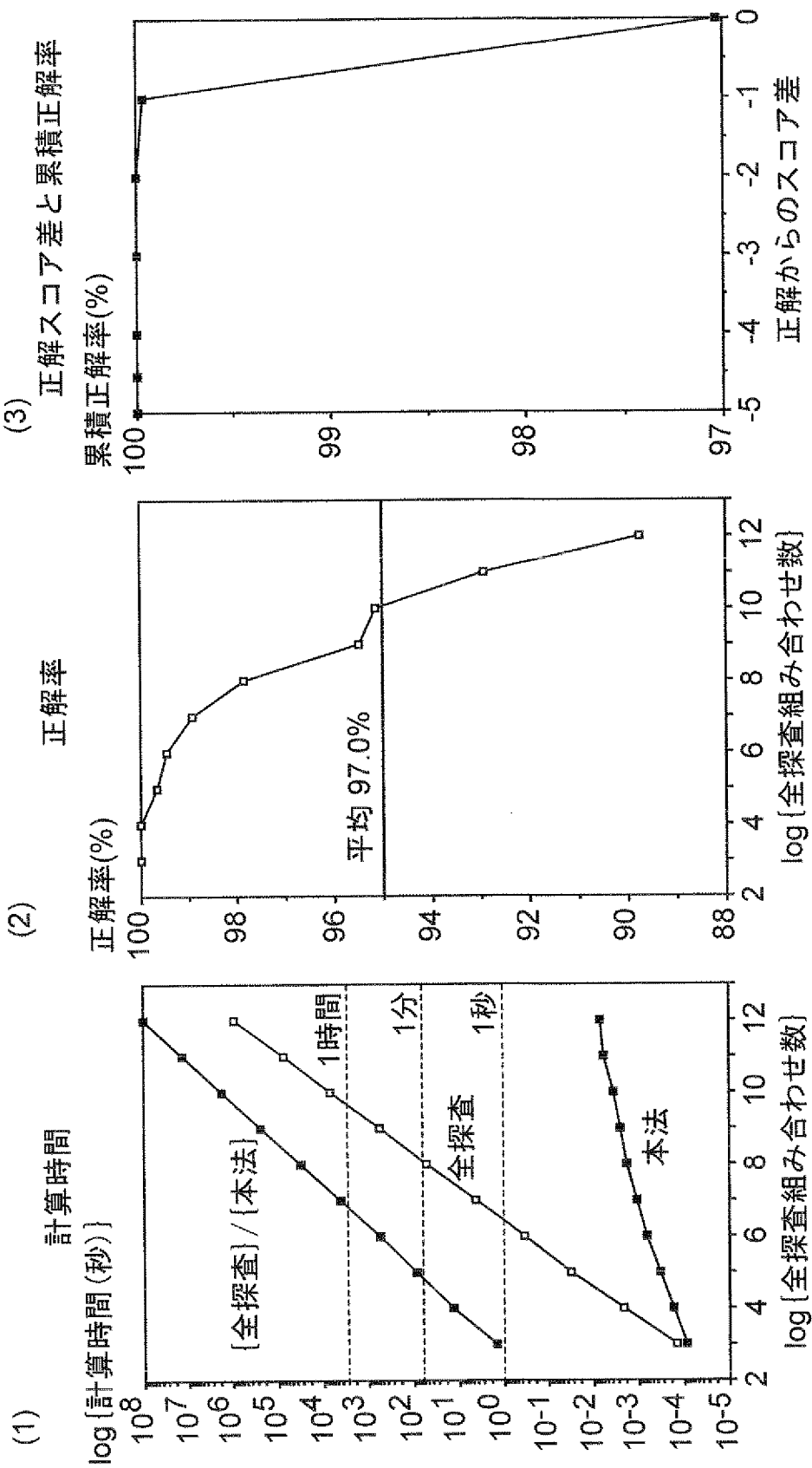
[図10]



[図11]

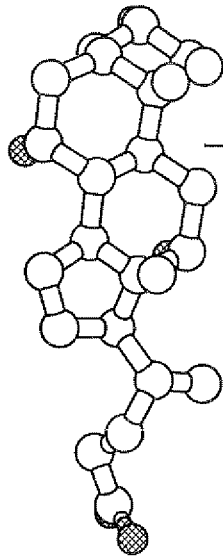


[図12]

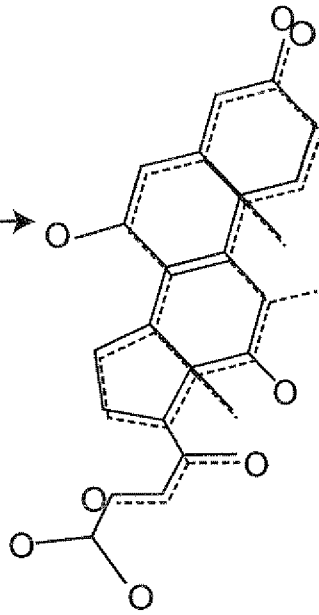
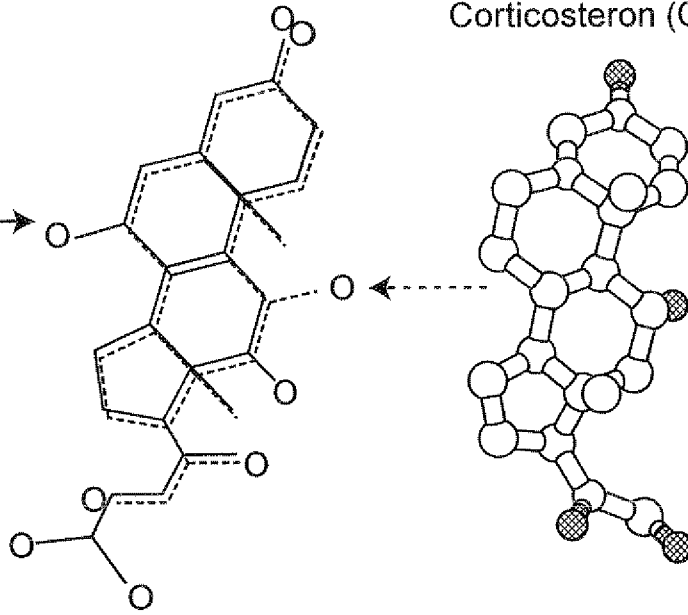


[圖13]

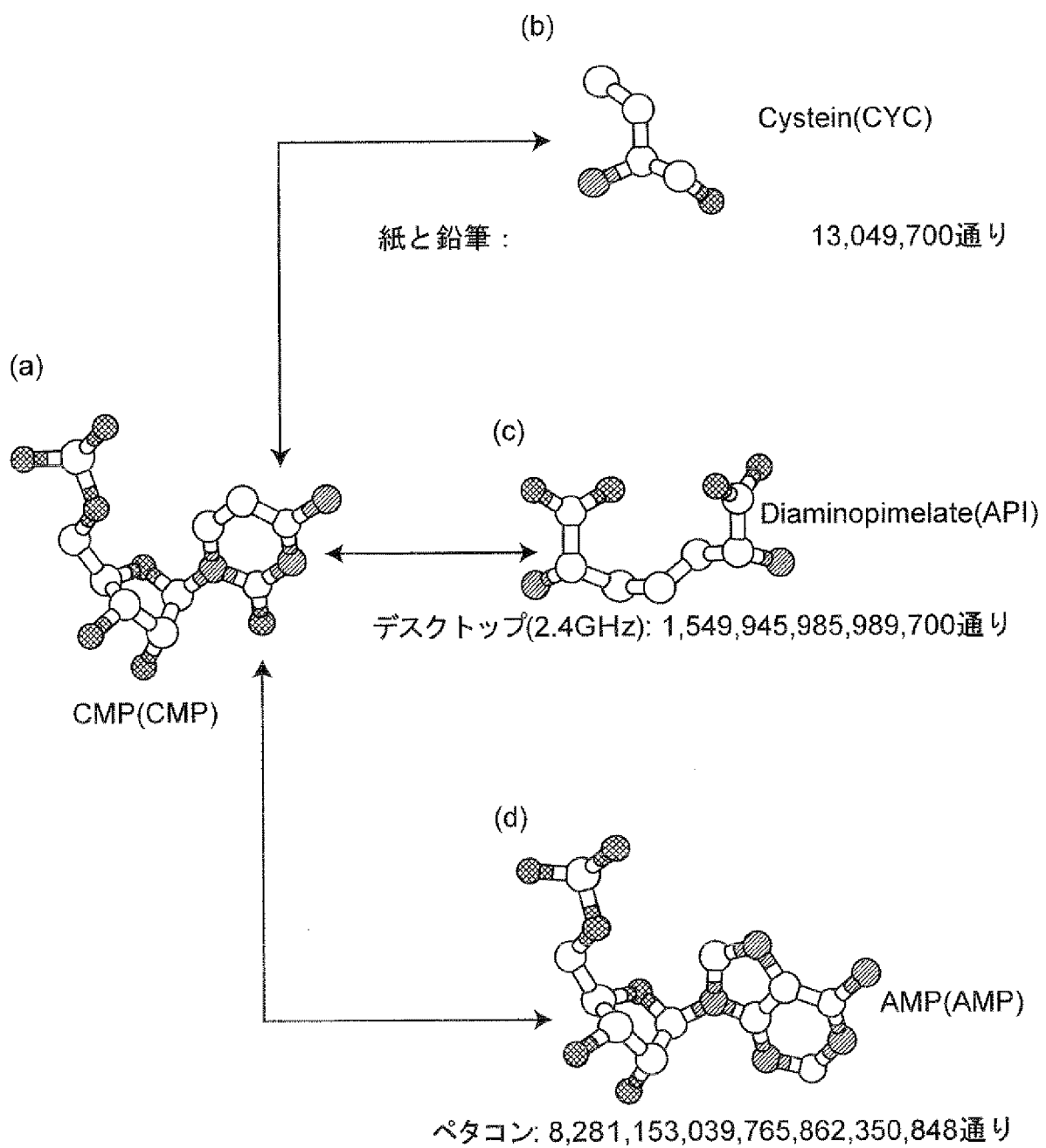
Cholic acid (CHD)



Corticosteron (COR)



[図14]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/053280

A. CLASSIFICATION OF SUBJECT MATTER

C40B30/02(2006.01)i, G06F17/30(2006.01)i, G06F17/50(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C40B30/02, G06F17/30, G06F17/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2011
Kokai Jitsuyo Shinan Koho	1971-2011	Toroku Jitsuyo Shinan Koho	1994-2011

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JSTPlus/JMEDPlus/JST7580 (JDreamII), PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-118594 A (Ishihara Sangyo Kaisha, Ltd.), 15 April 2004 (15.04.2004), entire text; all drawings (Family: none)	1-7
A	HATTORI, M., Heuristics for chemical compound matching, Genome informatics. International Conference on Genome Informatics, 2003, Vol.14, p.144-153	1-7
A	TAKAHASHI, Y., Recognition of largest common structural fragment among a variety of chemical structures, Analytical Science, 1987, Vol. 3, p. 23-28	1-7

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
16 March, 2011 (16.03.11)Date of mailing of the international search report
29 March, 2011 (29.03.11)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/053280

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RAYMOND, J.W., Maximum common subgraph isomorphism algorithms for the matching of chemical structures, Journal of computer-aided molecular design, 2002.07, Vol. 16, No. 7, p.521-533	1-7
A	KOCH, I., Enumerating all connected maximal common subgraphs in two graphs, Theoretical Computer Science, 2001.01, Vol. 250, No.1-2, p. 1-30	1-7

A. 発明の属する分野の分類 (国際特許分類 (IPC))
 Int.Cl. C40B30/02(2006.01)i, G06F17/30(2006.01)i, G06F17/50(2006.01)i

B. 調査を行った分野
 調査を行った最小限資料 (国際特許分類 (IPC))
 Int.Cl. C40B30/02, G06F17/30, G06F17/50

最小限資料以外の資料で調査を行った分野に含まれるもの
 日本国実用新案公報 1922-1996年
 日本国公開実用新案公報 1971-2011年
 日本国実用新案登録公報 1996-2011年
 日本国登録実用新案公報 1994-2011年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)
 JSTPlus/JMEDPlus/JST7580(JDreamII), PubMed

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2004-118594 A (石原産業株式会社) 2004.04.15, 全文, 全図 (ファミリーなし)	1-7
A	HATTORI, M., Heuristics for chemical compound matching, Genome informatics. International Conference on Genome Informatics, 2003, Vol.14, p.144-153	1-7

C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
 「A」特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」口頭による開示、使用、展示等に言及する文献
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願日の後に公表された文献
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」同一パテントファミリー文献

国際調査を完了した日 16.03.2011	国際調査報告の発送日 29.03.2011
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 宮久保 博幸 電話番号 03-3581-1101 内線 3562

5 L 3 1 3 6

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	TAKAHASHI, Y., Recognition of largest common structural fragment among a variety of chemical structures, Analytical Science, 1987, Vol. 3, p. 23-28	1 - 7
A	RAYMOND, J.W., Maximum common subgraph isomorphism algorithms for the matching of chemical structures, Journal of computer-aided molecular design, 2002.07, Vol. 16, No. 7, p. 521-533	1 - 7
A	KOCH, I., Enumerating all connected maximal common subgraphs in two graphs, Theoretical Computer Science, 2001.01, Vol. 250, No. 1-2, p. 1-30	1 - 7