(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5484946号
(P5484946)

(45)発行日　平成26年5月7日(2014.5.7)　(24)登録日　平成26年2月28日(2014.2.28)

| (51)Int.Cl. | | F I | | |
|---|---|---|---|---|
| G06F 19/16 | (2011.01) | G06F | 19/16 | |
| G06F 17/30 | (2006.01) | G06F | 17/30 | 170F |
| G06F 17/50 | (2006.01) | G06F | 17/30 | 350C |
| | | G06F | 17/50 | 638 |

請求項の数 7　(全 21 頁)

(21)出願番号　特願2010-31526(P2010-31526)
(22)出願日　平成22年2月16日(2010.2.16)
(65)公開番号　特開2011-170444(P2011-170444A)
(43)公開日　平成23年9月1日(2011.9.1)
審査請求日　平成25年1月22日(2013.1.22)

(73)特許権者　503303466
学校法人関西文理総合学園
滋賀県長浜市田村町１２６６番地
(74)代理人　100101454
弁理士　山田　卓二
(74)代理人　100081422
弁理士　田中　光雄
(74)代理人　100125874
弁理士　川端　純市
(72)発明者　白井　剛
滋賀県長浜市田村町１２６６番地　長浜バイオ大学内

審査官　宮久保　博幸

最終頁に続く

(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)【特許請求の範囲】
【請求項１】
第１の分子Ａを構成する原子（Ａｉ，Ａｊ，・・・）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（Ｂｋ，Ｂｌ，・・・）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（Ａｉ，Ａｊ，・・・）と第２の分子Ｂの夫々の原子（Ｂｋ，Ｂｌ，・・・）との対応付け（ｍ（Ａｉ）＝Ｂｋ）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Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）を求める第１の算出手段と、
第１の分子Ａの全ての原子Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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Ａｊ、Ｂｌの全ての組につき、第１の類似指標Ｓ１（Ａｊ，Ｂｌ）を積算して算出する第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める算出手段であって、その原子Ａｉ、Ｂｋの対の各原子から等しい結合距離にある周囲の原子Ａｊ、Ｂｌが同じ元素

であれば、更に第１の類似指標Ｓ１（Ａｊ，Ｂｌ）に係数を掛けた上で積算する、第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める第２の算出手段と、

第１の分子Ａの全ての原子Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）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（Ａｉ、Ｂｋ）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（Ａｉ，Ｂｋ）を算出した際の、始点の原子（Ａｉ，Ｂｋ）の対から開始して、未対応の原子の対の中で最大のＳ３（Ａｊ，Ｂｌ）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｛Ａｉ，Ａｊ｝の組に対応する、第２の分子Ｂの｛Ｂｋ，Ｂｌ｝において、一方を他の原子Ｂｎ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】

$$M(A,B)=\sum_{i,j}E(Ai,Aj)+\sum_{i}N(Ai)$$

【表１】

| 項 | 値 | 条件 | |
|---|---|---|---|
| Ｅ（Ａｉ，Ａｊ） | 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、<br>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、<br>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、<br>Aiとm(Ai)、Ajとm(Aj)はそれぞれ同一元素であり、<br>Ai-Aj、m(Ai)-m(Aj)は結合次数が同じである。 | 実行mode＝３の場合 |
| | 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。 | |
| Ｎ（Ａｉ） | 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「Ｓ１（Ａｉ，Ｂｋ）」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「Ｓ２（Ａｉ，Ｂｋ）」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「Ｓ３（Ａｉ、Ｂｋ）」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 | Ai-Ajと、Bk-Blとは同じ結合距離にあり、AjとBl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 |
|---|---|---|
| $s_2(A_j,B_l)$ | $S_1(A_j,B_l)$×係数 | Ai-Ajと、Bk-Blとは同じ結合距離にあり、AjとBlとは同一元素である。 |
| | $S_1(A_j,B_l)$ | Ai-Ajと、Bk-Blとは同じ結合距離にあるが、AjとBl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(Ai,Bk)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Ｓ2(Ai,Bk)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、１２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（Ａｉ，Ａｊ，・・・）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（Ｂｋ，Ｂｌ，・・・）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（Ａｉ，Ａｊ，・・・）と第２の分子Ｂの夫々の原子（Ｂｋ，Ｂｌ，・・・）との対応付け（ｍ（Ａｉ）＝Ｂｋ）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Ａｉと第１の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）を求める第１の工程と、

第１の分子Ａの全ての原子Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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Ａｊ、Ｂｌの全ての組につき、第１の類似指標Ｓ１（Ａｊ，Ｂｌ）を積算する第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める工程であって、そのＡｉ、Ｂｋの対の各原子から等しい結合距離にある周囲の原子Ａｊ、Ｂｌが同じ元素であれば、更に第１の類似指標Ｓ１（Ａｊ，Ｂｌ）に係数を掛けた上で積算する、第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める第２の工程と、

第１の分子Ａの全ての原子Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）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（Ａｉ、Ｂｋ）を求める第３の工程と、

第３の工程にて最大のＳ３（Ａｉ，Ｂｋ）を算出した際の、始点の原子（Ａｉ，Ｂｋ）の対から開始して、未対応の原子の対の中で最大のＳ３（Ａｊ，Ｂｌ）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（Ａｉ，Ａｊ，・・・）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（Ｂｋ，Ｂｌ，・・・）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（Ａｉ，Ａｊ，・・・）と第２の分子Ｂの夫々の原子（Ｂｋ，Ｂｌ，・・・）との対応付け（ｍ（Ａｉ）＝Ｂｋ）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Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）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Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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Ａｊ、Ｂｌの全ての組につき、第１の類似指標Ｓ１（Ａｊ，Ｂｌ）を積算する第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める算出ステップであって、そのＡｉ、Ｂｋの対の各原子から等しい結合距離にある周囲の原子Ａｊ、Ｂｌが同じ元素であれば、更に第１の類似指標Ｓ１（Ａｊ，Ｂｌ）に係数を掛けた上で積算する、第２の類似指標Ｓ２（Ａｉ、Ｂｋ）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Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）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Ｓ２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Ｓ３（Ａｉ、Ｂｋ）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Ｓ３（Ａｉ，Ｂｋ）を算出した際の、始点の原子（Ａｉ，Ｂｋ）の対から開始して、未対応の原子の対の中で最大のＳ３（Ａｊ，Ｂｌ）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（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。

【発明の詳細な説明】

【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す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Ｃｙｓｔｅｉｎ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、超高速度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。
【先行技術文献】
【特許文献】
【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。
【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 Estimation of active confirmations of drugs by a new molecular superposing 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（Ａｉ，Ａｊ，・・・）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（Ｂｋ，Ｂｌ，・・・）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（Ａｉ，Ａｊ，・・・）と第２の分子Ｂの夫々の原子（Ｂｋ，Ｂｌ，・・・）との対応付け（ｍ（Ａｉ）＝Ｂｋ）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Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）を求める第１の算出手段と、

第１の分子Ａの全ての原子Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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Ａｊ、Ｂｌの全ての組につき、第１の類似指標Ｓ１（Ａｊ，Ｂｌ）を積算して算出する第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める算出手段であって、その原子Ａｉ、Ｂｋの対の各原子から等しい結合距離にある周囲の原子Ａｊ、Ｂｌが同じ元素であれば、更に第１の類似指標Ｓ１（Ａｊ，Ｂｌ）に係数を掛けた上で積算する、第２の類似指標Ｓ２（Ａｉ、Ｂｋ）を求める第２の算出手段と、

第１の分子Ａの全ての原子Ａｉと第２の分子Ｂの全ての原子Ｂｋとで形成される、原子Ａｉと原子Ｂｋの組の全てに関して、原子Ａｉ、Ｂｋ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（Ａｉ、Ｂｋ）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（Ａｉ、Ｂｋ）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（Ａｉ，Ｂｋ）を算出した際の、始点の原子（Ａｉ，Ｂｋ）の対から開始して、未対応の原子の対の中で最大のＳ３（Ａｊ，Ｂｌ）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（Ａｉ）｝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Ａｉと原子Ｂｋに関する指標Ｓ１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Ａｉと原子Ｂｋに関する指標Ｓ２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Ａｉと原子Ｂｋに関する指標Ｓ３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Ｓ１００８におけるグラフマッチスコアＭ０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Ｓ１０１２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｛Ａｉ｝に対応した｛ｍ（Ａｉ）｝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（Ａｉ）から分子Ｂ（Ｂｋ）への対応関係｛ｍ（Ａｉ）｝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（Ａｉ）｝、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（Ａｉ））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２．００Å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。

【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}＞1.44Å | 単結合 |
| | 1.44Å≧{原子間距離}＞1.25Å | 二重結合 |
| | 1.25Å≧{原子間距離} | 三重結合 |
| 炭素（Ｃ）－窒素（Ｎ） | {原子間距離}＞1.32Å | 単結合 |
| | 1.32Å≧{原子間距離}＞1.21Å | 二重結合 |
| | 1.21Å≧{原子間距離} | 三重結合 |
| 炭素（Ｃ）－酸素（Ｏ） | {原子間距離}＞1.25Å | 単結合 |
| | 1.25Å≧{原子間距離} | 二重結合 |
| 窒素（Ｎ）－窒素（Ｎ） | {原子間距離}＞1.30Å | 単結合 |
| | 1.30Å≧{原子間距離} | 二重結合 |
| 酸素（Ｏ）－酸素（Ｏ） | {原子間距離}＞1.25Å | 単結合 |
| | 1.25Å≧{原子間距離} | 二重結合 |
| その他の結合 | | 単結合 |

【００２５】

（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」であると設定する。

【００２６】

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。

【００２７】

（定義１）；「Ａｉ」は、分子Ａのｉ番目の原子であることを示す。「Ａｉ－Ａｊ」は、ＡｉとＡｊの結合を示す。

【００２８】

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

【００２９】

（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

【数１】

$$M(A,B)=\sum_{i,j}E(Ai,Aj)+\sum_{i}N(Ai)$$

数１の各項は、以下の通り定義される。なおＥ（Ａｉ，Ａｊ）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。

【表２】

| 項 | 値 | 条件 | |
|---|---|---|---|
| Ｅ（Ａｉ，Ａｊ） | 1 | Ai-Ajと、m(Ai)-m(Aj)とが、それぞれ直接結合する。 | 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、<br>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、<br>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、<br>Aiとm(Ai)、Ajとm(Aj)はそれぞれ同一元素であり、<br>Ai-Aj、m(Ai)-m(Aj)は結合次数が同じである。 | 実行mode＝３の場合 |
| | 0 | Ai-Ajと、m(Ai)-m(Aj)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（Ａｉ） | 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。 | |

【００３０】

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。

【００３１】

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（Ａｉ）}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。

【００３２】

[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（Ａｉ，Ｂｋ）について、以下の数２及び表３で定義される「Ｓ１（Ａｉ，Ｂｋ）」を求める。

【数２】

$$S_1(A_i,B_k)=\sum_{j}\sum_{l}s_1(A_j,B_l)$$

【表３】

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

【００３３】
Ｓ１（Ａｉ，Ｂｋ）は、原子Ａｉと原子Ｂｋ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。

【００３４】
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Ａｉから２の結合距離にある原子Ａｊと、原子Ａｉに対応する原子Ｂｋから２の結合距離にある原子Ｂｌとが同一元素であれば、ｓ１（Ａｊ，Ｂｌ）の値は“１”になる。原子Ａｉと原子Ｂｋ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Ｓ１（Ａｉ，Ｂｋ）は、対応する原子Ａｉ，Ｂｋ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。

【００３５】
[ステップＳ１００４]；次に、以下の数３及び表４で定義される「Ｓ２（Ａｉ，Ｂｋ）」を、全ての｛ｉ，ｋ｝の組について求める。

【数３】

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

【表４】

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

【００３６】
Ｓ２（Ａｉ，Ｂｋ）は、対応する原子Ａｉ、Ｂｋ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Ａｊ、Ｂｌ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Ｓ１（Ａｊ，Ｂｌ）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Ａｊ、Ｂｌが同じものであれば、更にＳ１（Ａｊ，Ｂｌ）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Ａｉ、Ｂｋ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。

【００３７】
例えば、図６に示される、原子Ａｉを含む分子Ａの一部、及び原子Ｂｋを含む分子Ｂの一部において、Ｓ２（Ａｉ，Ｂｋ）を検討する。原子Ａｉからある結合距離（図６では２）にある原子Ａｊと、原子Ｂｋからそれと等しい結合距離にある原子Ｂｌとの全ての対につき、ｓ２（Ａｊ，Ｂｌ）、即ちＳ１（Ａｊ，Ｂｌ）、又はＳ１（Ａｊ，Ｂｌ）×１２を積算する。特に、ＡｊとＢｌが同じ元素であれば、Ｓ１（Ａｊ，Ｂｌ）は所定数倍（ここでは１２倍）されて積算されて、Ｓ２が求められる。原子Ａｉ及び原子Ｂｋからの結合距離は、１から最大値（即ち、原子Ａｉ又は原子Ｂｋ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、Ｓ１（Ａｊ，Ｂｌ）は、原子Ａｊ，Ｂｌの対において、（図６のＡｍ、Ｂｎ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。

【００３８】
上記のＳ２（Ａｉ，Ｂｋ）では、対応する２つの原子Ａｉ，Ｂｋに関して、同じ結合距離の位置の原子の対（Ａｊ，Ｂｌ）のＳ１（Ａｊ，Ｂｌ）が積算されるが、（Ａｊ，Ｂｌ）が同じ元素であれば、Ｓ１（Ａｊ，Ｂｌ）が所定数倍（１２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Ｓ２（Ａｉ，Ｂｋ）は大きくなる。なお、係数「１２」は別の数値であってもよい。
【００３９】
[ステップＳ１００６]；次に、以下の数４及び表５で定義される「Ｓ３（Ａｉ，Ａｋ）」を、全ての｛ｉ，ｋ｝の組について求める。
【数４】

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

【表５】

| | 値条件 |
|---|---|
| $m_{0i,k}(A,B)$ | 原子(Ai,Bk)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Ｓ2(Ai,Bk)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） |

【００４０】
Ｓ３（Ａｉ，Ｂｋ）は、原子Ａｉ、Ｂｋ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Ｓ２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。
【００４１】
例えば、図７に示される、原子Ａｉを含む分子Ａの一部、及び原子Ｂｋを含む分子Ｂの一部において、Ｓ３（Ａｉ，Ｂｋ）を検討する。始点は、原子Ａｉ、Ｂｋの対である。原子Ａｉには、原子Ａｊ、Ａｐ、Ａｒが直接結合する。原子Ｂｋには、原子Ｂｌ、Ｂｑ、Ｂｓが直接結合する。｛Ａｊ、Ａｐ、Ａｒ｝と｛Ｂｌ、Ｂｑ、Ｂｓ｝とから形成され得る原子同士の（３×３＝９通りの）対のうちから、（Ａｊ、Ｂｌ）の対のＳ２が最大であるとすると、原子Ａｊと原子Ｂｌを対応付けすることになる。
【００４２】
次に、分子Ａにおいて対応付けが済んだＡｉ－Ａｊには、原子Ａｐ、Ａｒ、Ａｔ、Ａｖが直接結合する。分子Ｂにおいて対応付けが済んだＢｋ－Ｂｌには、原子Ｂｑ、Ｂｓ、Ｂｕ、Ｂｗが直接結合する。｛Ａｐ、Ａｒ、Ａｔ、Ａｖ｝と｛Ｂｑ、Ｂｓ、Ｂｕ、Ｂｗ｝とから形成され得る原子同士の（４×４＝１６通りの）対のうちから、（Ａｐ、Ｂｑ）の対のＳ２が最大であるとすると、原子Ａｐと原子Ｂｑ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Ａｉ、Ａｊ、Ａｐの対応付けが完了し、分子Ｂにおいては、原子Ｂｋ、Ｂｌ、Ｂｑの対応付けが完了する。
【００４３】
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。
【００４４】
上記のＳ３（Ａｉ，Ｂｋ）では、全ての｛Ａｉ，Ｂｋ｝の組み合わせの各々において、原子の対の始点｛Ａｉ，Ｂｋ｝の周囲から徐々に、Ｓ２（対応する原子Ａｉ、Ｂｋ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。

【００４５】

[ステップＳ１００８]；次に、ステップＳ１００６にて最大のＳ３（Ａｉ，Ｂｋ）を算出した際の、始点の原子（Ａｉ，Ｂｋ）の対応から開始して、未対応の原子の対の中で最大のＳ３（Ａｊ，Ｂｌ）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Ｍ０（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。

【００４６】

例えば、図８に示される、原子Ａｉを含む分子Ａの一部、及び原子Ｂｋ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×ｂ通りの）対のうち、原子Ａｉ、Ｂｋの対において、（ステップＳ１００６で求めた）Ｓ３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Ａｉ、Ｂｋの対が対応付けされる。

次に、Ａｉを除いた分子Ａを構成する（ａ－１）個の原子と、Ｂｋ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）×（ｂ－１）通りの対のうち、原子Ａｊ、Ｂｌの対において、Ｓ３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Ａｊ、Ｂｌの対が対応付けされる。このとき、ＡｊはＡｉと直接結合しているとは限らず、ＢｌはＢｊ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）。

【００４７】

次に、ＡｉとＡｊを除いた分子Ａを構成する（ａ－２）個の原子と、ＢｋとＢｌ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）×（ｂ－２）通りの対のうち、原子Ａｊ２、Ｂｌ２の対において、Ｓ３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Ａｊ２、Ｂｌ２の対が対応付けされる。

更に次に、ＡｉとＡｊとＡｊ２を除いた分子Ａを構成する（ａ－３）個の原子と、ＢｋとＢｌとＢｌ２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）×（ｂ－３）通りの対のうち、原子Ａｊ３、Ｂｌ３の対において、Ｓ３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Ａｊ３、Ｂｌ３の対が対応付けされる。

更に次に、ＡｉとＡｊとＡｊ２とＡｊ３を除いた分子Ａを構成する（ａ－４）個の原子と、ＢｋとＢｌとＢｌ２とＢｌ３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）×（ｂ－４）通りの対のうち、原子Ａｊ４、Ｂｌ４の対において、Ｓ３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Ａｊ４、Ｂｌ４の対が対応付けされる。

【００４８】

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Ｍ０を求める。

【００４９】

ステップＳ１００８では、ステップＳ１００６で求めた多数の（例えば、ａ×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。

【００５０】

[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のかどうかを確認するために行われる。

【００５１】

等しければ（ステップＳ１０１０・Ｙｅｓ）、ステップＳ１０１６にて原子対応｛ｍ（Ａｉ）｝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。

【００５２】

[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。

【００５３】

分子Ａにおけるひとつの結合した｛Ａｉ，Ａｊ｝の組に対応する、分子Ｂの｛Ｂｋ，Ｂｌ｝において、一方を他の原子Ｂｎ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Ｂｎ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。

【００５４】

例えば、図９に示される、原子Ａｉ、Ａｊ、Ａｍを含む分子Ａの一部、及び、原子Ｂｋ、Ｂｌ、Ｂｎ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。ＡｉとＢｋ、ＡｊとＢｌ、及び、ＡｍとＢｎが、対応付けられており、ＡｉとＡｊが結合しているとする。ここで、｛Ｂｋ、Ｂｌ｝のうちの一方であるＢｌと、Ｂｎとを入れ換え、ＡｊとＢｎを対応付け、同時に、ＡｍとＢｌを対応付ける。即ち、ｍ（Ａｊ）＝Ｂｌであったものをｍ（Ａｊ）＝Ｂｎとし、ｍ（Ａｍ）＝Ｂｎであったものをｍ（Ａｍ）＝Ｂｌ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Ｍ１（Ａ，Ｂ）を求める。

【００５５】

図９の例における原子Ｂｋを（図示しない）Ｂｐ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。

【００５６】

[ステップＳ１０１４]；算出したＭ１（Ａ，Ｂ）が、Ｍ０（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Ｍ１（Ａ，Ｂ）の変動が確認される。算出したＭ１（Ａ，Ｂ）が、Ｍ０（Ａ，Ｂ）より大きければ（ステップＳ１０１４・Ｙｅｓ）、Ｍ１（Ａ，Ｂ）の値がＭ０（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Ｍ１（Ａ，Ｂ）が求められる。

【００５７】

[ステップＳ１０１６]；算出したＭ１（Ａ，Ｂ）が、Ｍ０（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Ｍ０（Ａ、Ｂ）を出力して終了する。

【００５８】

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｛Ａｉ｝に対応した原子｛ｍ（Ａｉ）｝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。

【００５９】

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｛Ａｉ，Ａｊ，Ａｋ，Ａｌ｝が、同様に結合した分子Ｂの原子｛ｍ（Ａｉ），ｍ（Ａｊ），ｍ（Ａｋ），ｍ（Ａｌ）｝に対応し、かつ、結合Ａｊ－Ａｋと、ｍ（Ａｊ）－ｍ（Ａｋ）がいずれも回転可能結合であれば、分子Ｂのねじれ角｛ｍ（Ａｉ），ｍ（Ａｊ），ｍ（Ａｋ），　ｍ（Ａｌ）｝を、分子Ａの

対応するねじれ角｛Ａｉ，Ａｊ，Ａｋ，Ａｌ｝と同値にする。

【００６０】

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。

【００６１】

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２．４ＧＨｚ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。

【００６２】

以下の表６で、上述の計算によりクエリの原子｛Ａｉ｝に対応した｛ｍ（Ａｉ）｝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。

【表６】

| 分子 | 原子数 | スコア | 想定最大値 | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G39 | 20 | 40 | 40 | C1 | O1 | O1 | C2 | C3 | C4 | C5 | N5 | C1 | O1 | C1 | C6 | C7 | O7 | C8 | C9 | C8 | C8 | C9 | N4 |
| FDI | 21 | 38 | 42 | C1 | O1 | O1 | C5 | C4 | C3 | C2 | N3 | C3 | O3 | C3 | C1 | C6 | O1 | C1 | C1 | C2 | C2 | C1 | N1 C1 |
| 49A | 20 | 36 | 40 | C1 | O1 | O1 | C2 | C3 | C4 | C5 | N5 | C1 | O1 | C1 | C6 | O6 | O8 | C8 | C9 | C7 | O7 | N9 | N4 |
| ST6 | 18 | 34 | 36 | C1 | O1 | O2 | C1 | C2 | C3 | C4 | N4 | C4 | O4 | CM | C5 | C6 | O3 | C3 | CM | -- | -- | N3 | N3 |
| IEM | 19 | 35 | 38 | C1 | O1 | O1 | C2 | C3 | C4 | C5 | N5 | C1 | O1 | C1 | C6 | O6 | O7 | C7 | C8 | -- | C1 | C9 | O4 |
| G39 | 20 | 40 | 40 | C1 | O1 | O1 | C2 | C3 | C4 | C5 | N5 | C1 | O1 | C1 | C6 | C7 | O7 | C8 | C9 | C8 | C8 | C9 | N4 |
| OX3 | 18 | 34 | 37 | C1 | O1 | O9 | C1 | C1 | C1 | C1 | N8 | C7 | -- | C4 | C1 | C1 | O1 | C1 | C2 | C6 | C5 | C3 | - |
| 共通骨格 | | | | ** | ** | ** | ** | ** | ** | ** | ** | ** | | ** | ** | | ** | ** | ** | | | ++ | |

【００６３】

図１１は、クエリの原子クエリの原子｛Ａｉ｝に対応した｛ｍ（Ａｉ）｝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。

【００６４】

７．アルゴリズム性能評価

本実施形態に係る高速グラフマッチ探索アルゴリズムの性能評価を行った。

【００６５】

（７．１　アルゴリズム性能評価（１））

本実施形態は、多項式時間アルゴリズム未知のＮＰ困難問題に近似解を与えるものである。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(＝正解)を求め、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解と比較を行った。図１２（１）は、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であり、下方から本実施形態による計算時間、全探査による計算時間、及び、本実施形態による計算時間に対する全探査による計算時間の比を示している。本実施形態に係るアルゴリズムは、グラ

フの探査範囲では$10^{-4}$～$10^{-3}$秒で計算が可能である。全探査による場合は、$10^{-4}$～$10^{6}$秒を要するものである。

【００６６】

図１２（２）は、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である。グラフの探査範囲では、平均９７％の割合で正解を発見した。更に、図１２（３）は、正解スコア差と累積正解率の関係をグラフ化したものである。誤答した場合でも、正解とのスコア差は最大２点であった。これら図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。

【００６７】

（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 C. & 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。

【００６８】

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。

【００６９】

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。

【表７】

実行時間差とスコア差｛Δ（本実施形態－simcomp）｝の平均

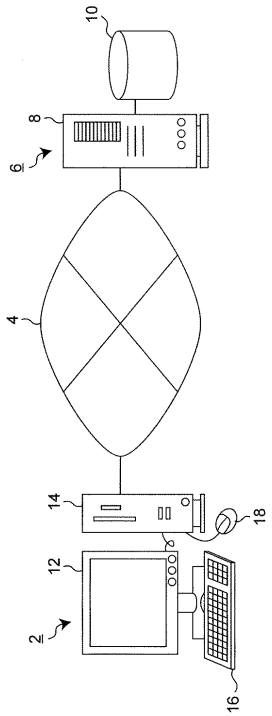
| Ｒｍａｘ | $<\Delta t_{sec}>_{all}$ | $<\Delta S>_{all}$ | $<\Delta t_{sec}>_{partial}$ | $<\Delta S>_{partial}$ |
|---|---|---|---|---|
| $1.5\times10^{4}$ | 0.096 | 2.43 | -0.048 | 0.16 |
| $1.5\times10^{5}$ | -0.127 | 2.13 | -0.505 | 0.04 |
| $1.5\times10^{6}$ | -1.612 | 2.00 | -4.779 | 0.01 |
| $1.5\times10^{7}$ | -12.912 | 1.98 | -23.020 | 0.01 |
| $1.5\times10^{8}$ | -95.416 | 1.94 | -95.723 | 0.01 |

【符号の説明】

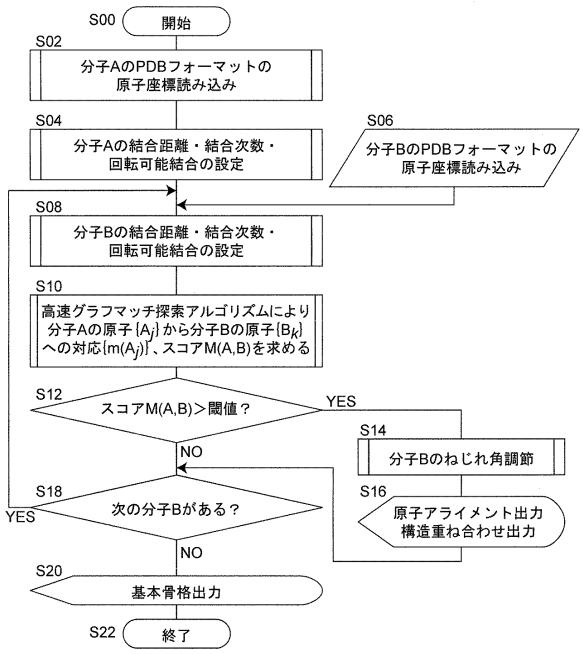
【００７０】

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。

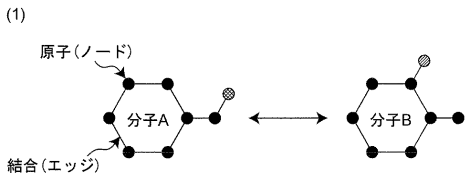
【図１】


10
8
6
4
14
18
12
2
16

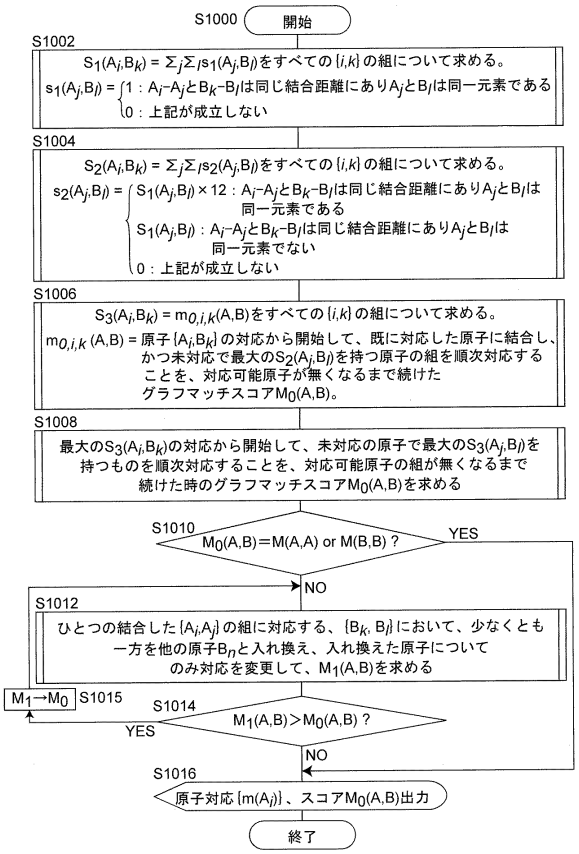
【図２】


S00
開始
S02
分子AのPDBフォーマットの
原子座標読み込み
S04
分子Aの結合距離・結合次数・
回転可能結合の設定
S06
分子BのPDBフォーマットの
原子座標読み込み
S08
分子Bの結合距離・結合次数・
回転可能結合の設定
S10
高速グラフマッチ探索アルゴリズムにより
分子Aの原子{Aj}から分子Bの原子{Bk}
への対応{m(Aj)}、スコアM(A,B)を求める
S12
スコアM(A,B)＞閾値？
YES
NO
S14
分子Bのねじれ角調節
S16
原子アライメント出力
構造重ね合わせ出力
S18
次の分子Bがある？
YES
NO
S20
基本骨格出力
S22
終了

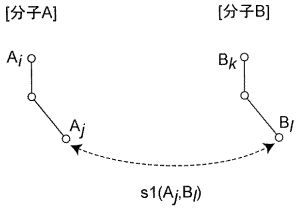
【図３】


(1)
原子（ノード）
分子A
分子B
結合（エッジ）
(2)
M(A,A) = 16,
M(B,B) = 16,
M(A,B) = 14

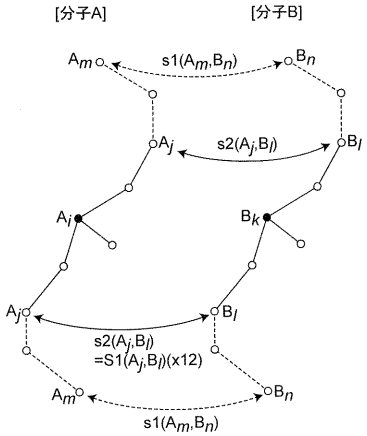
【図４】


S1000
開始
S1002
S1(Ai,Bk) = ΣjΣlS1(Aj,Bl)をすべての{i,k}の組について求める。
s1(Aj,Bl) = 1：Ai−AjとBk−Blは同じ結合距離にありAjとBlは同一元素である
0：上記が成立しない
S1004
S2(Ai,Bk) = ΣjΣlS2(Aj,Bl)をすべての{i,k}の組について求める。
s2(Aj,Bl) = S1(Aj,Bl)×12：Ai−AjとBk−Blは同じ結合距離にありAjとBlは同一元素である
S1(Aj,Bl)：Ai−AjとBk−Blは同じ結合距離にありAjとBlは同一元素でない
0：上記が成立しない
S1006
S3(Ai,Bk) = m0,i,k(A,B)をすべての{i,k}の組について求める。
m0,i,k (A,B) = 原子{Ai,Bk}の対応から開始して、既に対応した原子に結合し、かつ未対応で最大のS2(Aj,Bl)を持つ原子の組を順次対応することを、対応可能原子が無くなるまで続けたグラフマッチスコアM0(A,B)。
S1008
最大のS3(Ai,Bk)の対応から開始して、未対応の原子で最大のS3(Aj,Bl)を持つものを順次対応することを、対応可能原子の組が無くなるまで続けた時のグラフマッチスコアM0(A,B)を求める
S1010
M0(A,B)＝M(A,A) or M(B,B)？
YES
NO
S1012
ひとつの結合した{Ai,Aj}の組に対応する、{Bk, Bl}において、少なくとも一方を他の原子Bnと入れ換え、入れ換えた原子についてのみ対応を変更して、M1(A,B)を求める
M1→M0
S1015
S1014
M1(A,B)＞M0(A,B)？
YES
NO
S1016
原子対応{m(Ai)}、スコアM0(A,B)出力
終了

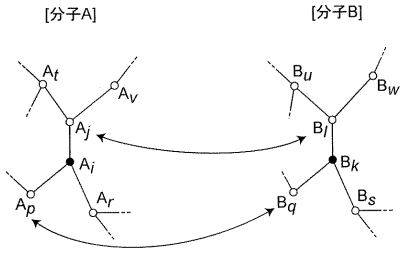
【図５】



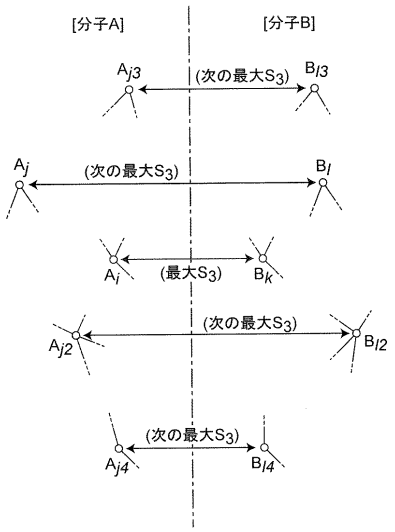

【図６】



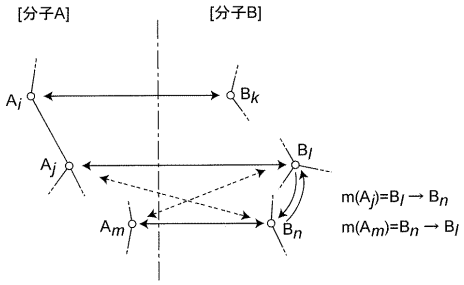

【図７】



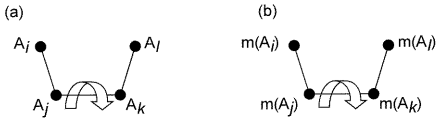

【図８】



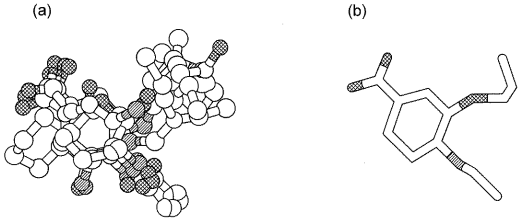

【図９】



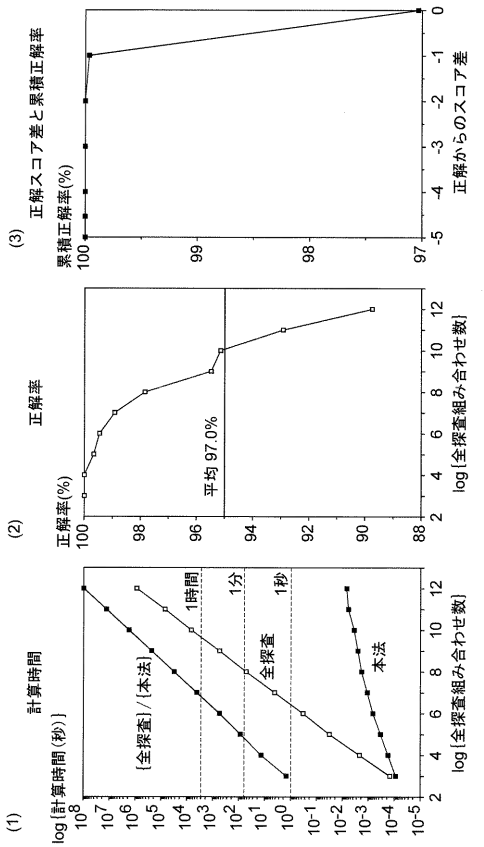

【図１０】



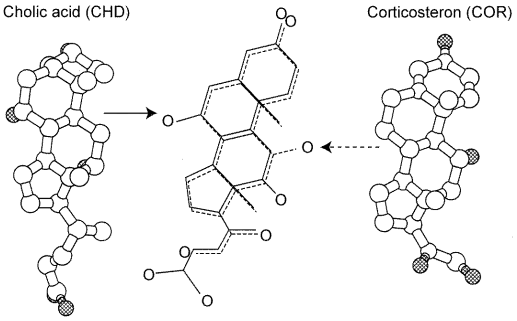

【図１１】



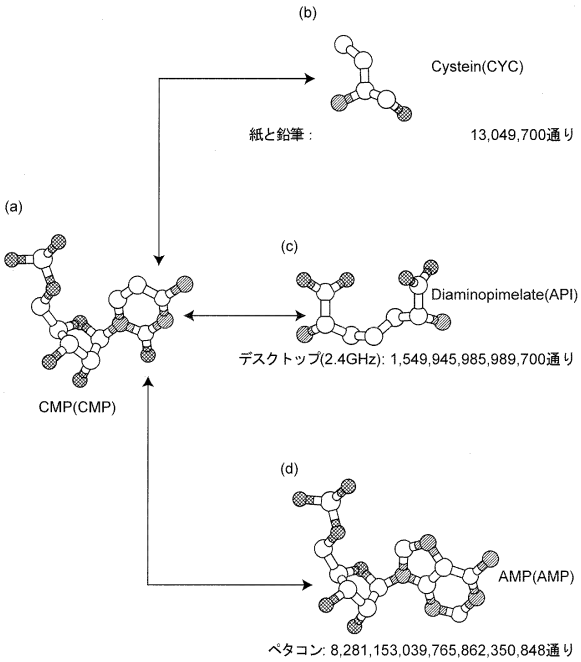

【図１２】



【図１３】


Cholic acid (CHD)
Corticosteron (COR)

【図１４】


(b)
Cystein(CYC)
紙と鉛筆 :
13,049,700通り
(a)
(c)
Diaminopimelate(API)
デスクトップ(2.4GHz): 1,549,945,985,989,700通り
CMP(CMP)
(d)
AMP(AMP)
ペタコン: 8,281,153,039,765,862,350,848通り

フロントページの続き

(56)参考文献 特開２００４－１１８５９４（ＪＰ，Ａ）
HATTORI, M., Heuristics for chemical compound matching, Genome informatics. International Conference on Genome Informatics, 日本, ２００３年, Vol.14, p.144-153
TAKAHASHI, Y., Recognition of largest common structural fragment among a variety of chemical structures, Analytical Science, １９８７年, Vol. 3, p. 23-28
RAYMOND, J.W., Maximum common subgraph isomorphism algorithms for the matching of chemical structures, Journal of computer-aided molecular design, ２００２年 7月, Vol. 16, No. 7, p. 521-533
KOCH, I., Enumerating all connected maximal common subgraphs in two graphs, Theoretical Computer Science, ２００１年 1月, Vol. 250, No. 1-2, p. 1-30

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｆ １９／１０
Ｇ０６Ｆ １７／３０
Ｇ０６Ｆ １７／５０
ＪＳＴＰｌｕｓ／ＪＭＥＤＰｌｕｓ／ＪＳＴ７５８０（ＪＤｒｅａｍＩＩＩ）
ＰｕｂＭｅｄ