

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2009年5月22日 (22.05.2009)

PCT

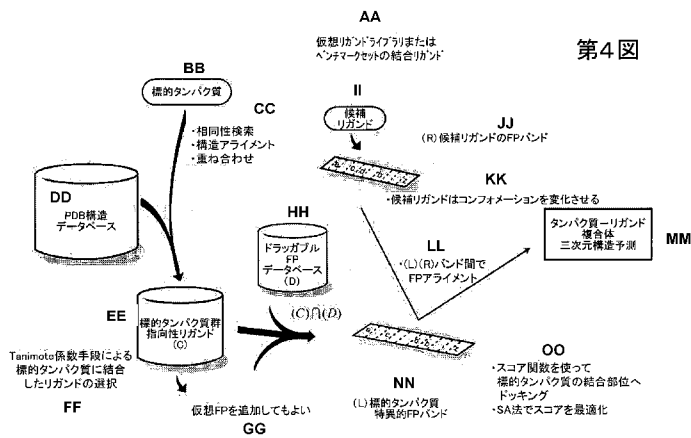
(10) 国際公開番号
WO 2009/064015 A1

- (51) 国際特許分類:
C40B 30/02 (2006.01) G06F 19/00 (2006.01)
- (21) 国際出願番号: PCT/JP2008/070973
- (22) 国際出願日: 2008年11月12日 (12.11.2008)
- (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 2007-293751
2007年11月12日 (12.11.2007) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社インシリコサイエンス (In-Silico Sciences, Inc.) [JP/JP]; 〒1450065 東京都大田区東雪谷二丁目15番9号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 梅山秀明 (UMEYAMA, Hideaki) [JP/JP]; 〒2790011 千葉県浦安市美浜1-7-1002 Chiba (JP). 高谷大輔 (TAKAYA, Daisuke) [JP/JP]; 〒2790012 千葉県浦安市入船2-3-608 Chiba (JP). 志鷹真由子 (SHITAKA, Mayuko) [JP/JP]; 〒1890001 東京都東村山市秋津町4-40-95 Tokyo (JP). 加納和彦 (KANOU, Kazuhiko) [JP/JP]; 〒1200003 東京都足立区東和2-11-16 Tokyo (JP). 寺師玄記

[続葉有]

(54) Title: IN SILICO SCREENING SYSTEM AND IN SILICO SCREENING METHOD

(54) 発明の名称: インシリコスクリーニング装置、および、インシリコスクリーニング方法



第4図

- AA BOUND LIGANDS IN VIRTUAL LIGAND LIBRARY OR BENCHMARK SET
- BB TARGET PROTEIN
- CC ·HOMOLOGY SEARCH
·STRUCTURAL ALIGNMENT
·OVERLAPPING
- DD PDB STRUCTURE DATABASE
- EE TARGET PROTEIN GROUP-DIRECTED LIGANDS (C)
- FF SELECTING LIGAND BOUND TO TARGET PROTEIN BY TANIMOTO COEFFICIENT METHOD
- GG VIRTUAL FP MAY BE ADDED
- HH DRUGGABLE FP DATABASE (D)
- II CANDIDATE LIGANDS
- JJ FP BAND OF (R) CANDIDATE LIGAND
- KK CANDIDATE LIGAND ALTERING CONFORMATION
- LL FP ALIGNMENT BETWEEN (L) AND (R) BANDS
- MM PRESUMPTION OF THREE-DIMENSIONAL STRUCTURE OF PROTEIN-LIGAND COMPLEX
- NN (L) TARGET PROTEIN-SPECIFIC FP BAND
- OO ·DOCKING INTO BINDING SITE OF TARGET PROTEIN USING SCORE FUNCTION
·OPTIMIZING SCORE BY SA METHOD

(57) Abstract: A system provided with a compound database which has been prepared by extracting the compound fingerprints with respect to multiple atoms in compounds for the individual candidate compounds, characterized by comprising extracting the compound fingerprints of binding compounds, which have been already known as binding to a family protein having the same or similar steric structure to a target protein, together with the three-dimensional coordinates converted into the coordinate system of the target protein to thereby prepare a binding compound fingerprint set, and, with respect to a candidate compound memorized in the compound database, computing the steric structure as described above of the candidate compound to the target protein so as to optimize the interaction score based on the square average deviation of the compound fingerprint unit having been computed by referring the three-dimensional coordinate of the binding compound fingerprint set as the base.

(57) 要約: 本発明は、化合物中の複数の原子に係る化合物指紋を候補化合物ごとに抽出して作成された化合物データベースを備え、標的タンパク質と立体構造が同一または類似するファミリータンパク質に結合することが既

知の結合化合物について、標的タンパク質の座標系に変換した三次元座標とともに化合物指紋を抽出して結合化合物指紋セットを作成し、化合物

[続葉有]

WO 2009/064015 A1



(TERASHI, Genki) [JP/JP]; 〒1520034 東京都目黒区
緑が丘 2 - 1 1 - 4 - 2 0 3 Tokyo (JP). 小松克一
郎 (KOMATSU, Katsuichiro) [JP/JP]; 〒1450065 東京
都大田区東雪谷 2 - 1 5 - 9 Tokyo (JP).

MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE,
SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(74) 代理人: 酒井宏明 (SAKAI, Hiroaki); 〒1006020 東京
都千代田区霞が関三丁目 2 番 5 号 霞が関ビルディ
ング 酒井国際特許事務所 Tokyo (JP).

(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, 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 (AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE,
SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

(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, 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, JP, KE, KG, KM,
KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA,

添付公開書類:
— 国際調査報告書

明 細 書

インシリコスクリーニング装置、および、インシリコスクリーニング方法

5 技術分野

この発明は、インシリコスクリーニング装置、および、インシリコスクリーニング方法に関するものである。

背景技術

10 従来、試薬提供会社等から発売される医薬品該当化合物や試薬化合物等の化合物が存在する。また、化合物と相互作用する高分子として、質量分析を主体とする各種実験等で確かめられた高分子や、例えばNatureやScienceに代表される雑誌に収録された文献などにより社会で認知された高分子等のように、創製された医薬品等の化合物と相互作用して、動植物の病気状態や疾患状態を治癒、症状軽減または現状維持等をもたらす、薬物標的タンパク質や薬物標的核酸

15 や薬物標的糖質や薬物標的脂質等の標的高分子が存在する。

標的高分子に対する低分子化合物ドッキングとインシリコスクリーニングを行うに当たっては、従来、上述のような医薬品候補化合物等の化合物の膨大な数が納められた化合物データベースの各化合物を、例えば、タンパク質を主体とする

20 標的高分子タンパク質にドッキング相互作用をさせ、現実中存在する何十万個に相当する化合物が標的タンパク質と直接相互作用する座標配置（コンフォメーション）を決定し、相互作用エネルギーやそれに相当するスコア値を獲得していた。そして、当該スコア値を安定さの指標にして大きな方から小さい方に並べ、化合物—薬物標的タンパク質の相互作用の順番を決定していた。

25 例えば、KuntzらのDock (Ewing et al著 J Comput Aided Mol Des. 2001 15 (5) 411-28参照) や、GoodsellらのAutoDock (Goodsell et a

1 著 J. Mol. Recognit 1996 9 1-5参照) や、G
a r e t h らの GOLD (J o n e s e t a l 著 J. Mol. Bio
l. 1997 267, 727-748参照)、R a r e y らの Flex X、
N i c o l a s らの Fragment Potential 等の従来方法におい
5 ては、上述のスコア値の計算のために、それぞれの方法に標的高分子である標的
タンパク質のリガンド結合環境の格子情報や、化合物と標的高分子間のベクトル
を重視する化合物の多点情報を用いて計算を行っていた。

すなわち、標的タンパク質の生物学的環境等の何らかの工夫があるにしても、
格子情報や多点情報等の情報に基づいて、化合物の原子と標的高分子タンパク質
10 を構成する原子との古典物理学的原子間ポテンシャル式から相互作用エネルギー
等を計算して、化合物のコンフォメーションや相互作用の結合の強さに関する
順番をスコア値で決定していた。また、相互作用の順番を決めるために相互作用
している種々の化合物のコンフォメーションをクラスタリング等の手法を用いて
順番を決める工夫等を行っていた。

15 しかしながら、従来のインシリコスクリーニング方法においては、タンパク質
ーリガンド複合体を精度よく予測することに注目されており、直接、ヒットする
化合物を数多く選出することとは一致しないという問題点を有していた。

また、従来のインシリコスクリーニング方法においては、古典物理学的なポテ
ンシャル関数を用いて非経験的な予測を行っており、生物化学的な実験等の情報
20 を考慮に入れた予測効率の高いスクリーニングができないという問題があった。

本発明は、上記に鑑みてなされたもので、タンパク質と化合物との結合を精度
よく予測することができる一方で、ヒットする化合物を数多く選出することがで
き、また、予測効率を高めることができる、インシリコスクリーニング装置、お
よび、インシリコスクリーニング方法を提供することを目的とする。

25

発明の開示

本願発明者は、X線解析、NMR、電子線解析、高分解能電子顕微鏡写真等の

実験によって得られた、化合物と標的高分子との相互作用を示す膨大な三次元座標情報が公開データベースに登録されていることや、近年のコンピュータの性能向上とバイオインフォマティクスの進歩等に鑑み、従来のような一般的で古典物理学的なインシリコスクリーニング方法を行う代わりに、標的高分子タンパク質に結合した種々の化合物の集団的重なり状態等のバイオインフォマティクス情報を利用して、人の叡智を基礎にした半経験的な化合物のインシリコスクリーニングを実行することが可能であるとの着想を得た。

本発明は、上記着想に基づいて本願発明者により鋭意検討された結果、完成したものであり、標的タンパク質に結合する候補化合物のスクリーニングを行う、記憶部と制御部を少なくとも備えたインシリコスクリーニング装置であって、上記記憶部は、化合物中の複数個の原子に係る化合物指紋として、原子タイプと原子間結合規則とを含む化学記述子を、上記候補化合物ごとに抽出して作成された化合物データベース、を備え、上記制御部は、上記標的タンパク質と立体構造が同一または類似するファミリータンパク質に結合することが既知の結合化合物について、上記標的タンパク質の座標系に変換した三次元座標とともに上記化合物指紋を抽出して結合化合物指紋セットを作成する化合物指紋作成手段と、上記化合物データベースに記憶された上記候補化合物について、上記結合化合物指紋セットの上記三次元座標を基底として算出した上記化合物指紋単位の二乗平均偏差を基礎とする相互作用スコアが最適化されるように、当該候補化合物の上記標的タンパク質に対する上記立体構造を演算する最適化手段と、を備えたことを特徴とする。

すなわち、本発明によれば、タンパク質と化合物との結合を精度よく予測することができる一方で、ヒットする化合物を数多く選出することができ、また、生物化学的な実験等の情報を考慮に入れた半経験的なスクリーニングを行うことができ、さらに、予測効率を高めることができる。

以上のように、本発明は、三次元の化合物指紋セットを用いるバイオインフォマティクス技術を、古典物理学的エネルギー手法を用いた低分子化合物と高分子

タンパク質とのドッキングと同等の性能を発揮させるようにした点で従来手法とは異なっている。特に、X線解析、NMR、電子線解析、高分解能電子顕微鏡解析などの技術が格段に進歩していることを考えると、標的高分子タンパク質に結合した化合物の分子の数は膨大に増加すると予測されるため、本発明は高い効果を発揮する。

また、本発明は、上記記載のインシリコスクリーニング装置において、化合物に結合したタンパク質の立体構造およびアミノ酸配列を記憶するタンパク質データベース装置に接続され、上記制御部は、上記標的タンパク質の上記アミノ酸配列との相同性に基づいて、上記ファミリータンパク質および上記結合化合物を上記タンパク質データベース装置から検索する相同性検索手段、を更に備え、上記化合物指紋作成手段は、上記相同性検索手段により検索された上記ファミリータンパク質に結合する上記結合化合物について、上記標的タンパク質の座標系に変換した上記三次元座標とともに上記化合物指紋を抽出して上記結合化合物指紋セットを作成すること、を特徴とする。

ここで、本発明の一例として具体例を示すと、本発明は、標的高分子の中でも、標的タンパク質の立体構造に類似しているファミリー高分子セットに種々の低分子化合物が結合した集団的コンフォメーションを抽出するときの条件として、ファミリー高分子セットを取り出すときに、当該標的タンパク質の配列を照会（クエリー）配列として、P S I - B l a s t 等による相同性（H o m o l o g y）検索によって検出する。そして、本発明は、検出されたタンパク質の中で、該当するとして検索され、タンパク質-リガンド複合体（P r o t e i n - L i g a n d c o m p l e x）として低分子リガンドを含んでいた場合、C E（原子の種類を意識しないタンパク質同士の構造の重ね合わせ操作）等を用いて、標的タンパク質に重ね合わせる。そして、本発明は、その構造の類似性を表す Z - S c o r e が所定の値（例えば 3.7 以上）となった場合、検索された類似タンパク質に結合したリガンドを類似タンパク質の座標系から標的タンパク質の座標系にリガンド座標と共に変換して、リガンドだけ抜き出すことができるようになる。

ここで、CEは、原子の種類を意識しないタンパク質同士の構造の重ね合わせ操作を行うが、同様の機能を持つプログラムでも代用可能である。また、本発明は、当該標的タンパク質の配列を照会（クエリー）配列として、PSI-Blast等による相同性検索によって高いホモロジーを持つ配列のみが得られた場合は、原子の種類を意識したタンパク質同士の構造の重ね合わせ操作のプログラムを使用してもよい。また、本発明は、相同性検索において、PSI-Blastに限らず、配列をクエリーとして相同性検索ができ、その配列類似性の評価を定量的にできるソフトプログラムなら、どのような相同性検索プログラムを適用してもよい。

10 また、本発明は、上記記載のインシリコスクリーニング装置において、上記化合物指紋作成手段は、上記ファミリータンパク質と上記標的タンパク質との構造重ね合わせにより、当該ファミリータンパク質に結合する上記結合化合物の上記三次元座標を上記標的タンパク質の座標系に変換し、変換された上記三次元座標とともに上記化合物指紋を抽出して上記結合化合物指紋セットを作成すること、
15 を特徴とする。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記化合物指紋作成手段は、上記結合化合物と異なる他の上記化合物を参照して構造重ね合わせを行い、当該結合化合物と当該他の上記化合物の原子間をまたがる上記化合物指紋を抽出して上記結合化合物指紋セットに追加する新規化合物指紋追加手段、を更に備えたことを特徴とする。

本発明の一例として具体例を示すと、結合化合物指紋セットの具体例としては、標的高分子の中でも標的タンパク質の立体構造に類似しているファミリー高分子タンパク質セットに結合した種々低分子化合物データベースである「CElib」(FP (fingerprint) set extracted from collected ligands in the binding site (結合部位のリガンド集合から抽出された化合物指紋セット))として構成してもよい。このCElibには、標的タンパク質の座標系における座標とSyby

l原子タイプ (a t o m - t y p e) および、単結合、二重結合、芳香環結合等といった結合規則情報を含んでいる。ここで、本発明は、標的タンパク質に対する低分子化合物の探索の狙いの必要に応じて、C E l i b に任意の F P (f i n g e r p r i n t : 「化合物指紋」のことをいう。以下に同じ。) を加えても良い。

すなわち、既存の標的タンパク質の立体構造に類似しているファミリー高分子セットに集団的に結合した種々低分子化合物から F P を抽出する代わりに、本発明では、普通の一般に存在する化合物分子と F P の類似性を保持したまま、種々低分子化合物の中で原子の種類を入れ替える。そして、本発明は、「c i r c l e」等のような安定性を評価できるプログラムを用いて標的タンパク質との相互作用エネルギーを計算し、相互作用をより安定にする少し構造の違った “M o d i f i e d F P” (改正 F P) を得る。そして、本発明は、標的タンパク質に対して局所エネルギー的に安定な改正 F P を使って、あたかもタンパク質同士の構造の重ね合わせ操作の結果として得られた集団的に結合した種々低分子化合物から得た F P のように捕らえ、それを新たな F P として、上述の発明において行われたように、F P の重ね合わせにその対象 F P として採用する。

上記発明では、タンパク質とリガンドとのドッキングにおいて、従来使用されてきた物理化学的相互作用関数の代わりに、三次元座標を含む化合物指紋セットというバイオインフォマティクスを用いたリガンドコンフォメーションを得る。そして、本発明では、既存の標的タンパク質の立体構造に類似しているファミリー高分子タンパク質セットに集団的に結合した種々低分子化合物から F P を抽出する代わりに、種々低分子化合物の中で違う分子化合物を参照して、普通の一般に存在する分子の F P に似た複数化合物結合三次元化合物指紋セットを創作する。そして、本発明は、創作した化合物指紋セットを、あたかもタンパク質同士の構造の重ね合わせ操作の結果として得られた集団的に結合した種々低分子化合物から得た F P のように捕らえ、それを新たな F P として、上記発明において行われたように、F P の重ね合わせにその対象 F P を採用する。

すなわち、上記発明は、ファミリー高分子セットに集団的に結合した種々低分子化合物を完全に分解して、従来なら物理学的公式が基底となるドッキング計算の代わりに、ばらばらにした種々低分子化合物FPをドッキングの基底としたものである。本発明は、さらに、既存の標的タンパク質の立体構造に類似しているファミリー高分子タンパク質セットに集団的に結合した種々低分子化合物のコンフォメーションの存在は標的タンパク質のファミリータンパク質と相互作用した最安定構造に近いという事実の熟慮から生まれたものであり、従来手法と異なり高い効果を有し有用である。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記化合物指紋作成手段は、タニモト係数に基づき上記結合化合物と類似する上記化合物について、当該結合化合物と当該化合物の原子間で原子の種類を入れ替え、上記標的タンパク質に対する相互作用エネルギーを算出して当該結合化合物の上記化合物指紋よりも局所エネルギー的に安定な上記化合物指紋を作成して上記結合化合物指紋セットに追加する新規化合物指紋追加手段、を更に備えたことを特徴とする。

本発明の一例として具体例を示すと、本発明は、CElibから、標的タンパク質ファミリー高分子タンパク質セットと結合した種々低分子化合物について、各々のファミリー高分子タンパク質とリガンドの複合体を対象にして、リガンドとの相互作用が安定になるように、Circleプログラム等の相互作用算出プログラムを利用する。本発明は、fingerprint (fp) 単位、すなわち化学記述子単位に原子の種類や結合の種類を改良変更し、それを新たなfingerprint (fp) 単位、即ち化学記述子単位として、それを新たなFPとして、上記発明において行われたように、FPの重ね合わせにその対象FPとして採用するものとする。

また、標的高分子の中でも、標的タンパク質の立体構造に類似しているファミリー高分子タンパク質セットに結合した種々低分子化合物データベースであるCElibのFPがドッキングscore (スコア) を決めるのに大きな貢献をし

ている。そこで、本発明では、上記発明において、標的高分子タンパク質に結合する理想的な低分子リガンドのドッキング構造が実験的に解析済みである場合、その結合に理想的な低分子リガンドをリード化合物として、相互作用エネルギーがよくなるようにいろいろな置換基を付加したり、理想的な低分子リガンドに化合物指紋の定量化関数であるタニモト係数が非常に似た、すなわち1に近い任意の低分子リガンドを見つけたりする場合に、FP領域をその実験的に解析済みの理想的な低分子リガンドの周りの領域（例えば4または5オングストローム）に限定する。これにより、本発明は、それら化学構造が似ているタニモト係数が非常に似た化合物のドッキング構造とそのscore（スコア）を容易に計算できる。これは、結合化合物のリード最適化（lead optimization）または化合物の新規（de novo）デザインであり、上述の発明でのFPの役割との組み合わせにおいて、従来手法と異なり高い効果を有し有用である。

また、従来、標的タンパク質に種々低分子化合物の部分的なものであるベンゼン環等の官能基を結合させ、物理学的に安定な部分的構造を得て、それらの結果を標的タンパク質と分子内自由回転を多く含有する種々低分子化合物との相互作用を計算するとき、そのドッキングコンブオーメーションの発生を少なくすることが一般的に行われていた。本発明では、バイオインフォマティクスの手法である「circle」のような安定性を評価できるプログラムを用いて標的タンパク質との相互作用エネルギーを計算して、改正FPを創作している。この点に関し、文献等の公知物は発見できず、本発明のように、ドッキングの計算の基底として、FPの重ね合わせを採用するとき、ドッキングの計算の改正FPを基底にすることを報告している従来手法はなく、従来手法と異なり高い効果を有し有用である。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記結合化合物は、公知のドッキングアルゴリズムにより上記標的タンパク質に対して安定なコンフォメーションを持つと予測された化合物であること、を特徴とする。

本発明の一例として具体例を示すと、本発明では、従来一般的に行われている

方法である水素結合や疎水性相互作用や静電相互作用といった物理的なポテンシャル関数を用いた第一原理的アプローチ (Ab-initio Approach) を採用する。例えば、本発明は、正解構造を隠したブラインド・テスト (blind test) によって正解構造に rmsd 2.0 以下で予測できる割合が保証されているような、DOCKやAutoDockやGOLDなど既存のドッキングソフトを使ってドッキング計算で安定コンフォメーションが高いスコアを持つと予測された低分子化合物の三次元座標から抽出したFP (finger print) を追加する。

また、本発明は、標的タンパク質と種々低分子化合物の相互作用のスコア化によって得られたコンフォメーションを、DOCKやAutoDockやGOLDなど既存のドッキングソフトの初期コンフォメーションとして用いてもよい。これにより、上記発明において、得られた初期コンフォメーションが簡便に得られるのに加えて、実験を再現する精度が高いので、他のソフトプログラムとの組み合わせによって、有用な結果を得られる。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記最適化手段は、上記化合物指紋単位に上記二乗平均偏差を基礎とした、上記候補化合物の、上記標的タンパク質との衝突具合、上記標的タンパク質の相互作用領域における存在割合、および、上記標的タンパク質との直接相互作用割合を考慮に入れた関数に基づいて、上記相互作用スコアを計算する相互作用スコア計算手段、を更に備えたことを特徴とする。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記最適化手段は、上記相互作用スコアをメトロポリス法に基づいて判定し、判定結果にしたがって基底となる上記化合物指紋を変更、増加、または減少させることにより、上記相互作用スコアを最適化させること、を特徴とする。

本発明の一例として具体例を示すと、本発明のメトロポリス判定は、前回のスコアより、今回のスコアが大きいならば候補リガンドの構造を採用し、スコアが小さいならば、採用確率 P_{accept} を計算して、 P_{accept} に従って棄

却するか採用するかを決定してもよい。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記最適化手段は、上記相互作用スコアの最適化過程において、上記候補化合物のコンフォメーションを繰り返し変化させ、シミュレティッドアニーリング法に基づいて、当該候補化合物の上記コンフォメーション毎に当該候補化合物を剛体として繰り返し並進または回転させる構造変換手段、を更に備え、上記最適化手段は、上記構造変換手段により並進または回転された上記コンフォメーション毎の上記候補化合物について上記相互作用スコアを計算すること、を特徴とする。

本発明の一例として具体例を示すと、標的タンパク質の立体構造に類似しているファミリー高分子セットに結合した種々低分子化合物の幾つかの三次元座標情報を含んだFPに、仮想化合物ライブラリーから、標的タンパク質にライブラリー低分子化合物をドッキングして、相互作用の最適なコンフォメーションを探すために、繰り返し、モンテカルロ的シミュレティッド・アニーリング (simulated annealing) によって、スコアが最高になるように数学的計算を行う。

さらに具体的には、まず、本発明は、候補リガンドの回転可能な二面角をランダムに変更することにより、コンフォメーションを変化させ、そのコンフォメーションの変化した候補リガンドの座標を用いる。そして、本発明は、標的タンパク質のファミリータンパク質に結合した結合化合物セット由来のFPバンドからランダムに10個のFPを選ぶ。そして、本発明は、選択されたfpバンドから候補リガンドおよび、ライブラリーリガンドからFP原子座標セットをランダムに選択する。そして、本発明は、この状態をフィンガープリント (FP) アライメントとし、その対応関係で、最小二乗フィッティングを行う。本発明は、そのときの重ね合わせの二乗平均偏差 (rmsd) と重ね合わせ後の候補リガンドの原子座標を用いて相互作用スコアを計算する。そして、本発明は、二回目以降は前回の状態を記憶しておき、リガンド原子のコンフォメーションを保ったまま、すなわち剛体並進、回転を行う。そして、本発明は、一つのFPの増加、減少、

および、原子座標セットの対応関係の変更、追加を行う。本発明は、このステップを例えば10000回行う。ここで、シミュレティッドアニーリングの温度は、30Kからはじめ、0.07Kまで下げてもよい。このように、本発明は、一つのコンフォメーションのスコアの最大値を計算し、初期に発生した1000個の
5
コンフォメーションについて比較し、スコアが最大の構造をタンパク質ーリガンド複合体構造として予測し出力する。このとき、1000個のコンフォメーションを当該スコア順位付けする過程は、遺伝的アルゴリズム等を使用することにより、計算時間や、最大値の探索において工夫してもよい。

また、本発明は、上記記載のインシリコスクリーニング装置において、上記最
10
適化手段は、上記相互作用スコアを以下の数式(1)に基づいて算出すること、を特徴とする。

[数1]

$$\begin{aligned}
 FPAScore &= F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule}) \\
 &= \text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) \\
 &\quad \times \text{fp_volume}(\text{molecule}) \\
 &\quad \times \text{fp_contact_surface}(\text{molecule}) \quad (1)
 \end{aligned}$$

(ここで、上記FPAScoreは上記相互作用スコアを表し、上記F(aligned_fp, fp_rmsd, molecule)は、上記結合化合物と上記候補化合物間の上記化合物指紋単位のアライメント度および上記二乗平均偏差、
15
ならびに、上記候補化合物の上記標的タンパク質に対する上記立体構造を変数とする関数であり、上記BaseScore(aligned_fp, fp_rmsd)は、上記化合物指紋単位の一一致度および密集度を示す指標であり、上記fp_volume(molecule)は、上記結合化合物指紋セットの上記三次元座標からなる空間を上記候補化合物が占める割合、および、上記標的タンパク質との衝突具合を示す指標であり、上記fp_contact_surface(molecule)は、上記候補化合物の上記標的タンパク質との接触度、
20
および、上記結合化合物指紋セットの上記三次元座標への帰属度を示す指標であ

る。)

5 以上のように、これら上記に述べた発明における数学的計算は、従来における物理学的相互作用関数で標的タンパク質と仮想化合物ライブラリー低分子化合物の相互作用を計算していたところを、バイオインフォマティクスの情報を使って半経験的に計算している点が従来手法と異なり、さらに構造予測の成功率は世界で認められているドッキングソフトプログラムに優れるとしても、決して劣ることはないという高い効果を発揮する。また、情報の蓄積が、半経験的バイオインフォマティクス手法の相互作用計算の結果を良いほうに導くので、従来手法と異なり有用である。

10 また、本発明は、上記記載のインシリコスクリーニング装置において、上記数式(1)における、上記 $BaseScore(aligned_fp, fp_rmsd)$ は、以下の数式(2)に基づいて算出され、

[数2]

$$BaseScore(aligned_fp, fp_rmsd) = \frac{RawScore(aligned_fp)}{1 + \ln(fp_rmsd^{k1} + 1)} \quad (2)$$

15 (ここで、上記 $RawScore(aligned_fp)$ は、上記結合化合物と上記候補化合物間でアライメントされた上記化合物指紋における原子の数に基づく指標であり、上記 fp_rmsd は、上記二乗平均偏差である。)

上記 $fp_volume(molecule)$ は、以下の数式(6)に基づいて算出され、

20 [数3]

$$fp_volume(molecule) = \ln \frac{1.0 + nafp^{k2}}{1.0 + nap^{k3}} \quad (6)$$

(ここで、上記 $nafp$ は、上記結合化合物指紋セットの上記三次元座標に基づく固有格子点領域に上記候補化合物の上記三次元座標が占有する格子点の数であり、上記 nap は、上記標的タンパク質の上記立体構造における原子の固有格子点領域に上記候補化合物の上記三次元座標が属する格子点の数であり、上記 $k2$

25

および上記 k_3 は、任意の定数である。)

上記 $fp_contact_surface(molecule)$ は、以下の数式 (7) に基づいて算出されること、を特徴とする。

[数 4]

$$fp_contact_surface(molecule) = \frac{\sum_{i=1}^n density_of_atom(atom(i))}{total_density_of_atom(molecule)} \quad (7)$$

(ここで、上記 n は、上記候補化合物の原子の数であり、上記 $atom(i)$ は、上記候補化合物の i 番目の原子の上記三次元座標であり、上記 $density_of_atom(atom(i))$ は、当該原子の上記三次元座標が上記結合化合物指紋セットの上記化合物指紋に属している場合に、当該化合物指紋の上記原子と所定の距離で接触している上記標的タンパク質の原子の数と、当該化合物指紋の同一格子点に属する上記結合化合物の原子の数との和を返す関数であり、上記 $total_density_of_atom(molecule)$ は、上記 $density_of_atom$ の分布を降順に並べ換えたものを上記候補化合物の原子の数だけ順に足し合わせた数である。)

本発明の一例として具体例を示すと、本発明は、上記の内容の中で k_2 、 k_3 の値を明確にするために、EGFRやVEGFRなどの固有の標的タンパク質に対して既知活性化合物を探し、 k_2 、 k_3 を最適化する。そして、本発明は、その値がたとえば、EGFRの阻害剤のインシリコスクリーニングにおいて、 $k_2 = 2.0$ 、 $k_3 = 1.0$ 、となるようなインシリコスクリーニングを行う方法である。上記発明に係るインシリコスクリーニングによって、EGFRやVEGFRなどの固有の標的タンパク質に適合する化合物を的確にリストアップすることは、抗がん剤の新薬開発に直結するので、従来手法と異なり高い効果を有し有用である。

従来、GOLDのようなドッキングソフトプログラムは、生物学的に重要な水素結合に参加する原子を点やベクトルとして、遺伝子アルゴリズムにおいて、良

いセットを選ぶ工夫をしている。このような点やベクトルは、上記発明で記載のような標的タンパク質の立体構造に類似しているファミリー高分子セットに種々低分子化合物が結合した集団的コンフォメーションを抽出するときの条件としての部品である三次元化学記述子のFPとは違っている。この発明では、上述の発明において、相互作用しているコンフォメーションにおいて生物学的に重要な水素結合等を構成する原子点やベクトルのセットを取り込む場合に、 fp_rmsd 値を以下の式とすれば、生物学的に重要な水素結合または疎水結合またはファンデルワールス相互作用に参加する原子を上述の発明に矛盾なく含ませることが出来ることを特徴とする。

すなわち、本発明では、 $fp_rmsd + distance_rmsd\ in\ indicative\ atom\ set\ composed\ of\ important\ points\ vectors$ の式を $fp_rmsd^{**k1} + distance_rmsd^{**k4}$ ($**k1 \ll **k4$ はFPの寄与が小さい： $**k1 \gg **k4$ はFPの寄与を重視)の形に拡張してもよく、 $distance_rmsd^{**k4}$ としてもよい。ここで、 $distance_rmsd$ は、標的タンパク質とドッキングする低分子化合物との相互作用において、リガンド原子が標的タンパク質のリガンド結合部位における生物学的に重要な水素結合または疎水結合またはファンデルワールス相互作用する場合、標的タンパク質のリガンド結合部位における理想座標、標的タンパク質の生物学的に重要な原子、もしくは、その近傍の原子から発生させたベクトルの終点座標との最小二乗誤差として定義される。

また、本発明では、種々低分子化合物において、化合物の殆どがアミノ酸残基のつながったペプチドの場合、ペプチド基が多いために、 fp の対応関係が複雑になるので、スコアの計算過程で過小評価して、上記発明における $RawScore$ についての上記数式において、ペプチドの部分のFPの式に該当する部分をゼロ等の過小評価の数字にしてもよい。

すなわち、本発明は、FPを基底にして、標的タンパク質とドッキングする低

分子化合物との相互作用を計算する方法に、標的高分子である標的タンパク質のリガンド結合環境の格子情報、化合物と標的高分子間のベクトルを重視する化合物の多点情報、標的タンパク質の生物学的環境を表す化合物から標的タンパク質に向かうベクトル等の何らかの工夫を行う。その上で、本発明は、化合物の種々

5 原子と標的高分子タンパク質を構成する種々原子との古典物理学的原子間ポテンシャル式から相互作用エネルギー等を計算する方法を包含して、融合させるような、上記発明の拡張発明であり、化合物のコンフォメーションや相互作用の結合の強さに関係する順番をスコア値で決めることに関して、従来手法と異なり高い効果を有し有用である。

10 また、本発明は、記憶部と制御部を少なくとも備えたインシリコスクリーニング装置において実行される、標的タンパク質に結合する候補化合物のスクリーニングを行うインシリコスクリーニング方法であって、上記記憶部は、化合物中の複数個の原子に係る化合物指紋として、原子タイプと原子間結合規則とを含む化学記述子を、上記候補化合物ごとに抽出して作成された化合物データベースを備

15 えており、上記制御部において実行される、上記標的タンパク質と立体構造が同一または類似するファミリータンパク質に結合することが既知の結合化合物について、上記標的タンパク質の座標系に変換した三次元座標とともに上記化合物指紋を抽出して結合化合物指紋セットを作成する化合物指紋作成ステップと、上記化合物データベースに記憶された上記候補化合物について、上記結合化合物指紋

20 セットの上記三次元座標を基底として算出した上記化合物指紋単位の二乗平均偏差を基礎とする相互作用スコアが最適化されるように、当該候補化合物の上記標的タンパク質に対する上記立体構造を演算する最適化ステップと、を含むことを特徴とする。

以上、この発明によれば、タンパク質と化合物との結合を精度よく予測することが

25 ができる一方で、ヒットする化合物を数多く選出することができ、また、生物化学的な実験等の情報を考慮に入れた半経験的なスクリーニングを行うことができ予測効率を高めることができるという効果を奏する。

図面の簡単な説明

第1図は、本発明が適用される本インシリコスクリーニング装置の構成の一例を示すブロック図であり、第2図は、インシリコスクリーニング装置100の処理の一例を示すフローチャートであり、第3図は、従来のドッキングソフトと、タンパク質-リガンド複合体の多数のX線構造やNMR構造を効果的に用いたバイオインフォマティクスによる本実施例に係るドッキング方法を示す状況図であり、第4図は、本実施例(ChooseLD)によるタンパク質-リガンド・ドッキングの原理構成図であり、第5図は、FP(fingerprint)の作り方を一例として示す図であり、第6図は、本実施例で用いた原子の文字列一覧を示す図表であり、第7図は、タニモト係数による化合物間の類似性算出方法を示す模式図であり、第8図は、標的タンパク質の結合部位にリガンドをドッキングさせる場合のFPを一例として示す模式図であり、第9図は、たどった経路から原子座標を得て、FPバンドに登録する過程を一例として示す図であり、第10図は、本実施例におけるFPバンドの絞り込みステップ(method step of shrinking fingerprint band)を一例として示す図であり、第11図は、座標ベクトル同士に対応関係を与える過程を一例として示す模式図であり、第12図は、nafpとnapの具体例を原子数が31のリガンドを用いて示した図であり、第13図は、標的タンパク質の活性部位近傍におけるFPライブラリー由来のリガンドの位置を一例として示した図であり、第14図は、シミュレティッドアニーリング過程を一例として示す概念図であり、第15図は、FPAScoreを算出するためのFPアライメントおよび最小自乗フィッティングを模式的に示した図であり、第16図は、EGFRインシリコスクリーニングにおける計算時間の分布を示す図であり、第17図は、ベンチマークの概要を一例として示す図であり、第18図は、PDBへの登録数の年度分布を表す図であり、第19図は、予測と実験結果間でのrmsdを要約したテーブルであり、第20図は、85セットにおける予測成功率一覧(k1と

Tc Rangeの関係)を示す図表であり、第21図は、10位までにrmsd 2.0以下で予測できる割合を示す図表であり、第22図は、10位までにrmsd 2.5 (Close)以下で予測できる割合を示す図表であり、第23図は、成功とみなす正解構造とのrmsdを2.0 Å以外でも行った場合を示す図表であり、第24図は、ChooseLDと比較して、Dock, AutoDockおよびGOLDのベンチマークの結果を示す図表であり、第25図は、85セットにおけるFPAScoreの予測構造と実験構造とのrmsdが2.0 Å以下における各々標的タンパク質との衝突個数の分布を示す図であり、第26図は、85セットベンチマークにおける予測成功構造の個数分布を示す図であり、第27図は、各ターゲットにおける全10回のドッキング試行における成功個数を示す図であり、第28図は、133セットのベンチマークにおけるDOCK, AutoDock, GOLD予測構造のrmsd分布の結果と、ChooseLD法の結果を示す図であり、第29図は、133セットのベンチマークにおけるDOCK, AutoDock, GOLD予測構造のrmsd分布の結果と、ChooseLD法の結果を示す図であり、第30図は、各ターゲットにおける全10回のドッキング試行における成功個数を示す図であり、第31図は、各ターゲットにおける全10回のドッキング試行における成功個数を示す図であり、第32図は、Tc範囲で限定されたFPライブラリーにおいてFPAScoreで順位付けされた分布内に実験構造とのrmsdが2.0 Å以下の構造が得られる確率を示す図であり、第33図は、Tc範囲で限定されたFPライブラリーにおいてFPAScoreで順位付けされた分布内に実験構造とのrmsdが2.0 Å以下の構造が得られる確率を示す図であり、第34図は、予測成功構造の衝突個数の分布を示す図であり、第35図は、FPライブラリーに用いるリガンドのTc範囲の上限値をさらに低くし、0.16, 0.24, 0.36に下限値を0.08にした場合の性能および、前述したTc範囲、すなわち上限値0.56, 0.76, 0.96、下限値0.08の予測成功率を示す図であり、第36図は、1DR1について予測されたタンパク質-リガンド構造を示す図であり、第37図

は、4ESTについて予測されたタンパク質ーリガンド構造を示す図であり、第38図は、GOLDが失敗したがChooseLDは予測に成功したターゲットを示す1CDGについての図であり、第39図は、GOLDが失敗したがChooseLDは予測に成功したターゲットを示す1DR1についての図であり、第40図は、GOLDが失敗したがChooseLDは予測に成功したターゲットを示す1LDMについての図であり、第41図は、GOLDが失敗したがChooseLDは予測に成功したターゲットを示す4ESTについての図であり、第42図は、133セット中における90ターゲットにおける予測成功率を示す図表であり、第43図は、ドッキングソフト間の予測に成功した標的タンパク質のPDBIDの類似度をTc(タニモト係数)で算出した図表であり、第44図は、90ターゲット中の一つの標的タンパク質に対する各ドッキングソフトの予測の成否分布を示す図表であり、第45図は、DOCKが失敗したがChooseLDは予測に成功したターゲットを示す1HYTについての図であり、第46図は、DOCKが失敗したがChooseLDは予測に成功したターゲットを示す1PHGについての図であり、第47図は、DOCKが失敗したがChooseLDは予測に成功したターゲットを示す1TMNについての図であり、第48図は、1位だけではなく10位までにrmsd2.0の構造が採取できる割合を示す図であり、第49図は、1位だけではなく10位までにrmsd2.5(Close)の構造が採取できる割合を示す図であり、第50図は、成功と定義するrmsdを変化させた場合を示す図表であり、第51図は、本実施例による処理の結果を示す図表であり、第52図は、EGFRからの細胞内シグナル伝達経路を示した図であり、第53図は、EGFRのアミノ酸配列のアライメントを示す図であり、第54図は、構築されたEGFRのモデルを示す図であり、第55図は、入手した11個の阻害剤の平面構造を示す図であり、第56図は、FPAScoreで定義されたk2値を0.5から5.0の範囲に変更した際の収穫率折れ線グラフを示す図であり、第57図は、FPAScoreにおけるk3値を0.5から2.0の範囲に変更した際の収穫率折れ線グラフを示す図であり、第58図

は、T_c上限値を1.00とし、T_c下限値の範囲を0.08から0.32まで、0.08刻みで変化させた場合の、それぞれのT_c範囲におけるインシリコスクリーニングの結果を示す図であり、第59図は、PDBに登録されているタンパク質—リガンド複合体構造既知のPDB IDとそのリガンドの順位付けを示す図
5 であり、第60図は、第59図のリガンドIDと化合物名を対応付ける図であり、第61図は、Kinaseのインシリコスクリーニングによる絞り込みの結果の上位10位のタンパク質—リガンド複合体を示す図であり、第62図は、第61図を別角度から見た図であり、第63図は、TGF- α 結合ドメイン近傍を表した図であり、第64図は、MDL Comprehensive Medicinal Chemistry (MDL CMC) Libraryを用いたEGFRのTGF- α 結合ドメインに対するインシリコスクリーニングの結果を示す図であり、第65図は、MDL ACD Libraryを用いた同インシリコスクリーニングの結果を示す図であり、第66図は、KRN633 (IC₅₀ = 1.16 nm/L) の平面構造を示す図であり、第67図は、KRN95
15 1 (IC₅₀ = 0.16 nm/L) の平面構造を示す図であり、第68図は、KRN633のVEGFR2活性部位近傍へのドッキングに用いたFPライブラリーに所属するリガンドにおいてドッキングに使用されたリガンドの上位10個を示した図であり、第69図は、KRN633について、ChooseLD法を10回実行し、予測された構造10個をVEGFR2の活性部位近傍の立体構造とともに示した図であり、第70図は、KRN951のVEGFR2活性部位近傍へのドッキングに用いたFPライブラリーに所属するリガンドにおいてドッキングに使用されたリガンドの上位10個を示した図であり、第71図は、KRN951について、ChooseLD法を10回実行し、予測された構造10個をVEGFR2の活性部位近傍の立体構造とともに示した図であり、第72図は、1
25 33セットを用いたChooseLD法のドッキング性能試験の結果得られたT_c下限値を0.08に固定し、T_c上限値を変化させた時の予測成功率について、横軸にT_c 上限値、縦軸に成功率としたグラフを示す図であり、第73図は、

enoyl acyl carrier proteinの立体構造を示した図であり、第74図は、MDL Comprehensive Medicinal Chemistry (MDL CMC) Libraryを用いて、enoyl acyl carrier proteinのインシリコスクリーニングを行った結果のFPAScoreの上位10構造を示す図であり、第75図は、AMPK_{homo}GAMMA1と2V9J_Eのアミノ酸配列のアライメントを示した図であり、第76図は、リガンドが受容体全体に結合したCMC医薬品の結果リストを示す図であり、第77図は、1位から10位までの2V9J_E受容体への結合状態を図に示した図である。

10

発明を実施するための最良の形態

以下に、本発明にかかるインシリコスクリーニング装置およびインシリコスクリーニング方法の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。

15 [本発明の概要]

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。

現在、X線解析、NMR実験、電子線解析実験、高分解能電子顕微鏡写真等の実験によって、ペプチドや低分子化合物や金属等の種々の化合物が標的高分子と直接相互作用をしている状態を示す、約四万に至る数の三次元座標がPDB (Protein Data Bank) に登録されている。また、コンピュータの性能とバイオインフォマティクスの進歩により、種々化合物が結合した、標的高分子タンパク質に立体構造が類似しているファミリー高分子タンパク質セットは、SCOP等のウェブサイトやCASPで優秀な成績を示している本出願人が製作したプログラム等により容易に得られ、抽出できる。

25

この状態をふまえ、本願発明者は、従来一般的に古典物理学的に決められている標的高分子タンパク質に対する直接結合する当該化合物のコンフォメーション

やその時得られるスコア値を利用した当該相互作用エネルギーの結果から化合物のインシリコスクリーニングの順番を決める手法の代わりに、標的高分子タンパク質に結合した種々化合物の集団的重なり状態を利用して、バイオインフォマテイクスを代用できれば、人の叡智を基礎にした化合物のコンフォメーションやその時得られるスコア値を利用した相互作用エネルギーの結果から化合物のインシリコスクリーニングにより順番を決めることが可能となるはずであるとの着想を得た。

本発明は、上記着想に基づいて本願発明者により鋭意検討された結果完成されたものであり、概略的に、以下の基本的特徴を有する。すなわち、本発明は、記憶部と制御部を少なくとも備えた、標的タンパク質に結合する候補化合物のスクリーニングを行うインシリコスクリーニング装置であって、記憶部は、化合物中の複数個の原子に係る化合物指紋として、原子タイプと原子間結合規則とを含む化学記述子を、候補化合物ごとに抽出して作成された化合物データベースを備える。

ここで、「化合物指紋」(フィンガープリント:FP)とは、より具体的には、化合物中の原子2個、3個ないし4個等の原子の原子タイプと原子間結合規則を内包した化学記述子である。「原子タイプ」は、一例として、Sybyl原子タイプ(atom-type)や「原子価タイプ」(Valence-type)等である。「原子間結合規則」は、原子間の化学結合の状態を表すものであり、例えば、単結合や二重結合や芳香環結合等の結合規則や、分子軌道法による分類等を示すものである。

つづいて、本発明のスクリーニング装置は、標的タンパク質と立体構造が同一または類似するファミリータンパク質に結合することが既知の結合化合物について、標的タンパク質の座標系に変換した三次元座標とともに化合物指紋を抽出して結合化合物指紋セットを作成する。すなわち、標的タンパク質の座標系において、その立体構造に結合した化合物集団の集団的コンフォメーションを収集し、三次元座標を対応付けて化合物指紋を抽出する。

ここで、「標的タンパク質と立体構造が同一または類似するファミリータンパク質」は、標的タンパク質自体でもよく、標的タンパク質の一部の構造（例えば、活性部位やリガンド結合部位など）と同一または類似するタンパク質でもよく、標的タンパク質の立体構造を解析して活性部位を指定することなく同一または類似するタンパク質を用いてもよい。安定コンフォメーションが高いスコアを持つようにするために、従来のDOCKやAutoDockやGOLDなど既存のドッキングソフトを使つてのドッキング計算では、予め当該標的タンパク質の立体構造を解析して活性部位を指定する必要があった。しかし、本発明では、これらに比べて、従来手法とは異なり高い効果を有し、文献等の学習を通じて活性部位を指定する必要がないので有用である。

また、標的タンパク質のアミノ酸配列をクエリー配列として、化合物に結合したタンパク質の立体構造およびアミノ酸配列を記憶するタンパク質データベース等から相同性検索を行つて、標的タンパク質との構造重ね合わせにより構造の類似性を表す指標が一定値以上となったタンパク質をファミリータンパク質としてもよい。また、ここで、「タンパク質に結合することが既知の結合化合物」には、X線構造解析やNMR構造解析等により実験的にタンパク質-化合物複合体の立体構造が確認されたものでもよい。また、結合化合物は、単にタンパク質に結合することが既知であればよく、公知のドッキングアルゴリズム（DOCKやAutoDockやGOLD等）や任意の座標発生プログラム（Corinaなど）等により標的タンパク質に対して安定なコンフォメーションを持つと予測された化合物でもよい。

また、ここで、本インシリコスクリーニング装置は、結合化合物の三次元座標を標的タンパク質の座標系に変換するために、ファミリータンパク質と標的タンパク質との構造重ね合わせ操作を行い、ファミリータンパク質に結合した結合化合物をファミリータンパク質の座標系から標的タンパク質の座標系に結合化合物の座標と共に変換してもよい。例えば、構造重ね合わせ操作は、原子の種類を考慮しないタンパク質同士の構造の重ね合わせアルゴリズム（CE等）によって実

行してもよく、標的タンパク質とファミリータンパク質との相同性が高い場合には、原子の種類を考慮した構造重ね合わせを行ってもよい。

また、化合物指紋の抽出は、結合化合物から直接抽出することに限らず、標的タンパク質に対する候補化合物の探索の狙いの必要に応じて任意の化合物指紋を加えてもよい。例えば、結合化合物と異なる他の化合物を参照して構造重ね合わせを行い、結合化合物と他の上記化合物の原子間をまたがる新たな化合物指紋を作成して結合化合物指紋セットに加えてもよく、タニモト係数に基づき結合化合物と類似する化合物について、結合化合物と当該化合物の原子間で原子の種類を入れ替え、安定性を評価できるプログラム（「circle」等）を用いて標的タンパク質に対する相互作用エネルギーを算出して結合化合物の化合物指紋よりも局所エネルギー的に安定な化合物指紋を「改正化合物指紋（Modified FP）」として新たに作成して結合化合物指紋セットに追加してもよい。すなわち、標的タンパク質との結合に理想的な低分子化合物をリード化合物として、相互作用エネルギーがよくなるようにいろいろな置換基を付加したり、理想的な低分子化合物に化合物指紋の定量化関数であるタニモト係数が非常に似た、すなわち1に近い任意の低分子化合物を見つける場合に、化合物指紋領域を、実験的に解析済みの理想的な低分子化合物の周りの領域である4または5オングストロームに限定する。これにより、それら化学構造が似ているタニモト係数が非常に似た化合物のドッキング構造とその相互作用スコアを容易に計算できる。

つづいて、本発明のインシリコスクリーニング装置は、化合物データベースに記憶された候補化合物について、座標固定の結合化合物指紋セットの三次元座標を基底として算出した化合物指紋単位の二乗平均偏差（ $rmsd: root-mean-square-deviation$ ）を基礎とする相互作用スコアが最適化されるように、候補化合物の標的タンパク質に対する立体構造を演算する。

すなわち、この最適化過程において、本インシリコスクリーニング装置は、一例として、候補化合物のコンフォメーションを繰り返し変化させ、候補化合物のコンフォメーション毎に候補化合物を剛体として繰り返し並進または回転させ、

二乗平均偏差を基礎として算出した相互作用スコアをメトロポリス法に基づいて判定し、判定結果にしたがって候補化合物の化合物指紋を変更、増加、または減少させる。ここで、化合物指紋をいくつかランダムに抽出して、基底となる座標固定の結合化合物指紋セットを選択してもよい。また、候補化合物の回転可能な

5 二面角をランダムに変更することによりコンフォメーションを変化させる代わりに、遺伝子アルゴリズム等のように以前のコンフォメーションを記憶して候補化合物の構造を変化させてもよい。

また、上記最適化過程における相互作用スコアの計算は、一例として、化合物指紋単位に二乗平均偏差を基礎とした、候補化合物の、標的タンパク質との衝突

10 具合、標的タンパク質の相互作用領域における存在割合、および、標的タンパク質との直接相互作用割合を考慮に入れた関数に基づいて計算される。相互作用スコアは、より具体的には、以下の数式(1)に基づいて算出される。

[数5]

$$\begin{aligned}
 FPAScore &= F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule}) \\
 &= \text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) \\
 &\quad \times \text{fp_volume}(\text{molecule}) \\
 &\quad \times \text{fp_contact_surface}(\text{molecule}) \quad (1)
 \end{aligned}$$

15 (ここで、FPAScoreは相互作用スコアであり、 $F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule})$ は、結合化合物と候補化合物間の化合物指紋単位のアライメント度および二乗平均偏差、ならびに、候補化合物の標的タンパク質に対する立体構造を変数とする関数であり、 $\text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd})$ は、化合物指紋単位の一

20 致度および密集度を示す指標であり、 $\text{fp_volume}(\text{molecule})$ は、結合化合物指紋セットの三次元座標からなる空間を候補化合物が占める割合、および、標的タンパク質との衝突具合を示す指標であり、 $\text{fp_contact_surface}(\text{molecule})$ は、候補化合物の標的タンパク質との接触度、および、結合化合物指紋セットの三次元座標への帰属度を示す指標である。)

以上が、本発明の処理の概要である。このように、最適化手法に従って計算された相互作用スコアに基づいて、候補化合物の標的タンパク質に対する相互作用の順位が決定され、化合物データベースから有意な候補化合物が推定できるので、タンパク質と化合物との結合を精度よく予測することができる一方で、ヒットする化合物を数多く選出することができ、また、生物化学的な実験等の情報を考慮に入れた半経験的なスクリーニングを行うことができ予測効率を高めることができる。

すなわち、本発明は、標的タンパク質の立体構造に同一または類似しているファミリータンパク質に集団的に結合した種々の低分子化合物（結合化合物）のコンフォメーションが、標的タンパク質と相互作用した最安定構造に近いことを考察した結果なされたものである。さらに、本発明は、結合化合物と候補化合物とを対比する際に取り扱いやすい化合物指紋を単位として適切な相互作用スコアのスコア付けを行って最適化することにより、従来手法よりも予測効率を高めた半経験的なインシリコスクリーニングを行うことができる。

15 [インシリコスクリーニング装置の構成]

まず、本インシリコスクリーニング装置の構成について説明する。図1は、本発明が適用される本インシリコスクリーニング装置の構成の一例を示すブロック図であり、該構成のうち本発明に関係する部分のみを概念的に示している。

図1においてインシリコスクリーニング装置100は、概略的に、インシリコスクリーニング装置100の全体を統括的に制御するCPU等の制御部102、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通信制御インターフェース部104、入力装置112や出力装置114に接続される入出力制御インターフェース部108、および、各種のデータベースやテーブルなどを格納する記憶部106を備えて構成されており、これら各部は任意の通信路を介して通信可能に接続されている。更に、このインシリコスクリーニング装置100は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク300に通信可能に接続されている。

記憶部106に格納される各種のデータベースやテーブル（候補化合物DB106a～医薬品化合物DB106c）は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ等を格納する。

- 5 これら記憶部106の各構成要素のうち、候補化合物DB106aは、インシリコスクリーニングの候補となる化合物（「候補化合物」と呼ぶ。）ごとに化合物指紋を抽出して作成された候補化合物データベース手段である。

- また、結合化合物指紋セット106bは、標的タンパク質と立体構造が同一または類似するタンパク質（「ファミリータンパク質」と呼ぶ。）に結合することが既知の化合物（「結合化合物」と呼ぶ。）について、標的タンパク質の座標系に変換した三次元座標とともに化合物指紋を抽出して作成された結合化合物指紋セットを記憶する結合化合物指紋記憶手段である。

- また、医薬品化合物DB106cは、既知の医薬品化合物について化合物指紋を抽出して作成された医薬品化合物指紋セットを記憶する、MDL CMC Library等の医薬品化合物データベースである。すなわち、医薬品化合物DB106cは、医薬品データベースを使って化合物情報を引き出すために、薬物吸収や薬物代謝や薬物排泄や薬物毒性等を指標にして、化合物指紋の整理の基底としての基礎データ単位を使って、予め整理した薬物吸収や薬物代謝や薬物排泄や薬物毒性に特化した結合化合物指紋セット106bを作成するために用いられる。

- また、図1において、通信制御インターフェース部104は、インシリコスクリーニング装置100とネットワーク300（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部104は、他の端末と通信回線を介してデータを通信する機能を有する。

- また、図1において、入出力制御インターフェース部108は、入力装置112や出力装置114の制御を行う。ここで、出力装置114としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下におい

ては出力装置 1 1 4 をモニタとして記載する場合がある)。また、入力装置 1 1 2 としては、キーボード、マウス、記録媒体読取装置等を用いることができる。この入力装置 1 1 2 を介して、インシリコスクリーニングの対象となる標的タンパク質や候補化合物が入力される。

- 5 また、図 1 において、制御部 1 0 2 は、OS (Operating System) 等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 1 0 2 は、機能概念的に、化合物指紋作成部 1 0 2 a、最適化部 1 0 2 b、スクリーニング結果出力部 1 0 2
10 c、相同性検索部 1 0 2 d を備えて構成されている。

化合物指紋作成部 1 0 2 a は、候補化合物や結合化合物や医薬品化合物等の化合物から化合物指紋を抽出する化合物指紋作成手段である。例えば、化合物指紋作成部 1 0 2 a は、入力装置 1 1 2 を介して入力された候補化合物について化合物指紋を抽出して候補化合物指紋セットを作成し候補化合物 DB 1 0 6 a に格納
15 する。また、化合物指紋作成部 1 0 2 a は、取得した医薬品化合物から化合物指紋を抽出して医薬品化合物指紋セットを作成し医薬品化合物 DB 1 0 6 c に格納する。

また、化合物指紋作成部 1 0 2 a は、ファミリータンパク質に結合することが既知の結合化合物について、原子の三次元座標を標的タンパク質の座標系に変換
20 し、変換した三次元座標とともに化合物指紋を抽出して結合化合物指紋セット 1 0 6 b を作成する。すなわち、化合物指紋作成部 1 0 2 a は、標的タンパク質の座標系において、その立体構造に結合した化合物集団の集団的コンフォメーションを収集し、三次元座標に対応付けて化合物指紋を抽出する。換言すれば、化合物指紋作成部 1 0 2 a は、標的タンパク質に結合した化合物集団から化合物指紋
25 と呼ばれる原子 2 個、3 個、または 4 個等の原子の原子タイプと原子間の結合規則を内包した化学記述子を化合物記述子の三次元座標を伴って、できる限り多く抽出し、それらを記憶部 1 0 6 の中にデータベースの表として収納することによ

り結合化合物指紋セット106bを作成する。

ここで、化合物指紋作成部102aは、結合化合物の三次元座標を標的タンパク質の座標系に変換するために、ファミリータンパク質と標的タンパク質との構造重ね合わせ操作を行い、ファミリータンパク質に結合した結合化合物の三次元座標を、（ファミリータンパク質の座標系から）標的タンパク質の座標系に変換してもよい。例えば、化合物指紋作成部102aは、原子の種類を考慮しないタンパク質同士（標的タンパク質とファミリータンパク質）の構造重ね合わせアルゴリズム（CE等）によって構造重ね合わせ操作を行ってもよく、標的タンパク質とファミリータンパク質との相同性が高い場合には、原子の種類を考慮した構造重ね合わせを行ってもよい。

また、化合物指紋作成部102aは、結合化合物から直接、化合物指紋を抽出することに限らず、標的タンパク質に対する候補化合物の探索の狙いの必要に応じて任意の化合物指紋を結合化合物指紋セット106bに加えてもよい。ここで、化合物指紋作成部102aは、図1に示すように、新規化合物指紋追加部102eを備えて構成される。すなわち、新規化合物指紋追加部102eは、結合化合物から直接抽出される化合物指紋以外の新規な化合物指紋を作成し結合化合物指紋セット106bに追加する新規化合物指紋追加手段である。例えば、新規化合物指紋追加部102eは、結合化合物と異なる他の化合物を参照して構造重ね合わせを行い、結合化合物と他の上記化合物の原子間をまたがる新たな化合物指紋を作成して結合化合物指紋セット106bに加えてもよい。また、新規化合物指紋追加部102eは、タニモト係数に基づき結合化合物と類似する化合物について、結合化合物と当該化合物の原子間で原子の種類を入れ替え、安定性を評価できるプログラム（「circle」等）を用いて標的タンパク質に対する相互作用エネルギーを算出して結合化合物の化合物指紋よりも局所エネルギー的に安定な化合物指紋を改正化合物指紋（Modified FP）として新たに作成して結合化合物指紋セット106bに追加してもよい。

最適化部102bは、候補化合物DB106aに記憶された候補化合物につい

て、結合化合物指紋セット106bに記憶された化合物指紋の三次元座標を基底として化合物指紋単位 $(rmsd)$ を算出し、当該二乗平均偏差を基礎とする相互作用スコアが最適化されるように、候補化合物の標的タンパク質に対する立体構造を演算する最適化手段である。例えば、最適化部102bは、

5 生成した候補化合物の当該コンフォメーションおよび標的タンパク質に対する三次元座標ごとに、二乗平均偏差を基礎として算出した相互作用スコアをメトロポリス法に基づいて判定し、判定結果にしたがって候補化合物の化合物指紋を変更、増加、または減少させる。ここで、最適化部102bは、結合化合物指紋セット106bから化合物指紋をいくつかランダムに抽出して、基底となる座標固定の

10 結合化合物指紋セットを選択してもよい。ここで、最適化部102bは、図1に示すように、相互作用スコア計算部102f、構造変換部102gを備えて構成されている。

相互作用スコア計算部102fは、最適化部102bによる最適化過程において、化合物指紋単位に二乗平均偏差を基礎とした、候補化合物の、標的タンパク質との衝突具合、標的タンパク質の相互作用領域における存在割合、および、標的タンパク質との直接相互作用割合を考慮に入れた関数に基づいて、相互作用スコアを計算する相互作用スコア計算手段である。なお、相互作用スコア計算部102fによる相互作用スコアの計算の具体例については、以下の処理の説明で詳細に述べる。

15

また、構造変換部102gは、最適化部102bによる最適化過程において、候補化合物のコンフォメーションを繰り返し変化させ、シミュレティッドアニーリング法に基づいて、当該候補化合物のコンフォメーション毎に当該候補化合物を剛体として繰り返し並進または回転させる構造変換手段である。また、構造変換部102gは、候補化合物の回転可能な二面角をランダムに変更することによりコンフォメーションを変化させる代わりに、遺伝子アルゴリズム等のように以前のコンフォメーションを記憶して候補化合物の構造を変化させてもよい。

20

25

スクリーニング結果出力部102cは、最適化部102bにより最適化された

相互作用スコアに基づいて、候補化合物の標的タンパク質に対する相互作用順位を決定して、インシリコスクリーニング結果を出力する結果出力手段である。

5 同源性検索部 102d は、標的タンパク質のアミノ酸配列との同源性に基づいて、ファミリータンパク質および結合化合物をタンパク質データベース装置から検索する同源性検索手段である。すなわち、同源性検索部 102d は、結合化合物を取得するために、標的タンパク質のアミノ酸配列をクエリー配列として、外部システム 200 等のタンパク質データベースに照会することにより同源性検索を行い、標的タンパク質に対して同源性を有するタンパク質に結合した構造が既知の結合化合物を取得する。

10 図 1 に示すように、本インシリコスクリーニング装置 100 は、アミノ酸配列情報やタンパク質立体構造情報に関する外部データベースや、配列や立体構造のアライメント等を行う外部プログラム等を提供する外部システム 200 と、ネットワーク 300 を介して通信可能に接続して構成されてもよい。なお、ネットワーク 300 は、インシリコスクリーニング装置 100 と外部システム 200 とを
15 相互に接続する機能を有し、例えば、インターネット等である。

すなわち、図 1 において、外部システム 200 は、ネットワーク 300 を介して、インシリコスクリーニング装置 100 と相互に接続され、アミノ酸配列情報やタンパク質立体構造情報に関するタンパク質データベース等の外部データベース (PDB や P S I - B l a s t 等) や、配列や立体構造のアライメント等を行う外部プログラム等を提供する機能を有する。ここで、タンパク質データベースには、X線構造解析や NMR 構造解析等により実験的にタンパク質-化合物複合体の立体構造が確認されたものに限らず、単にタンパク質に結合することが既知の化合物が保存されてもよい。この場合、上述の化合物指紋作成部 102a は、
20 公知のドッキングアルゴリズム (DOCK や A u t o D o c k や G O L D 等) や任意の座標発生プログラム (C o r i n a など) 等により、標的タンパク質に対して安定なコンフォメーションを持つ結合化合物の構造を予測して結合化合物指紋セット 106b の作成に利用する。

[インシリコスクリーニング装置100の処理]

次に、このように構成された本実施の形態における本インシリコスクリーニング装置100の処理の一例について、以下に図2を参照して詳細に説明する。図2は、インシリコスクリーニング装置100の処理の一例を示すフローチャートである。

図2に示すように、まず、相同性検索部102dは、入力装置112を介して入力された標的タンパク質のアミノ酸配列に基づいて、外部システム200等のタンパク質データベースから特定の化合物（結合化合物）と結合した立体構造が既知のファミリータンパク質を相同性検索する（ステップSA-1）。

そして、化合物指紋作成部102aは、標的タンパク質の構造と、結合化合物を伴ったファミリータンパク質の構造とを重ね合わせる（ステップSA-2）。ここで、化合物指紋作成部102aは、原子の種類を考慮しないタンパク質同士の構造重ね合わせを行ってもよく、標的タンパク質とファミリータンパク質との相同性が所定値以上で高い場合には、原子の種類を考慮した構造重ね合わせを行ってもよい。

そして、化合物指紋作成部102aは、結合化合物の三次元座標を、ファミリータンパク質の座標系から標的タンパク質の座標系に変換する（ステップSA-3）。

そして、化合物指紋作成部102aは、標的タンパク質の座標系に変換した結合化合物の三次元座標とともに、結合化合物から化合物指紋を抽出して記憶部106に格納することにより結合化合物指紋セット106bを作成する（ステップSA-4）。ここで、新規化合物指紋追加部102eは、標的タンパク質に対する候補化合物の探索の狙いの必要に応じて任意の化合物指紋（「Modified FP」等）を加えてもよい。また、化合物指紋作成部102aは、結合化合物指紋セット106bに記憶された化合物指紋セットと医薬品化合物DB106cに記憶された化合物指紋セットとの積集合を求めることにより医薬品化合物に似た構造の絞り込みをかけてもよい。

そして、最適化部102bは、候補化合物DB106aに記憶された候補化合物についての相互作用スコアの計算の基底となる座標固定の化合物指紋を結合化合物指紋セット106bから選出する（ステップSA-5）。

5 5 最適化部102bは、候補化合物について、選出した化合物指紋の座標固定の三次元座標を基底として化合物指紋単位の二乗平均偏差を算出して最小自乗フィッティングを行い、当該二乗平均偏差を基礎とする相互作用スコアが最適化されるように、候補化合物の標的タンパク質に対する立体構造を演算する（ステップSA-6）。すなわち、最適化部102bは、相互作用スコア計算部102fの処理により、結合化合物指紋セット106bから任意に選ばれた、標的
10 10 タンパク質の座標固定の化合物指紋を基底として化合物指紋同士の三次元座標の二乗平均偏差を基礎とした相互作用スコアを算出する。そして、最適化部102bは、相互作用スコアを指標として、構造変換部102gの処理により変換された候補化合物のコンフォメーションおよび標的タンパク質に対する構造が最適化されるように、メトロポリス法を基本にしたシミュレティッドアニーリング法を
15 15 実行する。

そして、スクリーニング結果出力部102cは、最適化部102bにより最適化された相互作用スコアに基づいて、候補化合物DB106a中の候補化合物の、標的タンパク質に対する相互作用順位を決定して、インシリコスクリーニングの結果を出力装置114に出力する（ステップSA-7）。例えば、スクリー
20 20 ング結果出力部102cは、最適化部102bにより各候補化合物ごとに得られた最高点の相互作用スコアについて降順に候補化合物群を並べ替えて出力する。

以上で、インシリコスクリーニング装置100の処理が終了する。

[相互作用スコアの算出]

つぎに、相互作用スコア計算部102fによる相互作用スコアの計算方法の一例を以下に説明する。相互作用スコア計算部102fは、化合物指紋単位の二乗平均偏差を基礎とした、候補化合物の、標的タンパク質との衝突具合、標的タン
25 25 パク質の相互作用領域における存在割合、および、標的タンパク質との直接相互

作用割合を考慮に入れた関数に基づいて相互作用スコアを計算する。より具体的には、相互作用スコアは、以下の数式（1）に基づいて算出される。

[数6]

$$\begin{aligned} FPAScore &= F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule}) \\ &= \text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) \\ &\quad \times \text{fp_volume}(\text{molecule}) \\ &\quad \times \text{fp_contact_surface}(\text{molecule}) \quad (1) \end{aligned}$$

- 5 (ここで、FPAScoreは相互作用スコアであり、F(aligned__fp, fp_rmsd, molecule)は、結合化合物と候補化合物間の化合物指紋単位のアライメント度および二乗平均偏差、ならびに、候補化合物の標的タンパク質に対する立体構造を変数とする関数であり、BaseScore(aligned__fp, fp_rmsd)は、化合物指紋単位の一
- 10 致度および密集度を示す指標であり、fp_volume(molecule)は、結合化合物指紋セットの三次元座標からなる空間を候補化合物が占める割合、および、標的タンパク質との衝突具合を示す指標であり、fp_contact_surface(molecule)は、候補化合物の標的タンパク質との接触度、および、結合化合物指紋セットの三次元座標への帰属度を示す指標である。)
- 15 更に具体的には、上記数式（1）における各項は、本実施の形態において以下の数式に基づいて算出される。

<BaseScore(aligned__fp, fp_rmsd)の項>

この項は、化合物指紋単位の一

[数7]

$$20 \quad \text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) = \frac{\text{RawScore}(\text{aligned_fp})}{1 + \ln(\text{fp_rmsd}^{k1} + 1)} \quad (2)$$

(ここで、RawScore(aligned__fp)は、結合化合物と候補化合物間でアライメントされた化合物指紋における原子の数に基づく指標であり、fp_rmsdは、二乗平均偏差である。)

上式の $RawScore(aligned_fp)$ は、具体的には以下の数式 (3) により算出される。

[数 8]

$$RawScore(aligned_fp) = \sum_{i=1}^n assigned_score(i) \quad (3)$$

5 (ここで、 $assigned_score(i)$ は、 i 番目にアライメントされた化合物指紋にあらかじめ与えられた以下の式に基づくスコアである。)

更に詳細には、 $assigned_score(i)$ は、以下の数式 (4) で求められる。

[数 9]

$$assigned_score(i) = \begin{cases} \sum_{j=1}^{total_atom(i)} Case1_S + \ln(n_neighbor_atom(i)+1) \\ \sum_{j=1}^{total_atom(i)} Case2_S + \ln(n_neighbor_atom(i)+1) \\ \sum_{j=1}^{total_atom(i)} Case3_S + \ln(n_neighbor_atom(i)+1) \end{cases} \quad (4)$$

10

15

(ここで、 $total_atom(i)$ はその i 番目にアライメントされた化合物指紋を構成する原子の数であり、例えば、4 原子からなる化合物指紋の場合は 4 である。 $Case1_S$, $Case2_S$, $Case3_S$ は、下記で述べる条件を満たした場合与えられるスカラー値である。 $n_neighbor_atom(i)$ は後述するが i 番目の原子セットに近接する同じ化合物指紋に属する原子の数である。)

例えば、 $Case1_S$ については、結合化合物指紋セットに存在する一つの結合化合物に対して、深さ優先探索 ($depth_first_search$)

(“C アルゴリズム全科 基礎からグラフィクスまで ISBN 4-7649-0239-7 近代科学社” 参照) を 4 原子まで行う (例えば、 $C-ar-N-ar-C-ar-C-ar$ 等の化合物指紋)。本実施の形態では、4 原子までで探

20

索を終えているので、環構造の数は考慮されない。すなわちベンゼン環とナフタレン環は区別されない。探索に成功した場合、化合物指紋を構成する各原子にスコア (Case 1_S) が与えられる。ここでは、一つの原子あたりのスカラー値を 5.0 とする。すなわち、4 原子で構成される化合物指紋には 20.0、3 5 原子なら 15.0 と与えられる。

また、Case 2_S は、Case 1 で得られた化合物指紋を用いて新たな化合物指紋が作成された場合であって、ある一定の距離で重なり合う任意の二つの化合物指紋をえらび、原子を仮想的な結合で結び、新たな化合物指紋を作成し、各原子にある一定のスコアのことである。デフォルトは 2.5 を用いてもよい。

10 また、Case 3_S は、生物化学的情報や、エネルギー計算により原子の存在の可能性がある場合に与えられる任意のスカラー値である。ここで、Case 3_S は、トレーニングセットを使用した検証計算では用いていない。

ここで、上記の Case 1_S, Case 2_S の作成過程で得られた化合物指紋は結合規則情報と原子タイプの識別できる既知医薬品データベースから得ら
15 れる化合物指紋セットに属していなければならない。また、Case 1_S と Case 2_S, Case 3_S の作成過程において、同一化合物指紋に属する座標間において、原子座標セットとそのほかの原子との距離が $dist$ (デフォルトは 1.0 Å) 以内にある原子の個数の自然対数を fp の座標のスコアに加算する。なお、結合化合物において、化合物の殆どがアミノ酸残基のつながったペプ
20 チドの場合、ペプチド基が多く化合物指紋の対応関係が複雑になるので、その対応関係を相互作用スコアの計算過程で過小評価して、Raw Score についての上記の数式において、ペプチドの部分の化合物指紋の数式 (3) に該当する部分をゼロ等の過小評価の数字にしてもよい。

上記数式 (2) の右辺分母は、以下の数式 (5) で求められる。

25 [数 10]

$$1.0 + \ln(fp_rmsd**k1 + 1.0) \quad (5)$$

(ここで、 \ln は、ログナチュラルである。 k_1 は、最適化した結果として4.0を用いる。 fp_rmsd は、最小二乗重ね合わせの時の $rmsd$ である。 k_1 は、 fp の重ね合わせの精度をどこまで厳密にするかをきめるスケール因子であり、大きくした場合に、 $rmsd$ が大きく(悪く)、すなわち数式(3)の $RawScore$ (スコア)が小さくなるような定数である。)

< $fp_volume(molecule)$ の項 >

この項は、結合化合物指紋セットの三次元座標からなる空間を候補化合物が占める割合、すなわち結合化合物指紋セットより得られた化合物指紋からなる空間をどの程度満たしているか、および、標的タンパク質との衝突を評価する関数である。

[数11]

$$fp_volume(molecule) = \ln \frac{1.0 + nafp^{k_2}}{1.0 + nap^{k_3}} \quad (6)$$

(ここで、 $nafp$ (Number of Ligand Atom covering Fingerprint)は、結合化合物指紋セットの三次元座標に基づく固有格子点領域に候補化合物の三次元座標が占有する格子点の数であり、 nap (Number of Ligand Atom covering Protein)は、標的タンパク質の立体構造における原子の固有格子点領域に候補化合物の三次元座標が属する格子点の数であり、 k_2 および k_3 は、それぞれ係数であり、標的タンパク質の生物化学的情報、誘導適合の程度等によって変更可能な任意の定数であり、本実施の形態では、デフォルトに1.0を用いる。)

< $fp_contact_surface(molecule)$ の項 >

この項は、候補化合物の標的タンパク質との接触度、および、結合化合物指紋セットの三次元座標への帰属度を考慮した関数である。

[数12]

$$fp_contact_surface(molecule) = \frac{\sum_{i=1}^n density_of_atom(atom(i))}{total_density_of_atom(molecule)} \quad (7)$$

- (ここで、 n は、候補化合物の原子の数であり、 $atom(i)$ は、候補化合物の i 番目の原子の三次元座標であり、 $density_of_atom(atom(i))$ は、当該原子の三次元座標が結合化合物指紋セットの化合物指紋に属している場合に、化合物指紋の原子と所定の距離で接触している標的タンパク質の原子の数と、化合物指紋の同一格子点に属する結合化合物の原子の数との和を返す関数であり、 $total_density_of_atom(molecule)$ は、 $density_of_atom$ の分布を降順に並べ換えたものを候補化合物の原子の数だけ順に足し合わせた数である。)
- 10 更に詳細には、 $density_of_atom(atom(i))$ は、以下の数式(8)で表される。

[数13]

$$density_of_atom(atom(i)) = \begin{cases} 0 \\ \ln(nfpcontact + natom + hi) \end{cases} \quad (8)$$

- この式において、もし、候補化合物を構成する原子の座標が、結合化合物指紋
15 セット由来の化合物指紋に属していない場合は0となり、属している場合は上記の式に従い、スコアが計算される。

- すなわち、 $nfpcontact$ は、化合物指紋に属している原子とある一定の距離(デフォルトは、3.8)で接触している候補化合物の原子の個数である。また、 $natom$ は、同一格子点に属する結合化合物セット由来の化合物を構成
20 する原子の数となる。同じ結合化合物であって、PDBのIDコードが違う場合について、適宜変更可能となるが、本実施の形態では重複を許して数える。また、 hi は、特に重要な生化学的情報がある場合に使用するものであり、デフォルトでは0を用いる。すなわち、「Circle」などの3D-1D法によって、標

的タンパク質との安定的な接触が示唆された場合に導入されるModified FP (改正FP) によって生じる。

次に、total_density_of_atom (molecule) の数式について以下に記述する。

5 [数14]

$$total_density_of_atom(molecule) = \sum_{i=1}^{total} sort_density_of_atom(i) \quad (9)$$

(ここで、totalは、化合物の原子 (moleculeのatom) 数である。また、sort_density_of_atomは、density_of_atomの分布を大きい方から順に並べ替えたものである。つまり、分子が
10 大きいと大きい数値が加算されるのでtotal_density_of_atomは大きくなる。)

以上で、相互作用スコア計算部102fによる相互作用スコアの計算方法の一例の説明を終える。

[シミュレティッドアニーリングによる相互作用スコアの最大化]

15 つづいて、上述した相互作用スコアの計算方法により計算された相互作用スコアに基づいて、最適化部102bによるシミュレティッドアニーリングに従って候補化合物のコンフォメーションおよび配置を最適化する処理の一例について以下に説明する。

最初に、構造変換部102gは、候補化合物の回転可能な二面角をランダムに変更することにより、コンフォメーションを変化させる。本実施の形態では、コンフォメーション変化は、1000回行う。この数は多ければ多いほど良い結果が得られる可能性があるが、バーチャルな候補化合物DB106aに含まれる多くの低分子化合物についてドッキング計算を行う必要があるため、有限な回数
20 の大きさとする必要があり、候補化合物の回転自由度に依存するとしても予備計算ではこの回数で十分と考える。なお、初期のコンフォメーションは、候補化合物
25 DB106aに登録された、ファミリータンパク質に対する結合コンフォメーション

ョンとしてもよい。最適化部102bは、この変化させたコンフォメーション毎に、以下の処理で候補化合物の座標を用いる。

そして、最適化部102bは、結合化合物指紋セット106bの化合物指紋バンド (f p b a n d s) から、ランダムに10個の化合物指紋を選ぶ。なお、
 5 10個に満たない場合は、化合物指紋バンドの最大数の半分を用いる。より具体的には、選択された化合物指紋バンドから、候補化合物および結合化合物指紋セット106bの化合物指紋の原子座標をランダムに選択する。この状態を、フィンガープリント・アライメント (f i n g e r p r i n t a l i g n m e n t) と呼ぶ。そして、その対応関係で、最小二乗フィッティングを行い、そのとき
 10 の重ね合わせの自乗平均偏差 (r m s d) と重ね合わせ後の候補化合物の原子座標を用いて、上記の式により相互作用スコアを計算する。

そして、構造変換部102gは、繰り返し二回目以降は前回の状態を記憶部106に記憶しておき、候補化合物のコンフォメーションを保ったまま、すなわち候補化合物を剛体として、並進、回転を行い、一つの化合物指紋の増加、減少、
 15 および、原子座標セットの対応関係の変更、追加を行う。本実施の形態では、このステップを10000回おこなう。

この過程において、最適化部102bは、メトロポリス (M e t r o p o l i s) 判定を行う。すなわち、最適化部102bは、前回の相互作用スコアより、今回の相互作用スコアが大きいならば、当該候補化合物の配置を採用 (a c c e p t)
 20 p t) し、反対に、相互作用スコアが小さいならば、以下の数式に基づき採用確率 (P a c c e p t) を計算する。

[数15]

$$\Delta \text{Score} = \text{Score}(\text{今回}) - \text{Score}(\text{前回})$$

$$25 \quad P_{\text{accept}} = \exp \frac{\Delta \text{Score}}{T}$$

すなわち、採用確率 P a c c e p t の範囲は、 $0 < P a c c e p t \leq 1$ とな

るので、最適化部102bは、このとき同時に $0 \leq r \leq 1$ の範囲の一樣乱数を発生させ、 $r < \text{P a c c e p t}$ ならば、相互作用スコアが前回よりも小さい場合も採用する。なお、シミュレティッドアニーリング（焼きなまし）過程において、T（温度）は、30Kからはじめ、0.07Kまで下げる。

- 5 このようにして、最適化部102bは、一つのコンフォメーションの相互作用スコアの最大値を計算し、初期に発生させた1000個のコンフォメーションについて比較し、相互作用スコアが最大の構造を、最適な標的タンパク質-候補化合物複合体（P r o t e i n - L i g a n d c o m p l e x）構造として予測する。このとき、1000個のコンフォメーションを順位付けする過程において、
- 10 ランダムにコンフォメーションを発生させる代わりに遺伝的アルゴリズム等を利用するなどして、以前のコンフォメーションを記憶して何らかのアルゴリズムでリガンド構造を変えていき、計算時間や最大値の探索において工夫を行ってもよい。1000回の計算過程で、リガンドコンフォメーションの順番を決めるために、GOLDプログラムで採用されているような遺伝子アルゴリズム等を使って、
- 15 計算時間の短縮やリガンドコンフォメーションがより真実に近づく可能性のある最小スコアを得ることができる。

以上で、シミュレティッドアニーリングによる相互作用スコアの最大化の説明を終える。

[タニモト指数]

- 20 化合物指紋セットを作る際に、化合物間の類似を計る尺度として、たとえば、タニモト係数（T c）が0.08以上の低分子化合物のセットを用いてもよい。S y b y l原子タイプのような各々の化合物の化合物指紋である化学記述子から化合物指紋（f p）を決める場合、タニモト係数（T c）は下記のように算出する。

- 25 [数16]

$$Tc = \frac{a}{a+b+c}$$

(ここで、aは、化合物指紋が、結合化合物と候補化合物の両方のFPバンド (fp bands) に存在する個数であり、b, cは、fpが片方のFPバンドにのみ存在する個数である。)

同じことを集合 (assembly) を使って説明すると、A, BをそれぞれのFPバンドが持つ化合物指紋の集合とするならば、以下の式になるともいえる。

[数17]

$$Tc = \frac{\text{number_of_fp}(A \cap B)}{\text{number_of_fp}(A \cup B)}$$

(ここで、number_of_fp (assembly) は、ある集合 (assembly) に所属する化合物指紋の数である。)

10 以上で、タニモト指数の説明を終える。

[実施例1]

次に、本発明が適用される本実施の形態の実施例1について、以下に図3～図29を参照しながら詳細に説明する。なお、以下の実施例においては、結合化合物指紋セット106bを「CElib」(FP (fingerprint) set extracted from collected ligands in the binding site) という名称で呼ぶことがある。

[リガンドドッキングについての生物学的情報を半経験的に選択する方法の開発 (Development of choosing biological information semi-empirically on the Ligand Docking)]

近年、計算機の速度の向上により、医薬品開発の分野においてタンパク質の立体構造予測法、およびその立体構造の評価 [参考文献: Terashi G, Takeda-Shitaka M, Kanou K, Iwadate M, Takaya D, Hosoi A, Ohta K, Umeyama H
25 Proteins, 2007, 69 (S8): 98-107] は改良されている。例えば、タンパク質の立体構造の予測法の一つであるホモロジーモデリン

グ (Homology Modeling) は、PDB (Protein Data Bank) [参考文献: Westbrook et al Nucleic Acids Res. 2003 Jan 1; 31 (1) : 489-91] へ登録される構造の増加と、膜タンパク質を除いて参照する鋳型 (Template) の増加と、および、CASP (the Critical Assessment of Techniques for Protein Structure Prediction) におけるブラインドテスト (blind test) によって、その予測精度は上昇している [参考文献: Takeda-Shitaka, M., Terashi, G., Takaya, D, Kanou, K., Iwadate, M., Umeyama, H. Protein structure prediction in CASP6 using CHIMERA and FAMS. Proteins 61, 122-127 (2005)]。そして、当該ホモロジーモデリングは、立体構造予測法の適用範囲は変異 (mutation) の影響による活性変化の予測 [参考文献: 中町祐司, 河野誠司, 矩口真理子, 野口依子, 木下承皓, 加納和彦, 寺師玄記, 竹田一志鷹真由子, 近藤信一, 熊谷俊一, P04-08 “Ala54Thr および Ala249Glu 変異 Antithrombin のコンピュータ・モデリング解析”]、ドラッグデザイン [参考文献: Takeda-Shitaka, M., Takaya, D., Chiba, C., Tanaka, H., & Umeyama, H. Curr. Med. Chem. 11, 551-558 (2004)] などに広がっている。

また、PDBへ登録されるタンパク質の立体構造の増加と共に、タンパク質-リガンド複合体 (Protein-Ligand complex) のX線構造解析結果も増加しており、一つのファミリータンパク質内において、解析済の複数のX線構造が存在することも多い [参考文献: Edgar R. Wood et al CANCER RESEARCH 2004 64 6652-6659, 参考文献: Jennifer et al J. Bio. Chem.

2002 Vol. 277, No. 48, 46265-46272]。
また、前述のCASPにおいても、タンパク質の結合部位 (binding site) の残基を予測する試験を行うなど [参考文献: Lopez, G, Rojas, A, Tress, M, Valencia, A Proteins, 2007, 69 (S8) : 165-174]、タンパク質-リガンド複合体 (Protein-Ligand complex) の予測精度の向上の重要性は高まりつつある。

一方で、近年、疾病原因タンパク質の実験的決定が盛んであり (参考文献: Nature など)、そのタンパク質を阻害する阻害剤の設計の必要性はますます高まっている。

阻害剤の設計のための有力な方法として、標的タンパク質の立体構造に基づいた阻害剤設計 (SBDD) があり、タンパク質-リガンド複合体 (Protein-Ligand complex) 予測ソフト (いわゆるドッキングソフト) を用いたインシリコ (In-silico) スクリーニングが行われている。ここで、図3は、従来のドッキングソフトと、タンパク質-リガンド複合体の多数のX線構造やNMR構造を効果的に用いたバイオインフォマティクスによる本実施例に係るドッキング方法を示す状況図である。

図3に示すように、既存のドッキングソフトにおいては、AutoDock [参考文献: Goodsell et al J. Mol. Recognit 1996 9 1-5], DOCK [参考文献: Ewing et al J Comput Aided Mol Des. 2001 15 (5) 411-28], GOLD [参考文献: Gareth et al J. Mol. Biol. 1997 267, 727-748] などは水素結合、疎水性相互作用、静電相互作用といった古典物理的なポテンシャル関数を用いた第一原理的アプローチ (Ab-initio Approach) を採用している。さまざまな検証によって、これらの既存のソフトはよい精度でドッキングできている (たとえば正解構造を隠したブラインドテスト (blind test)) によ

- って正解構造に $rmsd \leq 2.0$ 以下で予測できる割合を検証されている) [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1609-1618, 参考文献: Michael et al J. Med. Chem. 2007, 50, 726-741]。
- 5 また、回転可能な結合が多い化合物を精度よくドッキングするために、リガンド結合部位 (ligand binding site) に予め化合物のフラグメントを、ポテンシャル関数をもちいて配置しておくといった方法も考案されている [参考文献: Budin et al Biol Chem. 2001 382 (9), 1365-72]。
- 10 既存のドッキングソフトを用いて仮想化合物ライブラリーから、標的タンパク質に阻害剤候補化合物をドッキングし、タンパク質-リガンド複合体 (Protein-Ligand complex) の構造を予測した後、ヒット化合物 (Hit Compound) を選ぶために、既知のタンパク質-リガンド複合体 (Protein-Ligand complex) の構造からタンパク質とリ
- 15 ガンド間の距離、古典物理学的エネルギーの計算などを行い、相互作用情報を抽出し、ヒット化合物を数多く選ぶための再評価を行う試みも多く報告されている [参考文献: Sukumaran et al Eur. J. Med. Chem. 2007, 42, 966-976, 参考文献: Zhan et al J. Med. Chem. 2004, 47, 337-344]。
- 20 しかし、上記一連の研究が示していることは、既存のドッキングソフトはよい精度でタンパク質-リガンド複合体 (Protein-Ligand complex) を予測できるものの、そのことが直接にヒット化合物 (Hit Compound) を数多く仮想化合物ライブラリーから選ぶこととは一致しない (直結しない) ことを意味している。
- 25 すなわち、現在、タンパク質-リガンド複合体 (protein-ligand complex) の構造を精度よく予測できる一方で、なおかつ、バーチャルライブラリーからヒット化合物 (Hit Compound) を多く検出できる

システムを開発することが非常に要請されており、創薬において必要不可欠である。

そのような状況の中、本願発明者は、タンパク質-リガンド複合体 (Protein-Ligand complex) 相互作用の評価に古典物理学的なポテンシャル関数を用いず、PDBに登録されている相互作用既知のタンパク質-リガンド複合体 (Protein-Ligand complex) の生物化学的情報から効率的に有効な情報を選び出し、ドッキングを行いタンパク質-リガンド複合体 (Protein-Ligand complex) の構造を予測し、かつ、ヒット化合物 (Hit Compound) を多く検出できるシステム ChooseLD (Choose information Semi-Empirically on the Ligand Docking) を開発した。

また、本願発明者の方法では、タンパク質-リガンド複合体 (Protein-Ligand complex) の相互作用評価において、古典物理学的ポテンシャル関数を使用していない。したがって、本発明の方法は、相互作用の物理学的エネルギーが最適化されているとは言えないタンパク質-リガンド複合体 (Protein-Ligand complex) 構造の最適化において物理的なアプローチであるCHARMM [参考文献: Brooks, R. B., Bruccoleri, E. R., Olafson, D. B., States, J. D., Swaminathan, S. & Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations J. Comp. Chem. 4 187-217 (1983)], AMBER [参考文献: Case, A. D., Cheatham III, E. T., Darden, T.,ohlke, H., Luo, R., Merz Jr., M. K., Onufriev, A., Simmerling, C., Wang, B. & Woods, J. R. The Amber Biomole

ular Simulation Programs J Comput Chem 26 1668-1688 (2005)] および、量子化学 [参考文献: Fedorov, G. D. & Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method J. Phys. Chem. 111 6904-6914 (2007)] が有効に機能すると期待している。

[本実施例1の概要]

10 ここで、本実施例の概要について、以下に図4を用いて説明を行う。図4は、本実施例(ChooseLD)によるタンパク質-リガンド・ドッキングの原理構成図である。ここで、本実施例において、ライブラリーリガンド(LIBRARY LIGANDS)は結合化合物の集合に相当し、CELibは結合化合物指紋セット106bに相当する。

15 ここで、図4において、各円柱は、データの集合を表しており、楕円は入力情報、長方形は出力構造を示している。平行四辺形は、化学記述子としての化合物指紋(FP: fingerprint)である。すべての過程は計算機(インシリコスクリーニング装置100)上でおこなわれるため、入力する情報は電子情報としてのファイルである。すなわち、PDB形式に代表されるような形式で記載された標的タンパク質の三次元座標ファイル、ドッキングされるリガンドの三次元座標ファイルを想定している。

25 図4において、矢印はおもにデータの集合の絞り込みや入力情報の修飾といった変換操作を意味しており、変換操作には詳細な条件を指定することができる。ただし、これらの変換操作には既定の値を定めており、入力情報がファイル形式的に、かつ、入力されたタンパク質の座標が物理化学的に正常であるならば、全自動で出力を得ることができる。すなわち、標的タンパク質の三次元座標ファイルとドッキングされる候補リガンドの三次元座標ファイルを入力したならば、タ

タンパク質ーリガンド複合体構造の三次元座標ファイルが出力されるということである。タンパク質の三次元座標および、アミノ酸配列は、ホモロジー検索、結合化合物指紋セット106bに相当するFPライブラリーの構築、ドッキング計算のためのタンパク質立体構造の三次元座標として用いられ、ターゲットの候補リガンドは、候補化合物に相当し、候補タンパク質特異的FPバンド、リガンドの三次元コンフォメーション探索に使用される。

すなわち、図4に示すように、まず、本実施の形態に係るインシリコスクリーニング装置100は、相同性検索部102dの処理により、標的タンパク質についてPDB等のタンパク質構造データベースに対して、相同性検索を行い、化合物指紋作成部102aの処理により、相同なタンパク質と構造アライメントにより重ね合わせ(fitting)を行い、標的タンパク質の座標系に変換した三次元座標とともに化合物指紋を抽出して、結合化合物指紋セット106bに相当する標的タンパク質指向性リガンド群(C)を作成する。

そして、インシリコスクリーニング装置100は、標的タンパク質指向性リガンド群(C)を、医薬品化合物DB106cに相当する医薬品的(druggable)FPデータベース(D)に照会し、積集合 $(C) \cap (D)$ として標的タンパク質特異的FPバンド(L)を得る。ここで、標的タンパク質指向性リガンド群(C)には、新規化合物指紋追加部102eの処理により、Modified FP等の仮想FPを追加していてもよい。

つづいて、インシリコスクリーニング装置100は、仮想リガンドライブラリーまたはベンチマークセットの、標的タンパク質とドッキングを行うリガンド(docked ligand)である候補リガンドから化合物指紋を抽出し、候補化合物DB106aに相当する候補リガンドのFPバンド(R)を作成する。

そして、インシリコスクリーニング装置100は、構造変換部102gの処理により、候補リガンドのコンフォメーションを変化させ、標的タンパク質指向性リガンド(C)と候補リガンドのFPバンド(R)間でFPアライメントを行う。

そして、インシリコスクリーニング装置100は、最適化部102bの処理に

より、相互作用スコア関数を用いて標的タンパク質の結合部位に候補リガンドをドッキングさせる場合に、シミュレティッドアニーリング (SA) 法を用いて相互作用スコアを最適化させながら、標的タンパク質-候補リガンド複合体の三次元構造予測を行う。以上が本実施例の概要である。

5 [ライブラリーリガンド]

ライブラリーリガンド (LIBRARY LIGANDS) とは、結合化合物の集合に相当するものである。すなわち、インシリコスクリーニング装置 100 は、PSI-Blast [参考文献: Altschul et al Nucl e i c A c i d s R e s . 1 9 9 7 2 7 (1 7) 3 3 8 9 - 4 0 2]
10 による相同性 (Homology) 検索によって検出されたタンパク質の中で、それがタンパク質-リガンド複合体 (Protein-ligand complex) であった場合、立体構造アライメント発生プログラムであるCE [参考文献: Shindyalov et al Protein Engineering 1998 11 (9) 739-747] を用いて、標的タンパク質と
15 相同タンパク質間のアライメントを行い、最小二乗法 (least square fitting) により標的タンパク質に重ね合わせる。そして、ライブラリーリガンドは、その最小自乗フィッティングによる Z-Score が 3.7 以上となった場合、結合リガンドを標的タンパク質の座標系に変換し、結合リガンドだけを抜き出したものである。

20 なお、本実施例では、Z-Score 3.7 未満は、結合化合物として使用されない。この数値の根拠はCEによると、" 3.7-4.0 - twilight zone where some similarities of biological significance can be seen ;
" (生物学的意義の共通性を見出せるか中間的な領域) とあるためであり 3.7
25 以上を採用した。ホモロジー検索の最低ホモロジーは、本実施例では、相同性 (Homology) 0.1% 以上とした。つまりホモロジー検索で検出された類似タンパク質のほとんどがCEによって重ね合わされることになる。

[FPの定義およびFPバンドの構築]

FPバンド (f p b a n d) の作り方について、以下に図5を参照して詳細に説明する。ここで、本実施例で使用する化合物指紋 (f p : f i n g e r p r i n t) を定義する前に、化合物指紋の解釈について説明する。化合物指紋 (f i n g e r p r i n t, 以下、「FP」という。) は、ケムインフォマティクスの分野において、化合物の特徴を表すベクトルや化合物間の類似性算出のために使用される計算機上の表現法の一つである (Swamidass, S. J. & Baldi, P. Mathematical Correction for Fingerprint Similarity Measurements to Improve Chemical Retrieval J. Chem. Inf. Model. 47, 952-964 (2007))。

本実施例では、FPの正確な解釈を目的としていないが、混乱をさけるために下記の用語に統一する。一つの分子を原子型 (または原子タイプ)、原子結合の順番などを考慮した組み合わせを要素に持つベクトルで表現した場合、ベクトルの要素を「FP」、ベクトルを「FPベクトル」とした。本実施例では、ベクトルの要素に、単に原子型の文字列表記以上の情報が付加されている場合があるが、その付加情報も分子を表現する特徴の一つであると解釈し、そのベクトルの要素を意味する場合も「FP」とし、そのFPを要素に持つベクトルを通常の「FPベクトル」と区別して「FPバンド」とした。このことは、「FPバンド」が「FPベクトル」における各要素が原子型であるといった性質も併せて持つことになる事を意味する。ここで、図5は、FP (f i n g e r p r i n t) の作り方を一例として示す図である。

本実施例であるChooseLD法では、相互作用が既知のタンパク質—リガンド複合体構造を用いて、自由エネルギーの最小化を満たすようにドッキングする未知のリガンド構造を予測することを目的として、この目的を達成するために、相互作用が既知のリガンドから部分的な結合自由エネルギーを保持した部品であるFP (f i n g e r p r i n t) を定義した。図5に一例として示す化学物質

の物質名は、AZD2171 (Cancer Res 2005; 65: (10), May 15, 2005) である。図5に示すように、与えられた結合規則情報を用いて原子をたどることによりFPを作成する。たどる原子の数は、2, 3, 4個である(この数には理由があるので後述する)。各々の囲んだ線は算出されるFPを意味している。aで示すFPは、2個の原子をたどった場合であり、bで示すFPは、3個の原子をたどった例である。cとdで示すFPは、それぞれ4個の場合であり、同じ原子を通過しているが、この場合も許容される。eで示すFPは、異なる座標であるが同じ原子種をたどっており、後述の相互作用スコア関数のFPの重複度が加算される。

すなわち、図5の化合物の結合上の線を囲んだ部分は、ChooseLD法および、化合物の類似性の比較でも用いられるFPの原子型表記を意味する。化合物の上の任意の原子を基点として深さ優先探索法を用い(Chiba et al C algorithm ZENKA 1995 ISBN4-7649-0239-7)、与えられたリガンドの原子間結合情報に従い原子を通過するが、通過する結合の数は、1, 2, 3本とした。すなわち、ベンゼン環とナフタレン環からは同じ原子型表記が構築されることになり、環構造違いは区別されない。一つの原子は、Sybyl Atom Type (Tripos Inc., 1699 South Hanley Road, St Louis, MO 63144-2913, USA (<http://www.tripos.com>))を用いて表現され、これらにはAMBER99 (J. Comput. Chem. 26, 1668-1688 (2005))を参考にした原子量、原子半径、結合可能数が定義されている。この時点では、FPの原子型のみを考慮しており、通過した原子座標は考慮していない。ここで、図6は、本実施例で

25 [タニモト係数による化合物間の類似性算出]

タニモト係数による化合物間の類似性算出方法について以下に説明する。ここで、図7は、タニモト係数による化合物間の類似性算出方法を示す模式図である。

- 本実施例では、化合物間の類似性を算出するためにタニモト係数（以下、 T_c ）を導入した（*J. Chem. Inf. Comput. Sci.* 40, 163-166 (2000)）。一般に、 T_c は二つのビット、すなわち0もしくは1、からなるベクトルの類似度を数値化したものである。図7に示すように、本実施例では、対象となる一つの低分子化合物に対して、上記で導入したFP構築法を用いてFPベクトルを作成し、ベクトル上に定義されたFPが存在するなら1を、存在しないなら0を与えた。このようにして作成された同じ長さで、かつ、対応する成分は同じFPを意味する二つのベクトルから化合物間の類似性を評価した。
- 10 T_c は、下記の数式によって算出した。ここで、両方のベクトルの対応するビットが共にonの場合、 a に1が加算され、片方のベクトルのみビットがonならば **b** もしくは **c** に1が加算される。すなわち、お互いにoffの場合の **d** は加算されず、 T_c 算出において考慮しないことになる。例えば、図7に示した2つのビット列間では、 $a=9$ 、 $b+c=7$ であり、 $T_c=9/(9+7)=0.5625$ となる。
- 15

[数18]

$$T_c = \frac{a}{a+b+c}$$

- 本実施例では、FPバンド（*fp bands*）は、結合化合物のライブラリーリガンド（*LIBRARY LIGANDS*）に属する低分子化合物の集合から得て、集合を形成する低分子化合物由来の、ある二つのFPバンド（*fp bands*）を比較する際は、タニモト係数（ T_c ）が0.08以上でなければならないこととした。換言すれば、上記数式において、 a は、FPが両方のFPバンドに存在する個数（*the number of fp existing in each fp bands*）である。また、 b 、 c は、FPが片方のF
- 20

Pバンドにのみ存在する個数 (the number of fp existing in the other fp band) である。

同じことを集合 (assembly) を使って説明すると、A、Bをそれぞれのバンドが持つFPの集合とするならば、以下のように表せる。

5 [数19]

$$T_c = \frac{\text{number_of_fp}(A \cap B)}{\text{number_of_fp}(A \cup B)}$$

ここで、number_of_fp (assembly) はある集合 assembly に所属する fp の数である。

[FPライブラリーの構築]

- 10 FPライブラリーとは、結合化合物の集合に相当し、本実施例の Choose LD法で用いられるFPの原子型表記の入手源であり、さらに構築されたFPに登録される原子座標の起源となるリガンド群のことである。通常、標的タンパク質の一次構造、すなわちアミノ酸配列をクエリーとしたホモロジー検索等で検出
- 15 されたファミリータンパク質から収集するが、ファミリータンパク質に限らず標的タンパク質の活性部位等の標的部位に結合すると考えられるリガンド、もしくはタンパク質、ペプチド等であっても、必要であれば追加可能である。

- 本実施例の Choose LD法では、主にファミリータンパク質からFPライブラリーを構築した。PSI-Blast (Nucleic Acids Res. 27, 3398-3402 (1997)) によるホモロジー検索によ
- 20 って検出された三次元座標構造が既知のタンパク質において、タンパク質—リガンド複合体であった場合、CE (Protein Engineering 11, 739-747 (1998)) を用い、標的タンパク質とファミリータンパク質との立体構造アライメントをおこなう。CEは、二つのタンパク質をアミノ酸配列類似性によらず、立体構造的に類似した部分を用いてアライメントを

おこなうアルゴリズムを実装したプログラムであり、他の立体構造アライメントのプログラムには、D a l i (J . M o l . B i o l . 2 3 3 , 1 2 3 - 1 3 8 (1 9 9 3)) , T O P O F I T (P r o t e i n S c i e n c e 1 3 , 1 8 6 5 - 1 8 7 4 (2 0 0 4)) 等が存在する。これらの主な違いを記述すると、C E はアミノ酸配列をN末端から順に重ね合わせる等の改良により、高速に結果を得ることが可能であるが、対象タンパク質にドメインスワッピング等が存在する場合には精度よくアライメントすることが難しく、その場合アミノ酸配列の順番に依存しないアライメントを行うD a l i 等をもちいた方が精度がよい。

10 本実施例のC h o o s e L D法では、P S I - B l a s tで検出されたファミリータンパク質を主に重ね合わせることから、計算時間が短いC Eをもちいた。C Eが出力するアライメントを用い、最小自乗フィッティングにより標的タンパク質に重ね合わせた。C EのアライメントのZ - S c o r eが3.7以上となった場合、結合リガンドを標的タンパク質の座標系に変換し、結合リガンドだけ抜き出した。すなわち、本実施例では、標的タンパク質と構造的に類似しているタンパク質のみがファミリータンパク質として使用されることになる。

[F Pバンドの構築]

20 F Pバンドは付加情報として、一つもしくは複数の原子座標を関連付けたF Pのベクトルであり、F Pライブラリーに属する結合リガンドの集合から得る。得られた集合(F Pライブラリー)に属する結合リガンドには、標的タンパク質の座標系における座標、および、S y b y l原子タイプ(A t o m T y p e)で表現される原子型および、単結合、二重結合、芳香環結合といった結合規則情報を含む。ここで、図8は、標的タンパク質の結合部位にリガンドをドッキングさせる場合のF Pを一例として示す模式図である。図8において、幾つかの幾何学的図形(長方形や菱形や楕円)で構成された半透明の部分は、各種のF Pを表している。

“I n t r a - m o l e c u l e F P”(図8の長方形)は、リガンド分子

内の情報のみを用いて構築されたFPのことであり、FPライブラリーに属する一つのリガンドの内部のみから得られた原子型情報と結合情報を持ちいて作成されたFPのことである。一つのFPは、リガンド分子内の一つの原子を起点として前述したFPの原子型表記の構築法に基づき、結合している原子を1, 2または3回通過して、図8のような分岐のしない最大4つの原子を構成する。本実施例で最も小さいFPは2原子からなる。一回のFP構築の試行の中で、一度たどった原子はその試行中に二回通過することは無く、通過する結合が無くなった場合は、その時点でのFPの原子型表記と原子座標をFPバンドに登録する。そのFPがすでにFPバンドに登録されている場合は除外するのではなく、一つのFPに複数の原子座標に登録する。ここで、図9は、たどった経路から原子座標を得て、FPバンドに登録する過程を一例として示す図である。

図9において、下の行列は原子座標を意味しており、その行数はFPを構成する原子の個数を表現している。例えば、4行3列からなる行列ならば、そのFPに4つの原子座標を含んでいることを表す。

“Modified FP” (図8の菱形) は、与えられた結合情報と近接する原子同士を仮想的な結合と仮定して作成されるFPのことである。結合している原子および、実際には結合していないが、特に指定が無い限りは1 Å以内に原子が存在すれば仮想的な結合と判定し、結合を1, 2もしくは3回通過して、分岐のしない最大4つの原子からなるFPを構築する。本実施例では、最も小さいFPは2原子からなる。“Intra-molecule FP”の構築の操作と同様に一回のFP作成の試行の中で、一度たどった原子は二度通過することはなく、通過する結合が無くなった場合は、その時点でのFPの原子型表記と原子座標をFPバンドに登録する。これにより、リガンド分子内の結合に加え、リガンド分子間の結合を含んだFPを作成しているため、実際には存在しないようなFPが得られる。すなわち、物理化学的に存在し得ないような結合のFP (例えば、N. a m, N. a m, N. a m, N. a mのようなFP) が構築されることが考えられる。

そこで、本実施例では、物理化学的に存在する医薬品の三次元座標データベースであるMDL Comprehensive Medicinal Chemistry (MDL CMC) Library (医薬品化合物DB106cに相当する。) から、ドラッグライクなFPベクトルを作成し、FPライブラリーより得たFPバンドのFPベクトル部分と比較し、両方に含まれているFPの原子型表記が標的タンパク質特異的FPバンドに残るようにする。任意のFP (fingerprint) を使う計算の過程で、医薬品データベースや化合物データベースを使って、化合物情報を引き出すことにより、この元になるデータベースを薬物吸収や薬物代謝や薬物排泄や薬物毒性等を指標にして、fingerprint (FP) 等を整理の基底としての基礎データ単位を使って、予め整理した薬物吸収や薬物代謝や薬物排泄や薬物毒性に特化した医薬品データベースや化合物データベースを作成して、同じ一連の操作を行う。

具体的には、リガンドライブラリー由来のFPベクトルと、医薬品ライブラリー由来のFPベクトルとの積集合を求めることにより、医薬品化合物DB106cに存在するFPのみがFPバンドに登録され、医薬品化合物DB106cに存在しないFPは本実施例では無視されて、結合化合物指紋セットI06bが構築される。ここで、図10は、本実施例におけるFPバンドの絞り込みステップ (method step of shrinking fingerprint band) を一例として示す図である。

図10に示すように、MDL CMC Libraryより得たFPバンド (A) と、標的タンパク指向性リガンド群より得たFPバンド (B) を比較し、両方にFPが存在する場合以外は、(A) あるいは (B) のFPバンドから取り除く (図10のX印で表される)。この結果、ライブラリーリガンド由来のFP (Library Ligand FP) には、必ず座標が存在することになる。

以上で、本実施例におけるFPバンドの構築方法の説明を終える。なお、本実施例においては、すべてのFPバンド構築の過程において、一つの原子が複数のFPに所属することは許容される。また、FPバンドに得られたFPがすでに登

録されていたなら、FPの座標が追加され、存在しない場合は、FPバンドに新たなFPを追加し、座標を追加する。また、一つの原子が複数のFPに所属することは許容される。ドッキングのターゲットとなる候補リガンド (docked ligand) に対しても同様の操作を行い、候補リガンド由来のFPバンド
5 (fp bands of docked ligand) が作成される。

[FPバンドのアライメント]

FPバンドには原子セットの座標が関連付けられており、二つのFPバンドを比較する際は、単に原子型だけを用いるのではなく関連付けられた座標も用いる。すなわち、FPバンドのアライメントは、候補リガンドから得られたFPバンド
10 と、結合リガンドのFPライブラリーから得られたFPバンドとの比較を行うことを意味する。比較は、以下の(1)、(2)の過程を経ておこなわれる。

(1) FPを構成する原子型表記の文字列の完全一致の比較

ドッキングさせる候補リガンドから得られたFPバンド由来のFPベクトル (ビット列(1)) と、結合化合物を含むFPライブラリーから得られたFPバンド由来のFPベクトル (ビット列(2)) において、FPの有無をビット化し、
15 双方のビットがonである組み合わせを選択する(図7参照)。

(2) 選択されたFPに登録されている原子の座標ベクトル同士に対応関係を与える過程

図11は、座標ベクトル同士に対応関係を与える過程を一例として示す模式図
20 である。一つのFPは、ドッキングされる候補リガンド分子由来の原子座標ベクトル(1)と、FPライブラリーの結合リガンド由来の原子座標ベクトル(2)からなり、この原子座標間に対応関係を与える。

これら二つの過程(1)、(2)を行うことが、本実施例におけるFPのアライメントである。また、「FPアライメントが異なる」とは、

- 25
1. 二つのビットが共にonであるFPの総数
 2. 対応させるFPの種類
 3. FP内部における座標の対応関係

のうち少なくとも一つが異なることを意味する。すなわち、「FPアライメントを変化させる」とは、これらのうち少なくとも一つを変化させることを意味する。

「少なくとも一つ」という意味は、FPの原子型が変化した場合、変化前のFPの座標の対応関係が消失し、変更後のFPにおいて対応関係を与え直すため、必然的に座標の対応関係も変化するからである。

[相互作用スコア (FPAScore)]

本実施例における相互作用スコアFPAScoreについて、以下に詳細に説明する。FPAScore (fingerprint alignment score) は、本実施例において、FPが部分結合自由エネルギーの集合であるというChooseLD法の仮定に基づき、FPAScoreが高いほど、相互作用が既知のファミリータンパク質—結合リガンド複合体構造を満たすように定義した。FPAScoreは、FPの重ね合わせの精度と、アライメントに用いたFPの数、FPの密集度、および、タンパク質—リガンド複合体相互作用を同時に考慮し、標的タンパク質—候補リガンド複合体構造を評価する。本実施例では、前述の操作で得られたFPバンドの最適なアライメントを探索することにより、最適な標的タンパク質—候補リガンド複合体を予測した。

すなわち、本実施例において、相互作用スコアFPAScoreは、以下の数式として定義した。ここで、aligned_fpは、アライメントされたFP、fp_rmsdは、そのアライメントを用いた最小自乗フィッティングによって算出されたrmsd、moleculeは、候補リガンドが標的タンパク質にドッキングした後の複合体の座標を意味する。各項については以下に詳細に説明する。

[数20]

$$\begin{aligned} \text{FPA Score} &= F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule coordinate.}) \\ &= \text{BaseScore}(\text{fp_rmsd}, \text{aligned_fp}) \\ &\quad * \text{fp_volume}(\text{molecule}) \end{aligned}$$

* fp_contact_surface(molecule)

< 1. BaseScore (fp_rmsd, aligned_fp) の項
>

- 5 この項は、FPの一致度および密集度を考慮した関数として定義されたものであり、すなわち、既知のFPの使用強度を評価する関数であり、以下の数式で表せる。

[数 2 1]

$$10 \quad \text{BaseScore}(\text{fp_rmsd}, \text{aligned_fp}) = \frac{\text{raw_score}(\text{aligned_fp})}{1.0 + \ln(\text{fp_rmsd}^{**k1} + 1.0)}$$

- ここで、lnは、ログナチュラル（自然対数）である。また、k1は、FPの重ね合わせの精度をどこまで厳密にするかをきめるスケール因子である。アライメントされたFPの重ね合わせのrmsdが大きい場合、分母が大きくなりBaseScoreが小さくなる。FPの一致度が大きくとも、そのFPに登録されているFPの原子座標の重なるの精度を示すrmsdが大きい(悪い)場合を排除することを意味する。本実施例では、k1を4.0とした。fp_rmsdは、そのアライメントを用いた最小自乗フィッティングによって算出されたrmsd
15
20 である。aligned_fpは、そのときのfpの対応関係、すなわちアライメントされたFPである。

- ここで、上記数式において、raw_score(aligned_fp)は、以下の式で表せる。ここで、assigned_score(i)は、i番目にアライメントされたFPにあらかじめ与えられるスコアである。nは、アライメントされたFPの総数である。アライメントされたFPとは、標的タンパク質特
25 異的FPバンドにおける原子型と原子座標のセットを意味している（上記「FP

バンドのアライメント」および図11参照)。すなわち、FPのアライメントにおいてFPが同じ原子型であっても、原子座標が異なっていれば異なるFPを意味する。

[数22]

$$\text{raw_score}(\text{aligned_fp}) = \sum_{i=0}^n \text{assigned_score}(i)$$

ここで、上記数式において、 $\text{assigned_score}(i)$ は、 i 番目にアライメントされたFPにあらかじめ与えられたスコアであり、以下の数式で表せる。このスコアは、CElib等のリガンドライブラリーより得られたFPに対して下記のように与えられる。

[数23]

$$\text{assigned_score}(i) = \left\{ \begin{array}{l} \text{(total_atom} \\ \text{(} \sum_{j=0}^{\text{Case1_S} + \ln(N+1)} \dots \text{case1} \\ \text{(} \quad j=0 \\ \text{(total_atom} \\ \text{(} \sum_{j=0}^{\text{Case2_S} + \ln(\text{Neighbor_atom} + 1)} \dots \text{case2} \\ \text{(} \quad j=0 \end{array} \right.$$

ここで、上記数式の $\text{total_atom}(i)$ は、FPを構成する原子座標の個数を表す。Case1_S, Case2_S, Case3_S (上記せず) は、あらかじめFPを構成する原子に与えられるスコアであり、それぞれ下記の場合に用いられる。

Case1_Sは、前述の“*Intra-molecule FP*”を構成した場合に各原子に与えられるスコアである。特に指定が無い場合は5.0を用いる。例えば、探索に成功した場合は、FPを構成する各原子にスコアCase1_S (デフォルト5.0を用いた) が与えられ、4原子で構成されるFPには2

0. 0、3原子なら15. 0点が与えられる。

次に、Case 2__Sについて述べる。前述の“Modified FP”を構築した場合に各原子に与えられるスコアである。特に指定が無い場合は2. 5を用いる。

5 最後に、Case 3__Sについて記述すると、生物化学的情報やエネルギー計算（「circle」など）により原子の存在の可能性がある場合に与えられる任意のスカラー値のことである。Case 3__Sは、本実施例では用いておらず、ベンチマークセットを使用したドッキング性能(結合モード予測性能)検証計算、およびインシリコスクリーニング性能で用いていない。

10 本実施例では、Case 1__S, Case 2__S, Case 3__Sの和のスコアに加え、FPライブラリーに属する原子の密集度の自然対数値をスコアに加えた。これはFPに属する原子座標セットの原子と1. 0 Å以内にあるその他のFPに属する原子座標セットの原子個数 ($n_neighbor_atom(i)$) の自然対数をFPのスコアに加算するものであり、この項は密集しているFP
15 を優遇する項であるといえる。すなわち、case 1とcase 2において、同一FPに属する座標間において、距離がdist (デフォルト 1. 0 Å) 以内にある原子座標セットの原子個数 ($Neighbor_atom$) の自然対数をFPの座標のスコアに加算することとした。

<2. fp_volume (molecule) の項>

20 この項は、アライメントされたFPを用いて候補リガンドが標的タンパク質にドッキングした後、その複合体構造を評価する関数である。すなわち、ドッキング後の候補リガンドの分子座標がFPライブラリーの結合リガンドから得られたFPからなる空間を占有する個数（すなわち、FPライブラリー由来のFPからなる空間をどの程度満たしているか）と標的タンパク質との衝突を評価する関数
25 であり、以下の数式で表せる。ここで、moleculeは、候補リガンドのドッキング後の原子座標を表す。

[数24]

61

$$\text{fp_volume}(\text{molecule}) = \ln \frac{1.0 + \text{nafp} ** k2}{1.0 + \text{nap} ** k3}$$

5 ここで、nafp (Number of Ligand Atom covering Fingerprint) は、ライブラリーリガンド (LIBRARY LIGAND) を構成する低分子の原子を用いて作成された固有格子点領域に分子 (molecule) の座標が占有する個数、すなわち候補リガンドが FPライブラリーを構成する結合リガンド原子を用いて作成された固有格子点領域の座標を占有する個数である。nafpにより、候補リガンド分子 (molecule) が座標固定のFP (fingerprint) をどれだけ満たしているか表している。nap (Number of Ligand Atom covering Protein) は、標的タンパク質の原子座標より作成される固有格子点領域にmolecule (ドッキング後の候補リガンド分子) の座標が属する数であり、標的タンパク質の構成原子との衝突具合を表現している。

15 また、k2, k3は、それぞれ係数であり、特に指定のない場合 (デフォルト) では、それぞれ1.0を用いるが、それぞれ標的タンパク質の生物化学的情報、誘導適合の程度によって変更可能である。すなわち、k2は、その標的タンパク質のファミリータンパク質の結合リガンド集団の空間を占有する領域を重視する定数であり、係数が増大するならば、大きなリガンドが大きなスコアを得ることができるようになる。k2値は、標的タンパク質の結合領域の大きさによっても、グループ化できる可能性がある。また、k3は、その標的タンパク質の占有する領域に候補リガンドが衝突することの許容度因子であり、候補リガンド原子と標的タンパク質原子の衝突を重視する係数である。k3値が大きくなれば、標的タンパク質と候補リガンドとの衝突を許さないことになる。k3について、タンパク質 (protein) の活性部位の柔らかさなどをグループ化できる可能性がある。ここで、図12は、nafpとnapの具体例を原子数が31のリガンド

を用いて示した図である。

図12に示すように、候補リガンドにおいて標的タンパク質と衝突する原子数が10個、FPライブラリー由来の格子点に原子が21個所属し、 k_2 値、 k_3 値が1.0であるならば、 $fp_volume(molecule)$ の項は
 $\ln(22/11) = 0.693$ という値になる。この項の関数の性質上、 n_{afp} が31から30、即ち衝突の個数が0個から1個における変化率がもっとも大きい。またリガンド原子の半分近くが衝突している場合は、負値となるため非常に採用されにくくなる。すなわち、FPAScoreにおいて、経験的物理関数である分子間引力—反発項を表現するレナードジョーンズポテンシャルに対応する
 10 ものとして定義されている。なお、EGFRを標的タンパク質として用いたインシリコスクリーニング性能についての項で、 k_2 値、 k_3 値の最適化の一例の結果を後述する。

<3. $fp_contact_surface(molecule)$ の項>

次に、 $fp_contact_surface$ の項は、候補リガンドのドッキング後の構造に対してその原子座標の標的タンパク質への接触度、および、その座標のFPライブラリーへの帰属度を考慮する関数であり、以下の数式で表せる。ここで、 $molecule$ は候補リガンドのドッキング後の原子座標、 $atom(i)$ は、そのドッキング後の*i*番目の原子座標、 n は原子数を意味する。すな
 15 わち、この式は、上述の fp_volume の数式と同様に、候補リガンドが標的タンパク質へドッキングした後の複合体構造に対して計算され、候補リガンド原子座標の標的タンパク質の表面との接触度、およびFPライブラリーから得られたFP原子に対しての候補リガンド原子座標の帰属度を考慮する関数である。

[数25]

$$fp_contact_surface(molecule) = \frac{\sum_{i=0}^n density_of_atom(atom(i))}{total_dense_of_atom(molecule)}$$

上記数式において、density_of_atomは、以下の数式で表せる。
 ここで、nhpcontactは、FPライブラリーに属しているFPの原子座
 標と、特に指定が無い限り（デフォルトでは）3.8 Å以下で接触している標的
 タンパク質の原子の個数であり、natomは同一格子点に属するFPライブラ
 5 リー由来の結合リガンド化合物の原子の数となる。このとき、同じ原子型のリガ
 ンド分子が複数に存在していてもよく、同じリガンド分子であって、PDBのI
 Dコードが違う場合についても、本実施例ではすべて取り込む。hiは特に重要
 な生化学的情報がある場合は使用する変数であり、特に指定の無い場合は（デフ
 オルトでは）0を用いるが、CIRCLE (Terashi G, Taked
 10 a-Shitaka M, Kanou K, Iwadate M, Tak
 aya D, Hosoi A, Ohta K, Umeyama H Pr
 oteins, 2007, 69 (S8) : 98-107) 等の3D-1Dスコ
 ア値によってファミリータンパク質に依存しないFP (Modified FP
 やCreative FP等) を入れた場合に使用されることを想定している。
 15 下記の数式は、リガンド原子座標xがFPライブラリーから得られたFPに属し
 ていない（3.8 Å以下で接触していない）場合は0となり、属している場合は
 上記の式に従い、スコアが計算される。

[数26]

$$\text{density_of_atom}(x) = 0 \text{ or } \ln(\text{nhpcontact} + \text{natom} + \text{hi})$$

20 図13は、標的タンパク質の活性部位近傍におけるFPライブラリー由来のリ
 ガンドの位置を一例として示した図である。図13に示すように、標的タンパク
 質近傍で楕円（一点鎖線の円）に囲まれた付近のFPは、標的タンパク質に接し
 ているので、nhpcontactが優遇される。さらに、黒円付近は、FPラ
 イブラリー由来の結合リガンド原子が密集しておりnatomが優遇される。す
 25 なわち、これらの部分に、ドッキングされた候補リガンドの原子座標が近接した
 場合、上記数式によりスコアが優遇されることになる。

また、上記fp_contact_surfaceの数式において、tota

l_dense_of_atom(molecule)は、以下の数式で表せる。
 ここで、totalは、候補リガンド分子の原子数である。また、sort_density_of_atom
 のスカラー値の分布を大きい方から順に並べ替えたものである。つまり、候補リ
 5 ガンド分子が大きいとtotal_dense_of_atomは大きくなる。

[数27]

$$\text{total_dense_of_atom}(\text{molecule}) = \sum_{i=0}^{\text{total}} \text{sort_density_of_atom}(i)$$

10

以上で、本実施例における相互作用スコアFPAScoreの説明を終える。

[シミュレティッドアニーリングによる相互作用スコアの最大化およびコンフォ
 メーションチェンジ]

つぎに、上記のとおり定義されたFPAScore関数を最大化するために、
 15 本実施例におけるシミュレティッドアニーリング（以下、「SA」とよぶ。）の
 実行方法について、図14を参照して説明する（J. Mol. Graphics
 Mod. 18, 258-272, 305-306 (2000)）。
 図14は、シミュレティッドアニーリング過程を一例として示す概念図である。

最初に、候補リガンドのコンフォメーション変化から、その構造におけるFP
 20 AScoreが最大となるドッキング構造を得るまでのステップ1. ~3. の1
 サイクルについて述べる。

<ステップ1.>

まず、ドッキング対象となる候補リガンド(docked ligand)に
 存在する回転可能な二面角をランダムに変更することにより、コンフォメーショ
 25 ンを変化させる。本実施例では、候補リガンド原子のファンデルワールス半径は
 AMBER99を参考にした値を使用した。

<ステップ2.>

コンフォメーションの変化した候補リガンドを剛体として用いて、リガンド結合部位 (the binding site) にドッキングさせる。以下の並進回転操作は、ステップ 1. で発生させた一つのコンフォメーションについて行われる。

- 5 まず、前述したFPバンドからランダムに10個のFPの原子型を選ぶ。10個に満たない場合は、FPバンドのFPベクトルのサイズの最大数の半分を用いた。そして、選択されたFPに登録されている原子座標セットをランダムに選択する。これをアライメントされたFPとし、その対応関係で最小自乗フィッティングをおこない候補リガンドの原子座標とFPライブラリー由来の原子座標間の
- 10 rmsdを計算し、このとき得られる並進および回転行列を標的リガンドに対して作用させ、ひとつの標的タンパク質—候補リガンド複合体構造を得る。そして、アライメントされたFP、rmsd、標的タンパク質—候補リガンド複合体構造を用いてFPAScoreを算出する。ここで、図15は、FPAScoreを算出するためのFPアライメントおよび最小自乗フィッティングを模式的に示した
- 15 図である。

- 図15に示すように、FPバンドのアライメントの項で上述したようにFPアライメントは (D) , (E) の各FPの型ごとの座標行列の間で行われ、<1>リガンドライブラリー由来のFPベクトル (D) と、候補リガンド由来のFPベクトル (E) において、双方のビットがonである組み合わせが選択される。この
- 20 選択過程で一致しなかったFPはアライメントから除かれる。<2>そして、1つのFPにおいて、候補リガンド分子由来の原子座標ベクトル (1) と、FPライブラリーの結合リガンド由来の原子座標ベクトル (2) との座標間の対応付けを行い、最小自乗フィッティングに基づいて相互作用スコアを計算する。

- シミュレティッドアニーリングによる状態変化は、FPの変更、増加、減少過程
- 25 である。すなわち、当該状態変化は、そのFPに属する座標を、ドッキングさせる候補リガンド由来のFP、および、リガンドライブラリー由来のFPから選ぶ過程を繰り返すことによって行われる。そして、シミュレティッドアニーリン

5
10
15
20
25

グは、アライメントされたFPに対して、FPの原子型を一つ増加もしくは保持し、FPに登録されている原子座標セットの対応関係の変更もしくは追加と、FPの減少を行い、アライメントを変化しFPAScoreを最大化する。一つのFPから一つ以上の原子座標セットが選ばれること、もしくは座標があるのにも関わらず、FPAスコアが減少した場合は、メトロポリス判定が行われ、採用されれば状態を保つ。すなわち、SA過程においてメトロポリス判定が行われ、前回のスコアより、今回のスコアが大きいならば採用し、そうでない場合は、以下の数式に基づき採用確率Pacceptを計算する。このとき同時に $0 \leq r < 1$ の範囲の一様乱数を発生させ $r < Paccept$ ならば、スコアが低い場合も採用する。本実施例では、T（温度）は30.0Kからはじめ、0.07Kまで下げた。このようにして、一つのコンフォメーションに対してFPAScoreの最大値を計算する。

[数28]

$$15 \quad \Delta \text{Score} = \text{Score}(\text{after}) - \text{Score}(\text{before})$$

$$P_{\text{accept}} = \exp(\Delta \text{Score}/T)$$

このように得られたFPバンドを用いてSA法によりFPAScoreを最適化する。なお、本実施例において、SAは10,000回行った。

20 <ステップ3.>

一つのコンフォメーションに対して上記ステップ2で得られた最大のFPAScoreを、その構造とともに記憶部の構造プールに保存する。

以上が、一つのコンフォメーションについてのFPAScore最大化のための1サイクルの処理である。

25 <ステップ4.>

本実施例においては、コンフォメーションの変化を1000回行うことと設定したので、1000回に満たない場合、上述のステップ1.～3.を再試行する

よう制御する。なおコンフォメーション発生回数は多ければ多いほど良い結果が得られる可能性があるが、バーチャルな化合物データベースに含まれる多くの低分子化合物についてドッキング計算をする必要があり、有限な回数の大ききで止めねばならず、化合物の回転自由度に依存するとしても本実施例の予備計算ではこの回数で十分であった。

発生させた1000個のコンフォメーションのそれぞれについて、相互作用スコアFPAScoreの最大値が計算された場合、サイクルの繰り返し処理を終了し、構造プールに保存された1000個のコンフォメーションの最大FPAScoreを比較し、スコアが最大のドッキング構造を、当該候補リガンドについての最適なコンフォメーションとして標的タンパク質-候補リガンド複合体 (Protein-Ligand complex) の予測構造を出力する。

[結果と考察 (材料)、方法関連]

本実施例について以下に「結果と考察 (材料)」を述べる。本実施例で記述したFPライブラリーの構築には、Perl (<http://www.perl.com/>)、Ruby (<http://www.ruby-lang.org/>)、bash (<http://www.gnu.org/software/bash/>)等のシェル、スクリプト言語を組み合わせて開発した。また、本実施例の方法で記述したドッキングされる候補リガンドのコンフォメーションを変化し、FPAScoreを最大化するようなタンパク質-リガンド複合体構造を探索するアルゴリズムはC/C++で記述した。コンパイラーはIntel (登録商標) C++ Compiler 10.0を用いた。使用した計算機の構成について述べると、OSはRed Hat Linux、Scientific Linux、CPUはPentium4, Core2Duo, Opteron、メモリーは512M, 1024M, 2048Mと計算機の構成の異なるメモリー非共有型計算機クラスターを最大200台用いた。参考に計算時間を記述すると、後述するEGFRのkinaseドメインに対して、MDL Available Chemicals Directory (MDL ACD)

Library (Symyx Technologies, Inc. Corporate Address: 3100 Central Expressway, Santa Clara, CA 95051) の20,000化合物のインシリコスクリーニングを行った場合、一つの標的タンパク質に対して
5 一つ候補リガンドをドッキングする1CPUあたりの計算実行時間の中央値は10.2分、平均値は18.6分であった。最小計算時間は、4.8分、最長計算時間は1077分であった。ここで、図16は、EGFRインシリコスクリーニングにおける計算時間の分布を示す図である。

図16のEGFRインシリコスクリーニングにおける計算時間の分布に示すよ
10 うに、ドッキングされるリガンドによっては非常に時間がかかる場合がある。この原因の一つには、内部衝突をさけるようなコンフォメーションの探索が難しいリガンドをドッキングする場合は考えられ、これは回転可能な結合をランダムで選択していることが原因であり、分子内衝突が起こりにくいように回転する必要があることがわかった。また、本実施例のChooseLDの計算時間は、標的
15 タンパク質の大きさ、FPライブラリーに含まれるリガンドの数および、リガンドの分子量、候補リガンドの分子量、回転可能な結合の数に依存し、標的タンパク質のリガンド結合部位を絞り込み、FPライブラリーの絞り込みを行えば、より速く予測構造を得ることが可能であった。

本実施例では、ChooseLDのドッキング性能を試験するために、Protein Data Bank (Nucleic Acids Res. 31,
20 489-491 (2003)) からタンパク質-リガンド複合体構造を入手した。用いたベンチマークについて図17および図18を参照して説明する。図17は、ベンチマークの概要を一例として示す図である。また、図18は、PDBへの登録数の年度分布を表す図である。

図17に示すように、使用したベンチマークセットの数は、それぞれリガンドを有する218種のタンパク質である。85種のPDB構造(図17の左)は、スコア方程式(score equation)を作成するために使用された。

また、133種のPDB構造(図17の右)は、他のドッキング法(DOCK, AUTODOCK, GOLDなど)と比較するために使用された(以下にPDB IDを示す)。

85 PDB structures; 1G9V 1GKC 1GM8 1G
5 PK 1HNN 1HP0 1HQ2 1HVY 1HWI 1HWW 1IA
1 1IG3 1J3J 1JD0 1JJE 1JLA 1K3U 1KE5
1KZK 1L2S 1L7F 1LPZ 1LRH 1M2Z 1MEH
1MMV 1MZC 1N1M 1N2J 1N2V 1N46 1NAV 1
OF1 1OF6 1OPK 1OQ5 1OWE 1OYT 1P2Y 1P
10 62 1PMN 1Q1G 1Q41 1Q4G 1R1H 1R55 1R5
8 1R90 1S19 1S3V 1SG0 1SJ0 1SQ5 1SQN
1T40 1T46 1T9B 1TOW 1TT1 1TZ8 1U1C
1U4D 1UML 1UNL 1UOU 1V0P 1V48 1V4S 1
VCJ 1W1P 1W2G 1X8X 1XM6 1XOQ 1XOZ 1Y
15 6B 1YGC 1YQY 1YV3 1YVF 1YWR 1Z95 2BM
2 2BR1 2BSM
133 PDB structures; 1AAQ 1ABE 1ACJ 1
ACK 1ACM 1ACO 1AEC 1AHA 1APT 1ASE 1A
TL 1AZM 1BAF 1BBP 1BLH 1BMA 1BYB 1CB
20 S 1CBX 1CDG 1CIL 1COM 1COY 1CPS 1CTR
1DBB 1DBJ 1DID 1DIE 1DR1 1DWD 1EAP
1EED 1EPB 1ETA 1ETR 1FEN 1FKG 1FKI 1
FRP 1GHB 1GLP 1GLQ 1HDC 1HDY 1HEF 1H
FC 1HRI 1HSL 1HYT 1ICN 1IDA 1IGJ 1IM
25 B 1IVE 1LAH 1LCP 1LDM 1LIC 1LMO 1LNA
1LPM 1LST 1MCR 1MDR 1MMQ 1MRG 1MRK
1MUP 1NCO 1NIS 1PBD 1PHA 1PHD 1PHG 1

POC 1RDS 1RNE 1ROB 1SLT 1SNC 1SRJ 1S
 TP 1TDB 1TKA 1TMN 1TNG 1TNI 1TNL 1TP
 H 1TPP 1TRK 1TYL 1UKZ 1ULB 1WAP 1XID
 1XIE 2ADA 2AK3 2CGR 2CHT 2CMD 2CTC
 5 2DBL 2GBP 2LGS 2MCP 2MTH 2PHH 2PK4 2
 PLV 2R07 2SIM 2YHX 3AAH 3CLA 3CPA 3G
 CH 3HVT 3PTB 3TPI 4CTS 4DFR 4EST 4FA
 B 4PHV 5P2P 6ABP 6RNT 6RSA 7TIM 8GCH

図17の2つの円は、PDBIDをタンパク質-リガンド複合体の特徴ごとに
 10 分類したものであり、それらすべてのPDBIDを示している。図中の右の円の
 集合は、医薬品開発の標的タンパク質となりえるが、結合しているリガンドは医
 薬品的な化合物、ペプチド、糖鎖等と多様性に富んでいる。一方、左の円のPD
 BIDは、右の円と同様に医薬品開発のターゲットとなるタンパク質が選ばれて
 いるが、右の円のPDBIDとは異なり、医薬品的なリガンドで構成されている。
 15 より詳しく記述すると、右の円の集合は、リガンドの分子構造を用いて、ヘテロ
 アトムの有無、水素供与体、受容体、および疎水基等の有無、リピンスキーのル
 ールオブファイブ (Adv Drug Deliv Rev 46 (1-3
), 3-26.) を満たしているかといった判定基準で医薬品的であるリガン
 ドと判定されたものを最終的には手動で選定したというものである (J. Me
 20 d. Chem. 50, 726-741 (2007))。

すなわち、これらベンチマークセットの内訳は、85ベンチマークセットは、
 PDBに2000年8月11日より後に登録されたものの中から創薬のターゲッ
 トになる標的タンパク質を選び、ドッキングするべきリガンドもヘテロアトムを
 有するか、水素供与体、受容体、および疎水基等を有するか、ピンスキーの5ル
 ールを満たしているかといった判定基準で医薬品的なリガンドと判定されたもの
 25 を最終的には手動で選んだものを集めたものである。また、一方、理研ベンチマ
 ーク [参考文献: Onodera et al J. Chem. Inf. M

odel. 2007, 47, 1609-1618] は、GOLD [参考文献: Gareth et al J. Mol. Biol. 1997 267, 727-748] のベンチマークを使っている。このベンチマークは、上述したように、2000年8月以前にPDBに登録された標的タンパク質を用いている。しかし、このベンチマークでは、GOLD以外に、AutoDock, DOCKを比較しているため、このベンチマークの結果と比較することは、ChooseLDのドッキングソフトの中における位置づけを知るには非常に有用であると考

5 えた。上述した二つのベンチマークにおいてはPDB IDに重複はない。そこで、85セットでChooseLDのデフォルトパラメータの決定を行い、理研ベンチマークで、そのパラメータにおけるChooseLDの性能評価を行った。ここで、図18は、85セット（左の円）および、133セット（右の円）で提案されたPDB IDが登録された年を横軸に、その年の合計登録数を縦軸にプロットした図である。

10

これらのベンチマークセットへの登録年は図18に示すように分布している。

15 図18の2つのベンチマークセットのタンパク質-リガンド複合体の集団の色が示す事柄を記述すると、グラフの左側の山は、標的タンパク質が医薬品的 (drugable: 薬剤開発の対象となりうる標的タンパク質という意味) であり、リガンドは、種々の低分子化合物である場合の登録年の分布を表している (Green plane: 133 benchmark set Gold Benchmark (Jones et al. J. Mol. Biol. 1997, 267, 727-748) (Onodera et al. J. Chem. Inf. Model. 2007, 47, 1609-1618))。また、グラフの右側の山は、標的タンパク質とリガンドは共に医薬品的な (drugable) 化合物である場合の登録年の分布を表している (Blue plane: 85 benchmark set (Hartshom et al. J. Chem. 2007, 50, 726-741))。

20

25 黒ラインは、それぞれの平均PDB数を表しており、平均値は、緑で9.5、青

で14.2である (Black line: average of number of PDB of each (green, blue) plane. Average value are 9.5 and 14.2 for the green and blue plane, respectively. 5))。

ここで、図19は、予測と実験結果間でのrmsdを要約したテーブルである (Table. Summary of r. m. s deviation between predictions and experimental results)。結合モード予測構造の精度を評価するために、予測構造と実験構造のrmsdを計算した。rmsdが大きい場合、予測構造と実験構造とのずれが大きいことを意味しており、すなわち予測の失敗を意味する。そこで、予測構造を正解と見なすrmsdの上限値を設定した。図19の表はJonesらによって行われた結合モード予測構造と実験構造のrmsdと人間の感覚、すなわちGood, Close, Errors, Wrongの関係を示したものである。rmsdが2.0 Å以下なら予測構造が実験構造にくらべて良い、すなわちGoodとなる。rmsdが2.5 Å以下なら実験構造に近い予測構造を含んでおり、かつ、よい予測構造が含まれているということになる。すなわちCloseとなる。そこで、rmsdが2.0 Å以下の予測構造が得られた場合を予測の成功と定義した。rmsdが、2.0以上2.5以下なら、ビジュアルでの評価Good, Close, Errors, Wrongである (Jones et al. J. Mol. Biol. 1997 267; 727-748より抜粋)。すなわち、rmsdが2.0 Å以下ならリガンドモデルとして正解に比べて良い。rmsdが2.5 Å以下ならリガンドモデルとして正解に比べて似ている (Close) と良い (Good) の両方を含む。

25 [結果と考察 (1) : FPA関数におけるk1最適化 (Optimized k1 in FPA Score function)]

上述したように、FPAScoreのk1値はFPライブラリーに登録されて

いる原子座標と候補リガンドの原子座標の一致度を調節する係数である。k 1 値はターゲットに応じて変更可能であるが、大量の標的タンパク質に対してインシリコスクリーニングを行う場合や、他の研究者に使用されることを考慮すると最適なパラメーターを決定することは本手法を採用する判断材料の一つとなることから、ChooseLD法のドッキング性能試験においては最適値を85セット [参考文献: Michael et al J. Med. Chem. 2007, 50, 726-741] を使い、FPAScore関数のk 1の最適値を決定した。

85セットはドラッグライクな標的タンパク質を多く集めており、GOLD [参考文献: Gareth et al J. Mol. Biol. 1997 267, 727-748] の性能評価を行っている。これは、85セットはPDBIDが133セットと重複しないため、すなわち、この最適化の過程において、85セットは、133セットの情報を使用していないためである。また85セットはGOLDのベンチマークのみを行っており、GOLDの成功率はCorinaの構造を標的タンパク質にドッキングさせた場合、75.2±0.4%であり、実験構造のリガンド構造を用い結合部位を6 Åと定義した場合80.5±0.5%であり、実験構造のリガンド構造を用い、結合部位を4 Åと定義した場合86.9±0.3%であり、X線結晶構造中に存在する結晶水を含めた場合98.6±0.1%であった (J. Med. Chem. 50, 726-741 (2007))。すなわち、GOLDだけの評価を行っている場合、既存のドッキングソフトの中におけるChooseLDの位置づけを知ることができないので、85セットはk 1値の最適化に使用した。ここでは、FPAスコア (Score) で記述したk 1の最適化をおこなった。

ドッキングの条件は下記に述べる通りである。他のベンチマークと同様に、リガンド結合部位の探索範囲を狭める等の利点があるため、リガンド結合部位を定義した。すなわち、ChooseLDのドッキング性能試験のベンチマークは、タンパク質のリガンド結合部位のアミノ酸残基を予測するものではなく、リガ

ド結合部位における候補リガンドの配座の正確性を試験することである。結合部位 (binding site) の大きさは、タンパク質-リガンド複合体 (Protein-Ligand complex) の正解構造のリガンドの各原子から 4 Å とした。また、FP ライブラリーに含まれているリガンドの候補リガンドとの類似性の及ぼす影響を試験するため FP ライブラリーに属するリガンドとの Tc を計算し、FP ライブラリーに含まれるリガンドを限定した。ドッキングするリガンドとライブラリーリガンド (LIBRARY LIGANDS) に属するリガンドとの Tanimoto 係数を薬剤様 FP (Drug Like Fingerprint) を用いて計算し、fp bands の Tc Range は、最大値を 0.96, 0.76 そして、0.56、最小値を 0.08 とした。

初期コンフォメーションは二面角をランダムに回転させ、初期リガンドから最も rmsd の大きい構造を結合部位 (binding site) から十分に離れたものを使用した。そのリガンドを用いて、一つのターゲットに対して 10 回のドッキングをおこなった。85 セットのうち、84 セットをドッキングすることができた。ここで、図 20 は、85 セットにおける予測成功率一覧 (k1 と Tc Range の関係) を示す図表である。

図 20 の表の k1 は、FPAScore で述べた係数のことである。その下の数値は、計算をおこなった k1 値である。Tc Range は、最大値を 0.96, 0.76, そして 0.56、最小値を 0.08 とした。カラムの中の数値は成功率 (%) であり、平均 (average) は、上記の範囲の平均値である。

この結果、k1 = 4.0 の時の平均値が最も成功率が 62.1% と最も高く、次に 6.0, 3.0, 5.0, 2.0 の順で成績がよかった。k1 値が 1.0 の場合は、すべての TC Range において、そのほかの k1 値の成功率より悪かった。k1 値が 4.0 と 6.0 の場合はほぼ同等であったが、わずかに平均値においてまさる 4.0 を最適値として 133 種 [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1

609-1618] のベンチマークはこの数値を用いた。

ここで、図21は、10位までに $rmsd \leq 2.0$ Å以下で予測できる割合を示す図表である。図21の右図は、その時の成功率をプロットしたものであるが、採用される $FPAScore$ による順位を増やすにつれ、予測成功構造を得られる確率が上昇することが示された。すなわち、 $FPAScore$ 上位の予測構造を一つ用いるのではなく複数用いる場合、正解に近い構造を得られる確率が上がることになる。すなわち、 $FPAScore$ 上位の予測構造を分子動力学計算や量子化学計算による複合体構造の最適化における初期構造にも複数用いた方がよいと考えられる。成功とみなす実験構造との $rmsd$ を 2.0 Åとした場合、10位までに最大82.9%予測に成功することが示された。

また、図22は、10位までに $rmsd \leq 2.5$ (Close) Å以下で予測できる割合を示す図表である。図22に示すように、成功とみなす実験構造との $rmsd$ を 2.5 Åとした場合、10位までに最大87.6%予測に成功することが示された。

また、図23は、成功とみなす正解構造との $rmsd$ を 2.0 Å以外でも行った場合を示す図表である。図23の右図は、横軸に成功と見なす実験構造との $rmsd$ 、縦軸に予測成功率をプロットしたものである。

上述のように、 2.5 Åでは、約7割成功としたが、85セットベンチマークにおけるGOLDの予測成功率の一つである *Corina* で発生させたリガンド、すなわち、実験構造のコンフォメーションを用いない場合の結合モード予測の成功率75.2% (参考文献 (Michael et al. J. Med. Chem. 2007, 50, 726-741)) と同等の成功率を得るには、 $Tc Range$ を $0.56-0.08$ では $3.2-3.3$ 、 $0.76-0.08$ ならば 2.8 を、 $0.96-0.08$ ならば $2.6-2.7$ を用いる必要があることが示された。なお、一般的な共有結合長である 1.5 Åを成功と定義した場合では、約4割の予測に成功したことになる。ファンデルワールス相互作用の限界値にちかい 3.5 Å以内では約8割の予測に成功したことになる。ここで、

図24は、ChooseLDと比較して、Dock、AutoDockおよびGOLDのベンチマークの結果を示す図表である。

図24は、Onodera et al [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1609-1618] によるベンチマークでCorinaによる座標発生に失敗したターゲット、DOCKまたはGOLDで失敗したターゲットをのぞいた116種のPDB IDの結果を示す図である。図24の成功率 (success rate) は、rmsd 2.0 Åか、それより良い構造の割合を示している。

ここで、ドッキング方法 (Docking method) は、各ドッキングソフト (Docking soft) の名前を意味している。ChooseLDは、3つのTc Rangeについて性能評価をおこなっている。GOLD GOLDScoreSTD, GOLDScoreLib, GOLD ChemScoreSTD, AutoDock、そしてDOCKの値は、CorinaとMINIの平均値とし、各ドッキングソフトの成功率において標準偏差を細い棒で示している。

図24のグラフに示すように、本実施例のChooseLDのrmsd 2.0 Åか、それより良い構造を予測する性能 (成功率) は、Tc Rangeが0.96から0.08の場合、GOLDとほぼ対等である。Tc Rangeが0.76から0.08の場合、GOLDとほぼ対等か少し劣る。Tc Rangeが0.56から0.08の場合、GOLDには及ばないが、DOCK, AutoDockよりよい、ということが示された。

ここで、図25は、85セットにおけるFPAScoreの予測構造と実験構造とのrmsdが2.0 Å以下における各々標的タンパク質との衝突個数の分布を示す図である。衝突0個の構造が75.0%であり、衝突1個の構造が17.3%であるため合計が、計92.3%となっていることから、FPAScoreの衝突判定関数は、経験的物理関数であるレナードジョーンズ型関数の衝突判定に相当するものとして機能していることが示された。

図26および図27は、各ターゲットにおける全10回のドッキング試行における成功個数を記したものである。図26は、85セットベンチマークにおける予測成功構造の個数分布を示している。なお、図26の「*1」は、予測成功個数が5から10個のPDB IDの個数の全体に占める割合を表している。すべてのTc範囲において、10回成功と10回失敗の割合が大きい。また、10回中5回成功したターゲットは62.7~65.5%であった。また、Tc範囲の上限値を小さくしていくと、10回とも失敗する個数が増える傾向が示された。これはChooseLD法が、FPライブラリーとして既知のタンパク質ーリガンド複合体構造に依存しているため、FPライブラリーに属するリガンドが減ると精度が落ちるためと考えられる。

[結果と考察(2) (Result and Discussion (2)) :
133種のベンチマークの結果]

Onodera et al [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1609-1618] によって、各ドッキングソフトを提供されている状態に近い状態でベンチマークが行われている。彼らによると標的タンパク質は、GOLD [参考文献: Gareth et al J. Mol. Biol. 1997 267, 727-748] のベンチマークに使用されているタンパク質ーリガンド複合体 (Protein-Ligand complex) は、133種の中でGOLD, DOCKでドッキングすることができなかったターゲット、および、Corinaで三次元座標を発生できなかったターゲットをのぞいた計116種を用いられている。なお、除かれたPDB IDは、1TPH, 1TRK, 1XID, 4FAB, 6RSA, 1BBP, 1CTR, 1HYT, 1PHG, 1POC, 1SNC, 1TMN, 1CDG, 1DR1, 1LDM, 4CTS, 4ESTである (Virtual Screening J. Chem. Inf. Model. 47, 1609-1618 (2007))。

各々のドッキングソフトのパラメータは各々ドッキングソフトで提供されているパラメータを使用しており、パラメータをターゲット用に最適化していない。パラメータの最適化を行えば、もちろん成功率は変わると考えられるが、それは、ChooseLDにおいても同様であり、ChooseLD法においても、標的タンパク質に応じて変更可能パラメータ k_1 , k_2 , k_3 値が定義されているので、最適化の余地が残っている。そこで、ChooseLDの性能評価には方法の項で述べられた値と85セットで最適化をおこなった k_1 値、すなわち4.0を用いた。

ここで、ChooseLDが使用するドッキングの条件は各ターゲットにおいて、以下のように定めた。

1. 結合部位 (binding site)

結合部位 (binding site) は、従来のベンチマーク [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1609-1618] に類似してネイティブ (Native) のタンパク質-リガンド複合体 (Protein-Ligand complex) のリガンド (ligand) の各原子から半径5.0 Å以内の距離に存在するProteinの原子の球とした。

2. リガンドのコンフォメーション変化

133セットのベンチマークではドッキングするリガンドを3つ用意している。すなわち、Corinaで発生させたリガンドと、Corinaで発生したリガンドのうちエネルギー最小構造 (以下MINIとする) のものと、そしてPDBに登録されている状態の構造との3つであり、これらをそれぞれ116の標的タンパク質に対して1000個の予測を行っている (Virtual Screening J. Chem. Inf. Model. 47, 1609-1618 (2007))。ChooseLD法のドッキング性能試験においてはコンフォメーションをランダムに変化させ、実験構造のタンパク質-リガンド複合体のリガンドから最もrmsdが大きい構造で、かつ上記で定義されたリガ

ド結合部位から十分に離れた状態のリガンドを使用した。すなわち、実験構造をそのまま用いずに各116の標的タンパク質に対して10回の予測を行ったことになり、133セットを用いたベンチマークとほぼ同条件で行ったことになる。これらの過程においてリガンドに水素が存在した場合は取り除かれる。

5 3. リガンドとのタニモト係数の範囲

使用するライブラリーリガンド (LIBRARY LIGAND) は、候補リガンド (docked ligand) とTcの範囲で、その最大値である0.96, 0.76および0.56は、それぞれ、ドッキングリガンドと非常によく似ている化合物が存在するもの、似ている化合物が存在するもの、少し似ている化合物が存在するもの、に該当する。そこで、Tcの範囲には、0.96-0.08 (つまり答えを含まない), 0.76-0.08および0.56-0.08に該当するものを用いた。

4. Onodera et alは、一つのリガンドに対して1000回ドッキングを行っている [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1609-1618]。今回のChooseLDの性能評価では、候補リガンド (docked ligand) を10回ドッキングした。すなわち1160回のドッキングを各々のTc Rangeで行い計3480回のドッキングをおこなった。一回のドッキング試行において予測されたドッキング構造とネイティブなタンパク質-リガンド複合体 (Native Protein-Ligand Complex) のリガンドとのrmsdが2.0 Åか、それより良いならば成功とした。

図28および図29は、133セットのベンチマークにおけるDOCK, AutoDock, GOLD予測構造のrmsd分布の結果と、ChooseLD法の結果を示す図である。Docking methodは各ドッキングソフトの名前を意味している。ChooseLDは3つのTc範囲について性能評価をおこなっている。GOLDは、GOLDScoreSTD ('Standard Default Settings' with GOLDScore), GOL

DScoreLib ('Library Screening Settings' with GOLDScore), GOLDChemScoreSTD ('Standard Default Settings' with ChemScore)の3つのパラメーター(Virtual Screening J. Chem. Inf. Model. 47, 1609-1618 (2007)), AutoDockそしてDOCKの値は、CorinaとMINIの平均値とした。このグラフよりChooseLDのrmsdが2.0 Å以下の構造を予測する性能は、Tc範囲が0.96~0.08ならGOLDとほぼ対等である。Tc範囲が0.76~0.08ならGOLDとほぼ対等か少し劣り、Tc範囲が0.56~0.08ならば、DOCK, AutoDockよりよいことが示された。

図30および図31は、各ターゲットにおける全10回のドッキング試行における成功個数を記したものである。なお、図30中の「*1」は、予測成功個数が5から10個のPDB IDの個数の全体に占める割合を示している。85セットと同様に、10回成功と10回失敗の割合の二極化がおきているが、10回失敗の数をもっとも多いことが示された。また、85セットに比べて、10回成功率が20%近く下がっている。これらのことから、133セットは85セットに比べて、ドッキングが難しいターゲットが多く含まれているものと考えられる。85セットの医薬品的な化合物は分子量、回転可能な結合数、水素供与体、水素受容体の数がリピンスキーの5ルール等で限定されているので、その絞り込みの影響によりドッキングしやすい化合物が多く含まれるためであると考えられる。

図32および図33は、Tc範囲で限定されたFPライブラリーにおいてFPAScoreで順位付けされた分布内に実験構造とのrmsdが2.0 Å以下の構造が得られる確率を示す図である。すなわち、順位が1の場合は前述したその他のドッキングソフトとの比較の成功率と一致する。この結果も85セットと同様に全体の成功率が低下している。

図34は、予測成功構造の衝突個数の分布を示す図であり、133セットにお

ける予測構造と実験構造との $rmsd$ が 2.0 \AA 以下の構造における各々標的タンパク質との衝突個数の分布を示す。衝突0個の構造が 56.0% で衝突1個の構造が 28.7% であり、計 84.6% となっており、 $FPA\ Score$ の衝突判定関数は経験的物理関数であるレナードジョーンズ型関数の衝突判定に相当するものとして機能していることが示された。85セット、133セットとも同様の傾向を示したことから、衝突判定は十分機能していると考えられる。

図35は、FPライブラリーに用いるリガンドの Tc 範囲の上限値をさらに低くし、 0.16 , 0.24 , 0.36 に下限値を 0.08 にした場合の性能および、前述した Tc 範囲、すなわち上限値 0.56 , 0.76 , 0.96 、下限値 0.08 の予測成功率を示す図である。 Tc の上限値を低くした場合は、 $0.24-0.08$ で133セットベンチマークにおけるDOCK (21.1%) と同程度の予測精度であり、 $0.36-0.08$ で133セットベンチマークにおけるAutoDock (26.6%) と同程度の予測精度であることが示された。

(GOLDとの比較)

15 理研ベンチマークでGOLDが失敗したが、本願発明者の方法ではドッキングでき、かつ、 $rmsd$ 2.0 以下であった例を2例示す。

ここで、図36は、1DR1について予測されたタンパク質-リガンド構造を示す図である (Predicted protein-ligand complex structure for 1DR1)。

20 図36における条件や値等は以下のとおりである。

PDBID: 1DR1

TITLE: CHICKEN LIVER DIHYDROFOLATE REDUCTASE

DOCKED LIGANDSD: NADP

25 RMSD: 1.743

FPA: Score 1295.553

CYAN (図中央のシアン (淡い水色)) : 実験 (X線結晶解析) 構造 (Ans

w e r) (以下も同じ。)

GREEN (図中央の濃い緑) : 予測のリガンド構造 (P r e d i c t e d l i g a n d S t r u c t u r e) (以下も同じ。)

The o t h e r (その他) : 結合部位 (t h e b i n d i n g s i t e) (以下も同じ。)

すなわち図36は、PDB ID ; 1DR1に対する本実施例の予測構造を示している。これはGOLDが予測に失敗した標的タンパク質、すなわち133セットのベンチマークから除外されたターゲットである (V i r t u a l S c r e e n i n g J. C h e m. I n f. M o d e l. 47, 1609-1618 (2007))。本実施例のChooseLDは、予測構造と実験構造のrmsdが1.74 Åであり、予測に成功した。これは、リガンドに存在する環構造がFPライブラリーにも多く含まれていたためであると考えられる。

また、図37は、4ESTについて予測されたタンパク質-リガンド構造を示す図である (P r e d i c t e d p r o t e i n - l i g a n d c o m p l e x s t r u c t u r e f o r 4 E S T)。

図37における条件や値等は以下のとおりである。

PDB ID: 4EST

TITLE: CRYSTAL STRUCTURE OF THE COVALENT COMPLEX FORMED BY A

20 PEPTIDYL ALPHA, ALPHA-DIFLUORO
-BETA-KETO AMIDE WITH PORCINE

PANCREATIC ELASTASE AT 1.78-
ANGSTROMS RESOLUTION

25 DOCKED LIGAND: INHIBITOR ACE-*ALA-*P
RO-*VAL-*DIFLUORO-*N-*PHENYLETHYLACE
TAMIDE

RMSD : 1.729

FPAScore : 451.291

すなわち、図37は、PDBID ; 4ESTに対する本実施例の予測構造を示して
5 しており、これはGOLDが予測に失敗した標的タンパク質であり、133セッ
トのベンチマークから除外されたターゲットである (Virtual Screening J. Chem. Inf. Model. 47, 1609-
1618 (2007))。ChooseLDは、予測構造と実験構造のrmsdが1.73Åであり予測に成功した。これはドッキングするリガンドがペプチ
ド性リガンドであったこともあり、FPライブラリーに含まれるペプチド性リガ
10 ンドの主鎖の炭素、窒素、酸素が主に使用されたためであると考えられる。

[結果と考察(2)] (Result and Discussion (2.1)) : 予測された構造結果 (result of predicted structure)

既存のドッキングソフト (GOLD, DOCK) が失敗したすべてのドッキ
15 グの例を4例示す。

ここで、図38~図41は、GOLDが失敗したがChooseLDは予測に成功したターゲットを示す図である。

図38における条件等は以下のとおりである。

1CDG

20 TITLE :

NUCLEOTIDE SEQUENCE AND X-RAY STRUCTURE OF

CYCLODEXTRIN GLYCOSYLTRANSFERASE FROM BACILLUS CIRCULANS STRAIN

25 251 IN A MALTOSE-DEPENDENT CRYSTAL FORM

また、図39における条件等は以下のとおりである。

1 DR1

2. 2 ANGSTROMS CRYSTAL STRUCTURE OF C
HICKEN LIVER DIHYDROFOLATE

REDUCTASE COMPLEXED WITH NADP+ AND
5 BIOPTERIN

また、図40における条件等は以下のとおりである。

1 LDM

REFINED CRYSTAL STRUCTURE OF DOGFISH
M4 APO-LACTATE DEHYDROGENASE

10 また、図41における条件等は以下のとおりである。

4 EST

Title CRYSTAL STRUCTURE OF THE COVA
LENT COMPLEX FORMED
BY A PEPTIDYL ALPHA, ALPHA-DIFLUORO-B
15 ETA-KETO AMIDE WITH PORCINE
PANCREATIC ELASTASE AT 1.78-ANGSTROM
S RESOLUTION

(GLIDEを含めた比較)

20 Glide (J. Med. Chem. 47, (2004) 1739
-1749) はフレキシブルリガンドドッキングソフトであり、本実施例の方法
の中でGOLD等との予測精度の比較を行っている。図42は、133セット
中における90ターゲットにおける予測成功率を示す図表である。但し、上表の
予測成功率の算出法は各ドッキングソフトによって異なる。すなわち、GOLD
は各ターゲットに対して遺伝的アルゴリズムによる最適化を20回行った場合の
25 結果 (the best of GA 20 run) ([http://www.
ccdc.cam.ac.uk/products/life_sciences/
validate/gold_validation/value.htm](http://www.ccdc.cam.ac.uk/products/life_sciences/validate/gold_validation/value.htm))

1) であり、ChooseLDは各ターゲットに対して10回ドッキングを行いFPAScore上位2個選び、ベストの構造を選択した。Glideのドッキング性能の検証には記載が無いのでGOLDに準ずると考えられる。133セットのベンチマークの結果において、GOLDの予測成功率が45%程度であった
5 事実からドッキング条件および予測構造の選択法によって予測成功率が大幅に変動すると考えられる。

(予測成功標的タンパク質の分布)

図43は、ドッキングソフト間の予測に成功した標的タンパク質のPDBIDの類似度をTc (タニモト係数) で算出した図表である。ここで、133セット
10 の中における90セットにおける、それぞれの標的タンパク質に関して、両方のドッキングソフトが予測に成功した場合、Tc計算式のaを加算し、片方のみが予測に成功したのならbもしくはcを加算する。

図43に示すように、Glide, GOLD, FlexX (J. Mol. Biol. 261, 470-489 (1996)) 間のTcが0.61~
15 0.65であるのに対してChooseLDとその他のドッキングソフト間では0.47~0.55程度であった。予測成功率はGOLD, Glide, ChooseLD間で顕著な差が無いことも考えるとChooseLDはその他のドッキングソフトに比較し、予測が成功する標的タンパク質の分布に独自性があることが示された。

20 また、図44は、90ターゲット中の一つの標的タンパク質に対する各ドッキングソフトの予測の成否分布を示す図表である。一方のドッキングソフトが予測可能なターゲットは多く存在し、現状では、すべての標的タンパク質の予測に成功するドッキングソフトは無いということになった。このような背景のもと、複数のドッキングソフトを用いることを前提に、ドッキングソフトのスコアによって予測構造を選択するのではなく、予測された標的タンパク質—リガンド複合体
25 構造から、水素結合等のタンパク質との相互作用情報を持ちいて、より実験構造に近い予測構造を選択する研究が多く行われている (European Jou

Journal of Medicinal Chemistry 42, 966
-976 (2007)、J. Med. Chem. 47, 337-34
4 (2004)。

ここで、図45～図47は、DOCKが失敗したがChooseLDは予測に
5 成功したターゲットを示す図である。

ここで、図45における条件等は以下のとおりである。

1HYT

RE-DETERMINATION AND REFINEMENT OF T
HE COMPLEX OF

10 BENZYL SUCCINIC ACID WITH THERMOLYSI
N AND ITS RELATION

TO THE COMPLEX WITH CARBOXYPEPTIDAS
E A

また、図46における条件等は以下のとおりである。

15 1PHG

CRYSTAL STRUCTURES OF METYRAPONE-AND
PHENYLIMIDAZOLE-INHIBITED

COMPLEXES OF CYTOCHROME P450-CAM

また、図47における条件等は以下のとおりである。

20 1TMN

BINDING OF N-CARBOXYMETHYL DIPEPTID
E INHIBITORS TO THERMOLYSIN

DETERMINED BY X-RAY CRYSTALLOGRAPHY.

A NOVEL CLASS OF TRANSITION-STATE AN
25 ALOGUES FOR ZINC PEPTIDASES

[結果と考察(3) (Result and Discussion(3)) :
許可されたrankrangeの結果 (result of accepted

rankrange)]

図48は、1位だけではなく10位までにrmsd 2.0の構造が採取できる割合を示す図である。図48に示すように、10位まで採取すると、6割以上がrmsd 2.0以下でドッキング可能である。

- 5 また、図49は、1位だけではなく10位までにrmsd 2.5 (Close)の構造が採取できる割合を示す図である。

[結果と考察(4) (Result and Discussion (4)) :
成功とみなせるrmsdの結果 (result_rmsd_regard_as
_success)]

- 10 成功と定義するrmsdを変化させる。理研ベンチマークとの比較の際には、成功と定義する予測構造の正解構造とのrmsdを2.0 Åとしたが、そのほかの数値(1.5, 2.5, 3.0, そして3.5)を示す。3.5 Åであれば、その予測リガンド構造はほぼリガンド結合部位の近傍に存在すると考え、その構造を分子動力学や、量子化学計算の初期構造として用いることができるためである。図50は、成功と定義するrmsdを変化させた場合を示す図表である。

- 15 図50に示すように、3.5 Å以内に予測できた構造は、Tc Range 0.56-0.08 (即ち少し似ているリガンドがライブラリーに存在する場合)において68.9%であった。つまり、類似した化合物の実験構造が存在すれば、この精度でドッキング構造が少なくともリガンド結合部位近傍に予測可能であることを意味している。

また、Tc Range 0.96-0.08 (即ち、かなり似たリガンドがライブラリーに存在する場合)においては、7割の程度がリガンド結合部位に存在することを示している。

- 25 ここで、ドッキングの成功の定義としてのrmsd 2.0という数値は、様々なベンチマーク [参考文献: Gareth et al J. Mol. Biol. 1997 267, 727-748], [参考文献: Michael et al J. Med. Chem. 2007, 50, 726-7

41], [参考文献: Onodera et al J. Chem. Inf. Model. 2007, 47, 1609-1618] などにおける基本的な評価基準である。しかし、実際には、rmsdが2.0より大きいケースでも、MD, QMなどの最適化を行えば、精度よくタンパク質-リガンド複合体 (Protein-Ligand complex)の構造を予測し得る。即ち、これらの成功と定義するrmsdを示すことは、MD, QM研究者が複合体構造の最適化のための初期構造を選ぶ際の、有用なデータとなる。つまり、最適化にかかる時間 (shot time 100ps, long time 1ns and so on) または最適化するリガンド結合部位の範囲 (5 Å, 10 Å and so on) を見積もる場合の参考になると考える。

[結果と考察 (5) (Result and Discussion (5)) : 理想的な方法 (method_ideal)]

主に考察 (Discussion) を以下に再び図8を参照しながら記述する。

すなわち、本実施例では、リガンドのパーツであるFPのコンフォメーションが相互作用した構造として最も安定であるとの仮定をたてる。本実施例のFPの標的タンパク質との相互作用とは、タンパク質と近い距離にあるFPを疎水性相互作用、水素結合相互作用および、ファンデルワールス相互作用といったエンタルピー的相互作用と解釈し、また、タンパク質と遠い距離にあるFPを溶媒との相互作用といったエントロピー的相互作用と解釈する。

つまり、本実施例においては、最終的にFPのコンフォメーションを使って、基底として化合物 (Chemical Compound) が最も安定なドッキング構造をとる場合、タンパク質リガンド相互作用において最も安定な自由エネルギーをとると言うことに相当すると仮定されている。

すなわち、重なるのよい類似タンパク質由来の結合リガンド (ligand) 群から抽出したFP配置は、タンパク質との相互作用の自由エネルギーを含んでいる。

ここで、一つの標的タンパク質がある場合、多くのリガンドを集めるためにホ

モロジューまたは、e-valueの低い類似タンパク質を利用するが、これら機能的分類に縛られない広義のファミリータンパク質は活性部位近傍が少しの構造変化と、アミノ酸残基の変化を伴い、ファミリータンパク質から抽出したFPが自由エネルギー安定の仮定を満たされない可能性も当然考えられる。

- 5 そのため、この欠点を補う必要があり、ファミリータンパク質から抽出したFPを、標的タンパク質との相互作用においてより自由エネルギーが安定になるFPに変えて、“Modified FP”とし信頼性の少し落ちたFPとして採用する。これには3D-1D法のProgramを修正して対応する。このModified FPの作成を標的タンパク質に対して行えば、まだ見つからない新規骨格のリガンドを考慮したことに相当し、標的タンパク質に結合した既知のリガンドよりも活性の高い化合物を見つけられる可能性がある。

- 15 一方、複数の結合化合物の原子相互作用の共通領域のFPは、ファミリータンパク質が似たような複数の化合物と結合するという重なりを重視しており、生物化学的情報やエネルギー計算により原子の存在の可能性がある場合に与えられる
- 15 “Creative FP”よりも実験情報を反映したFPを得ることができる
と考える。

[他の方法 (MD, QM) のタンパク質-リガンド複合体の最適化 (Protein-Ligand Complex Optimize for other method (MD, QM))]

- 20 従来の古典物理学的エネルギーによって予測されたタンパク質-リガンド複合体 (Protein-Ligand complex) の構造に対して、既知のタンパク質-リガンド複合体 (Protein-Ligand complex) の構造の情報を用いて、上記の方法で得られたドッキング構造の順位付け、クラスタリングが行われている [参考文献: Zhan et al J. Med. Chem. 2004, 47, 337-344]。これらのことは、既存のドッキングソフトによる出力は、実験情報を確実に反映していない構造を出力することを意味している。
- 25

一方で、予測されたタンパク質-リガンド複合体 (Protein-Ligand complex) の構造をAMBER, CHARMMを用いたMD (参考文献:それぞれ、Case, A. D., Cheatham III, E. T., Darden, T., Gohlke, H., Luo, R. 5, Merz Jr., M. K., Onufriev, A., Simmerling, C., Wang, B. & Woods, J. R. The Amber Biomolecular Simulation Programs J Comput Chem 26 1668-1688 (2005), Brooks, R. B, Bruccoleri, E. R. 10, Olafson, D. B., States, J. D., Swaminathan, S. & Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations J. Comp. Chem. 4 187-217 (1983)) またはQM (参考文献: Kamiya K, Sugawara Y, Ume- 15 yama H. J. Comput. Chem. 2003, 24, 826-841) で最適化する試みもなされている。これらのMDやQMなどの方法では、ドッキングやインシリコスクリーニングを行うのは計算量が大きすぎるため、真 (Nativeと言う意味で) のタンパク質-リガンド複合体 (Protein-Ligand complex) の構造からある程度近い位置にリガンドをドッキングし、それを初期構造とする必要がある。

その初期構造を得るために既存のドッキングソフトを用いるのだが、前に述べた物理エネルギーを主体にしているため、物理エネルギーによる最適化を繰り返すことになる。

25 一方、本実施例による手法は、既知のタンパク質-リガンド複合体 (Protein-Ligand complex) の情報を主に使用しておりバイオインフォマティクスの観点と物理エネルギーによる観点を考慮することが可能であり、

また、本実施例で用いたPDBの構造情報といったバイオインフォマティクス情報は、年ごとに蓄積されるものなので、医学的に興味あるタンパク質-リガンド複合体 (Protein-Ligand Complex) は多くの研究者によって研究され、これらの予測構造の最適化にも有用であると考えられる。

5 [結論 (Conclusion) : 本手法の性能]

図51は、本実施例による処理の結果を示す図表である。図51に示すように、本実施例による手法を用いれば、T85 setをドラッグブルタンパク質 (Druggable-protein) に対して、ドラッグ様リガンド (Drug like ligand) をドッキングした場合、Tc Rangeが0.56-0.08, 0.76-0.08, 0.96-0.08の場合それぞれ、Goodの構造を得る確率は、58.9, 62.1そして65.2%であり、Closeの構造を得る確率は、それぞれ、68.6, 72.1, 72.4%であった。

また、ドラッグブル標的タンパク質 (Druggable-Target Protein) に対して、様々なリガンド (ligand) をドッキングした場合の性能は、Tc Rangeが0.56-0.08, 0.76-0.08, 0.96-0.08の場合それぞれ、Goodの構造を得る確率は、40.1, 44.8, そして46.4%であり、Closeの構造を得る確率はそれぞれ、53.2, 57.8そして59.3%であった。そしてこれらの性能は既存のドッキングソフトとほぼ同等の性能であることを示した。

20 標的タンパク質とリガンドが共に医薬品的な (druggable) 化合物を含んだトレーニング計算の結果から、標的タンパク質と任意のリガンドの相互作用スコアが10番目までのコンフォメーションを考察すれば、標的タンパク質全体の83% (図21の0.96-0.08、10位までの値) に対して、正解に対して良いモデルを与えるという2.0 Åの範囲の答えを含んだリガンド構造が一つは見つかるので、目視をして良い構造を探す価値があるということになる。

25 一方、標的タンパク質と任意のリガンドの相互作用スコアが10番目までのコンフォメーションを考察すれば、標的タンパク質全体の88% (図22の0.9

6-0.08、10位までの値) に対して、正解に対して良いモデルと似ているモデルを与えるという2.5 Åの範囲の答えを含んだリガンド構造が一つは見つかるので、目視をして良い。構造または似ているモデル構造を探す価値があるということになる。

- 5 また、標的タンパク質はドラッグابل (drugable) であり、リガンドは種々の低分子化合物を含んだトレーニング計算の結果から、標的タンパク質と任意のリガンドの相互作用スコアが10番目までのコンフォメーションを考察すれば、標的タンパク質全体の65% (図48の0.96-0.08、10位までの値) に対して、正解に対して良いモデルを与えるという2.0 Åの範囲の答えを含んだリガンド構造が一つは見つかるので、目視をして良い構造を探す価値があるということになる。

- 10 一方、標的タンパク質と任意のリガンドの相互作用スコアが10番目までのコンフォメーションを考察すれば、標的タンパク質全体の76% (図49の0.96-0.08、10位までの値) に対して、正解に対して良いモデルと似ているモデルを与えるという2.5 Åの範囲の答えを含んだリガンド構造が一つは見つかるので、目視をして良い構造または似ているモデル構造を探す価値があるということになる。

- 15 従来、物理学的相互作用関数で当該標的タンパク質と仮想化合物ライブラリー低分子化合物の相互作用を計算していたところ、本実施例は、バイオインフォマティクスの情報を使って半経験的に計算している点で従来手法と異なっており、さらに構造予測の成功率は世界で認められているドッキングソフトプログラムGOLDと比べて優れる高い効果もあり、また、年々高まっている情報の蓄積が、半経験的バイオインフォマティクス手法の当該相互作用計算の結果を良いほうに導くので、有用性も大きく従来手法と異なる効果を奏する。

- 25 また、本実施例は、標的タンパク質と種々の低分子化合物との相互作用のスコア化によって得られたコンフォメーションを、分子動力学計算式を内包したドッキングプログラムであるDOCKやAutoDockやGOLDにおいて、また、

分子動力学計算プログラムであるA m b e rやC h a r mなど既存のドッキングソフトの初期コンフォメーションとして用いることができる。これは、本実施例において得られた初期コンフォメーションが簡便に得られるのに加えて、実験を再現する精度が高いので、他のソフトプログラムとの組み合わせによって、有用な結果を得られる。

また、本実施例は、標的タンパク質の立体構造に類似しているファミリー高分子タンパク質セットに結合した種々の低分子化合物データベースであるC E l i b (F P (f i n g e r p r i n t) s e t e x t r a c t e d f r o m c o l l e c t e d l i g a n d s i n t h e b i n d i n g s i t e (結合部位のリガンド集合から抽出された化合物指紋セット)) を基に、任意のF P (f i n g e r p r i n t) を使う計算の過程で、標的タンパク質の立体構造を解析して活性部位を指定することを必要としない方法とすることができる。従来手法では、安定コンフォメーションが高いスコアを持つようにするために、D O C KやA u t o D o c kやG O L Dなど既存のドッキングソフトを使つてのドッキング計算において、予め標的タンパク質の立体構造を解析して活性部位を指定することをする必要があったが、これらに比べて、本実施例は、従来手法と異なる高い効果を有し、文献等の学習を通じて活性部位を指定する必要がなく有用である。

[結論]

本実施例による方法は、バイオインフォマティクスの観点から既知のタンパク質リガンド複合体 (P r o t e i n - L i g a n d C o m p l e x) の相互作用情報を定義したスコアを用いて的確にドッキングシミュレーションに反映することに成功した。

従来からも既存のドッキングソフトの出力を既知のタンパク質リガンド複合体 (P r o t e i n - L i g a n d C o m p l e x) の情報をドッキングシミュレーションに加えることにより、精度を上げるころみは行われているが、これらの方法では、研究者の叡智と実践に依存しており一般性がない。

本実施例による手法は、相同性 (Homology) 検索および立体構造重ね合わせを自動でおこない、さらに、本手法で提示されたスコア関数を用いることにより、精度よくドッキング構造を得ることができた。

これらのことにより、研究者によるヒューマンインタベーションを多く必要とせず広く使用できる。また、本手法で提示されたスコア関数は既存のドッキングソフトと組み合わせることも可能である。

すなわち、本実施例による方法は下記の三点においてきわめて有用である。

本実施例による手法は、バイオインフォマティクスの観点から既知のタンパク質-リガンド複合体 (Protein-Ligand Complex) の相互作用情報を的確にドッキングシミュレーションに反映できるところが従来手法とは異なる。更に、本実施例による手法は、リガンドに適切な物理量、距離拘束といったパラメータを受容体との相補性、および既知リガンドのコンフォメーションおよび原子種を考慮して自動的に付加できるという高い効果を発揮し、当然これらのことは新医学的、生物学的に重要な標的タンパク質とリガンドの相互作用のバイオインフォマティクス情報は年ごとに蓄積するので新規骨格医薬品もしくは類似骨格の探索にきわめて有用である。さらにテーラーメイド医療時代の到来で実験情報が豊富な標的タンパク質のドラッグデザイン (Drug Design) が必要になるので、本実施例による方法は、きわめて有用である。

[実施例2]

実施例2として、EGFR (Epidermal growth factor receptor) を標的タンパク質とした場合の k_2 と k_3 の最適化とインシリコスクリーニングについて以下に説明する。ここで、図52は、EGFRからの細胞内シグナル伝達経路を示した図である。

上記実施例1のChooseLD法において定義されたFPAScoreスコアの k_2 、 k_3 値は標的タンパク質に応じて最適化可能な係数として定義した。そこで、標的タンパク質に対して有効に機能するかどうか検証を行った。上皮増殖因子受容体ファミリーであるEGFRは癌治療において、重要な阻害標的とな

っている (J. Biol. Chem. 277 46265-46272 (2002), Cell 125 1137-1149 (2006))。そのため、EGFRを標的タンパク質として用いて、インシリコスクリーニングをおこなった。

5 (EGFRの立体構造構築)

EGFRのアミノ酸配列はNCBI (Wheeler, D. L. et al., Nucleic Acids Res. (2007) Nov 27) ACCESSION ID P00533を使用し、鋳型をPDBID 1M17のA鎖とした。アライメントは図53に示すものを使用した。図53は、EGFRのアミノ酸配列のアライメントを示す図である。

ホモロジーは約99%であり、立体構造を予測するよりはむしろ、1M17のC末端の残基欠損を補うことを目的としている。上記アライメントを用いてホモロジーモデリングソフトFAMS Ligand & Complex (Proteins, Suppl 7 122-127 (2005))を用いてモデルを構築した。ここで、図54は、構築されたEGFRのモデルを示す図である。

CIRCLEスコア (Terashi, G. et al. Proteins, (2007))は71.367であった。なお、鋳型の1M17_Aのスコア82.110はであった。CIRCLEスコアはPDB等から得られた実験構造座標データベースに所属するタンパク質のX線構造から得られた統計的ポテンシャルであり、スコアが正の方向に大きいほど既知のタンパク質X線構造の環境を満たしていることになり、すなわち、X線構造に近いモデルであると言える。

(EGFR特異的FPライブラリーの構築)

実施例2のChooseLD法に従い得られたFPライブラリーとして用いたリガンドのPDBIDは下記の通りである。

1AD5, 1AGW, 1BYG, 1E9H, 1FGI, 1FIN, 1FPU, 1FVV, 1GAG, 1H1P, 1H1Q, 1H24, 1H25, 1H26, 1H

27, 1I44, 1IEP, 1IR3, 1JPA, 1JQH, 1K3A, 1KS
W, 1M17, 1M52, 1MP8, 1MQB, 1OEC, 1OGU, 1OI9,
1OIU, 1OPJ, 1OPK, 1OPL, 1PF8, 1PKG, 1QCF, 1
QMZ, 1QPC, 1QPD, 1QPE, 1QPJ, 1ROP, 1RQQ, 1S
5 M2, 1SNU, 1T46, 1U4D, 1U54, 1U59, 1UWH, 1UW
J, 1VYW, 1XBB, 1XBC, 1XKK, 1Y57, 1Y6A, 1Y6B,
1YKR, 1YOL, 1YOM, 1YVJ, 1YWN, 2B54, 2B7A, 2
BDF, 2BDJ, 2BKZ, 2BPM, 2C0I, 2C0O, 2C0T, 2C
4G, 2C5N, 2C5O, 2C5P, 2C5T, 2C5V, 2C5X, 2DQ
10 7, 2E2B, 2ETM, 2EVA, 2EXM, 2F4J, 2FB8, 2FGI,
2FOO, 2G1T, 2G2F, 2G2H, 2G2I, 2G9X, 2GNF, 2
GNG, 2GNH, 2GNI, 2GQG, 2GS6, 2GS7, 2H8H, 2H
CK, 2HEN, 2HIW, 2HK5, 2HWO, 2HWP, 2HYY, 2HZ
0, 2HZ4, 2HZI, 2HZN, 2I0V, 2I0Y, 2I1M, 2I40,
15 2ITN, 2ITO, 2ITP, 2ITQ, 2ITT, 2ITU, 2ITV, 2
ITW, 2ITX, 2ITY, 2ITZ, 2IVS, 2IVT, 2IVU, 2I
VV, 2IW6, 2IW8, 2IW9, 2J0J, 2J0K, 2J0L, 2J0
M, 2J5F, 2J6M, 2NRU, 2NRY, 2OF2, 2OF4, 2OFU,
2OFV, 2OG8, 2OIQ, 2OJ9, 2OO8, 2OSC, 2OZO, 2
20 POC, 2P2H, 2P2I, 2P4I, 2SRC, 2UUE

(IC50既知化合物の入手)

BIOMOL (<http://www.biomol.com/>) のWebサ
イトからEGFRを競合的に阻害し、IC50値が既知の化合物の平面構造を1
1個入手した。図55は、入手した11個の阻害剤の平面構造を示す図である。
25 図55において、その化合物の平面構造に対応付けて、IC50値を示している。
これらの化合物の三次元座標は、Chem3Dを用いて立体構造を発生させたの
ち、Chem3D付属のエネルギー最小化計算を行ったものを使用した。

(EGFRのインシリコスクリーニングのためのk₂, k₃値を最適化)

FPAScoreのk₂値を0.5から5.0の範囲で変更し、MDL Comprehensive Medicinal Chemistry (MDL CMC) Library (Symyx Technologies, Inc. Corporate Address: 3100 Central Expressway, Santa Clara, CA 95051) をEGFRに活性のないダミー化合物と仮定し、既知の阻害剤がそれらの化合物に比べて上位に順位付けされるかといった実験を行った。

図56は、FPAScoreで定義されたk₂値を0.5から5.0の範囲に変更した際の収穫率折れ線グラフを示す図である。このときk₃値は1.0とした。randomの直線は、ランダムに母集団から化合物を選択した場合に既知阻害剤が得られる推定順位の直線であり、この直線からより下位に折れ線が描けるのならば、FPAScoreでの順位付けにおいて上位に阻害剤を検出できる能力が高いということになり、すなわちインシリコスクリーニングの性能がよいことを意味する。k₂値が0.5, 1.0, 5.0の場合、化合物の出現順位が6から折れ線が上昇を始めている。k₂値が2.0, 3.0の折れ線を比較すると、9, 10位において2.0の線がより、収穫率が良好であった。そこで、k₂値を2.0とした。

図57は、FPAScoreにおけるk₃値を0.5から2.0の範囲に変更した際の収穫率折れ線グラフを示す図である。このときk₂値は1.0とした。どのk₃値においても、おおむね同様の直線を得られたが、k₃値が0.5, 2.0の場合は、順位が10, 11において、折れ線が上昇しているためk₃値1.0を最適値とした。

(Tc 下限値の最適化)

FPライブラリーにふくめるリガンドのTcの下限値を設定した。Tcの下限値を限定することにより、ドッキングリガンドに類似しない化合物を除外できる。収穫率折れ線が最適になるようなTc下限値を決定した。

図58は、Tc上限値を1.00とし、Tc下限値の範囲を0.08から0.32まで0.08刻みで変化させた場合の、それぞれのTc範囲におけるインシリコスクリーニングの結果を示す図であり、活性既知化合物の出現個数が横軸、FPAScoreによる順位が縦軸となっている。Tc下限値が0.24の場合
5 において、出現個数1から6個においてx軸にこのような良好な折れ線となっていることから、この値を最適なTc下限値とした。なお、Tc下限値0.32時における折れ線は出現個数2個付近から急激に上昇している。これは、Tc下限値による絞り込みでFPライブラリーに使用すべきリガンドを除外してしまったためであると考えられ、インシリコスクリーニングにおいて、単にドッキングリ
10 ガンドと類似しているFPをもったリガンドだけを含めたとしても成功しないということを意味していると考ええる。

図59は、PDBに登録されているタンパク質—リガンド複合体構造既知のPDB IDとそのリガンドの順位付けを示す図である。図60は、図59のリガンドIDと化合物名を対応付ける図である。図59に示すように、順位付けを行っ
15 たりガンドには、EGFR阻害剤も含まれる。これらのリガンドはFPライブラリーに含まれているので、これら由来のFPがFPアライメントの際に主に使用され、FPAScoreが高くなり上位にランクインしたと考えられる。Tc下限値が0.24のインシリコスクリーニングにおいて、0.08の場合と比較して、これらのリガンド出現順位が分散しているが、タンパク質—リガンド複合体
20 構造が解明されていないEGFRに対するIC50既知の化合物はTc下限値0.24の時が良好な収穫率カーブを描いていたことから、Tc下限値0.24が最適であると考えた。

(インシリコスクリーニングの結果)

$k_2 = 2.0$, k_3 値 = 1.0, Tc下限値 = 0.24とした時のEGFRイ
25 ンシリコスクリーニングの結果を以下に示す。上位100構造において、97構造がリン酸原子を含むATP誘導体であった。そこで、下記の絞り込みをおこなった。

- (1) 分子量350以上800以下の分子、リンを含む分子を除外
- (2) 重要な水素結合をしない分子を除外 (METの主鎖の窒素)
- (3) タンパク質原子とリガンド原子の衝突2.0 Å以下が存在するドッキングリガンド分子を除外

5 図61および図62は、Kinaseのインシリコスクリーニングによる絞り込みの結果の上位10位のタンパク質ーリガンド複合体を示す図である。なお図62は図61を別角度から見たものである。キナーゼ (Kinase) ドメインの空間内における立体構造相補性を満たし、かつ、相互作用に重要な水素結合を満たす構造がFPAScoreによるランキングに存在したことになり、本実施

10 例のChooseLD法がインシリコスクリーニングによる阻害剤探索にも有用であることが示された。なお、これらの試薬は購入可能であり、活性値を測定することが可能である。しかしながら、FPAScoreによる順位付けは、標的タンパク質の活性阻害の強さを直接あらわしているスコアではないため、FPに与えるスコアをFP構築法に依存して一律に与えるのではなく、結合定数の大き

15 さも反映できるようなスコアに改良することも可能であると考え。

[適用例]

上記実施例1, 2にかかるChooseLD法を様々な標的タンパク質に対して適用した結果を以下に示す。これらの結果は、実験による証明が必要である。一例目は、EGFRの二量体形成阻害剤探索に関するものである。二例目は、V

20 EGF2に対するKRN633, KRN951の複合体構造の予測に関し、タンパク質ーリガンド複合体構造の予測はX線構造解析による証明が必要である。三例目は、マラリアに対するインシリコスクリーニングに関しても、結合実験による証明が必要である。

(EGFRのTGF α 結合ドメイン阻害剤のインシリコスクリーニング)

25 図52で示したように、EGFRは二量体を形成することにより、シグナルを伝達することが知られている (Nat. Rev. Cancer. 4, 361-370 (2004))。リガンドとしてEGFRに結合するTrans

forming Growth Factor α (TGF- α) はEGFRが複合体を形成するために必要なペプチドである。すなわち、EGFRのTGF- α 結合ドメインの阻害剤開発は創薬のターゲットとなる。そこで、Choose LD法を用いて、EGFRのTGF- α 結合ドメインに対するインシリコスクリーニングを行った。EGFRの立体構造はPDB ID ; 1MOXをもちいた。TGF- α 結合ドメイン近傍にTGF類似体のペプチドをFAMS Ligand & Complex (Proteins 61, 122-127 (2005))を用いてモデリングしその側鎖を切り出した。

図63は、TGF- α 結合ドメイン近傍を表した図であり、黄色はTGF α 類似体のペプチドから側鎖のみを切り出したものであり、これをChoose LD法のFPライブラリーとして用いた。これは、ペプチド性の阻害剤がFPAScore上位に出現することを防ぐ目的で行われた。

図64は、MDL Comprehensive Medicinal Chemistry (MDL CMC) Libraryを用いたEGFRのTGF- α 結合ドメインに対するインシリコスクリーニングの結果を示す図であり、図65は、MDL ACD Libraryを用いた同インシリコスクリーニングの結果を示す図である。これにより、本実施例によって、タンパク質—タンパク質相互作用の情報を用いたドッキングが可能であることが示された。

(VEGFR2 (Vascular endothelial growth factor Receptor-2) に対するKRN633, KRN951の複合体構造の予測)

VEGFR2は、血管新生に関与するキナーゼ (kinase) であり、肺癌などの癌発症時に異常発現するタンパク質の一つであり、このタンパク質を特異的に阻害する化合物は癌の治療薬となる。阻害剤としてKRN633 (Mol. Cancer. Ther. 3, 1639-1649 (2004)), KRN951 (Cancer Res. 66, 9134-9142 (2006)) が知られている。しかし、これらの複合体構造は2007年12月時

点において、X線結晶構造解析がなされていない。そこで、VEGFR2とKRN633の複合体、および、VEGFR2とKRN951の複合体構造を予測した。ここで、図66は、KRN633 ($IC50 = 1.16 \text{ nm/L}$)の平面構造を示す図であり、図67は、KRN951 ($IC50 = 0.16 \text{ nm/L}$)の平面構造を示す図である。

VEGFR2の立体構造はPDBID 2P2HのA鎖を用いた。KRN633, KRN951のドッキングについての条件を記載すると、FPライブラリーに用いたリガンドはPSI-Blastによるホモロジー検索により入手し、ドッキングに使用されたFPライブラリーの上位10個はKRN633では、PDBID: 2HZN_A, 1YWN_A, 2J5F_A, 2IVU_A, 2H8H_A, 2OH4_A, 1GAG_A, 1FPU_A, 2COI_A, 2P4I_Aであり、KRN951においては、2IOV_A, 2HZN_A, 2OH4_A, 1FGI_A, 1YWN_A, 1FPU_A, 2OFU_A, 2COI_A, 2H8H_A, 2FGI_Aとなった。

図68～図71は、VEGFR2の活性近傍の立体構造を示した図である。タンパク質側の赤いリボンは α -ヘリックス、シアンのリボンは β シートを意味する。図68は、KRN633のVEGFR2活性部位近傍へのドッキングに用いたFPライブラリーに所属するリガンドにおいてドッキングに使用されたリガンドの上位10個の集合を表しており、図70は、同様に、KRN951のFPライブラリーに用いたFPライブラリーに所属するリガンドにおいてVEGFR2活性部位近傍へのドッキングに使用されたリガンドの上位10個の集合を表している。図69は、KRN633について、ChooseLD法を10回実行し、予測された構造10個をVEGFR2の活性部位近傍の立体構造とともに示している。FPライブラリーのリガンドの中でKRN633との類似度にTcを用いた場合、最高値は0.45であった。10回の試行において、ほぼ同様の構造を得ることができた。図71は、同様にKRN951について、ChooseLD法を10回実行し、予測された構造10個をVEGFR2の活性部位近傍の立体

構造とともに示している。予測構造の10個中8個がほぼ同じ構造であった。FPライブラリーのリガンドの中でKRN951との類似度にT_cを用いた場合、最高値は0.29であった。

(VEGFR-2のドッキング予測成功率の算出)

5 KRN633, KRN951の予測複合体構造の信頼性を評価するために、FPライブラリーに含まれるドッキングリガンドのT_c最大値を用いて、133セットから算出された統計的な成功率を算出した。図72は、133セットを用いたChooseLD法のドッキング性能試験の結果得られたT_c下限値を0.08に固定し、T_c上限値を変化させた時の予測成功率について、横軸にT_c上限値、縦軸に成功率としたグラフを示す図である。

すなわち、グラフにT_c上限値を内挿することによってChooseLD法適用時の予測成功精度を統計的に算出することが可能である。ただし、この統計的予測成功率は、標的タンパク質の立体構造、アミノ酸配列を考慮していない。KRN633のドッキングで用いたFPライブラリーに含まれるリガンドの中、T_cが最大のもは、0.45であったことから、0.36と0.56の時の予測成功率を用いて予測成功率を内挿すると、34.7%となった。KRN951も同様に、0.24と0.36の時の予測成功率から、推定予測成功率は、24.3%となった。133セットでの予測成功率で最も予測成功率が高かったGOLD Score STDが46.0%、DOCKは21.1%、AutoDockは26.6%であり、KRN633はAutoDockよりよく、GOLDには及ばない精度で予測でき、KRN951に関しては、AutoDockと同程度の精度で予測できたと考えられた。

25 (Plasmodium falciparum enoyl acyl carrier protein reductase に対する低分子 (NAD) が介在した状態でのドッキング)

Plasmodium falciparumのenoyl acyl carrier protein はマラリア熱の病原タンパク質の一つであり、脂

質合成に関与するタンパク質であるが、この脂質合成経路はヒトには存在しないため、このタンパク質の機能を阻害することはマラリア熱治療につながると考えられている (J. Biol. Chem. 277, 13106-13114 (2002))。

5 図73は、enoyl acyl carrier proteinの立体構造を示した図である。また、図73に示すように、このタンパク質を阻害する化合物としてトリクロサン等が存在し、複数の阻害剤とのX線結晶構造解析が行われており (J. Biol. Chem. 277, 13106-13114 (2002))、これらの阻害剤はNADを介して結合する。これらをFPライブラリーとして用いることにより、新規阻害剤のリード化合物探索のためのインシリコスクリーニングを実行した。

図74は、MDL Comprehensive Medicinal Chemistry (MDL CMC) Libraryを用いて、enoyl acyl carrier proteinのインシリコスクリーニングを行った結果のFPAScoreの上位10構造を示す図である。上側の円で囲まれている部分がインシリコスクリーニングによる結果であり、下側の円で示すNADの占める空間を考慮したドッキングが行われている。なお、MDL Available Chemicals Directory (MDL ACD) Libraryに対してもインシリコスクリーニングを実行しているが、本実施例の ChooseLD法によれば、NADやH₂O等の標的タンパク質の活性部位近傍に存在する低分子を考慮したドッキングが可能であることを示すことができた。

(結論)

本実施例では、新しく定義したFPAScoreをシミュレティッドアニーリングで最適化する方法を用いるバイオインフォマティクスに基づいたリガンドドッキングとインシリコスクリーニング法、ChooseLD法を開発した。また、85セットにおけるk1値の最適化を行うことにより、ハイスループットスクリーニング等に用いられることを想定した最適値を4.0に決定した。このk1値

をもちいた場合で、133セットにおいて、rmsdが2.0Å以下で実験構造を予測できる割合を指標としたとき、本実施例のChooseLD法のドッキング性能は既存の古典的物理関数を用いてドッキングを行うGOLDと同程度であり、Tc上限値が低い場合はDOCK, AutoDockと同程度であった。このことは、ファミリータンパク質由来のリガンドから構築したFPライブラリーに含まれるリガンドからFP構築法によって得られたFPが、自由エネルギーの低くなるような座標であるという仮定が正しかったことを示している。

しかし、従来の既存のドッキングソフトが自由エネルギー最小の構造を必ずしも探索できないことから、従来手法にはまだ改良の余地があることも示している。

また、133セットにおいて、予測に成功したPDBIDの分布の観点からChooseLD法とGlide, GOLD, FlexXと比較をおこない、PDBIDの分布の類似度をTcによって算出した結果、予測に成功するターゲットに独自性があり、本実施例であるChooseLD法と従来法との併用でインシリコスクリーニングの精度が上昇する可能性を示した。さらに、上述のように、本実施例2では、FPAScoreのk2値, k3値が標的タンパク質に応じて、最適化可能な変数であることをEGFRのkinaseドメインを標的タンパク質として用いて示した。これらの結果から、本実施例2のChooseLD法におけるFPAScoreのk1, k2, k3値は標的タンパク質に応じて最適化することにより、より多くの阻害剤、およびリード化合物が、インシリコスクリーニングスクリーニングされると考えられた。

[実施例3]

実施例3について以下に説明する。実施例3では、AMPKhomogamma1酵素の阻害薬（アンタゴニスト）並びに作動薬（アゴニスト）を開発する目的で、インシリコスクリーニングを行った。

まず、AMPKhomogamma1酵素を標的タンパク質として、そのアミノ酸配列の相同性検索を行ない99.7%のホモロジーをもつ2V9J_E（2V9JのE鎖）を鋳型として次のリガンドを含めてFAMS Ligandを用

いてAMPK h o m o G A M M A 1をモデリングした。ここで、図75は、AMPK h o m o G A M M A 1と2V9J__Eのアミノ酸配列のアライメントを示した図である。その結果、結合リガンドは、2V8Q__Eの3個のリガンドAMP__E__1327、AMP__E__1328、AMP__E__1329、2V92__E
5 の3個のリガンドATP__E__1327、ATP__E__1328、AMP__E__1329、2V9J__Eの3個のリガンドと2個のマグネシウムATP__E__1327、ATP__E__1328、AMP__E__1329、MG__E__1330、MG__E__1331、2QRE__Eの1個のリガンドAMZ__E__1002であった。

10 つぎに、2V9J__E以外のリガンドは、CEによるフィッティング（原子の種類を意識しないタンパク質同士の構造重ね合わせ）で2V9J__Eの座標系に重ね合わせた。2V9J__Eモデルの3ヶ所のATP（AMP）結合部位の中からMGイオンに依存しないAMP__E__1329サイトに絞って阻害剤並びに作動薬のスクリーニングを実施することにした。

15 本実施例のChooseLDを実施するに当たり、AMP__E__1329の結合部位から18Å以内のアミノ酸残基を切り出し2V9J__Eの受容体モデルとした。またChooseLDスクリーニング時には、受容体結合サイト以外のリガンドとMGイオンは補欠分子（C o f a c t o r）として受容体に含めた。また、本実施例のChooseLDのFPには、受容体結合部位のリガンド分子からリン酸基（PO3）を除いた3個のA d e n o s i n eと1-（5-A m i n o
20 o-4-carboxamide-1H-imidazole-yl）-riboseを使うことにしたが、リン酸基部分は候補化合物の官能基には向かない。そのため、リン酸をそのままFPにするのではなく、リン酸基の酸素原子と水素結合しているH i s 1 5 1とH i s 2 9 8（鋳型タンパク質の2V9J__EではH i s 1 5 0とH i s 2 9 7）ペアの相対的な距離を計算し、構造的なずれをG D T _ T S（0.5Å, 1.0Å, 1.5Å, 2.0Å）で計算し70%以上（
25 変更可能）G D T _ T Sの残基ペアであり、残基ペアから3.0Å以内（変更可

能)に存在するリガンドを95%NR_PDBからHETATMとして抽出した。なお、このとき2アミノ酸残基ではなく3アミノ酸残基を指定することも可能である。

GDT_TSはネイティブ構造に対してXÅ以下で重ねられる残基の割合を示す。その結果1061個のリガンドを取り出すことができた。これらのリガンドについて、2V9J_E受容体との衝突をチェックすることにより18個のリガンドあるいはリガンドの一部をFPに追加して合計22個のFPによりCMC (Comprehensive Medicinal Chemistry, 2006. 1, Elsevier MDL) データベースのスクリーニングを行った。

受容体側とリガンドとの原子衝突(2.0Å1原子以下、2.2Å3原子以下、2.4Å5原子以下)、リガンド分子量200から500まで、リガンドLogP-1から5まで、リガンドの環の数、水素供与原子、水素受容原子、それぞれ0から5などに設定した。ここで、図76は、リガンドが受容体全体に結合したCMC医薬品の結果リストを示す図である。

ここで、図77は、この中の1から10位までの2V9J_E受容体への結合状態を集合的に表した図である。緑色のボールアンドスティックモデルは2つのHIS残基を、黄色のスティックモデルは、3個のAdenosineと1-(5-Amino-4-carboxamide-1H-imidazole-yl)-riboseを示す。その間に10個の医薬品がドッキングされている。さらに3個のAdenosineと1-(5-Amino-4-carboxamide-1H-imidazole-yl)-riboseに加えてCMCスクリーニングで取れてきた医薬品化合物27個をFingerPrintとして計31個のFPを用いてACD (Available Chemicals Directory, 2008. 1, Elsevier MDL) のスクリーニングを行いAMPKhomogamma1酵素の阻害薬(アンタゴニスト)並びに作動薬(アゴニスト)の候補化合物を得ることもできる。

[他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲および本願発明の概要に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

5 例えば、インシリコスクリーニング装置100がスタンドアローンの形態で処理を行う場合を一例に説明したが、インシリコスクリーニング装置100とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するように構成してもよい。

10 また、実施の形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

15 このほか、上記文献中や図面中で示した処理手順、制御手順、具体的名称、各処理の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

 また、インシリコスクリーニング装置100に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

20 例えば、インシリコスクリーニング装置100の各装置が備える処理機能、特に制御部102にて行われる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該CPUにて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。また、外部システム200は、WEBサーバやASPサーバ等として構成していてもよく、
25 そのハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成していてもよい。また、外部システム200の各機能は、外部システム200のハードウェア構成中

のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

尚、プログラムは、後述する記録媒体に記録されており、必要に応じてインシリコスクリーニング装置100に機械的に読み取られる。すなわち、ROMまたはHDなどの記憶部106などは、OS (Operating System) 5 として協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAMにロードされることによって実行され、CPUと協働して制御部を構成する。また、このコンピュータプログラムは、インシリコスクリーニング装置100に対して任意のネットワーク300を介して接続された外部システム200等のアプリケーションプログラムサーバに記憶されていてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。 10

また、本発明に係るプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、15 光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating System) 20 に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。 25

記憶部106に格納される各種のデータベース等は、RAM、ROM等のメモ

リ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやデータベースやウェブページ用ファイル等を格納する。

- 5 また、インシリコスクリーニング装置100は、既知のパーソナルコンピュータ、ワークステーション等の情報処理装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

- 10 更に、装置の分散・統合の具体的形態は図示するものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。

産業上の利用可能性

- 15 標的高分子タンパク質にどのような当該化合物が有意に相互作用して、ドッキングするかの情報は新しい医薬品開発の要であり、またテーラーメイド医療というのは少なくとも、一つのアミノ酸残基の置換に対応して、従来ではきかなかつた医薬品の開発をすることなので、標的高分子タンパク質に結合した化合物の情報は実験決定済み化合物数においてその数が豊富であり、新薬の開発は非常に加速されるので、本願発明において述べたインシリコスクリーニング装置およびインシリコスクリーニング方法の産業上利用可能性は非常に大きい。

請求の範囲

1. 標的タンパク質に結合する候補化合物のスクリーニングを行う、記憶部と制御部を少なくとも備えたインシリコスクリーニング装置であって、
 - 5 上記記憶部は、

化合物中の複数個の原子に係る化合物指紋として、原子タイプと原子間結合規則を含む化学記述子を、上記候補化合物ごとに抽出して作成された化合物データベース、

を備え、
 - 10 上記制御部は、

上記標的タンパク質と立体構造が同一または類似するファミリータンパク質に結合することが既知の結合化合物について、上記標的タンパク質の座標系に変換した三次元座標とともに上記化合物指紋を抽出して結合化合物指紋セットを作成する化合物指紋作成手段と、
 - 15 上記化合物データベースに記憶された上記候補化合物について、上記結合化合物指紋セットの上記三次元座標を基底として算出した上記化合物指紋単位の二乗平均偏差を基礎とする相互作用スコアが最適化されるように、当該候補化合物の上記標的タンパク質に対する上記立体構造を演算する最適化手段と、

を備えたことを特徴とするインシリコスクリーニング装置。
 - 20
2. 請求項1に記載のインシリコスクリーニング装置において、

化合物に結合したタンパク質の立体構造およびアミノ酸配列を記憶するタンパク質データベース装置に接続され、

上記制御部は、

- 25 上記標的タンパク質の上記アミノ酸配列との相同性に基づいて、上記ファミリータンパク質および上記結合化合物を上記タンパク質データベース装置から検索する相同性検索手段、

を更に備え、

上記化合物指紋作成手段は、

上記相同性検索手段により検索された上記ファミリータンパク質に結合する上記結合化合物について、上記標的タンパク質の座標系に変換した上記三次元座標
5 とともに上記化合物指紋を抽出して上記結合化合物指紋セットを作成すること、
を特徴とするインシリコスクリーニング装置。

3. 請求項1に記載のインシリコスクリーニング装置において、

上記化合物指紋作成手段は、

10 上記ファミリータンパク質と上記標的タンパク質との構造重ね合わせにより、
当該ファミリータンパク質に結合する上記結合化合物の上記三次元座標を上記標
的タンパク質の座標系に変換し、変換された上記三次元座標とともに上記化合物
指紋を抽出して上記結合化合物指紋セットを作成すること、
15 を特徴とするインシリコスクリーニング装置。

4. 請求項1に記載のインシリコスクリーニング装置において、

上記化合物指紋作成手段は、

上記結合化合物と異なる他の上記化合物を参照して構造重ね合わせを行い、当
該結合化合物と当該他の上記化合物の原子間をまたがる上記化合物指紋を抽出し
20 て上記結合化合物指紋セットに追加する新規化合物指紋追加手段、
を更に備えたことを特徴とするインシリコスクリーニング装置。

5. 請求項1に記載のインシリコスクリーニング装置において、

上記化合物指紋作成手段は、

25 タニモト係数に基づき上記結合化合物と類似する上記化合物について、当該結
合化合物と当該化合物の原子間で原子の種類を入れ替え、上記標的タンパク質に
対する相互作用エネルギーを算出して当該結合化合物の上記化合物指紋よりも局

所エネルギー的に安定な上記化合物指紋を作成して上記結合化合物指紋セットに追加する新規化合物指紋追加手段、

を更に備えたことを特徴とするインシリコスクリーニング装置。

- 5 6. 請求項1に記載のインシリコスクリーニング装置において、
上記結合化合物は、公知のドッキングアルゴリズムにより上記標的タンパク質
に対して安定なコンフォメーションを持つと予測された化合物であること、
を特徴とするインシリコスクリーニング装置。

- 10 7. 請求項1に記載のインシリコスクリーニング装置において、
上記最適化手段は、
上記化合物指紋単位に上記二乗平均偏差を基礎とした、上記候補化合物の、上
記標的タンパク質との衝突具合、上記標的タンパク質の相互作用領域における存
在割合、および、上記標的タンパク質との直接相互作用割合を考慮に入れた関数
15 に基づいて、上記相互作用スコアを計算する相互作用スコア計算手段、
を更に備えたことを特徴とするインシリコスクリーニング装置。

8. 請求項1に記載のインシリコスクリーニング装置において、
上記最適化手段は、
20 上記相互作用スコアをメトロポリス法に基づいて判定し、判定結果にしたがっ
て上記候補化合物の基底となる上記化合物指紋を変更、増加、または減少させる
ことにより、上記相互作用スコアを最適化させること、
を特徴とするインシリコスクリーニング装置。

- 25 9. 請求項1に記載のインシリコスクリーニング装置において、
上記最適化手段は、
上記相互作用スコアの最適化過程において、上記候補化合物のコンフォメーシ

ョンを繰り返し変化させ、シミュレティッドアニーリング法に基づいて、当該候補化合物の上記コンフォメーション毎に当該候補化合物を剛体として繰り返し並進または回転させる構造変換手段、

を更に備え、

- 5 上記最適化手段は、上記構造変換手段により並進または回転された上記コンフォメーション毎の上記候補化合物について上記相互作用スコアを計算すること、を特徴とするインシリコスクリーニング装置。

10. 請求項1に記載のインシリコスクリーニング装置において、
10 上記最適化手段は、
上記相互作用スコアを以下の数式(1)に基づいて算出すること、

[数1]

$$\begin{aligned}
 FPAScore &= F(\text{aligned_fp}, \text{fp_rmsd}, \text{molecule}) \\
 &= \text{BaseScore}(\text{aligned_fp}, \text{fp_rmsd}) \\
 &\quad \times \text{fp_volume}(\text{molecule}) \\
 &\quad \times \text{fp_contact_surface}(\text{molecule}) \quad (1)
 \end{aligned}$$

- (ここで、上記FPAScoreは上記相互作用スコアを表し、上記F(aligned_fp, fp_rmsd, molecule)は、上記結合化合物と上記候補化合物間の上記化合物指紋単位のアライメント度および上記二乗平均偏差、
15 ならびに、上記候補化合物の上記標的タンパク質に対する上記立体構造を変数とする関数であり、上記BaseScore(aligned_fp, fp_rmsd)は、上記化合物指紋単位の一貫度および密集度を示す指標であり、上記fp_volume(molecule)は、上記結合化合物指紋セットの上記三
20 次元座標からなる空間を上記候補化合物が占める割合、および、上記標的タンパク質との衝突具合を示す指標であり、上記fp_contact_surface(molecule)は、上記候補化合物の上記標的タンパク質との接触度、および、上記結合化合物指紋セットの上記三次元座標への帰属度を示す指標であ

る。)

を特徴とするインシリコスクリーニング装置。

11. 請求項10に記載のインシリコスクリーニング装置において、

5 上記数式(1)における、

上記BaseScore(aligned_fp, fp_rmsd)は、以下の数式(2)に基づいて算出され、

[数2]

$$BaseScore(aligned_fp, fp_rmsd) = \frac{RawScore(aligned_fp)}{1 + \ln(fp_rmsd^{k1} + 1)} \quad (2)$$

10 (ここで、上記RawScore(aligned_fp)は、上記結合化合物と上記候補化合物間でアライメントされた上記化合物指紋における原子の数に基づく指標であり、上記fp_rmsdは、上記二乗平均偏差である。)

上記fp_volume(molecule)は、以下の数式(6)に基づいて算出され、

15 [数3]

$$fp_volume(molecule) = \ln \frac{1.0 + nafp^{k2}}{1.0 + nap^{k3}} \quad (6)$$

(ここで、上記nafpは、上記結合化合物指紋セットの上記三次元座標に基づく固有格子点領域に上記候補化合物の上記三次元座標が占有する格子点の数であり、上記napは、上記標的タンパク質の上記立体構造における原子の固有格子点領域に上記候補化合物の上記三次元座標が属する格子点の数であり、上記k2
20 および上記k3は、任意の定数である。)

上記fp_contact_surface(molecule)は、以下の数式(7)に基づいて算出されること、

[数4]

115

$$fp_contact_surface(molecule) = \frac{\sum_{i=1}^n density_of_atom(atom(i))}{total_density_of_atom(molecule)} \quad (7)$$

(ここで、上記nは、上記候補化合物の原子の数であり、上記atom(i)は、上記候補化合物のi番目の原子の上記三次元座標であり、上記density_of_atom(atom(i))は、当該原子の上記三次元座標が上記結合化合物指紋セットの上記化合物指紋に属している場合に、当該化合物指紋の上記原子と所定の距離で接触している上記標的タンパク質の原子の数と、当該化合物指紋の同一格子点に属する上記結合化合物の原子の数との和を返す関数であり、上記total_density_of_atom(molecule)は、上記density_of_atomの分布を降順に並べ換えたものを上記候補化合物の原子の数だけ順に足し合わせた数である。)

を特徴とするインシリコスクリーニング装置。

12. 記憶部と制御部を少なくとも備えたインシリコスクリーニング装置において実行される、標的タンパク質に結合する候補化合物のスクリーニングを行うインシリコスクリーニング方法であって、

上記記憶部は、

化合物中の複数個の原子に係る化合物指紋として、原子タイプと原子間結合規則とを含む化学記述子を、上記候補化合物ごとに抽出して作成された化合物データベースを備えており、

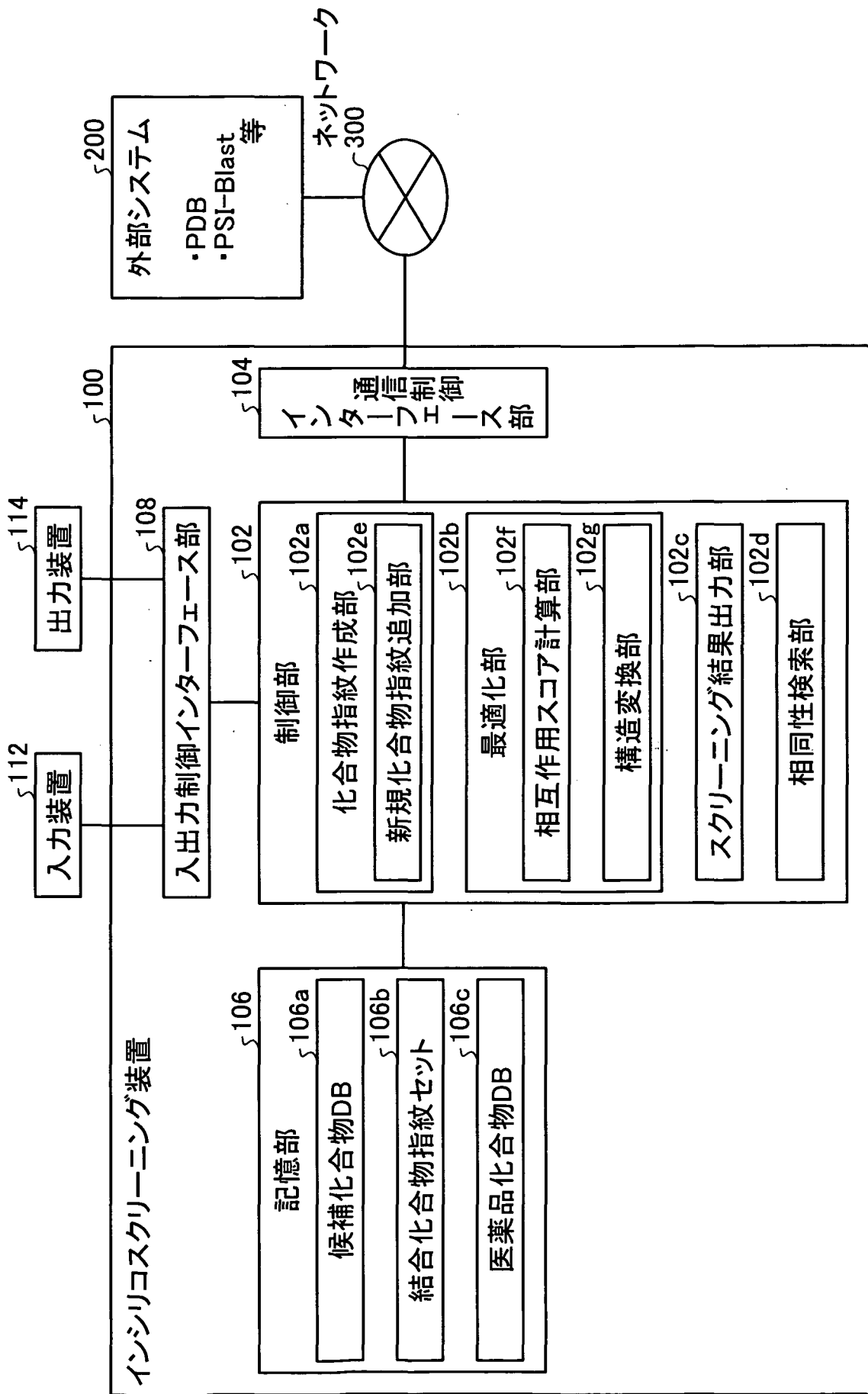
上記制御部において実行される、

上記標的タンパク質と立体構造が同一または類似するファミリータンパク質に結合することが既知の結合化合物について、上記標的タンパク質の座標系に変換した三次元座標とともに上記化合物指紋を抽出して結合化合物指紋セットを作成する化合物指紋作成ステップと、

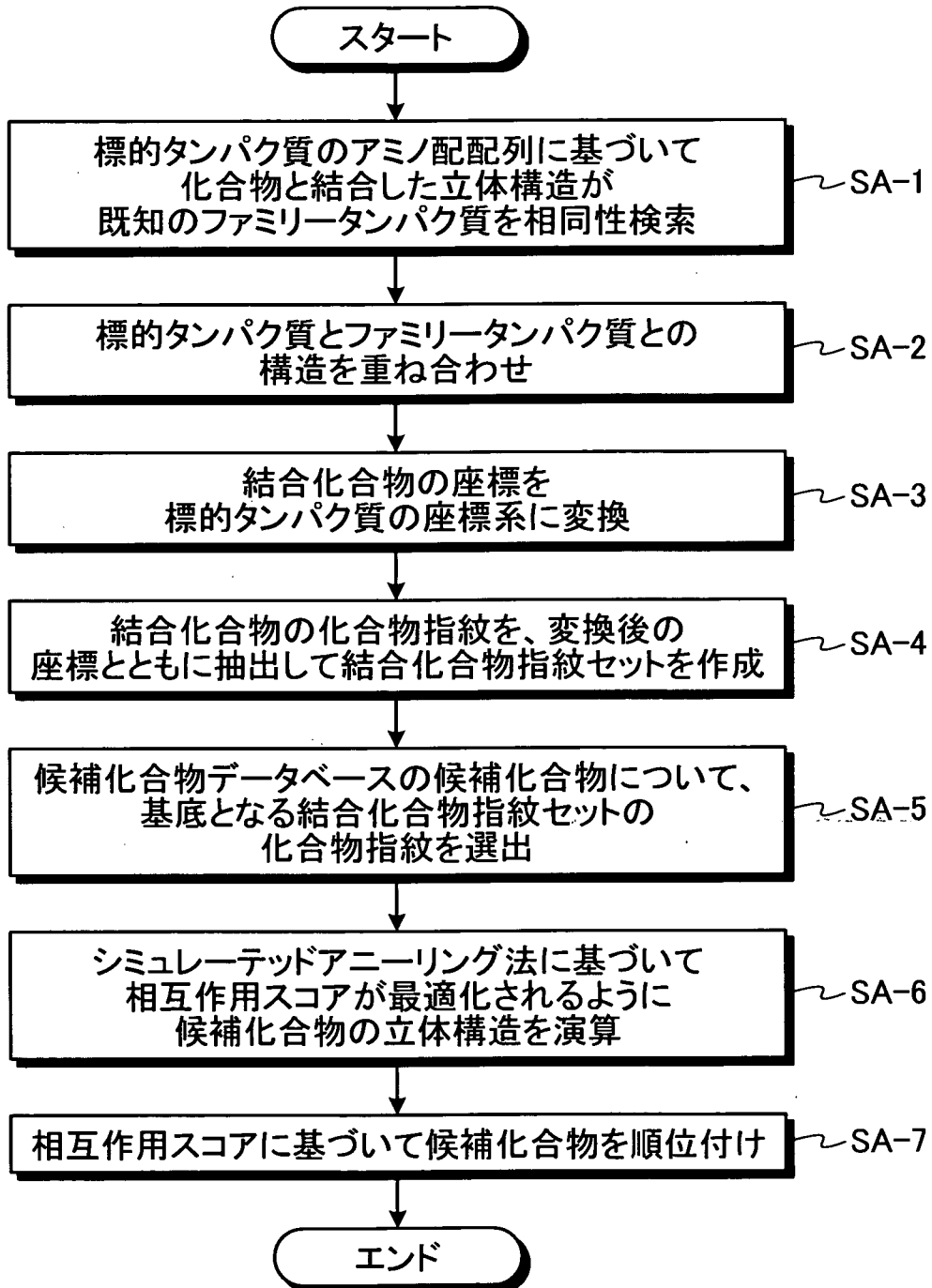
上記化合物データベースに記憶された上記候補化合物について、上記結合化合

物指紋セットの上記三次元座標を基底として算出した上記化合物指紋単位の二乗平均偏差を基礎とする相互作用スコアが最適化されるように、当該候補化合物の上記標的タンパク質に対する上記立体構造を演算する最適化ステップと、
を含むことを特徴とするインシリコスクリーニング方法。

第1図



第2図



第3図

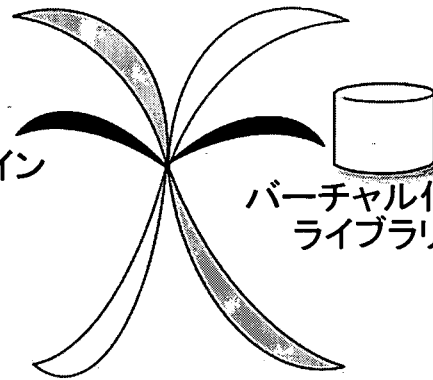
アプローチと材料

- バイオインフォマティクスアプローチ
- ・タンパク質構造予測
 - ・X線構造

- 第一原理的アプローチ
- ・極性相互作用
 - ・LJポテンシャル等

インシリコスクリーニング

- ・疾患関連タンパク質
- ・テーラーメイドドラッグデザイン



バーチャル化合物
ライブラリー

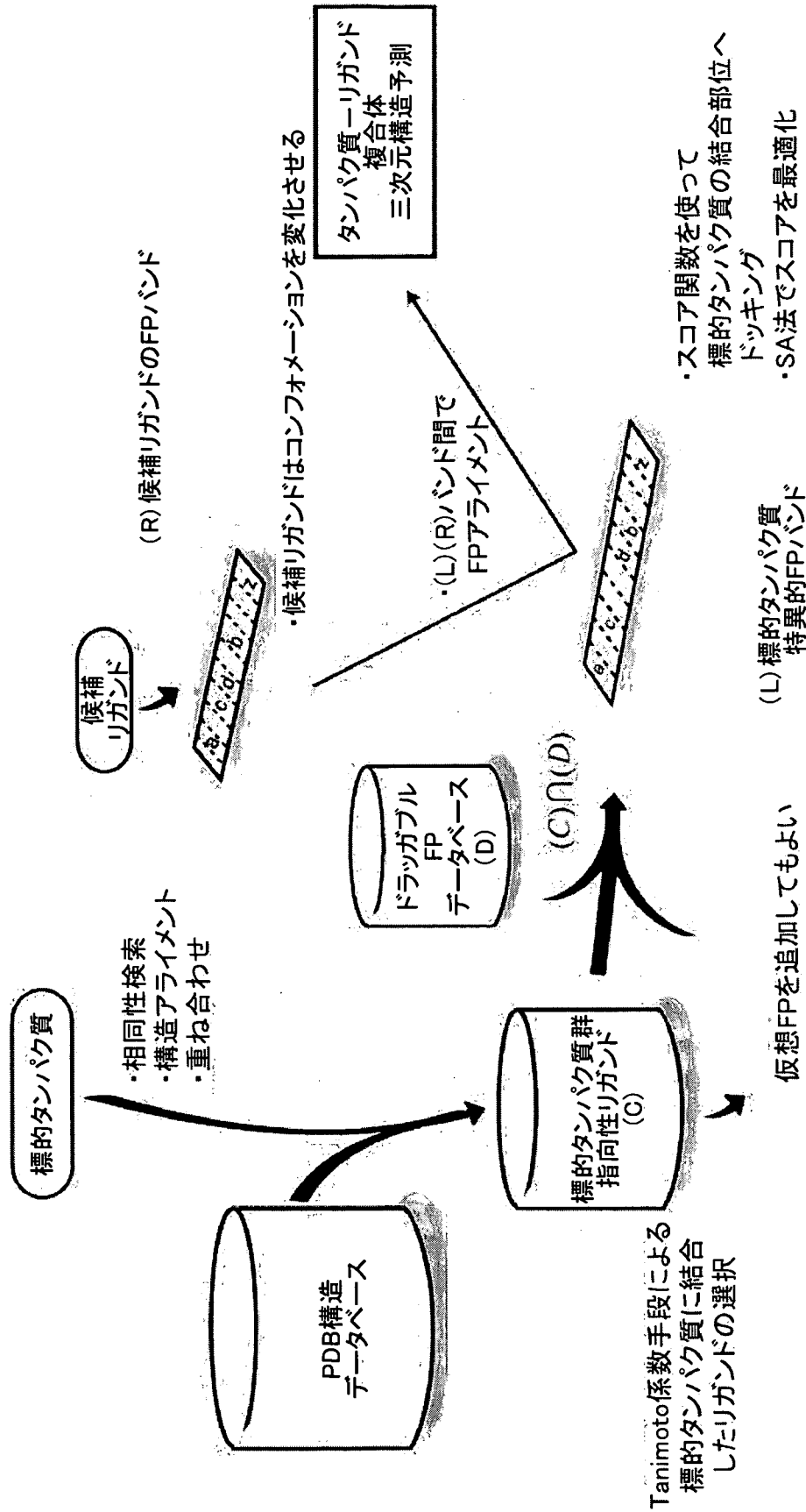
解法

AutoDock, DOCK,
FlexX, GOLD

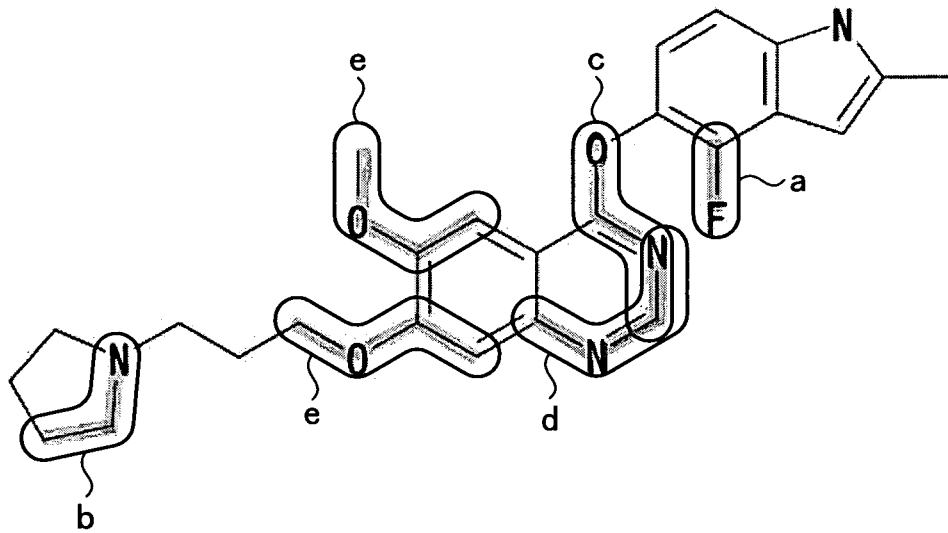
ChooseLD

第4図

仮想リガンドライブまたは
ベンチマークセットの結合リガンド



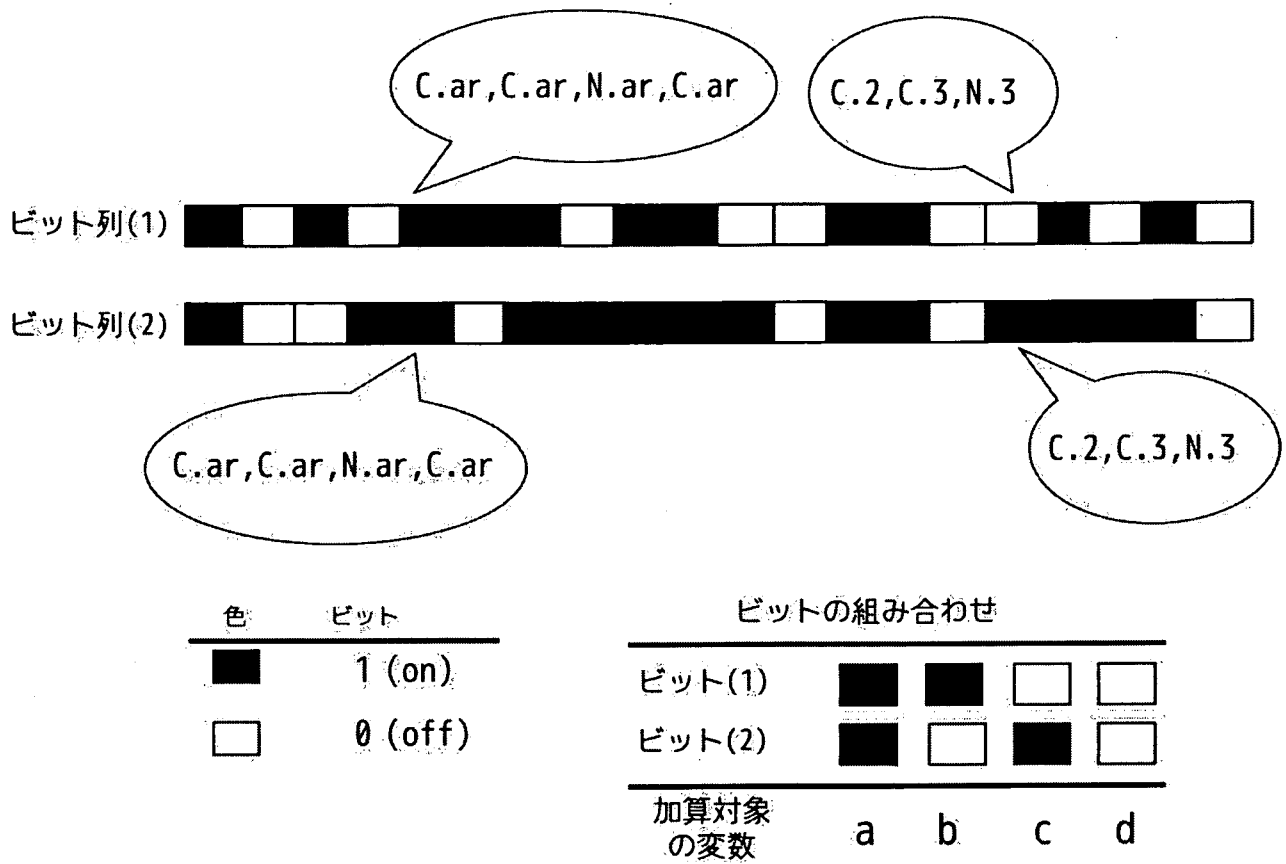
第5図



第6図

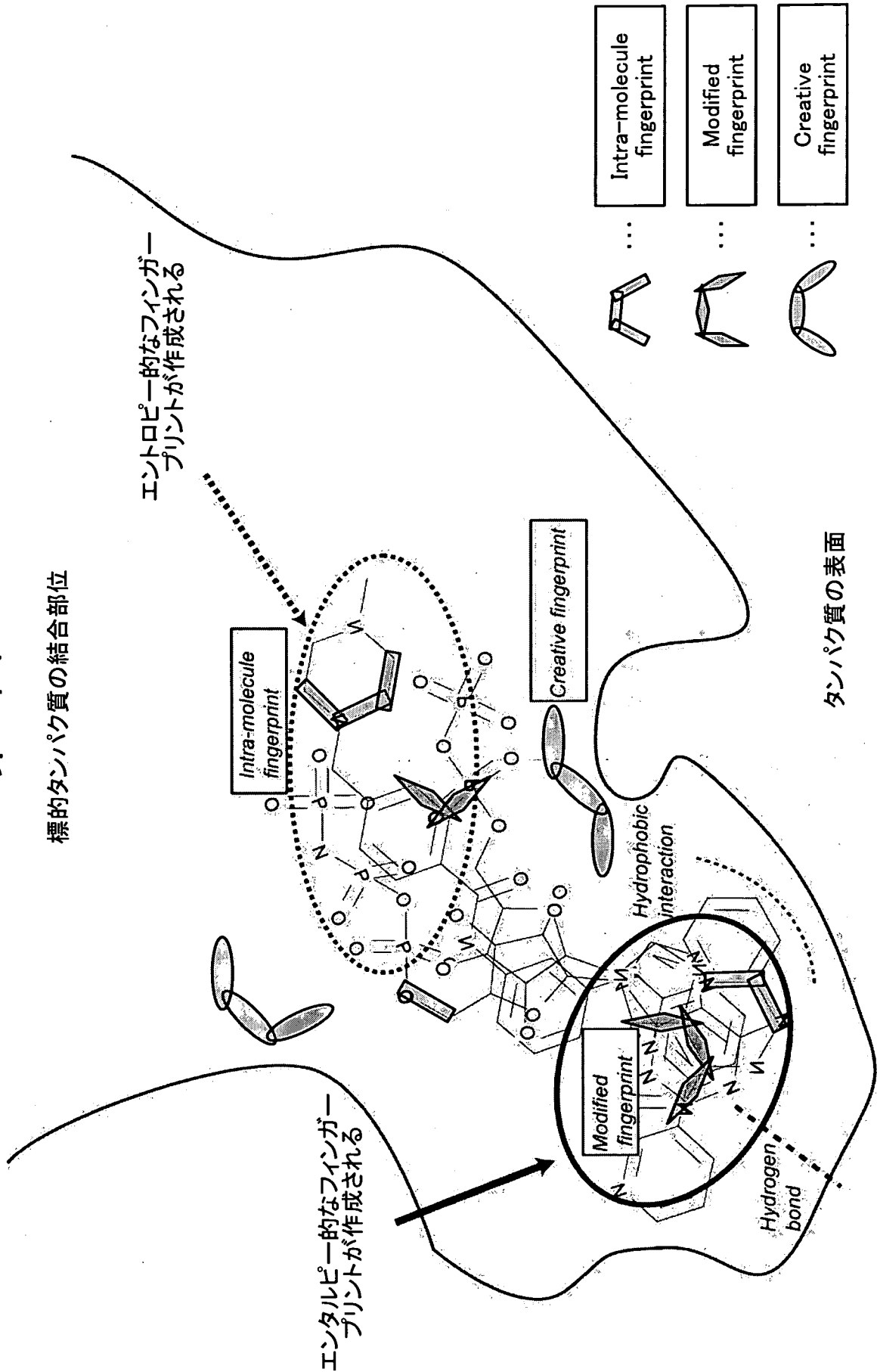
文字列	名前	文字列	名前
H	hydrogen	Si	silane
LP	lone pair	P.3	phosphorous sp3
Li	lithium	S.3	sulfur sp3
B	boron	S.2	sulfur sp2
C.3	carbon sp3	S.o	sulfoxide sulfur
C.2	carbon sp2	S.o2	sulfone sulfur
C.1	carbon sp1	Cl	chlorine
C.ar	carbon aromatic	K	potassium
C.cat	carbocation	Ca	calcium
C.co2	carbon sp2 carbonyl	Mn	manganese
N.3	nitrogen sp3	Fe	iron
N.2	nitrogen sp2	Co	cobalt
N.1	nitrogen sp	Ni	nickel
N.ar	nitrogen aromatic	Zn	zinc
N.am	nitrogen amide	Ge	germanium
N.pl3	nitrogen trigonal pla	Se	selenium
N.4	nitrogen sp3 positive	Br	bromine
O.3	oxygen sp3	Sr	strontium
O.2	oxygen sp2	Sn	tin
O.co2	oxygen in carboxylate	Te	tellurium
O.hoh	oxygen water	I	iodine
F	fluorine	Ba	barium
Na	sodium	Pb	lead (IV)
Mg	magnesium		

第7図

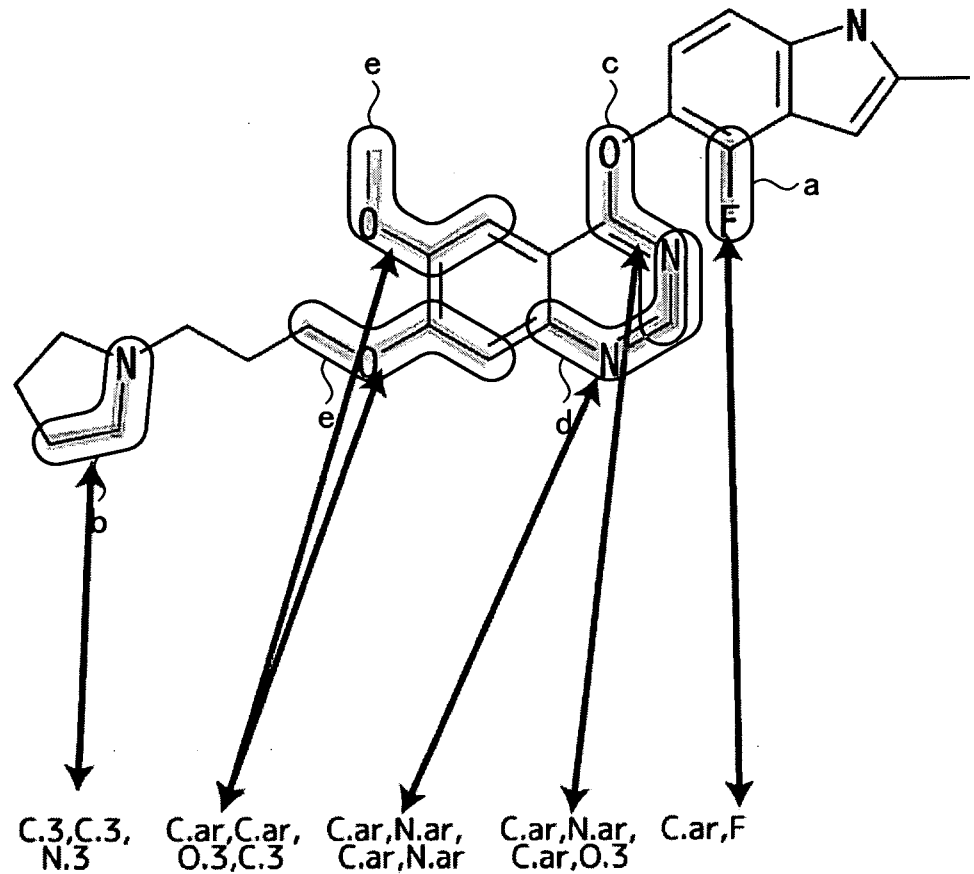


第8図

標的タンパク質の結合部位



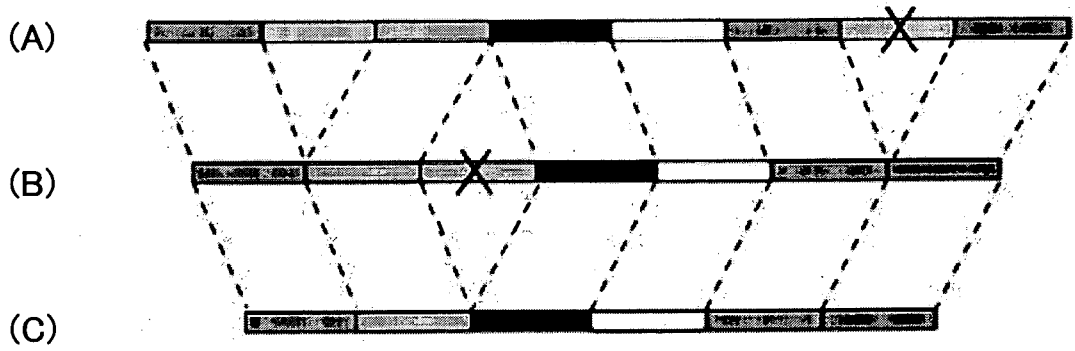
第9図



FPAScore算出
のためのFP band

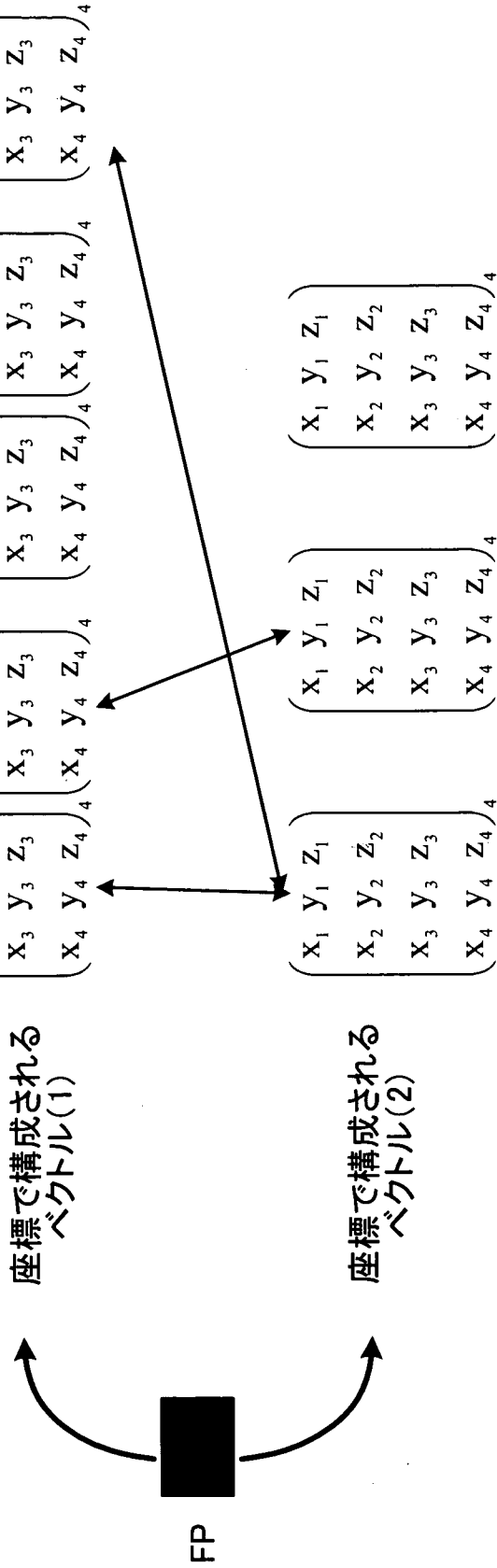
$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix}_3$	$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix}_4$	$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix}_4$	$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix}_2$	
	$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix}_4$			

第10図

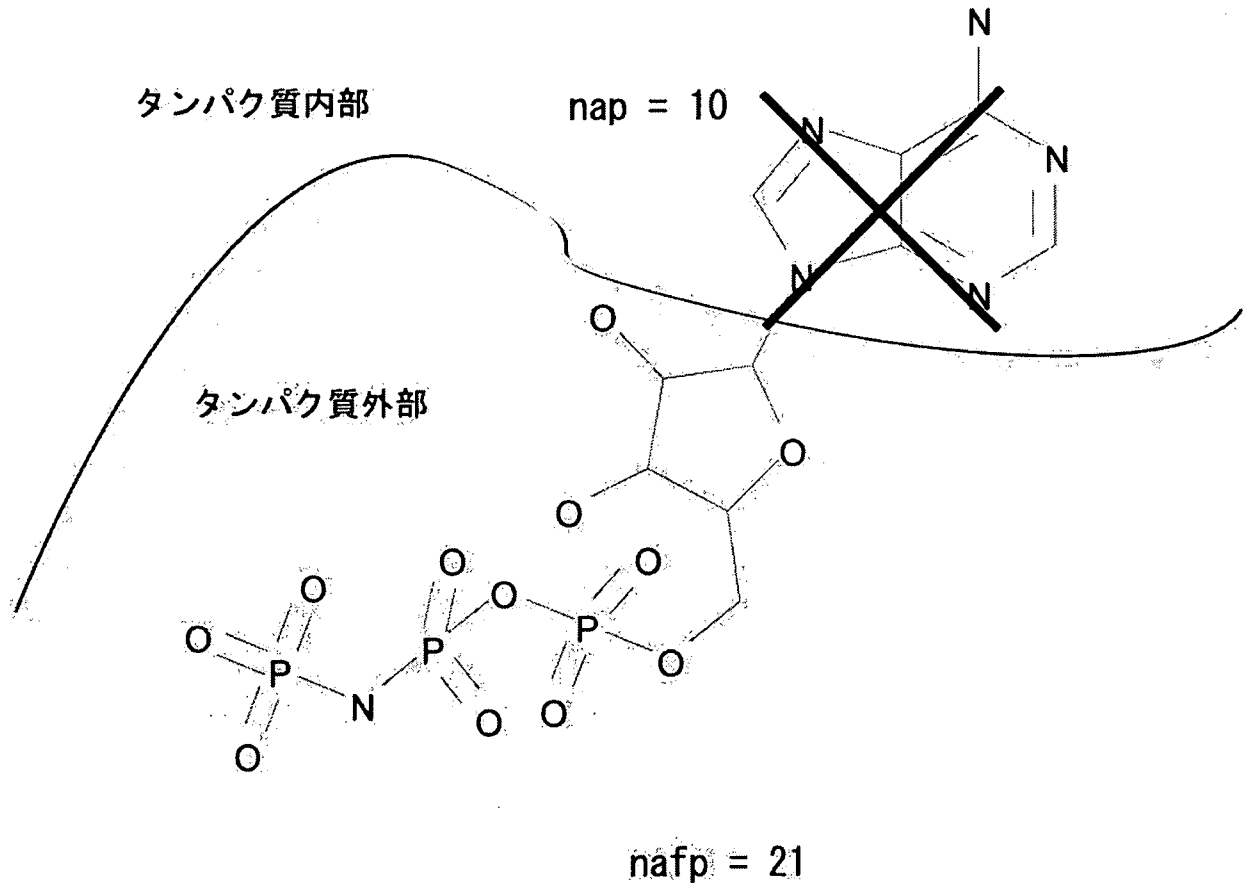


第11図

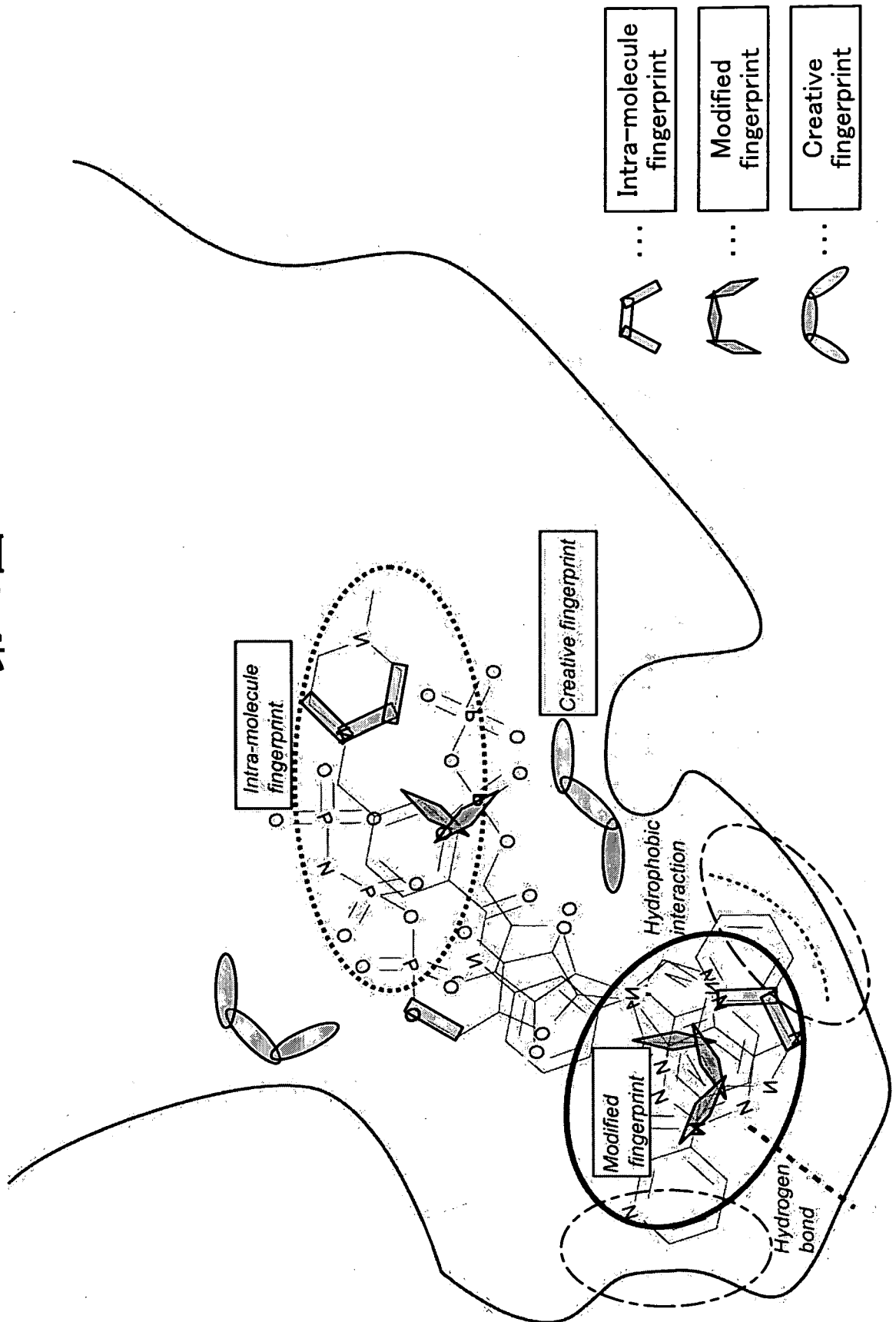
一つのFPIには2つのベクトルが登録されている



第12図

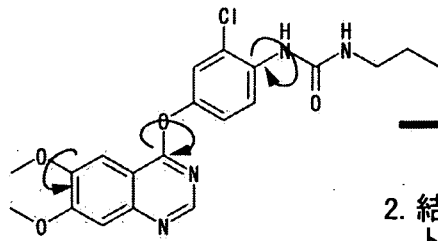


第13図



第14図

1. ランダムにコンフォメーションを変化 (1,000回)



2. 結合部位に
ドッキング

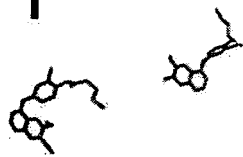
剛体を並進・回転



KRNG33 (VEGF inhibitor Nakamura et al³)

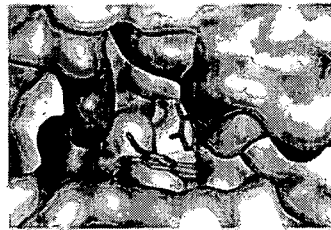
4. 1~3のステップを再試行

3. ドッキング構造を
構造プールに保存



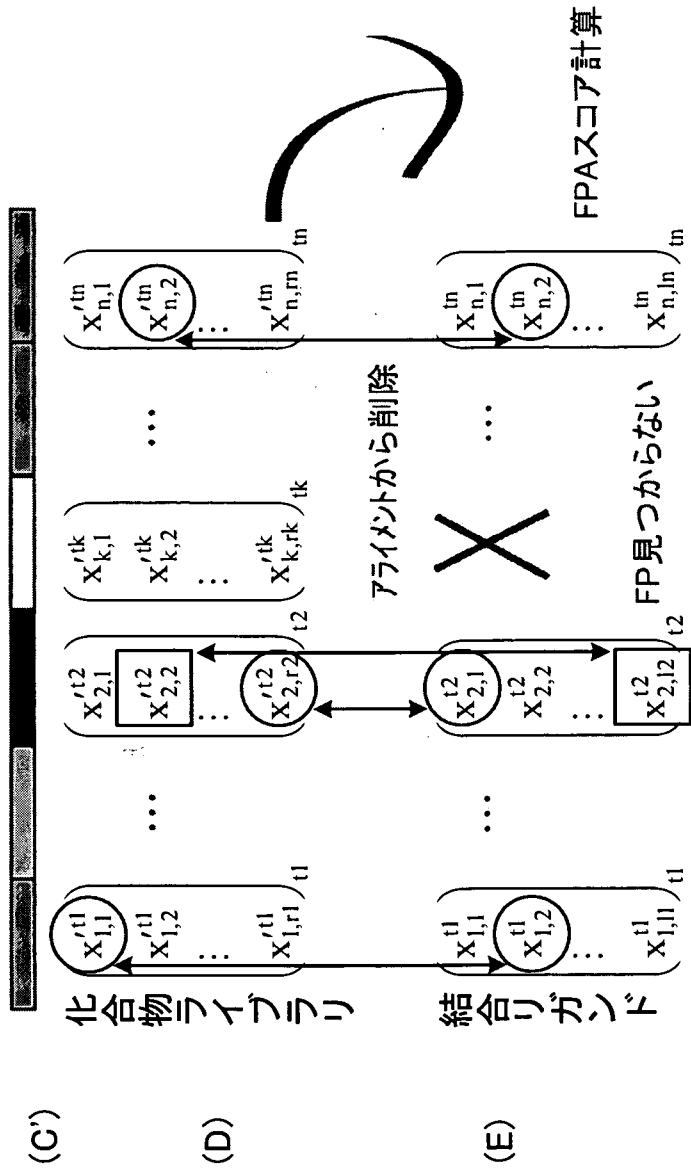
Structure pool

シミュレートアニーリングを10,000ステップ



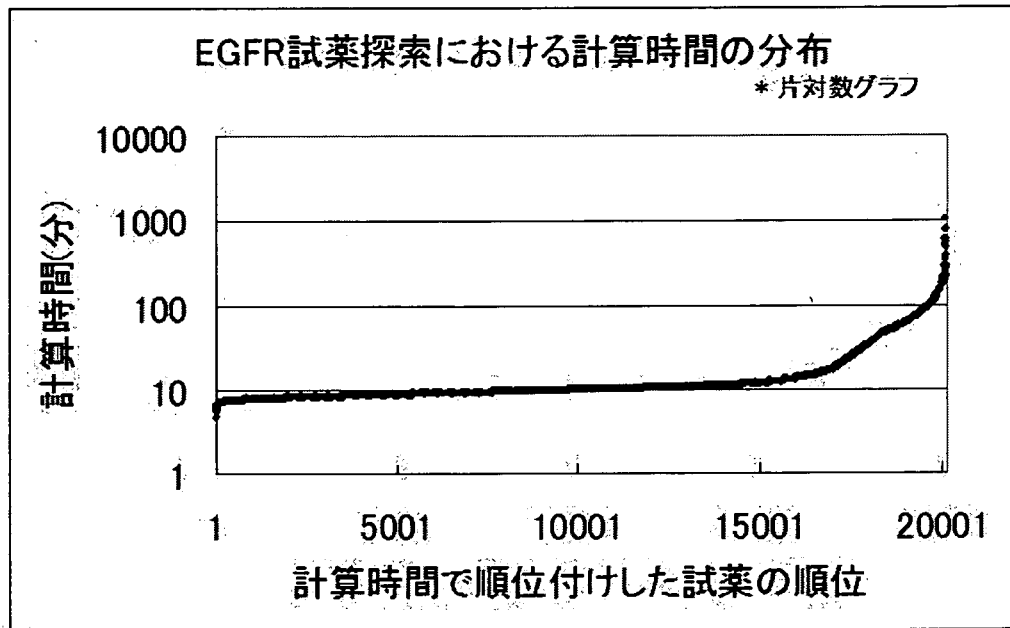
最後に、最大スコアを有する
ドッキング構造を選出

第15図

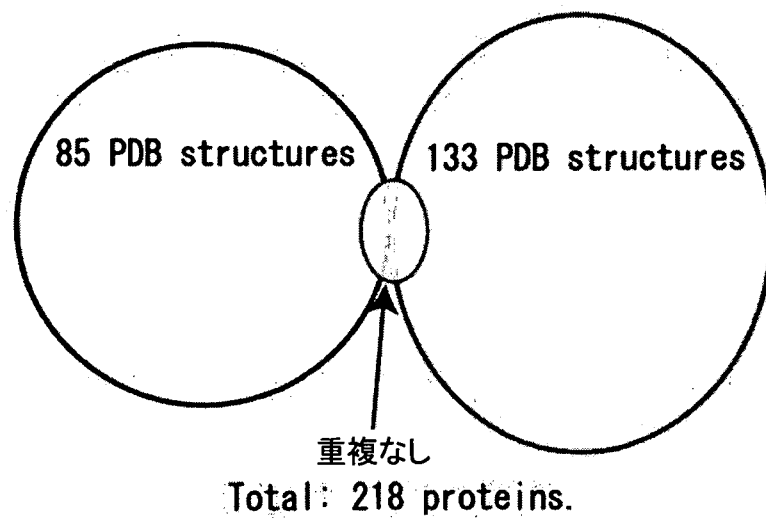


$$x_{a,b}^t = \begin{pmatrix} x_{a,b}^1 & y_{a,b}^1 & z_{a,b}^1 \\ \vdots & \vdots & \vdots \\ x_{a,b}^m & y_{a,b}^m & z_{a,b}^m \\ \vdots & \vdots & \vdots \\ x_{a,b}^t & y_{a,b}^t & z_{a,b}^t \end{pmatrix}$$

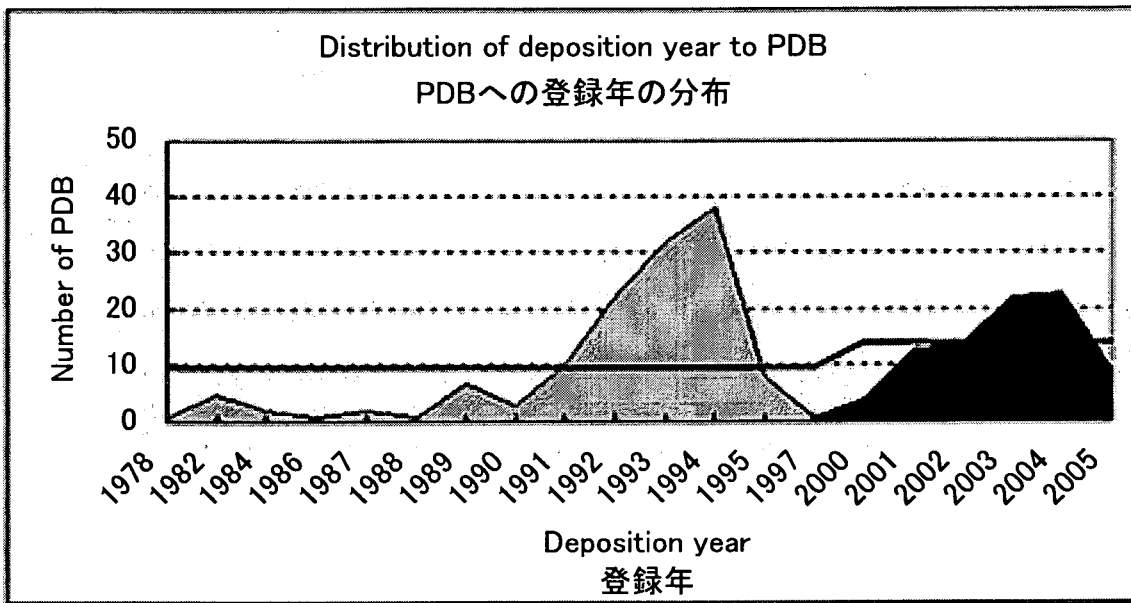
第16図



第17図



第18図



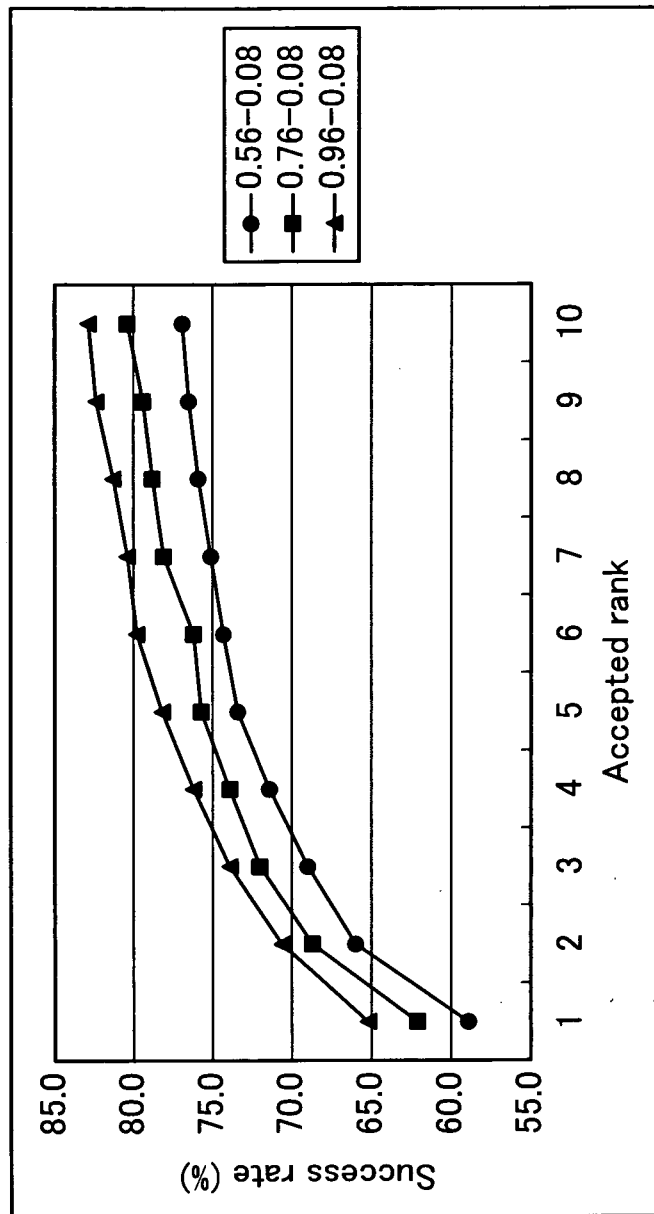
第19図

r.m.s.	Total	Good	Close	Errors	Wrong
≤ 0.5	8	8	0	0	0
> 0.5, ≤ 1.0	27	24	3	0	0
> 1.0, ≤ 1.5	20	7	13	0	0
> 1.5, ≤ 2.0	11	2	9	0	0
> 2.0, ≤ 2.5	2	0	2	0	0
> 2.5, ≤ 3.0	3	0	2	1	0
> 3.0	28	0	1	8	19

第20図

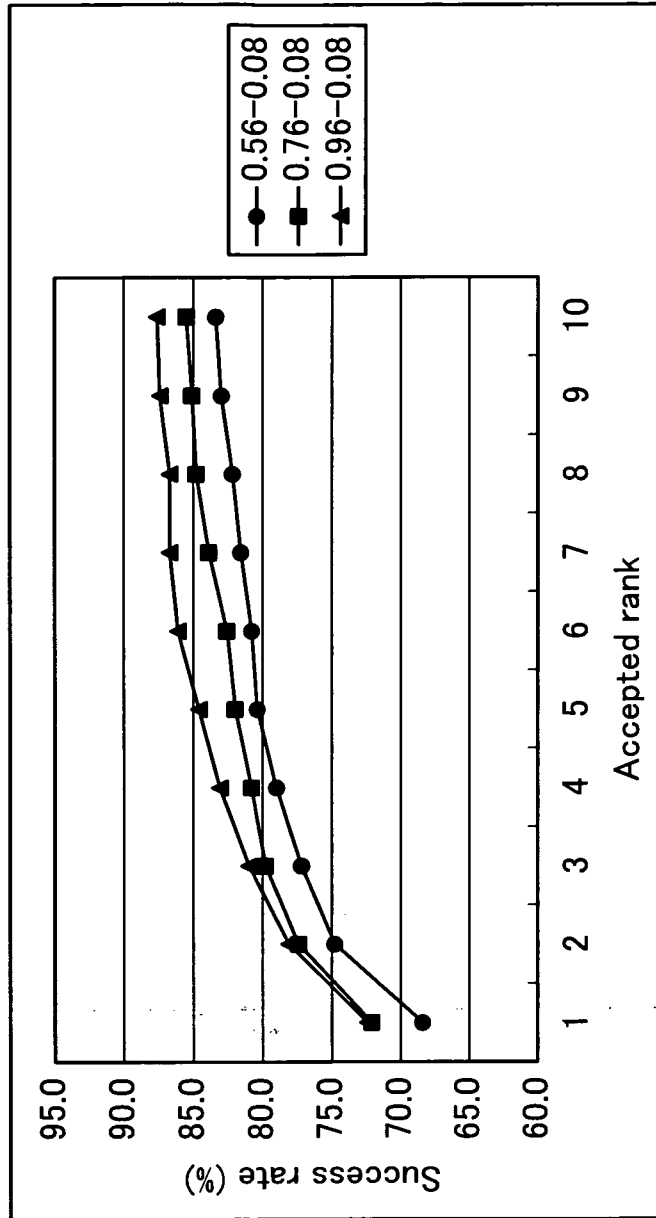
TC Range		k1					
		1.0	2.0	3.0	4.0	5.0	6.0
	0.56-0.08	46.0	54.6	57.6	58.9	56.3	58.6
	0.76-0.08	50.0	61.0	60.7	62.1	59.4	62.1
	0.96-0.08	55.6	62.1	64.4	65.2	65.8	64.8
	average(%)	50.5	59.2	60.9	62.1	60.5	61.8

第21図



	Tc Range		
	0.56-0.08	0.76-0.08	0.96-0.08
1	58.9	62.1	65.2
2	66.0	68.7	70.6
3	69.0	72.0	73.9
4	71.4	73.9	76.2
5	73.4	75.7	78.2
6	74.3	76.2	79.8
7	75.1	78.1	80.4
8	75.9	78.8	81.3
9	76.5	79.4	82.4
10	76.9	80.4	82.9

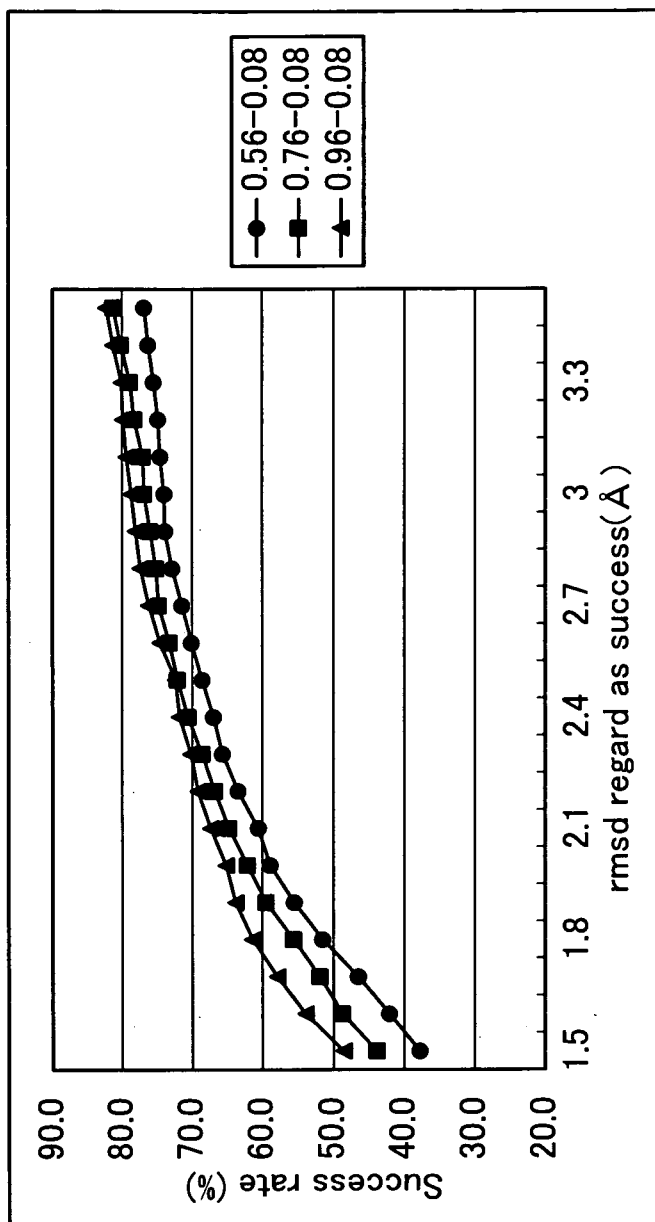
第22図



	Tc Range		
	0.56-0.08	0.76-0.08	0.96-0.08
1	68.4	72.1	72.4
2	74.8	77.4	78.1
3	77.2	79.8	81.0
4	79.0	80.8	83.1
5	80.4	82.0	84.6
6	80.8	82.6	86.1
7	81.6	83.9	86.7
8	82.2	84.8	86.7
9	83.0	85.1	87.4
10	83.4	85.5	87.6

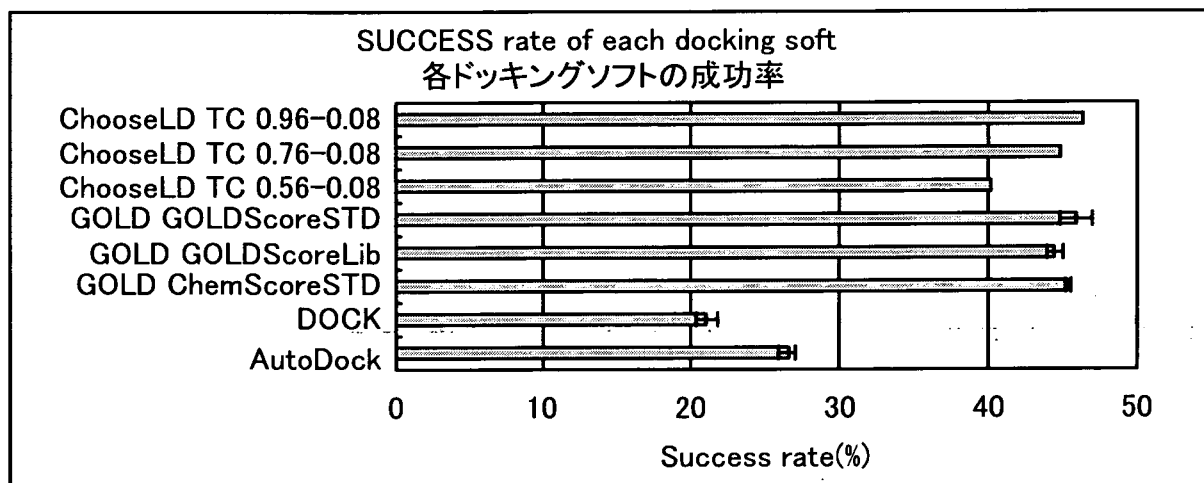
第23図

	Tc Range		
	0.56-0.08	0.76-0.08	0.96-0.08
1.5	37.8	43.9	48.5
1.6	42.1	48.7	53.9
1.7	46.5	51.9	57.9
1.8	51.5	55.6	61.4
1.9	55.5	59.5	63.8
2	58.9	62.1	65.2
2.1	60.6	64.8	67.4
2.2	63.5	66.8	69.2
2.3	65.7	68.6	70.2
2.4	67.0	70.6	71.8
2.5	68.6	72.1	72.4
2.6	70.1	73.3	74.6
2.7	71.5	74.8	76.2
2.8	72.9	75.2	77.4
2.9	73.9	76.0	78.1
3	74.0	76.9	78.7
3.1	74.6	77.1	79.4
3.2	74.9	78.3	79.9
3.3	75.5	78.9	80.1
3.4	76.3	80.2	81.3
3.5	76.9	81.3	82.3

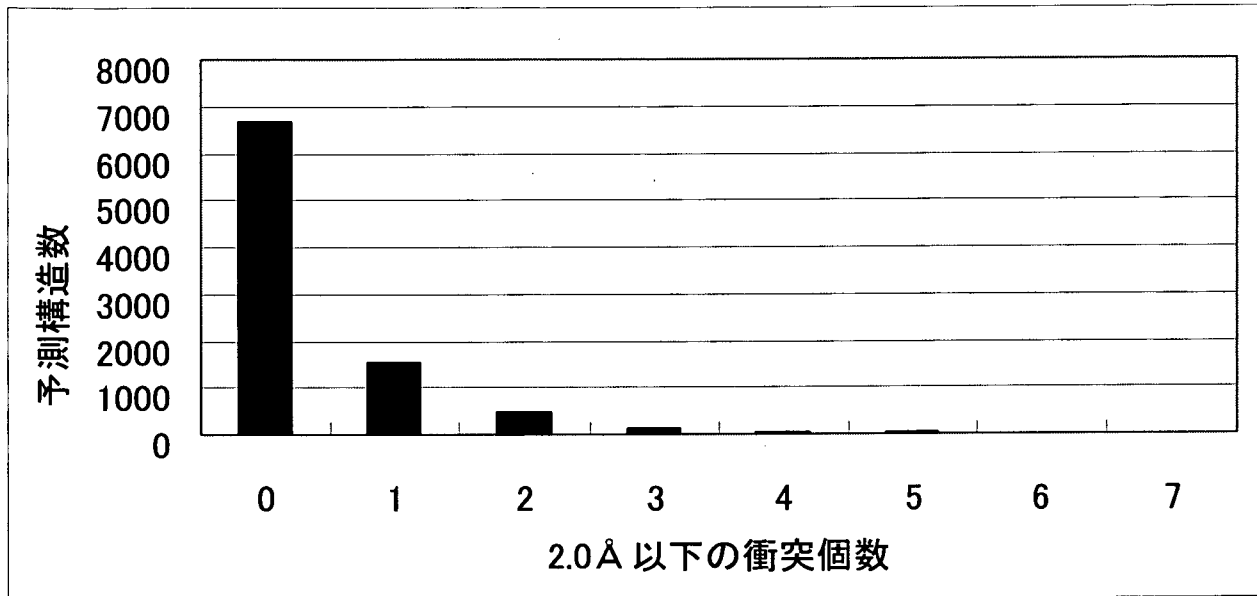


第24図

Docking soft	Corina	MINI	success rate (%)	SD(Corina, MINI)
AutoDock	26.2	27	26.6	0.57
DOCK	21.6	20.6	21.1	0.71
GOLD ChemScoreSTD	45.5	45.3	45.4	0.14
GOLD GOLDScoreLib	44.1	44.9	44.5	0.57
GOLD GOLDScoreSTD	45.2	46.7	46.0	1.06
ChooseLD TC 0.56-0.08	n/a	n/a	40.1	n/a
ChooseLD TC 0.76-0.08	n/a	n/a	44.8	n/a
ChooseLD TC 0.96-0.08	n/a	n/a	46.4	n/a



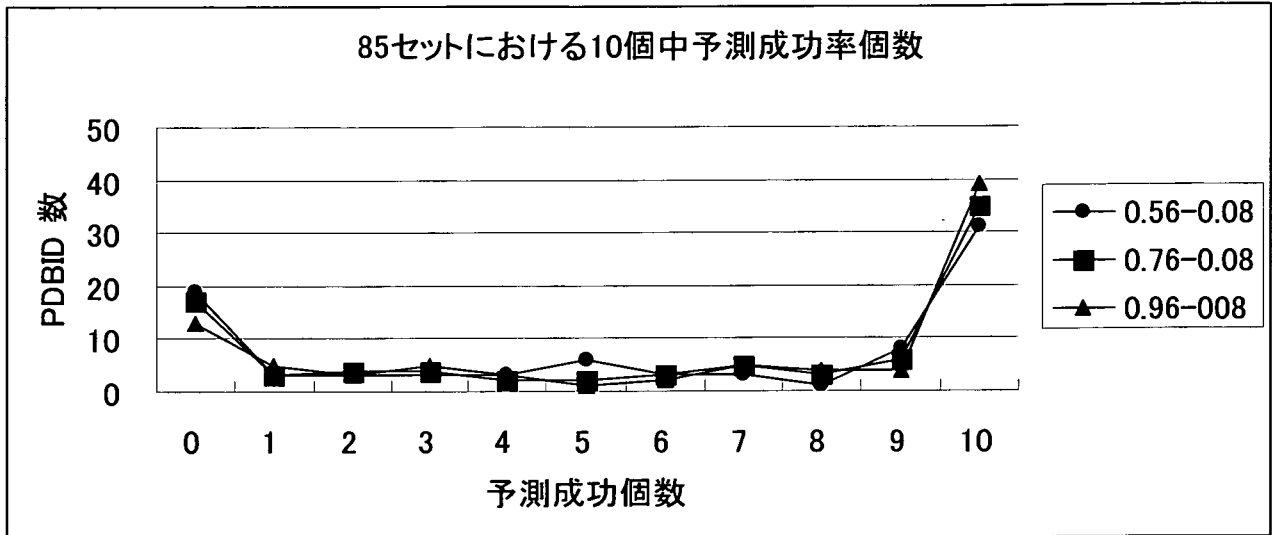
第25図



第26図

Tc 範囲	85 セットにおける 10 個中予測成功個数											合計	*1
	0	1	2	3	4	5	6	7	8	9	10		
0.56-0.08	19	3	3	3	3	6	3	3	1	8	31	83	62.7%
0.76-0.08	17	3	4	4	2	2	3	5	3	6	35	84	64.3%
0.96-0.08	13	5	3	5	3	1	2	5	4	4	39	84	65.5%

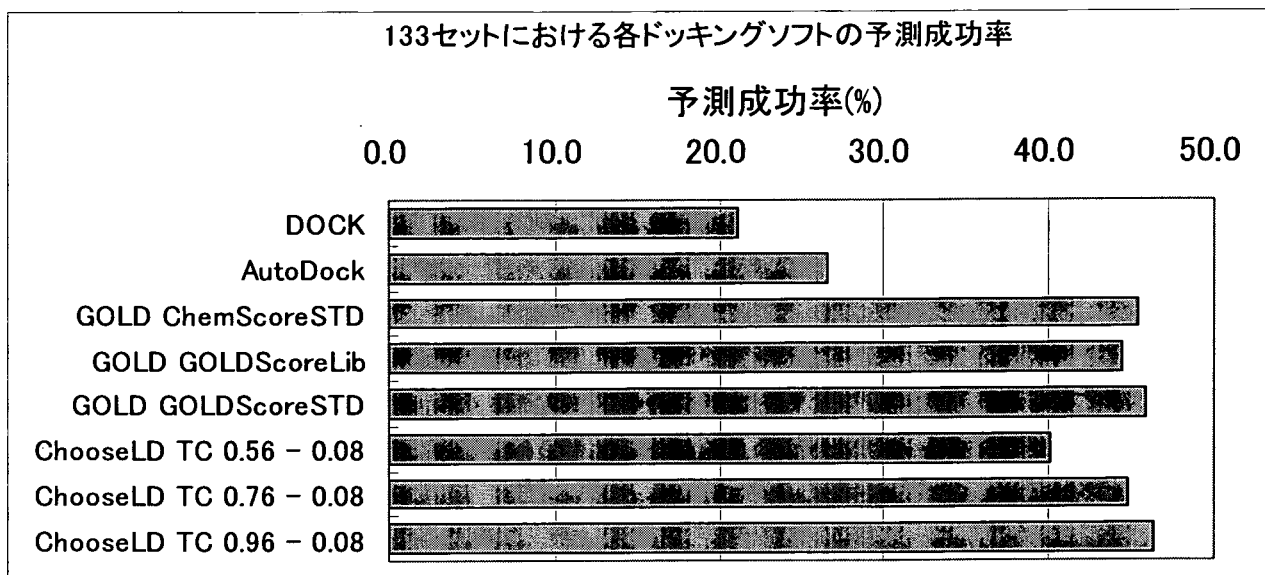
第27図



第28図

ドッキングソフト	予測成功率 (%)			ドッキングソフト	予測成功率 (%)
	Corina	MINI	平均		
DOCK	21.6	20.6	21.1	ChooseLD TC 0.56 - 0.08	40.1
AutoDock	26.2	27.0	26.6	ChooseLD TC 0.76 - 0.08	44.8
GOLD ChemScoreSTD	45.5	45.3	45.4	ChooseLD TC 0.96 - 0.08	46.4
GOLD GOLDScoreLib	44.1	44.9	44.5		
GOLD GOLDScoreSTD	45.2	46.7	46.0		

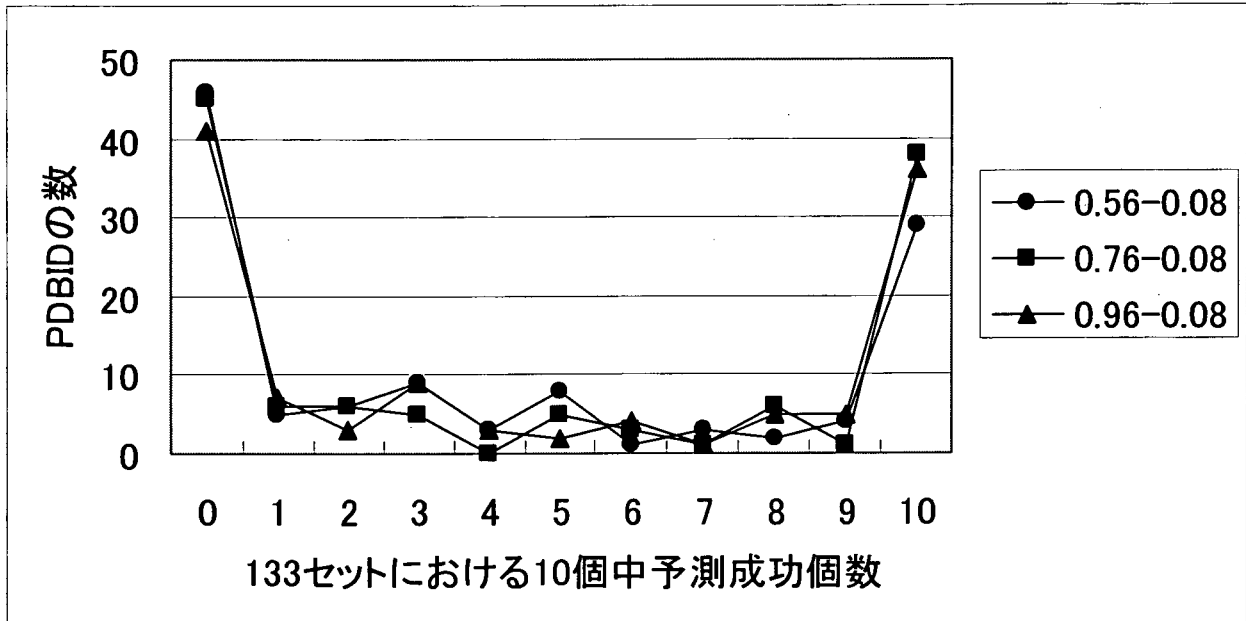
第29図



第30図

Tc 範囲	133 セットにおける 10 個中予測成功個数										合計	*1	
	0	1	2	3	4	5	6	7	8	9			10
0.56-0.08	46	5	6	9	3	8	1	3	2	4	29	116	40.5%
0.76-0.08	45	6	6	5	0	5	3	1	6	1	38	116	46.6%
0.96-0.08	41	7	3	9	3	2	4	1	5	5	36	116	45.7%

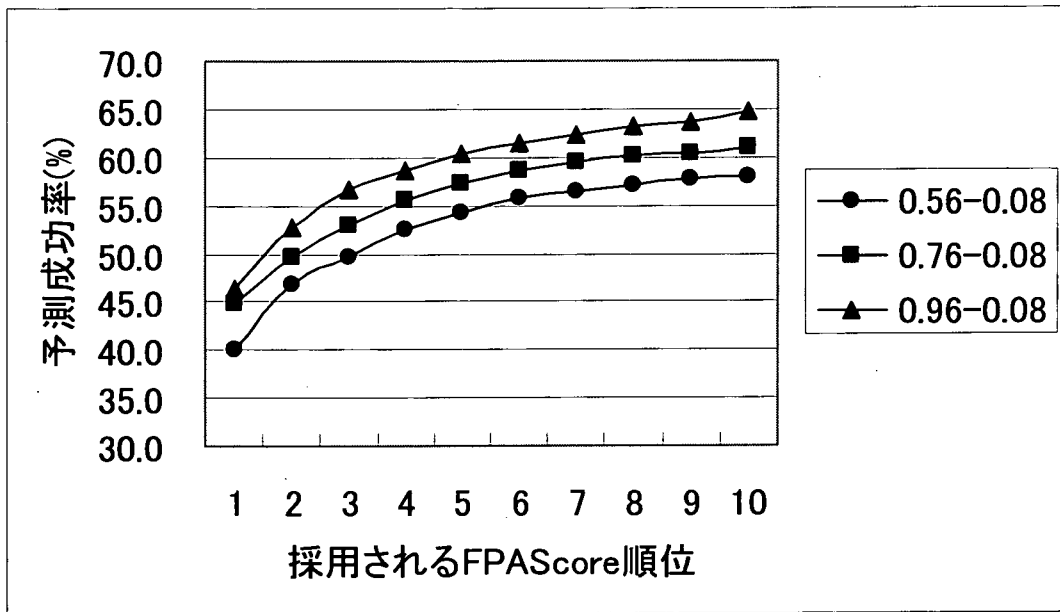
第31図



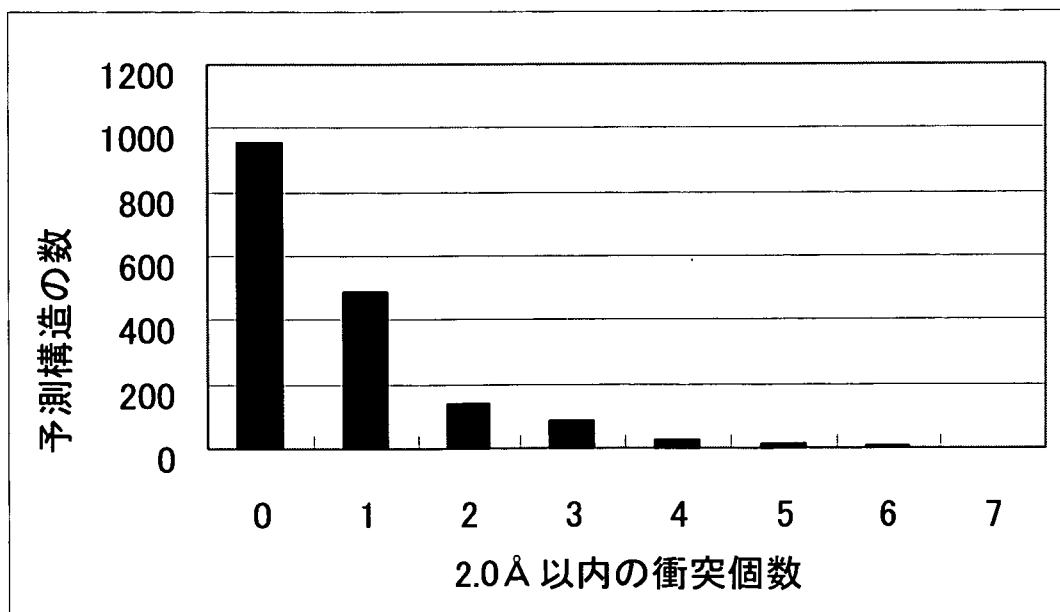
第32図

Tc 範囲	採用される FPAScore 順位									
	1	2	3	4	5	6	7	8	9	10
0.56-0.08	40.1	46.9	49.7	52.4	54.2	55.9	56.5	57.2	57.8	58.0
0.76-0.08	44.8	49.7	52.9	55.6	57.2	58.5	59.5	60.1	60.3	61.0
0.96-0.08	46.4	52.7	56.6	58.7	60.3	61.6	62.3	63.3	63.7	64.7

第33図



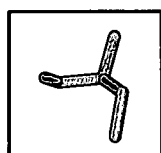
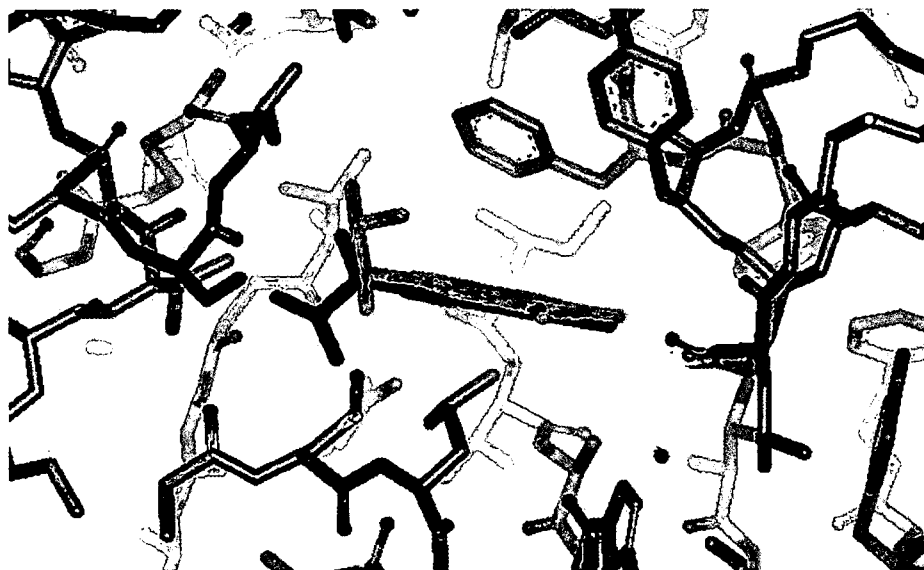
第34図



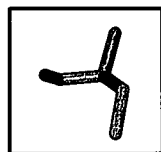
第35図

T _c 範囲	予測成功率(%)	T _c 範囲	予測成功率(%)
0.16 - 0.08	12.6	0.56 - 0.08	40.1
0.24 - 0.08	20.8	0.76 - 0.08	44.8
0.36 - 0.08	29.2	0.96 - 0.08	46.4

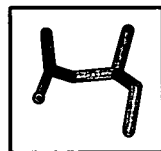
第36図



:リガンドの正解構造(中央のシアン)

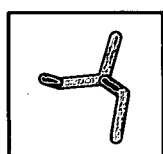
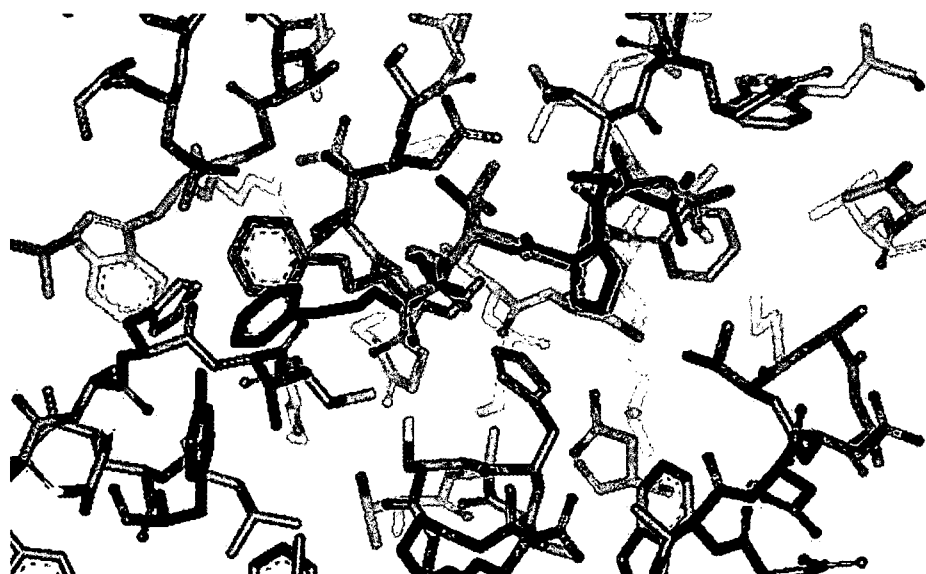


:リガンドの予測構造(中央の緑)

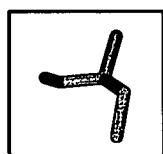


:タンパク質の結合部位(周辺部分)

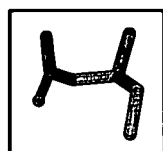
第37図



:リガンドの正解構造(中央のシアン)

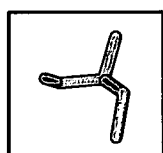
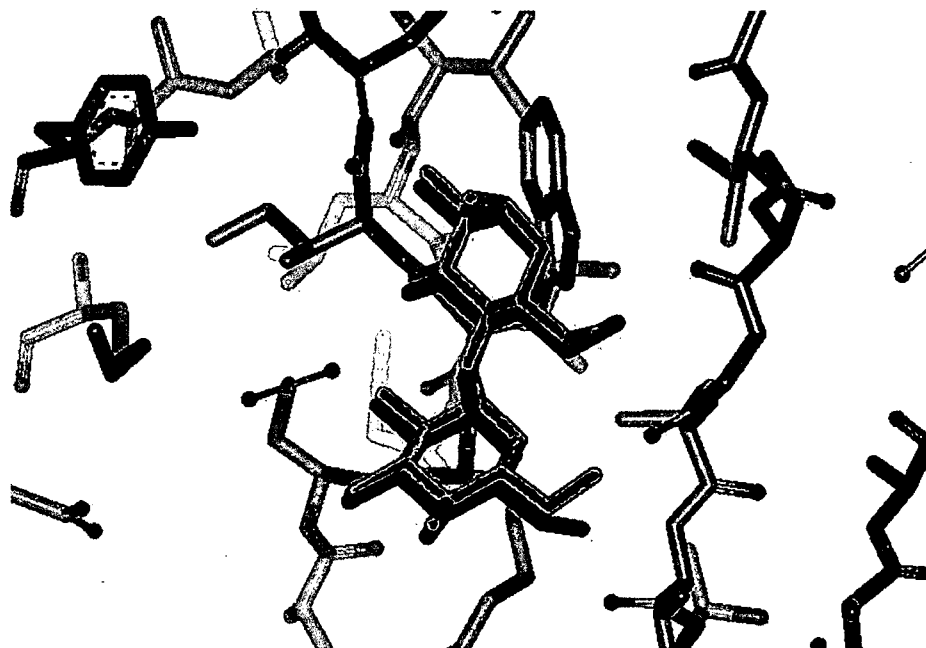


:リガンドの予測構造(中央の緑)

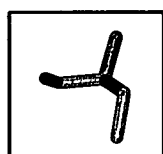


:タンパク質の結合部位(周辺部分)

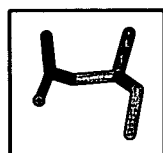
第38図



:リガンドの正解構造(中央のシアン)

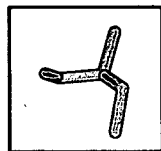
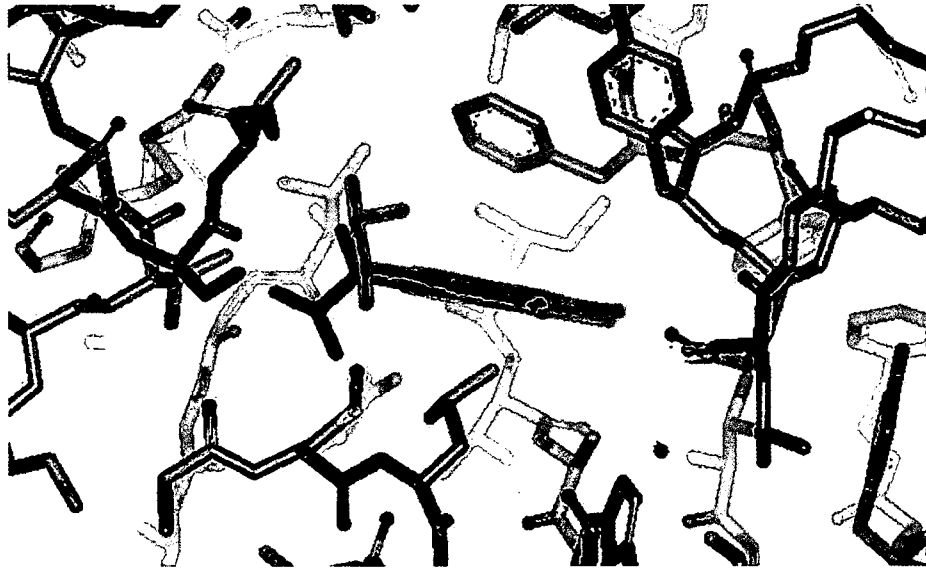


:リガンドの予測構造(中央の緑)

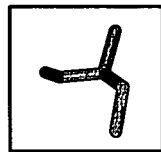


:タンパク質の結合部位(周辺部分)

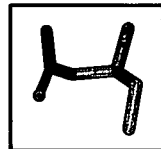
第39図



:リガンドの正解構造(中央のシアン)

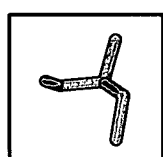
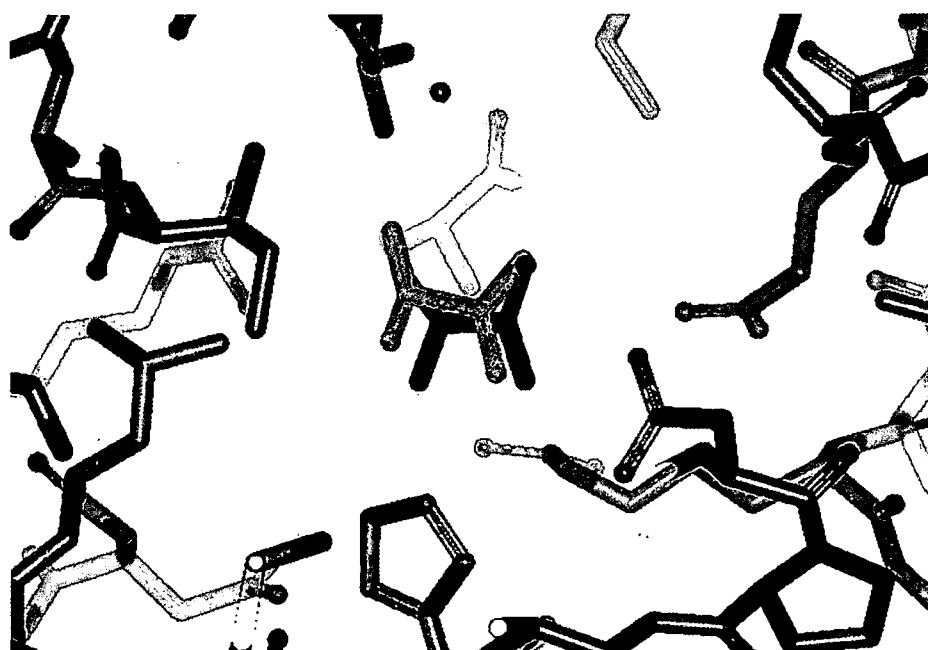


:リガンドの予測構造(中央の緑)

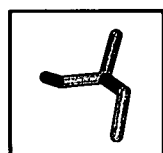


:タンパク質の結合部位(周辺部分)

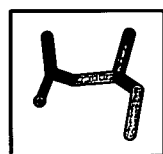
第40図



:リガンドの正解構造(中央のシアン)

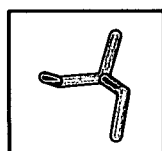
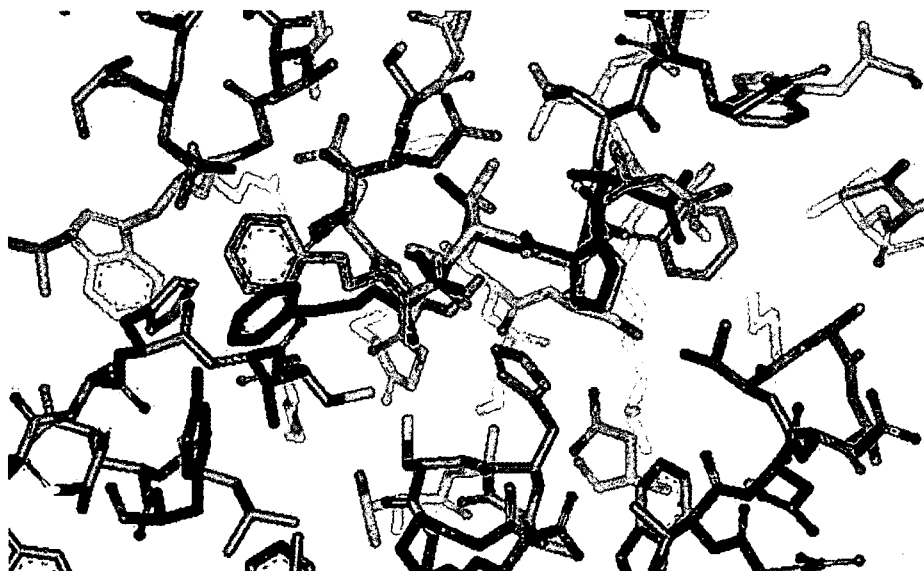


:リガンドの予測構造(中央の緑)

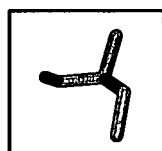


:タンパク質の結合部位(周辺部分)

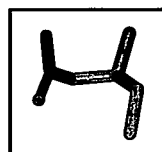
第41図



:リガンドの正解構造(中央のシアン)



:リガンドの予測構造(中央の緑)



:タンパク質の結合部位(周辺部分)

第42図

C	予測成功率(%)
ChooseLD 0.96-0.08	62.2
ChooseLD 0.76-0.08	61.1
ChooseLD 0.56-0.08	60.0
Glide	72.2
GOLD	66.7

第43図

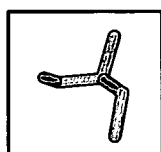
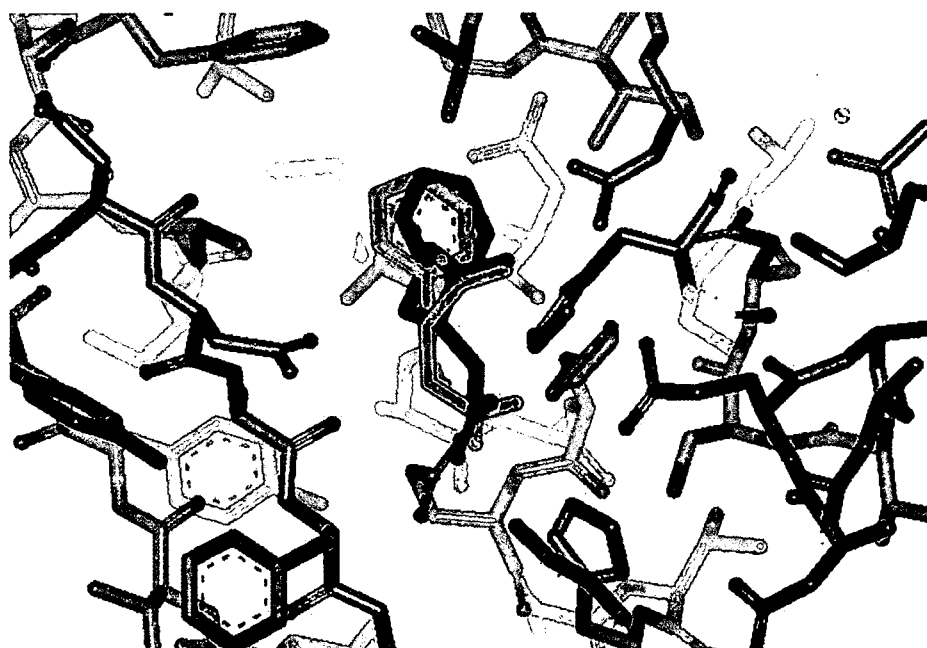
	Glide	GOLD	FlexX	ChooseLD
Glide	1.00	0.65	0.61	0.55
Gold	0.65	1.00	0.66	0.50
FlexX	0.61	0.66	1.00	0.47
ChooseLD	0.55	0.50	0.47	1.00

第44図

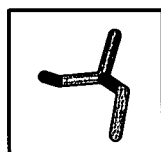
	Glide 成功	Glide 失敗
ChooseLD 成功	42	12
ChooseLD 失敗	23	13

	GOLD 成功	GOLD 失敗
ChooseLD 成功	38	16
ChooseLD 失敗	23	13

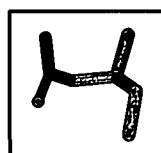
第45図



:リガンドの正解構造(中央のシアン)

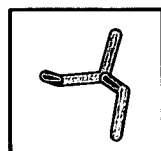
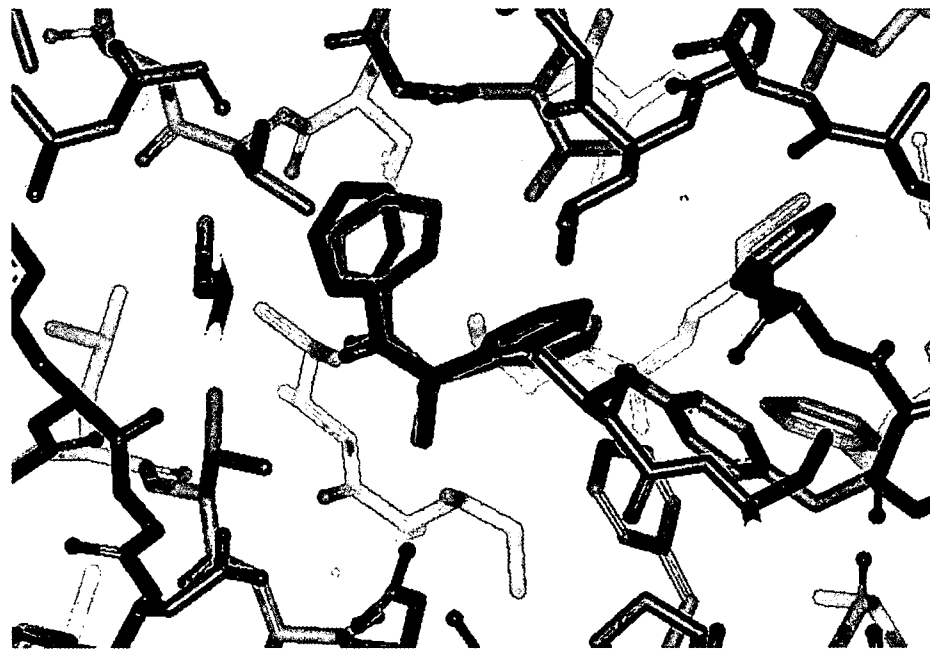


:リガンドの予測構造(中央の緑)

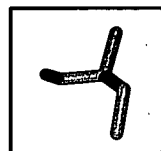


:タンパク質の結合部位(周辺部分)

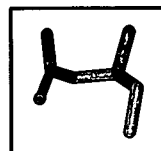
第46図



:リガンドの正解構造(中央のシアン)

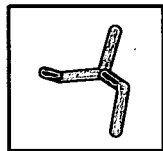
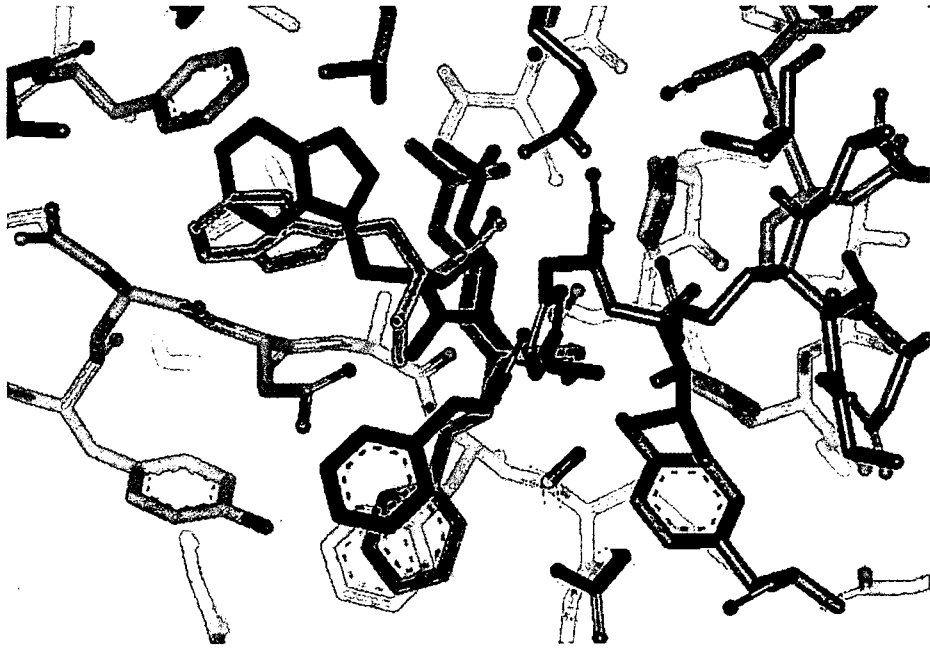


:リガンドの予測構造(中央の緑)

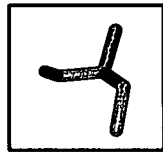


:タンパク質の結合部位(周辺部分)

第47図



:リガンドの正解構造(中央のシアン)

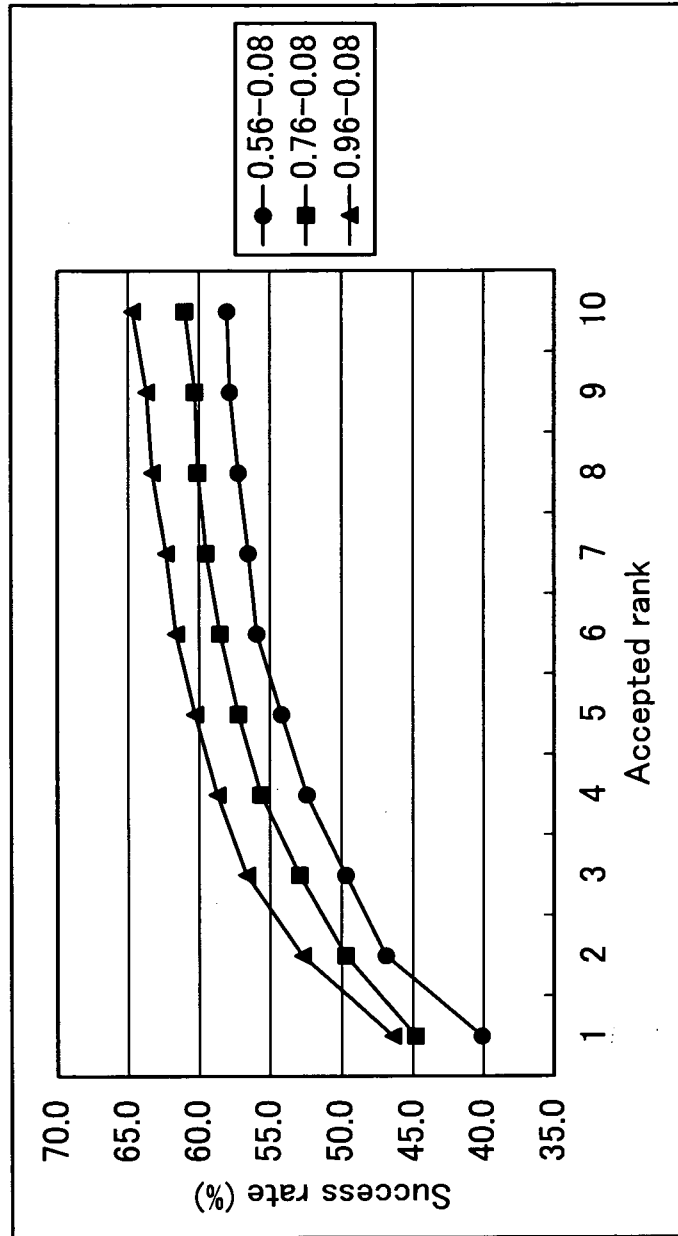


:リガンドの予測構造(中央の緑)



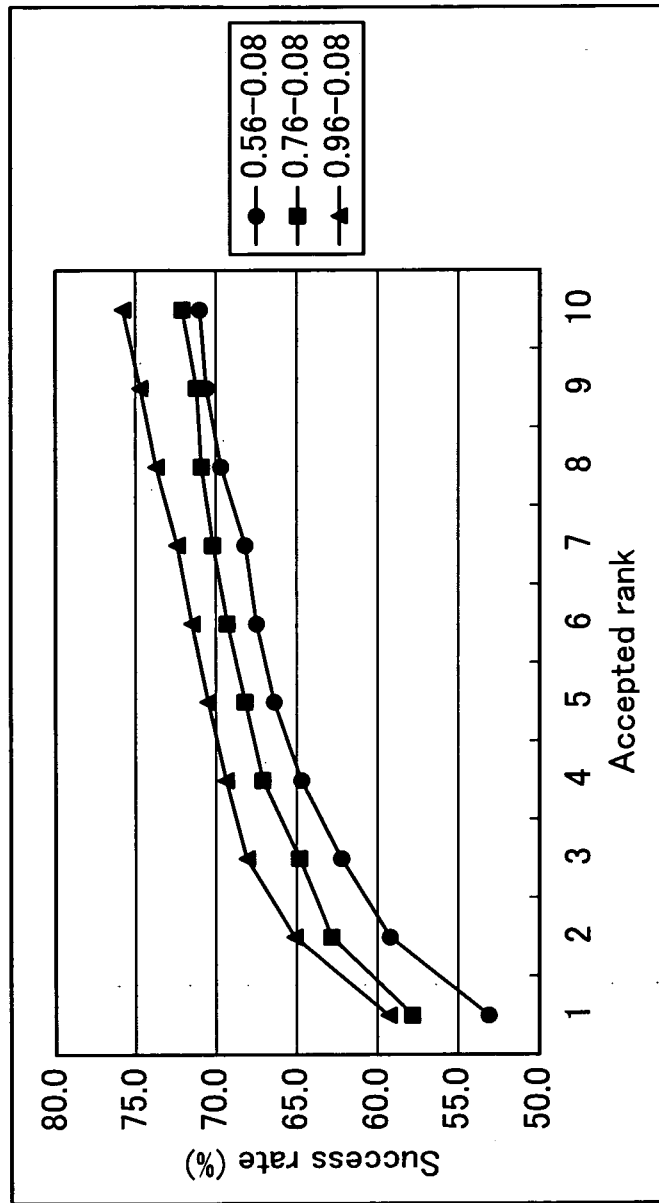
:タンパク質の結合部位(周辺部分)

第48図



	Tc Range		
	0.56-0.08	0.76-0.08	0.96-0.08
1	40.1	44.8	46.4
2	46.9	49.7	52.7
3	49.7	52.9	56.6
4	52.4	55.6	58.7
5	54.2	57.2	60.3
6	55.9	58.5	61.6
7	56.5	59.5	62.3
8	57.2	60.1	63.3
9	57.8	60.3	63.7
10	58.0	61.0	64.7

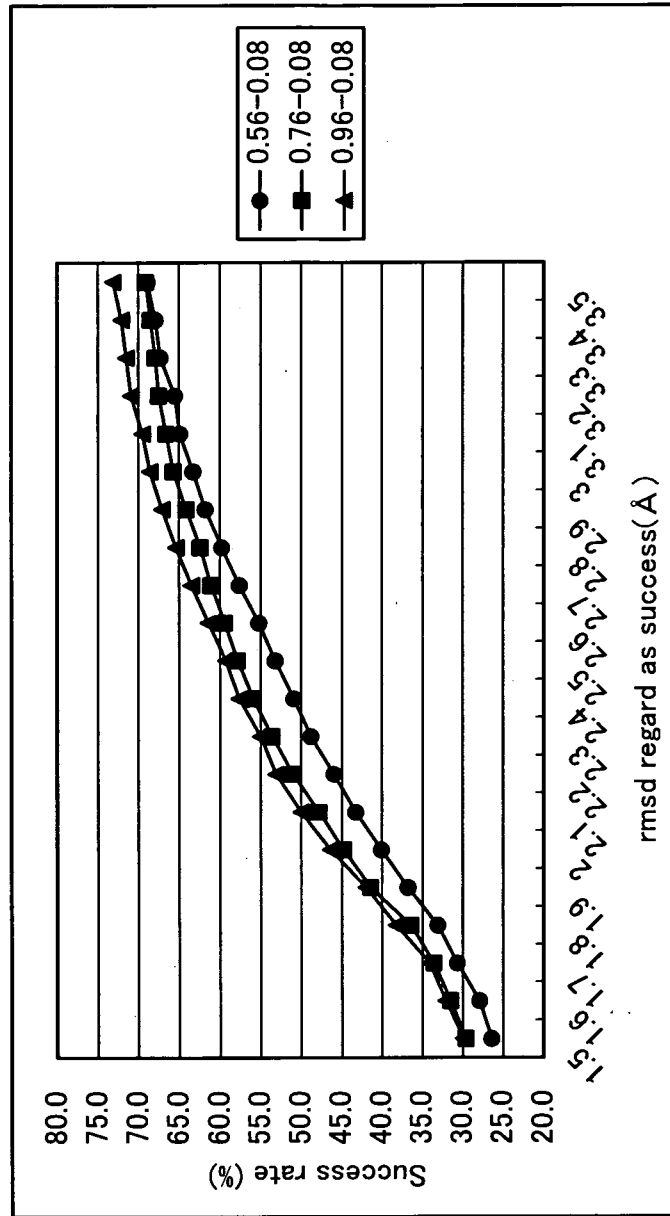
第49図



	Tc Range		
	0.56-0.08	0.76-0.08	0.96-0.08
1	53.1	57.8	59.3
2	59.2	62.8	65.1
3	62.2	64.8	68.1
4	64.7	67.1	69.4
5	66.4	68.2	70.5
6	67.5	69.3	71.5
7	68.2	70.2	72.4
8	69.7	70.9	73.7
9	70.6	71.2	74.7
10	71.0	72.1	75.8

第50図

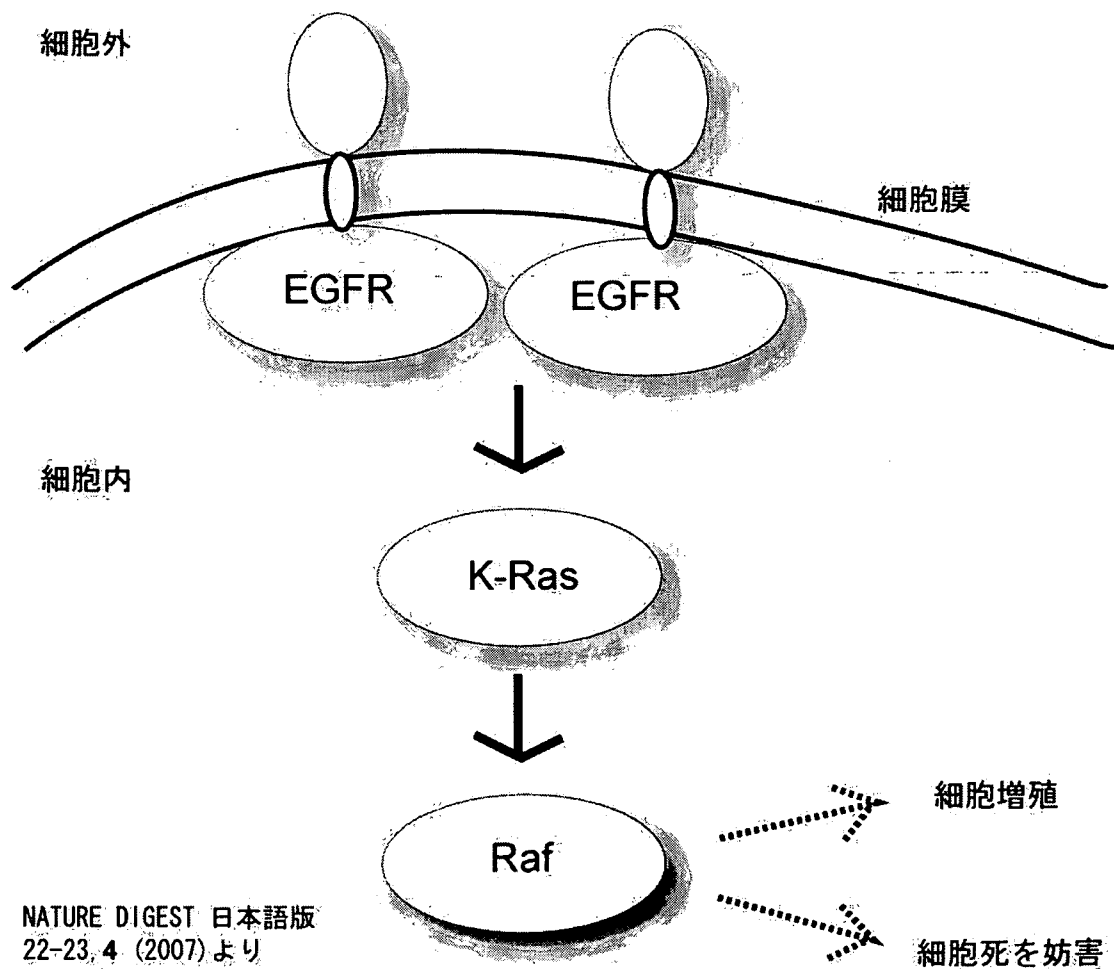
	Tc Range		
	0.56-0.08	0.76-0.08	0.96-0.08
1.5	26.4	29.6	29.9
1.6	27.9	31.5	32.1
1.7	30.7	33.6	34.1
1.8	33.1	36.4	38.2
1.9	36.8	41.4	42.0
2.0	40.1	44.8	46.4
2.1	43.3	47.8	50.0
2.2	46.0	51.0	53.1
2.3	48.8	53.6	55.0
2.4	50.9	55.9	57.6
2.5	53.2	57.8	59.3
2.6	55.2	59.4	61.4
2.7	57.6	61.1	63.5
2.8	59.8	62.4	65.4
2.9	61.8	64.1	67.1
3.0	63.3	65.7	68.6
3.1	64.9	66.6	69.5
3.2	65.5	67.5	70.9
3.3	67.3	67.9	71.5
3.4	67.9	68.5	72.1
3.5	68.9	69.1	73.1



第51図

		rmsd 2.0(Good)	rmsd 2.5(Close)
Druglike-ligand vs Druggable-protein(85 set)	Tc Range 0.56-0.08	58.9	68.6
	Tc Range 0.76-0.08	62.1	72.1
	Tc Range 0.56-0.08	65.2	72.4
Various-ligand vs Druggable-protein(133 set)	Tc Range 0.56-0.08	40.1	53.2
	Tc Range 0.76-0.08	44.8	57.8
	Tc Range 0.56-0.08	46.4	59.3

第52図



第53図

MULTIPLE MATCHING Version in2mlmt

>EGFR_HUMAN696		HOMOLOGY	MATCH	MISMATCH	INSERTION	DELETION
>1M17_A	312	98.7%	308	4	0	12

GEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPVAKELREATSPKANKEILDEAYVMASVDNPHVC
 GEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPVAKELREATSPKANKEILDEAYVMASVDNPHVC

 cccc111111cccc222222cccccc222222cccccc222222cccc111111111111cccccc22
 ccccccccccc4444222222c4444222222c4444ccc222222cccccc111111111111cccc44cc

RLLGI CLTSTVQLITQLMPFGCLLDYVREHKDNI GSQYLLNWCVQI AKGMNYLEDRLVHRDLAARNVLVKTPQHVKITD
 RLLGI CLTSTVQLITQLMPFGCLLDYVREHKDNI GSQYLLNWCVQI AKGMNYLEDRLVHRDLAARNVLVKTPQHVKITD

 222222cccc222222cccccc111111cccccc11111111111111111111ccc1111111122cccc2222c
 c22222244422222c44422111111c111cc11111111111111111111c224444c1112222442222cc

FGLAKLLGAEKEYHAEGGKVPKWMALSI LHR IYTHQSDVWSYGVTWELMTFGSKPYDGI PASE ISS ILEKGERLPQ
 FGLAKLLGAEKEYHAEGGKVPKWMALSI LHR IYTHQSDVWSYGVTWELMTFGSKPYDGI PASE ISS ILEKGERLPQ

 cc11cccccccc22ccccccc1cc111111cccc1111111111111111cccccccccc11111111cccccc
 444422c4444cc22cccccc4444c111111c22c11111111111111114444444444411111111cccccc

PPICTIDVYIMVKCWMIDADSRPKFRELIEFSKMARDPQRYLVIQGDERMHLPSPDTSNFYRALMDEEDMDDVDDADE
 PPICTIDVYIMVKCWMIDADSRPKFRELIEFSKMARDPQRYLVIQGDERMHLMDEE-----DMDDVDDADE

 cccc111111111111cc111ccc111111111111c11111cccccccccccccccccccc1111cccccc
 444cc1111111111c444111ccc111111111111c111ccc444444ccccccc..... c4444cc111

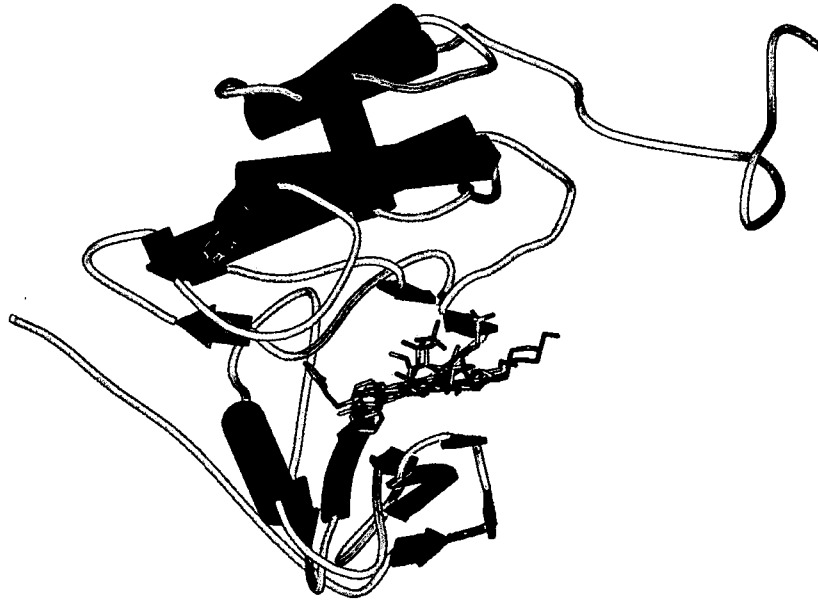
YLIP
 YLIP

 cccc
 cccc

//

* 一致したアミノ酸残基 + 一致しないアミノ酸残基

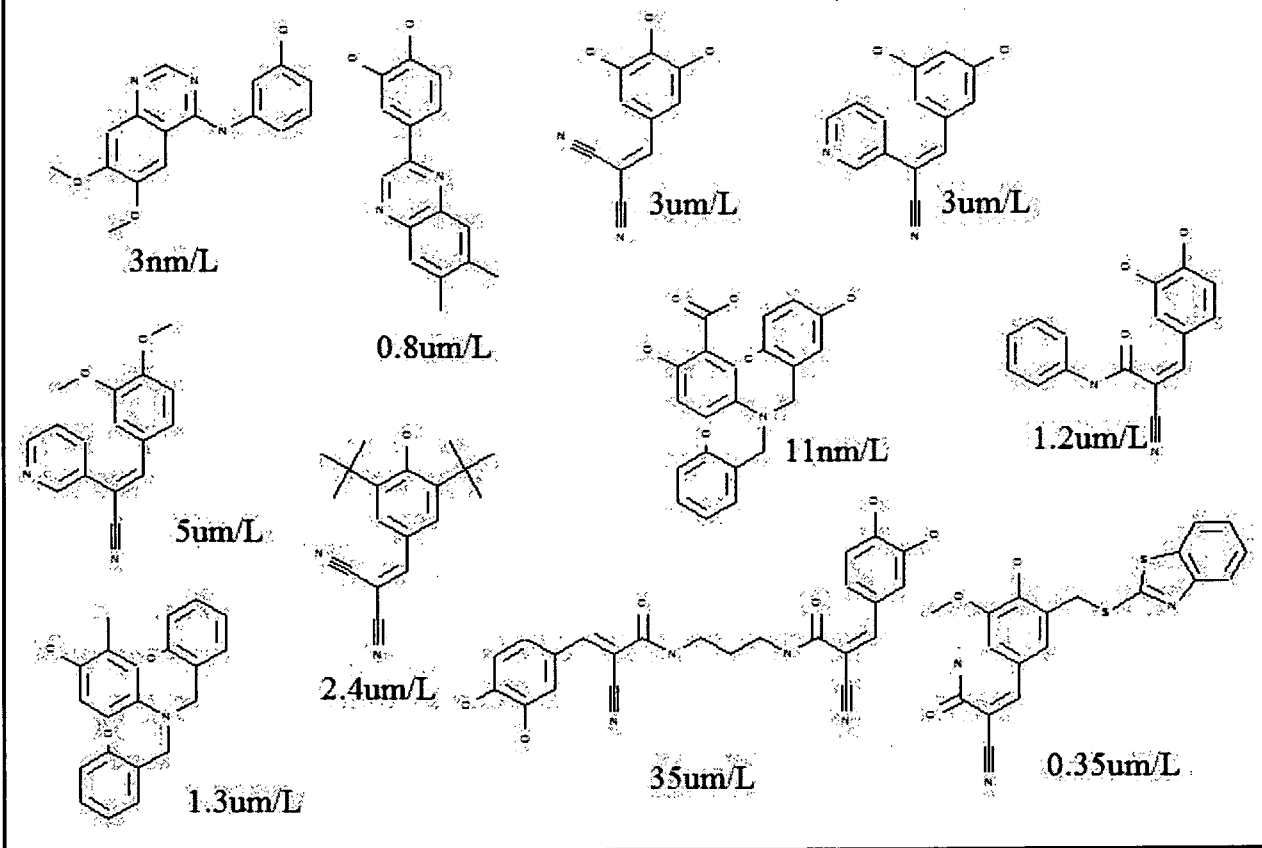
第54図



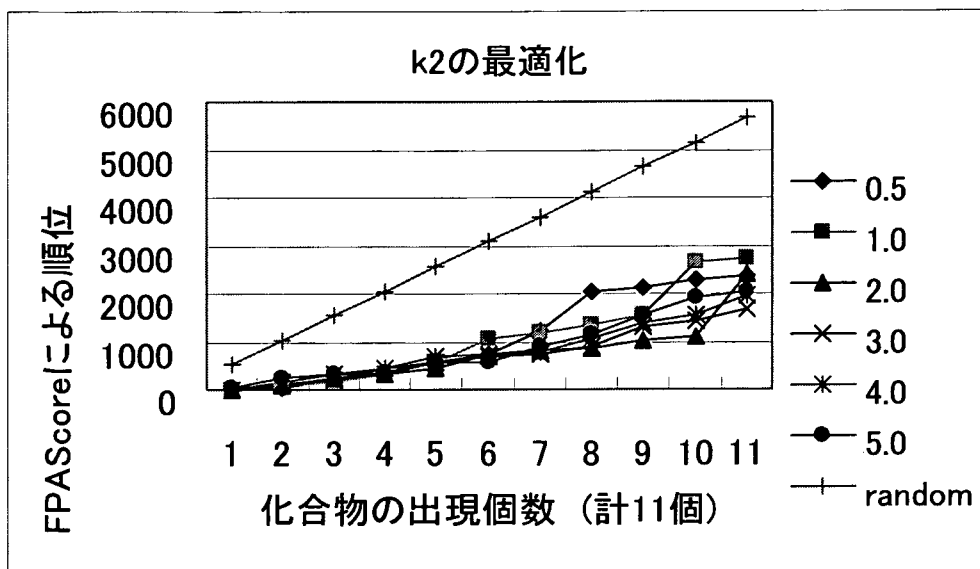
EGFRのモデル; 鑄型として1M17を使用

第55図

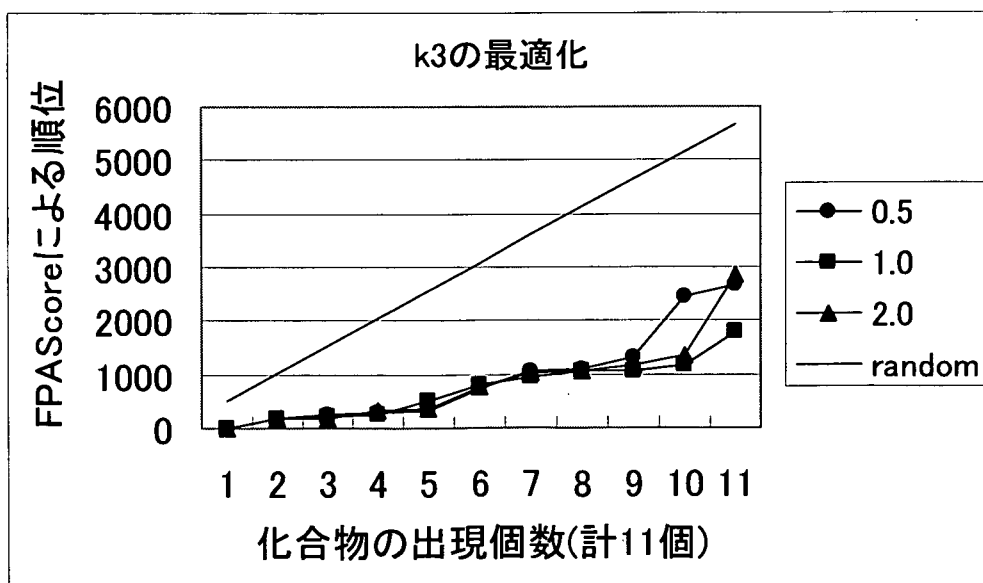
IC50既知化合物一覽(EGFR inhibitor 11個)



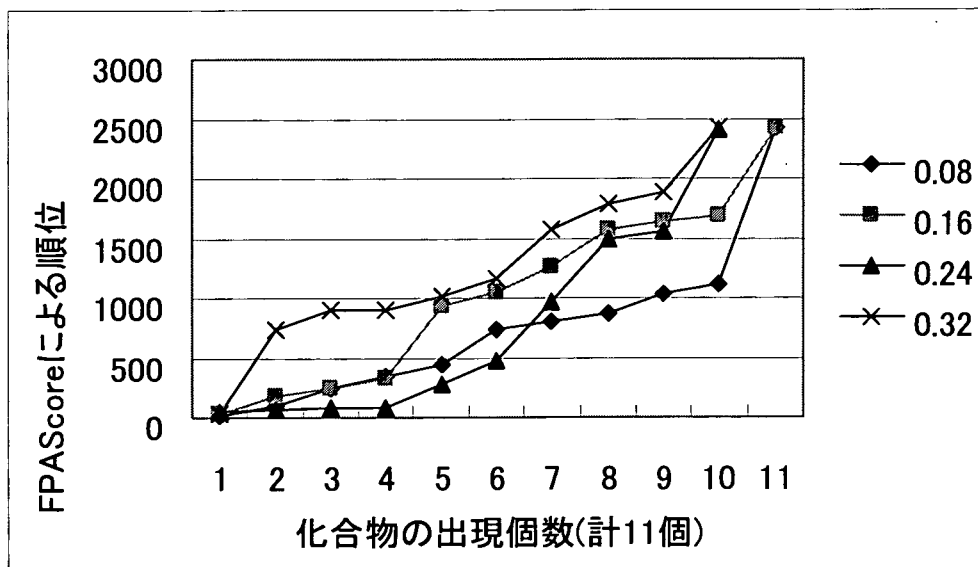
第56図



第57図



第58図



第59図

順位	FPAScore	PDBID とリガンド ID	順位	FPAScore	PDBID とリガンド ID
1	2834.83	2ITQ;STU	6	3733.392	2ITQ;STU
2	2430.896	2J5F;DJK	17	2846.782	2GS7;ANP
3	2414.404	2ITO;IRE	23	2671.382	1XKK;FMM
4	2178.805	2GS7;ANP;	29	2484.615	2J5F;DJK
5	2133.221	1M17;AQ4	42	2363.591	2ITO;IRE
8	1986.738	1XKK;FMM	48	2308.081	2ITP;AEE
81	1407.896	2ITP;AEE	105	2075.384	1M17;AQ4

K1=4.0, k2=2.0, k3=1.0 Tc 上限値 1.0 Tc 下限値 0.08

K1=4.0, k2=2.0, k3=1.0 Tc 上限値 1.0 Tc 下限値 0.24

第60図

リガンド ID 対応する化合物名

AEE AEE788

ANP PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER

AQ4 ERLOTINIB

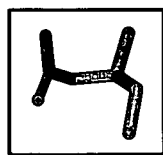
DJK N-[4-(3-BROMO-PHENYLAMINO)-QUINAZOLIN-6-YL]- ACRYLAMIDE

FMM Lapatinib

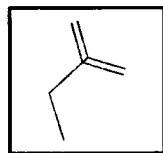
IRE IRESSA

STU STAUROSPORINE

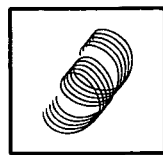
第61図



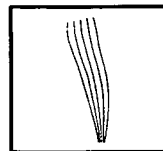
:リガンド(中央付近)



:タンパク質の結合部位(周辺部分)

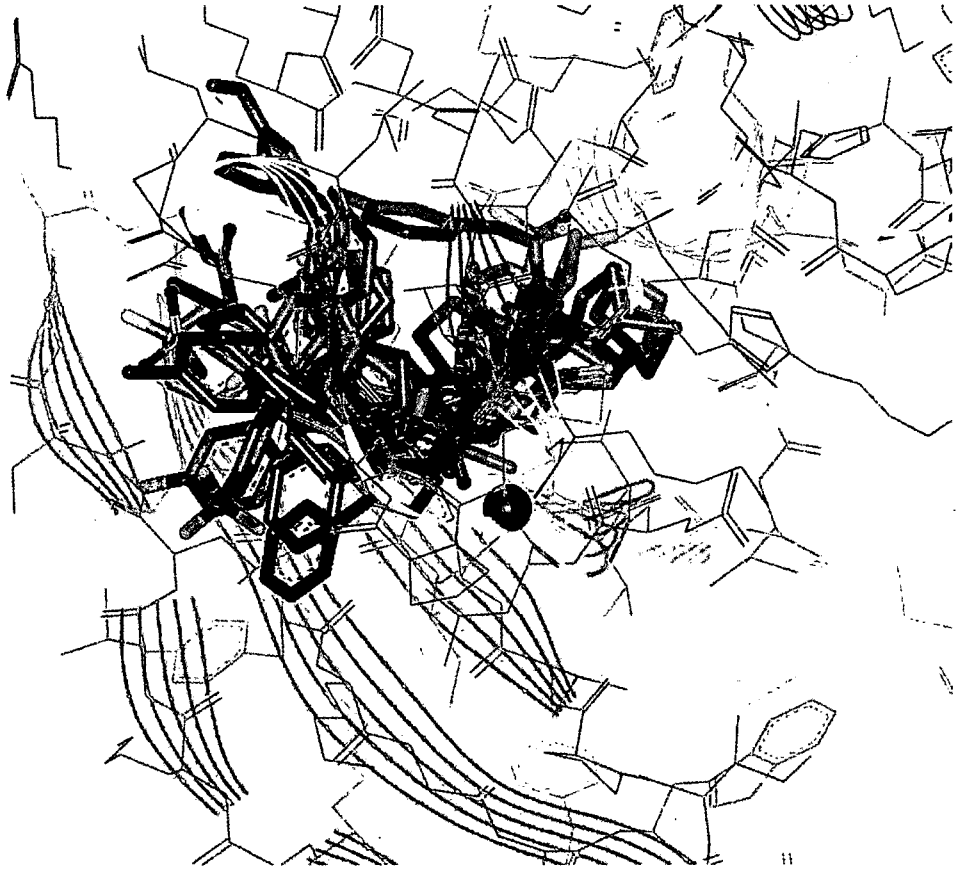


: α ヘリックス(赤)

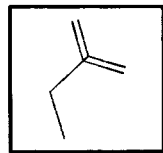


: β シート(シアン)

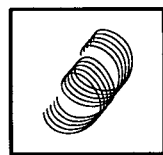
第62図



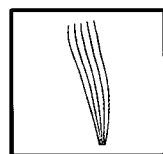
: リガンド(中央付近)



: タンパク質の結合部位(周辺部分)

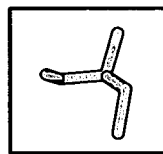
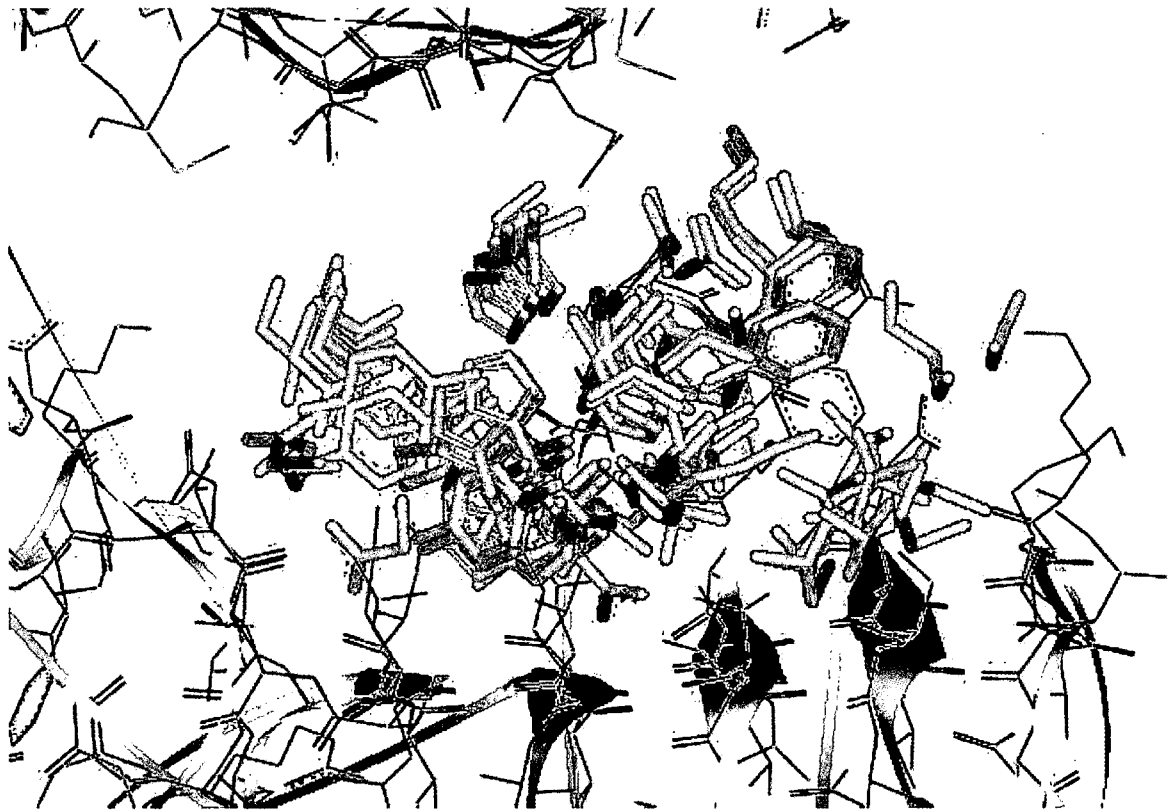


: α ヘリックス(赤)

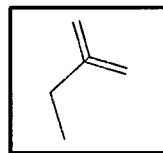


: β シート(シアン)

第63図



: TGF- α 類似体側鎖(中央の黄)

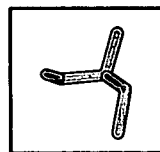
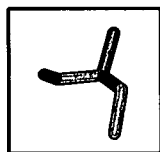
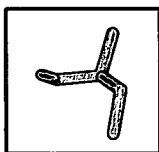
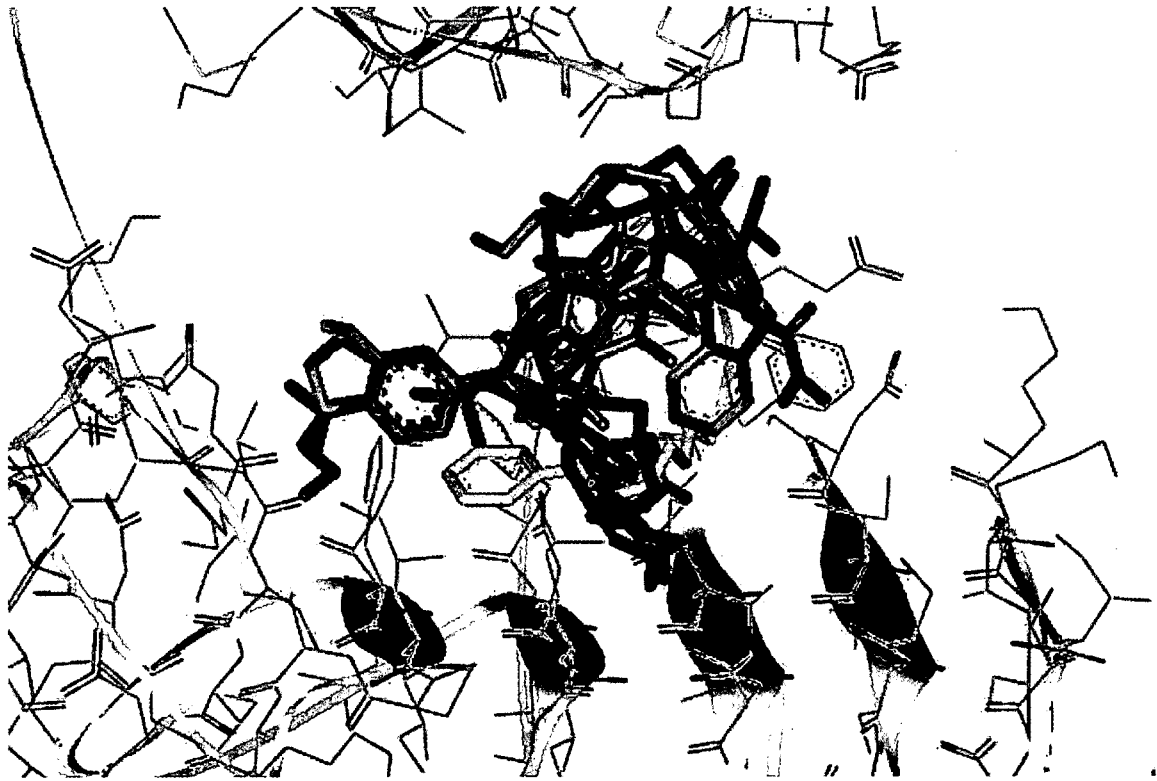


: EGFR結合部位(周辺部分)

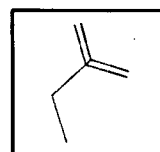
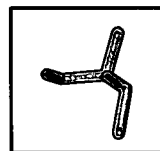
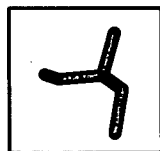
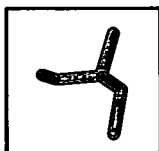


: β シート(シアン)

第64図



:リガンド(中央付近)

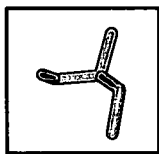
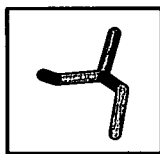
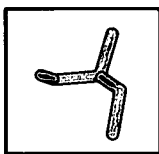
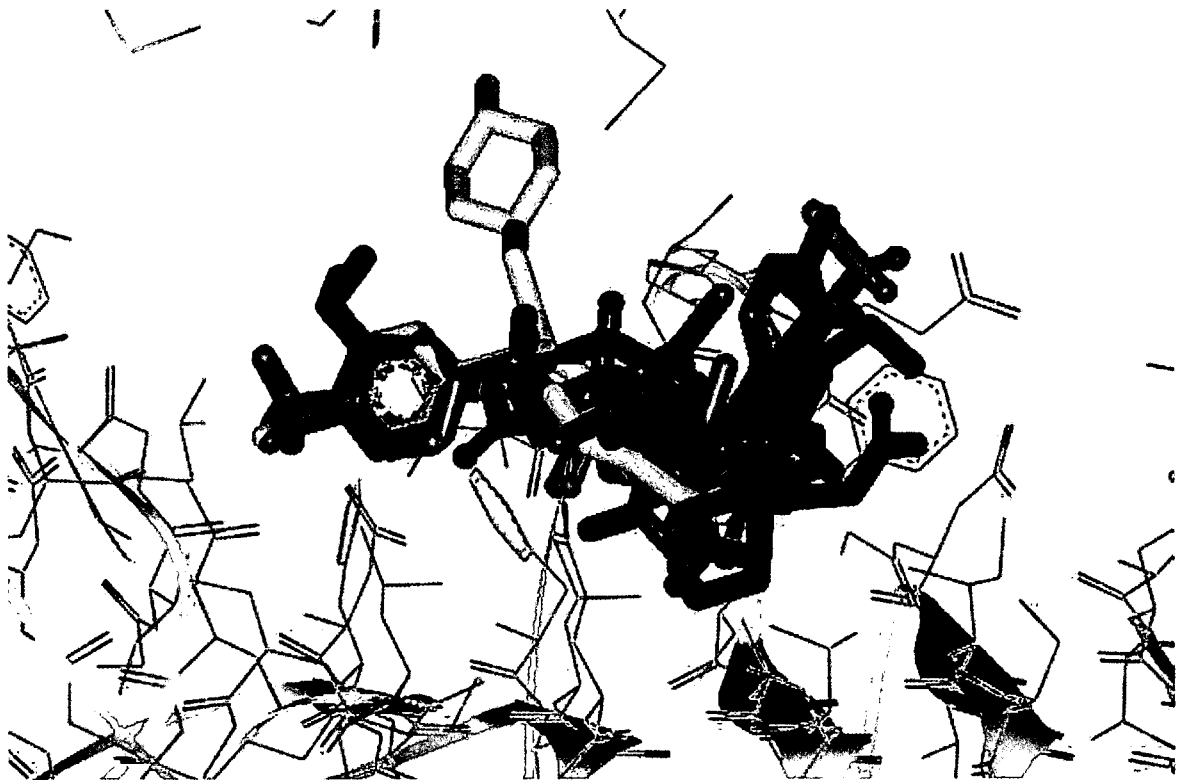


:タンパク質の結合部位(周辺部分)

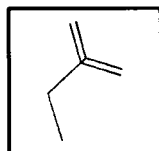
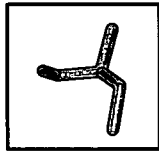
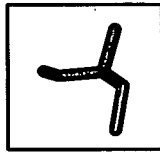
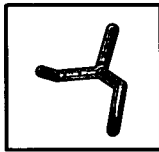


:βシート(シアン)

第65図



:リガンド(中央付近)

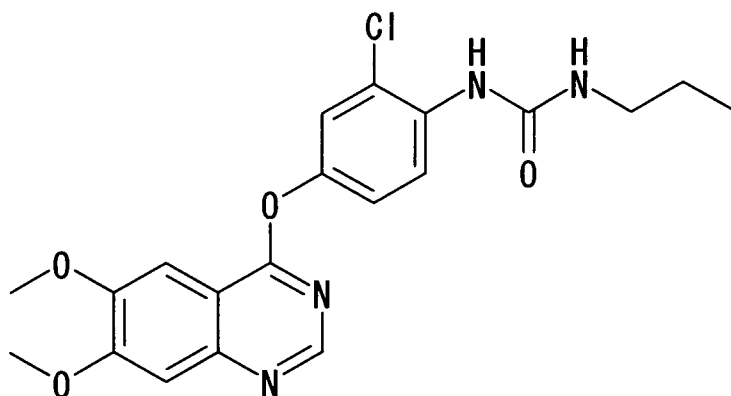


:タンパク質の結合部位(周辺部分)

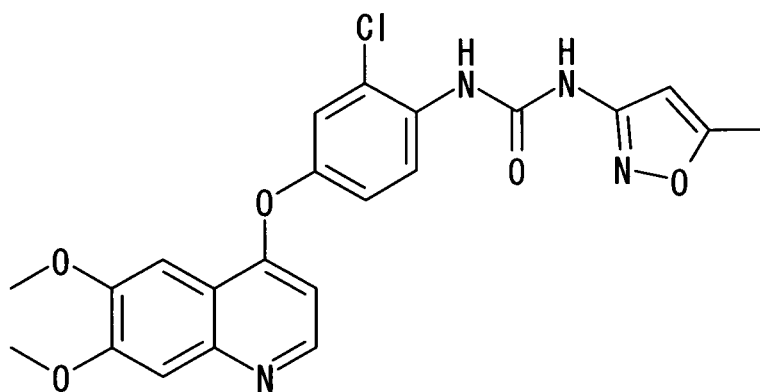


:βシート(シアン)

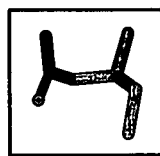
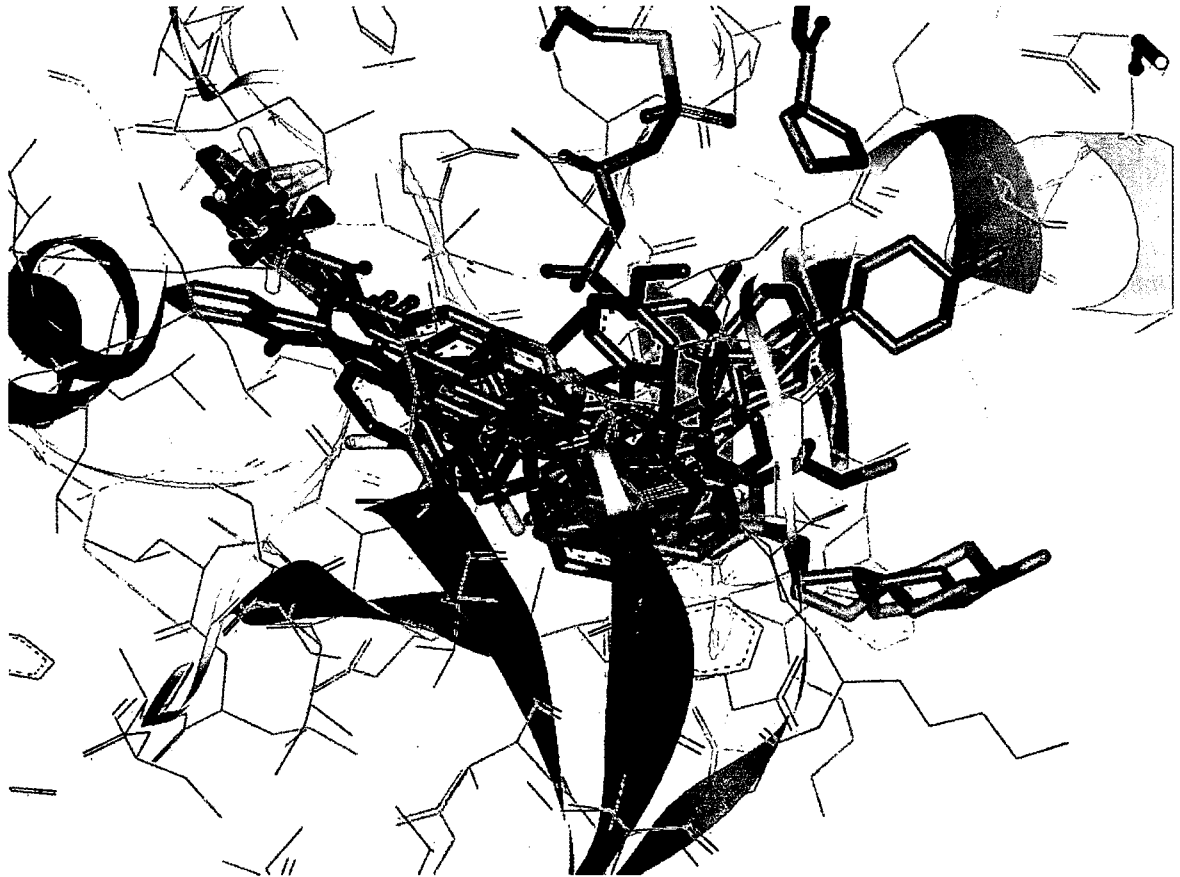
第66図



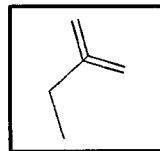
第67図



第68図



: リガンド(中央付近)



: タンパク質の結合部位(周辺部分)



: α ヘリックス(赤)

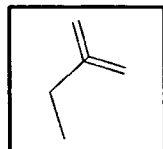


: β シート(シアン)

第69図



: リガンド(中央付近)



: タンパク質の結合部位(周辺部分)



: α ヘリックス(赤)

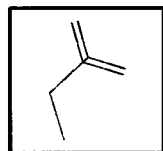


: β シート(シアン)

第70図



: リガンド(中央付近)



: タンパク質の結合部位(周辺部分)

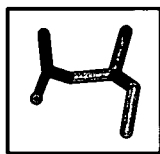


: α ヘリックス(赤)

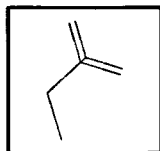


: β シート(シアン)

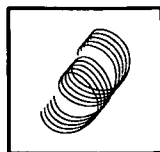
第71図



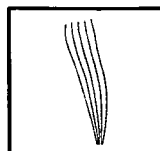
:リガンド(中央付近)



:タンパク質の結合部位(周辺部分)



: α ヘリックス(赤)



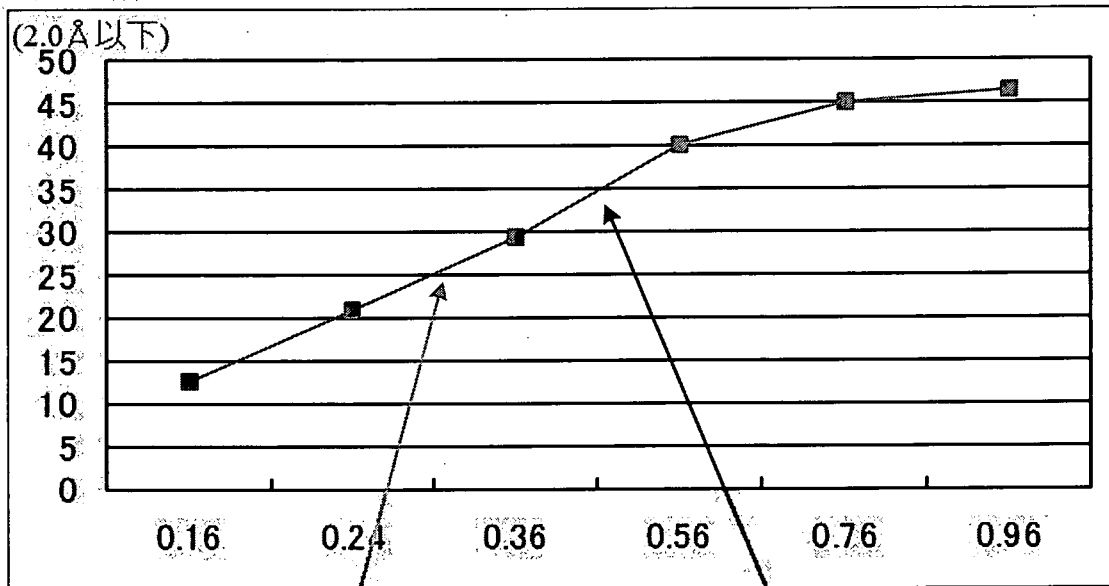
: β シート(シアン)

第72図

VEGFR-2のドッキング予測成功率の算出

T133から学習したターゲットによらない統計的な成功率

ドッキング予測成功率

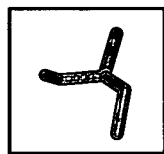
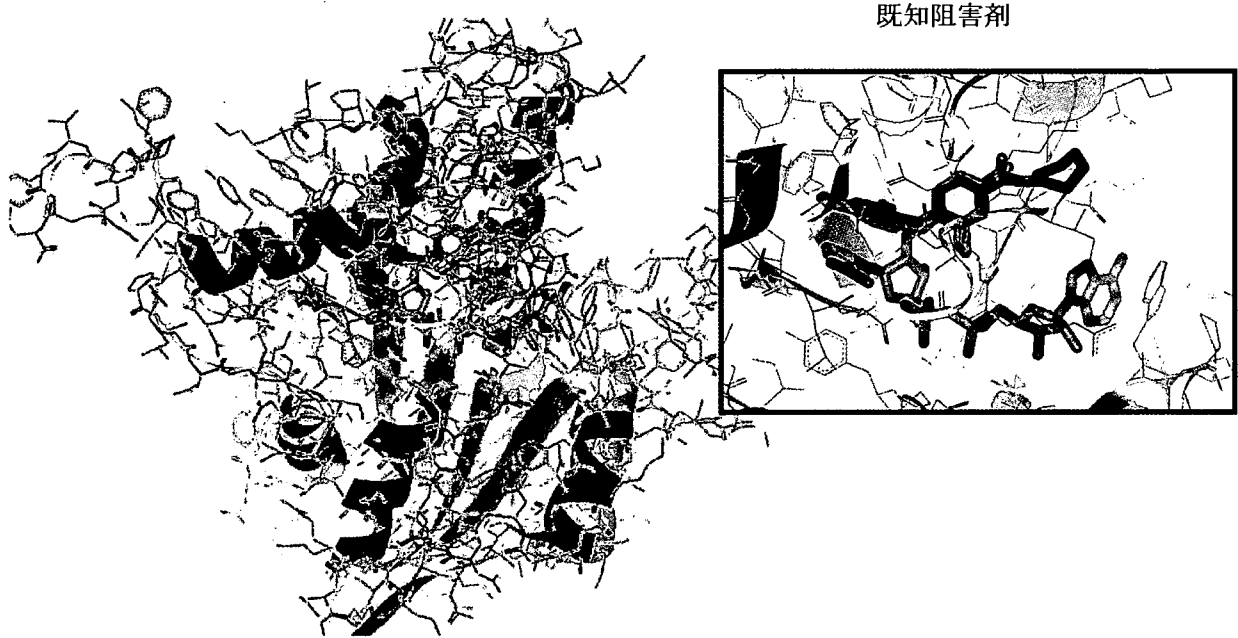


Tc 上限値

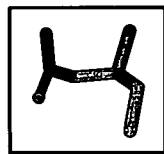
VEGFR-2 KRN633 (FP Library Tc 最大値:0.45[横軸] 予測成功率 34.7%[縦軸])

VEGFR-2 KRN951 (FP Library Tc 最大値:0.29[横軸] 予測成功率 24.3%[縦軸])

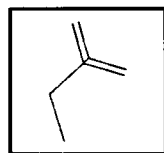
第73図



: 阻害剤(中央上側)



: NAD(中央下側)



: タンパク質の結合部位(周辺部分)

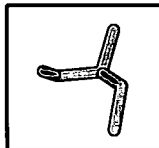
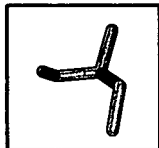
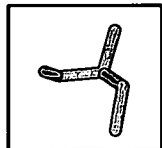


: α ヘリックス(赤)

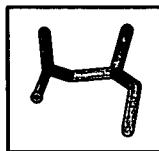
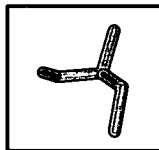
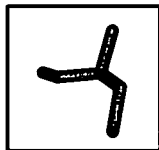
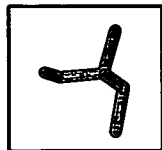


: β シート(シアン)

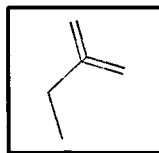
第74図



: リガンド(中央上側)



: NAD(中央下側)



: タンパク質の結合部位(周辺部分)



: β シート(シアン)

第75図

>AMPKhomoGAMMA1	304	homology	match	mismatch	insertion	deletion
>2V9J_E	304	98.7%	300	4	0	0

SNNSVYTSFMKSHRCYDLIPTSSKLVVFDTS LQVKKAFFALVTNGVRAAPLWDSKKQSFVGMTITDFINILHRYYSAL
 SNSSVYTTFMKSHRCYDLIPTSSKLVVFDTS LQVKKAFFALVTNGVRAAPLWDSKKQSFVGMTITDFINILHRYYSAL

VQIYELEEHK IETWREVYLQDSFKPLVCI SPNASLFDAVSSLIRNKIHRLPVIDPESGNTLYILTHKRILKFLKLFITF
 VQIYELEEHK IETWREVYLQDSFKPLVCI SPNASLFDAVSSLIRNKIHRLPVIDPESGNTLYILTHKRILKFLKLFITF

PKPEFMSKSLEELQI GTYANI AMVRTTTPVYVALGIFVQHRVSALPVVDEKGRVVDIYSKFDVINLAAEKTYNLDVSVT
 PKPEFMSKSLEELQI GTYANI AMVRTTTPVYVALGIFVQHRVSALPVVDEKGRVVDIYSKFDVINLAAEKTYNLDVSVT

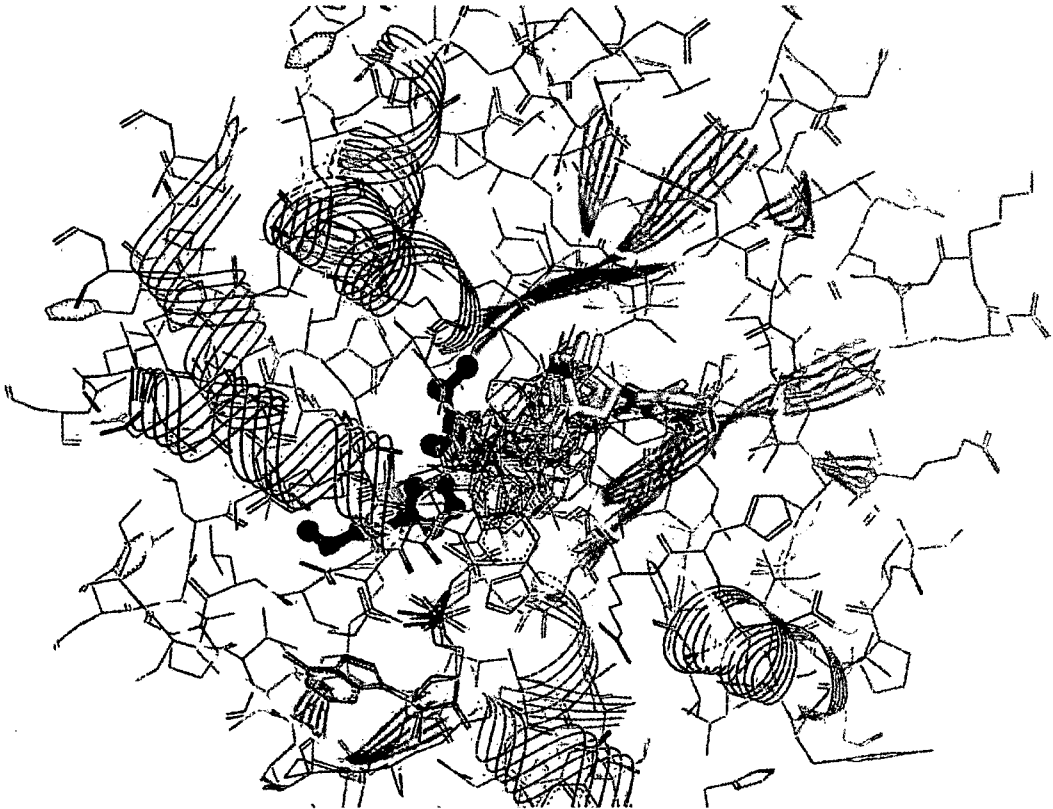
KALQHRSHYFEGVLKCYLHETLETI INRLVEAEVHRLVVVDENDVVKGIVSLSDILQALVLTGG
 KALQHRSHYFEGVLKCYLHETLEAI INRLVEAEVHRLVVVDEHDVVKGIVSLSDILQALVLTGG

第76図

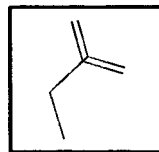
AMPKhomoGAMMA1のCMCスクリーニング結果

Rank	chk	Docked ligand	original ligand	FPAS core	colli	name	png	formula	Rings	M.W.	logP	Acc	Dnr	tky L.J.
1(14)	<input type="checkbox"/>	CMC2006v1 00004868(1)	-1	901.267	1,2,4	EPIROPRI M	png	C ₁₉ H ₂₃ N ₅ O ₂	3	330.24	3.98	5	4	-29.31
2(137)	<input type="checkbox"/>	CMC2006v1 00005124(1)	-1	724.98	1,2,2	ROSIGLIT AZONE	png	C ₁₈ H ₁₉ N ₃ O ₃	3	338.28	2.82	4	4	-7.75
3(199)	<input type="checkbox"/>	CMC2006v1 00007786(1)	-1	693.542	0,1,2	BUTANIXI N	png	C ₁₆ H ₁₇ N ₂ O ₂	2	252.19	2.6	1	3	-2.7
4(247)	<input type="checkbox"/>	CMC2006v1 00006998(1)	-1	675.763	0,1,5	PROPIOM AZINE	png	C ₂₀ H ₂₅ N ₂	3	316.3	3.48	3	1	-48.82
5(260)	<input type="checkbox"/>	CMC2006v1 00006302(1)	-1	672.265	0,1,2	BUTANSE RIN	png	C ₂₄ H ₂₆ FN ₃ O ₃	4	397.28	3.14	3	4	-36.29
6(389)	<input type="checkbox"/>	CMC2006v1 00005803(1)	-1	632.201	0,0,1	ANISOPIR OL	png	C ₂₁ H ₂₈ FN ₂ O ₂	3	331.24	3.69	3	3	-31.39
7(414)	<input type="checkbox"/>	CMC2006v1 00005171(1)	-1	624.97	0,2,5	AZAPERO NE	png	C ₁₉ H ₂₃ FN ₃	3	305.23	3.22	2	3	-36.02
8(451)	<input type="checkbox"/>	CMC2006v1 00002706(1)	-1	618.414	0,2,4	PIPRATEC OL	png	C ₁₉ H ₂₅ N ₂ O ₄	3	320.22	2.18	3	4	-36.03
9(469)	<input type="checkbox"/>	CMC2006v1 00005967(1)	-1	615.724	0,2,2	BUTROPIP AZONE	png	C ₂₀ H ₂₄ FN ₂	3	303.23	3.83	2	2	-21.13
10(671)	<input type="checkbox"/>	CMC2006v1 00005328(1)	-1	582.259	0,2,5	CARPHEN AZINE	png	C ₂₄ H ₃₃ N ₃ O ₂	4	394.35	4.25	4	2	-15.16
11(756)	<input type="checkbox"/>	CMC2006v1 00006323(1)	-1	567.542	0,1,2	MEFECLO RAZINE	png	C ₂₀ H ₂₆ CIN ₂ O ₂	3	335.69	3.94	4	3	-35.22
12(836)	<input type="checkbox"/>	CMC2006v1 00006505(1)	-1	557.475	1,1,1	MCMC000 10966	png	C ₂₄ H ₂₃ N ₃ O ₂	5	362.28	4.11	4	2	-38.47
13(894)	<input type="checkbox"/>	CMC2006v1 00006361(1)	-1	551.318	1,2,3	MCMC000 11316	png	C ₂₈ H ₃₀ N ₅ O ₃	5	454.34	3.8	5	5	-64.06
14(904)	<input type="checkbox"/>	CMC2006v1 00004094(1)	-1	550.754	1,3,5	LESOPITR ON	png	C ₁₅ H ₂₂ CIN ₆	3	299.66	2.15	4	3	-34.77
15(974)	<input type="checkbox"/>	CMC2006v1 00004397(1)	-1	543.014	1,2,3	TRAZODO NE	png	C ₁₉ H ₂₃ CIN ₅	4	349.7	2.58	5	2	-19.05
16(1258)	<input type="checkbox"/>	CMC2006v1 00008285(1)	-1	512.867	0,2,5	LOFEPRA MINE	png	C ₂₆ H ₂₈ CIN ₂	4	391.75	4.43	2	2	-30.95
17(1426)	<input type="checkbox"/>	CMC2006v1 00004887(1)	-1	499.07	0,0,0	KETANSE RIN	png	C ₂₂ H ₂₂ FN ₃ O ₃	4	373.26	2.36	3	4	-11.87
18(1519)	<input type="checkbox"/>	CMC2006v1 00004722(1)	-1	491.994	0,0,1	SPIPERON E	png	C ₂₃ H ₂₇ FN ₃ O ₂	4	369.27	3.76	3	3	-16.87
19(2047)	<input type="checkbox"/>	CMC2006v1 00006678(1)	-1	453.939	1,2,3	AMIPERON E	png	C ₂₄ H ₂₉ CIFN ₂ O ₂	3	402.73	4.72	2	4	-16.26
20(2209)	<input type="checkbox"/>	CMC2006v1 00002983(1)	-1	444.529	1,2,4	ECABAPID E	png	C ₂₀ H ₂₅ N ₃ O ₄	2	346.24	2.69	5	4	-21.95
21(2710)	<input type="checkbox"/>	CMC2006v1 00006571(1)	-1	417.289	1,3,5	PROSTAG LANDIN	png	C ₂₀ H ₃₅ O ₅	1	322.23	1.93	3	5	-29.73
22(3012)	<input type="checkbox"/>	CMC2006v1 00004756(1)	-1	402.001	1,1,2	2-AMINO- 4H-3,1- BENZOXA ZINONES	png	C ₁₄ H ₈ ClIN ₂ O ₂	3	390.52	4.26	3	3	-29.57
23(3069)	<input type="checkbox"/>	CMC2006v1 00005798(1)	-1	398.616	0,3,5	PROGEST ERONE	png	C ₂₁ H ₃₀ O ₂	4	287.25	4.72	0	2	-5.42
24(3577)	<input type="checkbox"/>	CMC2006v1 00006639(1)	-1	368.772	0,1,4	PROTIZINI C ACID	png	C ₁₇ H ₁₆ NO ₃	3	298.26	2.85	3	3	-10.23
25(3787)	<input type="checkbox"/>	CMC2006v1 00005686(1)	-1	356.823	0,1,2	TRIMEGES TONE	png	C ₂₂ H ₃₀ O ₃	4	315.26	4.15	0	3	-10.59
26(3915)	<input type="checkbox"/>	CMC2006v1 00004045(1)	-1	350.149	0,2,5	DEPRODO NE PROPION	png	C ₂₄ H ₃₂ O ₅	4	371.28	3.55	1	5	268.09
27(4885)	<input type="checkbox"/>	CMC2006v1 00003905(1)	-1	277.556	1,1,2	QUINAZOS IN	png	C ₁₇ H ₂₄ N ₅ O ₂	3	306.22	2.34	5	4	-40.89

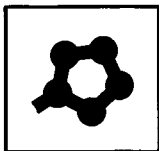
第77図



:リガンド(中央付近)



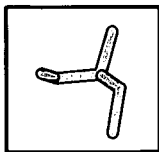
:タンパク質の結合部位
(周辺部分)



:HIS残基(緑)



: α ヘリックス(赤)



:Adenosine等(黄)



: β シート(シアン)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/070973

A. CLASSIFICATION OF SUBJECT MATTER

C40B30/02(2006.01)i, G06F19/00(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, G06F19/00

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 1996-2009

Kokai Jitsuyo Shinan Koho 1971-2009 Toroku Jitsuyo Shinan Koho 1994-2009

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

PubMed, JSTPlus (JDreamII), JMEDPlus (JDreamII)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2002/041184 A1 (Kyorin Pharmaceutical Co., Ltd.), 23 May, 2002 (23.05.02), Full text; all drawings & US 2004/0038429 A1 & AU 2403402 A	1-12
A	JP 2002-530727 A (Glaxo Group Ltd.), 17 September, 2002 (17.09.02), Full text; all drawings & US 2002/0052694 A1 & EP 1153358 A & WO 2000/025106 A2	1-12
P, X	TAKAYA, D., Bioinformatics based Ligand-Docking and in-silico screening, Chemical & pharmaceutical bulletin, 2008.05, Vol.56, No.5, p.742-4	1-12

 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
15 January, 2009 (15.01.09)Date of mailing of the international search report
27 January, 2009 (27.01.09)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/070973

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MPAMHANGA, C.P., Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions, Journal of chemical information and modeling, 2006, Vol.46, No.2, p.686-98	1-12
A	SCIABOLA, S., Pharmacophoric fingerprint method (TOPP) for 3D-QSAR modeling: application to CYP2D6 metabolic stability, Journal of chemical information and modeling, 2007.02, Vol.47, No.1, p.76-84	1-12
A	KELLY, M.D., Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design, Journal of chemical information and computer sciences, 2004, Vol.44, No.6, p.1942-51	1-12

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, G06F19/00(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, G06F19/00

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 1971-2009年
 日本国実用新案登録公報 1996-2009年
 日本国登録実用新案公報 1994-2009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)
 PubMed, JSTPlus(JDreamII), JMEDPlus(JDreamII)

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 請求の範囲の番号
A	WO 2002/041184 A1 (杏林製薬株式会社) 2002.05.23, 全文, 全図 & US 2004/0038429 A1 & AU 2403402 A	1-12
A	JP 2002-530727 A (グラクソ グループ リミテッド) 2002.09.17, 全文, 全図 & US 2002/0052694 A1 & EP 1153358 A & WO 2000/025106 A2	1-12

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。

<p>* 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</p>	<p>の日の後に公表された文献 「T」国際出願日又は優先日後に公表された文献であって出願と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</p>
---	---

国際調査を完了した日 15.01.2009	国際調査報告の発送日 27.01.2009
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 請求の範囲の番号
P, X	TAKAYA, D., Bioinformatics based Ligand-Docking and in-silico screening, Chemical & pharmaceutical bulletin, 2008.05, Vol.56, No.5, p.742-4	1 - 1 2
A	MPAMHANGA, C.P., Knowledge-based interaction fingerprint scoring: a simple method for improving the effectiveness of fast scoring functions, Journal of chemical information and modeling, 2006, Vol.46, No.2, p.686-98	1 - 1 2
A	SCIABOLA, S., Pharmacophoric fingerprint method (TOPP) for 3D-QSAR modeling: application to CYP2D6 metabolic stability, Journal of chemical information and modeling, 2007.02, Vol.47, No.1, p.76-84	1 - 1 2
A	KELLY, M.D., Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design, Journal of chemical information and computer sciences, 2004, Vol.44, No.6, p.1942-51	1 - 1 2