

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-18793

(P2005-18793A)

(43) 公開日 平成17年1月20日(2005.1.20)

(51) Int. Cl.⁷

G06F 19/00

G06F 17/30

F I

G06F 19/00

1 0 0

G06F 19/00

Z T D

G06F 17/30

1 7 0 F

テーマコード (参考)

5 B 0 7 5

審査請求 未請求 請求項の数 12 O L 外国語出願 (全 32 頁)

(21) 出願番号 特願2004-190416 (P2004-190416)
 (22) 出願日 平成16年6月28日 (2004. 6. 28)
 (31) 優先権主張番号 60/483, 267
 (32) 優先日 平成15年6月26日 (2003. 6. 26)
 (33) 優先権主張国 米国 (US)

(71) 出願人 504249101
 アクセルリーズ・ソフトウェア・インコー
 ポレーテッド
 アメリカ合衆国・カリフォルニア・921
 21・サン・ディエゴ・スクラントン・ロ
 ード・9685

(74) 代理人 100064908
 弁理士 志賀 正武

(74) 代理人 100089037
 弁理士 渡邊 隆

(74) 代理人 100101465
 弁理士 青山 正和

(74) 代理人 100108453
 弁理士 村山 靖彦

最終頁に続く

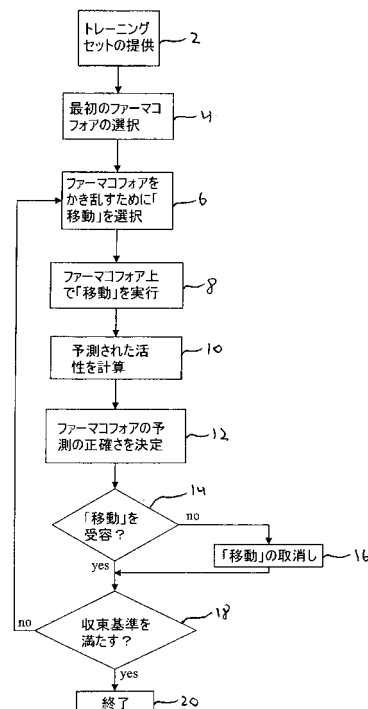
(54) 【発明の名称】 ファーマコフォアモデルの生成及び使用

(57) 【要約】

【課題】 ファーマコフォアモデルを生成する方法及び装置を提供すること。

【解決手段】 ファーマコフォアモデルを生成する方法及び装置は、モデルに存在する分子性状と欠けていると定義された性状との両方によって特徴付けられる。このように、特定のターゲットに対して活性を示す立体バルクのような性状と活性を促進する官能基のような性状とを考慮したモデルを開発することができる。活性を示す性状は、公知の活性な分子を公知の不活性な分子と比較することによって識別することができる。不活性な分子に存在するが活性な分子に欠けている性状は、ファーマコフォアモデルで定義することができる。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

ファーマコフォアを形成する方法であって、
第 1 選択分子性状を示すような第 1 位置を形成する段階と；
第 2 選択分子性状を欠くような第 2 位置を形成する段階と；
を備え、前記第 2 位置は、
一つ又は複数のターゲットに対して活性を示す第 1 分子を最初のバージョンのファーマコフォアに整列させる段階と；
前記一つ又は複数のターゲットに対してより少ない活性を示す第 2 分子を前記最初のバージョンに整列させる段階と；
前記第 1 分子の一つ又は複数の分子性状と相反する前記第 2 分子の分子性状を前記第 2 位置として認定する段階と；
によって決定されることを特徴とする方法。

10

【請求項 2】

請求項 1 記載の方法において、
前記第 2 選択分子性状は、立体バルクであることを特徴とする方法。

【請求項 3】

請求項 1 記載の方法において、
前記第 2 分子性状は、選択された原子の官能基であることを特徴とする方法。

【請求項 4】

請求項 1 記載の方法において、
前記第 2 分子性状は、荷電部分であることを特徴とする方法。

20

【請求項 5】

請求項 1 記載の方法において、
前記第 2 分子性状は、選択された原子型であることを特徴とする方法。

【請求項 6】

請求項 5 記載の方法において、
前記第 2 分子性状は、選択された組の原子型であることを特徴とする方法。

【請求項 7】

ファーマコフォアを形成する方法であって、
第 1 選択分子性状を示すような第 1 位置を形成する段階と；
第 2 選択分子性状を欠くような第 2 位置を形成する段階と；
を備え、前記第 2 位置は、
一つ又は複数のターゲットに対して活性を示す第 1 分子を前記一つ又は複数のターゲットに対してより少ない活性を示す第 2 分子に整列させる段階と；
前記第 1 分子の一つ又は複数の分子性状と相反する前記第 2 分子の分子性状を前記第 2 位置として認定する段階と；
によって決定されることを特徴とする方法。

30

【請求項 8】

ファーマコフォアに欠けているような性状を形成する方法であって、
一つ又は複数のターゲットに対して活性を示す第 1 分子を前記一つ又は複数のターゲットに対してより少ない活性を示す第 2 分子に対して整列させる段階と；
前記第 1 分子の一つ又は複数の分子性状と相反する前記第 2 分子の分子性状を前記性状として認定する段階と；
を備えることを特徴とする方法。

40

【請求項 9】

ファーマコフォアに欠けているような性状を形成する方法であって、
一つ又は複数のターゲットに対して不活性な分子を最初のバージョンの前記ファーマコフォアに対して整列させる段階と；
前記最初のバージョンの一つ又は複数の分子性状と相反する前記分子の分子性状を前記

50

性状として認定する段階と；
を備えることを特徴とする方法。

【請求項 10】

一つ又は複数のターゲットに対して活性を有することが期待された分子実体のファーマコフォアモデルを最適化する方法であって、

前記ターゲットに対して前記活性を示す第 1 分子を最初のバージョンの前記ファーマコフォアモデルと整列させる段階と；

前記ターゲットに対して前記活性を示さない第 2 分子を前記最初のバージョンの前記ファーマコフォアモデルと整列させる段階と；

両方が前記ファーマコフォアモデルと整列されたときに前記第 1 分子の分子性状と相反する前記第 2 分子の分子性状を認定する段階と；

認定された分子性状が欠けているという条件を含むように前記ファーマコフォアモデルを更新する段階と；

を備えることを特徴とする方法。

【請求項 11】

特定の特性を示す分子のファーマコフォアモデルを形成する方法であって、

第 1 組の分子性状を存在すると定義し、かつ第 2 組の分子性状を欠けていると定義する段階を備え、

分子における前記第 2 組の分子性状の存在は、前記分子に前記特性を示させないようにし、

前記第 2 組の分子性状は、前記特定の特性を示す分子を前記特定の特性を示さない分子と比較することによって決定されることを特徴とする方法。

【請求項 12】

分子スクリーニングで使用するためのファーマコフォアを生成する装置であって、

活性が知られている 1 組の誘引分子の分子構造を格納するメモリと；

ファーマコフォアモデルを生成しかつ該モデルを前記メモリに格納するよう構成されたファーマコフォア生成モジュールであって、該ファーマコフォア生成モジュールは活性分子性状存在モジュール及び不活性分子性状存在モジュールを備え、前記活性分子性状存在モジュールは存在が活性に寄与する前記ファーマコフォア内の介在物に対して分子性状を形成し、かつ前記不活性分子性状存在モジュールは存在が活性を抑制する、前記ファーマコフォア内で欠けていると示されるべき分子性状を形成し、欠けていると示されるべき分子性状は、異なる活性を有するトレーニングセットに 2 つの分子構造を整列させ、かつその他の分子構造における一つ又は複数の分子性状と相反する、前記 2 つの分子構造のうちの一つにおける分子性状を認定することによって決定される、ファーマコフォア生成モジュールと；

前記メモリから前記トレーニングセット内の分子構造と前記ファーマコフォアとを検索し、かつ前記分子構造と前記ファーマコフォアとの間の類似点を決定するよう構成された分子 - ファーマコフォア比較モジュールと；

前記類似点に基づいて前記分子構造に対応する分子の活性を評価するよう構成された活性予測モジュールであって、評価された活性が、前記ファーマコフォアモデルを生成するときに前記ファーマコフォア生成モジュールによって使用される活性予測モジュールと；
を備えることを特徴とする装置。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、計算化学に関する。より詳細には、この発明は、ファーマコフォアモデル (pharmacophore model) の生成及び使用に関する。

【背景技術】

【0002】

いくつかの計算方法は、研究者が分子の活性及び / 又は実験的特性を予測するときに用

10

20

30

40

50

いるのに利用できる。これら方法のいくつかは、一つ又は複数のファーマコフォアの生成を含む。ファーマコフォアは、特定の活性 / 特性、例えば位置 A における水素結合ドナー、位置 B における疎水基、等と相関関係にある分子の性状 (feature) の 3 次元表現である。ひとたび所望の活性 / 特性に対応するファーマコフォアが形成されれば、一つ又は複数の分子は、スクリーンされた分子のどれが、ファーマコフォアによって定義される性状と著しく重なる性状 (特徴) を有するかを決定することによって、活性 / 特性に対してスクリーンされ得る。ファーマコフォアの形成及びファーマコフォアの制御されたデータベース検査の概観は、非特許文献 1 に与えられている。

【0003】

最も初期のファーマコフォアは、結合機構の最も重要な性状を理解しようとする試みにおいて、リガンドの 3D 構造及び / 又は関連する結合位置、の直接の研究からの研究から手動で開発された。一つのそのような例は、非特許文献 2 に記述されている CNS ファーマコフォアである。

10

【0004】

1980 年代後半、活性を有する構造及び / 又は機能分子性状を首尾よく相関させるファーマコフォアを定義するプロセスを自動化しようとする試み、DISCO として知られているプログラムが作成された。このプログラムは、1 組の活性化化合物に関して似たような空間関係に位置する共通の構造及び / 又は機能性状に対して自動化された検査を実行した。

【0005】

いくつかの注意を受けたファーマコフォア生成の一つの側面は、ファーマコフォアの形成において「排除体積」を含めようとする試みだった。この問題点は、特定の分子は活性に要求される構造及び / 又は機能性状を含むことができるが、分子のある部分は、立体干渉がターゲットに対する結合を防止するように必要な機能性状に関係して位置させることができるために現れる。このように、原子を含むことが許されていないファーマコフォアの性状を生み出す活性の周囲に領域を定義することが役に立つことがわかった。

20

【0006】

ターゲット結合ポケット (target binding pocket) の構造が知られている場合には、改善されたファーマコフォアは、結合ポケットの内部表面に対応する排除体積領域を選択することによって開発されてきた。例えば、非特許文献 3 の第 18 章及び第 20 章を参照されたい。加えて、ファーマコフォア生成プログラム ALLADIN は、ユーザーが、ファーマコフォアの排除体積領域を定義する球の点格子又は組を定義することを可能にした。

30

【0007】

排除体積はファーマコフォアへ組み込まれてきたが、これらの体積を首尾よく組み込むのに要求されるかなりの量のユーザー相互作用が残っている。その上、多くの場合、結合ポケット構造の情報は利用できない。排除体積領域を形成する好ましい方法、特により自動化された排除体積の定義を可能にする方法は、当技術で有益である。

【非特許文献 1】Greene et. al, Chemical Function Queries for 3D Database Search, J. Chem. Inf. Comput. Sci. 34, 1297-1308 (1994)

40

【非特許文献 2】Lloyd et. al, A Common Structural Model for Central Nervous System Drugs and their Receptors, J. Med. Chem. 29 453-462 (1986)

【非特許文献 3】“ pharmacophore Perception, Development, and Use in Drug Design, ” edited by Osman F. Guner, International University Line, ISBN 0-9636817-6-1 (2000)

【発明の開示】

【課題を解決するための手段】

【0008】

一実施例は、第 1 選択分子性状を示すような第 1 位置を形成する段階と ; 第 2 選択分子性状を欠くような第 2 位置を形成する段階と ; を備え、前記第 2 位置は、1) 一つ又は複

50

数のターゲットに対して活性を示す第 1 分子を最初のバージョンのファーマコフォアに整列させる段階と； 2) 前記一つ又は複数のターゲットに対してより少ない活性を示す第 2 分子を前記最初のバージョンに整列させる段階と； 3) 前記第 1 分子の一つ又は複数の分子性状と相反する前記第 2 分子の分子性状を前記第 2 位置として認定する段階と； によって決定される、ファーマコフォアを形成 (define) する方法である。

【 0 0 0 9 】

他の実施例は、第 1 選択分子性状を示すような第 1 位置を形成する段階と；第 2 選択分子性状を欠くような第 2 位置を形成する段階と；を備え、前記第 2 位置は、 1) 一つ又は複数のターゲットに対して活性を示す第 1 分子を前記一つ又は複数のターゲットに対してより少ない活性を示す第 2 分子に整列させる段階と； 2) 前記第 1 分子の一つ又は複数の分子性状と相反する前記第 2 分子の分子性状を前記第 2 位置として認定する段階と； によって決定される、ファーマコフォアを形成する方法である。

10

【 0 0 1 0 】

他の実施例は、一つ又は複数のターゲットに対して活性を示す第 1 分子を前記一つ又は複数のターゲットに対してより少ない活性を示す第 2 分子に対して整列させる段階と；前記第 1 分子の一つ又は複数の分子性状と相反する前記第 2 分子の分子性状を前記性状として認定する段階と；を備える、ファーマコフォアに欠けているような性状を形成する方法である。

【 0 0 1 1 】

他の実施例は、一つ又は複数のターゲットに対して不活性な分子を最初のバージョンの前記ファーマコフォアに対して整列させる段階と；前記最初のバージョンの一つ又は複数の分子性状と相反する前記分子の分子性状を前記性状として認定する段階と；を備える、ファーマコフォアに欠けているような性状を形成する方法である。

20

【 0 0 1 2 】

他の実施例は、前記ターゲットに対して前記活性を示す第 1 分子を最初のバージョンの前記ファーマコフォアモデルと整列させる段階と；前記ターゲットに対して前記活性を示さない第 2 分子を前記最初のバージョンの前記ファーマコフォアモデルと整列させる段階と；両方が前記ファーマコフォアモデルと整列されたときに前記第 1 分子の分子性状と相反する前記第 2 分子の分子性状を認定する段階と；認定された分子性状が欠けているという条件を含むように前記ファーマコフォアモデルを更新する段階と；を備える、一つ又は複数のターゲットに対して活性を有することが期待された分子実体のファーマコフォアモデルを最適化する方法である。

30

【 0 0 1 3 】

他の実施例は、第 1 組の分子性状を存在すると定義し、かつ第 2 組の分子性状を欠けていると定義する段階を備え、分子における前記第 2 組の分子性状の存在は、前記分子に前記特性を示させないようにし、前記第 2 組の分子性状は、前記特定の特性を示す分子を前記特定の特性を示さない分子と比較することによって決定される、特定の特性を示す分子のファーマコフォアモデルを形成する方法である。

【 0 0 1 4 】

他の実施例は、分子の分子性状がファーマコフォアモデルに存在すると定義される対応する性状から特定の距離内にある場合に、評価された活性を増加させる段階と；分子の分子性状がファーマコフォアモデルで排除されると定義される領域から特定の距離内にある場合に、評価された活性を減少させる段階と；を含む分子の活性を評価する方法である。

40

【 0 0 1 5 】

他の実施例は、特定の特性を示すよう予測された分子のファーマコフォアモデルが格納されたメモリであって、ファーマコフォアモデルは、一つ又は複数の分子性状とこれらの個々の空間位置とをないとは定義する、メモリと；候補分子をファーマコフォアモデルと整列させ、かつ一つ又は複数の分子性状が候補分子に存在するか否かを決定することによって候補分子をファーマコフォアモデルと比較するよう構成されたプロセッサと；を備えるインシリコ分子スクリーニング装置 (in silico molecular screening system) である。

50

【 0 0 1 6 】

他の実施例は、活性が知られている 1 組の誘引分子の分子構造を格納するメモリと；ファーマコフォアモデルを生成しかつ該モデルを前記メモリに格納するよう構成されたファーマコフォア生成モジュールであって、該ファーマコフォア生成モジュールは活性分子性状存在モジュール及び不活性分子性状存在モジュールを備え、前記活性分子性状存在モジュールは存在が活性に寄与する前記ファーマコフォア内の介在物に対して分子性状を形成し、かつ前記不活性分子性状存在モジュールは存在が活性を抑制する、前記ファーマコフォア内で欠けていると示されるべき分子性状を形成し、欠けていると示されるべき分子性状は、異なる活性を有するトレーニングセットに 2 つの分子構造を整列させ、かつその他の分子構造における一つ又は複数の分子性状と相反する、前記 2 つの分子構造のうちの 10 一つにおける分子性状を認定することによって決定される、ファーマコフォア生成モジュールと；前記メモリから前記トレーニングセット内の分子構造と前記ファーマコフォアとを検索し、かつ前記分子構造と前記ファーマコフォアとの間の類似点を決定するよう構成された分子 - ファーマコフォア比較モジュールと；前記類似点に基づいて前記分子構造に対応する分子の活性を評価するよう構成された活性予測モジュールであって、評価された活性が、前記ファーマコフォアモデルを生成するときに前記ファーマコフォア生成モジュールによって使用される活性予測モジュールと；を備える、分子スクリーニングで使用するためのファーマコフォアを生成する装置である。

10

【 0 0 1 7 】

他の実施例は、テスト分子の分子構造及びファーマコフォアモデルを格納するメモリと；メモリからファーマコフォアモデル及び分子構造を検索し、かつ分子構造とファーマコフォアとの間の類似点を決定するよう構成された分子ファーマコフォア比較モジュールであって、類似点は、ファーマコフォアに存在すると定義された分子性状とファーマコフォアに存在しないと定義された分子性状とに基づいている、分子ファーマコフォア比較モジュールと；類似点に基づいて分子の活性を評価するよう構成された活性予測モジュールであって、評価された活性は、分子が、存在しないと定義された分子性状を含む場合に、減少する、活性予測モジュールと；を備えるテスト分子の活性を評価する装置である。

20

【 発明を実施するための最良の形態 】

【 0 0 1 8 】

上述したように、モデルに含まれねばならないものに完全に基づくファーマコフォアは、所望の活性又は特性を持たないトレーニングセット (training set) の分子に存在する分子性状によって引き起こされる不活性に対する寄与を無視する。このように、どんな分子性状が含まれねばならないかだけでなく、どんな性状が欠けていなければならないかに基づいて組み立てられたファーマコフォアを生成及び使用するアルゴリズムが必要とされている。

30

【 0 0 1 9 】

いくつかの実施例では、ある分子構造又は性状が欠けていることによって少なくとも部分的に特徴づけられるファーマコフォアを形成するアルゴリズムが与えられる。そのようなファーマコフォアの有用性の一つの例は、分子の形状 (立体バルク) と分子ターゲットの形状との間に不一致があるときである。例えば、特定の分子があるターゲットに対して活性であるような分子の種類を一般的に特徴づける官能基を有することができる一方で、分子は、分子がターゲットと首尾よく結合することを禁止する付加的な立体バルクを有することができる。官能基の存在と選択された領域における立体バルクの欠如との両方によって定義されるファーマコフォアは、ファーマコフォアを用いた分子活性予測 / 分子特性予測の正確さを高める。

40

【 0 0 2 0 】

[ファーマコフォアモデルの生成]

ファーマコフォアの生成及び使用の一般的な側面は、Osman F. Gunerによる " Pharmacophore Perception, Development, and Use " , International University Lineに記述されている。一般的に、ファーマコフォアの生成は、一つ又は複数の候補ファーマコフォア

50

の繰り返しの進歩を含む。この一般的なファーマコフォアの生成方法は、カルフォルニア州サンディエゴのAccelrys Softwareによって開発され売買されているプログラムC A T A L Y S Tに組み入れられてきた。このプログラムは、数値活性データが知られている分子のトレーニングセットの中の一つ又は複数の最も活性な分子で共通の性状構成を探す「構成段階 (constructive phase)」で始まる。次いで、このプログラムは、「控除段階 (subtractive phase)」を用いる。ここでは、あまりにも多くのあまり活性でないトレーニングセットの分子をも包含する場合には、構成段階で見つかった共通の性状構成は、さらなる考慮から排除される。最終的に、構成段階及び控除段階を乗り切るファーマコフォア候補は、「最適化段階」でかき乱される。このプロセスでは、ファーマコフォアの形成は、かき乱され、すなわち「移動され」、分子のトレーニングセットに関してファーマコフォアの予測の正確さに対する「移動」の効果が決定される。一実施例では、移動が予測の正確さを改善する場合には、移動が受け入れられる。移動が予測の正確さを改善しない場合には、様々な種類の可能な公知のモンテカルロシミュレーションによる決定基準に基づいて受容又は拒絶され得る。これらのプロセスは、上記Gunerの第10章及び第26章により詳細に記述されている。

10

20

30

【0021】

一実施例では、この最適化プロセスの一般的なアウトラインは、図1に示されている。モデルは、分子活性を予測するよう構成されている。ブロック2では、分子のトレーニングセットが与えられる。トレーニングセットは通常、活性として分類される分子と不活性として分類される分子とを含む。トレーニングセットにおける分子の活性は、例えば分子の実験的な分析によって決定される。分子の活性は従来、溶液 (IC_{50}) 内のターゲットの50%に結合することが要求される分子の分子濃度として、又は通常ナノモルからミリモルの範囲にある IC_{50} を用いた $- \log (IC_{50})$ として、定義される。かくして、ターゲットに対してより小さな結合親和力を有する分子は大きな IC_{50} とより小さな $- \log (IC_{50})$ とを有する。「活性」及び「不活性」との用語は相対的なものであるので、分子のどちらかの部類への分類は、様々な方法で行うことができる。例えば、最も低い IC_{50} (最も高い $- \log (IC_{50})$) を有する一つ又は複数の分子は「活性」として定義することができ、残りの分子は「不活性」として定義することができる。他の選択として、トレーニングセットの「不活性」分子は、最も低い IC_{50} を有するトレーニングセットの分子よりも大きなある閾値量の IC_{50} を有するものとして定義することができる。

【0022】

ブロック4では、最初のファーマコフォアが選択される。C A T A L Y S T プログラムでは、この選択は、構成段階及び控除段階で行われるが、ファーマコフォア候補を生成するあらゆる方法を用いることができる。一実施例では、最初のファーマコフォアは、高い活性を有するトレーニングセットにおける一つの分子の選択に基づく。一実施例では、選択された分子の分子構造全体が、最初のファーマコフォアとして用いられる。他の実施例では、分子構造の部分集合が、例えば全ての水素原子を除去することによって用いられる。さらに他の実施例では、選択された分子に存在する特定の官能基のみが、最初のファーマコフォアに含まれる。

40

【0023】

ブロック6では、ファーマコフォアをかき乱すために「移動」が選択される。いくつかの実施例では、「移動」は、官能基又は原子を加えること又は除去することを含む。他の実施例では、ファーマコフォア内の官能基又は原子は変形 (translate) される。好適な実施例では、「移動」はまた、欠けていなければならない添加性状 (adding feature)、除去性状 (removing feature)、又は変形性状 (translating feature) を含むことができる。これらの性状は、立体バルク (steric bulk) 又は荷電原子 (charged atom) を含むことができる。ブロック8では、「移動」は、性状を変更するためにファーマコフォアで実行される。ブロック10では、トレーニングセットの分子に対する予測された活性が、ファーマコフォアを各分子と比較することによって計算される。ブロック12では、予

50

測された活性が、ファーマコフォアの予測の正確さを決定するために公知の活性と比較される。いくつかの実施例では、この比較は、上述のGunerの第26章で定義されたようなコストを計算することを含む。

【0024】

決定ブロック14では、ファーマコフォア上で実行される「移動」を受け入れるか否かが決定される。いくつかの実施例では、「移動」は、ファーマコフォアの予測の正確さを改善する場合に受け入れられる。いくつかの実施例では、「移動」は、MCメトロポリスの受入基準を満たす限り、予測の正確さを改善しない場合にさえ受け入れられる。いくつかの実施例では、「移動」を受け入れるべきかの決定は、コストの変化に基づく。例えば、コストの変化のシミュレートされたアニール機能が、受入基準として用いられ得る。「移動」が拒絶される場合には、移動はブロック16でなされず、ブロック18でアルゴリズムが進行する。「移動」が受け入れられた場合には、アルゴリズムは直ちにブロック18へ進む。

10

【0025】

決定ブロック18では、特定の収束基準が満たされるかが決定される。いくつかの実施例では、収束基準は、特定の閾値に到達する予測の正確さに基づく。他の実施例では、収束基準は、追加の「移動」を用いた予測の正確さにおいて著しい進歩がないことに基づく。さらに他の実施例では、収束基準は、特定の閾値よりも下に下がる予測の正確さにおける進歩に基づく。収束基準が満たされるならば、ブロック20においてアルゴリズムが停止し、結果のファーマコフォアは、活性が知られていない分子の活性を予測するために用いることができる。収束基準が満たされない場合には、アルゴリズムは、ファーマコフォアの予測の正確さをさらに改善するための試みにおいて他の「移動」を選択するために、ブロック6へ戻る。

20

【0026】

いくつかの実施例では、最適化アルゴリズムが作動するときnトップのファーマコフォアが格納される。例えば、各「移動」が受け入れられた後、アルゴリズムは、新しいファーマコフォアがトップリストに現在格納されている最も悪いファーマコフォアよりも優れた予測の正確さを有するかを決定することができる。予測の正確さが優れているならば、新しいファーマコフォアがリストへ追加され、リスト上の最も悪いファーマコフォアが捨てられる。アルゴリズムの完了時、トップリストから一つ又は複数のファーマコフォアが、活性が知られていない分子の活性を予測するのに用いるために選択され得る。

30

【0027】

[排除体積の決定]

いくつかの実施例では、(立体バルクのない)排除体積をファーマコフォアへ追加するための「移動」は、トレーニングセット内の最も活性な分子の一つをトレーニングセット内の不活性な分子の一つに整列させることによって決定される。一実施例では、2つの分子がまた、現在のファーマコフォアに整列される。整列された活性分子内の全ての原子からの閾値距離よりも大きい、整列された不活性分子内のあらゆる原子は、不活性の、低い活性に責任を負っている可能性がある。かくしてこれら原子の位置は、排除体積を加えるための候補位置として用いることができる。他の実施例では、一つのあまり活性でない分子のみが現在のファーマコフォアに整列される。ファーマコフォアに存在すると定義される性状からの閾値距離よりも大きく離れて整列された分子内の原子は、排除体積を加えるための候補位置として用いられる。排除体積を加えることによって、原子が排除体積に侵入するいくつかの分子の予測された活性が減少する。

40

【0028】

排除体積を決定するアルゴリズムが図2のフローチャートに示されている。ブロック50では、トレーニングセット内の分子が活性又は不活性であるとして分類される。一実施例では、分類はユーザー定義される。他の実施例では、分類は、ある閾値よりも上又は下の活性を有する分子に基づいて決定される。一実施例では、不活性な分子は、次の基準を満たす分子として定義される：

50

$\log(\text{候補不活性分子のIC}_{50}) - \log(\text{最も活性な化合物のIC}_{50}) > \text{閾値}$
ここで、閾値は、ユーザー定義され、初期設定値 3.5 を有する。

【0029】

ブロック 52 では、現在の仮定とされるファーマコフォアに対して最も高い適合点数を有する不活性な分子が選択される。適合点数は、以下に述べられるようにして決定することができる。ブロック 54 では、活性であるとして分類されるトレーニングセット内の一つ又は複数の分子がファーマコフォアに整列される。分子及びファーマコフォアを整列させるための手順は、以下に述べられる。活性な分子における原子の位置が、活性な原子のリストを作成するために用いられる。一実施例では、排除体積アルゴリズムは、活性な分子が、選択された不活性な分子の適合点数よりも大きな、ファーマコフォアに対する整列適合点数を有する場合に実行されるのみである。ブロック 56 では、選択された不活性な分子がファーマコフォアと整列させられる。不活性な分子の位置は、不活性な原子のリストを生成するために用いられる。いくつかの実施例では、特定の原子のみが活性原子リスト及び不活性原子リストに含まれる。例えば、水素でない原子のみが含まれ得る。ブロック 58 では、活性原子リスト内の全ての原子からの閾値距離よりも遠い、不活性原子リスト内の原子が認定される。いくつかの実施例では、閾値距離はユーザー選択される。いくつかの実施例では、閾値距離は、1.2 オングストロームの初期設定値を有する。ブロック 60 では、一つ又は複数の排除体積位置が、ブロック 58 で認定される原子の位置から選択される。いくつかの実施例では、これらの位置は、認定された原子から無作為に選択される。最後に、ブロック 62 で、排除体積がファーマコフォアに加えられる。いくつかの実施例では、排除体積は、特定の半径を有する、ブロック 58 で選択される位置に中心を置かれた球として定義される。いくつかの実施例では、半径は 1.2 オングストロームである。

10

20

【0030】

ひとたび排除体積がファーマコフォアに加えられるれば、後の「移動」は、一つ又は複数の排除体積を除去することができるか、又はこれら排除体積を他の位置へ移すことができる。

【0031】

[分子の整列]

分子は、現在知られている様々な方法を用いて、互いに及び/又はファーマコフォアに整列され得る。文献の大部分はそのような方法を述べており、多くは、DISCO及びCATALYSTのような商業的に利用できる製品に組み込まれてきた。2つの分子の整列は、例えば共通のファーマコフォアへ両方を整列させることによって実行され得る。ファーマコフォアに対する分子の整列は、分子とファーマコフォアとの間の重複部分の特徴づける適合値を定義することによって実行され得る。いくつかの実施例では、適合値は、ファーマコフォアが存在するとして示す性状とファーマコフォアが欠けているとして示す性状との両方の整列によって特徴づけられる。いくつかの実施例では、ファーマコフォア内の性状は、上記Gunerの第26章に記述されたようなファーマコフォアモデルで相対的な重要性を示す、指定された重量値である。いくつかの実施例では、1.0の初期設定重量が全ての性状に割り当てられる。

30

40

【0032】

ファーマコフォアの各性状は、3次元位置を特定する一つ又は複数の位置束縛によって定義され得る。その上、各位置束縛は、各位置束縛についての誤差を定義する特定半径の球と結びつけられてきた可能性がある。

【0033】

いくつかの実施例では、適合値は、次の式によって決定される。

【0034】

【数 1】

$$\text{適合} = \sum_i \text{重量}(f_i) [1 - SSE(f_i)]$$

【0035】

ここで、各 f_i は、ファーマコフォアに存在する性状であり、重量 (f_i) は i 番目の性状に割り当てられる重量であり、 $SSE(f_i)$ は【数 2】によって定義される。

【0036】

【数 2】

$$SSE(f_i) = k \sum_j \left(\frac{D(c_{i,j})}{T(c_{i,j})} \right)^2$$

10

【0037】

ここで、各性状 f_i は、各性状に対して異なる可能性のある j 個の位置束縛を有し、 $c_{i,j}$ は、各性状 f_i に対する位置束縛であり、 $D(c_{i,j})$ は、性状 f_i における位置束縛 $c_{i,j}$ の対応する中心からのテスト分子内の原子位置の変位であり、 $T(c_{i,j})$ は、位置束縛球（誤差）の半径であり、 k は 1 又は $1/j$ のどちらかとすることができる。いくつかの実施例では、テスト分子及びファーマコフォアは、適合を最大にする分子の位置及び配向を見つけることによって整列される。当技術で知られている多くの適合アルゴリズムのうちのいずれもが、適合を最大にする際に用いられ得る。

20

【0038】

上で示した適合値は、ファーマコフォアに欠けているとして定義される性状を考慮に入れるよう調整され得る。例えば、ファーマコフォアが排除体積を含むならば、ファーマコフォアに対してテストされている分子が、排除領域の内側にあらゆる原子の $v d W$ （ファンデルワールス）体積をもたない場合、適合点数は変わらないままとすることができる。適合は、テスト分子が排除領域の内側に原子 $v d W$ 体積を含む場合にゼロと定義され得る。あるいは、テスト分子の排除体積と原子 $v d W$ 体積との間の所定量の重複部分が許され得る。この場合、適合点数は、重複部分の量に依存している量によって決められ得る。いくつかの実施例では、水素原子は、適合値を排除体積との重複部分に対して調整するとき

30

【0039】

排除体積に対して適合値を調整する一手段において、テスト分子内の原子と排除体積の中心との間の距離 d が決定され得る。 $d < x t$ 、ここで x は特定の排除体積因子であり、 t は排除体積の誤差（半径）に加えた原子のファンデルワールス半径である、ならば、テスト分子の原子は排除体積内なので、適合はゼロになるよう調整される。 $x t < d < t$ ならば、原子と排除体積との間の許される重複部分を明らかにするために、適合は【数 3】倍することができる。 $d > t$ ならば、適合は、原子がファーマコフォアの排除体積内にないので、変化しないままとすることができる。ファーマコフォアに欠けているとして定義される分子性状を明らかにするために他の基準案及び調整案を用いることもできる。

40

【0040】

【数 3】

$$\left(\frac{d - xt}{t - xt} \right)^2$$

【0041】

[予測された活性の計算]

上述したように、分子の活性は、活性をモデルファーマコフォアと比較することによって予測することができる。そのような予測は、上述のような最適化されたファーマコフォ

50

アを生成するプロセスの一部として、又は活性が知られていない分子の活性を予測するために、誘引分子 (training molecule) と共に用いることができる。いくつかの実施例では、予測された活性は、例えば上述の方法によってテスト分子とファーマコフォアとの間の類似点を決定することによって計算される。テスト分子とファーマコフォアとの間の類似性がより高いことにより、予測された活性がより高くなる。一実施例では、活性は次の式によって評価される。

$$\text{活性} = 10 \times \exp [- (\text{適合} + \text{遮断})]$$

ここで、活性は分子に対して予測された IC_{50} であり、適合は上で定義されたようなものであり、遮断は、予測された活性と分子のトレーニングセットの実際の活性との間の相関を最大にするための回帰分析を用いて決定される。

10

【0042】

〔ファーマコフォア装置〕

上述のアルゴリズムは、メモリ及びプロセッサを備えた汎用のコンピュータ装置で実行することができる。一つのそのような実施例は図3に描かれている。図3の装置は、メモリ100を備える。メモリ100は、一つ又は複数のファーマコフォアモデルと、一つ又は複数の誘引分子及び/又は一つ又は複数のテスト分子の分子構造と、を格納するために用いることができる。ファーマコフォア生成モジュール102は、メモリ100から誘引分子の構造を検索し、最適化されたファーマコフォアを構成するように作動する。ファーマコフォア生成モジュール102は、ファーマコフォアに含まれるべき性状を決定する活性分子性状存在モジュール104を備える。ファーマコフォア生成モジュール102はまた、ファーマコフォアに欠けているとして定義される排除体積のような性状を決定する不活性分子性状存在モジュール106をも備える。決定を行うとき、ファーマコフォア生成モジュール102は、トレーニングセット分子とファーマコフォアとの間の類似点を決定する分子-ファーマコフォア比較モジュール108を使用する。ファーマコフォア生成モジュール102はまた、分子-ファーマコフォア比較モジュール108によって生成された結果に基づいてトレーニングセット分子の予測された活性を計算する活性予測モジュール110をも使用することができる。

20

【0043】

活性予測モジュール110はまた、活性が不明の活性分子を予測するために用いることもできる。この実施例では、分子-ファーマコフォア比較モジュール108は、構造がメモリ100に格納されている、分子とファーマコフォアとの間の類似点を決定する。活性予測モジュールはしたがって、予測された活性を計算するためにこの決定を使用することができる。

30

【0044】

上述のアルゴリズムはいくつかの優位点を有する。一つは、ファーマコフォアの予測の正確さを改善する排除体積が、広大なユーザー相互作用なしに、又はターゲットの結合位置を知ることなしに、自動化された方法で定義され得るということである。これら方法が、ファーマコフォアモデルに欠けているとして定義される性状を付加的に定義することを含むように拡張することができる、というのが他の優位点である。例えば、排除体積の代わりに、不活性な分子が、活性な分子及び/又はファーマコフォア候補と整列され得るとともに、電気を帯びた領域、ある官能基、又は結合親和性を妨げることもあり得る特定の原子種のような、他の特定の性状の存在に対してスクリーン(遮蔽)され得る。これら他の種類の性状はしたがって、上述のファーマコフォア最適化プロセスにおいてファーマコフォア生成の一部としてテストされ得る。これにより、以前に用いられた方法からファーマコフォアを生成する際の柔軟性が著しく拡張される。

40

【0045】

この出願は、2003年6月26日に提出された米国特許仮出願第60/483,267号明細書に対する優先権を主張する。

【図面の簡単な説明】

【0046】

50

【図 1】ファーマコフォアモデルを生成するアルゴリズムを示すフローチャートである。

【図 2】ファーマコフォアモデル内に排除体積を定義するアルゴリズムを示すフローチャートである。

【図 3】ファーマコフォアモデルを生成及び使用する装置を示す図である。

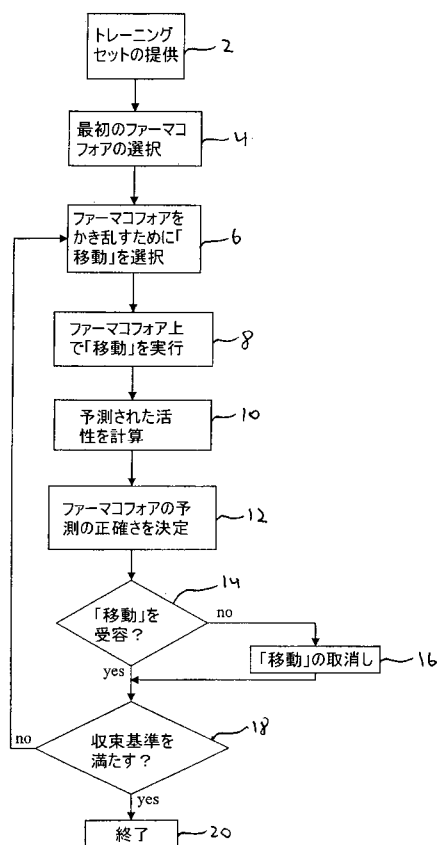
【符号の説明】

【 0 0 4 7 】

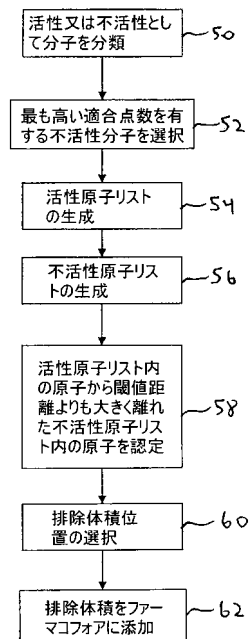
1 0 0 メモリ
1 0 2 ファーマコフォア生成モジュール
1 0 4 活性分子性状存在モジュール
1 0 6 不活性分子性状存在モジュール
1 0 8 分子 - ファーマコフォア比較モジュール
1 1 0 活性予測モジュール

10

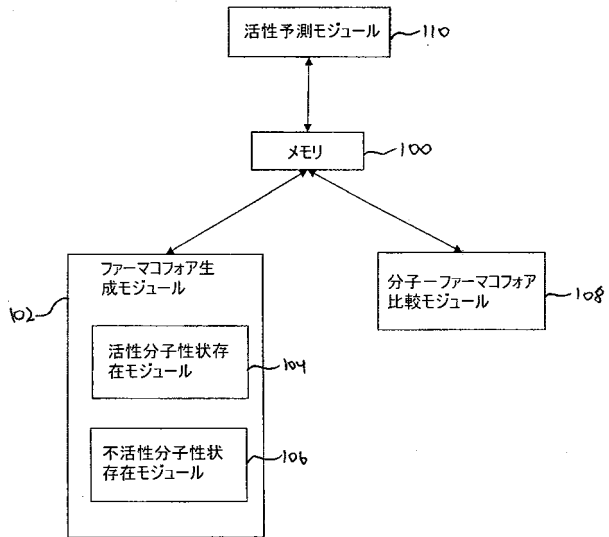
【図 1】



【図 2】



【 図 3 】



フロントページの続き

(74)代理人 100110364

弁理士 実広 信哉

(72)発明者 ジョナサン・マー・サッター

アメリカ合衆国・カリフォルニア・9 2 0 8 1・ヴィスタ・カベゾン・プレイス・6 1 4

(72)発明者 アリスター・ジェイ・メイナード

アメリカ合衆国・カリフォルニア・9 2 0 3 7・ラ・ジョラ・ヴィラ・ラ・ジョラ・ドライブ・#
3・8 6 5 6

(72)発明者 マーヴィン・ウォルドマン

アメリカ合衆国・カリフォルニア・9 2 1 3 0・サン・ディエゴ・カミニト・ミラ・デル・マール
・1 2 5 7 4

Fターム(参考) 5B075 UU18

【 外国語明細書 】

PHARMACOPHORE MODEL GENERATION AND USE

Related Application

[0001] This application claims priority to U.S. Provisional Application No. 60/483,267, filed on June 26, 2003.

Background of the Invention

Field of the Invention

[0002] The present invention relates to computational chemistry. More specifically, the present invention relates to the generation and use of pharmacophore models.

Description of the Related Art

[0003] Several computational methods are available for researchers to use in predicting the activities and/or experimental properties of molecules. Some of these methods include the generation of one or more pharmacophores. Pharmacophores are 3-dimensional representations of features of molecules that correlate with a specified activity/property, e.g. a Hydrogen bond donor at location A, a hydrophobic group at location B, etc. Once a pharmacophore corresponding to a desired activity/property is defined, one or more molecules can be screened for the activity/property by determining which of the screened molecules have features that significantly overlap the features defined by the pharmacophore. An overview of pharmacophore definition and pharmacophore directed database searching is provided in Greene et. al, Chemical Function Queries for 3D Database Search, *J. Chem. Inf. Comput. Sci.* 34, 1297-1308 (1994).

[0004] The earliest pharmacophores were manually developed from direct researcher study of 3D structures of ligands and/or associated binding sites in an attempt to understand the most important features of the binding mechanism. One such example is a CNS pharmacophore described in Lloyd et. al, A Common Structural Model for Central Nervous System Drugs and their Receptors, *J. Med. Chem.* 29 453-462 (1986).

[0005] In the late 1980s, a program known as DISCO was created which attempted to automate the process of defining pharmacophores that successfully correlate structural and/or functional molecular features with activity. This program performed an

automated search over a set of active compounds for common structural and/or functional features positioned in similar spatial relationships.

[0006] One aspect of pharmacophore generation that has received some attention has been the attempt to include "excluded volumes" in the pharmacophore definition. This issue arises because a particular molecule may contain the structural and/or functional features required for activity, but some portion of the molecule may be located relative to the necessary functional features such that steric interference prevents binding to the target. Thus, it has been found useful to define regions around the activity producing features of the pharmacophore which are not allowed to contain atoms.

[0007] In cases where the structure of the target binding pocket is known, improved pharmacophores have been developed by choosing an excluded volume region corresponding to the inner surface of the binding pocket. See, for example, Chapters 18 and 20 of "Pharmacophore Perception, Development, and Use in Drug Design," edited by Osman F. Güner, International University Line, ISBN 0-9636817-6-1 (2000). In addition, the pharmacophore generating program ALLADIN allowed the user to define a point grid or set of spheres defining excluded volume regions of pharmacophores.

[0008] Although excluded volumes have been incorporated into pharmacophores, there remains a significant amount of user interaction required to successfully incorporate them. Furthermore, in many cases, no binding pocket structure information is available. Better methods of defining excluded volume regions would be beneficial in the art, especially methods that allow more automated excluded volume definition.

Summary of the Invention

[0009] One embodiment is a method of defining a pharmacophore comprising: defining a first location as exhibiting a first selected molecular feature; and defining a second location as lacking a second selected molecular feature, wherein the second location is determined by: 1) aligning a first molecule that exhibits an activity against one or more targets to an initial version of a pharmacophore; 2) aligning a second molecule that exhibits less activity against the one or more targets to the initial version; and 3) identifying as the second location a molecular feature of the second molecule that is inconsistent with one or more molecular features of the first molecule.

[0010] Another embodiment is a method of defining a pharmacophore comprising: defining a first location as exhibiting a first selected molecular feature; and

defining a second location as lacking a second selected molecular feature, wherein the second location is determined by: 1) aligning a first molecule that exhibits an activity against one or more targets to a second molecule that exhibits less activity against the one or more targets; and 2) identifying as the second location a molecular feature of the second molecule that is inconsistent with one or more molecular features of the first molecule.

[0011] Another embodiment is a method of defining a feature as absent in a pharmacophore comprising: aligning a first molecule that exhibits an activity against one or more targets to a second molecule that exhibits less activity against the one or more targets; and identifying as the feature a molecular feature of the second molecule that is inconsistent with one or more molecular features of the first molecule.

[0012] Another embodiment is a method of defining a feature as absent in a pharmacophore comprising: aligning a molecule that is inactive against one or more targets to an initial version of the pharmacophore; and identifying as the feature a molecular feature of the molecule that is inconsistent with one or more molecular features of the initial version.

[0013] Another embodiment is a method of optimizing a pharmacophore model of a molecular entity expected to have activity against one or more targets; the method comprising: aligning a first molecule that exhibits the activity against the target with an initial version of the pharmacophore model; aligning a second molecule that does not exhibit the activity against the target with the initial version of the pharmacophore model; identifying a molecular feature of the second molecule that is inconsistent with the molecular features of the first molecule when both are aligned with the pharmacophore model; and updating the pharmacophore model to include a requirement that the identified molecular feature be absent.

[0014] Another embodiment is a method of defining a pharmacophore model of a molecule exhibiting a particular property, the method comprising defining a first set of molecular features as present and a second set of molecular features as absent, wherein the presence of the second set of molecular features in a molecule inhibits the molecule from exhibiting the property, the second set of molecular features determined by comparing a molecule exhibiting the particular property with a molecule not exhibiting the particular property.

[0015] Another embodiment is a method of estimating the activity of a molecule comprising: increasing the estimated activity if a molecular feature of the molecule is within a specified distance from a corresponding feature defined as present in a pharmacophore model; and decreasing the estimated activity if a molecular feature of the molecule is within a specified distance from a region defined as excluded in the pharmacophore model.

[0016] Another embodiment is an *in silico* molecular screening system comprising: a memory having stored therein a pharmacophore model of molecules predicted to exhibit a particular property, wherein the pharmacophore model defines one or more molecular features and their respective spatial positions as absent; and a processor configured to compare candidate molecules to the pharmacophore model by aligning the candidate molecules with the pharmacophore model and determining whether or not the one or more molecular features are present in the candidate molecules.

[0017] Another embodiment is a system for generating a pharmacophore for use in molecular screening comprising: a memory storing molecular structures of a set of training molecules for which activity is known; a pharmacophore generation module configured to generate a pharmacophore model and store the model in the memory; the pharmacophore generation module comprising an active molecular feature presence module and an inactive molecular feature presence module, wherein the active molecular feature presence module defines molecular features for inclusion in the pharmacophore whose presence contributes to activity and the inactive molecular feature presence module defines molecular features to be designated in the pharmacophore as absent whose presence inhibits activity, wherein molecular features to be designated as absent are determined by aligning two molecular structures in the training set that have different activities and identifying a molecular feature in one of the two molecular structures that is inconsistent with one or more molecular features in the other molecular structure; a molecule-pharmacophore comparison module configured to retrieve a molecular structure in the training set and the pharmacophore from the memory and determine similarity between the molecular structure and the pharmacophore; and an activity-prediction module configured to estimate activity of the molecule corresponding to the molecular structure based on the similarity, wherein the estimated activity is used by the pharmacophore generation module in generating the pharmacophore model.

[0018] Another embodiment is a system for estimating activity of a test molecule comprising: a memory storing a pharmacophore model and a molecular structure of the test molecule; a molecule-pharmacophore comparison module configured to retrieve the pharmacophore model and the molecular structure from memory and determine similarity between the molecular structure and the pharmacophore, wherein the similarity is based on molecular features that are defined as present in the pharmacophore and molecular features that are defined as absent in the pharmacophore; and an activity prediction module configured to estimate activity of the molecule based on the similarity, wherein the estimated activity is decreased if the molecule contains the molecular features that are defined as absent.

Brief Description of the Drawings

[0019] FIGURE 1 is a flowchart illustrating an algorithm for generating pharmacophore models.

[0020] FIGURE 2 is a flowchart illustrating an algorithm defining excluded volumes in a pharmacophore model.

[0021] FIGURE 3 illustrates a system for generating and using pharmacophore models.

Detailed Description of the Preferred Embodiment

[0022] As discussed above, pharmacophores based entirely on what must be *included* in the model ignores contributions to inactivity caused by molecular features present in the molecules of the training set that do not have the desired activity or properties. Thus, there is a need for algorithms that generate and make use of pharmacophores that are constructed based not only on what molecular features must be included, but also what features must be *absent*.

[0023] In some embodiments, an algorithm is provided that defines a pharmacophore that is at least partially characterized by the absence of some molecular structure or feature. One example of the utility of such a pharmacophore is when there is an incompatibility between a molecule's shape (steric bulk) and the shape of a molecular target. For example, while a particular molecule may have functional groups that generally characterize a class of molecules as active for some target, the molecule may have additional steric bulk that prohibit the molecule from successfully binding with the

target. A pharmacophore that is defined by both the presence of the functional groups and the absence of steric bulk in selected regions enhances the accuracy of molecular activity/property predictions using the pharmacophore.

[0024] Pharmacophore Model Generation

[0025] General aspects of pharmacophore generation and use are described in "Pharmacophore Perception, Development, and Use" by Osman F. Güner, International University Line. Generally, pharmacophore generation involves iterative improvement of one or more candidate pharmacophores. This general pharmacophore generation method has been incorporated into the program CATALYST, developed and marketed by Accelrys Software of San Diego, California. This program begins with a "constructive phase" that looks for common feature arrangements in one or more of the most active molecules in a training set of molecules for which numerical activity data is known. It then uses a "subtractive phase" where common feature arrangements found in the constructive phase are eliminated from further consideration if they also cover too many of the less active molecules of the training set. Finally, the pharmacophore candidates that survive the constructive and subtractive phases are perturbed in an "optimization phase." In this process, the definition of a pharmacophore is perturbed or "moved" and the effect of the "move" on pharmacophore predictive accuracy with respect to the training set of molecules is determined. In one embodiment, if the move improves predictive accuracy, the move is accepted. If the move does not improve predictive accuracy, it may be accepted or rejected based on a variety of possible known Monte Carlo simulation decision criteria. These processes are described in more detail in Chapters 10 and 26 of Güner, *supra*.

[0026] In one embodiment, the general outline of this optimization process is illustrated in Figure 1. The model is designed to be predictive of molecular activity. At block 2 a training set of molecules is provided. The training set typically includes molecules that are classified as active and molecules that are classified as inactive. The activity of the molecules in the training set are determined such as by experimental assays of the molecules. The activity of a molecule is conventionally defined as the molar concentration of the molecule required to bind to 50% of the target in solution (IC_{50}), or as the $-\log(IC_{50})$ with the IC_{50} typically being in the nanomolar to millimolar range. Thus, a molecule with less binding affinity for the target has a larger IC_{50} and a smaller $-\log(IC_{50})$. Because the terms "active" and "inactive" are relative, the classification of a

molecule to either category can be made in a variety of ways. For example, one or more of the molecules having the lowest IC_{50} (highest $-\log(IC_{50})$) may be defined as "active" and the remaining molecules may be defined as "inactive." As another alternative, an "inactive" molecule of the training set may be defined as one that has an IC_{50} of a certain threshold amount greater than the molecule of the training set with the lowest IC_{50} .

[0027] At block 4, an initial pharmacophore is selected. In the CATALYST program, this is done in the constructive and subtractive phases, but any method of generating a pharmacophore candidate may be used. In one embodiment, the initial pharmacophore is based on a selection of one of the molecules in the training set that has high activity. In one embodiment, the entire molecular structure of the selected molecule is used as the initial pharmacophore. In other embodiments, a subset of the molecular structure is used, such as by removing all hydrogen atoms. In still other embodiments, only specified functional groups present in the selected molecule are included in the initial pharmacophore.

[0028] At block 6, a "move" is selected to perturb the pharmacophore. In some embodiments, the "move" includes adding or removing a functional group or atom. In other embodiments, functional groups or atoms in the pharmacophore are translated. In advantageous embodiments, the "move" may also include adding, removing, or translating features that must be absent. These features can include steric bulk or charged atoms. At block 8, the "move" is performed on the pharmacophore to change it. At block 10, predicted activities for the molecules in the training set are calculated by comparing the pharmacophore with each molecule. At block 12, the predicted activities are compared with the known activities to determine the predictive accuracy of the pharmacophore. In some embodiments, this comparison comprises calculating a cost as defined in the above mentioned Chapter 26 of Güner.

[0029] At decision block 14, it is determined whether or not to accept the "move" performed on the pharmacophore. In some embodiments, the "move" is accepted if it improves the predictive accuracy of the pharmacophore. In some embodiments, the "move" is accepted even if it does not improve predictive accuracy as long as it meets a MC Metropolis acceptance criterion. In some embodiments, determination of whether to accept the "move" is based on the change in cost. For example, a simulated anneal function of the change in cost may be used as the acceptance criterion. If the "move" is

rejected, it is undone at block 16 and the algorithm proceeds at block 18. If the “move” is accepted, the algorithm immediately advances to block 18.

[0030] At decision block 18, it is determined whether a specified convergence criterion is met. In some embodiments, the convergence criterion is based on the predictive accuracy reaching a specified threshold. In other embodiments, the convergence criterion is based on there being no significant improvement in predictive accuracy with additional “moves.” In still other embodiments, the convergence criterion is based on the improvement in predictive accuracy dropping below a specified threshold. If the convergence criterion is met, the algorithm stops at block 20 and the resulting pharmacophore can be used to predict the activity of molecules for which the activity is unknown. If the convergence criterion is not met, the algorithm returns to block 6 to select another “move” in an attempt to further improve the predictive accuracy of the pharmacophore.

[0031] In some embodiments, the n top pharmacophores are stored as the optimization algorithm operates. For example, after each accepted “move,” the algorithm can determine whether the new pharmacophore has better predictive accuracy than the worst pharmacophore currently stored in the top list. If the predictive activity is better, the new pharmacophore is added to the list and the worst pharmacophore on the list is discarded. Upon completion of the algorithm, one or more pharmacophores from the top list can be selected for use in predicting activities of molecules whose activities are not known.

Determination of Excluded Volume

[0032] In some embodiments, a “move” for adding excluded volume (absence of steric bulk) to a pharmacophore is determined by aligning one of the most active molecules in the training set to one of the inactive molecules in the training set. In one embodiment, the two molecules are also aligned to the current pharmacophore. Any atoms in the aligned inactive molecule that are greater than a threshold distance from all the atoms in the aligned active molecule could be responsible for the low activity of the inactive. The locations of these atoms may thus be used as candidate locations for adding excluded volumes. In another embodiment, only one less active molecule is aligned to the current pharmacophore. Atoms in the aligned molecule that are greater than a threshold distance from features defined as present in the pharmacophore are used as candidate

locations for adding excluded volume. Adding an excluded volume will decrease the predicted activity of some molecules whose atoms encroach on the excluded volumes.

[0033] An algorithm for determining excluded volume is illustrated in the flowchart of Figure 2. At block 50, the molecules in the training set are classified as being active or inactive. In one embodiment, the classification is user defined. In another embodiment, the classification is determined based on those molecules having an activity above or below some threshold. In one embodiment, inactive molecules are defined as those molecules for which the following criterion is met:

$$\log(IC_{50} \text{ of candidate inactive molecule}) - \log(IC_{50} \text{ of the most active compound}) > \text{threshold}$$

where *threshold* is user defined and has a default value of 3.5.

At block 52, the inactive molecule having the highest fit score to the currently hypothesized pharmacophore is selected. The fit score may be determined as described below. At block 54, one or more molecules in the training set that are classified as active are aligned to the pharmacophore. A procedure for aligning a molecule and a pharmacophore is described below. The co-ordinates of the atoms of the active molecules are used to create an active atom list. In one embodiment, the excluded volume algorithm is only pursued if the active molecules have an alignment fit score to the pharmacophore greater than the fit score of the selected inactive molecule. At block 56, the selected inactive molecule is aligned with the pharmacophore. The co-ordinates of the inactive molecule are used to generate an inactive atom list. In some embodiments, only specified atoms are included in the active and inactive atom lists. For example, only non-hydrogen atoms may be included. At block 58, the atoms in the inactive atom list that are further than a threshold distance from all of the atoms in the active atoms list are identified. In some embodiments, the threshold distance is user selected. In some embodiments, the threshold distance has a default value of 1.2 Angstroms. At block 60, one or more excluded volume locations are selected from the locations of the atoms identified at block 58. In some embodiments, the locations are randomly selected from the identified atoms. Finally, at block 62, excluded volume is added to the pharmacophore. In some embodiments, the excluded volume is defined as a sphere centered on the locations

selected in block 58 having a specified radius. In some embodiments, the radius is 1.2 Angstroms.

[0034] Once excluded volume is added to a pharmacophore, later “moves” may remove the one or more excluded volumes or translate them to other locations.

Alignment of Molecules

[0035] Molecules may be aligned to one another and/or to a pharmacophore using a variety of currently known methods. A large body of literature describes such methods, many of which have been incorporated into commercially available products such as DISCO and CATALYST. Alignment of two molecules may be performed, for example, by aligning both to a common pharmacophore. Alignment of a molecule to a pharmacophore may be performed by defining a fit value that characterizes the overlap between a molecule and a pharmacophore. In some embodiments, the fit value is characterized by both alignment of features that the pharmacophore designates as being present and features that the pharmacophore designates as being absent. In some embodiments, features in the pharmacophore are assigned weight values that indicate their relative importance in the pharmacophore model as described in Chapter 26 of Güner, *supra*. In some embodiments, a default weight of 1.0 is assigned to all features.

[0036] Each feature of a pharmacophore may be defined by one or more location constraints that specify 3-dimensional coordinates. Furthermore, each location constraint may have associated with it a sphere of specified radius that defines a tolerance about each location constraint.

[0037] In some embodiments, the fit value is determined by the following formula:

$$fit = \sum_i weight(f_i)[1 - SSE(f_i)]$$

where each f_i is a feature that is present in the pharmacophore, $weight(f_i)$ is the weight assigned to the i -th feature, and $SSE(f_i)$ is defined by:

$$SSE(f_i) = k \sum_j \left(\frac{D(c_{i,j})}{T(c_{i,j})} \right)^2$$

where each feature f_i has j location constraints, which can be different for each feature, $c_{i,j}$ are the location constraints for each feature f_i , $D(c_{i,j})$ is the displacement of atom positions in the test molecule from the corresponding centers of location constraints $c_{i,j}$ in feature f_i ,

and $T(c_{ij})$ is the radius of the location constraint sphere (tolerance) and k may be either 1 or $1/j$. In some embodiments, a test molecule and the pharmacophore are aligned by finding the position and orientation of the molecule that maximizes *fit*. Any of the many fitting algorithms known in the art may be used in maximizing *fit*.

[0038] The above-indicated *fit* value may be adjusted to take into account features that are defined as being absent in the pharmacophore. For example, if a pharmacophore contains an excluded volume, the fit score may be left unaffected if the molecule being tested against the pharmacophore does not have any atom vdW (van der Waals) volume inside the excluded region. The fit may be defined as zero if the test molecule includes an atomic vdW volume inside the excluded region. Alternatively, a defined amount of overlap between the excluded volume and an atomic vdW volume of the test molecule may be allowed. In this case, the fit score may be scaled by an amount that is dependent on the amount of overlap. In some embodiments, hydrogen atoms are ignored in adjusting the *fit* value for overlap with an excluded volume.

[0039] In one implementation of adjusting the *fit* value for an excluded volume, a distance d between an atom in the test molecule and the center of the excluded volume may be determined. If $d < xt$, where x is a specified excluded volume factor and t is the tolerance (radius) of the excluded volume plus the van der Waals radius of the atom, then *fit* is adjusted to be zero because an atom of the test molecule is within the excluded volume. If $xt < d < t$, then *fit* may be multiplied by:

$$\left(\frac{d - xt}{t - xt} \right)^2$$

to account for allowed overlap between an atom and the excluded volume. If $d > t$, *fit* may be left unchanged because the atom is not within an excluded volume of the pharmacophore. Other criteria and adjustment schemes to account for molecular features that are defined as absent in the pharmacophore may also be used.

Calculation of Predicted Activities

[0040] As discussed above, the activity of a molecule can be predicted by comparing it with a model pharmacophore. Such prediction may be used with training molecules as part of the process of generating an optimized pharmacophore as described above or to predict the activity of a molecule for which activity is not known. In some embodiments, the predicted activity is calculated by determining the similarity between

the test molecule and the pharmacophore such as by the methods described above. Higher similarity between the test molecule and the pharmacophore leads to a higher predicted activity. In one embodiment, activity is estimated by the following formula:

$$activity = 10 \exp[-(fit + intercept)]$$

Where *activity* is the predicted IC₅₀ for the molecule, *fit* is as defined above and *intercept* is determined using a regression analysis to maximize correlation between predicted activities and actual activities of the training set of molecules.

Pharmacophore System

[0042] The algorithms described above may be implemented in a general purpose computer system comprising a memory and a processor. One such embodiment is depicted in Figure 3. The system of Figure 3 comprises a memory 100. The memory 100 can be used to store one or more pharmacophore models as well as the molecular structures of one or more training molecules and/or one or more test molecules. Pharmacophore generation module 102 operates to retrieve the structures of training molecules from memory 100 and construct an optimized pharmacophore. The pharmacophore generation module 102 comprises an active molecular feature presence module 104 that determines features that are to be included in the pharmacophore. The pharmacophore generation module 102 also comprises an inactive molecular feature presence module 106 that determines features such as excluded volumes that are defined as absent in the pharmacophore. In making its determinations, the pharmacophore generation module 102 makes use of a molecule-pharmacophore comparison module 108 that determines the similarity between the training set molecules and a pharmacophore. The pharmacophore generation module 102 can also make use of an activity prediction module 110 that calculates predicted activity of the training set molecules based on the results produced by the molecule-pharmacophore comparison module 108.

[0043] The activity prediction module 110 can also be used to predict the activity molecules for which activity is unknown. In this embodiment, the molecule-pharmacophore comparison module 108 determines the similarity between the molecule and a pharmacophore, whose structures are stored in memory 100. The activity prediction module can then make use of this determination to calculate predicted activity.

[0044] The above described algorithms have several advantages. One is that excluded volumes which improve pharmacophore predictive accuracy can be defined in an automated way without extensive user interaction or knowledge of target binding sites. It is another advantage that the methods can be extended to incorporate additional definitions of features defined as absent in a pharmacophore model. For example, instead of excluded volumes, inactive molecules could be aligned with active molecules and/or a pharmacophore candidate and be screened for the presence of other specific features such as charged regions, certain functional groups, or specific atom types that may also interfere with binding affinity. These other types of features could then be tested as part of pharmacophore generation in the above described pharmacophore optimization process. This significantly extends the flexibility of pharmacophore generation from methods used previously.

1. A method of defining a pharmacophore comprising:
 - defining a first location as exhibiting a first selected molecular feature; and
 - defining a second location as lacking a second selected molecular feature,wherein said second location is determined by:
 - aligning a first molecule that exhibits an activity against one or more targets to an initial version of a pharmacophore;
 - aligning a second molecule that exhibits less activity against said one or more targets to said initial version; and
 - identifying as said second location a molecular feature of said second molecule that is inconsistent with one or more molecular features of said first molecule.
2. The method of Claim 1, wherein said second selected molecular feature comprises steric bulk.
3. The method of Claim 1, wherein said second molecular feature comprises a selected atomic functional group.
4. The method of Claim 1, wherein said second molecular feature comprises a charged moiety.
5. The method of Claim 1, wherein said second molecular feature comprises a selected atom type.
6. The method of Claim 5, wherein said second molecular feature comprises a selected set of atom types.
7. A method of defining a pharmacophore comprising:
 - defining a first location as exhibiting a first selected molecular feature; and
 - defining a second location as lacking a second selected molecular feature,wherein said second location is determined by:
 - aligning a first molecule that exhibits an activity against one or more targets to a second molecule that exhibits less activity against said one or more targets; and
 - identifying as said second location a molecular feature of said second molecule that is inconsistent with one or more molecular features of said first molecule.

8. A method of defining a feature as absent in a pharmacophore comprising:
 - aligning a first molecule that exhibits an activity against one or more targets to a second molecule that exhibits less activity against said one or more targets; and
 - identifying as said feature a molecular feature of said second molecule that is inconsistent with one or more molecular features of said first molecule.
9. A method of defining a feature as absent in a pharmacophore comprising:
 - aligning a molecule that is inactive against one or more targets to an initial version of said pharmacophore; and
 - identifying as said feature a molecular feature of said molecule that is inconsistent with one or more molecular features of said initial version.
10. A method of optimizing a pharmacophore model of a molecular entity expected to have activity against one or more targets; said method comprising:
 - aligning a first molecule that exhibits said activity against said target with an initial version of said pharmacophore model;
 - aligning a second molecule that does not exhibit said activity against said target with said initial version of said pharmacophore model;
 - identifying a molecular feature of said second molecule that is inconsistent with the molecular features of said first molecule when both are aligned with said pharmacophore model; and
 - updating said pharmacophore model to include a requirement that said identified molecular feature be absent.
11. A method of defining a pharmacophore model of a molecule exhibiting a particular property, said method comprising defining a first set of molecular features as present and a second set of molecular features as absent, wherein the presence of the second set of molecular features in a molecule inhibits the molecule from exhibiting said property, wherein said second set of molecular features is determined by comparing a molecule exhibiting the particular property with a molecule not exhibiting the particular property.
12. A system for generating a pharmacophore for use in molecular screening comprising:
 - a memory storing molecular structures of a set of training molecules for which activity is known;

a pharmacophore generation module configured to generate a pharmacophore model and store said model in said memory; the pharmacophore generation module comprising an active molecular feature presence module and an inactive molecular feature presence module, wherein said active molecular feature presence module defines molecular features for inclusion in said pharmacophore whose presence contributes to activity and said inactive molecular feature presence module defines molecular features to be designated in said pharmacophore as absent whose presence inhibits activity, wherein molecular features to be designated as absent are determined by aligning two molecular structures in said training set that have different activities and identifying a molecular feature in one of the two molecular structures that is inconsistent with one or more molecular features in the other molecular structure;

a molecule-pharmacophore comparison module configured to retrieve a molecular structure in said training set and said pharmacophore from said memory and determine similarity between said molecular structure and said pharmacophore; and

an activity-prediction module configured to estimate activity of the molecule corresponding to said molecular structure based on said similarity, wherein said estimated activity is used by said pharmacophore generation module in generating said pharmacophore model.

1 Abstract

Methods and systems for generating pharmacophore models are characterized both by molecular features that are present in the model and features that are defined as absent. Thus, models can be developed that take into account features such as steric bulk that inhibit activity for a specified target as well as features such as functional groups that promote activity. Features that inhibit activity can be identified by comparing known active molecules with known inactive molecules. Features that are present in the inactive molecules but absent in the active molecules can be defined in a pharmacophore model.

2 Representative Drawing

Fig. 1

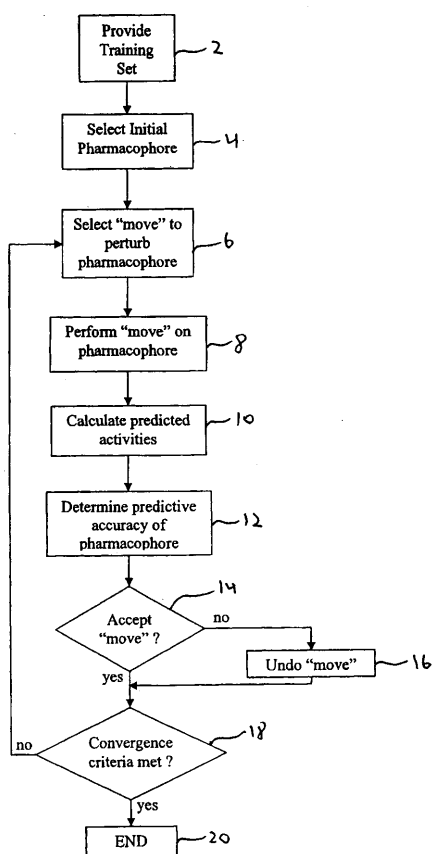


FIG. 1

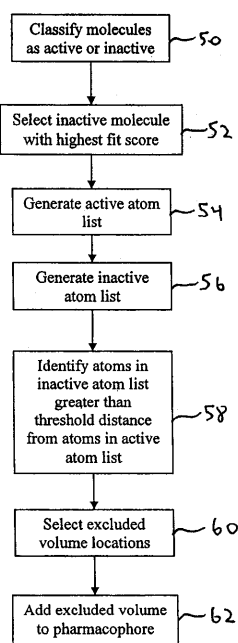


FIG. 2

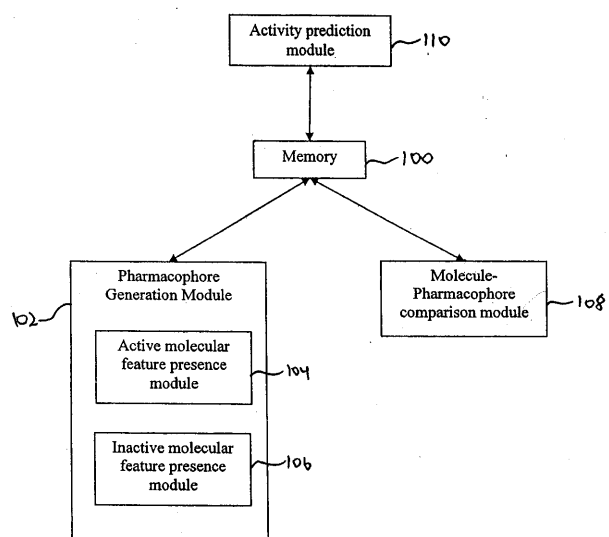


FIG. 3