

(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2017-123169

(P2017-123169A)

(43) 公開日 平成29年7月13日(2017.7.13)

(51) Int.Cl.
G06F 19/12 (2011.01)F I
G O 6 F 19/12

テーマコード (参考)

審査請求 有 請求項の数 72 O L 外国語出願 (全 79 頁)

(21) 出願番号 特願2016-256619 (P2016-256619)
 (22) 出願日 平成28年12月28日(2016.12.28)
 (62) 分割の表示 特願2014-211968 (P2014-211968)
 の分割
 原出願日 平成16年10月14日(2004.10.14)
 (31) 優先権主張番号 60/511,474
 (32) 優先日 平成15年10月14日(2003.10.14)
 (33) 優先権主張国 米国 (US)

(71) 出願人 506125867
 バーセオン
 アメリカ合衆国 カリフォルニア州 94
 538, フレモント, カト ロード 4
 8820-100B
 (74) 代理人 100076428
 弁理士 大塚 康徳
 (74) 代理人 100115071
 弁理士 大塚 康弘
 (74) 代理人 100112508
 弁理士 高柳 司郎
 (74) 代理人 100116894
 弁理士 木村 秀二
 (74) 代理人 100130409
 弁理士 下山 治

最終頁に続く

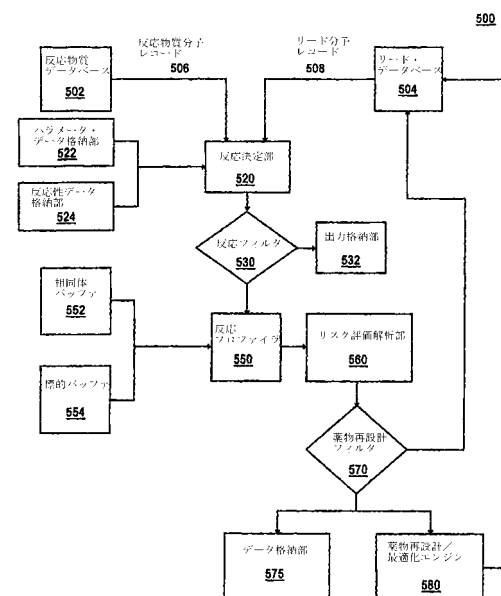
(54) 【発明の名称】 リード分子交差反応の予測・最適化システム

(57) 【要約】 (修正有)

【課題】 リード候補生体分子と、生体高分子であることが多い潜在的な反応物質分子の間の有害交差反応を予測して、対象とする行基または生物学的な状態に関する安全性および薬効の点でリード候補をよりよく特徴づけ、他の無菅家菜生体分子との有害なかつ／またはその他の点で不足の交差反応を回避する。

【解決手段】 計算システムで、適切な環境において反応をモデル化して、リード候補分子と潜在的な反応物質分子の間の反応プロファイルを求める。次いで、各リードと複数の潜在的な反応物質分子についての複数の反応プロファイルに基づいて、リードごとにリスク評価を生成し、予測される有害な交差反応をなくすために、反復してリード候補を再設計し最適化する。

【選択図】 図5



・モデリング・システム

【特許請求の範囲】

【請求項 1】

1 つまたは複数の所望の標的生体分子と反応することが知られているリード候補生体分子と潜在的な反応物質生体分子との潜在的な交差反応を、分子の組合せを計算によりモデル化し、関連する生物学的かつ化学的な注釈を分析することに基づいて予測しモデル化して反応プロファイルを生成する方法であって、

計算によるモデル化システムに前記リード候補生体分子のデータ表現を提供することを含み、前記データ表現は、前記リード候補の化学的かつ構造的な情報を含み、前記方法はさらに、

計算によるモデル化システムに前記潜在的な反応物質生体分子のデータ表現を提供することを含み、前記データ表現は、前記潜在的な反応物質の化学的かつ構造的な情報を含み、前記方法はさらに、

ある環境状況における分子の組合せの計算によるモデル化に基づいて、前記リード候補生体分子と前記潜在的な反応物質生体分子の間の潜在的な交差反応について、1 つまたは複数の反応性の値を推定することと、

前記潜在的な交差反応に関連する 1 つまたは複数の反応性の値の関数として反応の尺度を構築することと、

前記リード候補および前記潜在的な反応物質に関連する前記反応の尺度、ならびに関連するバイオインフォマティクスおよび化学的な注釈の分析結果の関数として反応プロファイルを構築することを含む、方法。

【請求項 2】

前記環境は、前記リード候補と前記潜在的な反応物質の反応部位の物理的かつ化学的な周囲の状況である、請求項 1 に記載の方法。

【請求項 3】

前記環境は、前記リード候補と、前記リードに対する所望の、または既知の標的生体分子との反応部位の物理的かつ化学的な周囲の状況である、請求項 1 に記載の方法。

【請求項 4】

前記環境は、複数の原子、イオン、および／または他の分子を含む溶媒の（暗示的または明示的な）表現を含み、前記溶媒中に前記反応部位が埋め込まれている、請求項 1 に記載の方法。

【請求項 5】

前記環境は、ヒトその他の哺乳動物の身体の 1 つまたは複数の構成要素を含む、請求項 1 に記載の方法。

【請求項 6】

前記潜在的な反応物質生体分子は、生体高分子、あるいは生体高分子の 1 つまたは複数の部分を含み、前記 1 つまたは複数の部分は、1 つまたは複数の既知の、または予測される活性部位を含む、請求項 1 に記載の方法。

【請求項 7】

前記リード候補は、合成化合物、化学基、ペプチド、医薬化合物、タンパク質系薬物、または炭水化物を含む、請求項 1 に記載の方法。

【請求項 8】

前記潜在的な交差反応に関連する反応の尺度を形成するのに用いる前記反応性の値の 1 つは、前記反応の結合親和性または自由エネルギーの計算による推定値である、請求項 1 に記載の方法。

【請求項 9】

前記潜在的な交差反応に関連する反応の尺度を形成するのに用いる前記反応性の値の 1 つは、並列処理パイプラインを利用して得られる、請求項 1 に記載の方法。

【請求項 10】

前記潜在的な交差反応に関連する前記反応性の値の一部または全部は、あらかじめ生成されるか、またはオフラインで導出され、コンピュータ可読媒体上にデジタル表現で格納

10

20

30

40

50

され、次いで、前記反応の尺度の形成時に取り出される、請求項 1 に記載の方法。

【請求項 1 1】

前記潜在的な交差反応に関連する反応の尺度を形成するのに用いる前記反応性の値の 1 つは、前記 2 つの生体分子間の定量的な構造活性の関係の事前予測の結果である、請求項 1 0 に記載の方法。

【請求項 1 2】

前記潜在的な交差反応に関連する反応の尺度を形成するのに用いる前記反応性の値の 1 つは、1 つまたは複数の実験室でのアッセイから得られた測定値または結果である、請求項 1 0 に記載の方法。

【請求項 1 3】

前記 1 つ（または複数）の実験室でのアッセイは、*in vitro*、細胞ベース、または *in vivo* のいずれかにかかわらず、スクリーニングその他の機能アッセイを含む、請求項 1 2 に記載の方法。

【請求項 1 4】

前記潜在的な交差反応に関連する反応の尺度は、複数の反応性の値の 1 次結合である、請求項 1 に記載の方法。

【請求項 1 5】

前記潜在的な交差反応に関連する反応の尺度は、複数の反応性の値の非線形関数である、請求項 1 に記載の方法。

【請求項 1 6】

前記潜在的な交差反応に関連する反応の尺度は、1 つまたは複数の反応性の値に基づくスコアまたは等級である、請求項 1 に記載の方法。

【請求項 1 7】

前記システムに反応フィルタを導入して、関連する反応の尺度が判定基準を満たさない潜在的な交差反応物質を、スクリーニングによる除去その他の方法でさらなる検討から除外する、請求項 1 に記載の方法。

【請求項 1 8】

前記判定基準は、1 つ（または複数）の入力された反応の尺度の大きさに適用する数値的な閾値である、請求項 1 7 に記載の方法。

【請求項 1 9】

前記判定基準は、前記システムによって反応の尺度が生成されるときに動的に適用される、請求項 1 7 に記載の方法。

【請求項 2 0】

前記判定基準は、個々のリード候補生体分子を扱うすべての反応の尺度が生成された後でのみ適用される、請求項 1 7 に記載の方法。

【請求項 2 1】

前記判定基準は、1 つまたは複数の所望の、または既知の標的と反応するリード候補生体分子について計算または実験により得られた対応する反応の尺度との相対的な比較に基づくものである、請求項 1 7 に記載の方法。

【請求項 2 2】

前記判定基準は、前記リード候補生体分子と潜在的な反応物質の反応に関連する結合親和性（または等価の反応の尺度）に閾値を適用することであり、前記親和性閾値は、（おそらくは 1 未満の）所定の定数と、前記リード候補生体分子と選択された標的との反応に関連する計算または測定された結合親和性（または等価の反応の尺度）とを掛けたものとして表現される、請求項 2 1 に記載の方法。

【請求項 2 3】

前記反応プロファイルの構築で用いる前記潜在的な反応物質生体高分子に関連する前記バイオインフォマティクスその他の化学的な注釈は、前記潜在的な反応物質に関連するソース、機能、相同性、および／またはファミリーその他の分類の情報に係るデータを含む、請求項 6 に記載の方法。

10

20

30

40

50

【請求項 2 4】

前記潜在的な反応物質についての前記ソース注釈は、前記潜在的な反応物質が得られ、由来することが期待され、また、見つかる可能性がある生物、種、器官、組織、細胞型、突然変異体または野生型、および／または系統に関する情報を含む、請求項 2 3 に記載の方法。

【請求項 2 5】

前記潜在的な反応物質についての前記機能注釈は、前記潜在的な反応物質に関わる生物学的な、さらには病気の経路その他の反応および過程、あるいは、前記潜在的な反応物質についての既知の、さらには予測される生物学的な機能に関する情報を含む、請求項 2 3 に記載の方法。

10

【請求項 2 6】

前記潜在的な反応物質についての前記相同性情報は、同じまたは異なる生物の他の生体高分子との構造および／または配列に基づく相同体との関連物を含み、タンパク質についての前記構造に基づく相同性は、保存領域、タンパク質モチーフ、さらには 3 次構造の類似性に基づいて確立され、生体高分子についての前記配列に基づく相同性は、1 次アミノ酸またはその基礎となるゲノミクス配列の類似性に基づくものである、請求項 2 3 に記載の方法。

【請求項 2 7】

前記潜在的な反応物質についての前記相同性情報は、前記リード候補生体分子と反応または結合することが知られている、予測される、かつ／または望まれる 1 つまたは複数の標的との構造および／または配列に基づく相同性を含む、請求項 2 6 に記載の方法。

20

【請求項 2 8】

前記潜在的な反応物質についての構造および／または配列に基づく相同体は、前記潜在的な反応物質についての追加の機能注釈を推測または予測するためのものであり、前記相同体についてのソース情報は、前記潜在的な反応物質に関連するソース情報に類似していることも、異なることもある、請求項 2 6 に記載の方法。

【請求項 2 9】

前記反応プロファイルの構築で用いる前記リード候補に関連する前記化学注釈は、前記リード候補に関連する化学的かつ構造的な組成、反応の化学的性質、相同性、および／または化学的または構造的な分類情報に関するデータを含む、請求項 1 に記載の方法。

30

【請求項 3 0】

前記リード候補についての前記化学的かつ構造的な組成の注釈は、化学式、化学名、1 つまたは複数の化学目録に関する指標記述子、生体分子の化学的かつ物理的な特性、キラル中心および立体化学についての知識、または 2 次元または 3 次元の表現での構造情報に関する情報を含む、請求項 2 9 に記載の方法。

【請求項 3 1】

前記リード候補についての前記反応の化学的性質の注釈は、前記リード候補に関わることが知られている、またはその可能性がある 1 つまたは複数の化学反応または生物学的なプロセスに関する反応データを含む、請求項 2 9 に記載の方法。

【請求項 3 2】

前記化学反応または生物学的なプロセスの範囲は、前記リードとの構造的かつ／または化学的な類似性が高い他の生体分子に関わることが知られている、またはその可能性がある化学反応または生物学的なプロセスを含むように拡張される、請求項 3 1 に記載の方法。

40

【請求項 3 3】

前記リードに関わる 1 つまたは複数の化学反応または生物学的なプロセスに関する前記反応データは、前記リードが結合その他の方法で反応することが既に知られている 1 つまたは複数の標的についての情報との関連物を含む、請求項 3 1 に記載の方法。

【請求項 3 4】

前記リードに関わる 1 つまたは複数の化学反応または生物学的なプロセスに関する前

50

記反応データは、アドメトックスその他の薬物動態学のプロファイルの1つまたは複数の情報構成要素を含む、請求項31に記載の方法。

【請求項35】

前記リードに関わる1つまたは複数の化学反応または生物学的なプロセスに関係する前記反応データは、1つまたは複数の実験室でのアッセイから得られた1つまたは複数の測定値を含む、請求項31に記載の方法。

【請求項36】

前記リード候補についての前記相同性情報は、他の生物分子との構造に基づく相同体および／または化学的な相同体との関連物を含み、前記構造に基づく相同性および／または化学的な相同性は、2次元または3次元のファルマコフォア・モデル、化合物のクラスタリングおよび分類で用いる様々な分子記述子、または定量的な構造活性の関係の類似性を用いることに基づいて確立される、請求項29に記載の方法。

10

【請求項37】

前記潜在的な反応物質と、前記リード候補に関連する1つまたは複数の所望の、または既知の標的との構造的または化学的な類似性に関する情報は、前記潜在的な反応物質を扱う潜在的な交差反応についての前記反応プロファイルに反応フィルタを適用するとき、あるいは、前記反応プロファイルを構築するときに考慮される、請求項1に記載の方法。

【請求項38】

前記リード候補と、前記潜在的な反応物質についての1つまたは複数の既知の、または予測される結合剤との構造的または化学的な類似性に関する情報は、前記リードを扱う潜在的な交差反応についての前記反応プロファイルに反応フィルタを適用するとき、あるいは、前記反応プロファイルを構築するときに考慮される、請求項1に記載の方法。

20

【請求項39】

前記リード候補生体分子は、複数の潜在的な反応物質生体分子に関して検討され、1組の反応プロファイルが構築され、それぞれ異なる反応プロファイルは、前記リード候補生体分子と個々の潜在的な反応物質生体分子の間の潜在的な交差反応に対応する、請求項1に記載の方法。

【請求項40】

リード候補生体分子についてのリスク評価は、複数の潜在的な反応物質生体分子に関して前記リード候補について生成された1組の反応プロファイルに基づいて評価される、請求項39に記載の方法。

30

【請求項41】

リード候補生体分子についての前記リスク評価は、新しい潜在的な反応物質が識別され、前記モデリング・システムで利用可能になったときに定期的または不定期に更新される、請求項40に記載の方法。

【請求項42】

前記リスク評価は、複数の定量的な反応プロファイルの1次結合その他の機能的な合成物である、請求項40に記載の方法。

【請求項43】

前記リスク評価は、複数の定量的な反応プロファイルに基づいて生成された定性的な分類または等級である、請求項40に記載の方法。

40

【請求項44】

前記リスク評価は、複数の定量的な反応プロファイルの統計分析に基づいて生成された信頼性のスコアまたは確率あるいは統計量である、請求項40に記載の方法。

【請求項45】

リード候補生体分子についての前記リスク評価を構成するのに用いる前記統計分析では、1組の反応プロファイルすべての平均値または中心値、あるいは、前記1組の反応プロファイルの百分率の上位選択範囲を用いる、請求項44に記載の方法。

【請求項46】

リード候補生体分子についての前記リスク評価を構成するのに用いる前記統計分析はベ

50

イズ・リスク分析である、請求項 44 に記載の方法。

【請求項 47】

前記リスク評価は、コンピュータ学習手法から出力された予測を表す、請求項 40 に記載の方法。

【請求項 48】

前記リスク評価を生成するのに用いる前記コンピュータ学習手法は、ニューラル・ネットワークに基づくものである、請求項 47 に記載の方法。

【請求項 49】

前記リスク評価を生成するのに用いる前記コンピュータ学習手法は、サポート・ベクトル・マシンに基づくものである、請求項 47 に記載の方法。

10

【請求項 50】

複数の潜在的な反応物質生体分子の選択は、2次元または3次元の構造データベースの全部または一部からなされる、請求項 39 に記載の方法。

【請求項 51】

複数の潜在的な反応物質生体分子の選択は、関連する生物学的かつ／または薬物動態学的な経路の事前知識に関わる、請求項 39 に記載の方法。

【請求項 52】

前記潜在的な反応物質は生体高分子であり、複数の潜在的な反応物質生体高分子の選択は、前記リード候補と、1つまたは複数の所望の、または既知の標的との構造上かつ／または配列上の類似性に関わる、請求項 39 に記載の方法。

20

【請求項 53】

複数の潜在的な反応物質生体高分子の選択には、前記リード候補の1つまたは複数の所望の、または既知の標的の、タンパク質または遺伝子のファミリー、あるいは、他の機能またはバイオインフォマティックスの分類結果の他のメンバが含まれる、請求項 52 に記載の方法。

【請求項 54】

潜在的な反応物質間の化学的かつ／または構造的な類似性の考慮を含む1組のクラスタリング基準に基づいて、前記複数の潜在的な反応物質生体分子にクラスタリング操作を適用して、クラスタまたはグループの集合を生成することと、

各クラスタまたは各グループから1つの潜在的な反応物質を選択して、クラスタの代表とすることと、

30

前記リード候補と各クラスタ代表の間の潜在的な交差反応に対応する1組の反応の尺度を生成することと、

前記得られた1組の反応の尺度に反応フィルタを適用して、前記リードとの高い反応性を示すクラスタ代表のサブセットを選択することと、

合格したクラスタ代表を含む潜在的な反応物質からなる各クラスタを選択し直すことと、

前記選択し直したクラスタのあらゆるメンバについての詳細な個々の反応の尺度および後続の反応プロファイルを生成することとをさらに含む、請求項 39 に記載の方法。

【請求項 55】

40

前記潜在的な反応物質は生体高分子であり、前記クラスタリング基準は、配列相同性、構造相同性、および／または他のバイオインフォマティックスまたはプロテオミックスの分類基準に関わる、請求項 54 に記載の方法。

【請求項 56】

1つまたは複数の所望の標的との結合特異性を改善し、前記リード候補との潜在的に有害な交差反応に関与することが予測される、またはその可能性がある1つまたは複数の潜在的な反応物質との反応性を低下させるために前記リード候補の最適化および再設計を行う別の方法が採用され、この方法は、

1つまたは複数の所望の標的と結合するために前記リードが必要とする、前記リード候補に関連する分子のコアまたは枠組を識別することと、

50

前記枠組の1つまたは複数の連結部位に追加または置換し得る分子群その他の構成要素を識別することと、

前記リードと、前記リードが強く反応することが望まれる1つまたは複数の所望の標的との間の1つ（または複数）の反応を、前記リードおよび所望の各標的に関連する化学基間でエネルギー的に有利な分子相互作用を保存または促進する点で分析することと、

前記リードと、前記リードが強く反応することが望まれない1つまたは複数の潜在的な反応物質との間の1つ（または複数）の反応を、前記リードおよび潜在的な各反応物質に関連する化学基間でエネルギー的に有利な分子相互作用を低減させるか、かつ／またはエネルギー的に不利な分子相互作用を促進する点で分析することと、

前記リードと1つまたは複数の所望の標的との分子相互作用の分析に従って、かつ、前記リードと1つまたは複数の潜在的な反応物質との分子相互作用の分析に従って、局所分子構造の追加または置換を行うための前記枠組の1つまたは複数の連結部位を選択することと、

10

前記リードと1つまたは複数の所望の標的との分子相互作用の分析に従って、かつ、前記リードと1つまたは複数の潜在的な反応物質との分子相互作用の分析に従って、前記選択した連結部位の1つまたは複数において局所分子構造に対して追加または置換する1つまたは複数の分子群その他の構成要素を選択することと、

前記枠組の選択した連結部位と選択した分子群の追加または置換との組合せに基づいて、前記リード候補の複数の構造変異体を構築することと、

前記リードの新しい各構造変異体を、1つまたは複数の潜在的な反応物質および1つまたは複数の所望標的との反応性に従って評価することとをさらに含む、請求項1に記載の方法。

20

【請求項57】

前記リードの新しい各構造変異体を評価することは、前記リード候補の前記新たに生成された構造変異体を、新しい入力リード候補として、請求項1に記載の方法に基づいて潜在的な交差反応をモデル化するシステムに提示し直すことを含む、請求項56に記載の方法。

【請求項58】

前記リード候補について既に構築された反応プロファイルおよび／または前記リード候補に関連するリスク評価の評価に基づいて、可能な再設計および最適化のためにリード候補に優先順位を付ける、請求項56に記載の方法。

30

【請求項59】

リードについての対応するリスク評価により、前記リード候補と1つまたは複数の潜在的な反応物質を扱う潜在的な有害交差反応に関連するリスクが許容差内にあり、かつ、所定の制限を超えないことが示唆されるとき、前記リード候補を提示して可能な再設計および最適化を行う、請求項58に記載の方法。

【請求項60】

前記リード候補について複数の構造的に別個の枠組を識別し、前記リードの前記可能な再設計および最適化中に、それらを前記リードの複数の構造変異体を生成するための可能なシードとして用いる、請求項56に記載の方法。

40

【請求項61】

前記枠組の1つまたは複数の連結部位に対する可能な追加または削除に関して識別された前記1組の分子群は、有機化学で見られる様々な化学基、アミノ酸側鎖、残基、さらには個々の原子またはイオンを含む、請求項56に記載の方法。

【請求項62】

前記リードと1つまたは複数の潜在的な反応物質の間の1つ（または複数）の反応の分析は、水素結合の相互作用するドナー基またはアクセプタ基の改変によって不安定になった場合、前記リードと前記潜在的な反応物質の反応性を低下させることがあるが、前記リードと1つまたは複数の所望の標的の反応性を過度に損なわない1つまたは複数の分子間水素結合を識別することを含む、請求項56に記載の方法。

50

【請求項 6 3】

前記リードと 1 つまたは複数の潜在的な反応物質の間の 1 つ（または複数）の反応の分析は、相互作用する疎水基の改変によって不安定になった場合、前記リードと前記潜在的な反応物質の反応性を低下させることがあるが、前記リードと 1 つまたは複数の所望の標的の反応性を過度に損なわない 1 つまたは複数の分子間疎水性接触部を識別することを含む、請求項 5 6 に記載の方法。

【請求項 6 4】

前記リードと 1 つまたは複数の潜在的な反応物質の間の 1 つ（または複数）の反応の分析は、相互作用する静電的な、またはイオン化された基の改変によって不安定になった場合、前記リードと前記潜在的な反応物質の反応性を低下させることがあるが、前記リードと 1 つまたは複数の所望の標的の反応性を過度に損なわない 1 つまたは複数の分子間塩橋を識別することを含む、請求項 5 6 に記載の方法。

10

【請求項 6 5】

前記リードと 1 つまたは複数の潜在的な反応物質の間の 1 つ（または複数）の反応の分析は、化学的または構造的に改変された場合、前記リードと前記潜在的な反応物質の反応性を低下させることがあるが、前記リードと 1 つまたは複数の所望の標的の反応性を過度に損なわない高い局所立体相補性を示すいずれかの分子の 1 つまたは複数の相互作用する化学基を識別することを含む、請求項 5 6 に記載の方法。

【請求項 6 6】

前記リード候補の複数の構造変異体を構築することは、前記リード候補の 1 つまたは複数の基または構成要素に関連する立体化学的な性質を改変することをさらに含む、請求項 5 6 に記載の方法。

20

【請求項 6 7】

前記リード候補の複数の構造変異体を構築することは、前記リード候補の 1 つまたは複数の基または構成要素に関連するイオン化状態を改変することをさらに含む、請求項 5 6 に記載の方法。

【請求項 6 8】

前記リード候補の複数の構造変異体を構築することに関わるプロセス、前記構造変異体を、請求項 1 に記載の方法に基づいて潜在的な交差反応をモデル化するシステムに入力として提示し直すこと、ならびに、その後で、新しい各入力構造変異体と 1 つまたは複数の潜在的な反応物質ごとに反応プロファイルおよび／またはリスク評価を構築することを、何らかの終端条件その他の基準が満たされるまで反復的に繰り返す、請求項 5 7 に記載の方法。

30

【請求項 6 9】

潜在的な交差反応をモデル化する前記計算システムは、前記計算プラットフォームを実施するソフトウェアを含むプログラム可能な汎用コンピュータ、専用ハードウェア、ファームウェア、あるいはこれらの組合せの 1 つまたは複数を用意する、請求項 1 に記載の方法。

【請求項 7 0】

反応性の評価および関連する生物学的かつ化学的な注釈の分析に基づいて、前記潜在的な反応物質分子の中から、既知の標的と異なる 1 つまたは複数の新しい潜在的な標的を識別することをさらに含む、請求項 1 に記載の方法。

40

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、概ね、バイオインフォマティクス、プロテオミクス、分子モデリング、コンピュータ支援分子設計（CAMD）に関し、より詳細には、コンピュータ支援薬物設計（CADD）および分子の組合せの計算によるモデル化に関する。

【背景技術】

【0002】

本願は、2003 年 10 月 14 日出願の「Lead Molecule Cross

50

Reaction Prediction and Optimization Process」という名称の米国仮出願第60/511474号の優先権を主張し、その正式出願である。事実上、この内容全体を参照により本明細書に組み込む。

【0003】

本開示は、（本願と同日に出願のプラカッシュ他の「Method and Apparatus for Analysis of Molecular Configurations and Combinations」という名称の米国特許出願第10/967085号（弁理士整理番号第021986-000310US）（以下、「プラカッシュ I」と称する）本発明の譲受人に譲渡された出願／特許である米国特許第8036867号に関係するものである。

10

【0004】

事実上、これらの出願／特許のそれぞれの開示全体を参照により本明細書に組み込む。

【0005】

従来方式の創薬プロセスおよびそれらの制限を説明すると、本発明を理解するのに有用であろう。

【0006】

何らかの生物学的な異常の処置または治療する新薬を創り出すことは、時間とコストがかかるプロセスであり、典型的には、1つの薬物当たり平均して12年の年月および8億ドルの費用がかかり、場合によっては、完成するのにおそらくは最大で15年以上の年月および10億ドルの費用がかかる。このプロセスは、薬物自体または最終的な薬物分子の前駆体として働く潜在的な化学化合物の識別、評価、最適化を行うためのウェット・ラボによる試験／実験、様々な生化学および細胞に基づくアッセイ、動物モデル、および計算ツールの形式での計算によるモデリングを含み得る。

20

【0007】

創薬プロセスの目標は、潜在的な分子の相互作用または組合せを介して、生体内の1つまたは複数の生体分子（すなわち、薬物の「標的」）の機能に影響を及ぼす化学化合物またはリガンド、すなわち、結合剤、生体分子を識別し特徴付けることであり、これらは通常、生体高分子である。本明細書では、生体高分子という用語は、タンパク質、核酸（DNAまたはRNA）、ペプチドまたはヌクレオチドの配列、あるいはこれらの断片の任意の部分のうち1つまたは複数を含む巨大分子を指す。本明細書では、生体分子という用語は、生体高分子、炭化水素、ホルモンその他の無機または有機の分子あるいは化合物のうち1つまたは複数を含む化学的な実体を指す。これらには、合成化合物、医薬化合物、薬物様化合物、または天然化合物、あるいはこれらの任意の部分または断片が含まれるが、これらに限定されるものではない。標的分子は、典型的には、病気に関係する標的のタンパク質または核酸であり、その機能、構造、および／または化学的活性の変化に影響を及ぼし、それによって患者の病気その他の疾患の処置の助けとなることが望まれるものである。他の場合には、標的は、ウィルス、細菌、または寄生生物など、病気の原因となる生物内に見つかる生体分子であり、薬物の影響を受けると、この感染性生物の生存または活動が影響を受けることになるものである。他の場合には、標的は、癌細胞などの異常または有害な細胞の生体分子である。他の場合には、標的は、アレルギー反応その他の望まれない免疫学的または生物学的な応答を誘発し得る抗原その他の環境的な化学作用物質である。

30

40

【0008】

リガンドは、典型的には、薬効、低毒性、膜透過性、溶解性、化学／代謝安定性などの点で所望の薬物様の特性を備えた小型分子による薬物または化学化合物として知られているものである。他の場合には、リガンドは、生物学的なもの、例えば、タンパク質系またはペプチド系の注入薬物とすることもできるし、あるいは別の完全なタンパク質とすることもさえできる。他の場合には、リガンドは、標的酵素の化学的な基質とし得る。リガンドは、標的に共有結合させることさえでき、また実際には、タンパク質の一部、例えば、タンパク質の2次構造成分、活性部位を含むタンパク質ドメインまたは活性部位近傍のタン

50

パク質ドメイン、適切なタンパク質4次構造のタンパク質サブユニットなどとすることもできる。

【0009】

背景技術の考察の残りの部分を通じて、具体的に区別しない限り、（潜在的な）分子の組合せは1つのリガンドおよび1つの標的を扱い、これらのリガンドおよび標的は別々の化学的な実体であり、リガンドは化学化合物と仮定し、標的は、典型的には、（突然変異型または野生型の）生物学的なタンパク質とする。近年、標的としての核酸（DNA/RNAとも）の頻度は、遺伝子治療および病原微生物学が進歩するにつれ増加する可能性があることに留意されたい。また、「分子複合体」という用語は、適切な（しばしば水を含む）環境中で標的とリガンドが相互作用するときの両者の結合状態を指す。「潜在的な」分子複合体は、確率は低いが発生し得る結合状態、したがって、通常の条件下で実際に形成されることもあり、されないこともある結合状態を指す。

10

【0010】

創薬プロセス自体は、典型的には、（1）標的の実証、（2）リードの創製／最適化、（3）臨床前試験、および（4）臨床的な治験および承認の4つの異なるサブプロセスを含む。

【0011】

標的の実証には、病気に関連する1つまたは複数の標的の確定が含まれ、通常、完了するのに2年半かかる。この標的実証段階の結果には、生物内の標的分子の存在または活動により、治療または処置を試みる病気が発症するか、悪化するか、または、それに寄与するある種の作用が生じるか、または影響を受けることを確定することが含まれる。場合によっては、実験的な方法により、標的についての天然の結合剤または基質を確定することもできる。

20

【0012】

リードの創製は、典型的には、リード化合物、すなわち、標的分子に結合し得るリガンドの識別に関わるものである。これらのリガンドは、標的の機能の活性化、不活性化、触媒作用、または抑制によって標的の作用を改変し得るものである。この場合、リードは、薬物適用プロセスで使用する適切なリガンドの候補とみなされる。リードの最適化は、所望の標的との結合親和性を向上させ、選択性を増大させ、毒性、溶解性、および代謝性の基本的な問題に対処するために、リード候補を化学的かつ構造的に改善して薬物前駆体にするに関わるものである。リードの創製およびリードの最適化を合わせて完了するのに、典型的には約3年かかり、さらなる考察を行うための1つまたは複数の化学的に異なるリードを得ることができる。

30

【0013】

臨床前試験では、生化学的アッセイおよび動物モデルを用いて、薬物の吸収および膜透過性、分布、代謝、排泄、毒性、副作用、および必要とされる投薬量に関する様々な薬物動態学因子について、選択されたリードを試験する。この臨床前試験は約1年かかる。臨床前試験期間後、臨床の治験および承認には、さらに6～8年以上かかり、その間、人間の被験者に対して安全および薬効について薬物候補を試験する。今日の創薬の途方もない経費は、1つには、最適化したリードの多くが依然として、副作用その他の理由から臨床前試験および臨床試験で不合格になることからくるものである。実際、臨床治験を通過し、最終的に承認される薬物の数は、創薬プロセスの初期に実証された標的タンパク質について開始されるプロジェクトの数と比較して極めて少ない。

40

【0014】

合理的な薬物設計は一般に、有効なリード候補の創製および最適化を設計する基礎として、薬物の標的についての構造的な情報を利用し（構造に基づく設計）、かつ／またはこれらの天然リガンドを利用する（リガンドに基づく設計）。構造に基づく合理的な薬物設計は一般に、標的構造の3次元モデルを利用する。標的のタンパク質または核酸の場合、このような構造は、X線結晶解析／NMRその他の測定手順の結果であることもあるし、相同性モデリング、タンパク質モチーフおよび保存されているドメインの解析、および／

50

またはタンパク質の折畳みまたは核酸の等価物の計算モデリングから得ることもできる。多数の膜関連標的タンパク質、例えば、GPCRおよびイオン・チャネルなどを考慮するときに利用可能なものは、モデルにより構築された構造だけであることが多い。リガンドの構造は、同様のやり方で生成することもできるし、その代わりに、リガンドが生体高分子でない場合には、物理学および化学の基本原則を用いて最初から周知の2次元化学表現から構築することもできる。

【0015】

合理的な薬物設計は、標的ーリガンドの分子間の相互作用および組合せの計算によりモデル化してリードを最適化し、所望の薬物様の生物学的かつ薬物動態学的な特性を計算により予測することまで、いくつかの計算コンポーネントのいずれかを使用することを組み込むことができる。合理的な薬物設計の状況で計算によるモデル化を利用することは、概ね、生物学的な「ウェット」ラボ試験などにおけるしばしば時間集約的であり、コストのかかる労力をかけないことによって、必要とされる時間を短縮し、薬物研究開発への集中度およびその効率を改善したいという動機に基づいている。

【0016】

リードを創製する状況での標的ーリガンドの分子的な組合せの計算によるモデル化は、分子ライブラリの大規模な *in silico* スクリーニング（すなわち、ライブラリ・スクリーニング）に関わることがある。これは、これらのライブラリが、1つまたは複数の構造データベースとして仮想的に生成され格納されているか、コンビナトリアル・ケミストリおよび有機合成によって構築されているかにかかわらず、対象とする標的分子に関する計算による生物活性の予測（または等価な尺度）に基づいて、選択されたりガンドのサブセットをランク付けする計算法を利用して行われる。

【0017】

本明細書を通じて、「結合様式」という用語は、最小結合エネルギー（すなわち、最大結合親和性）での、またはその近傍の結合状態での潜在的な分子複合体の3次元分子構造を指す。ここで、（「結合自由エネルギー」またはそれと正反対の概念である「結合親和性」と置き換えられることがある）「結合エネルギー」という用語は、潜在的な分子複合体が形成された後の分子系の自由エネルギーの変化、すなわち、リガンドおよび標的の非結合状態から（潜在的な）結合状態への移行を指す。「系のポーズ」という用語も、結合様式を指すのに用いられることがある。ここで、自由エネルギーという用語は、概ね、構成要素原子間の物理的な相互作用、および分子のそれら自体の間の結合（すなわち、分子間および分子内の相互作用）の結果、かつ、これらを取り巻く環境、すなわち、1つまたは複数の分子間の反応部位の物理的かつ化学的な周囲状況によるエンタルピーおよびエントロピーの両方の作用を指す。自由エネルギーの例は、平衡統計力学のカノニカル集合またはグランド・カノニカル集合で見られるギブスの自由エネルギーである。

【0018】

一般に、所与の標的ーリガンド対の最適結合自由エネルギーは、化学的な平衡状態にある2つの分子間での組合せまたは潜在的な分子複合体の形成の確度に直接関連しているが、実際には、この結合自由エネルギーは、単一の結合様式ではなく、（推定）複合構造の集合を記述している。ただし、計算によるモデル化では、通常、自由エネルギーの変化は、最小エネルギーに対応する単一構造によって左右されると仮定する。これは、強い結合剤（ pK が約0.1～10ナノモル）について確かに当てはまるが、弱い結合剤（ pK が約10～100マイクロモル）には疑問の余地がある。通常は、支配的な構造が結合様式であるとみなされる。場合によっては、関連する系の状態がエネルギーの点でほぼ縮退しているとき、2つ以上の代替結合様式を考慮する必要があることがある。

【0019】

結合親和性は、創薬および合理的な薬物設計に直接関わるものである。というのは、2つの分子の相互作用、例えば、生物学的なプロセスまたは経路の一部であるタンパク質と、この生物学的なプロセスまたは経路の改変を目標としようとする薬物候補との相互作用がしばしば、この薬物候補がどの程度その目的に役立つかを示す助けになるからである。

さらに、結合様式が決定可能な場合、標的に対する薬物の作用をよりよく理解することができる。このような理解は、例えば、リガンドの1つまたは複数の特徴をさらに改変して、その（標的に関する）薬効、（他の標的生体高分子に関する）結合特異性、または他の化学的かつ代謝的な特性を改善することが望ましいときに有用であることがある。

【0020】

標的分子とリガンドの親和性を測定または推定する実験室的な方法がいくつか存在する。しばしば、最初に標的を分離し、次いで、*in vitro*でリガンドと混合し、ハイスループット・スクリーニングに関連する生化学アッセイおよび機能アッセイなどで分子間相互作用を実験的に評価することがある。しかし、このような方法は、標的が分離しやすく、リガンドが製作しやすく、分子間相互作用が容易に測定される場合に最も有用であるが、標的を容易に分離することができない、あるいは、分離により生物学的なプロセスまたは病気の経路が妨げられる、あるいは、十分な量のリガンドを合成するのが難しい場合、あるいは、特定の標的またはリガンドがあらかじめ良好に特徴付けられない場合にはより問題である。後者の場合、標的とリガンドのあらゆる可能な組合せについて何千回もの、または何百万回もの実験が必要とされ、そのため、実験室的な方法を利用することは実現不可能なことがある。

【0021】

まず、標的（さらに、タンパク質ファミリーのメンバなどの関連の標的）および／またはこの標的に対する1つまたは複数の周知の天然の結合剤または基質の様々な化学的かつ生物学的な特性の特殊な知識を利用して、実験室での処理に必要とされる組合せの数を減らすことによってこの障害を解決するためのいくつかの試みがなされているが、依然としてほとんどの場合、実際的ではなく、コストがかかりすぎる。実験室の環境で実際に分子を組合せ、実験結果を測定する代わりに別の手法は、コンピュータを使用して、2つ以上の分子（すなわち、*in silico*でモデル化した分子の組合せ）間の分子間相互作用のシミュレーションまたは特徴付けを行うことである。計算による方法を用いて分子の組合せおよび相互作用を評価することは通常、構造に基づくものであるか、リガンドに基づくものであるか、あるいはその両方であるかにかかわらず、合理的な薬物設計の1つまたは複数の段階に関係する。

【0022】

所与の標的－リガンド対について、潜在的な分子の組合せの性質および／または確度を計算によりモデル化する場合、結合様式および親和性の実際の計算による予測は、慣習的に、（a）計算システムにより、リガンドと標的についての最適結合様式を予測しようと試みる「ドッキング」、および（b）計算システムにより、計算された結合様式に関連する結合親和性を推定しようと試みる「スコアリング」の2つの部分で実施される。ライブラリ・スクリーニング中に、スコアリングを利用して、標的分子に関して1つのリガンドと別のリガンドの相対的な結合親和性を予測し、それによって、これらのリガンドの優先順位を付け、また、結合の確率を割り当てることもできる。

【0023】

ドッキングは、本質的に決定論的であるか、確率論的であるかにかかわらず、探索アルゴリズムまたは機能最適化アルゴリズムに関わり、その意図は、好ましい親和性を有する1つまたは複数の系のポーズを見付けることである。

【0024】

スコアリングは、親和性関数をより精度よく推定することに関わり、ここで、親和性は、1つまたは複数の実験的な表現、分子力学に基づく表現、量子力学に基づく表現、または知識に基づく表現の組合せ、すなわち、スコアリング関数の形で表される。個々のスコアリング関数自体を組合せ、様々な式を使用してよりロバストなコンセンサス・スコアリング方式を形成する。実際、多くの異なるドッキング戦略およびスコアリング方式が、今日の計算による薬物設計の状況で採用されている。

【0025】

計算によるドッキングおよび／またはスコアリングの別の重要な適用領域は、仮想ライ

10

20

30

40

50

ブラリ・スクリーニングの範囲を超えて、*in silico*リード最適化プロセスにおけるものである。この場合、リード候補と標的の結合親和性に関して、計算によりリード候補をより厳密に検査する。大部分の生体分子は、それらが、標的の通常の結合相手、あるいは別の天然化合物またはアンタゴニスト、さらには別の薬物を含めて、同じまたは近傍の活性部位において標的タンパク質と相互作用し得る他の生体分子を凌駕する場合にのみ潜在的な薬物として適切なものとなる。ハイスループット・スクリーニングの結果、識別されるのは、通常、結合親和性がマイクロモルまたはそれよりも悪い（すなわち、 IC_{50} が約 $1 \sim 100$ マイクロモルの）リード候補のみである。リードの最適化は、1つ（または複数）のリード候補を改善または改変して、マイクロモル未満、さらにはナノモル（すなわち、 IC_{50} が約 10^{-9} ）の薬物を創製することに関わるものである。リードの最適化を行うために改変されたリードの結合親和性を推定する典型的な計算による方法には、 $QSAR$ [53、54]、 QM 、 MM 、または QM/MM によるシミュレーション [55]、摂動論を利用する系の自由エネルギーの変化の推定 [56、57]、および構造に基づく分子設計で用いる他の方法 [58] が含まれる。

10

【0026】

ただし、潜在的なリード候補が1つまたは複数の所望の標的生体高分子に良好に結合しても、すなわち、良好な生体活性を示しても、候補の分子は最終的に、代謝、毒性、望まれない副作用、宿主の分布および意図する標的部位への送達、細胞間および細胞内の輸送、ならびに排泄に関する別の厳格な要件を満たさなければならない。潜在的な薬物分子としてのリード候補の存続可能性を評価し、それによっておそらくは、市販するまでのスケジュールが短くなり、研究開発の労力の無駄が少なくなるように、計算によるモデル化を採用して、いわゆるアドメトックス ($ADME/Tox$ 、吸収-分布-代謝-排泄-毒性) プロファイル [59] が生成される。

20

【0027】

*in silico*でアドメトックス・プロファイルを生成する際に、多くの尺度が関わっている。これらの尺度には、生物濃縮性、バイオ・アベイラビリティ、代謝、 pK_a 、発癌性、突然変異性、 $\log D$ 、 n -オクタノール/水分配係数または $\log P$ 、水溶性、血液脳関門その他の膜に関する透過性、腸内吸収、皮膚感受性、さらには、他の周知の薬物または有機化合物との化学的かつ構造的な類似性についての実験的に導かれた予測または推定が含まれる。ある種の尺度は、生体分子の物理的またはエネルギー的な特性、例えば、 pK_a 、 $\log D$ 、 $\log P$ などのより正確な数値計算または数値予測に関わるものである [60]。他の尺度は、典型的には、知識またはルールに基づく手法、例えば、既知の生物学的な経路および形質転換に基づく代謝特性の予測、リビンスキーの実験的な「5の法則 (*rule of five*)」に基づく分布特性の定性的な予測 [61]、およびケモインフォマティクスの知識データベースを利用して発癌性、突然変異性、奇形発生性、皮膚感受性などを評価する毒性の予測に関わるものである。他の尺度は、例えば、血液脳関門に関して透過性である既知の生体分子、あるいは、肝臓または腎臓などの様々な器官の機能におそらくは悪影響を及ぼす既知の生体分子との化学的かつ構造的な類似性に基づくものである。

30

【0028】

しかし、これらの計算ツールがすべて揃っていても、人間（または他の）被験体を扱う臨床前試験および臨床治験における安全性および薬効の審理ために失敗する率が依然として高く、コストおよび市場機会を失うことの両方の点で不安定である。これらの失敗の主な理由は、所望の標的に結合する特異性が欠如しているために、宿主生物内でリード候補と（しばしば他の生体高分子である）他の生体分子との間に有害な交差反応が存在することである。これらの有害な交差反応によりしばしば、薬物候補の薬効が低減し、望まれない副作用が生じ、さらには、病気または死に至ることがある。

40

【0029】

例えば、リード候補は、所望の標的タンパク質を抑制すると同時に、遺憾ながら、肝臓内で1つまたは複数のタンパク質あるいは細胞表面受容体にも結合し、それによって、化

50

学的な不均衡が生じ、さらには、肝硬変などの深刻な病状が生じることがある。別の例では、リード候補は、宿主生物の免疫系によって病原体その他の抗原と誤って識別されることがあり、それによって、アレルギー反応その他の免疫疾患が生じることがある。皮肉にも、臨床治験に合格し承認された多くの市販薬でさえ、しばしば、不眠症、吐き気、脱水症、勃起障害、および眠気などの様々な望まれない副作用を伴う。抗癌薬または強力な抗細菌作用物質などの他の薬物は、例えば、様々な癌に用いる化学治療薬、H I V 複合製剤など、基となる病気がそうしなければ致死性であるために医学会でしか許容されないはるかに深刻な副作用を有する。

【0030】

潜在的な有害交差反応についてこのように前もって知識が得られれば、臨床前試験および臨床治験だけでなく、望まれない副作用がより少なく、薬効がより高い薬物候補を選択する際に関わる時間およびコストを削減することができ、それによって、その薬物の商業化に直接的な利益がもたらされることになる。

【0031】

2003年9月30日現在、タンパク質データ・バンク(PDB)のウェブ・サイト(<http://www.rcsb.org/pdb/>)で公表されている情報によれば、20,501種類のタンパク質構造、948種類のタンパク質/核酸複合体、ならびに1233種類の核酸(および18種類の炭化水素)の合計22,700種類の巨大分子構造がPDBにあり、これらが約3年で倍になる率で増加すると予測されている[62]。これらの約83%はX線回折法で得られ、15%はNMRで得られ、残りの部分は他の実験的な方法で得られる。ほぼ20000種類のタンパク質構造が、1000種類を超える個々の種由来のものであり、そのうち、4767種類がヒト由来のものであり、2411種類が大腸菌由来のもの、1030種類がマウス由来のものなどである。一方、ほぼ1900種類が合成によるものであり、追加の4658種類の実体については、種が明確に同定されていない。種ごとの完全な内訳については、<http://www.biochem.ucl.ac.uk/bsm/pdbsum/species/>を参照されたい。PDBで利用可能なタンパク質構造情報のさらなる解析および分類は、[65]および[66]に出ている。ヒト以外のタンパク質についてPDBで利用可能なタンパク質構造はしばしば、特に、マウス、ブタ、ニワトリ、線虫、酵母菌などの他の生物由来の多くのタンパク質が、構造および機能の両方の点でヒトのタンパク質に相同であることを考慮すると、薬物設計のための豊富な情報を有する。さらに、ある種の病気の標的は、ウィルスまたは細菌のタンパク質など、ヒト以外のタンパク質に関わるものである。構造プロテオミクスおよび機能ゲノミクスにおける新しい進歩により、利用可能な質の高い機能注釈付き3次元タンパク質構造の数が増え続けている。また、相同性モデル化、タンパク質モチーフ、およびタンパク質スレッディングにおける別の進歩により、実験により得られたタンパク質および核酸の3次元構造が継続して補完されている。そのため、多数の薬物候補について良好に特徴付けられた潜在的な交差反応物質の一覧は既にかんがりのものになりつつあり、今後、増え続ける一方であろう。

【0032】

本発明より以前は、コンピュータ・ベースのシステムで、リード分子(すなわち、潜在的な薬物候補)の有害反応を正確かつ効果的に計算する体系的な方法はなかった。

【0033】

近年、「逆ドッキング」と称する方法では、はるかに限定された範囲の問題、すなわち、単一の薬物様分子についての代替標的の識別に注目しており、これは、チェン、ワイ・ゼット、ジー、ディー・ジー、「Ligand-Protein Inverse Docking and Its Potential Use in the Computer Search of Protein Targets of a Small Molecule」、Proteins、第43巻、217~226頁、2001年、さらに、チェン、ワイ・ゼット、ジー、ディー・ジー、ウン、シー・ワイ、「Computational Method for Drug Target Search a

nd Application in Drug Discovery」、Journal of Theoretical and Computational Chemistry、第1巻、第1号、213～224頁、2002年に記載されている（以下、「チェン他」と称する）。これらをすべて全体として、参照によりここに組み込む。この限定された範囲でさえ、これらの方法は、これから説明するように、大きな欠点を有する。チェン他の方法は、概ねUCSFのソフトウェアであるドッキング・ツールDOCK [5、6、7]に関連する「sphgen」アルゴリズムと、既知の薬物候補と潜在的な代替標的の相互作用を評価するワン他 [33]に記載のハイブリッド・ドッキング法とを組み合わせたものを本質的に利用する極めて特殊な幾何学アルゴリズムに依存している。当技術分野では、このような手法は、ほとんどのリガンドについて、結合エネルギーの正確な測定はおろか、正確な結合様式を予測する際にあまり当てにならないことが既の実証されている。

10

【0034】

このような方法は、結合エネルギーを簡単な分子力学により予測することに依拠しており、これは、当技術分野では、ほとんどの系について自由エネルギーの変化を誤差 $\pm 3 \text{ kcal/mol}$ で予測するために知られているものである [34～39]。2 kcal/mol の差でさえ、分子系の解離定数または阻害定数のほぼ30倍の差になる。そのため、チェン他の方法が、系の結合エネルギーの絶対変化を良好に推定すると期待することはできない。このことは実際に、著者ら自身が公表したデータを厳密に調べることによって裏付けられる。チェン他（2002年）の表1では、9個の系のうち4個（1hvr、4phv、1dhf、3cpa）の結合自由エネルギーの実験値が一般に利用可能である（<http://www-mitchell.ch.cam.ac.uk/pld/energy.php>で公開されており、それぞれ、 -13.13 、 -12.86 、 -3.08 、および -5.39 kcal/mol である）。これは、それぞれ -70.2 、 -94.51 、 -48.67 、および -40.63 kcal/mol の報告されている予測値と全く対照的である。さらに、1hvrおよび4phvのpdbエントリのリガンドは、同じHIV-1プロテアーゼ標的タンパク質に対して類似の結合親和性を有することが実験的に測定されているが、対応する予測値は、34%よりも大きく外れており、かなりの誤差がある。

20

【0035】

さらに、チェン他の方法は、潜在的な標的タンパク質の関連する活性部位をあらかじめ正確に特徴付けることに依拠している。このような事前情報は、タンパク質については、それらの天然結合剤またはアゴニスト／アンタゴニストに関係してしばしば入手可能であるが、様々な生体高分子と、他の特定の標的タンパク質について設計されたリード候補との間の潜在的な交差反応を調べるときには当てはまらないことが多い。

30

【0036】

さらに、チェン他の方法では、リガンド原子の数の多項式関数を、これらリガンド原子のドッキング・プロトコルによってpdb構造のサブセットに対して生成された結果に合わせることに基づく実験的な閾値を用いて、ある種の偽陽性候補を除外する。リガンド原子の数に基づくこのような実験的な尺度が、現在までに集計された結合自由エネルギーの実験観察データと良好に相関しないことが当業者には理解されよう。

40

【0037】

最後に、上記で列挙した理由から、チェン他の方法は、潜在的候補と少なくとも1つの他の「競合する」リガンドとの複合体を扱う少なくとも1つのpdbエントリが存在することを必要とし、それによって、問題のリガンドについての結合親和性の予測をより良好に校正して、その適用性をさらに限定する。しかし、pdb構造の多くは、関連する結合リガンドを含まず、他の多くは、リード候補に関連し得る部位と全く異なる活性部位でリガンドに結合している。さらに、多くの系では、標的タンパク質は、非結合状態から結合状態に移行するときに大きな配座変化、すなわち、「誘導適合」を受ける。そのため、受容体の柔軟性を適切にモデル化する仕組みが欠如している場合、結合配座状態を利用する

50

ことは多くの系では不適切である。カルボキシペプチダーゼ A と複合体を作る様々な小型ペプチドおよびカルモジュリンとタモキシフェンの複合体の例があり、これらはいずれも、チェン他の結果の一覧に含まれるものである。これらの要因があいまって、チェン他が公開した方法によってスクリーニングし得る潜在的な標的の範囲が大きく制限されることがある。

【0038】

チェン他の方法は、複数の潜在的なタンパク質標的に対して所与の薬物候補を高速に処理するために、予測の精度および確かさをともに犠牲にする。これは、リード候補の追加の 2 次治療標的を「探り出す」ことを意図している。あらかじめ特徴付けられた関連活性部位において少なくとも 1 つの競合リガンドと潜在的な標的が複合するという要件により、代替標的生体高分子の大きな集合に対して潜在的な望まれない副作用および他の有害な交差反応を識別することに関してこれらの候補が極めて魅力のないものになっている。

10

【0039】

さらに、後でより詳細に論じるように、現在までの先行技術は、有害な交差反応を予測し、創薬プロセスの改善に適用することに関する他の 3 つの極めて重要な問題に対処していない。第 1 の問題は、公共ドメインおよび所有権下にあるドメインに存在する巨大分子標的についての機能、ソース、他の構造との相同性、および系統分類に関する豊富な注釈情報を利用して、潜在的な有害交差反応についてよりよく判定する方法である。第 2 の問題は、リード分子の集合の有効な比較評価を、それらの潜在的な交差反応プロファイルに基づいて生成する方法である。第 3 の問題は、結合特異性が欠如しているために 1 つまたは複数の潜在的な有害交差反応を示すリード候補を再構築または再設計して、所望の治療標的との結合と潜在的に有害な交差反応物質との結合をいかに大きく区別するかを推測ために、高度な計算によるモデル化の結果を用いて、生体分子を臨床治験で存続可能な薬物候補にする方法である。

20

【0040】

創薬プロセスを大きく改善する際の潜在的な交差反応の事前知識の重要性は既に上記で説明した。このような知識は、明らかに、副作用が少なく、低コストで、臨床試験に合格する確度がより高いよりよい薬物を設計する助けとなる。本発明では、この極めて重要な要求に対処するシステムおよび方法を説明する。

【0041】

参照文献および先行技術

30

本発明の分野における先行技術は大量に文書化されている。以下、それらを概観することを試みる。

【0042】

ドゥルース [1] は、創薬の現況をうまく概観している。[2] では、アバギャンおよびトトロフは、ハイスループットのドッキングおよびスコアリングならびにその応用例の状況を示している。ラム他 [3] はさらに、あるタンパク質ファミリーに対して複数のコンビナトリアル・ライブラリを設計し、ドッキングし、仮想的にスクリーニングする一般手法を教示している。最後に、ワスコイチ他 [4] は、複数のコンピュータを使用し、特定の標的に対する大きなリガンド・ライブラリの仮想スクリーニングを、リガンドを群にして特性のコンピュータに割り当てることによって加速することを記載している。

40

[1] ジェイ・ドゥルース、「Drug Discovery: A Historical perspective」、Science、第 287 巻、1960~1964 頁、2000 年

[2] ルーベン・アバギャン、マキシム・トトロフ、「High-throughput docking for lead generation」、Current Opinion in Chemical Biology、第 5 巻、375~382 頁、2001 年

[3] ラム、エム・エル、バーディック、ケイ・ダブリュー、トバ、エス、ヤング、エム・エム、スキルマン、エイ・ジー他、「Design, docking, and eva

50

luation of multiple libraries against multiple targets」、Proteins 第42巻、296～318頁、2001年

[4] ワスコイチ、ビー、パーキンス、ティー・ディー・ジェイ、サイクス、アール・エイ、リー、ジェイ、「Large-scale virtual screening for discovering leads in the post-genomic era」、IBM Systems Journal、第40巻、第2号、2001年

【0043】

ドッキング・シミュレーションを実施するのに現在使用されているソフトウェア・ツールの例がいくつかある。これらの方法は、a) 幾何学的ハッシュ法、ポーズ・クラスタリングを利用する表面相関またはグラフ・パターン・マッチングに基づく剛体パターン・マッチング・アルゴリズム、b) 段階的構成 (incremental construction) または「配置し結合する」演算子を含めて断片に基づく方法、c) モンテ・カルロ法、シミュレーテッド・アニーリング、または遺伝子 (またはミーム) アルゴリズムを用いることを含めて確率最適化法、d) 分子動力学シミュレーション、またはe) これらから導出したハイブリッド戦略を利用することを含めて、広範な計算技術に関わるものである。

【0044】

最初のドッキング・ソフトウェア・ツールは、1982年にUCSFで開発され (バージョン1.1)、現在はバージョン5.0までである (拡張機能は、段階的構成を含む) DOCKと呼ばれるグラフに基づく剛体パターン・マッチング・アルゴリズムであった [5、6、7]。グラフに基づくパターン・マッチング・アルゴリズムの他の例には、(GRID [9]、FLOG [10]、およびLIGIN [11] を使用する) CLIX [8] が含まれる。

[5] ショイチェット、ビー・ケイ、ボディアン、ディー・エル、クンツ、アイ・ディー、「Molecular docking using shape descriptors」、J. Comp. Chem.、第13巻、第3号、380～397頁、1992年

[6] メン、イー・シー、シュベント、ディー・エイ、ブラニー、ジェイ・エム、アイ・ディー・クンツ、「Orientational sampling and rigid-body minimization in molecular docking」、Proteins: Structure, Function, and Genetics、第17巻、266～278頁、1993年

[7] ユーイング、ティー・ジェイ・エイ、クンツ、アイ・ディー、「Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening」、J. Computational Chemistry、第18巻、第9号、1175～1189頁、1997年

[8] ローレンス、エム・シー、デイビス、ビー・シー、「CLIX: A Search Algorithm for Finding Novel Ligands Capable of Binding Proteins of Known Three-Dimensional Structure」、Proteins、第12巻、31～41頁、1992年

[9] カステンホルツ、エム・エイ、バスター、エム、クルシアーニ、ジー、ハークスマ、イー・イー・ジェイ、フォックス、ティー、「GRID/CPCA: A new computational tool to design selective ligands」、J. Medicinal Chemistry、第43巻、3033～3044頁、2000年

[10] ミラー、エム・ディー、カースリー、エス・ケイ、アンダーウッド、ディー・ジ

10

20

30

40

50

エイ、シェリダン、アール・ピー、「FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure」、J. Computer-Aided Molecular Design、第8巻、第2号、153～174頁、1994年

[11] ソボレフ、ブイ、ウェイド、アール・シー、ブレンド、ジー、エデルマン、エム、「Molecular docking using surface complementarity」、Proteins、第25巻、120～129頁、1996年。他の剛体パターン・マッチングによるドッキング・ソフトウェア・ツールには、形状に基づく相関法であるFTDOCK [12] およびHEX [13]、幾何学的ハッシュ法であるFischer他 [14]、またはポーズ・クラスタリング法であるレアリー他 [15] が含まれる。

[12] カチャルスキー、カツィール、イー、シャリフ、アイ、アイゼンスタイン、エム、フリーセム、エイ・エイ、アフラロー、シー、バクサー、アイ・エイ、「Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques」、Proceedings of the National Academy of Sciences of the United States of America、第89巻、第6号、2195～2199頁、1992年

[13] リッチー、ディー・ダブリュー、ケンプ、ジー・ジェイ・エル、「Fast Computation, Rotation, and Comparison of Low Resolution Spherical Harmonic Molecular Surfaces」、J. Computational Chemistry、第20巻、第4号、383～395頁、1999年

[14] フィッシャー、ディー、ノレル、アール、ウルフソン、エイチ、ヌシノフ、アール、「Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition」、Proteins、第16巻、278～292頁、1993年

[15] レアリー、エム、ウェフィング、エス、レンガウアー、ティー、「Placement of medium-sized molecular fragments into active sites of proteins」、J. Computer-Aided Molecular Design、第10巻、41～54頁、1996年

【0045】

一般に、剛体パターン・マッチング・アルゴリズムは、標的およびリガンドがともに剛体であり（すなわち、可撓性でない）、したがって、小型の剛体分子（または分子の断片）が、簡単なタンパク質に、ほぼ剛体の明確な活性部位でドッキングするのに適したものであり得ると仮定する。そのため、このクラスのドッキング・ツールは、de novoリガンド設計、コンビナトリアル・ライブラリ設計、または1つのリガンド当たり複数の配座異性体を含む分子ライブラリの直裁的な剛体スクリーニングに適していることがある。

【0046】

段階的構成に基づくドッキング・ソフトウェア・ツールには、（EMBLからライセンスを受けた）Tripos社のFlexX [16、17]、Hammerhead [18]、（任意選択として）DOCKバージョン4.0 [7]、およびリーチ他のものぐさバックトラッキング・アルゴリズム [19] が含まれる。de novoリガンド設計の状況で段階的構成を利用するプログラムには、（Accelrys社の）LUDI [20] およびGrowMol [21] が含まれる。「配置し結合する」戦略に基づくドッキング

・ソフトウェア・ツールには、デジャレー他[22]が含まれる。

[16] クレイマー、ビー、レアリー・エム、レンガウアー、ティー、「Evaluation of the FlexX incremental construction algorithm for protein-ligand docking」、Proteins、第37巻、228～241頁、1999年

[17] レアリー・エム、クレイマー、ビー、レンガウアー・ティー、クリーブ、ジー、「A Fast Flexible Docking Method Using An Incremental Construction Algorithm」、J. Mol. Biol.、第261巻、470～489頁、1996年

[18] ウェルチ、ダブリュー、ルーパー、ジェイ、ジェイン、エイ・エヌ、「Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites」、Chemical Biology、第3巻、449～462頁、1996年

[19] リーチ、エイ・アール、クンツ、アイ・ディー、「Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites」、J. Comp. Chem.、第13巻、730～748頁、1992年

[20] ボーム、エイチ・ジェイ、「The computer program LUDI: a new method for the de novo design of enzyme inhibitors」、J. Computer-Aided Molecular Design、第6巻、61～78頁、1992年

[21] ボハチェク、アール・エス、マクマーティン、シー、「Multiple Highly Diverse Structures Complementary to Enzyme Binding Sites: Results of Extensive Application of a de Novo Design Method Incorporating Combinatorial Growth」、J. American Chemical Society、第116巻、5560～5571頁、1994年

[22] デジャレー、アール・エル、シェリダン、アール・ピー、ディクソン、ジェイ・エス、クンツ、アイ・ディー、ベンカタラガバン、アール、「Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape」、J. Med. Chem.、第29巻、2149～2153頁、1986年

【0047】

段階的構成アルゴリズムを利用して、良好に特徴付けられた活性部位を有する剛体標的分子への可撓性リガンドのドッキングをモデル化することができる。このようなアルゴリズムは、1つまたは複数の標的に対して可撓性リガンドのライブラリをスクリーニングするときに用いられる。このようなアルゴリズムは、比較的計算集約的ではなく、その結果、確率最適化に基づく多くの競合アルゴリズムよりも精度が低いことが多い。しかし、FlexXでさえ、標的-リガンドの1つの組合せを処理するのに1～2分未満程度かかることがあり、そのため、依然として、ライブラリのサイズによっては（例えば、数千万以上の化合物の場合）計算が厄介なことがある。段階的構成アルゴリズムではしばしば、1つまたは複数のスコアリング関数を使用して、計算中に出くわす異なる系のポーズの値を求め、ランク付けする。最近、FlexXは、ある種の活性部位回転異性体のユーザ定義集合を用いることによって、標的分子の活性部位の部分的な可撓性を補償することを試みるために、FlexE[23]に拡張された。

[23] クラウセン、エイチ、ブニング、シー、レアリー・エム、レンガウアー・ティー、「FlexE: Efficient Molecular Docking Considering Protein Structure Variations」、J. Molecular Biology、第308巻、377～395号、2001年

【0048】

確率最適化に基づく計算によるドッキング・ソフトウェア・ツールには、(MolSoft社の)ICM[24]、(Schrodinger社の)GLIDE[25]、および(Accelrys社の)LigandFit[26](これらはすべて、変形モンテ・カルロ技法に基づくものである)、ならびにシミュレーテッド・アニーリングに基づく(Scripps Instituteの)AutoDockバージョン2.5[27]が含まれる。遺伝子アルゴリズムまたはミーム・アルゴリズムに基づく他のものには、GOLD[28、29]、DARWIN[30]、および(やはりScrippsの)AutoDockバージョン3.0[31]が含まれる。

[24] アバガン、アール・エイ、トトロフ、エム・エム、クツネットソフ、ディー・エヌ、「Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins」、J. Comp. Chem.、第15巻、488～506頁、1994年

[25] ハルグレン、ティー・エイ、マーフィー、アール・ビー、フリースナー、アール・エイ、ベアード、エイチ・エス、フライ、エル・エル、ポラード、ダブリュー・ティー、バンクス、ジェイ・エル、「Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening」、J. Med. Chem.、第47巻、第7号、1750～1759頁、2004年

[26] ルーティ、ビー・エイ、ワッサーマン、ゼット・アール、スタウテン、ピー・エフ・ダブリュー、ホッジ、シー・エヌ、ザチャリアス、エム、マッカモン、ジェイ・エイ、「Molecular Mechanics/Grid Method for the Evaluation of Ligand-Receptor Interactions」、J. Comp. Chem.、第16巻、454～464頁、1995年

[27] グッドセル、ディー・エス、オルソン、エイ・ジェイ、「Automated Docking of Substrates to Proteins by Simulated Annealing」、Proteins: Structure, Function, and Genetics、第8巻、195～202頁、1990年

[28] ジョーンズ、ジー、ウィレット、ピー、グレン、アール・シー、「Molecular Recognition of Receptor Sites using a Genetic Algorithm with a Description of Desolvation」、J. Mol. Biol.、第245巻、43～53頁、1995年

[29] ジョーンズ、ジー、ウィレット、ピー、グレン、アール・シー、リーチ、エイ、テイラー、アール、「Development and Validation of a Genetic Algorithm for Flexible Docking」、J. Mol. Biol.、第267巻、727～748頁、1997年

[30] テイラー、ジェイ・エス、バーネット、アール・エム、Proteins、第41巻、173～191頁、2000年

[31] モリス、ジー・エム、グッドセル、ディー・エス、ハリディ、アール・エス、ヒューイ、アール、ハート、ダブリュー・イー、ブリュー、アール・ケイ、オルソン、エイ・ジェイ、「Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function」、J. Comp. Chem.、第19巻、1639～1662頁、1998年

【0049】

確率最適化に基づく方法を用いて、可撓性のリガンドの標的分子へのドッキングをモデル化することができる。このような方法では一般に、分子力学に基づく親和性関数の定式化を利用し、様々な戦略を採用して、有利な系の1つまたは複数のエネルギー極小値を探

す。このような方法は、より計算集約的であるが、それに伴って競合する段階的構成による方法よりもロバスト性が高いことが多い。このような方法は本質的に確率論的なので、実行するたびに、またはシミュレーションするたびに、しばしば異なる予測値が得られることがある。従来、確率最適化法を利用する大部分のドッキング・ソフトウェア・ツールは、標的がほぼ剛体である（すなわち、活性部位における水素結合のドナー基とアクセプタ基は回転し得る）と仮定する。というのは、そうしないと、組合せによる複雑さが急激に増大し、そのため、適度な時間内に問題をロバストに解くことが難しくなるからである。

【0050】

標的ーリガンドの組合せを計算によりモデル化する状況では、分子動力学シミュレーションも用いられている。これには、（モンテ・カルロ法とともに）ディフォーラ他[32]およびルーティ他[16]に示す実施形態が含まれる。原理的に、分子動力学シミュレーションは、タンパク質の可撓性を任意の程度にモデル化することができる。一方、分子動力学シミュレーションは、多数のきめ細かい時間ステップの値を求めることを必要とし、そのため、極めて時間集約的であることが多い（1つの標的ーリガンドの組合せ当たり数時間程度、さらには数日程度）。分子動力学シミュレーションは、しばしば、有効な経路を選択するのに使用者との対話も必要とする。したがって、リードを発見するのに分子動力学シミュレーションを用いることが適しているのは、小数の見込みのあるリード候補を扱う予測される複合体を極小にする場合である。

[32] ディフォーラ、エイ、ベレンドセン、エイチ・ジェイ・シー、ロッカターノ、ディー、「Molecular Dynamics Simulation of the Docking of Substrates to Proteins」、Proteins、第19巻、174～182頁、1994年

【0051】

ハイブリッド法は、選択した低エネルギー・リガンドの配座を高速にスクリーニングするために剛体パターン・マッチング技法の利用を必要とすることがあり、その後、モンテ・カルロ法によって残ったポーズの捻れ最適化を行い、最後にさらに、いくつか選択した（潜在的に）可撓性のタンパク質活性部位と組み合わせたリガンド構造に分子動力学による改善を加える。このタイプのドッキング・ソフトウェア戦略の例は、ワン他[33]である。

[33] ワン、ジェイ、コールマン、ピー・エイ、クンツ、アイ・ディー、Proteins、第36巻、1～19頁、1999年。標的ーリガンドの親和性を推定し、ライブラリ・スクリーニングにより異なるリガンドの優先順位を付け、また、結合様式を予測するために中間ドッキング・ポーズをランク付けるのに用いられる、ソフトウェアの形で実施されるスコアリング関数の例がいくつかある。従来、スコアリング関数は、a) 実験的なスコアリング関数、b) 分子力学に基づく表現、または、c) 知識に基づくスコアリング関数、あるいはこれらから導出されるハイブリッド方式の3つの別個の範疇に分かれる。

【0052】

（標的ーリガンドの組合せに適用される）実験的に導出されるスコアリング関数は、当初、しばしばQSARの研究で用いられる自由エネルギーの線形な関係から考え出された。初期の例は、（LUDIで使用される）ベーム他[20、34]のものである。他の実験的なスコアリング関数には、（FlexXで使用される）SCORE[35]、ChemScore[36]、PLP[37]、Fresno[38]、およびGlideScoreバージョン2.0+（GLIDEが使用するChemScoreの改変形態）[39]が含まれる。

[34] ベーム、エイ・ジェイ、「The Development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure」、J. Comput-Aided Mol. Des.、第8巻、2

10

20

30

40

50

43～256頁、1994年

[35] ワン、アール、ガオ、ワイ、ライ、エル、「A new empirical method for estimating the binding affinity of a protein-ligand complex」、J. Molecular Modeling、第4巻、379頁、1998年

[36] エルドブリッジ、エム・ディー、マリー、シー・ダブリュー、オートン、ティー・アール、パオリニ、ジー・ブイ、ミー、アール・ピー、「Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes」、J. Computer-Aided Molecular Design、第11巻、425～445頁、1997年

[37] ゲルハー、ディー・ケイ、ブージダ、ディー、レイト、ピー・エイ、「Rational Drug Design: Novel Methodology and Practical Applications」、Parrill, L.、Reddy, M. R. 監修、American Chemical Society、米国ワシントン D. C.、292～311頁、1999年

[38] ログナン、ディー、ローエモラー、エス・エル、ホルム、エイ、ブース、エス、シンケ、ブイ・ジェイ、Medicinal Chemistry、第42巻、4650～4658頁、1999年

[39] ハルグレン、ティー・エイ、マーフィー、アール・ピー、フリースナー、アール・エイ、ベアード、エイチ・エス、フライ、エル・エル、ボラード、ダブリュー・ティー、バンクス、ジェイ・エル、「Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening」、J Med Chem.、第47巻、第7号、1750～1759頁、2004年

【0053】

一般に、実験的なスコアリング関数は、特に、大きな化合物ライブラリをスクリーニングする状況では、今日使用されているスコアリング関数の大部分を含む。基本的な前提は、実験的なエネルギー・モデルにそれぞれ関連する数値的な重みを掛け、（いわゆる）「マスタ・スコアリング計算式」で表される1組の相互作用成分の1つを表すようにし、これらの1次結合を較正することである。この場合、前記計算式は、分子の組合せの結合自由エネルギーを良好に近似しようと試みる。これらの数値的な重み係数は、1組の学習用の標的ーリガンド複合体用に構成された実験による結合自由エネルギー・データに合わせることで得ることができる。

【0054】

分子力学に基づくスコアリング関数は、当初、AMBER [40、41]、OPLS [42]、MMFF [43]、およびCHARMM [44]のように、分子力学の力場の状況で分子モデリングに利用するために開発された。分子力学に基づくスコアリング関数の例には、化学的かつエネルギーに基づくスコアリング関数である（AMBERに基づく）DOCKバージョン4.0 [7]、GOLD [28、29]、（実験による重みを伴う）AutoDockバージョン3.0 [31]、およびFLOG [10]で使用される目的関数がともに含まれる。

[40] パールマン、ディー・エイ、ケース、ディー・エイ、コールドウェル、ジェイ・シー、ロス、ダブリュー・エス、チャーザムIII、ティー・イー、ファーガソン、ディー・エム、ザイベル、ジー・エル、シン、ユー・シー、ウィナー、ピー、コールマン、ピー・エイ、AMBER 4.1、University of California, San Francisco、1995年

[41] コーネル、ダブリュー・ディー、シーブラック、ピー、ベイリー、シー・アイ、グールグ、アイ・アール、メルツ、ケイ・エム、ファーガソン、ディー・エム、スペルマ

イヤー、ディー・シー、フォックス、ティー、コールドウェル、ジェイ・ダブリュー、コールマン、ピー・エイ、「A second-generation force field for the simulation of proteins, nucleic acid, and organic molecules」、J. American Chemical Society、第117巻、5179～5197頁、1995年

[42] ヨルゲンセン、ダブリュー・エル、ティラドーリベス、ジェイ、J. American Chemical Society、第110巻、1657～1666頁、1988年

[43] ハルグレン、ティー・エイ、「Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94」、J. Comp. Chem.、第17巻、490～519頁、1996年

[44] ブルックス、ビー・アール、ブルッコレリ、アール・イー、オルソン、ビー・ディー、ステイツ、ディー・ジェイ、スワミナサン、エス、カープラス、エム、「CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations」、J. Comp. Chem.、第4巻、187～217頁、1983年

【0055】

一般に、分子力学に基づくスコアリング関数は、多くの確率最適化に基づくドッキング・プログラムで使用される目的関数に極めて類似していることがある。このような関数は、典型的には、分子力学の1つまたは複数の力場（例えば、AMBER、MMFF、OPLSなど）に基づいて、様々な属性（例えば、電荷、質量、vdW半径、結合平衡定数など）の原子（または化学基）レベルのパラメータ化を必要とする。場合によっては、他の分子モデリング・ソフトウェア・パッケージの利用法に基づいて、リガンドについての関連するパラメータを割り当てることもできる。例えば、MOPAC [45]、AMPAC [46]、またはAMSOL [47]を利用することによってリガンドの部分的な電荷が割り当てられる。これらのパラメータは、分子内相互作用（すなわち、分子の自己エネルギー）、ならびに静電気などの遠距離相互作用も含み得る。場合によっては、リガンド-標的試験複合体を再生するのに最適化された数値的な重みによって、やはりエネルギー項の組合せを実現することができる。

[45] スチュワート、ジェイ・ジェイ・ピー、「Quantum Chemistry Program Exchange」、第10巻、86頁、1990年

[46] リオタード、ディー・エイ、ヒーリー、イー・エフ、ルイス、ジェイ・エム、デュワー、エム・ジェイ・エス、「Quantum Chemistry Program Exchange-no. 506」、QCPE Bulletin、第9巻、123頁、1989年

[47] AMSOL-version 6.5.1、ジー・ディー・ホーキンス、ディー・ジェイ・ギーセン、ジー・シー・リンチ、シー・シー・チェンバース、アイ・ロッシ、ジェイ・ダブリュー・ストーラー、ジェイ・リー、ディー・リナルディ、ディー・エイ・リオタード、シー・ジェイ・クレイマー、ディー・ジー・トルラー、University of Minnesota, Minneapolis、1997年

【0056】

知識に基づくスコアリング関数は、当初、液体をモデル化するための平均力統計力学法のポテンシャルによって考え出された。例には、DrugScore [48]、PMF [49]、およびBLEEP [50]が含まれる。

[48] ゴールケ、エイチ、ヘンドリック、エム、クリーブ、ジー、「Knowledge-based Scoring Function to Predict protein-Ligand Interactions」、J. Mol. Biol.、第295巻、337～356頁、2000年

[49] ミューゲ、アイ、マーティン、ワイ・シー、「A general and fast scoring function for protein-ligand interactions—a simplified potential approach」、J. Med. Chem.、第42巻、791～804頁、1999年

[50] ミッチェル、ジェイ・ビー・オー、ラスコウスキー、アール・エイ、アレックス、エイ、ソートン、ジェイ・エム、「BLEEP—Potential of Mean Force Describing Protein-Ligand Interactions II. Calculation of Binding Energies and Comparison with Experimental Data」、J. Comp. Chem.、第20巻、1165～1176頁、1999年

10

【0057】

一般に、知識に基づくスコアリング関数では、親和性関数を分割する必要はないが、関連する分子複合体の3次元構造の大きなデータベースを使用する必要がある。通常は、結合親和性が実験的に既知である分子複合体の1組のデータに回帰する必要もない。これらの方法は、所与の間隔の2つの原子の相互作用が有利なほど、無秩序なバルク媒体内では予想値よりも相互作用が頻繁に発生するという基礎仮定に基づいている。このような方式は、「逆ボルツマン」方式と称することがあるが、実際には、巨大分子における局所的な最適化された構造およびタンパク質の折畳みの存在は、距離に依存するペアワイズ・プリファレンス分布が厳密にボルツマンである必要がないことを意味する。例えば、溶媒和作用を近似するための溶媒接触可能表面積など、他の分子記述子に基づくシングレット・プリファレンスの概念を導入することも可能である。

20

【0058】

ハイブリッド・スコアリング関数は、異なる型の1つまたは複数のスコアリング関数を合成したものとし得る。一例は、分子力学／実験的ハイブリッド関数であるVALIDATE [51] である。スコアリング関数の他の組合せは、分子の各組合せごとに複数の関数の値を求めることができ、1組のルールまたは統計的基準、例えば、各スコアリング関数の上位10%の順位一覧で生じる（共通部分に基づく）状態、平均順位が高い（平均に基づく）状態などに基づいて何らかの形式の「コンセンサス」決定がなされるコンセンサス・スコアリングの概念を含み得る。コンセンサス・スコアリングの考察の有用な検討は、ピサント他 [52] に出ている。

30

[51] ヘッド、アール・ディー、スミス、エム・エル、オブリー、ティー・アイ、ウォーラー、シー・エル、グリーン、エス・エム、マーシャル、ジー・アール、「VALIDATE: A New Method for Receptor-Based Prediction of Binding Affinities of Novel Ligand」、J. American Chemical Society、第118巻、3959～3969頁、1996年

[52] ピサント、シー、フォルカーズ、ジー、ログナン、ディー、「Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations」、J. Med. Chem.、第43巻、4759～4767頁、2000年

40

【0059】

計算リソースの制約のためにリードの最適化中に特に使用する計算による他の方法には、QSAR（定量的構造活性相関）、QM/MMシミュレーション、および自由エネルギー摂動論が含まれる。これらのうち後者2つが、結合様式が既に良好に決定されている場合には、様々な複合体について結合親和性をより高精度に予測するにはより適切である。構造に基づく設計の状況でリードの最適化に適切な上記その他の方法には、ボハチェック他 [58] で検討されているものが含まれる。

[53] 「3D QSAR in Drug Design: Recent Advances」、(第1～3巻)、ヒューゴ・クビニー（監修）、ゲルト・フォルカーズ（監修

50

)、イボンヌ・シー・マーティン (監修)

[54] 「Exploring QSAR: Hydrophobic, Electronic, and Steric Constants (ACS Professional Reference Book)」、コーウィン ハンシュ (監修)、アルバート・レオ (監修)、ディー・エイチ・フークマン (監修)

[55] マーフィー、アール・ビー、フィリップ、ディー・エム、フリースナー、アール・エイ、「Frozen Orbital QM/MM methods for density functional theory」、Chem. Phys. Letters、第321巻、113~120頁、2000年

[56] アクビスト、ジェイ、メディーナ、シー、サミュエルソン、ジェイ・イー、「A new method for predicting binding affinity in computer-aided drug design」、Prot Eng、第7巻、385頁、1994年

[57] ジョーンズ・ハーツォグ、ディー・ケイ、ヨルゲンソン、ダブリュー・エル、「Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method」、J Med Chem、第40巻、1539頁、1997年

[58] ボハチェク、アール・エス、マクマーティン、シー、グイダ、ダブリュー・シー、「The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective」、Medicinal Research Reviews、第16巻、3頁、1996年

【0060】

リードの最適化はしばしば、ウェット・ラボ、または計算手段、あるいはその両方のいずれによるかにかかわらず、リード候補についての様々な薬物動態学特性の評価を含む。これには、Log D、Log Pなどの様々な尺度を含むアドメトックス・プロファイル、ならびに、実験に基づくリピンスキーの「5のルール」が含まれる。

[59] エキンズ、エス、ローズ、ジェイ、「In silico ADME/Tox: the state of the art」、J Mol Graph Model、第20巻、第4号、305~309頁、2002年

[60] ゴース、エイ、ケイ、ビスワナダン、ブイ・エヌ、ウェンドロスキー、ジェイ・ジェイ、「Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of ALOGP and CLOGP methods」、J. Phys. Chem、第102巻、3762~3772頁、1998年

[61] リピンスキー、シー・エイ、ロンバード、エフ、ドミニー、ビー・ダブリュー、フィーニー、ピー・ジェイ、「Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings」、Advanced Drug Delivery Reviews 1997、第23巻、3~25頁、1997年

【0061】

タンパク質データ・バンク (PDB) などの誰でもアクセス可能なデータベースに存在する様々な化合物と複合体を作るタンパク質に関して多くの実験的に得られた構造が存在する。GenBankおよびSwissProtなど、タンパク質および核酸の配列についての公共のデータベースも存在する。このようなデータベースの複数の解析を利用して、冗長性、配列類似性、発現した遺伝子の数などを含めて、データの様々な特性が推定されている。

10

20

30

40

50

[62] パーマン、エイチ・エム、ウェストブルック、ジェイ、フェン、ゼット、ジリランド、ジー、バート、ティー・エヌ、ワイシグ、エイチ、シンジャロフ、アイ、エヌ、ボーン、ピー・イー、「The Protein Data Bank」、Nucleic Acids Research、第28巻、235～242頁、2000年

[63] www.ncbi.nlm.nih.gov/Genbank/

[64] バリロック、エイ、ベックマン、ビー、「The SWISS-PROT protein sequence data bank, recent developments」、Nucleic Acids Res、第21巻、3093～3096頁、2003年

[65] ロスト、ビー、サンダー、シー、「Bridging the protein sequence structure gap by structure predictions」、Annual Review Biophys Biomol Struct、第25巻、113～136頁、1996年 10

[66] ワン、ジー、ダンブラック、アール・エル、「PISCES: a protein culling server」、Bioinformatics、投稿中、2002年

[67] エイ・エル・ホプキンス、シー・アール・グルーム、「The Druggable Genome」、Nature Reviews Drug Discovery、第1巻、727～730頁、2002年

[68] マーフィー、アール・ビー、フィリップ、ディー・エム、フリースナー、アール・エイ、「Frozen Orbital QM/MM methods for density functional theory」、Chem. Phys. Letters、第321巻、113～120頁、2000年 20

【0062】

構造データベースに関係する標的タンパク質および化合物について、構造的かつ化学的な情報をデジタル的に表現するための様々なファイル・フォーマットが存在する。例には、pdbフォーマット、(Tripos社の)mol2フォーマット、およびSMILE Sフォーマットが含まれる。

[69] ウェストブルック、ジェイ、フィッツジェラルド、ピー・エム、「Structural Bioinformatics」、ピー・イー・ボーン(監修)、エイチ・ワイシグ(監修)、John Wiley & Sons社、米国ニュージャージー州Hoboken所在、161～179頁、2003年 30

[70] <http://www.tripos.com/custResources/mol2Files/>

[71] <http://www.daylight.com/dayhtml/smiles/smiles-intro.html>

[72] クラーク、エム、クレイマー、アール・ディー、オブデンボッシュ、エヌ、ブイ、「Validation of the General Purpose Tripos 5.2 Force Field」、J. Comp. Chem.、第10巻、982～1012頁、1989年 40

[73] スタウテン、ピー・エフ・ダブリュー、フロメル、シー、ナカムラ、エイチ、サンダー、シー、「An effective solvation term based on atomic occupancies for use in protein simulations」、Molecular Simulation、第10巻、第2～6号、97～120頁、1993年

[74] <http://www2.chemie.uni-erlangen.de/software/corina/index.html>

[75] タカハシ、ワイ、サトー、ワイ、ササキ、エス、「Recognition of largest common structural fragment among a variety of chemical structures」、Ana 50

lytical Sciences、第3巻、23～28頁、1987年

[76] マクレガー、エム・ジェイ、パライ、ピー・ブイ、「Clustering of large databases of compounds using MDL Keys as structural descriptors」、J. Chem. Inf. Comput. Science、第37巻、頁443～、1997年

【先行技術文献】

【特許文献】

【0063】

【特許文献1】米国仮出願第60／511474号

【特許文献2】米国特許第 号

10

【非特許文献】

【0064】

【非特許文献1】<http://www.rcsb.org/pdb/>

【非特許文献2】<http://www.biochem.ucl.ac.uk/bsm/pdbsum/species/>

【非特許文献3】チェン、ワイ・ゼット、ジー、ディー・ジー、「Ligand-Protein Inverse Docking and Its Potential Use in the Computer Search of Protein Targets of a Small Molecule」、Proteins、第43巻、217～226頁、2001年

20

【非特許文献4】チェン、ワイ・ゼット、ジー、ディー・ジー、ウン、シー・ワイ、「Computational Method for Drug Target Search and Application in Drug Discovery」、Journal of Theoretical and Computational Chemistry、第1巻、第1号、213～224頁、2002年

【非特許文献5】<http://www-mitchell.ch.cam.ac.uk/pld/energy.php>

【非特許文献6】ジェイ・ドゥルース、「Drug Discovery: A Historical perspective」、Science、第287巻、1960～1964頁、2000年

30

【非特許文献7】ルーベン・アバギャン、マキシム・トトロフ、「High-throughput docking for lead generation」、Current Opinion in Chemical Biology、第5巻、375～382頁、2001年

【非特許文献8】ラム、エム・エル、バーディック、ケイ・ダブリュー、トバ、エス、ヤング、エム・エム、スキルマン、エイ・ジー他、「Design, docking, and evaluation of multiple libraries against multiple targets」、Proteins 第42巻、296～318頁、2001年

【非特許文献9】ワスコイチ、ビー、パーキンス、ティー・ディー・ジェイ、サイクス、アール・エイ、リー、ジェイ、「Large-scale virtual screening for discovering leads in the post-genomic era」、IBM Systems Journal、第40巻、第2号、2001年

40

【非特許文献10】ショイチェット、ビー・ケイ、ボディアアン、ディー・エル、クンツ、アイ・ディー、「Molecular docking using shape descriptors」、J. Comp. Chem.、第13巻、第3号、380～397頁、1992年

【非特許文献11】メン、イー・シー、シュベント、ディー・エイ、ブラニー、ジェイ・エム、アイ・ディー・クンツ、「Orientational sampling an

50

d rigid-body minimization in molecular docking」、Proteins: Structure, Function, and Genetics、第17巻、266～278頁、1993年

【非特許文献12】ユーイング、ティー・ジェイ・エイ、クンツ、アイ・ディー、「Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening」、J. Computational Chemistry、第18巻、第9号、1175～1189頁、1997年

【非特許文献13】ローレンス、エム・シー、デイビス、ピー・シー、「CLIX: A Search Algorithm for Finding Novel Ligands Capable of Binding Proteins of Known Three-Dimensional Structure」、Proteins、第12巻、31～41頁、1992年

【非特許文献14】カステンホルツ、エム・エイ、バスター、エム、クルシアーニ、ジー、ハークスマ、イー・イー・ジェイ、フォックス、ティー、「GRID/CPCA: A new computational tool to design selective ligands」、J. Medicinal Chemistry、第43巻、3033～3044頁、2000年

【非特許文献15】ミラー、エム・ディー、カースリー、エス・ケイ、アンダーウッド、ディー・ジェイ、シェリダン、アール・ピー、「FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure」、J. Computer-Aided Molecular Design、第8巻、第2号、153～174頁、1994年

【非特許文献16】ソボレフ、ブイ、ウェイド、アール・シー、ブレンド、ジー、エデルマン、エム、「Molecular docking using surface complementarity」、Proteins、第25巻、120～129頁、1996年

【非特許文献17】カチャルスキー・カツィール、イー、シャリフ、アイ、アイゼンスタイン、エム、フリーセム、エイ・エイ、アフラーロ、シー、バクサー、アイ・エイ、「Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques」、Proceedings of the National Academy of Sciences of the United States of America、第89巻、第6号、2195～2199頁、1992年

【非特許文献18】リッチー、ディー・ダブリュー、ケンプ、ジー・ジェイ・エル、「Fast Computation, Rotation, and Comparison of Low Resolution Spherical Harmonic Molecular Surfaces」、J. Computational Chemistry、第20巻、第4号、383～395頁、1999年

【非特許文献19】フィッシャー、ディー、ノレル、アール、ウルフソン、エイチ、ヌシノフ、アール、「Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition」、Proteins、第16巻、278～292頁、1993年

【非特許文献20】レアリー、エム、ウェフィング、エス、レンガウアー、ティー、「Placement of medium-sized molecular fragments into active sites of proteins」、J. Computer-Aided Molecular Design、第10巻、41～54

10

20

30

40

50

頁、1996年

【非特許文献21】クレイマー、ビー、レアリー・エム、レンガウアー、ティー、「Evaluation of the FlexX incremental construction algorithm for protein-ligand docking」、Proteins、第37巻、228～241頁、1999年

【非特許文献22】レアリー・エム、クレイマー、ビー、レンガウアー・ティー、クリーブ、ジー、「A Fast Flexible Docking Method Using An Incremental Construction Algorithm」、J. Mol. Biol.、第261巻、470～489頁、1996年

【非特許文献23】ウェルチ、ダブリュー、ルーパー、ジェイ、ジェイン、エイ・エヌ、「Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites」、Chemical Biology、第3巻、449～462頁、1996年

【非特許文献24】リーチ、エイ・アール、クンツ、アイ・ディー、「Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites」、J. Comp. Chem.、第13巻、730～748頁、1992年

【非特許文献25】ボーム、エイチ・ジェイ、「The computer program LUDI: a new method for the de novo design of enzyme inhibitors」、J. Computer-Aided Molecular Design、第6巻、61～78頁、1992年

【非特許文献26】ボハチェク、アール・エス、マクマーティン、シー、「Multiple Highly Diverse Structures Complementary to Enzyme Binding Sites: Results of Extensive Application of a de Novo Design Method Incorporating Combinatorial Growth」、J. American Chemical Society、第116巻、5560～5571頁、1994年

【非特許文献27】デジャレー、アール・エル、シェリダン、アール・ピー、ディクソン、ジェイ・エス、クンツ、アイ・ディー、ベンカタラガバン、アール、「Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape」、J. Med. Chem.、第29巻、2149～2153頁、1986年

【非特許文献28】クラウセン、エイチ、ブニング、シー、レアリー・エム、レンガウアー・ティー、「FlexE: Efficient Molecular Docking Considering Protein Structure Variations」、J. Molecular Biology、第308巻、377～395号、2001年

【非特許文献29】アバギャン、アール・エイ、トトロフ、エム・エム、クツネットソフ、ディー・エヌ、「Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins」、J. Comp. Chem.、第15巻、488～506頁、1994年

【非特許文献30】ハルグレン、ティー・エイ、マーフィー、アール・ビー、フリースナー、アール・エイ、ベアード、エイチ・エス、フライ、エル・エル、ボラード、ダブリュー・ティー、バンクス、ジェイ・エル、「Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening」、J. Med. Chem.、第47巻、第7号、1750～1759頁、2004年

10

20

30

40

50

【非特許文献31】ルーティ、ビー・エイ、ワッサーマン、ゼット・アール、スタウテン、ビー・エフ・ダブリュー、ホッジ、シー・エヌ、ザチャリアス、エム、マッカモン、ジェイ・エイ、「Molecular Mechanics/Grid Method for the Evaluation of Ligand-Receptor Interactions」、J. Comp. Chem.、第16巻、454～464頁、1995年

【非特許文献32】グッドセル、ディー・エス、オルソン、エイ・ジェイ、「Automated Docking of Substrates to Proteins by Simulated Annealing」、Proteins: Structure, Function, and Genetics、第8巻、195～202頁、1990年

10

【非特許文献33】ジョーンズ、ジー、ウィレット、ピー、グレン、アール・シー、「Molecular Recognition of Receptor Sites using a Genetic Algorithm with a Description of Desolvation」、J. Mol. Biol.、第245巻、43～53頁、1995年

【非特許文献34】ジョーンズ、ジー、ウィレット、ピー、グレン、アール・シー、リーチ、エイ、テイラー、アール、「Development and Validation of a Genetic Algorithm for Flexible Docking」、J. Mol. Biol.、第267巻、727～748頁、1997年

20

【非特許文献35】テイラー、ジェイ・エス、バーネット、アール・エム、Proteins、第41巻、173～191頁、2000年

【非特許文献36】モリス、ジー・エム、グッドセル、ディー・エス、ハリディ、アール・エス、ヒューイ、アール、ハート、ダブリュー・イー、ブリュー、アール・ケイ、オルソン、エイ・ジェイ、「Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function」、J. Comp. Chem.、第19巻、1639～1662頁、1998年

【非特許文献37】ディノーラ、エイ、ベレンドセン、エイチ・ジェイ・シー、ロッカターノ、ディー、「Molecular Dynamics Simulation of the Docking of Substrates to Proteins」、Proteins、第19巻、174～182頁、1994年

30

【非特許文献38】ワン、ジェイ、コールマン、ビー・エイ、クンツ、アイ・ディー、Proteins、第36巻、1～19頁、1999年

【非特許文献39】ベーム、エイ・ジェイ、「The Development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure」、J. Comput-Aided Mol. Des.、第8巻、243～256頁、1994年

40

【非特許文献40】ワン、アール、ガオ、ワイ、ライ、エル、「A new empirical method for estimating the binding affinity of a protein-ligand complex」、J. Molecular Modeling、第4巻、379頁、1998年

【非特許文献41】エルドブリッジ、エム・ディー、マリー、シー・ダブリュー、オートン、ティー・アール、パオリニ、ジー・ブイ、ミー、アール・ピー、「Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes」、J. Computer-Aided

50

Molecular Design、第11巻、425～445頁、1997年

【非特許文献42】ゲルハー、ディー・ケイ、ブージダ、ディー、レイト、ピー・エイ、
「Rational Drug Design: Novel Methodology
and Practical Applications」、Parrill, L.、R
eddy, M. R. 監修、American Chemical Society、米国
ワシントンD. C.、292～311頁、1999年

【非特許文献43】ログナン、ディー、ローエモラー、エス・エル、ホルム、エイ、ブー
ス、エス、シンケ、ブイ・ジェイ、Medicinal Chemistry、第42
巻、4650～4658頁、1999年

【非特許文献44】ハルグレン、ティー・エイ、マーフィー、アール・ビー、フリースナ
ー、アール・エイ、ベアード、エイチ・エス、フライ、エル・エル、ポラード、ダブリュ
ー・ティー、バンクス、ジェイ・エル、「Glide: a new approach
for rapid, accurate docking and scoring. 2
. Enrichment factors in database screenin
g」、J Med Chem.、第47巻、第7号、1750～1759頁、2004年

【非特許文献45】パールマン、ディー・エイ、ケース、ディー・エイ、コールドウェル
、ジェイ・シー、ロス、ダブリュ・エス、チャーザムIII、ティー・イー、ファーガソ
ン、ディー・エム、ザイベル、ジー・エル、シン、ユー・シー、ウィナー、ピー、コール
マン、ピー・エイ、AMBER 4.1、University of Califor
nia, San Francisco、1995年

【非特許文献46】コーネル、ダブリュ・ディー、シーブラック、ピー、ベイリー、シ
ー・アイ、グールグ、アイ・アール、メルツ、ケイ・エム、ファーガソン、ディー・エム
、スペルマイヤー、ディー・シー、フォックス、ティー、コールドウェル、ジェイ・ダブ
リュウ、コールマン、ピー・エイ、「A second-generation for
ce field for the simulation of proteins,
nucleic acid, and organic molecules」、J. Am
erican Chemical Society、第117巻、5179～5197頁
、1995年

【非特許文献47】ヨルゲンセン、ダブリュ・エル、ティラドーリベス、ジェイ、J.
American Chemical Society、第110巻、1657～166
6頁、1988年

【非特許文献48】ハルグレン、ティー・エイ、「Merck Molecular F
orce Field. I. Basis, Form, Scope, Parameteri
zation, and Performance of MMFF94」、J. Comp
. Chem.、第17巻、490～519頁、1996年

【非特許文献49】ブルックス、ビー・アール、ブルッコレリ、アール・イー、オルソン
、ビー・ディー、ステイツ、ディー・ジェイ、スワミナサン、エス、カープラス、エム、
「CHARMM: A Program for Macromolecular Ene
rgy, Minimization, and Dynamics Calculatio
ns」、J. Comp. Chem.、第4巻、187～217頁、1983年

【非特許文献50】スチュワート、ジェイ・ジェイ・ピー、「Quantum Chem
istry Program Exchange」、第10巻、86頁、1990年

【非特許文献51】リオタード、ディー・エイ、ヒーリー、イー・エフ、ルイス、ジェイ
・エム、デュワー、エム・ジェイ・エス、「Quantum Chemistry Pr
ogram Exchange-no. 506」、QCPE Bulletin、第9巻
、123頁、1989年

【非特許文献52】AMSOL-version 6.5.1、ジー・ディー・ホーキン
ス、ディー・ジェイ・ギーセン、ジー・シー・リンチ、シー・シー・チェンバース、アイ
・ロッシ、ジェイ・ダブリュ・ストローラー、ジェイ・リー、ディー・リナルディ、ディ
ー・エイ・リオタード、シー・ジェイ・クレイマー、ディー・ジー・トルラー、Uni

10

20

30

40

50

versity of Minnesota, Minneapolis, 1997年

【非特許文献53】ゴールケ、エイチ、ヘンドリック、エム、クリーブ、ジー、「Knowledge-based Scoring Function to Predict protein-Ligand Interactions」、J. Mol. Biol.、第295巻、337～356頁、2000年

【非特許文献54】ミュゲ、アイ、マーティン、ワイ・シー、「A general and fast scoring function for protein-ligand interactions—a simplified potential approach」、J. Med. Chem.、第42巻、791～804頁、1999年

【非特許文献55】ミッチェル、ジェイ・ビー・オー、ラスコウスキー、アール・エイ、アレックス、エイ、ソントン、ジェイ・エム、「BLEEP—Potential of Mean Force Describing Protein-Ligand Interactions II. Calculation of Binding Energies and Comparison with Experimental Data」、J. Comp. Chem.、第20巻、1165～1176頁、1999年

【非特許文献56】ヘッド、アール・ディー、スミス、エム・エル、オブリー、ティー・アイ、ウォラー、シー・エル、グリーン、エス・エム、マーシャル、ジー・アール、「VALIDATE: A New Method for Receptor-Based Prediction of Binding Affinities of Novel Ligand」、J. American Chemical Society、第118巻、3959～3969頁、1996年

【非特許文献57】ビサンツ、シー、フォルカーズ、ジー、ログナン、ディー、「Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations」、J. Med. Chem.、第43巻、4759～4767頁、2000年

【非特許文献58】「3D QSAR in Drug Design: Recent Advances」、(第1～3巻)、ヒューゴ・クビニー (監修)、ゲルト・フォルカーズ (監修)、イボンヌ・シー・マーティン (監修)

【非特許文献59】「Exploring QSAR: Hydrophobic, Electronic, and Steric Constants (ACS Professional Reference Book)」、コーウィン・ハンシュ (監修)、アルバート・レオ (監修)、ディー・エイチ・フークマン (監修)

【非特許文献60】マフィー、アール・ビー、フィリップ、ディー・エム、フリースナー、アール・エイ、「Frozen Orbital QM/MM methods for density functional theory」、Chem. Phys. Letters、第321巻、113～120頁、2000年

【非特許文献61】アクビスト、ジェイ、メディーナ、シー、サミュエルソン、ジェイ・イー、「A new method for predicting binding affinity in computer-aided drug design」、Prot. Eng.、第7巻、385頁、1994年

【非特許文献62】ジョーンズ・ハーツォグ、ディー・ケイ、ヨルゲンソン、ダブリュー・エル、「Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method」、J. Med. Chem.、第40巻、1539頁、1997年

【非特許文献63】ボハチェク、アール・エス、マクマーティン、シー、グイダ、ダブリュー・シー、「The Art and Practice of Structure-Based Drug Design: A Molecular Modeling

10

20

30

40

50

Perspective」、Medicinal Research Reviews、第16巻、3頁、1996年

【非特許文献64】エキンズ、エス、ローズ、ジェイ、「In silico ADME/Tox: the state of the art」、J Mol Graph Model、第20巻、第4号、305～309頁、2002年

【非特許文献65】ゴース、エイ、ケイ、ビスワナダン、ブイ・エヌ、ウェンドロスキー、ジェイ・ジェイ、「Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of ALOGP and CLOGP methods」、J. Phys. Chem、第102巻、3762～3772頁、1998年

【非特許文献66】リピンスキー、シー・エイ、ロンバード、エフ、ドミニー、ビー・ダブリュー、フィーニー、ピー・ジェイ、「Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings」、Advanced Drug Delivery Reviews 1997、第23巻、3～25頁、1997年

【非特許文献67】バーマン、エイチ・エム、ウェストブルック、ジェイ、フェン、ゼット、ジリランド、ジー、バート、ティー・エヌ、ワイシグ、エイチ、シンジャロフ、アイ、エヌ、ボーン、ピー・イー、「The Protein Data Bank」、Nucleic Acids Research、第28巻、235～242頁、2000年

【非特許文献68】www.ncbi.nlm.nih.gov/Genbank/

【非特許文献69】バリロック、エイ、ベックマン、ビー、「The SWISS-PROT protein sequence data bank, recent developments」、Nucleic Acids Res、第21巻、3093～3096頁、2003年

【非特許文献70】ロスト、ビー、サンダー、シー、「Bridging the protein sequence structure gap by structure predictions」、Annual Review Biophys Biochem Struct、第25巻、113～136頁、1996年

【非特許文献71】ワン、ジー、ダンブラック、アール・エル、「PISCES: a protein culling server」、Bioinformatics、投稿中、2002年

【非特許文献72】エイ・エル・ホブキンス、シー・アール・グルーム、「The Druggable Genome」、Nature Reviews Drug Discovery、第1巻、727～730頁、2002年

【非特許文献73】マーフィー、アール・ビー、フィリップ、ディー・エム、フリースナー、アール・エイ、「Frozen Orbital QM/MM methods for density functional theory」、Chem. Phys. Letters、第321巻、113～120頁、2000年

【非特許文献74】ウェストブルック、ジェイ、フィッツジェラルド、ピー・エム、「Structural Bioinformatics」、ピー・イー・ボーン（監修）、エイチ・ワイシグ（監修）、John Wiley & Sons 社、米国ニュージャージー州 Hoboken 所在、161～179頁、2003年

【非特許文献75】<http://www.tripos.com/custResources/mol2Files/>

【非特許文献76】<http://www.daylight.com/dayhtml/smiles/smiles-intro.html>

【非特許文献77】クラーク、エム、クレイマー、アール・ディー、オブデンボッシュ、エヌ、ブイ、「Validation of the General Purpose

10

20

30

40

50

Tripos 5.2 Force Field」、J. Comp. Chem.、第10巻、982～1012頁、1989年

【非特許文献78】スタウテン、ピー・エフ・ダブリュー、フロメル、シー、ナカムラ、エイチ、サンダー、シー、「An effective solvation term based on atomic occupancies for use in protein simulations」、Molecular Simulation、第10巻、第2～6号、97～120頁、1993年

【非特許文献79】<http://www2.chemie.uni-erlangen.de/software/corina/index.html>

【非特許文献80】タカハシ、ワイ、サトー、ワイ、ササキ、エス、「Recognition of largest common structural fragment among a variety of chemical structures」、Analytical Sciences、第3巻、23～28頁、1987年

【非特許文献81】マクレガー、エム・ジェイ、バライ、ピー・ブイ、「Clustering of large databases of compounds using MDL Keys as structural descriptors」、J. Chem. Inf. Comput. Science、第37巻、頁443～、1997年

【発明の概要】

【課題を解決するための手段】

【0065】

10

20

本発明の態様は、リード候補生体分子と、生体高分子であることが多い潜在的な反応物質分子の間の有害交差反応を予測して、対象とする病気または生物学的な状態に関する安全性および薬効の点でリード候補をよりよく特徴付け、リード候補分子のさらなる最適化を容易にして、所望の標的との結合特異性を改善し、他の無関係な生体分子との有害な、かつ／またはその他の点で不測の交差反応を回避する方法に係るものである。

【0066】

計算システムでは、適切な環境内で反応がモデル化され、それによってリード候補分子と潜在的な反応物質分子の間の反応プロファイルが求められる。この方法は、リード分子と潜在的な反応物質分子との間の反応について反応の尺度を取得することと、リードおよび潜在的な反応物質分子についての他の変量、例えば、関連するソース、機能、および相同体の注釈、モデル予測、ならびに（適用可能な場合には）実験結果を取得することと、反応の尺度と他の関連する注釈および変量とを組み合わせることで反応プロファイルを構築することを含む。この反応プロファイルは、リード候補と潜在的な反応物質分子が組み合わせられる確度、ならびに、予測される交差反応が望ましくなく、さらには有害である確度をともに反映するものである。このプロセスは同じリード候補分子について、分子ライブラリまたは分子データベースからの複数の潜在的な反応物質分子に対して繰り返すことができ、リード候補について交差反応概要プロファイルが生成され、それによって、安全性および薬効の点でこのリードがよりよく特徴付けられる。

30

【0067】

本発明の別の態様は、リード候補の再設計および最適化を、おそらくは本質的に反復して行って、予測される有害な交差反応を回避する方法を含む。この方法は、リード候補を所望の標的分子に結合する際に必要とされる共通の分子のコアまたは枠組を識別することと、次いで、リード候補分子の、追加、除去、または置換し得る様々な分子構成要素を識別して、予測される関連結合様式および1つ（または複数）の反応の尺度を調べることによって、リードと所望の標的分子との結合親和性を大きく低下させることなく、潜在的な有害反応物質分子との結合親和性を低下させ、かつ／または、ある種の分子間水素結合を阻害することを含む。次いで、リード候補の新たに生成された構造的な変形について、分子ライブラリまたは分子データベースからの複数の潜在的な反応物質分子に対して新しい概要反応プロファイルを生成することによって、リード候補の構造的な変異体の1つまたは複数の有害な交差反応の確率が十分に低いと考えられるまで、新しい最適化サイクル

40

50

を反復することができる。

【図面の簡単な説明】

【0068】

【図1A】本発明の様々な実施形態の説明で用いる標的生体高分子の例として示すタンパク質（ジヒドロ葉酸還元酵素）の1次アミノ酸配列を示す図である。

【図1B】本発明の様々な実施形態の説明で用いる標的生体高分子の例として示すタンパク質（ジヒドロ葉酸還元酵素）の2次構造表現を示す図である。

【図1C】本発明の様々な実施形態の説明で用いる標的生体高分子の例として示すタンパク質（ジヒドロ葉酸還元酵素）の溶媒接触可能表面を示す図である。

【図1D】本発明の様々な実施形態の説明で用いる標的生体高分子の例として示すタンパク質（ジヒドロ葉酸還元酵素）を扱うp d bエントリ4 d f rについてのX線結晶解析による構造データのp d bファイルの表現を示す図である。

【図2A】本発明の様々な実施形態の説明で用いるリード候補の例として示す分子（メトトレキサート）の化学式および立体S M I L Eを示す図である。

【図2B】本発明の様々な実施形態の説明で用いるリード候補の例として示す分子（メトトレキサート）の2次元構造表現を示す図である。

【図2C】本発明の様々な実施形態の説明で用いるリード候補の例として示す分子（メトトレキサート）の配座の「玉と棒」による3次元表現を示す図である。

【図2D】本発明の様々な実施形態の説明で用いるリード候補の例として示す分子（メトトレキサート）の同じ配座についての構造データのT r i p o s m o l 2ファイル表現を示す図である。

【図3A】本発明の様々な実施形態の説明で用いる複合体構造の例として示すジヒドロ葉酸還元酵素に結合したメトトレキサートを扱う分子複合体の活性部位の溶媒接触表面図である。

【図3B】本発明の様々な実施形態の説明で用いる複合体構造の例として示すジヒドロ葉酸還元酵素に結合したメトトレキサートを扱う分子複合体の活性部位内で結合したメトトレキサートを示す複合体の3次元表現図である。

【図3C】本発明の様々な実施形態の説明で用いる複合体構造の例として示すジヒドロ葉酸還元酵素に結合したメトトレキサートを扱う分子複合体のリガンド（メトトレキサート）と標的（ジヒドロ葉酸還元酵素）に対して相互作用する基の2次元模式表現図である。

【図3D】本発明の様々な実施形態の説明で用いる複合体構造の例として示すジヒドロ葉酸還元酵素に結合したメトトレキサートを扱う分子複合体の標的タンパク質の活性部位残基の4つの基の表を示す図である。

【図4A】本発明のいくつかの実施形態で用いるのに、標的および／または潜在的な反応物質のレコードをどのように構築し得るかの例を示す図である。

【図4A-2】本発明のいくつかの実施形態で用いるのに、標的および／または潜在的な反応物質のレコードをどのように構築し得るかの例を示す図である。

【図4A-3】本発明のいくつかの実施形態で用いるのに、標的および／または潜在的な反応物質のレコードをどのように構築し得るかの例を示す図である。

【図4B】本発明のいくつかの実施形態で用いるのに、リード候補のレコードをどのように構築し得るかの例を示す図である。

【図4B-2】本発明のいくつかの実施形態で用いるのに、リード候補のレコードをどのように構築し得るかの例を示す図である。

【図4B-3】本発明のいくつかの実施形態で用いるのに、リード候補のレコードをどのように構築し得るかの例を示す図である。

【図5】様々な潜在的な反応物質とリード候補の潜在的な交差反応を求め、リード候補を再設計し最適化して1つまたは複数の所望の標的との結合特異性を改善するのに使用し得る、本発明の実施形態による計算モデリング・システムのブロック図である。

【図6A】本発明の様々な実施形態の説明で用いる複合体構造の別の例として示す大腸菌チミジル酸シンターゼに結合するメトトレキサートを扱う分子複合体を示す図で、ジヒドロ

10

20

30

40

50

葉酸還元酵素と複合体を作るメトトレキサートについて図 3 A で示すものに対応する。

【図 6 B】本発明の様々な実施形態の説明で用いる複合構造の別の例として示す大腸菌チミジル酸シンターゼに結合するメトトレキサートを扱う分子複合体を示す図で、ジヒドロ葉酸還元酵素と複合体を作るメトトレキサートについて図 3 B で示すものに対応する。

【図 6 C】本発明の様々な実施形態の説明で用いる複合構造の別の例として示す大腸菌チミジル酸シンターゼに結合するメトトレキサートを扱う分子複合体を示す図で、ジヒドロ葉酸還元酵素と複合体を作るメトトレキサートについて図 3 C で示すものに対応する。

【図 7 A】本発明の様々な実施形態の説明で用いる複合構造の別の例として示すウサギ細胞質セリンヒドロキシメチル転移酵素に結合するフォリン酸を扱う分子複合体を示す図で、ジヒドロ葉酸還元酵素と複合体を作るメトトレキサートについて図 3 A で示すものに対応する。

【図 7 B】本発明の様々な実施形態の説明で用いる複合構造の別の例として示すウサギ細胞質セリンヒドロキシメチル転移酵素に結合するフォリン酸を扱う分子複合体を示す図で、ジヒドロ葉酸還元酵素と複合体を作るメトトレキサートについて図 3 B で示すものに対応する。

【図 7 C】本発明の様々な実施形態の説明で用いる複合構造の別の例として示すウサギ細胞質セリンヒドロキシメチル転移酵素に結合するフォリン酸を扱う分子複合体を示す図で、ジヒドロ葉酸還元酵素と複合体を作るメトトレキサートについて図 3 C で示すものに対応する。

【図 8 A】メトトレキサートの枠組分子構造の 2 次元表現を、注釈を付けた構造連結部位とともに示す図である。

【図 8 A-2】メトトレキサートに適用する薬物再設計／最適化プロセスに関わる本発明の一実施形態の一部として、既存の枠組に置換または追加し得る化学基の表である。

【図 8 B】メトトレキサートに適用する薬物再設計／最適化プロセスに関わる本発明の一実施形態の一部として生成された様々な異なる構造の例を示す図である。

【図 8 B-2】メトトレキサートに適用する薬物再設計／最適化プロセスに関わる本発明の一実施形態の一部として生成された様々な異なる構造の例を示す図である。

【発明を実施するための形態】

【0069】

本発明は、本開示を読んだ後で明らかになるように、広く適用される。本発明による計算システムの実施形態の説明では、いくつかの可能な変形形態のみを説明する。他の応用例および変形形態は当業者には明らかであり、そのため、本発明を、これらの実施例と同じ程度に狭く解釈すべきではなく、添付の特許請求の範囲に従って解釈すべきである。

【0070】

次に、例として本発明の実施形態を説明するが、本発明はこれらに限定されるものではない。本発明の有用性は広く、本発明を多くの異なる状況で利用し得ることを理解されたい。

【0071】

本発明は、おそらくは 1 つまたは複数の所望の標的に関連する 1 つまたは複数のリード候補と、1 組の潜在的な反応物質との交差反応性を推定して、これらのリード候補の有用性の 1 つまたは複数の態様を評価することに関係するものである。

【0072】

本発明では、標的は、生体分子、生体分子の一部、1 つまたは複数の生体分子の化合物、あるいは他の生体反応性作用物質である。これらは生体高分子であることが多く、その環境において活動が改変されることが望まれるものである。例えば、タンパク質、ポリペプチド、および核酸を含めて、生体高分子は標的の例である。これらの標的の活動の改変は、他の相互作用の前またはその最中に、標的の活動を不活性化（抑制作用）か、標的の活動を高めるか、または他のやり方でその活動を改変（触媒作用）ことを含む得る。一実施形態では、標的は、ヒトの体内で生成されるか、またはヒトの体内に導入され、病気その他の悪影響を及ぼすタンパク質とすることができ、所望の改変は、このタンパ

10

20

30

40

50

ク質の関連する活性部位に小型の生体分子を競合結合させることによって、このタンパク質の活動を抑制することである。別の実施形態では、この標的タンパク質自体は、望まれない病気または悪影響を直接引き起こすものではないが、その機能に影響を及ぼすことによって、何らかの他のタンパク質（例えば、酵素、抗体など）または生体分子に関わる反応をより良好に制御することができ、それによって、処置を必要とする病状が軽減される。

【0073】

別の実施形態では、細胞表面受容体にリード候補をドッキングすることができ、それによって、ウィルス・タンパク質などの病原体生体高分子が細胞に取り付くことを妨げて病気を防ぐ。別の例では、リード候補は、標的タンパク質の化学的な改変（例えば、リン酸化、ジスルフィド結合の切断、金属イオン配位の崩壊など）、または遺伝子治療の場合には、DNA中の1つまたは複数の塩基対の実際の置換を促進することができる。別の例では、所望の改変は、標的を改変して、標的を扱う機能経路をより容易に識別し、かつ／またはよりよく特徴付けることである。例えば、リード候補によって分子標識を担持するように標的を改変し、それによって、この標的がどこにいるのか、標的がその環境とどのように相互作用するのかを検出するのを容易にすることができる。

【0074】

図1a～図1dに、タンパク質の例を示す。このタンパク質は、p d b エントリ 4 d f r で示す結晶構造を有する。このタンパク質のジヒドロ葉酸還元酵素は、DNA合成に必要なデオキシチミジル酸の産生サイクルの一部としてテトラヒドロ葉酸の再生に関わる標的である。図1aの項目110に、アミノ酸ごとの3文字表記を用いてこのタンパク質の1次アミノ酸配列を示す。図1bの絵120に、4 d f r タンパク質分子の2次構造の模式的な表現を、βシート、αヘリックス、ヘアピンなどの様々な2次構造構成要素の場所を示す表の集合125とともに示す。図1cの絵130に、半径1.4 Åのプローブ球を用いたジヒドロ葉酸還元酵素タンパク質の溶媒接触可能表面の表現を示す。この絵で、異なる色は異なる表面原子を示す（例えば、項目131の白は水素、項目132の黒は炭素、項目133のネズミ色は酸素、など）。

【0075】

図1dに、タンパク質データ・バンク[62、69]から得られたジヒドロ葉酸還元酵素タンパク質の入力された構造配座について、X線結晶解析に基づく化学構造のp d b ファイルの表現150を示す。図1dでは、p d b ファイルの表現150は、全般ヘッダ部分142、原子タイプおよび配位情報からなる部分144、および結合連結性情報に関する部分146を含む。ヘッダ部分142は、標的生体分子の身元、ソース、構造的構成要素、配列データその他の特徴に関する所望の任意の注釈その他の情報を含み得る。部分144に、省略記号を用いて、ジヒドロ葉酸還元酵素の単一鎖の1280個の非水素原子の一覧を示す。原子ごとに、化学の種類（例えば、原子）および3つの空間座標が含まれる。例えば、原子9の行は、それが、名称がNの窒素原子であり、A鎖のI L E 残基中であり、残基I Dが2であり、指定直交座標系における（x、y、z）座標が（23.920、57.222、6.665）であることを示す。部分144は、標的に関連する金属イオン、構造水、あるいは他の追加の原子または分子などの様々な異種原子のエントリも含み得ることに留意されたい。部分144では、項目148で示す行に、構造水に対応する酸素、さらには塩素原子についてのいくつかの行エントリの例を示す。

【0076】

P D B ファイル150の部分146に、P D B ファイルのいわゆる「連結レコード」のサブセットを示す。各行のエントリは、各原子に関連する結合部の一覧を記述している。例えば、この部分の第1行には、原子1が原子2に結合していることが示され、第2行には、原子2が原子1、3、および5に結合していることが示されている。この例で水素がどのように欠落しており、したがって、原子ごとの結合連結部は完全ではないことがあることにも留意されたい。当然のことながら、水素原子の位置が既に指定されている場合には、このP D B ファイルの表現の完全な変異体が可能であるが、化学構造がX線結晶解析

などの実験的な観察から得られる多くの場合、水素の位置は極めて不確かであるか、完全に欠落していることがある。ただし、水素原子の位置を予測し、生理 pH の仮定に基づいてタンパク質を正しく水素化し得る標準計算ツールが存在する。

【0077】

ファイル・フォーマットの他の例には、MDL フォーマット、Tripos 社の mol2 フォーマット [70]、SMILES 規約 [71] などが含まれる。スキーマとエンティティの様々な関係を扱う様々な 3 次元生体分子構造データベースの形で、標的の化学的かつ構造的な情報を格納することもできる。

【0078】

リード候補は、標的との関連生物活性の事前評価に基づいて選択された生体分子、生体分子の一部、1 つまたは複数の生体分子の化合物、あるいは、他の生体反応性作用物質である。事前評価の目的は通常、適切な環境にあるリード候補が、標的と反応して標的の活動を所望のとおりに変更するかどうかを判定することである。生物活性の評価は、無数の生化学アッセイ、ウェット・ラボ実験、様々な測定プロセス、結合親和性の計算による予測、ハイスループット・スクリーニング、仮想ライブラリ・スクリーニング、アドメトックス・プロファイル、QSAR、分子モデリングなどを必要とすることがある。リード候補の例には、小型の分子リガンド、ペプチド、タンパク質、タンパク質の一部、合成化合物、天然化合物、有機分子、炭水化物、残基、無機分子、イオン、個々の原子、遊離基、および他の化学的に活性な要素が含まれる。リード候補から、所望の変更を作り出すために投与または使用し、また、望ましくない変更について検査または試験するために使用する薬物または化合物の基礎を形成することができる。「リード」および「薬物候補」という用語は、「リード候補」という用語と交換可能に用いられる。

【0079】

図 2 a ~ 図 2 d に、図 1 a ~ 図 1 d に示すジヒドロ葉酸還元酵素に対する周知のナノモル結合剤である化合物メトトレキサートを示す。メトトレキサートは、タンパク質ジヒドロ葉酸還元酵素を抑制し、それによって間接的にデオキシチミジル酸の産生を阻害し、DNA 合成に関わる基本細胞プロセスを止めて悪性癌細胞の成長を遅らせることによっていくつものタイプのヒトの癌を治療する抗癌薬として化学療法に用いられる。図 2 a の項目 210 に、メトトレキサートの化学式を示す。図 2 a の項目 215 に、メトトレキサートについての立体 SMILES [71] を示す。図 2 b の絵 220 に、非水素原子のみを含むメトトレキサート分子の 2 次元構造表現を示す。

【0080】

図 2 c の項目 230 に、図 2 a の項目 210 で列挙した化学式 $C_{20}H_{22}N_8O_5$ に従ってメトトレキサート分子の配座 230 を「玉と棒」で表現したものを示す。この分子は、原子 240 および結合部 260 の集合からなる。項目 243 で示す小型の黒い原子は炭素原子を表す。項目 245 で示すより小さい白い原子は水素原子を表し、わずかに大きい暗い色の原子（項目 247）は酸素原子であり、より大型の白い原子（項目 249）は窒素原子である。

【0081】

図 2 c を継続して参照すると、項目 253 は、ベンゼン環 (C_6H_4) を含む円を示し、項目 255 は、カルボキシル基 (COO^-) を含む円を示し、項目 257 は、メチル基 (CH_3) を含む別の円を示している。項目 263 は、ベンゼン環 256 と、メチル基 257 を含むエステル基とを連結する共有結合を示す。項目 265 は、炭素原子 243 とカルボキシル基 255 を連結する共有結合を示す。最後に、項目 267 は、メチル基 257 と窒素原子 249 を連結する共有結合を示す。

【0082】

図 2 d に、図 2 c に示すメトトレキサートについて入力された配座についての様々な構造的かつ化学的な情報を含む Tripos mol2 ファイルを示す。列 270 には、各原子ごとに指数が列挙されている。列 273 には、原子ごとに（一意でないことがある）原子名が列挙されている。列 275、277、および 279 にはそれぞれ、原子ごとに内

10

20

30

40

50

部座標系の座標 x 、 y 、 z が列挙されている。列 280 には、原子ごとに、ハイブリッド形成状態、化学的種類、結合連結性、水素結合能力、芳香族性、ならびに、場合によっては化学基に関する情報を体系化した *Tripos* 力場 [72] による SYBYL 原子タイプが列挙されている。列 282 および 285 には、原子ごとに（タンパク質、核酸などに関連する）残基 ID および残基名が列挙されている。部分 290 には、この分子サブセットにおけるすべての結合部が列挙されている。列 291 には、結合部ごとに結合指数が列挙されており、列 292 および 293 には、結合部によって連結された 2 つの原子の原子指数が列挙されており、列 295 には、単結合、二重結合、三重結合、非局所的結合、アミド結合、芳香結合、または他の特殊な共有結合であり得る結合のタイプが列挙されている。他の実施形態では、このような情報は、塩橋または水素結合などの非共有結合を表すこともある。この例では、水素原子がこの時点でどのように含まれているかに留意されたい。

10

【0083】

標的の表面上またはその近傍でリード候補が標的と結合または安定に接触する相互作用の座を通常、活性部位と称する。これは、結合ポケット、結合部位、または結合キャビティとも称する。

【0084】

図 3 a ~ 図 3 d は、図 1 a ~ 図 1 d のジヒドロ葉酸還元酵素タンパク質と図 2 a ~ 図 2 d のメトトレキサート分子によって形成される分子複合体に関係するものである。図 3 a に、メトトレキサートとの反応に関連するジヒドロ葉酸還元酵素の活性部位を示す。図 3 a では、この活性部位は、明るい灰色のタンパク質の全溶媒接触可能表面（項目 320）に埋め込まれ、暗い灰色で陰影が付けられた溶媒接触可能表面の一部（項目 310）として示されている。図 3 b に、白球 325 の集合で表すリード候補（メトトレキサート）に結合した複峰性溶媒接触可能表面で表された（暗い灰色の部分は活性部位 310 であり、明るい灰色の部分 320 はタンパク質の残りの部分である）標的タンパク質（ジヒドロ葉酸還元酵素）を含む分子複合体を示す。図 3 b で明らかなように、メトトレキサート分子は、図 3 a に示す活性部位に緊密に収まっている。

20

【0085】

図 3 c に、メトトレキサートの様々な化学基とジヒドロ葉酸還元酵素の様々な残基との相互作用の 2 次元模式表現を示す。項目 340 は、p d b エントリ 4 d f r の第 1 鎖についてのジヒドロ葉酸還元酵素の活性部位内のメトトレキサート分子 (M T X) の 2 次元表現を示す。項目 331、333、334、335、336、337、および 339 は、メトトレキサートの様々な基との疎水性接触に関わるタンパク質残基を表す。項目 350、352、354、356、および 358 はそれぞれ、活性部位内のジヒドロ葉酸還元酵素の触媒残基 A s p 27、I l e 5、I l e 94、A r g 52、および A r g 57 を表す。これらはそれぞれ、メトトレキサートとの 1 つまたは複数の細胞間水素の形成に関わるものである。この水素結合自体は、対をなす水素結合ドナー基とアクセプタ基の間の破線で示す。

30

【0086】

図 3 d に、メトトレキサート分子と相互作用するジヒドロ葉酸還元酵素タンパク質の活性残基の集合の文字列表現を示す。ここで、4 つの活性部位残基にグループ分けし、それらを、項目 371 (A P T)、372 (A N M)、373 (A A B)、および 374 (A G L) と示す。項目 375 は、A P T グループについての概要説明を表し、項目 377 は、A P T グループに含まれる個々の残基の一覧である。A N M、A A B、および A G L のグループについて、類似の情報が入手可能である。

40

【0087】

図 3 a ~ 図 3 d に示す分子複合体についての情報は、様々な手段によってデジタル的に表してもよい。これらの手段には、図 1 d および図 2 d の例で示した上記 p d b および m o l 2 のファイル・フォーマットが含まれるが、これらに限定されるものではない。

【0088】

50

潜在的な反応物質は、リードが標的と反応すると予想される環境と類似の環境で生じるか否かにかかわらずリードと反応し得る生体分子、生体分子の一部、1つまたは複数の生体分子の複合体、あるいは、他の生体反応性作用物質である。本明細書では、標的と宿主生物を共有する1つまたは複数の潜在的な反応物質の活動をリードが改変することを「交差反応」と称する。

【0089】

リード候補とこのような潜在的な反応物質との交差反応により、潜在的な反応物質の動作の、望ましくないことが多い不測の改変が生じることがある。しかし、このような潜在的な反応物質は、すぐには病気に直接関係しない生体高分子であることが多く、普通なら宿主生物にとって正常な機能を実施し、また、不可欠な機能さえ実施するので、このような交差反応により望ましくない副作用が生じることがある。例えば、メトトレキサートは、ジヒドロ葉酸還元酵素の強力な抑制剤であると同時に、肺、腎臓、および神経系内の複数のタンパク質の活動に影響を及ぼす毒性副作用を引き起こすことが知られている。別の例では、肝臓内のタンパク質との望ましくない相互作用を起こし、それによって肝臓障害、さらには、肝硬変などの深刻な病状を誘発する多くの化合物が知られている。

【0090】

潜在的な反応物質は、ヒト、哺乳動物、昆虫、植物、細菌などを含めて、様々な生物由来のものとし得る。潜在的な反応物質は、様々な異なる種、器官、組織、細胞型などから得ることもできる。これらの構造的な情報は、様々な実験手段および測定プロセスによって得ることもできるし、計算によるモデル化および構造的な予測によって得ることもできる。

【0091】

1組の潜在的な反応物質は、本質的に実験的なもの、計算によるもの、または生物学的なものであるかに関わらず、様々なソースから知ることができ、かつ／または得ることができる。これらのソースは、PDB [62]、GenBank [63]、SwissProt [64] など、1つまたは複数の公共の、または所有権下にあるプロテオミクスまたはゲノミクスのデータベースを含み得る。これらに関して、例えば、種／生物、分子サイズおよび分子量、塩基の非冗長配列、識別または予測される機能、疎水性、pKa、回転可能な結合の数、作用部位、タンパク質のファミリーまたはスーパー・ファミリーによる分類など、周知の、または導出された注釈に基づいて、様々なクエリ選択を行うことができる。このような組は、周知のあらゆるタンパク質、ヒト、またはそうでないものさえ含み得る。

【0092】

他のソースには、機能ゲノミクス研究、あるいは周知のアミノ酸配列からの、さらには周知のDNA/RNA配列から直接行われる構造モデル予測が含まれる。潜在的な反応物質は、タンパク質チップ、遺伝子またはプロテオミクスによる発現研究、配列レベルまたは構造レベルのいずれかでの比較相同性分析、機能経路または信号経路の研究、あるいは他の方法を利用する様々な生化学アッセイその他の実験の結果として識別することもできる。

【0093】

例として「標的分子」と称するが、標的は上記で説明したように単分子に限定されないこと、特にことわらない限り、標的分子に適用されることは他のタイプの標的にも適用されることを理解されたい。同じことが、リード分子と称するリードその他のリード候補および潜在的な反応物質分子と称する潜在的な反応物質にも当てはまる。

【0094】

この時点で、「環境」という用語を詳細に論じることが価値があろう。というのは、それが本発明の実施形態に関係するからである。背景技術のところで既に述べたように、「環境」という用語は概ね、分子間の反応部位の物理的かつ化学的な周囲の状況を指す。本発明のある種の実施形態では、環境は、ヒトの体内、ヒト以外の体内、実験室環境、または他の生物学的に適切な環境であり得る。他の実施形態では、環境は、身体全体よりも小

10

20

30

40

50

さな部分、例えば、特定の器官、循環系、または神経系と指定されることがある。反応部位の例には、様々な細胞内領域、細胞核内、血流中、消化管内、様々な膜、および細胞表面が含まれる。

【0095】

標的の環境はしばしば、他の分子、1つまたは複数の分子の化合物、および／または他の生体反応性作用物質または不活性構成要素など、他の構成要素を含む。ある種の実施形態では、環境は、溶媒を含む複数の原子、イオン、および／または分子を含むことがあり、この溶媒中に反応部位が内部に埋め込まれる。溶媒の原子、イオン、または分子の例は、標的または潜在的な反応物質とリードとの反応部位のところ、またはその近傍に存在する水、アルコール、アルカン、塩、糖、脂肪酸、酸、塩基、ヒドロニウムイオン (H^+)、遊離基、ペプチドその他の生体分子、さらには生体高分子を含み得る。他の実施形態では、環境は、誘電強度、イオン化傾向、透過性、化学ポテンシャル、静電ポテンシャルなど、様々な特性をモデル化するのに用いる連続体溶媒和モデルも含み得る。他の実施形態では、環境は、その化学的な構成要素だけでなく、光その他の利用可能な放射の量、温度、圧力、水、または他の関連する環境条件などによってさらに特徴付けることができる。

【0096】

本発明では、標的とリードの反応部位の環境は、潜在的な反応物質とリードの反応部位の環境と極めて異なることがあることに留意されたい。実際、反応部位は、同じ細胞型、組織、さらには生物内とする必要はない。場合によっては、標的とリードの反応部位の環境は既知であり、例えば筋組織内注入など、様々な手段によって生物のある種の部分にリードを閉じ込めることができる。この場合、これら2つの環境、すなわち、標的ーリードの反応部位の一方の環境と、潜在的な反応物質ーリードの反応部位の他方の環境とを同じに、少なくとも極めて類似したものにし得る。このような場合、標的とリードの反応部位に関連する環境と大きく異なる環境に存在する潜在的な反応物質を除外することが可能なこともある。しかし、ほとんどの場合、環境を極めて特定なものとしなくてよい。というのは、リード、標的、および／または潜在的な反応物質は、生物の消化管、血流その他の区域を通して移動し得るからである。

【0097】

背景技術の項で既に論じたように、結合親和性は、様々なエンタルピ成分およびエントロピ成分に関わることがある。エンタルピ成分は、分子内相互作用および分子間相互作用の両方からの寄与を含み得る。これらの相互作用には、静電気、vdW、脱溶媒、水素結合、金属ーリガンドの相互作用、および分子内歪みが含まれるが、これらに限定されるものではない。正確にモデル化するのがより難しいことが多いエントロピ成分は、結合後の配座エントロピの低下および疎水作用などの溶媒に関係する現象を含み得るが、これらに限定されるものではない。ドッキングおよびスコアリングの状況でのポテンシャル・エネルギー、および自由エネルギーのエントロピ成分の標準的なモデル化の考察については、[6、17、26、29、31]を参照されたい。

【0098】

自由エネルギーへのエンタルピおよびエントロピの寄与はともに、物理モデルに適用する1つまたは複数の計算による方法を利用することによって、予測または近似することができる。結合エネルギーの例として、メトトレキサートとジヒドロ葉酸還元酵素の複合体について実験的に求めた結合エネルギーは、 -55.34 KJ/mol であり、これは、解離定数 K_d が、標準生理条件下で 1.97×10^{-10} であることを示唆している（この場合も、<http://www-mitchell.ch.cam.ac.uk/pld/energy.php>）。

【0099】

モデルの例には、量子力学による計算、古典分子力学、統計力学、および実験および／または知識に基づく方式、あるいはこれらの任意の組合せに基づくものが含まれる。標準的な計算による方法の例には、分子動力学シミュレーション、統計力学シミュレーション、機能最適化ルーチン（モンテ・カルロ法、遺伝子アルゴリズム、動的計画法、シミュレ

ーテッド・アニーリングなど)、自由エネルギー摂動論、QM/MMシミュレーションなどが含まれる。

【0100】

分子力学モデルまたは分子動力学シミュレーションの状況では、パラメータ、エネルギー・パラメータは概ね、1つまたは複数の標準エネルギー成分の計算の状況で特定の物理的かつ化学的な属性を表す構成要素の原子、結合部、および／または化学基に関連する。エネルギー・パラメータの割当ては、化学基、ポリペプチド中のアミノ酸などの分子構造、タンパク質中の α ヘリックスまたは β シートなどの2次構造、あるいは全体としての分子の状況での所与の相互作用に関わる1つまたは複数の原子または結合部の化学的な身元および／またはこれら1つ(または複数)の原子または1つ(または複数)の結合部の場所のみに依存し得る。一例は、タンパク質のリジン残基の主鎖上のペプチド結合に関わる窒素原子に割り当てられた静電相互作用に関する電荷の値である。別の例は、タンパク質のアルギニン残基の側鎖中の窒素原子に割り当てられた電荷の値である。この割り当てられた電荷の値は、リジン残基中の上記窒素に関連するものと大きく異なることがある。

【0101】

典型的には、これらのパラメータは、全原子力場または融合原子力場、例えば、AMBER [40、41]、OPLS [42]、MMFF [43]、CHARMM [44]などを用いることによって、環境中の原子または原子群、あるいは標的／リード／潜在的な反応物質の結合部、ならびに、他の原子または分子に割り当てられる。エネルギー・パラメータは、計算実行時に先だって1つまたは複数の分子に割り当て、何らかのコンピュータ可読フォーマット(すなわち、フラット・ファイル、データベース・テーブルなど)に格納し、次いで、必要に応じて取り出すこともできるし、実際には、実際の計算自体の間に「計算中に」割り当てることができる。

【0102】

本発明では、標的、リード、および潜在的な反応物質分子についての化学的かつ構造的な情報、関連の注釈、および関連するエネルギー・パラメータを分子レコードに記憶することができる。これらの分子レコードは、複数のデータベース・レコードとして、または別々のファイルとして、あるいは他のデータ構造でデジタル的に表すことができる。特定の実施形態では、レコードは、データベースから読み取られ、個々のファイルとして格納され、これら個々のファイルは、このファイルに含まれるデータの演算を行うために本発明の他の部分に渡される。

【0103】

これらのレコードがそれらの標的／リード／潜在的な反応物質を表すやり方はいくつかのものがあり得る。一実施形態では、ある分子のレコードは、この分子中のすべての原子について、これらの原子が配置される空間的な場所、様々な物理的相互作用に対応するエネルギー・パラメータ、原子群についての(例えば分子のあまり関連しない部分、例えば活性部位から離れたタンパク質の内側の部分についての)近似、ならびにこの分子についての他のデータ、例えば、原子の種類、原子の位置、および結合部の一覧を含む。図4aおよび図4bに、本発明に関連する標的、リード、および潜在的な反応物質分子についての関連するデータの記憶部を列挙する一実施形態によるサンプル・レコードを示す。

【0104】

例えば、図4aに、本発明の一実施形態に従って、pdbエントリ4dfrに関連する標的タンパク質のジヒドロ葉酸還元酵素を標的分子レコード400で表す。図4aでは、項目403は、このタンパク質に関連する1つまたは複数の名称を表し、項目405は、このタンパク質に関わるソース情報、機能、および反応／プロセスに係するバイオインフォマティクスおよびプロテオミクスの注釈を表す。項目407は、このタンパク質の機能的／構造的な分類およびファミリー情報を表し、配列および／または構造の比較に基づく相同体との関連も含み得る。

【0105】

図4aを継続して参照すると、項目410は、このタンパク質のアミノ酸配列を、予測

される２次構造要素の注釈とともに表している。ここで、項目４１１、４１２、および４１３はそれぞれ、 β シート・ストランド、 α ヘリックス、および様々なターンおよび／またはヘアピンを表す。項目４１４は、このタンパク質についての構造注釈、例えば、引用文献、結晶分解能、および鎖情報を含めて実験的な構造データに関する有効な p d b ヘッダに見られるものを表す。

【０１０６】

項目４１５は、塩イオン、構造水酸素、またはこのタンパク質構造に関連する他の分子など、（P D B規格〔６２〕に関して定義された）個々のタンパク質原子または異種原子の化学的な身元および空間座標に関連するデータ、すなわち、p d bフォーマット・ファイルのA T O MまたはH E T A T Mのレコードに類似のブロックを表す。本発明のある種の実施形態による標的分子レコードの項目４１５に格納されるデータの例については、図１dに関連する記載に戻って参照されたい。本発明の他の実施形態では、ある種の形式の計算処理を用いて、欠落した、または「壊れた」残基および原子を検出し、次いで、推定される場所に欠落した構造構成要素を配置することによってその構造を修復する。このような修復機能を実施し得る現在入手可能なソフトウェア・ツールの例には、T r i p o s社のS Y B Y L（商標）、U C S FのC h i m e r a（商標）、およびW h a t I f（商標）が含まれる。

10

【０１０７】

標的レコードが、結合した複合体に対応する場合、リガンドに関連する異種原子を含めることもできるし、含めなくてもよいことに留意されたい。一実施形態では、図４aの項目４１５の構造データは、未処理の測定値を含み得る。別の実施形態では、データの値は、任意の別の計算によるモデル化に対するプリプロセッサとして働く標準エネルギー正規化ルーチンの結果として生成される。別の実施形態では、標的分子レコードは、実際には、未処理構造データおよび最終処理済み構造データならびに様々な中間状態を含み得る。

20

【０１０８】

項目４２３は、p d bまたはm o l 2連結レコードに類似のデータ・ブロックを表し、様々な化学結合、すなわち共有結合、ジスルフィド結合を列挙し、タンパク質内水素結合または構造水との水素結合さえ含み得る。本発明のある種の実施形態による標的分子レコードの項目４２３に格納されるデータの例については、図１dに関連する記載に戻って参照されたい。

30

【０１０９】

一実施形態では、項目４２５は、全電荷および部分電荷（列４２７）、質量（列４２８）、v d w半径（列４２９）、およびレナード・ジョーンズ１２－６ポテンシャルに従うウェル深さパラメータ（列４３０）を含めて、タンパク質原子（列４２６）に関連する非結合エネルギー・パラメータを含むフィールドである。典型的には、非結合エネルギー・パラメータの最終的な選択は、様々なエネルギー成分に関する相互作用モデルによって決まる。例えば、溶媒和／疎水作用についてのモデルは、スタウテン他〔７３〕の断片溶媒和モデルに従う原子溶媒和パラメータを必要とすることがある。別の例として、レナード・ジョーンズ１２－１０ポテンシャルに従う水素結合のドナー基とアクセプタ基の相互作用をモデル化するために水素結合パラメータを必要とすることがある。

40

【０１１０】

一実施形態では、項目４３１は、適切な捻れの変化に関連する二面エネルギーに関して、タンパク質結合部または隣接するタンパク質結合部の基に関連する結合エネルギー・パラメータを含むフィールドである。ここで、列４３２は、ピッツァ・ポテンシャルに基づく捻れエネルギー・モデルとともに用いる様々なパラメータ（例えば、位相、振幅、移動度など）を表す。典型的には、結合エネルギー・パラメータの最終的な選択は、様々なエネルギー成分に関する相互作用モデルによって決まる。例えば、結合伸張、結合角の曲げ、不適切な捻れ、さらにはリング摂動に関係する結合エネルギー成分を表すために追加のパラメータを必要とすることがある。

【０１１１】

50

一実施形態では、図4aの項目425および431に示す非結合および結合のエネルギー・パラメータはともに、Amber96力場[40、41]を用いて割り当てられる。ただし、項目425および431に関連して、エネルギー・パラメータを割り当てるのに選択した方法およびエネルギー・パラメータ自体の性質は、本発明の一実施形態に係る単なる例であることを理解されたい。別の実施形態では、これらのエネルギー・パラメータは、計算実行時に「計算中」に割り当てられ、項目425および431は不要である。

【0112】

最後に、項目433は、標的タンパク質に結合する様々な周知の生体分子（天然またはそうでないもの）との関連および入手可能な場合には任意の関連する反応データを含み得る任意選択のフィールドである。

10

【0113】

本発明に関して使用する各リード候補生体分子ごとにリードのレコードも提供される。例えば、リード候補生体分子のメトトレキサートは、本発明の一実施形態に従って、図4bのリード候補生体分子レコードによって表す。図4bでは、項目434は、リード候補に関連する1つまたは複数の名称を表し、435は、化学式、分子量、化学名、およびCAS登録番号を含めて、この化合物の化学的性質に関連する様々な化学注釈を表す。項目437は、この生体分子の化学的かつ／または構造的な分類を表し、生体分子に関して用いられる任意の数の標準化合物分類子を含み得る。項目437はさらに、化学的かつ構造的な比較に基づく他の化合物相同体との関連を含み得る。

20

【0114】

継続して図4bを参照すると、項目445は、mol2フォーマット・ファイルに従う個々の原子の化学的な身元および空間座標に関連するデータのブロックを表している。これは通常、リード候補にイオン化状態を割り当て、リード候補を水素化した後のものである。項目453は、pdbまたはmol2の連結レコードに類似のデータ・ブロックを表し、例えば共有結合などの様々な化学結合の一覧である。本発明のある種の実施形態によるリード候補分子レコードの項目445および453に格納されるデータの例については、図2dに関連する記載に戻って参照されたい。項目455および457は、標的分子に割り当てられるものに類似のやり方で、非結合および結合のエネルギー・パラメータについての別々のデータ・ブロックを表す。本発明のある種の実施形態によるリード候補分子レコードの項目455および457に格納されるデータの例については、図4aの項目425および431に関連する記載に戻って参照されたい。

30

【0115】

項目459は、リードが結合することが知られている様々な周知の標的分子との関連、ならびに入手可能な場合には、関連する反応データを含み得る任意選択のフィールドである。この例では、メトトレキサートを扱う周知の複合体についてのpdb受付コードが列挙されている。最後に、項目460は、リード候補についてのアドメトックス・プロファイルを表す別の任意選択のフィールドであり、これには、臨床前試験または臨床試験の結果として周知の副作用の一覧が含まれることもある。

【0116】

別の実施形態では、適切な計算ツールを用いて構造データが生成されることがある。別の実施形態では、仮想化合物ライブラリに含めることに従って、insilicoで構造データが生成されることがある。別の実施形態では、コンビナトリアル・ケミストリ・ライブラリから構造データを取得することができる。別の実施形態では、構造データは、本質的に2次元または3次元であり得るが、典型的には、2次元情報は、適切な構造生成計算ツールによって変換されることになる。このようなソフトウェア・ツールの例は、CORINAプログラム[74]である。

40

【0117】

本発明の一実施形態では、潜在的な反応物質のレコードは、図4aに関して説明した標的分子レコードのものに極めて類似した形式をとり得る。別の実施形態では潜在的な反応

50

物質のレコードの形式は、実際には、特に潜在的な反応物質が生体高分子であるときには、標的レコードのものと形式が同じであることさえある。しかし、典型的には、注釈情報も周知の結合剤もあまり多くなく、したがって、入手可能な反応データもほとんどないか、または全くないことがある。さらに、場合によっては、構造データは、直接測定したものではなく、構造モデル予測の結果であることがある。これは、GPCRまたはイオン・チャンネル受容体など、多数の膜関連タンパク質を考慮するときにしばしば当てはまる。大きなタンパク質ファミリーの類似の受容体を考慮するときには、モデルにより構築された構造もしばしば用いられる。

【0118】

本発明の一実施形態では、標的、リード候補、および潜在的な反応物質は、それらが互いに、かつそれらの環境と相互作用する場合、完全な分子（または、2つ以上の分子の化合物）として表され、またモデル化される。別の実施形態では、1つまたは複数の計算プロセスを速めるために、計算プロセスではこのような分子の無関係な、または不活性な部分を除くし、実際の分子が普通なら備える数よりも原子および結合部の数が少ない部分的な分子として標的、リード候補、および／または潜在的な反応物質をモデル化することがある。例えば、一実施形態では、標的タンパク質の1つまたは複数の結合ポケットのところに、またはその近傍に存在するものを除くすべてのタンパク質原子を無視することができる。別の実施形態では、潜在的な水素結合ドナーに取り付いた極水素以外のすべての水素原子を無視することができる。別の実施形態では、これらの水素原子を無視するのではなく、それらの非水素原子の相手とともに1つにまとめる。別の実施形態では、水素原子についての類似の処置が、リード候補および／または潜在的な反応物質中の原子に当てはまることがある。

【0119】

本発明の一実施形態では、タンパク質の内部の表面から離れた残基を含む原子は、それぞれ別個の残基に従って合わせてグループ化され、計算のための1つのユニットとして各グループを扱うことによって、エネルギー・パラメータの割当ておよび後続の計算によるモデル化が高速化される。1つ（または複数）の分子の一部を適切に除去し、代表的なグループによって他の部分を置換することが適切に行われると、計算結果に大きな影響を与えずに、必要とされる計算の数をかなり少なくすることができる。

【0120】

一実施形態では、リードの送達および／またはリードと標的または潜在的な反応物質との反応部位の場所を考慮に入れる。例えば、リードが経口摂取される薬物の場合、計算プロセスでのリードの表現は、このリードが消化系を通過した後に残るリードの部分とし得る。別の例では、潜在的な反応物質が膜結合タンパク質であり、反応部位が細胞表面である場合、この潜在的な反応物質の表現は、この細胞の外側に存在するタンパク質の部分だけを含み得る。

【0121】

本明細書では、「反応」という用語は、反応物質の1つまたは複数の作用を変化させる何らかの化学的または物理的な相互作用を指す。例えば、病気または障害を引き起こす何らかの望ましくない生物学的なプロセスに関与するタンパク質である標的を考える。この場合、この病気または障害の治療には、原因タンパク質と反応し、それによってこの特定の望ましくない作用についてこのタンパク質を不活性にする何らかのリード生体分子、例えば小型の分子リガンドを必要とし得る。

【0122】

反応は、反応物質、環境、または反応物質とそれらの環境との相互作用を改変することである。交差反応は、2つ以上の反応物質、すなわち、リード候補と潜在的な反応物質の間の反応である。この潜在的な反応物質は、標的と異なる別の生体高分子であることが多く、不測の、おそらくは望ましくないものでさえある。反応物質が別の反応物質に結合するときに有害な交差反応が生じ、それによって作用が変化し、また、局所的な環境、組織の一部、器官、さらには生物全体に対する悪影響につながる応答が生成される。このよう

な有害な交差反応の例は、メトトレキサートがジヒドロ葉酸還元酵素以外の肝臓内のタンパク質と結合し、それによって、化学療法処置で一般的な脱毛症、吐き気、および肝臓障害のような有毒副作用が誘発されることである。上記で説明したように、アドメトックス・プロファイルは、吸収、分布、代謝、排泄、および毒性に関係する反応を評価しようと試みるものである。

【0123】

本明細書では、反応性は、1つまたは複数の分子間の化学活性を、それらの環境と組み合わせて表す用語として用いられる。反応性の値は、本質的に、定性的（完全に反応する、極めて反応性が高い、あまり反応性が高くない）、定量的（結合自由エネルギー、エンタルピー、反応の度合いまたは速度、解離定数または阻害定数の推定、実験的に導出される信号強度、ハイスループット・スクリーニングの結果、Q S A R予測[53]、[54]など）、観察的（ある種の特徴を示す、ある作用を誘発する、観察される症状、徴候）、または実験的（知識またはルールに基づく予測、類似の分子の組合せからの示度）なものとし得る。交差反応性は、先に論じたように、潜在的に有害な、またはそうでない交差反応の反応性を指す。

10

【0124】

反応の尺度は、1つまたは複数の反応性の値の関数である。本発明の一実施形態では、反応の尺度は、結合親和性の推定値を表す1つの反応性の値によって決まる。別の実施形態では、反応の尺度は、複数の反応性の値によって決まる。例えば、これらの反応性の値の1つは、結合親和性の推定値を表し、別の反応性の値は、Q S A R解析の結果を表す。別の実施形態では、反応の尺度は、1つまたは複数の反応性の値の多項式関数である。別の実施形態では、反応の尺度は、1つまたは複数の反応性の値の非線形関数である。これは、例えば、結合エネルギーの推定値に基づくボルツマン確率分布である。交差反応の尺度は、先に論じたように、潜在的に有害な、またはそうでない交差反応に関する反応の尺度を指す。

20

【0125】

リード候補と潜在的な反応物質を扱う組合せに割り当てられる反応プロファイルは、反応の尺度および他の変数の関数である。ある種の実施形態では、反応プロファイルは、様々な生物学的かつ／または化学的な注釈を含み得る。このような注釈は、リピンスキーの「ルール・オブ・ファイブ」に基づく分類、標的／リード／潜在的な反応物質分子のソース／機能／相同体注釈、アドメトックス・プロファイルの結果、反応部位の場所、log Pなどを含み得るが、これらに限定されるものではない。この反応プロファイルは、定性的なもの（例えば、等級または分類）、あるいは定量的なもの（例えば、複合スコアまたは数学関数）とし得る。交差反応プロファイルは、先に論じたように、潜在的に有害な、またはそうでない交差反応に関する反応の評価を指す。

30

【0126】

ソース注釈は、標的／リード／潜在的な反応物質分子の宿主生物、器官、組織、膜、細胞型、野生型の知識を含み得る。

【0127】

機能注釈は、例えば、タンパク質についての機能の知識、タンパク質をコードする遺伝子の機能、あるいはこの生体分子に対応する生物経路または反応サイクルの知識を含めて、標的／リード／潜在的な反応物質分子についての機能の知識または予測を含み得る。

40

【0128】

相同体注釈は、クエリ標的／リード／潜在的な反応物質に何らかの形で類似すると考えられる他の生体分子、すなわち「クエリ分子」に関連する情報を含み得る。本発明の一実施形態では、クエリ分子がタンパク質のとき、相同体関連物は、クエリ分子のアミノ酸配列と、周知のペプチドまたはタンパク質に対応するアミノ酸配列の基準集合または基準データベースとの間の配列レベルの相同性に基づいて割り当てることができる。別の実施形態では、相同体関連物は、クエリ分子をコードする核酸配列（該当する場合）と、遺伝子、ミトコンドリア遺伝子、mRNA、tRNA、リボソームRNAなどに対応する周知の

50

核酸配列に対応する核酸配列の基準集合または基準データベースとの間の配列レベルの相同性に基づいて割り当てることができる。別の実施形態では、基準集合または基準データベースを含む上記アミノ酸および核酸の配列は、様々な異なる種に由来することがあり、生物学的な機能の点で既に特徴付けられていることもあり、そうでないこともある生体高分子に対応することがある。

【0129】

一実施形態では、クエリ分子がポリペプチドまたはタンパク質のとき、相同体関連物は、周知の公共の、および所有権下にあるタンパク質との、タンパク質モチーフ、2次構造要素、保存領域、および入手可能な場合にはタンパク質3次構造に基づく構造相同性に基づいて割り当てることができる。別の実施形態では、クエリ分子がタンパク質でなく、例えば、一般有機化合物である場合、様々な化学的かつ／または構造的なクラスタリングおよび分類の技術〔75、76〕を利用して、周知の化学構造との構造相同性に基づいて相同体関連物を割り当てることができる。

10

【0130】

本発明の別の実施形態では、同じ種内での、または異なる種にまたがる相同体注釈を用いて、相同基準分子のファミリー分類、系統樹その他の機能注釈に基づいてクエリ分子についての追加の機能注釈を推測することができる。

【0131】

リスク評価は、1つまたは複数の潜在的に有害な交差反応の確率を反映したリード候補ごとの概要評価であり、各リード候補と1組の潜在的な反応物質分子との反応に対応する1組の反応プロファイルに基づくものである。リスク評価の値が大きいことは、1つまたは複数の標的分子についての所望の作用部位と同じ環境においてか、あるいは、例えば消化管内など、薬物送達／分配プロセスの何らかの他の生物学的な段階においてかにかかわらず、リード候補が、1つまたは複数の毒性副作用のために安全でないか、複数の望まれない副作用のために望ましくないか、あるいは、様々な潜在的な重要反応物質との反応性が高いために薬効がないかを示唆している。

20

【0132】

リード候補についてリスク評価の値を求めることは、このリード候補が1つまたは複数の所望の標的分子と反応する際の反応データを必要とすることもある。リスク評価の大小に応じて、リード候補を改変して、1つまたは複数の所望の標的との結合特異性を増大させることが望ましいことがある。一実施形態では、リスク評価の値が大きすぎる場合、リード候補全体を破棄することがある。所与のリード候補との反応プロファイルを入手可能な潜在的な反応物質分子が多いほど、リスク評価が正確かつ堅固になる可能性が高くなる。リスク評価は、新しい潜在的な反応物質分子が入手可能になったときに更新することもできる。

30

【0133】

図5を参照すると、1つまたは複数のリードと潜在的な反応物質の集合との間の交差反応の値を求め、対応する反応プロファイルを生成し、各リード候補ごとのリスク評価を形成し、所望の場合には、薬物の再設計および最適化の手順に入って、ある種のリードの選択性を改善するコンピュータ援用モデリング・システム500が示されている。

40

【0134】

図5では、項目502は、潜在的な反応物質分子のデータベースまたはライブラリを表す。一実施形態では、潜在的な各反応物質分子に関連する情報は、図4aに示す形式の分子レコードに格納され、処理中にこれらのレコードは、システム500の他の部分に渡される。図5では、これらの潜在的な反応物質分子レコードを506と示す。潜在的な反応物質データベース502および関連する潜在的な反応物質分子レコード506に関する様々な以下の実施形態は、リード・データベース504および関連するリード分子レコード508に等しく良好に適用されることに留意されたい。

【0135】

本明細書では「潜在的な反応物質分子」と称するが、ある種の実施形態では、潜在的な

50

反応物質は単分子に限定されないことを理解されたい。潜在的な反応物質データベースは、商業的かつ／または学術的な状況で一般に用いられる、ファイル・システム上に格納されるフラット・ファイルの集合、リレーショナル・データベースその他のタイプの情報データベースとし得る。

【0136】

一実施形態では、データベース502は、タンパク質データ・バンクの既知の非冗長タンパク質をすべて表し、周期的に更新して最新の情報に保つことができる。別の実施形態では、データベース502は、GenBank、SwissProtなどにおけるゲノム情報の様々な翻訳物を表すアミノ酸配列に基づいてモデルにより構築されたタンパク質構造を含み得る。別の実施形態では、データベース502に格納される構造の一部は、保存領域、結合部位、個々の鎖などを含むものなど、様々な生体高分子の関連するサブセットを表すことができる。

10

【0137】

別の実施形態では、データベース502を構成する潜在的な反応物質分子の選択は、1つまたは複数の標的分子の事前知識または特徴付け、ならびに可能なリード候補分子が関わる反応予想部位の局所環境の影響を受ける。一実施形態では、潜在的な反応物質分子は、2次構造または3次構造、あるいはアミノ酸配列のレベルで、1つまたは複数の所望の分子の全部または一部に相同なタンパク質その他のポリペプチドを表すことがある。一実施形態では、潜在的な反応物質分子は、1つまたは複数の所望の標的分子を扱う生物経路、またはその代わりに、1つまたは複数の所望の標的分子についての薬物候補の分配、排泄、代謝、および吸収に関係する基本的な生物学的プロセスに関連する生物経路に関わることが知られている生体分子を表すことがある。データベース502を形成するために潜在的な反応物質分子を選択し得る多くのやり方があるが、これらはしばしば、所望の1つ（または複数）の標的分子およびそれらに付随する生物経路に関する1種または複数種の考察を必要とすることになる。別の実施形態では、潜在的な反応物質データベース502は、交差反応について試験すべき様々なリード候補が対象とする1つまたは複数の標的分子についての分子レコードも含む。図4aおよび図4bに関して既に述べたように、潜在的な反応物質に関する分子レコードおよび標的生体分子に関する分子レコードは、本質的に極めて類似したものになることがある。

20

【0138】

図5では、項目504は、リード候補分子のデータベースまたはライブラリを表す。一実施形態では、各リード候補に関連する情報は、図4bに示す形式の分子レコードに格納され、処理中にこれらのレコードは、システム500の他の部分に渡される。図5では、これらのリード候補分子レコードを508と示す。このリード候補データベースは、商業的かつ／または学術的な状況で一般に用いられる、ファイル・システム上に格納されるフラット・ファイルの集合、リレーショナル・データベースその他のタイプの情報データベースとし得る。

30

【0139】

データベース504を構成するリード候補分子は、多くのやり方で選択することができる。一実施形態では、リード候補分子は、1つまたは複数の所望の標的生体分子に適用する様々なリード創製プロセスおよび／または他のリード最適化プロセスの結果である。先に論じたように、このようなリード創製および最適化のプロセスは、合理的な薬物設計、QSAR、構造に基づく設計、化合物または仮想ライブラリのin silicoスクリーニング、機能アッセイ、マイクロ・フルイディクス、他の細胞に基づくアッセイまたは生化学アッセイなどを含むハイスループット・スクリーニング、あるいはこれらの任意の組合せを含み得る。次いで、このようなプロセスに合格したリード候補をデータベース504のメンバとして組み込むことができる。別の実施形態では、リード候補分子は、普通なら臨床前検査に進む1組のリードを表すことがある。

40

【0140】

別の実施形態では、データベース504に組み込まれるリード候補分子は、1つまたは

50

複数の所望の標的生体分子についての構造的かつ化学的な多様性および任意の事前特徴付けに基づいて、既存の仮想的な、またはそうでない化合物ライブラリからのサブセットとして選択し得る。別の実施形態では、リード候補分子の集合は、シード構造、すなわちコアまたは枠組として使用することができ、これに、分子構造生成技術および／またはコンビナトリアル・ケミストリ技術用の標準の計算ツールを適用して、元のコア構造に基づく別の代表的な構造を備えた追加の化合物の集合を生成する。データベース504を形成するために有力なリード候補分子を選択し得る多くのやり方があり、これらはしばしば、対象とする1つ（または複数）の標的分子およびそれらに付随する生物経路に関する1種または複数種の考察を必要とすることが当業者には理解されよう。実際、多くの場合、データベース504中のリード候補分子は、何らかの他のやり方で、1つまたは複数の所望の標的分子の挙動を抑制し、これらに触媒作用を及ぼし、また影響を及ぼすと予想される。

10

【0141】

一実施形態では、データベース504は、後でより詳細に論じる薬物の再設計および最適化の手順の結果として生成される構造についての分子レコードも含み得る。

【0142】

別の実施形態では、データベース502および504に格納される分子レコードおよび関連するデータを様々な情報圧縮方式を利用して圧縮し、それによって、分子の原子および結合部に関連する潜在的に無数の数値パラメータを格納するコストを下げるができる。別の実施形態では、データベースから関連するデータにアクセスするとき、データベースの処理時または要求時に、その実行中にこれらのデータを解凍することができる。

20

【0143】

データベース502および504中の分子レコードを、バイオインフォマティックスの様々な化学的、構造的その他の技術によってクラスタ化して、データベースの一部選択およびクエリを処理しやすく、また理解かつ再検討しやすくすることもできる。一実施形態では、あらかじめデータベース502をクラスタ化し、クラスタの代表を選択してモデリング・システム500の別の処理ステップによって処理し、それによって全体の計算時間を速くする。次いで、リード候補とクラスタ化された潜在的な反応物質の1組の代表について生成された結果を再検討し、大きな反応性を示したクラスタの代表について、データベース502から対応するクラスタ全体を取り出し、モデリング・システム500によるさらなる処理を開始し、それによって、問題の1つ（または複数）のリード候補についてのリスク評価が完了する。このようにして、多数の潜在的な反応物質に対して多数のリードをスクリーニングすることができ、次いで、1つまたは複数のリードと有望なクラスタの個々のメンバとの間の個々の処理を行うことができる。このような実施形態により、所与の労力で、1組の可能なリード候補と潜在的な反応物質をより効率的に扱うことができる。

30

【0144】

図5を再度参照すると、処理ブロック520は、反応決定部を表している。

【0145】

1つまたは複数の潜在的な反応物質のレコード506を、1つまたは複数の可能なリード候補のレコード508とともに反応決定部520に提供する。あるいは、反応決定部520は、処理中に状況の必要に応じてデータベース502および504から直接、追加の分子レコードを先取りすることができる。反応決定部520は、1つまたは複数のリード候補レコード508および1つまたは複数の潜在的な反応物質のレコード506について、1組の反応の尺度を決定する。

40

【0146】

反応の尺度はそれぞれ、1対のレコードに対応する。一方のレコードはリード候補に対応し、もう一方のレコードは潜在的な反応物質に対応する。この反応決定部は、各対を順に処理することもできるし、複数の対を並列に処理することもできる。典型的な実施形態では、1つのリード候補のレコードおよび複数の潜在的な反応物質分子が反応決定部に提示され、反応決定部は、リード候補分子および1つの潜在的な反応物質分子からなる対ご

50

とに反応の尺度を生成する。別の実施形態では、1つの潜在的な反応物質のレコードおよび複数のリード候補のレコードが反応決定部520に提示され、その後反応決定部520は、リード候補分子および1つの潜在的な反応物質分子からなる対ごとに反応の尺度を生成する。

【0147】

反応の尺度を構築するために、反応決定部520は、コンピュータ記録可能媒体上の格納部522から取り出した追加の入力パラメータ・データに基づいて1つまたは複数の反応性の値を直接生成することもできるし、コンピュータ記録可能媒体上の格納部524から1つまたは複数の既に計算した反応性の値を取り出すこともできる。一実施形態では、この反応決定部は、格納部522に置かれた様々な力場パラメータ・データ、例えばAMBER96力場に関連するパラメータを使用する分子力学法に従って、リード候補と1つ（または複数）の潜在的な反応物質レコードとの間の結合自由エネルギーについての推定値を計算する。別の実施形態では、反応決定部520は、直接にはモデリング・システム500の一部でさえないことがある別の処理モジュールによって事前に、または並列に計算した反応性の値を取り出す。例えば、リード候補および1つ（または複数）の潜在的な反応物質分子についてのQSAR計算をオフラインで計算し、次いで、ファイル・システムまたはコンピュータ・メモリなどのコンピュータ記録可能媒体上に適切に格納し、その後、関連する反応の尺度を形成するときに、反応決定部が取り出すことができる。別の実施形態では、反応決定部520内にQSAR計算用の専用処理要素を埋め込む。

【0148】

反応決定部520は、1つまたは複数のリード候補と潜在的な反応物質対の反応の尺度を表すデータを、潜在的な反応物質ーリード候補用反応フィルタ530に出力する。反応フィルタ530は、入力された反応の尺度に基づく様々な判定基準を適用して、特定の潜在的な反応物質ーリード候補対をモデリング・システム500の後続の段階で利用可能にすべきかどうかを決定する。原理的には、リード候補ー潜在的な反応物質対の多くは、反応性をわずかしき示さないか、または全く示さないことがあり、そのため、対応する反応の尺度は極めて小さい。したがって、場合によっては、反応の確率が低く見える、すなわち、結合親和性が悪く見える対をフィルタリングして除去することが望ましいことがある。

【0149】

一実施形態では、反応フィルタ530は、反応の尺度の大きさが所定の閾値よりも大きいリード候補ー潜在的な反応物質対しかシステム500の後続の段階に通過させない。別の実施形態では、反応フィルタ530は、反応決定部520によってすべての関連する対が処理されるまで待ち、これらの対応する反応の尺度によってランク付けを行い、次いで、上位X%（Xは、1と100の間の選択した数）を通過させる。別の実施形態では、反応フィルタ530は、所定の閾関数を適用する。この所定の閾関数は、入力される関数の引き数として1つまたは複数の反応の尺度を取り、次いで、変換された反応の尺度、例えば、1つ（または複数）の反応の尺度の多項式関数または指数関数に閾論理を適用する。

【0150】

別の実施形態では、特定のリード候補ー潜在的な反応物質対についての反応の尺度をよりよく判定するために、反応フィルタ530は、リード候補が反応かつ／または結合すると予想される1つまたは複数の所望の標的分子と反応した同じリード候補についてのあらかじめ集計された関連する反応データを使用する。例えば、関連する反応データが入手可能な場合、反応フィルタ530は、反応の尺度が、リード候補と所望の標的分子についての反応の尺度に定数を掛けたもの未満であるリード候補ー反応物質対をフィルタリングして除去することができる。ここで、前記定数は通常、ゼロと1の間の値である。別の実施形態では、反応フィルタ530は、代わりに、潜在的な反応物質と結合することが知られている1つまたは複数の生体分子、すなわち、この潜在的な反応物質に関わる周知の結合剤または複合体と反応した潜在的な反応物質についてのあらかじめ集計された関連する反応データを使用する。しかし、多くの場合、このような追加の反応データは入手可能でな

いことがあり、そのため、反応フィルタ 530 は必ずしもそれを要求しない。

【0151】

他の実施形態では、反応フィルタ 530 を通過することができないリード候補—潜在的な反応物質対も、それらの対応する反応の尺度その他のデータを転送して、1 つまたは複数のデータベース・テーブルまたはフラット・ファイル中のコンピュータ記録可能媒体（項目 532）に格納する。別の実施形態では、すべてのリード候補—潜在的な反応物質対は、反応フィルタを通過することができ、反応フィルタ 530 は単に、関連する反応データを取り込み、その結果をデータ構造 532 に丸ごと格納する機構として働く。

【0152】

ある種の実施形態では、反応フィルタは、本発明に不可欠ではないとみなされることがあり、すべてのリード候補—潜在的な反応物質対は、モデリング・システム 500 の後続の段階へと通過する。

【0153】

反応フィルタ 530 は、反応プロファイル 550 へと通過するリード候補—潜在的な反応物質対に対応する反応データを出力する。反応プロファイル 550 は、前に定義したように、530 から入力される対応する反応の尺度その他の変量に基づく各リード候補—潜在的な反応物質対ごとの反応プロファイルの構築を司る。先に述べたように、これらの他の変量は、例えば、リピンスキーの「ルール・オブ・ファイブ」に基づく分類、関連する標的／リード／1 つ（または複数）の潜在的な反応物質分子のソース／機能／相同体注釈、アドメトックス・プロファイルの結果、反応部位の場所、log P、log D など

【0154】

潜在的な反応物質分子は、相同体注釈について前に説明した目的で、他の生体分子に関連する相同体も含み得る。一実施形態では、これらの相同体関連物は、潜在的な反応物質分子データベース 502 に記録として格納された他の潜在的な反応物質分子を指す。別の実施形態では、潜在的な反応物質分子の記録は、図 5 の格納バッファ 552 で表す専用のデータ構造またはファイルにバッファされる。別の実施形態では、バッファ 552 は、潜在的な反応物質データベース 502 に直接関連しないコンピュータ記録可能媒体上のデータベースまたはデータ構造からロードされる。

【0155】

別の実施形態では、異なる格納バッファ 554 は、1 つ（または複数）のリード候補が結合することが知られている 1 つまたは複数の標的生体分子を格納するためのデータの構造またはファイルを表す。このような実施形態では、バッファ 554 に格納されたデータを用いて、1 つ（または複数）の標的およびリード候補を扱う分子の組合せの反応の尺度を生成することができる。次いで、得られた反応の尺度と、このリードと 1 つまたは複数の潜在的な反応物質に関連する反応の尺度とを比較することができる。これは例えば、結合親和性についての相対比較である。一実施形態では、この 1 つ（または複数）の比較では、1 つまたは複数の定量的なリード—標的反応の尺度を用いて数値的な閾値を生成し、それによって、同じリードと潜在的な反応物質についての定量的な反応の尺度をさらにフィルタリングまたは重み付けする必要があることがある。別の実施形態では、格納バッファ 554 は、識別された標的分子に相同の生体分子の記録も含む。一実施形態では、これらの標的記録および／または相同標的記録を格納し、潜在的な反応物質データベース 502 から、適切な相互参照によってリード候補データベース 504 中で直接アクセス可能である。一実施形態では、バッファ 554 は、データベース 502 または 504 に直接関連しない入力される様々なデータベースまたはデータ構造からロードされる。

【0156】

バッファ 554 に格納されるものなどの情報は、一般に有用であるが、本発明に本質的なものではなく、このようなあらかじめ識別された標的分子が存在しない場合、バッファ

554は本質的に残っているだけで使用しない。本発明の別の実施形態では、個々のリード候補－潜在的な反応物質対に対応する反応プロファイルを構築する際に、反応プロファイル550はバッファ552および554のいずれも使用しない。

【0157】

反応プロファイルは、例えば等級などの本質的に定性的なもの、または、例えばスコアまたは関数値などの本質的に定量的なものとし得る。反応プロファイル550は、反応フィルタ530から提示されたリード候補－潜在的な反応物質対についてのみ反応プロファイルを生成することを想起されたい。一般に、リード候補－潜在的な反応物質対の反応プロファイルを形成するとき、1つ（または複数）の反応の尺度、リード候補および潜在的な反応物質についての注釈、ならびに、アドメトックス・プロファイルまたは様々な生化学アッセイの結果など、他の特性を含めて、入手可能な反応データが多いほど、反応プロファイルが質の高い、すなわち、適切かつ有意義なものである確度が高くなる。

10

【0158】

一般に、反応の尺度は、リード候補と潜在的な反応物質の間の交差反応の強さを反映するが、それ自体では、その交差反応が有害であるか否か、すなわち、有毒であり、また、望まれない副作用を誘発するかどうかについてほとんどわからない。また、反応の尺度はそれ自体では、1つまたは複数の所望の標的生体分子との意図する反応部位におけるリード候補のバイオ・アベイラビリティが、対象の病状または病気処置するのに効果があるほど十分なのかどうかは直接予測されない。

【0159】

20

一実施形態では、ソース、機能、および相同体の注釈を含めて、リード候補および潜在的な反応物質に対応する反応の尺度の大きさおよび注釈データに基づいて、反応プロファイルに信頼度または信頼性スコアを割り当てる。例えば、潜在的な反応物質のソースおよび機能が周知である場合、例えば、ヒトインスリン分子であり、反応の尺度が大きい場合、反応プロファイルの信頼性スコアは高くなり、それによって、反応は強くかつ有害な可能性がある確率が高いことを反映することになる。別の例では、潜在的な反応物質は、脳における周知の機能のタンパク質であり、いかなる交差反応も悪性である可能性が高いが、このリード候補に付随する注釈情報は、このリードが血液脳関門を貫通することができず、そのため、このリードと潜在的な反応物質が互いに遭遇する可能性は少ないことを示すことがある。

30

【0160】

別の例では、潜在的な反応物質自体は良好に特徴付けられないが、マウスの周知の癌遺伝子に関係するマウスタンパク質との構造的かつ配列のレベルで相同性が高いことが示される。先に説明したように、相同性注釈を用いて機能を推測することができる。したがって、リード候補が、上記マウス相同体潜在反応物質について高い反応の尺度を示す場合、このリードはヒト宿主生物においても発癌性を示す確度が高いと推論し得るはずである。一方、潜在的な反応物質との反応が潜在的に危険であっても、反応の尺度が小さいので信頼性スコアが比較的低下することがある。というのは、リードとの何らかの反応が生じることが疑わしいことがあるからである。

【0161】

40

別の例では、潜在的な反応物質が、ヒト免疫系に関わるタンパク質ファミリーのメンバーと分類される場合、反応の尺度が大きいと、このリードに対する免疫系の応答が潜在的に有害である確度が高く、おそらくは、アレルギー反応を誘発することになると推測される。別の例では、信頼性スコアが高いことは、様々な注釈および分類によって様々な消化プロセスに関わる生体分子と特徴付けられた潜在的な反応物質との強い相互作用のためにバイオ・アベイラビリティが低下することを反映していることがある。別の例では、信頼性スコアが高いことは、薬効が欠如していることを反映していることがある。というのは、潜在的な反応物質が、所望の標的との意図する反応部位の同じ局所的な環境に存在する酵素であり、この潜在的な反応物質がリードと極めて強く結合して、標的と競合し、それによってリード候補の作用を実質的に低下させるからである。

50

【0162】

別の実施形態では、信頼性スコアの代わりに、1つ（または複数）の反応の尺度の大きさ、ならびに潜在的な反応物質およびリード候補分子についての周知の入手可能な情報を反映する等級その他の定性的な分類を用いる。別の実施形態では、リード候補についてのアドメトックス・プロファイルが、このリードの様々な代謝、吸収、分布、排泄、および有毒な特性に関して既知であるか、あるいは、別の分子との構造的または機能的な類似性に基づいて外挿されるとき、このアドメトックス・プロファイルを用いて反応プロファイルの構築を補足することができる。例えば、リードが生物の消化系を無傷で通過して血流に効率的に吸収される確度についての知識、血液脳関門などの主要な膜を貫通する透過性、発癌性などを用いて、特定の潜在的な反応物質との反応が生じ、またそれが有害であるかどうかを推測することができる。

10

【0163】

反応プロファイル 550 は、各リード候補—潜在的な反応物質対ごとに、反応プロファイルリスク評価解析部 560 に出力する。リスク評価解析部 560 の役割は、入力された反応プロファイルに基づいて、リード候補分子に関わる 1 組の反応プロファイルを解析した後で、リード候補ごとのコンセンサスまたは概要評価を形成することである。ある種の実施形態では、リード候補ごとのリスク評価は、この 1 組の反応プロファイルに基づく定量的なスコアまたはランク付けである。本発明の一実施形態では、リード候補に関わる 1 組の定量的な反応プロファイルを合計してリスク評価を生成する。別の実施形態では、リスク評価は、これらの反応プロファイルの 1 次結合である。別の実施形態では、リスク評価は、多項式関数または非線形関数など、定量的な反応プロファイルの関数である。

20

【0164】

別の実施形態では、リスク評価は、ニューラル・ネットワークまたはサポート・ベクトル・マシンなどのコンピュータ学習手法から出力された予測を表す。この場合、定量的な反応プロファイルは入力として働き、学習アルゴリズムは、好都合なリードおよび不都合なリードを含めて、副作用および薬効の点で既に特徴付けられた化合物のデータの組について訓練される。別の実施形態では、リスク評価は、上位 N 個の反応プロファイルの平均として求められ、それらの対応する信頼性尺度に基づいてランク付けされる。ここで、N は、例えば正の小整数である。別の実施形態では、リスク評価は、定性的な反応プロファイルに基づいて得られたコンセンサス等級その他の定性的な分類である。別の実施形態では、リスク評価は、複数の反応プロファイルに統計解析を適用することに基づいて生成された信頼性のスコアまたは確率である。この統計解析には、多変量解析（例えば、z スコア、カイ自乗など）またはベイズ解析（例えば、ベイズ・リスク）が含まれるが、これらに限定されるものではない。

30

【0165】

このシステムの他の構成要素の場合と同様に、リスク評価解析部 560 は、より多くの潜在的な反応物質および／またはより多くの反応データが入手可能になったときに、所与のリード候補についてのリスク評価を更新することができる。一実施形態では、リスク評価解析部 560 は、リード候補ごとに、専用のデータ構造にリスク評価を出力して、記録可能なコンピュータ媒体上に格納し、後で再検討し解析することができる。

40

【0166】

この考察をよりよく例示するために、再度、図 2 a～図 2 d に示し、図 4 b に示す分子レコードにも表すリード候補であるメトトレキサートを考える。メトトレキサートは、図 1 a～図 1 d に示す標的分子であるジヒドロ葉酸還元酵素（DHFR）を抑制する。メトトレキサートは、実際には、ある種の白血病を含めて様々な癌を治療する化学療法で使用する癌治療薬である。モデリング・システム 500 を使用して、化合物メトトレキサートについてのリスク評価を生成すると、いくつかの結果を観察することができる。

【0167】

第 1 の結果では、メトトレキサートは、大腸菌由来のチミジル酸シンターゼを表す潜在的な反応物質とかなり強く結合することが観察されることがある。大腸菌チミジル酸シン

50

ターゼ（E C T S）は、病原性大腸菌内でデオキシウリジン 5' - リン酸をデオキシチミジン 5' - リン酸に変換することに関係する *de novo* DNA 合成における不可逆ステップを制限することに関わるタンパク質である。図 6 a ~ 図 6 c に、大腸菌チミジル酸シンターゼと、図 3 a ~ 図 3 c に示すものに類似しているが、ここでは P D B エントリ 1 a x w から得られた構造データを有するフォーマットのメトトレキサートとの分子複合体の表現を示す。E C T S に関連する反応の尺度は、D H F R に関連する反応の尺度よりも弱いことがあるが、依然としてかなり大きい可能性がある。E C T S - メトトレキサート対についての反応プロファイルを構築するとき、E C T S に付随する注釈情報を調べると、T S の大腸菌変異体は、ヒトの変異体に極めて相同していることがわかる。E C T S に関連する機能注釈をさらに調べると、癌を治療する際、これは実際に有益な標的であり、同様に抑制することが示唆される。というのは、腫瘍の成長に必須である新しい D N A 合成をさらに下方制御することによって、腫瘍細胞成長がさらに制限されることもあるからである。

【0168】

第 2 の結果では、メトトレキサートは、ウサギの細胞質セリンヒドロキシメチル転移酵素（または S H M T）を表す第 2 の潜在的な反応物質に、ある種の親和性で結合することが観察されることがある。この転移酵素は、とりわけ、メチオニンの産生、セリンの可逆産生、およびグリシンの産生を含めて、複数の反応の触媒作用に関わる酵素である。ヒトの S H M T をコードするヒト・ミトコンドリア遺伝子との配列相同性が高いことは、メトトレキサートが、S H M T に関わる反応経路に影響を及ぼすことがあることを示唆している。メトトレキサートの濃度または投薬量が多いと、細胞内のある種のアミノ酸合成経路に悪影響を及ぼすことがあるが、メトトレキサートの関連するアドメトックス・プロファイル調べると、メトトレキサートが外部細胞膜を貫通する可能性はかなり低く、そのため、細胞表面受容体との相互作用が、アミノ酸合成およびプリン産生において S H M T が関与する様々なミトコンドリア・プロセスを大きく妨げる可能性は少ないことがわかる。そのため、ウサギ S H M T - メトトレキサート対について得られた反応プロファイルは、それ自体では、ヒトにおいて有害な交差反応が生じる確率が高いことを示さないことになる。図 7 a ~ 図 7 c に、ウサギ細胞質セリンヒドロキシメチル転移酵素とフォリン酸（メトトレキサートに類似の化学相同体）の分子複合体の表現を、図 3 a ~ 図 3 c に示すものに類似のフォーマットで示す。この構造データは、P D B エントリ 1 k 1 2 から得られたものである。

【0169】

第 3 の結果では、メトトレキサートは、ヒト、あるいはマウス、ウサギ、またはイヌなどの他の生物の腎臓におけるある種のタンパク質または酵素を表す 1 つまたは複数の潜在的な反応物質に適度な反応強度で結合することが観察されることがある。ソース／機能／相同体注釈、ならびに、排泄に関する潜在的な懸念を示唆するアドメトックス・プロファイルに基づいて、この反応プロファイルは、メトトレキサートの投薬量が多いと、腎機能に有害である確率が高いことを反映していることがある。

【0170】

このように、例えば、メトトレキサートについての概要リスク評価は、（a）S H M T を扱う生物経路が阻害される可能性は低い、（b）チミジル酸シンターゼとの反応により腫瘍細胞の成長がさらに制限される可能性が高い、（c）メトトレキサートの濃度が高いことに関係する腎臓の問題および／または障害が潜在的に存在するという予測を含み得る。これらはすべて、様々な実施形態に関して先に論じたように、定量的なスコアまたはランク付けの形式で表すことができ、それによって、メトトレキサートのリスク評価が体系的にまとめられる。

【0171】

図 5 に描いた実施形態では、リスク評価解析部 5 6 0 は、薬物再設計フィルタ 5 7 0 に、リード候補ごとのリスク評価も出力する。

【0172】

10

20

30

40

50

薬物再設計フィルタ570は、入力されたリスク評価を調べて、1つまたは複数の所望の標的生体分子に対する選択性を増すために、さらに、リード候補についてのリスク評価が示す識別された潜在的な交差反応の確度または強度を小さくするためにリード候補を改変することができるか、あるいは改変すべきかどうかを決定する。

【0173】

本発明の一実施形態では、薬物再設計フィルタ570が用いる決定プロセスは、使用者が入力する様々なパラメータによって決まる。一実施形態では、これらの使用者入力パラメータは、例えば、反応の尺度が所定の数値的な閾値を上回る潜在的な交差反応の数をどれだけ許容するかなど、本質的に定量的なものとし得る。別の実施形態では、これらの使用者入力パラメータは、例えば、定性的なリスク評価に関連する許容差など、本質的に定性的なものとし得る。

10

【0174】

別の実施形態では、本質的に定性的であるか、定量的であるかにかかわらず、これらの使用者入力パラメータを用いて、低リスクのリード候補を無視し、他のすべてのリードにフィルタ570を通過させることができる。別の実施形態では、使用者が、あらゆる潜在的に有害な、または望まれない副作用がなくなるようにこのようなリードが改変される可能性が小さいとみなすことがあるので、高リスクの候補も無視される。別の実施形態では、フィルタ570は、使用者が指定した数のリードだけを、これらのリードをそれらのリスク評価に基づいて、使用者が指定した順序に従って仕分けした後で通過させることができる。

20

【0175】

薬物再設計フィルタ570は、1つまたは複数の所望の標的分子と反応するリード候補の反応データの知識も含む。例えば、このリード候補の所望の標的との反応は比較的弱く、多くの潜在的な有害交差反応を示すことが知られている場合、このリード候補は、薬物再設計プロセスの労力に見合わないと考えられ、そのため、フィルタ570によって遮断し得る。

【0176】

典型的には、薬物再設計フィルタ570の意図は、可能な化学的かつ構造的な改変を行って、このリードと1つ（または複数）の所望の標的生体分子との選択性を増すこと、すなわち、1つ（または複数）の標的が付随していない潜在的な反応物質との反応性は低くするが、1つ（または複数）の所望の標的分子自体との高い反応性を維持することから利益を得ることができるリード候補を識別することである。この意図はさらに、所望の標的との反応性を改善することも潜在的に含み得る。別の実施形態では、フィルタ570は、可能な化学的かつ構造的な改変を行って、1つ（または複数）の所望の標的分子との反応性を高く維持しながら、有益な潜在反応物質との反応性を増大させることから利益を得ることができるリード候補を識別することもできる。例えば、メトトレキサートの場合、リードを改変して、メトトレキサートの元の標的であるヒト・ジヒドロ葉酸還元酵素との高反応性を保持しながら、ヒト・チミジル酸シンターゼにより良好に結合させ、それによって、癌治療特性がさらに優れた改変リード候補を生成することは可能であり、かつ望ましいことがある。このような場合、問題の潜在的な反応物質は、潜在的な標的と分類し直すことができる。薬物再設計フィルタは、リード候補のサブセットを選択し、これらを薬物再設計／最適化エンジン580に送って、可能な化学的かつ構造的な改変を行うのと同時に、このリード候補の残りの部分を専用のデータ格納部575に送ってさらなる再検討および解析を行う。

30

40

【0177】

図5では、薬物再設計／最適化エンジン580は、薬物再設計フィルタ570が提示した1つまたは複数のリード候補の再設計による可能な改変を探す処理ユニットを表す。これらの改変は、典型的には、リード候補の化学的かつ構造的な変更であり、それによって、このリード候補と1つまたは複数の生体分子との反応性が改変される。例えば、これらの改変では、リードと、リスク評価解析部560で生成されたこのリードのリスク評価で

50

識別された1つまたは複数の潜在的な反応物質との反応性を減少または低下させ、同時に、1つまたは複数の所望の標的分子との反応性の度合いを満足に維持すること、すなわち、リード候補の特異性を増大させることを意図し得る。別の実施形態では、この意図は、化学的かつ構造的な改変を導入して、1つまたは複数の潜在的に有益な交差反応物質との反応性を増大させることである。別の実施形態では、これらの改変は、リードと1つ（または複数）の所望の標的分子との結合親和性を増大させて、薬効を改善し、かつ／または投薬量の要件を減らし、同時に、有害な交差反応の危険性を低く維持することを意図している。

【0178】

以下の考察の目的のために、構造構成要素という用語を、化学基、さらには場合によっては個々の原子と定義する。有効な構造構成要素は、例えば、アセチル基、メチル基、カルボキシル基、アミン基、ヒドロキシル、芳香環および複素環、リンおよび硫黄を含む様々な基などを含み得る。他の有効な構造構成要素は、DNAまたはRNAにおけるものなど、1つまたは複数の核酸、あるいは、化学的な修飾および／または開裂の結果改変されたアミノ酸を含めて1つまたは複数のアミノ酸、あるいは、炭水化物、脂肪酸、さらには、 β シートまたは α ヘリックスなどのタンパク質2次構造の構成要素を含む。他の有効な構造構成要素は、有機化合物に一般に見られるものなど、例えば、炭素、酸素、窒素、硫黄、リン、および水素などの個々の原子、または、塩素、臭素、フッ素などのハロゲン、あるいは遷移金属または鉄、マグネシウムなどの他の金属イオンを含み得る。

【0179】

リード候補構造の可能な改変には、新しい構造構成要素の追加、既存のリード候補構造を含む構造構成要素の置換および／または削除、あるいはこれらの任意の組合せを含む。新しい構造構成要素の追加または既存の構造構成要素の改変により、リード候補の反応性を劇的に変えることができる。

【0180】

例えば、主要な水素結合のドナー基またはアクセプタ基を取り除くと、潜在的な反応物質との反応が不安定になることがある。一方、特定の水素結合のドナー基またはアクセプタ基を追加すると、1つまたは複数の標的分子との結合親和性が増大することがある。他の基を追加すると、立体構造の衝突または反発が導入されることがあり、そのため、好ましい結合様式が阻害され、そうでない場合にはエネルギー的に有利なファンデルワールスおよび静電気による相互作用が低減することから、1つまたは複数の分子との反応性が大きく劣化する。

【0181】

構造構成要素の変化により、リードのイオン化状態または電荷分布が変化することがあり、それによって、リードと、1つ（または複数）の標的分子および1つ（または複数）の潜在的な反応物質の電荷および双極子モーメントによって生成された静電ポテンシャル場との相互作用が大きく影響を受ける。疎水基および／または親水基の導入または削除は、リードの様々な溶媒和特性だけでなく、リードとその溶媒環境、例えば水溶液との相互作用も改変し得る。構造構成要素の変化により、溶媒環境内で分離されているか、標的分子または潜在的な反応物質に近接しているかにかかわらず、リードのエネルギー的に最も有利な配座の性質が変化し得る。構造構成要素の追加、置換、および／または削除により、リガンドの電子密度も影響を受けることがあり、それによって、リードと、1つまたは複数の標的分子または潜在的な反応物質との量子力学的な相互作用が影響を受ける。大部分の実施形態では、元のリードの様々な改変によって得られた分子構造はそれぞれ、適切な原子価および／または正しいイオン化状態という意味で正当なはずであり、ほとんどの場合、化学的に合成可能なはずである。

【0182】

図8aの絵800に、メトトレキサートの構造の2次元表現を表示する。一方、絵820に、メトトレキサートおよびその構造的な仲間の枠組分子構造を、様々な構造構成要素を追加、置換、さらには削除し得る7つの異なる連結部位で強調して示す。図8では、各

連結部位は、802、803、804、806、807、808、および809と示す。5つの連結部位（802、803、804、806、および807）では、記号「R」は、いくつかの可能な構造構成要素のいずれかを示す。6番目の連結部位は、陰影を付けた領域808で表す。ここで、連結部位808は、810と標示したアミド結合の先の分子を置換または削除し得ることを表す。7番目かつ最後の連結部位は、さらに大きな陰影を付けた領域809で表す。ここで、連結部位809は、811と標示した結合の先の分子を置換または削除し得ることを表す。

【0183】

図8aを継続して参照すると、表830に、代替構造構成要素として働き得る1組の分子群または断片の例が列挙されている。表830は、個々の原子種、さらにはアミノ酸、例えば、タンパク質中に一般に見られる20個の残基によってさらに増強することもできる。

10

【0184】

図8bに、絵800のメトトレキサートの最終構造に対して、表830からの代替構造構成要素を、図8aの絵820に示す1つまたは複数の連結部位のところで追加または置換することによって得られた6つの構造の例（項目840、850、860、870、880、および890）を示す。例えば、構造840は、メトトレキサートのメチル基を連結部位804のところで個々の水素原子に置換した結果である。実際には、構造840は、アミノプテリンという名称のジヒドロ葉酸還元酵素の別の周知の結合剤である。しかし、このような些細な構造の改変では、生じ得る有害交差反応の危険性が小さくなる可能性は小さい。

20

【0185】

構造850は、連結部位806および807のカルボキシル基をメチル基で置換した結果である。構造860に、803のところの平面三角形アミン基を sp^2 酸素で置換して、葉酸として一般に知られる化合物が得られたところを示す。構造870に、連結部位808のアミノ基の先の構造構成要素をすべて、20個の標準アミノ酸の1つであるセリン残基で一括置換し、さらに、アミノプテリンの場合と同様に804のところのメチル基を水素原子で置換したものを示す。構造の例880に、連結部位809の先の構造構成要素をすべてエステル基で置換し、さらに、葉酸の場合と同様に803の平面三角形アミン基を sp^2 酸素で置換したものを示す。構造880は、ビオプテリンとしてより一般に知られているものである。最後に、構造890に、連結部位808のエステル基を置換し、さらに、804のところのメチル基をプテリジン環で置換したものを示す。

30

【0186】

この時点で、このような手順の組合せの態様を論じることは価値があろう。N個の異なる代替構造構成要素およびM個の連結部位では、削除を数えなければ M^N 個の可能な構造がある。例えば、 $N=25$ 、 $M=3$ の場合、 25^3 個、すなわち15,625個の可能な構造がある。あるいは、 $N=15$ 、 $M=6$ の場合、11,390,625個の可能な構造がある。ここで、所与の連結部位に、これらの代替構造構成要素のいずれをも配置し得るとは限らないことが多い。さらに、これら可能な組合せの多くは、アドメトックス・プロファイルの点で薬物にふさわしくない選択をもたらすこともあるし、実際、その化合物を化学的に合成し得ないこともある。さらに、あらゆる可能な組合せをそのまま探す必要はないことがあり、また、得策でないことが多い。上記の数字の例は、例として示したものである。ある種の実施形態では、所与の時間で改変可能なのは、可能な連結部位の一部のみ、例えば、6つの可能な連結部位の3つだけである。これは、 $N=15$ では、異なる構造は67,500個だけであることを示唆している。

40

【0187】

所望の作用に応じて、薬物再設計／最適化エンジン580は、いくつかの構造的な組合せを効率よく探し、通常は1つまたは複数の所望の標的分子との反応性を高く維持し、同時に、リードと1つまたは複数の潜在的な反応物質との反応性を改変する（通常は低くする）新しい最適な構造を1つまたは複数選択する必要がある。

50

【0188】

一実施形態では、薬物再設計／最適化エンジン580は、標的の結合キャビティと交差反応物質の結合キャビティを比較して異なる部分を識別する計算方法を採用する。次いで、キャビティごとに代表形状関数を生成し、その後、前記形状関数の相互の重なり合いを計算することによって、大雑把な評価を高速に行うことができる。重なり合いが少ない領域は異なると考えられる。潜在的な反応物質の結合キャビティの異なる領域を取り出すと、既存のリード候補構造の可能な改変形態について異なる座のところの、またはその近傍の潜在的な連結部位を識別することができる。次いで、リード分子の識別された連結部位のところの1つまたは複数の化学基を追加または改変することができ、それによって、選択された異なる領域の1つまたは複数のところで、反発または立体構造の衝突が生じるか、あるいは、水素結合の発生率および／または強度が低下する。正しく行われると、選択された座は、潜在的な反応物質と標的分子では異なるので、1つ（または複数）の所望の標的分子との反応性は保存されたまま、リードと潜在的な反応物質の反応の強さが弱くなり得る。識別される領域には、対象とする様々な部分間の類似性の尺度を計算するとき、分子の静電的な特徴を考慮に入れることもできる。別の実施形態では、1つ（または複数）の標的分子の結合キャビティの1つまたは複数の異なる領域を選択することができ、これら異なる領域と緊密に接触するリードの構造構成要素に適切な改変を加えて、このリードと1つ（または複数）の標的との反応性を高め、リードと1つまたは複数の潜在的な反応物質との結合相互作用の改変は最小限に抑える。こうすると、このような改変されたリードは、所与の標的とする病状をより低い投薬量で処置することができ、そのため、有害副作用がなくなる。

10

20

【0189】

別の実施形態では、リード候補データベース504からの試験中のリードは、ミニ・リード・ライブラリ自体として扱うことができ、新しいリードを生成する化学的に許容可能なルールに基づいて異なるリードの一部を交換することができる。すなわち、問題のリードの様々な連結部位で試験される代替構造構成要素は、1つまたは複数の他の被検リードから（部分的に）得ることができる。1つのリードがある種の潜在的な反応物質と交差反応しないか、または比較的低い親和性で反応する場合、そのリードの一部を使用するか、それを現在のものと交換して、改変されたリードを生成することができる。特に、標的に良好に結合するリードの部分を保存することができ、他の何らかの部分、（問題の交差反応物質と反応しない）別のリードからの部分と置き換える場合、1つ（または複数）の標的分子に結合する際の特異性を増大させることができる。

30

【0190】

新しい摂動構造が、元のリード候補構造の主要な枠組またはコアからどのように生成されるかにかかわらず、この新しい構造は、再設計プロセスで参照する潜在的な反応物質ならびに1つ（または複数）の所望の標的分子との反応の点で評価しなければならない。一実施形態では、新しい構造は、1つまたは複数の生化学アッセイまたはウェット・ラボ・アッセイから得られる結果に基づいて評価することができる。本発明の別の実施形態では、この評価は、新しい構造と1つ（または複数）の関連する潜在的な反応物質および1つ（または複数）の標的分子ごとに、好ましい結合様式および結合親和性の予測に基づく計算手段によって実現される。

40

【0191】

一実施形態では、計算による予測は、標準形状相補性法、分子力学パラダイム、分子動力学、および／または量子力学シミュレーションに基づく従来技術、Q S A R、自由エネルギー摂動論、さらには、様々な実験的に導出された、または知識に基づくスコアリング関数を使用することを含めて、先の背景技術の項および詳細な技術説明のところで既に論じたいくつかの従来技術の1つを利用して得られる。

【0192】

一実施形態では、静電相互作用は、距離に依存する誘電体を伴う改変されたクーロン・ポテンシャルによってモデル化される。別の実施形態では、分子および局所環境に関わる

50

静電相互作用および静電脱溶媒作用は、ポアソン-ボルツマン方程式の数値解法を利用して計算される。別の実施形態では、静電脱溶媒作用および疎水相互作用はともに、断片ボリュームに基づくモデルを利用してモデル化される。別の実施形態では、静電脱溶媒作用は、改変型ボルン溶媒和モデルを利用してモデル化される。別の実施形態では、ファンデルワールス相互作用は、改変型レナード・ジョーンズ・ポテンシャルによってモデル化される。別の実施形態では、水素結合および金属-リガンド相互作用はともに、改変型レナード・ジョーンズ・ポテンシャルまたは改変型モース・ポテンシャルによってモデル化される。

【0193】

別の実施形態では、標準の分子力学による表現を用いて、結合伸張、結合の曲げ、適切な捻れおよび不適切な捻れ、ならびに様々なリング変換による配座の変化をモデル化する。別の実施形態では、実験的な定式化を利用して、親油性かつ親水性の相互作用を評価する。別の実施形態では、実験的な定式化を利用して、結合後のリードおよび／またはタンパク質の側鎖のエントロピーの低下をモデル化する。別の実施形態では、様々な基底関数の組を用いて、量子力学の手法における適切なハミルトニアン固有状態および固有値を解く。別の実施形態では、このハミルトニアンを簡略化するためにある種の近似を加えて、分子の電子密度を表す量子力学波動関数の解法を高速化することができる。

【0194】

本発明の一実施形態では、上記計算を完了するために必要なエネルギー・パラメータは、先に説明したように、例えば、AMBER [40、41]、OPLS [42]、または Merck Molecular Mechanics 力場 [43] のものなど、分子力場の一部として提供される。別の実施形態では、これらのエネルギー・パラメータは、アニーリング分子動力学シミュレーションその他の最適化手順に従って、経時的または反復してアニーリングすることができる。これら最適化手順には、遺伝子アルゴリズム、ミーム・アルゴリズム、シミュレーテッド・アニーリング・アルゴリズムなどが含まれるが、これらに限定されるものではない。既に説明したように、これらのエネルギー・パラメータの一部または全部は、データベース 502 および 504 中のリード候補および潜在的な反応物質のついての分子レコード中に格納することができる。

【0195】

別の実施形態では、プラカッシュ I の方法および装置を利用して、正確かつ効率的に新しい構造をそれぞれ評価する。

【0196】

一実施形態では、薬物再設計／最適化エンジン 580 から新たに生成された構造は、リード・データベース 504 に再提示され、潜在的な交差反応をモデル化するためにシステム 500 によって再度処理される。ある種の実施形態では、交差反応をモデル化するプロセスの後で、薬物の再設計および最適化が行われる。リードの新しい構造的な変形の創製は、何らかの終端条件その他の基準が満足されるまで反復して繰り返すことができる。このことを、図 5 に、項目 580 と項目 504 を結ぶ矢印で示す。

【0197】

図 8b のメトトレキサートに対して改変された可能な構造 850 では、図 8a の絵 820 の連結部位 806 および 807 のカルボキシル基が取り除かれ、それらの代わりに、水素結合を形成することができない疎水メチル基が用いられる。これら 2 つのカルボキシル基は、ジヒドロ葉酸還元酵素のある種の表面残基に対して強い水素結合を形成しないが、その代わりに、まず間違いなく溶媒水分子と結合するので、これらを削除すると、メトトレキサートと所望の標的であるジヒドロ葉酸還元酵素との結合がわずかに不安定になることがある。しかし、これらの新しいメチル基は、取り除かれたカルボキシル基に立体構造がいくらか類似しているので、この変化により、メトトレキサートと所望の標的であるジヒドロ葉酸還元酵素との結合が過度に不安定にならないはずである。一方、これらを取り除き、疎水基と置き換えることにより、例えば、ヒトの腎臓における潜在的な反応物質との結合はかなり不安定になり、それによって、リスク評価解析部 560 に関して論じたよ

10

20

30

40

50

うに、様々な排泄プロセスの望まれない変化が小さくなる。

【0198】

次いで、1つまたは複数の潜在的な反応物質ならびに1つまたは複数の標的生体分子との反応性の所望の変化（例えば、潜在的な有害交差反応物質との反応性の低下、1つ（または複数）の所望の標的との反応性の維持または増大など）を示すこれら新しい構造についての分子レコードは、戻されてリード候補データベース504に格納され、新しい再設計または最適化された構造として適切に注釈が付けられる。その後、これらの新しい最適化された候補構造を、モデリング・システム500による新たな処理に再提示して、データベース502中の潜在的な反応物質との有害な交差反応の確率が十分に低くなり、それによって、これら1つ（または複数）の新しい構造がよりよい薬物候補になる確度が改善されるようにする。これらの技術を利用して、初期リード候補データベースにない望ましいリードを見付けることができるはずである。

10

【0199】

潜在的な反応物質に対してスクリーニングを行い、リードごとにリスク評価を生成し、薬物の再設計および最適化を実施し、再度スクリーニングするこのプロセスは、特異性が高く、毒性が低く、他の有益な特性を備えた十分に望ましいリード候補構造が得られるまで、複数回実施することができることを理解されたい。

【0200】

モデリング・システム500は、ソフトウェア、ハードウェア、ファームウェア、またはこれらの何らかの組合せの形で実施し得る。システム500は、ローカル・システムまたは分散システムとすることもできるし、1つまたは複数の汎用コンピュータ上で動作させることもでき、必要に応じて、ネットワーク接続部など、このシステムの周りにデータを伝達する様々なデータ通信チャネルを含み得る。

20

【0201】

一実施形態では、モデリング・システム500は、専用のマイクロプロセッサ、ASIC、またはFPGA上で実施し得る。別の実施形態では、モデリング・システム500は、複数のマイクロプロセッサ、ASIC、FPGAを扱う電子基板またはシステム基板上で実施し得る。別の実施形態では、モデリング・システム500は、1つまたは複数の電子装置内に収容された複数の基板上で、または複数の基板にまたがって実施し得る。別の実施形態では、モデリング・システム500は、1つまたは複数の電子基板上に1つまたは複数のマイクロプロセッサ、ASIC、またはFPGAを含み、かつネットワークを介して接続された複数の装置にまたがって実施し得る。反応決定部520および／または薬物再設計／最適化エンジン580に関連するものなど、様々な計算ステップを高速化することを意図して専用のハードウェア処理ユニットを使用することがあり、プラカッシュIに記載のものに類似のシステムを必要とすることがある。

30

【0202】

ある種の実施形態では、モデリング・システム500は、解析で使用し、また、解析によって生成される必要とされる様々なデータ要素を格納する1つまたは複数の記憶媒体装置も含み得る。

【0203】

あるいは、ある種の実施形態では、これらの記憶媒体装置の一部または全部を、ネットワークその他の方法でモデリング・システム500に接続して外部に配置し得る。外部記憶媒体装置の例は、1つまたは複数のデータベース・サーバまたはファイル・システムを含み得る。1つまたは複数の基板を扱う実施形態に関わるある種の実施形態では、モデリング・システム500は、計算プロセスの助けとなる1つまたは複数のソフトウェア処理コンポーネントも含み得る。あるいは、ある種の実施形態では、これらソフトウェア処理コンポーネントの一部または全部を、モデリング・システム500にネットワークその他の方法で接続して外部に配置し得る。

40

【0204】

反応決定部520に関して先に説明したように、システム500は、一度に1つのリー

50

ド候補および潜在的な交差反応物質に対して、すなわち順に動作することもできるし、複数の対に対して並列に動作することもできる。典型的な実施形態では、1つのリード候補レコードおよび複数の潜在的な反応物質分子がシステム500に提示され、このリード候補についてのリスク評価が、リード候補と潜在的な反応物質の対ごとに、関連する1組の反応プロファイルに基づいて確立され、次いで、新しいリード候補レコードおよび同じ1組の複数の潜在的な反応物質分子がシステム500に提示され、関連するすべてのリード候補についてのリスク評価が生成されるまで、このプロセスが繰り返される。

【0205】

任意選択で、580の薬物再設計／最適化プロセスの開始は、1つのリード候補分子についてのリスク評価の完了後、あるいは、関連するすべてのリード候補分子についての処理が完了した後とし得る。別の実施形態では、1組の潜在的な反応物質分子に対する1組のリード候補についての結果が生成されるが、新しい潜在的な反応物質分子が利用可能になると、すべての潜在的な反応物質に対してリード候補を再計算する代わりに、データベースの最近の更新に関わった潜在的な反応物質のみを、さらには、潜在的な反応物質データベース中の既存の構造に非相同の潜在的な反応物質のみをシステム500に提示して、このリード候補についての反応プロファイルおよびリスク評価を更新する。既に論じたように、薬物再設計／最適化エンジン580で新しい候補構造を構築するときには、この手順全体を反復手順で複数回実施し得る。

【0206】

ある種の実施形態では、各リードおよび個々の潜在的な反応物質ごとに反応の尺度を個々に計算する代わりに、様々な配列相同性、構造的または化学的な類似性、および／または機能バイオインフォマティクスまたはプロテオミクスの分類基準に基づいて認められた方法に従ってこれらの潜在的な反応物質をクラスタ化またはグループ化し、各クラスタまたはグループからの代表を用いて、特定のリードに対する反応の尺度を生成することができる。次いで、確度の高い反応性を示す潜在的な反応物質クラスタのみをメンバごとに個々に処理して、このリードに関して詳細な個々の反応の尺度およびその後の反応プロファイルを生成し、それによって、計算の全体的なコストおよび時間を低減することができる。

【0207】

ある種の潜在的な交差反応は有害でなく、運良く好都合なことがあり、予期せぬ利益が得られることがあるので、モデリング・システム500を使用すると、このリード候補についての可能な代替標的が識別されることもある。

【0208】

要約すると、潜在的に有害な交差反応を計算により予測し、リードおよびその変異体を選択し最適化する新規のシステムを説明してきた。リード候補は、1つまたは複数の所望の標的分子と反応することが知られている複数のリード候補分子から選択することができる。標的は、病気の経路におけるタンパク質とし、リード分子は、この病気に対処する潜在的な薬物候補とし得る。このリード分子は、小型分子薬物の構成要素、またはタンパク質に基づく薬物の構成要素とし得る。このリード分子と1組の潜在的な反応物質分子のそれぞれとの1つまたは複数の反応性の値を求め、反応の尺度を計算する。これらの潜在的な反応物質分子は、リードと1つ（または複数）の標的分子との予想される反応部位に関連するものと同じ環境または異なる環境の、ヒトその他の生物内に存在する、様々な器官／組織／細胞型からの複数の反応物質分子から選択し得る。反応の尺度は、反応の結果としてのこれらの分子およびそれらの環境からなる系の自由エネルギーの変化、すなわち、リードと潜在的な反応物質（または標的）分子との間の結合親和性の関数とし得る。

【0209】

この反応の尺度を、リード候補および潜在的な反応物質についてのソース／機能／相同体注釈などの他の反応データ、ならびに、リード候補および1つ（または複数）の所望の標的分子についての対応するアドメトックス・プロファイルおよび1つ（または複数）の反応の尺度の計算値などの他の尺度とともに用いて、このリード候補－潜在的な反応物質

10

20

30

40

50

対についての反応プロファイルを確立する。リード候補—潜在的な反応物質対についての反応プロファイルは、指定した細胞型、指定した組織、指定した器官、および／または指定した生物における反応物質分子の発生率および／または生物学的な機能、リード分子に関わる送達、分布、吸収、排泄、および代謝の様々な機構の作用、潜在的な反応物質および1つ（または複数）の標的分子に関連する機能経路情報、および反応物質分子の関連する配列に基づく、モチーフの、かつ／または構造的な相同体を含めて、1つまたは複数の追加の基準も含み得る。

【0210】

1組の潜在的な反応物質分子についてのリード候補ごとの1組の反応プロファイルは照合され、リードごとのリスク評価を形成するためにさらに処理される。リードのリスク評価ならびに使用者による様々な入力その他の決定基準に応じて、1つ（または複数）の所望の標的との結合特異性を増大させ、有害な交差反応の確度を低下させ、かつ／または、処置で用いる際に必要とされる投薬量を減少させるために、リード候補を化学的かつ構造的に改変することは価値がある。そのために、このシステムでは、リード、1つまたは複数の潜在的な反応物質、および1つ（または複数）の標的分子の間での予測される結合様式および結合親和性の点で所望の反応挙動を示す1つまたは複数の構造が得られるまで、分子の1つまたは複数の連結部位における様々な非コア分子群の追加、置換、および／または削除を利用して、選択されたリード候補構造の協調的かつ効率的な再設計または最適化も行う。先に論じたように、様々な実施形態は、このような再設計プロセスで生成される新しい潜在的な構造を評価するための新規技術および従来技術とともに包含する。

【0211】

上記の説明は、例示的なものであり、限定的なものではない。この開示を検討した後なら、当業者には本発明の多くの変形形態が明らかになるであろう。したがって、本発明の範囲は、上記の説明を参照して決定すべきではなく、添付の特許請求の範囲を、それらの均等物すべてとともに参照して決定すべきである。

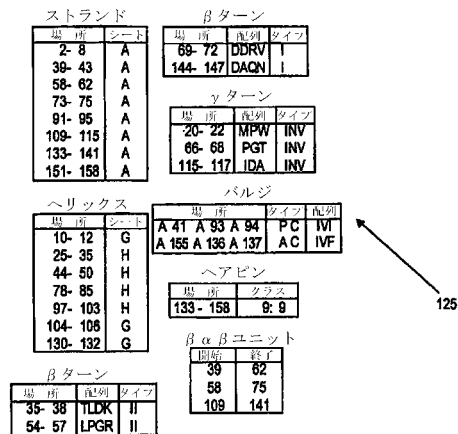
【図1A】

MET ILE SER LEU ILE ALA ALA LEU ALA VAL ASP ARG VAL
ILE GLY MET GLU ASN ALA MET PRO TRP ASN LEU PRO ALA
ASP LEU ALA TRP PHE LYS ARG ASN THR LEU ASP LYS PRO
VAL ILE MET GLY ARG HIS THR TRP GLU SER ILE GLY ARG
PRO LEU PRO GLY ARG LYS ASN ILE ILE LEU SER SER GLN
PRO GLY THR ASP ASP ARG VAL THR TRP VAL LYS SER VAL
ASP GLU ALA ILE ALA ALA CYS GLY ASP VAL PRO GLU ILE
MET VAL ILE GLY GLY GLY ARG VAL TYR GLU GLN PHE LEU
PRO LYS ALA GLN LYS LEU TYR LEU THR HIS ILE ASP ALA
GLU VAL GLU GLY ASP THR HIS PHE PRO ASP TYR GLU PRO
ASP ASP TRP GLU SER VAL PHE SER GLU PHE HIS ASP ALA
ASP ALA GLN ASN SER HIS SER TYR CYS PHE LYS ILE LEU
GLU ARG ARG

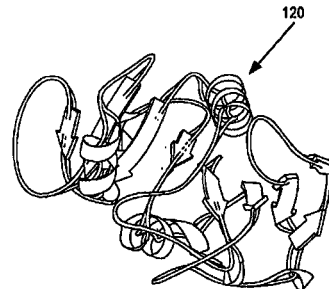
110

ジヒドロ葉酸還元酵素のアミノ酸配列

【図1B】

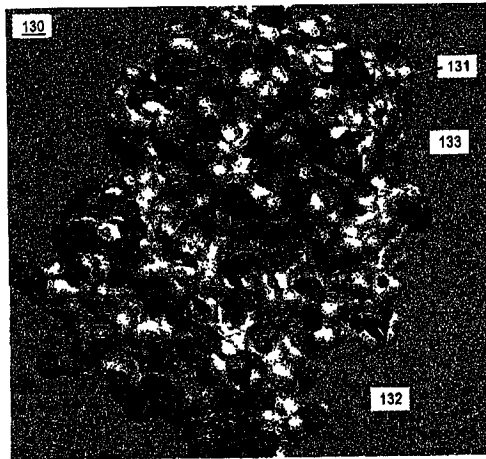


125



ジヒドロ葉酸還元酵素の2次構造

【図 1 C】



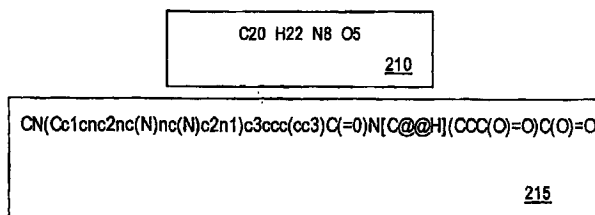
ジヒドロ葉酸還元酵素の溶媒接触表面

【図 1 D】

HEADER OXIDO-REDUCTASE4DFR 4/25
 COMPND DIHYDROFOLATE REDUCTASE (EC:1.5.1.3) COMPLEX WITH
 METHOTREXATE
 SOURCE (ESCHERICHIA COLI b), STRAIN/MB144285
 REMARK RESOLUTION: 1.7 ANGSTROMS.
 SEQRES 1 A 159 MET ILE SER LEU ILE ALA ALA LEU ALA VAL ASP ARG VAL
 9 A 159 ILE GLY MET GLU ASN ALA MET PRO TRP ASN LEU PRO ALA
 142 HELIX 1 HBA LEU A 24 THR A 35 1
 HELIX 2 HCA GLY A 43 ILE A 50 1
 SHEET 8 HFG GLY B 96 LEU B 104 1
 SHEET 1 S1A B THR A 73 VAL A 75 0
 SHEET 2 S1A B LYS A 58 SER A 63 1 O ASN A 59 N THR A 73
 SHEET 8 S1B B ASP B 132 HIS B 141-1 N GLU B 134 O GLU B 157
 ATOM 1 N MET A 1 24.293 59.579 4.215 1.00 39.20
 ATOM 2 CA MET A 1 25.127 58.554 4.868 1.00 37.00
 ATOM 3 C MET A 1 24.186 58.457 6.297 1.00 35.60
 ATOM 4 O MET A 1 23.827 58.402 6.804 1.00 34.50
 ATOM 5 CB MET A 1 26.311 59.103 5.385 1.00 40.80
 ATOM 6 CG MET A 1 27.346 58.263 5.966 1.00 46.90
 ATOM 7 SD MET A 1 28.097 59.298 7.167 1.00 52.50
 ATOM 8 CE MET A 1 28.875 60.685 6.576 1.00 53.80
 ATOM 9 N ILE A 2 23.920 57.222 6.865 1.00 32.60
 ATOM 10 CA ILE A 2 23.091 56.867 7.886 1.00 30.00
 ATOM 1269 N ARG A 159 33.592 49.480 7.054 1.00 49.80 1
 ATOM 1270 CA ARG A 159 34.659 49.036 6.032 1.00 68.70 1
 ATOM 1271 C ARG A 159 33.951 47.783 5.282 1.00 62.30 1
 ATOM 1272 O ARG A 159 33.704 48.253 3.980 1.00 67.20 1
 ATOM 1273 CB ARG A 159 35.899 48.487 6.547 1.00 57.00 1
 ATOM 1274 CG ARG A 159 38.216 47.712 7.812 1.00 58.60 1
 ATOM 1280 OXT ARG A 159 33.513 46.685 5.892 1.00 64.50 1
 TER 1281 ARG A 159
 HETATM 1315 CL CL A 2 11.750 52.951 20.253 1.00 28.60
 HETATM 1316 O HOH A 402 14.724 49.484 22.590 99 25.30
 HETATM 1317 O HOH A 403 22.932 59.466 28.571 1.02 39.50
 TER
 CONNECT 1 2
 CONNECT 2 1 3 5
 CONNECT 3 2 4 9
 CONNECT 4 3
 CONNECT 5 2 6
 CONNECT 6 5 7
 CONNECT 7 6 8
 CONNECT 8 7
 CONNECT 9 3 10
 CONNECT 10 9 11 13
 CONNECT 1250 1249
 CONNECT 1251 1249
 CONNECT 1252 1243 1253
 CONNECT 1253 1252 1254 1256
 CONNECT 1254 1253 1255 1258
 CONNECT 1255 1254
 CONNECT 1256 1253 1257
 CONNECT 1257 1256
 CONNECT 1258 1254

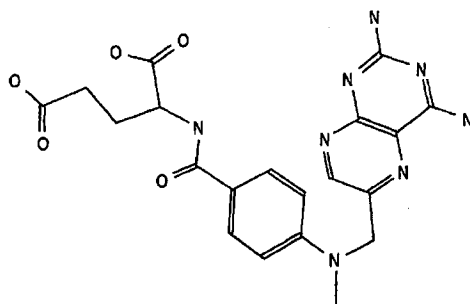
ジヒドロ葉酸還元酵素を扱うエントリ
4 d f r の PDB ファイル

【図 2 A】



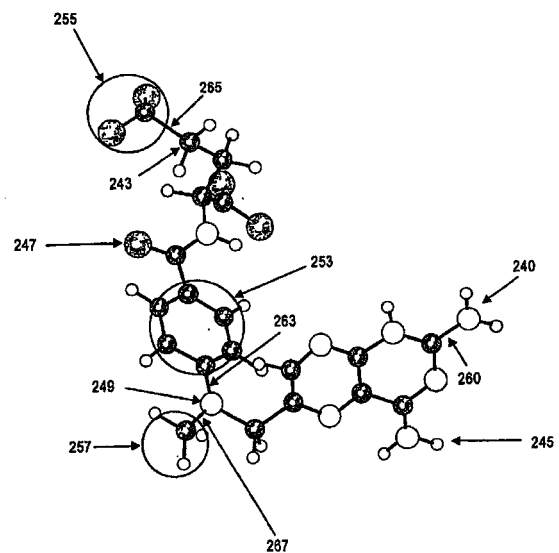
メトトレキサートの化学式および立体 SMILE

【図 2 B】



メトトレキサートの 2 次元概略表現

【図 2 C】



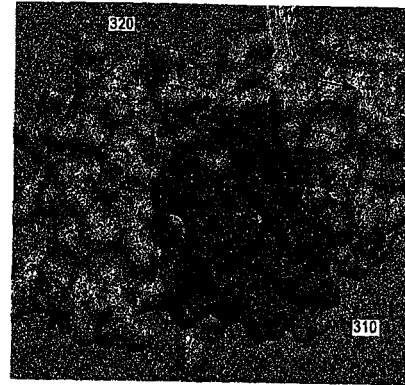
メトトレキサートの玉と棒による図

【図 2 D】

270	273	275	277	279	280	282	285
@<TRIPOS>MOLECULE							
MTX							
56	58	1					
@<TRIPOS>ATOM							
1	C7	19.781	59.256	25.628	C.ar	1	MTX
2	N8	21.096	59.176	25.562	N.ar	1	MTX
3	C8A	21.608	58.594	24.363	C.ar	1	MTX
4	N1	22.983	58.667	24.488	N.ar	1	MTX
5	C2	23.468	58.215	23.282	C.ar	1	MTX
6	NA2	24.797	58.223	23.208	N.2	1	MTX
7	HAE	25.253	57.932	22.367	H	1	MTX
8	HAD	25.34	58.52	23.994	H	1	MTX
9	N3	22.792	57.819	22.23	N.ar	1	MTX
10	C4	21.459	57.803	22.068	C.ar	1	MTX
...							
35	CG	20.891	67.636	27.32	C.3	1	MTX
36	CD	19.921	68.524	28.357	C.3	1	MTX
37	OE2	19.441	69.469	27.489	O.3	1	MTX
38	HAC	19.949	69.421	26.629	H	1	MTX
...							
53	H013	21.845	66.934	25.507	H	1	MTX
54	H014	21.798	67.729	27.734	H	1	MTX
55	H015	20.557	66.703	27.454	H	1	MTX
56	H016	20.325	66.4	29.283	H	1	MTX
@<TRIPOS>BOND							
1	1	2	ar				
2	1	16	ar				
3	2	3	ar				
4	3	4	ar				
5	3	14	ar				
6	4	5	ar				
7	5	6	1				
8	5	9	ar				
9	6	7	1				
10	6	8	1				
...							
54	34	52	1				
55	34	53	1				
56	35	54	1				
57	35	55	1				
58	36	56	1				
@<TRIPOS>SUBSTRUCTURE							
MTX 1							
291	292	293	295				

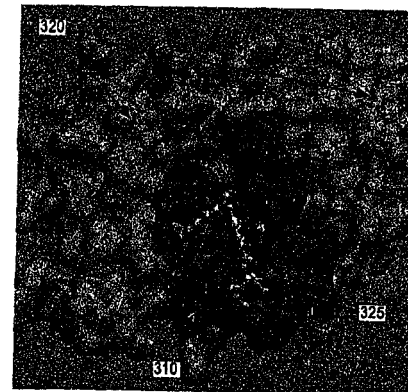
メトトレキサートの
Tripos mol2表現

【図 3 A】



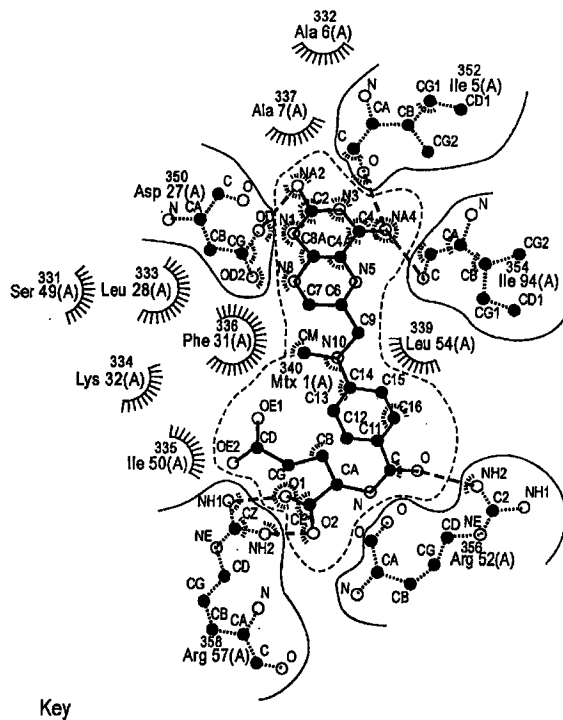
ジヒドロ葉酸還元酵素の活性部位

【図 3 B】



ジヒドロ葉酸還元酵素の活性部位内のメトトレキサート

【図 3 C】



Key

- リガンド結合
- 非リガンド結合
- 水素結合
- His 53 1つ（または複数）の疎水性接触に関わる非リガンド残基
- 1つ（または複数）の疎水性接触に関わる対応する原子

【図 3 D】

- ▲ APT
部位「APT」は、A分子内のメトトレキサート抑制剤のプテリジンと相互作用する残基を含む。
371 1つ（または複数）の活性部位残基 :-
ILEA 5 ALA A 6 ALA A 7 TRP A 22
ASPA 27 LEU A 28 PHEA 31 ILEA 94
THRA 113 HOHA 403 HOHA 405 HOHA 412 } 377
- ▲ ANM
部位「ANM」は、A分子内のメトトレキサート抑制剤のN(10)メチルと相互作用する残基を含む。
372 1つ（または複数）の活性部位残基 :-
- ▲ SERA 49
- ▲ AAB
部位「AAB」は、A分子内のメトトレキサート抑制剤のP-アミノベンゾイルと相互作用する残基を含む。
373 1つ（または複数）の活性部位残基 :-
LEU A 28 PHEA 31 ILEA 50 ARG A 52
LEU A 54
- ▲ AGL
部位「AGL」は、A分子内のメトトレキサート抑制剤のグルタミン酸と相互作用する残基を含む。
374 1つ（または複数）の活性部位残基 :-
LEU A 28 PHEA 31 LYS A 32 LEU A 54
ARG A 57

メトトレキサートとジヒドロ葉酸還元酵素を扱うPDBエントリー4 d f r
との活性部位情報

【図 4 A】

403	名称 ジヒドロ葉酸還元酵素
	他の名称: テトラヒドロ葉酸脱水素酵素、DHFR、還元酵素、7, 8-ジヒドロ葉酸還元酵素、NADPH-ジヒドロ葉酸還元酵素
	ソース 大腸菌(B)、系統 Mb_1428、抗メトトレキサート突然変異体 他のソース: ldr1 (ニワトリの肝臓)、lai9 (カンジダアルビカンス)、lboz (ヒト)
405	分子機能 グリシンおよびプリン de novo 合成、dna 前駆体合成、および dUMP から dTMP への変換に関する本質的なステップ
	病気との関連 DHFR の欠乏による巨赤芽球性貧血
	経路 グリシンおよびプリン de novo 合成、dna 前駆体合成、および dUMP から dTMP への変換に関する本質的なステップ
	反応/プロセス 5, 6, 7, 8-テトラヒドロ葉酸 + NADP (+) = 7, 8-ジヒドロ葉酸 + NADPH (葉酸を 5, 6, 7, 8-テトラヒドロ葉酸に還元する)
	分類/ファミリー ・機能: アクセプタとして NAD (+) または NADP (+) を伴うドナーの CH-NH 基に作用するオキシドレクターゼ ・構造: α/β タンパク質 ・ファミリー: ジヒドロ葉酸還元酵素ファミリー
407	配列相同体 DR (e) ジヒドロ葉酸還元酵素、真核型、DR (p) ジヒドロ葉酸還元酵素、原核型 ドメイン: laoea、タンパク質: DR (e)、種: 酵母 (カンジダアルビカンス)、ドメイン: ldyr、タンパク質: DR (e)、種: 真菌 (ニューモシステス・カリニ)、ドメイン: lkmva、タンパク質: DR (e)、種: ヒト (ホモ・サピエンス)、ドメイン: 8dfr、タンパク質: DR (e)、種: ニワトリ (ガルスガルス)、ドメイン: ldiga、タンパク質: DR (p)、種: 高度好熱菌、ドメイン: lvdra、タンパク質: DR (p)、種: 高度好熱菌、ドメイン: ldf7a、タンパク質: DR (p)、種: ヒト結核菌、ドメイン: lra9、タンパク質: DR (p)、種: 大腸菌、ドメイン: 3dfr、タンパク質: DR (p)、種: カセイ菌
	2次構造相同体 チオレドキシ、GTP 結合タンパク質、ジヒドロ葉酸還元酵素、ヌクレオチドキナーゼ、β-ラクタマーゼ、リシン様タンパク質、サブチラーゼ、ペリプラズム結合タンパク質-糖、ホスホフルクトキナーゼ、乳酸/リンゴ酸脱水素酵素、グリセルアルデヒドリン酸脱水素酵素
	機能相同体 プテリジン還元酵素、チミジル酸シンターゼ (複数種における二機能性ドメイン)

タンパク質レコード

【図 4 A - 3】

426	↓	427	↓	428	↓	429	↓	430	↓
425	4dfr	Amber	電荷	質量	VdW	VdW			
	原子	原子			半径	深さ			
	N1	N3	0.1592	14.01	1.8240	0.1700			
	CA	CT	-0.0221	12.01	1.9080	0.1094			
	C	C	0.6123	12.01	1.9080	0.0860			
	O	O	-0.5713	16.00	1.6612	0.2100			
	CB	CT	0.0865	12.01	1.9080	0.1094			
	CG	CT	0.0334	12.01	1.9080	0.1094			
	SD	S	-0.2774	32.06	2.0000	0.2500			
	CE	CT	-0.0341	12.01	1.9080	0.1094			
	N	N3	0.0311	14.01	1.8240	0.1700			
	CA	CT	0.0257	12.01	1.9080	0.1094			
	C	C	0.6123	12.01	1.9080	0.0860			
	O	O	-0.5713	16.00	1.6612	0.2100			
	CB	CT	0.1885	12.01	1.9080	0.1094			
	CG1	CT	-0.0387	12.01	1.9080	0.1094			
	CG2	CT	-0.3720	12.01	1.9080	0.1094			
	CD1	CT	-0.0908	12.01	1.9080	0.1094			
	...								
431	ビクワ・デ・テンシヤル: (PK/IDIVF) * (1 + cos(PN*phi - PHASE))								
	4dfr 結合	Amber 結合	IDIVF	PK	PHASE	PN			
	H-N3-CT-CT	X-N3-CT-X	9	1.40	0.0	3.0			
	C-CT-CT-CT	X-CT-CT-X	9	1.40	0.0	3.0			
	CT-CT-CT-S	X-CT-CT-X	9	1.40	0.0	3.0			
	CT-CT-S-CT	X-CT-S-X	3	1.00	0.0	3.0			
	CT-S-CT-H	X-S-CT-X	3	1.00	0.0	3.0			
	H-N3-CT-CT	X-N3-CT-X	9	1.40	0.0	3.0			
	C-CT-CT-CT	X-CT-CT-X	9	1.40	0.0	3.0			
	...								
433	天然結合剤								
	葉酸、NADP、テトラヒドロ葉酸、ジヒドロ葉酸								
	他の周知の結合剤								
	ビオプテリン、アミノプテリン、メトトレキサート、トリメトプリン								

タンパク質レコード

【図 4 A - 2】

410	↑	412	↑	411	↑	413	↑
	αヘリックス			βシート			ヘアピン
414	たんばく質コード 4dfr						
	引用文献 ボリン、ジェイ・ディー、フィルマン、マッシュューズ、ディー・エイ、ハムリン、アール・シー、クラウト、ジェイ、[Crystal structure of e. Coli and Lactobacillus casei dihydrofolate reductase refined 1.7 Å resolution. General Features and Binding of methotrexate.], J. Biol. Chem., 257 [22], 13650~6頁, 1982年11月25日						
	方法 X線回折、分解能 1.7 Å						
415	類 非対称単位 of 2つの独立したタンパク質分子を A および B と示す。それぞれ 150 残基を含む。						
423	原子座標 ...						
	化学結合 ...						

【図 4 B】

434	分子名 メトトレキサート
	商標名 アメトプレリン (Amethopterin)、メキサート (Mexate)、フォレックス (Folex)
	化学式 C20 H22 N8 O5
435	分子量 453.4416
	化学名 N-[4-[[[2,4-diaminopteridin-6-yl)methyl](methyl)amino]benzoyl]-L-glutamic acid
	ステレオ smile CN(Cc1cnc2nc(N)nc(N)c2n1)c3ccc(cc3)C(=O)N(C)C(CCC(=O)O)C(=O)O
	構造的/化学的な分類 標準αアミノ酸; 非高分子 メトトレキサートは、抗代謝剤として知られる薬物の一般群に属する。メトトレキサートは、骨肉腫、乳癌、頭部癌、および頸部癌を含めていくつかのタイプの癌を治療するのに用いる。
437	薬物投与経路 静脈内投与と化学治療薬、リウマチ関節炎には経口または筋肉内注射
	作用/過程 メトトレキサートは、核酸の合成を阻害することによって細胞による DNA および RNA の産生を妨げ、それによって、癌細胞の成長を阻害する。
445	原子座標 ...
453	化学結合 ...

リード候補レコード

【図 4 B - 2】

4dfr 原子	Amber 原子	電荷	質量	VdW 半径	VdW 深さ
N1	NT	0.1571	14.01	1.8240	0.1700
C2	CM	-0.0211	12.01	1.9080	0.0860
NA2	NT	0.2746	14.01	1.8240	0.1700
N3	N2	-0.3253	14.01	1.8240	0.1700
C4	CM	0.0724	12.01	1.9080	0.0850
NA4	NT	0.3867	14.01	1.8240	0.1700
C4A	CD	-0.1444	12.01	1.9080	0.0860
N5	NC	-0.0303	14.01	1.8240	0.1700
C6	CM	-0.0100	12.01	1.9080	0.0860
C7	CM	-0.0582	12.01	1.9080	0.0860
N8	NC	0.0543	14.01	1.8240	0.1700
C8A	CM	-0.1038	12.01	1.9080	0.0860
C9	CT	-0.0544	12.01	1.9080	0.1094
N10	N2	0.0215	14.01	1.8240	0.1700
CM	CT	-0.0805	12.01	1.9080	0.1094
C11	CA	-0.1230	12.01	1.9080	0.0860
C12	CA	-0.0146	14.01	1.9080	0.0860
C13	CA	-0.1936	12.01	1.9080	0.0860
C14	CA	-0.0392	12.01	1.9080	0.0860
C15	CA	-0.1980	12.01	1.9080	0.0860
C16	CA	-0.0365	12.01	1.9080	0.0860
C	C	0.2424	12.01	1.9080	0.0860
O	O	-0.4167	16.00	1.8612	0.2100
N	N	0.0791	14.01	1.8240	0.1700
CA	CT	-0.1038	12.01	1.9080	0.1094
CT	C	0.4000	12.01	1.9080	0.0860
O1	O2	-0.7173	16.00	1.8612	0.2100
O2	O2	-0.5761	16.00	1.8612	0.2100
CB	CT	-0.0899	12.01	1.9080	0.1094
CG	CT	-0.2188	12.01	1.9080	0.1094
CD	C	0.4453	12.01	1.9080	0.0860
OE1	O2	-0.6396	16.00	1.8612	0.2100
OE2	O2	-0.5993	16.00	1.8612	0.2100

$$\text{ビッファ・ポテンシャル: } (PK/IDIVF) * (1 + \cos(PN * \phi - \text{PHASE}))$$

4dfr 結合	Amber BOND	IDIVF	PK	PHASE	PN
O2-C-CT-N	X-C-CT-X	4	0.00	0.0	2.0
O2-C-CT-CT	X-C-CT-X	4	0.00	0.0	2.0
C-CT-CT-CT	X-CT-CT-X	9	1.40	0.0	3.0
CT-CT-CT-N	X-CT-CT-X	9	1.40	0.0	3.0
CT-CT-N-C	CT-CT-N-C	1	0.53	0.0	1.0
CA-C-N-CT	X-C-N-X	4	10.00	180.0	2.0
N-C-CA-CA	X-C-CA-X	4	14.50	180.0	2.0
CA-CA-N2-CT	X-CA-N2-X	4	9.60	180.0	2.0
H-CT-N2-CT	X-CT-N2-X	6	0.00	0.0	3.0
CM-CT-N2-CT	X-CT-N2-X	6	0.00	0.0	3.0
CM-CM-CT-N2	X-CM-CT-X	6	0.00	0.0	3.0

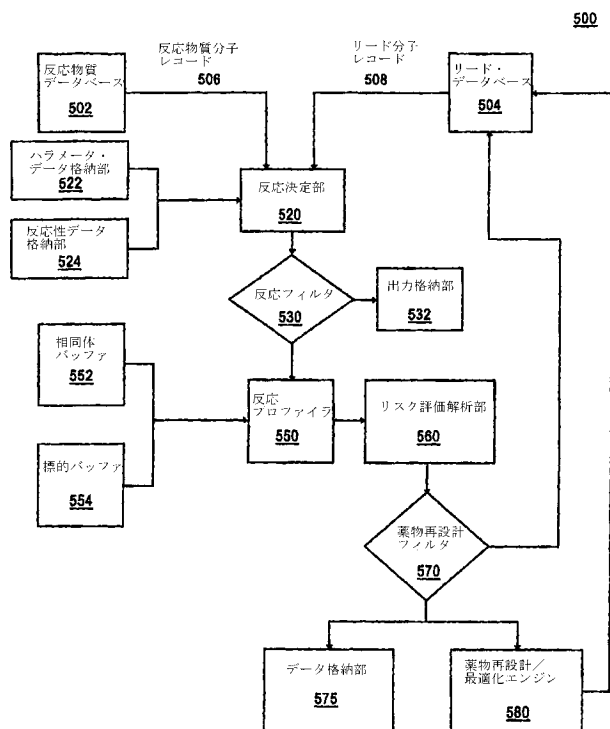
リード候補レコード

【図 4 B - 3】

459	MTXが結合する他の標的 1ao8, 1axw, 1dlg, 1ddr, 1dds, 1df7, 1dhj, 1dhj, 1dls, 1dra, 1drb, 1dre, 1e7w, 1mxf, 1p33, 1ra3, 1rb3, 1rg7, 1rx3, 1tdr, 2drc, 3cd2, 3dfr, 3drc
460	アドメトックス 高吸収率 経口または静脈注射（抗癌）あるいは筋肉内注射（抗リウマチ関節炎）により 摂取し得る数日の期間を通じて代謝的に安定 投薬量が中程度または高いと有毒 特に腎排泄に関して薬物の交差反応の危険性が高い 薬物とトリメトプリムの相互作用が致死性である可能性あり 正常細胞と癌／異常細胞を選択的に区別しない
	副作用 生死に関わるもの：じんましん、発疹、激しいかゆみ 一般的なもの：黒色便／血が混じる嘔吐、のどの痛み、発熱、口の糜爛、悪寒、 異常な出血または打撲傷、腹部の痛み、吐き気、嘔吐 不定期なもの：発作、めまい、眠気、頭痛、錯乱、かすみ目、息切れ、関節痛、 血尿、黄疸、下痢、皮膚の発赤、背痛、咳、発疹、男性の性的 不全、にきび、おでき、脱毛症、敏感肌 まれなもの：けいれん、痛みを伴う尿排泄

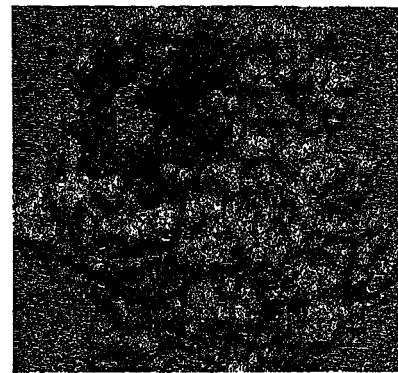
リード候補レコード

【図 5】



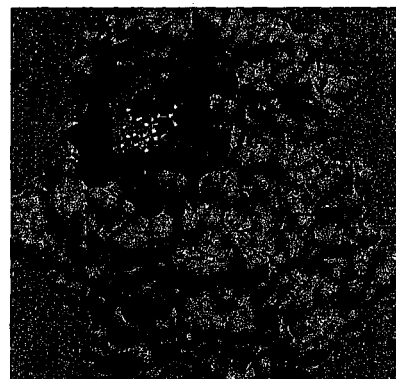
モデリング・システム

【図 6 A】



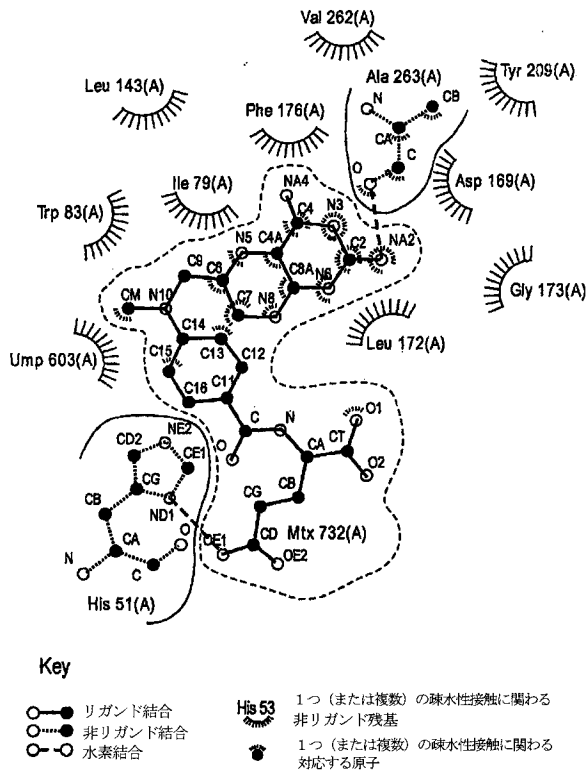
大腸菌チミジル酸シンターゼの活性部位

【図 6 B】

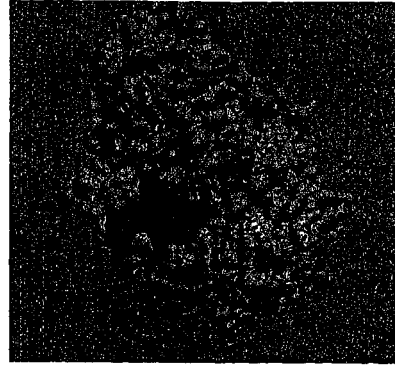


大腸菌チミジル酸シンターゼの活性部位におけるメトトレキサート

【図 6 C】

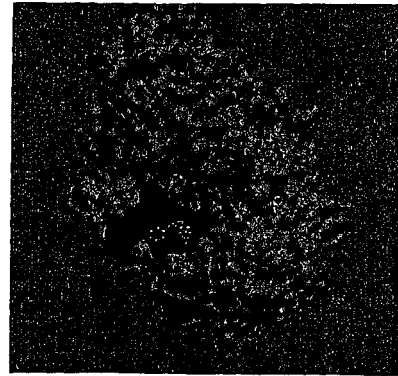


【図 7 A】



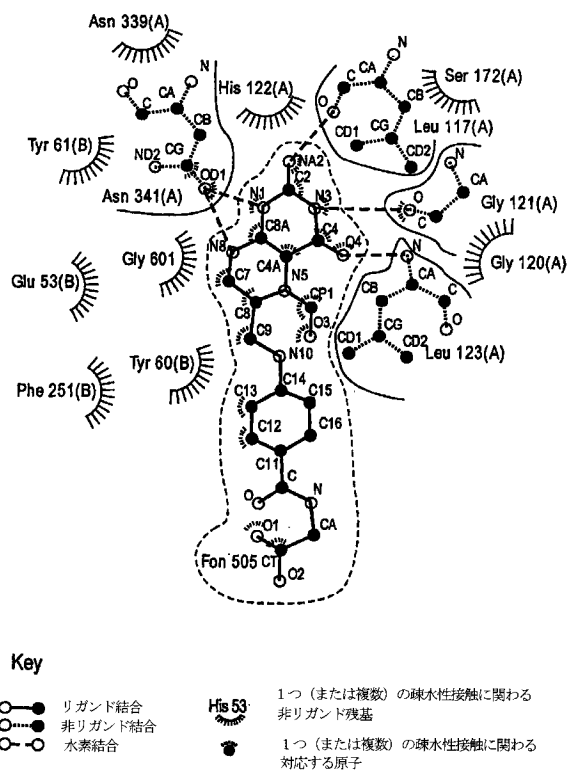
セリンヒドロキシメチル転移酵素の活性部位

【図 7 B】

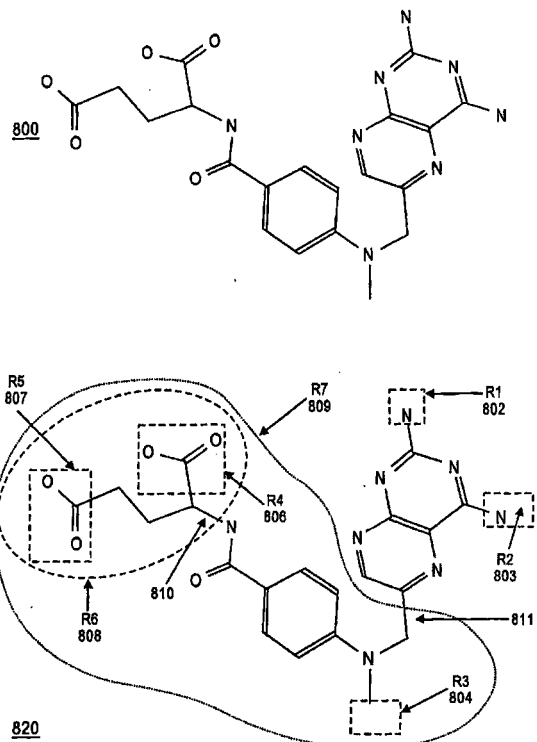


セリンヒドロキシメチル転移酵素の活性部位におけるフォリン酸

【図 7 C】



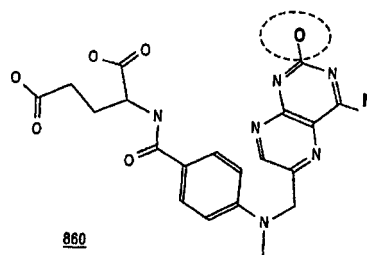
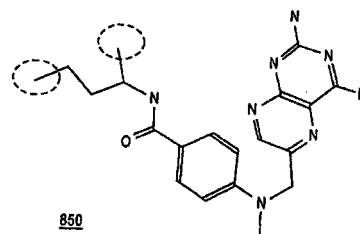
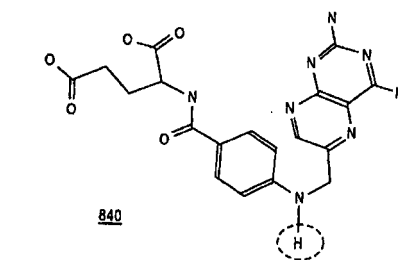
【図 8 A】



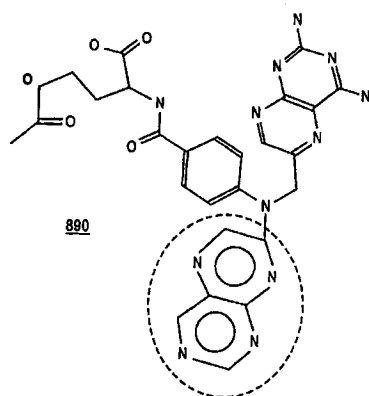
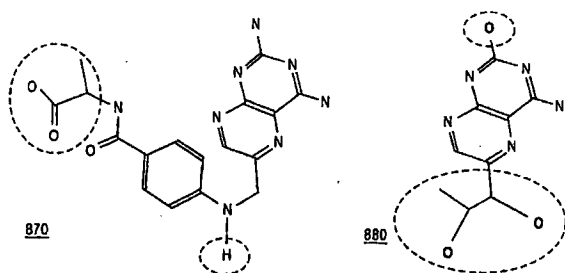
【図 8 A - 2】

化学的な実体	例
アルキル ($R = C_nH_{2n+1}$)	H, CH_3
エステル ($R-C(=O)-OR'$)	ギ酸メチル
ペプチド	セリン、フェニルアラニン
核酸塩基	チミン、グアニン
芳香	ベンゼン、プテリジン
...	...
元素	酸素、窒素、硫黄、塩素

【図 8 B】



【図 8 B - 2】



【手続補正書】

【提出日】平成29年1月27日(2017.1.27)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

1つまたは複数の所望の標的生体分子と反応することが知られているリード候補生体分子と、複数の潜在的な反応物質生体分子との潜在的な交差反応を予測し、モデル化する方法であって、

計算機モデル化システムにて、前記リード候補生体分子の化学的かつ構造的な情報を含む前記リード候補生体分子のデータ表現を受け取るステップと、

前記計算機モデル化システムにて、前記潜在的な反応物質生体分子の化学的かつ構造的な情報を含む、前記潜在的な反応物質生体分子それぞれのデータ表現を受け取るステップと、

前記計算機モデル化システムを使用して、指定の環境内で発生する、前記リード候補生体分子と各潜在的な反応物質生体分子との間の潜在的な交差反応それぞれについて、反応性を表す1つまたは複数の反応性値を推定するステップと、

前記計算機モデル化システムを使用して、前記潜在的な交差反応の各々に関連する反応性値の関数として、複数の反応尺度を計算するステップと、

前記計算機モデル化システムを使用して、それぞれが単一の信頼度を含む1つまたは複数の反応プロファイルを計算するステップであって、この信頼度は、各反応尺度の関数であり、かつ前記反応尺度に対応する前記潜在的な反応物質生体分子に関連するバイオインフォマティクスおよび／または化学的な1つまたは複数の注釈の関数でもあり、前記1つまたは複数の反応プロファイルは、リード候補生体分子と潜在的な反応物質生体分子の組み合わせに応じて、当該組み合わせに関連する反応尺度および関数を割り当てることにより計算される、ステップと、

前記反応プロファイルのうちの少なくとも2つの反応プロファイルに基づいて、1つまたは複数の有害反応が前記リード候補生体分子に作用する可能性を反映するリスク評価を行うステップと、

前記リスク評価をコンピュータ記録可能媒体に出力するステップと
を具備する方法であって、

潜在的な交差反応をモデル化するための前記計算機モデル化システムは、計算プラットフォームを実装するためのソフトウェア、専用ハードウェア、ファームウェア、またはこれらの組合せを含む、1つまたは複数のコンピュータを備え、

前記反応プロファイルの構築に用いられ、前記潜在的な反応物質に関連付けられた前記バイオインフォマティクスおよび化学的な注釈は、前記潜在的な反応物質に関連付けられたデータであって、ソース、機能、相同性、化学特性、実験室試験、反応部位、及びファミリー情報の内の1つ以上に少なくとも関係している潜在的な反応物質に関するデータを含み、

また、前記リスク評価は、前記反応プロファイルのうちの前記少なくとも2つの反応プロファイルのそれぞれの信頼度の関数である定量的値を含む、ことを特徴とする方法。

【請求項 2】

前記環境は、前記リード候補生体分子と前記潜在的な反応物質生体分子との反応部位の物理的かつ化学的な周囲の状況であり、

前記環境は、

前記リード候補生体分子と、前記リード候補生体分子に対する所望の、または既知の標的
的生体分子との反応部位の物理的かつ化学的な周囲の状況であり、

複数の原子、イオン、および／または他の分子を含む溶媒の、暗示的または明示的な表現であって、前記溶媒中に前記反応部位が埋め込まれている、前記溶媒の表現を含み、または、

ヒトもしくは他の哺乳動物の身体の、1つまたは複数の構成要素を含む、請求項1に記載の方法。

【請求項3】

前記潜在的な反応物質生体分子の1つは、生体高分子、あるいは生体高分子の1つまたは複数の部分を含み、前記1つまたは複数の部分は、1つまたは複数の活性部位を含む、請求項1に記載の方法。

【請求項4】

前記リード候補生体分子は、合成化合物、化学基、ペプチド、医薬化合物、タンパク質系薬物、または炭水化物を含む、請求項1に記載の方法。

【請求項5】

前記潜在的な交差反応に関連する反応の反応尺度を形成する際に用いる前記反応性値の1つは、前記交差反応の結合親和性または自由エネルギーの計算による推定値である、請求項1に記載の方法。

【請求項6】

前記潜在的な交差反応に関連する前記反応性値の一部または全部は、あらかじめ生成されるか、またはオフラインで導出され、コンピュータ可読媒体上にデジタル表現で格納され、その後、前記反応尺度の形成時に取り出される、請求項1に記載の方法。

【請求項7】

前記潜在的な交差反応に関連する反応尺度を形成するのに用いる前記反応性値のうちの少なくとも1つは、前記2つの生体分子間の定量的な構造活性の関係の事前予測（QSAR）の結果である、請求項6に記載の方法。

【請求項8】

前記潜在的な交差反応に関連する反応尺度を形成するのに用いる前記反応性値のうちの少なくとも1つは、1つまたは複数の実験室試験から得られた測定値または結果である、請求項6に記載の方法。

【請求項9】

前記1つまたは複数の実験室試験は、in vitro、細胞ベース、またはin vivoのいずれかにかかわらず、スクリーニングその他の機能アッセイを含む、請求項8に記載の方法。

【請求項10】

前記潜在的な交差反応に関連する反応尺度は、複数の反応性の値の1次結合である、請求項1に記載の方法。

【請求項11】

前記潜在的な交差反応に関連する反応尺度は、複数の反応性の値の非線形関数である、請求項1に記載の方法。

【請求項12】

前記潜在的な交差反応に関連する反応尺度は、1つまたは複数の反応性の値に基づくスコアまたは等級であり、

前記反応プロファイルを計算する前記ステップの前に、反応フィルタを用いて、関連する反応尺度が判定基準を満たさない潜在的な交差反応を、以降の検討から除外するステップをさらに具備する、請求項1に記載の方法。

【請求項13】

前記反応プロファイルを計算する前記ステップの前に、反応フィルタを用いて、関連する反応尺度が判定基準を満たさない潜在的な交差反応を、以降の検討から除外するステップをさらに具備する、請求項1に記載の方法。

【請求項14】

前記判定基準は、前記入力された反応尺度の大きさに適用する数値的な閾値である、請

求項 1 3 に記載の方法。

【請求項 1 5】

前記判定基準は、前記計算機モデル化システムによって反応尺度が生成されるときに、動的に適用される、請求項 1 3 に記載の方法。

【請求項 1 6】

前記判定基準は、個々のリード候補生体分子を特徴づけるすべての反応尺度が生成された後に適用される、請求項 1 3 に記載の方法。

【請求項 1 7】

前記判定基準は、1 つまたは複数の標的と反応するリード候補生体分子について計算または実験により得られた対応する反応尺度との相対的な比較に基づくものである、請求項 1 3 に記載の方法。

【請求項 1 8】

前記判定基準は、前記リード候補生体分子と潜在的な反応物質の反応に関連する前記反応尺度に閾値を適用することであり、前記親和性の閾値は、所定の定数と、前記リード候補生体分子と選択された標的との反応に関連する前記計算または測定された反応尺度とを掛けたものとして定式化される、請求項 1 7 に記載の方法。

【請求項 1 9】

前記潜在的な反応物質生体分子についての前記ソースの注釈は、前記潜在的な反応物質生体分子が得られ、由来することが期待され、また、見つかる可能性がある生物、種、器官、組織、細胞型、突然変異体または野生型、および／または系統に関する情報を含む、請求項 1 に記載の方法。

【請求項 2 0】

前記潜在的な反応物質生体分子についての前記機能の注釈は、前記潜在的な反応物質生体分子に関わる生物学的な、さらには病気の経路その他の反応および過程、あるいは、前記潜在的な反応物質生体分子についての既知の、さらには予測される生物学的な機能に関する情報を含む、請求項 1 に記載の方法。

【請求項 2 1】

前記潜在的な反応物質生体分子についての前記相同性の情報は、同じまたは異なる生物の他の生体高分子との構造および／または配列に基づく相同体との関連物を含み、タンパク質についての前記構造に基づく相同性は、保存領域、タンパク質モチーフ、または 3 次構造の類似性に基づいて確立され、生体高分子についての前記配列に基づく相同性は、1 次アミノ酸またはその基礎となるゲノミクス配列の類似性に基づくものである、請求項 1 に記載の方法。

【請求項 2 2】

前記潜在的な反応物質生体分子についての前記相同性の情報は、前記リード候補生体分子と反応または結合する 1 つまたは複数の標的との構造および／または配列に基づく相同性を含む、請求項 2 1 に記載の方法。

【請求項 2 3】

前記リスク評価は、前記バイオインフォマティクスおよび／または化学的な注釈に基づいて行われるものであり、

前記リード候補生体分子に関連する前記注釈は、前記リード候補生体分子に関連する化学的および構造的な組成、反応の化学的性質、相同性、および／または化学的または構造的な分類情報に関するデータを含む、請求項 1 に記載の方法。

【請求項 2 4】

前記リード候補生体分子についての前記化学的および構造的な組成の注釈は、化学式、化学名、1 つまたは複数の化学目録に関する指標記述子、生体分子の化学的かつ物理的な特性、キラル中心および立体化学についての知識、または 2 次元または 3 次元の表現での構造情報に関する情報を含む、請求項 2 3 に記載の方法。

【請求項 2 5】

前記リード候補生体分子についての前記反応の化学的性質の注釈は、前記リード候補生

体分子に関わる 1 つまたは複数の化学反応または生物学的なプロセスに係する反応データを含む、請求項 2 3 に記載の方法。

【請求項 2 6】

前記化学反応または生物学的なプロセスの範囲は、前記リード候補生体分子との構造的かつ／または化学的な類似性が高い他の生体分子に関わる化学反応または生物学的なプロセスを含むように拡張される、請求項 2 5 に記載の方法。

【請求項 2 7】

前記リード候補生体分子に関わる前記 1 つまたは複数の化学反応または生物学的なプロセスに係する前記反応データは、前記リード候補生体分子が結合するか、または反応する 1 つまたは複数の標的についての情報との関連物を含む、請求項 2 5 に記載の方法。

【請求項 2 8】

前記リード候補生体分子に関わる 1 つまたは複数の化学反応または生物学的なプロセスに係する前記反応データは、アドメトックスその他の薬物動態学のプロファイルの 1 つまたは複数の情報構成要素を含む、請求項 2 5 に記載の方法。

【請求項 2 9】

前記リード候補生体分子に関わる 1 つまたは複数の化学反応または生物学的なプロセスに係する前記反応データは、1 つまたは複数の実験室試験から得られた 1 つまたは複数の測定値を含む、請求項 2 5 に記載の方法。

【請求項 3 0】

前記リード候補生体分子についての前記相同性の情報は、他の生物分子との構造に基づく相同体および／または化学的な相同体との関連物を含み、前記構造に基づく相同性および／または化学的な相同性は、2 次元または 3 次元のファルマコフォア・モデル、化合物のクラスタリングおよび分類で用いる様々な分子記述子、または定量的な構造活性の関係の類似性を用いることに基づいて確立される、請求項 2 3 に記載の方法。

【請求項 3 1】

前記潜在的な反応物質生体分子と、前記リード候補生体分子に関連する 1 つまたは複数の標的との前記構造的または化学的な類似性に関する情報が、前記潜在的な反応物質生体分子を特徴づける潜在的な交差反応についての前記反応プロファイルに前記反応フィルタを適用する前記ステップにおいて、あるいは、前記反応プロファイルを計算する前記ステップにおいて、考慮される、請求項 1 に記載の方法。

【請求項 3 2】

前記リード候補生体分子と、前記潜在的な反応物質生体分子の 1 つまたは複数の結合剤との構造的または化学的な類似性に関する情報が、前記リード候補生体分子を扱う潜在的な交差反応についての前記反応プロファイルに前記反応フィルタを適用する前記ステップにおいて、あるいは、前記反応プロファイルを構築する前記ステップにおいて、考慮される、請求項 1 に記載の方法。

【請求項 3 3】

リード候補生体分子についての前記リスク評価は、定期的に、あるいは、新しい潜在的な反応物質が識別されて前記計算機モデル化システムで利用可能になったときに更新される、請求項 1 に記載の方法。

【請求項 3 4】

前記リスク評価は、複数の定量的な反応プロファイルの、1 次結合若しくはその他の機能的な合成で表現される、請求項 1 に記載の方法。

【請求項 3 5】

前記リスク評価は、複数の定量的な反応プロファイルに基づいて生成された定性的な分類または等級である、請求項 1 に記載の方法。

【請求項 3 6】

前記リスク評価は、複数の定量的な反応プロファイルの統計分析に基づいて生成された信頼性のスコアまたは確率あるいは統計量である、請求項 1 に記載の方法。

【請求項 3 7】

リード候補生体分子についての前記リスク評価を構成するのに用いる前記統計分析では、1組の反応プロファイルすべての平均値または中心値、あるいは、前記1組の反応プロファイルの百分率の上位選択範囲を用いる、請求項36に記載の方法。

【請求項38】

リード候補生体分子についての前記リスク評価を構成するのに用いる前記統計分析は、ベイズ・リスク分析である、請求項36に記載の方法。

【請求項39】

前記リスク評価は、コンピュータ学習手法から出力された予測を表す、請求項1に記載の方法。

【請求項40】

前記リスク評価を生成するのに用いる前記コンピュータ学習手法は、ニューラル・ネットワークに基づくものである、請求項39に記載の方法。

【請求項41】

前記リスク評価を生成するのに用いる前記コンピュータ学習手法は、サポート・ベクトル・マシンに基づくものである、請求項39に記載の方法。

【請求項42】

複数の潜在的な反応物質生体分子の選択は、2次元または3次元の構造データベースの全部または一部からなされる、請求項1に記載の方法。

【請求項43】

前記複数の潜在的な反応物質生体分子を、関連する生物学的かつ／または薬物動態学的な経路の知識ベースを用いて事前に選択する、請求項1に記載の方法。

【請求項44】

前記潜在的な反応物質は、生体高分子であって、

前記潜在的な反応物質生体高分子のそれぞれのデータ表現を受け取る前記ステップの前に、前記潜在的な反応物質生体高分子を、前記リード候補生体分子中の、1つまたは複数の所望の標的との、または、1つまたは複数の既知の標的との、構造上かつ／または配列上の類似性に基づいて選択するステップをさらに含む、請求項1に記載の方法。

【請求項45】

前記複数の潜在的な反応物質生体高分子の前記選択するステップは、前記リード候補生体分子の1つまたは複数の標的の、タンパク質または遺伝子のファミリーによる分類メンバを、あるいは、他の機能上またはバイオインフォマティックス上の分類メンバを、含める、請求項44に記載の方法。

【請求項46】

潜在的な反応物質間の化学的かつ／または構造的な類似性の考慮を含む1組のクラスタリング基準に基づいて、前記複数の潜在的な反応物質生体分子にクラスタリング操作を適用して、クラスタまたはグループの集合を生成するステップと、

前記クラスタまたはグループの各々から、クラスタ代表として、1つの潜在的な反応物質を選択するステップと、

前記リード候補生体分子と前記クラスタ代表の各々との潜在的な交差反応について、対応する1組の反応尺度を生成するステップと、

前記得られた1組の反応尺度に、反応フィルタを適用して、前記リード候補生体分子との高い反応性を示すクラスタ代表のサブセットを選択するステップと、

合格したクラスタ代表を含む潜在的な反応物質からなる各クラスタを選択し直すステップと、

前記選択し直されたクラスタのあらゆるメンバの各々についての詳細な個々の反応尺度および後続の反応プロファイルを生成するステップと
をさらに含む、請求項1に記載の方法。

【請求項47】

前記潜在的な反応物質は生体高分子であり、

前記クラスタリング基準は、配列相同性、構造相同性、および／または他のバイオイン

フォーマティックスまたはプロテオミックスの分類基準に関わる、請求項 4 6 に記載の方法。

【請求項 4 8】

1 つまたは複数の所望の標的との結合特異性を改善し、前記リード候補生体分子との潜在的に有害な交差反応に関与することが予測される、またはその可能性がある 1 つまたは複数の潜在的な反応物質生体分子との反応性を低下させる目的で、前記リード候補生体分子を最適化し再設計する複数のステップをさらに具備し、

前記複数のステップは、

前記リード候補生体分子が 1 つまたは複数の所望の標的と結合するために必要とされ、前記リード候補生体分子に関連付けられた分子のコアまたは枠組を識別するステップと、

前記枠組の 1 つまたは複数の連結部位に追加可能なまたは置換し得る、分子群又は他の構成要素を識別するステップと、

前記リード候補生体分子と、このリード候補生体分子が強く反応することが望まれる 1 つまたは複数の所望の標的との間に起こる 1 つまたは複数の反応を、前記リード候補生体分子および前記所望の標的の各々に関連付けられた化学基間でエネルギー的に有利な分子相互作用を保存し又は促進する観点で分析するステップと、

前記リード候補生体分子と、前記リード候補生体分子が強く反応することが望まれない 1 つまたは複数の潜在的な反応物質生体分子との間に起こる 1 つまたは複数の反応を、前記リード候補生体分子および潜在的な各反応物質生体分子に関連付けられた化学基間でエネルギー的に有利な分子相互作用を低減させる観点か、かつ／またはエネルギー的に不利な分子相互作用を促進する観点で分析するステップと、

前記リード候補生体分子と 1 つまたは複数の所望の標的との分子相互作用に対する前記分析に従って、並びに、前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子との分子相互作用に対する前記分析に従って、局所分子構造に追加すべき、または置換すべき前記枠組の 1 つまたは複数の連結部位を選択するステップと、

前記リード候補生体分子と 1 つまたは複数の所望の標的との分子相互作用に対する前記分析に従って、並びに、前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子との分子相互作用に対する前記分析に従って、前記選択した連結部位の 1 つまたは複数において局所分子構造に対して追加すべき、または置換すべき 1 つまたは複数の分子群並びに他の構成要素を選択するステップと、

選択した前記枠組の前記連結部位と、選択し追加され又は置換された分子群との組合せに基づいて、前記リード候補生体分子の複数の構造変異体を構築するステップと、

前記リード候補生体分子の新たに構築された構造変異体各々を、当該各構造変異体の、1 つまたは複数の潜在的な反応物質生体分子に対する反応性、および、1 つまたは複数の所望標的との反応性、とに従って評価するステップとを含む、請求項 1 に記載の方法。

【請求項 4 9】

前記リード候補生体分子の新たに構築された構造変異体を評価する前記ステップは、前記リード候補生体分子の前記新たに構築された構造変異体を、新しい入力リード候補として、前記計算機モデル化システムに提示し直すことを含む、請求項 4 8 に記載の方法。

【請求項 5 0】

前記リード候補生体分子は、前記リード候補生体分子について既に構築された反応プロファイルおよび／または前記リード候補生体分子に関連するリスク評価に基づいて、可能な再設計および最適化のためにリード候補生体分子に優先順位を付けられる、請求項 4 8 に記載の方法。

【請求項 5 1】

前記リード候補生体分子は、リード候補生体分子についての対応するリスク評価により、前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子とを扱う潜在的な有害交差反応に関連するリスクが許容差内にあり、かつ、所定の制限を超えないことが示唆されると、可能な再設計および最適化が提示される、請求項 5 0 に記載の方法。

【請求項 5 2】

前記リード候補生体分子について複数の構造的に別個の枠組が識別され、前記リード候補生体分子の前記可能な再設計および最適化中に、それらを前記リード候補生体分子の複数の構造変異体を生成するための可能なシードとして用いられる、請求項 48 に記載の方法。

【請求項 53】

前記枠組の 1 つまたは複数の連結部位に対する可能な追加または削除に関して識別された前記 1 組の分子群は、有機化学で見られる様々な化学基、アミノ酸側鎖、残基、さらには個々の原子またはイオンを含む、請求項 48 に記載の方法。

【請求項 54】

前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子の間の 1 つまたは複数の反応の分析は、水素結合の相互作用するドナー基またはアクセプタ基の改変によって不安定になった場合、前記リード候補生体分子と前記潜在的な反応物質生体分子の反応性を低下させることがあるが、前記リード候補生体分子と 1 つまたは複数の所望の標的の反応性を過度に損なわない 1 つまたは複数の分子間水素結合を識別することを含む、請求項 48 に記載の方法。

【請求項 55】

前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子の間の 1 つまたは複数の反応を分析する前記ステップは、分子間疎水性接触部が相互作用する疎水基の改変によって不安定になった場合には、前記リード候補生体分子と前記潜在的な反応物質生体分子との反応性を低下させることはあっても、前記リード候補生体分子と 1 つまたは複数の所望の標的との反応性を過度に損なわない、前記 1 つまたは複数の分子間疎水性接触部を識別するステップをさらに含む、請求項 48 に記載の方法。

【請求項 56】

前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子の間の 1 つまたは複数の反応を分析する前記ステップは、分子間塩橋が相互作用する静電的な基またはイオン化された基の改変によって不安定になった場合には、前記リード候補生体分子と前記潜在的な反応物質生体分子との反応性を低下させることはあっても、前記リード候補生体分子と 1 つまたは複数の所望の標的との反応性を過度に損なわない前記 1 つまたは複数の分子間塩橋を識別するステップをさらに含む、請求項 48 に記載の方法。

【請求項 57】

前記リード候補生体分子と 1 つまたは複数の潜在的な反応物質生体分子の間の 1 つまたは複数の反応を分析する前記ステップは、局所立体相補性が化学的または構造的に改変された場合には、前記リード候補生体分子と前記潜在的な反応物質生体分子との反応性を低下させることはあっても、前記リード候補生体分子と 1 つまたは複数の所望の標的との反応性を過度に損なわない高い局所立体相補性を示すいずれかの分子の 1 つまたは複数の相互作用する化学基を識別するステップをさらに含む、請求項 48 に記載の方法。

【請求項 58】

前記リード候補生体分子の複数の構造変異体を構築する前記ステップは、前記リード候補生体分子の 1 つまたは複数の基または構成要素に関連する立体化学的な性質を改変するステップをさらに含む、請求項 48 に記載の方法。

【請求項 59】

前記リード候補生体分子の複数の構造変異体を構築する前記ステップは、前記リード候補生体分子の 1 つまたは複数の基または構成要素に関連するイオン化状態を改変するステップをさらに含む、請求項 48 に記載の方法。

【請求項 60】

前記リード候補生体分子の複数の構造変異体を構築することに関わるプロセス、前記構造変異体を、請求項 1 に記載の方法に基づいて潜在的な交差反応をモデル化するシステムに入力として提示し直すステップと、

その後で、新しい各入力構造変異体と 1 つまたは複数の潜在的な反応物質生体分子ごと
に反応プロファイルおよび／またはリスク評価を構築するステップと、

何らかの終端条件その他の基準が満たされるまで上記ステップを反復的に繰り返すステップと
を含む、請求項 49 に記載の方法。

【請求項 61】

反応性の評価および関連する生物学的かつ化学的な注釈の分析に基づいて、前記潜在的な反応物質生体分子の中から、1つまたは複数の新しい潜在的な標的を識別するステップをさらに含む、請求項 1 に記載の方法。

【請求項 62】

少なくとも 1 つの反応プロファイルは、
前記リード候補生体分子と前記対応する反応尺度によって与えられる前記潜在的な反応物質生体分子の間の交差反応の確度と、
前記対応する潜在的な反応物質生体分子に関連するバイオインフォマティックスおよび／または化学的な前記注釈に基づく交差反応が有害である確度とに基づく信頼度またはスコアを含む、請求項 1 に記載の方法。

【請求項 63】

前記信頼度はさらに、前記リード候補生体分子が前記標的の生物体内の前記対応する潜在的な反応物質生体分子に遭遇する確度に基づく、請求項 62 に記載の方法。

【請求項 64】

前記確度は数値として表される、請求項 62 に記載の方法。

【請求項 65】

1つまたは複数の所望の標的生体分子と反応することが知られているリード候補生体分子と、複数の潜在的な反応物質生体分子との間での潜在的な交差反応を予測しモデル化する方法を実行するように、計算機によるモデル化システムのプロセッサに指示するプログラム命令を記憶して格納するコンピュータプログラム可読媒体であって、前記プログラム命令は、

前記モデル化システムに前記リード候補生体分子のデータ表現を提供するプログラムステップであって、前記データ表現は、前記リード候補生体分子の化学的かつ構造的な情報を含む、前記プログラムステップと、

計算によるモデル化システムに前記潜在的な反応物質生体分子のそれぞれのデータ表現を提供するプログラムステップであって、前記データ表現は、前記潜在的な反応物質生体分子の化学的かつ構造的な情報を含む、前記プログラムステップと、

ある環境状況における前記リード候補生体分子と潜在的な反応物質生体分子それぞれの間のそれぞれの潜在的な交差反応について、1つまたは複数の反応性の値を取得するプログラムステップと、

複数の反応尺度を、それぞれ、各潜在的な交差反応に関連する前記反応性値の関数として構築するプログラムステップと、

各反応尺度の関数であり、かつ前記反応尺度に対応する前記潜在的な反応物質生体分子に関連する 1つまたは複数のバイオインフォマティックスおよび／または化学的な注釈の関数である単一の信頼度をそれぞれの反応プロファイルが含む 1つまたは複数の反応プロファイルを構築し、前記 1つまたは複数の反応プロファイルは、リード候補生体分子と潜在的な反応物質生体分子の組み合わせに応じて、該組み合わせに対応する反応尺度および関数を割り当てることにより計算される、プログラムステップと、

前記反応プロファイルのうちの少なくとも 2 つの反応プロファイルに基づいて、1つまたは複数の有害反応が前記リード候補生体分子に作用する可能性を反映するリスク評価を評価するプログラムステップと、

前記リスク評価をコンピュータ記録可能媒体に出力するプログラムステップと
を含む、コンピュータプログラム可読媒体。

【請求項 66】

前記反応尺度に対応する前記潜在的な反応物質生体分子に関連する前記 1つまたは複数のバイオインフォマティックスおよび／または化学的な注釈を取得するためにデータベ

スのクエリを実行するステップをさらに含む、請求項 1 に記載の方法。

【請求項 67】

前記反応プロファイルの前記単一の値は数値スコアであり、前記注釈に基づいて反応尺度を増減することにより数値スコアを取得するステップをさらに含む、請求項 1 に記載の方法。

【請求項 68】

再検討のため前記リスク評価の結果を使用者に提供するステップをさらに含む、請求項 1 に記載の方法。

【請求項 69】

反応プロファイルの前記単一の値は、定性的な分類である、請求項 1 に記載の方法。

【請求項 70】

前記化学的な特性は、log P または pKa または分子量の内の 1 または複数を含むことを特徴とする請求項 1 に記載の方法。

【請求項 71】

前記実験室試験は、実験室での測定、吸収・分布・代謝・排泄 (ADME) 試験、毒物学試験 (Tox) の 1 または複数を含む、請求項 1 に記載の方法。

【請求項 72】

前記反応部位に関係する前記データは、潜在的な反応物質とリード候補とが生命体内の何処で互いに反応するかを示す、請求項 1 に記載の方法。

フロントページの続き

(74)代理人 100134175

弁理士 永川 行光

(72)発明者 キタ、デビット

アメリカ合衆国 カリフォルニア州 9 5 0 3 5 ミルピタス、グランド テトン ドライブ 1
6 7 4

(72)発明者 フォードール、エニコ

アメリカ合衆国 カリフォルニア州 9 4 5 3 9 フレモント、クリアビュー テラス 3 3 2 0

(72)発明者 ブラカシュ、アディティヨ

アメリカ合衆国 カリフォルニア州 9 4 5 3 9 フレモント、クリアビュー テラス 3 3 2 0

【外国語明細書】
2017123169000001.pdf