

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2010年3月11日(11.03.2010)

PCT

(10) 国際公開番号
WO 2010/026738 A2

- (51) 国際特許分類:
G06F 19/00 (2006.01) C12Q 1/48 (2006.01)
- (21) 国際出願番号: PCT/JP2009/004286
- (22) 国際出願日: 2009年9月1日(01.09.2009)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2008-228434 2008年9月5日(05.09.2008) JP
- (71) 出願人(米国を除く全ての指定国について): 国立大学法人神戸大学(NATIONAL UNIVERSITY CORPORATION KOBE UNIVERSITY) [JP/JP]; 〒6578501 兵庫県神戸市灘区六甲台町1番1号 Hyogo (JP).
- (72) 発明者; および
- (75) 発明者/出願人(米国についてのみ): 高岡裕 (TAKAOKA, Yutaka) [JP/JP]; 〒6578501 兵庫県神戸市灘区六甲台町1番1号国立大学法人神戸大学内 Hyogo (JP). 三浦研爾(MIURA, Kenji) [JP/JP]; 〒6578501 兵庫県神戸市灘区六甲台町1番1号国立大学法人神戸大学内 Hyogo (JP). 西尾久英(NISHIO, Hisahide) [JP/JP]; 〒6578501 兵庫県神戸市灘区六甲台町1番1号国立大学法人神戸大学内 Hyogo (JP).
- (74) 代理人: 庄司隆, 外(SHOJI, Takashi et al.); 〒1030004 東京都中央区東日本橋3-4-1 秋場ビル6階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- 添付公開書類:
— 国際調査報告なし; 国際調査報告を受け取り次第公開される。(規則 48.2(g))



WO 2010/026738 A2

(54) Title: METHOD FOR PREDICTING ENZYME ACTIVITY BY COMPUTER SIMULATION

(54) 発明の名称: 酵素活性をコンピュータを用いたシミュレーションにより予測する方法

(57) Abstract: Disclosed is a method for predicting enzyme activity based on the three-dimensional structure of proteins, even without actually measuring enzyme activity using proteins. This is accomplished by performing a docking simulation between a protein and a substrate by using a computer, and then predicting the enzyme activity of the protein by substituting the docking simulation results into equation 1 or 3.

(57) 要約: 本発明は、タンパク質を用いて実際に酵素活性を測定しなくても、タンパク質の立体構造に基づいて酵素活性を予測可能とすることを課題とする。かかる課題は、コンピュータによりタンパク質と基質とのドッキングシミュレーションを行い、かかるドッキングシミュレーションの結果を式1または3に代入して、タンパク質の酵素活性を予測することにより解決される。

明 細 書

発明の名称：

酵素活性をコンピュータを用いたシミュレーションにより予測する方法

技術分野

- [0001] 本発明は、特定のタンパク質Aの酵素活性をコンピュータを用いたシミュレーションにより数式を用いて予測する方法に関する。また本発明は、かかる酵素活性予測方法を実行可能な記録媒体、および装置に関する。さらに、前記酵素活性予測方法、前記記録媒体、または装置を用いて予測された結果を用いた、酵素の基質適合性の判断方法、および、基質の生体への投与間隔等を評価する方法に関する。
- [0002] 本出願は、参照によりここに援用されるところの日本出願特願2008-228434号優先権を請求する。

背景技術

- [0003] 近年、タンパク質の構造解析に、コンピュータを用いた構造シミュレーションが取り入れられるようになってきている。タンパク質の3次元構造の決定などが、コンピュータを用いて行われており、X線結晶解析により決定された3次元構造データからなるデータベースや、ホモロジーモデリング等で決定した3次元構造データからなるデータベースが構築されている（非特許文献1）。タンパク質の3次元構造データは、これらのデータベースから容易に取得することができる。
- [0004] タンパク質の3次元構造データを利用および解析して、タンパク質立体構造構築の原理や、タンパク質の作用メカニズムなどの研究が進んでいるところである。タンパク質の3次元構造データを利用した解析として、タンパク質と低分子化合物であるリガンドとのドッキングシミュレーションが挙げられる（非特許文献2）。これらのシミュレーションに関連するソフトウェア群は、複数開発されており、一般的に利用可能なレベルになっている（非特許文献2～6）。その結果、容易にタンパク質構造シミュレーションを研究

に用いることが可能になってきた。

[0005] ドッキングシミュレーションは、タンパク質とリガンドとの結合能の予測や、化合物ライブラリーから薬剤候補化合物を探索するスクリーニングなどに利用されている（特許文献1，2）。

[0006] タンパク質には生体内で種々の化学反応を触媒する酵素が含まれる。

生体内の酵素に、UDP-グルクロン酸転移酵素（EC 2.4.1.17）（以下「UGT」と称する）がある。UGTはグルクロン酸抱合を触媒する。グルクロン酸抱合により、内因性物質、生体外から投与された薬剤、食物中の化学物質、環境汚染物質などの排出パスウェイが提供されるため、UGTは一次代謝産物の排泄において重要な酵素である。UGTのうち、UGT1A1はビリルビンのグルクロン酸抱合に関与する主要な酵素であることが知られている。UGT1A1遺伝子に変異が生じた場合には、UGT1A1のビリルビンに対する酵素活性が消失または深刻に低下し、黄疸を主症状とするGilbert症候群やCrigler-Najjar症候群の原因となる。UGT1A1の変異型は多数あり、変異部位によって、ビリルビンに対する酵素活性のレベルが異なっている（非特許文献7，8）。また、同じ変異部位を持つ場合でも、変異型UGT1A1の抱合活性は、薬剤によって異なる。

[0007] UGT1A1にみられるように、タンパク質自体の構造や、触媒対象である基質に応じて、酵素の活性レベルは変化する。特定のタンパク質の特定の基質に対する酵素活性について情報を得たい場合は、組換えタンパク質を作製するなどして、実際に酵素活性を測定しなくてはならず、手間と時間がかかってしまうという問題がある。

先行技術文献

特許文献

[0008] 特許文献1：特開2005-181104号公報

特許文献2：特開2007-272627号公報

非特許文献

[0009] 非特許文献1：Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS,

Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, Shen MY, Kelly L, Melo F, Sali A: MODBASE: Nucleic Acids Research 2006, 1(34):D291-D295.

非特許文献2: Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ: J Comput Chem 1998, 19:1639-1662.

非特許文献3: Guex N, Peitsch MC: SWISS-MODEL and the Swiss-PdbViewer: Electrophoresis 1997, 18:2714-2723.

非特許文献4: Ren P, Ponder JW: J Phys Chem B 2003, 107:5933-5947.

非特許文献5: DeLano, W.L. The PyMOL Molecular Graphics System. (2008) DeLano Scientific LLC, Palo Alto, CA, USA., インターネット<URL: <http://www.pymol.org>>

非特許文献6: Molecular Operating Environment (moe), version 2008.10. Chemical Computing Group, Inc. Montreal, Quebec, Canada, 2008, インターネット<URL: <http://www.chemcomp.com>>

非特許文献7: Udomuksorn W, Eliot DJ, Lewis BC, Mackenzie PI, Yoovathaworn K, Miners JO: Pharmacogenetics & Genomics 2007, 17(12):1017-1029.

非特許文献8: Yamamoto K, Sato H, Fujiwara Y, Doida Y, Bamba T: Biochimica et Biophysica Acta 1998, 1406:267-273.

発明の概要

発明が解決しようとする課題

[0010] 本発明は、タンパク質を用いて実際に酵素活性を測定しなくても、タンパク質の立体構造データに基づいて酵素活性を予測可能とすることを課題とする。

課題を解決するための手段

[0011] 本発明者らは上記課題を解決するために鋭意検討を重ね、コンピュータによりタンパク質と基質との分子シミュレーションを行い、なかでもドッキングシミュレーションの結果を数式を用いて解析することにより、タンパク質

の酵素活性を予測することが可能であることを見出し、本発明を完成した。

[0012] すなわち、本発明は以下よりなる。

1. タンパク質Aの酵素活性をコンピュータを用いたシミュレーションにより予測する方法であって、

酵素活性がタンパク質Aへの基質の進入の向きにより規定されるようなタンパク質であり、

タンパク質Aの酵素活性 f が以下の式1により算出される方法；

[数1]

$$f = g \times E \times \left(\frac{n}{N} \right)^\beta + a \quad \dots \text{式1}$$

式中、 g は基質ごとに固有の定数であり、 E はタンパク質Aと補酵素とのドッキングの酵素活性に対する寄与度であり、 β はタンパク質への基質の進入の向きの酵素活性に対する寄与度であり、 a は生体内環境による影響を表す定数であり、 N と n は、タンパク質Aと基質とのドッキングシミュレーションにより得られる値であり、 N はタンパク質Aと基質とのドッキングシミュレーションの総回数であり、 N は2以上であり、 n は酵素反応を受け得る向きで基質がタンパク質Aに進入した回数である。

2. タンパク質Aの立体構造データが、タンパク質Aとは別のタンパク質Bの立体構造データに基づいて計算されるものであり、 β が以下の式2によって算出される前項1に記載の方法；

[数2]

$$\beta = \frac{N_w}{n_w} \quad \dots \text{式2}$$

式中、 N_w と n_w は、タンパク質Bと基質とのドッキングシミュレーションにより得られる値であり、 N_w はタンパク質Bと基質とのドッキングシミュレ-

シヨンの総回数であり、 N_w は2以上であり、 n_w は酵素反応を受け得る向きで基質がタンパク質Bに進入した回数である。

3. タンパク質Bの酵素活性に対するタンパク質Aの相対的な酵素活性をコンピュータを用いたシミュレーションにより予測する方法であって、酵素活性がタンパク質Aへの基質の進入の向きにより規定されるようなタンパク質であり、

タンパク質Aの相対的な酵素活性 f' が以下の式3により算出される方法；

[数3]

$$f' = \frac{g \times E \times \left(\frac{n}{N}\right)^\beta + a}{g \times E \times \left(\frac{n_w}{N_w}\right)^\beta + a} \times 100 \quad \dots \text{式3}$$

式中、 g は基質ごとに固有の定数であり、 E はタンパク質Aと補酵素とのドッキングの酵素活性に対する寄与度であり、 a は生体内環境による影響を表す定数であり、 β はタンパク質への基質の進入の向きの酵素活性に対する寄与度であり、次の式2により表され；

[数4]

$$\beta = \frac{N_w}{n_w} \quad \dots \text{式2}$$

N と n は、タンパク質Aと基質とのドッキングシミュレーションにより得られる値であり、 N はタンパク質Aと基質とのドッキングシミュレーションの総回数であり、 N は2以上であり、 n は酵素反応を受け得る向きで基質がタンパク質Aに進入した回数であり、 N_w と n_w はタンパク質Bと基質とのドッキングシミュレーションにより得られる値であり、 N_w はタンパク質Bと基質とのドッキングシミュレーションの総回数であり、 N_w は2以上であり、 n_w

は酵素反応を受け得る向きで基質がタンパク質Bに進入した回数である。

4. Eが、下記の式16または式17により表される、前項1～3のいずれか1に記載の方法：

[数5]

$$\left(\frac{l}{L} \right)^{\gamma} \quad \dots \text{式 16}$$

(式16中、 γ はタンパク質への補酵素の進入の向きの酵素活性に対する寄与度であり、次の式21により表され、

[数6]

$$\gamma = \frac{L_w}{l_w} \quad \dots \text{式 21}$$

Lとlは、タンパク質Aと補酵素とのドッキングシミュレーションにより得られる値であり、Lはタンパク質Aと補酵素とのドッキングシミュレーションの総回数であり、Lは2以上であり、lは酵素反応が進行し得る向きで補酵素がタンパク質Aに進入した回数であり、 L_w と l_w はタンパク質Bと補酵素とのドッキングシミュレーションにより得られる値であり、 L_w はタンパク質Bと補酵素とのドッキングシミュレーションの総回数であり、 L_w は2以上であり、 l_w は酵素反応が進行し得る向きで補酵素がタンパク質Bに進入した回数である)；

[数7]

$$\left(\frac{m}{m_w} \right)^{\delta} \quad \dots \text{式 17}$$

(式17中、mはタンパク質Aと補酵素とのドッキングモデルをクラスタリングして得られたクラス内のモデル数であり、 m_w はタンパク質Bと補酵素

とのドッキングモデルをクラスタリングして得られたクラスタ内のモデル数であり、 δ はクラスタ内のモデル数の酵素活性への寄与度を表す。)

5. 式1または式3において、

[数8]

$$\frac{n}{N} \quad \text{および/または} \quad \frac{n_w}{N_w}$$

について、Sigmoid関数を導入する、前項1～4のいずれか1に記載の方法。

6. 2以上のタンパク質Aについてドッキングシミュレーションを行い、 g と a が、ドッキングシミュレーションにより得られた計算値 y と、測定された値 y' との二乗誤差を最小にする値であり、下記の式4を用いて算出される、前項2～5のいずれか1に記載の方法。

[数9]

$$(g, a) = \underset{g, a}{\operatorname{argmin}} \left\{ (y_w - y'_w)^2 + (y_{A1} - y'_{A1})^2 + \dots + (y_{Ap} - y'_{Ap})^2 \right\} \quad \dots \text{式4}$$

式中、 y_w と y'_w はタンパク質Bについての値であり、 y_{A1} と y'_{A1} 、 y_{Ap} と y'_{Ap} はタンパク質Aについての値であり、 p は2以上の数を表す。

7. タンパク質と基質とのドッキングシミュレーションが以下の工程を含む前項1～6のいずれか1に記載の方法：

(a) タンパク質Bの立体構造データを入手し、タンパク質Aの立体構造データをタンパク質Bの立体構造データに基づいて計算し、

(b) タンパク質AまたはBと補酵素とのドッキングシミュレーションを行い、熱力学的に安定なドッキングモデルを決定し；

(c) タンパク質AまたはBと基質とのドッキングのグリッドを設定し；

(d) タンパク質Bと基質とのドッキングシミュレーションを N_w 回を行い、 N_w は2以上であり、酵素反応を受け得る向きで基質がタンパク質Bに進入した回数 n_w を計数し、

(e) タンパク質Aと基質とのドッキングシミュレーションを N 回を行い、 N

は2以上であり、酵素反応を受け得る向きで基質がタンパク質Aに進入した回数 n を計数する。

8. 工程 (b) の後に次の工程 (b-1) を行い;

工程 (b-1) タンパク質AまたはBと補酵素とのドッキングモデルをクラスタリングし、クラスタリングして得られたクラスタ内の2以上のモデルについてinduced fitを行い、

工程 (d) において、induced fit後の各モデルについて、基質とのドッキングシミュレーションを行う、

前項1~7のいずれか1に記載の方法。

9. タンパク質Aが変異型タンパク質である前項1~7のいずれか1に記載の方法。

10. タンパク質AがUDP-グルクロン酸転移酵素であり、補酵素がUDP-グルクロン酸である、前項4~9のいずれか1に記載の方法。

11. タンパク質Aが g と a の算出のために用いた変異型タンパク質以外の変異型タンパク質であり、算出された g と a の値を用いて酵素活性の予測を行う、前項6~10のいずれか1に記載の方法。

12. 前項1~11のいずれか1に記載の方法を実行するために、コンピュータを下記的手段として機能させるプログラムを担持する記録媒体:

(1) 入力されたアミノ酸配列情報に基づいて、タンパク質Aの立体構造データを計算する手段、

(2) タンパク質Aの立体構造データと、ドッキングシミュレーションの対象となる基質の立体構造データを記憶する手段、

(3) 前記記憶された、タンパク質Aの立体構造データと基質の立体構造データを用いて、タンパク質Aと基質とのドッキングシミュレーションを行うシミュレーション手段、

(4) シミュレーションにより得られた結果を記憶する手段、

(5) 記憶されたシミュレーション結果に基づいて、タンパク質Aの酵素活性を算出する手段、

(6) 算出された酵素活性を表示する手段。

13. 前項1～11のいずれか1に記載の方法を実行するために、下記の手段を担持する装置；

(1) 入力されたアミノ酸配列情報に基づいて、タンパク質Aの立体構造データを計算する手段、

(2) タンパク質Aの立体構造データと、ドッキングシミュレーションの対象となる基質の立体構造データを記憶する手段、

(3) 前記記憶された、タンパク質Aの立体構造データと基質の立体構造データを用いて、タンパク質Aと基質とのドッキングシミュレーションを行うシミュレーション手段、

(4) シミュレーションにより得られた結果を記憶する手段、

(5) 記憶されたシミュレーション結果に基づいて、タンパク質Aの酵素活性を算出する手段、

(6) 算出された酵素活性を表示する手段。

14. 前項1～11のいずれか1に記載の方法、前項12に記載の記録媒体、または前項13に記載の装置を用いて、2以上の基質について、基質ごとのタンパク質Aの酵素活性を予測し、得られた2以上の予測結果に基づいて目的の基質を選択する、基質適合性の判定方法。

15. 前記基質が生体に投与される薬剤であって、前項14に記載の方法を用いて、薬剤の投与量および/または投与間隔、投与頻度を評価する方法。

発明の効果

[0013] 本発明によれば、実際にタンパク質を用いて酵素活性の測定をすることなく、タンパク質の立体構造データに基づいて酵素活性を予測することができる。例えば組換えタンパク質を作製し、実際に酵素活性の測定をする場合には、数週間の時間を要するが、本発明の方法を用いることにより、数時間で酵素活性の予測をすることが可能である。また、本発明により予測された酵素活性は、実際にタンパク質を用いて測定して得られた値と相関性が見られ、信頼性が高い。

図面の簡単な説明

- [0014] [図1] コンピュータを用いて計算された各種UGT1A1の立体構造を示す。
(実施例1)
- [図2] 各種UGT1A1とUDPGAとのドッキングモデルを示す。(実施例2)
- [図3] 各種UGT1A1とUDPGAとのドッキングにおけるドッキングエネルギーの分布を示す。(実施例2)
- [図4] 基質のUGT1A1への進入の2つの向きを示す。(実施例3)
- [図5] *in vitro*実験における変異型UGT1A1の抱合活性を示す。(実施例4)
- [図6] 変異型UGT1A1の抱合活性の*in vitro*測定値と、シミュレーションに基づく計算値との比較を示す。(実施例6)
- [図7] 野生型UGT1A1と34種の変異型UGT1A1の抱合活性のシミュレーションに基づく計算値を示す。(実施例6)
- [図8] UGT1A1とビリルビンとのドッキングシミュレーションのためのグリッドを示す(a: 抱合反応空間の断面を表示した図、b: 抱合反応空間の正面から見た図)。(実施例7)
- [図9] 変異型UGT1A1の抱合活性の*in vitro*測定値(文献報告)と、シミュレーションに基づく計算値との比較を示す。(実施例7)
- [図10] 変異型UGT1A1の抱合活性の*in vivo*測定値(文献報告)と、シミュレーションに基づく計算値との比較を示す。(実施例8)
- [図11] 変異型UGT1A1の抱合活性の*in vitro*測定値(文献報告)と、シミュレーションに基づく計算値との比較を示す。(実施例9)
- [図12] 変異型UGT1A1の抱合活性の*in vivo*測定値(文献報告)と、シミュレーションに基づく計算値との比較を示す。(実施例10)
- [図13] 変異型UGT1A1の抱合活性の*in vivo*測定値(文献報告)と、シミュレーションに基づく計算値との比較を示す。(実施例11)
- [図14] 水分子を付加して構造計算した各種UGT1A1とUDPGAとのド

ッキングシミュレーションにおけるドッキングエネルギーの分布を示す。（実施例 1 2）

[図15]水分子を付加して構造計算した各種UGT 1 A 1と基質とのドッキングシミュレーションの結果を示す。（実施例 1 2）

[図16]水分子を付加して構造計算した各種UGT 1 A 1と基質とのドッキングシミュレーションの結果と、各種UGT 1 A 1の抱合活性のin vitro測定値との比較を示す。（実施例 1 2）

[図17]水分子を付加して構造計算した変異型A r l 6と、GTP γ SもしくはGDPとのドッキングシミュレーションの結果を示す。（実施例 1 2）

[図18]induced fitを行った場合の、各種UGT 1 A 1と基質とのドッキングシミュレーションの結果を示す。（実施例 1 3）

[図19]induced fitを行った場合の、各種UGT 1 A 1と基質とのドッキングシミュレーションの結果と、各種UGT 1 A 1の抱合活性のin vitro測定値との比較を示す。（実施例 1 3）

[図20]クラスタ内の全モデルに対してinduced fitを行った各種UGT 1 A 1と基質とのドッキングシミュレーションの結果と、各種UGT 1 A 1の抱合活性のin vitro測定値との比較を示す。（実施例 1 3）

[図21]式 1 0における項RにSigmoid関数を導入して算出した、各種UGT 1 A 1の抱合活性の計算値と、各種UGT 1 A 1の抱合活性のin vitro測定値との比較を示す。（実施例 1 5）

[図22]式 1 0における項RにSigmoid関数を導入して算出した、野生型UGT 1 A 1と3 4種の変異型UGT 1 A 1の抱合活性の計算値を示す。（実施例 1 5）

[図23]各種UGT 1 A 1の抱合活性のin vitro測定値と、シミュレーションに基づく計算値との比較を示す。（実施例 1 6）

発明を実施するための形態

[0015] 本発明は、特定のタンパク質Aの酵素活性をコンピュータを用いたシミュレーションにより予測する方法に関する。一般にタンパク質の酵素活性は、

基質の種類に応じて違いがある。本発明においては、下記の式 1 を用いることにより、特定の基質に対する特定のタンパク質 A の酵素活性を予測することができ、基質、タンパク質 A とともに、種々の種類を選択して組み合わせ、解析を行うことができる。

[数10]

$$f = g \times E \times \left(\frac{n}{N} \right)^\beta + a \quad \cdots \text{式 1}$$

[0016] 本発明において解析の対象となるタンパク質 A は、酵素活性がタンパク質 A への基質の進入の向きにより規定されるようなタンパク質である。一般的にタンパク質の酵素活性のレベルは、タンパク質の量と、タンパク質の性質により規定される。生体内では、タンパク質の量はタンパク質の合成・分解のバランスにより制御され、タンパク質が他の生体分子と可逆的に作用したり、立体構造が変化することによりタンパク質の性質が変化する。本発明において予測される「酵素活性」は、タンパク質の立体構造に基づくタンパク質の性質（触媒能）であり、かかる立体構造に基づく触媒能は、補酵素とのドッキングエネルギー、補酵素のタンパク質への進入の向き、基質の進入口の断面の大きさ、基質とのドッキングエネルギー、基質と触媒部位との距離、基質のタンパク質への進入の向きなど、種々の要因により規定されると考えられる。これらの要因のうち、特に基質のタンパク質への進入の向きにより酵素活性（タンパク質の立体構造に基づく触媒能）が規定されるタンパク質について、式 1 を用いた本発明により酵素活性（タンパク質の立体構造に基づく触媒能）を予測することが可能である。

[0017] かかるタンパク質 A は、いかなる反応を触媒する酵素であってもよく、細胞膜や、細胞質内等のいかなる場所に存在する酵素であってもよい。タンパク質 A は、補酵素から基質への原子団の転移を触媒する転移酵素であることが好ましい。中でも、グリコシル基転移酵素が好ましく、グリコシル基転移酵素としては、例えば UDP-グルクロン酸転移酵素（以下「UGT」と称

する)、ラクトースシンターゼ (lactose synthase) (EC 2.4.1.22) が挙げられる。UGTは配列相同性と遺伝子構造によってUGT 1 A、UGT 2 A、UGT 2 Bの3つのサブファミリーに区別される。サブファミリーUGT 1 Aは染色体2 q 3 7に位置し、9種類のアイソフォーム、UGT 1 A 1、UGT 1 A 3、UGT 1 A 4、UGT 1 A 5、UGT 1 A 6、UGT 1 A 7、UGT 1 A 8、UGT 1 A 9、UGT 1 A 10が存在する。これらのアイソフォームは選択的スプライシングにより変化する最初のエキソンと、共通する4つのエキソンとからなる。9種類のアイソフォームの1種である、UGT 1 A 1が本発明において解析されるタンパク質として特に好ましい。

[0018] 本発明では、タンパク質Aと基質のドッキングシミュレーションをコンピュータ上で行うことが必須である。ドッキングシミュレーションに必要なタンパク質Aの立体構造データは、公知のデータベースからダウンロードするか、X線結晶構造解析、核磁気共鳴法などにより入手するか、または、公知のタンパク質Bの立体構造データに基づいて計算して入手することができる。公知のタンパク質Bの立体構造データは、公知のデータベースからダウンロードするか、X線結晶構造解析などにより入手することができる。

[0019] タンパク質Bは、タンパク質Aの立体構造データを計算するための基準となるタンパク質であり、タンパク質Aとは異なるタンパク質であるが、タンパク質Aと同様の触媒作用を有するものである。またタンパク質Bは、タンパク質Aと高いアミノ酸相同性(25~30%以上、好ましくは40%以上)を有しており、かつタンパク質Aと構造類似性を有しているものが好ましい。

[0020] タンパク質Aは、基準となるタンパク質(タンパク質B)とアミノ酸配列が相違するものである。タンパク質Aとタンパク質Bの関係としては、例えば、変異型タンパク質と野生型タンパク質の関係や、互いにアイソフォームの関係などが挙げられる。さらに具体的には、変異型タンパク質と野生型タンパク質との関係では、タンパク質Aが変異型UGT 1 A 1であり、タンパク質Bは野生型UGT 1 A 1であり、アイソフォームの関係では、タンパク

質AがUGT1A6であり、タンパク質BがUGT1A1である場合などが例示される。

[0021] タンパク質Aとタンパク質Bの関係が変異型タンパク質と野生型タンパク質である場合、タンパク質Aのアミノ酸配列は、タンパク質Bのアミノ酸配列においてアミノ酸が置換、欠失、挿入、および/または付加されているものである。置換、欠失、挿入、および/または付加されているアミノ酸の個数は1以上であればよい。また、変異型タンパク質には、現在同定されていないあらゆる変異部位を持つタンパク質が含まれる。現在同定されていない変異部位は、天然に存在する変異部位であってもよいし、天然には存在しない人工的な変異部位であってもよい。

[0022] タンパク質Aと基質のドッキングシミュレーションについて説明する。

まず、タンパク質Aの立体構造データを入手する。タンパク質Aの立体構造データは、公知のデータベースからダウンロードする、X線結晶構造解析、核磁気共鳴法などにより入手する、または、公知のタンパク質Bの立体構造データに基づいて計算して入手することができる。公知のタンパク質Bの立体構造データは、公知のデータベースからダウンロードする、または、X線結晶構造解析などにより入手することができる。データベースとしては、PDB (<http://www.pdbj.org/>)、MODBASE (<http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi>, Morris GM et al.: J Comput Chem 1998, 19:1639-1662) が例示される。

[0023] タンパク質Aの立体構造データは、自体公知の方法およびソフトウェアを用いて、公知のタンパク質Bの立体構造データに基づいて計算することが可能である。例えば、タンパク質Bの立体構造データを用いてホモロジーモデリングなどの手法を用いて計算すればよい。立体構造の計算に用いるソフトウェアとしては、公知のものを使用可能であり、今後開発されるものも使用可能である。以下に具体的に例示して、立体構造データの構築の方法を説明する。タンパク質Aが1アミノ酸置換の変異型UGT1A1であり、タンパク質Bが野生型UGT1A1である場合を例示する。

[0024] 野生型UGT1A1の立体構造データを、MODBASE (accession number: Q5DT03) よりダウンロードする。ダウンロードした野生型UGT1A1の立体構造データに、PyMOLプログラム (<http://pymol.sourceforge.net/>, DeLano WL: DeLano Scientific, Palo alto, CA, USA; 2002) を用いて適切に水素原子を追加し、SWISS-PDBViewerプログラム (<http://spdbv.vital-it.ch/>, Gue x N, Peitsch MC: Electrophoresis 1997, 18: 2714-2723) を用いて、変異型UGT1A1の立体構造データを作成する。例えば、公知の変異部位を持つ、G71R変異型UGT1A1 (71番目のグリシンがアルギニンに置換)、F83L変異型UGT1A1 (83番目のフェニルアラニンがロイシンに置換)、I322V変異型UGT1A1 (322番目のイソロイシンがバリンに置換) などについて、立体構造データを作成することができる。TINKERパッケージ (Ren P, Ponder JW: J Phys Chem B 2003, 107:5933-5947) のminimizeプログラムとAMBER99力場パラメータによるエネルギー最小化計算を、RMS勾配が例えば0.3になるまで行い、各タンパク質の立体構造を計算する。力場等は、後の工程で用いるドッキングシミュレーションのソフトウェアに応じて適宜設定可能である。計算されたG71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1の三次元構造を図1に示す。

[0025] タンパク質Aの立体構造データを計算する際、UGT1A1の場合と同様に水分子を付加せずに計算をおこなってもよいが、水分子を付加して計算を行ってもよい。例えば、タンパク質Aが細胞膜に存在するタンパク質の場合は、水分子を付加せずに、上述の手法で計算することが好ましい。タンパク質Aが細胞質に存在するタンパク質の場合は、上述の手法に加えて、水分子を付加して立体構造データを計算することが好ましい。タンパク質Aの存在する箇所や、触媒反応の起こる部位などについて、疎水性度および親水性度を考慮することにより水分子付加の要否を決定して、立体構造データの計算を行うことが可能である。なお、タンパク質Aへの水分子の付加は、MOE (Chemical Computing Group Inc.) 等のプログラムを用いて行うことができる。

[0026] 次に、ドッキングプログラムにより、タンパク質Aの立体構造データを用いてドッキングシミュレーションを行う。基質や補酵素のデータは、既存のデータベース（例えば、ChemIDPlus (<http://chem.sis.nlm.nih.gov/chemidplus/>) など）からダウンロードすることができる。ドッキングシミュレーションに用いるソフトウェアとしては、公知のものや今後開発されるものを使用可能であり、Dock (<http://dock.compbio.ucsf.edu/>)、AutoDock (<http://scripps.edu/mb/olson/doc/autodock/>)、GOLD (http://ccdc.cam.ac.uk/products/life_sciences/gold/)、MOEなどが例示される。それぞれのソフトウェアの使用方法に従って、ドッキングシミュレーションを行えばよい。

[0027] タンパク質Aが補酵素を必要とする酵素（例えば転移酵素）である場合は、まず補酵素とタンパク質Aとのドッキングシミュレーションを行い、安定なドッキングモデルを決定する必要がある。基質とのドッキングシミュレーションには、補酵素とタンパク質Aとの安定なドッキングモデルを用いればよい。補酵素とタンパク質Aとの安定なドッキングモデルは、1つであっても複数であってもよい。1つの安定なドッキングモデルを用いる場合は、後述する手法にて最も安定なドッキングモデルを選択して用いることができる。複数の安定なドッキングモデルを用いる場合は、後述する手法にて、補酵素とタンパク質Aとのドッキングモデルについて階層型クラスタリングを行って得られるクラスタ（最も安定な集合C）内の2以上のドッキングモデル、好ましくは全ドッキングモデルを用いることができる。

[0028] 変異型UGT1A1について、AutoDock 4プログラム (Morris GM et al.: J Comput Chem 1998, 19:1639-1662) を用いたドッキングシミュレーションを例示しながら説明する。

UGTはグルクロン酸抱合に関与する酵素であり、補酵素であるUDP-グルクロン酸（以下「UDPGA」と称する）から、基質（例えば、生体内物質であるビリルビンや、薬剤であるイリノテカン）にグルクロン酸を転移させる機能を持つ。UGT1A1と基質とのドッキングシミュレーションを行うために、まず、UGT1A1とUDPGAのドッキングシミュレーショ

ンを行う。

[0029] UDPGAの立体構造データは、ChemIDPlusに登録されているデータをダウンロードして用いることができる (registry number: 2616-64-0)。5種類のUDPGAの立体構造データが存在するが、これら5種類の立体構造データを用いてドッキングシミュレーションを行えばよい。

以下にAutoDockを用いてドッキングシミュレーションを行った例を用いて説明する。まず、UDPGAの配置を探索するマップを、グリッド間隔0.375 Å、60×60×60ポイントの立方体としてAutoGridプログラムを用いて生成する。グリッド探索アルゴリズムは、ラマルク型遺伝アルゴリズムを使用すればよい。その他のパラメータはAutoDock 4のデフォルト値を使用すればよい。

[0030] 変異型UGT1A1と5種類のUDPGAの組み合わせ毎に、複数回、例えば10～100回ずつドッキングシミュレーションを実行する。10回ずつシミュレーションを行った場合は、1種の変異型UGT1A1について、50個の計算結果が得られる。得られた50個の計算結果のうち、抱合反応可能な向きでUDPGAがドッキングしているものについて、ドッキングエネルギーの平均値を計算する。ドッキングエネルギー ΔG は下記の式6で計算される。

[数11]

$$\Delta G = \Delta G_{\text{inter}} + \Delta G_{\text{intra}} + \Delta G_{\text{torsion}} - \Delta G_{\text{unbound}} \quad \dots \text{式 6}$$

[0031] 変異型UGT1A1とUDPGAとのドッキングシミュレーションにて得られた50個の結果を、分子間エネルギー、UDPGAの分子内部エネルギー、およびアンバウンドエネルギーのそれぞれについて、低い順に順位づけを行う。各結果について分子間エネルギー、UDPGAの分子内部エネルギー、およびアンバウンドエネルギーの3種類の順位の順位和を算出し、順位和が最小のドッキング結果を最も安定なドッキングモデルとして選択し、変異型UGT1A1とUDPGAのドッキングモデルとして決定することがで

きる。この時、ドッキングモデルに含まれる水素結合は、PyMOLを用いて検出すればよい。図2に各種UGT1A1とUDPGAとのドッキングモデルを示す。

- [0032] 変異型UGT1A1と基質のドッキングシミュレーションを行うために、グリッドを決定する必要がある。
- [0033] 変異型UGT1A1とUDPGAとのドッキングモデルから、ドッキングエネルギーが低い一群のシミュレーション結果を、群平均法（2つのクラスターを順次統合していく階層型クラスタリングの一種）によるクラスタリングで、グループ化して定義する。
- [0034] 群平均法では、全てのクラスター間の距離を計算し、最も距離の小さい2つのクラスターを統合する。クラスターXとYの間の距離 $d(X, Y)$ は以下の式7のように定義される。

[数12]

$$d(X, Y) = \frac{1}{|X||Y|} \sum_{x \in X} \sum_{y \in Y} |x - y| \quad \dots \text{式7}$$

- [0035] 変異型UGT1A1と抱合反応可能な向きで結合しているUDPGAとの分子間エネルギー、分子内部エネルギー、アンバウンドエネルギーの順位和を指標として群平均法でクラスタリングを行い、「相対的に低いドッキングエネルギーを有する数種類のドッキング構造」の集合（本明細書にて「最も安定な集合C」とも称する。）を求める。

ここで、最も安定な集合Cの至適な粒度を決定することで、グループ化が可能である。クラスターXの粒度 $G(X)$ を、抱合反応可能なUDPGA全体の集合をSとした次式8により計算する。

[数13]

$$G(X) = \frac{\max_{x,y \in X} |x - y|}{\max_{z,w \in S} |z - w|} \quad \dots \text{式8}$$

[0036] 階層型クラスタリングでは、様々な粒度のクラスタリング結果が得られる。粒度の決定のために、野生型、G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1と、UDPGAのドッキングシミュレーション結果を使って、群平均法によるクラスタリングを行い、全てのクラスタの粒度を計算してその分布を解析する。例えば、5%水準で、平均と比して有意に大きい粒度を除外するものとしたので、棄却域の値は0.56であった。つまり、粒度0.56以下のクラスタのうち最大のものを、グループとして定義可能であった。

[0037] 最も安定な集合Cは、抱合反応可能なドッキング結果から群平均法で得られたクラスタをgac(S)と定義して、次の式9のように計算する。

[数14]

$$C = \operatorname{argmax}_{X \in T} (G(X)) \quad \dots \text{式9}$$

式中、 $T = \{x \mid x \in \text{gac}(S), \min(S) \in x, G(x) < 0.56\}$

上記式9により定義されたクラスタの位置から分子の揺らぎを加味した抱合反応の場を決定することができる。

変異型UGT1A1と基質とのドッキングシミュレーションにおいて、基質の大きさに合わせて、グリッドの最奥の位置を決定すればよい。基質が、アセトアミノフェンなどの比較的小さな化合物である場合は、グリッドは1つ設定すればよいが、ビリルビンなどの比較的大きな化合物である場合は、グリッドを複数、例えば3つ設定すればよい。グリッドの個数は、基質の大

きさだけではなく、タンパク質の反応の場の立体構造も加味して決定される。

- [0038] また、かかるクラスタリングにより得られたクラスタ（最も安定な集合C）に含まれる補酵素とUGT1A1とのドッキングモデルのうち、2以上を選択して、基質とのドッキングシミュレーションに用いることができる。好ましくは、最も安定な集合Cに含まれる全ドッキングモデルについて、基質とのドッキングシミュレーションを行う。
- [0039] 定めたグリッド位置を用いて、UGT1A1と基質（例えば、アセトアミノフェン（以下「AAP」と称する）、エストラジオール（以下「E2」と称する）、ビリルビン等）とのドッキングシミュレーションを行う。AAPとE2の立体構造データはChemIDPlusに登録されているものを用いればよい。ドッキングシミュレーションは、ドッキングモデルごと及び／又は基質ごとに複数回、例えば10～1000回行えばよい。100回程度でもよい。ドッキングシミュレーションごとに、基質のUGT1A1への進入の向きを評価する。基質の変異型UGT1A1への進入の向きが、抱合反応可能な向きであった場合を計数する。
- [0040] 基質のタンパク質への進入の向きは、基質のUGT1A1への進入の向きを例にすると、図4に示す2種類が考えられる。＜向きI＞では、基質のヒドロキシル基がUDPGAに向いており、グルクロン酸抱合反応の進行が可能である。＜向きII＞では、基質のヒドロキシル基がUDPGAの反対側に向いており、グルクロン酸抱合反応の進行が困難である。本発明では、基質が酵素反応を受け得る向きでタンパク質に進入した回数、すなわちUGT1A1の場合は、基質が＜向きI＞でUGT1A1に進入したドッキングシミュレーションの回数を計数する。
- [0041] また、タンパク質Aと補酵素とのドッキングモデルを得た後、基質とのドッキングシミュレーションの前に、induced fit（誘導結合）を行ってもよい。本発明においてinduced fitとは、Flexible Dockingを行うことによりタンパク質の活性部位の立体構造を変化させることを意味する。Flexible Dockin

gは、MOE (Chemical Computing Group Inc.) 等のプログラムにより行うことができる。

[0042] induced fitは、タンパク質Aと補酵素とのドッキングモデルを得た後、基質とのドッキングシミュレーションに用いるドッキングモデルを対象に行う。1種のドッキングモデルに対して行ってもよいし、2以上の複数のドッキングモデルに対して行ってもよいが、上述の手法でクラスタリングされた「最も安定な集合C」のクラスタに含まれる全てのドッキングモデルに対して行うことが好ましい。induced fitを行った後の補酵素とのドッキングモデルを用いて、基質とのドッキングシミュレーションを行い、基質が酵素反応を受け得る向きでタンパク質に進出した回数を計数する。

[0043] ドッキングシミュレーションにより得られた結果を用いて、タンパク質Aの酵素活性を予測する。タンパク質Aの酵素活性 f は、以下の式10を用いて一般的に算出することができる。

[数15]

$$f = g \times E \times R + a \quad \dots \text{式10}$$

g は基質ごとに固有の定数であり、 E はタンパク質Aと補酵素とのドッキングの酵素活性に対する寄与度であり、 R は基質が酵素反応を受け得る向きで進出した割合であり、 a は、タンパク質Aの酵素反応が進行している環境、例えば生体内環境による影響を表す定数である。

[0044] R はドッキングシミュレーションにより得られた結果を用いて以下の式11を用いて表される。

[数16]

$$R = \left(\frac{n}{N} \right)^\beta \quad \dots \text{式11}$$

式中、 N と n は、タンパク質Aと基質とのドッキングシミュレーションにより得られる値であり、 N はタンパク質Aと基質とのドッキングシミュレーションの総回数であり、 N は2以上であり、 n は酵素反応を受け得る向き（UGT1A1の場合は、〈向きI〉）で基質がタンパク質Aに進入した回数である。 β はタンパク質への基質の進入の向きの酵素活性に対する寄与度である。例えば、タンパク質Aが変異型であり、タンパク質Aの立体構造データをタンパク質Bの立体構造データに基づいて計算した場合、 β はタンパク質Bへの基質の進入の向きの酵素活性に対する寄与度とすればよい。

[0045] したがって、本発明の方法では下記の式1を用いて、ドッキングシミュレーションから得られた結果を用いてタンパク質Aの酵素活性 f を予測することができる。

[数17]

$$f = g \times E \times \left(\frac{n}{N} \right)^{\beta} + a \quad \dots \text{式1}$$

[0046] また、タンパク質Bの酵素活性に対するタンパク質Aの相対的な酵素活性 f' は、以下の式3により算出される。

[数18]

$$f' = \frac{g \times E \times \left(\frac{n}{N} \right)^{\beta} + a}{g \times E \times \left(\frac{n_w}{N_w} \right)^{\beta} + a} \times 100 \quad \dots \text{式3}$$

[0047] 式中、 N_w と n_w は、タンパク質Bと基質とのドッキングシミュレーションにより得られる値であり、 N_w はタンパク質Bと基質とのドッキングシミュレーションの総回数であり、 N_w は2以上であり、 n_w は酵素反応を受け得る向きで基質がタンパク質Bに進入した回数である。

[0048] また、 β は以下の式 2 によって算出することができる。

[数19]

$$\beta = \frac{N_w}{n_w} \quad \dots \text{式 2}$$

[0049] タンパク質 A と補酵素とのドッキングにおける、補酵素の向きやドッキングエネルギーなどが酵素活性に寄与していないか、もしくは、関与が極めて小さい場合は、 $E = 1$ とすることができる。ドッキングエネルギーが酵素活性に関与していない場合とは、例えばドッキングシミュレーションにより得られたドッキングエネルギー結果が、測定された値 y' と相関がない場合を意味する。この場合、式 1 と式 3 はそれぞれ以下の式 12 と式 13 により表される。

[数20]

$$f = g \times \left(\frac{n}{N} \right)^\beta + a \quad \dots \text{式 12}$$

[数21]

$$f' = \frac{g \times \left(\frac{n}{N} \right)^\beta + a}{g \times \left(\frac{n_w}{N_w} \right)^\beta + a} \times 100 \quad \dots \text{式 13}$$

[0050] また、式 1 における E (補酵素とタンパク質 A とのドッキングの酵素活性に対する寄与度) は、補酵素の酵素への進入の向きが酵素活性に対して寄与する場合は、以下の式 16 または式 17 のいずれかを代入することも可能である。好ましくは、式 17 を代入する。

[数22]

$$\left(\frac{l}{L} \right)^{\gamma} \quad \dots \text{式 16}$$

(式 16 中、 γ はタンパク質への補酵素の進入の向きの酵素活性に対する寄与度であり、次の式 21 により表され、

[数23]

$$\gamma = \frac{L_w}{l_w} \quad \dots \text{式 21}$$

L と l は、タンパク質 A と補酵素とのドッキングシミュレーションにより得られる値であり、 L はタンパク質 A と補酵素とのドッキングシミュレーションの総回数であり、 L は 2 以上であり、 l は酵素反応が進行し得る向きで補酵素がタンパク質 A に進入した回数であり、 L_w と l_w はタンパク質 B と補酵素とのドッキングシミュレーションにより得られる値であり、 L_w はタンパク質 B と補酵素とのドッキングシミュレーションの総回数であり、 L_w は 2 以上であり、 l_w は酵素反応が進行し得る向きで補酵素がタンパク質 B に進入した回数である) ;

[0051] [数24]

$$\left(\frac{m}{m_w} \right)^{\delta} \quad \dots \text{式 17}$$

(式 17 中、 m はタンパク質 A と補酵素とのドッキングモデルをクラスタリングして得られたクラスタ内のモデル数であり、 m_w はタンパク質 B と補酵素とのドッキングモデルをクラスタリングして得られたクラスタ内のモデル数であり、 δ はクラスタ内のモデル数の酵素活性への寄与度を表す。 δ は、ドッキングシミュレーションの結果に基づいて式 17 を用いて算出した計算値

y と測定された値 y' との二乗誤差を最小にするように、下記式 23 を用いて算出することができる。

[数25]

$$\delta = \underset{\delta}{\operatorname{argmin}} \left\{ (y_w - y'_w)^2 + (y_{A1} - y'_{A1})^2 + \cdots + (y_{Ap} - y'_{Ap})^2 \right\} \quad \cdots \text{式 23}$$

式 23 中、 y_w と y'_w はタンパク質 B についての値であり、 y_{A1} と y'_{A1} 、 y_{Ap} と y'_{Ap} はタンパク質 A についての値であり、 p は 2 以上の数を表す。))

[0052] 式 1、3、12、13 等において、 g と a は、予測対象となる基質について予め設定された数値を用いることができる。あるいは、 g と a の値が不明である場合は、2 以上のタンパク質 A についてドッキングシミュレーションを行った結果得られる計算値 y と、測定された値 y' を用いて、 g と a の値を算出することができる。例えば、計算値 y と、測定された値 y' との二乗誤差を最小にするように、下記の式 4 を用いて、 g と a を算出することができる。測定された値 y' は、大きな値から小さな値を満遍なく選択して用いることが好ましい。例えば、タンパク質 B に対する相対活性で示すと、100%~1%未満まで幅広く用いることが好ましい。測定された値 y' が偏っている場合、例えば相対活性では 100%~50% の値である場合は、 $a=0$ として酵素活性を算出することもできる。なお、 g と a は、喫煙等の生体外環境により影響をうけるであろう生体内環境を加味した値を設定することも可能である。

[数26]

$$(g, a) = \underset{g, a}{\operatorname{argmin}} \left\{ (y_w - y'_w)^2 + (y_{A1} - y'_{A1})^2 + \cdots + (y_{Ap} - y'_{Ap})^2 \right\} \quad \cdots \text{式 4}$$

式中、 y_w と y'_w はタンパク質 B についての値であり、 y_{A1} と y'_{A1} 、 y_{Ap} と y'_{Ap} はタンパク質 A についての値であり、 p は 2 以上の数を表す。

[0053] 上記式 4 における測定された値 y' は、ドッキングシミュレーションに基づく計算値ではなく、何らかの実測値を意味する。測定された値 y' は例え

ば、組換えタンパク質を用いたアッセイで得られた実験値、論文などの文献報告や臨床検査結果などから予測される値を含む。例えばUGT1A1のビリルビンに対する抱合活性は、ビリルビンの血中濃度から推定することができる。推定された値を測定された値 y' として使用することができる。

[0054] 変異型UGT1A1の、AAP、E2もしくはビリルビンに対する酵素活性の予測を例示して説明する。

まず、AAP、E2、ビリルビンについて、各々 g と a を算出する。G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1の各々について、組換えタンパク質の酵素活性を測定することにより、測定された値 y' を得ることができる。組換えタンパク質の作製方法、および酵素活性の測定方法は自体公知の方法により行うことができる。例えば、実施例に記載の方法により行えばよい。また、ビリルビンに対する測定された値 y' として、文献報告のビリルビン抱合活性（例えば、Yamamoto K et al.: Biochem Biophys Acta 1998, 1406:267-273, Udomuksorn W et al.: Pharmacogenetics & Genomics 2007, 17:1017-1029, Ciotti M et al.: Biochimica et Biophysica Acta 1998, 1407:40-50) を使用することもできる。

[0055] 下記の式5を用いて、AAP、E2、ビリルビンについて、測定された値 y' を代入して、二乗誤差を最小にする g と a の値を算出する。

[数27]

$$(g, a) = \underset{g, a}{\operatorname{argmin}} \left\{ (y_w - y'_w)^2 + (y_{G71R} - y'_{G71R})^2 + (y_{F83L} - y'_{F83L})^2 + (y_{I322V} - y'_{I322V})^2 \right\} \cdots \text{式5}$$

その結果、AAPについて、 $g = 1.1055$ 、 $a = 0.0722$ 、E2について、 $g = 5.9410$ 、 $a = 1.2548$ 、ビリルビンについて、 $g = 47.58$ 、 $a = 0.04$ の値を算出することができる。

[0056] 上記算出された g と a を用いて、変異型UGT1A1の酵素活性を予測することができる。式3に g 、 a を代入し、ドッキングシミュレーションにより得られる N と n を利用して、変異型UGT1A1の算出することができる。

。34種類の変異型UGT1A1の、AAPおよびE2についての酵素活性を算出した結果を図7に示す。

[0057] さらに、式1および式3における

[数28]

$$\frac{n}{N} \quad \text{および/または} \quad \frac{n_w}{N_w}$$

について、Sigmoid関数を導入することも可能である。Sigmoid関数とは、下記式19で表されるS字型の関数である。

[数29]

$$\zeta_p(x) = \frac{1}{1 + e^{-px}} \quad \dots \text{式19}$$

[0058] 本発明の式1において、Sigmoid関数を導入した計算式は、下記式20にて表すことができる。

[数30]

$$f = g \times E \times \left(\frac{1}{1 + e^{-t \left(\frac{n}{N} - d \right)}} \right)^\beta + a \quad \dots \text{式20}$$

ここでtは、基質の向きに対する酵素活性の感受性を表すものであり、tの値が大きい程基質の向きの僅かな差異が酵素活性を大きく変化させることとなる。tの値は、2以上のタンパク質Aについてドッキングシミュレーションを行った結果得られる計算値yと、測定された値y'を用いて算出することができ、以下のdの算出方法と同様にして求めることができる。dは、シミュレーションの結果を上記の式に代入して得られる計算値yと、in vitro 解析等による測定値y'との二乗誤差を最小化する値として、下記の式22を用

いて算出することができる。

[数31]

$$d = \underset{d}{\operatorname{arg\,min}} \left\{ (y_w - y'_w)^2 + (y_{A1} - y'_{A1})^2 + \dots + (y_{Ap} - y'_{Ap})^2 \right\} \quad \dots \text{式22}$$

[0059] 本発明は、タンパク質Aの酵素活性を予測する方法を実行するプログラムを担持する記録媒体、およびタンパク質Aの酵素活性を予測する方法を実行する手段を担持する装置にも及ぶ。記録媒体に担持されるプログラムは、コンピュータを以下の手段として機能させるものであり、装置は、以下の手段を含むものである。

(1) 入力されたアミノ酸配列情報に基づいて、タンパク質Aの立体構造データを計算する手段、

(2) タンパク質Aの立体構造データと、ドッキングシミュレーションの対象となる基質の立体構造データを記憶する手段、

(3) 前記記憶された、タンパク質Aの立体構造データと基質の立体構造データを用いて、タンパク質Aと基質とのドッキングシミュレーションを行うシミュレーション手段、

(4) シミュレーションにより得られた結果を記憶する手段、

(5) 記憶されたシミュレーション結果に基づいて、タンパク質Aの酵素活性を算出する手段、

(6) 算出された酵素活性を表示する手段。

[0060] また、本発明は、タンパク質Aの酵素活性を予測する方法、前記記録媒体、または前記装置を用いた、基質適合性の判断方法にも及ぶ。本発明における式1または3を用いて、2以上の基質について、基質ごとの特定のタンパク質Aの酵素活性を予測し、得られた2以上の予測結果に基づいて目的の基質を選択する。例えば、タンパク質Aが特定の変異型UGT1A1であり、基質が薬剤である場合に、本方法により、当該変異型UGT1A1による抱合活性が最も高い薬剤を検出し、選択することができる。本方法により生体内での薬剤の効能と代謝とのバランスを鑑み、薬剤を選択することも可能と

なる。

- [0061] さらに、本発明はタンパク質Aの酵素活性を予測する方法、前記記録媒体、または前記装置を用いて、当該薬剤の投与量および／または投与間隔、投与頻度を評価する方法にも及ぶ。基質が特定の薬剤であり、当該薬剤が生体に投与される場合であって、生体が例えばUGT1A1に変異を持つ場合に、かかる変異型UGT1A1の当該薬剤に対する酵素活性を予測し、予測結果に基づいて、当該薬剤の投与量および／または投与間隔、投与頻度を評価することが可能である。例えば、薬剤に対する変異型UGT1A1の抱合活性が低い場合は、投与量、投与頻度を低くすることを検討することが可能である。

実施例

- [0062] 以下、実施例により本発明を説明するが、本発明はこれらに限定されるものではない。

- [0063] (実施例1) UGT1A1の立体構造データの計算

野生型UGT1A1の立体構造データを、MODBASEからダウンロードした (accession number: Q5DT03)。PyMOLプログラムを用いて水素原子を追加し、SWISS-PDB Viewerプログラムを用いて、G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1の各変異型のデータを作成した。これらのデータを用いて、TINKERパッケージのminimizeプログラムとAMBER99力場パラメータによるエネルギー最小化計算をRMS勾配が0.3になるまで行い、各変異型の立体構造を求めた。計算された各変異型の三次元構造を図1に示す。

- [0064] (実施例2) 各種UGT1A1とUDPGAとのドッキングシミュレーション

AutoDock 4プログラムを用いて、各種UGT1A1 (野生型UGT1A1、G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1) とUDPGAのドッキングシミュレーションを行った。UDPGAの立体構造データは、ChemIDPlusに登録されている5個のデー

タを用いた (registry number: 2616-64-0)。UDPGAの配置を探索するマップを、グリッド間隔0.375 Å、60×60×60ポイントの立方体としてAutoGridプログラムを用いて生成した。グリッド探索アルゴリズムは、ラマルク型遺伝アルゴリズムを使用した。その他のパラメータはAutoDock 4のデフォルト値を使用した。

各種UGT1A1と5個のUDPGAの組み合わせ毎に、10回ずつドッキングシミュレーションを実行した。計算結果は、1種の変異型UGT1A1について、50個得られた。各々の変異型UGT1A1について、得られた50個の計算結果のうち、抱合反応可能な向きUDPGAのドッキングエネルギーの平均値を計算した。なお、ドッキングエネルギー ΔG は下記の式6で計算した。

[数32]

$$\Delta G = \Delta G_{\text{inter}} + \Delta G_{\text{intra}} + \Delta G_{\text{torsion}} - \Delta G_{\text{unbound}} \quad \dots \text{式 6}$$

[0065] 各変異型UGT1A1とUDPGAとのドッキングシミュレーションにて得られた50個の結果を、分子間エネルギー、UDPGAの分子内部エネルギー、およびアンバウンドエネルギーのそれぞれについて、低い順に順位づけを行った。分子間エネルギー、UDPGAの分子内部エネルギー、およびアンバウンドエネルギーの3種類の順位の順位和が最小のドッキング結果を最も安定なドッキングモデルとして選択し、それぞれの変異型UGT1A1とUDPGAのドッキングモデルに決定した。この時、ドッキングモデルに含まれる水素結合は、PyMOLを用いて検出した。

[0066] 図2にドッキングモデルの構造を示す。UDPGAのウラシル環は、野生型では357番目のグルタミン(Q)と、G71R変異型UGT1A1では42番目のセリン(S)、173番目のヒスチジン(H)、375番目のセリン(S)、396番目のアスパラギン酸(D)と、F83L変異型UGT1A1では374番目のグリシン(G)との相互作用が見られた。I322V変異型UGT1A1ではウラシル環と相互作用するアミノ酸は見られな

った。UDPGAのグルクロン酸部分は、野生型UGT1A1では396番目のアスパラギン酸（D）と、G71R変異型UGT1A1では310番目のメチオニン（M）、312番目のセリン（S）、393番目のロイシン（L）との相互作用が見られた。F83L変異型UGT1A1においては、グルクロン酸部分と相互作用するアミノ酸は見られなかった。I322V変異型UGT1A1においては、153番目のフェニルアラニン（F）とグルクロン酸部分との相互作用が見られた。なお、UDPGA反応の場の疎水性度には大きな違いはなかった。

[0067] 抱合反応可能な向きUDPGAと各UGT1A1のドッキングエネルギーの分布を図3に示す。野生型とG71R変異型とのドッキングエネルギー、および、F83L変異型とI322V変異型とのドッキングエネルギーに有意差は無かった。F83L変異型は、野生型およびG71R変異型よりもドッキングエネルギーが有意に高かった。また、I322V変異型は野生型およびG71R変異型よりもドッキングエネルギーが有意に高かった。

[0068] （実施例3）UGT1A1とUDPGAとの複合体における基質のドッキングシミュレーション

UGT1A1と基質とのドッキングシミュレーション条件のうち、グリッドを決定するために、まずドッキングモデルのうちからドッキングエネルギーが低い一群のシミュレーション結果を、群平均法（2つのクラスタを順次統合していく階層型クラスタリングの一種）によるクラスタリングで、グループ化して定義した。群平均法では、全てのクラスタ間の距離を計算し、最も距離の小さい2つのクラスタを統合する。クラスタXとYの間の距離 $d(X, Y)$ は以下の式7のように定義された。

[数33]

$$d(X, Y) = \frac{1}{|X||Y|} \sum_{x \in X} \sum_{y \in Y} |x - y| \quad \dots \text{式7}$$

[0069] UGT1A1と抱合反応可能な向きUDPGAの分子間エネルギー、分

子内部エネルギー、アンバウンドエネルギーの順位和を指標として群平均法でクラスタリングを行い、「相対的に低いドッキングエネルギーを有する数種類のドッキング構造」の集合を求めた。

ここで、最も安定な集合Cの至適な粒度を決定することで、グループ化が可能である。クラスタXの粒度G(X)を、抱合反応可能なUDPGA全体の集合をSとした次の式8により計算した。

[数34]

$$G(X) = \frac{\max_{x,y \in X} |x - y|}{\max_{z,w \in S} |z - w|} \quad \dots \text{式8}$$

[0070] 階層型クラスタリングでは、様々な粒度のクラスタリング結果が得られる。

粒度の決定のために、野生型、G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1と、UDPGAのドッキングシミュレーション結果を使って、群平均法によるクラスタリングを行い、全てのクラスタの粒度を計算してその分布を解析した。5%水準で、平均と比して有意に大きい粒度を除外するものとしたので、棄却域の値は0.56であった。つまり、粒度0.56以下のクラスタのうち最大のものを、グループとして定義可能であった。

[0071] 最も安定な集合Cは、抱合反応可能なドッキング結果から群平均法で得られたクラスタをgac(S)と定義して、次の式9のように計算した。

[数35]

$$C = \operatorname{argmax}_{X \in T} (G(X)) \quad \dots \text{式9}$$

式中、 $T = \{x \mid x \in \text{gac}(S), \min(S) \in x, G(x) < 0.56\}$

式9により定義されたクラスタの位置から分子の揺らぎを加味した抱合反応の場を決定した。そして、各変異型UGT1A1とAAPもしくはE2とのドッキングシミュレーションにおいて、各基質の大きさに合わせたグリッドの最奥の位置を決定した。かかるグリッド位置を用いて、UGT1A1のグルクロン酸抱合反応の場への各基質の進入方向を解析した。AAPとE2の分子構造データはChemIDPlusに登録されているものを用い、ドッキングシミュレーションを基質ごとに100回実行した。

[0072] 図4に基質の2つの向きを示す。向きIは基質のヒドロキシル基がUDPGAに向いており、グルクロン酸抱合反応の進行が可能である。向きIIは基質のヒドロキシル基がUDPGAの反対側に向いており、グルクロン酸抱合反応の進行が困難である。

表1に向きIおよび向きIIの各々でドッキングした回数を示す。

[表1]

	野生型		変異型 G71R		変異型 F83L		変異型 I322V	
	向き I	向き II	向き I	向き II	向き I	向き II	向き I	向き II
AAP	87	13	81	19	6	94	89	11
E2	98	2	4	96	10	90	95	5
N=100								

AAPでは、野生型、G71R変異型、I322V変異型では大部分のドッキング結果が向きIであった。F83L変異型では大部分のドッキング結果が、向きIIであった。

E2では、野生型、I322V変異型では大部分のドッキング結果において、向きIであった。G71R変異型、F83L変異型では、大部分のドッキング結果が、向きIIであった。

[0073] (実施例4) UGT1A1のin vitroでの酵素活性測定

ヒト肝臓cDNAライブラリーから、PCR増幅によりヒトUGT1A1 cDNAを抽出し、pENTER™/D-TOPOベクター (Invitrogen, Carlsbad, CA, USA) に挿入した。Site-directed mutagenesis法を用いて、遺伝子変異をクローンcDNAに導入した。正常型および各変異型のcDNA配列を、組み換

えを用いて発現ベクター—pcDNA-DEST40 Gateway™ (Invitrogen, Carlsbad, CA, USA) に移植した。各種UGT1A1の発現ベクターをルシフェラーゼレポーターベクター (pGL3-vector) とともにLipofectamine™2000を用いてCOS-7細胞に導入した。48時間後に細胞を採取し、0.1M Tri-HCl 70μl で均質化し、ルシフェラーゼとUGT1A1のアッセイの酵素ソースとして使用した。TD-20/20 luminometer (Promega, Madison, WI, USA) を用いてルシフェラーゼ活性を計測し、UGT1A1の酵素活性の標準化に供した。E2のグルクロン酸抱合反応を、UGT Reaction Mix (BD Gentest, Franklin Lakes, NJ, USA) を使用して分析した。反応生成物を遠心した後、LC/MS/MS解析に供し、E2とグルクロン酸の抱合体 (E2G) の量を測定した。AAPについても同様の手法を用いて抱合体 (AAPG) の測定を行った。

Analyst 1.3.1ソフトウェアでAPI-3000™LC/MS/MSシステム (Applied Biosystems-MDS Sciex, Tronto, Canada) を操作し、データ取得と解析を行った。

[0074] AAPおよびE2を基質とした時の各変異型UGT1A1のin vitro抱合活性を図5に示す。F83L変異型のAAPに対する酵素活性は、野生型と比較して有意に低下していた (n=5, p<0.005)。G71R変異型およびF83L変異型のE2に対する酵素活性は、野生型と比較して有意に低下していた (n=5, p<0.005)。F83L変異型のE2に対する酵素活性は、G71R変異型のE2に対する酵素活性よりも有意に低かった (n=5, p<0.05)。

[0075] (実施例5) UGT1A1酵素活性を予測するための数式の作成

UGT1A1の酵素活性は、(i) UGT1A1とUDPGAのドッキング、と(ii) UGT1A1の抱合反応空間への基質の進入、の積で規定され、プロセス(i)をE、プロセス(ii)をRと定義することで、抱合活性fは次式10で表される。

[数36]

$$f = g \times E \times R + a \quad \dots \text{式10}$$

g は基質に固有の定数、 a は内因性 UGT 1 A 1 等の生体内環境による影響を表す定数である。ここでは、実施例 2 と実施例 4 の結果から、UDP G A と UGT 1 A 1 とのドッキングエネルギーの抱合反応活性への関与は認められないと考えられたため、 $E = 1$ を代入した。

[0076] プロセス (i i) では、基質の進入方向の関与を反映させる。変異型 UGT 1 A 1 と基質とのドッキングシミュレーションを N 回行い、そのうち基質が抱合反応可能な向きである結果が n 回であった場合、プロセス (i i) の R は次式 1 1 で定義される。

[数37]

$$R = \left(\frac{n}{N} \right)^{\beta} \quad \dots \text{式 11}$$

β は、以下の式 2 で表される。式中、 N_w は野生型 UGT 1 A 1 と基質とのドッキングシミュレーションの総回数であり、 n_w は全 N_w 回中、基質が野生型 UGT 1 A 1 に抱合反応可能な方向で進入した回数である。

[数38]

$$\beta = \frac{N_w}{n_w} \quad \dots \text{式 2}$$

E と R を式 1 0 に代入すると、酵素活性の計算式は下記式 1 2 で表される。

[数39]

$$f = g \times \left(\frac{n}{N} \right)^{\beta} + a \quad \dots \text{式 12}$$

[0077] ここで定数 g 、 a は、計算式と in vitro 実験結果の二乗誤差が最小となるように基質ごとに異なる値を設定する。

[0078] また、式 12 を用いて、変異型 UGT1A1 の、野生型の酵素活性に対する相対的な酵素活性 f' は次の式 13 のように計算できる。

[数40]

$$f' = \frac{g \times \left(\frac{n}{N} \right)^\beta + a}{g \times \left(\frac{n_w}{N_w} \right)^\beta + a} \times 100 \quad \dots \text{式 13}$$

[0079] (実施例 6) AAP と、E2 に対する酵素活性の予測

上記実施例 5 の式により算出された計算値 y と、野生型および G71R 変異型、F83L 変異型、I322V 変異型の UGT1A1 の各基質に対する抱合活性の実験値 y' (実施例 4) を用いて、二乗誤差を最小化する g , a を次式 5 で計算することができる。

[数41]

$$(g, a) = \underset{g, a}{\operatorname{argmin}} \left\{ (y_w - y'_w)^2 + (y_{G71R} - y'_{G71R})^2 + (y_{F83L} - y'_{F83L})^2 + (y_{I322V} - y'_{I322V})^2 \right\} \quad \dots \text{式 5}$$

式 5 に AAP および E2 の in vitro 実験結果を代入し、 g , a の値を求めた。その結果、AAP について、 $g = 1.1055$ 、 $a = 0.0722$ 、E2 について、 $g = 5.9410$ 、 $a = 1.2548$ の値が算出された。

[0080] まず G71R 変異型、F83L 変異型、I322V 変異型の in vitro での抱合活性の測定結果 (in vitro 測定値) と、式 3 を用いた抱合活性の算出結果 (計算値) を比較した。結果を図 6 に示す。式 3 を用いて算出した抱合活性 (計算値) は in vitro の抱合活性 (in vitro 測定値) を良く再現可能であることがわかった。

[0081] 式 3 を用いて、現在報告されている他の 31 種類の変異型 UGT1A1 について、ドッキングシミュレーションを行い、抱合活性を式 3 を用いて算出した。全 34 種類の変異型 UGT1A1 に関する算出結果を、図 7 に示した

[0082] (実施例7) ビリルビンに対する酵素活性の予測

ビリルビンを基質とした時の抱合活性を式3を用いて算出した。

まず、ビリルビンとUGT1A1のドッキングシミュレーションを行った。基質(ビリルビン)の分子が大きいため、基質進入方向のグリッドとして、グリッドA~Cの3つを設定した。3つのグリッドを図8に示す。野生型、およびG71R変異型、F83L変異型、I294T変異型とビリルビンとのドッキングシミュレーションを行い、基質の進入の向きを解析した。グリッドA~Cのそれぞれについて100回ずつドッキングシミュレーションを行った。各グリッドについてNとnは、各々 N_A 、 N_B 、 N_C 、 n_A 、 n_B 、 n_C として計数した。Nを N_A と N_B と N_C の和(すなわち300回)、nを n_A と n_B と n_C の和として算出した。その結果を表2に示す。

[表2]

	野生型	G71R	F83L	I294T
向きI (n)	91	46	34	59

N=300

[0083] 上記結果と、野生型および文献で報告されているG71R変異型、F83L変異型、I294T変異型のin vitro酵素活性の値 y' ($y'_w=1.00$, $y'_{G71R}=0.32$, $y'_{F83L}=0.05$, $y'_{I294T}=0.50$)を用いて二乗誤差を最小化するg, aを計算したところ、 $g=44.06$ 、 $a=0.17$ であった。in vitro酵素活性は、G71R変異型については、Yamamoto K et al.: Biochem Biophys Acta 1998, 1406:267-273を、F83L変異型についてはUdomuksorn W et al.: Pharmacogenetics & Genomics 2007, 17:1017-1029を、I294T変異型については、Ciotti M et al: Biochimica et Biophysica Acta 1998, 1407:40-50を参照した。これらの文献では、酵素活性は正常値(野生型)に対する割合(相対活性)で示されており、G71R変異型は32%、F83L変異型は5%、I294T変異型は50%である。

これらの結果を用いて式3により、各変異型UGT1A1のビリルビンに

対する相対抱合活性を算出した。

[0084] G 7 1 R 変異型、F 8 3 L 変異型、I 2 9 4 T 変異型の in vitro での抱合活性値 (in vitro 測定値 (文献報告)) と、式 3 を用いた抱合活性の算出結果 (計算値) を比較した。結果を図 9 に示す。式 3 を用いて算出した抱合活性 (計算値) は in vitro の抱合活性 (in vitro 測定値 (文献報告)) を良く再現可能であることがわかった。

[0085] (実施例 8) ビリルビンに対する酵素活性の予測 2

実施例 7 にて得られた g と a の値を用いて、R 3 3 6 L 変異型、N 4 0 0 D 変異型、W 4 6 1 R 変異型についてビリルビンに対する活性を算出した。

まず、各変異型 UGT 1 A 1 とビリルビンとのドッキングシミュレーションを行った。ドッキングシミュレーションの結果を表 3 に示す。

[表3]

	野生型	R336L	N400D	W461R	
向き I (n)	91	88	58	42	N=300

これらの結果と実施例 7 にて得られた g と a の値を用いて、抱合活性を式 3 を用いて算出した。

[0086] 次に、文献に記載のデータから R 3 3 6 L 変異型、N 4 0 0 D 変異型、W 4 6 1 R 変異型のビリルビン抱合活性の値の範囲を算出した。

C r i g l e r - N a j j a r 症候群 I 型 (C N - I) 患者に見られる変異型 UGT 1 A 1 のビリルビン抱合活性は、野生型の 0 ~ 1 0 % と算出される (Yamamoto K et al.: Biochem Biophys Acta 1998, 1406:267-273)。ホモ型の W 4 6 1 R 変異型 (T A 6 / T A 6) が、C N - I 患者にて発見されている (Maruo Y, et al.: J Pediatr Gastroenterol Nutr 2003, 37(5):627-30)。よって、ホモ型の W 4 6 1 R 変異型の酵素活性は、野生型の 0 ~ 1 0 % と計算された。

[0087] C r i g l e r - N a j j a r 症候群 I I 型 (C N - I I) および G i l b e r t 症候群 (G S) 患者に見られる変異型 UGT 1 A 1 のビリルビン抱

合活性は、野生型の26～66%と算出される (Udomuksorn W et al.: Pharmacogenetics & Genomics 2007, 17:1017-1029; Yamamoto K et al.: Biochem Biophys Acta 1998, 1406:267-273; Seppen J, et al.: J Clin Invest 1994, 94(6):2385-2391)。ホモ型のN400D変異型が、CN-I I患者に見られる (Labrune P et al.: Hum Mutat 2002, 20(5):399-401)。よって、ホモ型のN400D変異型の酵素活性は、野生型の26～66%と計算された。

[0088] ヘテロ型のR336L変異型は、CN-I I患者に見られる (Servedio V et al.: Hum Mutat 2005, 25(3):325)。Servedio V et al.に記載のCN-I I患者ではプロモータ領域にTA7/TA7の変異も確認されている。TA7/TA7変異のある患者では、ビリルビン抱合活性が野生型と比べて50%に低下することが報告されている (Peterson et al.: J Nutr 2005, 135:1051-1055)。これらの報告から、R336L型変異の染色体あたりの抱合活性低下をx (%)とすると、下記の式14が成り立つ:

[数42]

$$26 < 100 - x - 50 < 66 \quad \cdots \text{式14}$$

式14から以下の式15が得られた。

[数43]

$$-16 < x < 24 \quad \cdots \text{式15}$$

したがって、ホモ型のR336L変異型の抱合活性は野生型の52～132% (平均値92%)と計算された。

[0089] 表3に記載のドッキングシミュレーションの結果から算出した値 (計算値) と、文献から得られた値 (in vivo測定値 (文献報告)) の比較を図10に示す。この結果、高い精度で正しい抱合活性を導出可能なことが分かった。UGT1A1は肝臓で働く酵素であり、ヒト生体内でのUGT1A1抱合活性は飲酒や喫煙の影響を受ける可能性が高いことを考慮すると、本発明に用

いられる式により得られる値は非常に精度が高いと考えられた。

[0090] (実施例 9) ビリルビンに対する酵素活性の予測 3

in vitro酵素活性の値として、G 7 1 R 変異型、F 8 3 L 変異型、I 2 9 4 T 変異型ではなく、G 7 1 R 変異型、P 2 2 9 Q 変異型、I 2 9 4 T 変異型のものを用いた以外は、実施例 7 と同様にして、g と a を算出した。P 2 2 9 Q 変異型の値は、Udomuksorn W et al: Pharmacogenetics and genomics 2007, 17(12):1017-29を参照した。この文献では、P 2 2 9 Q 変異型の正常値(野生型)に対する割合(相対活性)は、6 1 %と示されている。各変異型UGT 1 A 1 とビリルビンとのドッキングシミュレーションを行った結果を、表 4 に示す。

[表4]

	野生型	G71R	P229Q	I294T	
向き I (n)	91	46	37	59	N=300

上記結果と、野生型および文献で報告されているG 7 1 R 変異型、P 2 2 9 Q 変異型、I 2 9 4 T 変異型のin vitro酵素活性の値 y' ($y'_w=1.00$, $y'_{G71R}=0.32$, $y'_{P229Q}=0.61$, $y'_{I294T}=0.50$) を用いて二乗誤差を最小化する g, a を計算したところ、 $g = 29.36$ 、 $a = 0.40$ であった。

これらの結果を用いて式 3 により、各変異型UGT 1 A 1 のビリルビンに対する相対抱合活性を算出した。

[0091] G 7 1 R 変異型、P 2 2 9 Q 変異型、I 2 9 4 T 変異型のin vitroでの抱合活性値(in vitro測定値(文献報告))と、式 3 を用いた抱合活性の算出結果(計算値)を比較した。結果を図 1 1 に示す。

[0092] (実施例 1 0) ビリルビンに対する酵素活性の予測 4

R 3 3 6 L 変異型、N 4 0 0 D 変異型、W 4 6 1 R 変異型について、実施例 8 にて得られたドッキングシミュレーションの結果と、実施例 9 にて得られた g と a の値を用いて、抱合活性を式 3 を用いて算出した。

[0093] ドッキングシミュレーションの結果から算出した値(計算値)と、実施例

8に記載の文献から得られた値（in vivo測定値（文献報告））の比較を図12に示す。

[0094]（実施例11）ビリルビンに対する酵素活性の予測5

R336L変異型、N400D変異型、W461R変異型について、 $g = 29.36$ 、 $a = 0$ を代入して、抱合活性（相対活性）を式3を用いて算出した。

[0095] 算出した結果（計算値）と、実施例8に記載の文献から得られた値（in vivo測定値（文献報告））の比較を、図13に示す。

[0096]（実施例12）構造計算時の水分子付加の影響

水分子を付加して、又は水分子を付加せず（Gas Phase）に、UGT1A1（細胞膜タンパク質）およびGタンパクであるAr16（細胞質に存在）のタンパク質2種類について、シミュレーションを行った。

水分子を付加する場合は、TINKERパッケージの代わりにMOEを用いた以外は実施例1と同様にして立体構造を計算した（野生型UGT1A1、G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1：野生型Ar16、T31R変異型Ar16、G169A変異型Ar16、L170W変異型Ar16）。また、水分子を付加しない場合は、TINKERパッケージを用いて、実施例1と同様にして立体構造を計算した。

[0097] その後UGT1A1について、水分子を付加した場合はMOE DockもしくはAutoDock 4プログラムを用いて、水分子を付加していない場合はAutoDock 4プログラムを用いて、実施例2と同様にして、UDPGAとのドッキングシミュレーションを行った。抱合反応可能な向き of UDPGAと各UGT1A1のドッキングエネルギーの分布を図14に示す。

[0098] 次に、水分子を付加した各変異型UGT1A1と基質（AAPもしくはE2）とのドッキングシミュレーションまたは、水分子を付加した各変異型Ar16と基質（GTPγSもしくはGDP）とのドッキングシミュレーションを行った。基質とのドッキングシミュレーションは、MOE DockもしくはAutoDock4を用いて、実施例3と同様にして行った。ドッキングシミュレーショ

ンの結果を図15 (UGT1A1) と図17 (Ar16) に示す。図15の a は、ドッキングプログラムにMOE Dockを使用し、b はAutoDock4を用いた結果である。図17の結果は、ドッキングプログラムにAutoDock4を用いた結果である。

[0099] 各種UGT1A1について、基質が向きIで進入した回数とin vitro測定値との比較を図16に示す。図16の a は、水分子付加した場合に基質とのドッキングシミュレーションをMOE Dockを用いて行った結果であり、b は水分子を付加した場合に基質とのドッキングシミュレーションをAutoDock4を用いて行った結果である。

[0100] UGT1A1では、水分子を付加して構造計算を行った場合、基質と各種UGT1A1のドッキングシミュレーションの結果（基質が向きIで各種UGT1A1に進入した回数）がin vitroの抱合活性（in vitro測定値）に相関しないことがわかった。また、ドッキングシミュレーションにMOE DockおよびAutoDock4のいずれを使用した場合でも同様に相関がみられず、基質がAAPおよびE2のいずれの場合でも相関がみられなかった。

また、Ar16では水分子を付加して構造計算を行った結果、基質（GTP γ SまたはGDP）と各種Ar16のドッキングシミュレーションの結果（GTP γ SまたはGDPが各種Ar16に結合し得る向きで進入した回数）と、in vitroのAr16実験値（Kobayashi et. al, BBRC 381, 439-442, 2009および東京大学大学院薬学研究科機能薬学教室生理化学教室の紺谷圏二先生の実験データ）とに高い相関が見られることを確認した。なおAr16は酵素タンパク質ではないが、酵素タンパク質について酵素反応が進むためには基質との結合が必須であることから、Ar16の基質との結合能は酵素タンパク質の触媒能と同義に考えることができる。

[0101] UGT1A1は、細胞膜に存在するタンパク質であり、触媒反応が酵素内部で起こると考えられる。一方、Ar16は細胞質に存在するタンパク質であり、GTPとの結合がタンパク質表面で起こると考えられる。これらの結果から、対象となる酵素タンパク質の存在する箇所や、触媒反応の起こる部

位などの疎水性度および親水性度を考慮して、構造計算時における水分子付加の可否を決定可能であることが示唆された。

[0102] (実施例 13) induced fitの影響

(1) 水分子を付加せずに各種 UGT 1A1 の立体構造データの計算を行い、UDPGA とのドッキングシミュレーションを行った。構造計算に MOE 又は TINKER を使用し、ドッキングシミュレーションに MOE Dock 又は AutoDock4 を使用して、実施例 1 および 2 と同様の手法でシミュレーションを行った。UDPGA とのドッキングモデル 1 つに対して、MOE を用いて induced fit を行い、基質 (AAP または E2) とのドッキングシミュレーションを行った。

[0103] 結果を図 18 および図 19 に示す。図 18 の a は、構造計算に MOE を用いてドッキングシミュレーションに MOE Dock を用いた結果であり、図 18 の b は、構造計算に TINKER を用いてドッキングシミュレーションに AutoDock4 を用いた結果である。図 19 の a は構造計算に MOE を用いてドッキングシミュレーションに MOE Dock を用いた結果と in vitro 測定値との比較であり、図 19 の b は、構造計算に TINKER を用いてドッキングシミュレーションに AutoDock4 を用いた結果と in vitro 測定値との比較である。

[0104] induced fit を行わなかった結果 (図 19 のグラフ、実線) と比較すると、induced fit を行った場合 (図 19 のグラフ、点線) では、基質と各種 UGT 1A1 のドッキングシミュレーションの結果 (基質が向き I で各種 UGT 1A1 に進入した回数) が in vitro の抱合活性 (in vitro 測定値) に相関しないことがわかった。本実施例では、UDPGA とのドッキングモデルを 1 種類 (最も安定なドッキングモデル) としている。induced fit のシミュレーションに対する影響をさらに検討するため、UDPGA とのドッキングモデルについてクラスタリングを行い (実施例 3 参照)、クラスタリングの結果得られたクラスタ (最も安定な集合 C) 内の全ドッキングモデルについて、induced fit を行い影響を検討した。

[0105] (2) 実施例 1 および 2 と同様にして、TINKER を使用して水分子を付加せずに各種 UGT 1A1 の立体構造データの計算を行い、AutoDock4 で UDPGA

とのドッキングシミュレーションを行った。次にUDPGAとのドッキングモデルについてクラスタリングを行い（実施例3参照）、クラスタリングの結果得られた最も安定な集合C内の全ドッキングモデルについて、MOEを用いてinduced fitを行い、AutoDock4を用いて基質（AAPまたはE2）とのドッキングシミュレーションを行った。これらの結果を用いて、式3（gとaの値は、実施例6のものを使用）により、各変異型UGT1A1のAAPもしくはE2に対する相対抱合活性を算出した。

[0106] 結果を図20に示す。最も安定な集合C内の全ドッキングモデルについてinduced fitを行った場合（図20（B））、induced fitなしの場合（図20（C））およびクラスタ内の1種類のモデルについてinduced fitを行った場合（図20（A））と比較して、式3を用いて算出した抱合活性（計算値）がin vitroの抱合活性（in vitro測定値）と高い相関を持つことがわかった。従って、UDPGAとのドッキングモデルのクラス内の複数のモデル、好ましくは全てのモデルについて、induced fitを行い、基質とのドッキングシミュレーションを行うことが好ましいことが、示唆された。

[0107] （実施例14）補酵素の各種UGT1A1への進入の向きの抱合能への影響
補酵素のドッキングエネルギーがUGT1A1の抱合能に影響をしないことを実施例2および4にて確認した。補酵素の各種UGT1A1への進入の向きが、抱合能へ影響するかについて検討を行うため、式1について3種類のin silico抱合能計算式を導出した。すなわち、E=1（補酵素の向きの寄与が皆無）、E=下記式16（補酵素が酵素反応を受け得る向きで各種UGT1A1へ進入した回数の寄与）、

[数44]

$$\left(\frac{I}{L} \right)^{\gamma} \quad \cdots \text{式16}$$

E=下記式17（補酵素とのシミュレーションを行った後、クラスタリングにより選択されたモデル数の寄与）

[数45]

$$\left(\frac{m}{m_w} \right)^\delta \quad \cdots \text{式 17}$$

である。L、l、 γ 、m、 m_w 、 δ については本明細書にて定義したとおりである。

[0108] 実施例13の手法により、UDPGAとのドッキングモデルのクラスタ内の全てのモデルについて、induced fitを行い、基質とのドッキングシミュレーションを行った。基質とのドッキングシミュレーションの結果を用いて、各計算式を用いてG71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1、R336L変異型UGT1A1、H376R変異型UGT1A1、P387S変異型UGT1A1の抱合活性（計算値）を算出した（gとaは実施例6のものを使用、 δ は0.37を用いた）。得られた計算値とin vitro抱合能（基質はAAPまたはE2）との二重誤差（式18）を算出し、各計算式の抱合能の予測精度を評価した。

[数46]

$$diff = (y_w - y'_w)^2 + (y_{A1} - y'_{A1})^2 + \cdots + (y_{Ap} - y'_{Ap})^2 \quad \cdots \text{式 18}$$

[0109] E=1（補酵素の寄与が皆無）の場合は、二重誤差が0.137062であり、E=式16（補酵素が酵素反応を受け得る向きで各種UGT1A1へ進入した回数の寄与）の場合は0.117838、E=式17（補酵素とのシミュレーションを行った後、クラスタリングにより選択されたモデル数の寄与）の場合は0.053017であった。補酵素とのドッキングシミュレーションを行った後、クラスタリングにより選択されたモデル数を計算式に使用した場合に、最も二重誤差が減少した。従って、補酵素のオリエンテーションが抱合能に関与しており、これを考慮することにより、さらに正確に抱合能を予測可能となることが示唆された。

[0110]（実施例15）基質のオリエンテーションについてSigmoid関数を用いた場合

の影響

Sigmoid関数とは下記式19で表されるS字型の関数であり、実数xに対して(0, 1)の値域を持つ単調増加関数である。pをゲインと呼び、関数の形状に影響する。

[数47]

$$\zeta_p(x) = \frac{1}{1 + e^{-px}} \quad \dots \text{式19}$$

抱合能の計算式である式1において、式10における項R (UGT1A1の抱合反応空間への基質の進入)にSigmoid関数を適用した計算式を作成した(式20)。

[数48]

$$f = g \times E \times \left(\frac{1}{1 + e^{-t \left(\frac{n}{N} - d \right)}} \right)^\beta + a \quad \dots \text{式20}$$

[0111] G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1、R336L変異型UGT1A1、H376R変異型UGT1A1、P387S変異型UGT1A1について、実施例13の手法により、UDPGAとのドッキングモデルのクラス内の全てのモデルについて、induced fitを行い、基質とのドッキングシミュレーションを行った。基質とのドッキングシミュレーションの結果を用いて、実施例14の3種類のEのうち式17を上記式20と組み合わせた計算式により、各種UGT1A1の抱合活性(計算値)を算出した(gとaは実施例6のものを使用。d、tは実施例13の結果と、野生型およびG71R変異型、F83L変異型、I322V変異型のin vitro酵素活性の値を用いて、二重誤差を最小化する値を計算した。その結果、t=7.00、d=0.54であった。δは0.37を用いた)。なお、各種UGT1A1 (G71R変異型、F83L変異型、I322V変異型

、R336L変異型、H376R変異型、P387S変異型)のin vitro抱合能は、実施例4の手法と同様にして測定した。さらに、現在報告されている他の28種類の変異型UGT1A1について、本実施例と同様にしてドッキングシミュレーションを行い、その結果を用いて式17および式20を用いて、抱合活性を測定した。

[0112] 図21に、in vitro測定値とシミュレーションに基づく計算値との比較を示す。また図22に、野生型UGT1A1と、全34種の変異型UGT1A1のシミュレーションに基づく計算値の結果を示す。

計算式にSigmoid関数を使用した場合、in vitroの抱合活性を良く再現可能であることがわかった(図21および図22)。よって、式10における項R(UGT1A1の抱合反応空間への基質の進入)へSigmoid関数を適用することが好ましいことが示唆された。

[0113] (実施例16)AAPと、E2に対する酵素活性の予測

実施例15と同様の手法を用いて、ドッキングシミュレーションを行った。かかるドッキングシミュレーションの結果を用いて、実施例14の3種類のEのうちE=1を式3と組み合わせた計算式により、G71R変異型UGT1A1、F83L変異型UGT1A1、I322V変異型UGT1A1、R336L変異型UGT1A1、H376R変異型UGT1A1、P387S変異型UGT1A1の抱合活性(計算値)を算出した(gとaは実施例6のものを使用)。なお、各種UGT1A1(G71R変異型、F83L変異型、I322V変異型、R336L変異型、H376R変異型、P387S変異型)のin vitro抱合能は、実施例4の手法と同様にして測定した。

[0114] これらの結果を図23に示す。

産業上の利用可能性

[0115] 本発明の方法を用いれば、特定の基質に対する酵素活性が予測できない変異型タンパク質についての、酵素活性を予測することができ有用である。薬剤代謝に重要なUGT1A1の変異型を例に説明すると、天然に存在する変異部位の解析結果は、将来的に個人のゲノム解析結果が得られた場合、薬剤

投与のための情報として有益である。また、上記酵素活性 f または f' は、タンパク質の立体構造自体に由来する酵素活性であり、上記数式を使用して、その他の環境要因等を含めるような数式を作成し、リスクファクターを加味した患者ごとの薬剤投与計画の作成のための参考情報を得ることも可能である。さらに、特定のタンパク質について、人工変異部位を含めて、網羅的に変異部位を持つタンパク質について酵素活性の予測を行うことにより、酵素の触媒作用におけるタンパク質の重要部位の決定が可能となり、創薬ターゲットに利用可能である。

符号の説明

- [0116] 1 野生型
2 P 3 4 Q 変異型
3 H 3 9 D 変異型
4 G 7 1 R 変異型
5 F 8 3 L 変異型
6 L 1 7 5 Q 変異型
7 C 1 7 7 R 変異型
8 R 2 0 9 W 変異型
9 V 2 2 5 G 変異型
10 P 2 2 9 Q 変異型
11 G 2 7 6 R 変異型
12 E 2 9 1 V 変異型
13 A 2 9 2 V 変異型
14 I 2 9 4 T 変異型
15 G 3 0 8 E 変異型
16 I 3 2 2 V 変異型
17 Q 3 3 1 R 変異型
18 R 3 3 6 L 変異型
19 R 3 3 6 Q 変異型

- 20 R 3 3 6 W 変異型
- 21 W 3 5 4 R 変異型
- 22 Q 3 5 7 R 変異型
- 23 R 3 6 7 G 変異型
- 24 A 3 6 8 T 変異型
- 25 S 3 7 5 F 変異型
- 26 H 3 7 6 R 変異型
- 27 G 3 7 7 V 変異型
- 28 S 3 8 1 R 変異型
- 29 P 3 8 7 S 変異型
- 30 G 3 9 5 V 変異型
- 31 N 4 0 0 D 変異型
- 32 A 4 0 1 P 変異型
- 33 R 4 0 3 C 変異型
- 34 K 4 2 8 E 変異型
- 35 W 4 6 1 R 変異型

- (A) induced fit有り、1モデル
- (B) induced fit有り、クラスタ内の全モデル
- (C) induced fitなし
- (D) UDPGAが酵素反応を受け得る向きで各種UGT1A1に進入した回数
- (E) クラスタ内のモデル数
- (F) 基質が向きIで各種UGT1A1に進入した回数（向きIの回数／シミュレーションの全回数）

請求の範囲

[請求項1] タンパク質Aの酵素活性をコンピュータを用いたシミュレーションにより予測する方法であって、
 酵素活性がタンパク質Aへの基質の進入の向きにより規定されるようなタンパク質であり、
 タンパク質Aの酵素活性 f が以下の式1により算出される方法；

[数1]

$$f = g \times E \times \left(\frac{n}{N} \right)^\beta + a \quad \dots \text{式1}$$

式中、 g は基質ごとに固有の定数であり、 E はタンパク質Aと補酵素とのドッキングの酵素活性に対する寄与度であり、 β はタンパク質への基質の進入の向きの酵素活性に対する寄与度であり、 a は生体内環境による影響を表す定数であり、 N と n は、タンパク質Aと基質とのドッキングシミュレーションにより得られる値であり、 N はタンパク質Aと基質とのドッキングシミュレーションの総回数であり、 N は2以上であり、 n は酵素反応を受け得る向きで基質がタンパク質Aに進入した回数である。

[請求項2] タンパク質Aの立体構造データが、タンパク質Aとは別のタンパク質Bの立体構造データに基づいて計算されるものであり、 β が以下の式2によって算出される請求の範囲第1項に記載の方法；

[数2]

$$\beta = \frac{N_w}{n_w} \quad \dots \text{式2}$$

式中、 N_w と n_w は、タンパク質Bと基質とのドッキングシミュレーションにより得られる値であり、 N_w はタンパク質Bと基質とのドッ

キングシミュレーションの総回数であり、 N_w は2以上であり、 n_w は酵素反応を受け得る向きで基質がタンパク質Bに進入した回数である。

[請求項3]

タンパク質Bの酵素活性に対するタンパク質Aの相対的な酵素活性をコンピュータを用いたシミュレーションにより予測する方法であって、

酵素活性がタンパク質Aへの基質の進入の向きにより規定されるようなタンパク質であり、

タンパク質Aの相対的な酵素活性 f' が以下の式3により算出される方法；

[数3]

$$f' = \frac{g \times E \times \left(\frac{n}{N}\right)^\beta + a}{g \times E \times \left(\frac{n_w}{N_w}\right)^\beta + a} \times 100 \quad \dots \text{式3}$$

式中、 g は基質ごとに固有の定数であり、 E はタンパク質Aと補酵素とのドッキングの酵素活性に対する寄与度であり、 a は生体内環境による影響を表す定数であり、 β はタンパク質への基質の進入の向きの酵素活性に対する寄与度であり、次の式2により表され；

[数4]

$$\beta = \frac{N_w}{n_w} \quad \dots \text{式2}$$

N と n は、タンパク質Aと基質とのドッキングシミュレーションにより得られる値であり、 N はタンパク質Aと基質とのドッキングシミュレーションの総回数であり、 N は2以上であり、 n は酵素反応を受け

得る向きで基質がタンパク質Aに進入した回数であり、 N_w と n_w はタンパク質Bと基質とのドッキングシミュレーションにより得られる値であり、 N_w はタンパク質Bと基質とのドッキングシミュレーションの総回数であり、 N_w は2以上であり、 n_w は酵素反応を受け得る向きで基質がタンパク質Bに進入した回数である。

[請求項4] Eが、下記の式16または式17により表される、請求の範囲第1項～第3項のいずれか1に記載の方法：

[数5]

$$\left(\frac{l}{L} \right)^\gamma \quad \dots \text{式16}$$

(式16中、 γ はタンパク質への補酵素の進入の向きの酵素活性に対する寄与度であり、次の式21により表され、

[数6]

$$\gamma = \frac{L_w}{l_w} \quad \dots \text{式21}$$

L と l は、タンパク質Aと補酵素とのドッキングシミュレーションにより得られる値であり、 L はタンパク質Aと補酵素とのドッキングシミュレーションの総回数であり、 L は2以上であり、 l は酵素反応が進行し得る向きで補酵素がタンパク質Aに進入した回数であり、 L_w と l_w はタンパク質Bと補酵素とのドッキングシミュレーションにより得られる値であり、 L_w はタンパク質Bと補酵素とのドッキングシミュレーションの総回数であり、 L_w は2以上であり、 l_w は酵素反応が進行し得る向きで補酵素がタンパク質Bに進入した回数である)

；

[数7]

$$\left(\frac{m}{m_w} \right)^\delta \quad \dots \text{式17}$$

(式17中、 m はタンパク質Aと補酵素とのドッキングモデルをクラスタリングして得られたクラスタ内のモデル数であり、 m_w はタンパク質Bと補酵素とのドッキングモデルをクラスタリングして得られたクラスタ内のモデル数であり、 δ はクラスタ内のモデル数の酵素活性への寄与度を表す。)

[請求項5]

式1または式3において、

[数8]

$$\frac{n}{N} \quad \text{および/または} \quad \frac{n_w}{N_w}$$

について、Sigmoid関数を導入する、請求の範囲第1項～第4項のいずれか1に記載の方法。

[請求項6]

2以上のタンパク質Aについてドッキングシミュレーションを行い、 g と a が、ドッキングシミュレーションにより得られた計算値 y と、測定された値 y' との二乗誤差を最小にする値であり、下記の式4を用いて算出される、請求の範囲第2項～第5項のいずれか1に記載の方法。

[数9]

$$(g, a) = \underset{g, a}{\operatorname{argmin}} \left\{ (y_w - y'_w)^2 + (y_{A1} - y'_{A1})^2 + \dots + (y_{Ap} - y'_{Ap})^2 \right\} \quad \dots \text{式4}$$

式中、 y_w と y'_w はタンパク質Bについての値であり、 y_{A1} と y'_{A1} 、 y_{Ap} と y'_{Ap} はタンパク質Aについての値であり、 p は2以上の数を表す。

[請求項7]

タンパク質と基質とのドッキングシミュレーションが以下の工程を含

む請求の範囲第1項～第6項のいずれか1に記載の方法：

(a) タンパク質Bの立体構造データを入手し、タンパク質Aの立体構造データをタンパク質Bの立体構造データに基づいて計算し、

(b) タンパク質AまたはBと補酵素とのドッキングシミュレーションを行い、熱力学的に安定なドッキングモデルを決定し；

(c) タンパク質AまたはBと基質とのドッキングのグリッドを設定し；

(d) タンパク質Bと基質とのドッキングシミュレーションを N_w 回を行い、 N_w は2以上であり、酵素反応を受け得る向きで基質がタンパク質Bに進入した回数 n_w を計数し、

(e) タンパク質Aと基質とのドッキングシミュレーションを N 回を行い、 N は2以上であり、酵素反応を受け得る向きで基質がタンパク質Aに進入した回数 n を計数する。

[請求項8] 工程(b)の後に次の工程(b-1)を行う；

工程(b-1) タンパク質AまたはBと補酵素とのドッキングモデルをクラスタリングし、クラスタリングして得られたクラスタ内の2以上のモデルについてinduced fitを行い、

工程(d)において、induced fit後の各モデルについて、基質とのドッキングシミュレーションを行う、

請求の範囲第1項～第7項のいずれか1に記載の方法。

[請求項9] タンパク質Aが変異型タンパク質である請求の範囲第1項～第7項のいずれか1に記載の方法。

[請求項10] タンパク質AがUDP-グルクロン酸転移酵素であり、補酵素がUDP-グルクロン酸である、請求の範囲第4項～第9項のいずれか1に記載の方法。

[請求項11] タンパク質Aが g と a の算出のために用いた変異型タンパク質以外の変異型タンパク質であり、算出された g と a の値を用いて酵素活性の予測を行う、請求の範囲第6項～第10項のいずれか1に記載の方法

。

[請求項12] 請求の範囲第1項～第11項のいずれか1に記載の方法を実行するために、コンピュータを下記的手段として機能させるプログラムを担持する記録媒体：

(1) 入力されたアミノ酸配列情報に基づいて、タンパク質Aの立体構造データを計算する手段、

(2) タンパク質Aの立体構造データと、ドッキングシミュレーションの対象となる基質の立体構造データを記憶する手段、

(3) 前記記憶された、タンパク質Aの立体構造データと基質の立体構造データを用いて、タンパク質Aと基質とのドッキングシミュレーションを行うシミュレーション手段、

(4) シミュレーションにより得られた結果を記憶する手段、

(5) 記憶されたシミュレーション結果に基づいて、タンパク質Aの酵素活性を算出する手段、

(6) 算出された酵素活性を表示する手段。

[請求項13] 請求の範囲第1項～第11項のいずれか1に記載の方法を実行するために、下記的手段を担持する装置：

(1) 入力されたアミノ酸配列情報に基づいて、タンパク質Aの立体構造データを計算する手段、

(2) タンパク質Aの立体構造データと、ドッキングシミュレーションの対象となる基質の立体構造データを記憶する手段、

(3) 前記記憶された、タンパク質Aの立体構造データと基質の立体構造データを用いて、タンパク質Aと基質とのドッキングシミュレーションを行うシミュレーション手段、

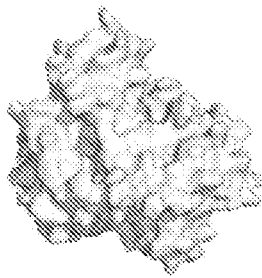
(4) シミュレーションにより得られた結果を記憶する手段、

(5) 記憶されたシミュレーション結果に基づいて、タンパク質Aの酵素活性を算出する手段、

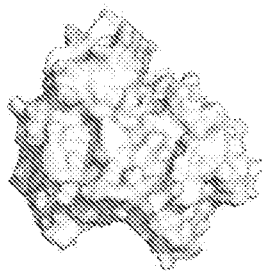
(6) 算出された酵素活性を表示する手段。

- [請求項14] 請求の範囲第1項～第11項のいずれか1に記載の方法、請求の範囲第12項に記載の記録媒体、または請求の範囲第13項に記載の装置を用いて、2以上の基質について、基質ごとのタンパク質Aの酵素活性を予測し、得られた2以上の予測結果に基づいて目的の基質を選択する、基質適合性の判定方法。
- [請求項15] 前記基質が生体に投与される薬剤であって、請求の範囲第14項に記載の方法を用いて、薬剤の投与量および／または投与間隔、投与頻度を評価する方法。

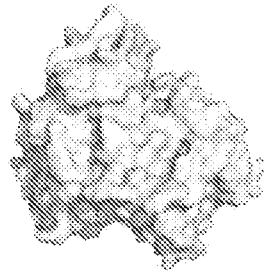
[図1]



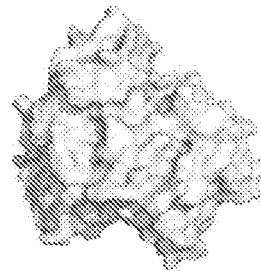
野生型



G 7 1 R 変異型

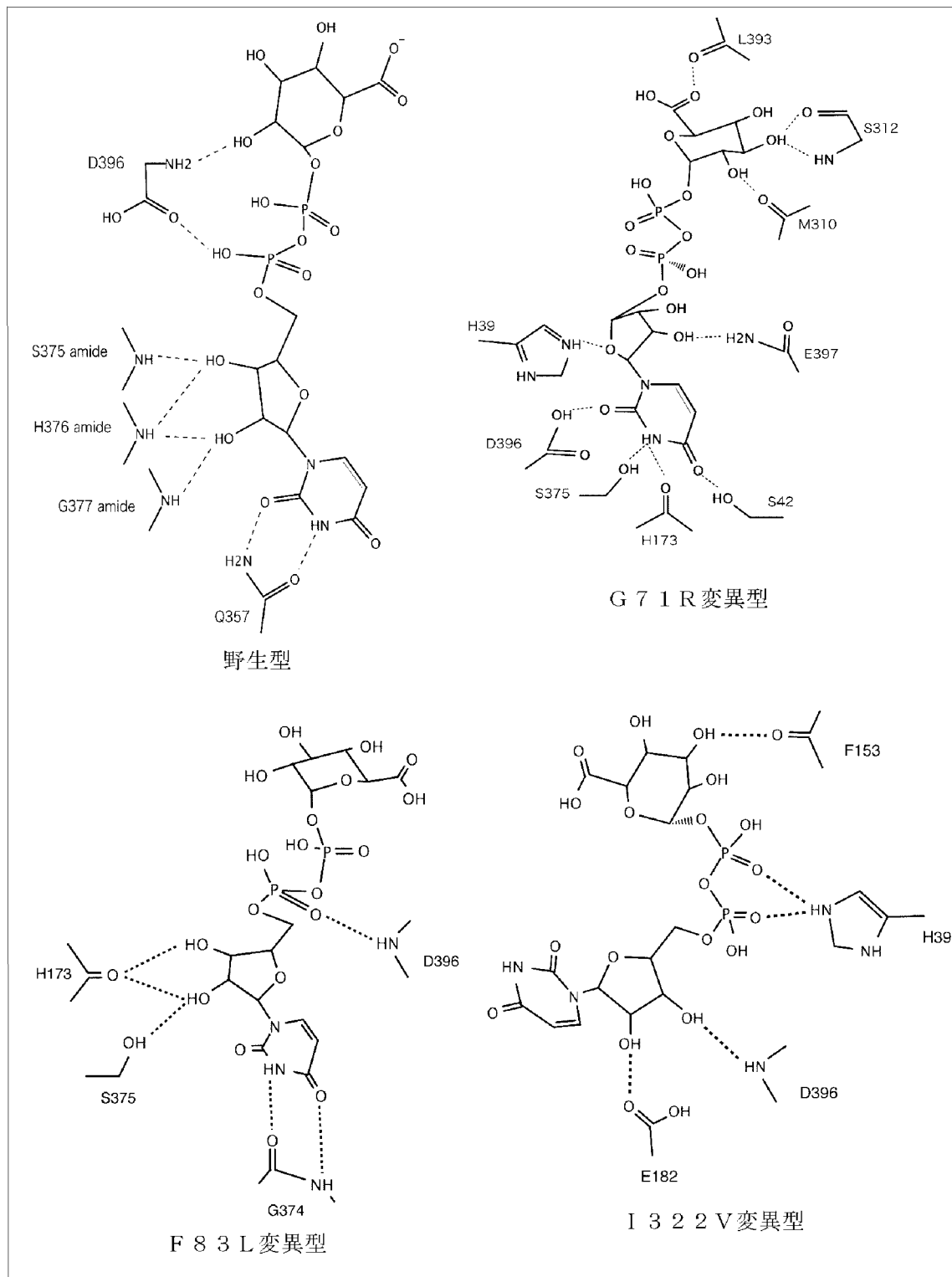


F 8 3 L 変異型

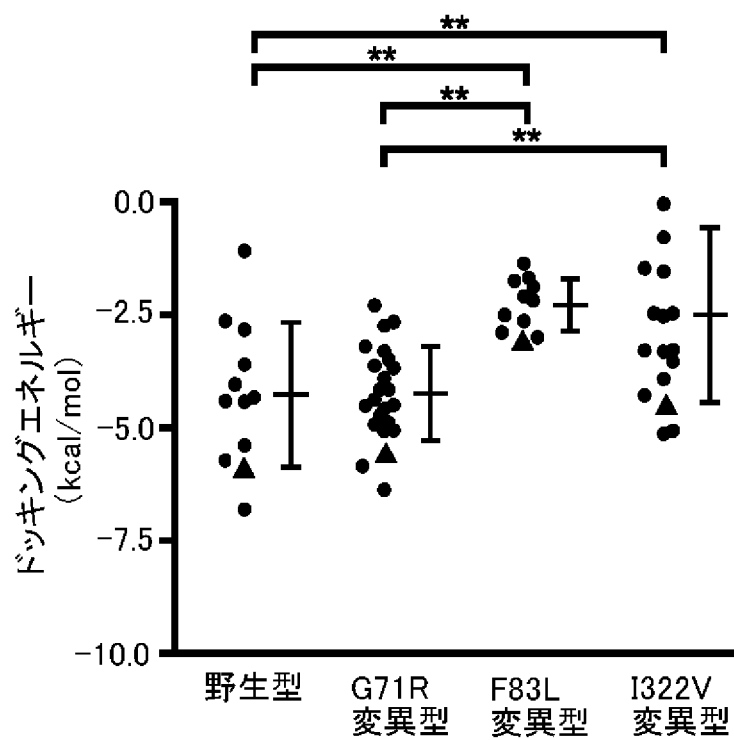


I 3 2 2 V 変異型

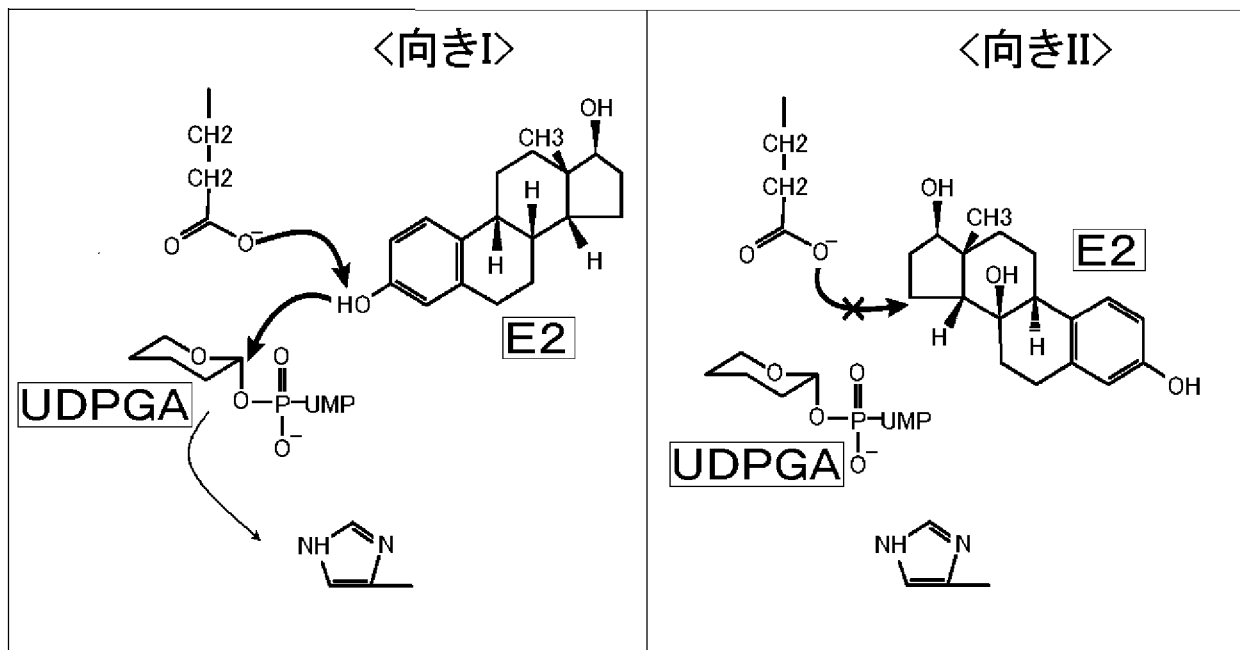
[図2]



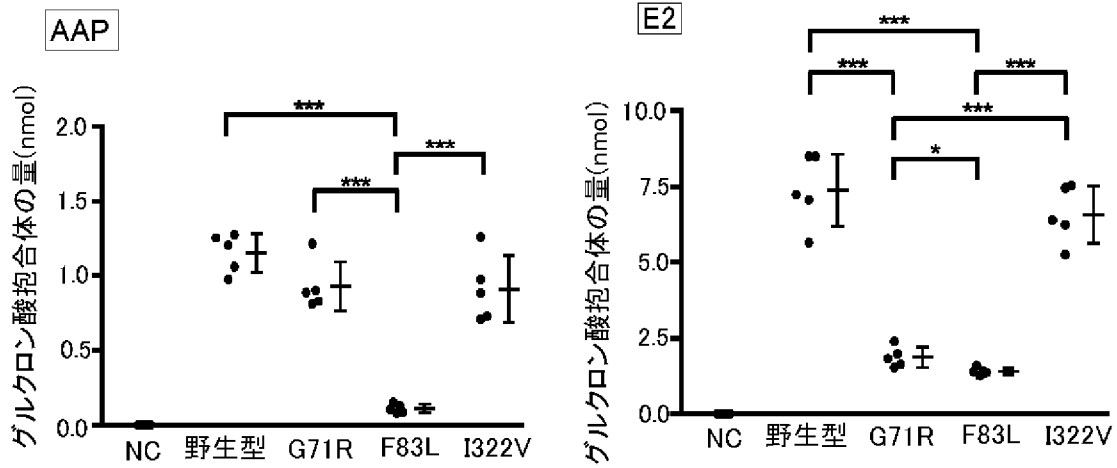
[図3]



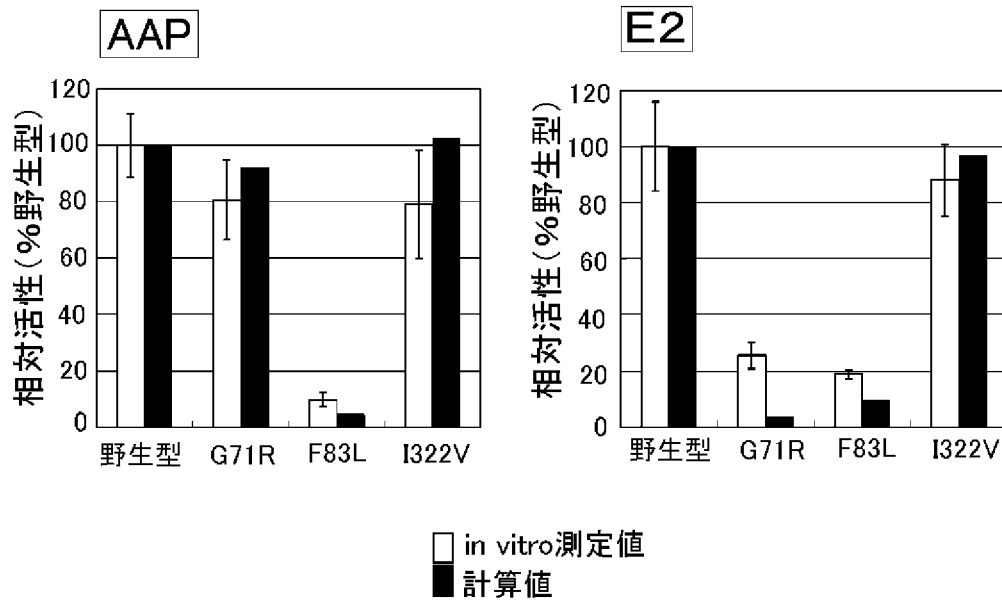
[図4]



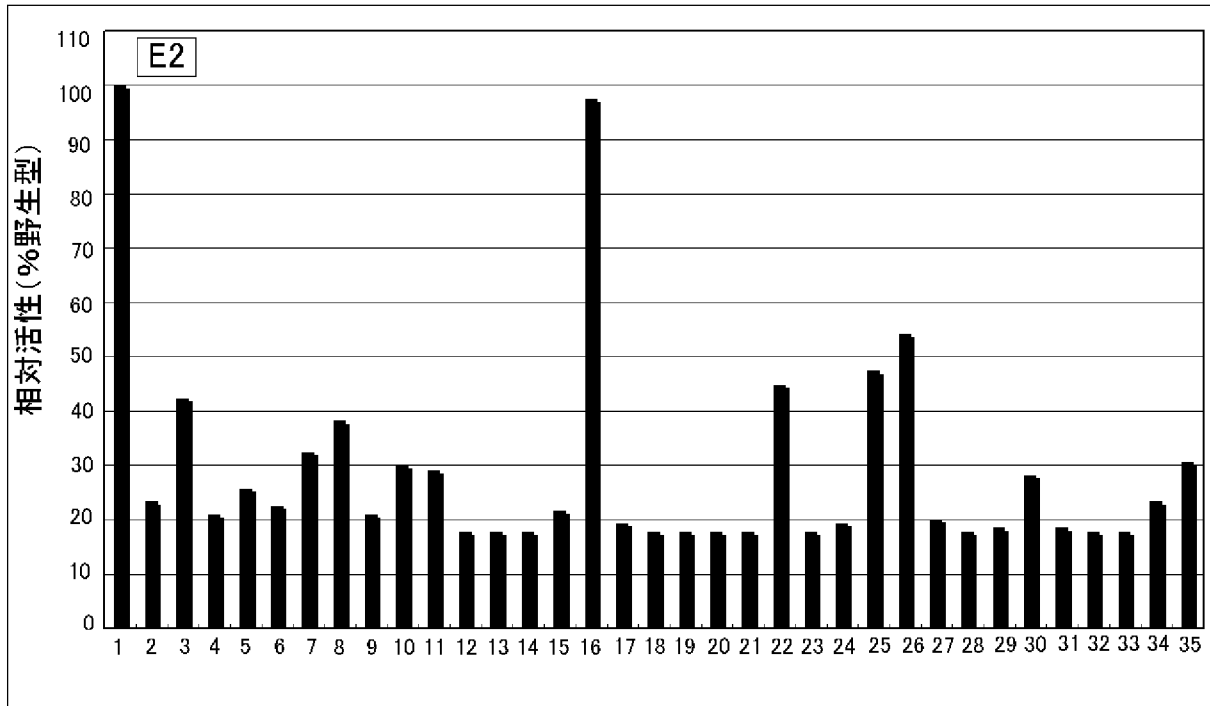
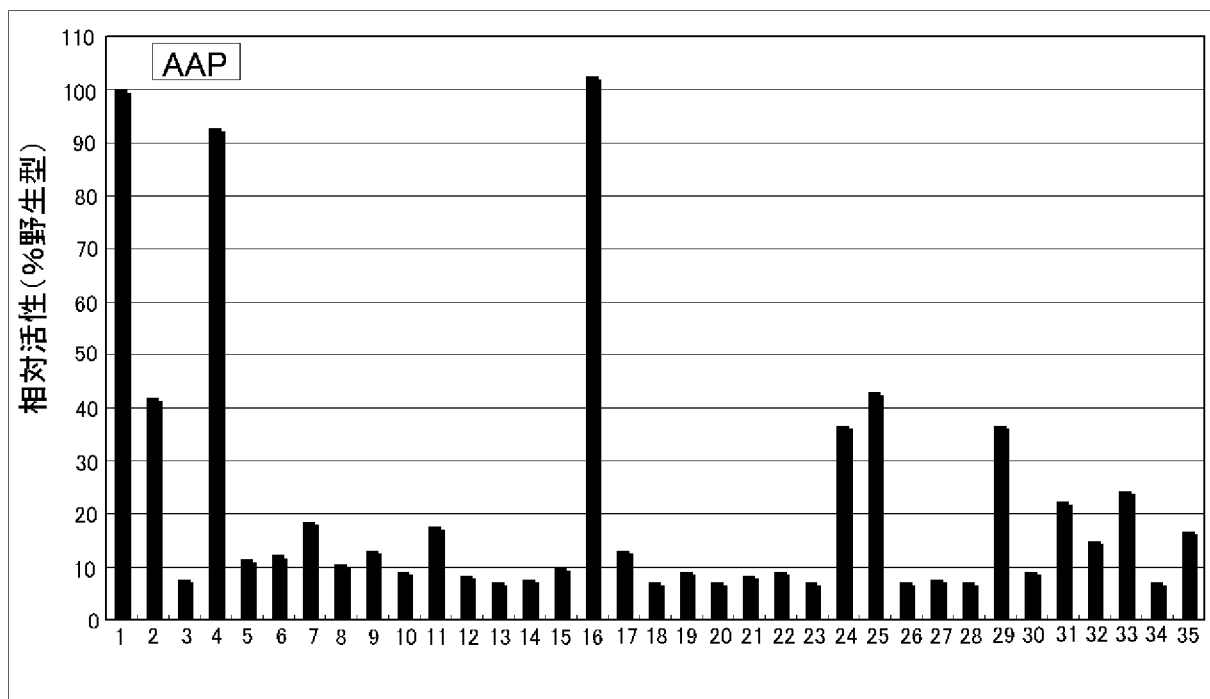
[図5]



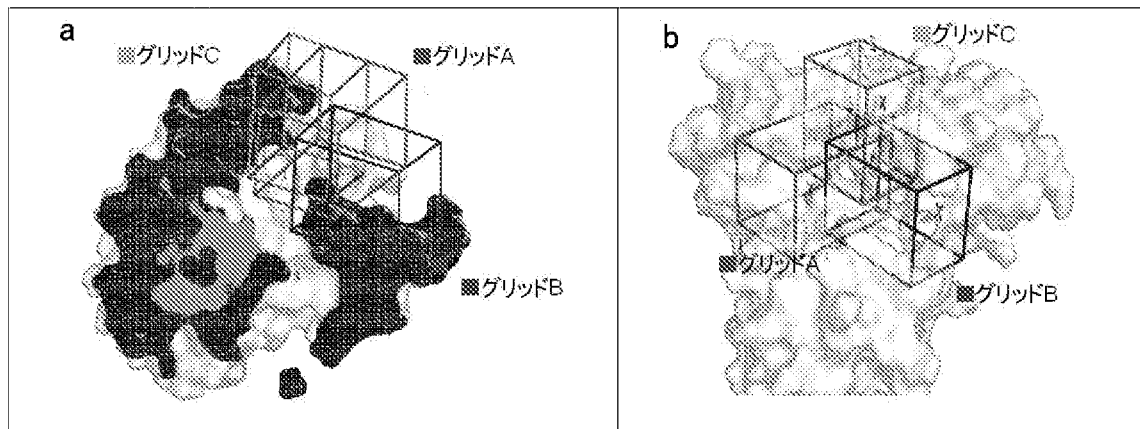
[図6]



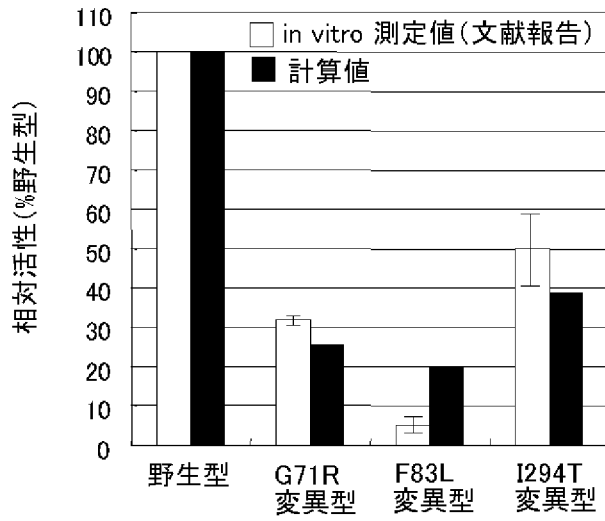
[図7]



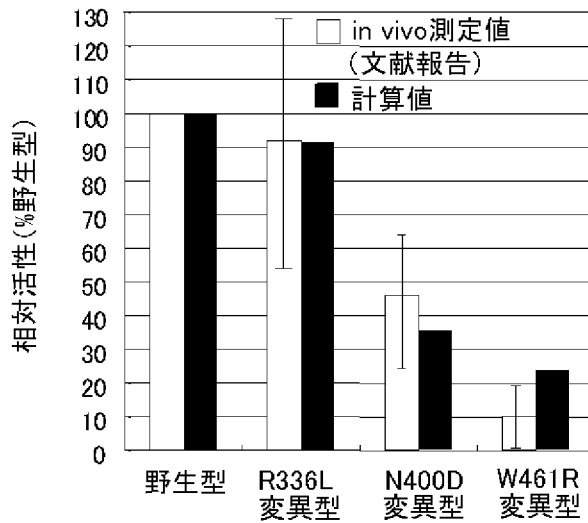
[図8]



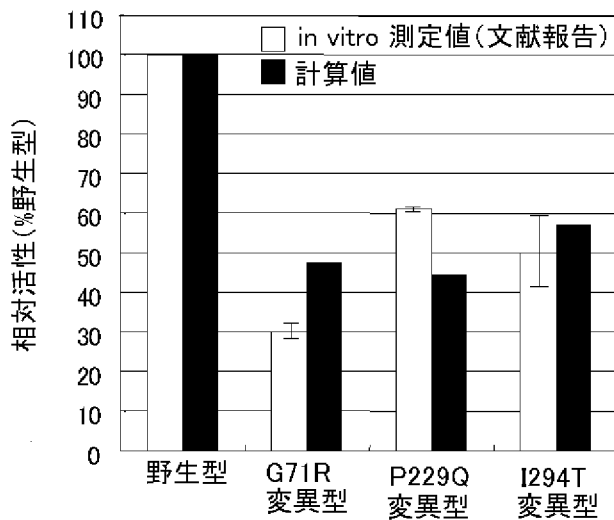
[図9]



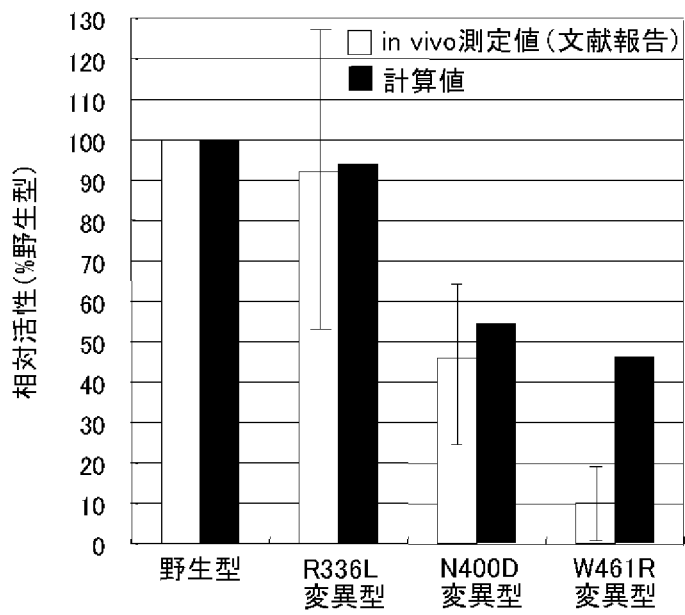
[図10]



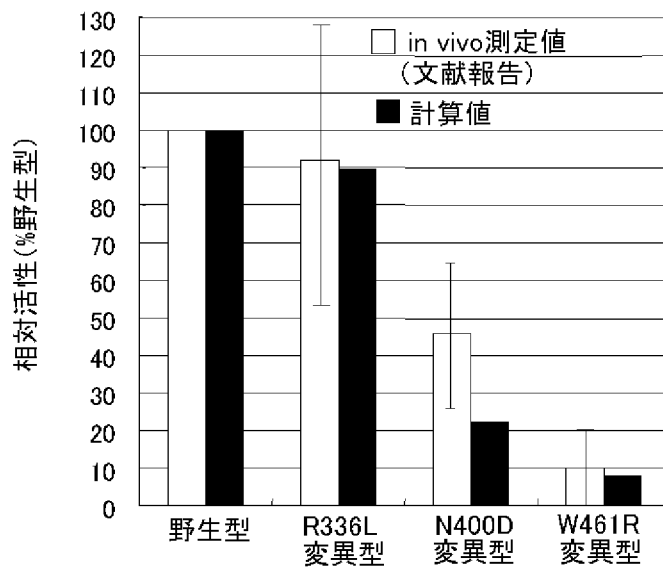
[図11]



[図12]

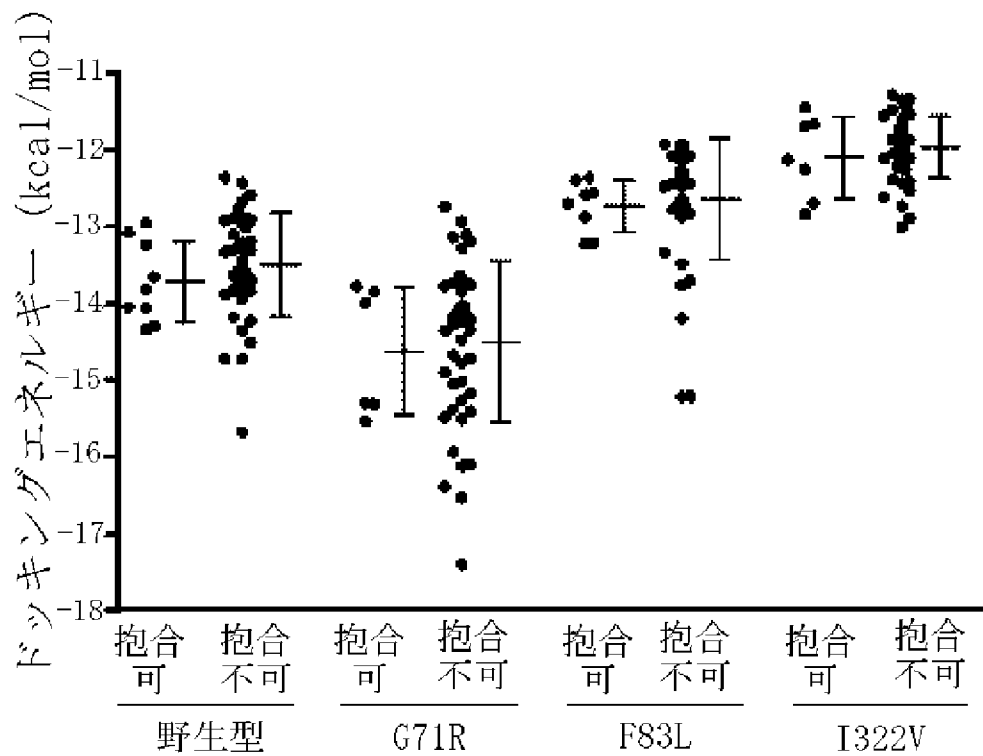


[図13]

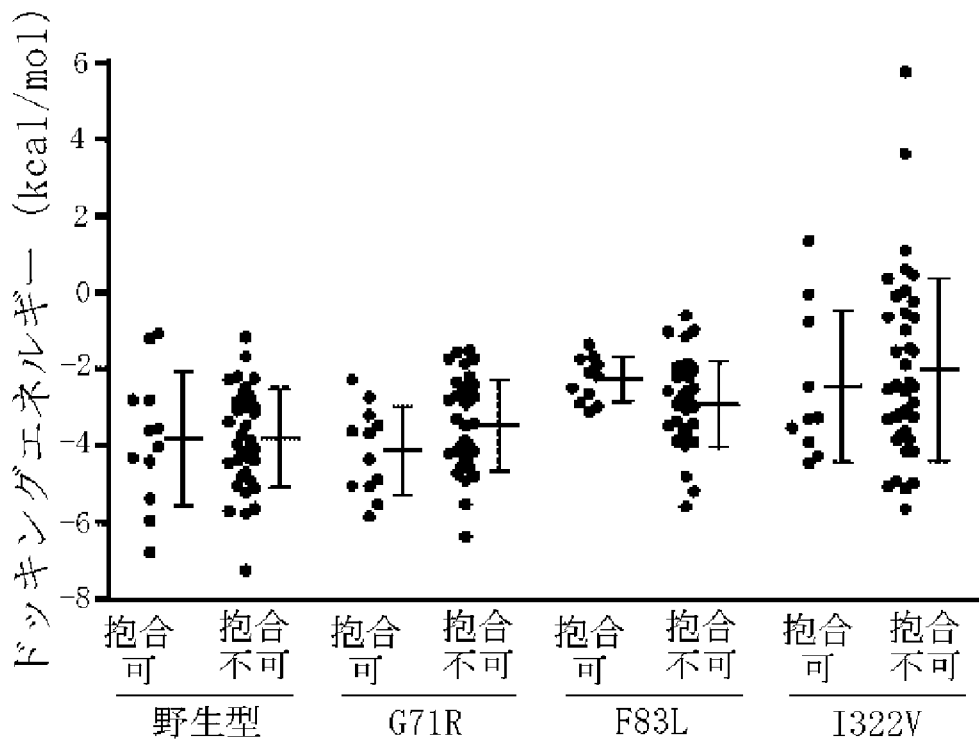


[図14]

水分子付加



水分子付加無し

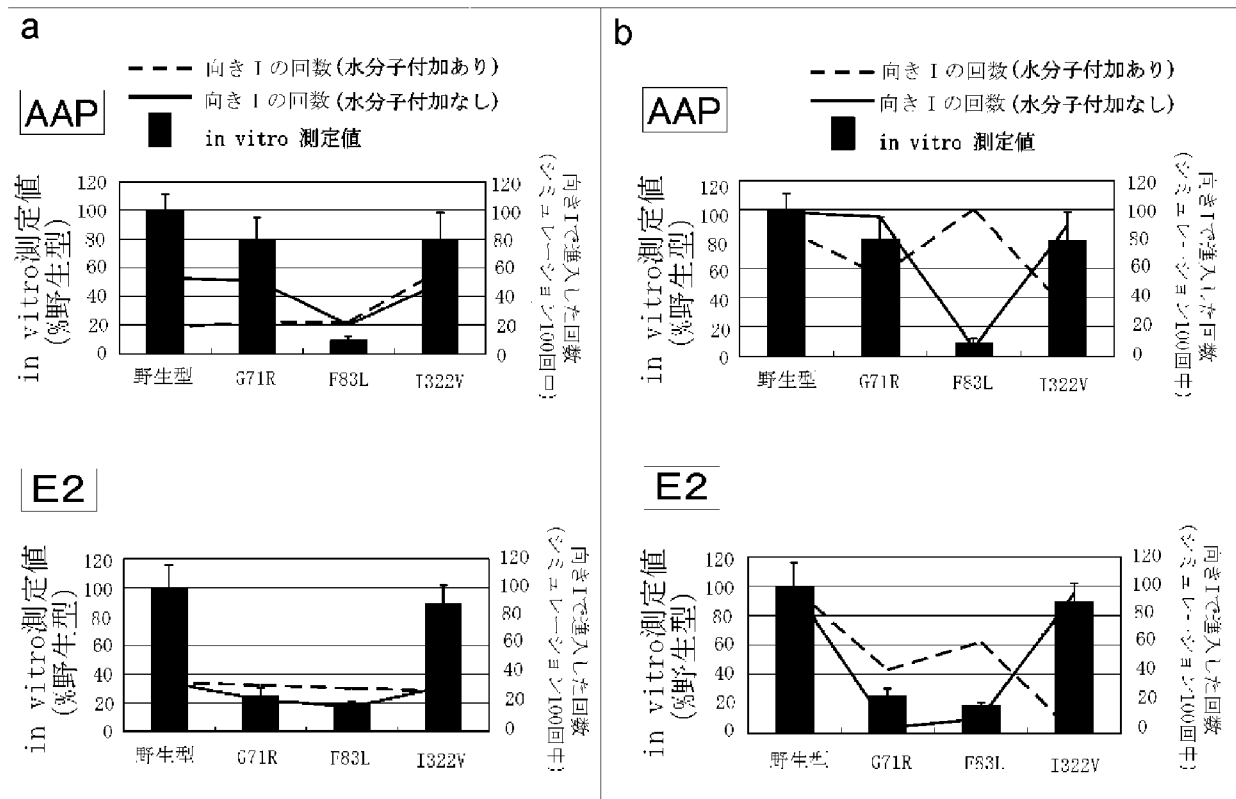


[図15]

a		AAP			
		野生型	G71R	F83L	I322V
		18	22	22	58
E2		AAP			
		野生型	G71R	F83L	I322V
		34	32	30	28

b		AAP			
		野生型	G71R	F83L	I322V
		87	55	100	32
E2		AAP			
		野生型	G71R	F83L	I322V
		98	43	62	0

[図16]



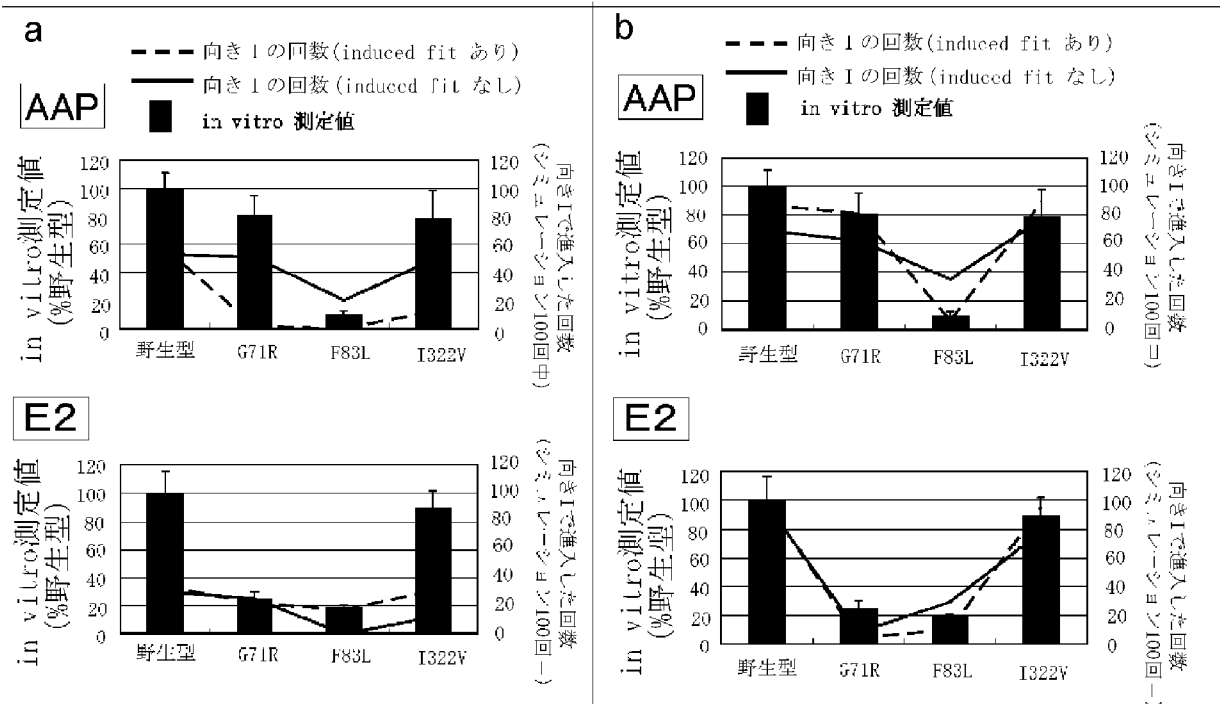
[図17]

GTPγS			
野生型	T31R	G169A	L170W
15	0	6	3
GDP			
野生型	T31R	G169A	L170W
10	26	10	4

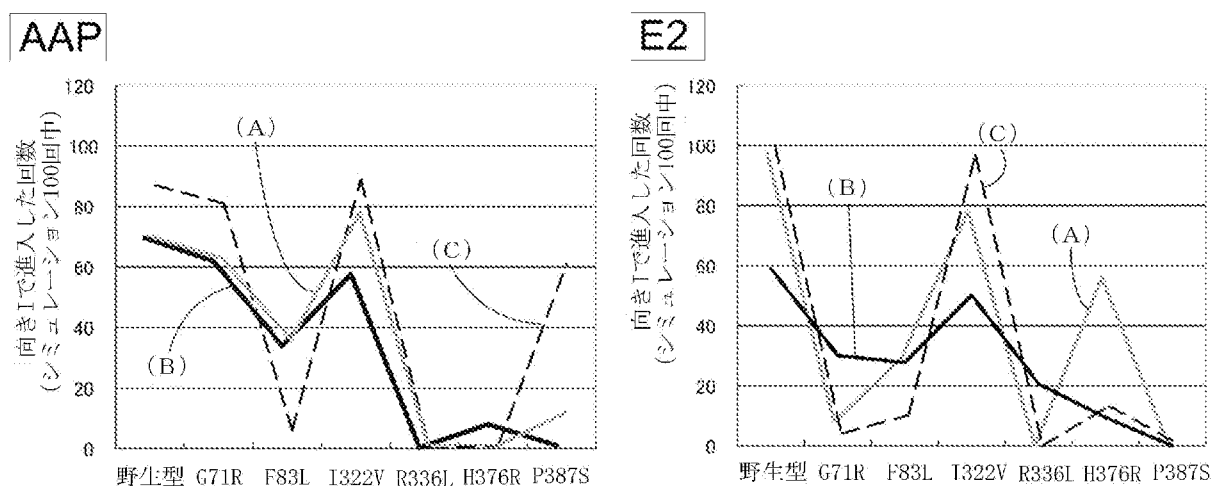
[図18]

a					b				
AAP					AAP				
野生型	G71R	F83L	I322V		野生型	G71R	F83L	I322V	
57	2	0	12		69	62	35	77	
E2					E2				
野生型	G71R	F83L	I322V		野生型	G71R	F83L	I322V	
30	25	0	12		97	8	29	78	

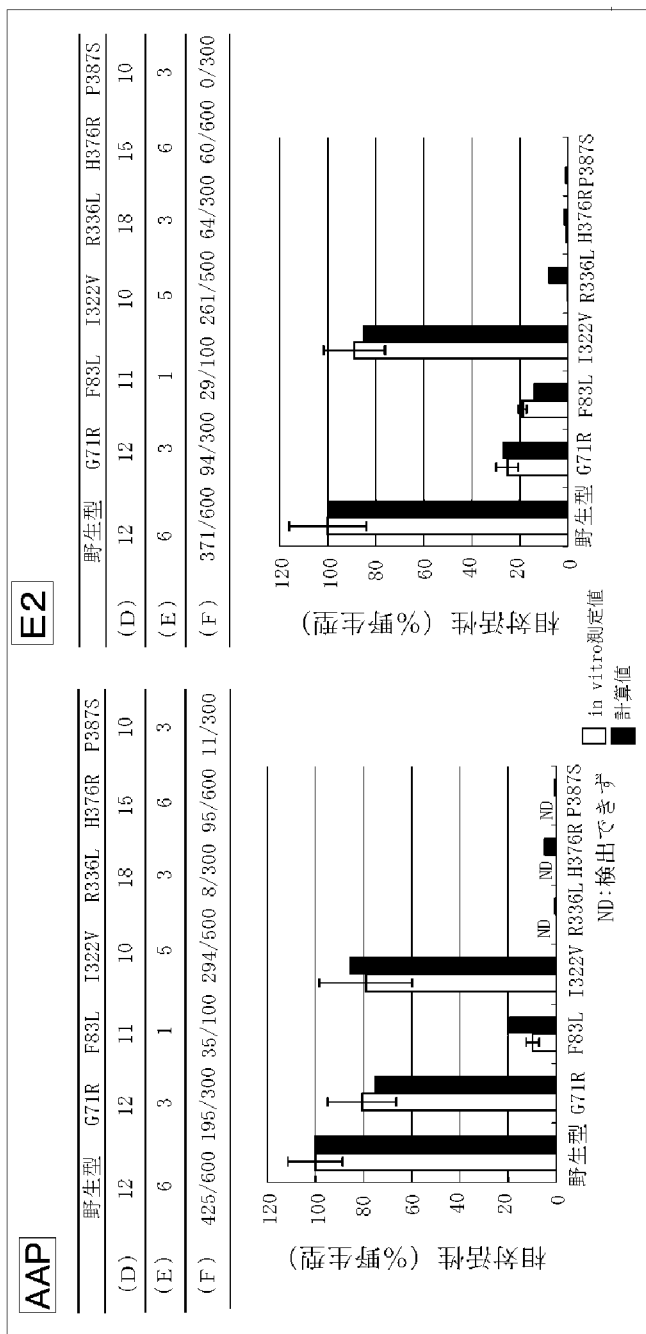
[図19]



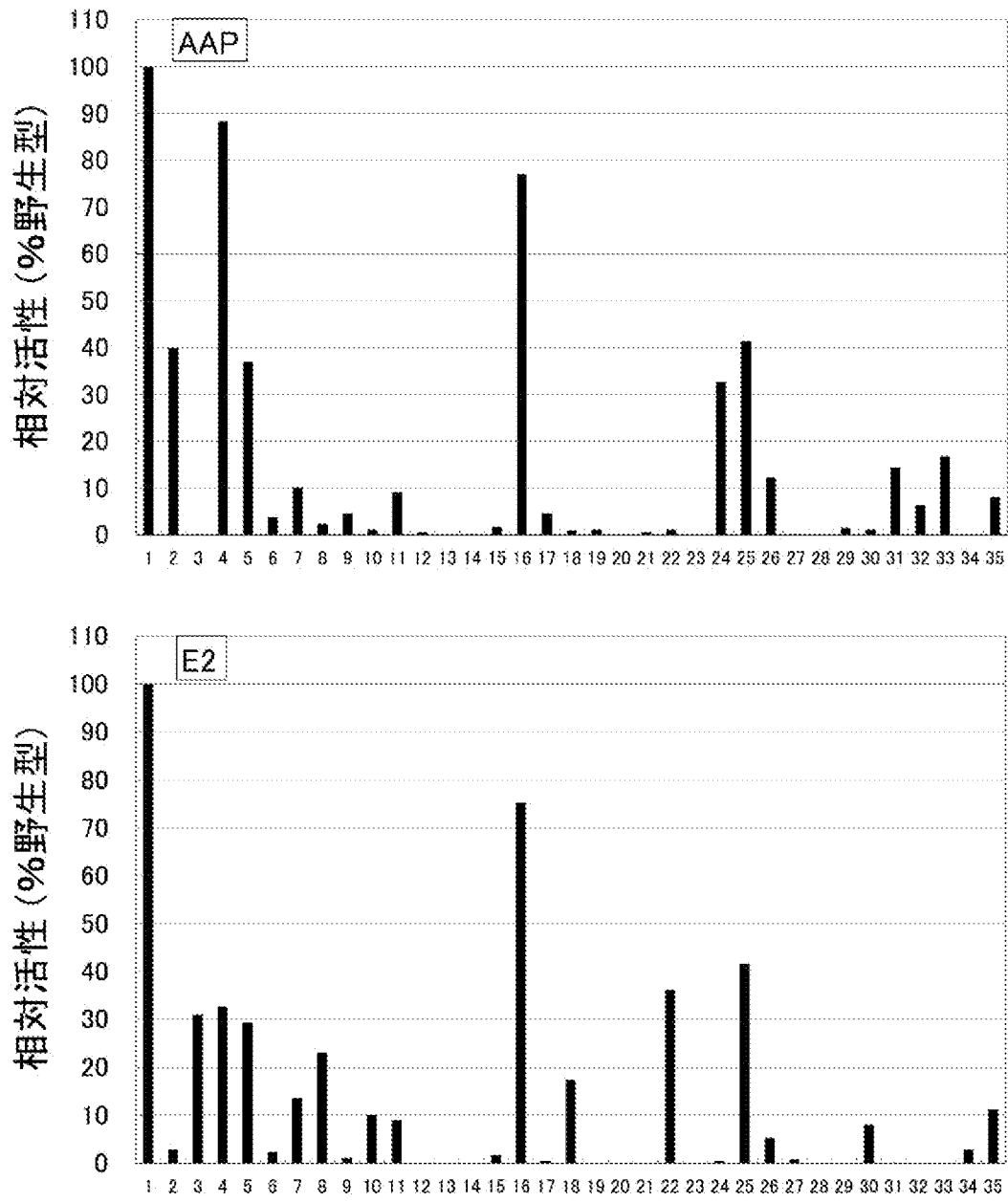
[図20]



[図21]



[図22]



[図23]

