

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7438545号
(P7438545)

(45)発行日 令和6年2月27日(2024.2.27)

(24)登録日 令和6年2月16日(2024.2.16)

(51)国際特許分類	F I	
G 1 6 B 30/00 (2019.01)	G 1 6 B 30/00	
C 0 7 K 14/00 (2006.01)	C 0 7 K 14/00	Z N A
C 1 2 P 21/02 (2006.01)	C 1 2 P 21/02	C

請求項の数 18 (全44頁)

(21)出願番号	特願2020-566712(P2020-566712)	(73)特許権者	500579280
(86)(22)出願日	令和1年5月30日(2019.5.30)		トラスティーズ・オヴ・ダートマス・カレッジ
(65)公表番号	特表2021-525917(P2021-525917 A)		アメリカ合衆国 ニューハンプシャー州
(43)公表日	令和3年9月27日(2021.9.27)		0 3 7 5 5 ハノーヴァー ロープ フェリー ロード 1 1
(86)国際出願番号	PCT/US2019/034670	(74)代理人	100080791
(87)国際公開番号	WO2019/232222		弁理士 高島 一
(87)国際公開日	令和1年12月5日(2019.12.5)	(74)代理人	100136629
審査請求日	令和4年1月18日(2022.1.18)		弁理士 鎌田 光宜
(31)優先権主張番号	62/678,588	(74)代理人	100125070
(32)優先日	平成30年5月31日(2018.5.31)		弁理士 土井 京子
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	100121212
			弁理士 田村 弥栄子
		(74)代理人	100174296

最終頁に続く

(54)【発明の名称】 3次又は4次構造モチーフを用いた計算タンパク質設計

(57)【特許請求の範囲】

【請求項1】

標的構造を複数の構造モチーフに分解する工程と、
構造データベースで前記複数の構造モチーフの各々に対する複数の構造マッチを同定する工程と、

前記複数の構造マッチの各々を用いて配列 - 構造関係への少なくとも1つの非局所エネルギー寄与の値を推測する工程と、

少なくとも1つの候補アミノ酸配列を発生させる工程であって、前記候補アミノ酸配列が設計可能な性質を有する(たとえば、前記標的構造の結合パートナーにフォールド可能である)、工程と、

前記複数の構造マッチの各々を用いて配列 - 構造関係への少なくとも1つの局所エネルギー寄与の値を取得する工程と、

を含む、アミノ酸配列の *in silico* 設計法。

【請求項2】

前記少なくとも1つの非局所エネルギー寄与が、前記複数の構造モチーフの1つ内の単一設計位置を取り囲む骨格のコンティグストレッチに由来する、請求項1に記載の方法。

【請求項3】

前記少なくとも1つの非局所エネルギー寄与が、前記複数の構造モチーフの1つ内の単一設計位置に配列近接ではなく空間近接する骨格に由来する、請求項1に記載の方法。

【請求項4】

前記少なくとも1つの非局所エネルギー寄与が、前記複数の構造モチーフの1つ内の結合残基対に由来する、請求項1に記載の方法。

【請求項5】

前記少なくとも1つの局所エネルギー寄与が、前記複数の構造モチーフの1つ内の単一設計位置の骨格角度に由来する、請求項1～4のいずれか一項に記載の方法。

【請求項6】

前記骨格角度が 角、 角、又は 角である、請求項5に記載の方法。

【請求項7】

前記標的構造がタンパク質の3次構造である、請求項1～6のいずれか一項に記載の方法。

【請求項8】

前記標的構造がタンパク質複合体の4次構造である、請求項1～6のいずれか一項に記載の方法。

【請求項9】

標的構造を複数の構造モチーフに分解する工程と、
構造データベースで前記複数の構造モチーフの各々に対する複数の構造マッチを同定する工程と、

エネルギー寄与の階層に従って前記複数の構造マッチの各々を用いて配列 - 構造関係へのエネルギー寄与の値のセットを逐次推測する工程であって、前記階層が、

i . 前記複数の構造モチーフの1つ内の単一設計位置について少なくとも1つの局所エネルギー寄与、

i i . 前記単一設計位置を取り囲む骨格のコンティグストレッチ、

i i i . 前記単一設計位置に配列近接ではなく空間近接する骨格、及び

i v . 前記単一設計位置を含む結合残基対、

の少なくとも2つを含む、工程と、

設計可能な性質を有する（たとえば、前記標的構造の結合パートナーにフォールド可能である）少なくとも1つの候補アミノ酸配列を発生させる工程と、

を含む、アミノ酸配列の *in silico* 設計法。

【請求項10】

前記階層が、

v . 前記単一設計位置を含む残基のトリプレット

をさらに含む、請求項9に記載の方法。

【請求項11】

前記少なくとも1つの局所エネルギー寄与が、前記複数の構造モチーフの1つ内の単一設計位置の骨格角度に由来する、請求項9又は請求項10に記載の方法。

【請求項12】

前記少なくとも1つの局所エネルギー寄与が、前記複数の構造モチーフの1つ内の単一設計位置の埋没状態に由来する、請求項9又は請求項10に記載の方法。

【請求項13】

前記標的構造がタンパク質の3次構造である、請求項9～12のいずれか一項に記載の方法。

【請求項14】

前記標的構造がタンパク質複合体の4次構造である、請求項9～12のいずれか一項に記載の方法。

【請求項15】

標的構造にフォールド可能なアミノ酸配列の *in silico* 設計のための命令がエンコードされた非一時コンピューター可読記憶媒体であって、前記命令がプロセッサにより実行可能であり且つ請求項1～14のいずれか一項に記載の方法を含む、非一時コンピューター可読記憶媒体。

【請求項16】

10

20

30

40

50

請求項 1 ~ 14 のいずれか一項において発生させた前記候補アミノ酸配列をコードする核酸配列を提供することと、

前記核酸配列を宿主細胞に導入することと、

前記候補アミノ酸配列を発現させることと、

を含む、標的構造の結合パートナーにフォールドするタンパク質の製造法。

【請求項 17】

前記候補アミノ酸配列が前記標的構造の前記結合パートナーにフォールドするかを決定することをさらに含む、請求項 16 に記載の方法。

【請求項 18】

前記タンパク質が、酵素、抗体、レセプター、輸送タンパク質、ホルモン、成長因子、及びそれらの断片からなる群から選択される、請求項 16 に記載の方法。

10

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本特許出願は、2018年5月31日出願の米国仮特許出願第62/678,588号（その全内容は参照により本明細書に完全に組み込まれる）に基づく優先権を主張する。

【0002】

連邦政府委託の研究又は開発

本発明は、国立科学財団（National Science Foundation）により付与されたDMR1534246及び国立衛生研究所（National Institutes of Health）により付与されたP20GM113132に基づいて政府の支援を受けて行われた。政府は本発明に関する一定の権利を有する。

20

【0003】

本開示は、計算タンパク質設計に関し、特定的には、あらかじめ定義された構造に又は標的構造の結合パートナーにフォールド可能なタンパク質を設計するための方法、デバイス、及びシステムに関する。

【背景技術】

【0004】

計算タンパク質設計（CPD）は、あらかじめ定義された構造（標的）にフォールドするアミノ酸配列を見いだすタスクである。1990年代中期に最初に定式化されたCPDへの現代のアプローチの背景にある基本概念は、物理原理から基本タンパク質現象（たとえば、フォールディング及び結合）のアミノ酸配列決定基をキャプチャーすることである。具体的には、目標は、根底にある原子間相互作用をモデリングすることにより標的構造中のいずれかのタンパク質配列の自由エネルギーを近似することである。そうするための計算手順は、スコアリング機能といわれる。スコアリング機能が手に入れば、所与の標的にとくに有利なエネルギーを有する配列を探索することによりCPDを実施可能である。

30

【0005】

実際には、多くの課題が伝統的CPDの確度を制限して究極的に低ロバスト性をもたらす。設計との関連では、正確な自由エネルギーを計算するのに十分な詳細レベルでタンパク質構造の物理をモデリングすることは、現在のところ実現不可能である。そのため、予測能力を強く制限する物理ベーススコアリング機能では、かなりの近似を行わなければならない。代替案として、いくつかの基本物理現象は、知識ベースポテンシャル（統計ポテンシャルとしても知られる）を介して経験的にモデリング可能である。この場合、特異的な構造上の特徴（たとえば、互いに特定距離にある2つの特異的原子）の有利性を導出するために原子相互作用のエネルギーを評価する代わりに、頻度の高いものほど有利であると仮定することにより、既知のタンパク質構造中のこうした特徴の頻度を測定してその経験的有利性を定量する。たとえば、骨格二面角、原子間距離、及び充填密度、結合配向、残基埋没状態、及び残基間接触など、単純な構造上の特徴が統計ポテンシャルを構築するために活用されてきた。物理ベース機能、統計機能、又はハイブリッドエネルギー機能に

40

50

依拠するかにかかわらず、CPDの基本的問題が残る。つまり、原子間相互作用の詳細は現実には究極的に配列 - 構造関係（すなわち、どの配列が所与の構造にフォールドするか）を具体化するが、それにもかかわらずそれはこうした関係から取り出される非常に多くの工程である。そのため、原子論的現象のモデリングの誤差がたとえ少量であっても、合わさるとアミノ酸配列の究極的予測ではかなりの誤差になる可能性がある。既存のポテンシャルの誤差が小さくなく且つランダムでもなく、むしろ、コンフィギュレーションエントロピー、アンフォールド状態の自由エネルギー、又は溶媒の存在などの寄与が多くの場合完全に失われることに関連して、大きく且つ系統的であるという事実により、このことはさらに悪化する。実は、基本的原子間相互作用及び他のエネルギー寄与が加成性であるという基本仮定でさえも、単なる近似にすぎない。たとえば、所与のコンフィギュレーションアンサンブルのタンパク質配列の自由エネルギーは、とくに溶媒の効果を考慮したとき、その原子間相互作用の加法関数ではないことが知られている。

10

【発明の概要】**【発明が解決しようとする課題】****【0006】**

そのため、CPDの成功率が有意に高くなるようにスコアリング機能問題に対処する新たな方法を提供するタンパク質設計へのアプローチの必要性が当技術分野に存在する。

【課題を解決するための手段】**【0007】**

本開示は、根底にある原子論的物理をモデリングすることにより間接的に導くのではなく、既存のタンパク質構造から配列対構造関係を直接観測することに基づく新たなCPD法を提供する。タンパク質構造は、天然アミノ酸の配列で実現可能であるという意味で、ある特定の骨格ジオメトリーのみが許容される（すなわち設計可能である）準離散空間を表す。2次、3次、及び4次構造状況をキャプチャーするタンパク質データバンク（PDB）の各残基を取り囲む局所骨格構造モチーフは、系統的に特徴付けられてきた（1）。本明細書では総称して「TERM」（3次モチーフの略であるが、以上に述べたようにこのモチーフは2次、3次、及び4次構造をキャプチャーする）というこうしたモチーフは、非関連タンパク質に渡って天然で高度に再使用される。たとえば、わずか約600TERMあればサブ分解能で既知の構造ユニバースの50%を記述するのに十分である（1）。構造空間のこの明白な縮重のおかげで、TERMは、配列 - 構造関係の基本ルールを効果的にキャプチャーする。この理由は、各モチーフが多くの場合何千もの異なる配列 / 構造状況のPDBで何回も現れることにある。こうした多くのマッチ配列の解析により、対応するTERMにより表される構造フラグメントの配列決定基を抽出可能である。

20

30

【0008】

本明細書に提供されるアプローチには、当該技術の現状を凌駕する少なくとも3つの利点が存在する。第1に、本明細書に記載の方法は、天然タンパク質で観測される配列 - 構造関係の証明済みルールに基づいて配列を設計する。すなわち、設計手順に向けて考慮されるTERMマッチ配列はすべて、標的構造の一部である対応する骨格コンフォメーションを真に形成することが、アプリアリに分かる。既知のビルディングブロックからのこのタイプの設計は、既存の方法よりもかなり高い成功率を期待できることを意味する（このことは本明細書に開示される確証試験で観測された）。第2に、同様に既存のタンパク質構造に基づく統計スコアリング機能との関連で、本明細書に記載の方法は、距離や角度などの基本的な構造上の特徴の優先度間に加成性及び非依存性を仮定しない。その代わりに、TERMベース配列 - 構造優先度を直接観測することにより、本方法は、複数の寄与の集団作用を（黙示的に）説明する。最後に、TERMベースアプローチは、タンパク質がスタティック分子ではなくコンフォメーションアンサンブルとして室温で存在することを認識する新規な方法を提供する。この理由は、配列統計（及び究極的にはスコアリング機能）がTERMマッチにより表される構造アンサンブル、すなわち、構造データベース（たとえば、天然タンパク質を含む構造データベース）に見いだされる類似の骨格コンフィギュレーションの厳密ではないが近似インスタンスから生じることにある。そのため、T

40

50

ERMベース設計は、特定フリーズ骨格コンフィギュレーションだけでなくタンパク質構造状態のより適切な表現である近似コンフィギュレーションのアンサンブルにも適合可能なアミノ酸配列の同定を可能にする。骨格フレキシビリティをモデリングする必要性に対処するアプローチは、既存のCPD法との関連で提案されてきたが、かなりの計算コストを招くほか、まさに背景のセクションで考察したスコアリング確度（究極的にはロバスト性）の制限を受ける。

【0009】

一態様では、本開示は、ホリスティック原子論的定義構造環境との関連で配列統計を得ることに基づくタンパク質設計へのアプローチを提供する。このアプローチは、少なくとも基本的構造記述子の加成性の仮定の必要性だけでなくタンパク質構造の天然縮重の認識及び利用の必要性も回避するので有利である。実は、このアプローチの優れた性能は、タンパク質構造ユニバースがある特定の骨格ジオメトリーのみが許容される（すなわち設計可能である）準離散空間を表すというその認識に（少なくとも部分的に）起因しうる。そのため、本開示は、精密に定義された詳細構造環境の統計を活用するタンパク質設計へのアプローチを提供する。

【0010】

他の一態様では、本開示は、アミノ酸配列の *in silico* 設計法を提供する。ある特定の実施形態では、本方法は、標的構造を複数の構造モチーフに分解する工程と、構造データベースで複数の構造モチーフの各々に対する複数の構造マッチを同定する工程と、複数の構造マッチの各々を用いて配列 - 構造関係への少なくとも1つの非局所エネルギー寄与の値を推測する工程と、少なくとも1つの候補アミノ酸配列を発生させる工程と、を含む。ある特定の実施形態では、候補アミノ酸配列は設計可能な性質を有する。ある特定の実施形態では、候補アミノ酸配列は、標的構造の結合パートナーにフォールド可能なタンパク質である。ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置を取り囲む骨格のコンティグストレッチ（たとえば、 $(i - n) \sim (i + n)$ ）であり、式中、 i は所与の位置であり、且つ n は制御可能パラメーターである）に由来する。ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置に配列近接ではなく空間近接する骨格に由来する。ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与は、複数の構造モチーフの1つ内の結合残基対に由来する。ある特定の実施形態では、本方法は、複数の構造マッチの各々を用いて配列 - 構造関係への少なくとも1つの局所エネルギー寄与の値を取得する工程をさらに含む。いくつかのかかる実施形態では、少なくとも1つの局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置の骨格角度に由来する。いくつかのかかる実施形態では、骨格角度は 角、 角、又は 角である。ある特定の実施形態では、標的構造はタンパク質の3次構造である。ある特定の実施形態では、標的構造はタンパク質複合体の4次構造である。

【0011】

さらに他の一態様では、本開示は、アミノ酸配列の *in silico* 設計法を提供する。ある特定の実施形態では、本方法は、標的構造を複数の構造モチーフに分解する工程と、構造データベースで複数の構造モチーフの各々に対する複数の構造マッチを同定する工程と、エネルギー寄与の階層に従って複数の構造マッチの各々を用いて配列 - 構造関係へのエネルギー寄与の値のセットを逐次推測する工程であって、階層が、 (i) 複数の構造モチーフの1つ内の単一設計位置について少なくとも1つの局所エネルギー寄与、 $(i i)$ 単一設計位置を取り囲む骨格のコンティグストレッチ、 $(i i i)$ 単一設計位置に配列近接ではなく空間近接する骨格、及び $(i v)$ 単一設計位置を含む結合残基対、の少なくとも2つを含む、工程と、少なくとも1つの候補アミノ酸配列を発生させる工程と、を含む。ある特定の実施形態では、候補アミノ酸配列は、標的構造の結合パートナーにフォールド可能なタンパク質である。ある特定の実施形態では、階層は高次寄与をさらに含む。ある特定の実施形態では、階層は、 (v) 単一設計位置を含む残基のトリプレットをさらに含む。ある特定の実施形態では、少なくとも1つの局所エネルギー寄与は、複数の構

10

20

30

40

50

造モチーフの1つ内の単一設計位置の骨格角度に由来する。ある特定の実施形態では、少なくとも1つの局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置の埋没状態に由来する。ある特定の実施形態では、標的構造はタンパク質の3次構造である。ある特定の実施形態では、標的構造はタンパク質複合体の4次構造である。

【0012】

さらに他の一態様では、本開示は、標的構造の結合パートナーにフォールド可能なアミノ酸配列の *in silico* 設計のための命令がエンコードされた非一時コンピュータ可読記憶媒体を提供する。命令は、プロセッサにより実行可能であり且つ本明細書に開示される方法を含む。

【0013】

さらに他の一態様では、本開示は、標的構造の結合パートナーにフォールドするタンパク質の作製方法を提供する。ある特定の実施形態では、本方法は、本明細書に開示される *in silico* 設計法により発生させた候補アミノ酸配列をコードする核酸配列を提供することと、核酸配列を宿主細胞に導入することと、候補アミノ酸配列を発現させることと、を含む。ある特定の実施形態では、本方法は、候補アミノ酸配列が標的構造の結合パートナーにフォールドするかを決定することをさらに含む。

【0014】

さらに他の一態様では、本開示は、本明細書に開示される方法により生成されたタンパク質を提供する。

【0015】

本明細書に記載の態様のいずれかのある特定の実施形態では、タンパク質は、酵素、抗体、レセプター、輸送タンパク質、ホルモン、成長因子、及びそれらの断片からなる群から選択される。

【0016】

本明細書に記載の態様のいずれかのある特定の実施形態では、タンパク質は標的構造の設計変異体である。いくつかのかかる実施形態では、標的構造は、蛍光タンパク質、Gタンパク質共役レセプター (GPCR)、及びPDZドメイン含有タンパク質からなる群から選択される。

【0017】

本明細書に記載の態様のいずれかのある特定の実施形態では、標的構造は蛍光タンパク質である。いくつかのかかる実施形態では、蛍光タンパク質は赤色蛍光タンパク質 (RFP) である。

【0018】

本明細書に記載の態様のいずれかのある特定の実施形態では、標的構造はGタンパク質共役レセプター (GPCR) である。いくつかのかかる実施形態では、GPCRは、 β -1アドレナリン作動性レセプターなどのアドレナリン作動性レセプターである。

【0019】

本明細書に記載の態様のいずれかのある特定の実施形態では、標的構造はPDZドメイン含有タンパク質である。いくつかのかかる実施形態では、PDZドメイン含有タンパク質は、 Na^+/H^+ 交換体レギュラトリー因子2 (NHERF-2) (E3KARP、SIP-1、及びTKA-1とも呼ばれる) である。いくつかのかかる実施形態では、PDZドメイン含有タンパク質は、膜関連グアニル酸キナーゼ (MAGI-3) である。

【0020】

本明細書に記載の態様のいずれかのある特定の実施形態では、標的構造の結合パートナーは、PDZドメインに結合するタンパク質又は他の分子である。いくつかのかかる実施形態では、標的構造の結合パートナーは、リゾホスファチジン酸レセプター2 (LPA2) である。

【0021】

本発明のこれらの及び他の目的は、以下のパラグラフに記載される。これらの目的は、本発明の範囲を狭くするものとみなすべきでない。

10

20

30

40

50

【 0 0 2 2 】

本発明をより良く理解するために、以下の図面に示される実施形態を参照しうる。

【 図面の簡単な説明 】

【 0 0 2 3 】

【 図 1 】 図 1 は、本技術の模範的実施形態に係るフローチャートを示す。

【 図 2 - 1 】 図 2 A 及び 2 B は、本技術の模範的実施形態に係るフローチャートを示す。

【 図 2 - 2 】 図 2 A 及び 2 B は、本技術の模範的実施形態に係るフローチャートを示す。

【 図 3 】 図 3 は、本技術の模範的実施形態に係るフローチャートを示す。

【 図 4 - 1 】 図 4 は、模範的計算タンパク質設計法の模式図である。

【 図 4 - 2 】 図 4 は、模範的計算タンパク質設計法の模式図である。

【 図 5 】 図 5 は、模範的標的構造 m C h e r r y の全表面再設計を示す。左側パネルは、設計の変動が許容された 6 4 表面位置を灰色球として示す。中間パネル及び右側パネルは、真空静電位を疑似色で表してそれぞれ元の m C h e r r y 及び再設計変異体の表面を示す。

【 図 6 】 図 6 は、m C h e r r y タンパク質のサイズ排除クロマトグラムを示す。トップパネルは、野生型 m C h e r r y 及び m C h e r r y - L O V 2 融合タンパク質（後者は W a n g ら（ 2 ）により記載される）を含有する標準のクロマトグラムを示す。ボトムパネルは、野生型と同一の体積近くに溶出することを示す単独再設計 m C h e r r y 変異体のクロマトグラムを示す。標準に基づいて、ダイマータンパク質は、点線で表される体積で溶出すると予想されるので、設計オリゴマー化の可能性は排除される。そのため、サイズ排除クロマトグラフィーは、設計 m C h e r r y タンパク質が溶液中でモノマーであることを示す。

【 図 7 】 図 7 は、m C h e r r y タンパク質の吸収スペクトルを示す。トップパネルは、野生型及び再設計 m C h e r r y タンパク質の吸収スペクトル（それぞれ左側及び右側の Y 軸に吸光度値が示される）を比較して両者が類似のスペクトル形状を呈することを示す。ボトムパネルは、等価タンパク質濃度で測定された両タンパク質の蛍光スペクトルを比較する。再設計 m C h e r r y タンパク質は、フルオロフォアの光の性質を保存する。

【 図 8 】 図 8 は、m C h e r r y 及び模範的設計変異体の化学変性を示す。フォールド度は、5 8 7 n m のクロモフォア吸光度によりモニターした。クロモフォアは、水に暴露されると迅速に加水分解するので、これは感度の良い構造メトリックを構成する。データはヒル式に当てはめられ、半変性濃度が凡例に記される。

【 図 9 】 図 9 は、1 アドレナリン作動性レセプター G P C R（P D B エントリー 4 B V N）の結晶構造を示し、細胞外境界及び細胞質膜境界の近似位置が赤線及び青線で表される（左側パネル）。中間パネル及び右側パネルは、それぞれ野生型 G P C R 及びその再設計カウンターパートの真空中の静電表面電位を示す（同一配向で）。

【 図 1 0 - 1 】 図 1 0 A ~ 1 0 D は、B a k e r らが設計研究で標的とした 4 つの異なるトポロジーを例示する（ 3 ）。図 1 0 E ~ 1 0 F は、本明細書に記載の模範的設計法を用いて計算された X 軸上の各設計の長さを規格化したスコア（そのそれぞれの骨格について）と、Y 軸上の各配列の実験的導出安定性スコアと、の相関を示す。散布プロットの点の色はデータ密度を表し、赤色は最高密度を表し、且つ青色は最低密度を表す。平均曲線は、スコアの 1 0 進行ウィンドウの安定性スコアを平均することにより得られた丸と共に黒線で示される。図 1 0 I ~ 1 0 L は、それぞれ図 1 0 E ~ 1 0 F の場合と同一のプロットを示すが、ロゼッタ法を用いて計算されたスコアを X 軸上に示す。いずれの場合も、本明細書に開示される模範的設計法を用いて計算されたスコアが呈する相関は、ロゼッタ法を用いて計算されたスコアの場合を有意に凌駕する。事実上、ロゼッタでは 4 つの場合のうち 3 つで、相関は、誤った兆候であるか又は統計的に有意でないかのどちらかである（「 X 」で表されるパネル）。一方、本明細書に開示される模範的設計法では、相関は、常に適正な兆候であり且つ統計的にかなり有意である（黒色チェックマークで表される）。そのため、本明細書に開示される T E R M ベース法により計算された統計エネルギーは、設計品質を表す。

10

20

30

40

50

【図10-2】図10A~10Dは、Bakerらが設計研究で標的とした4つの異なるトポロジーを例示する(3)。図10E~10Fは、本明細書に記載の模範的設計法を用いて計算されたX軸上の各設計の長さを規格化したスコア(そのそれぞれの骨格について)と、Y軸上の各配列の実験的導出安定性スコアと、の相関を示す。散布プロットの点の色はデータ密度を表し、赤色は最高密度を表し、且つ青色は最低密度を表す。平均曲線は、スコアの10進行ウィンドウの安定性スコアを平均することにより得られた丸と共に黒線で示される。図10I~10Lは、それぞれ図10E~10Fの場合と同一のプロットを示すが、ロゼッタ法を用いて計算されたスコアをX軸上に示す。いずれの場合も、本明細書に開示される模範的設計法を用いて計算されたスコアが呈する相関は、ロゼッタ法を用いて計算されたスコアの場合を有意に凌駕する。事実上、ロゼッタでは4つの場合のうち3つで、相関は、誤った兆候であるか又は統計的に有意でないかのどちらかである(「X」で表されるパネル)。一方、本明細書に開示される模範的設計法では、相関は、常に適正な兆候であり且つ統計的にかなり有意である(黒色チェックマークで表される)。そのため、本明細書に開示されるTERMベース法により計算された統計エネルギーは、設計品質を表す。

10

【図10-3】図10A~10Dは、Bakerらが設計研究で標的とした4つの異なるトポロジーを例示する(3)。図10E~10Fは、本明細書に記載の模範的設計法を用いて計算されたX軸上の各設計の長さを規格化したスコア(そのそれぞれの骨格について)と、Y軸上の各配列の実験的導出安定性スコアと、の相関を示す。散布プロットの点の色はデータ密度を表し、赤色は最高密度を表し、且つ青色は最低密度を表す。平均曲線は、スコアの10進行ウィンドウの安定性スコアを平均することにより得られた丸と共に黒線で示される。図10I~10Lは、それぞれ図10E~10Fの場合と同一のプロットを示すが、ロゼッタ法を用いて計算されたスコアをX軸上に示す。いずれの場合も、本明細書に開示される模範的設計法を用いて計算されたスコアが呈する相関は、ロゼッタ法を用いて計算されたスコアの場合を有意に凌駕する。事実上、ロゼッタでは4つの場合のうち3つで、相関は、誤った兆候であるか又は統計的に有意でないかのどちらかである(「X」で表されるパネル)。一方、本明細書に開示される模範的設計法では、相関は、常に適正な兆候であり且つ統計的にかなり有意である(黒色チェックマークで表される)。そのため、本明細書に開示されるTERMベース法により計算された統計エネルギーは、設計品質を表す。

20

30

【図10-4】図10A~10Dは、Bakerらが設計研究で標的とした4つの異なるトポロジーを例示する(3)。図10E~10Fは、本明細書に記載の模範的設計法を用いて計算されたX軸上の各設計の長さを規格化したスコア(そのそれぞれの骨格について)と、Y軸上の各配列の実験的導出安定性スコアと、の相関を示す。散布プロットの点の色はデータ密度を表し、赤色は最高密度を表し、且つ青色は最低密度を表す。平均曲線は、スコアの10進行ウィンドウの安定性スコアを平均することにより得られた丸と共に黒線で示される。図10I~10Lは、それぞれ図10E~10Fの場合と同一のプロットを示すが、ロゼッタ法を用いて計算されたスコアをX軸上に示す。いずれの場合も、本明細書に開示される模範的設計法を用いて計算されたスコアが呈する相関は、ロゼッタ法を用いて計算されたスコアの場合を有意に凌駕する。事実上、ロゼッタでは4つの場合のうち3つで、相関は、誤った兆候であるか又は統計的に有意でないかのどちらかである(「X」で表されるパネル)。一方、本明細書に開示される模範的設計法では、相関は、常に適正な兆候であり且つ統計的にかなり有意である(黒色チェックマークで表される)。そのため、本明細書に開示されるTERMベース法により計算された統計エネルギーは、設計品質を表す。

40

【図10-5】図10A~10Dは、Bakerらが設計研究で標的とした4つの異なるトポロジーを例示する(3)。図10E~10Fは、本明細書に記載の模範的設計法を用いて計算されたX軸上の各設計の長さを規格化したスコア(そのそれぞれの骨格について)と、Y軸上の各配列の実験的導出安定性スコアと、の相関を示す。散布プロットの点の色はデータ密度を表し、赤色は最高密度を表し、且つ青色は最低密度を表す。平均曲線は

50

、スコアの10進行ウィンドウの安定性スコアを平均することにより得られた丸と共に黒線で示される。図10I~10Lは、それぞれ図10E~10Fの場合と同一のプロットを示すが、ロゼッタ法を用いて計算されたスコアをX軸上に示す。いずれの場合も、本明細書に開示される模範的設計法を用いて計算されたスコアが呈する相関は、ロゼッタ法を用いて計算されたスコアの場合を有意に凌駕する。事実上、ロゼッタでは4つの場合のうち3つで、相関は、誤った兆候であるか又は統計的に有意でないかのどちらかである(「X」で表されるパネル)。一方、本明細書に開示される模範的設計法では、相関は、常に適正な兆候であり且つ統計的にかなり有意である(黒色チェックマークで表される)。そのため、本明細書に開示されるTERMベース法により計算された統計エネルギーは、設計品質を表す。

10

【図11】図11A~11Dは、それぞれ、ヒトPin1WWドメインの変異体(PDBエントリー2ZQTを用いてモデリングした)、ヒトYes関連タンパク質65WWドメインの変異体(PDBエントリー4REXを用いてモデリングした)、ピリンヘッドピースヘリカルサブドメインの変異体(残基42~76、PDBエントリー1VIIを用いてモデリングした)、及び周辺サブユニット結合ドメインファミリーメンバーBBLの変異体(モデリングされたPDBエントリー2WXC)に対応する。各データ点は、単一配列変異体に対応し、その熱力学的安定性が本明細書に記載の模範的設計法を用いて計算されたそのスコアに対してプロットされる。熱力学的安定性は、図11A、11C、及び11Dではアンフォールディングの自由エネルギーにより表され、且つ図11Bでは明白な融解温度により表される。最良当てはめ線は、二重平方加重関数によるロバスト線形回帰を用いて生成される。ピアソン相関は、各パネルのタイトルに示される。テューキーフェンスアプローチを用いて同定された外れ値の点は、赤色輪郭で記され相関係数の計算に含まれない。そのため、本明細書に開示されるTERMベース法により計算されたスコアは、熱力学的安定性と相関する。

20

【図12】図12は、新規なPDZ結合モードを設計する手順を示す。すべてのパネルで、N2P2は緑色で示され、且つ結合ペプチド(PDBエントリー2HE4由来)は黒色で示される。図12Aは、完成TERM(シアン色スティック)を示し、一方のセグメントは結合ペプチドにオーバーラップし、且つ他方のセグメントは結合ポケットの外側でN2P2表面領域との接触(赤色で記された接触位置)を形成する。図12Bは、ライブラリーの他のTERMを用いて完成TERMと元の結合ペプチドとを結合する複数の手段を示す。図12Cは、設計配列と共に最終骨格テンプレートを示す。

30

【図13】図13は、N2P2(左側)及びM3P6(右側)に対する設計ペプチドのFPベース阻害アッセイによるプロットを示す。阻害定数はプロット上に示される。

【図14】図14Aは、Rocklinら(3)により標的とされたdenovo設計構造の骨格を示す。図14Bは、この骨格に対して本明細書に開示される模範的設計法を用いて設計された配列の構造モデルを示す(配列はボトムに示される)。全40位置は、いずれかの天然アミノ酸を取り入れることが許容された。図14Cは、標的骨格(緑色)とBakerらによる対応する設計の実験的決定構造(シアン色)(3)との重ね合わせを示す。この構造(PDBコード5UP5)は、構造予測法HHPred(4)により生成された設計配列のトップヒットである。2番ヒットはPDBエントリー1UTAであり、その関連部分(シアン色)は、標的骨格(緑色)上に重ね合わせて図14Dに示される)。そのため、本明細書に開示される模範的設計法は、denovo発生構造の設計に適用可能である。

40

【発明を実施するための形態】

【0024】

本詳細な説明は、特定使用の要件に最適となりうるように当業者が本発明をその多くの形態で適合させ適用しうるように、本発明、その原理、及びその実用的適用を当業者に熟知してもらうことを意図するにすぎない。本説明及びその具体例は、例示目的のみを意図するにすぎない。したがって、本発明は、本特許出願に記載の実施形態に限定されるものではなく、各種変更を行いうる。

50

【 0 0 2 5 】

少なくとも一態様では、本開示は、アミノ酸配列の設計法を提供する。本方法は、標的構造の3次構造モチーフ又は4次構造モチーフなどの適切に定義された構造モチーフ（すなわち、1つ以上のディスジョイント骨格セグメントを含む、構造から切り出された骨格フラグメント）への構造マッチから少なくとも1つの非局所擬エネルギー寄与の値を推測することを含む。ある特定の実施形態では、設計アミノ酸配列は、標的構造の結合パートナーにフォールドするタンパク質である。

【 0 0 2 6 】

ある特定の実施形態では、非局所擬エネルギー寄与は、自己骨格寄与、近傍骨格寄与、対寄与、及び/又はトリプレット（若しくは高次）寄与である。

10

【 0 0 2 7 】

ある特定の実施形態では、非局所擬エネルギー寄与の値は、構造マッチの配列統計から推測される。好ましい実施形態では、構造マッチ内の配列統計は、構造モチーフ内に含有されるアミノ酸位置により駆動される（たとえば、アミノ酸対は、対応する位置対が構造モチーフ内に含有される場合に限り配列統計に影響を及ぼす）。

【 0 0 2 8 】

ある特定の実施形態では、構造マッチは、構造データベースを検索することにより得られる。いくつかのかかる実施形態では、構造データベースは、タンパク質データバンク（PDB）である。他のかかる実施形態では、構造データベースは、たとえば膜貫通タンパク質のみを含有する専用データベースである。

20

【 0 0 2 9 】

ある特定の実施形態では、標的構造は複数の構造モチーフに分解される。いくつかのかかる実施形態では、標的構造はタンパク質であり、且つ構造モチーフは2次及び3次構造モチーフを含む。いくつかのかかる実施形態では、標的構造はタンパク質複合体であり、且つ構造モチーフは2次、3次、及び/又は4次構造モチーフを含む。ある特定の実施形態では、標的構造の所与の残基 i に対する構造モチーフは、自己骨格（たとえば、残基 $i - 2 \sim i + 2$ ）及び近傍骨格（たとえば、 i が接触を形成可能な全残基を取り囲む骨格）を含む。

【 0 0 3 0 】

ある特定の実施形態では、本方法は、構造マッチから少なくとも1つの局所擬エネルギー寄与の値を推測することをさらに含む。いくつかのかかる実施形態では、局所擬エネルギー寄与は、所与のアミノ酸残基 i の二面角及び/又は埋没状態からの寄与である。そのため、ある特定の実施形態では、本方法は、非局所擬エネルギー寄与及び局所擬エネルギー寄与の各々の値のセットを推測することを含む。いくつかのかかる実施形態では、擬エネルギー寄与は、階層：（1）局所擬エネルギー寄与及び（2）非局所擬エネルギー寄与に従って推測される。たとえば、階層は、（ i ）構造マッチ内の単一アミノ酸残基（たとえば、所与の残基 i ）について少なくとも1つの局所擬エネルギー寄与、（ $i i$ ）単一アミノ酸残基を取り囲む骨格のコンティグストレッチ（たとえば、（ $i - n$ ） \sim （ $i + n$ ））、ただし、 i は所与の位置であり、且つ n は制御可能パラメーターである）、（ $i i i$ ）単一アミノ酸残基に配列近接ではなく空間近接する骨格（たとえば、 i が接触を形成可能な全残基を取り囲む骨格）、及び/又は（ $i v$ ）単一設計位置を含む結合残基対、の少なくとも2つを含みうる。他の例として、階層は、（ i ）標的構造の特定設計位置のアミノ酸の骨格二面角、たとえば、 θ 、 ϕ 、及び/又は ψ 、（ $i i$ ）特定設計位置のアミノ酸の埋没状態、（ $i i i$ ）単一アミノ酸残基を取り囲む骨格のコンティグストレッチ、（ $i v$ ）設計位置に配列近接ではなく空間近接する骨格、及び/又は（ v ）設計位置のアミノ酸を含む結合残基対、からの擬エネルギー寄与を含みうる。高次寄与を後で階層に含めることにより、かかる寄与は、低次寄与によりすでに記載されるものを凌駕する（必要な範囲内でのみ）補正子としてのみ使用される。こうして、擬エネルギー寄与は階層で考慮され、前のものによりまだキャプチャーされていないものを記述するためにのみ各次のタイプの寄与が導入される。ある特定の実施形態では、階層中の最初の寄与は、統計ノイ

30

40

50

ズに比較的影響されずに最高信頼度効果が最初にキャプチャーされるように最強配列統計に関連するものであるので、局所寄与及び非局所寄与の階層的考慮は有益である。

【0031】

好ましい実施形態では、高次擬エネルギー寄与は必要と時のみ考慮される（すなわち、両者が観測を等しく記載する場合、低次擬エネルギー寄与のみを含むモデルは、高次寄与をも含むものよりも好ましい）。いくつかのかかる実施形態では、高次擬エネルギー寄与は、低次寄与に対する補正子として作用する。たとえば、対エネルギーは、自己寄与で満身に記載されない配列統計の態様を記述するためにのみ必要とされる。

【0032】

本明細書に開示される各種態様では、構造モチーフとくに3次及び/又は4次構造モチーフに基づくタンパク質設計は、標的構造のフリーズ骨格コンフィギュレーションだけでなく近似コンフィギュレーションのアンサンブルにも適合可能なアミノ酸配列、すなわち、タンパク質構造状態の適切な表現の選択を可能にする。

10

【0033】

A. 計算タンパク質設計

図1は、たとえば標的構造の結合パートナーにフォールドするタンパク質などのアミノ酸配列の設計法100のフロー図を示す。ボックス102に示されるように、標的構造は、複数の2次、3次、又は4次構造モチーフに分解される。かかる分解は、(i)標的構造の結合残基及び/又は(ii)標的構造の残基-骨格作用のグラフ表現によりガイドしうる。たとえば、各2次、3次、又は4次構造モチーフは、標的構造の結合残基を表すグラフの結合サブグラフを表す1アミノ酸残基以上のセットを取り囲んで形成される。ある特定の实施形態では、標的構造は、標的構造の記述に必要なわずかな3次(又は4次)構造モチーフに分解される。

20

【0034】

ボックス104に示されるように、3次(又は4次)構造モチーフが同定されたら、構造マッチを同定するために構造データベースが検索される。構造データベースは、たとえば全PDB又はPDBのフィルタードサブセットでありうる。構造データベースは、たとえばローカル及び/又はリモートメモリーに記憶しうる。構造データベースに記憶されたデータは、いずれかの好適なフォーマットでありうる。ある特定の实施形態では、構造データベースを検索するためにMASTERなどの検索エンジンが利用される。ある特定の实施形態では、検索エンジンは、2次、3次(又は4次)構造モチーフをクエリーとして取り込んで、所与の根平均二乗偏差(RMSD)閾値内でクエリーにマッチする断片のすべてを構造データベースから返す。構造マッチを含有する結果セットは、RMSDの増加などにより順序付けうる。

30

【0035】

ボックス106では、局所擬エネルギー寄与が推測される。局所擬エネルギー寄与は、標的中の所与の位置の単一アミノ酸の骨格二面角(すなわち、 ϕ 角、 ψ 角、若しくは χ 角)又は所与の目標位置の単一アミノ酸の埋没状態に関連しうる。局所擬エネルギー寄与は、PDB内の対応する構造環境の配列統計から推測しうる。

【0036】

ボックス108では、非局所擬エネルギー寄与が推測される。非局所擬エネルギー寄与は、単一設計位置を取り囲む骨格のコンティグストレッチ、単一設計位置に配列近接ではなく空間近接する骨格、及び/又は単一設計位置を含む結合残基対に関連しうる。非局所擬エネルギー寄与は、適切に構築されたTERMへの構造マッチの配列統計から推測しうる。

40

【0037】

ボックス110では、最適アミノ酸配列又はアミノ酸配列のセットが選択されている。最適アミノ酸配列又はアミノ酸配列のセットを選択するために、さまざまな最適化法を使用可能である。たとえば、設計問題への拘束の導入を可能にする整数線形プログラミング(ILP)アプローチを使用しうる(たとえば、配列対称性の拘束、荷電/極性残基の数

50

の拘束、又はいくつかの出発配列と比べて突然変異した残基の制限など)。他の例として、自己無撞着平均場 (SCMF) 又はビリーフプロパゲーション (BP) の技術を使用しうる。さらに他の例として、シミュレーテッドアニーリングモンテカルロ (MC) を使用しうる。

【0038】

図2Aは、構造マッチの配列統計及び環境から擬エネルギー寄与を推測する方法200のフロー図を示す。

【0039】

ボックス202では、局所擬エネルギー寄与が推測される。局所擬エネルギー寄与は、構造マッチ内の単一設計位置の骨格角度、たとえば、角、角、及び/若しくは角、並びに/又は単一設計位置の埋没状態に由来しうる。局所擬エネルギー寄与は、構造マッチの配列統計から推測しうる。

10

【0040】

ボックス204では、少なくとも1つの非局所擬エネルギー寄与が推測される。たとえば、少なくとも1つの非局所擬エネルギー寄与は、単一設計位置を取り囲む骨格のコンティグストレッチに由来しうる。

【0041】

後続の非局所擬エネルギー寄与は、ブロック204により示されるように推測しうる。後続の非局所擬エネルギー寄与は、たとえば、単一設計位置に配列近接ではなく空間近接する骨格、単一設計位置を含む結合残基対、及び/又は単一設計位置を含む残基トリプレットでありうる。

20

【0042】

最適アミノ酸配列又はアミノ酸配列のセットは、ブロック208により示されるように選択される。最適アミノ酸配列又はアミノ酸配列のセットを選択するために、さまざまな最適化法、たとえば、限定されるものではないが以上に記載したILP、SCMF、BP、又はMCのアプローチを使用可能である。

【0043】

ある特定の実施形態では、たとえば図2Aに示される実施形態では、複数の非局所擬エネルギー寄与がブロック204により示されるように推測される。たとえば、複数の非局所擬エネルギー寄与は、(i) 単一設計位置を取り囲む骨格のコンティグストレッチ、(ii) 単一設計位置に配列近接ではなく空間近接する骨格、(iii) 単一設計位置を含む結合残基対、及び/又は(iv) 単一設計位置を含む残基トリプレットに由来しうる。いくつかのかかる実施形態では、上述した寄与(i)~(iv)の各々は特定順序で計算される。しかしながら、かかる実施形態では、後続の寄与は、すでに説明及び観測されたものとの差のみを説明しなければならない。そのため、階層中の後続の寄与は、おそらく徐々に小さくなりうるとともに、記述すべきものが多く残されていない場合には有意性のないものに近づくことさえもありうる。たとえば、後続の寄与は、最終にはゼロ又は実質的にゼロになりうるとともに、その場合にはほとんど計算されなかったようになる。

30

【0044】

図2Bは、構造マッチの配列統計及び環境から擬エネルギー寄与を推測する方法200のフロー図を示す。

40

【0045】

ボックス202では、局所擬エネルギー寄与が推測される。局所擬エネルギー寄与は、構造マッチ内の単一設計位置の骨格角度、たとえば、角、角、及び/若しくは角、並びに/又は単一設計位置の埋没状態に由来しうる。局所擬エネルギー寄与は、構造マッチの配列統計から推測しうる。

【0046】

ボックス204では、第1の非局所擬エネルギー寄与が推測される。たとえば、第1の非局所擬エネルギー寄与は、単一設計位置を取り囲む骨格のコンティグストレッチに由来しうる。

50

【 0 0 4 7 】

決定ダイヤモンド206により示されるように、代替応答は、いずれかの位置優先度が未解明のままであるかに依存して現れる。位置優先度が未解明の場合、ブロック204により示されるように後続の非局所擬エネルギー寄与が推測される。後続の非局所擬エネルギー寄与は、たとえば、単一設計位置に配列近接ではなく空間近接する骨格、単一設計位置を含む結合残基対、及び/又は単一設計位置を含む残基トリプレットでありうる。位置優先度が未解明のままでない場合、ブロック208により示されるように最適アミノ酸配列又はアミノ酸配列のセットが選択される。最適アミノ酸配列又はアミノ酸配列のセットを選択するために、さまざまな最適化法、たとえば、限定されるものではないが以上に記載したILP、SCMF、BP、又はMCのアプローチを使用可能である。

10

【 0 0 4 8 】

図3は、構造マッチの配列統計及びマッチング環境から擬エネルギー寄与を推測する方法300のフロー図を示す。

【 0 0 4 9 】

ボックス302では、局所擬エネルギー寄与が推測される。局所擬エネルギー寄与は、構造マッチ内の単一設計位置の骨格角度、たとえば、角、角、及び/若しくは角、並びに/又は単一設計位置の埋没状態に由来しうる。局所擬エネルギー寄与は、構造マッチの配列統計から推測しうる。ボックス304では、単一設計位置を取り囲む骨格のコンティグストレッチから非局所擬エネルギー寄与が推測される(すなわち自己骨格寄与)。ボックス306、単一設計位置に配列近接ではなく空間近接する骨格から非局所擬エネルギー寄与が推測される(すなわち近傍骨格寄与)。ボックス308では、単一設計位置を含む結合残基対から非局所擬エネルギー寄与が推測される(すなわち結合対寄与)。ボックス310では、単一設計位置を含む残基トリプレットから非局所擬エネルギー寄与(すなわちトリプレット又は他の高次の寄与)が任意選択的に推測される。

20

【 0 0 5 0 】

こうして、擬エネルギー寄与は階層で推測され、前のものによりまだキャプチャーされていないものを記述するためにのみ各次のタイプの寄与が導入される。

【 0 0 5 1 】

図4は、3次/4次構造モチーフに基づく模範的計算タンパク質設計法の模式図を示す。図4に示されるように、標的構造は、(a)グラフGとして示されるその結合残基及び(b)グラフBとして示される残基-骨格作用のグラフ表現によりガイドされて、2次/3次/4次構造モチーフに分解しうる。各構造モチーフへの構造マッチは、構造データベースから同定しうる。標的構造中の配列-構造関係を支配する擬エネルギー寄与の値を導出するために、構造マッチにより示唆される配列アライメントを使用しうる。擬エネルギー寄与の値が与えられれば、コンビナトリアル最適化を用いて最適アミノ酸配列又は最適アミノ酸配列のライブラリーを生成しうる。

30

【 0 0 5 2 】

ある特定の実施形態では、図1~4との関連で記載されるアクティビティの少なくとも一部分は、1つ以上の特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)、ディスクリートロジックを介して、及び/又は1つ以上のサーバー若しくはコンピューター、たとえば、プロセッサとメモリーとを有するコンピューティングデバイスにより実行可能ソフトウェアを用いて、実現しうる。プロセッサは、いずれかのカスタムメイド又は市販のプロセッサ、たとえば、Intel Corporation製のCoreシリーズ、vPro、Xeon、若しくはItaniumプロセッサ、又はAdvanced Micro Devices, Inc.製のPhenom、Athlon、Sempron、若しくはOpteronシリーズプロセッサなどでありうる。プロセッサはまた、ユニゾンで動作する複数の並列又は分散プロセッサを相当しうる。

40

【 0 0 5 3 】

メモリー中のソフトウェアは、1つ以上の個別のプログラム又はアプリケーションを含

50

みうる。プログラムは、論理機能を実現するための実行可能命令の順序付けられたリストを有しうる。ソフトウェアは、サーバー又はコンピューターの好適なオペレーティングシステム、たとえば、Apple, Inc. 製の macOS、OS X、Mac OS X、及び iOS、Microsoft Corporation 製の Windows、Windows Phone、及び Windows 10 Mobile、Unix オペレーティングシステム、Unix デリバティブ（たとえば、BSD 又は Linux）、並びに Google, Inc. 製の Android を含みうる。オペレーティングシステムは、他のコンピュータープログラムの実行を本質的に制御し、スケジューリング、入出力制御、ファイル・データ管理、メモリー管理、及び通信制御、並びに関連サービスを提供する。

【0054】

一般的には、本実施形態に係るコンピュータープログラム製品又はコンピューター可読記憶媒体は、コンピューター可読プログラムコードを内蔵するコンピューター使用可能記憶媒体（たとえば、標準的ランダムアクセスメモリー（RAM）、光ディスク、ユニバーサルシリアルバス（USB）ドライブなど）を含み、コンピューター可読プログラムコードは、以下に記載の方法を実現するプロセッサ（たとえば、オペレーティングシステムに接続して動作するもの）により実行されるように適合化される。これに関連して、プログラムコードは、いずれかの所望の言語で実現しうるとともに、マシンコード、アセンブリコード、バイトコード、解釈可能ソースコードなど（たとえば、C、C++、Java、Actionscript、Objective-C、JavaScript、CSS、XML、及び / 又はその他を介して）として実現しうる。

【0055】

メモリーは、揮発性メモリーエレメント（たとえば、ランダムアクセスメモリー（RAM、たとえば、DRAM、SRAM、SDRAM など））及び不揮発性メモリーエレメント（たとえば、ROM、ハードドライブ、フラッシュドライブ、CDROM など）のいずれか1つ又はそれらの組合せを含みうる。それは電子型、磁気型、光学型、及び / 又は他の型の記憶媒体を組み込みうる。メモリーは、各種コンポーネントが互いに離れて位置するがそれでもなおプロセッサによりアクセスされる分散アーキテクチャーを有しうる。こうした他のコンポーネントは、ネットワーク上の他のどこかに位置するデバイス上又はクラウド配置で常駐しうる。

【0056】

サーバー又はコンピューターは、たとえば、ネットワークを介してデータを送信したり受信したりするトランシーバーを含みうる。トランシーバーは、無線接続及び / 又は有線（たとえば Ethernet）接続を介してデータを受信したり伝送したりするように適合化しうる。トランシーバーは、IEEE 802.11 規格又は他の規格に準拠して機能しうる。より特定的には、トランシーバーは、サーバー又はコンピューターを追加のデバイス又はコンポーネントに通信接続するために1つ以上のセルサイト又はベースステーションを含むワイドエリアネットワークで通信するように構成された WWAN トランシーバーでありうる。さらに、トランシーバーは、サーバー又はコンピューターをローカルエリアネットワーク及び / 又はパーソナルエリアネットワークたとえば Bluetooth ネットワークに接続するように構成された WLAN 及び / 又は WPAN トランシーバーでありうる。

【0057】

A1. 標的構造の分解及び構造マッチの同定

少なくとも一態様では、本開示は、計算タンパク質設計法を提供する。本方法は、標的構造を複数の構造モチーフに分解することを含む。ある特定の実施形態では、標的構造はタンパク質の3次構造である。ある特定の実施形態では、標的構造はタンパク質複合体の4次構造である。

【0058】

ある特定の実施形態では、複数の構造モチーフは、標的構造中の各残基及び各結合残基対をカバーする。たとえば、すべての残基及びすべての結合残基対は、複数の構造モチー

10

20

30

40

50

フ中の少なくとも1つの構造モチーフによりカバーしうる。

【0059】

ある特定の実施形態では、標的構造を複数の構造モチーフに分解する工程は、標的構造中の結合残基を同定することを含む。かかる結合残基は、直接的又は間接的な物理的相互作用を介して又は実験的証拠を介して互いに影響を及ぼすアミノ酸をホスト可能な位置対を見いだすことにより標的構造中で同定しうる。いくつかの実施形態では、所与の構造内の結合残基を同定するために接触度が使用される。

【0060】

たとえば、所与の位置対 i 及び j が接触を形成可能であるかを決定する一方法は、最初に、骨格と衝突しない両位置のすべての可能な回転異性体（すべてのアミノ酸の）を見いだし、次いで、密に接近する非水素原子を有する i 及び j で回転異性体の組合せの加重分率すなわち接触度を計算する。

10

【0061】

接触度を計算する模範式は以下の通りである。

【0062】

【数1】

$$c(i, j) = \frac{\sum_{a \in AA} \sum_{b \in AA} \sum_{r_i \in R_i(a)} \sum_{r_j \in R_j(a)} I_{ij}(r_i, r_j) Pr(a) Pr(b) p(r_i) p(r_j)}{\sum_{a \in AA} \sum_{b \in AA} \sum_{r_i \in R_i(a)} \sum_{r_j \in R_j(a)} Pr(a) Pr(b) p(r_i) p(r_j)}$$

20

【0063】

式中、 $R_i(a)$ は、位置 i のアミノ酸 a の側鎖回転異性体のセットであり（骨格と衝突する回転異性体の廃棄後）、 $I_{ij}(r_i, r_j)$ は、両回転異性体 r_i 及び r_j が互いの存在に強く影響を及ぼす可能性があるか（3 以内の非水素原子対を有するか）の指標となるバイナリー変数であり、 $Pr(a)$ は、構造データベース中のアミノ酸 a の頻度であり、且つ $p(r_i)$ は、回転異性体 r_i の確率である。回転異性体及びその確率は、いずれかの骨格ライブラリーから採用可能である。たとえば、Dunbrackらは骨格依存ライブラリーを開発した（Shapovalov MV & Dunbrack RL, Jr. (2011) A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19(6): 844-858）。構築によって値 $c(i, j)$ は0~1で変動し、大きな数は、より平衡を保って互いに影響を及ぼす位置対に対応する。

30

【0064】

ある特定の実施形態では、設計計算の目的で位置対が結合されているとみなすべきであることを同定するために接触度カットオフが使用される。たとえば、接触度カットオフは、約0.01~約0.2又は代替的に約0.01~0.1又は代替的に約0.01~0.05でありうる。いくつかのかかる実施形態では、接触度カットオフは約0.01である。他のかかる実施形態では、接触度カットオフは約0.05である。

40

【0065】

ある特定の実施形態では、標的構造を複数の構造モチーフに分解する工程は、(i) 標的構造の結合残基及び/又は(ii) 標的構造の残基-骨格作用のグラフ表現によりガイドされる。模範的グラフ G 及び B は図4に示される。グラフ G では、ノードは残基を表し、且つエッジは結合を示し、そしてエッジ加重は任意選択的に結合の強度を表す。グラフ B では、ノードは残基を表し、且つ有向エッジ $a \rightarrow b$ は、 b の骨格が a のアミノ酸選択に影響を及ぼしうることを示す。

50

【 0 0 6 6 】

ある特定の実施形態では、(i) 標的構造の結合残基及び / 又は (i i) 標的構造の残基 - 骨格作用のグラフ表現から導出されるサブグラフは、構造モチーフを同定する。いくつかのかかる実施形態では、複数の構造モチーフ中の各構造モチーフは、結合残基のグラフ表現の結合サブグラフを表す 1 残基以上のセットを取り囲んで形成される。

【 0 0 6 7 】

ある特定の実施形態では、2 次構造モチーフは、残基 (i - n) ~ (i + n) (ただし、n は制御可能パラメーターである) を含むように所与の残基 i を取り囲んで定義され、我々はこれを i のシングルトンモチーフと呼ぶ。たとえば、n は 1 ~ 10、たとえば 1、2、3、4、5、6、7、8、9、又は 10 でありうる。いくつかのかかる実施形態では、n は 1 である。他のかかる実施形態では、n は 2 である。

10

【 0 0 6 8 】

ある特定の実施形態では、3 次又は 4 次構造モチーフは、所与の残基 i を取り囲んで、より好ましくは残基 i の局所骨格 (たとえば、(i - n) ~ (i + n)、ただし、i は所与の位置であり、且つ n は制御可能パラメーターである) を取り囲んで定義される。たとえば、構造モチーフを同定するプロセスは、孤立した残基 i (たとえば 1 ノードサブグラフ) を含むうるとともに、残基 i が有向エッジを有するいくつかの又はすべてのノード (グラフ B に関して、かかるセットは (i) と呼びうる) を考慮しうる。

【 0 0 6 9 】

ある特定の実施形態では、構造モチーフは、標的構造の結合残基のグラフ表現 (たとえばグラフ G) で各エッジに対して定義される。いくつかのかかる実施形態では、構造モチーフは、対をなすさらには関連するシングルトンモチーフの各残基を含む。

20

【 0 0 7 0 】

少なくとも一態様では、本開示は、計算タンパク質設計法を提供する。本方法は、構造データベースで複数の構造モチーフの各々に対する複数の構造マッチを同定することを含む。

【 0 0 7 1 】

ある特定の実施形態では、構造データベースは Protein Data Bank (P D B) である。他のかかる実施形態では、構造データベースは、たとえば膜貫通タンパク質などのある特定のタンパク質のみを含有する専用データベースである。

30

【 0 0 7 2 】

いくつかのかかる実施形態では、高品質フィルターが構造データベースに適用される。たとえば、高品質フィルターは、高品質構造データのみを検索に利用可能であることを保証しうる。模範的高品質フィルターは、特定分解能たとえば 2 . 6 Å 又はそれよりも良好な分解能で X 線結晶解析により解明されたエントリーのみを利用可能にする。いくつかのかかる実施形態では、冗長性フィルターが構造データベースに適用される。たとえば、冗長性フィルターは、不要な繰返しを除去してデータベース検索の計算時間を節約しうる。模範的冗長性フィルターは、過度に冗長な生物学的ユニット、たとえば、すでに含まれる生物学的ユニットに対する特定配列 (%) アイデンティティを有するものを除去する。特定配列 (%) アイデンティティは、たとえば、 > 30 %、> 40 %、> 50 %、> 60 %、> 70 %、> 80 %、又は > 90 % でありうる。

40

【 0 0 7 3 】

ある特定の実施形態では、複数の構造マッチは、構造データベースを検索することにより得られる。構造データベースを検索する模範的検索エンジン MASTER は、Zhou J & Grigoryan G (2014) Rapid search for tertiary fragments reveals protein sequence - structure relationships . Protein Science 24 (4) : 508 - 524 に記載されている。ある特定の実施形態では、クエリーは、小さな根平均二乗偏差 (R M S D) で構造モチーフの骨格にアライメントするデータベースからの骨格部分構造を包含する。いくつかのかかる実施形態では、水素原子は、R M S D を

50

計算するとき除外される。いくつかのかかる実施形態では、検索結果は、RMSDの増加により順序付けられる。

【0074】

ある特定の実施形態では、複数の構造マッチは、ある特定の閾値未満のRMSDを有する構造マッチを含む。模範的なサイズ及び複雑性依存RMSDカットオフ関数は以下の通りである。

【0075】

【数2】

$$RMSD_{cut} = \sigma_m \sqrt{d/N}$$

$$d = N \left(1 - \frac{2}{N(N-1)} \sum_k^{n_k} \sum_{i=1}^{n_k} \sum_{j=i+1}^{n_k} e^{[(i-j)/L]} \right)$$

【0076】

ただし、 d は、モチーフの自由度の有効数であり、 n_k は、モチーフの k 番目のコンティグメントの長さであり、 N は、モチーフの全長（すなわち $N = \sum_k n_k$ ）であり、 L は、相関長（同一ポリペプチド鎖の残基間の空間相関度を記述するパラメーター）であり、且つ σ_m は、プラトーパラメーターである。ある特定の実施形態では、 L は約20であり、且つ σ_m は約1.0である。

【0077】

ある特定の実施形態では、複数の構造マッチは N マッチを含む。ただし、 N は、後続の擬エネルギー計算に必要な所望のサンプルサイズに基づいて選択可能である。たとえば、 N は、少なくとも100、少なくとも200、少なくとも300、少なくとも400、少なくとも500、少なくとも1000、少なくとも1500、又は少なくとも2000でありうる。いくつかのかかる実施形態では、 N は200である。いくつかのかかる実施形態では、 N は1000である。

【0078】

ある特定の実施形態では、構造マッチは、冗長性に関してスクリーニングされる。いくつかのかかる実施形態では、構造マッチは、配列冗長性に関してスクリーニングされる。いくつかのかかる実施形態では、構造マッチは、構造冗長性に関してスクリーニングされる。

【0079】

たとえば、配列冗長性のスクリーニングは、マッチ m 中の各ディスジョイントセグメントを取り囲む局所配列ウィンドウを考慮に入れることと、ニードルマン・ブンシュアルゴリズム及びBLOSUM62マトリックスを介してアライメントすることにより、これらとすでに得られているマッチ（ μ ）の各々の対応する局所配列断片とを比較することと、を含みうる。局所配列ウィンドウは、 m の起源の構造中の15先行残基及び15後続残基を有する対象セグメントとして定義可能である。いくつかのかかる実施形態では、いずれかの局所配列ウィンドウアライメントが約 10^{-3} 未満、代替的に約 10^{-4} 未満、代替的に約 10^{-5} 未満、又は代替的に約 10^{-6} 未満の p 値を有する場合、マッチ m はマッチ μ に対して冗長であるとみなしうる。アライメント p 値は、アライメントスコアに基づいて計算しうるとともに、同一長さの配列（データベースアミノ酸頻度を用いて選択される）間のアライメントが同様に又はより良好にスコア付けされる確率を表しうる。

【0080】

他の例として、構造冗長性のスクリーニングは、対応するクエリー

【0081】

10

20

30

40

50

【数 3】

$$N_m^{near}$$

【0082】

にアライメントする残基のいずれかに結合されるマッチ m の起源の構造中のすべての残基を同定することと、その近接残基のうちいくつが μ の近接残基に良好にアライメントするか（特定閾値未満の骨格 RMSD を有するものとして定義される）を m 及び μ の両方がクエリーモチーフに最適にアライメントされるとき配向で計算することにより、マッチ m とすでに得られているマッチ μ の各々を比較することと、を含みうる。これとの関連で、マッチ m とすでに得られているマッチ μ との間の構造環境類似性を計算する模範的関数は以下の通りである。

10

【0083】

【数 4】

$$S_{m,\mu} = N_{m,\mu}^{near} / (0.5 \cdot [N_m^{near} + N_\mu^{near}] + 1)$$

20

【0084】

いくつかのかかる実施形態では、 $S_{m,\mu}$ が特定カットオフ超である場合、マッチ m はマッチ μ に対して冗長であるとみなしうる。たとえば、特定カットオフは、少なくとも 0.1、少なくとも 0.2、又は少なくとも 0.3 でありうる。いくつかのかかる実施形態では、特定カットオフは 0.2 である。

【0085】

A2. 擬エネルギー寄与の計算

少なくとも一態様では、本開示は、3次又は4次構造モチーフへの複数の構造マッチの各々で配列-構造関係への少なくとも1つの非局所エネルギー寄与の値を推測する方法を提供する。

30

【0086】

ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置を取り囲む骨格のコンティグストレッチに由来する（すなわち自己骨格寄与）。ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置に配列近接ではなく空間近接する骨格に由来する（すなわち近傍骨格寄与）。ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与は、複数の構造モチーフの1つ内の結合残基対に由来する（すなわち対寄与）。ある特定の実施形態では、少なくとも1つの非局所エネルギー寄与の値は、構造モチーフ及びその構造マッチを解析することにより設計計算を実施しつつオンザフライで計算される。

40

【0087】

ある特定の実施形態では、本方法は、複数の構造マッチの各々を用いて配列-構造関係への少なくとも1つの局所エネルギー寄与の値を取得することをさらに含む。ある特定の実施形態では、少なくとも1つの局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置の骨格角度に由来する。いくつかのかかる実施形態では、骨格角度は 角、角、又は 角である。ある特定の実施形態では、少なくとも1つの局所エネルギー寄与は、複数の構造モチーフの1つ内の単一設計位置の埋没状態に由来する。ある特定の実施形態では、少なくとも1つの局所エネルギー寄与の値は、データベースに基づいてあらかじめ計算される。

50

【 0 0 8 8 】

ある特定の実施形態では、本方法は、エネルギー寄与の階層に従って複数の構造マッチの各々を用いて配列 - 構造関係へのエネルギー寄与の値のセットを逐次推測することを含み、階層は、

i . 複数の構造モチーフの 1 つ内の単一設計位置について少なくとも 1 つの局所エネルギー寄与、

i i . 単一設計位置を取り囲む骨格のコンティグストレッチ、

i i i . 単一設計位置に配列近接ではなく空間近接する骨格、

i v . 単一設計位置を含む結合残基対、及び

v . 単一設計位置を含む残基トリプレット

の少なくとも 2 つを含む。

10

【 0 0 8 9 】

A 2 A . 骨格角度

ある特定の実施形態では、本方法は、少なくとも 1 つの局所エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、局所擬エネルギー寄与は、骨格（ファイ）及び（プサイ）二面角の異なるアミノ酸の傾向を記述する。いくつかのかかる実施形態では、骨格及び二面角の異なるアミノ酸の傾向を記述する擬エネルギー寄与は、エネルギー寄与の階層の第 1 層である。

【 0 0 9 0 】

ある特定の実施形態では、及び骨格角度からの擬エネルギー寄与は、/ 位相空間をピン（たとえば $10^\circ \times 10^\circ$ のピン）にスプリットすることと、その及び角の値に基づいて構造データベース中の各残基を対応するピンに帰属することと、により推測される。骨格二面角ピン B_i に関連するアミノ酸 a の擬ポテンシャルの値を計算する模範的関数は以下の通りである。

20

【 0 0 9 1 】

【数 5】

$$E^{\varphi\psi}(a|B_i^{\varphi\psi}) = -\ln(f(a, B_i^{\varphi\psi}))$$

30

【 0 0 9 2 】

ただし、 $f(a, B_i)$ は、構造データベース中のタンパク質内でアミノ酸 a がこのピンに見いだされる頻度であり、

【 0 0 9 3 】

【数 6】

$$f(a, B_i^{\varphi\psi}) = N(a, B_i^{\varphi\psi}) / \sum_{aa=1}^{20} N(aa, B_i^{\varphi\psi})$$

40

【 0 0 9 4 】

$N(aa, B_i)$ は、アミノ酸 aa がピン B_i に見いだされる回数である。

【 0 0 9 5 】

ある特定の実施形態では、本方法は、少なくとも 1 つの局所エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、局所擬エネルギー寄与は、異なる骨格（オメガ）二面角に対するアミノ酸の優先度を記述する。いくつかのかかる実施形態で

50

は、異なる骨格 二面角に対するアミノ酸の優先度を記述する擬エネルギー寄与は、エネルギー寄与の階層の第2層である（たとえば、局所擬エネルギー寄与が骨格（ファイ）及び（プサイ）二面角の異なるアミノ酸の傾向を記述することを熟慮した後でのみ考慮される）。

【0096】

ある特定の実施形態では、二面角からの擬エネルギー寄与は、位相空間をビンにスプリットすることと、その角の値に基づいて構造データベース中の各残基を対応するビンに帰属することと、により推測される。角は、部分二重結合特性を有するペプチド結合の周りで定義されるので、角は典型的には平面的であり、最も一般的には180°近くの値を有するが（トランスペプチド結合）、一般に（ただし排他的ではない）Pro又はGlyアミノ酸で約0°の値も現れる（シスペプチド結合）。そのため、いくつかのかかる実施形態では、本方法は、角の非一様ビンングを含み、ビン幅は少なくとも1°であるが、各ビンに十分な数の構造データベース残基を有するのに必要とされる程度に大きい。

10

【0097】

角ビン B_i に関連するアミノ酸 a の擬ポテンシャルの値を計算する模範的関数は以下の通りである。

【0098】

【数7】

$$E^\omega(a|B_i^\omega) = -\ln\left(\frac{N(a, B_i^\omega) + \epsilon_\omega}{N_e(a, B_i^\omega) + \epsilon_\omega}\right)$$

20

【0099】

式中、 $N(a, B_i)$ は、アミノ酸 a がビン B_i に見いだされる回数であり、 $N_e(a, B_i)$ は、既知の擬エネルギー寄与（たとえば / エネルギー）に基づいて、 a がビンに見いだされると予想される回数であり、且つ ϵ_ω は、低占有ビンからの過剰統計ノイズを防止する擬カウントとして作用する。いくつかのかかる実施形態では、 ϵ_ω は1である。

30

【0100】

$N_e(a, B_i)$ の模範的関数は以下の通りである。

【0101】

【数8】

$$N_e(a, B_i^\omega) = \sum_{k \in B_i^\omega} \frac{\exp\left(-E^{\varphi\psi}(a|B^{\varphi\psi}(k))\right)}{\sum_{aa \in AA} \exp\left(-E^{\varphi\psi}(aa|B^{\varphi\psi}(k))\right)}$$

40

【0102】

式中、外側の和は、ビン B_i に分類されるすべての天然残基にわたり、内側の和は、集合 AA で表されるすべての天然アミノ酸にわたり、且つ B () は、残基 a が分類される / ビンである。内側の部分は、ビン中の各残基の / 環境で a を観測する予想確率を表す（すべての可能なアミノ酸にわたる）。以上の式の予想による補正は、 E が E に対する補正子としてのみ作用し、データでまだ説明されていないもののみを説明することを保証する。

【0103】

50

A 2 B . 埋没状態

ある特定の実施形態では、本方法は、少なくとも1つの局所エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、局所擬エネルギー寄与は、残基の一般的环境（すなわち埋没状態）に由来する。いくつかのかかる実施形態では、残基の埋没状態からの擬エネルギー寄与は、エネルギー寄与の階層の後続寄与である（たとえば、骨格及び二面角の異なるアミノ酸の傾向を記述する局所擬エネルギー寄与と、異なる骨格二面角に対するアミノ酸の優先度を記述する局所擬エネルギー寄与と、を熟慮した後でのみ考慮される）。

【0104】

ある特定の実施形態では、埋没状態からの擬エネルギー寄与は、構造データベース中の全残基について環境記述子 e を計算することと、 e に従って残基をビンングすることと、により推測される。単一体（自己）寄与として残基の埋没状態からの寄与をキャプチャーするために、環境記述子は配列非依存環境記述子でありうる。

10

【0105】

環境ビン B_i^e に関連するアミノ酸 a の擬ポテンシャルの値を計算する模範的関数は以下の通りである。

【0106】

【数9】

$$E^e(a|B_i^e) = -\ln\left(\frac{N(a, B_i^e) + \varepsilon_e}{N_e(a, B_i^e) + \varepsilon_e}\right)$$

20

【0107】

式中、 $N(a, B_i^e)$ は、アミノ酸 a がビン B_i^e に見いだされる回数であり、 $N_e(a, B_i^e)$ は、既知の擬エネルギー寄与（たとえば / エネルギー及び エネルギー）に基づいて、 a がビンに見いだされると予想される回数であり、且つ ε_e は、低占有ビンからの過剰統計ノイズを防止する擬カウントとして作用する。いくつかのかかる実施形態では、 ε_e は1である。

30

【0108】

$N_e(a, B_i^e)$ の模範的関数は以下の通りである。

【0109】

【数10】

$$N_e(a, B_i^e) = \sum_{k \in B_i^e} \frac{\exp\left(-E^{\varphi\psi}(a|B^{\varphi\psi}(k)) - E^\omega(a|B^\omega(k))\right)}{\sum_{aa \in AA} \exp\left(-E^{\varphi\psi}(aa|B^{\varphi\psi}(k)) - E^\omega(aa|B^\omega(k))\right)}$$

40

【0110】

式中、外側の和は、環境ビン B_i^e に帰属されるすべての天然残基にわたり、且つ B () は、残基 がマップされる ビンである。以上の式の予想による補正は、階層の早期に考慮された擬エネルギー寄与（たとえば、 E 及び / 又は E ）によりすでに説明されるものに対する補正子としてのみ E^e が作用することを保証する。

【0111】

さまざまな配列非依存環境記述子 e を使用しうる。一実施形態では、配列非依存環境記

50

述子は、残基を取り囲む容積が非占有でその回転異性体に利用可能な程度を決定するために、所与の位置及びその周囲のすべての天然アミノ酸のすべての可能な回転異性体を考慮する「残基自由度」でありうる。所与の残基 i の自由度の模範的関数 $F(i)$ は以下の通りである。

【 0 1 1 2 】

【数 1 1】

$$F(i) = \sqrt{\frac{V_{i,0.5}^2 + V_{i,2}^2}{2}}$$

10

【 0 1 1 3 】

式中

【 0 1 1 4 】

【数 1 2】

$$V_{i,\tau} = \frac{\sum_{a \in AA} \sum_{r_i \in R_i(a)} I(p_c(r_i) < \tau)}{\sum_{a \in AA} |R_i(a)|}, \text{ 及び}$$

20

$$p_c(r_i) = \sum_{j \neq i} \sum_{b \in AA} \sum_{r_j \in R_j(b)} I_{ij}(r_i, r_j) Pr(b) p(r_j)$$

【 0 1 1 5 】

式中、 $R_i(a)$ は、位置 i のアミノ酸 a の側鎖回転異性体のセットであり（骨格と衝突する回転異性体の廃棄後）、 $I_{ij}(r_i, r_j)$ は、両回転異性体 r_i 及び r_j が互いの存在に強く影響を及ぼす可能性があるか（3 以内の非水素原子対を有するか）の指標となるバイナリー変数であり、 $Pr(a)$ は、構造データベース中のアミノ酸 a の頻度であり、且つ $p(r_i)$ は、回転異性体 r_i の確率であり、また、 $p_c(r_i)$ は、回転異性体 r_i の「衝突確率質量」（すなわち、他の位置の回転異性体とどの程度衝突する可能性があるか）である。

30

【 0 1 1 6 】

A 2 C . 自己骨格

ある特定の実施形態では、本方法は、少なくとも 1 つの非局所擬エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、非局所擬エネルギー寄与は、所与の位置で単一設計位置を取り囲む骨格のコンティグストレッチに由来する（すなわち自己骨格寄与）。いくつかのかかる実施形態では、自己骨格寄与は、エネルギー寄与の階層の後続寄与である（たとえば、1 つ以上の局所擬エネルギー寄与を考慮した後でのみ考慮される）。

40

【 0 1 1 7 】

ある特定の実施形態では、自己骨格寄与は、位置 p を取り囲む骨格の局所コンティグストレッチが、 / 、 、及び埋没状態の優先度によりすでにキャプチャーされるものを超えて、そのアミノ酸優先度をどのようにモジュレートするかをキャプチャーする。

【 0 1 1 8 】

ある特定の実施形態では、自己骨格寄与は、位置 p 及びその周囲コンティグ骨格断片を含む構造モチーフ T_p を標的構造から切り出すことと、構造データベースで T_p への構造マッチを同定することと、により推測される。構造マッチのセットを M_p という。

【 0 1 1 9 】

50

位置 p のアミノ酸 a の自己骨格寄与の値を計算する模範的関数は以下の通りである。

【 0 1 2 0 】

【数 1 3 】

$$E_p^o(a|D) = -\ln \left(\frac{N(a, M_p) + \varepsilon_o}{N_e(a, M_p) + \varepsilon_o} \right)$$

【 0 1 2 1 】

10

式中、 $N(a, M_p)$ は、アミノ酸 a が構造マッチ M_p のセット内の p に対応する位置で観測される回数であり、且つ $N_e(a, M_p)$ は、既知の擬エネルギー寄与（たとえば、 E^{ψ} 、 E^{ω} 、及び E^e 又は環境エネルギー）に基づいて a がこの位置にあると予想される回数であり、且つ ε_o は、擬カウントとして作用する。いくつかのかかる実施形態では、 ε_o は 1 である。

【 0 1 2 2 】

$N_e(a, M_p)$ の模範的関数は以下の通りである。

【 0 1 2 3 】

【数 1 4 】

20

$$N_e(a, M_p) = \sum_{m \in M_p} \frac{\exp \left(-E^{\psi}(a|B^{\psi}(m_p)) - E^{\omega}(a|B^{\omega}(m_p)) - E^e(a|B^e(m_p)) \right)}{\sum_{aa \in AA} \exp \left(-E^{\psi}(aa|B^{\psi}(m_p)) - E^{\omega}(aa|B^{\omega}(m_p)) - E^e(aa|B^e(m_p)) \right)}$$

【 0 1 2 4 】

式中、外側の和は、 M_p 中のマッチにわたり、 m_p は、 T_p 中の位置 p にアライメントするマッチ m 中の残基であり、且つ $B^e(m_p)$ は、マッチ m の起源の構造中のその周囲に基づいて m_p が属する環境ピンである。

30

【 0 1 2 5 】

A 2 D . 近傍骨格

ある特定の実施形態では、本方法は、少なくとも 1 つの非局所擬エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、非局所擬エネルギー寄与は、所与の位置で単一設計位置に配列近接ではなく空間近接する骨格に由来する（すなわち近傍骨格寄与）。いくつかのかかる実施形態では、近傍骨格寄与は、エネルギー寄与の階層の後続寄与である（たとえば、1 つ以上の局所擬エネルギー寄与及び自己骨格寄与を考慮した後でのみ考慮される）。

【 0 1 2 6 】

ある特定の実施形態では、近傍骨格寄与は、位置 p に配列近接ではなく密に空間近接する骨格セグメントの存在によりもたらされる位置 p のアミノ酸優先度のいずれかのさらなるモジュレーションをキャプチャーする。

40

【 0 1 2 7 】

ある特定の実施形態では、近傍骨格寄与は、位置 p 、その周囲コンティグ骨格セグメント、及び p に密に空間（配列ではない）近接する骨格セグメントを含む構造モチーフ $T'_{p,t}$ 、 t を標的構造から切り出すことと、構造データベースで $T'_{p,t}$ への構造マッチを同定することと、により推測され、下付き文字 t は、複数のかかる構造モチーフが可能であることを表す。構造マッチのセットを $M'_{p,t}$ という。

【 0 1 2 8 】

$T'_{p,t}$ 中のアミノ酸 a の近傍骨格寄与の値を計算する模範的関数は以下の通りである。

50

【 0 1 2 9 】

【 数 1 5 】

$$E'_{p,t}(a|D) = -\ln \left(\frac{N(a, M'_{p,t}) + \varepsilon_n}{N_e(a, M'_{p,t}) + \varepsilon_n} \right)$$

【 0 1 3 0 】

式中、 $N(a, M'_{p,t})$ は、アミノ酸 a が構造マッチ $M'_{p,t}$ のセット内の p に対応する位置で観測される回数であり、且つ $N_e(a, M'_{p,t})$ は、既知の擬エネルギー寄与（たとえば、 ψ 、 ω 、環境、及び ν 又は自己骨格エネルギー）に基づいて a がこの位置にあると予想される回数であり、且つ ε_n は、擬カウントとして作用する。いくつかのかかる実施形態では、 ε_n は 1 である。

10

【 0 1 3 1 】

$N_e(a, M'_{p,t})$ の模範的関数は以下の通りである。

【 0 1 3 2 】

【 数 1 6 】

$$N_e(a, M'_{p,t}) = \sum_{m \in M'_{p,t}} \frac{\exp(-E^{\psi}(a|B^{\psi}(m_p)) - E^{\omega}(a|B^{\omega}(m_p)) - E^{\nu}(a|B^{\nu}(m_p)) - E_p^{\nu}(a|m))}{\sum_{aa \in \mathcal{A}^2} \exp(-E^{\psi}(aa|B^{\psi}(m_p)) - E^{\omega}(aa|B^{\omega}(m_p)) - E^{\nu}(aa|B^{\nu}(m_p)) - E_p^{\nu}(aa|m))}$$

20

【 0 1 3 3 】

式中、外側の和は、 $M'_{p,t}$ 中のマッチにわたり、及び

【 0 1 3 4 】

【 数 1 7 】

$$E_p^{\nu}(a|m)$$

30

【 0 1 3 5 】

は、マッチ m の起源の構造に基づいて残基 m_p 中のアミノ酸 a の自己骨格擬エネルギーを表す。

【 0 1 3 6 】

A 2 E . 対

ある特定の実施形態では、本方法は、少なくとも 1 つの非局所擬エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、非局所擬エネルギー寄与は、標的構造中の結合残基対 (p, q) に由来する（すなわち対擬エネルギー寄与）。いくつかのかかる実施形態では、対寄与は、エネルギー寄与の階層の後続寄与である（たとえば、1 つ以上の局所擬エネルギー寄与、自己骨格寄与、及び ν 又は近傍骨格寄与を考慮した後でのみ考慮される）。

40

【 0 1 3 7 】

ある特定の実施形態では、対寄与は、位置 p 及び q を含む構造モチーフ

【 0 1 3 8 】

【 数 1 8 】

$$T''_{p,q}$$

50

【 0 1 3 9 】

を標的構造から切り出すことと、構造データベースで

【 0 1 4 0 】

【数 1 9 】

$$T''_{p,q}$$

【 0 1 4 1 】

への構造マッチを同定することと、により推測される。構造マッチのセットを

10

【 0 1 4 2 】

【数 2 0 】

$$M''_{p,q}$$

【 0 1 4 3 】

という。

【 0 1 4 4 】

それぞれ

20

【 0 1 4 5 】

【数 2 1 】

$$T''_{p,q}$$

【 0 1 4 6 】

の位置 p 及び q のアミノ酸 a 及び b の対寄与の値を計算する模範的関数は以下の通りである。

【 0 1 4 7 】

30

【数 2 2 】

$$E''_{p,q}(a, b|D) = -\ln \left(\frac{N(a, b, M''_{p,q}) + \varepsilon_p}{N_e(a, b, M''_{p,q}) + \varepsilon_p} \right)$$

【 0 1 4 8 】

式中、

【 0 1 4 9 】

【数 2 3 】

40

$$N(a, b, M''_{p,q})$$

【 0 1 5 0 】

は、アミノ酸 a 及び b が構造マッチ

【 0 1 5 1 】

【数 2 4 】

50

$$M''_{p,q}$$

【 0 1 5 2 】

のセット内の p 及び q に対応する位置で観測される回数であり、且つ

【 0 1 5 3 】

【 数 2 5 】

$$N_e(a,b,M''_{p,q})$$

10

【 0 1 5 4 】

は、既知の擬エネルギー寄与（たとえば、 Δ_p 、 Δ_q 、環境、自己骨格、及び Δ_p 又は近傍骨格エネルギー）に基づいて、(a, b) 対がこれらの位置にあると予想される回数であり、且つ Δ_p は、擬カウントとして作用する。いくつかのかかる実施形態では、 Δ_p は 1 である。

【 0 1 5 5 】

【 数 2 6 】

20

$$N_e(a,b,M''_{p,q})$$

【 0 1 5 6 】

の模範的関数は以下の通りである。

【 0 1 5 7 】

【 数 2 7 】

$$N_e(a,b,M''_{p,q}) = \sum_{m \in M''_{p,q}} \frac{\exp(-E_{lo}(a|m_p) - \Delta_p(a, M''_{p,q}))}{\sum_{aa \in AA} \exp(-E_{lo}(aa|m_p) - \Delta_p(a, M''_{p,q}))} \times \frac{\exp(-E_{lo}(b|m_q) - \Delta_q(a, M''_{p,q}))}{\sum_{aa \in AA} \exp(-E_{lo}(aa|m_q) - \Delta_q(a, M''_{p,q}))}$$

30

【 0 1 5 8 】

式中、簡潔さを期して、 $E_{lo}(a|m_p)$ は、マッチ m の位置 p にアライメントされた位置のアミノ酸 a に関連するこれまでに考慮されたすべての低次寄与の合計擬エネルギーを表し、

【 0 1 5 9 】

【 数 2 8 】

$$E_{lo}(a|m_p) = E^{\varphi\psi}(a|B^{\varphi\psi}(m_p)) + E^{\omega}(a|B^{\omega}(m_p)) + E^e(a|B^e(m_p)) + E_p^o(a|m) + \sum_t E'_{p,t}(a|m)$$

40

【 0 1 6 0 】

且つ

【 0 1 6 1 】

【 数 2 9 】

50

$$\Delta_p(a, M''_{p,q})$$

【 0 1 6 2 】

は、構造モチーフの個別結合位置の周辺アミノ酸分布を保存するために含めうる任意選択的調整エネルギーである。

【 0 1 6 3 】

A 2 F . トリプレット

ある特定の実施形態では、本方法は、少なくとも1つの非局所擬エネルギー寄与の値を推測することを含む。いくつかのかかる実施形態では、非局所擬エネルギー寄与は、標的構造中の残基トリプレット (p , q , r) に由来する (すなわちトリプレット擬エネルギー寄与) 。いくつかのかかる実施形態では、トリプレット寄与は、エネルギー寄与の階層の後続寄与である (たとえば、1つ以上の局所擬エネルギー寄与、自己骨格寄与、近傍骨格寄与、及び / 又は対寄与を考慮した後でのみ考慮される) 。

10

【 0 1 6 4 】

ある特定の実施形態では、トリプレット寄与は、位置 p 、 q 、 及び r を含む構造モチーフ

【 0 1 6 5 】

【数 3 0 】

$$T'''_{p,q,r}$$

20

【 0 1 6 6 】

を標的構造から切り出すことと、構造データベースで

【 0 1 6 7 】

【数 3 1 】

$$T'''_{p,q,r}$$

30

【 0 1 6 8 】

への構造マッチを同定することと、により推測される。構造マッチのセットを

【 0 1 6 9 】

【数 3 2 】

$$M'''_{p,q,r}$$

【 0 1 7 0 】

という。

【 0 1 7 1 】

それぞれ

【 0 1 7 2 】

【数 3 3 】

$$T'''_{p,q,r}$$

40

【 0 1 7 3 】

50

の位置 p 、 q 、及び r のアミノ酸 a 、 b 、及び c の対寄与の値を計算する模範的関数は以下の通りである。

【 0 1 7 4 】

【数 3 4】

$$E_{p,q,r}'''(a,b,c|D) = -\ln \left(\frac{N(a,b,c,M_{p,q,r}''') + \varepsilon_t}{N_e(a,b,c,M_{p,q,r}''') + \varepsilon_t} \right)$$

10

【 0 1 7 5 】

式中、

【 0 1 7 6 】

【数 3 5】

$$N(a,b,c,M_{p,q,r}''')$$

【 0 1 7 7 】

は、トリプレット (a 、 b 、 c) が構造マッチ

20

【 0 1 7 8 】

【数 3 6】

$$M_{p,q,r}'''$$

【 0 1 7 9 】

のセット内の (p 、 q 、 r) に対応する位置で観測される回数であり、且つ

【 0 1 8 0 】

【数 3 7】

30

$$N_e(a,b,c,M_{p,q,r}''')$$

【 0 1 8 1 】

は、(a 、 b 、 c) トリプレットが既知の擬エネルギー寄与 (たとえば、 ΔE 、 ΔE_{self} 、 $\Delta E_{neighbor}$ 、 $\Delta E_{backbone}$ 、 $\Delta E_{sidechain}$ 、及び ΔE_{loop} 又は対エネルギー) に基づいてこれらの位置にあると予想される回数であり、且つ ε_t は、擬カウントとして作用する。いくつかのかかる実施形態では、 ε_t は 1 である。

【 0 1 8 2 】

【数 3 8】

40

$$N_e(a,b,c,M_{p,q,r}''')$$

【 0 1 8 3 】

の模範的関数は以下の通りである。

【 0 1 8 4 】

【数 3 9】

50

$$N_s(a, b, c, M'''_{p,q,r}) = \sum_{m \in \mathcal{M}'_{p,q,r}} \frac{\exp\{-E_{10}(a, b, c | m_{p,q,r}) - \Delta_p(a, M'''_{p,q,r}) - \Delta_q(a, M'''_{p,q,r}) - \Delta_r(c, M'''_{p,q,r}) - \Delta_{p,q}(a, b, M'''_{p,q,r}) - \Delta_{p,r}(a, c, M'''_{p,q,r}) - \Delta_{q,r}(b, c, M'''_{p,q,r})\}}{\sum_{\alpha, \beta, \gamma \in \mathcal{A}} \exp\{-E_{10}(a, b, c | m_{p,q,r}) - \Delta_p(\alpha, M'''_{p,q,r}) - \Delta_q(\alpha, M'''_{p,q,r}) - \Delta_r(c, M'''_{p,q,r}) - \Delta_{p,q}(\alpha, b, M'''_{p,q,r}) - \Delta_{p,r}(\alpha, c, M'''_{p,q,r}) - \Delta_{q,r}(b, c, M'''_{p,q,r})\}}$$

【 0 1 8 5 】

式中、簡潔さを期して、 $E_{10}(a, b, c | m_{p,q,r})$ は、マッチ m の位置 p 、 q 、及び r にアライメントされた位置のアミノ酸 a に関連するこれまでに考慮されたすべての低次寄与の合計擬エネルギーを表し、

10

【 0 1 8 6 】

【数 4 0】

$$E_{10}(a, b, c | m_{p,q,r}) = \sum_{x=\{p,q,r\}} [E^{**}(aa_x | B^{**}(m_x)) + E^*(aa_x | B^*(m_x)) + E^e(aa_x | B^e(m_x)) + E^i(aa_x | m) + \sum_t E'_{x,t}(aa_x | m)] + \sum_{\substack{x,y=\{p,q,r\} \\ x \neq y}} E''_{x,y}(aa_x, aa_y | m)$$

【 0 1 8 7 】

及び

20

【 0 1 8 8 】

【数 4 1】

$$\Delta_{p,q}(a, b, M'''_{p,q,r})$$

【 0 1 8 9 】

は、

【 0 1 9 0 】

【数 4 2】

30

$$T'''_{p,q,r}$$

【 0 1 9 1 】

中の位置対のペアワイズアミノ酸分布を拘束するために含めうる任意選択的調整エネルギーである。

【 0 1 9 2 】

A 3 . タンパク質最適化

少なくとも一態様では、本開示は、標的構造の結合パートナーにフォールド可能なアミノ酸配列又はアミノ酸配列のライブラリーを決定する方法を提供する。アミノ酸配列のライブラリーは、たとえば、多くとも約 50%、代替的に多くとも約 60%、代替的に多くとも約 70%、代替的に多くとも約 80%、又は代替的に多くとも約 90% の配列同一性を互いに有するアミノ酸配列のセットを含みうる。ある特定の実施形態では、アミノ酸配列のセットは、コアジェネリック配列の変異体を含む。

40

【 0 1 9 3 】

ある特定の実施形態では、標的構造の結合パートナーにフォールド可能なアミノ酸配列又はアミノ酸配列のライブラリーを決定するために最適化アプローチが使用される。たとえば、擬エネルギー寄与のすべての値を計算し、任意選択的に自己、対、及び可能な限り高次の擬エネルギー寄与の表に体系化したら、最適化アプローチのホストを用いて最適ア

50

ミノ酸配列を推測可能である。ある特定の実施形態では、整数線形プログラミング（ILP）アプローチが使用される。ILPアプローチは、設計問題への拘束の導入を可能にする（たとえば、配列対称性の拘束、荷電/極性若しくは疎水性残基の数の拘束、又はいくつかの出発配列と比べて突然変異した残基の制限）。ある特定の実施形態では、代替最適化法、たとえば、自己無撞着平均場（SCMF）又はシミュレーテッドアニーリングモンテカルロ（MC）が使用される。ある特定の実施形態では、絶対グローバル最適配列の同定は必要でなく、いずれかの近似～最適配列があれば十分である。

【0194】

B. タンパク質発現

ある特定の態様では、本明細書に記載の方法の産物は、発現さらに *in vitro* 及び/又は *in vivo* 実験手順を用いた最適化が推奨されるアミノ酸配列又はアミノ酸配列のライブラリー若しくはセットである。

10

【0195】

さらなる態様では、本開示は、本明細書に提供される計算設計タンパク質をコードする核酸配列を提供する。かかる核酸配列は、コードタンパク質の発現及び/又は精製を促進するのに有用な追加の配列、たとえば、限定されるものではないが、ポリA配列、改変コザック配列、及びエピトープタグをコードする配列、移出シグナル、及び分泌シグナル、核局在化シグナル、及び形質膜局在化シグナルをさらに含む。

【0196】

ある特定の実施形態では、核酸配列は、ベクター（たとえば、プラスミド、コスミド、ウイルス、バクテリオファージ、又は遺伝子工学で従来から使用されている他のベクター）に含有される。いくつかのかかる実施形態では、ベクターは、好適な宿主細胞でコード領域の適正発現を可能にする発現制御エレメントを含む。計算設計タンパク質をコードする核酸配列に作動可能に連結された「制御エレメント」は、計算設計タンパク質の発現を誘発可能なさらなる核酸配列である。たとえば、制御エレメントは、さまざまな構成的プロモーター、たとえば、限定されるものではないが、CMV、SV40、RSV、若しくはアクチン、又は誘導性プロモーター、たとえば、限定されるものではないが、テトラサイクリン若しくはステロイドにより駆動されるプロモーター、のいずれかを含む。制御エレメントは、その発現をダイレクトする機能がある限り、タンパク質をコードする核酸配列とのコンティグである必要はない。そのため、たとえば、プロモーター配列と核酸配列との間に非翻訳であるが転写される介在配列を存在させることが可能であり、それでもなおプロモーター配列は、コード配列に「作動可能に連結された」とみなすことが可能である。他のかかる制御配列としては、限定されるものではないが、開始シグナル、ポリAデニル化シグナル、終止シグナル、及びリボソーム結合部位が挙げられる。ある特定の実施形態では、ベクターは、好適な宿主細胞内で且つ好適な条件下でベクターの選択を可能にするマーカー遺伝子などのさらなる遺伝子を含む。核酸分子を含むベクターの構築のための、適切に選択された宿主細胞内へのベクターの導入のための、又は核酸分子の発現を誘発若しくは達成するための、核酸分子の構築方法は、当技術分野で周知である。

20

30

【0197】

他の一態様では、本開示は、本明細書に開示される核酸又はベクターを含む宿主細胞を提供する。宿主細胞は、原核細胞又は真核細胞のどちらかでありうる。宿主細胞は、一過的又は安定的にトランスフェクト可能である。原核細胞及び真核細胞への発現ベクターのかかるトランスフェクションは、当技術分野で公知のいずれかの技術、たとえば、限定されるものではないが、標準的細菌トランスフォーメーション、リン酸カルシウム共沈、電気穿孔、又はリボソーム媒介、DEAEデキストラン媒介、ポリカチオン媒介、若しくはウイルス媒介のトランスフェクションを介して達成可能である。

40

【0198】

さらなる態様では、本開示は、計算設計タンパク質の生成方法を提供する。本方法は、(a) タンパク質の発現を助長する条件下でタンパク質をコードする核酸配列を含む宿主細胞を培養する工程と、(b) 任意選択的に発現タンパク質を回収する工程と、を含む。

50

そのため、ある特定の実施形態では、計算設計タンパク質の生成方法は、少なくとも1つのアミノ酸配列を設計及び選択することと、発現系でアミノ酸配列を発現させることにより計算設計タンパク質を生成することと、を含む。ある特定の実施形態では、アミノ酸配列は、標的構造の結合パートナーにフォールド可能なタンパク質である。

【0199】

いくつかのかかる実施形態では、本方法は、少なくとも1つの候補アミノ酸配列を *in silico* で発生させることと、候補アミノ酸配列をコードする核酸配列を宿主細胞に導入することと、候補アミノ酸配列を発現させること、を含む。いくつかのかかる実施形態では、本方法は、候補アミノ酸配列が標的構造の結合パートナーにフォールドするかを決定することをさらに含む。かかる決定は、タンパク質結合を評価するための生化学的及び/又は生物物理学的方法をはじめとする既知の方法により行うことが可能である。

10

【0200】

ある特定の実施形態では、計算設計タンパク質は、酵素、抗体、レセプター、リガンド、輸送タンパク質、ホルモン、成長因子、及びそれらの断片である。いくつかのかかる実施形態では、抗体はヒト抗体である。いくつかのかかる実施形態では、計算設計タンパク質は、一本鎖抗体（たとえば一本鎖 Fv）である。いくつかのかかる実施形態では、計算設計タンパク質は、抗原結合抗体フラグメント、たとえば、Fab又はFab'フラグメントである。

【0201】

C. 定義

本明細書で用いられる場合、「接触度」とは、所与の位置対 *i* 及び *j* が接触を確立するのに有利と思われる状況を意味する。接触度は、「結合残基」を同定するために使用可能である。

20

【0202】

本明細書で用いられる場合、「結合残基」とは、たとえば標的構造中のアミノ酸残基対を意味し、この対では一方の残基のアミノ酸のアイデンティティーは、他方の残基のアミノ酸のアイデンティティーに依存する。

【0203】

本開示では、選言命題の使用は、連言命題を含むことが意図される。定冠詞又は不定冠詞の使用は、カーディナリティーを表すことが意図されない。特定的には、「the」オブジェクト又は「a」及び「an」オブジェクトへの参照は、可能な複数のかかるオブジェクトの1つを意味することも意図される。さらに、接続詞「or（又は）」は、相互排他的選択肢ではなく同時に存在する特徴を伝えるために用いられる。言い換えると、接続詞「or（又は）」は、「and/or（及び/又は）」を含むものと理解すべきである。「includes（～を含む）」、「including（～を含む）」、及び「include（～を含む）」という用語は包括的であり、それぞれ、「comprises（～を含む）」、「comprising（～を含む）」、及び「comprise（～を含む）」と同一範囲を有する。

30

【0204】

以上に記載の実施形態、とくにいずれかの「好ましい」実施形態は、実現可能例であり、単に本発明の原理を明確に理解すべく示される。本明細書に記載の技術の趣旨及び原理から実質的に逸脱することなく、以上に記載の実施形態に多くの変更及び修正を加えうる。修正はすべて、本開示の範囲内で本明細書に含まれるとともに以下の特許請求の範囲により保護されることが意図される。

40

【実施例】

【0205】

D. 実施例

以下の実施例は、単なる例示にすぎず、なんら本開示を限定的なものではない。

【0206】

実施例1：表面再設計（リサーフェイシング）

50

タンパク質表面（すなわち、溶媒に暴露される残基のセット）は、多くの生物物理学的性質、たとえば、溶解性、免疫原性、自己会合、アグリゲーション傾向、さらには安定性及びフォールド特異性を決定するうえで重要である。したがって、これら性質の1つ以上をモジュレートするために、その全体の構造及び機能を保存しつつ、所与のタンパク質の表面を単に再設計することが有用なこともある。本実施例は、赤色蛍光タンパク質（RFP）の表面を再設計（リサーフェイシング）するタスクを記述する。RFPは、可視スペクトルの赤色部分（約600nm）の近傍に集中する発光スペクトルを有して天然で蛍光を発するタンパク質である。他の蛍光タンパク質（FP）と同様に、RFPは、生物学的イメージングタグとして光学実験にきわめて有用である[1]。したがって、（多くの場合高濃度で）機能しなければならない環境（又は細胞型）に依存してRFPの表面残基をモジュレートすることが有用でありうる。

10

【0207】

設計テンプレートとしてRFP mCherry（PDBコード2H5Q[2]）の結晶構造を使用した。構造中の合計64位置を表面上にあるものとしてマニュアルで選択した（0.42超の自由度の値を有する位置におおよそ対応する）。これらは図5（左側パネル）に球として示される。この後、本明細書に記載の模範的TERMベース法を用いて、20天然アミノ酸間で変動する表面位置のすべてに対応する統計エネルギー表を計算し、残りの位置をPDBエントリー2H5Q中のそれらのアイデンティティーに固定した。したがって、得られたエネルギー表は、 $20^{64} = 2 \times 10^{83}$ 配列の配列空間を記述した。整数線形プログラミングを用いてこの空間全体にわたり最適化し、最低全統計エネルギースコアを有する単一配列を見いだした。得られた配列は、mCherryの出発配列と比較して表1に示される。元のmCherry構造及び得られた設計モデル構造の真空中表面静電位は、図5で比較される（中間パネル及び右側パネル）。明らかに、設計配列は、表面の静電的性質及び形状への有意な攪乱を呈する。事実上、64可変位置のうち合計48位置は設計で変化する。

20

【0208】

【表1】

表1. TERMベース設計配列は元の野生型mCherry配列と有意に異なる。

mCherry	MVSKGEEDNM	AIIKEEFMRFK	VHMEGSVNGH	<u>EFEIEGEGEG</u>	<u>RPYEGTQTAK</u>	
設計	MVSKGEEDNM	AIIKEEFMTFE	<u>VEMEGTVNGH</u>	<u>PFRIRGSGGG</u>	<u>DPYEGTQTAR</u>	
mCherry	LKVTKGGPLP	FAWDILSPQF	MYGSKAYVKH	PADIPDYLKL	SFPEGFKWER	
設計	<u>LEVVEGGPLP</u>	FAWDILSPQF	MYGSKAYVKH	PADIPDYLKL	SFPEGFTWTR	
mCherry	<u>VMNFEDGGVV</u>	<u>TVTQDSSLQD</u>	<u>GEFIYKVKLR</u>	<u>GTNFPDGPV</u>	<u>MQKKTMGWEA</u>	
設計	<u>TMEFEDGGTV</u>	<u>KVTQTSTLKD</u>	<u>GKFHYKVKLT</u>	<u>GSNFPDGPV</u>	<u>MQKKTMGWEA</u>	
mCherry	<u>SSERMYPEDG</u>	<u>ALKGEIKQRL</u>	<u>KLKDGGHYDA</u>	<u>EVKTTYKAKK</u>	<u>PVQLPGAYNV</u>	
設計	<u>SSTERMRPKDG</u>	<u>KLEGEIDQEL</u>	<u>RLKDGGYYRA</u>	<u>RVRTTYKAKK</u>	<u>PVQLPGAYTV</u>	
mCherry	<u>NIKLDITSHN</u>	<u>EDYTIVEQYE</u>	<u>RAEGRHSTGG</u>	<u>MDELYK</u>	(配列番号1)	
設計	<u>RIRLEITSHN</u>	<u>EDYTEVEQTE</u>	<u>TAKGEHSTGG</u>	<u>MDELYK</u>	(配列番号2)	

30

40

【0209】

設計で可変としてマークされた位置は下線付きであり、且つ設計配列で突然変異したものはボールド体でマークされる。

【0210】

設計を確認するために、配列をE.コリ（E.coli）中にクローニングし、続いて、標準的な分子生物学的及び生物物理学的技術を用いて発現及び精製を行った。

【0211】

50

高速タンパク質液体クロマトグラフィー（FPLC）は、タンパク質が（少なくとも10 μ Mの濃度で）天然mCherryとまったく同様に溶液中でモノマーであることを示した（図6参照）。

【0212】

48突然変異を保有するにもかかわらず且つ光学的性質の保存が設計拘束ではなかったという事実にもかかわらず（構造の保存のみ拘束）、設計は、依然として元のタンパク質に特有の桃色を呈した（図7参照、トップ）。さらに、設計タンパク質は、依然として蛍光を発し、発光スペクトルは、mCherryとほぼ同一の形状を呈した（図7参照、ボトム）。最後に、グアニジニウム塩酸塩（GuHCl）による化学変性は、タンパク質の構造が元のmCherryとほぼ同程度に良好にそのクロモフォアを保護することを明らかにした（それ自体きわめて安定な高度工学操作タンパク質）（図8）。そのため、あらゆる手段により、48位置が元のmCherryと異なる設計タンパク質は、出発構造さらには機能さえも保存した。かかる多様性を生成する能力は、一連の所望の性質を有するRFP又は他のタンパク質の変異体を迅速に工学操作するために容易に活用可能である。

10

【0213】

実施例2：膜タンパク質の可溶化のためのリサーフェイシング

とくに、リサーフェイシングアプローチは、水性溶液への溶解性に関して膜タンパク質を再設計するために使用可能である（5）。水溶性タンパク質は、膜貫通（TM）タンパク質よりも発現、精製、及び操作がかなり容易であるので、治療剤を標的化する課題がより容易になる。そのため、膜タンパク質の水溶性アナログを生成する能力は、Gタンパク質共役レセプター（GPCR）などの主要なバイオメディカル関連標的に対する薬剤及び抗体を同定するプロセスをかなり単純化可能であろう。

20

【0214】

この目的に合わせたTERMベース設計の使用は、水への可溶化時に溶媒暴露されるであろうTMタンパク質構造の表面上の脂質に面している位置を同定することと、以上の実施例1で利用した標準的手順を介してそれらを再設計することと、を含む。

【0215】

相互作用表面位置間のアミノ酸の組合せの特異的選択は、既知の水溶性タンパク質構造の類似の構造環境の配列統計を観測及び「学習」する結果として得られた。これは本明細書に開示される設計手順の一部でありうる。

30

【0216】

図9は、GPCR-1アドレナリン作動性レセプター（PDBコード4BVN、左側パネル参照）の結晶構造に本プロセスを適用した結果を示す。図9の中間パネルと右側パネルとを比較すると、脂質二重層との相互作用に理想な主に疎水性のものから、水との相互作用に好適な親水性のものに、タンパク質の表面が設計プロセスにより変換されたことは明らかである。そのため、本明細書に記載の方法は、水への溶解性に関してGPCRなどのタンパク質をリサーフェイスするのに有用である。

【0217】

実施例3：TERMベース法により計算された統計エネルギースコアは設計品質を示す

本実施例では、何千ものde novo設計タンパク質配列に関する既存の発表データを利用して、より良好な統計エネルギースコアがより高い設計成功率を表す傾向があるか及び設計タンパク質のより良好な品質に相関するかを決定した。特定的には、Bakerらにより発表されたデータを使用した。そこでは、フォールドされた安定なプロテアーゼ耐性構造を形成する能力に関して、4つの識別可能なトポロジー（図10A~10D参照）の合計約15,000de novo設計配列が高スループットで試験された（3）。これらの設計の各々は、Rosetta Designソフトウェアスイート（6）により所望の標的骨格と良好に適合可能であると予測された配列を呈したが、ほとんどの設計はフォールドに失敗した。

40

【0218】

本実施例は、本明細書に開示される設計法が成功又は失敗の設計をより良好に識別可能

50

であるかを試験しようとして試みたものである。この目的で、Bakerらにより投入された約15,000骨格構造の各々につき模範的設計法を用いて（それらの設計の各々につき1つずつ）（3）、標的モデルのいずれに対してもいずれかの天然アミノ酸配列の評価を可能にした。各設計配列に対してそのそれぞれの骨格で本明細書に開示される模範的設計法を用いてエネルギースコアを計算し、異なるトポロジーに渡って比較を容易にするために配列長で除算した。図10E~10Hは、4つのトポロジーの各々につき、得られたスコアと実験上の「安定性スコア」（Bakerらが高スループットで設計安定性を推定するために開発し熱力学的安定性に密に相関することが示されたプロテアーゼ耐性ベースメトリック）との間の相関を示す。明らかなように、TERMベーススコアと実験スコアとの間にロバストな相関が存在した（p値はすべての場合にきわめて有意である。図10E~10Hの凡例を参照されたい）。これとは対照的に、各配列に対して計算されたロゼッタスコア（同様にBakerらにより発表された）を考慮した場合、相関はすべての場合に有意により弱かった（図10I~10Lを参照されたい）。事実上、4つのトポロジーのうち3つでは、相関係数は、統計的に有意でないか（図10Kで0.1のp値）誤った符号であるか（図10J及び10Lで期待される負ではなく正の相関）のどちらかであった。

10

【0219】

Rosetta Designは、計算タンパク質設計で現在の最新技術に相当する（7）。そのため、TERMベーススコアリングは、既存の設計方法によりキャプチャーできない方法で構造-配列関係を統合することが、この結果から示唆される。さらに、ここで解析された約15,000設計配列は、TERMベーススコアリングではなくRosetta Designに基づいて最適化された。事実上、TERMベース最良スコアリング配列は、常にロゼッタベース設計と平均で84%異なっていた（すなわち、Rosettaベース選択配列とTERMベース選択配列との間で平均で位置の約16%が同一であったにすぎない）。それ自体の予測配列ランドスケープの最適性領域とは異なる配列でさえも定量的にスコア付けする本明細書に開示されるTERMベース法の能力は、本方法の一般性及びそれが定量する配列-構造関係のユニバーサル適用性をさらに確認する。

20

【0220】

図11は、本明細書に開示される模範的方法を用いて計算されたスコアが、4天然ドメインの120配列変異体を用いて熱力学的安定性に密に相関したことをさらに示す。これらは、高スループット実験安定性スコアの定量的性質を確立するためにRocklinらを使用したのと同じの変異体である（3）。TERMベーススコアと熱力学的実験との間の密な相関は、TERMベース法をさらに確認し、TERMベーススコアの最適化がロバストな汎用タンパク質設計戦略であることを示唆する。

30

【0221】

実施例4：新規な結合モードの設計

タンパク質-タンパク質相互作用は、生細胞の内部論理配線を効果的に提供し、細胞がその内部又は周囲のイベントをどのように感知しそれに反応するかを規定する。多くの細胞タンパク質-タンパク質相互作用は、専用のタンパク質相互作用ドメインによりコードされる。これらに包含されるものとしては、パートナータンパク質のC末端テールに特異的に結合し最後の6~10アミノ酸を特異的に認識するモジュールであるPDZドメインが挙げられる（8、9）。ヒトゲノム中には250超のPDZドメインが存在し、それらは細胞シグナリング及び局在化に広く関与する（8）。そのため、特異的PDZドメインを認識し阻害する分子は、大きなバイオメディカルニーズを呈する。しかしながら、PDZドメインの結合ポケットは構造的に保存され、多くのドメインがオーバーラップ結合特異性を呈するので、結合ポケットの外側のあまり保存されない領域を標的とすれば、より良好な阻害選択性を達しうる。

40

【0222】

本実施例では、2つのヒトPDZドメイン、すなわち、タンパク質NHERF-2の第2のPDZドメイン（N2P2）及びタンパク質MAGI-3の第6のPDZドメイン（

50

M3P6)を利用した。両ドメインは、リゾホスファチジン酸レセプター2(LPA22)のC末端を認識し、両方とも結腸癌に関連するシグナル伝達に参与する(10~13)。しかしながら、LPA22へのN2P2の結合は腫瘍形成活性を増強するが、M3P6の結合はそれを阻害する(12)。そのため、M3P6に勝るN2P2の選択的阻害は、結腸癌に対する可能性のある治療経路として妥当である(14)。

【0223】

両ドメインは天然で同一配列(LPA22のC末端)を認識するので、保存結合ポケットの外側でN2P2との接触を形成するように、TERMベースストラテジーを利用して既知のN2P2結合ペプチド(PDBエントリー2HE4中のN2P2の複合構造から取り出した)を伸長させた。ストラテジーは、N2P2の既存の構造を完成させるのに好適なマルチセグメントTERMを同定した。すなわち、TERMは、N2P2の表面領域に良好にアライメントするセグメントのサブセットを有し(インターフェースアンカー)、残りのセグメントは、推定インターフェースを形成し(インターフェースシード)、そしてTERM配列統計は、N2P2アンカー領域の配列に適合可能である(図12参照)。次いで、アンカー/シードの組合せをマニュアルで選択し(M3P6と対比して保存されない残基にマッピングするN2P2アンカー領域に基づく)、そして中程度に良好にオーバーラップするTERMにより既存の結合ペプチドに結合させた(図12参照)。最後に、得られた骨格構造(図12に示される)を本明細書に開示される模範的設計法を用いた設計に付し、実験的特徴付けに関して最適配列を選択した。

【0224】

精製設計ペプチドを市販品として入手し、N2P2及びM3P6の両方へのその親和性を我々の前の研究と同様に蛍光偏光(FP)阻害アッセイにより調べた(15)。N2P2への親和性は1 μ M程度であったが、M3P6との検出可能な相互作用は存在しなかったことが、図13により示される。比較すると、LPA2(N2P2及びM3P6の両方の天然パートナー)のC末端6マーのペプチドは、N2P2に約1/30の弱い結合をするが、N2P2及びM3P6に対してほぼ等しい親和性を呈する(15)。そのため、設計された新規な結合モードは、改善された親和性及び劇的に改善された選択性の両方を示す。

【0225】

実施例5：構造のde novo設計

本明細書に開示されるフレームワークは、既存のタンパク質フォールドに由来するかde novoで構築されたかにかかわらず、任意の構造に適用可能である。例として、図14Aは、Rocklinらが配列の設計に最近成功した計算発生骨格を示す(3)。この構造又はいずれかの他の新規な骨格は、以上に記載の方法を用いて設計可能である。この特異的骨格では、いずれかの天然アミノ酸を位置のいずれかで選択した場合(約10⁵の全配列空間)、図14Bに示される解が最適として選択された。設計配列のモデリングされた構造は、生物物理学的に適正であるように見えた(図14B参照)。さらに、モデリングされた配列と既知構造のタンパク質との間の離れた相同性を同定する能力に依拠する強力な構造予測法であるHHpred(4、16)に設計配列を付したところ、PDBエントリー5UP5が最近接マッチングとして明らかにされた(97%超の確率及び90%のアライメントカバレッジを有する)(Rocklinら(3)により設計された対応する配列のまさに実験における構造)(図14C参照)。重要なこととして、5UP5それ自体は、TERMベース配列統計においてクエリーが行われたタンパク質のデータベースで使用されなかった(それ自体de novo設計であるので、そのホモログはデータベース中にも存在しなかった)。これは、本明細書に開示される模範的方法を用いて設計された配列がたとえば標的構造にフォールドする可能性などの必要な特徴を有することを示唆する強力な証拠である。付随的に、HHpredにより明らかにされた第2のマッチであるPDBエントリー1UTAは、標的を強く連想させるフォールドを有する天然構造である(図14D参照)。

【0226】

参照文献

1. Mackenzie CO, Zhou J, & Grigoryan G (2016) Tertiary alphabet for the observable protein structural universe. *Proc Natl Acad Sci U S A* 113(47): E7438 - E7447.
2. Wang H, et al. (2016) LOVTRAP: an optogenetic system for photoinduced protein dissociation. *Nat Methods* 13(9): 755 - 758.
3. Rocklin GJ, et al. (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 357(6347): 168 - 175. 10
4. Meier A & Soeding J (2015) Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. *PLoS Comput Biol* 11(10): e1004343.
5. Perez-Aguilar JM, et al. (2013) A computationally designed water-soluble variant of a G-protein-coupled receptor: the human mu opioid receptor. *PLoS One* 8(6): e66009. 20
6. Leaver-Fay A, et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487: 545 - 574.
7. Alford RF, et al. (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* 13(6): 3031 - 3048.
8. Ivarsson Y (2012) Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett* 586(17): 2638 - 2647. 30
9. Lee HJ & Zheng JJ (2010) PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal* 8: 8.
10. Oh YS, et al. (2004) NHERF2 specifically interacts with LPA2 receptor and defines the specificity and efficiency of receptor-mediated phospholipase C-beta3 activation. *Mol Cell Biol* 24(11): 5069 - 5079. 40
11. Yun CC, et al. (2005) LPA2 receptor mediates mitogenic signals in human colon cancer cells. *Am J Physiol Cell Physiol* 289(1): C2 - 11.
12. Lee SJ, et al. (2011) MAGI-3 competes with NHERF-2 to negatively regulate LPA2 receptor signaling in colon cancer cells. *Gastroenterology* 140(3): 924 - 934.
13. Willier S, Butt E, & Grunewald TG (2013) Lysophosphatidic acid (LPA) signalling in 50

cell migration and cancer invasion: a focused review and analysis of LPA receptor gene expression on the basis of more than 1700 cancer microarrays. *Biol Cell* 105(8): 317-333.

14. Yoshida M, et al. (2016) Deletion of Na⁺/H⁺ exchanger regulatory factor 2 represses colon cancer progress by suppression of Stat3 and CD24. *Am J Physiol Gastrointest Liver Physiol* 310(8): G586-598.

10

15. Zheng F, et al. (2015) Computational design of selective peptides to discriminate between similar PDZ domains in an oncogenic pathway. *J Mol Biol* 427(2): 491-510.

16. Zimmermann L, et al. (2017) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol*.

【0227】

以上の詳細な説明及び付随する実施例は、単なる例示にすぎず、添付の特許請求及びその均等物によってのみ定義される本発明の範囲を限定するものとみなすべきではないと理解される。開示された実施形態に対する各種の変更及び修正は当業者には明らかであろう。かかる変更及び修正は、限定されるものではないが、化学構造、置換基、誘導體、中間体、合成、製剤、若しくは方法に関するもの、又は本発明の有用なかかる変更及び修正のいずれかの組合せを含めて、その趣旨及び範囲から逸脱することなく行いうる。

20

【0228】

以上で引用された参照文献（特許及び非特許）はすべて、参照により本特許出願に組み込まれる。それらの参照文献の考察は、それらの著者らによりなされた主張を単にまとめることが意図される。いずれかの参照文献（又はいずれかの参照文献の一部）が関連先行技術であること（又はまったく先行技術ではないこと）を容認するものではない。出願人は、引用された参照文献の正確性及び妥当性を検証する権利を留保する。

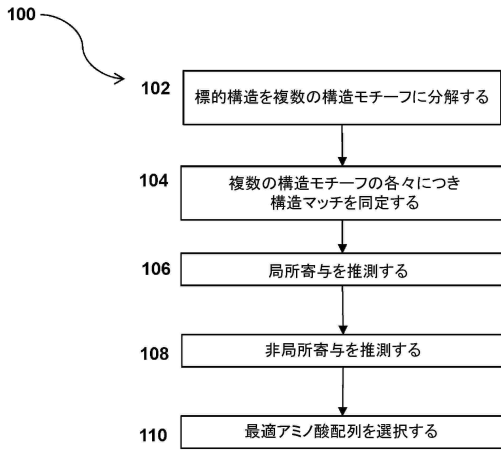
30

40

50

【図面】

【図 1】



【図 2 - 1】

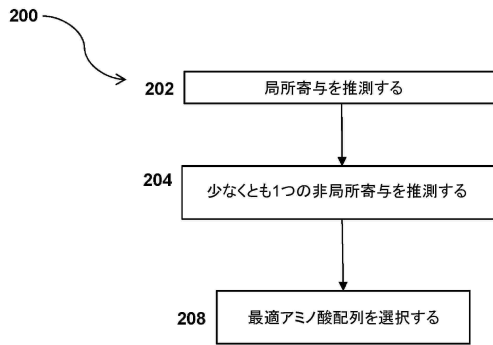


図2A

10

【図 2 - 2】

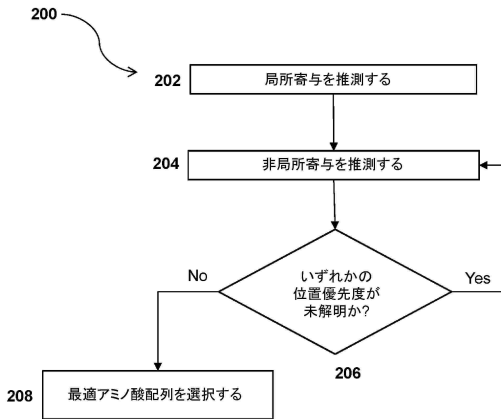
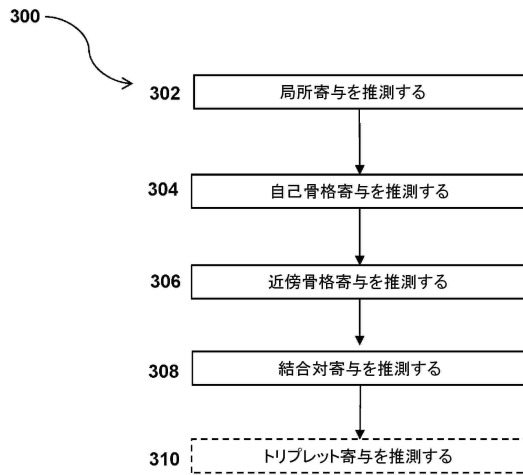


図2B

【図 3】



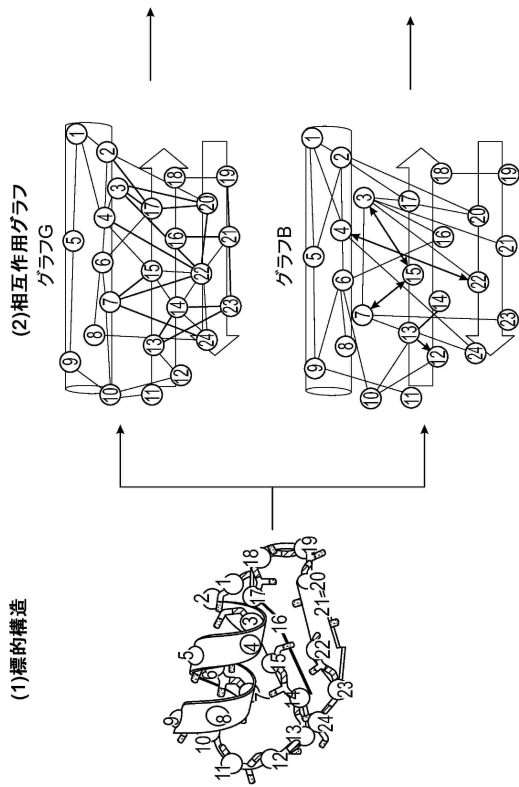
20

30

40

50

【 図 4 - 1 】



【 図 4 - 2 】

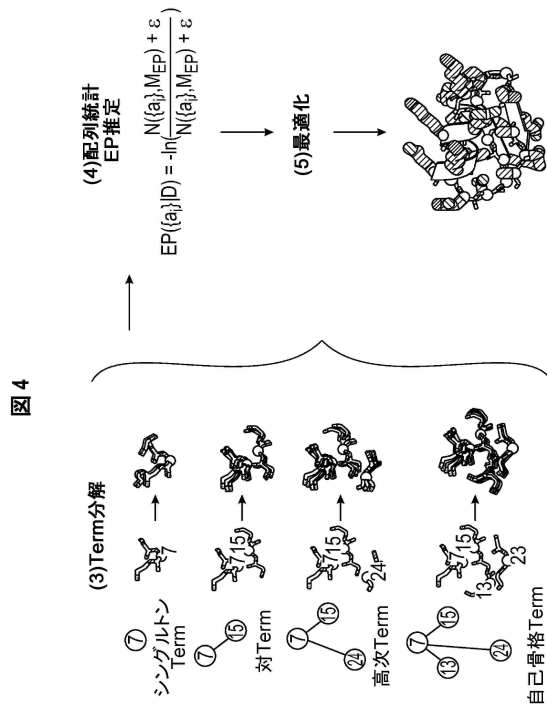
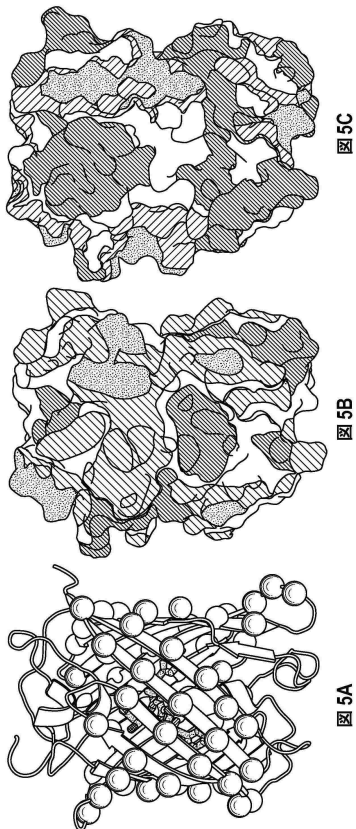


図 4 (続き)

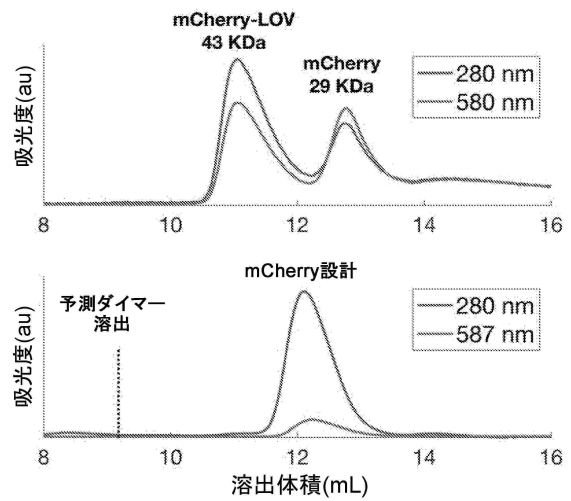
10

20

【 図 5 】



【 図 6 】

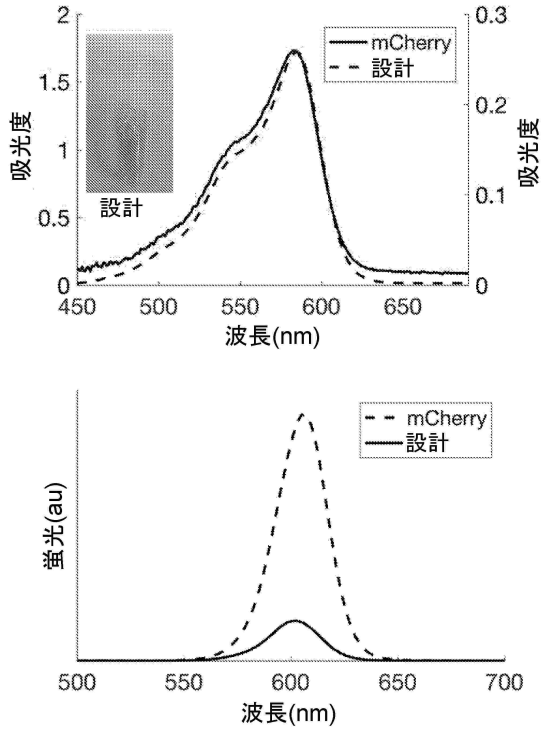


30

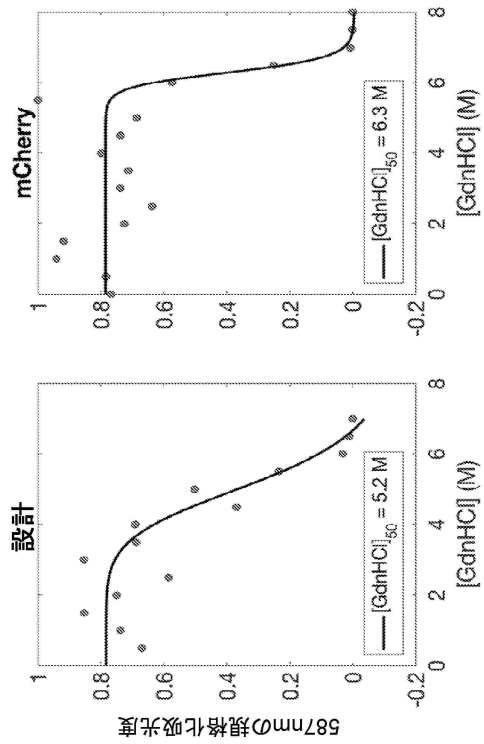
40

50

【 図 7 】



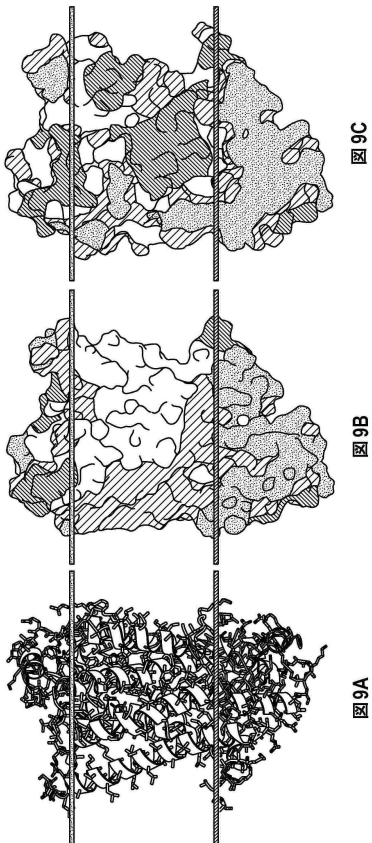
【 図 8 】



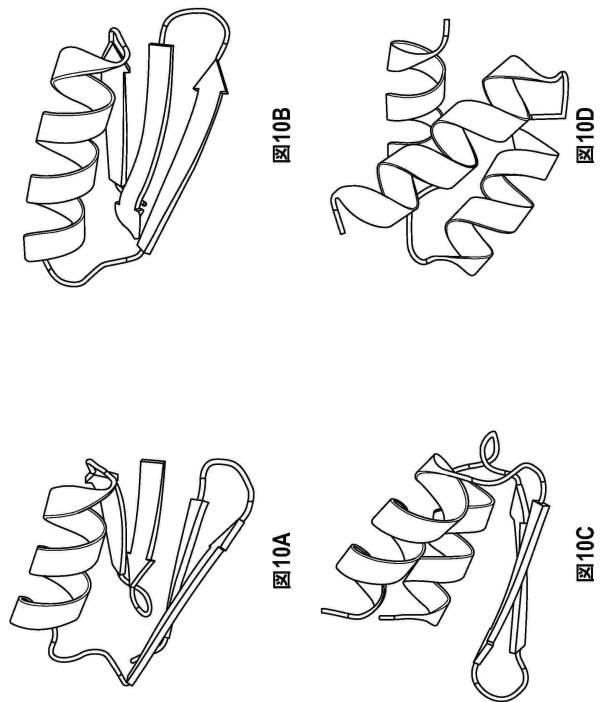
10

20

【 図 9 】



【 図 10 - 1 】



30

40

50

【図 10 - 2】

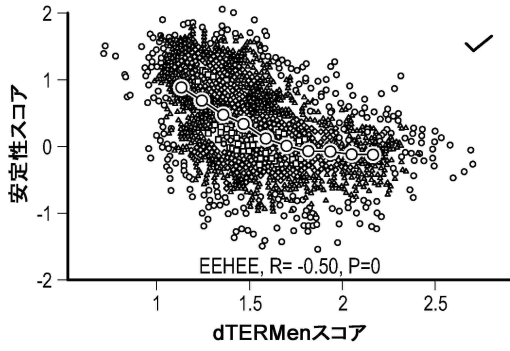


図 10E

【図 10 - 3】

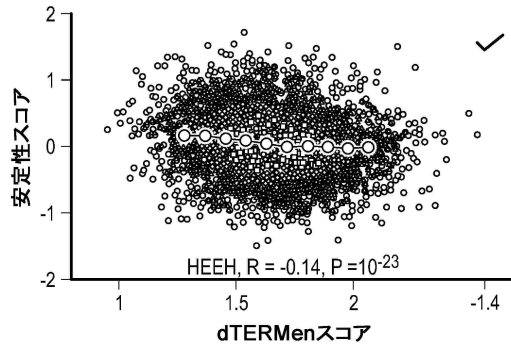


図 10G

10

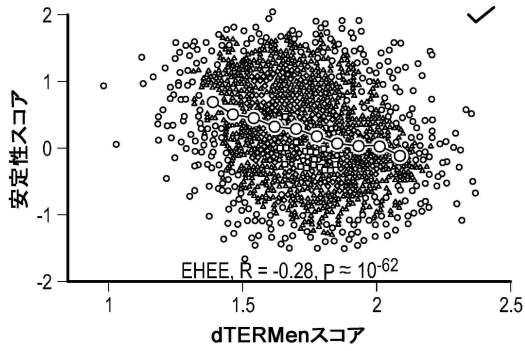


図 10F

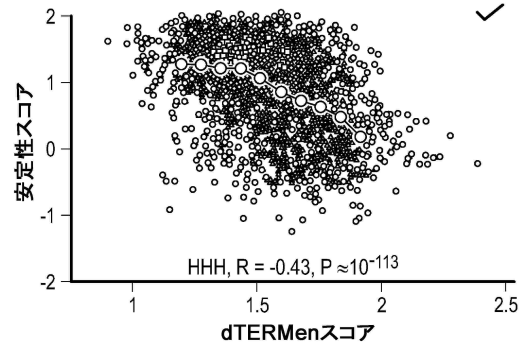


図 10H

20

【図 10 - 4】

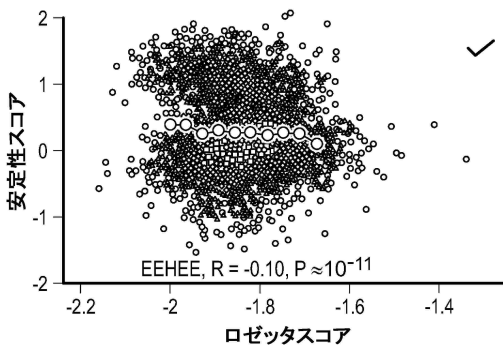


図 10I

【図 10 - 5】

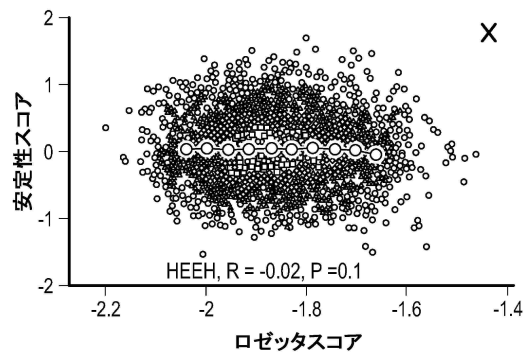


図 10K

30

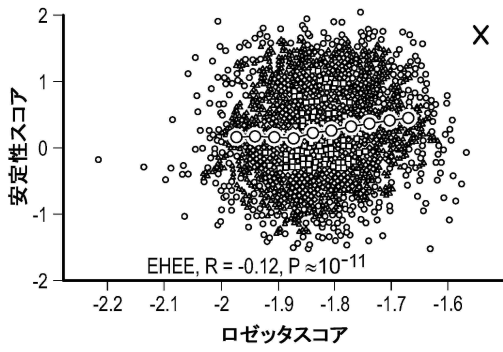


図 10J

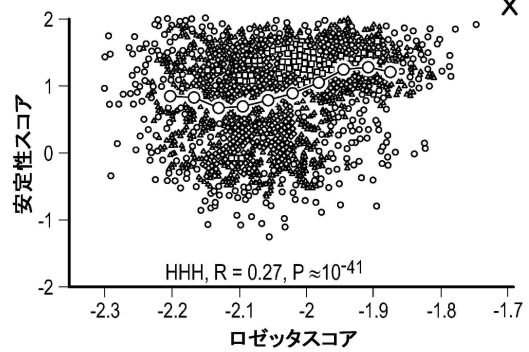
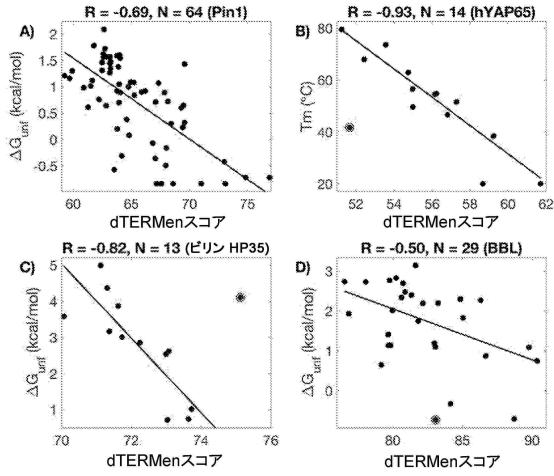


図 10L

40

50

【 図 1 1 】



【 図 1 2 】

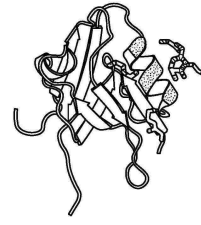


図 12A

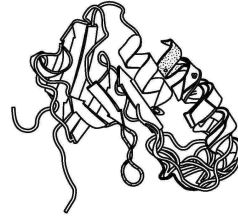
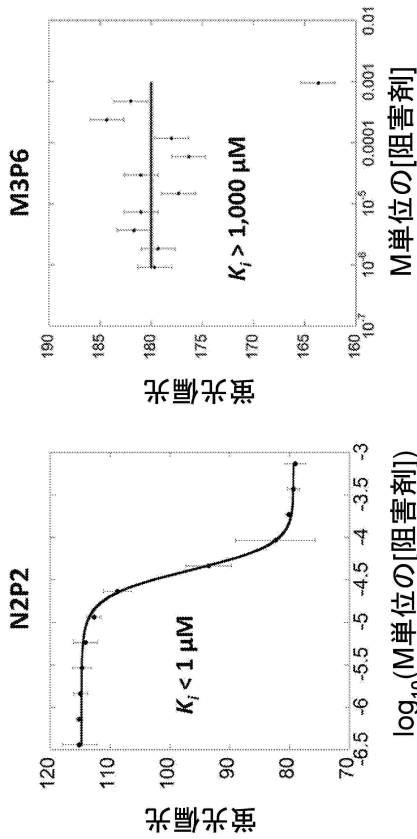


図 12B



図 12C

【 図 1 3 】



【 図 1 4 】

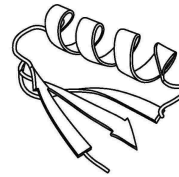


図 14A

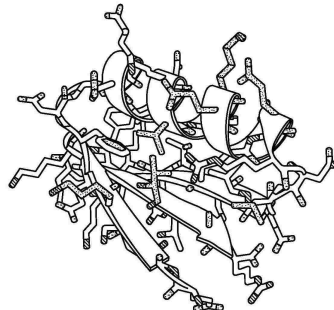


図 14B

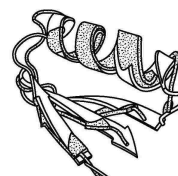


図 14C

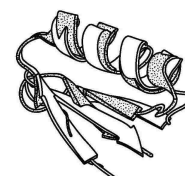


図 14D

10

20

30

40

50

【配列表】

0007438545000001.app

10

20

30

40

50

フロントページの続き

- 弁理士 當麻 博文
(74)代理人 100137729
弁理士 赤井 厚子
(74)代理人 100151301
弁理士 戸崎 富哉
(72)発明者 グレゴリアン、ゲヴォーグ
アメリカ合衆国、ニュー ハンプシャー州 03755、ハノーヴァー、カーリーア プレイス 5ビー
(72)発明者 チョウ、ジャンフ
アメリカ合衆国、ニュー ハンプシャー州 03784、ウェスト レバノン、セイチェム サークル
52
(72)発明者 マッケンジー、クレイグ
アメリカ合衆国、ニュー ハンプシャー州 03229、ホプキントン、メイン ストリート 27
8、アパートメント 2
審査官 渡邊 加寿磨
(56)参考文献 特開2004-33066(JP,A)
米国特許出願公開第2008/0059077(US,A1)
米国特許出願公開第2011/0275558(US,A1)
特表2017-526053(JP,A)
清水謙多郎, "タンパク質の構造予測手法の開発", [online], 国立研究開発法人 科学技術振
興機構, 2010年05月22日, [https://spc.jst.go.jp/hottopics/1006genome_biofunctional_an
alysis/r1006_shimizu.html](https://spc.jst.go.jp/hottopics/1006genome_biofunctional_analysis/r1006_shimizu.html), [検索日:2023.03.29]
(58)調査した分野 (Int.Cl., D B 名)
G 0 6 Q 1 0 / 0 0 - 9 9 / 0 0
G 1 6 B 5 / 0 0 - 9 9 / 0 0
G 1 6 Z 9 9 / 0 0
A 6 1 K 3 8 / 1 6
C 0 7 K 1 4 / 0 0
C 1 2 N 1 5 / 0 9
C 1 2 P 2 1 / 0 2