

技術動向サーベイ

- 作成者：熊谷 渉（O&M Design Gr., Project Design Div.）
- 担当日：2022年3月2日

対象論文

- タイトル：A backbone-centred energy function of neural networks for protein design [【URL】](#)
- 発行日：2022年2月9日
- 雑誌名：Nature, Vol.602, pp.523–528
- 著者：B. Huang^{1,2}, Y. Xu^{1,2}, X. Hu^{1,2}, Y. Liu¹, S. Liao¹, J. Zhang¹, C. Huang¹, J. Hong¹, Q. Chen¹, and H. Liu¹
- 所属
 - 1: University of Science and Technology of China (China)
 - 2: Contributed Authors
- 引用数：0？（8644 Accesses、2022年2月24日時点）
- キーワード：計算機、*de novo*タンパク質設計、分子設計

Abstract

タンパク質の主鎖構造は、そこに自律的に折り畳まれるアミノ酸配列が相当数存在すれば、設計可能であると言える。主鎖の設計可能性は、側鎖依存性、あるいは側鎖の種類に鈍感な分子間相互作用によって、主に支配されていることが示唆されており、主鎖を中心としたエネルギー表面の連続的なサンプリングと最適化に基づく新しい主鎖の設計（アミノ酸選択の準備のための）アプローチが示唆されている。しかし、この目的のために、十分に包括的で正確なエネルギー関数はまだ確立されていない。本論文では、ニューラルネットワーク形式のエネルギー項を用いた SCUBA（Side Chain-Unknown Backbone Arrangement）と名付けた統計モデルによって、この目標が達成されることを示す。この項は、カーネル密度推定とニューラルネットワークの学習からなる2段階のアプローチで学習され、既知のタンパク質構造における多次元高次相関を解析的に表現することができる。我々は、SCUBAを用いて高精度に主鎖を設計した9つの*de novo*タンパク質の結晶構造を報告する。そのうち4つは新規で非天然な全体構造を有している。SCUBAを用いた構造設計は、既存のタンパク質構造のフラグメントを使用しないため、設計可能な主鎖空間を広く探索することができ、*de novo*設計が可能なタンパク質の新規性と多様性を拡大することができる。主鎖を中心としたエネルギー表面のSCUBAによるモデリングは、設計可能なタンパク質の多様性を拡大する。

背景

- 計算機による*De novo*タンパク質の設計のほとんどはRosettaDesignを用いて実施されてきた。
- このアプローチは、既存構造のテンプレートを流用することで新規の主鎖を構築する。
 - Template-based : 既存の構造モジュール（またはテンプレート）間の相対的な形状をパラメトリックに変化させ、helixバンドルや繰り返しタンパク質を設計する
 - Template-free : 既存の構造からfragmentを組み立てる (Fragment Assembly)
- しかし、この方法は、テンプレートに依存するため、利用できる新しい構造の多様性が大きく制限され、設計可能な機能性の範囲が狭くなる可能性がある。

先行研究

- そこで、テンプレートフリーな設計アプローチとして、主鎖中心のエネルギー景観を明示的に表現し、これを設計に利用する方法がある。
 - この表現方法の進歩は遅い。*De novo*タンパク質設計タスクに必要なレベルで、関連する分子間相互作用を包括的かつ正確に表現する方法がないためだと考えられる。
- C_{α} 原子ベースの統計ポテンシャルは、天然のフラグメントをテンプレートとして使用せずに、新規ループの設計に成功しており、主鎖中心アプローチの有効性を示している。
 - C_{α} ポテンシャルは、局所的な主鎖コンフォメーションの正確さを重視している
 - 簡略化した主鎖エネルギー景観を探索したときは、天然の主鎖に似た平坦な局所解の存在を確認するだけだった^{4,22,23}。
- 特に、SCUBA (Side Chain-Unknown Backbone Arrangement) モデルは、主鎖設計に不可欠な要素を表現することで、*De novo*タンパク質設計に対応した網羅性と精度を達成することを目指している^{3,4,5}。
 - 主鎖の局所構造傾向や水素結合の形状、キラリと光る側鎖や密に配置された側鎖に必要な主鎖間スペースなど
- さらに、従来の統計ポテンシャルは、下記の課題があった。
 - 高次元空間におけるデータの相関を明確に表現できない
 - ポテンシャル関数は連続性の保証がない、あるいは勾配が容易に得られない

本論文の成果

- SCUBAとして、NC-NNを用いて統計ポテンシャルを学習する方法を提案した。
 - NC-NNはカーネル密度推定とニューラルネットワークを合わせたモデリング手法
- NC-NNベースのSCUBAを用いて、設計可能な主鎖空間を探索する、テンプレートフリーなタンパク質設計方法を提案した。

- 網羅性と精度の向上により、様々な設計仕様を満たす構造を有するタンパク質や、非天然なタンパク質を設計した。
- 本研究で得られた新しい構造のタンパク質は、自然界で観察されるよりも幅広い形状のタンパク質を設計するために、本研究のアプローチが有用であることを明確に示しています。
- 機能性タンパク質を設計する際、エネルギー関数駆動型のバックボーンサンプリングと最適化は、サンプリング技術の向上による構造空間の広範な探索と、機能関連の拘束を適用することによる設計構造の精密な制御の両方を容易に行うことができる。
- これらのアプローチにより、*De novo*タンパク質設計で利用可能な構造の多様性と機能性を大幅に拡大することができます。

Idea & Findings

本手法の特徴

- NC-NNは高次元空間における相関を忠実に保持することができる。
- NC-NNによって学習されたポテンシャルは、勾配駆動の構造サンプリングや最適化に適している。
 - 自動的に連続となり、解析的勾配を持つ
- 主鎖設計の段階で配列空間を探索する必要がない。

Extended Data Table 1: Summary of experimentally examined designs (Batch1,2)									
Batch	Name	Architecture	Number						
1	H2E4	2-helix and 4-strand bundle	33		1	H4	4-helix bundle	11	1
	EXTD	天然ドメインのループを4helixセグメントで置換	8		2	H2E4	2-helix and 4-strand bundle	30	2
	H4	4-helix bundle	8						

既存構造デザインの検証

- 既存構造で設計されたタンパク質を実験的に検証した。
 - H2E4、H4、EXTDのスケッチが含まれている
 - Batch1：主鎖最適化プロトコルのみで設計
 - Batch2：修正プロトコルで設計（主鎖最適化プロトコルにloopリサンプリング+最適化のステップを追加）

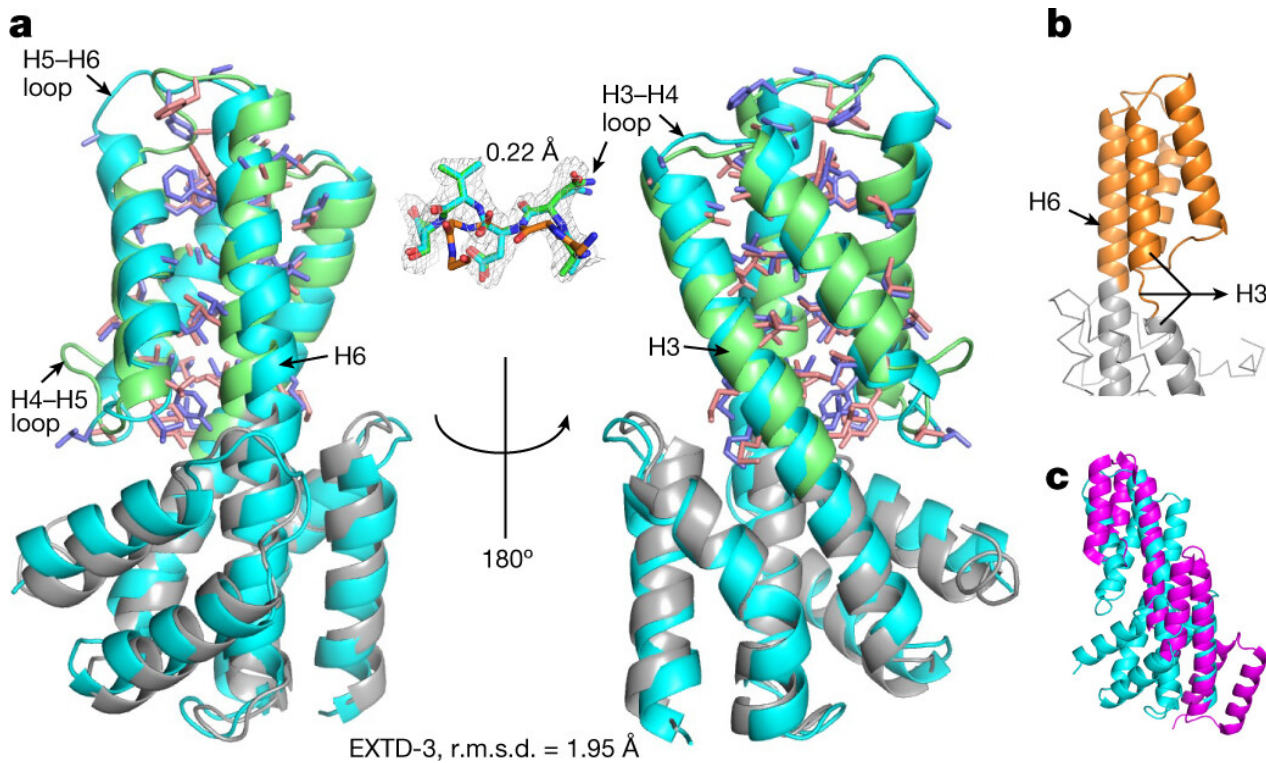


Fig. 2: *De novo*タンパク質EXT-D-3は、天然ではまだ観察されていない単一の剛体構造を形成するために、既存のパーツと新規設計されたパーツが統合される

Fig. 2

- Batch1のタンパク質は、主鎖最適化プロトコルを用いて設計された。
- H2E4とH4のデザイン（44個）には、結晶化に成功したものはなかった。
- EXT-Dデザイン（8個）の中の1つ（EXT-D-3）について、解明された結晶構造が得られた。

Fig. 2a

- 実験的に解明された構造（水色リボン）とEXT-D-3（灰色+緑色リボン）の重ね合わせ
 - 灰色リボン：既存部位、緑色リボン：新規設計部位
 - tepsin ENTH domain; PDB, 5WF2
 - スティック：側鎖（青色：実験、橙色：新規設計）
- 解明された構造では、設計上導入された3つのloop（H3-H4, H4-H5, H5-H6）のうち、H3-H4 loopは正確に一致している一方、他の2つは大きく外れている。

Fig. 2b

- 主鎖の初期構造
 - 灰色：既存、橙色：新規設計
- 主鎖H3は、helix-loop-helixセグメントとして存在する

Fig. 2c

- EXT-D-3（水色）と類似タンパク質（紫色）の重ね合わせ
 - EXT-D-3の既存部位と新規設計部位は、類似天然タンパク質の一部に一致する
 - PDB, 5H79; chain D
 - 一方、H3は、SCUBA駆動のSASD中に単一の連続helixに変化し、全体的には非天然な構造が得られた

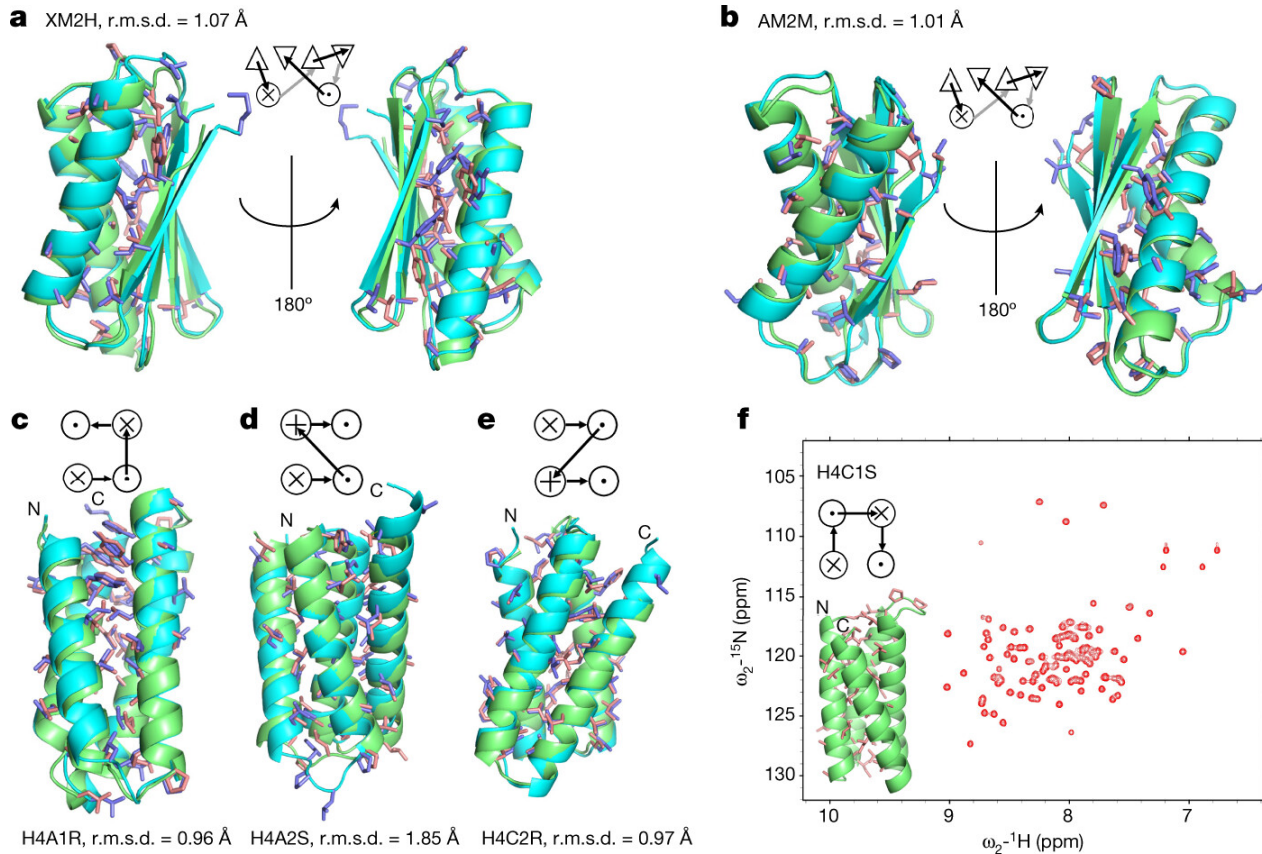


Fig. 3: 2層構造の α - β タンパク質や4-helix bundleタンパク質の設計に成功。

Fig. 3

- Batch1のタンパク質で明らかになった不一致さを受けて、Batch2のタンパク質（38個）は修正プロトコルで設計された。
 - H2E4とH4のスケッチに適用し、Rosettaバイアスフォワードfolding^{9,14}によって設計配列をフィルタリングした
- このうち、H2E4（12個）とH4（4個）がモノマーとして精製され、X線結晶構造または核磁気共鳴（NMR）異核単一量子コヒーレンス（HSQC）スペクトルによって、うまく折りたたまれていることが確認された。
 - X線結晶構造：H2E4モノマー（2個）、H4モノマー（3個）
 - E4NMR HSQCスペクトル：H2E4モノマー（10個）、H4モノマー（1個）

Fig. 3a-e

- H2E4モノマーとH4モノマーのX線結晶構造（水色リボン）と設計構造（緑色リボン）の重ね合わせ
 - a,b : H2E4モノマー（2個）、c,d,e : H4モノマー（3個）

Fig. 3f

- H4モノマー（1個）のE4NMR HSQCスペクトル

総合的な結果

- SCUBAで設計した主鎖は、X線構造とは原子レベル（主鎖間RMSD値0.96-1.85Å）で一致した。
- さらに、設計した構造中の19個のloopのうち16個は、実験的に精製したloopとサブÅ以下の精度で一致した。
- 特に、2つのH2E4は、2次構造が同じように配置されているにもかかわらず、同じloopのコンフォメーションを持っていなかった。
- 設計されたH2E4およびH4は、類似の天然タンパク質との配列同一性が低い。
 - 構造的に整列した天然タンパク質配列との同一性は3.5-25%（平均同一性は14%）を示し、天然配列と天然主鎖のために選択したABACUS2配列間の平均配列同一性（～33%）よりはるかに低い
- また、設計タンパク質は、天然主鎖にABACUS2配列を選択したタンパク質と同様に高い耐熱性を持っている。
- Batch2のタンパク質の設計成功率は約42%であった。
 - 実験的にテストされた38個のタンパク質のうち、16個がうまく折りたたまれた。この成功率には、設計後の計算機によるフィルタリングが寄与していることを認める
- これらの結果から、SCUBA+ABACUS2は、最先端のRosettaDesignと並んで、*De novo*タンパク質設計に有用な方法であることが示唆された。

Extended Data Table 1: Summary of experimentally examined designs (Batch3) | Batch | Name | Architecture | Number | | ----- | ----- | ----- | ----- | | 3 | Novel all-helical | helix (D12,D22,D47,D53) | 13 |

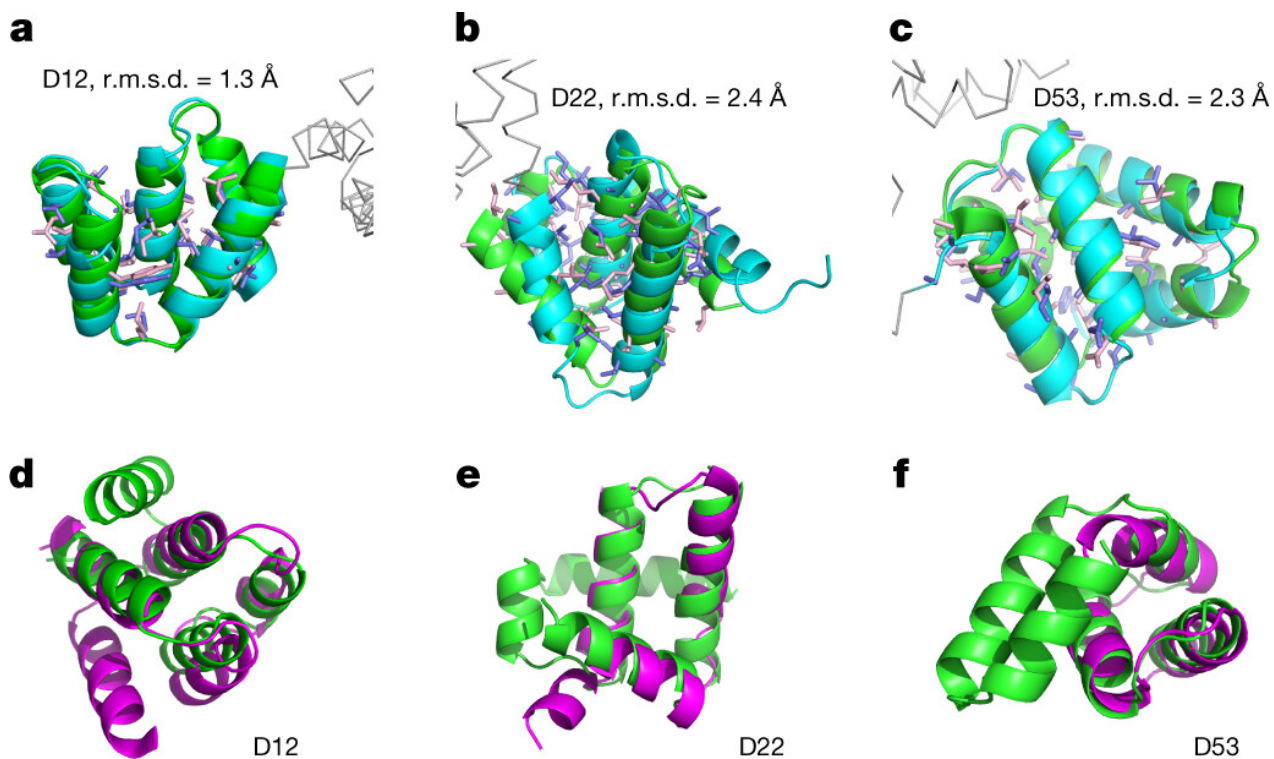


Fig. 4: 新規構造に折り畳まれる*De novo*タンパク質の設計に成功した構造

新規構造デザインの検証

- 実験的に調べた3番目のタンパク質は、計算によって見出された新しい構造を持っていた。
- これらのall-helical構造は、最初にランダムに配置されたhelical主鎖セグメントをSCUBA駆動型SASDで最適化した後、セグメントの切断と順序付けを行い、SCUBA最適化ループでセグメントを接続することで生成した。
- Batch3（13個）の中の3個（D12,D22,D53）はモノマーで精製され、うまく折りたたまれていることが確認された。
- この3つのタンパク質を（別々に）マンノース結合タンパク質（MBP）に融合させることで、タンパク質の結晶化が可能になった。

Fig. 4

Fig.4a-c

- 3つの設計モデル（緑色）と各結晶構造（水色）と重ね合わせ

Fig.4d-f

- 3つの設計モデル（緑）と類似天然部分構造（紫色）の重ね合わせ
 - 部分的に大きく差異があることから、新規性を示している

総合的な結果

- これらのタンパク質の設計におけるSCUBAの成功は、下記を強調するものである。
 - 現在PDBで知られているタンパク質構造は限られていること（可能なタンパク質構造の膨大さと比較すると）
 - SCUBA駆動設計はモデルの学習に使用される天然タンパク質構造に制限されないこと

疑問点・課題

- Fragment Assemblyなどを中心に、これまでの代表的な手法が網羅されていた。
- エネルギー関数の精度の違いに関する記述があった。非極性側鎖のコアパッキングなどが関係するかも。

手法の詳細（補足内容）

- 1章：背景
- 2章：NC-NNによる統計ポテンシャルの構築
- 3章：骨格中心SCUBAモデル
- 4章：SCUBAを用いたテンプレートフリータンパク質設計

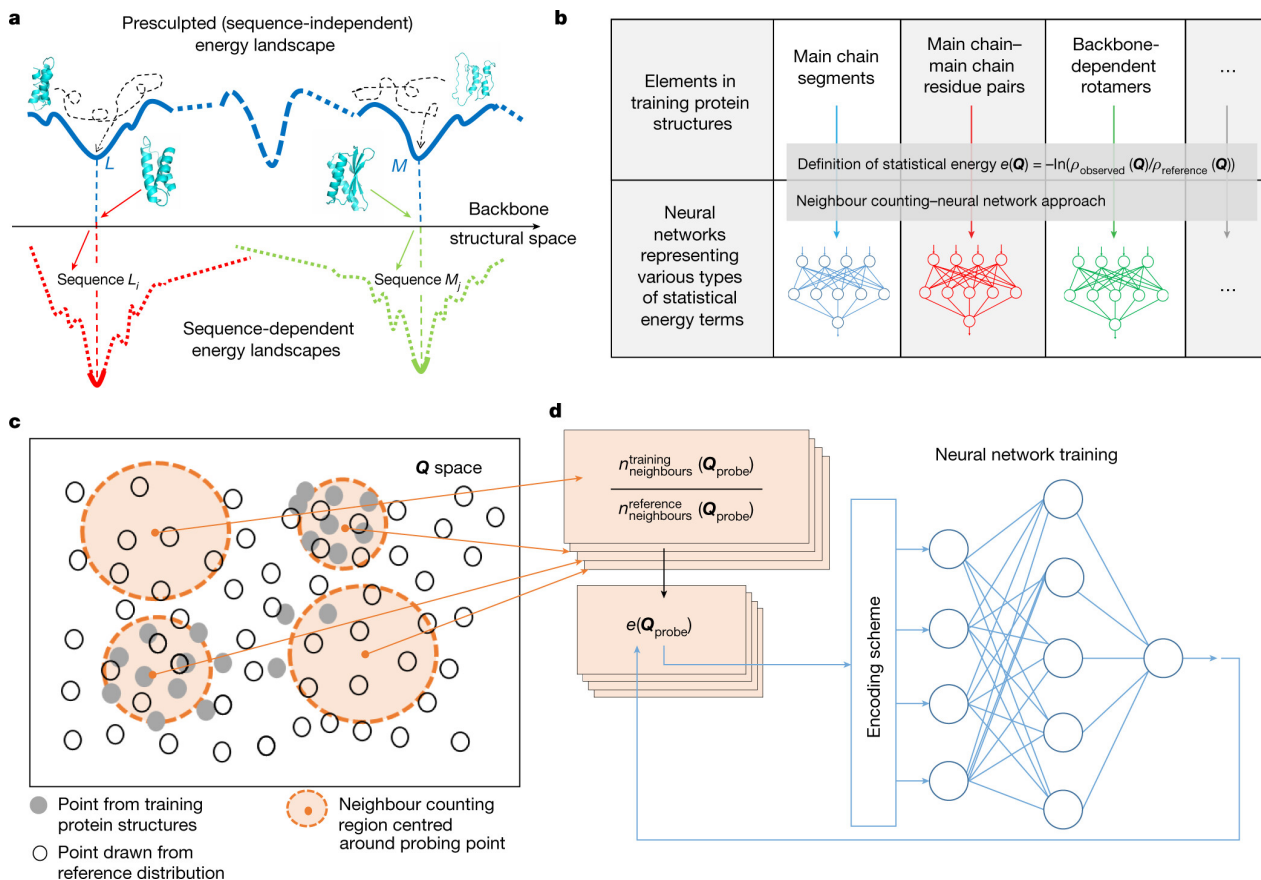


Fig. 1: Template-free protein design facilitated by explicit representation of the backbone-centred energy landscape.

1章：背景

そのために、NC-NNと名付けた一般的な手法で訓練した統計的エネルギー項（ポテンシャル）を用いて、様々な相互作用を表現します（図1b-d）。この名称は、まず生の構造データからカーネルベースの密度推定（つまり、隣接カウント）により統計的エネルギー値を推定し、次にポテンシャルを表現するためのニューラルネットワーク（完全連結3層パーセプトロン：図1dおよび補足方法）の学習を行うという2段階のプロセスを表しています。NC-NNは統計的ポテンシャルの構築における主な技術的課題を解決し、得られたポテンシャルは連続的であることに加え、構造サンプリングや最適化に使用する計算しやすい関数値（および微分）を提供することができます。は、実際の構造物の複雑で高次元かつ高相関の分布を表現することができる。のデータを忠実に再現することができます。

我々は、SCUBAを用いた確率論的動力学（SD）シミュレーション26とデータ駆動型固定バックボーンアミノ酸配列選択プログラムABACUS2（参考文献27,28）を併用し、SCUBAによる設計可能なバックボーン空間の計算探索により、様々な設計仕様を満たすトポロジカル構造を持つタンパク質や新規でゼロから設計されたタンパク質を設計しました。我々は、それぞれ独自に設計された配列と構造を持つ9つのde novoタンパク質のX線構造を報告する。

2章：NC-NNによる統計ポテンシャルの構築

形式的には、構造変数の集合（ Q とする）に依存する統計的ポテンシャルは、 $e(Q) = -\ln(p_{\text{observed}}(Q)/p_{\text{reference}}(Q))$ として定義することができる。ここで $p_{\text{observed}}(Q)$ は Q 空間に分布する観測データの確率密度、 $p_{\text{reference}}(Q)$ は理想化した参照系（つまり、相互作用 $e(Q)$ を欠いた系）の同じ変数の期待確率密度を示す。 $e(Q)$ を学習するためのニューラルネットワーク学習(NN)に続く近隣カウント(NC)のワークフローを図1c, dに示す。

NCステップでは、 Q 空間（図1cのボックスで表現）において選択された任意のプロブ点 Q_{probe} について、その一点エネルギー $e(Q_{\text{probe}})$ を、隣接する観測データ点の数 $n_{\text{neighbours}}^{\text{observed}}(Q_{\text{probe}})$ と隣接する参照データ点の数 $n_{\text{neighbours}}^{\text{reference}}(Q_{\text{probe}})$ との比として推定する。実構造物からの観測データ点と計算で求めた参照データ点を用いることで、カーネル法を用いてこの二つの数値を推定することができる。カーネルの半径は学習データの局所密度に基づいて適応的に選択することができ、推定エネルギーの分解能と統計的不確実性の間のトレードオフをバランスさせることができる。

NNステップ（図1d）では、異なるプロービングポイントにおけるNC推定エネルギーを用いて、統計的ポテンシャルを Q の解析的関数として符号化するためのニューラルネットワークを学習させる。NCステップでは1点のエネルギーしか推定できず、離散的でノイズの多い値が得られるため、NNステップは元のデータを参照せずに効率的に計算できる解析的ポテンシャルを得るために不可欠なステップです。

特徴をビンに分けて統計的ポテンシャルを学習する従来のアプローチと比較した場合のNC-NNの主な利点は、以下の通りである。NC-NNは高次元空間における相関を忠実に保持することができる

る。例えば、拡張データ図1は、NC-NNで学習したSCUBAエネルギー項の14変数による低次元投影図である。このデータはNC-NNによって捉えられた強い次元間相関を明確に示しており、観測された（主に好ましい）構成と計算で描かれた（主に好ましくない）構成を区別することができる。低次元、高次元ともに、NC-NNによって学習されたポテンシャルは、(1)自動的に連続となり、(2)解析的勾配を持つため、勾配駆動の構造サンプリングや最適化に適している。NC-NNは複雑なエネルギー面をニューラルネットワークで表現するという以前のアイデアに基づいているが29、生の構造データを用いて一点エネルギーを得るという点で、量子力学計算の利用とは根本的に異なることに注意されたい。

3章：骨格中心SCUBAモデル

SCUBAは、バックボーンを中心としたエネルギーランドスケープの様々な構造変数への依存性を包括的に検討するために考案されました。5種類のNC-NN学習されたエネルギー項を用いている（Extended Data Fig.2a）。学習データは、Protein Data Bank (PDB)31にある12,000以上の非冗長自然構造27,30から構成されている。ニューラルネットワークの項を通常の共有結合項および立体項と組み合わせて、解析的に原子直交座標に依存する有効エネルギー関数を形成した（詳細は補足方法を参照）。エネルギー関数の（座標に対する）負の勾配を力として使用できるため、SCUBAはタンパク質構造のニュートンまたはランジュバン分子力学（MD）またはSDシミュレーションの駆動に使用することができる。

バックボーン中心で開発されたが、SCUBAはバックボーンに依存する項と明示的な側鎖に依存する項の両方を含んでいる。側鎖が指定されていない場合、背骨の原子の位置が「不鮮明」であることがわかったため、側鎖を含めることにした（その結果、その後の配列選択で50%以上の位置でアラニンまたはグリシン残基が得られた）。側鎖の種類に依存しない有用なレベルを維持しながらこのばやけを最小限に抑えるため、SCUBAでは側鎖が関与する極性相互作用と溶媒和を省き、バックボーン依存のロータマーと立体充填項のみを用いて、明示的に側鎖を考慮しました。この方法でモデル化すると、一般的に選択された側鎖（例えば、以下に述べる「LVG」形式の側鎖）は、バックボーン間の空間を埋めるプレースホルダーとして機能し、最終的に最適化されたバックボーンは、密に充填されたキラル側鎖を収容するための適切な空間を持つことができるようになりました。

33種類の天然タンパク質のSCUBA駆動SDシミュレーションを用いて、SCUBAの様々なエネルギー項の相対的な重みを校正した。最終的に決定されたエネルギー重みをシミュレーションに使用した場合、中央値 主鎖の二乗平均平方根偏差（RMSD）が自然界と異なる。構造は、ネイティブの側鎖を使用した場合1.60Å、2.78Åでした。LVGで簡略化された側鎖（後述）を使用した場合（Extended Data Fig. 2b; 全体的な構造のコンパクトさを抑制した場合 シミュレーションの結果、r.m.s.d.の中央値は、天然配列、LVG配列を用いた場合は2.23Å）。

4章：SCUBAを用いたテンプレートフリータンパク質設計

バックボーン生成にSCUBA、配列選択にABACUS2（文献28）を用いたde novoタンパク質設計ワークフローを考案した。初期バックボーン（例えば、Extended Data Fig. 3に示した手順で得られ

たもの) から出発し、SCUBAによるシミュレーテッドアニーリングによるSDシミュレーション (SASD) を用いて、設計可能であると推定される最適化バックボーンを得ることができる。SASDは、側鎖を考慮しない段階と、側鎖を考慮する段階の2つの段階に分けて実施することができます。(また、一般的に選択された側鎖を考慮した場合(例えば、 α ヘリックスにはロイシンを、 β ストランドにはバリンを均質に使用し、ループには側鎖を使用せず、LVG配列となる)、LVG配列は、そのような二次構造のバックボーンになります。) LVGで単純化された側鎖は比較的特徴がなく、中程度の大きさで、対応する二次構造のタイプと一致する好ましいバックボーンコンフォメーションを持っており、バックボーン中心のモデルで適切な側鎖プレースホルダーを得ることができます。最終的な配列を選択する際に、このような側鎖で最適化されたバックボーンを微調整して特定の側鎖を収容するために、我々は反復配列選択-バックボーン緩和アプローチ6を使用して、ABACUS2 と SCUBA-drivenのSASDを逐次反復する。

SCUBA による SASD プロトコルを適用し、様々なトポロジカルアーキテクチャのスケッチの仕様を満たすバックボーンを最適化しました。スケッチ(つまり目標とする構造)は、球状タンパク質構造の「周期表」抽象化に基づいて定義された32, を用い、それに合わせて初期バックボーンを幾何学的に構築した(Supplementary Methods and Extended Data 図3)。SCUBAを用いたSASDは、スケッチの仕様を満たすバックボーンを一貫して生成するとともに、天然のバックボーンを高い精度で再現しました(Extended Data Fig. 4)。

目的の構造に酷似した初期構造(例えば、拡張データ図3のような初期構造)をSASDの出発点とすることで、構造空間での不必要な探索を避け、トポロジカルアーキテクチャを満たすタンパク質の設計を効率的に行うことができます。このような初期構造は任意であり、むしろ「不格好」である可能性さえあります。特に、「鎖」は鎖間水素結合を形成するために配置や向きを変える必要がない。なぜなら、その後のSCUBA駆動型SASDでは、水素結合が自動的に形成、切断、再構成され、最終的に β シート全体にわたる水素結合ネットワークが形成されるからです(拡張データ、図5a)。SASD 最適化プロトコルが初期バックボーンにほとんど制約を与えないことを考えると(デザインへの配慮は別として)仕様および/または効率)、拡張データFig.3に示した方法の例以外にも、多様な方法で初期構造を構築することができます。

SCUBA駆動型SSDでは、拡張データ図4と図5aの構造を見るとわかるように、バックボーンは初期構造に対して大きく進化しています。これは、拡張データ図5bの大きなr.m.s.d.値(6Åを超える)からも明らかである。この過程で、バックボーンの物理的な妥当性が大幅に改善された。このことをin silicoで実証するために、拡張データ図4のスケッチの初期および最適化バックボーンにABACUS2配列を選択し、残基あたりのABACUS2エネルギーとRosettaエネルギーを計算した(拡張データ図5c)。最適化されたバックボーンに対して決定されたエネルギーはと比較して、常に1~2エネルギー単位低くなっている。この結果は、設計のしやすさの向上と一致する。

側鎖を考慮しない) 最初のバックボーン最適化の後、さらにSASD最適化(配列選択前はLVG側鎖、配列選択後は配列特異的側鎖を使用)を行った結果、バックボーン構造は微妙に変化しました(LVG側鎖追加によるr.m.s.d.は約3Å、さらなる側鎖更新では1.1Å以下; 拡張データ図5bのスケッチの例参照)。このことは、側鎖の種類は最適化されたバックボーンに実質的に影響を与えな

いという考えを支持しています。しかし、意味のある配列選択を行うためには、第1段階からそれ以降の最適化段階までのバックボーンの調整が必要でした（Extended Data 図5d）。

初期プロトコルで設計した1つのde novoタンパク質について結晶構造を解いたが、実験結果から問題が明らかになった。このプロトコルは、設計可能な正確なループ構造を一貫して生成しないのです。つまり、バックボーン全体を一度に最適化すると、高エネルギーのローカルミニマムに捕捉されたループが含まれる可能性があり、ループの歪みが少ないことが設計可能なバックボーンの特徴であると理解されていることから、当然ながら設計の失敗の原因となります8, 10, 33. この問題を解決するために、ループの再サンプリングと最適化のステップ（SCUBA-driven SDも使用）を追加して、バックボーン最適化プロトコルを改良しました（Supplementary Methods）。2本らせん/4本鎖（H2E4）スケッチ（拡張データ図5a）に適用したところ、改良されたプロトコルはループ領域のエネルギーを大幅に改善し（拡張データ図6a）、Rosettaバイアス順折りたたみ10、15を用いた場合の最終配列の計算による折りたたみの成功率を改善しました（拡張データ図6b、c）。

以前のパート

- タンパク質のコンフォメーションエネルギーランドスケープは、高次元で、多くの局所的最適解（局所解）を持つ（多峰性）ため、複雑である。
 - x が各コンフォメーション、 E が各エネルギーとしたとき、 x に対する $E(x)$ の景観のこと。
- 予測や設計のために低エネルギーの領域を見つけるためには、効率的なサンプリング方法と正確なエネルギー関数が必要である。
- Box.1 左上図：勾配法（Gradient-based）
 - エネルギーが最も急速に減少する方向に進むように、エネルギー関数の導関数（ $dE(x)/dx$ ）を計算する。
 - dx は原子座標や骨格のねじれ角などの微小変位
 - 探索開始位置から最も近い局所解を探索するには効果的だが、多峰性では局所解に収束してしまい、大域的最適解を探索することは困難。
- Box.1 左下図：Monte Carloサンプリング（Sampling-based）
 - ランダムに選択されたコンフォメーションの移動と、局所解を回避するための脱出を使用する。
 - Local Searchは、近傍解を生成し、現在の解より改善するなら移動する（改善移動）ため、付近の局所解に必ず収束する（脱出不可）。
 - 近傍解への移動を「受理」と呼ぶ。
 - Metropolis-Hastings法では、悪化する場合でも確率的に受理するため、局所解の脱出が可能。
 - 改善する場合は、必ず受理する。

- 悪化する場合は、受理確率 p を満たす場合だけ受理する。受理確率 p はエネルギーの変化量に依存する。
- $$p = \frac{E(x_{\text{new}}) - E(x_{\text{now}})}{k_B T}$$
- Box.1 右下図：移動例
 - 移動例として、Fragment置換や側鎖ロータマー置換がある。
 - Fragment置換：現在のコンフォメーションの骨格セグメントを、Fragmentライブラリのコンフォメーションに置き換える。
- Box.1 右上図：MD
 - Monte Carloサンプリングに代わるものとして、MDがある。
 - 原子の位置と速度が与えられたとき、ポテンシャルエネルギーの勾配から各原子に働く力 F を計算し、ニュートンの第二法則（ $F = ma$ ）から加速度 a を導出する。
 - 非常に小さなステップ間隔（通常は数femt second（ 10^{-15} ）オーダー）で、新しい位置と速度を計算する。
 - 正確なエネルギー関数と十分に小さな時間ステップがあれば、長時間のMDによってエネルギーランドスケープの幅広いサンプリングが得られ、個々の分子が時間とともにどのように変化していくかを現実的に示すことができる。
 - MDに基づくモデリングの課題は、生物学的に適切な時間スケールに到達するためには、数百万から数兆の時間ステップの実行が必要なことで、場合によっては高性能ソフトウェアや特殊なスパコンを使う。
- タンパク質は、直鎖のアミノ酸配列がFoldingし、自然な立体構造（天然状態）を形成する。
 - 配列の情報は自由エネルギーのランドスケープにコード化されており、エネルギーが最も低い状態が天然状態に対応する。
- 構造予測手法では、この基礎概念に基づき、代替構造のサンプリングと、それらをエネルギーでランク付けし、最もエネルギーの低い状態を特定するようにスコアリングする。
- このアプローチにおける問題点は、成功を阻む最大の障害は、コンフォメーションが潜在的に取り得る空間が膨大なことである。
- この解決策は、コンフォメーション空間の全てを探索せず、効率的に探索することである。

Fig.1 a：エネルギーランドスケープの例

- エネルギーランドスケープは、フラットではなく漏斗状で、サンプリングが自然に近いコンフォメーションに導かれる。
- このときの探索方向は、局所的な残基相互作用や非局所的パッキング相互作用から生じることがある。
 - 局所的な残基相互作用：鎖の短い部分が特定の二次構造を形成するように偏る
 - 非局所的パッキング相互作用：グローバルな自然なFoldに到達する前でも形成可能

Fig.1 b : 立体構造の例

- 水溶性の球状タンパク質がFoldingする原動力は、疎水性側鎖を水から遠ざける力だと考えられている。
- タンパク質のコアに非極性側鎖が密に配置されていると、ファンデルワールス相互作用が強化され、エントロピー的に好ましくない内部空洞が排除される。
- さらに、このパッキングは、ねじれ角分布を制限する強い骨格と、側鎖のねじれ優先性に対応しながら達成されており、側鎖の柔軟性を各位置でロータマーの近傍にまで効果的に減少させている。
- 極性基はFolding中に埋没するため、タンパク質内の水素結合と塩橋【comment1】は、水との相互作用の損失を大きく補う。
- したがって、これらの相互作用は天然状態の安定性よりも、その特異性（天然状態と他の状態を識別する）に寄与する。

comment1 : 塩橋とは、カチオン性の残基（アルギニン、リシン）とアニオン性の残基（アスパラギン酸、グルタミン酸）の間に働く弱いイオン性相互作用のこと。タンパク質の高次構造を安定化させる。

配列探索

- 確率論的 : Simulated AnnealingやGenetic Algorithmなど
- ほとんどの手法は、側鎖の動きをPDBの高解像度で観察される一般的なコンフォメーション（またはロータマー）のセットに制限することで、側鎖最適化問題を簡略化している。
- 設計ソフトウェアRosettaは、Simulated Annealingを用いたMonte Carloサンプリングによって、ロータマー置換していき、低エネルギーの配列とロータマーを同定する。
 - 確率的アプローチの限界は、最もエネルギーの低い配列が同定されることが保証されないことである
- タンパク質設計パッケージOSPREYは、最低エネルギー配列を見つけるだけでなく、最低エネルギー配列のランク付けを行う手法も備えている。

参考になる文献
