

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1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マップを明示的に表現し、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マップを探索したときは、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]。
 - 主鎖の局所構造の傾向、水素結合の形状、(キラリティや密に配置された) 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。
- エネルギー関数駆動型的主鎖サンプリングと最適化は、機能性タンパク質設計においても、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と考えられる。
- *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期待される。

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ある。
 - 側鎖の種類は最適化された主鎖に実質的に影響を与えないという考えを支持している

NC-NNによる統計ポテンシャル

- NC-NN (Neighbour Counting-Neural Network) 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する。
 - NC : 構造データからカーネルベースの密度推定 (隣接するデータのカウンタ) により統計的エネルギー値を推定する
 - NN : 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) の学習を行う

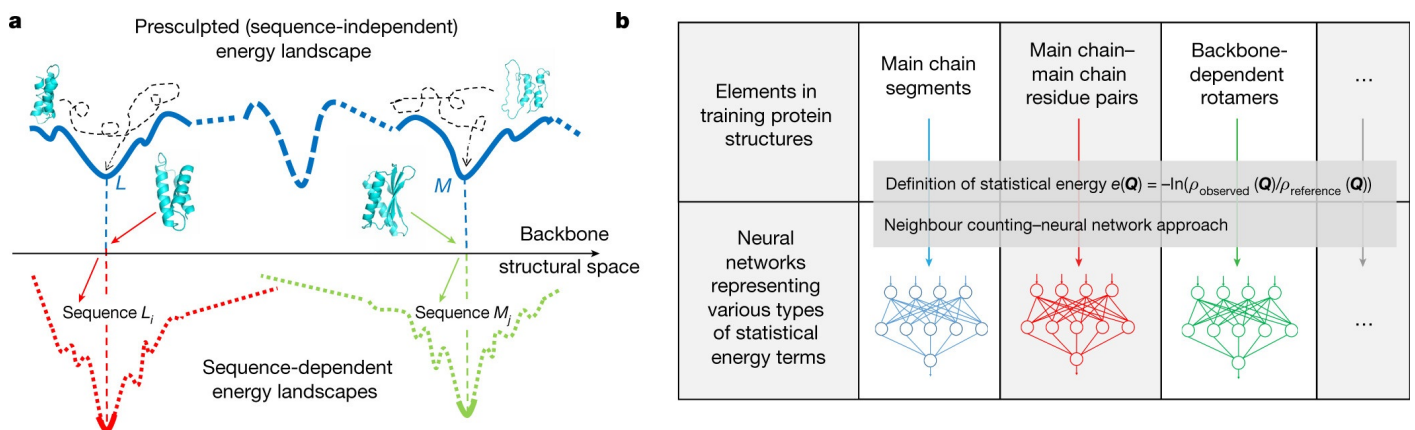
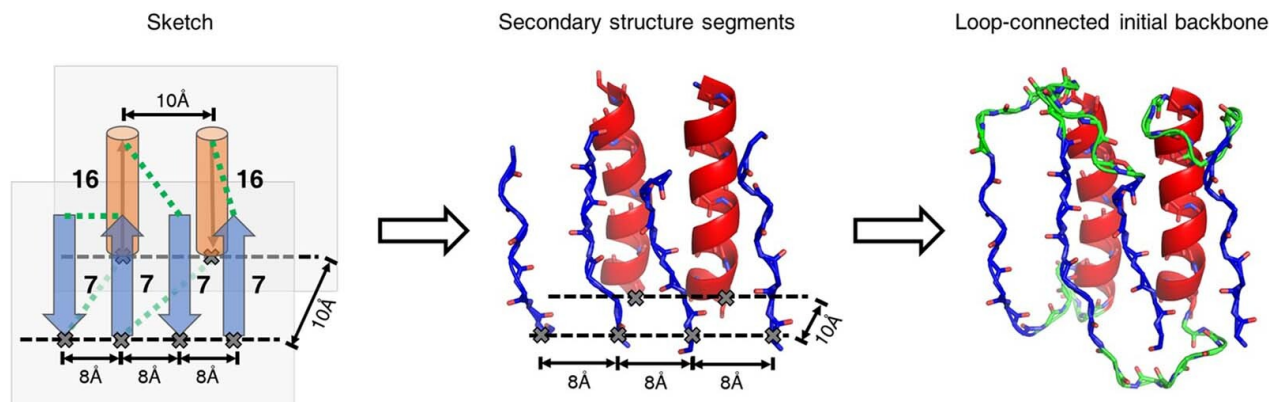


Fig. 1a,b: 主鎖エネルギーランドスケープの明示的表現によって促進されるテンプレートフリータンパク質設計



Extended Fig. 7: 既存のスケッチやトポロジカル構造による初期主鎖生成

SCUBAを用いた設計フレームワーク

- SCUBAは、主鎖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た。
 - NC-NNで学習した5種類のエネルギー項を加重和し、天然タンパク質で重みを校正
- 本手法は、下記で構成される設計フレームワークである。
 - 主鎖最適化：SCUBA駆動SASD
 - 主鎖評価：NC-NNベースのSCUBA
 - 主鎖サンプリング：確率論的動力学（SD）シミュレーション^[26]
 - 主鎖最適化：Simulated Annealing（SA）
 - アミノ酸配列選択：ABACUS2^[27,28]
- 初期主鎖を決め、SCUBA駆動SASDによって最適な主鎖を求める。
 - 初期主鎖は、スケッチ（目標構造）の二次構造を与え、loopで接続する手順で構成する
- その後、最適主鎖を微調整して特定の側鎖を収容するために、配列選択（ABACUS2）と主鎖緩和（SCUBA駆動SASD）を逐次反復することで、最終的な配列を選択する。
- 論文中では、loopのリサンプリングと組み合わせた改良プロトコルも提案している。

Extended 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）

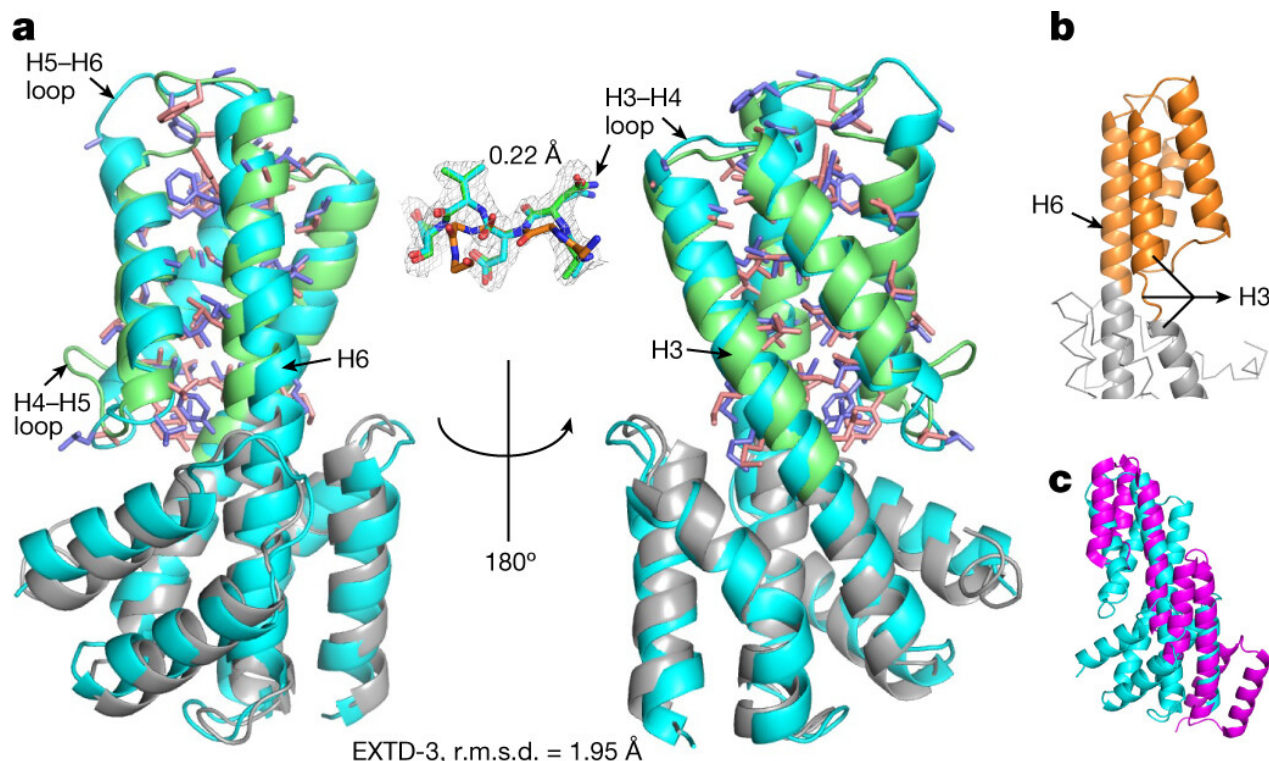


Fig. 2: *De novo*タンパク質EXTD-3は、天然ではまだ観察されていない単一の剛体構造を形成するために、既存のパーツと新規設計されたパーツが統合される

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。
- EXTDデザイン（8個）の中の1つ（EXTD-3）について、解明された結晶構造が得られた。

Fig. 2a

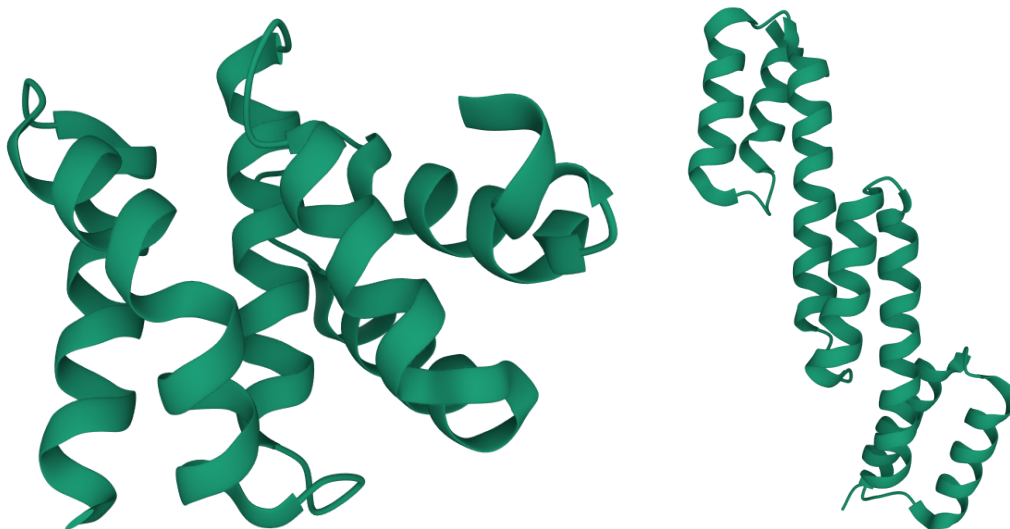
- 実験的に解明された構造（水色リボン）とEXTD-3（灰色+緑色リボン）の重ね合わせ
 - 灰色リボン：既存部位、緑色リボン：新規設計部位
 - tepsin ENTH domain; PDB, 5WF2（運搬体タンパク質）
 - スティック：側鎖（青色：実験、橙色：新規設計）
- 解明された構造では、設計上導入された3つのloop（H3-H4, H4-H5, H5-H6）のうち、H3-H4 loopは正確に一致している一方、他の2つは大きく外れている。

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（免疫系）
 - 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



App. Fig: 5WF2（左図）、5H79（右図）

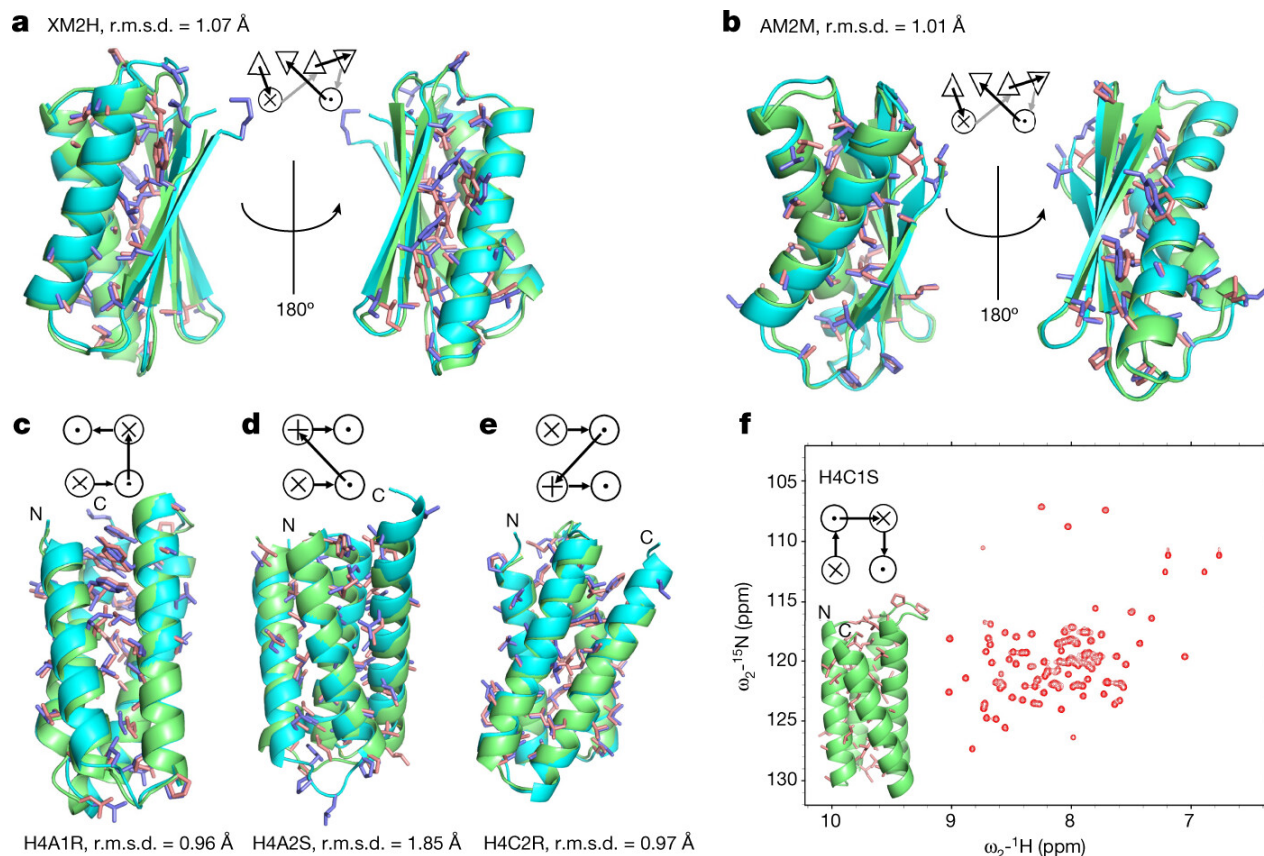


Fig. 3

- 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ABACUS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 Table 1: Summary of experimentally examined designs (Batch3)

Batch	Name	Architecture	Number
3	Novel all-helical	helix (D12,D22,D47,D53)	13

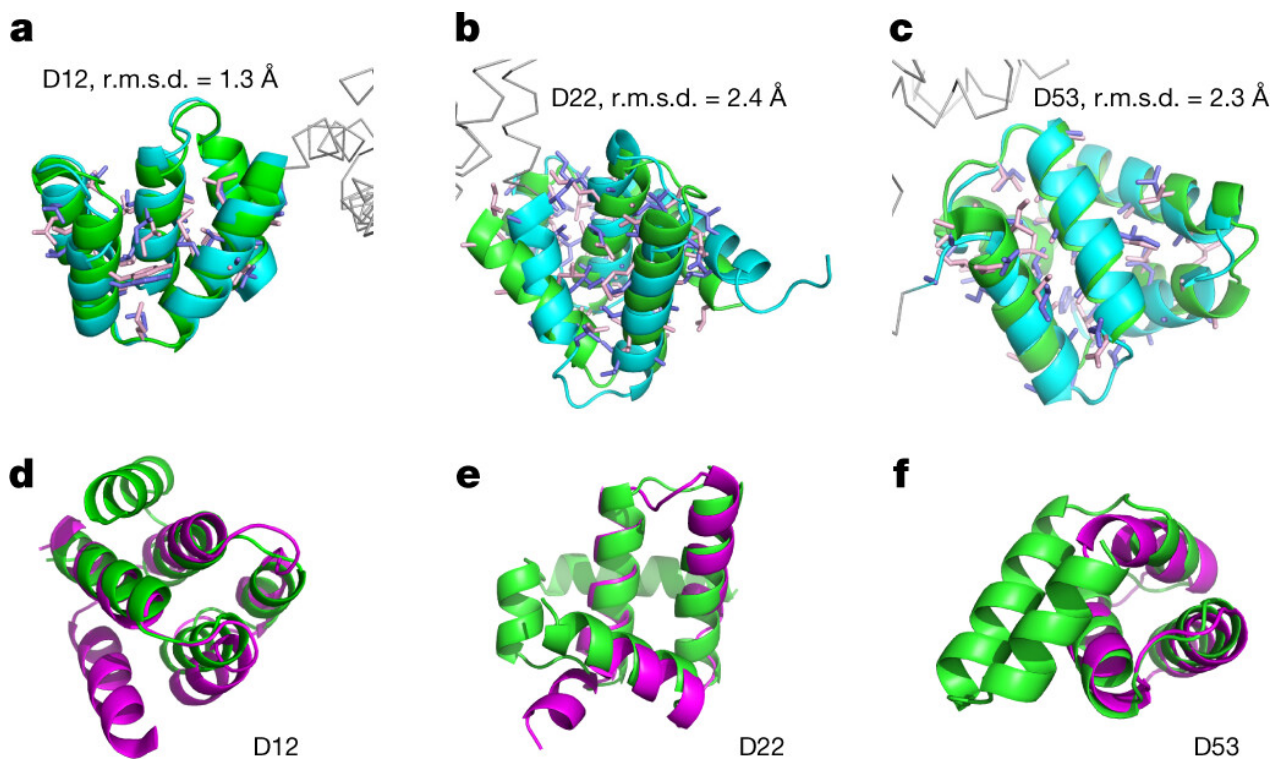


Fig. 4: 新規構造に折り畳まれる*De novo*タンパク質の設計に成功した構造

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（緑色）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）

- NC-NNによる統計ポテンシャルの構築
- 主鎖SCUBAモデル
- SCUBAを用いたテンプレートフリータンパク質設計

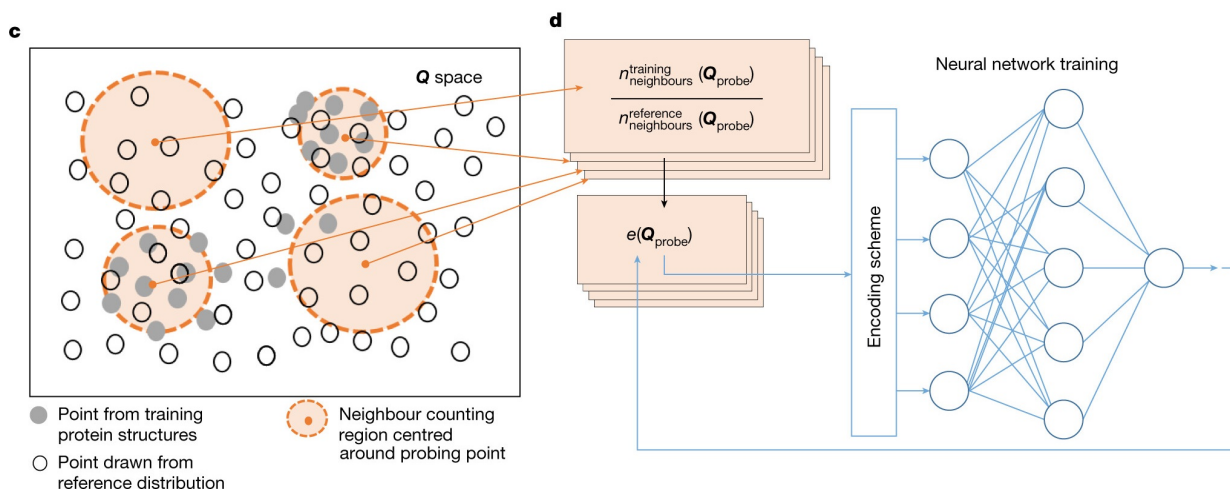


Fig. 1c,d: NC-NNの学習方法

NC-NNによる統計ポテンシャルの構築

- 構造変数の集合 Q に依存する統計的ポテンシャル $e(Q)$ は形式的に定義される。
 - $e(Q) = -\ln(\rho_{observed}(Q)/\rho_{reference}(Q))$
 - $\rho_{observed}(Q)$: Q 空間に分布する観測データの確率密度
 - $\rho_{reference}(Q)$: 理想化した参照系の同じ変数の期待確率密度（相互作用 $e(Q)$ を欠いた系）
- $e(Q)$ 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Fig. 1c, dに示す。

NCステップ

- NCステップでは、 Q 空間 (Fig. 1cのボックスで表現) において選択された任意のプローブ点 Q_{probe} について、その一点エネルギー $e(Q_{probe})$ を、隣接する観測データ点数／参照データ点数の比として推定する。
 - $e(Q_{probe}) = n_{neighbours}^{observed}(Q_{probe}) / n_{neighbours}^{reference}(Q_{probe})$
 - $n_{neighbours}^{observed}(Q_{probe})$: 隣接する観測データ点の数
 - $n_{neighbours}^{reference}(Q_{probe})$: 隣接する参照データ点の数
- 実際の構造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

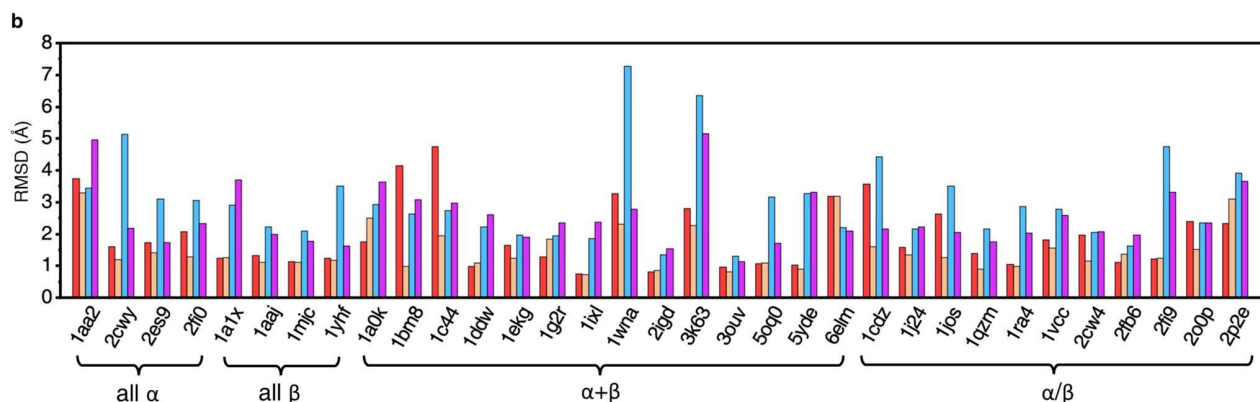
- NNステップ (Fig. 1d) では、異なるプロービングポイントにおけるNC推定エネルギーを用いて、統計的ポテンシャルを Q の解析的関数として符号化するためのニューラルネットワークを学習させる。
- NCステップでは1点のエネルギーしか推定できず、離散的でノイズの多い値が得られるため、NNステップは元のデータを参照せずに効率的に計算できる解析的ポテンシャルを得るために不可欠なステップである。

NC-NNの特徴

- 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。
- 実際の構造の複雑で高次元かつ高相関の分布のデータを忠実に再現することができる。
 - Extended Fig. 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
 - 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
- 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アイディアに基づいているが、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。

a

Type	Structural variables	Reference distribution	Sidechain involved?
$e_{NC-NN}^{\varphi-\psi-1}$	2 Ramachandran torsional angles of one residue, encoded by 12 input nodes.	uniform distribution in (φ, ψ) space	No
$e_{NC-NN}^{\varphi-\psi-5}$	8 Ramachandran torsional angles of a window of 5 consecutive residues, excluding the φ angle of the first and the ψ angle of the last residue, encoded by 64 input nodes.	multiplications of the observed 1-D or 2-D distributions of the Ramachandran angles of individual residues	No
$e_{NC-NN}^{site-pair}$	translations and rotations between the main chain atoms of two residues that are sequentially far-separated (sequence separation > 5) and spatially close (inter-Ca distance below 9.5 Å), and the two sets of four Ramachandran torsional angles, each set centered around one residue, encoded by 191 input nodes.	uniform distributions of inter-residue relative position and orientation, multiplied by the observed distributions of the four surrounding Ramachandran angles surrounding each residue. Sterically clashed configurations excluded	No
$e_{NC-NN}^{local-HB}$	internal geometries of a group of 6 atoms from two peptide units: $C_i, O_i, N_{i+1}, C_j, O_j$, and N_{j+1} in which i and j are sequential residue numbers and $j - i \leq 5$, encoded by 35 input nodes.	uniform distributions of the relative position and orientation between the two peptide units, excluding sterically clashed configurations	No
$e_{NC-NN}^{rotamer}$	sidechain and backbone rotatable torsional angles of one residue, each angle encoded by 8 input nodes.	uniform distributions of the sidechain torsional angles multiplied by the observed 2-D distribution of the Ramachandran angles	Yes



Extended Fig.2: NC-NN-learned components in SCUBA and simulations of natural protein structures by SCUBA.

主鎖中心SCUBAモデル

SCUBAによる主鎖エネルギー関数

- SCUBAは、主鎖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た。
 - 5種類のNC-NN学習されたエネルギー項を用いている
- 学習データは、PDBにある12,000以上の非冗長自然構造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。
- 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。

SCUBAモデル

- SCUBAは主鎖中心で開発されたが、主鎖に依存する項と側鎖に明示的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ぼ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、主鎖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、後述の「LVG」形式の側鎖）は、主鎖間の空間を埋めるブレースホルダーとして機能し、最終的に最適化された主鎖は、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った。

重み校正

- 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とき、主鎖のRMSD中央値は1.60Å（天然側鎖を使用）、2.78Å（LVGで簡略化された側鎖を使用）であった。
 - シミュレーションで全体的な構造のコンパクトさを抑制した場合は、RMSD中央値は天然配列で1.25Å、LVG配列で2.23Åであった（Extended Data Fig. 2b）。

SCUBAを用いたテンプレートフリータンパク質設計

ワークフロー

- 主鎖生成にSCUBA、配列選択にABACUS2を用いた*De novo*タンパク質設計ワークフローを考案した。
- 初期主鎖（例えば、Extended Fig. 3に示した手順で得られたもの）から出発し、SCUBAによるSimulated Annealing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な主鎖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る。
 - また、一般的に選択された側鎖を考慮した場合、LVG配列はそのような二次構造の主鎖になる
 - 例えば、 α helixにはロイシンを、 β strandにはバリンを均質に使用し、loopには側鎖を使用せず、LVG配列となる
- 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主鎖コンフォメーションを持っており、主鎖中心のモデルで適切な側鎖ブレースホルダーを得ることができる。

- 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主鎖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-主鎖緩和アプローチを使用して、ABACUS2 と SCUBA駆動のSASDを逐次反復する。

スケッチ

- 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主鎖を最適化した。
- スケッチ（目標構造）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ものを用い、それに合わせて初期主鎖を幾何学的に構築した（Extended Fig. 3）。
- SCUBAを用いたSASDは、スケッチの仕様を満たす主鎖を一貫して生成すると同時に、天然の主鎖を高い精度で再現した (Extended Fig. 4)。

初期スケッチの工夫

- 目的の構造に酷似した初期構造（例えば、Extended Fig. 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る。
- このような初期構造は任意であり、むしろ「不格好」である可能性もある。
- 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ため（Extended Fig. 5a）。
- SASD 最適化プロトコルが初期主鎖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）、Extended 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る。

初期主鎖の影響

- SCUBA駆動型SASDでは、Extended Fig. 4とExtended Fig. 5aの構造を見るとわかるように、主鎖は初期構造に対して大きく進化している。
 - これは、Extended Fig. 5bの大きなRMSD（6Åを超える）からも明らかである
- この過程で、主鎖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、Extended Fig. 4のスケッチの初期および最適化主鎖に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（Extended Fig. 5c）。
- 最適化された主鎖に対して決定されたエネルギーは比較して、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

主鎖の調整

- （側鎖を考慮しない）主鎖最適化の後、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、主鎖構造は微妙に変化した。
 - LVG側鎖追加によるRMSDは約3Å、さらなる側鎖更新では1.1Å以下（Extended Fig. 5bのスケッチ）
- このことは、側鎖の種類は最適化された主鎖に実質的に影響を与えないという考えを支持している。
- 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主鎖の調整が必要だった（Extended Fig. 5d）。

改良プロトコル

- 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loop構造を一貫して生成しない。
- つまり、主鎖全体を一度に最適化すると、高エネルギーの局所最適解に捕捉されたloopが含まれる可能性があり、loopの歪みが少ないことが設計可能な主鎖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る。
- この問題を解決するために、loopの再サンプリングと最適化のステップ（SCUBA駆動型SDも使用）を追加して、主鎖最適化プロトコルを改良した。
- 2本らせん/4本鎖（H2E4）スケッチ（Extended Fig. 5a）に適用したところ、改良されたプロトコルはloop領域のエネルギーを大幅に改善し（Extended Fig. 6a）、Rosettaバイアス順折りたたみ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た（Extended Fig. 6b, c）。

参考になる文献

- [33]: Bakerラボ, "What has de novo protein design taught us about protein folding and biophysics?," biorxiv (2020, 7) 【[URL](#)】[\[url_refer12\]](#)】