インシリコ解析ユニット

補助事業代表者（所属・職名）
寺田　透（東京大学大学院農学生命科学研究科・教授）

分子シミュレーションによる生体高分子の機能の予測と解析

補助事業分担者（所属・職名）
森脇　由隆（東京科学大学総合研究院難治疾患研究所・准教授）

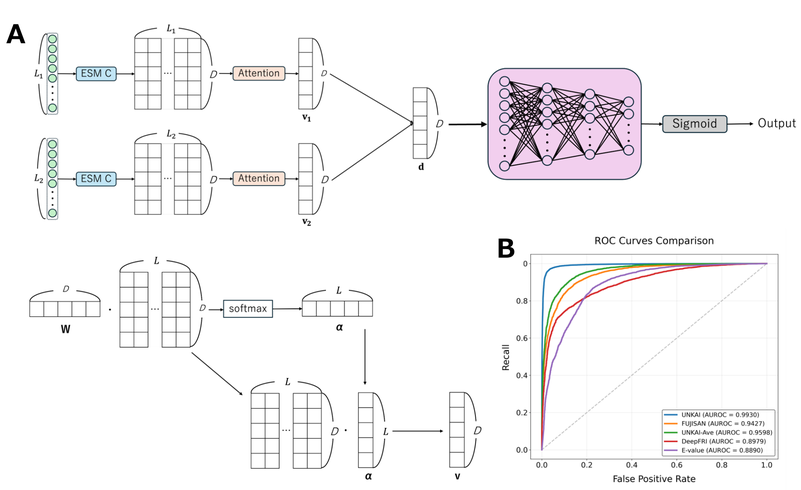
高度化における研究背景・目的

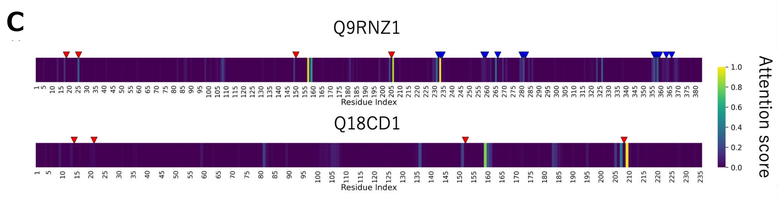
支援に資する技術開発および高度化を行うために、「予測構造を用いた基質特異性予測法の開発」、「物質輸送機構の解析法の開発」、「イオンチャネルに対する薬剤親和性予測法の高度化」、「予測構造を用いた酵素触媒機構解析法の開発」、「AlphaFoldを用いた単体構造および複合体構造予測の高度化」、「生成モデルを用いた新規タンパク質のデザインと検証」の各課題に取り組んだ。また、Phase Iの支援課題を含む支援課題を発展させた課題も高度化課題として取り組んだ。高度化における主な成果を以下に示す。

- Mirdita *et al.*, “ColabFold: making protein folding accessible to all” *Nat. Methods*, **19**, 679–682 (2022).
- Sawa *et al.*, “Comprehensive computational analysis of the SRK–SP11 molecular interaction underlying self-incompatibility in Brassicaceae using improved structure prediction for cysteine-rich proteins” *Comput. Struct. Biotechnol. J.* **21**, 5228–5239 (2023).
- Lusiany *et al.*, “Enhancement of SARS-CoV-2 Infection via Crosslinking of Adjacent Spike Proteins by N-Terminal Domain-Targeting Antibodies” *Viruses*, **15**, 2421 (2023).
- Fujita & Terada, “Enhanced prediction of protein functional identity through the integration of sequence and structural features” *Comput. Struct. Biotechnol. J.* **23**, 4124–4130 (2024).
- Zhu *et al.*, “Structure-function analysis of carrier protein-dependent 2-sulfamoylacetyl transferase in the biosynthesis of altemicidin” *Nat. Commun.* **15**, 10896 (2024).
- Furukawa *et al.*, “A dynamic structural unit of phase-separated heterochromatin protein 1α as revealed by integrative structural analyses” *Nucleic Acids Res.* **53**, gkaf154 (2025).
- Noguchi *et al.*, “Biosynthesis of the tetrahydroxynaphthalene-derived meroterpenoid furaquinocin via reductive deamination and intramolecular hydroalkoxylation of an alkene” *Chem. Sci.* **16**,7912–7920 (2025).
- Gao *et al.*, “Structural Basis for 3-Amino-3-carboxypropyl Transfer in Nocardicin Biosynthesis” *J. Am. Chem. Soc.* **147**, 33589–33596 (2025).
- Karasawa *et al.*, “ColabReaction: Accelerating Transition State Searches with Machine Learning Potentials on Google Colaboratory” *J. Chem. Inf. Model.* **65**, 11908–11914 (2025).
- Lin *et al.*, “Structural insights into glycoside hydrolase family 1 β-glucosidase: Selective oligosaccharide hydrolysis, synthesis, and product profiling” *J. Biol. Chem.* **302**, 111293 (2026).
- Teshirogi & Terada, “CGRig: A Rigid-Body Protein Model with Residue-Level Interaction Sites for Long-Time and Large-Scale Protein Assembly Simulation” *J. Phys. Chem. B*, **130**, 6290–6301 (2026).
- Moriwaki *et al.*, “High-throughput prediction of protein–protein interactions uncovers hidden molecular networks in biosynthetic gene clusters” *Comput. Struct. Biotechnol. J.* **35**, 0149 (2026).
- Miyoshi *et al.*, “Structural and Stability Analysis of de Novo Designed Proteins Incorporating a Plant Self-Incompatibility Motif” *Proteins*, in press.
- Ukai *et al.*, “UNKAI: A Protein Functional Identity Prediction Model Based on ESM-C Latent Representations and the Attention Mechanism” *Comput. Struct. Biotechnol. J.* **35**, 0206 (2026).

予測構造を用いた基質特異性予測法の開発

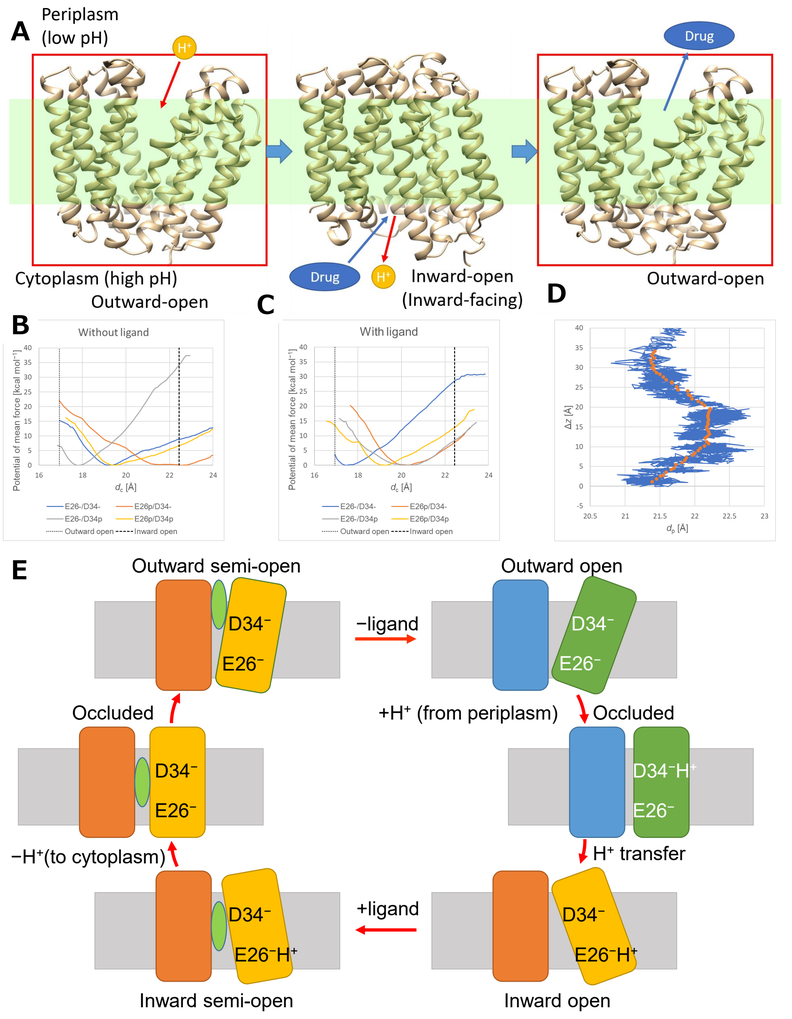
背景　冗長性を除いたアミノ酸配列データベースRefSeqに登録された配列数は2億を超えるのに対し、専門家によってその機能が確認されたタンパク質の数（Swiss-Protのエントリ数）は57万に過ぎない。本課題では、データベースに埋もれた有用な反応を触媒する酵素を見出すために、与えられた酵素が、どの基質に作用し、何を生成するかを予測する手法を開発する。我々は、アミノ酸配列全長の配列類似度に加えて、AlphaFoldによる予測構造の構造類似度を特徴量として、与えられた2つにタンパク質が同じ基質に同じ反応を触媒するかどうか（同じ機能を持つかどうか）を高い精度で判別する予測器FUJISANを開発した。しかし、この方法は、予測構造からの特徴量抽出に時間がかかるという問題があったため、アミノ酸配列の潜在表現を用いることで、より効率的に機能同一性予測を行うUNKAIの開発を行った。
方法　UNKAIのモデルを図Aに示す。まず、アミノ酸配列をタンパク質言語モデルESM Cの潜在表現に変換する。続いてAttention機構（図A下）を用いて配列長の方向に重み付き平均をとり、この差の絶対値を要素とするベクトルを多層パーセプトロンに入力し、0～1の値の出力を得る。このモデルを、Rhea IDが付与されたSwiss-Protのエントリのアミノ酸配列のすべてのペアのうち、BLASTのE-valueが10以下となるペアについて、Rhea IDが一致するペアを正例、一致しないペアを負例として、10万ペアずつランダムに選択し、訓練データ（70%）、検証データ（15%）、テストデータ（15%）に分割して使用した。
結果と考察　UNKAIのROC曲線を他の方法と比較したグラフを図Bに示す。E-valueのみに基づく方法や、深層学習に基づく方法（DeepFRI）に加えてFUJISANよりも高い予測性能を示した。Attention機構の重みの解析から、機能上重要なアミノ酸に対する重みが大きくなっていることが示された（図C）。





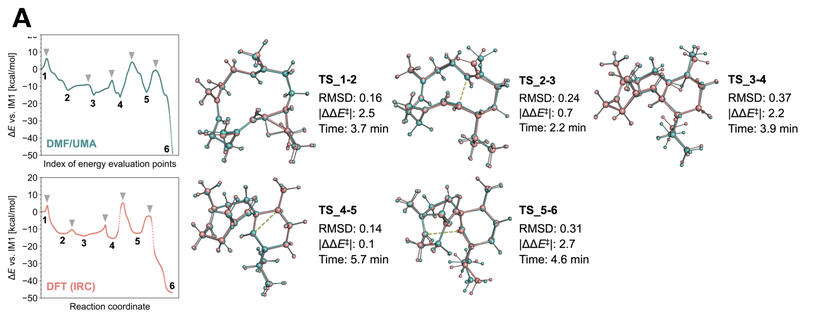
物質輸送機構の解析法の開発

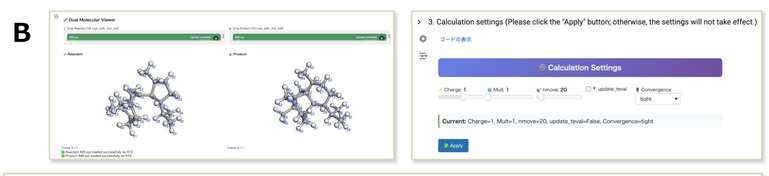
背景　本課題では、膜タンパク質の重要な機能の１つである、物質輸送のメカニズムを解析する方法を開発した。大腸菌に存在するMdfAは、Rocker-Switch機構により薬剤とH+を対向輸送すると考えられている（図A）。BINDS Phase Iの支援研究では、MDシミュレーションにより、膜貫通領域に存在するGlu26、Asp34のうち、Asp34をプロトン化すると、外開き構造、内開き構造のいずれからも、両側が閉じた閉塞構造に遷移することを明らかにした。しかし、Glu26の役割や、薬剤排出過程が未解明であったため、全薬剤輸送サイクルの解明を目指して高度化研究を行った。
方法　Glu26，Asp34のプロトン化状態と、薬剤の有無の計8種類の状態について、細胞質側の開き具合を反応座標としたmetadynamics法を用いたサンプリングを行い、続いてレプリカ交換アンブレラサンプリングを実施して、自由エネルギー地形を計算した。
結果と考察　Asp34のプロトン化によって、ペリプラズム側と細胞内側が閉じた閉塞構造が安定化することが確かめられた。薬剤非結合状態でこのプロトンがGlu26に移動すると、細胞内側が開いた内開き構造が安定化することが明らかとなった（図B）。この状態に薬剤が結合すると、細胞内側がやや閉じた構造が安定化し、さらにGlu26が脱プロトン化するとペリプラズム側も細胞内側も閉じた閉塞構造が安定化することが明らかとなった（図C）。続いて、閉塞構造から薬剤がペリプラズム側に排出される過程を明らかにするために、PaCS-MD法を用いて薬剤排出経路の探索を行い、レプリカ交換アンブレラサンプリングにより自由エネルギー地形を計算した。この結果、ペリプラズム側が少し（2 Å程度）開くだけで薬剤の放出が可能であること（図D）と、薬剤放出の自由エネルギー障壁は16 kcal mol^{-1}であることが示された。以上の結果から、図Eに示す輸送サイクルが明らかとなった。

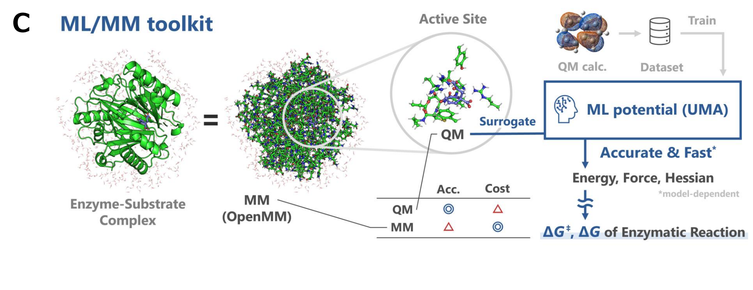


予測構造を用いた酵素触媒機構解析法の開発

背景　本課題では、まず、AlphaFoldによる酵素の予測構造を用いてQM/MM法により触媒機構を解析するパイプラインの開発を行った。QM/MM法では、QM計算に膨大な時間がかかり、多数のモデルを作成して試行錯誤を行うことができないという問題があった。この問題を解決するために、QM計算を機械学習（ML）ポテンシャルに置き換えた手法の開発を行った。
方法　MLポテンシャルとして2025年5月にMeta社から発表されたUMA（Universal Models for Atoms）を用いた。これにDirect MaxFlux法を組み合わせることで、効率的に遷移構造の探索を行った。
結果と考察　本手法をテルペンの環化反応の反応機構解析に適用し、精度を評価した。この結果、DFTによる遷移状態と比べて、平均RMSDは0.24 Å、遷移状態のエネルギーの平均絶対誤差は1.65 kcal mol^{-1}となり、十分な精度を持つことを確認した（図A）。本手法は、DFTによるScanに比べて100倍以上高速であることから、この方法を無料のクラウドサービスGoogle Colaboratoryに実装し、だれでも簡単に反応機構解析が行えるツールを開発した。これを用いると、①反応系と生成系の座標データのアップロード、②計算条件（全電荷、スピン多重度など）の設定、③計算の実行、の3ステップで反応解析を行うことができる。数分～20分程度で計算が完了すると、エネルギーダイアグラムとエネルギー極大点の構造が表示される（図B）。ここで得られた構造を、DFTなどにより最適化することで、正しい遷移状態の構造を得ることができる。私たちはさらに、従来のQM/MM法のQM領域の計算に機械学習ポテンシャルUMAを用いるML/MM toolkitを開発した（図C）。ここでは、ML領域とMM領域の境界に水素原子を付加するLink atom法を実装するとともに、解析的Hessianを用いた遷移状態探索、構造最適化、振動数解析を実装した。







AlphaFoldを用いた単体構造および複合体構造予測の高度化

背景　BINDS Phase Iの支援課題では、アブラナ科植物の自家不和合性に関わるタンパク質SRKによるSP11の認識メカニズムの研究を行った。同じハプロタイプ由来のSRKとSP11は強く結合して、受精を妨げることが知られている。この支援課題では一部のハプロタイプについてSRKとSP11の単体および複合体の立体構造予測を行い、MM-GBSA法により結合自由エネルギーを計算した。この結果に基づき、同じハプロタイプのタンパク質の親和性が高くなるメカニズムを明らかにした。ハプロタイプは200以上あるとされているため、他のハプロタイプの立体構造予測も試みたが、SP11はハプロタイプ間の配列一致度が低く、構造的な多様性も大きいため、AlphaFoldを用いても信頼度の高い予測は困難であった。そこで、本課題では、AlphaFoldに与える多重配列アラインメントを工夫することで、単体構造および複合体構造予測の精度を向上させる方法の開発を行った。
方法　AlphaFoldでは、クエリ配列に類似したアミノ酸配列を、メタゲノム解析データを含む膨大なアミノ酸配列データベースに対して検索し、多重配列アラインメントを作成するが、進化的類縁関係がなく、立体構造も類似していない配列が混じる可能性がある。そこで、文献から信頼性の高い98のハプロタイプのSP11のアミノ酸配列を収集し、ジスルフィド結合を形成するシステイン残基の位置が揃うように多重配列アラインメントを作成した（図上）。
結果と考察　このアラインメントを用いて、pLDDT値が高い予測構造を得ることができた。さらに、SRKの多重配列アラインメントも同様に作成し、信頼性の高いSRK-SP11複合体の予測構造を得た（図下）。MM-GBSA法により結合自由エネルギーを計算したところ、98の複合体モデルのうち57のモデルで、十分に低い結合自由エネルギー（十分に強い結合親和性）を与えることが示された。

