

PSILO による AlphaFold 構造の管理と解析

PSILO は、ウェブブラウザを介して利用するタンパク質立体構造データベースシステムです。分散する公共データやインハウスデータを一元管理し、独自の検索機能や解析機能により創薬研究を支援します。近年、AlphaFold をはじめとする深層学習モデルの出現により、予測構造が加速度的に増加しており、それらのデータを適切に管理して活用することの重要性が高まっています。本稿では、AlphaFold 構造に関連する PSILO の機能を紹介します。

■ はじめに：予測構造の急増とその課題

近年、AlphaFold をはじめとする深層学習ベースのタンパク質立体構造予測技術が急速に進歩し、膨大な数の高精度な予測構造が得られるようになりました。これにより、これらのタンパク質立体構造データをいかに効率よく整理・管理し、創薬・生命科学研究に活用するかが新たな課題となっています。

PSILO は、このような予測構造、および実験構造に関連するデータを一元的に取り扱い、検索・可視化・解析を可能にします。

■ PSILO による構造データの一元管理と検索機能

PSILO では、開発元が提供する v1 の AlphaFold 構造 (約 22 万件、一部の生物種を除く) をインポートして利用できます。v4 の構造もインポート可能です。

インポートされた AlphaFold などによる予測構造に加え、公共の PDB データやインハウスの構造データも、それぞれ別の論理データベースとして管理しつつ、それらを統合的に扱えます。これにより、分散しがちなタンパク質立体構造情報を、単一のインターフェイス上で一括して管理・活用できます。

PSILO の検索機能では、こうした複数のデータベースを横断的に検索でき、データの出所に関係なく一括で結果を得ることができます。例えば、タンパク質名やファミリー、結合ポケットの構造をクエリーとして、全データベースから包括的に情報を検索できます。これにより、既知構造と予測構造を同時に参照し、構造情報の全体像を効率的に把握できます。

■ AlphaFold 構造に対する専用の可視化・解析機能

PSILO には、AlphaFold などの予測構造に特化した可視化・評価機能が実装されています。

予測構造の代表的な評価指標として、各残基位置の信頼性を示す pLDDT や、相対位置関係の信頼性を示す PAE があります。PSILO では、AlphaFold Protein Structure Database¹⁾ と同様に pLDDT 値に応じて構造を色分け表示でき、高精度に予測された領域を視覚的に把握できます (図 1 左)。また、PAE プロットはグラデーションで表示され、ドメイン間の相対配置の妥当性を視覚的に評価できます (図 1 右)。図 1 の calmodulin-1 の例では、各ドメインは高い pLDDT 値を示す一方、それらをつなぐヘリックス部分は低い pLDDT 値を示し、信頼性が低いことがわかります。また、PAE プロットは市松模様を示

しており、ドメイン間の位置関係に不確かさがあることも読み取れます。

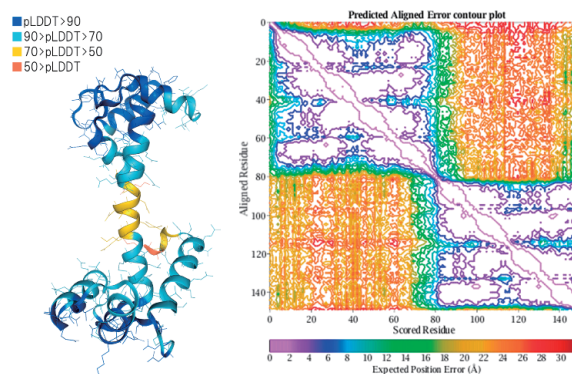


図1. AlphaFold構造の信頼性評価。Calmodulin-1 (AF-P0DP29-F1-v1) の構造の評価結果を表示。左図はpLDDT値による色分け、右図はPAEプロット。

■ 類似ポケット構造検索による新たな構造活用

PSILO では、予測構造に対しても結合ポケットを自動で検出し、それをクエリーとした類似ポケット構造検索²⁾ ができます (図 2)。これにより、ヒットした構造に結合している既知リガンドを活用したりリガンド設計や、ドラッグリポジショニングへと展開できます。また、類似したポケット構造を持つ他のタンパク質を特定することで、オフターゲット結合の可能性や副作用リスクの予測にも貢献します。

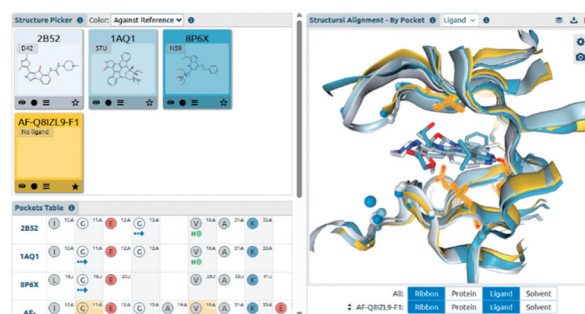


図2. Cyclin-dependent kinase 20 (AF-Q8IZL9-F1-v1) のポケット構造をクエリーとしてヒットした実験構造との重ね合わせ結果。左にAlphaFold構造 (黄色) と実験構造をリストで表示。右は重ね合わせ表示で、色は左のリストと対応。実験構造で相互作用が保存されているAlphaFold構造の残基をオレンジでハイライト。

1) <https://alphafold.ebi.ac.uk/>

2) Feldman, H. J.; Labute, P. *J. Chem. Inf. Model.* **2010**, 50 (8), 1466–1475.