

タンパク質立体構造データベースシステム PSILO

PSILO と MOE による SBDD データの管理と利用

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。クライオ電子顕微鏡法や AlphaFold などの普及により近年加速度的に増加する公共およびインハウスの立体構造データを統合し、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。本稿では、PSILO の SBDD データの管理機能とその利用例に関連する MOE のアプリケーションとともに紹介します。

■ SBDD の研究データと管理

SBDD (Structure-Based Drug Design) の研究では、膨大なデータが日々蓄積されています。例えば、タンパク質やリガンドの立体構造データ以外にも配列、電子密度マップ、実験や計算で得られたプロパティなどのデータがあります。

これらのデータを比較・検証するには事前に適切な処理が必要になります。通常、実験で得られた構造はそれぞれ座標が異なるので、基準の系に対して重ね合わせや電子密度の再計算を行う必要があります。また、実験や計算で得られたプロパティは、どの系に対する値かを管理する必要があります。

これらの前処理や更新を手作業で行うのは非常に煩雑であり、また、データの更新毎に作業が必要になります。PSILO は、これらの作業を自動化し、MOE を用いた SBDD の研究を強力に支援します。

■ PSILO による SBDD データの管理

PSILO には、類縁タンパク質のマルチプルアラインメントに基づいてタンパク質ファミリーデータベースを作成・管理する機能があります。これにより、あるファミリーに属するタンパク質の立体構造と関連データを 1 つのデータベースに集約し、SBDD データを効率的に管理・運用できます。MOE にも同様の機能 (MOE Project) がありますが、PSILO にはより多くの機能が実装されています。

データを集約する際には、必要な前処理を事前に定義して実行できます。例えば、構造の重ね合わせ、電子密度の計算、リガンドのトポロジーの修正、プロパティの推算、Protein Ligand Interaction Fingerprint (PLIF) の計算などを集約処理の際に自動で実行できます。PSILO ではさらに同一のファミリーとみなす E-value の閾値を設定したり、データ集約後のポストフィルターとして PSILO の検索クエリーを設定したりすることができ、よりプロジェクトの目的に合致した独自のファミリーデータベースを作成できます。

PSILO 上のファミリーデータベースはデフォルトで定期的に自動更新されるため、データの管理者は簡単にデータを管理できます。また、データの利用者は容量の大きいファミリーデータベースを個々にインストールせずに、常に PSILO サーバー上の最新のファミリーデータベースから必要なデータをクライアントマシン上のブラウザや MOE で利用できます (図 1)。



図 1. PSILO-MOE の利用イメージ

■ 標的別タンパク質ファミリーデータベース

MOE には、標準で 26 種類のタンパク質ファミリーデータベースが同梱されています。創薬研究で頻繁にターゲットとされる、キナーゼ、GPCR、抗体、核内受容体などの実験構造が、基準の系に対して既に重ね合わせされ、重要残基にラベルや配色がされた状態で提供されています。PSILO には、デフォルトでより多くのファミリーデータベース (39 種類) が提供されています (表 1)。

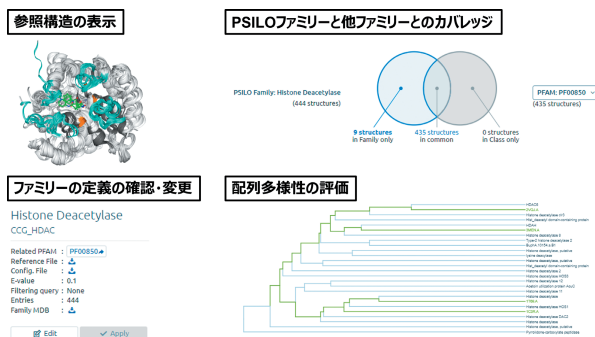
表 1. PSILO に登録されているタンパク質ファミリー。青背景はデフォルトで PSILO でのみ提供。

ABC Transporter	Histone Acyltransferase	Protein Phosphodiesterase
Antibody	Histone Deacetylase	Ras
Armadillo	I-Set	Retroviral Protease
Aspartic Protease	Ionotropic Glutamate Receptor	Retroviral Reverse Transcriptase
Bromodomain	Leucine-rich Repeat	RNA Recognition Motif
Calcium Channel	Lysozyme	Short-Chain Dehydrogenase
Carbonic Anhydrase	Major Histocompatibility Complex	Sodium Channel
Cytochrome P450	Matrix Metalloprotease	T-Cell Receptor
EF Hand	Nuclear Receptor	Trypsin-like Serine Protease
Globin	PDZ Domain	Two Pore Channel
GPCR Class A	Phosphoinositide 3-Kinase	Two Pore Domain Potassium Channel
Green Fluorescent Protein	Potassium Channel	WD40
Histone	Protein Kinase	Zinc Finger

■ PSILO によるタンパク質ファミリーの管理

PSILO のファミリー管理ページからは、ファミリーの定義設定の確認・変更、ファミリーデータベースのダウンロードなどができます。また、現在のファミリーデータベースと PFAM などの他のデータベースで定義されたファミリーとのカバレッジを確認したり、系統樹表示により参照配列の冗長性を確認したりすることで、作成

したファミリーデータベースの妥当性を評価できます (図 2)。



MOE による PSILO ファミリー検索

MOE には標的のタンパク質ファミリーデータベースに対してさまざまな条件で検索を行う機能 (Project Search) があり、例えば、キーワード、リガンド類似性、配列一致度などにより検索できます。Project Search は、PSILO 上のファミリーデータベースに対しても同様に実行でき、これにより目的の立体構造とそれに関連する情報を検索できます。

目的の立体構造について PLIF 計算が行われていれば、MOE の PLIF 解析機能により、タンパク質とリガンドの相互作用をバーコードとして可視化し、研究対象のタンパク質群の中から保存されている相互作用を検出したリ、類似の相互作用パターンをもつ構造を見つけたりすることができます (p. 10 図 3)。

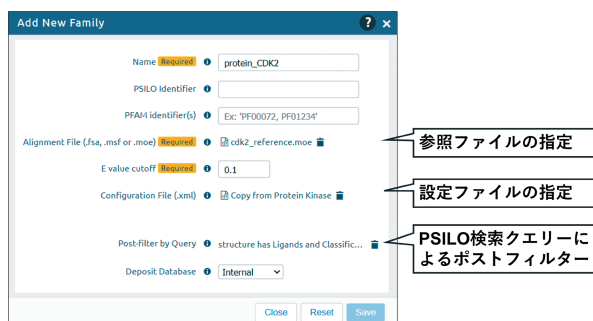
CDK2 ファミリーの作成例と利用例

PSILO には、デフォルトで Protein Kinase のデータベースが同梱されていますが、ここではその中でも CDK2 にフォーカスしたファミリーを作成するために、PSILO の機能を利用しました。また、作成したデータベースに対して MOE の Project Search による検索を行い、立体構造を絞り込みました。

CDK2 ファミリーの作成

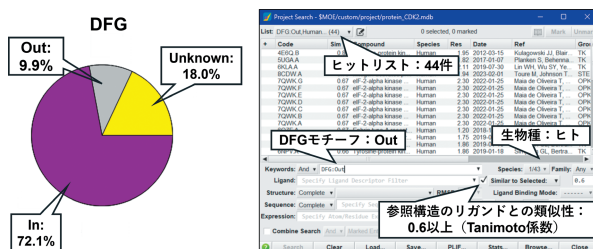
ファミリー作成画面は図 3 のようになっています。実用的なファミリーデータベースを作成するには、XML 形式の設定ファイルと MOE 形式の参照ファイルが必要です。設定ファイルはその場で作成することも、既存ファミリーのものをそのまま使うこともできます。ここでは Protein Kinase の設定ファイルを使用しました。参照ファイルには、CDK2 の結晶構造 (PDB ID: 1G5S) を基に、リガンド、構造の重ね合わせに使用する領域、ヒンジ、DFG モチーフ、G-loop などの重要な領域を定義したものを使用しました。さらに、PSILO 検索クエリーによるファミ

リー作成後のポストフィルタには、「リガンドを含む、かつ解像度 2.5 Å 未満、かつ Gene Ontology で protein phosphorylation に該当する構造」を指定しました。最終的に、Protein Kinase の約 1/3 の 4428 件のエントリーを含むファミリーデータベースを作成できました。



Project Search による CDK2 ファミリーの検索

各ファミリーに特有のプロパティの計算を前述の設定ファイルに記述すれば、そのプロパティをもとに Project Search で立体構造を検索できます。例えば、キナーゼでは DFG モチーフの状態が In か Out かで検索できます。また、その統計情報も確認できます (図 4 左)。PSILO で作成した CDK2 ファミリーに対して、「DFG モチーフが Out、かつヒト由来、かつ参照構造とリガンドが類似」という条件で検索すると、元のデータベースの 1/100 以下の 44 件にまで立体構造を絞り込むことができました (図 4 右)。



開発元の Webinar

PSILO のファミリーデータベース管理機能を紹介した Webinar 「PSILO: Streamlined Protein Family Analysis and Refinement」の録画を以下 URL から視聴いただけます。

<https://video.chemcomp.com/watch/ZzeJu5KQFQpHwAoLpju6ds>