

統合計算化学システム

MOE-QFSSバージョンアップ紹介

MOE-QFSS (Quick Federated Structure Search) は、複数の巨大な化合物データベースに対して高速かつ横断的に構造検索を行うためのMOEのアドオンツールです。リレーショナルデータベースを使用することなく、MOEのみで膨大な化合物データの中から部分構造と類似構造の高速検索が可能です。QFSSの新バージョンでは、公共データベースへのアクセスとMOE/web SOAPサーバーでの検索に対応しました。さらにQFSSのMOE/webアプリケーションを提供します。それによって目的や計算資源に応じたQFSSの実行環境を構築できます。ここではQFSSの概要と新機能、構造検索事例について紹介します。

■ QFSS概要

QFSSは、MOEウィンドウに表示した分子をクエリーとして、複数の巨大な化合物データベースに対して高速かつ横断的に、完全一致・部分構造・類似構造による検索が可能です。特徴的な母核を用いる場合、数千万構造に対して部分構造検索を数秒で行えます。化合物データベースとして、公共データベース、試薬ベンダーデータベース、インハウスデータベースを自由に設定できます。

検索結果はMOEのMDB形式で出力されるので、ヒット構造に対してSAR解析やドッキングシミュレーションなどシームレスな解析が可能です。QFSSは、高速な化合物検索を通して、分子設計やバーチャルスクリーニングの効率化を支援します。

■ QFSSのための部分構造フィンガープリント(SSFP)の定義

QFSSでの検索に用いるために、部分構造フィンガープリント(SSFP)を考案しました。試薬ベンダーとChEMBLから収集した化合物構造(約1104万構造)が持つ単環構造を抽出しフィンガープリント化しました。出現頻度の高い単環構造については、単環に結合する3連鎖の元素パターンの情報(CCCやCCO、CCN等の可能な組み合わせを機械的に列挙)を加えさらに細分化しました(図1)。それらの単環構造、一つの原子に結合する3連鎖の元素パターン、3連鎖にならない特徴的な原子団(C=O、C=C等)、環の大きさとその組成等の情報を含め計117,227種のSSFPを定義しました。

各化合物データベースについて、登録構造が持つ全SSFPを列挙したインデックスファイルを事前に作成し、検索時に参照することで高速な構造検索を実現しています。

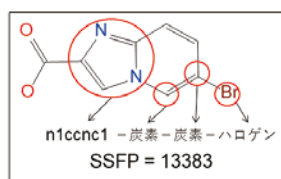


図1. SSFPの例。イミダゾール環に結合する炭素-炭素-ハロゲンの3連鎖情報を13383番のSSFPとして定義し、インデックスファイルとライブラリーのMDBファイルに関連付け検索に利用。

■ QFSSで利用できる化合物データベースの例

次のような化合物データベースをMDBファイルに変換することで検索対象として使用できます。この他にも、

MDBファイルであればユーザーが自由に追加できます。

○ MOEデータベース

・リードライク化合物データベース

○ 公共データベース

・ChEMBL (<https://www.ebi.ac.uk/chembl/>)

・SureChEMBL (<https://www.surechembl.org/search/>)

・ZINC (<http://zinc.docking.org/>)

・PubChem (<https://pubchem.ncbi.nlm.nih.gov/>)

○ 試薬ベンダーデータベース

・ナミキ商事株式会社 (<https://www.namiki-s.co.jp/>)

・キシダ化学株式会社 (<http://www.kishida.co.jp/>)

■ QFSSの機能

QFSSでは次のような検索が可能です。

- ・部分構造の超高速検索
- ・完全一致(Exact)とクエリー部分構造の完全一致(Semi-exact)検索
- ・MACCSキーを用いた類似構造検索
- ・SupersetとSubsetの出力
- ・分子量、logP、反応性、Lipinskiルールによるフィルター
- ・複数のクエリー分子を用いた連続検索
- ・2D母核置換

■ QFSSによる検索

MOEウィンドウにクエリーを読み込み、検索対象データベース、検索条件、フィルター、出力情報を指定し検索を実行します(図2)。出力結果はMOEのMDB形式で即座に得られます。ヒットした構造は、ライブラリーごとのIDが付加されています。結果に対して、MOEアプリケーションを用いた2次元/3次元の構造解析やIDを用いた公共データベース上での検索、試薬ベンダーからの購入などを速やかに行えます。

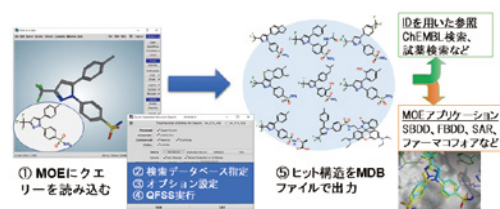


図2. QFSSの検索フロー。複数のライブラリー全体で1417万構造に対する部分構造検索(所要時間: 1.2秒)。使用PCスペック(CPU: Core i7-5500U 2.4GHz, Memory: 8GB, Disk: SSD)。

■バージョンアップ内容

新バージョンでは次のような機能が追加されます。

■ 公共データベースへのアクセス

公共データベースには、そのデータベースごとにさまざまな情報が登録されています。ChEMBLであれば標的タンパク質や活性値の情報、SureChEMBLであれば化合物の特許情報が収録されています。これらは、化合物のSAR解析や新薬設計を行う上で有益な情報となり得ます。

新バージョンでは、ChEMBL, SureChEMBL, ZINCの登録構造がヒットした場合、そのIDをダブルクリックすることで対応するウェブサイトへアクセスでき、上記のような情報を即座に得ることが出来ます。

■ MOE/web SOAPサーバー上での検索

検索対象データベースは、数GBから数十GBの容量を持つものが多くあります。MOE/web SOAPサーバーを利用することで、計算の実行とデータベースの保存を一つのサーバーに集約でき、ローカルマシンへの計算ならびに容量における負荷を削減できます(図3)。

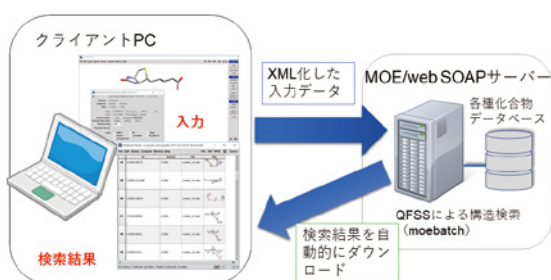


図3. MOE/web SOAPサーバーを用いた検索フロー

■ MOE/webアプリケーションの追加

QFSSがMOE/webアプリケーションとして追加されます。MOE/webは、ウェブブラウザを用いて、MOE/webサーバーにアクセスし、計算を実行する簡易的な計算モードです。利用者は、クライアントPCにMOEをインストールする必要がなく、ウェブブラウザを使用したよりシンプルな操作でQFSSによる構造検索が可能です(図4)。検索結果はMDBファイルやSDFファイルで保存できます。



図4. MOE/webによるQFSSの構造検索

■ QFSSを用いた応用事例

QFSSを用いた新規リガンド設計事例について紹介します。

■ 上皮成長因子受容体(EGFR)の阻害剤設計

EGFR阻害剤の一つであるErlotinibの母核構造(図5左の緑の楕円部分)をクエリーとして、ChEMBL23に登録されている172万構造についてQFSSを用いて部分構造検索を行いました。検索結果にLipinskiルールによるフィルターを適用し3435構造を得ました。ヒット構造をSARReportを用いて母核と置換基への自動分割と置換位置の割り当てを行い、母核と置換基をランダムに組み合わせたバーチャル化合物を1000個得ました。バーチャル化合物に対して、QSAR解析により pK_i 値を予測し、高活性と予測される構造をEGFRの複合体(PDB ID: 1M17)に対してErlotinibの母核を用いてテンプレートドッキングを行いました(図5右)。結果として、ChEMBLに登録された既知の活性化合物と類似の構造を構築できました。

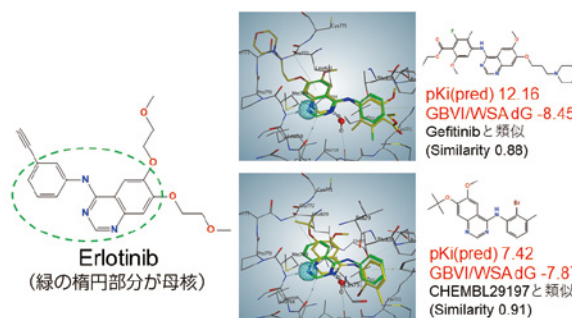


図5. QFSSとその他のMOEアプリケーションを用いた解析。左はクエリーおよびドッキング時のテンプレートとして用いたErlotinib。右は設計した既知の阻害剤の類似構造とその活性予測値およびドッキングスコア。

■ シクロオキシゲナーゼ-2(COX-2)阻害剤の2D母核置換

COX-2阻害剤の一つであるCHEMBL29920を用いて、QFSSにより2D母核置換を行いました。この検索では、指定した母核を入れ替え、その他の置換基を部分構造に有した構造を得ることができます。母核(図6赤丸)を置き換えた既知のCOX-2の阻害剤をヒットさせることが出来ました。

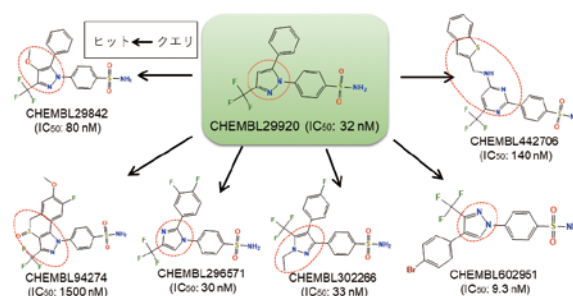


図6. COX-2阻害剤における2D母核置換

■ QFSSの入手方法

QFSSの新バージョンは2018年春にリリース予定です。QFSSをご希望の方はccg@molsis.co.jpまでご連絡ください。