

統合計算化学システム

MOE Extensions for KNIME Ver. 2.3リリース

KNIMEはドイツのKonstanz大学で開発されたワークフロー作成ツールです。さまざまなデータの処理を行うノードをつなげるだけで、複数の計算工程を自動処理する手続き(ワークフロー)を作成できます。MOE Extensions for KNIMEとは、MOEを用いた解析ノードの総称で、CCG社ではMOEの機能を組み込んだ約200のノードを開発し提供しています。KNIMEに含まれるデータマイニング等を行う既存の解析ノードとMOEの解析ノードを組み合わせることで、研究開発のための計算処理を誰でも簡単に実行することができるようになります。ここでは、MOE Extensions for KNIMEの特徴とVer. 2.3で追加された機能と活用例を紹介します。

■MOE Extensions for KNIME

MOE Extensions for KNIMEはさまざまなMOEの解析ノードの総称です。KNIME上で、MOEノードを組み合わせることでワークフローを構築することができ、SVLでスクリプトを作成するよりも簡単にわかりやすい計算プロトコルが作成できます。

ワークフローを構築することで計算の流れを視覚化でき、ノードの削除・追加だけで、計算手順を容易に変更できます。Ver. 2.3では、以下のようなMOEノードが搭載されています。

MOEノードの種類(抜粋)

- ・データの出入力に関するノード群 (29種類)
 - ・分子シミュレーションに関するノード群 (23種類)
 - ・SBDDおよびFBDD解析に関するノード群 (11種類)
 - ・QSARに関するノード群 (27種類)
 - ・タンパク質解析に関するノード群 (34種類)
 - ・分子フィンガープリントに関するノード群 (14種類)
- など、合計 約200種類

MOEノードの中には、標準のMOEには未搭載の計算ノードもあります。例えば、ECFP Descriptorsや Matched Molecular Pairsなどのノードがあります。

各ノードを処理の順につなぐことでデータの処理の流れを定義できます。図1は、化合物三次元化のワークフローで、化合物のデータベースファイルを読み込み、水素原子付加、構造最適化計算を行い、データベースファイルとして保存する例です。



図1 データの流れとデータ処理

MOEノード以外にも、KNIMEに搭載されている多くのノードを使い、多機能なワークフローを作成することが可能です。KNIMEの基本ノードに加え、RやWeka等の統計解析用ノード、JavaやPerl、Pythonなどのスクリプティングノード、BioSolveIT、ChemAxon等の計算ソフトウェアに対応したノードがあり、ソフトウェア間の連携を容易に

します。また、条件分岐やループによる繰り返し処理の機能をもつノードも利用することで、より複雑な処理が可能になります。

■MOEノードのカスタマイズ

Ver. 2.3から、MOEノードのカスタマイズが容易になりました。これまではSVLによるプログラミング後、コンパイルを行い、KNIMEへのインストールを行うといったいくつかの手順が必要でした。今バージョンから、SVLプログラムを指定のフォルダに配置するだけで、新しいノードが利用できます。コンパイルの必要がないため、デバッグ作業が容易になりました。ノードのエラーが出た際、KNIMEを閉じることなくSVLコードを変更して、動作を再確認することが可能です。図1の「Write MDB」ノードはワークフローを流れてきた分子データをMDBファイルとして保存するノードです。図2は、このノードをカスタマイズし部分電荷を計算する機能を追加した例です。ノードの設定パネルは、他のノードと同様のフォーマットで設計できます。

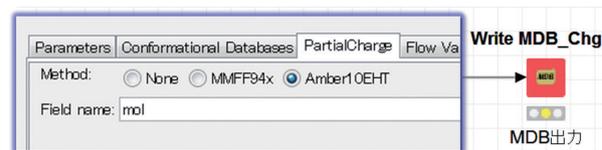


図2 MOEノードのカスタマイズ(部分電荷計算処理の追加)

■MOE/web SOAP機能の活用

MOE/webにはSOAP(Simple Object Access Protocol)機能があり、遠隔サーバー上のMOEで計算を実行させることができます。サーバーのみにインストールされているソフトウェアやデータベースを用いた計算、コストの高い計算などを行うのにMOE/web SOAP機能は有用です。Ver. 2.3では、PDB類似配列検索(図3)、ファーマコフォア検索、ホモロジーモデリング、Mogul I/F等のノードが用意されています。



図3 PDB類似配列検索のSOAPノードを用いた解析

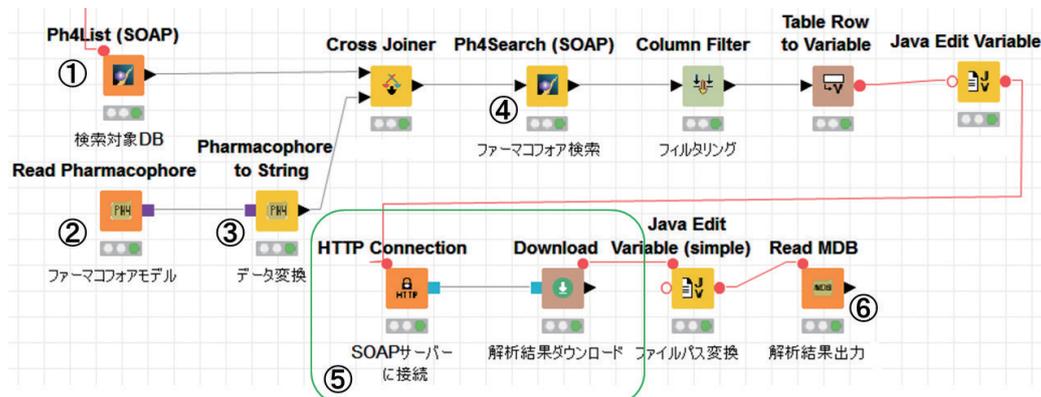


図4 SOAPノードを使ったファーマコフォア検索

SOAPによるファーマコフォア検索

SOAPノードを用いた応用例としてファーマコフォア検索を行うワークフローを示します。図4はサーバーに保存された配座解析済みリード化合物データベース(65万件、5 GB)に対して、入力ファーマコフォアモデルに適合する配座の検索をサーバー上で実行するためのワークフローです。検索結果として得られるデータベースをサーバーからダウンロードし、クライアント上でMOEデータベース(MDB)として保存します。①のノードではサーバーマシンにある検索対象データベースのファイルパス情報を取得します。②のノードではファーマコフォアモデルファイルを入力します。ワークフローを流れるデータは決まったフォーマットの分子構造データ、文字データ、数値データ、フロー変数(赤線)のため、ファーマコフォアファイルそのものを流すことはできません。そこで、③のノードでファーマコフォアファイルのデータを文字形式に変換します。④のノードでは、SOAPを通じてクライアントからサーバーに情報を転送し、サーバー上でファーマコフォア検索を行います。このノードから出力される情報は、データベース内の配座のファーマコフォアに適合した件数と検索結果のデータベースファイルが保存されているネットワーク上のアドレスです。⑤のノード群では、このアドレスをもとに計算結果のファイルをサーバーからダウンロードします。⑥のノードで検索結果の分子構造データをワークフローに読み込みます。

このワークフローで、EGFRのリガンドに適合するファーマコフォアモデルを②のノードに入力し計算を行った結果、324個のリード化合物構造(574配座)が得られました。

ここで得られた分子配座データは、ファーマコフォアに対してアラインメントされているため、受容体構造とファーマコフォア位置が合っていれば、受容体内でリガンド構造を最適化するだけでドッキング構造が得られます。ドッキング構造を得る場合は、図5のようにワークフローを改変します。⑦のノードにEGFR (PDB ID: 4ZAU) を

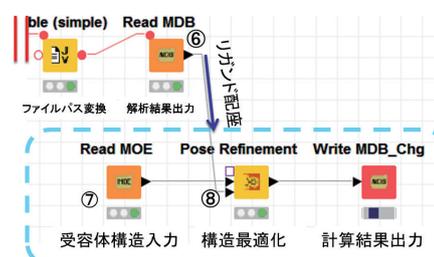


図5 ドッキング計算処理ノードの追加

受容体構造として入力し、⑧のノードで得られた配座を受容体に配置し、構造最適化を行います。その結果ドッキングスコアで324化合物中ランク1位の構造が、非小細胞肺癌の治療薬 (EGFR-TKI) であるErlotinibと類似するリード化合物がヒットしました (図6)。

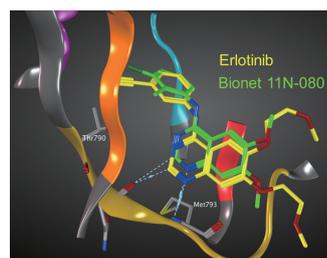


図6 ドッキング計算結果

まとめ

MOE Extensions for KNIMEには、次の利点があります。

1. 計算過程をワークフローとして保管できます。
2. 作成したワークフローは再利用が可能で、ワークフローの共有も可能です。
3. SVLプログラミングのみでノードが作成できます。
4. MOEとその他のソフトウェアとの連携ができます。

MOEの保守契約締結中の方を対象に、無償でMOE Extensions for KNIMEをご送付いたします。ご希望の方は、MOEサポートリクエスト窓口 cgg@rsi.co.jp までお問い合わせください。