

# KNIME を使った化合物の分類例

概要： ノードを使った例として、KNIME ノードを組み合わせたワークフローによって、100+の低分子化合物を化学的性質によって分類します。新規ノードの開発は不要で、全てドラッグアンドドロップによる操作のため、手間をかけずに解析ができます。

## 1. 化合物データセットの入手

解析対象化合物のデータセットとしては、どのようなデータでも結構です。ここでは例として BACE1 (beta secretase) と complex を形成している化合物を使用します。 PDB サイトより入手可。

## 2. KNIME の起動と拡張ノードのとりこみ

KNIME を [www.knime.org](http://www.knime.org) よりダウンロード後 KNIME を初めて起動する場合、図 1 の画面が表示されます。ここで”Get additional nodes”をクリックし、次画面にて”available software”のタブを選択、最初の行にチェックします (図 2)。Install ボタンをクリックし、その後のインストラクションに従えば、拡張ノードが自動的にインストールされます。

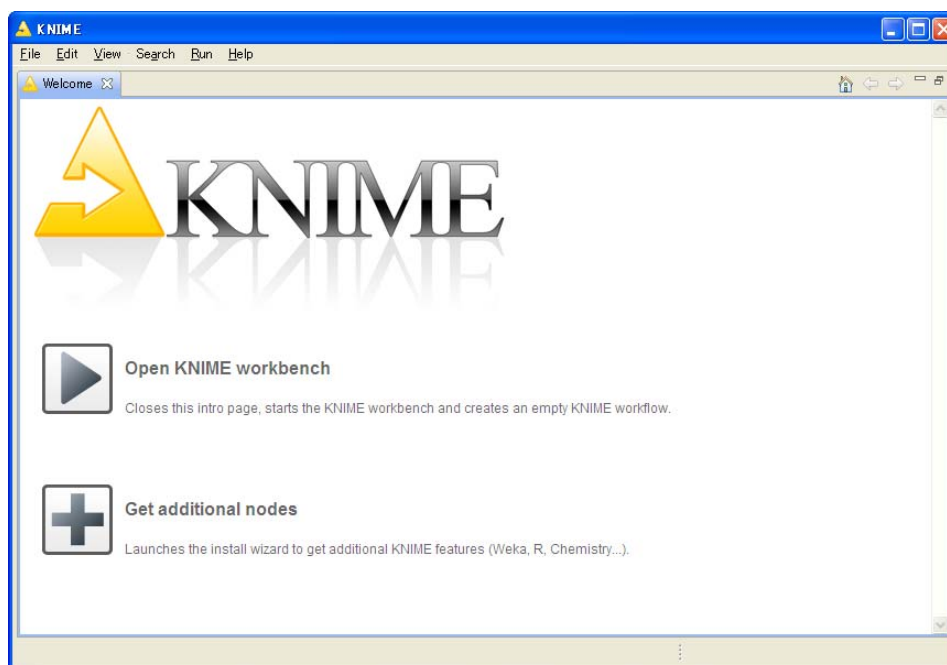


図 1 KNIME 起動画面

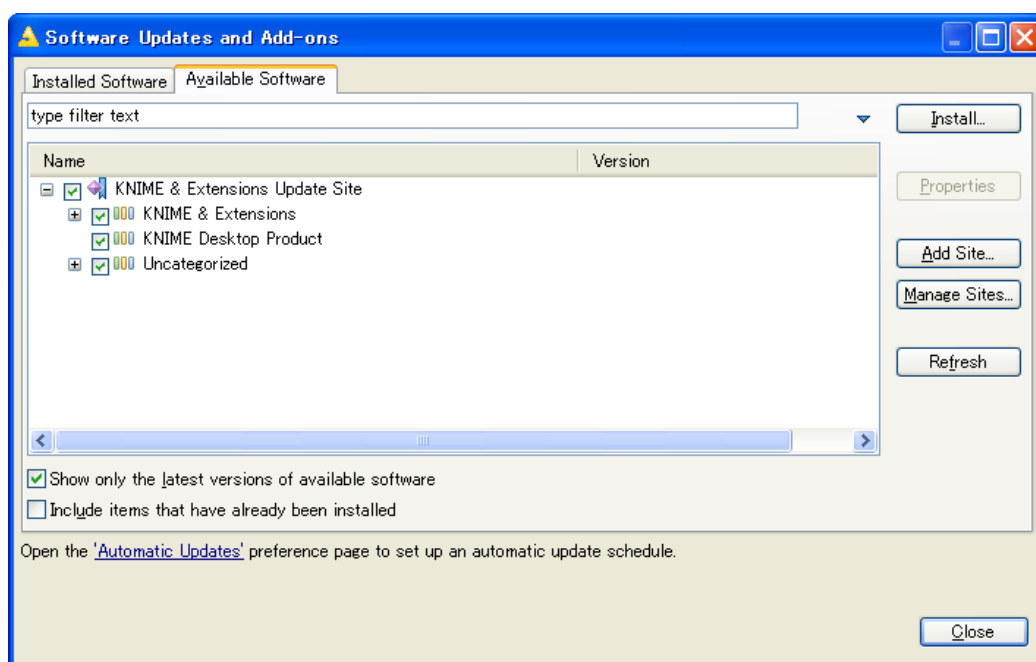


図 2 add-ons 画面

### 3. ワークフロー作成

node repository にあるノードをドラッグアンドドロップして図 3 のようなワークフローを作成します。Mol2 reader、Molecule to CDK、Molecular Properties、Fingerprints、XlogP、3D viewer はすべて chemistry の配下にあり、k-Means、Hierarchical Clustering は Mining >> clustering の配下にあります。これらのノードはマウスで接続できます。

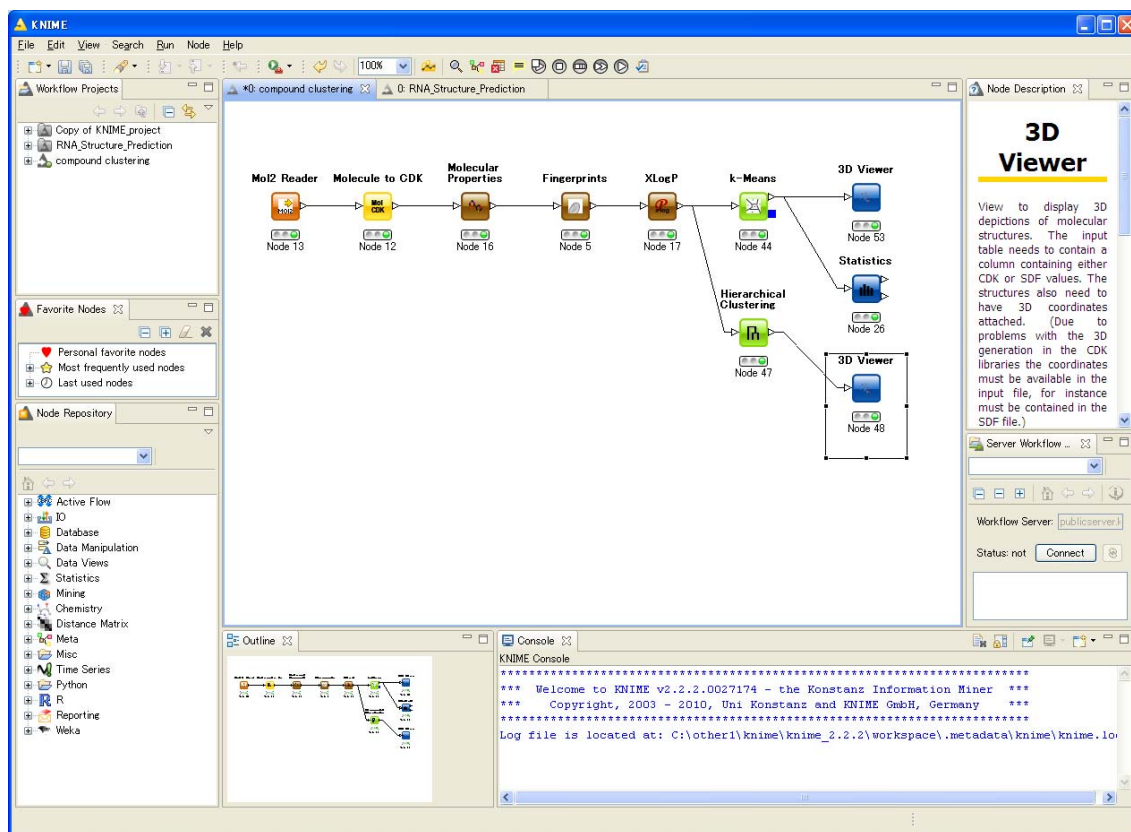


図 3 ワークフロー作成画面

#### 4. ノードの設定

化合物データセットを SDF フォーマットでダウンロードした場合、MOL2 フォーマットに変換して下さい (SDF フォーマットで読もうとするとエラーが発生)。Mol2 Reader では config にて MOL2 ファイルを指定します。Molecular Properties では config (ノード上で右クリックしてメニュー) にて、必要な property を指定します (bond 数、fragment 複雑度等多数あり、default では全ての property が指定となる)。K-means 及び Hierarchical Clustering では config にて clustering に必要な property 及び cluster 数を指定します (但し、Carbon connectivity index が入っているとエラーになるため除外して下さい)。その他のノードに関しては config 未設定であっても実行は可能です。

#### 5. ワークフローの実行

ノード設定完了後、全ノードの実行が可能となります。実行は、各ノード上で右クリック後”execute”を選択して行います。

## 6. 実行結果

k-means の先の 3D viewer では”execute and open view”を選択すると結果画面が表示されます (図4)。ここでは、クラスター番号が表示され、同じクラスター内の 3D イメージ、fingerprint、XlogP といった化合物情報を比較できます。また、Hierarchical Clustering にて”view”を選択すると、dendrogram が表示されます (図5)。

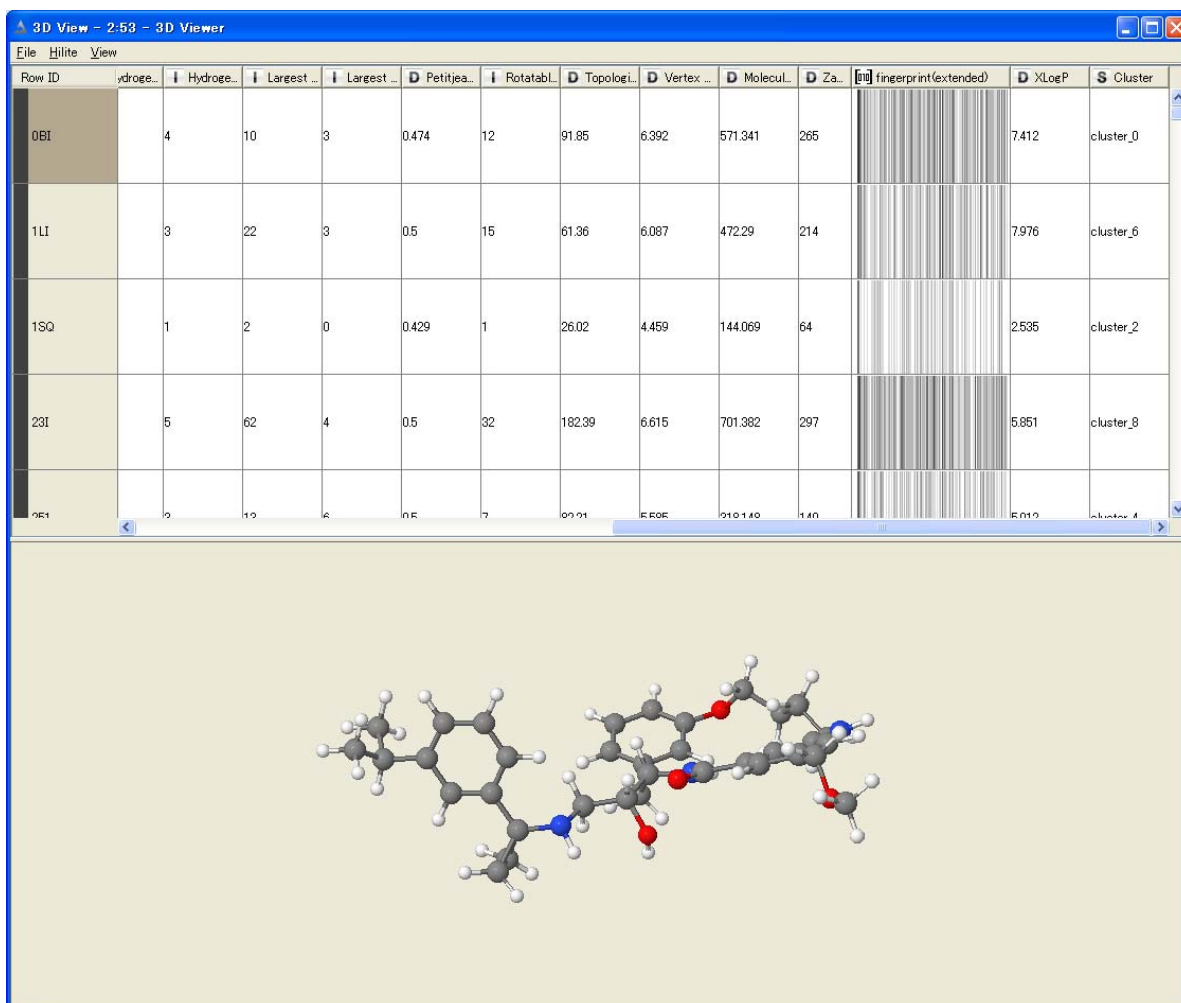


図4 k-means 結果表示画面

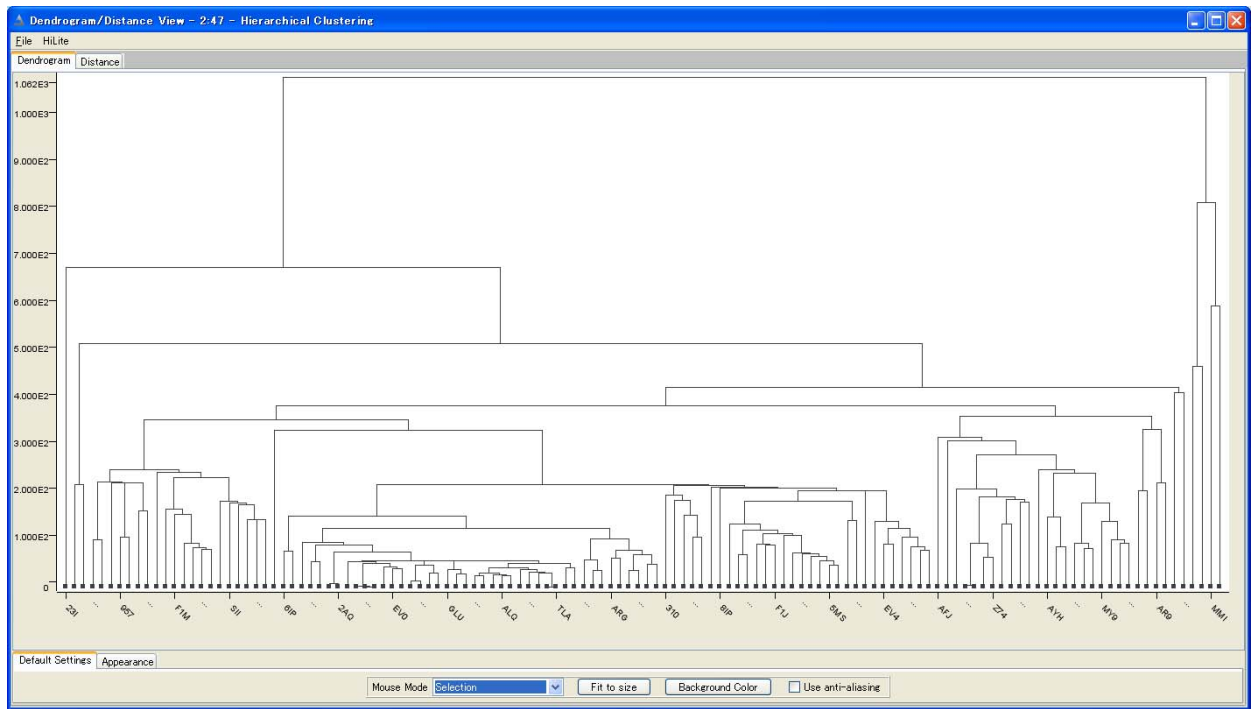


図 5 Hierarchical Clustering 結果表示画面