

# 創薬インフォマティクス ソフトウェア

## Dotmatics : 可視化解析ツールVortexによる 化合物の構造解析

Dotmatics Vortexは膨大な数のデータの解釈を容易にして研究者間で簡単に知見の共有ができるデータ可視化解析ツールです。前号に引き続き、Vortexによる化合物の多様性の可視化についてDotmatics社のポスター資料<sup>1)</sup>をもとに事例紹介します。

### はじめに

化合物を取り扱う研究における次のような課題の解決策となる解析事例について紹介します。

- ・データセットの間にはどんな差異があるのか?
- ・選択したサブセットは化合物の多様性を最大限にとっているか?
- ・良好なスクリーニング結果をもつ化合物の共通性は何か?

### 化合物の多様性の可視化

図1は、K-means 法により2つの化合物群が占める化合物空間を解析・図示した例です。5万件の化合物に対して10秒未満で計算が完了しました。各クラスターの重心となる化合物(DISTANCE=0)に類似性が高い化合物ほどDISTANCEの値が小さくなります。

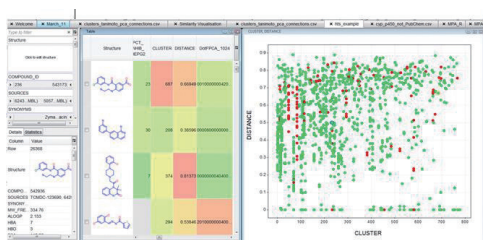


図1 2組の化合物群のK-meansクラスタリング、化合物群をそれぞれ赤と緑で表現

化合物空間を把握する手法として主成分分析があります。図2のように、それぞれのライブラリーが占める化学的空間を2Dプロットで比較できます。XlogPのような物性を色で表現することで物理化学的な空間を解析できます。

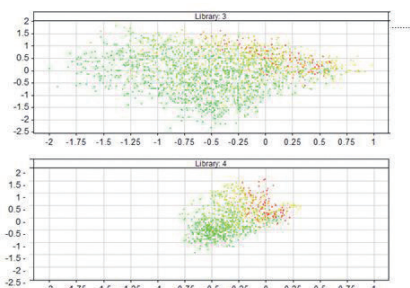


図2 各ライブラリーの主成分分析

データ相互の関係をみるのにCard Plotは便利です。類似の値や化学プロパティをもつカードを近くに配置したり線でつないだりして化合物の関係を表現します。

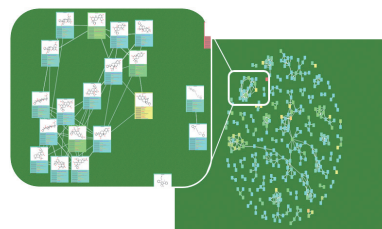


図3 Card Plot

化合物の多様性を見る他の方法として、図4に示した reduced scaffold tree dendrogramがあります。このビューでは、化学フィンガープリントではなく母核のツリーにすることで、母核や環構造、リンカー、修飾子に焦点を当てています。この例の上部はライブラリーごとに色分けされ、化合物の位置はreduced scaffold clusteringによって決まります。このビューから母核が一致する化合物を容易に同定できます。

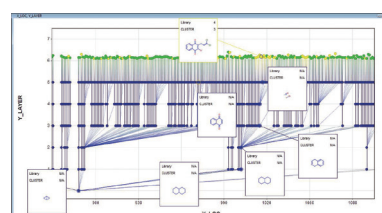


図4 Reduced scaffold treeによりデータセット内の共通の母核を強調表示

また、HTS(high throughput screening) データを有効活用することにより、化合物の多様性を理解することができます。図5はHTSデータセット内を共通の部分構造で検索し、部分構造ごとにヒット数を集計した結果です。これにより、スクリーニング結果が良好な部分構造やファーマコフォアを素早く見極めることができます。

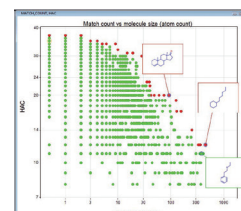


図5 データセット内の共通部分構造数を横軸、HAC (Heavy Atom Count)を縦軸にプロット

### 結論

Vortexにより巨大なデータセット内もしくは相互の関係を高速かつ視覚的に解釈することができ、十分な検証を重ねた上での意思決定が容易になります。

1) [http://www.dotmatics.com/wp-content/uploads/2015/04/ACINF34\\_VisualStructureComparePoster.pdf](http://www.dotmatics.com/wp-content/uploads/2015/04/ACINF34_VisualStructureComparePoster.pdf)