

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社の製品にはナレッジマネジメント、データストレージ、検索やレポート、データ解析と可視化といった多彩なツールがあります。これらツールにより研究者同士のコミュニケーションとコラボレーションを促進することで、研究をより効率的に進められる環境を実現できます。

その中のひとつ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の可視化解析ツールです。本号、次号の2回にわたり、Vortexによる解析事例についてDotmatics社のポスター¹⁾をもとに紹介します。

はじめに

化合物の多様性や相同性の解析はデータセットが巨大になるほど難しくなります。化学情報データの効率的な解析に不可欠な可視化機能には、高速かつ堅牢なクラスター解析と構造検索の技術が求められます。

super-structure検索および部分構造検索

データセットの中から特定の条件に基づいてデータを抽出する処理において、データセットが大規模であっても効率的に処理できることが重要です。図1は450万化合物から成るZINCリードライク・データセットに対するsuper-structure検索結果です。検索でマップされた化合物が強調表示されています。Vortexは1分間に1億4300万件の比較処理を行えるため、たいていの検索では瞬時に結果を得ることができます。

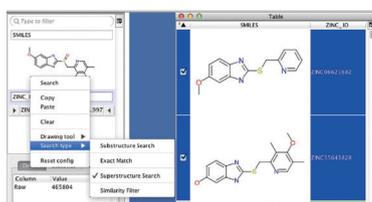


図1 super-structure検索結果。クエリーの構成要素となる構造式を同定できます。

共通の部分構造によって目的の物性をもつ化合物群を抽出したり、逆に望まない化合物群を排除する処理も頻繁に行われます。例として487のPAINS化合物²⁾のいずれかを部分構造にもつ化合物を、high throughput screening (HTS) 結果のデータセットから除去する処理を行いました。表1はその処理におけるVortexの検索速度を表しています。検索が完了すると該当する化合物はデータセットから除外されます。

表1 VortexにおけるPAINS部分構造検索の処理時間

Database	# molecules	PAINS identification
ChEMBL_20	1,456,020	5m 10s
ZINC	22,723,296	55m 10s
PubChem	67,441,980	3h 50m

類似構造フィルター

Vortexでは類似構造を素早く同定できます。図2は Tanimoto係数 < 0.5の化合物を除いた結果です。

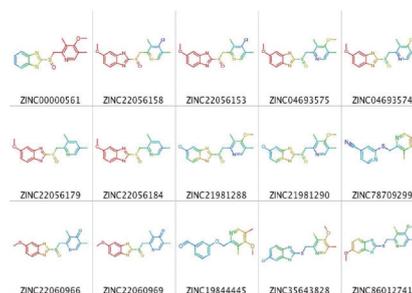


図2 化学類似性フィルター

分子の表現

分子の物性を色や形で表現することにより、分子の比較が容易になり、着目すべき構造や物性が端的にわかるようになります。図3は、分子どうしの共通の構造的特徴を識別したり、物理化学的特性への寄与度に応じて領域を色分けした例です。

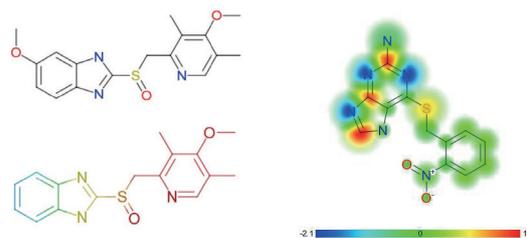


図3 左：類似構造解析（赤色は5以上、黄色は3以上の原子が一致する結合）。右：XlogPへの原子の寄与。

まとめ

Vortexにより、巨大なデータセットに対しても構造の比較処理を迅速に行え、訴求力の高い分子表現ができます。次号では化合物の多様性解析事例を紹介します。

- 1) http://www.dotmatics.com/wp-content/uploads/2015/04/ACINF34_VisualStructureComparePoster.pdf
- 2) Baell, J.B et. al., *J. Med. Chem.*, 2010(53), 2719-2740