

MassBankのRDF化による知識グラフ構築とスペクトルデータ解釈支援

岡昂輝¹、小川誠寛¹、亀谷優歩²、松沢佑紀²、津川裕司^{1,2}

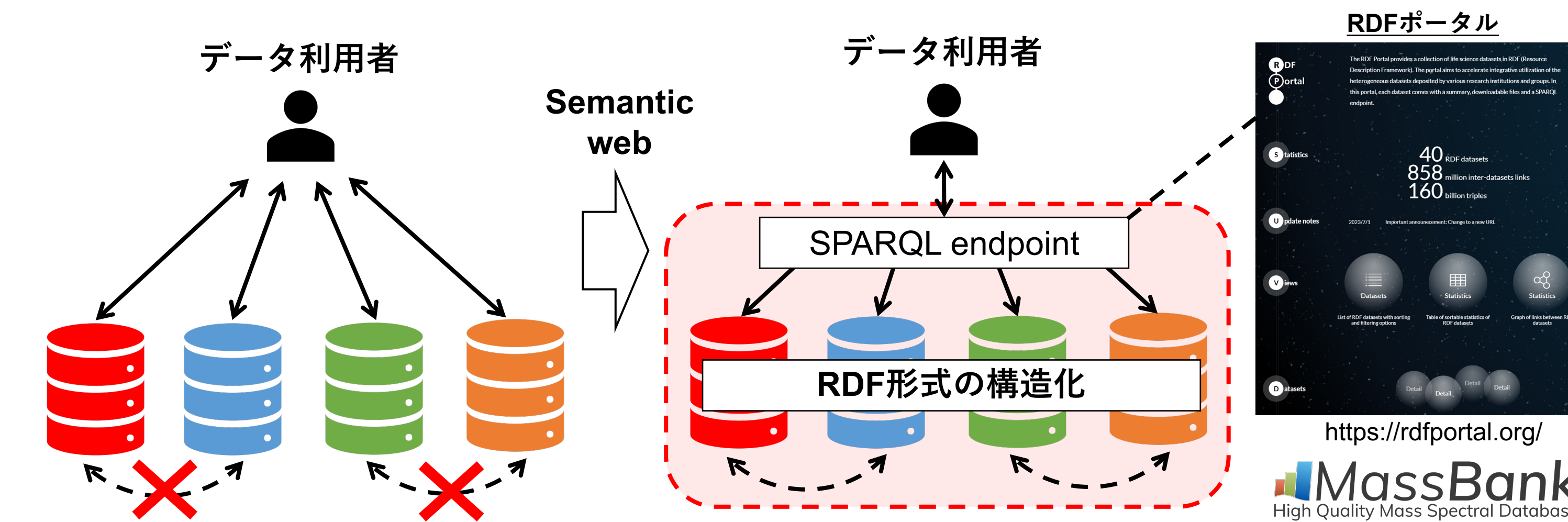
1. 東京農工大学、2. 京都大学



1. RDFを用いたデータエコシステム

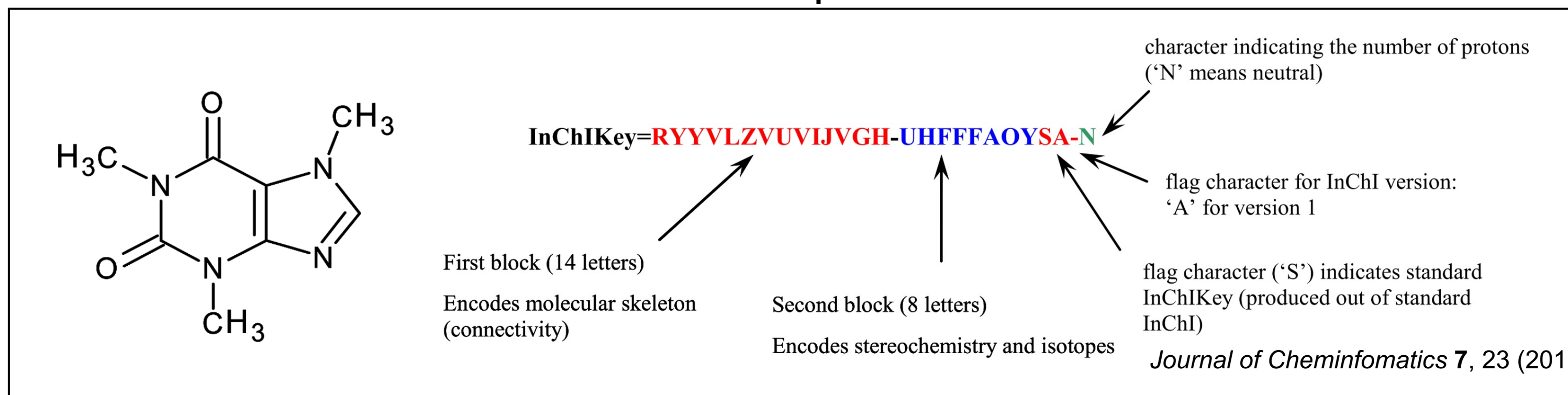
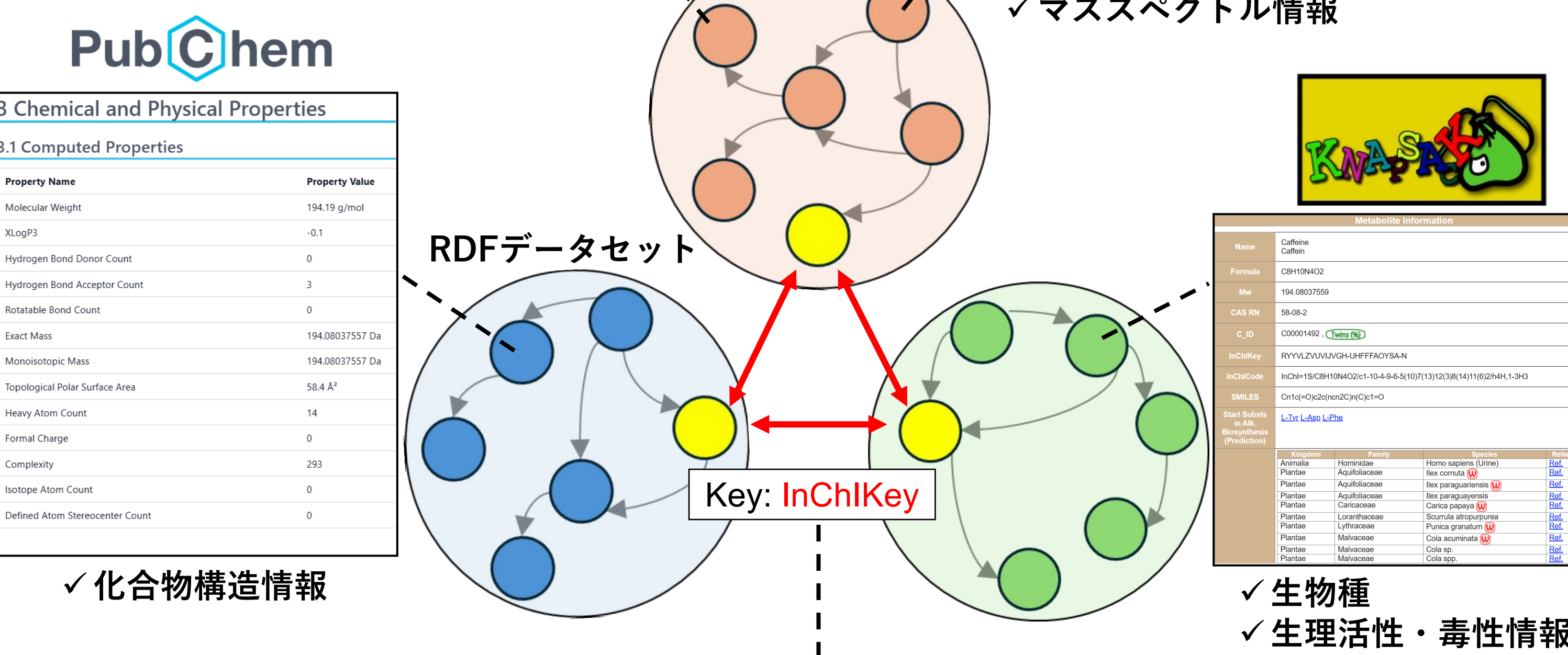
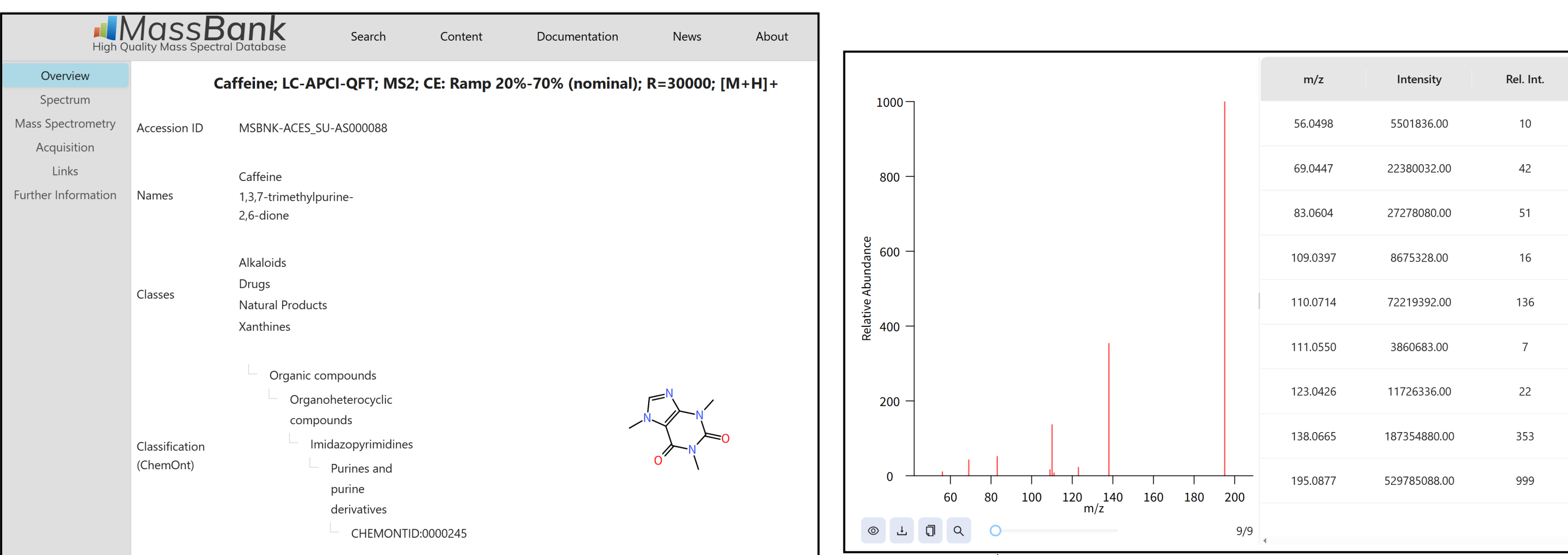
Semantic Web技術に基づく「データエコシステム」では、単一データベースの枠を超えて、データの再利用や統合解析の可能性が広がる。

MassBank-RDF : <https://github.com/systemsomicslab/MassBank-RDF>



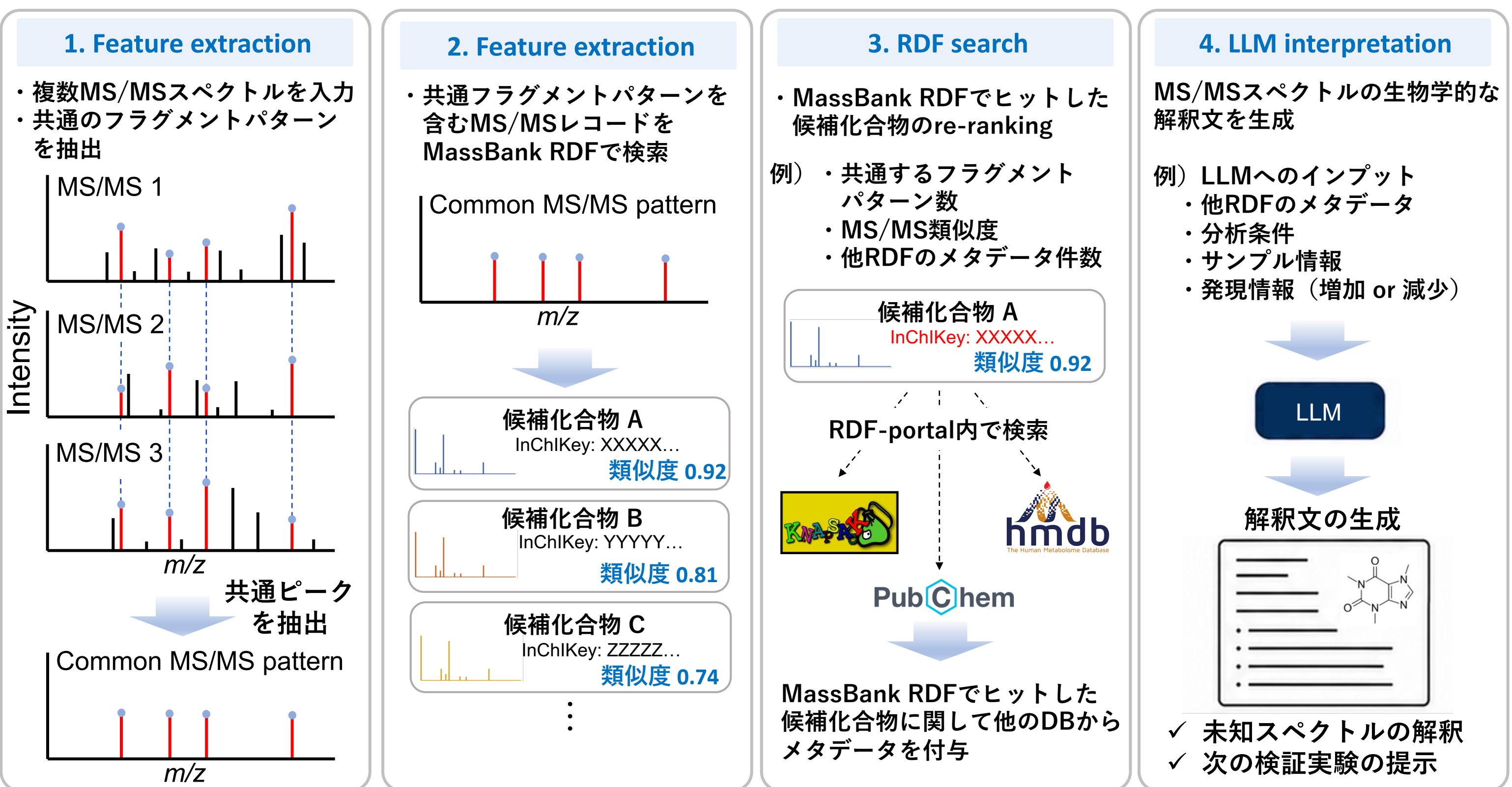
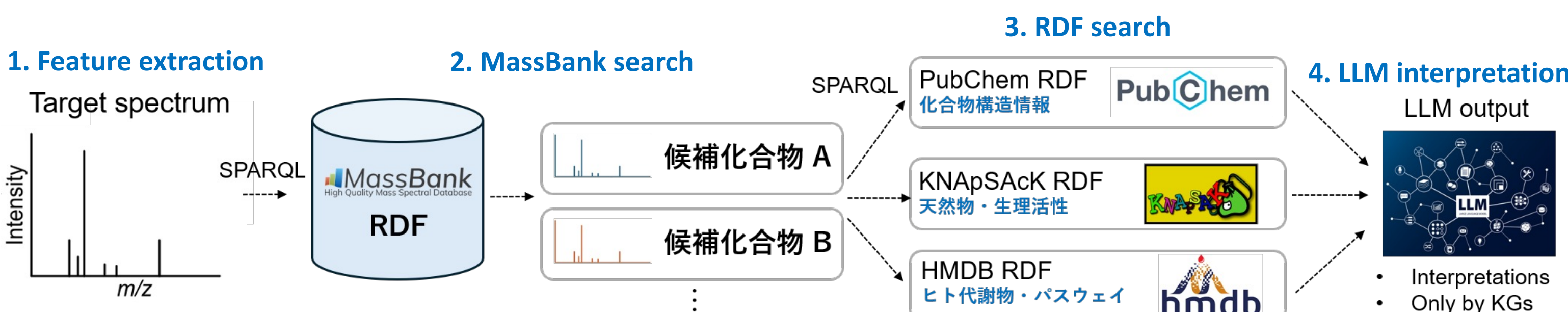
2. 質量分析データを活用した生命科学研究基盤

• MassBank: 低分子化合物のマススペクトルを蓄積したデータベース



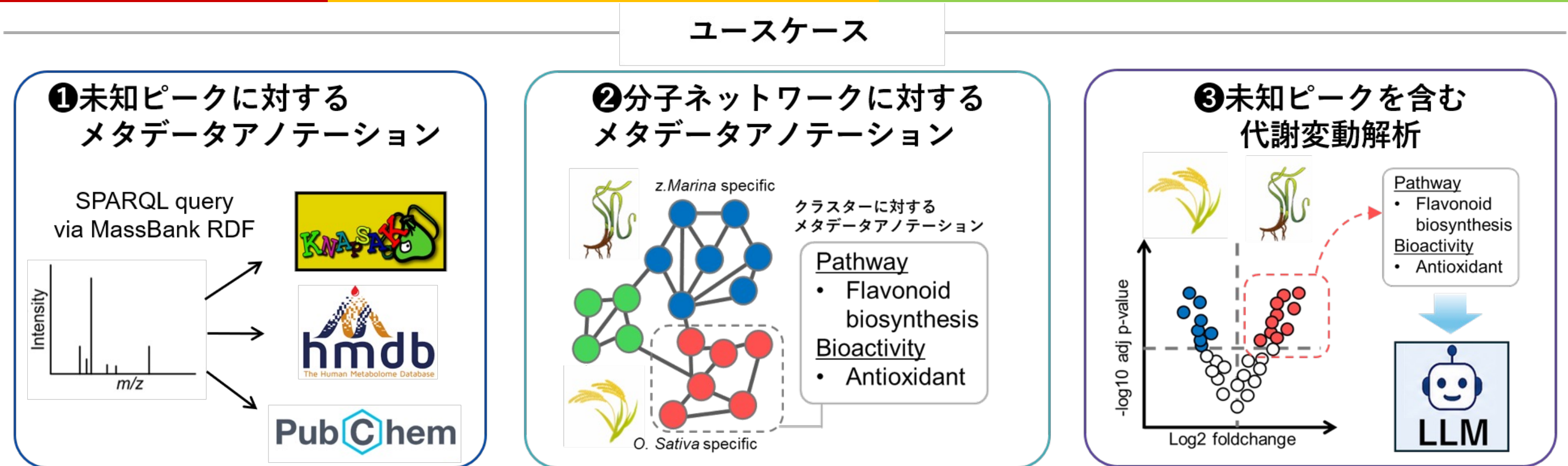
3. MassBank RDFを用いたスペクトル解釈基盤

MassBank RDFによって実現する環境：MS/MS特徴量へメタデータを付与する情報基盤



✓ MS/MSスペクトル情報と複数の生命科学データベースをリンクさせ、未知スペクトルに対するメタデータ付与・解釈基盤を構築する。

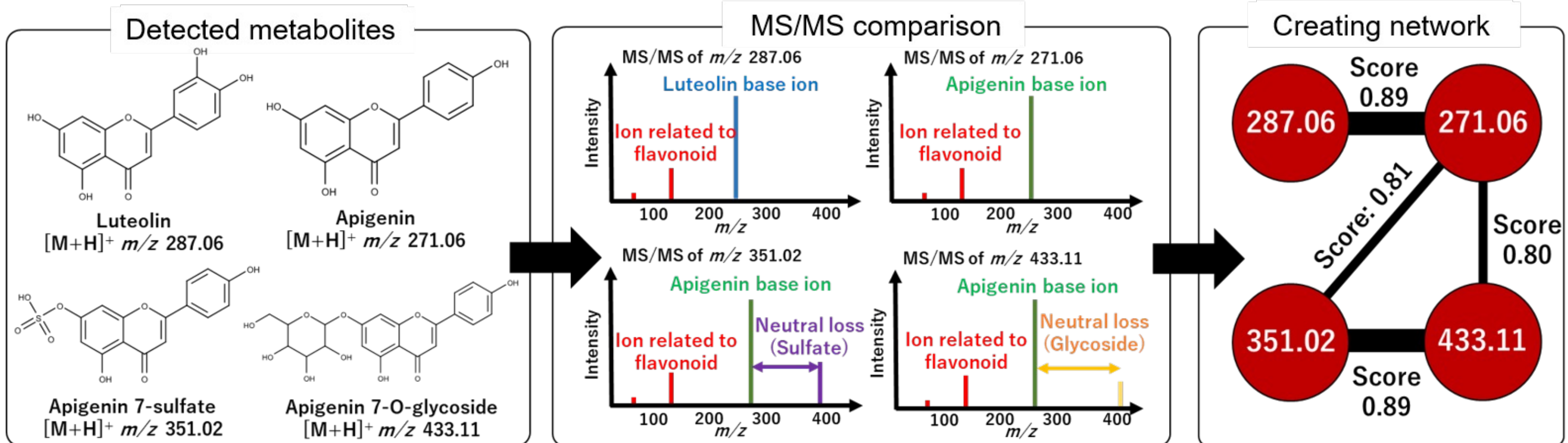
4. ユースケース



②分子ネットワークに対するメタデータアノテーション

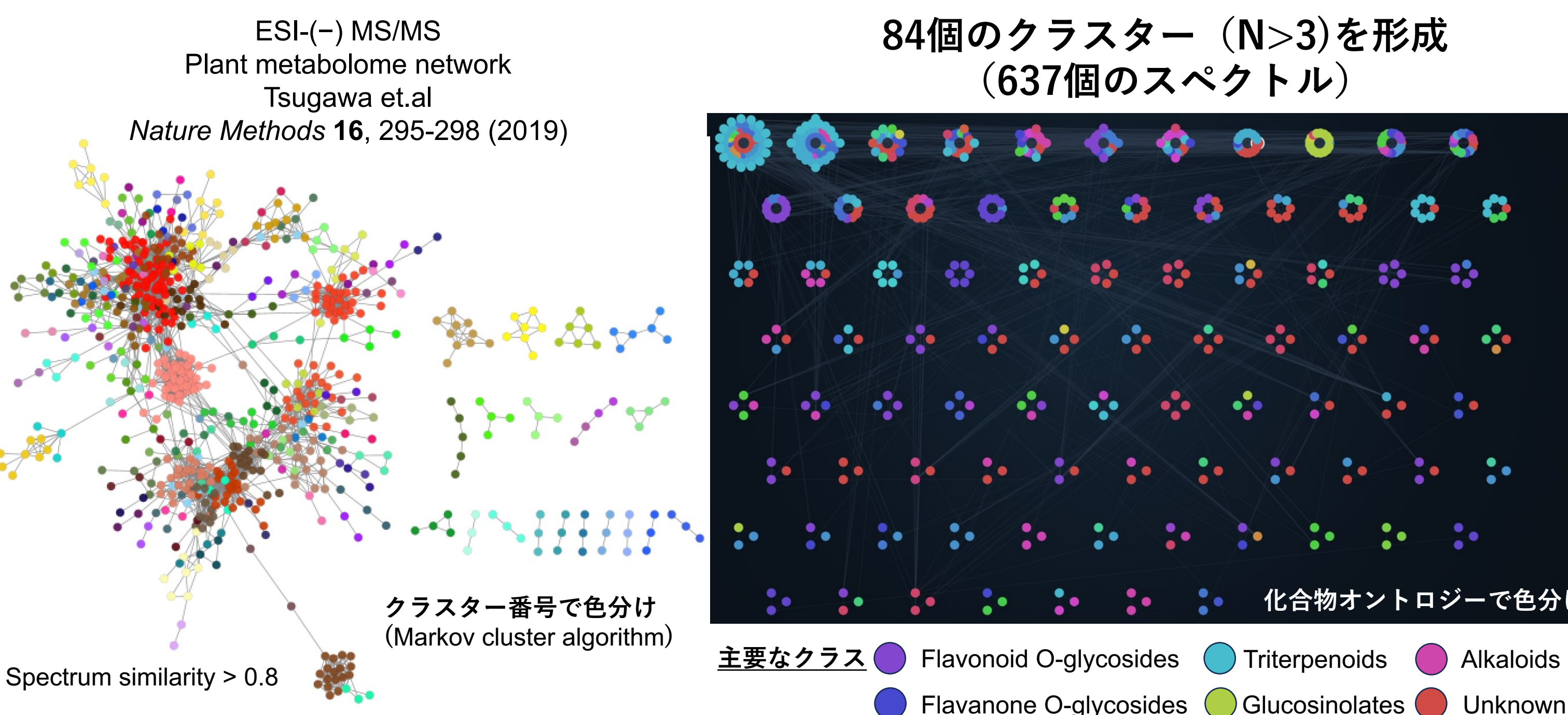
分子ネットワーク

MS/MSスペクトルの類似度（構造の類似性）をもとにネットワークを生成する
→Unknownスペクトルに紐づく既知代謝物情報から構造を推定する



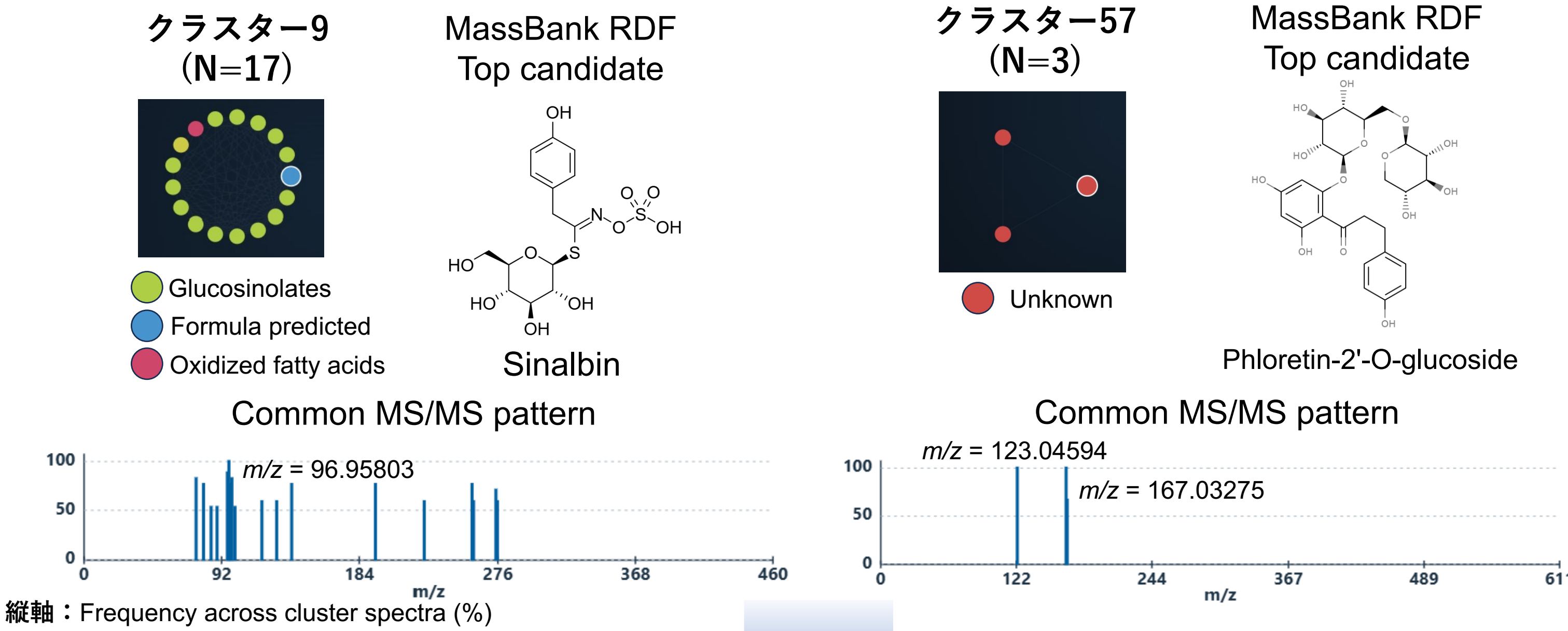
✓ 分子ネットワーク手法で生成される膨大な数の各クラスターにメタデータを付与
→分子ネットワーク手法による解釈を自動化する。

5. MassBank RDFによる分子ネットワークの自動解釈

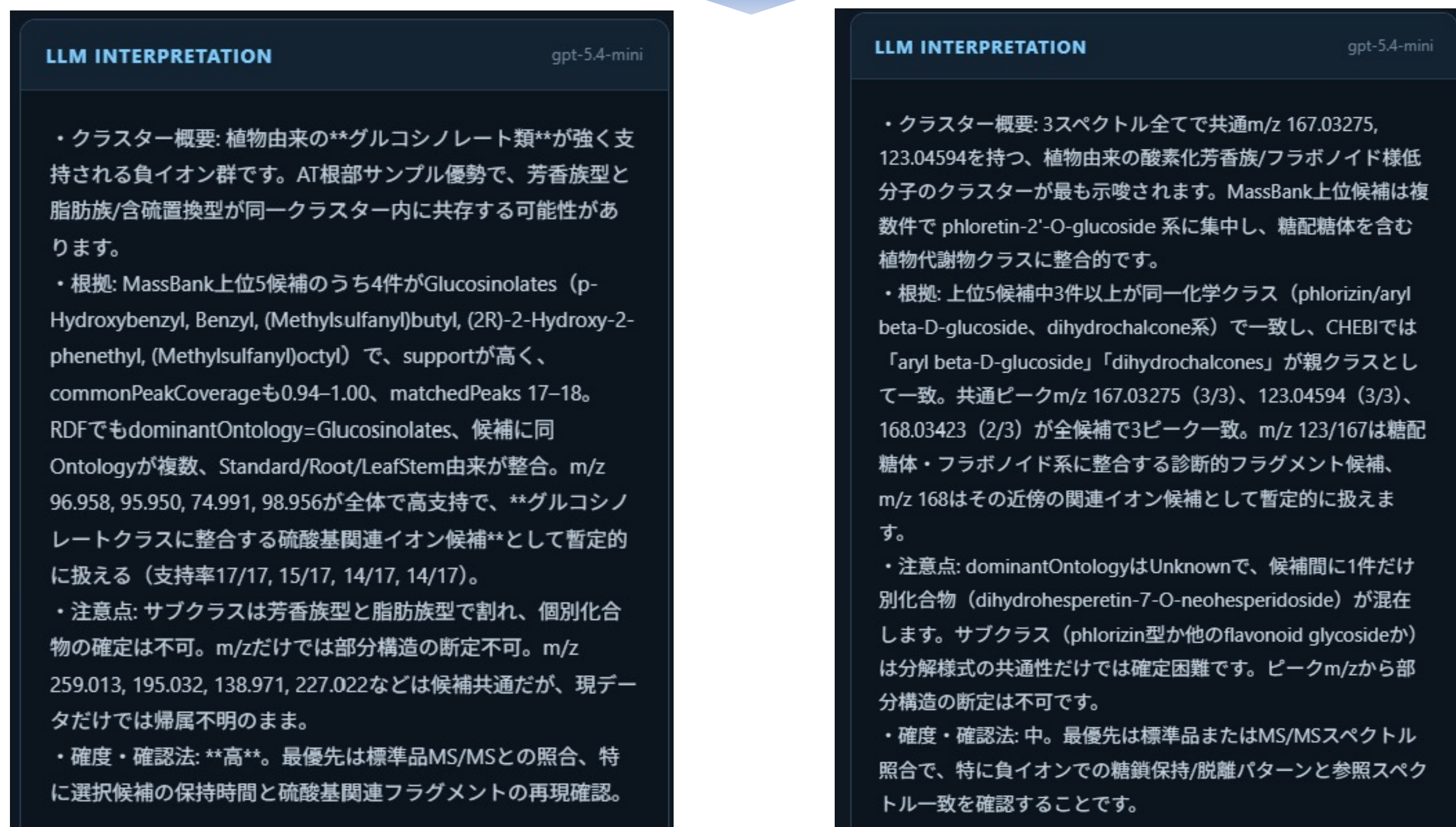


MassBank RDFを通じたメタデータアノテーション

61/84個のクラスターにメタデータを付与できた。



LLMによるメタデータ解釈



✓ MassBank RDF・生命科学データベースから取得したメタデータをLLMで統合し、分子ネットワークのクラスターを自動解釈する。

6. 謝辞

本研究開発はJST-NBDC (JPMJND2305) の支援を得て行われました。