

MassBank In silico: 低分子化合物の構造アノテーションツール

○鳥越大平¹⁾, 早坂亮祐²⁾, 高橋政友¹⁾, 廣瀬泰樹^{3,4)}, 秦康祐¹⁾, 津川裕司^{3,4)}, 平山明由²⁾, 松田史生⁵⁾, 和泉自泰¹⁾

¹⁾大阪大学大学院理学研究科, ²⁾慶應義塾大学先端生命科学研究所, ³⁾京都大学大学院薬学研究科,

⁴⁾東京農工大学大学院工学研究科, ⁵⁾大阪大学大学院情報科学研究科



I. OVERVIEW

Objective:

- To develop an ensemble-based annotation pipeline that integrates multiple in silico MS/MS prediction tools to assign structural information to MassBank spectra with higher confidence and transparency.

Methods:

- A unified-HILIC/AEX retention time prediction model was developed in Python/Scikit-learn and evaluated using a plasma sample (SRM-1950).
- The annotation tool was developed in a Python/Ubuntu/WSL2/Windows 11 environment.

Results:

- As a result, we constructed an integrated analysis pipeline combining CFM-ID, MetFrag, MS-FINDER, and SIRIUS.

II. INTRODUCTION

Objective

Structural elucidation of unknown metabolites

Approach

Construction of an in silico predicted mass spectral library (95,000 compounds) based on chemical structures. Development of an analytical pipeline that annotates high-quality MassBank Human spectral data with structural information



α Version Now Available

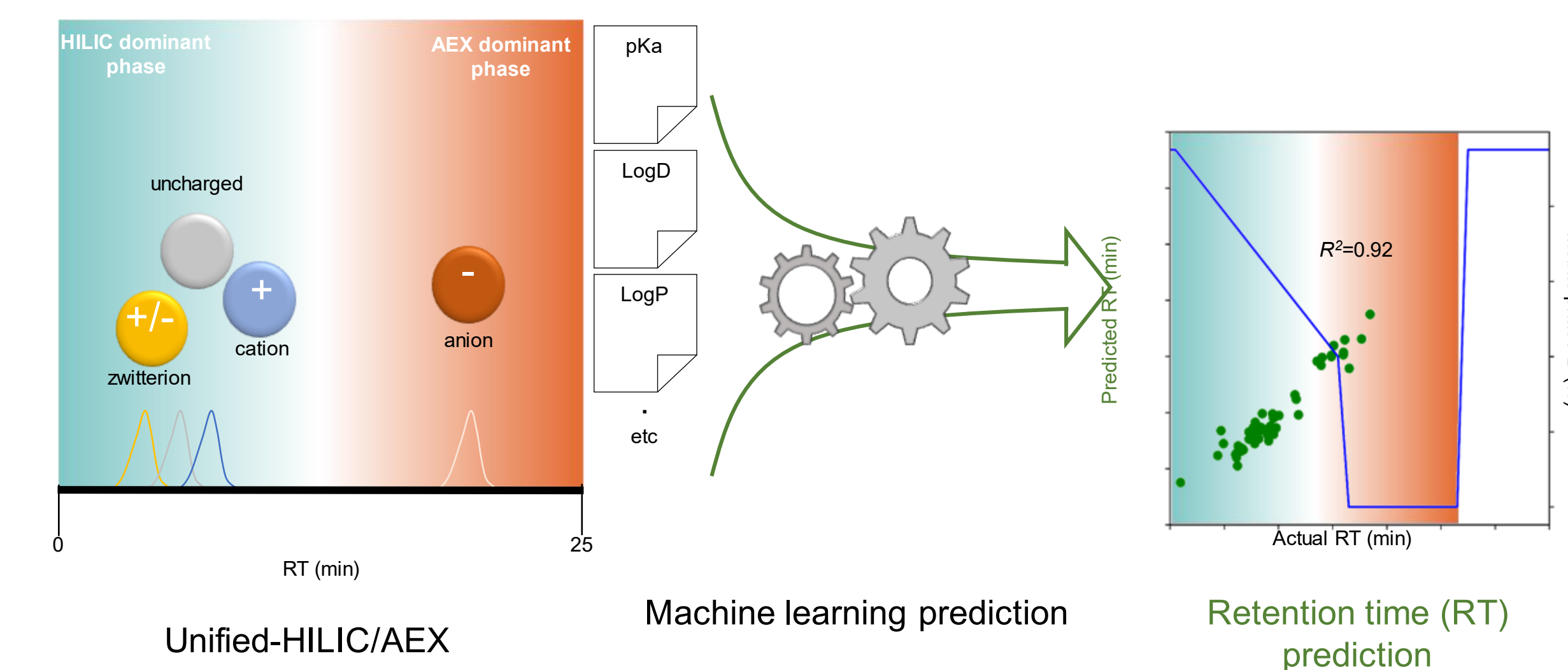
Stores in silico MS/MS prediction data for 95,000 metabolites based on conventional methods (including KEGG and HMDB database IDs and structures).

- Provides annotation functionality using in silico predicted spectra.
- Collects predicted MS/MS spectral data for metabolites lacking standard compounds.

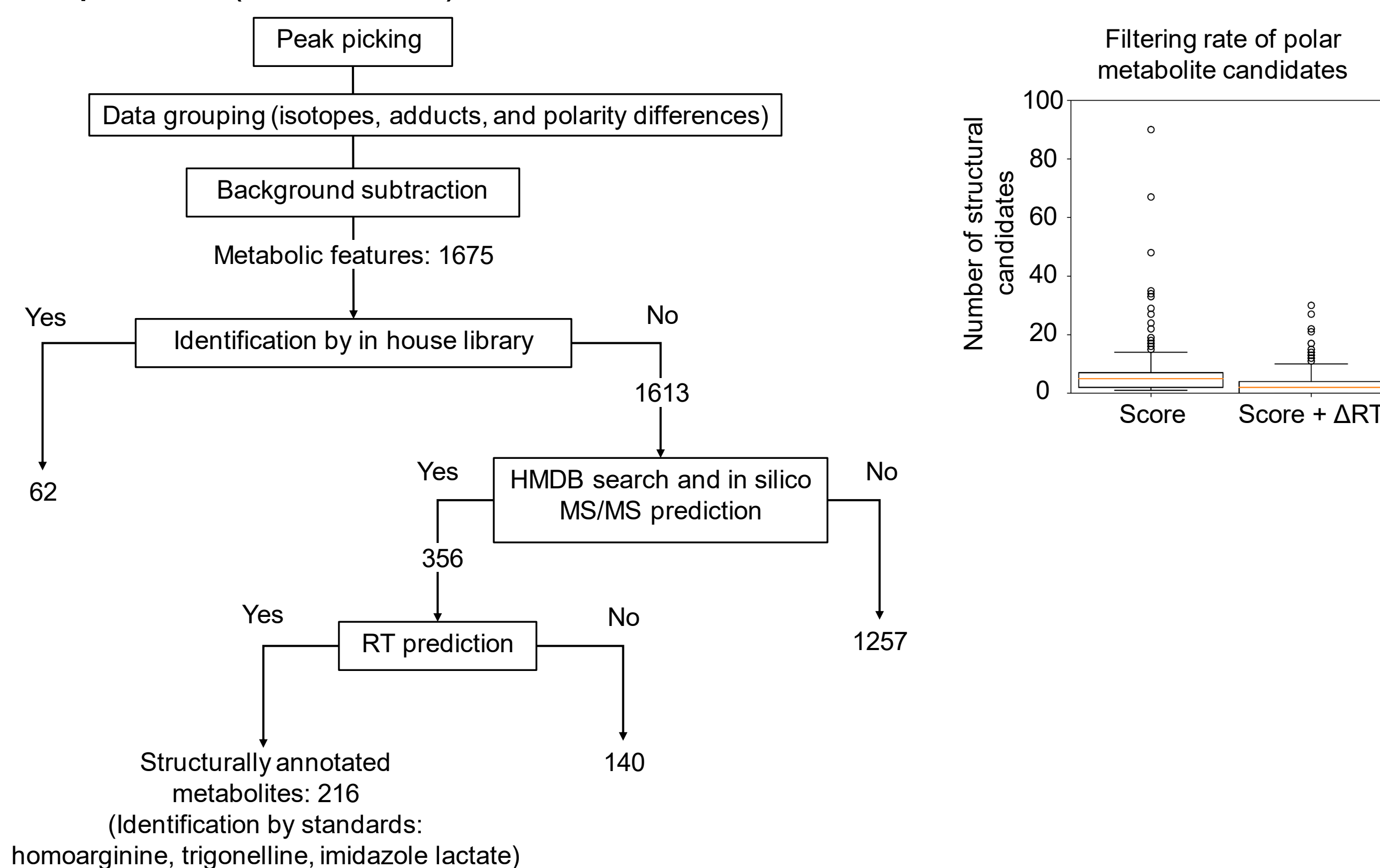
Recent advances in LC/MS-based metabolomics have enabled the detection of thousands of molecular features. However, 80–90% of them remain unidentified because a method that can be applied to all unknown metabolites has not yet been developed. To address this challenge, we developed an ensemble-based annotation approach that enables comprehensive structural inference for unknown metabolites.

III. RESULTS and DISCUSSION

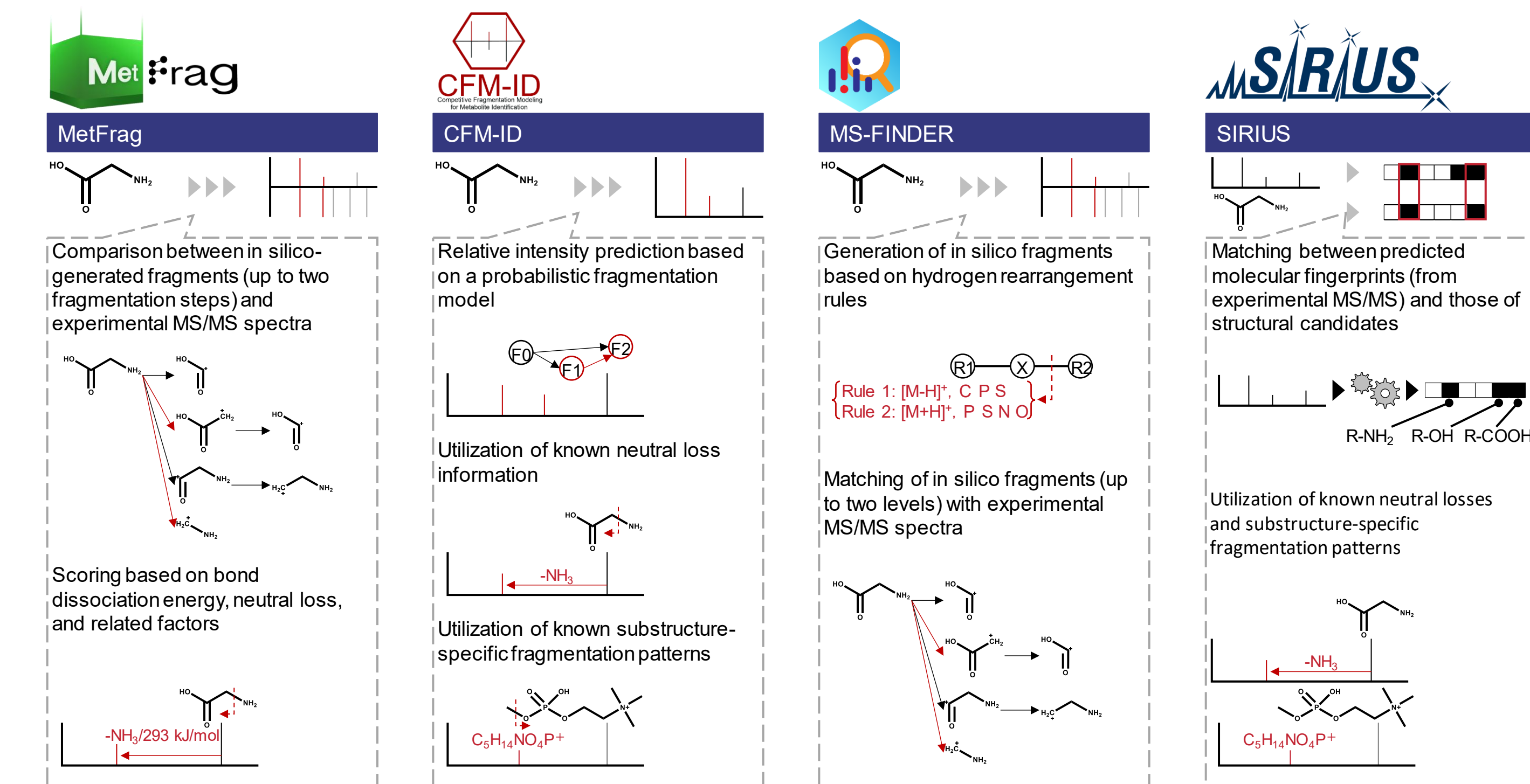
III-1. Unified-HILIC/AEX RT prediction model



III-2. Comprehensive structural annotation of polar metabolites in human plasma (SRM 1950)

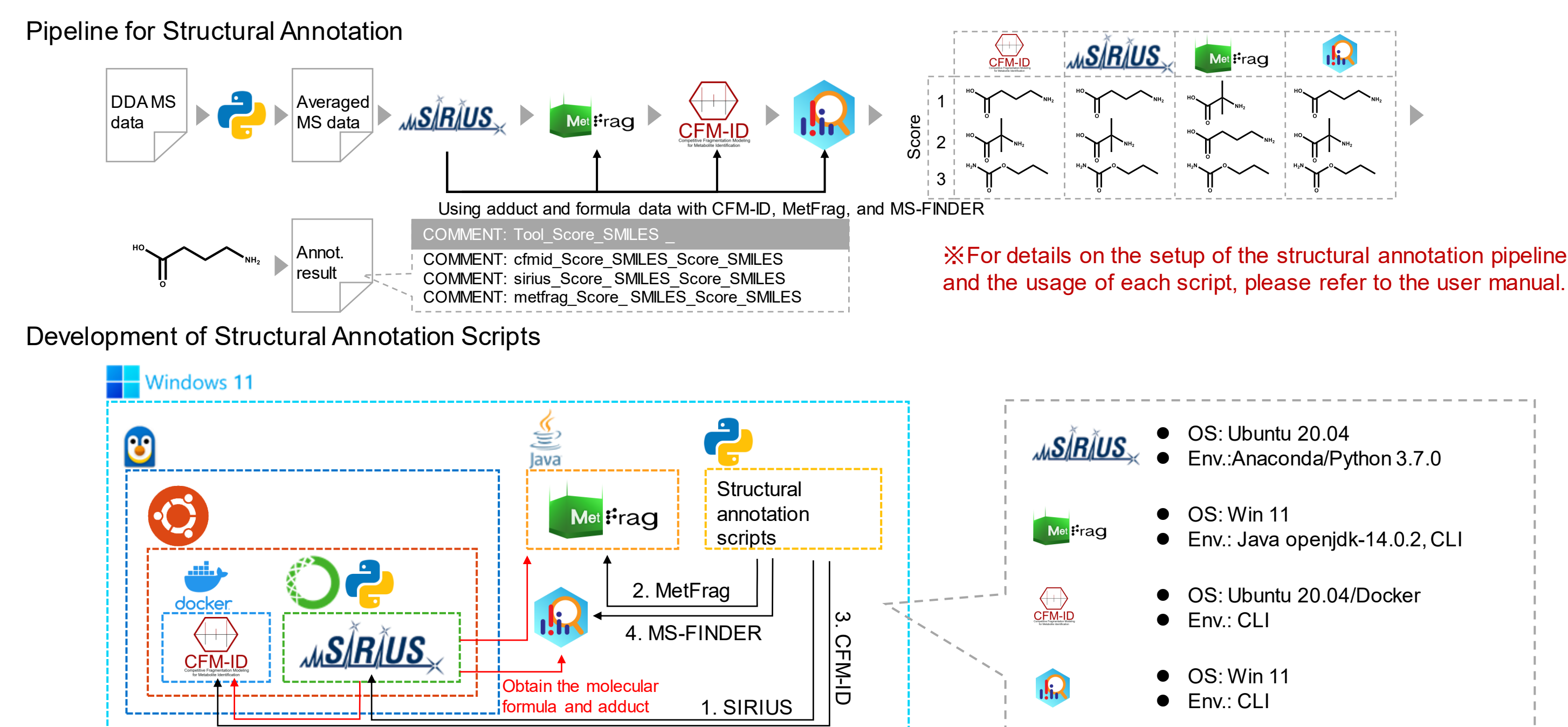


III-3. In silico structural annotation tools

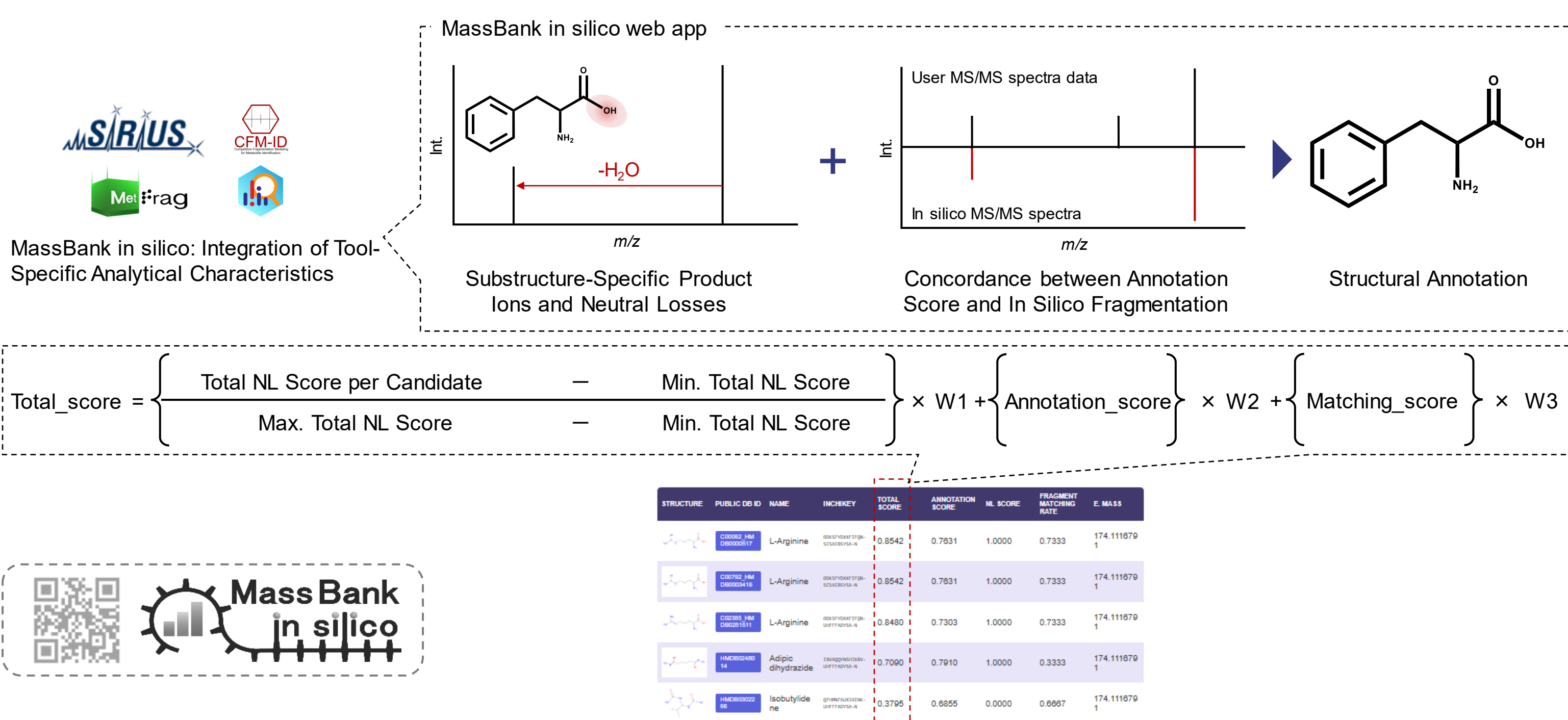


- Development of a novel structural annotation tool integrating features from existing tools
- Scoring based on bond dissociation energy and neutral loss
- Utilization of known substructure-specific fragmentation patterns
- Generation of in silico fragments based on hydrogen rearrangement rules

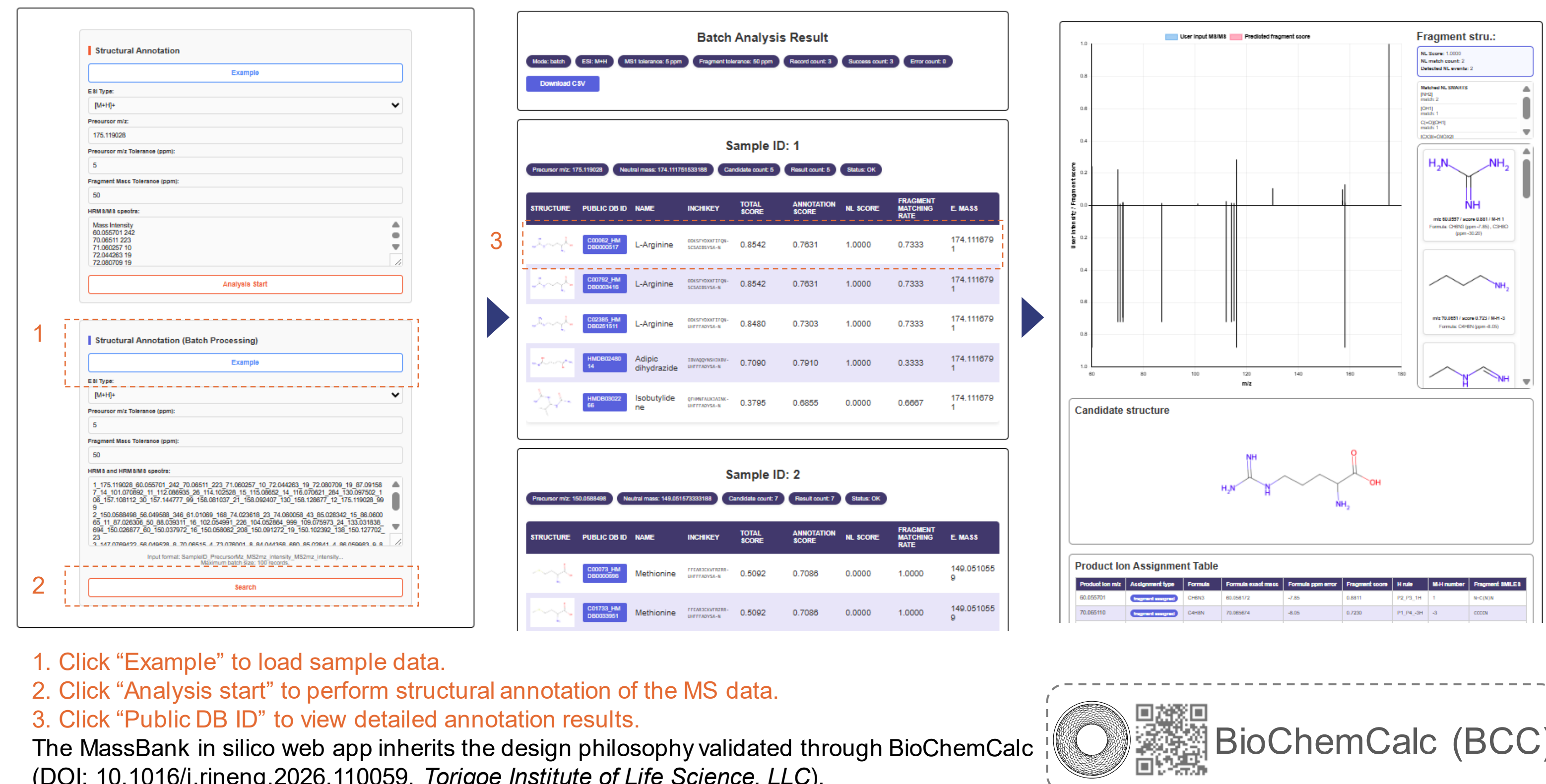
III-4. Workflow of structural annotation using MB in silico



III-5. MassBank in silico web app: Integration of Tool-Specific Analytical Characteristics



III-6. MassBank in silico Web App: Workflow and Analysis Examples



IV. CONCLUSIONS

- The median filtering rate of the structural candidates among the 356 metabolic features was 50%, which demonstrates the usefulness of the RT predictive filtering method.
- We constructed an integrated analysis pipeline combining CFM-ID, MetFrag, MS-FINDER, and SIRIUS.
- These results highlight the effectiveness of combining orthogonal strategies for improving structural annotation accuracy.
- We developed and released a new web app, MassBank in silico, which leverages the complementary strengths of existing annotation tools.
- As a future direction, we aim to apply retention time (RT) prediction to the annotation results to eliminate false positives.

Acknowledgement

This work was supported by JST-NBDC Life Science Database Integration Project “Shin-MassBank” Grant Number <JPMJND2305>. Torigoe T, et al., *Anal. Chem.*, 96(3), 1275-1283 (2024). doi: 10.1021/acs.analchem.3c04618

