Proposals for NORMAN Joint Programme of Activities 2024

Title	**Assessing the spectral quality and diagnostic information of MS2 data acquired with LC-ESI-HRMS**
Type of activity	Pilot Study
Leader	VU
Topic / activities	***Background / Justification for the proposed activity:*** Tandem mass spectra (MS2) are essential used for numerous applications, including the curation of spectra libraries and, more recently, the prediction of molecular (sub)structures and potential in-vivo or in-vitro toxicity of detected features. However, the quality of acquired MS2 data can have a large impact on the identification of features and the reliability of predictions. VU and KWR recently developed a machine learning algorithm to automatically assess the quality of MS2 spectra based on a set of mathematical features calculated from MS2 data[1]. **Aim:** The goal of this activity is to implement the recently developed algorithm to improve MS2 data from NORMAN members and support their activities. More specifically, we aim to apply the model to data from NORMAN (members), including large spectral libraries (e.g., MassBank), refine the model where needed (e.g., quality assessment criteria and model architecture), and finally evaluate its impact on identification and MS2-based toxicity prediction (e.g., MS2Tox and MLinvitroTox). ***Description of the proposed activity and expected outcomes for 2024:*** Activity 1: Gather MS2 data from NORMAN members and evaluate the model's output versus expert assessment using a quality score. Outcome: test the current model on more data and evaluate whether a scoring system (instead of the current binary output *poor* vs *good*) is more useful. Activity 2: Refine the model where needed based on outcomes from activity 1 (e.g., input variables, model architecture). Outcome: rollout a more advanced and robust version of the spectral quality assessment model based on activity 1. Activity 3: Apply the model on larger datasets, including libraries (e.g., MassBankEU), actual samples and data processed with different workflows/instruments. Outcome: test the model on actual data from different instruments and workflows. Activity 4: Evaluate the impact of the quality assessment model on the identification rate (e.g., based on spectral similarity scores) and on the outcomes of predictive approaches such as MS2Tox and MLinvitroTox. This can be carried out in parallel to the proposed JPA "Integration of computational toxicity driver prioritization tools to support non-target screening workflows in high-throughput effect-directed analysis". Outcome: Overview of model performances for various applications. ***Added value / Link with other NORMAN activities and / or other projects*** As mentioned above, this activity can be linked to the proposed activity "Integration of computational toxicity driver prioritization tools to support non-target screening workflows in high-throughput effect-directed analysis". Specifically, it can be used to determine whether improved predictions can be obtained when the quality of the MS2 spectra is assessed before hand. Furthermore, the proposed approach can be used to evaluate and curate libraries used in NORMAN (e.g., MassBankEU) and can potentially enhance identification rates. Results from the implementation of the spectral quality approach will be presented at a scientific conference (to be defined) and a dedicated (hybrid) workshop for interested parties will be organised in Amsterdam.
Participants	Eawag, SU, UniLU, KWR, VU,…
Proposed in-kind contribution	MS/MS data will be provided by participants. Data processing and computing will be provided in-kind by VU/KWR as part of ongoing (PhD) projects.
Contribution needed from NORMAN Association	Total: 9k€ (5k€ academic staff 0.1 FTE for 6 months for support during data collection, storage and processing; 2.5k€ for workshop organisation; 1.5k€ for travel and conference costs).

[1] S. Codrean et al., "Predicting the Diagnostic Information of Tandem Mass Spectra of Environmentally Relevant Compounds Using Machine Learning," *Analytical Chemistry* 95, no. 42 (October 9, 2023): 15810–17, https://doi.org/10.1021/acs.analchem.3c03470.