Appa: Agentic Preformulation Pathway Assistant

Julius Lange[1], Leonid Komissarov[2], Nicole Wyttenbach[1], Andrea Anelli[1]

Roche Pharma Research & Early Development (pRED), [1]Therapeutic Modalities, [2]Data & Analytics

The formulation of a small molecule in a drug development setting aims at enabling the delivery of a compound of choice within the desired dosage range and timing. This is usually achieved by combining several physico-chemical assays (e.g. solubility measurements) to probe which vehicles facilitates the diffusion of a molecule in a target environment.

Appa presents a unified, intuitive interface through an LLM agent for the task of preclinical drug formulation[1], enabling:

- Seamless access to experimental data and machine learning predictors

- Ability to respond to diverse scientific questions by chaining different tools

- A faster preformulation process through knowledge of experimental evidence, operating procedures and follow-up recommendations

- Clear decision making in the experiment planning
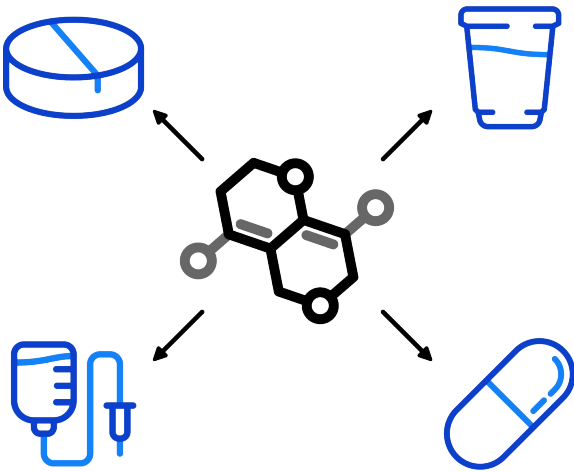
One substance – multiple formulations
Crucial for efficacy and activity of a drug

Example of a preformulation agentic pipeline
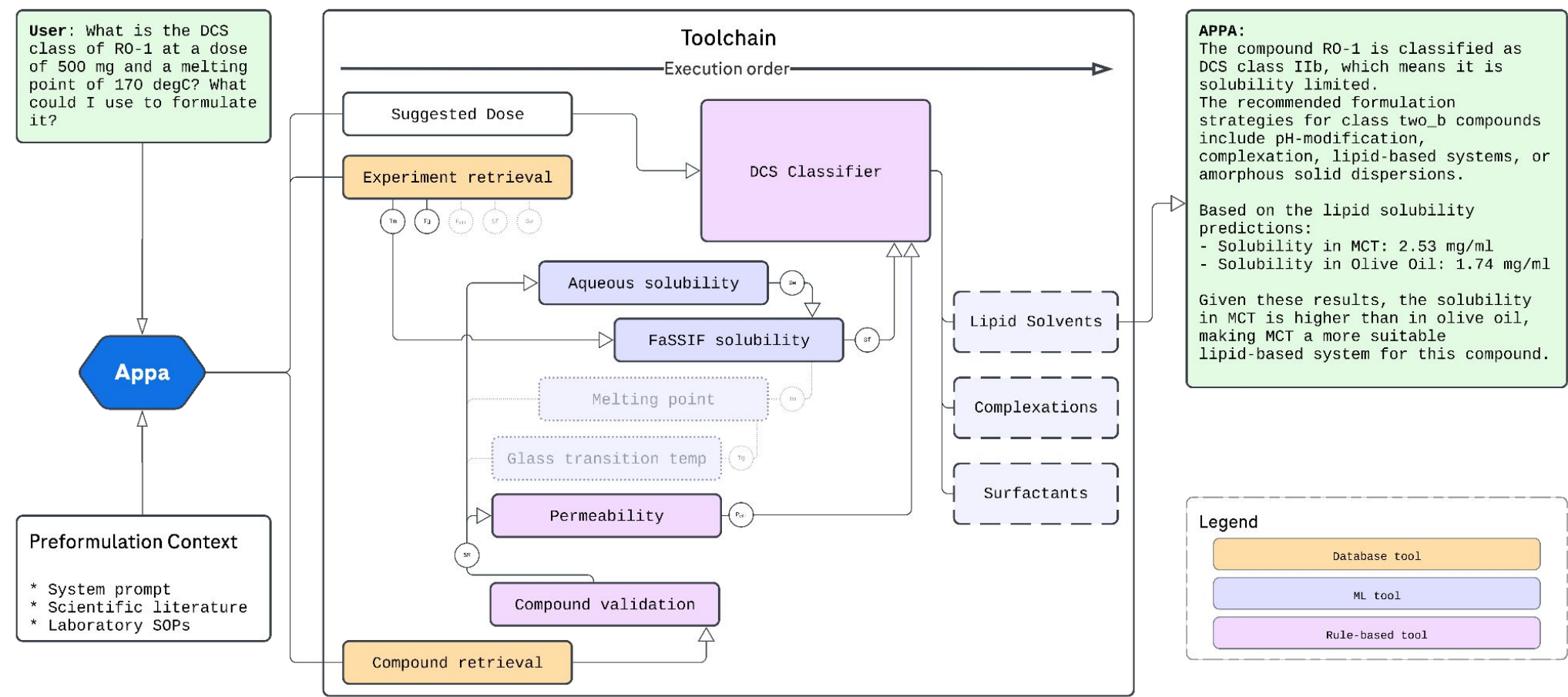
Evaluation

- **Task**: Classification of a substance based on the Developability Classification System[2] (DCS)

- Scientific knowledge and understanding of multiple chemical properties required

- Testing of 500 virtual compounds based on structures from the ChEMBL[3] database

query: What is the DCS class of SMILES [SMILES] at a dose of [DOSE] mg and a melting point of [MELTING POINT] degC?

Method	Accuracy	F1-micro
APPA	0.895	0.938
GPT-4o + Context	0.156	0.186

Future directions

The workflow introduced in this work focuses on small - yet fundamental - task in a preformulation pipeline, which is the vehicle selection. In the following, we will highlight the next steps along two separate axes:

More ML models for formulation route selections
Our next steps will cover both the integration of further machine learning methods to cover broader vehicle spaces and rational next screening steps.

Multi-agentic interface to cover automatic screening planning
We plan on introducing a second agent, whose focus would be aimed at translating the experimental design into lab orchestration operations - to automate the preparation and the execution of solubility assays.

References
1: Vargason et al " Nature biomedical engineering 5.9 (2021): 951-967.
2: Butler and Dressmann, *J. Pharm. Sci* (2010) **99** 12, 4940-4954
3: Zdrazil *et al.*, *Nucleic Acids Research* (2024) **52** D1, D1180-D1192

Go to publication