**Physics Informed One-stop AI Preformulation Prediction Platform**

**Nannan Wang1**, Defang Ouyang1,2.

State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences (ICMS), University of Macau, Macau, China1

Department of Public Health and Medicinal Administration, Faculty of Health Sciences (FHS), University of Macau, Macau, China2

**Background and aims.** The surge of AI-driven drug discovery has flooded development pipelines with hit compounds. However, these increasingly complex molecules face significant developability challenges and formulation design demands. High-throughput and one-stop preformulation investigation tools are urgently needed to reduce development attrition. Although some preformulation properties have been mentioned in molecular machine learning, existing studies are limited by data quality, chemical space coverage, and often overlook drug solid-state characteristics and pharmaceutics-specific issues. Against the mentioned backdrop, this study aims to develop a physics informed one-stop AI platform for preformulation prediction and developability investigation (Figure 1).

**Methods.** To begin with, a standardized data management program was created to harmonize multi-source datasets and ensure high data quality. The generative transformer-based algorithm, Tabular Prior-data Fitted Network (TabPFN)[1], was employed to construct robust and interpretable models. Crystal density prediction was incorporated to capture the impact of crystal structures on preformulation properties. Physics informed hierarchical modeling, task-specific ensemble learning, and tailored data management strategies were implemented to enhance model performance, interpretability, and robustness.

**Results.** First, 12 high-quality preformulation datasets were collected (Table 1). To adapt to pharmaceutical scenarios, the solubility datasets covered thermodynamic equilibrium solubility, kinetic solubility, and solubility in organic solvent and mixed solvents. Leveraging these datasets and tailored training strategies, 12 robust TabPFN models were developed, achieving state-of-the-art performance. Crystal density, linked to lattice energy, was incorporated into the prediction of solid-state-related properties, such as equilibrium solubility, enabling preformulation assessment across different crystal

structures. Subsequently, empirical physics formulas were employed to provide semi-quantitative evaluations of hard-to-obtain key properties, such as intrinsic solubility. Furthermore, an interpretable partial supervised learning model was further refined to conduct developability assessment[2]. Finally, a user-friendly web platform will be created for batch prediction and generating interpretable formulation descriptors (IF-Des) to benefit the AI drug development community.

**Conclusion.** The first physics informed AI preformulation prediction platform was established. It pioneeringly incorporates physical pharmaceutics information to enable one-stop preformulation exploration. As a vital part of AI drug development, it helps identify developability challenge, guide leads optimization, optimize formulation design, ultimately enhancing R&D efficiency and drug quality.

**Acknowledgements.** This work was financially supported by the Macau Science and Technology Development Fund (0071/2024/RIA1).

**References:**

[1] Hollmann, Noah, et al, Nature 637.8045 (2025): 319-326.

[2] Wang, Nannan, et al, Journal of Controlled Release 378 (2025): 619-636.

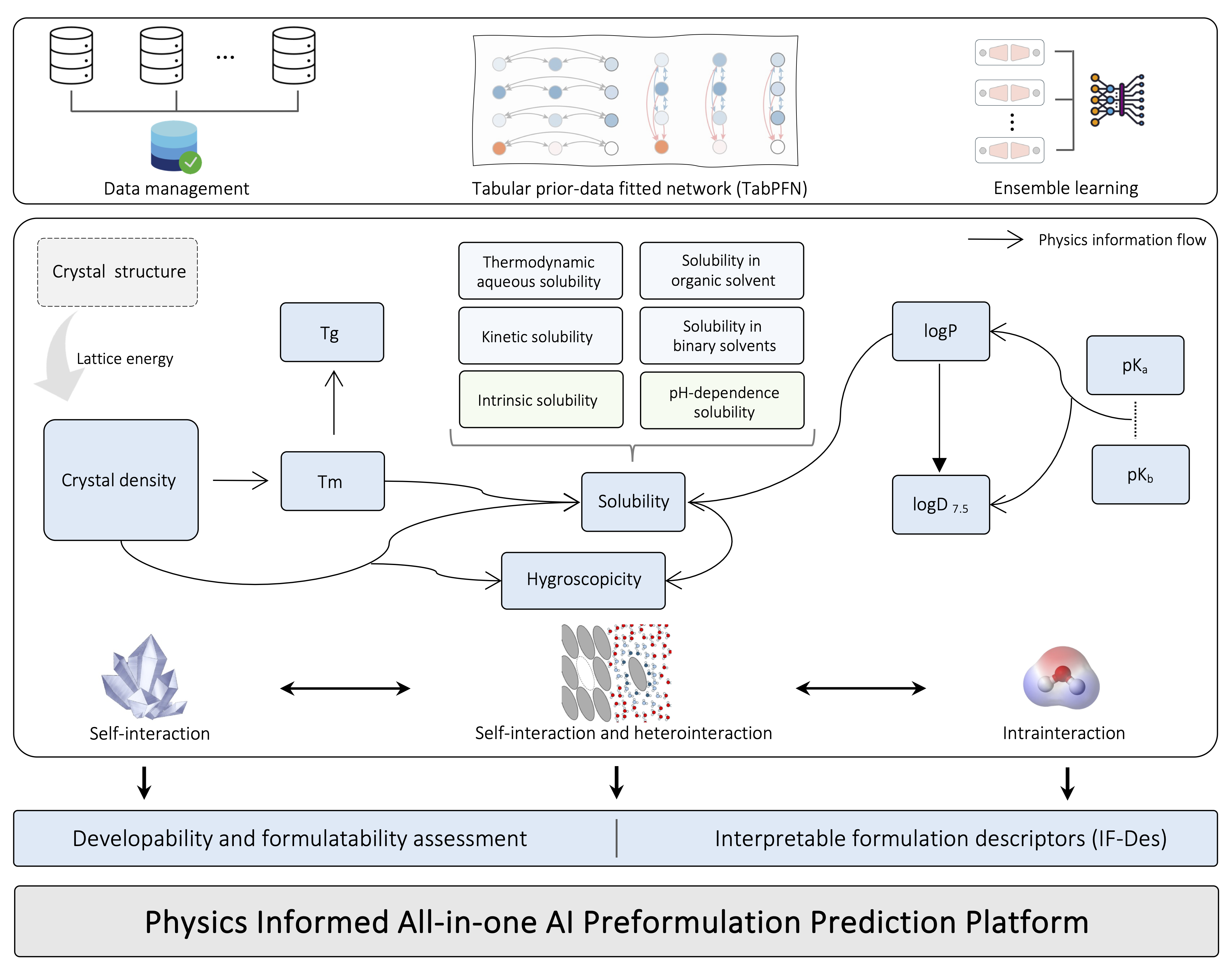


Figure 1 Graphical Abstract

|  |  |
| --- | --- |
| **Datasets** | **Date volume** |
| Thermodynamics equilibrium solubility | 15354 |
| Kinetic solubility | 52161 |
| Organic solvent solubility | 5031 |
| Mixed solvents solubility | 28101 |
| Crystal density | 171891 |
| Melting point | 273297 |
| Glass transition temperature | 1100 |
| Acidic dissociation constant (pKa) | 4352 |
| Basic dissociation constant (pKb) | 4121 |
| Logarithm of the octanol-water partition coefficient (logP) | 14717 |
| Logarithm of the octanol-water partition coefficient at pH 7.5 (logD7.5) | 5088 |
| Hygroscopicity | 635 |

Table 1 Datasets overview