Introduction

Introduction to Aqueous Solubility: Understanding the aqueous solubility of drug molecules is fundamental before building a machine learning model for it. Aqueous solubility refers to the maximum amount of a compound (solute) that can dissolve in a given volume of water at a specific temperature and pressure [1]. This property is heavily dependent on physicochemical factors such as hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), polar surface area (PSA), and material purity. However, it is also sensitive to experimental conditions, including buffer composition (pH), stirring time, sedimentation time, and phase separation techniques [1, 2].

Types of Solubility: Kinetic vs. Thermodynamic: To understand the data curation issues, we must distinguish between the two measurement modes:

- **Kinetic Solubility:** This is defined as the highest concentration of a compound that can remain in solution before precipitation occurs, typically measured when a pre-dissolved stock solution (often in DMSO) is added to an aqueous buffer. It represents a pseudo-equilibrium or metastable state [2, 3].
- **Thermodynamic (Equilibrium) Solubility:** This is the concentration of a compound in a saturated solution when the system has reached true equilibrium with excess solid present. This is the definitive physicochemical property of the molecule [2, 3]

Experimental Procedures:

- **The Shake-Flask Method (Gold Standard for Thermodynamic Solubility)** The shake-flask method is the standard reference for determining equilibrium solubility. In this procedure, an excess amount of the solid drug is added to a solvent in lined flasks. The samples are agitated (shaken) at a controlled temperature until an equilibrium-saturated solution is obtained (often requiring 24 to 72 hours). A portion of the supernatant is then filtered or centrifuged to remove the remaining solid and analysed using High-Performance Liquid Chromatography (HPLC-UV) [4]. While precise, this method is low-throughput, labor-intensive, and material-heavy.
- **Turbidity-Based Assays (Kinetic Solubility)** To increase throughput for early-stage screening, turbidity assays (or nephelometry) are used. A stock solution of the compound (dissolved in DMSO) is titrated into an aqueous buffer. A light-scattering detector monitors the solution; the point at which the solution becomes cloudy (turbid) indicates precipitation [5].
- **Potentiometric Titration (CheqSol)** The CheqSol method is a thermodynamic approach specifically for ionizable drugs. It determines solubility by monitoring changes in pH during titration. By adding acid or base to a solution containing the drug and its

solid form, the method detects the "Supersaturation Limit" and the "Subsaturation Limit" to calculate the equilibrium solubility [6,7].

Citations

1. Sorkun, M.C., Khetan, A. & Er, S. AqSolDB, a curated reference set of aqueous solubility and 2D descriptors for a diverse set of compounds. *Sci Data* **6**, 143 (2019).
2. Baka, Edit & Comer, John & Krisztina, Takacs-Novak. (2008). Study of equilibrium solubility measurement by saturation shake-flask method using hydrochlorothiazide as model compound. Journal of pharmaceutical and biomedical analysis. 46. 335-41. 10.1016/j.
3. Loftsson, Thorsteinn & Hreinsdóttir, Dagný. (2006). Determination of aqueous solubility by heating and equilibration: A technical note. AAPS PharmSciTech. 7. E4. 10.1208/pt070104.
4. Zarghampoor, Aynaz & Jouyban, Kimiya & Jouyban-Gharamaleki, Vahid & Jouyban, Abolghasem & Rahimpour, Elaheh. (2024). A Description on the Shake-Flask and Laser Monitoring-Based Techniques for Determination of the Drug's Solubility. Pharmaceutical Sciences
5. (2023). *Solubility determination of chemicals by nephelometry : 40 test items relevant to EU-NETVAL validation study of in vitro methods for the detection of thyroid hormone system disrupting chemicals*. Publications Office of the European Union.
6. Stuart, M. & Box, K. (2005). Chasing equilibrium: measuring the intrinsic solubility of weak acids and bases. *Analytical Chemistry*, 77(4), 983-990.
7. Etherson, K., Halbert, G., & Elliott, M. (2014). Determination of excipient based solubility increases using the CheqSol method. *International journal of pharmaceutics*, *465*(1-2), 202–209.