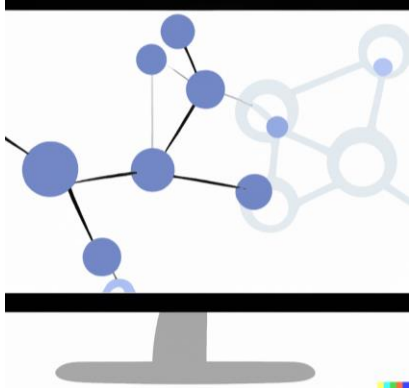


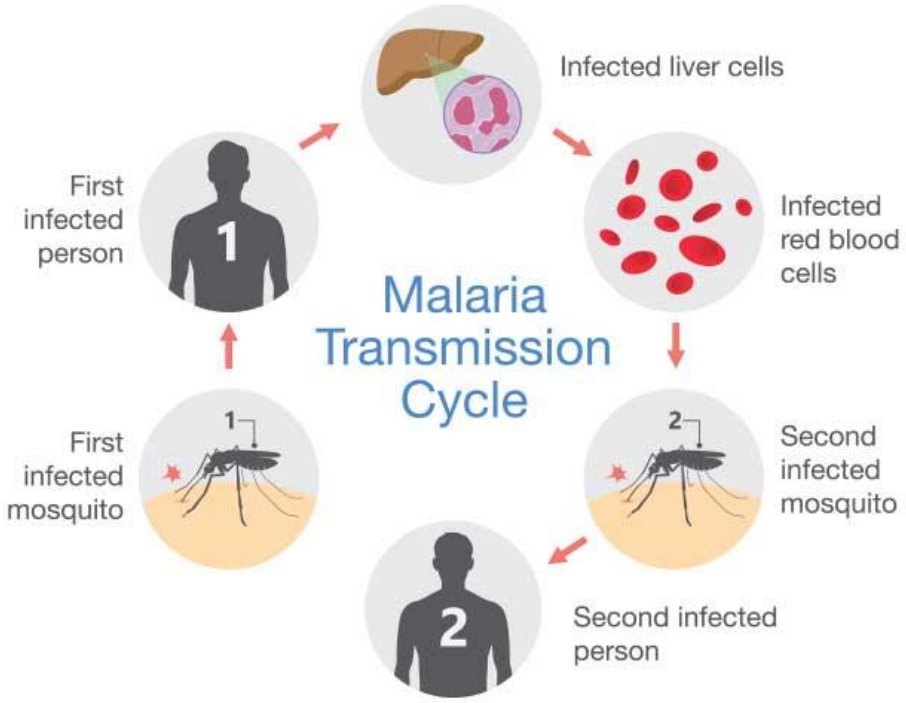
HS01: Exploring Antimalarial Drug Discovery Through Cheminformatics and QSAR

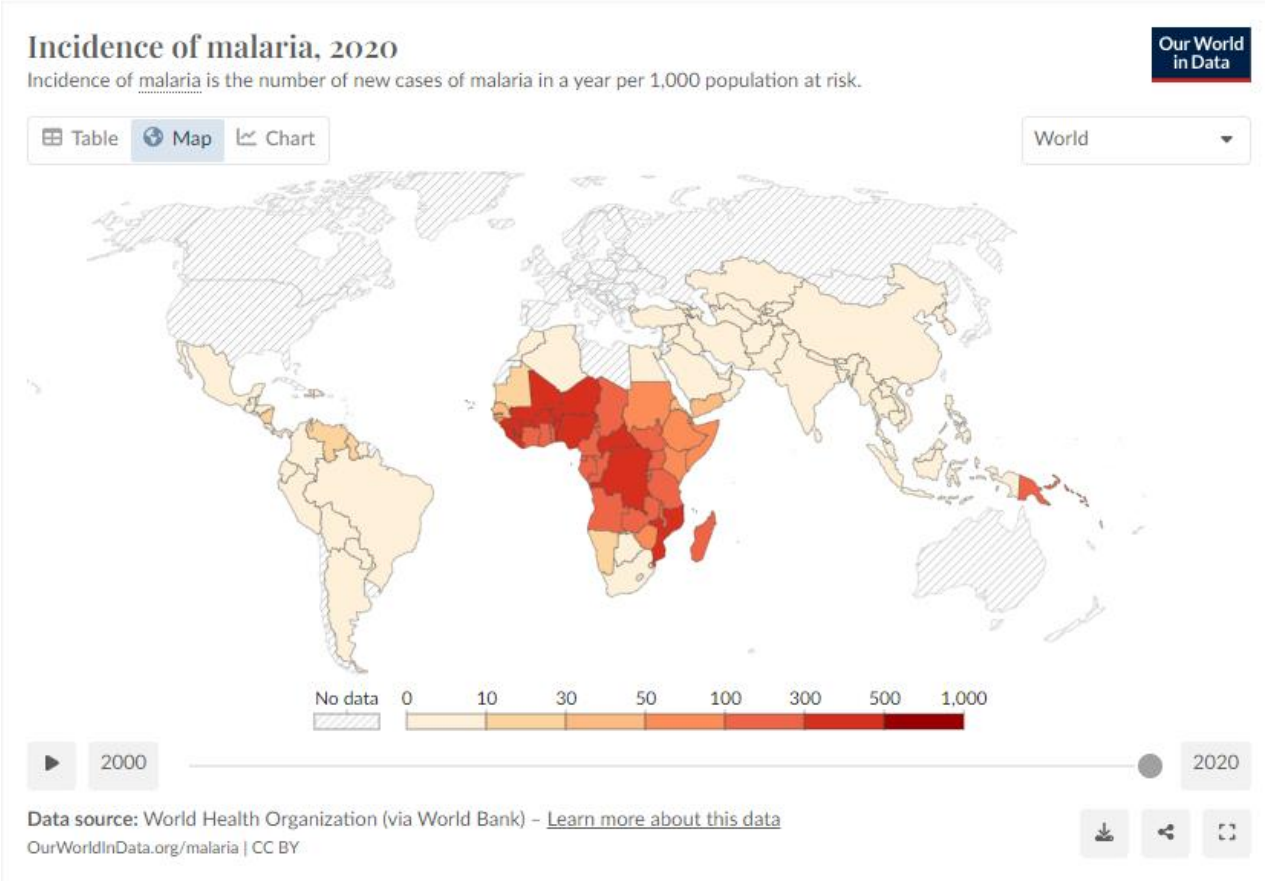
Yannick Djoumbou Feunang, PhD

02/27/2024

CAiSMD 2024: Computational Applications in Secondary Metabolite Discovery

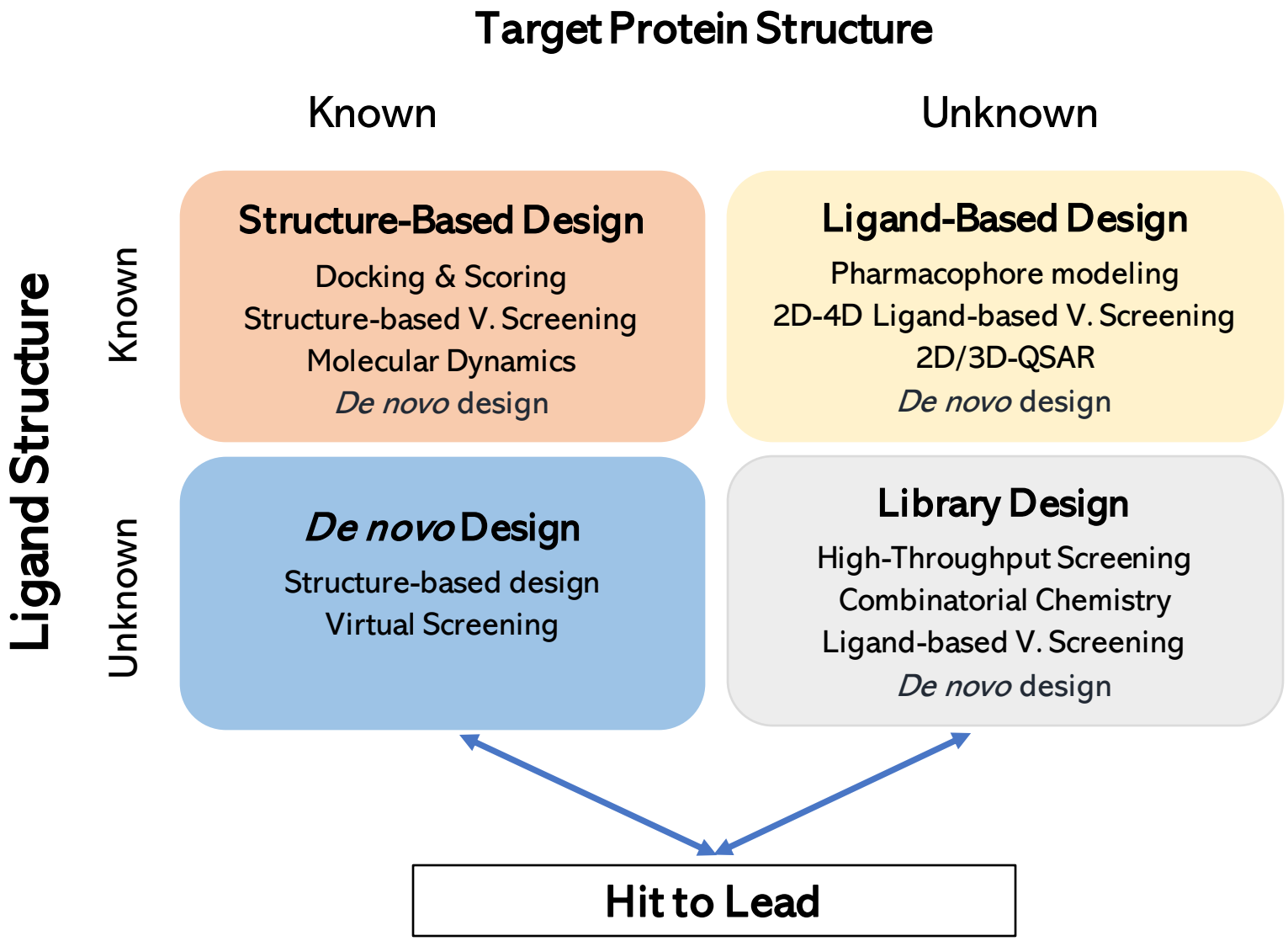
What is Malaria?





Over half a million people died from the disease each year in the 2010s

Computer-Aided Drug Design



The Drug Discovery Pipeline

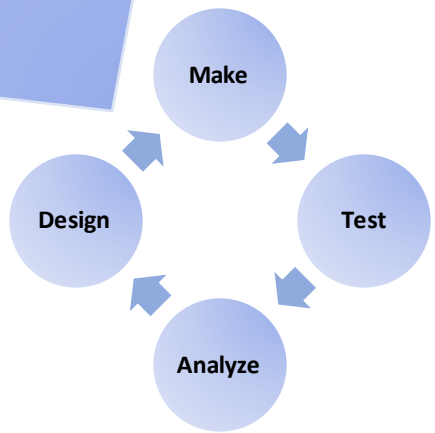
R&D: Molecule Discovery → Clinical Trials → Regulatory Approval: FDA → Product

Hit->Active->Lead Generation → Lead Optimization → Candidate Molecules ready for Trials

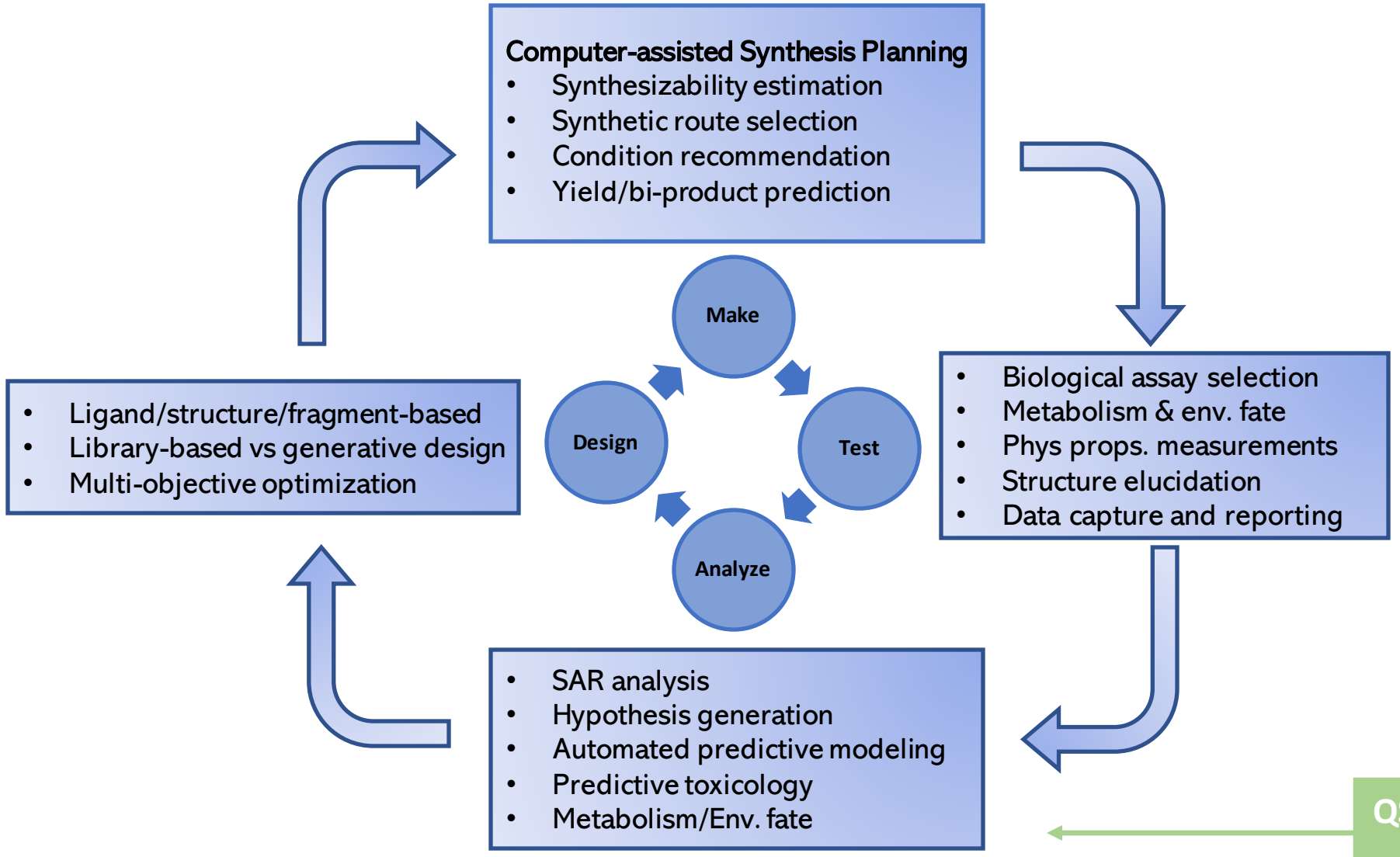
Find novel starting points for molecule discovery projects

Iterative molecule optimization until product goals are met or project discontinued

>=1 molecule ready for clinical trial testing.



Iterative DMTA Cycles

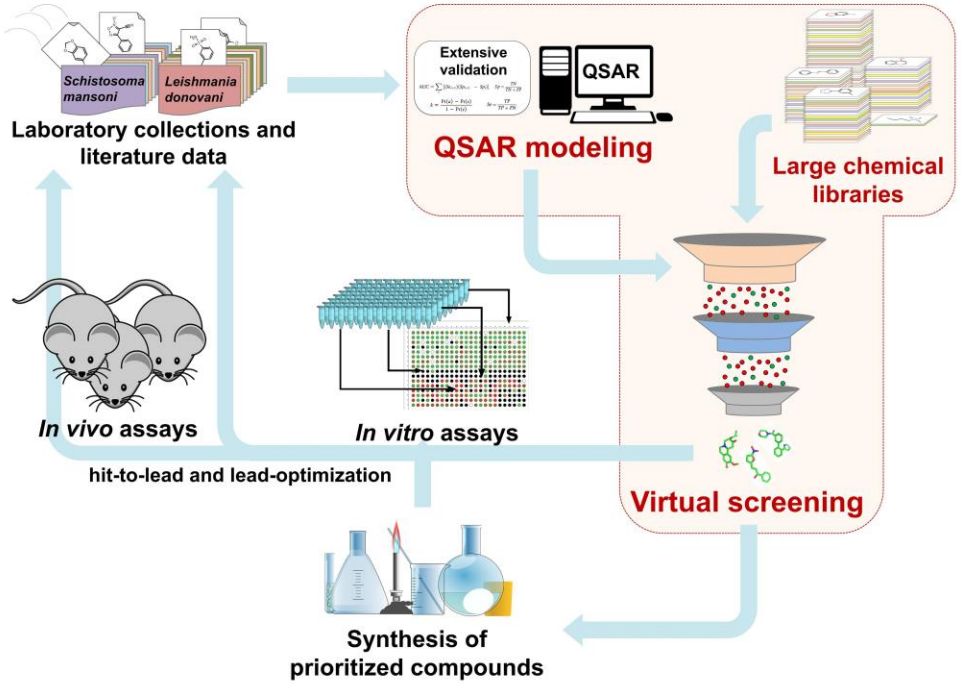


What is Cheminformatics?

- Cheminformatics is an interdisciplinary field that combines principles and techniques from chemistry, computer science, and information science to analyze, model, and understand chemical data.
- It finds applications in many fields of research, including, among others:
  - drug discovery,
  - chemical database management
  - materials science
  - chemical toxicology and environmental chemistry, and
  - pharmacogenomics, and personalized medicine

What is QSAR Modeling?

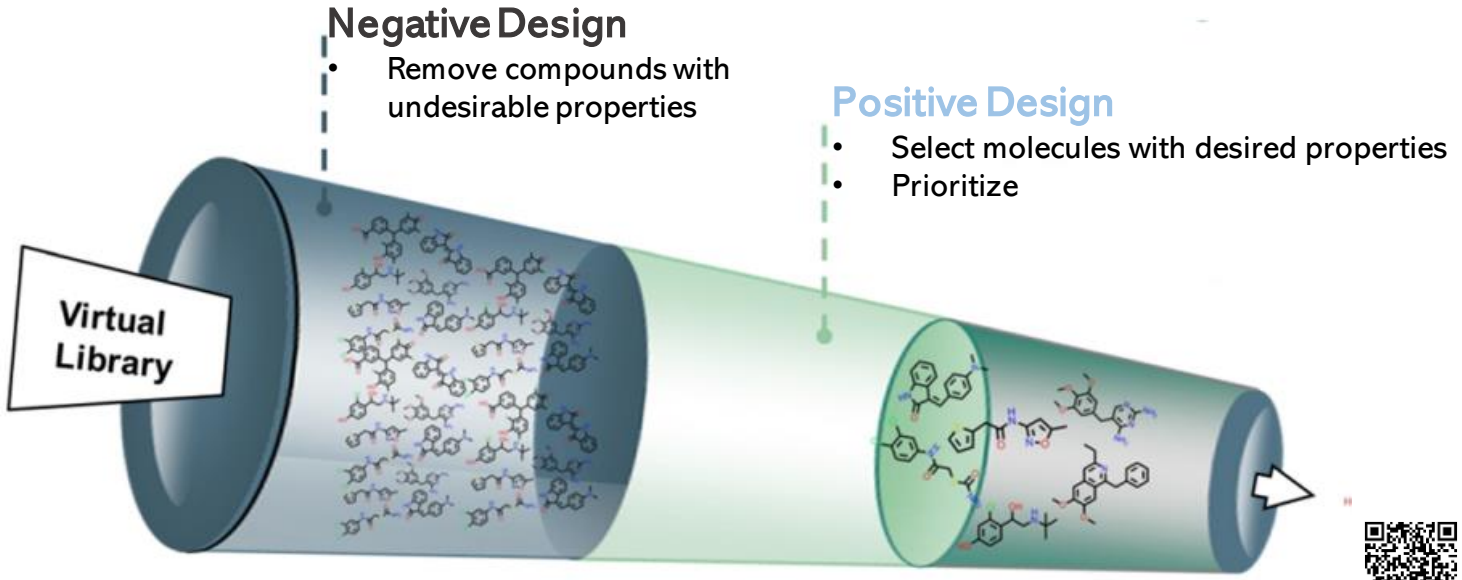
- **Quantitative Structure-Activity Relationship (QSAR)** modeling is a computational modeling method that aims to establish mathematical relationships between the structural properties of a molecule and its biological activity
- QSAR provides a cost-effective and efficient way to estimate the activity of new molecules, reducing reliance on expensive and time-consuming experiments.



Neves *et al.* 2018; doi.org/10.3389/fphar.2018.01275

What is QSAR Modeling?

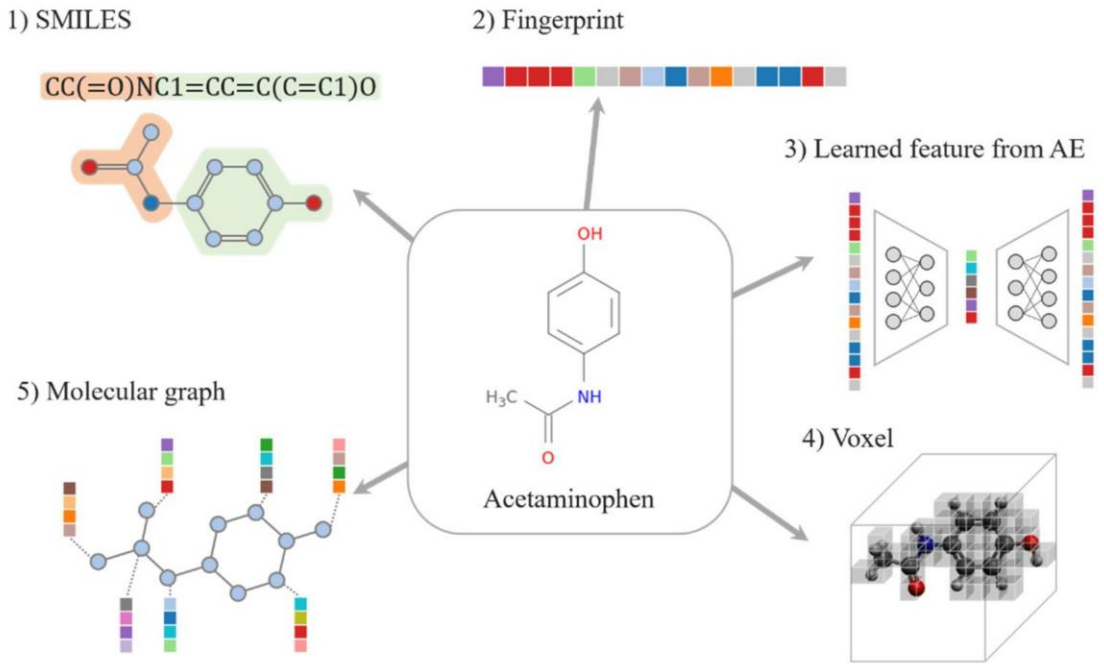
- It helps researchers prioritize promising candidates for further investigation and development, focusing resources on those most likely to be successful.
- QSAR modeling can be used at several stages of the discovery pipeline



Yang *et al.* (2020)

Molecular Representations

- QSAR models usually take mathematical representations of chemicals that can be derived from their structures.
- These can be either:
  - fixed (e.g.: fingerprints, phys. Chem. Properties)
  - Learned: from unsupervised or semi-supervised methods (e.g. deep learning)
  - combined



Classical operations

- **Data Collection and Preparation**
  - Gather the molecules and the target data; Clean and pre-process; descriptors selection
- **Data Splitting**
  - Divide data into train, validation, and test sets
- **Model Selection and Training**
  - Choose a suitable QSAR modeling technique:, and train the models
- **Model Validation**
  - Evaluate the model's performance on the validation set, using suitable metrics; refine the model
- **Model Application and Prediction**
  - Predict the biological activity of new molecules; interpret the model (optional)
- **External Validation and Reporting**
  - Optionally, perform external validation, on a completely independent dataset; document and report the results

Hand-on Session Activities

- We will provide a tutorial that explains with practical examples, the why and how of:
  - Molecular data cleaning, standardization, visualization and splitting
  - QSAR model design, training, and validation
  - QSAR prediction explanation
  - Applicability Domain definition and estimation
- The predictive models will aim at predicting growth inhibition for *P. Falciparum*.
- The code is shared, and can be executed easily
- The tutorial and material are available on GitHub

Relevant Literature

- Comprehensive
  - Djoumbou-Feunang *et al* (2023); Cheminformatics and artificial intelligence for accelerating agrochemical discovery
  - Kwon *et al.* (2019); Comprehensive ensemble in QSAR prediction for drug discovery
- Molecular Representations
  - David et al. (2020); Molecular representations in AI-driven drug discovery: a review and practical guide
- QSAR Data Curation
  - Free and open-source QSAR-ready workflow for automated standardization of chemical structures in support of QSAR modeling

Relevant Literature

- Gradient Boosting
  - Boldini et al. (2023); Practical guidelines for the use of gradient boosting for molecular property prediction
- Explainable AI
  - Jiménez-Luna et al. (2020); Drug discovery with explainable artificial intelligence
  - Rodríguez-Pérez et al. (2020); Interpretation of machine learning models using shapley values
- Domain of Applicability
  - Comparison of Different Approaches to Define the Applicability Domain of QSAR Models


Where?
What if?
When?
Why?
Could it?
How?
What?
Would it?
Ask A Question