

# Interview Presentation

Wei Ma

2025/05/

# 马 威

地址：南京市浦口区文德东路15号 手机：**18855488178** 邮箱：[mawei8178@foxmail.com](mailto:mawei8178@foxmail.com)

擅长于：**CADD/AIDD/药物发现一体化平台构建/AI大模型应用/数字化平台建设**



## 教育背景

2014.09—2018.06	安徽理工大学	药学	理学学士
2018.09—2021.06	华东理工大学	药学	药学硕士 (AIDD 方向)

## 实习经历

2019.09—2020.09	北京英飞智药科技有限公司	CADD研究员
-----------------	--------------	---------

## 工作经历

2021.07—至今	南京药石科技股份有限公司	AI 人工智能研究员
------------	--------------	------------

## 工作期间承担项目

- 主要职责**
- 负责AIDD (AI-driven drug discovery) 药物研发，包括活性和选择性小分子药物的发现及优化、性质筛选、虚拟筛选、基于片段和化学反应的分子生成；
  - 参与计算模拟一站式平台的前端开发，优化用户体验，提高药物研发的计算效率与降低计算工具使用门槛；
  - 应用大模型进行项目文档的自动翻译、文献文本数据提取与建库，提升文献分析和数据管理的自动化水平；
  - 负责数字化监控管理平台的搭建与应用（如：贵金属催化剂流转周期监控平台）

工作期间参与的具体项目包括但不限于以下：

# 一站式先导化合物计算筛选平台

项目目标:开发活性和选择性小分子化合物一站式计算筛选平台

## 集成模块

小分子的预处理、受体结构的准备、激酶选择性预测、蛋白-小分子复合物MD模拟等

### 一站式集成

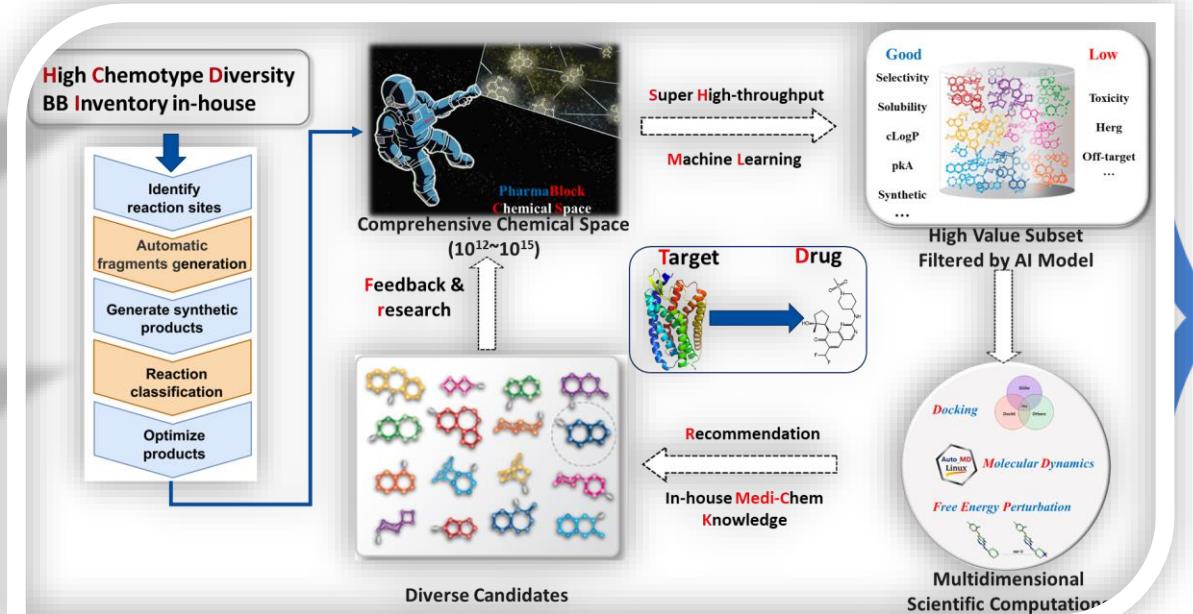
涵盖早期小分子药物发现的常用计算工具和多种人工智能模型

### 支持超高通量的筛选需求

轻松实现超高通量活性小分子筛选需求

### 应用便捷性和低成本

低学习成本指导湿实验团队研发；有效降低早期药物发现的成本



提高早期药物筛选命中率

发现新颖活性分子

降低早期药物发现的成本

对平台进行应用性测试并得到活性和选择性均较好的小分子化合物

# 高选择性靶标活性分子发现平台

## 模型功能: 快速高效预测小分子化合物激酶谱选择性情况

### 模型特点

快速对191个激酶进行活性预测，并对预测结果提供合理解释

#### 快速高效

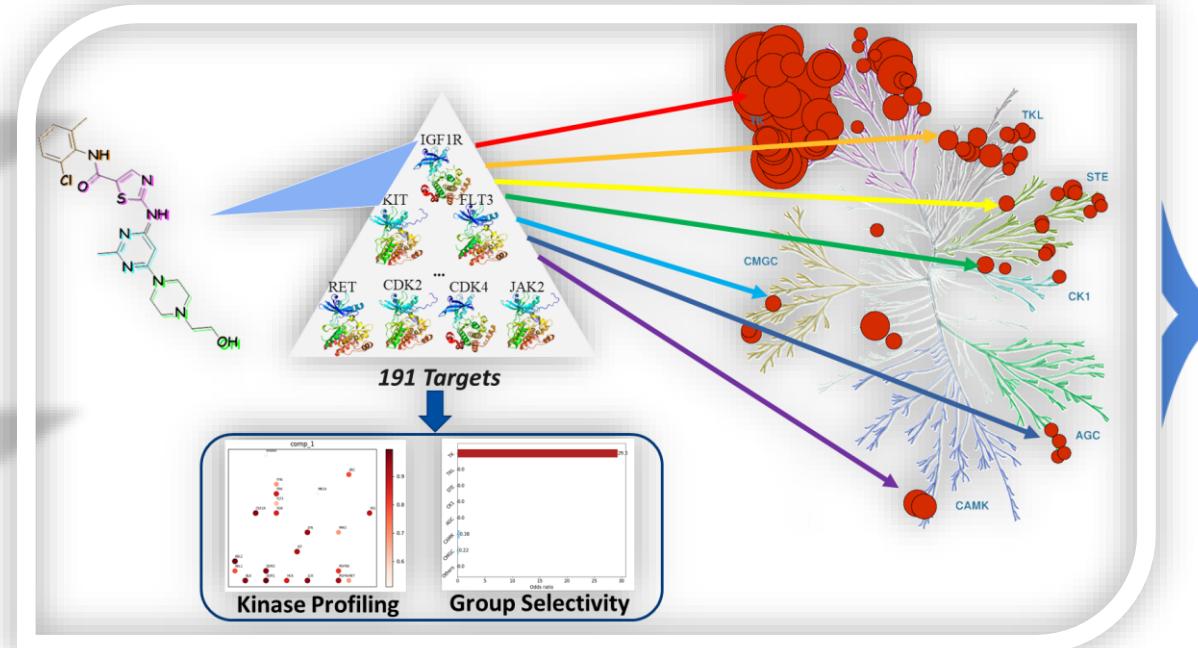
快速预测小分子化合物的激酶谱性质，帮助判断小分子的激酶选择性

#### 高通量和准确

可支持超大小分子化合物库的快速筛选，内部化合物测试较高的准确性

#### 易用性和超低的检测成本

上传化合物的smiles或结构式即可提交计算，模型计算预测成本几乎为零



#### 快速高效

#### 湿实验验证

#### 降低选择性筛选成本

对模型进行应用性测试并得到选择性较好的小分子化合物

# 化合物投票一致性筛选平台

模块功能: 化合物投票一致性筛选方法，提升虚拟筛选的命中率

## 模块特点

独创的投票一致性筛选算法，提升虚拟筛选命中率，降低假阳性

### 准确可靠

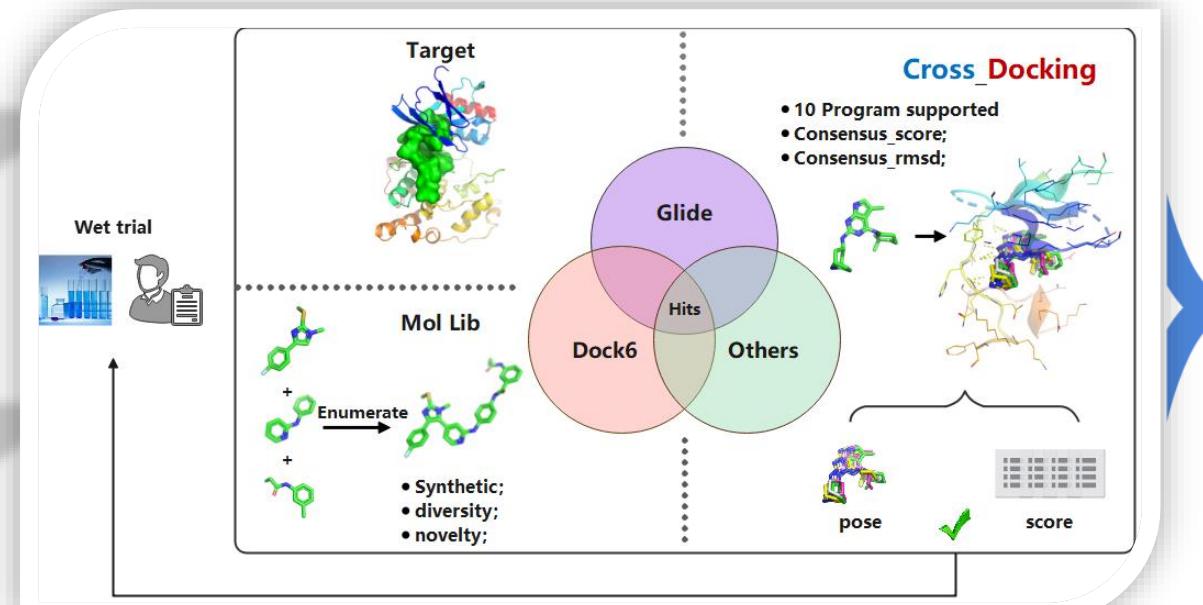
构建方法相比较于任一单一商业软件或开源软件的筛选效果有很大的提升

### 假阳性率低

筛选的平均命中率和富集效果是商业软件或开源软件的3.5倍

### 可拓展性与易用性

可灵活组合不同的筛选算法，提升筛选效率；非计算化学专业也可快速上手



### 科学可靠

假阳性率低

高命中率

对平台进行应用性测试并得到活性较好的小分子化合物

# Auto\_MD分子动力学模拟平台

模块功能:自动化一键式分子动力学模拟计算平台

## 模块特点

零成本高通量运行蛋白小分子复合物分子动力学模拟

### 零学习成本

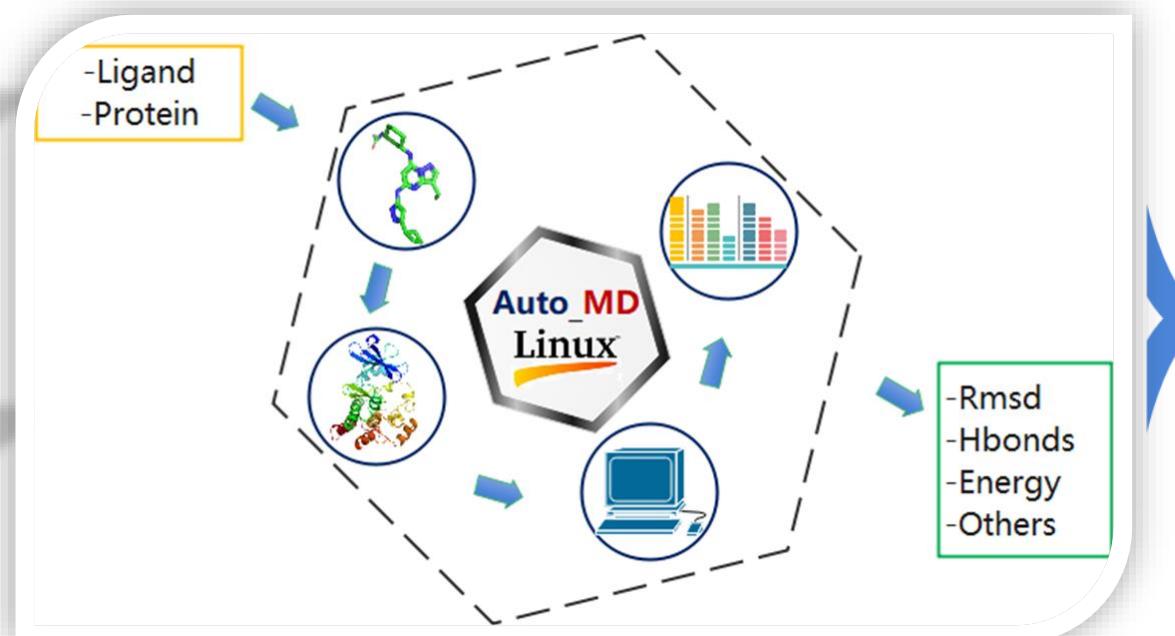
一键式提交MD任务，零门槛运行分子动力学模拟

### 高通量模拟支持

无需构建MD模拟参数文件，可满足高通量运行MD任务的需求

### 自动分析计算结果

网页端提交MD任务，自动呈现分析模拟结果



效率提升超3600倍

一键提交复杂MD计算

支持每周>100个体系的模拟

对模块进行应用性测试并得到活性和选择性较好的小分子化合物

# 机器学习超高通量筛选平台 (ML\_Screening)

模块功能:结合机器学习算法加速超大规模虚拟筛选

## 模块特点

采用主动学习的方式加速超大规模化合物库的虚拟筛选进程

### 加速筛选

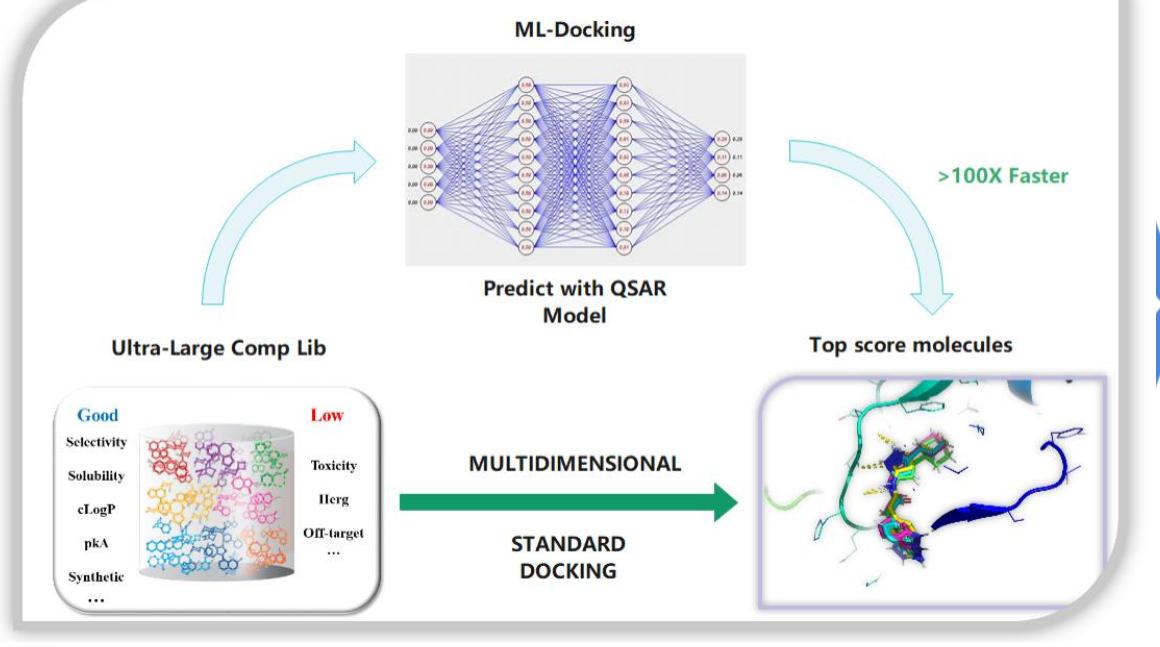
相比较与传统虚拟筛选方式速度提升超120倍以上

### 超高通量虚拟筛选支持

可满足**十亿至百亿级**化合物库的快速虚拟筛选

### 提升虚拟筛选命中率

超高通量筛选提升筛选的命中率及苗头化合物的新颖性



筛选速度提升超120倍

超高通量

低计算成本

对模块进行应用性测试并得到活性较好的小分子化合物

# 模型应用验证—CDK项目

- 针对指定靶标筛选潜在活性选择性分子；

Compounds\Enzyme Ki (nM)	CDK-2/CycE1	CDK-4/CycD3	CDK-6/CycD3	CDKb/c Selectivity
PB46-004-01	0.1	0.9	1	—
PB46-002-01	14	0.9	8	—
Molecule-10	>10000	-	-	—
Molecule-11	193	>10000	>10000	>51
Molecule-17	806	535	339	—
Molecule-18	321	>10000	>10000	>31
Molecule-20	827	>10000	>10000	>12
Molecule-21	1752	>10000	>10000	>5.7
Molecule-23	2001	>10000	>10000	>5.0
Molecule-25	4057	>10000	>10000	—
Molecule-73	3.7	56.2	—	—

1uM命中率： 50. 00%

2uM命中率： 62. 50%

5uM命中率： 87. 50%

# AI视觉识别高通量酶筛选项目

项目目标: 提升高通量生物酶筛选实验检测的效率和通量，同时降低实验检测的成本

## 解决思路

基于视觉识别算法和数据分析技术，将酶反应颜色差异转换为像素信号并快速排序

### 快速准确颜色信号识别

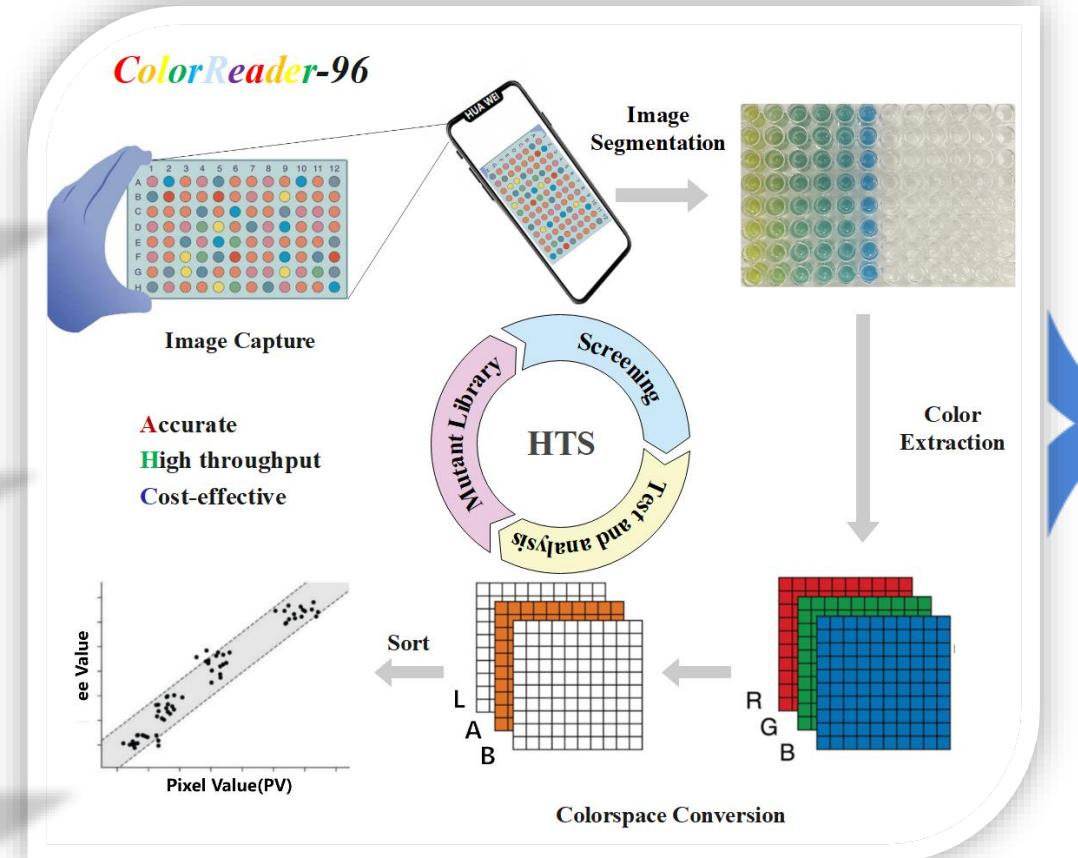
快速准确读取到96孔板每孔的准确颜色像素值，并给出top孔编号

### 支持超高通量的应用需求

可拓展性强，轻松实现超高通量酶筛选检测需求

### 便捷性和超低的检测成本

零学习成本，方便高效的指导湿实验；相比GC实验检测方法，模型计算预测成本几乎为零



效率提升/成本降低

5400倍

94.79%

可结合自动化实验平台实现自动监控酶反应进程

# 集团贵金属闭环管理的数字化平台建设

项目目标:建立数字化平台以监控贵金属闭环流转数据，提升库存、回收和生产管理效率

## 解决思路

理清涉及贵金属流转的各个关键环节和所需数据，建立闭环管理数字化监控平台

### 即时贵金属分布可视化

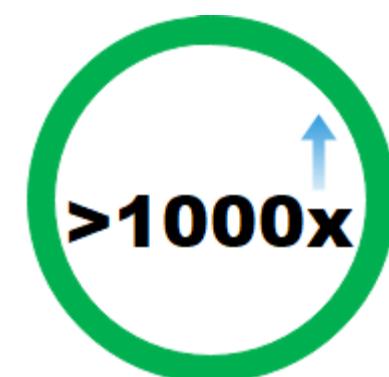
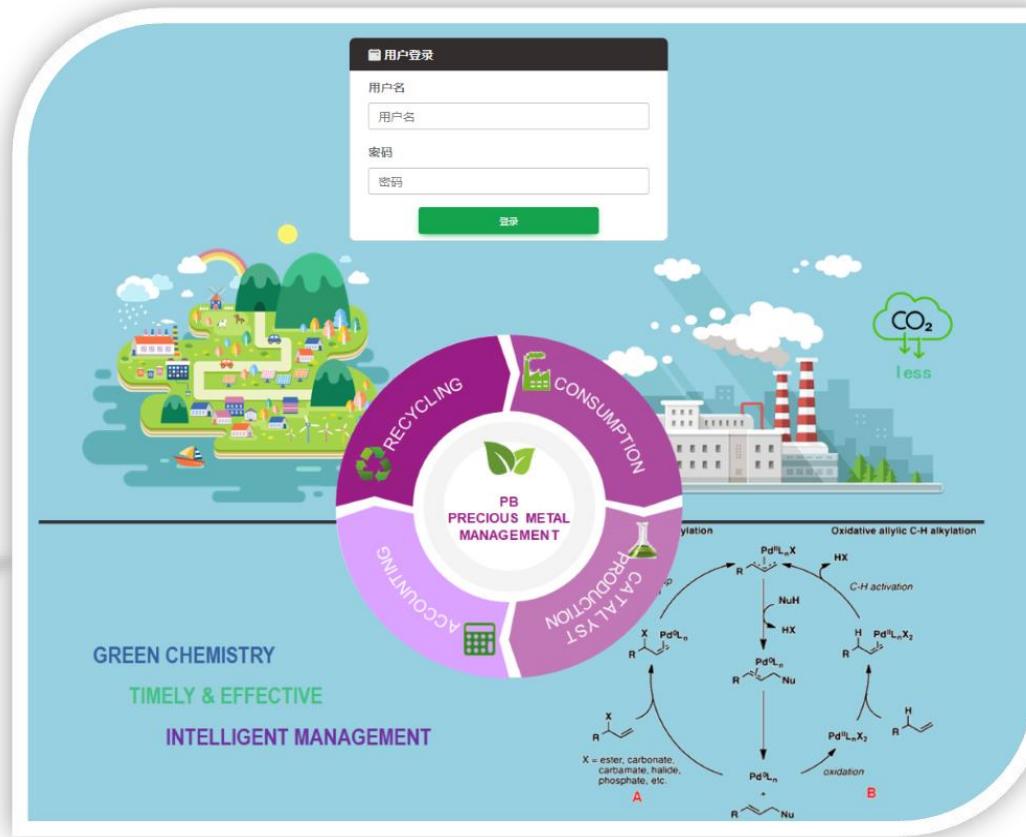
可即时展示贵金属催化剂的回收、生产、领用、库存变动状态

### 呆滞风险预警分析

对催化剂库存和领用进行风险等级划分，实时掌握呆滞废催化剂和呆滞库存的详细信息

### 贵金属流转状态追踪

对库存和领用催化剂状态进行详细信息监控



数据统计效率  
提升超1000倍

及时掌握贵金属流转状态；指导生产规划制定

可拓展至其他物料/供应链的闭环管理的数字化平台建设项目中

# 基于大模型的项目报告评分和翻译平台

平台介绍:实现本地化的基于AI大模型的项目报告质量评分及专业化翻译

## 业务专业性

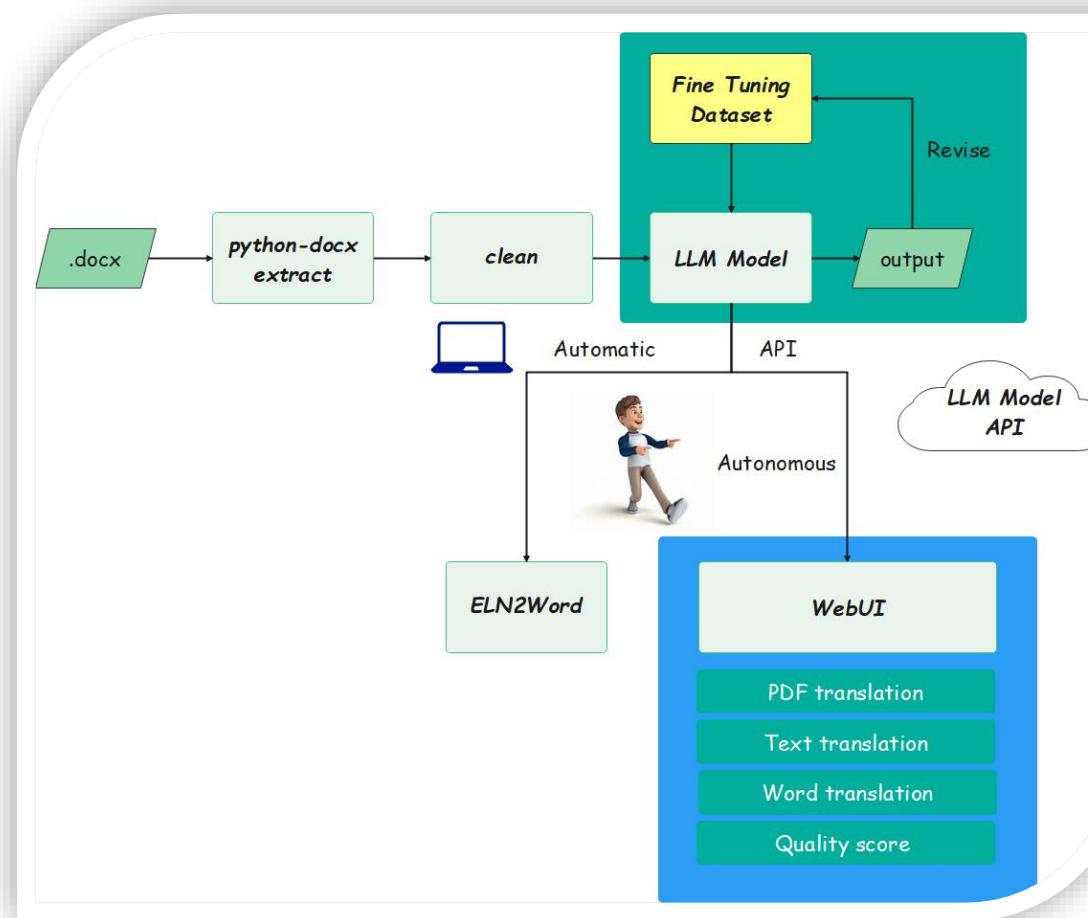
通过对大模型的微调，实现化学专业性和针对性的中英互译，同时对不符合业务场景的表述进行评分及纠正

## 用户友好

实现网页端功能部署，多格式支持(PDF/Word/Text)，用户友好的可视化界面

## 优秀的排版保留

翻译后的文件完整保留原始文件的排版、公式、图表、目录等细节

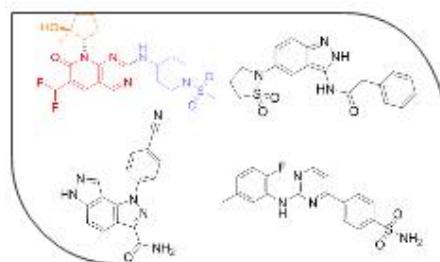


- 效率提升
- 超低成本
- 语境专业性

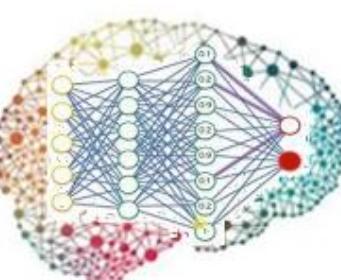
# Fragment-Based Library Generation

- Starting from the known molecular chemical space;
- From the unique 140000 compound fragment space of PB;
- Through precise filtering by AI models;

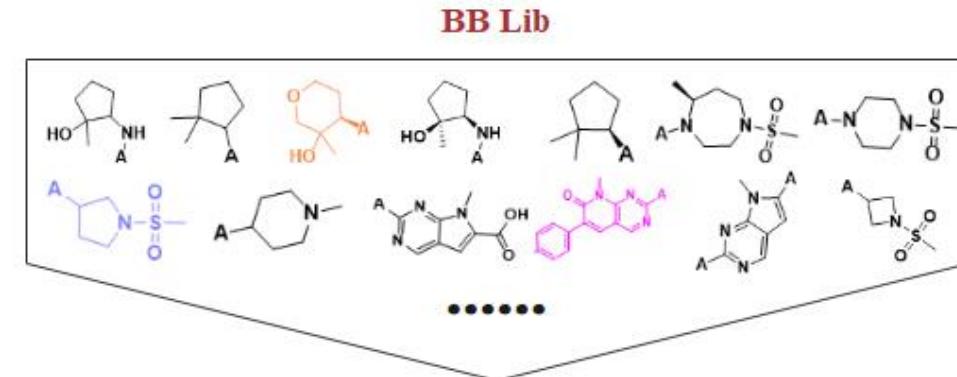
## Actives Molecule Mapping Algorithm (Fragment-Based Molecular Generation)



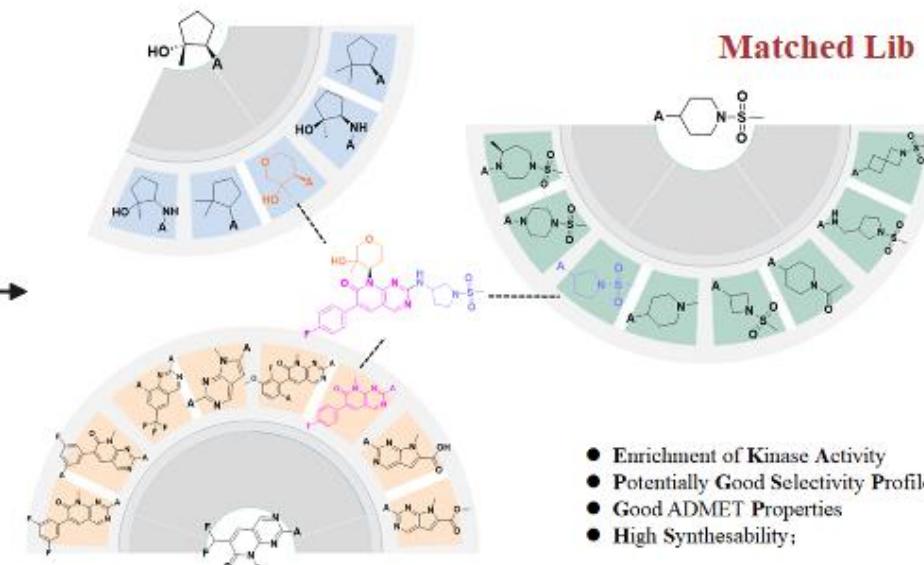
Actives



Mapping Algorithm



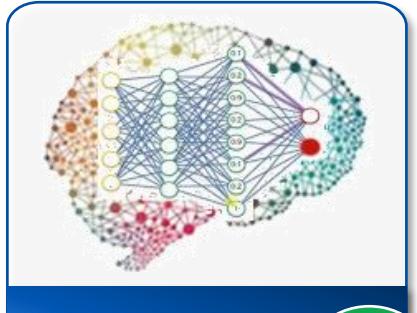
BB Lib



Matched Lib

- Enrichment of Kinase Activity
- Potentially Good Selectivity Profile;
- Good ADMET Properties
- High Synthesability;

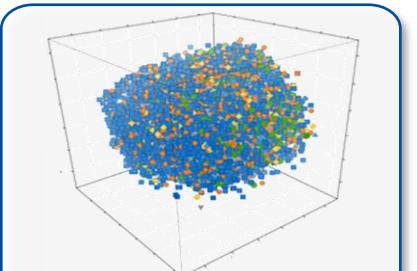
# Characteristics of the compound library



## AI-Driven

1

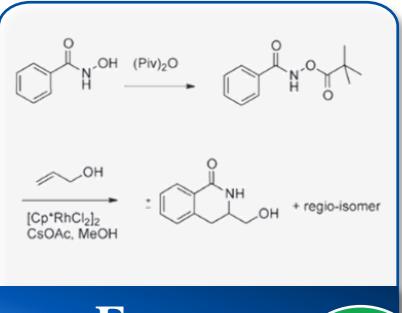
- Abundant AI algorithm model
- In-house data-driven



## Rich Diversity

5

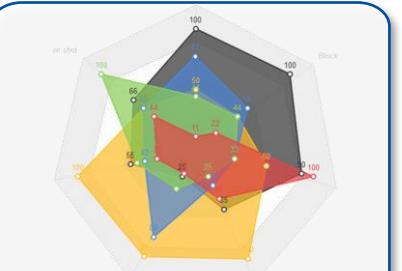
- 100K+ BBs, Diversity scaffolds
- Coverage of ultra-large chemical space(  $10^{14}$  )



## Easy Synthesis

2

- Enumeration based on reaction pathway templates
- 80%+ accessibility



## Excellent Properties

6

- Good kinase selectivity
- With excellent ADMET properties



## Customizable

3

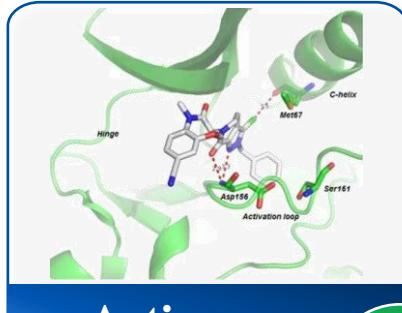
- Meeting different project requirements
- Flexible customization of libraries



## Fast Delivery

7

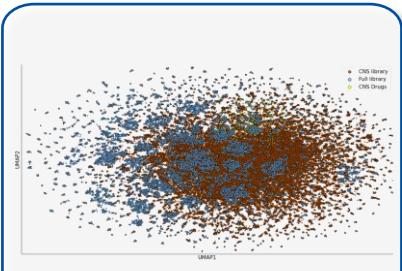
- Seamless connection between dry and wet test
- Top chemical synthesis



## Active Enrichment

4

- Spatial mapping of the active fragments
- Multiple screening + expert experience selection



## Extensible

8

- Lightweight with sufficient diversity for screening
- Expandable potential space

# Kinase activity-enriched library- 758110 molecules

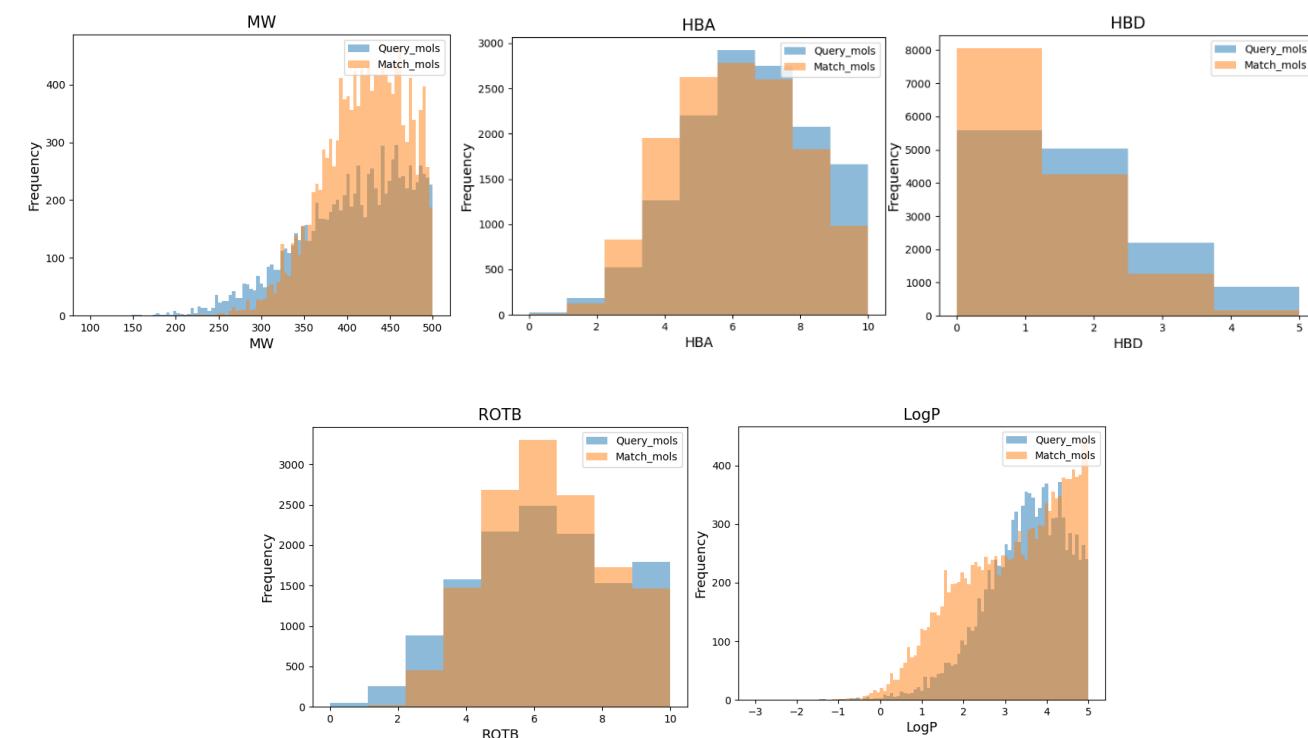
Catalog No.K0001

## 3. Analysis of drug-like parameters:

### Compounds compliant with Lipinski's Rules

PhysChem Properties	% Proportion
HBond donors <5	100
HBond acceptors <10	100
cLogP <5	100
MW <500	100
ROTB <10	100
ALERTS Group = 0	100

### Comparison of the RO5 distributions

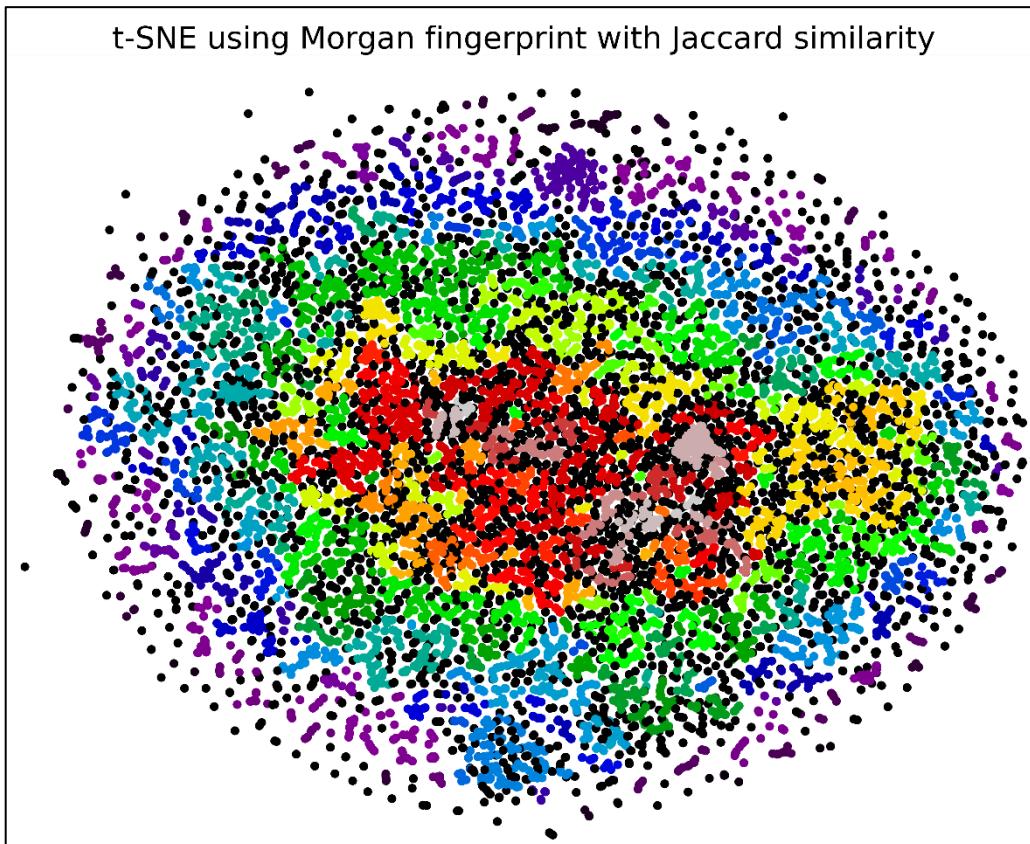


# Kinase activity-enriched library- 758110 molecules

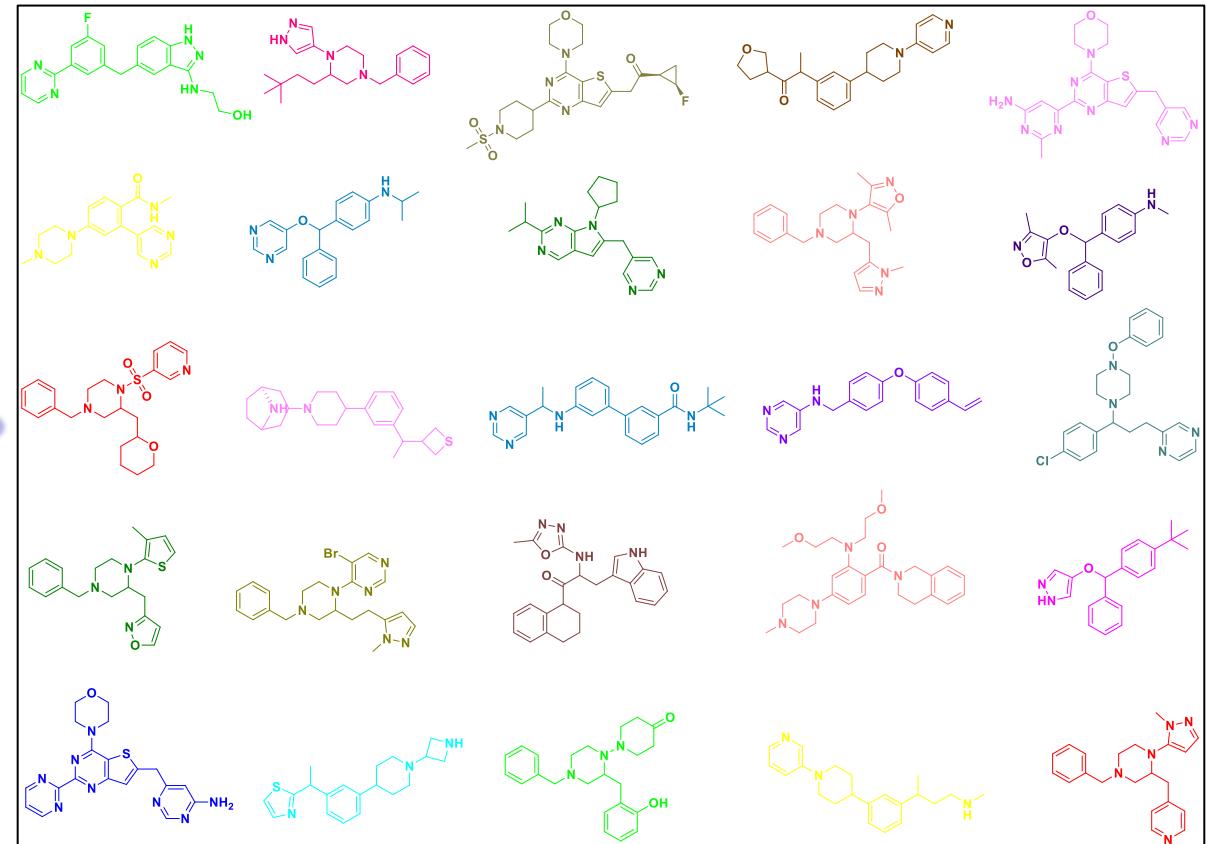
Catalog No.K0001

## 5. Structure sampling

### Compound library tsne dimensionality reduction



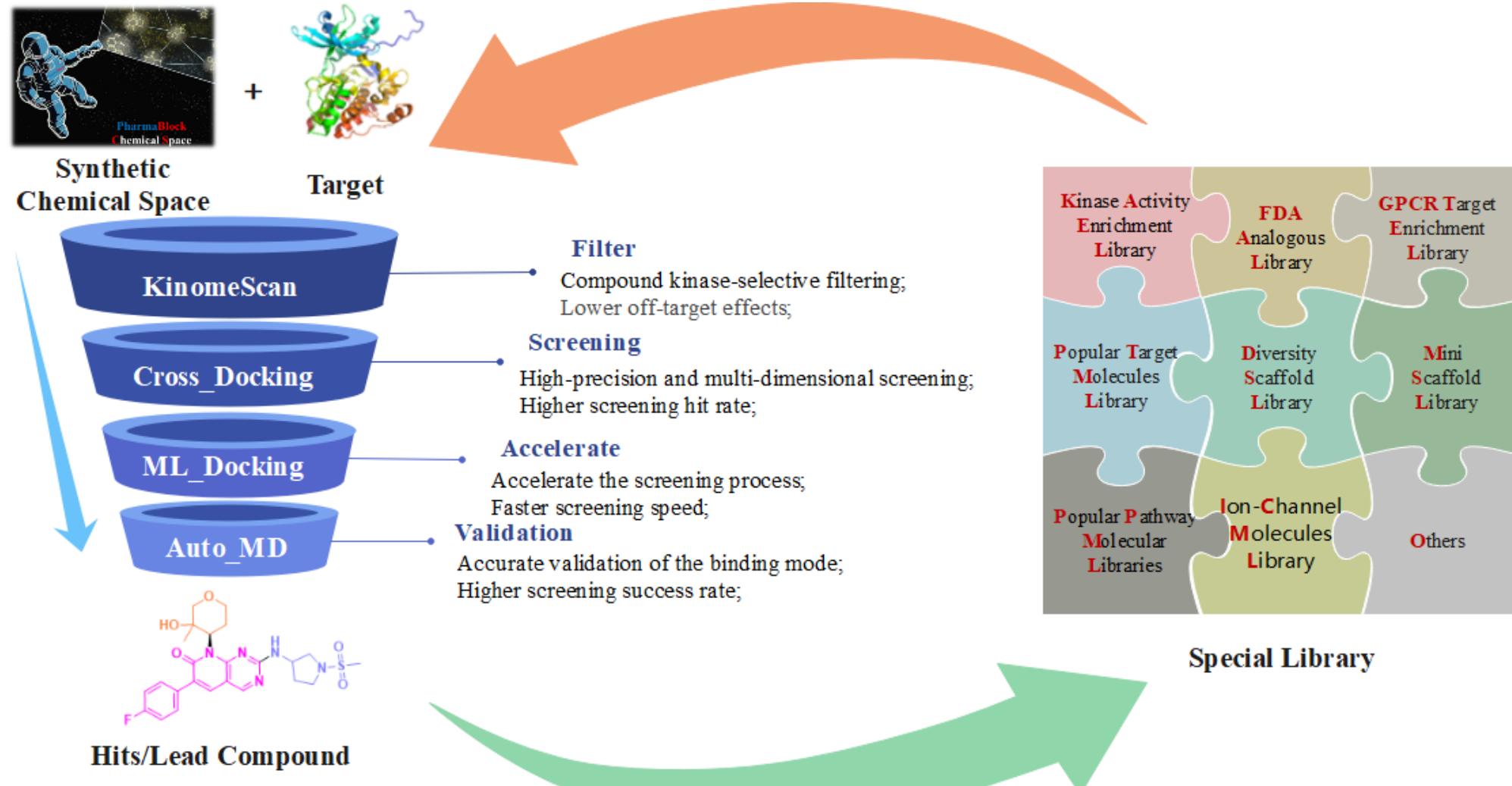
### Compound library structure sampling



# 特色子库构建

库类型	片段可及性	可合成性	状态	数据量
激酶活性富集库 (v1)	较高	较高	Done	75w
类FDA化合物库 (v1)	较高	较高	Done	116w
多样性骨架库 (v1)	较高	较高	Done	317w
高频片段重组库 (v1)	较高	较高	Done	306w
商业库片段源类FDA子库 (v2)	较高	较高	Done	144w
激酶片段重组库 (v3)	较高	较高	Done	146w
FDA片段重组库 (v3)	很高	很高	Done	103w

# Process of Lead compound screening



Seamless connection between sub library production and screening processes

# 参与的文章和专利

**DL-DockVS (共一) : Exploration on learning molecular docking with deep learning models**

—*Journal of Quantitative Biology*(2023/09)

**KinomePro-DL (一作) : The Development and Application of KinomePro-DL: A Deep Learning Based Online Small Molecule Kinome Selectivity Profiling Prediction Platform (KinomePro-DL预测平台)**

—*JCIM*(2024/09)

**专利 (参与) : 一种分子设计和结构优化方法、系统、装置及存储介质** — 2023/12

**Cross\_Docking(专利) : 多个程序投票一致性筛选算法, 一致性筛选分子并评价结合构象合理性;**

— *Preparing for submition*

# Thanks

Wei Ma

2025/05/