

基于AIGC和大模型打造分子设计平台

碳硅智慧 – 生命科学领域基础设施和服务提供商

我们将最先进的生命科学技术与人工智能等信息科学技术深度融合，面向新药研发，利用深度生成（AIGC）、预训练大模型、强化学习等人工智能先进技术，并深度融合物理计算及硬件自动化技术，通过提高新药研发领域生产数据、管理数据以及对数据进行AI建模的能力，将新药研发的各个环节数字化和智能化，形成干湿试验数据闭环，解决新药研发难题。

本资料版权归杭州碳硅智慧科技发展有限公司所有



20年经验AI专家+20年经验药学专家联合创立



邓亚峰
创始人&CEO

20年以上人工智能研发管理经验

- 曾任某互联网公司副总裁，人工智能研究院院长&搜索事业部总经理，带领400人团队
- 曾任格灵深瞳CTO
- 北京人工智能产业联盟副理事长
- 2021年中国人工智能年度十大风云人物
- 累计发明专利140余项（授权98项）



侯廷军
创始人&首席科学家

20年以上药物设计方法学和应用研究经验

- 浙江大学药学院求是特聘教授
- 全球学者学术影响力排行榜中国药学领域领军人物TOP3
- 国家万人计划科技创新领军人才
- 累计发表 SCI 学术论文400余篇，总引次数 (google)>2.5万次，H因子78

兼具前瞻性视角及实战经验的复合团队



人工智能+药学+化学+量子物理



谢昌谕 算法科学家

量子物理+AI 知名专家

- 浙江大学求是工程教授
- 前腾讯量子实验室理论组负责人(T12级研究员)
- 曾在麻省理工学院、多伦多大学、新加坡-麻省理工科研中心从事理论化学、量子计算、AIDD、生物传感器等方向研究。



宋剑飞 算法工程副总裁

14年人工智能和互联网领域研发经验

- 曾任微诊集智CTO、创新奇智算法负责人
- 在人脸识别、多模态预训练、知识图谱等领域曾取得多项世界第一。



潘培辰 癌症生物学专家

药物筛选/药理学/药物化学/癌症生物学/计算生物学

- 浙江大学 博士生导师;
- 发现了多种具有全新结构的抗肿瘤和抗炎先导化合物;
- 累计发表SCI论文40余篇, SCI引用超过1800余次, H因子为22; 获授权国家发明专利14项。



邹武新 副总裁 药物化学专家

超过20年药物化学工作经验

- 南开大学有机化学专业博士毕业; 20年+药化以及化学合成经验;
- 国内头部CRO公司工作十多年, 为国内外多家公司提供药化设计和化学合成服务;
- 负责和参与过的管线有多个项目推进到临床前候选化合物阶段和临床阶段; 具有丰富的项目推进和管理经验。



资深运营管理及销售



施慧 COO

多年医疗及互联网领域管理经验

- 前阿里健康互联网医疗总经理、好大夫副总裁、拜耳医药高级产品经理等;
- 中国生物技术协会生物医学信息分会委员;
- 中国医促会药学信息化分会委员;
- 中国药师协会患者教育分会委员。



何浩 商务副总裁

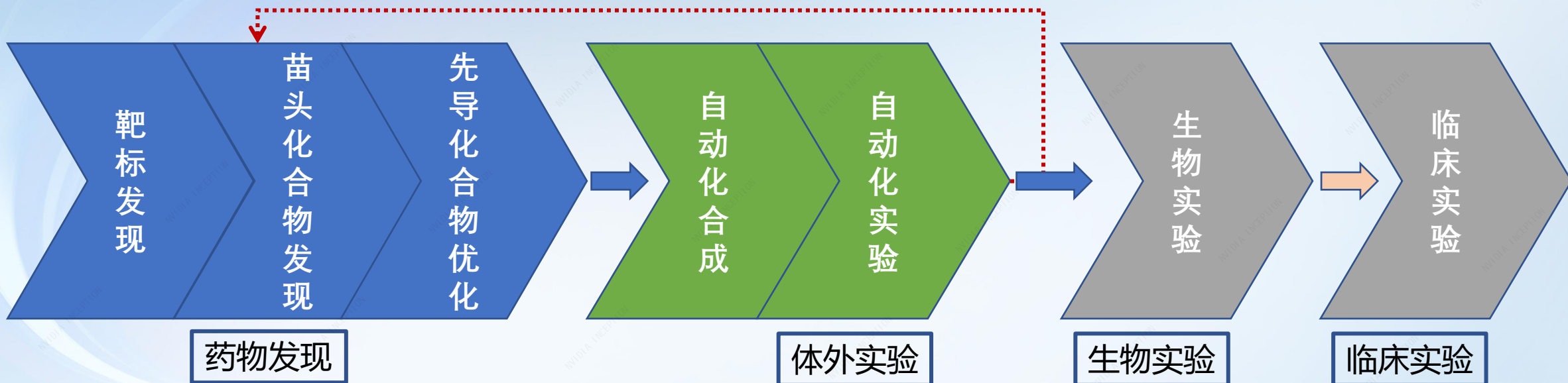
二十年以上医药软件领域商业经验

- 曾任Oracle、Veeva、百奥知、医渡云等国内外知名企业生命科学软件销售负责人;
- 20年以上制药研发、临床试验研究及医疗卫生领域的销售和市场经验。

碳硅智慧：攻克AIDD核心技术 变革药物研发范式

通过AI建模、物理计算和自动化流程，赋能专家，提高药物研发成功率及效率，

推动药物研发范式向AI建模+自动化试验+专家联合驱动转变

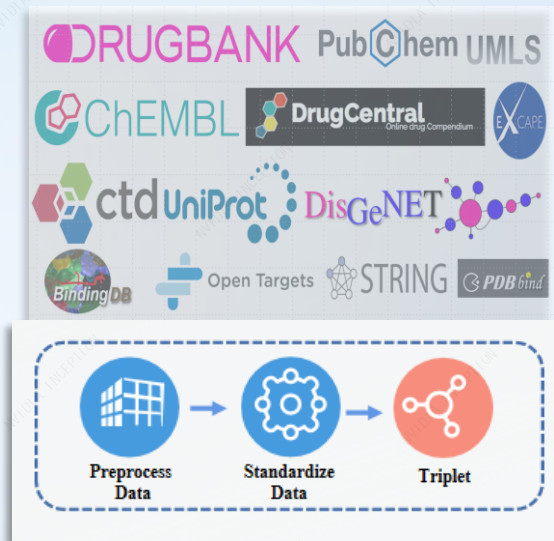
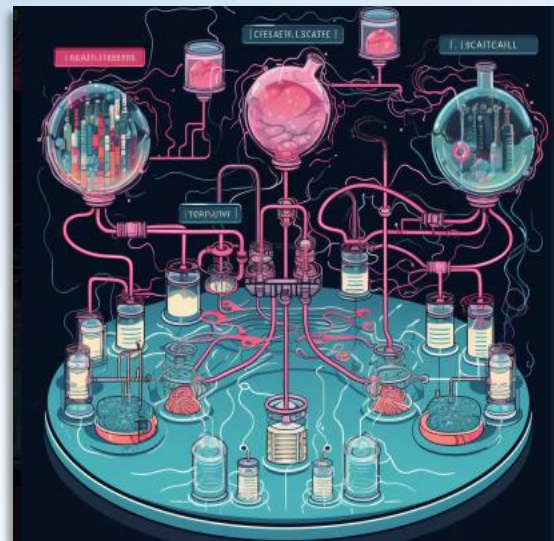
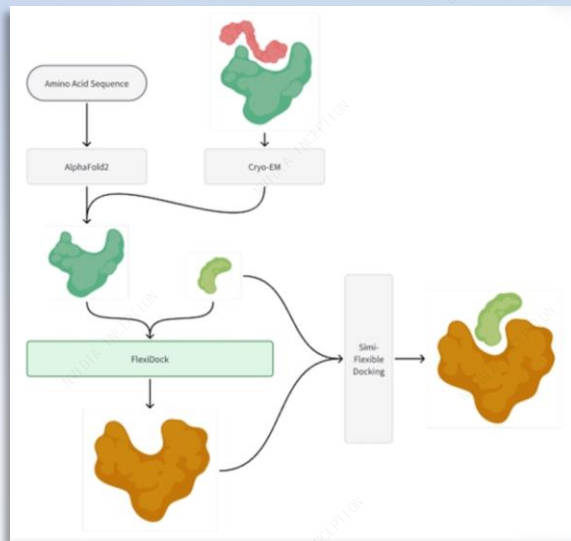
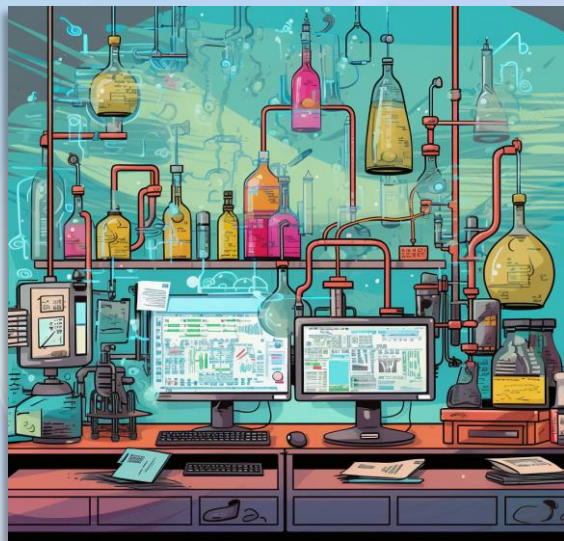


- 痛点一：苗头化合物发现依赖于通用百万级实体化合物库，新颖性不足，活性不足；
- 方案：研发的AI对接算法比传统对接算法快千倍，将筛选化合物库由百万扩展到亿级别；全新分子生成可产生针对该靶点的专有分子库，显著提高新颖性和活性；预测复合物结合构象，将蛋白质结构预测误差降低2-3倍，提高FIC成功率；

- 痛点二：先导化合物优化依赖于专家，结构同质化，且无法同时考虑多种因素，迭代慢，失败率高；
- 方案：研发基于大模型的多目标约束分子生成优化算法，进而结合业界领先的对接/重打分、成药性预测模型，打造分子工厂流程，可极大开拓分子化学空间，最终提高BIC项目的迭代效率和成功率；

- 痛点三：作为药物研发失败的重要原因，成药性预测不准，导致投入成本高且失败比例高；
- 方案：十年积累行业最全的成药性数据，并构建基于Transformer的预训练模型，打造预测终点最多、准确率最高的成药性预测模型；

- 痛点四：靶点同质化严重，靶点-疾病关系依赖于专家经验，三期临床失败率高；
- 方案：基于组学及临床数据，构建靶点、疾病、基因多模态模型，建模靶点-疾病关系，找到独特的成熟靶点，并为靶点推荐适应症；



分子工厂: 基于已有分子结构, 利用AIGC技术, 并结合自动化筛选和强化学习反馈, 优化设计新颖分子, 显著提高BIC项目的成功率和开发速度;

超高通量筛选: 拥有超高通量筛选能力 (单机亿级底库), 并结合全新分子生成和全柔性蛋白-分子结构预测, 可快速助力FIC项目成功;

ADMET成药性预测: 显著提高成药性预测准确率, 可降低药物研发后期的失败率;

靶点-疾病多模态大模型: 构建多模态模型, 发掘成熟靶点对应的新的适应症, 或者针对适应症推荐靶点, 以及预测靶点联用效果;

干湿试验闭环: 合作打造分子生成-自动化试验-类器官闭环;

基于AIGC和预训练大模型技术打造世界领先模型

领域	子任务	模型名称	领先性	核心技术	论文发表
活性	对接构象预测	KarmaDock	<u>相对传统对接，速度提高数千倍 (17ms vs 130s)</u>	基于等变神经网络和混合密度网络的构象预测和打分方法	<i>Nature Computational Science</i> , 2023
	对接构象预测	CarsiDock	精度相对传统方法由64.56%提高到92.28%	基于物理模拟数据预训练+大模型	投稿到 <i>Cell Systems</i> , 2023
	对接构象预测	FlexiDock	<u>考虑蛋白柔性，误差由2.32埃降低到0.77埃</u>	基于张量直积网络的蛋白质小分子全柔性结构预测	待发表
	重打分函数	RTMScore	虚拟筛选TOP1识别率由0.368提高到0.667	基于Transformer+混合密度网络的重打分预测	<i>J Med Chem</i> , 2022
	相对结合自由能预测	MSLD	<u>精度相近的情况下，速度相对FEP+提高10+倍</u>	Multisite λ -Dynamics	投稿到 <i>JACS Au</i> 2023
成药性	ADMET预测	MERT	<u>相对SOTA方法，分类准确率提高4%，回归任务RMSE降低0.2，支持70+种成药性终点</u>	基于多种预训练方法的Transformer模型	待发表
	ADMET可解释性	SME	支持成药性可解释，建模成药性性质相关的基团	基于GNN的可解释性模型	<i>SME: Nat Commun</i> . 2023 <i>MGA: Nucleic Acids Res</i>
分子生成优化	小分子生成	FragGPT	同时支持全新、RGroup、Linker、侧链、骨架跃迁全部五种任务	<u>基于片段分割将分子转换为1D信号，构建支持Prompt的GPT语言预训练大模型</u>	待发表
	小分子生成	ResGen	基于蛋白结构的全新分子生成，可预测键的类型	基于多尺度等变网络的蛋白口袋3D分子生成技术	<i>Nature Machine Intelligence</i> , 2023
	小分子生成	SurfGen	基于蛋白结构生成，改进版本支持RGroup、Linker、侧链、骨架跃迁4种任务	基于测地线距离表示的3D分子生成	<i>Nature Computational Science</i> , 2023
	多肽生成	AMP-Designer	多肽生成模型，生成成功率很高（生成10条多肽，有多条有活性）	基于Prefix Learning的GPT模型	投稿到 <i>Nature BioTechnology</i>
	AAV生成	AAVDiffusion	生成成功率达到92%+，显著高于传统方法	基于Diffusion模型的AAV序列生成	待发表
靶点-疾病关系	联合用药	KG-CombPred	联合用药的准确率提高38%	多模态+知识图谱	<i>Cell Reports Physical Science</i> , 2023

首个具有自主知识产权并供业界使用、物理与AI融合的AIDD平台 DrugFlow



小分子蛋白预训练
(粗排+精排)



独有数据
多任务预训练



预训练
AutoML



扩散模型
强化学习



- ✓ 世界领先的打分方法
- ✓ 世界领先的AI对接算法
- ✓ 十亿+底库高通量筛选
- ✓ 性能优化经典物理模拟方法

活性预测

成药性预测



- ✓ 17种物化、5种药化、21种成药性参数、27种毒性终点的系统性评价
- ✓ 速度、准确率业界领先
- ✓ 支持子结构可解释功能
- ✓ 频繁命中化合物和警示子结构预测



- ✓ 基于活性配体的分子生成
- ✓ 基于口袋结构的分子生成
- ✓ 同时支持全新分子、RGroup、Linker生成、骨架跃迁等模式

分子生成

AI建模



- ✓ 多种分子指纹表征
- ✓ 基于Auto ML的多种算法
- ✓ GNN、Transformer等深度学习方法
- ✓ 预训练模型 (小数据友好)

数千家注册用户，数百家重点测试用户，数十家签约客户

活性预测



活性预测

Inno-Docking

CarsiDock KarmaDock

AutoDock Vina

Inno-Rescoring

RTMScore

成药性预测



成药性预测

Inno-ADMET

ChemFH

Inno-SA

分子生成



从头生成

片段生长/分子优化

R-Group

Side-Chain

Linker-based

Scaffold-Hopping

AI建模



AI建模

Inno-QSAR

Inno-QSAR Application

应用场景



分子工厂：AI生成优化+筛选

虚拟筛选：超高通量筛选

其他工具



AI解谱

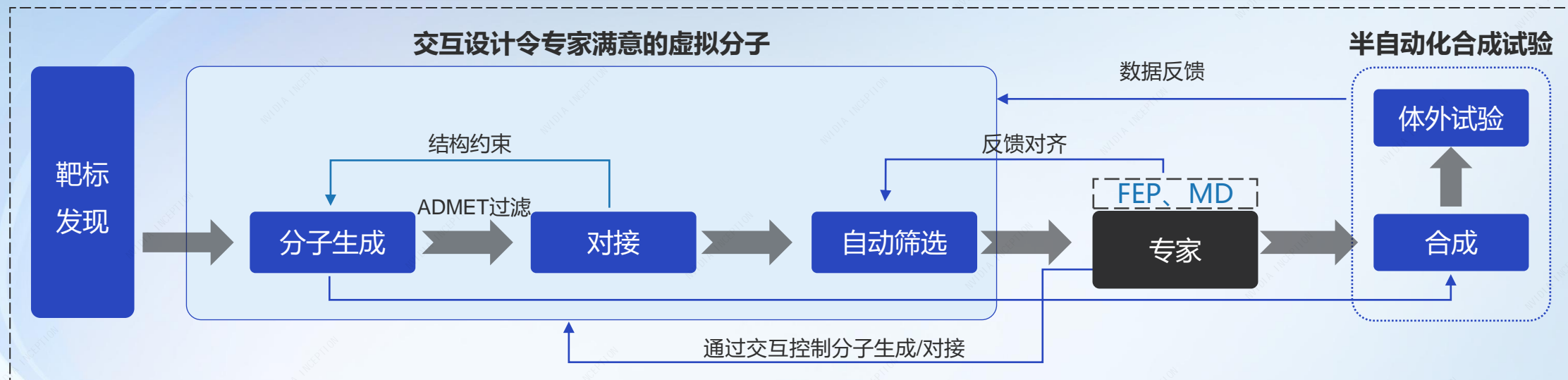
Inno-CReSS

Inno-CMGNet

配体预处理

蛋白预处理

DrugFlow分子设计流程 - 可降低50%早期研发时间, 降低70%早期研发成本



覆盖功能最全的分子生成模型

- ✓ 支持De novo、RGroup生成、Linker生成、骨架跃迁等多种分子生成方式
- ✓ 同时支持活性配体及口袋结构的生成
- ✓ 支持ADMET性质预测控制分子生成
- ✓ 基于点击化学和片段分割生成合成可及分子

最强复合物结构预测模型

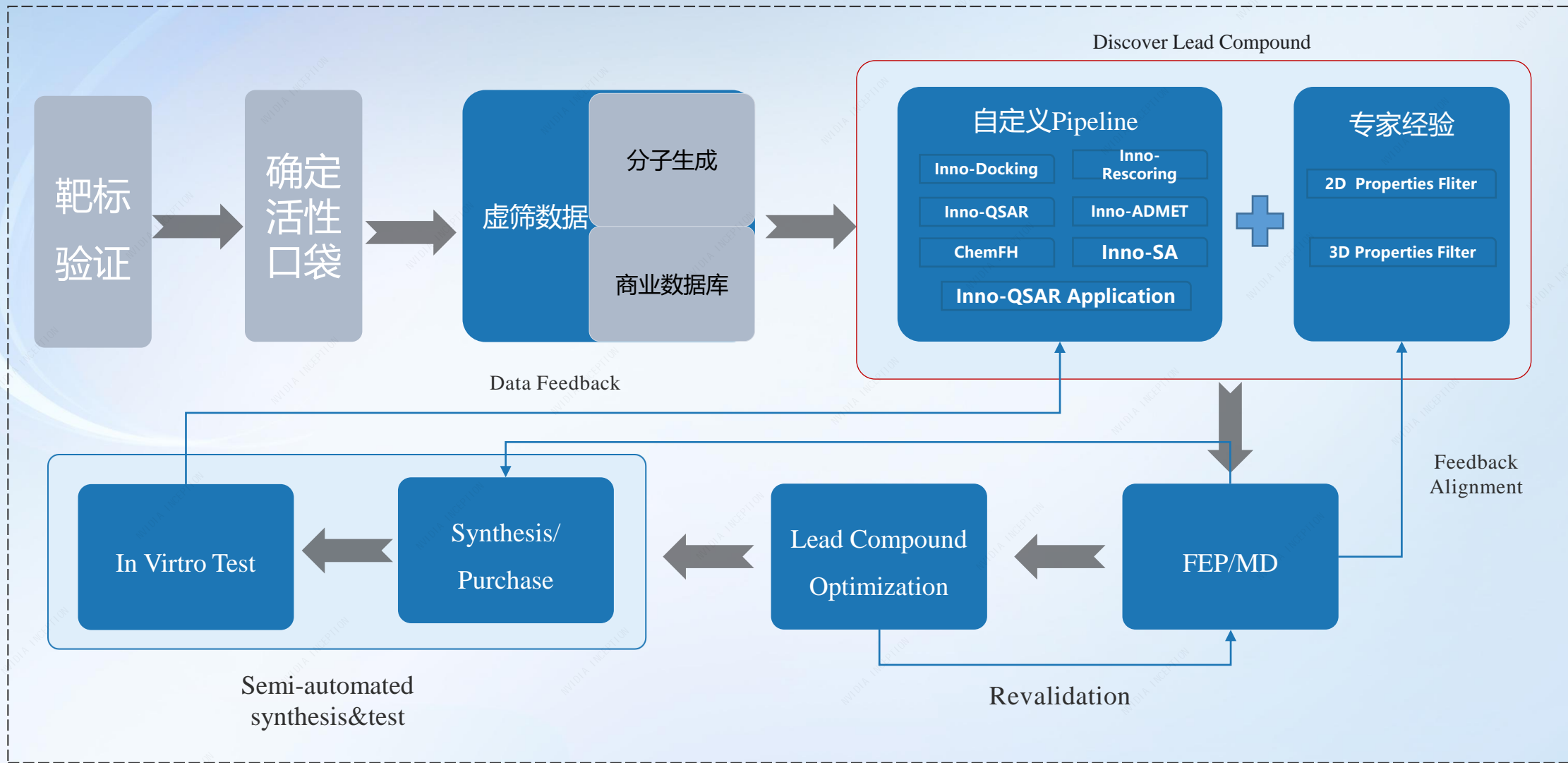
- ✓ 利用物理计算产生的模拟数据，帮助约束AI模型，快速精确预测小分子和蛋白结合构象
- ✓ 相对物理对接（半柔性对接需数分钟，全柔性需数小时），AI对接仅需要10多毫秒至数秒，且精度更高，未来可支持全柔性对接

利用强化学习技术与专家规则对齐

- ✓ 通过建模专家规则，保证筛选结果符合专家习惯
- ✓ 类似Chat GPT，专家可对过滤分子进行反馈，将筛选结果与专家做对齐 (RLHF)

基于AI Docking的超高速高通量筛选 (亿级底库单机天级完成)

高通量虚拟筛选, 单机亿级别底库2天完成筛选



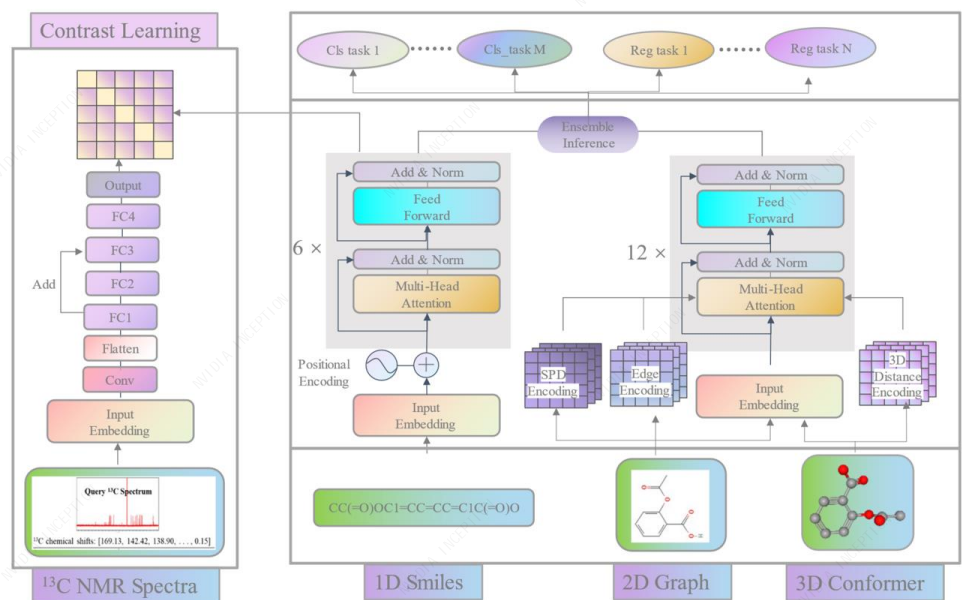
碳硅智慧分子生成模型



产品功能

Inno-ADMET主要用于系统评价药物成药性和毒性，基于收集的高质量ADMET实验数据，**可实现17种物化性质、11种药化性质(包括6种类药性性质)、21种成药性参数、27种毒性性质**的系统性评价，是目前国际上最为优秀的、计算速度最快的计算平台之一。

基于预训练的成药性预测技术



评测结果

- ✓ 从模型评估指标上来看，Inno-ADMET (MERT) **比其他平台准确率更高**；
- ✓ 从结果**可解释性**的角度来看，Inno-ADMET (MGA) 更具优势；

Metrics		Inno-ADMET (MERT)	Inno-ADMET (MGA)	SwissADME	admetSAR 2.0	pkCSM	vNN-ADME
Classification Model	AUC	0.902	0.863	0.877	0.848	0.841	0.85
	ACC	—	0.822	0.805	0.835	—	—
Regression Model	R2	0.828	0.763	—	0.705	0.823	0.79
	RMSE	0.5472	0.76	—	0.73	—	—
可解释性		×	✓	×	×	×	×

目前国际上最优秀的、计算速度最快的计算平台之一

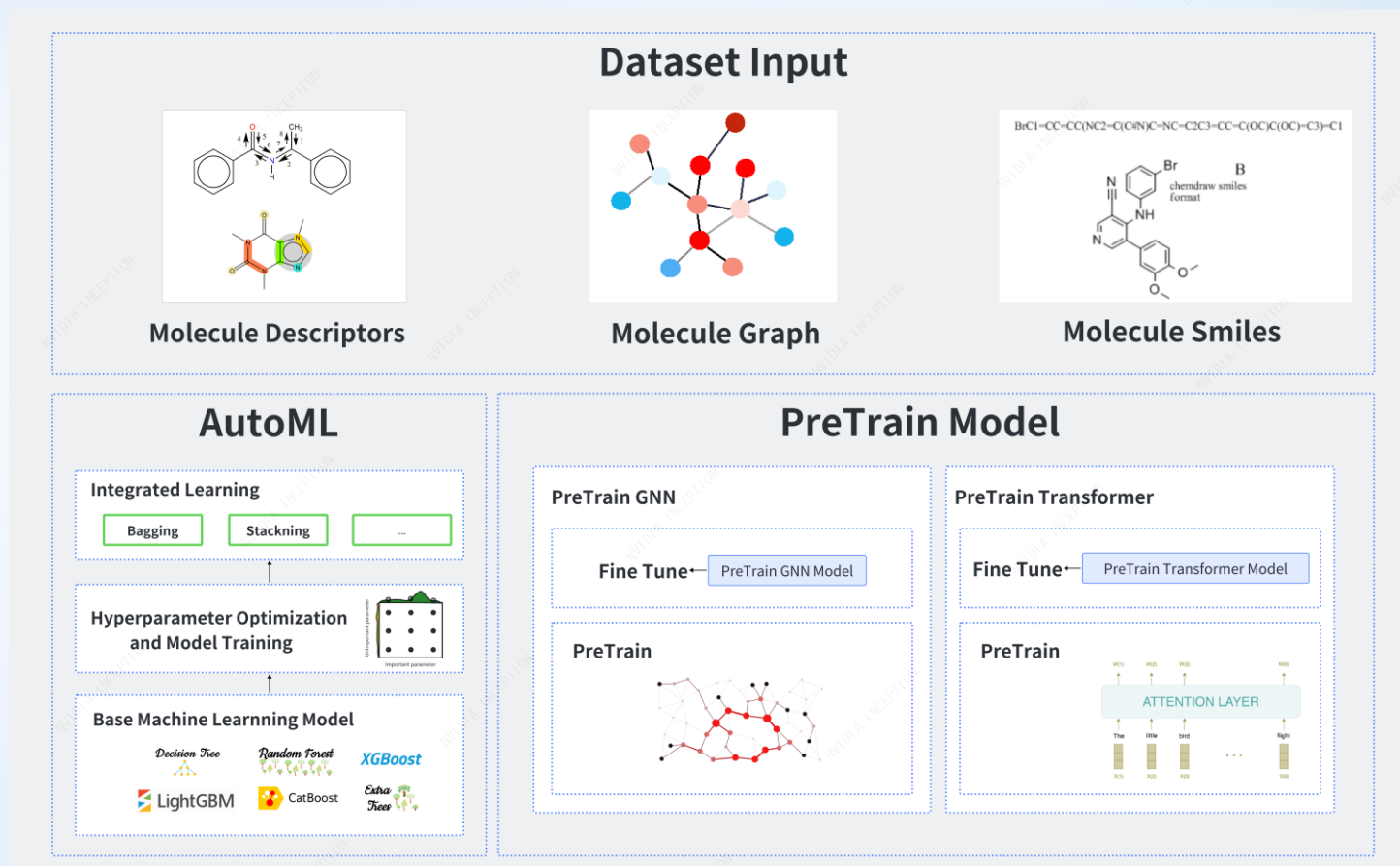
数据划分

数据来源于阿斯利康 Industry-scale application and evaluation of deep learning for drug target prediction 相关工作，原始数据来自ChEMBL和ExcapDB数据库。

从活性数据中进行抽取，分割如下：

- 分类数据由SMILES去重，从其中抽取 10000 条，训练集和测试集随机划分比例为 9:1 (9000:1000)，其中正负样本比例为1:1
- 回归数据集由 label 进行去重，抽取 8074 条，训练集和测试集随机划分，比例 9:1 (7266:808)

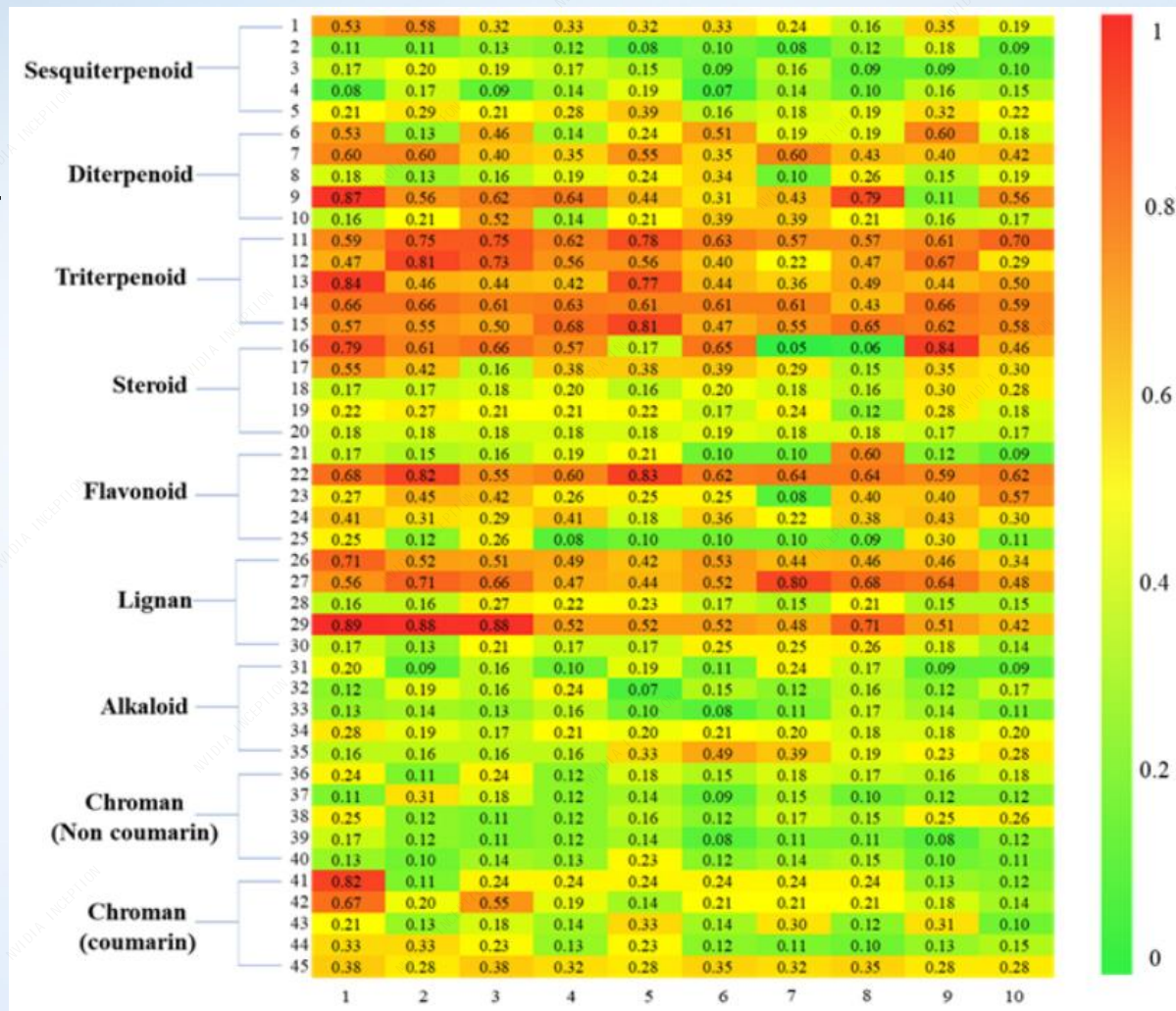
Model	AUC↑	F1↑	PRECISION↑	RECALL↑	ACCURACY↑	Training Time
PreTrain	0.9151	0.8276	0.8320	0.8284	0.828	~5 ~ 10 min
Non-PreTrain	0.855	0.764	0.7707	0.7655	0.796	~5 ~ 10 min
AutoML	0.9080	0.8288	0.8363	0.8214	0.829	~5 min (~10min)
XGBoost	0.8715	0.7943	0.8073	0.7817	0.796	~5 min (~10min)
RandomForest	0.8842	0.7897	0.8130	0.7678	0.794	~5 min (~10min)



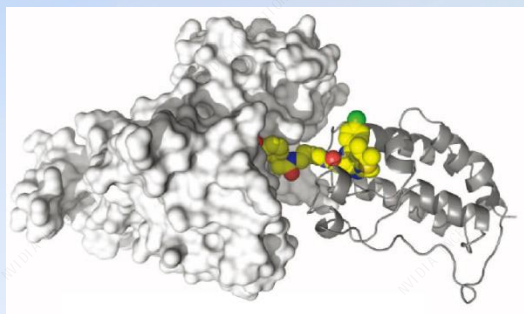
AI解谱-CReSS: 基于碳谱快速检索化合物分子结构

CReSS在外部测试集的碳谱结构检索任务中表现出卓越的性能。

- **速度快:** 对包含1040万个分子的参考结构库搜索了41,494条碳谱数据, 每条数据查询的处理速度仅为**0.114秒**。
- **准确率高:** top10召回率准确率达到**91.64%**。
- CReSS在三萜等多种类别的化合物上表现出优异的性能。



基于AI+物理模型发现先导化合物的成功案例



靶向ALK抑制剂

体外活性最强的Type-1¹/₂型ALK抑制剂，对多种ALK耐药突变体仍表现出很强的活性及高激酶选择性。

靶向AR同源二聚拮抗剂

靶向FgGpmk1抑制剂

靶向HBV衣壳组装调节剂

时间轴

2015

2017

2019

2021

2022



靶向TIE-2抑制剂

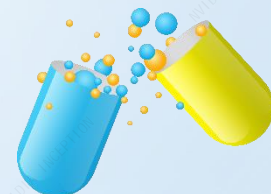
多类全新分子骨架，对阿托伐他汀诱导的脑出血具有很好的保护作用，毒性低。

靶向Top-1抑制剂

优异的体内外活性，与已上市药物拓扑替康相比具有更好的抗癌效果，具有治疗急性炎症疾病的潜力。

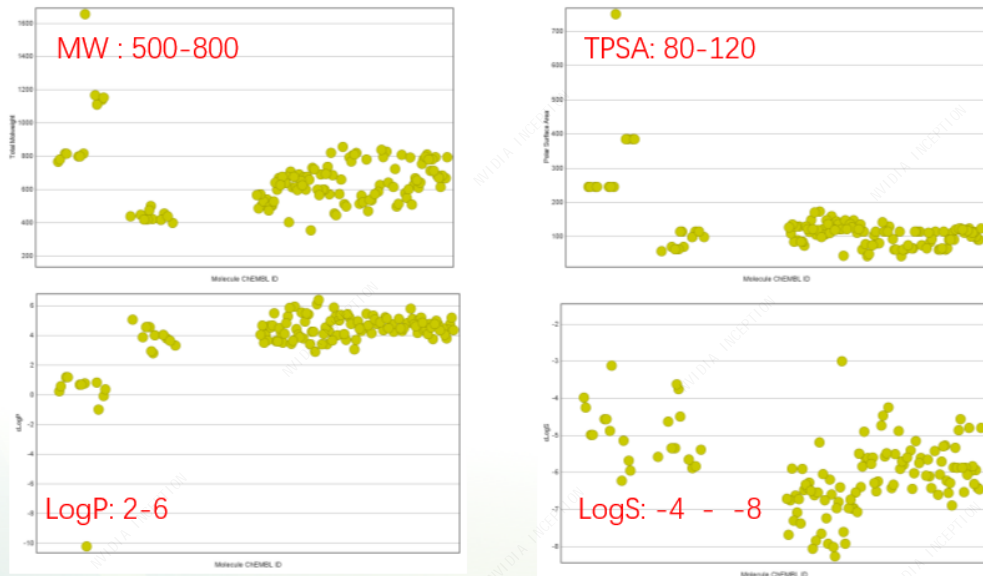
靶向GR拮抗剂

靶向AR-DNA结合域拮抗剂



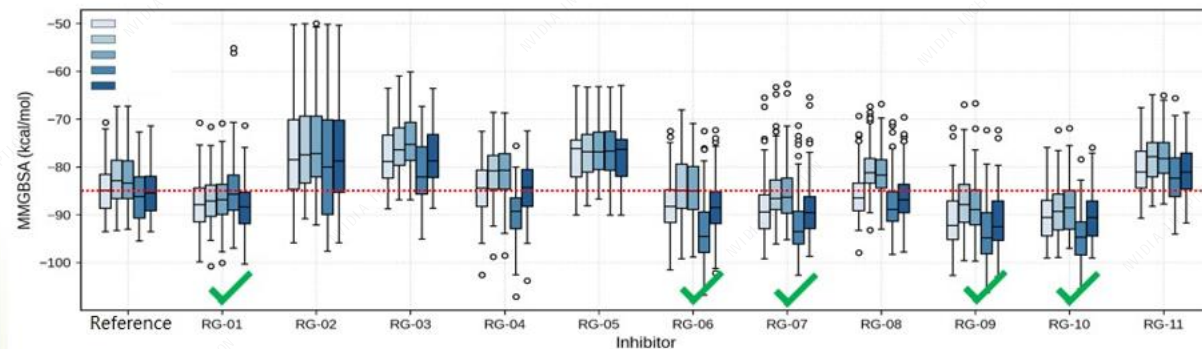
案例：合成少量分子可获得结构新颖且活性有显著优势的分子

1. 活性数据分析

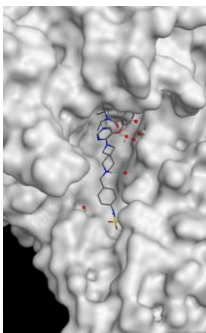


3. 基于RGroup、Linker生成进行分子生成

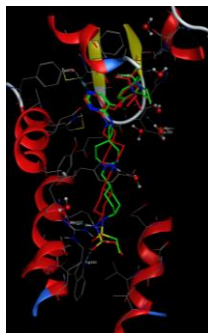
4. 通过动力学模拟进行选择分析确认潜在活性分子



2. 关键相互作用及对接算法分析



晶体结构研究



Carsidock reDocking

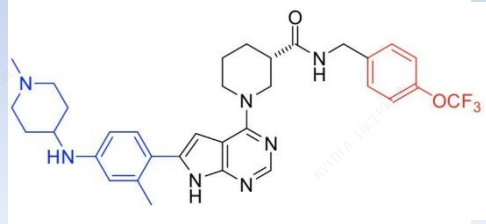
5. 合成并验证活性：仅合成10个分子，就得到结构新颖，且突变型活性高7倍的分子（AI生成分子活性占比60%）

Compound ID	IC ₅₀ (nM)	突变型1 IC ₅₀ (nM)	突变型2 IC ₅₀ (nM)
参考化合物	11.86	1421	280.2
RG-10	9.681	233.2	40.09
RG-06	32.91	513.8	53.8
D5-18-9	221.7	363.2	326
RG-07	364.3	590.7	384.9

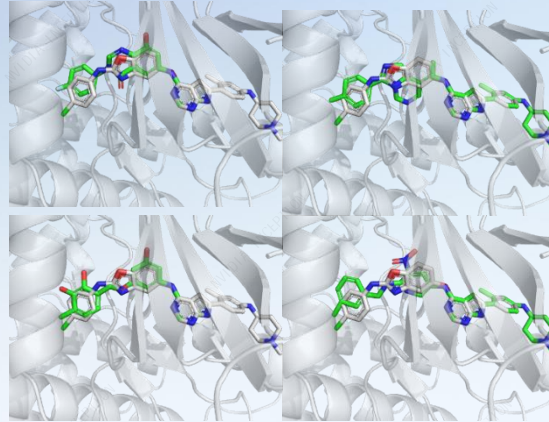
成功案例：优化生成的分子在CLIP1-LTK上验证有动物活性

AI分子设计

在所有癌症中，肺癌的死亡率最高，占癌症相关死亡的19%。而非小细胞肺癌 (NSCLC) 约占肺癌总数的85%，是最常见的肺癌类型。**CLIP1-LTK是2021年在Nature上发表**的NSCLC相关的新靶点。



化学家依据经验保留母核



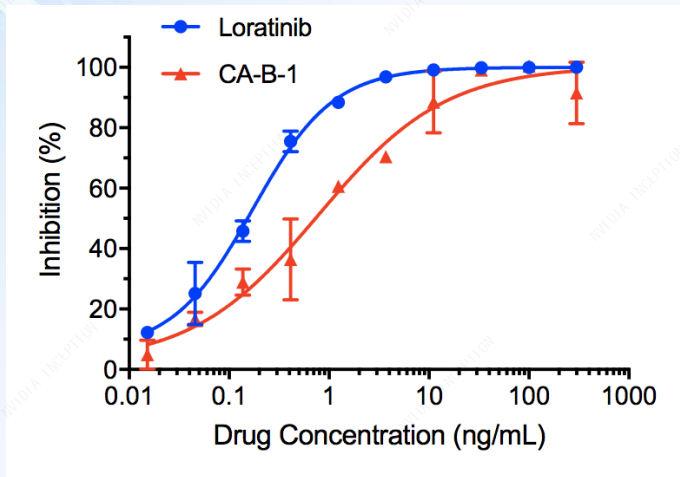
AI基于母核进行约束分子生成



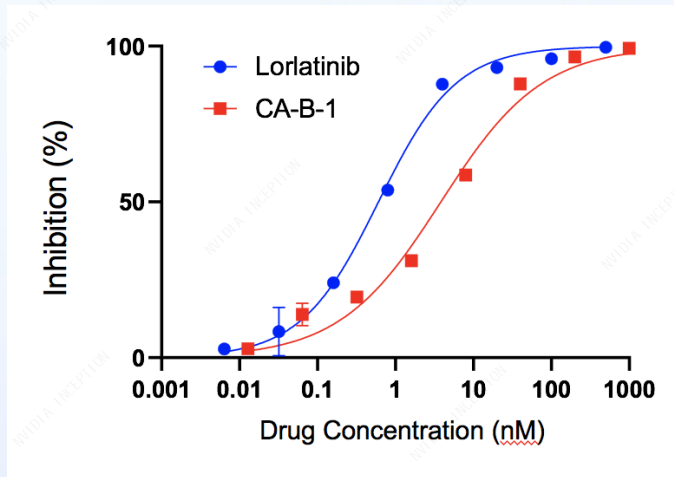
获得新的结果进行下游实验

试验结果验证

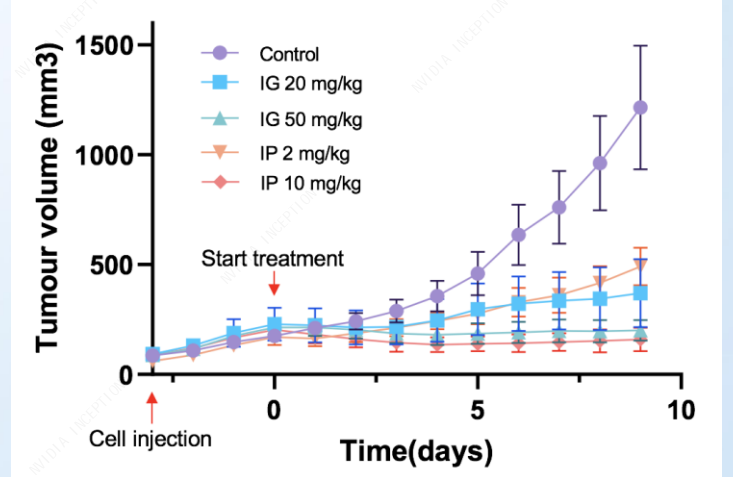
LTK蛋白抑制活性 $IC_{50} = 1.36 \text{ nM}$



Ba/F3-CLIP1-LTK细胞活性 $IC_{50} = 4.6 \text{ nM}$

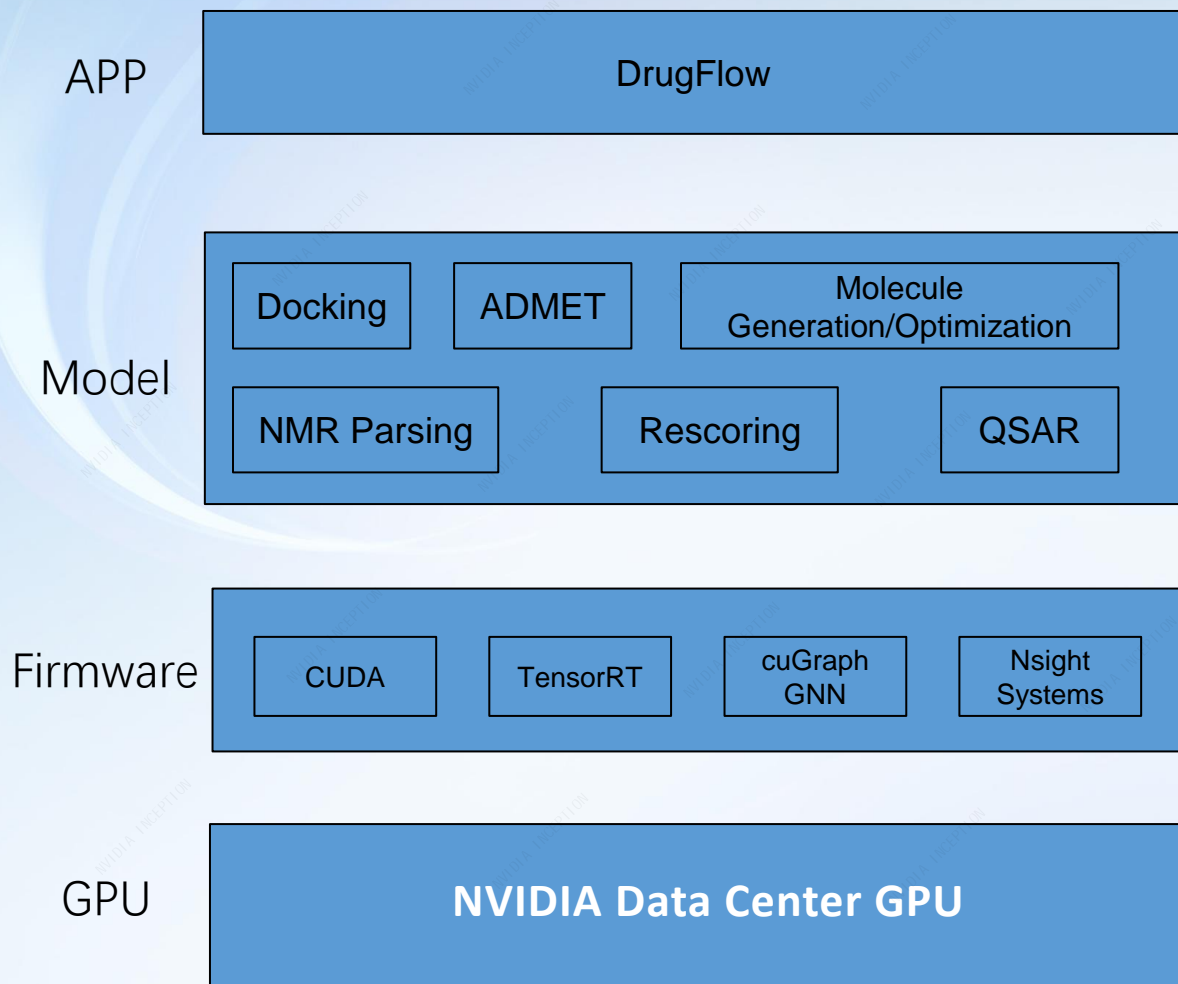


小鼠体内抗肿瘤活性



化合物CA-B-1具有较高的激酶抑制活性、细胞抑制活性，以及动物体内的抗肿瘤活性

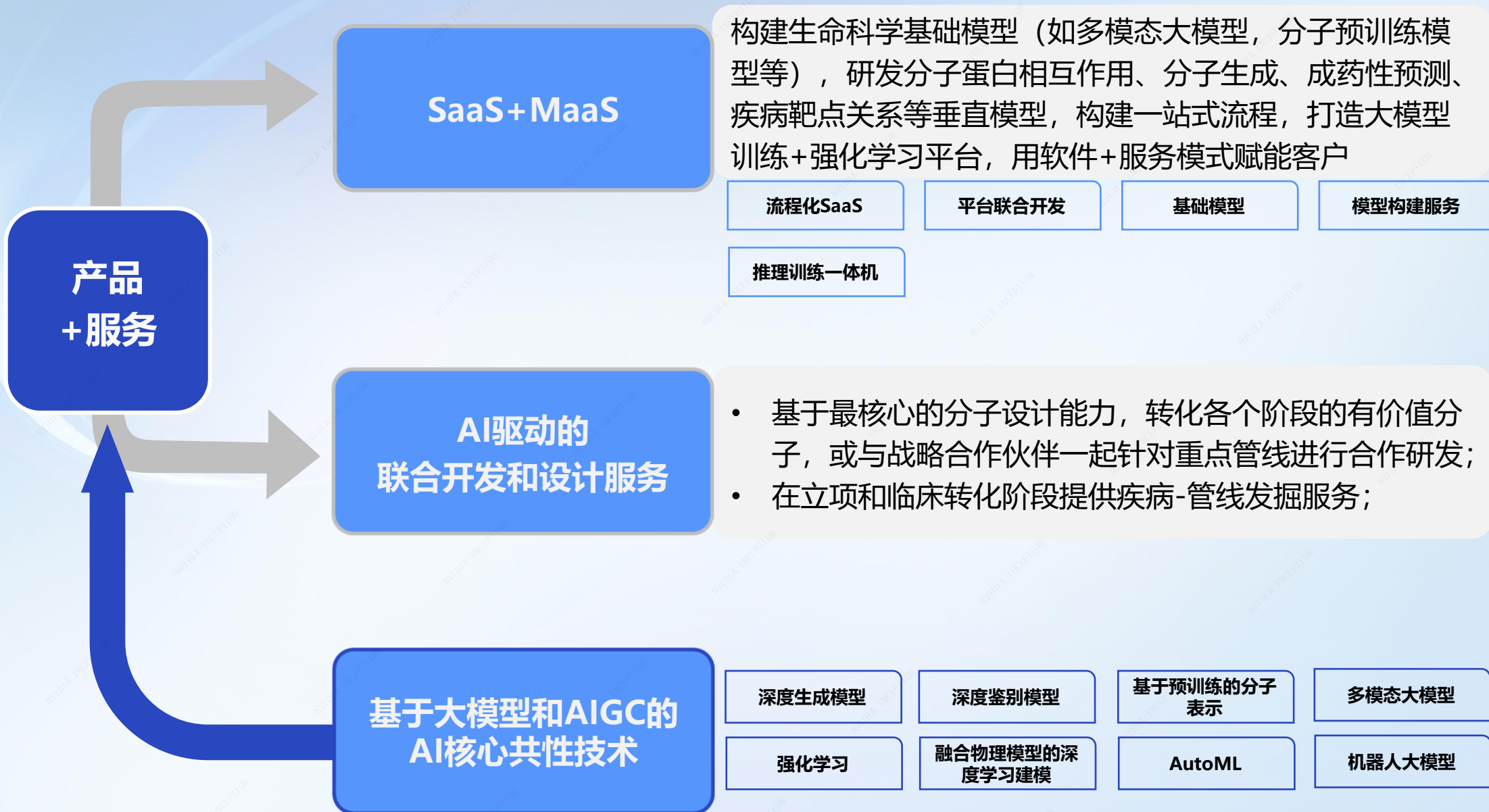
整个项目得到NVIDIA Inception项目的支持



模型运行速度

整个系统共包含100余个AI模型，通过CUDA优化和系统优化，最终绝大多数模型运行在GPU上，系统GPU使用率超过50%。部分模块优化后的运行速度如下：

模块	运行速度	功能	传统方法
CarsiDock	3秒	对接 (活性)	2分钟
KarmaDock	17ms	对接 (活性)	2分钟
MERT	2.35ms	ADMET性质预测	百毫秒
RTMScore	10ms	重打分	秒级
FragGPT	约500个/分钟	分子生成优化	60-100个/分钟



获得NVIDIA
2023初创企业展
示“荣耀企业称
号”（仅10家）

2023入选国家
科技型中小企
业

2022最具潜力
ScienceAI创
业企业

2023张江AI新
药研发联盟成
员单位

2023入选
NVIDIA 初创
企业加速计划

案例入选2022
创新药年度白
皮书

2022天使轮融
资5000万
RMB

2022最具创新
价值落地案例
TOP30

2023入选浙江
省科技型中小
企业

总结：项目优势



顶尖团队配置：AI+交叉融合的专家级团队

- AI算法、高性能计算、药学、化学、量子物理等跨学科专家级团队
- 核心团队具备20年以上行业经验；
- 基于预训练大模型和AIGC技术，在多个任务上显著超越传统模型；
- 一年内，发表包括4篇Nature子刊在内的20+余篇论文，并申请20+篇专利。



打造业内领先的AI平台，显著提升效率和成功率

- 仅用一年时间，推出业内首个AI驱动（结合物理）的具有自主知识产权的可商业化平台；
- 该平台已得到试验验证，AI可生成活性更优的全新结构化合物，且可显著提升早期药物发现的成功率，缩短研发时间。



已实现商业化落地，获业内专家、企业高度认可

- 数月内，软件平台已有上千家注册用户，百家重点POC用户，已签约知名企业 and 高校客户；
- 已与知名药企签约，针对目标靶点提供分子设计服务。



基于多模态语言和控制模型，构建行业基础设施

- 基于多模态模型，构建领域Agent，特别是建模疾病-靶点-基因关系；
- 基于控制大模型，打造低成本、通用机械臂/灵巧手，赋能研发领域。

Product : www.drugflow.com
Business: bd@carbonsilicon.ai



技术公众号: DrugFlow



公司公众号: CarbonSiliconAI

本资料版权归杭州碳硅智慧科技发展有限公司所有

