AI + 生物计算
— 生物医药

AI BioTech

网址：http://biocomputing.top/

Calvin Tang

179209347@qq.com

新药与新药研发

什么是药物

- 用来治疗、预防、诊断疾病的物质。

什么是新药

- 临床应用优于现有治疗效果的药物。
- 新化学实体
- 新靶点配基
- 新作用机制
- 新制剂形式

新药分类

- 化学药物
- 生物药物
- 天然产物（中药）
- 小分子药物
- 大分子药物

Calvin, QQ: 179209347 Mail: 179209347@qq.com

我国新药研发的发展阶段

跟踪仿制阶段	模仿创新阶段	原创新阶段
1950s-1990s	1990s - 现在	现在 – 未来
Copy	Me-too Me-better Fast follow	First in class

- 创新体系建设取得长足发展
- 一批具有自主知识产权的新药研发成功
- 生物医药产业持续高速增长

- 缺乏原创新药
- 缺乏原创理论
- 缺乏原创技术

Calvin, QQ: 179209347 Mail: 179209347@qq.com

药物发现的挑战和关键问题

传统药物研发存在的问题：

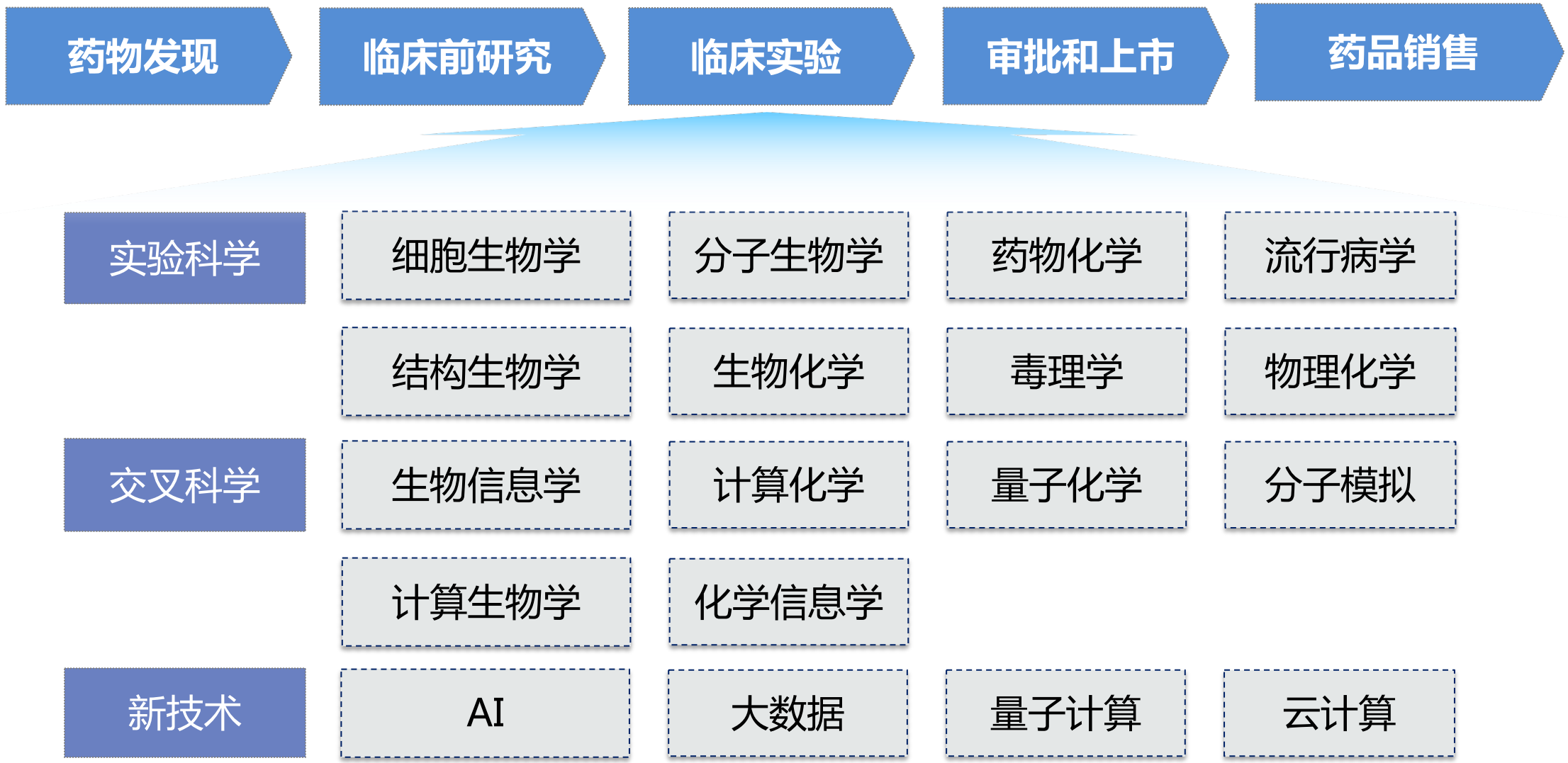
投入大	周期长	效率低	转化慢
研发成本达到新高，是10年前的10倍甚至20倍以上。5年前平均投入26亿美元。	新药发现阶段耗时多长，可谓难有“上限”，通常5-10年都很正常	传统方法以逐步实验筛选为主，失败率高。数据获取技术壁垒高，成本高，保密性强	多由大学和科研机构进行，后经过成果转化，被药企购买，转化很难慢。

AI+药物研发存在的问题：

商业模式不明确	高端复合型人才缺失	数据制约
目前多数企业发展依赖融资，对AI+药物研发技术创新企业来说，是自己做药物研发还是CRO模式，是需要结合自身发展做出适合的选择。	AI药物研发兼具信息科技和医药双重属性，既需要AI的人才也需要懂药物研发的人才。	AI训练模型需要优质的数据，新药研发领域的数据基本掌握在药企手里，公开的数据比较有限。

Calvin, QQ: 179209347 Mail: 179209347@qq.com

药物研发领域涉及的学科技术



Calvin, QQ: 179209347 Mail: 179209347@qq.com

AI 赋能药物研发九大场景

人工智能技术可作用于药物研发中的**五个阶段**：

- 药物发现
- 临床前研究
- 临床试验
- 药品生产
- 销售推广

九大主要场景：

主要应用于靶点发现、化合物合成、新适应症发现、化合物筛选、晶型预测、患者招募、优化临床试验设计、药品检查、学术推广等九大场景。

通过利用自然语言处理、深度学习、机器学习和图像识别等AI技术，来提升药物研发、生产和销售推广效率。

AI 技术在药物研发中的应用概况

药物阶段	应用环节	应用场景
药物发现	靶点发现	利用自然语言处理（NLP）技术检索分析海量的文献、专利和临床试验报告非结构化数据库，找出潜在的、被忽视的通路、蛋白和机制等与疾病的相关性，从而提出新的可供测试的假说，以发现新机制和新靶点
	先导化合物研究和化合物筛选	利用机器学习（或深度学习）技术学习海量化学知识及资料，建立高效的模型，快速过滤"低质量"化合物，富集潜在有效分子
	化合物合成	利用机器学习（或深度学习）技术学习海量已知的化学反应，之后预测在任何单一步骤中可以使用的化学反应，解构所需分子，得到可用试剂
临床前研究	新适应症发现	借助AI的深度学习能力和认知计算能力，将已上市或处于研发管线的药物与疾病进行匹配，发现新靶点，扩大药物的治
	晶型预测	晶型变化会改变固体化合物的物理及化学性质（如溶解度、稳定性、熔点等），导致药物在临床治疗、毒副作用、安全性方面的差异。这一多晶型现象会对药物研发造成干扰。可以利用认知计算实现高效动态配置药物晶型，预测小分子药
临床试验	临床试验设计	利用自然语言处理（NLP）技术检索过去临床试验中的成功和失败经验，使临床试验方案避免重复常见的遗漏、安全等
	患者招募	利用自然语言处理（NLP）技术提取患者数据，为临床试验匹配相应患者
药品生产	药品检查	计算机视觉检测压花、重影、划痕、分层等缺陷
药品销售	学术推广	为药械企业、医生、患者提供全流程的智能医学创新服务

目前，AI主要应用在药物发现阶段和临床前研究阶段，其中，靶点发现是AI+药物研发最热门的领域，按照应用场景的发展速度来看，未来药物合成或将成自动化程度最高的方向。

注：先导化合物研究是指将数以百万计的小分子化合物进行组合实验，以发现具有某种生物活性和化学结构的化合物，一般有两种思路，即高通量筛选和虚拟药物筛选。

我国主要AI+药物研发公司业务布局

重点企业	药物发现			临床前研究		临床试验		获批上市	商业化
	靶点发现	化合物筛选	化合物合成	晶型预测	新适应症发现	临床试验设计	患者招募	注册申报	学术推广
晶泰科技		■	■	■					
云势软件	■				■				■
深度智耀		■	■					■	
亿药科技	■	■							
宇道科创		■							
AccutarBio		■							
望石智慧		■	■						
矩坤智能	■	■			■				
零氪科技							■		■
百奥知						■			■
METiS	■	■	■						
分迪科技		■							
费米子	■	■	■						
智药科技	■	■	■						
元气知药	■	■	■						
赛恪科技		■							
火石数智									■

国内药物开发CRO企业及服务内容

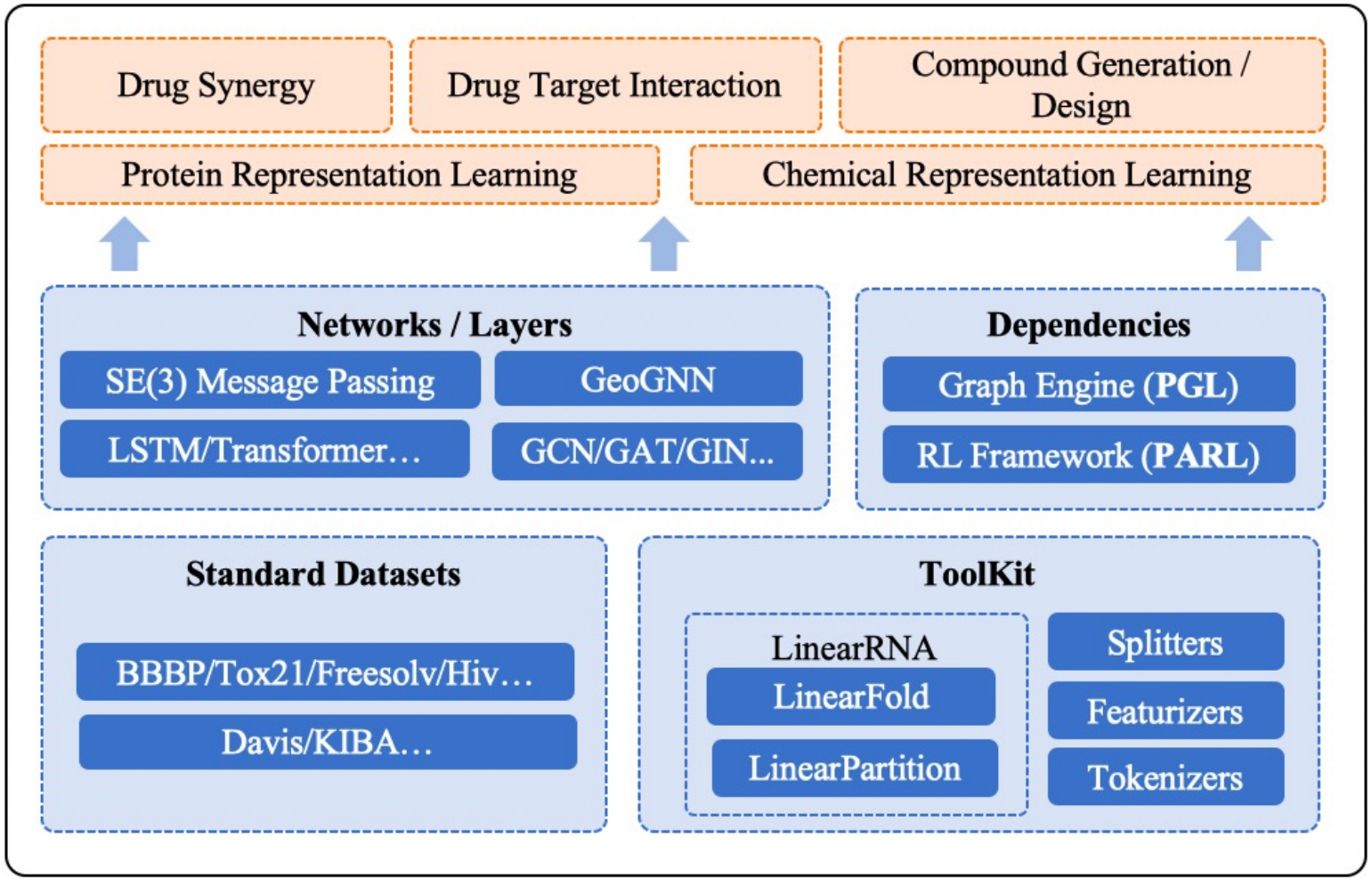
药明康德	基于片段的药物设计	基于结构的药物设计	虚拟筛选	高通量筛选	DNA编码化学物库技术
睿智化学	基于片段的药物设计	基于结构的药物设计	计算机辅助药物设计	多肽合成	杂交瘤和噬菌体技术
维亚生物	基于片段的药物设计	基于结构的药物设计	GPCR膜蛋白靶标技术平台	亲和选择质谱筛选平台	
药明生物	杂交瘤技术平台	噬菌体技术	双抗、ADC平台		
康龙化成	放射性标记化学合成技术、RadioTag技术	蛋白、多肽和小分子的聚乙二醇化、核苷酸共轭修饰			
药石科技	分子砌块技术				
成都先导	DNA编码化学物库技术				

AI辅助新药发现和研发的发展方向

药物理念	• 应用的目标是新药，药物的概念需要变化
理论基础	• AI技术基于药理学理论，理论创新至关重要
数据积累	• 数据库与大数据，药学与药理学数据的有限
技术方法	• AI起步与发展，数学模型，计算方法
研究策略	• 合理高效利用现有技术方法和资源

Calvin, QQ: 179209347 Mail: 179209347@qq.com

AI+药物研发技术框架案例：PaddleHelix



螺旋桨（PaddleHelix）是一个生物计算工具集，是用机器学习的方法，特别是深度神经网络，致力于促进以下领域的发展：

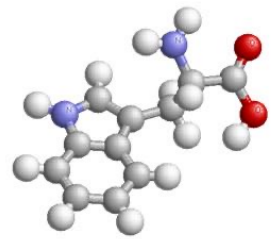
•**新药发现**。提供1)大规模预训练模型:化合物和蛋白质; 2)多种应用:分子属性预测,药物靶点亲和力预测,和分子生成。

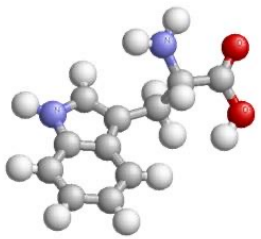
•**疫苗设计**。提供RNA设计算法,包括LinearFold和LinearPartition。

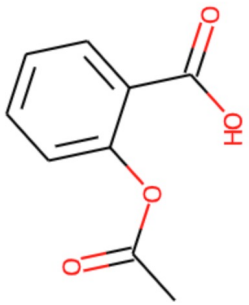
•**精准医疗**。提供药物联用的应用。

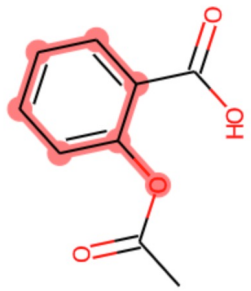
AI+药物研发案例：化学结构图像识别

实现分子结构“截图-粘贴-识别-展示-重新编辑”的自动化流程





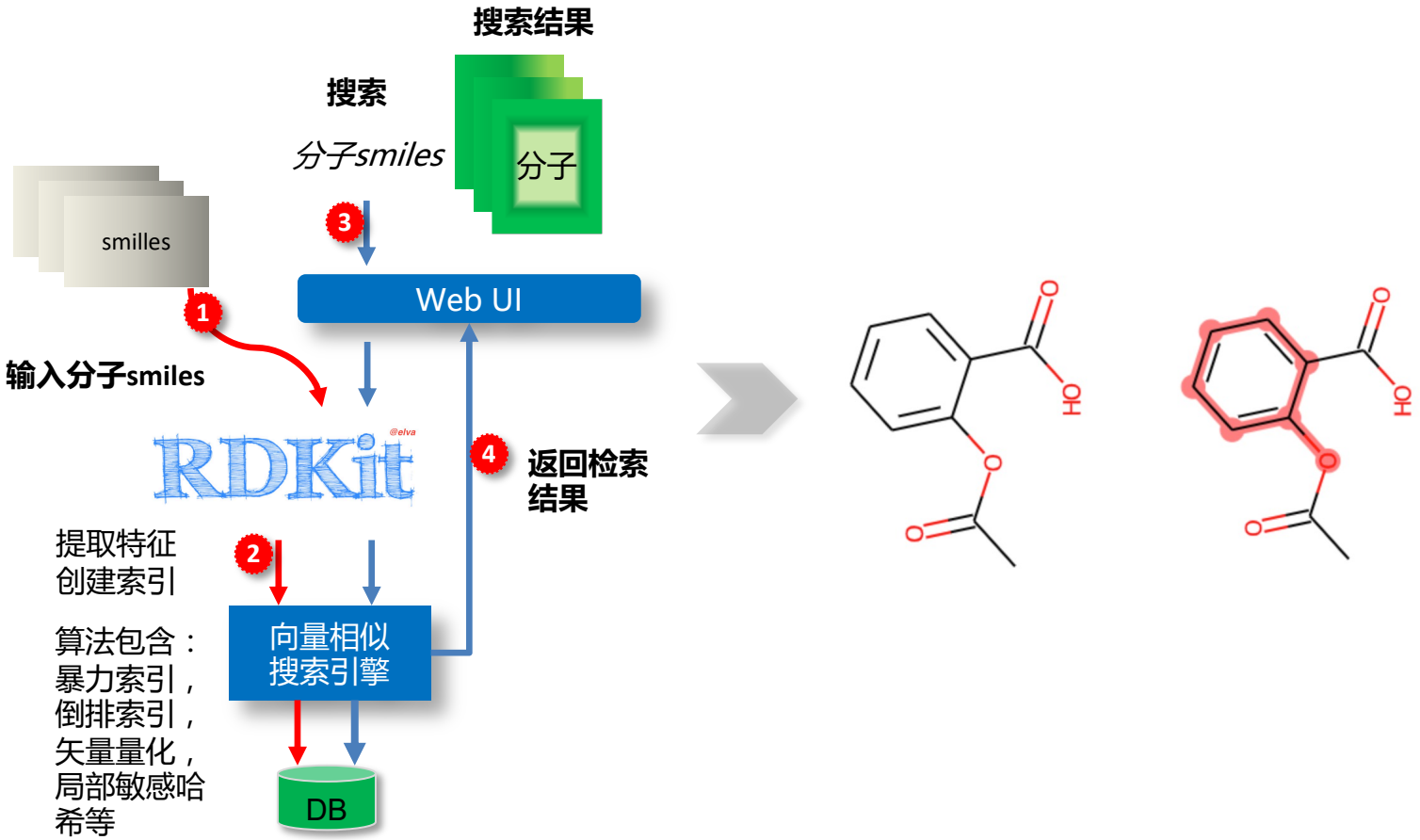




AI+药物研发案例：分子搜索

- **使用RDKit提取分子特征**

✓RDKit是一个用于化学信息学的开源工具包，基于对化合物2D和3D分子操作，利用机器学习方法进行化合物描述符生成，fingerprint生成，化合物结构相似性计算，2D和3D分子展示等。

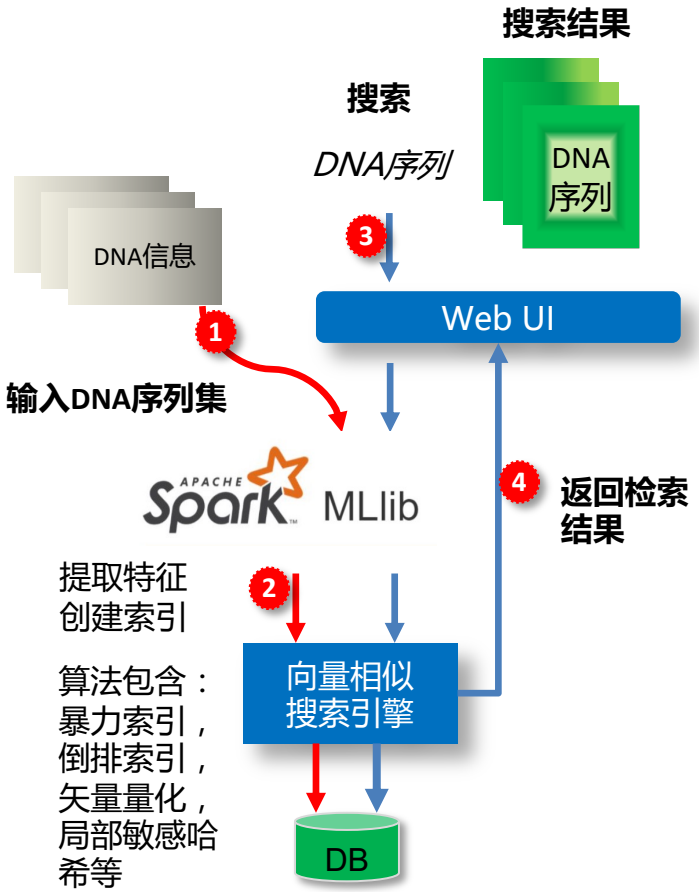


- 特征向量相似度搜索
- 单台服务器十亿级数据的毫秒级搜索
- 云原生，近实时搜索，支持分布式部署
- 随时对数据进行插入、删除、搜索、更新等操作

AI+药物研发案例：DNA序列搜索

DNA序列特征提取

- ✓ 将文本数据转化成特征向量的过程，比较常用的文本特征表示法为词袋法。
- ✓ CountVectorizer是属于常见的特征数值计算类，是一个文本特征提取方法。对于每一个训练文本，它只考虑每种词汇在该训练文本中出现的频率。
- ✓ CountVectorizer会将文本中的词语转换为词频矩阵，它通过fit函数计算各个词语出现的次数。
- ✓ CountVectorizer旨在通过计数来将一个文档转换为向量。当不存在先验字典时，Countvectorizer作为Estimator提取词汇进行训练。

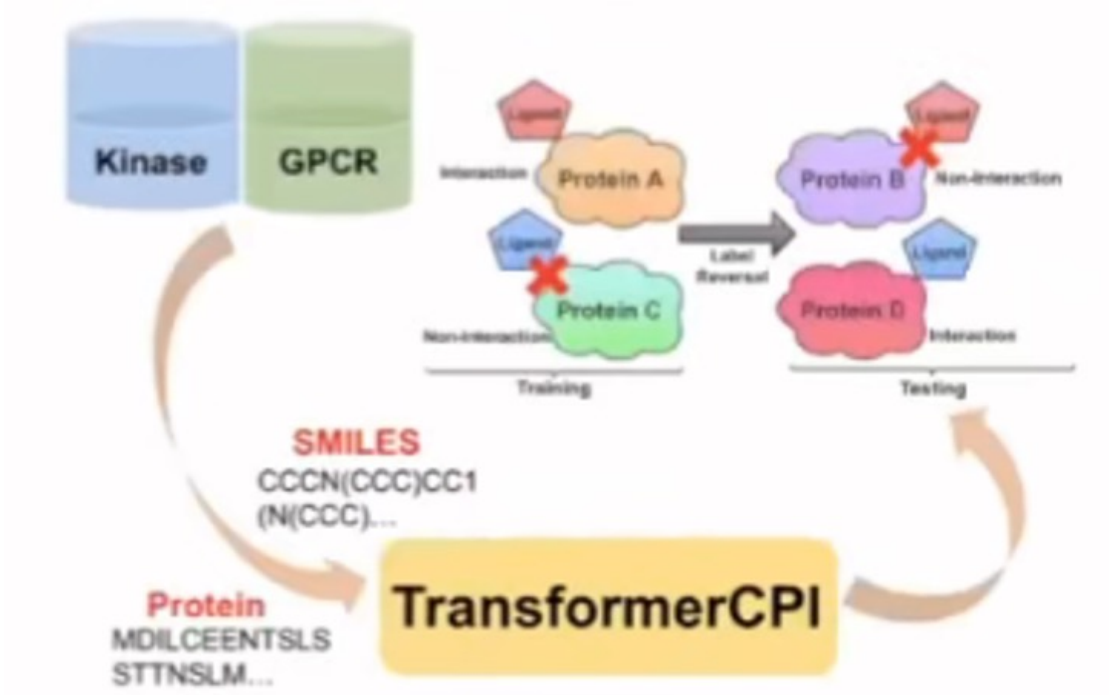


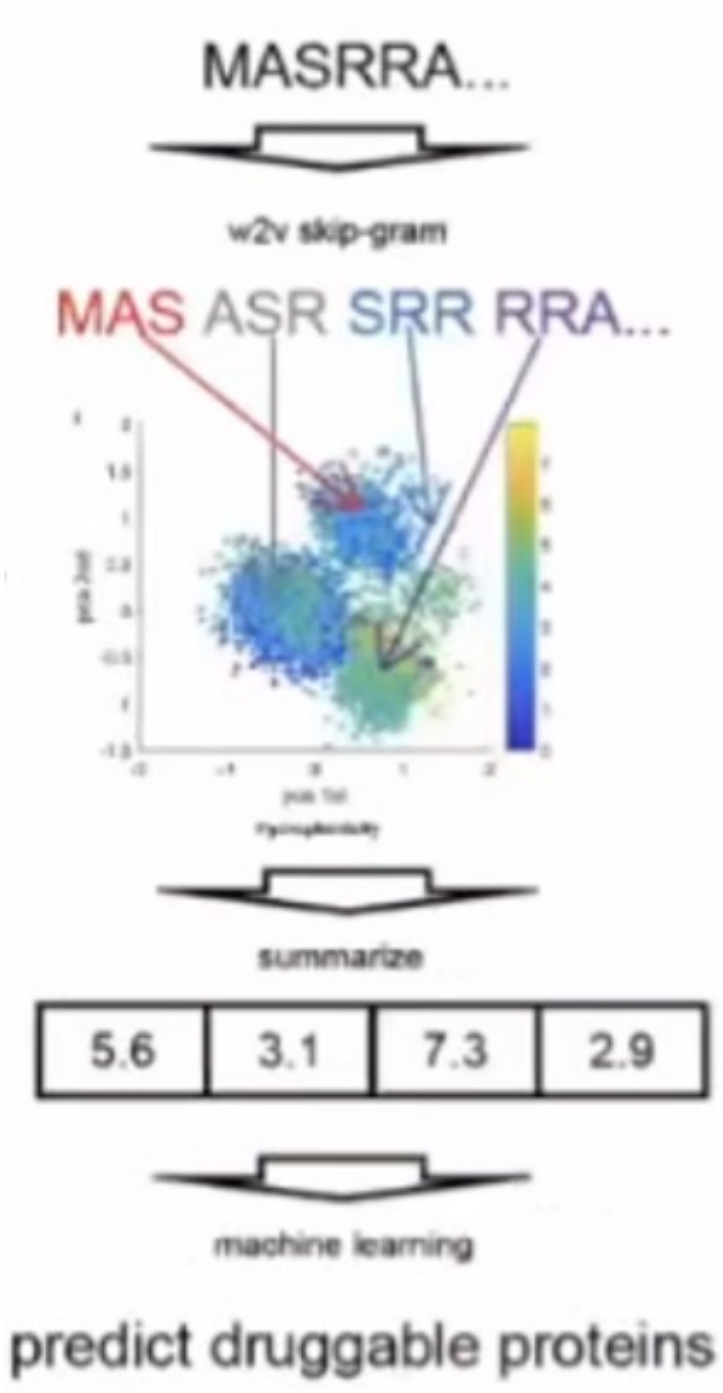

```
# 特征向量

+-----+--------------------+--------------------+
|label|            sequence|            features|
+-----+--------------------+--------------------+
|    4|[ATGC, TGCC, GCCC...|(336,[0,8,14,17,1...|
|    4|[ATGA, TGAA, GAAC...|(336,[0,1,2,3,5,7...|
|    3|[ATGT, TGTG, GTGT...|(336,[0,1,2,3,4,5...|
|    3|[ATGT, TGTG, GTGT...|(336,[0,1,2,3,4,5...|
|    3|[ATGC, TGCA, GCAA...|(336,[0,1,2,3,4,5...|
+-----+--------------------+--------------------+
```

- 特征向量相似度搜索
- 单台服务器十亿级数据的毫秒级搜索
- 云原生，近实时搜索，支持分布式部署
- 随时对数据进行插入、删除、搜索、更新等操作

AI+药物研发案例：靶点发现与确证

- 基于word2vec模型预测靶点可药性
- 基于蛋白质序列的靶点预测

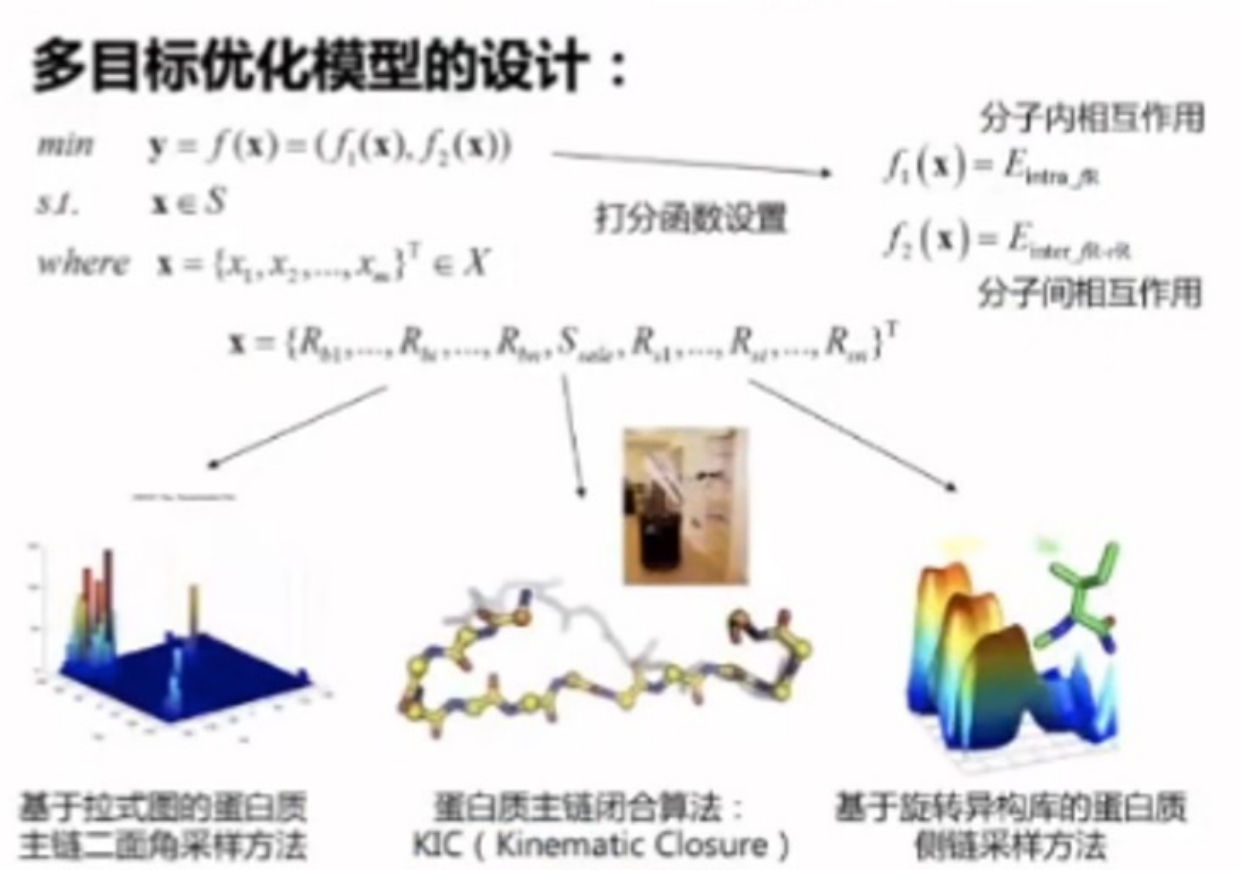


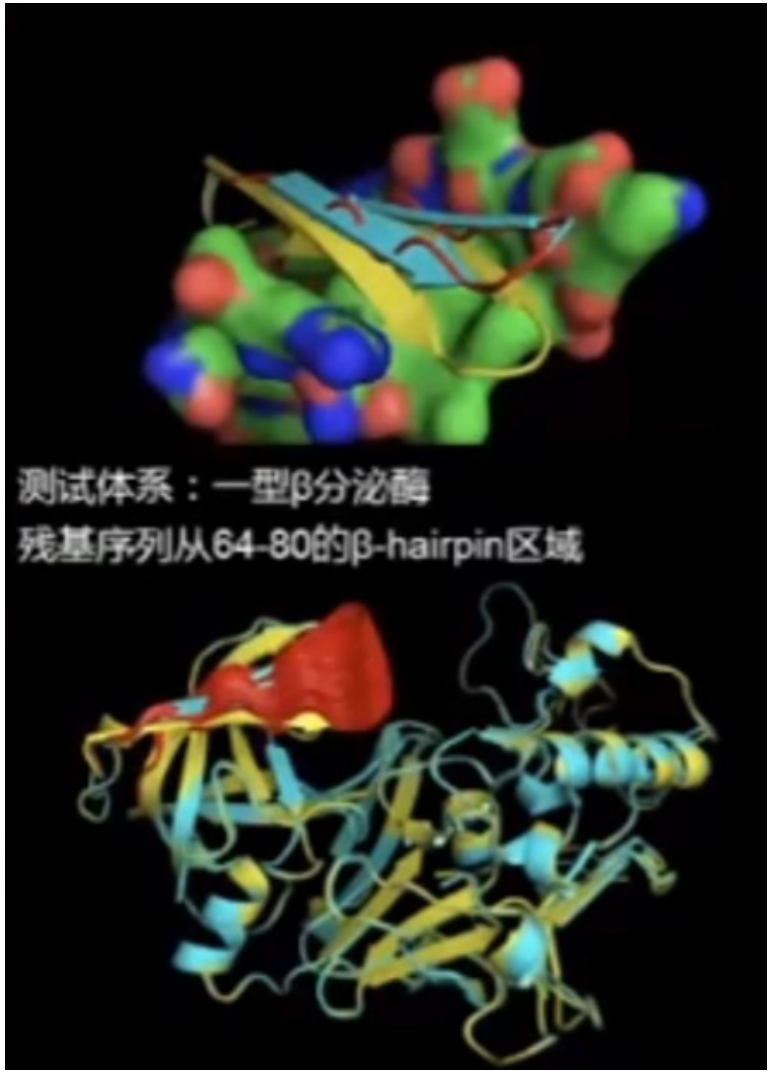


蒋华良院士-AI在创新药物研发中的作用

AI+药物研发案例：药物靶点相互作用分析

- 药物多靶点活性谱预测分析
- 大尺度蛋白质主链的构象采样及优化方法

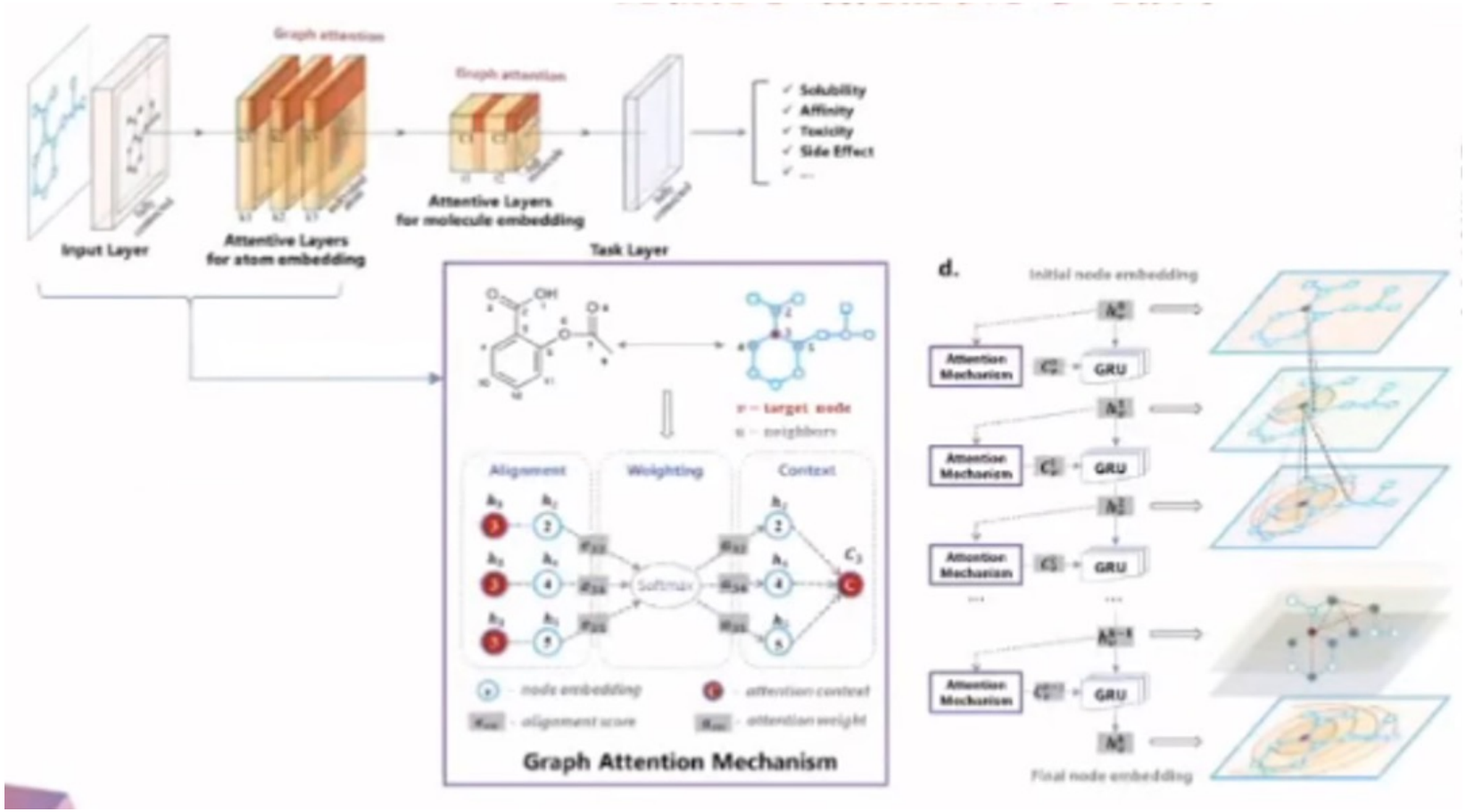




蒋华良院士-AI在创新药物研发中的作用

AI+药物研发案例：活性化合物发现与优化

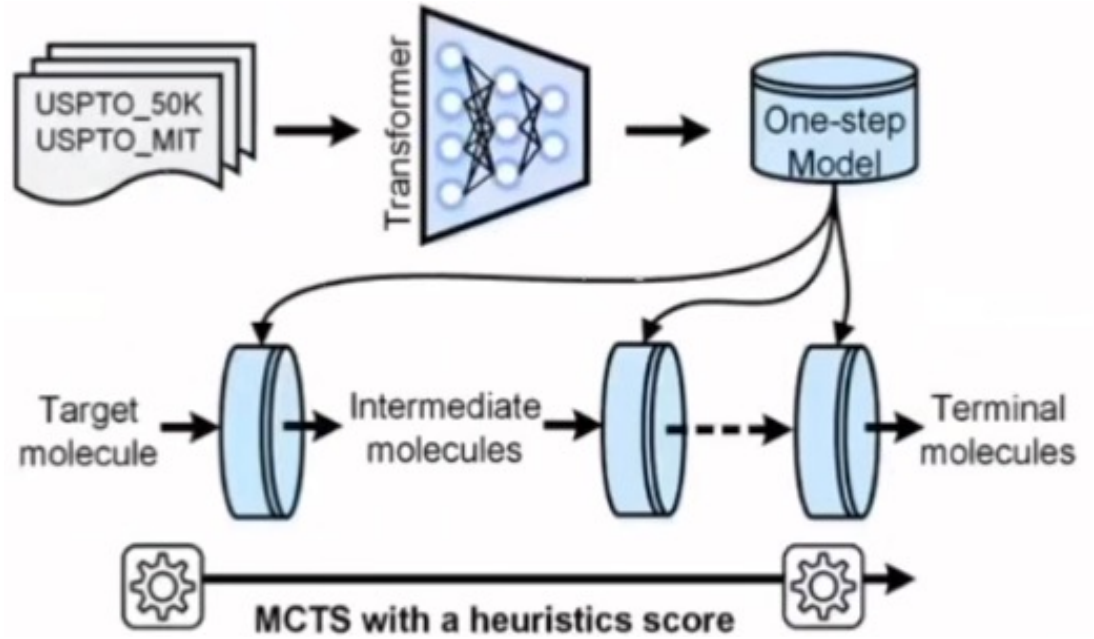
- 药物分子结构的表示学习技术
- 基于癌症基因组数据进行药物重定向
- 预测疾病模型对于抗癌药物的敏感度

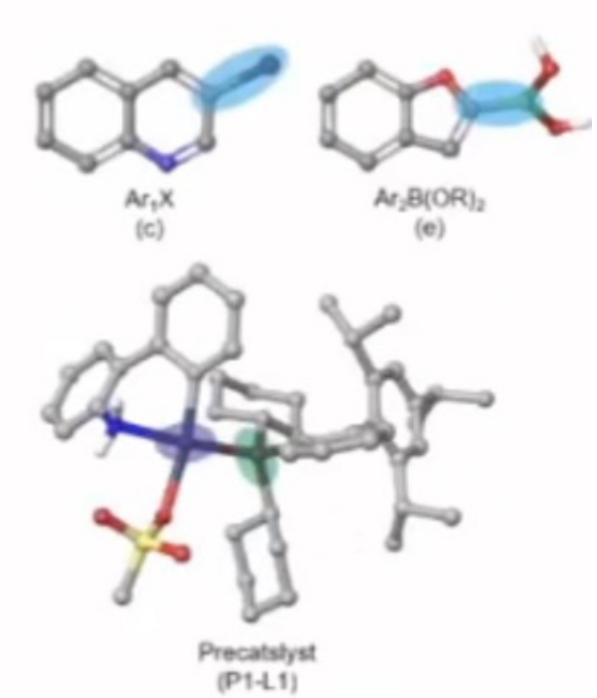


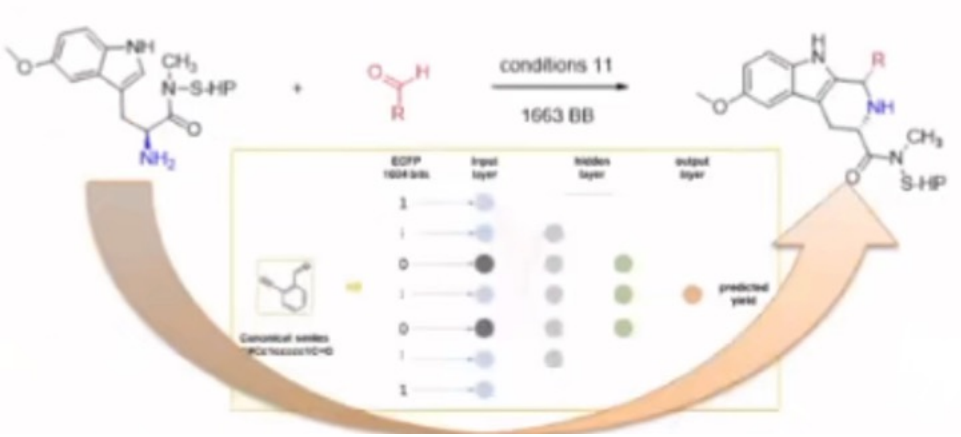
蒋华良院士-AI在创新药物研发中的作用

AI+药物研发案例：化学反应预测

- 不基于模版的逆反应路线预测
- 利用深度学习预测Suzuki-Miyaura偶联反应
- 利用深度学习优化Pictet-Spengler反应







蒋华良院士-AI在创新药物研发中的作用

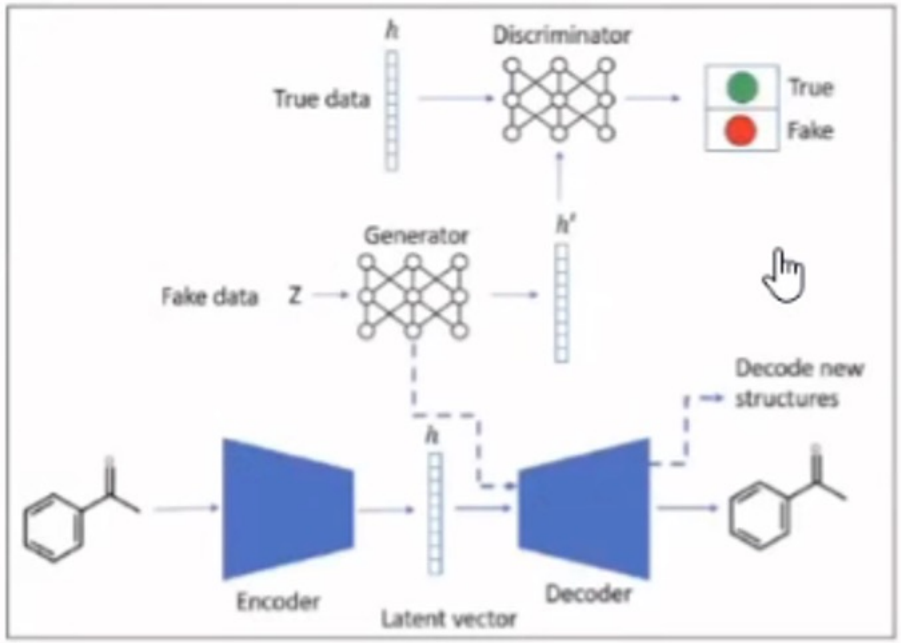
AI+药物研发案例：基于GAN分子生成

GAN (生成对抗网络)的思想是设置一个零和博弈,博弈中生成器用来生成样本, 判别器用来判断输入样本是否属于真实的训练样本。

- 使用fingerprints, concentration和Growth Inhibition percentage(GI)共同引导生成新化合物(新)

用GAN生成的图像

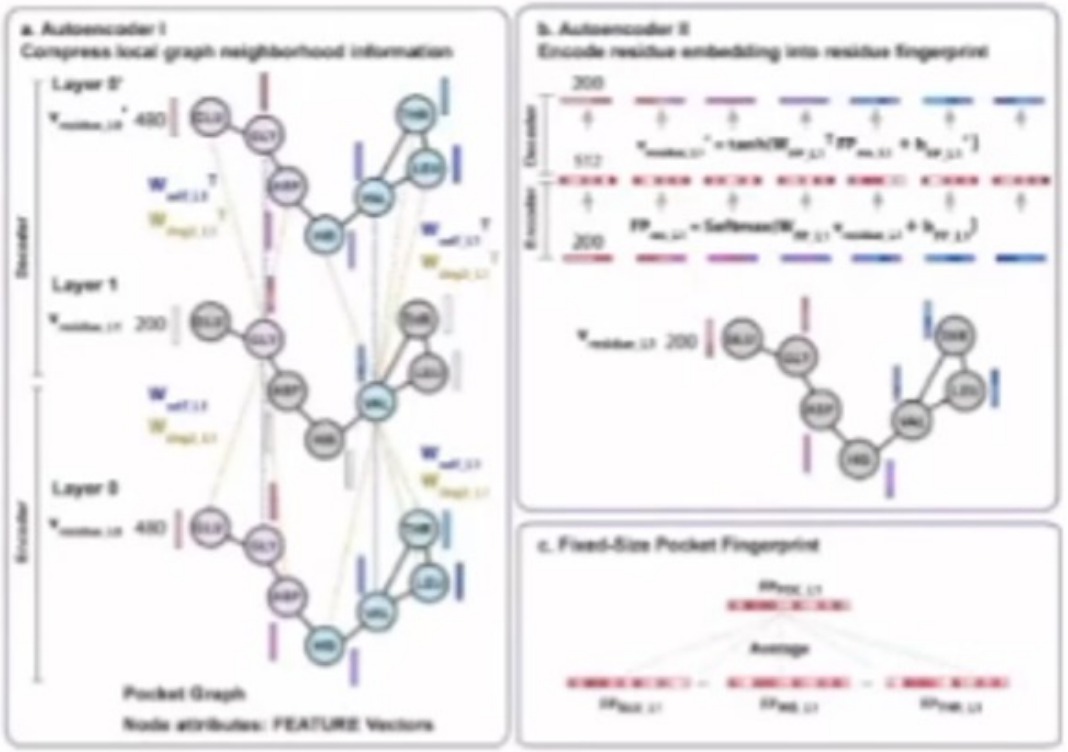
将GAN应用在药物分子生成



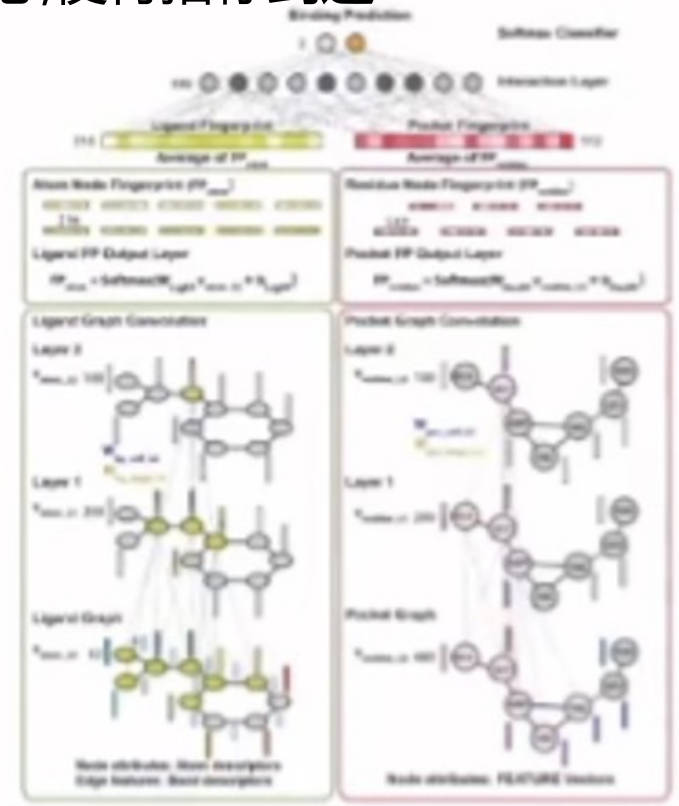
Calvin, QQ: 179209347 Mail: 179209347@qq.com

AI+药物研发案例：基于图神经网络药物靶点相互作用识别

- 图卷积神经网络(Graph Convolutional Network)是一种能对图数据进行深度学习的方法,本质目的就是用来提取拓扑图的空间特征。
- 算法改进：通过残差网络和attention机制来保留多层次的局部信息,使得指标到达90%



受体和配体的图形式

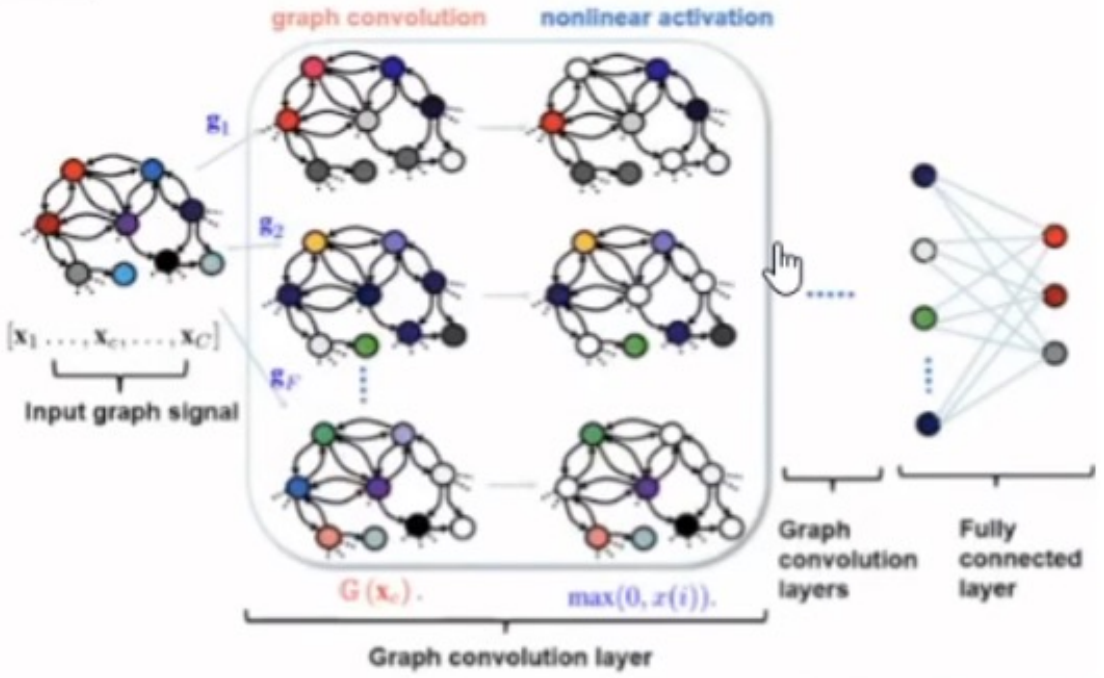


图神经网络用于分子对接

Calvin, QQ: 179209347 Mail: 179209347@qq.com

AI+药物研发案例：基于蛋白结构的药物筛选

对global, L1700, selleck三个化合物库,分别用随机森林模型, gcnn神经网络模型进行筛选,对两个筛选结果取交集，得到最终筛选结果。



Calvin, QQ: 179209347 Mail: 179209347@qq.com

AI+药物研发案例：药物再定位

基于AI技术的药物再定位研究

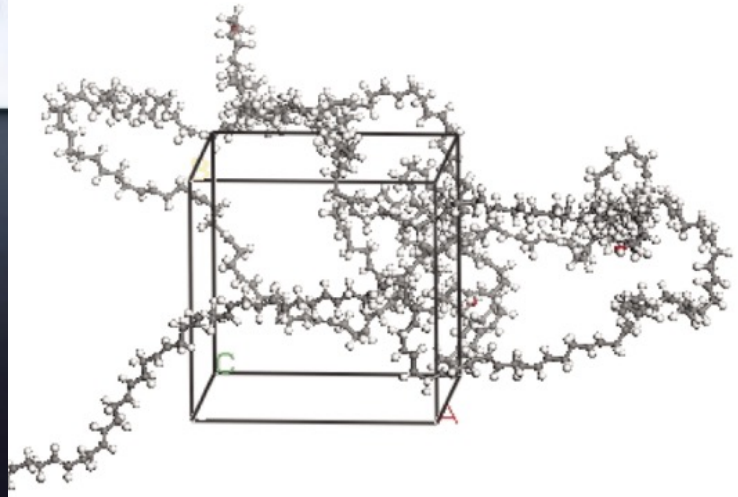
对全球上市药物筛选,寻找HDAC1抑制剂。借助人工智能的方法提高筛选的准确率。

- 采用随机森林算法(Random Forest)替换掉朴素贝叶斯,筛选出35种候选化合物。

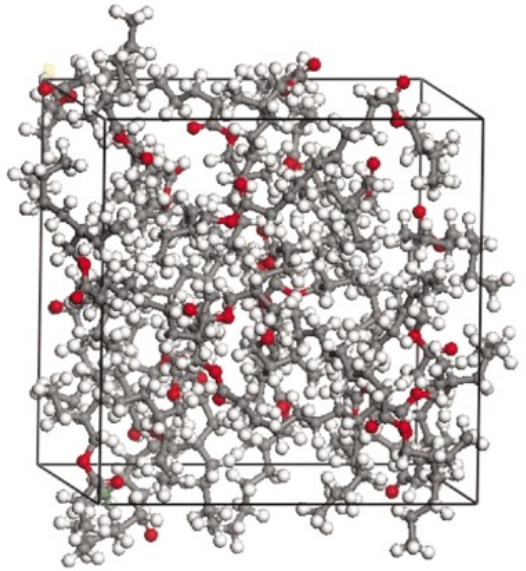
Calvin, QQ: 179209347 Mail: 179209347@qq.com

量子计算+药物研发案例：分子模拟

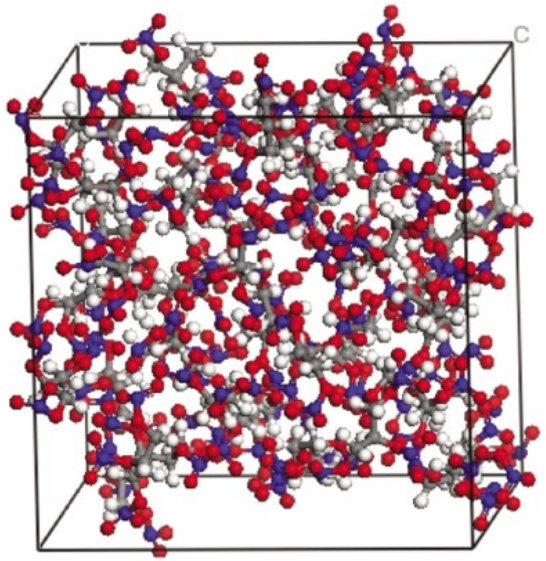
生物医药与量子计算的结合被社会各界普遍看好，是因为在所有待量子升级行业中，生物医药自身的科技水平本就很高， 高科技行业对新科技的接受度最高，也就是说生物医药具备接纳量子计算的天然环境。



a



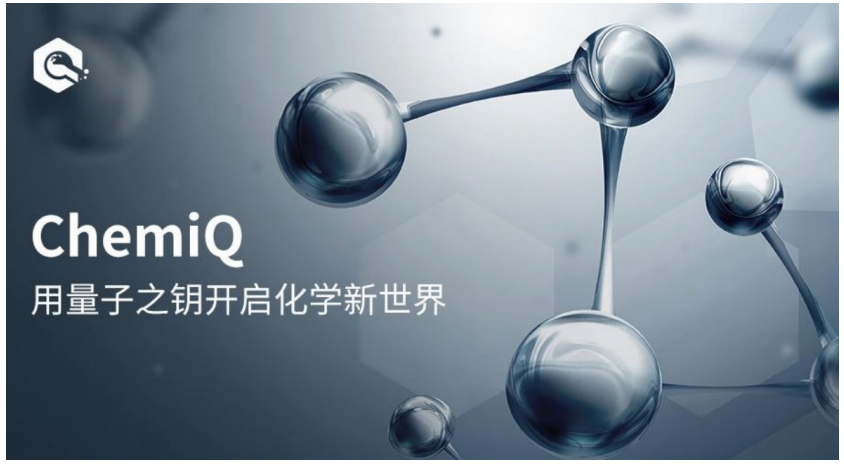
b



c

量子计算+药物研发案例：量子计算化学

本源量子计算化学软件ChemiQ，利用量子天然模拟优势，结合本源量子专业技术，使用真实或模拟量子计算机，能够可视化构建分子模型，快速揭示分子的电子结构，在加速化学合成、药物研发、材料设计、能源开发等方面具备广泛应用前景。



化学计算普遍性难点

大体系复杂分子态计算　过渡态理论的修正　高精度计算分子间弱相互作用　原子核量子效应的模拟

大尺度（纳米至微米）模拟及多尺度嵌套模拟　准简并体系的电子结构问题　……

- 变分量子算法
- 量子优化算法
- 量子虚实演化
- 量子机器学习
- …

ChemiQ量子化学计算

- 量子操作系统
- 量子计算编程语言
- 量子计算开发框架
- 量子机器学习框架
- ……

真实量子计算机

超导、半导…

量子虚拟机

单振幅、全振幅、部分振幅…

Calvin，QQ: 179209347 Mail: 179209347@qq.com

Thank You