

# 计算机辅助与 人工智能药物发现



INHIBITORS & AGONISTS



COMPOUND LIBRARIES



NATURAL PRODUCTS



RECOMBINANT PROTEINS



TECHNICAL SERVICE

# TOPSCIENCE

药物发现一站式平台 陶术

INHIBITORS & AGONISTS | COMPOUND LIBRARIES | TECHNICAL SERVICE  
RECOMBINANT PROTEINS | NATURAL PRODUCTS

[www.tsbiochem.com](http://www.tsbiochem.com)



# 计算机辅助药物设计

TOPIC   内容	PAGE   页码
技术概述	01
...	.....
常用方法	02
虚拟筛选	02
分子对接	06
分子动力学模拟	10
反向找靶	13
...	.....
筛选方案	16
...	.....
陶术生物优势	16

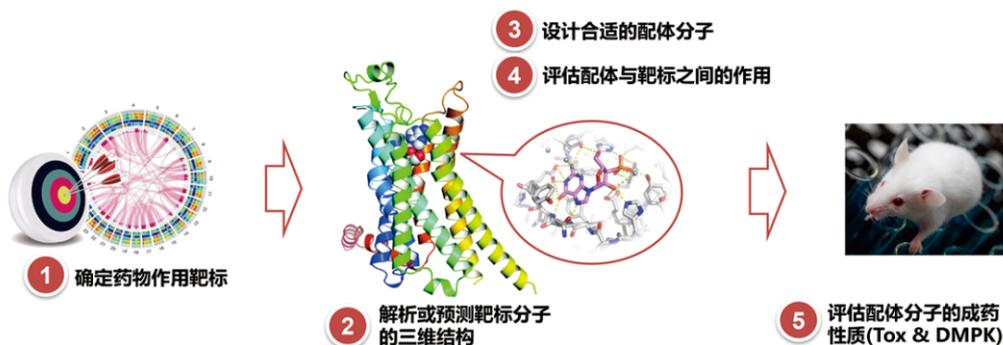


# SUMMARY

## 技术概述

### Computer-Aided Drug Design

计算机辅助药物设计 (Computer-Aided Drug Design, 简称 CADD) 是一类借助计算机技术来发现和优化活性先导化合物至候选药物的技术。CADD 技术可以助力药物研发早期阶段的多个环节, 包括确定药物作用靶标、解析靶标分子的三维结构、设计合适的配体分子、评估配体与靶标之间的相互作用以及评估化合物的毒性和 DMPK 等基础成药性质。CADD 技术的广泛应用显著提高了新药发现的效率, 降低了新药发现的成本。



## CADD 技术早期阶段取得的标志性成功案例

药物	治疗疾病	靶点	开发企业	上市时间
舒马曲坦 (Sumatriptan)	偏头痛	5-HT <sub>1D</sub> 受体	GSK (UK)	1993
多佐胺 (Dorzolamide)	青光眼	碳酸酐酶	Merck (US)	1995
沙奎那韦 (Saquinavir)	艾滋病	HIV 蛋白酶	Roche (UK)	1997
扎那米韦 (Zanamivir)	流感病毒	神经氨酸苷酶	Biota (Australia)	1999
伊马替尼 (Imatinib)	慢性髓性白血病	酪氨酸激酶 Abl	Novartis (Swiss)	2001
拉替拉韦 (Raltegravir)	艾滋病	整合酶	Merck (US)	2007

计算机辅助药物设计近年来得到人工智能技术的加持, 已逐步演化为人工智能药物设计 (Artificial Intelligence Drug Design, 简称 AIDD), 在药物发现领域中不断展现出新的面貌。

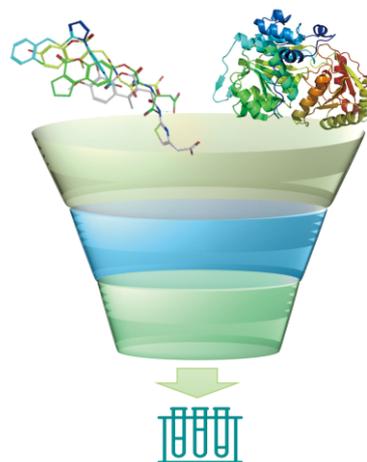


# METHOD 常用方法

Computer-Aided Drug Design

## 1、虚拟筛选

虚拟筛选是基于药物设计理论,利用计算机技术和专业应用软件,从大量化合物中筛选出具有潜在活性的候选化合物后,对其进行实验验证的一种方法,是现代药物发现中常用的先导化合物筛选技术之一。



### 虚拟筛选与高通量筛选特点对比

与传统的高通量筛选相比,虚拟筛选具有明显的优势。它能够综合考虑包括化合物的理化性质、ADMET 参数等多种因素,提供更全面的评估。虚拟筛选无需制备和测试海量的实体化合物,从而大幅提升实验的速度和效率。此外,虚拟筛选成功获得活性化合物的概率明显高于高通量筛选。



#### 高通量筛选

以分子/细胞水平为基础

实验室自动化设备

样品配置耗费人力

数十万级

高

慢



理论基础

设备需求

样品需求

样品数量

成本

时间



#### 虚拟筛选

以药物设计理论为基础

高性能计算机

不受样品限制

数百万-千万级

低

快

## 虚拟筛选的方法

虚拟筛选根据药物设计原理主要分为两种方法：**基于配体的筛选方法**和**基于受体的筛选方法**。

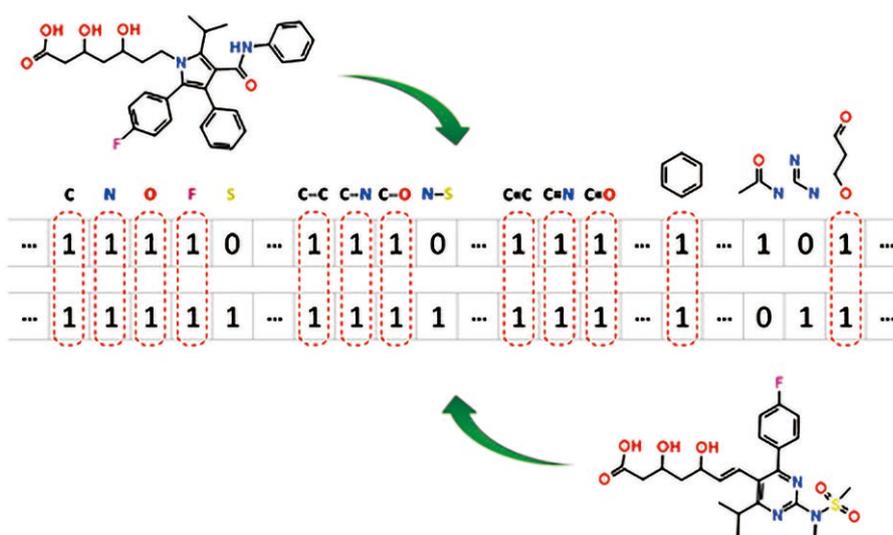
### 1) 基于配体的筛选:

利用已知活性分子的化学信息, 通过相似性搜索筛选潜在活性分子。

#### ● 基于 2D 结构相似性的筛选方法

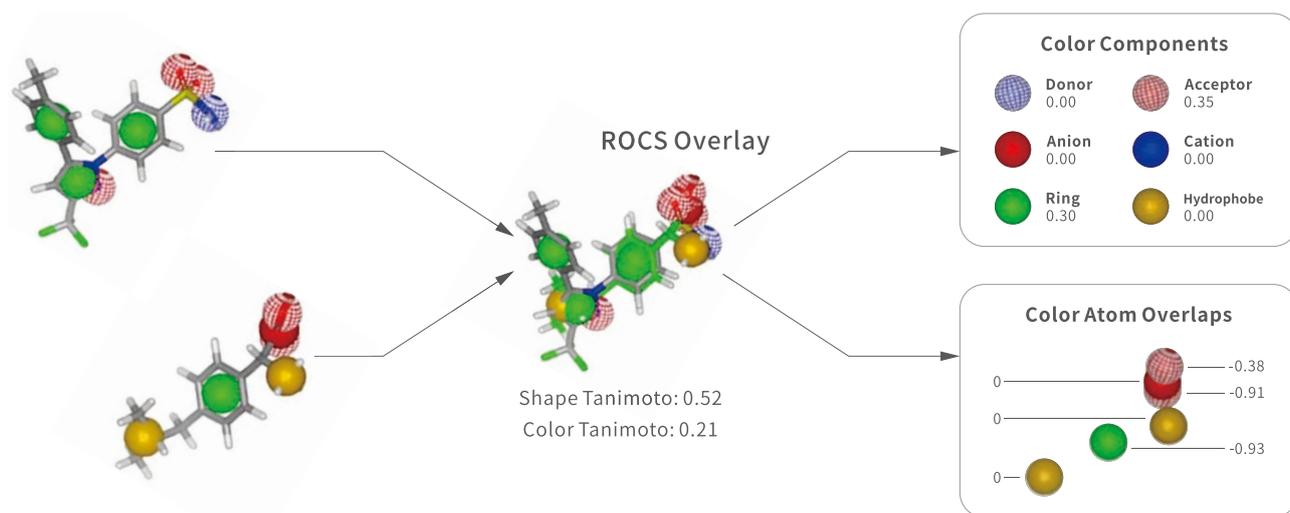
利用已知活性分子作为模板, 通过化学结构相似性搜索的方法筛选潜在活性分子。

$$\text{Tanimoto coeff.} = \frac{N_{A\&B}}{N_A + N_B - N_{A\&B}}$$



#### ● 基于 3D 结构相似性的筛选方法

利用分子的三维形状, 药效团特征计算待筛小分子化合物与参照化合物的相似度

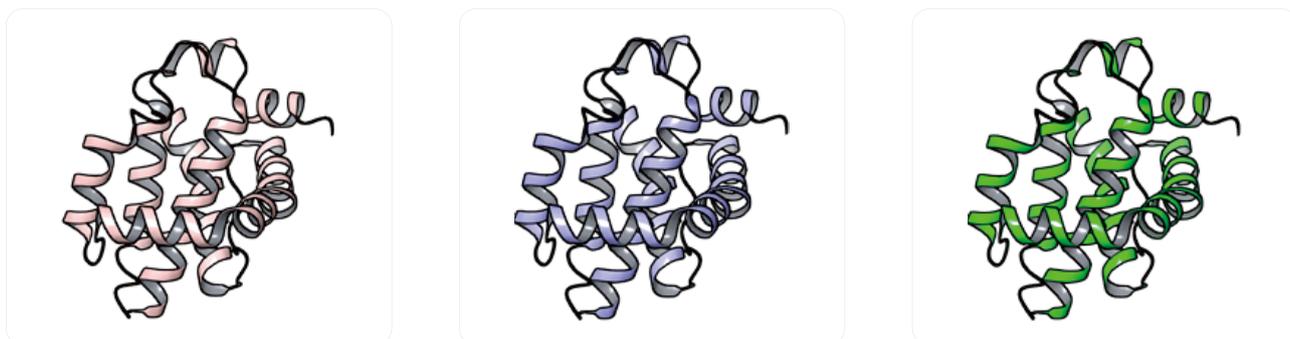


## 2) 基于受体的筛选:

基于受体的三维结构, 在结合位点处自动匹配小分子, 然后对可能的结合模式进行结合能计算, 从中挑选候选化合物。在此基础上衍生出多种虚拟筛选策略:

### ● 基于多构象的虚拟筛选

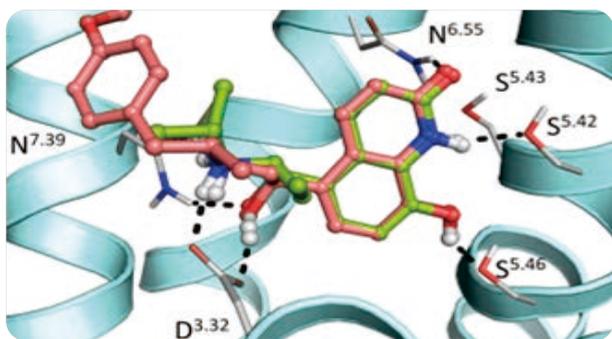
考虑到蛋白柔性的影响, 采用多个晶体结构或基于分子动力学模拟所得的若干个典型构象进行平行虚拟筛选可以有效地提高虚拟筛选的成功率。



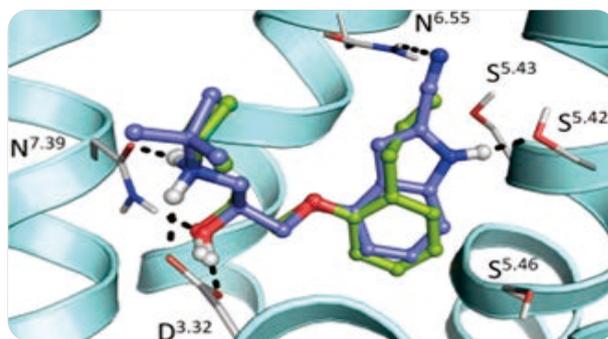
### ● 基于蛋白-配体分子作用指纹的虚拟筛选

蛋白-配体分子相互作用指纹 (Protein-Ligand Interaction Fingerprint, PLIF 或 IFP) 是一种数值化方法, 用于记录和描述蛋白质与配体之间不同类型的相互作用。该方法利用二进制编码的思路, 将复杂的分子间相互作用转化为简单的一维指纹形式。利用蛋白-配体分子作用指纹策略可以有效区分 GPCR 激动剂和拮抗剂。

#### Agonist: 2Y02 (chain A)



#### Antagonist: 2VT4 (chain A)



Ligand	Structure	D3.32	S5.42	S5.43	S5.46	N6.55	N7.39
Carmoterol	2VT4 chainA	■	■			■	■
Procaterol	2VT4 chainA		■				■
Cyanopindolol	2VT4 chainA	■	■		■	■	■
Alprenolol	2VT4 chainA	■			■	■	■

## 虚拟筛选合作成功案例

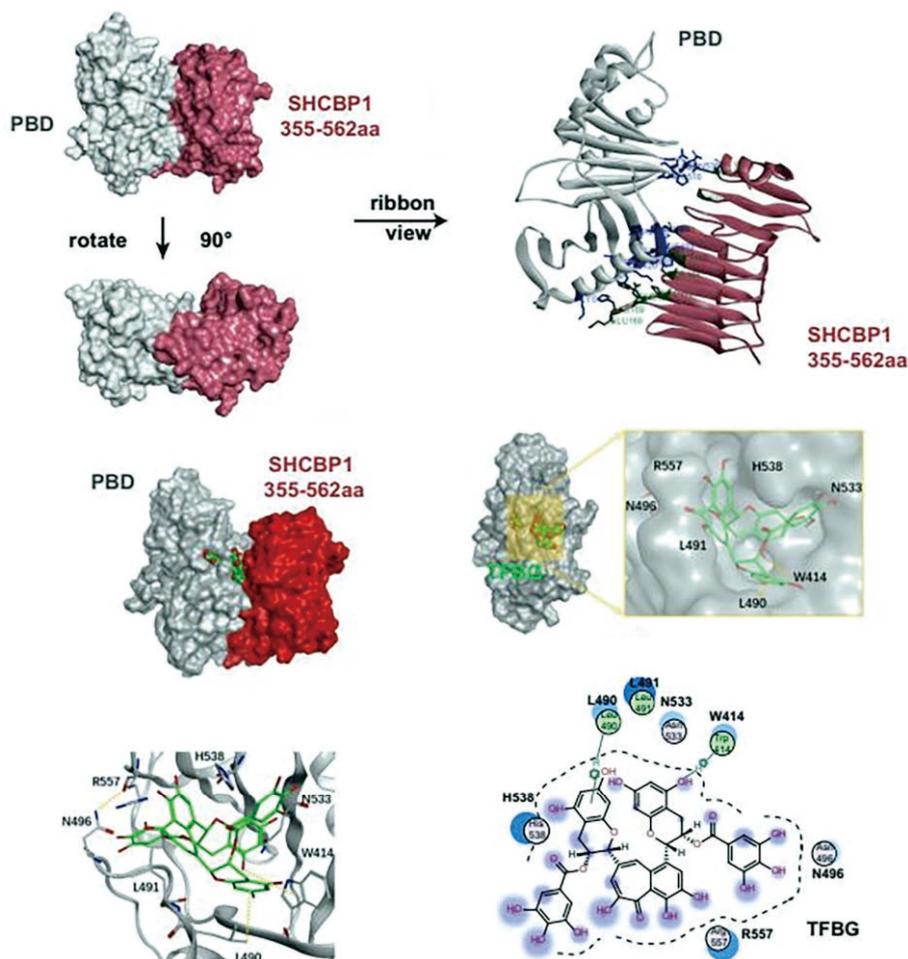
Hyperactivation of HER2-SHCBP1-PLK1 axis promotes tumor cell mitosis and impairs trastuzumab sensitivity to gastric cancer. Nat Commun. 2021 May 14;12(1):2812.

IF= 16.6

利用曲妥珠单抗治疗胃癌的过程中，HER2-SHCBP1-PLK1 轴的激活被认为是一个关键的促进肿瘤发生和耐药的因素。为了深入理解并克服这一挑战，研究团队通过结构域截短实验和免疫共沉淀，发现 SHCBP1 的 355-562 氨基酸区域是与 PLK1 结合所必需的，并发现与 SHCBP1 结合的是 PLK1 的 polo-box 结构域 (PBD) 而不是激酶结构域。

为了探索潜在的治疗策略，研究团队在陶术生物的协助下，利用同源建模和分子对接技术，得到了 SHCBP1 和 PLK1 的结合模式。并通过设计 PLK1 的蛋白质缺失突变和氨基酸突变实验进一步验证了这一结合模式。

基于这些结构信息，研究人员在陶术生物的协助下进行了虚拟筛选，成功发现 TFBG 作为一种潜在的抑制剂，通过抑制 SHCBP1 和 PLK1 的相互作用，展示了抗癌的功能。实验数据表明，TFBG 的 Kd 值低至 0.467  $\mu\text{M}$ ，对 SNU-216 胃癌细胞的增殖有明显的抑制作用 (IC<sub>50</sub>=23.42  $\mu\text{mol/ml}$ )。

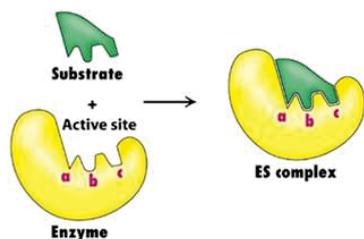


【命中分子 TFBG 与 SHCBP1-PLK1 复合物的结合模式及相互作用】

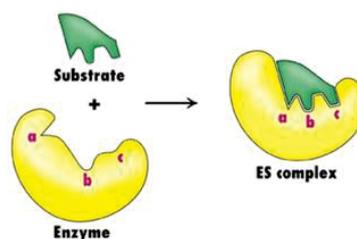
## 2、分子对接

分子对接是通过受体的特征以及受体和药物分子之间的相互作用方式来进行药物设计的方法。它主要研究分子间(如配体和受体)的相互作用,并预测其结合模式和亲和力。

### ➤ 分子对接理论基础



【锁-钥模型】



【诱导契合模型】

#### 1) 空间匹配:

空间匹配是分子间发生相互作用的基础。只有当分子间的几何形状相互匹配时,它们才有可能在物理上接近并发生进一步的相互作用。

#### 2) 能量匹配:

能量匹配是在空间匹配的基础上进行的进一步评估。它关注的是分子间相互作用的能量变化,特别是结合自由能的变化。

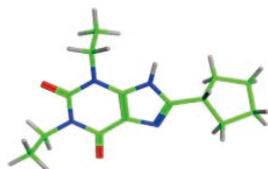
### ➤ 分子对接的一般流程

#### 01 靶标蛋白结构的选择与预处理



受体模型

#### 02 配体分子结构预处理



#### 03 分子对接过程



构象采样

- > 片段生长法
- > 系统搜索法
- > 遗传算法
- > 蒙特卡洛模拟算法

#### 04 分子对接结果评价

能量评价(打分函数)

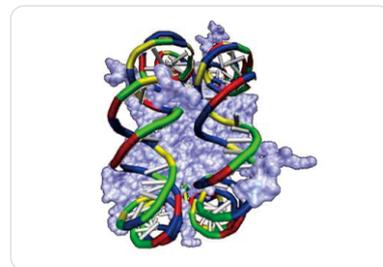
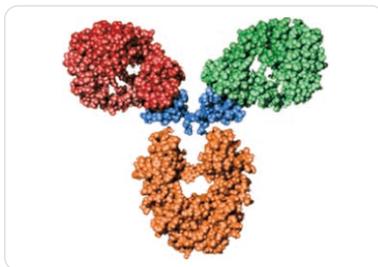
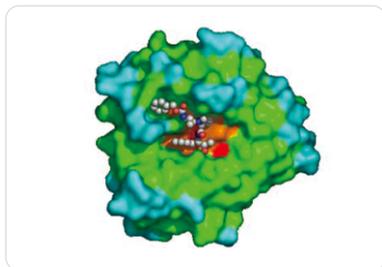
- > 经验型方法
- > 基于分子力场
- > 基于知识
- > 机器学习/深度学习

## ➤ 分子对接方法的分类

按照分子对接构象采样的原理可分为 **刚性对接**、**半柔性对接** 和 **柔性对接**。

对接类型	方法特点	计算量	应用
刚性对接	参与对接的分子构象不发生变化, 仅改变分子的空间位置与姿态	简化程度最高 计算量相对较小	适合于生物大分子之间的对接
半柔性对接	允许对接过程中小分子构象发生一定程度的变化, 但通常固定大分子的构象, 另外小分子构象的调整也可能受到一定程度的限制, 如固定某些非关键部位的键长、键角等	兼顾计算量与模型的预测能力	应用比较广泛
柔性对接	允许研究体系的构象发生自由变化	计算量非常大 消耗机时很多	适合精确考察分子间的识别

按照分子对接的对象类型可以分为 **蛋白-小分子对接**、**蛋白-蛋白对接** 和 **蛋白-核酸对接**。



## ➤ 分子对接案例

### 1) 蛋白-蛋白对接

蛋白-蛋白对接的常见策略主要有两种：

- **模板基础构建：**

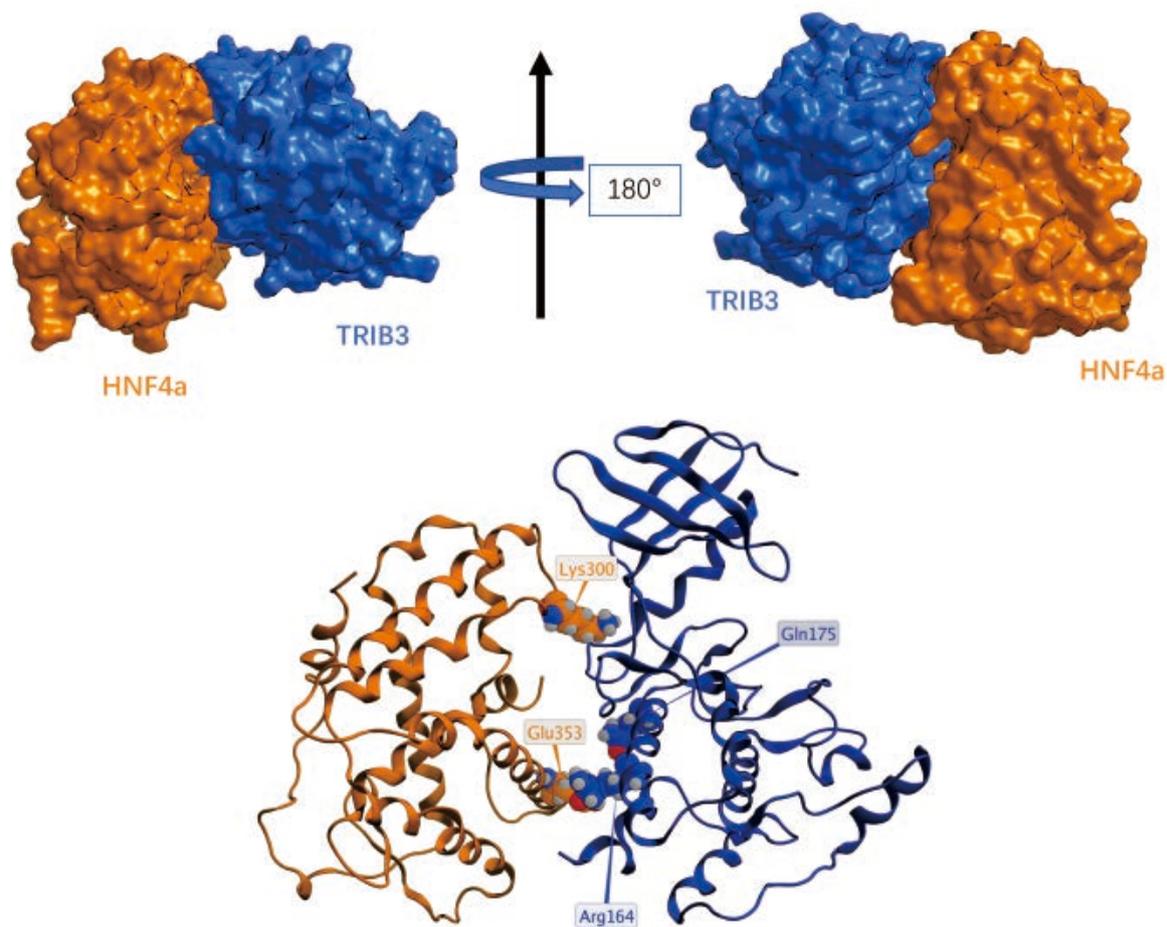
将已有的同源蛋白结构作为模板构建目标蛋白复合物, 随后对得到的结构进行优化, 以提高其结构的稳定性和可靠性。

- **构象采样与评分：**

固定一个蛋白, 对另外一个蛋白进行充分地构象采样, 对得到的构象采用打分函数进行排序, 通常选择得分较低(能量较低)的构象作为最优解。

📖 TRIB3-TRIM8 complex drives NAFLD progression by regulating HNF4α stability, Journal of Hepatology, 2024, ISSN 0168-8278.

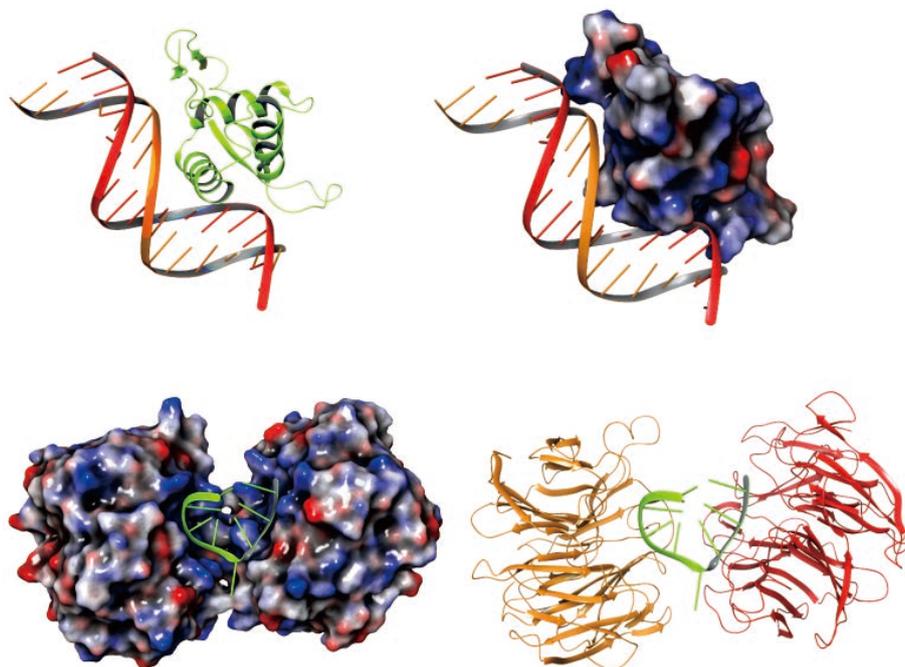
IF 25.7



【蛋白-蛋白对接案例：HNF4α-TRIB3 结合模式展示】

## 2) 蛋白-核酸对接

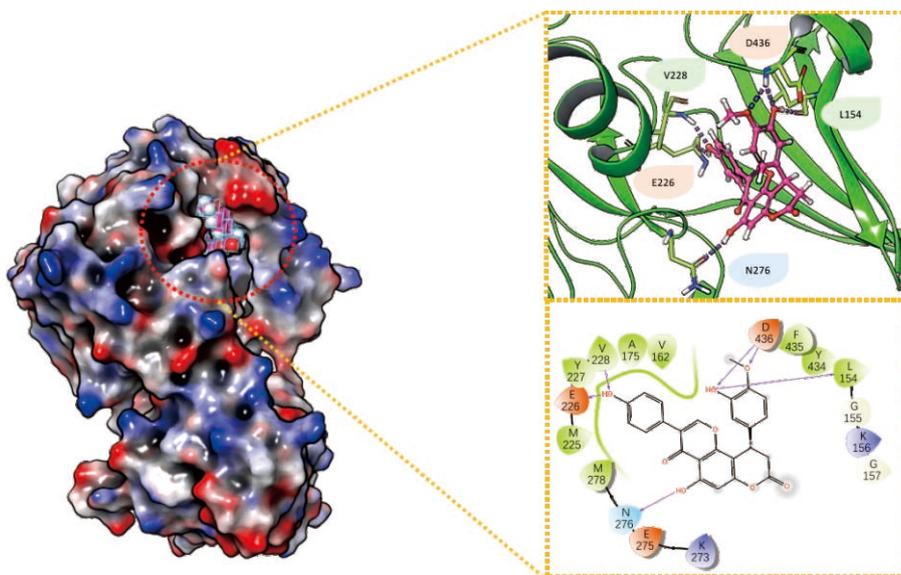
核酸相较于蛋白质具有更高的柔性和更强的极性。现有的蛋白-核酸复合物的结构数据远远少于蛋白-蛋白复合物，限制了可供参考的信息。因此，蛋白-核酸对接通常比蛋白-蛋白对接更具挑战性。目前常见的方法是采用打分函数计算核酸和蛋白之间的亲合性，从而得到蛋白-核酸的低能稳定构象。



【蛋白-核酸结合模式展示样图】

### 3) 蛋白-小分子对接

蛋白-小分子对接通常采用半柔性对接的方法。在对接过程中，蛋白通常固定不动，对小分子进行构象搜索以探索不同的结合可能性。通过打分函数对每种可能的构象进行评估和排序，选择最优的结合构象进行分析展示。由于小分子整体较小，需要考虑的因素相对较少，这使得蛋白-小分子对接在药物设计中得到广泛的应用。



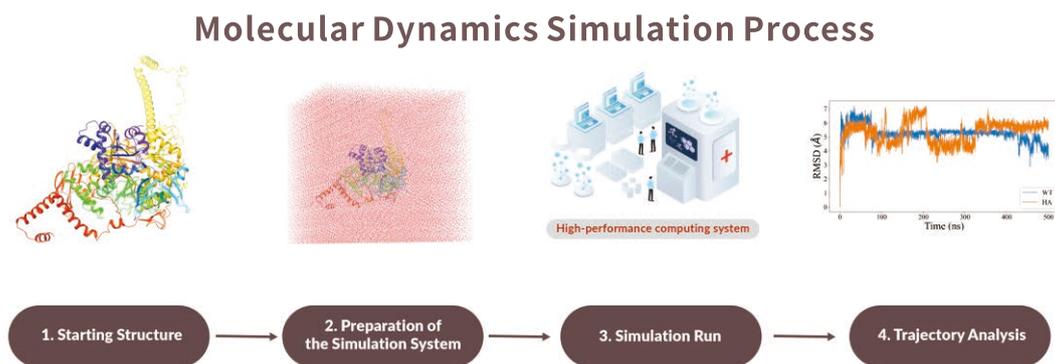
【蛋白-小分子对接案例：小分子配体结合模式展示】

### 3、分子动力学模拟

分子动力学模拟是一种计算模拟方法，用于通过模拟分子和原子的运动来研究物质的性质。它使研究人员能够以原子级分辨率观察分子在固定时间内的动态行为，从而揭示分子系统的结构、动力学和热力学性质。

在分子动力学模拟中，系统中每个分子的运动都是通过解牛顿运动方程来预测的，这些定律描述了分子之间的相互作用及其对分子运动的影响。通过这种方式，可以模拟不同环境条件下（如不同温度、压力或化学环境）分子系统的行为。

分子动力学模拟广泛应用于物理、化学、生物学和材料科学等领域。通过这种方法，科学家能够研究蛋白质折叠、药物分子与靶标的相互作用、纳米材料的性质以及许多其他复杂系统的微观机制。随着计算技术的发展，对更大系统和更长时间尺度的模拟成为可能，分子动力学模拟在科学研究中的应用将越来越广泛。



#### 可分析参数

##### 1) 骨架波动：

RMSD、RMSF

##### 2) 相互作用分析：

H-bonds、Ionic、hydrophobic、Salt bridges、Contact-Map

##### 3) 热点残基：

Alanine-scanning、Energy-decomposition

##### 4) 结合自由能：

MM/GB(PB)SA、TI、FEP

##### 5) 构象转变：

PCA、Energy-landscape plotting

##### 6) 理化性质：

Energy、volume、pressure、temperature、density monitoring

##### 7) 构象采样：

cluster analysis、dominant conformation identification

## ➤ 陶术生物分子动力学模拟业务

分子动力学模拟在生物和化学体系中有许多常见的应用, 包括:

### 1) 蛋白质折叠和结构稳定性研究

可以模拟蛋白质在原子水平上的运动和结构变化, 以理解其如何在不同条件下折叠成特定的结构, 以及结构的稳定性和动态性。

### 2) 酶的反应机理和催化机理

可以模拟酶与底物之间的相互作用, 研究酶催化反应的详细机制, 如过渡态的形成和解离。

### 3) 药物设计和分子对接

可以通过模拟分子与蛋白质靶标或受体的结合, 预测药物分子的结合模式和亲合力, 从而指导药物设计和优化。

### 4) 生物膜的结构和动态性

可以模拟生物膜的构建过程以及膜蛋白在膜上的结构和运动, 研究其在细胞内的功能和作用。

### 5) 溶剂效应和溶液化学

可以模拟溶剂对溶质 (如离子、小分子) 的包合和影响, 研究其在溶液中的行为和相互作用。

陶术生物分子动力学模拟业务致力于提供高效、精准分子动力学模拟服务, 利用先进的 GPU 加速技术, 我们能够快速模拟复杂的化学生物体系, 为客户提供最前沿的技术支持和专业的分析服务, 为客户的研究和开发提供强有力的支持, 帮助客户在分子动力学领域取得突破。

我们的分子动力学模拟业务包括但不限于以下内容:

**构象变化分析:** 精准模拟分子结构的动态变化, 揭示其在不同条件下的构象变化。

**相互作用分析:** 深入分析分子之间的相互作用, 帮助您理解分子机制及其功能。

**热点残基分析:** 识别和分析分子中的关键残基, 提供靶点药物设计的关键数据。

**结合自由能计算:** 准确计算分子间的结合自由能, 为药物筛选和优化提供理论依据。

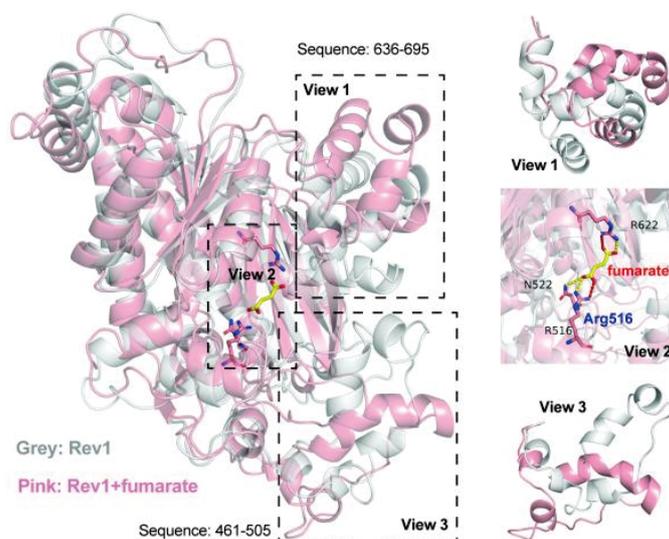
## 分子动力学成功合作案例

Tyrosine catabolism enhances genotoxic chemotherapy by suppressing translesion DNA synthesis in epithelial ovarian cancer. *Cell Metab.* 2023 Nov 7;35(11):2044-2059.e8.

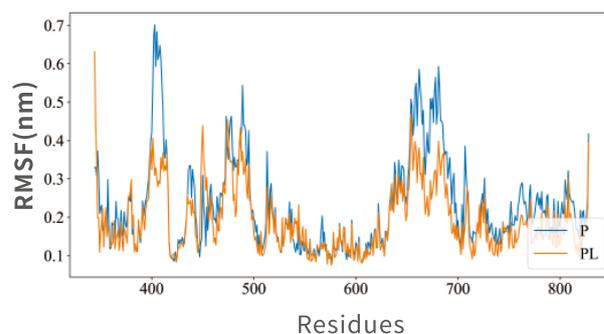
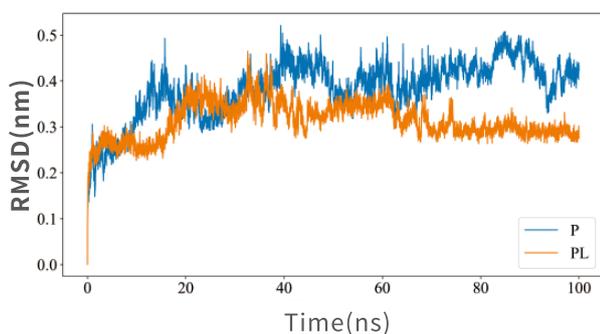
IF 29.0

顺铂等基因毒性化疗药物可通过干扰 DNA 损伤应答 (DDR) 机制阻碍 DNA 复制和细胞分裂, 引发肿瘤细胞死亡。然而, 患者易对化疗药物产生耐药性, 因此了解患者对基因毒性化疗药物的敏感性有助于优化治疗策略。

研究团队以卵巢癌为疾病模型, 发现酪氨酸代谢在化疗药物造成的 DNA 损伤中的作用。富马酸乙酰乙酸水解酶 (FAH) 在酪氨酸分解通路中催化最后一步反应, 影响了 REV1 和 REV7 之间的相互作用, 敲低 FAH 显著降低化疗引起的 DNA 损伤与细胞凋亡。研究团队通过分析 TCGA 数据库和卵巢癌患者样本, 发现肿瘤细胞内 Fumarate 的表达水平与化疗敏感性密切相关。



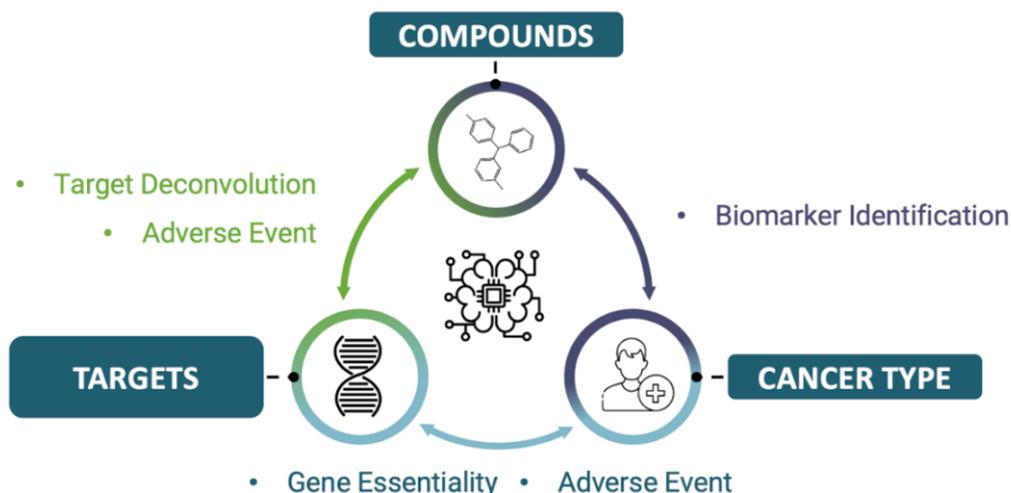
通过使用分子对接和分子动力学模拟的方法, 分析了 Fumarate 和 REV1 的结合模式和稳定性, 揭示了 FAH 代谢产物 Fumarate 在核内与 REV1 结合, 从而抑制跨损伤 DNA 合成的新机制。



【Fumarate 和 REV1 的模拟轨迹分析结果】

## 4、反向找靶

现代药物研发过程遵循“疾病-靶点-药物”的逻辑，靶标是药物与疾病之间的关键连接枢纽。识别和验证有效的疾病靶标可以显著提高药物研发的成功率，降低开发成本。传统的靶点发现方法往往费时费力，而且可能无法发现与疾病机制直接相关的关键靶标。为解决这些问题，反向找靶 (Target Fishing) 技术应运而生。



【抗肿瘤药物研发中的“疾病-靶点-药物”逻辑】

反向找靶技术通过分析化合物与已知蛋白的相互作用，预测化合物的潜在作用靶点。这项技术可以显著减少寻找靶标所需的实验数量，加快药物研发的进程。此外，它还有助于药物的再利用，并能用于构建活性结构字典，为药物设计和开发提供重要支持。

01

靶点解析



传统中药活性成分



表型筛选活性化合物

02

药物再利用



新适应症、老药新用

03

活性结构字典



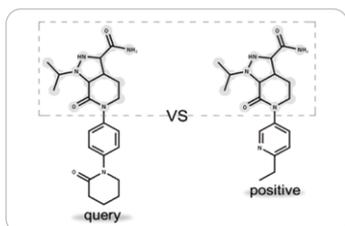
结构对照、辅助分子设计

【反向找靶技术常见用途】

## 反向找靶常见的方法和策略

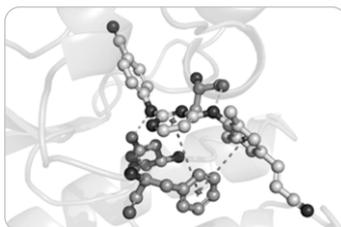
- ▶ **基于配体结构的算法**: 通过比对给定化合物与已知活性化合物的化学结构来预测其作用靶标, 例如: Swiss TargetPrediction, SEA
- ▶ **基于靶标结构的算法**: 通过分子对接等手段来分析给定化合物与靶标蛋白之间的相互作用, 进而预测其作用靶标。例如: TarFisDock
- ▶ **基于人工智能的算法**: 运用机器学习或深度学习来建立已知活性化合物及其作用靶标之间的关联, 据此预测给定化合物的作用靶标。例如: PPB2, DrugBan

### 1 基于配体结构的算法



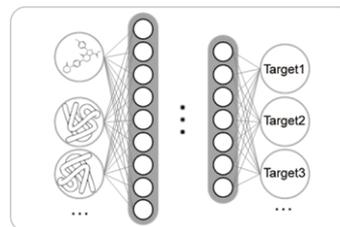
- 【优】快(数分钟), 准确度良好
- 【缺】不考虑活性悬崖

### 2 基于靶标结构的算法



- 【优】预测结果直观
- 【缺】慢(小时或天), 表现不稳定

### 3 基于人工智能的算法



- 【优】极快(数秒), 全靶点可预测
- 【缺】可解释性低, 表现不稳定

## 陶术生物自研算法的优势

复旦大学王任小研究员(兼任上海陶术 CTO)团队开发的多元组合式反向找靶算法 COMET, 集成了基于配体结构、基于靶点结构和基于人工智能的算法, 在精心挑选的人类靶标组中预测给定化合物的作用靶标。该算法有以下主要优势:



### 扎实的知识基础

COMET 算法的知识基础包括有明确疾病-靶点-药物关系的 2600 多个靶标的信息、靶标结构上 4.5 万个可能的结合部位以及约 100 万药物-靶标相互作用对数据。



### 预测准确性实现 SOTA

在多个大规模测试集上进行了检验: 与目前流行的其他算法相比, 预测精度稳定居第一, 且明显领先其他方法。例如, COMET 在 TOP15 靶标中至少包含一个真实靶标的成功率接近 80%。



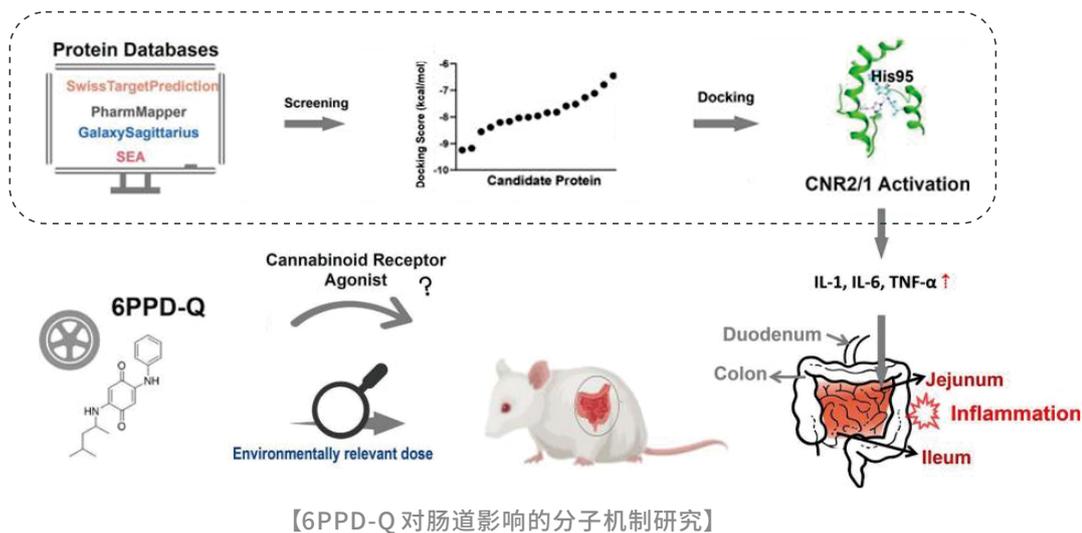
### 计算效率高, 方便易用

用户可以在网页上方便地输入给定化合物的结构, 并在较短的时间内获得详尽的预测结果。对于一个给定化合物, 完成对靶标组的扫描并给出预测结果, 平均耗时仅 10 分钟。

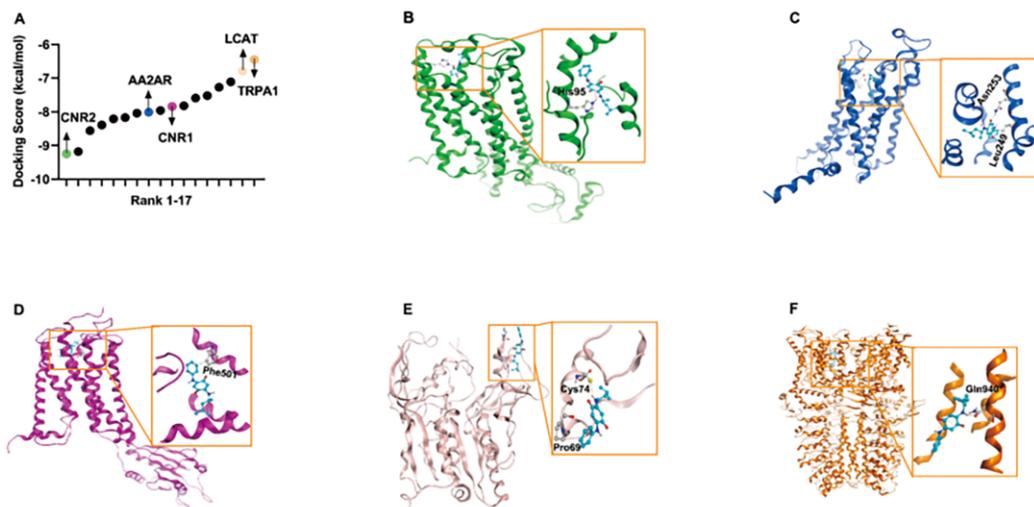
## 反向找靶成功案例

Environmentally realistic dose of tire-derived metabolite 6PPD-Q exposure causes intestinal jejunum and ileum damage in mice via cannabinoid receptor-activated inflammation. *Sci Total Environ.* 2024 Mar 25;918:170679. IF 9.8

N-(1,3-二甲基丁基)-1-苯基-苯二胺醌 (6PPD-Q) 是一种常见的轮胎添加剂 6PPD 的醌衍生物，在环境和人体中的存在被广泛报道。在该研究中，通过给成年 ICR 小鼠连续 21 天口服不同剂量的 6PPD-Q，评估 6PPD-Q 对肠道的影响。结果显示，6PPD-Q 对肠道屏障的破坏呈剂量依赖性，主要影响空肠和回肠，高剂量时还引起肠道炎症。



基于这些发现，陶术生物团队利用多种反向找靶技术，识别了若干最有可能的靶点，如 CNR2、AA2AR、CNR1、LCAT、TRPA1 等，并通过分子对接预测了其结合模式。通过分子动力学模拟，进一步验证了 6PPD-Q 和 CNR2 的结合稳定性及构象变化。



【6PPD-Q 与预测靶点的结合模式展示】

# CADD 筛选方案

Computer-Aided Drug Design

## 目前主要提供的 CADD 筛选方案

筛选方案	筛选周期
针对中药单体库 (L6810) 虚拟筛选赠送 53 个化合物 (10 $\mu$ L)	1-2 工作周
针对天然产物库 (L6000) 虚拟筛选赠送 53 个化合物 (10 $\mu$ L)	1-2 工作周
针对已知活性库 (L4000) 虚拟筛选赠送 53 个化合物 (10 $\mu$ L)	2-3 工作周
针对现货库 (T001 和 LF1000ava) 虚拟筛选赠送 53 个化合物 (10 $\mu$ L)	2-3 工作周
针对 40 万个化合物的骨架库 (EN1200) 虚拟筛选	2-3 工作周
针对 160 万个化合物的 D001 库虚拟筛选	2-3 工作周
针对 400 万个化合物的 D003 库虚拟筛选	2-4 工作周

除了上述提到的筛选方案, 客户也可以选择任意数量、规模和类型的化合物进行筛选, 详情可咨询陶术生物的工作人员。

# CADD 陶术生物优势

Computer-Aided Drug Design

陶术生物在药物发现和开发领域提供了高效且成本效益极佳的解决方案。其中, CADD 技术服务具有以下优势:

-  **专业团队领导:** 由复旦大学特聘研究员王任小亲自领导的专业团队。
-  **权威软件支持:** 拥有包括 MOE、StarDrop 在内的多款权威软件版权。
-  **海量化合物库选择:** 2600 万种化合物的数据库, 60 亿种快速定制化合物库, 上千种定制化合物库
-  **全面服务体验:** 免费的售前评估、实验过程中的沟通支持以及售后问题解答。
-  **一站式服务:** 提供从虚拟筛选到实体化合物供应再到验证实验的一站式服务。
-  **极具竞争力的价格:** 通过筛选化合物套餐大幅降低筛选成本, 提供性价比极高的服务。

# 人工智能驱动的药物发现

TOPIC | 内容

PAGE | 页码

技术概述

18

超大规模虚筛 AI 算法

19

蛋白/多肽药物设计

21

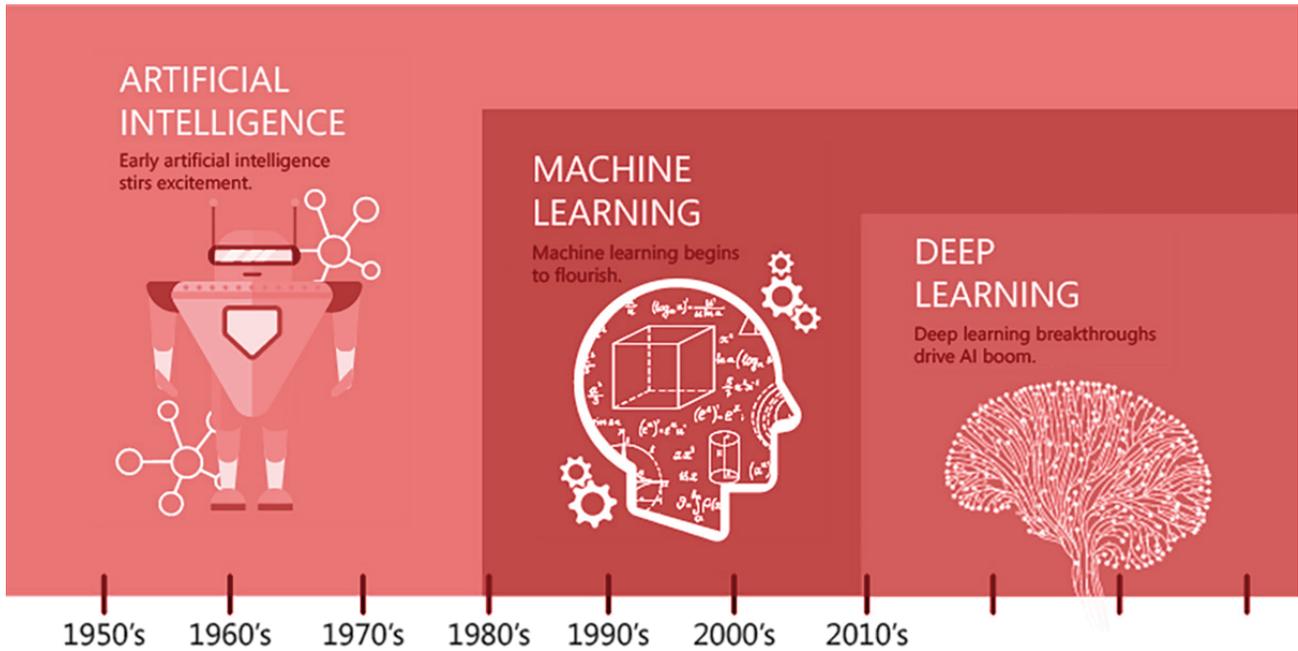


CADD & AIDD

# OVERVIEW

## 技术概述

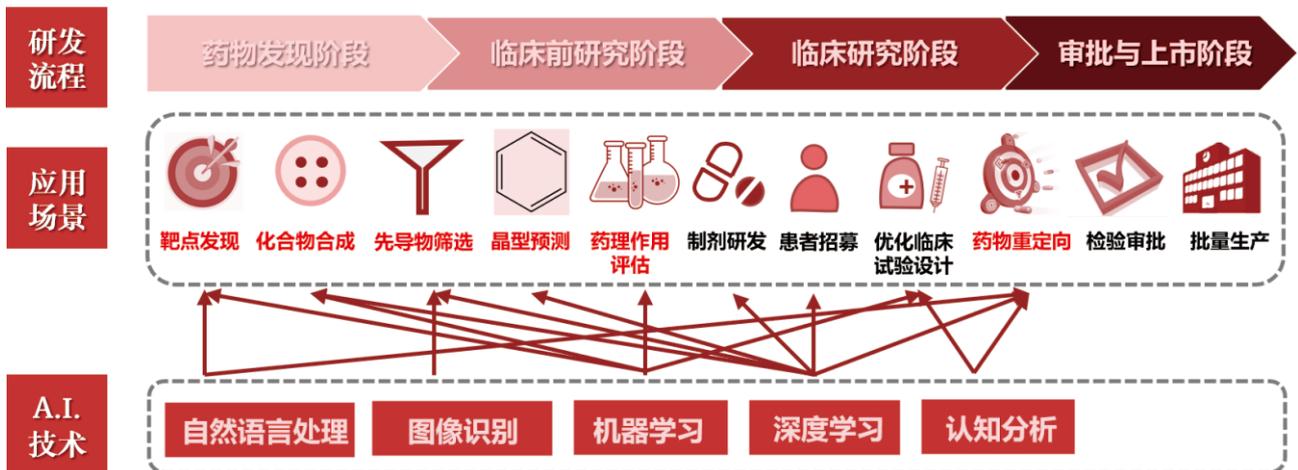
### Artificial Intelligence in Drug Discovery



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.

【人工智能技术发展史】

AIDD (人工智能驱动的药物发现) 利用 AI 技术来提高药物发现和设计过程的效率和成功率。这一领域结合了机器学习、数据挖掘和模拟技术, 以应对传统药物设计过程中的高成本、长周期和低成功率等挑战。随着机器学习和数据科学技术的发展, AIDD 逐渐成为药物发现领域的重要工具之一。

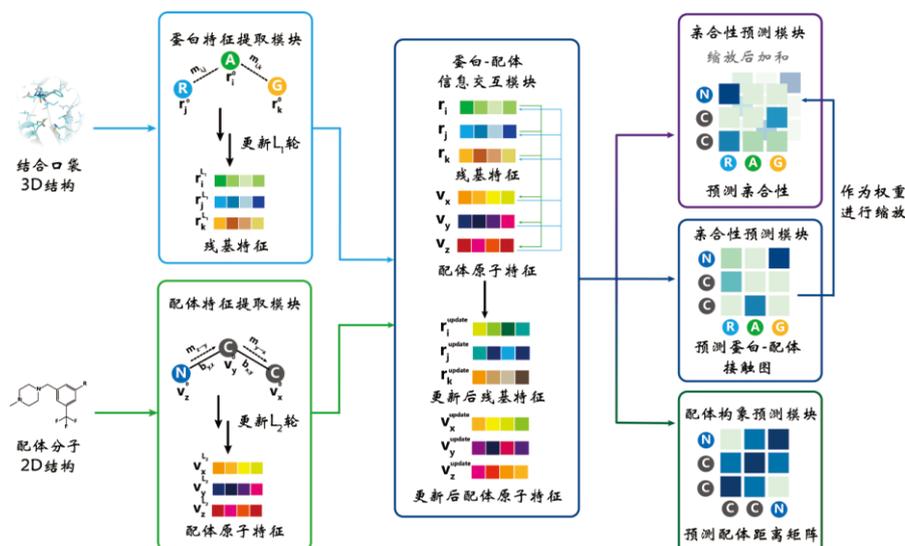


【AI 技术在药物研发中的应用】

# 实现超大规模虚拟筛选的 AI 算法

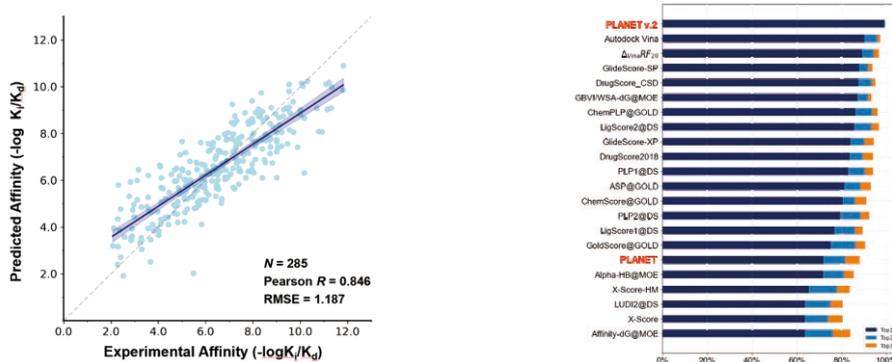
Artificial Intelligence in Drug Discovery

复旦大学王任小研究员（兼任上海陶术 CTO）团队开发了名为 PLANET（Protein-Ligand Affinity prediction NETwork）的 AI 算法，可用于执行超大规模虚拟筛选，目前已经发展到第二代，在保持了预测精度的同时大幅提升了计算效率。该算法目前授权给陶术生物 CADD 部门使用，为从事药物发现的广大用户提供技术服务。



【 PLANET 算法的基本架构设计 】

PLANET 模型的特点是无须依赖复合物三维结构进行计算，而是根据配体分子的二维结构和靶标蛋白结合部位直接预测靶标-配体亲合性，因此跳过了繁复的分子对接构象采样过程，在虚拟筛选任务中比传统的分子对接方法提高效率 1000 倍以上，因此可在常规硬件平台上实现 10 亿级化合物库的超大规模虚拟筛选。此外，PLANET 是一个多任务模型，除了给出靶标亲合性的预测数值，还能给出配体分子结合模式的信息，提供预测结果的可解释性。

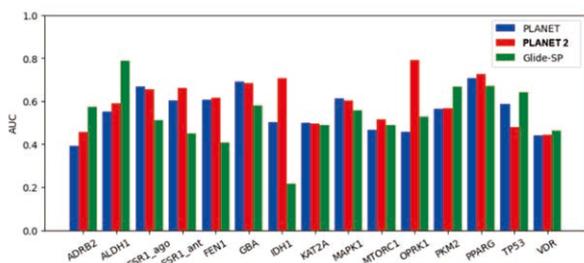


【 PLANET 二代算法在 CASF-2016 测试集上的测试结果 】

（左）靶标亲合性预测值和实验值的相关系数达到 0.846

（右）识别配体分子结合构象的成功率在 95% 以上。

在领域内若干标准测试集上测试的结果显示:对于定量预测靶标亲合性(即打分能力), PLANET 二代算法的预测精度位于目前各种深度学习打分函数中的先进水平;对于识别配体分子的正确结合构象(即对接能力)以及虚拟筛选中识别活性化合物(即筛选能力), PLANET模型的预测精度至少相当于最好的传统分子对接方法(如 AutoDock Vina 和 GLIDE)。



在 LIT-PCBA 测试集 15 个靶标上获得的虚拟筛选结果显示: PLANET 二代模型的平均 AUC 值为 0.599, 相比于 PLANET 一代模型( AUC=0.556)和 GLIDE ( AUC=0.536) 均有所提高。

## 陶术化合物库虚拟筛选的典型方案

为了满足不同规模和速度的需求, 陶术生物基于 PLANET 算法提供了以下三种化合物库筛选方案:

化合物库	化合物数量	筛选周期
REAL Database	60 亿	12~16 工作周
REAL Diversity Set	4820 万	6~8 工作周
TOPSCIENCE Database	2600 万	4~6 工作周

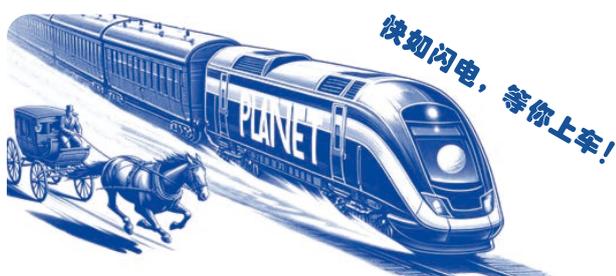
**[NOTE]** The current release of the REAL database comprises over 6B molecules which comply with Ro5 and Veber criteria, i.e. MW $\leq$ 500, SlogP $\leq$ 5, HBA $\leq$ 10, HBD $\leq$ 5, RotBonds $\leq$ 10, and TPSA $\leq$ 140;

The REAL Diversity Set compounds comply with the Ro5 and Veber criteria, i.e. MW $\leq$ 500, SlogP $\leq$ 5, HBA $\leq$ 10, HBD $\leq$ 5, RotBonds $\leq$ 10, and TPSA $\leq$ 140 and eliminate PAINS and toxicophores;

The TOPSCIENCE database has been updated to increase its size by 40%. Samples are available for purchase for 90% of the compounds in catalog.

## 陶术生物虚拟筛选方案的优势

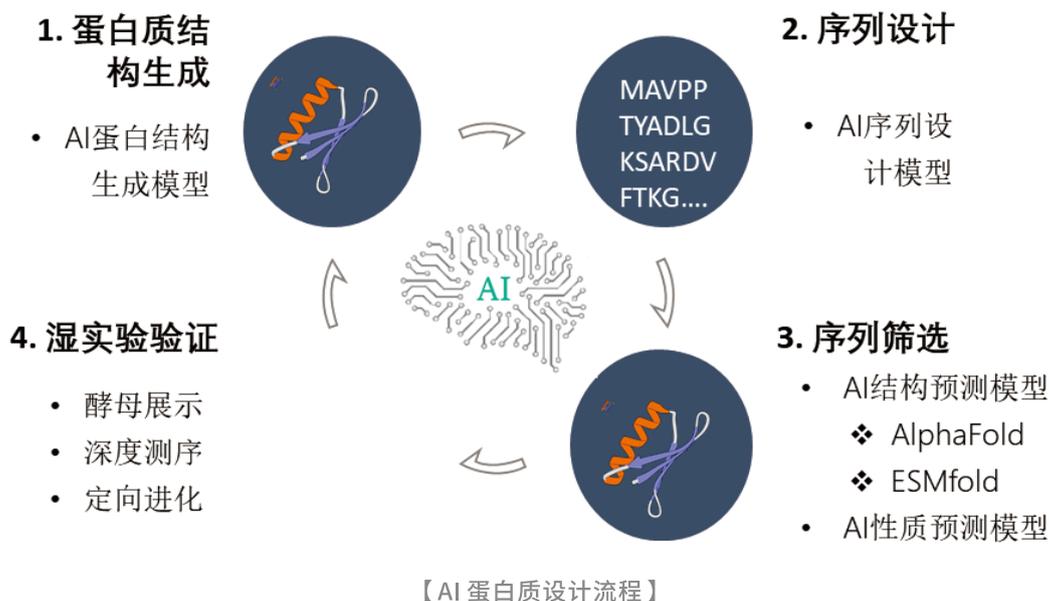
通过虚拟筛选发现活性化合物宛如沙里淘金, 淘的量越大收获就越多。上海陶术借助自研的高性能AI算法, 以极高的性价比实现超大规模虚拟筛选, 显著扩大了筛选范围, 提升发现活性化合物的几率!



# 蛋白/多肽药物设计

Artificial Intelligence in Drug Discovery

蛋白质药物是利用天然或重组蛋白质用于治疗、诊断或预防疾病的药物，由于蛋白质具有高度特异性和生物活性，它们在多种疾病的治疗中发挥了重要作用，主要包括单克隆抗体、激素和生长因子、酶、疫苗、融合蛋白和抗体药物偶联物等，用于癌症治疗、自身免疫疾病、感染性疾病等。多肽药物则是由氨基酸通过肽键连接形成的短链分子，通常包含 50 个或更少的氨基酸，具有特异性高、毒性低和副作用少的特点，主要包括天然多肽、合成多肽、修饰多肽等，用于代谢疾病、癌症治疗、心血管疾病等。未来，随着生物技术、工程技术、多肽修饰技术和药物递送系统的发展，蛋白质和多肽药物的设计和生产将更加高效，应用范围也将不断扩大，为更多患者提供有效的治疗选择。



传统蛋白质计算设计方法主要基于能量函数，通过搜索优化算法，寻找给定蛋白质骨架上能量最低的氨基酸序列。由于蛋白质序列组合空间巨大（对于一个 100 个氨基酸的蛋白质，可能的序列组合达到  $20^{100}$ ），通过优化算法难以找到全局能量最低点，并且能量函数准确度较低，造成传统计算蛋白质设计方法成功率较低。同时传统设计方法还面临蛋白质骨架生成的难题。

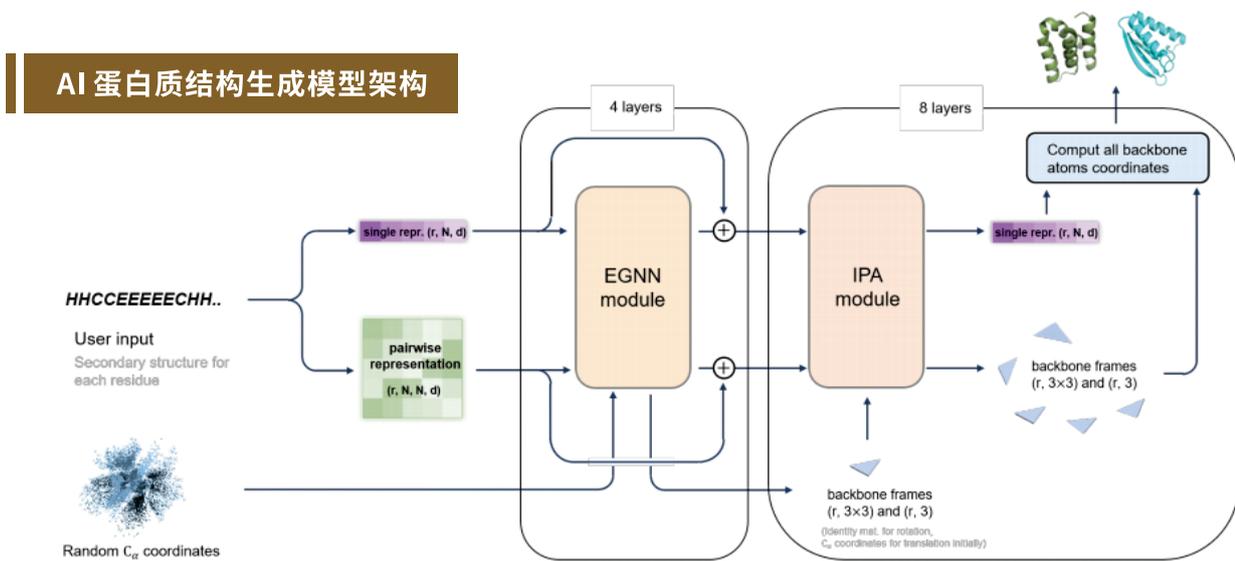
AI 蛋白质设计方法从全流程上实现了对传统设计方法的超越。在结构生成上，通过生成模型，可以从无到有产生全新的蛋白质三维骨架用于序列设计。在序列设计上，通过基于图网络的序列设计模型和蛋白质序列大语言模型，可以设计出具有很好可折叠性的序列。在序列筛选上，AlphaFold 等蛋白质结构预测模型以及稳定性、可溶性等性质预测模型可以准确、高效地对设计候选序列进行挑选。AI 方法的应用显著提高了蛋白质设计的效率，设计成功率相比传统方法提升一个数量级以上。

## 陶术生物蛋白/多肽药物设计优势

陶术生物和合作伙伴共同开发了蛋白质结构生成和序列设计 AI 模型。通过设计可以得到全新的结合特定靶标的蛋白和多肽,对蛋白质的物理化学性质进行优化(如热稳定性、可溶性等),调控蛋白的生理活性和蛋白-蛋白之间的相互作用强度。更进一步,这些技术可用于设计新的蛋白或多肽抑制剂和激动剂,为药物开发提供新的策略和工具。

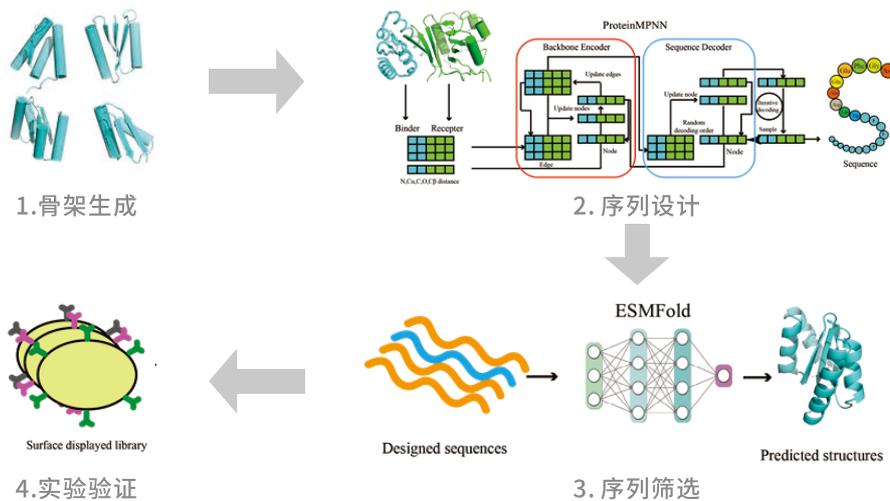
陶术生物在蛋白/多肽药物设计领域具有显著优势:

- 1) 独家专有算法:** 利用自主开发的算法,提供独特的解决方案,增强药物候选物的性能和应用潜力。
- 2) 先进的方法:** 采用行业领先的技术和最新的科研成果,确保药物设计的效率和创新性;



## 陶术生物蛋白/多肽药物设计成功案例

针对某跨血脑屏障靶标,通过AI蛋白质设计流程,成功获得数十个亲合力为nM级别的小蛋白,用于药物递送载体研发。



# 数据库 (Database)

---

TOPIC | 内容

PAGE | 页码

TOPSCIENCE Database

24

...

REAL Database

27

...

PDBbind Database

28

...



# TOPSCIENCE Database

## Artificial Intelligence in Drug Discovery

TOPSCIENCE Database 是由陶术生物团队整理和维护的开源数据库，收录了全球数百家供应商的小分子化合物产品数据，包含千万级的化合物结构，致力于为全球科研人员提供一个稳定可靠的数据库。

TOPSCIENCE Database 是一个综合性的化合物资源库，分为四个主要部分：筛选化合物数据库、天然产物类数据库、活性化合物数据库和片段化合物数据库。每个部分针对不同的科研需求，提供多样化的化合物选择。

化合物库优势	详细介绍
定期维护数据	确保所提供的化合物信息准确、更新及时。
规模庞大	拥有千万级的化合物规模，覆盖广泛的化学空间。
90%+ 现货供应	大多数化合物可以快速供应，筛选无后顾之忧。
完善的供应链	保障化合物的快速和安全交付。
成熟的报关流程	确保国际订单顺利通过海关。
货期较短	3-5周，从订单到交付时间短，加快研究进度。

## 筛选化合物数据库

筛选化合物数据库，可以进一步细分为三类：精选数据库、核心数据库、扩展数据库。

### ➤ 精选数据库

品牌编号	数量	特征
D006	1,634,293	源自美国的化合物品牌，总部位于圣地亚哥，拥有多样性库、大环库等多种热门化合物库。
D011	539,766	源自加拿大的化合物品牌，拥有2900多种化合物骨架结构，化合物规格较齐全且有对应价格。
D008	202,539	源自荷兰的化合物品牌，价格优势明显。
D009	1,412,794	源自美国的化合物品牌，在香港拥有发货中心，到货速度快，价格适中。
D019	44,157	源自德国的天然产物品牌，专注天然产物提取及类似物合成工作，产品质量稳定，价格较贵。
D003	4,076,937	源自乌克兰的化合物品牌，化合物研发能力较强，包含5个系列化合物，价格较贵。
D010	485,964	源自俄罗斯的化合物品牌，拥有多种天然产物及衍生物，价格偏贵。
D012	53,344	源自英国的化合物品牌，产品数量少而专，每种产品均有较大库存。

## 核心数据库

品牌编号	数量	特征
T001	17681	优势产品为已知活性化合物和天然产物,产品质量高,服务态度好,性价比高。
D001	1648191	全球最大的化合物品牌之一,拥有5000多种化合物骨架结构和100多种化合物库,性价比高。

## 扩展数据库

品牌编号	数量	特征
D020	102,619	源自英国的品牌,拥有 20 多年的有机合成经验。
D058	797,788	源自俄罗斯的化合物品牌,专业从事新型筛选化合物的生产与销售。
D104	26,848	源自俄罗斯的化合物品牌,拥有 20 多年的化合物供应经验,库存量充足,性价比较高。
D065	575,302	源自美国的品牌,20 多年来致力于类先导化合物及分子砌块的研发供应,价格较贵。
D015	1,226,425	源自乌克兰的年轻且发展迅速的化合物品牌,产品主要用于高通量筛选和药物发现,价格较高。
D059	61,009	源自乌克兰的化合物品牌,近 20 年来致力于提供高通量筛选化合物及相关服务,有价格优势。
D021	1,440,338	源自美国的化合物品牌,20 多年来设计独特的小分子化合物用于药物开发。
D013	279,081	源自加拿大的化合物品牌,专门从事特色化合物,生物化学药品和生物分析试剂的开发和生产。
D137	5,007	源自英国的化合物品牌,具有 30 多年的化合物供应和服务经验。
D111	439,901	源自俄罗斯的化合物品牌,20 多年来专门从事高质量新型化合物的合成和供应,用于高通量筛选。
D133	247,261	源自中国的 CRO 公司,具有丰富的化合物合成经验。
D140	372,596	源自乌克兰的新颖化合物供应公司,拥有小型灵活的团队和超过 20 年的化学品合成供应经验。
D062	867,806	源自美国的化合物供应公司,在多样性化合物和天然产物供应方面有着丰富的经验。
D144	2,244,464	源自乌克兰的化合物品牌,拥有约 200 万全新的类先导化合物,这些分子具有重要的药效基团。

## 天然产物类数据库

---

天然产物一直是先导化合物的重要来源,然而目前许多天然产物数据库中的结构难以获得实体化合物。为此, TOPSCIENCE Database 推出了专门的天然产物单体数据库,包含 16,627 种来源于多种动植物和微生物的天然产物,结构多样,且有实体供应,解决了筛选出活性分子却难以购买到实体的问题。每种化合物都经过 HNMR 或 LCMS 检测,保障化合物的品质;这些化合物的结构以 SDF 格式文件开源提供,是天然产物虚拟筛选的有力工具;陶术生物的产品现货供应和快速到货服务进一步提高了研究效率和便利性。

## 活性化合物数据库

---

活性化合物能够调节细胞、组织或个体的生物学反应,涵盖了从实验室的工具化合物到临床试验中的候选药物,乃至已经获批上市的药物分子。这些化合物具备与特定生物靶标如受体、酶、核酸等特异性相互作用的能力,能够促进或抑制这些分子的功能。

TOPSCIENCE Database 收集了约 11.7 万种活性化合物,为药物研发和生物医学研究提供了宝贵的资源。

## 片段化合物数据库

---

在药物发现领域,片段筛选技术展现出了独特的优势。同时在化学生物学方面,使用片段分子能够帮助发现新的生物学作用机制。片段分子具备更好的化学空间代表性,使得基于这些片段分子进行的药物筛选和药物设计能够获得更理想的药物分子。

TOPSCIENCE Database 片段化合物数据库集合十余家片段分子供应品牌,收录 46 万种片段化合物信息,涵盖桥环、螺环、共价片段等多种结构类别。片段化合物数据库具有良好的结构多样性和新颖性,且大多可以提供实体产品,为片段药物发现提供更多助力。

# REAL Database

Artificial Intelligence in Drug Discovery

虚拟筛选的成功很大程度上依赖于化合物库的选择;只有当库中含有充足的化合物种类时,才能提高发现活性小分子的机会。在这方面, REALdatabase 是一个优良选择。该数据库包含高达 60 亿种化合物,提供了广泛的化学空间,覆盖多种结构、性质和功能,极大地提高了筛选的成功率和命中分子的多样性。

## REALdatabase 的优势

- 1) 真实可查询、可合成,合成时间在 3-4 周左右,平均成功率超过 80%;
- 2) 全世界最大的超快可搜索化学空间之一;
- 3) 筛出典型先导化合物的平均搜索时间为 2 至 4 分钟,并且可以通过骨架搜索功能探索无 IP 区域。

## REALdatabase 细分化合物库

以 REALdatabase 为基础,还可根据不同的研究需求,通过一系列药物化学过滤规则构建出不同的 Real 化合物库,包括 Real 多样性库,Real 先导类似物库,Real 片段库,Real 共价库,Real 天然产物类似物库等。Real 化合物库的规模相对较小,对于计算能力相对有限的客户也是一种不错的选择。

### 1) Real 多样性库

包含 2,500 万个化合物,该库剔除了 PAINS 以及具有毒性的化合物;化合物间不存在 Tanimoto 相似度超过 0.6 的情况,保证了筛选所必需的多样性。

### 2) Real 天然产物类似物库

包含 7,230 万个化合物,利用 P. Ertl 等人发表的预测方法来比对化合物的天然产物相似性,从而获得的 Real 天然产物类似物库。

### 3) Real 片段库

包含了 1,290 万个片段化合物,该库采用了“3 规则”即 ( $MW < 300$ ,  $SlogP \leq 3$ ,  $HBA \leq 3$ ,  $HBD \leq 3$ ,  $rot B \leq 3$ ,  $TPSA \leq 60$ ),运用这个片段库可以找到更新颖的活性片段。

### 4) Real 共价库

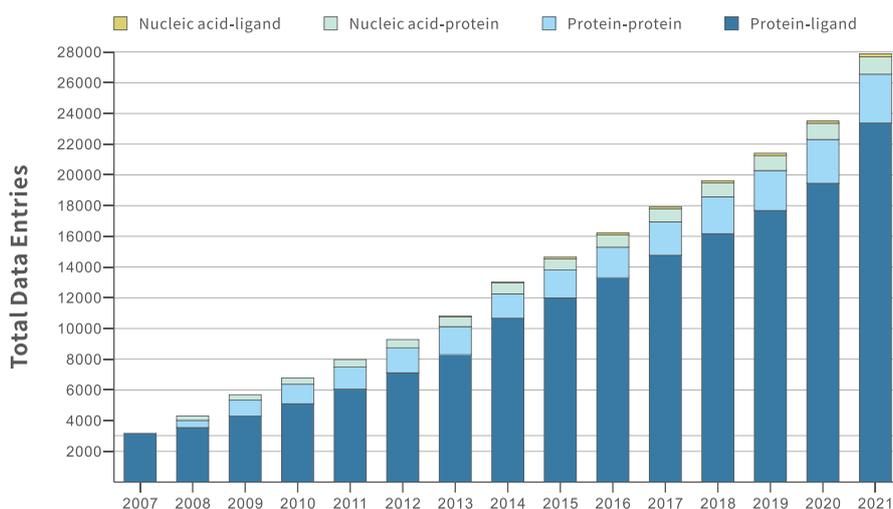
以目前广为研究的磺酰氟,丙烯酰胺和硼酸等共价靶头建立的化合物库,可用于筛选以共价作用与酶结合的共价抑制剂。相较于非共价药物,其有效浓度有望大幅度降低,有效作用时间大幅度增加。

# DATABASE

## The PDBbind database

### Artificial Intelligence in Drug Discovery

PDBbind 数据库由复旦大学王任小研究员（兼任上海陶术 CTO）团队创建，致力于系统地收集 Protein Data Bank 中各类复合物（蛋白-配体、蛋白-蛋白、蛋白-核酸和核酸-配体复合物）的三维结构及其匹配的亲合性实验数据。该数据库对于训练预测靶标-配体相互作用的人工智能算法等研究课题至关重要。该数据库自公开发布以来不断保持更新，在国内外同行中得到广泛应用，成为我国在药物设计领域中拥有自主知识产权的著名数据库之一。基于该数据库提供的高质量数据信息，国内外已有超过 1000 项研究成果公开发表。



【PDBbind 数据库的数据量逐年增长】

PDBbind 数据库自 2021 版起转变为付费订阅模式。复旦大学团队与上海陶术为此联合创建了 PDBbind+ 网站，作为发布 PDBbind 数据库最新版本和未来版本的唯一平台。



【PDBbind+ 网站的主界面】

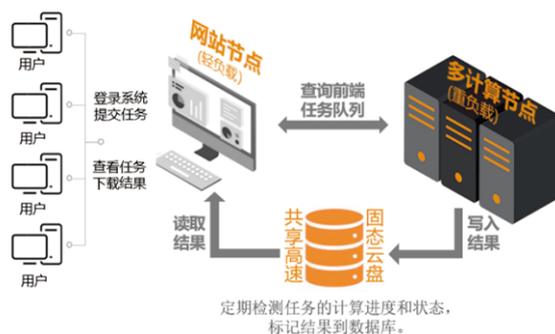
## PDBbind 数据库 2021 版

### 1) 规模显著扩展的亲合性实验数据

从原始文献中收集的复合物亲合性实验数据是 PDBbind 数据库的核心价值。最新的 2021 版本包含 27,408 个生物分子复合物的亲合性实验数据(其中包括 22920 个蛋白-配体复合物、3176 个蛋白-蛋白复合物、1141 个蛋白-核酸复合物以及 171 个核酸-配体复合物),相比 2020 版的数据量增长了约 20%。

### 2) 质量显著改善的复合物结构文件

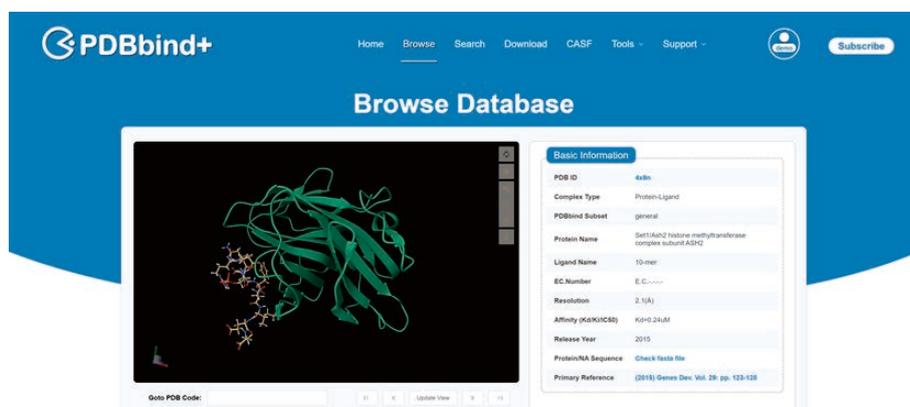
为了方便用户, PDBbind 数据库还提供经过细致处理的蛋白-配体复合物结构文件。在 2021 版本的开发过程中,我们采用了全新的工作流程来处理源自 Protein Data Bank 的原始结构文件,确保蛋白分子和配体分子在规定的文件格式中被准确且合理地表示。以往版本中的蛋白-配体复合物结构文件也采用新的工作流程进行了重制,解决了过去积累的种种琐碎问题。



【2021 版是 PDBbind 数据库历史上最重要的更新, 没有之一】

### 3) 功能显著增强的网络平台

PDBbind+ 网站基于先进的商业云服务器资源, 为用户提供稳定的网络连接和强大的在线计算服务。该平台提供结构可视化和分析的先进工具, 增强了用户体验, 并整合了反向找靶 方法等药物设计实用功能, 致力于打造成为综合性的药物设计服务平台。



【PDBbind+ 网站提供在线结构可视化的先进工具以及在线计算功能】

## CITATIONS

## 技术服务助力科研的部分成功案例

## Citations

- 1. Li J, Zheng C, Mai Q, Huang X, Pan W, Lu J, Chen Z, Zhang S, Zhang C, Huang H, Chen Y, Guo H, Wu Z, Deng C, Jiang Y, Li B, Liu J, Yao S, Pan C. Tyrosine catabolism enhances genotoxic chemotherapy by suppressing translesion DNA synthesis in epithelial ovarian cancer. *Cell Metab.* 2023 Nov 7;35(11):2044-2059.e8.
- 2. Shi W, Zhang G, Ma Z, Li L, Liu M, Qin L, Yu Z, Zhao L, Liu Y, Zhang X, Qin J, Ye H, Jiang X, Zhou H, Sun H, Jiao Z. Hyperactivation of HER2-SHCBP1-PLK1 axis promotes tumor cell mitosis and impairs trastuzumab sensitivity to gastric cancer. *Nat Commun.* 2021 May 14;12(1):2812.
- 3. Yang Y, Sun N, Lv J, Chen H, Wang H, Xu J, Hu J, Tao L, Fang M, Huang Y. Environmentally realistic dose of tire-derived metabolite 6PPD-Q exposure causes intestinal jejunum and ileum damage in mice via cannabinoid receptor-activated inflammation. *Sci Total Environ.* 2024 Mar 25;918:170679.
- 4. Xiao MC, Jiang N, Chen LL, Liu F, Liu SQ, Ding CH, Wu SH, Wang KQ, Luo YY, Peng Y, Yan FZ, Zhang X, Qian H, Xie WF. TRIB3-TRIM8 complex drives NAFLD progression by regulating HNF4 $\alpha$  stability. *J Hepatol.* 2024 May;80(5):778-791.
- 5. Hong X, Wei Z, He L, Bu Q, Wu G, Chen G, He W, Deng Q, Huang S, Huang Y, Yu C, Luo X, Lin Y. High-throughput virtual screening to identify potential small molecule inhibitors of the Z $\alpha$  domain of the adenosine deaminases acting on RNA 1(ADAR1). *Eur J Pharm Sci.* 2024 Feb 1;193:106672.
- 6. Du J, Liu N, Ma L, Liu R, Zuo D, Lan X, Yang J, Wei W, Peng X, Yu J. Antidepressant effect of the novel histone deacetylase-5 inhibitor T2943 in a chronic restraint stress mouse model. *Biomed Pharmacother.* 2024 Feb;171:116176.
- 7. Chen Y, Li L, Lan J, Cui Y, Rao X, Zhao J, Xing T, Ju G, Song G, Lou J, Liang J. CRISPR screens uncover protective effect of PSTK as a regulator of chemotherapy-induced ferroptosis in hepatocellular carcinoma. *Mol Cancer.* 2022 Jan 4;21(1):11.
- 8. Xie Y, Zhang L, Wang L, Chen B, Guo X, Yang Y, Shi W, Chen A, Yi J, Tang J, Xiang J. EphB1 promotes the differentiation and maturation of dendritic cells in non-small cell lung cancer. *Cancer Lett.* 2024 Feb 1;582:216567
- 09. Yu Z, Peng Y, Gao J, Zhou M, Shi L, Zhao F, Wang C, Tian X, Feng L, Huo X, Zhang B, Liu M, Fang D, Ma X. The p23 co-chaperone is a succinate-activated COX-2 transcription factor in lung adenocarcinoma tumorigenesis. *Sci Adv.* 2023 Jun 30;9(26):eade0387. doi: 10.1126/sciadv.ade0387. Epub 2023 Jun 30.

- 10. Bai YB, Yang XR, Li B, Zhou XZ, Wang WW, Cheng FS, Zhang JY. Virtual Screening and In Vitro Experimental Verification of LuxS Inhibitors for Escherichia coli O157:H7. *Microbiol Spectr.* 2023 Feb 21;11(2):e0350222.
- 11. Wang J, Ran T, Chen Y, Lu T. Bayesian machine learning to discover Bruton's tyrosine kinase inhibitors. *Chem Biol Drug Des.* 2020 Oct;96(4):1114-1122.
- 13. Wang W, Liu C, Li H, Tian S, Liu Y, Wang N, Yan D, Li H, Hu Q. Discovery of novel and potent P2Y14R antagonists via structure-based virtual screening for the treatment of acute gouty arthritis. *J Adv Res.* 2020 Feb 13;23:133-142.
- 14. Wang M, Hou S, Wei Y, Li D, Lin J. Discovery of novel dual adenosine A1/A2A receptor antagonists using deep learning, pharmacophore modeling and molecular docking. *PLoS Comput Biol.* 2021 Mar 19;17(3):e1008821.



# 上海陶术生物科技有限公司

抑制剂 & 激动剂 | 化合物库 | 天然产物 | 重组蛋白 | 技术服务

www.tsbiochem.com    Tel: 400 - 820 - 0310    Email: sales@tsbiochem.com

陶术生物所有产品和服务仅用于科学研究, 不能被用于人体, 我们不向个人提供产品和服务。



官方微信公众号



扫码浏览官网