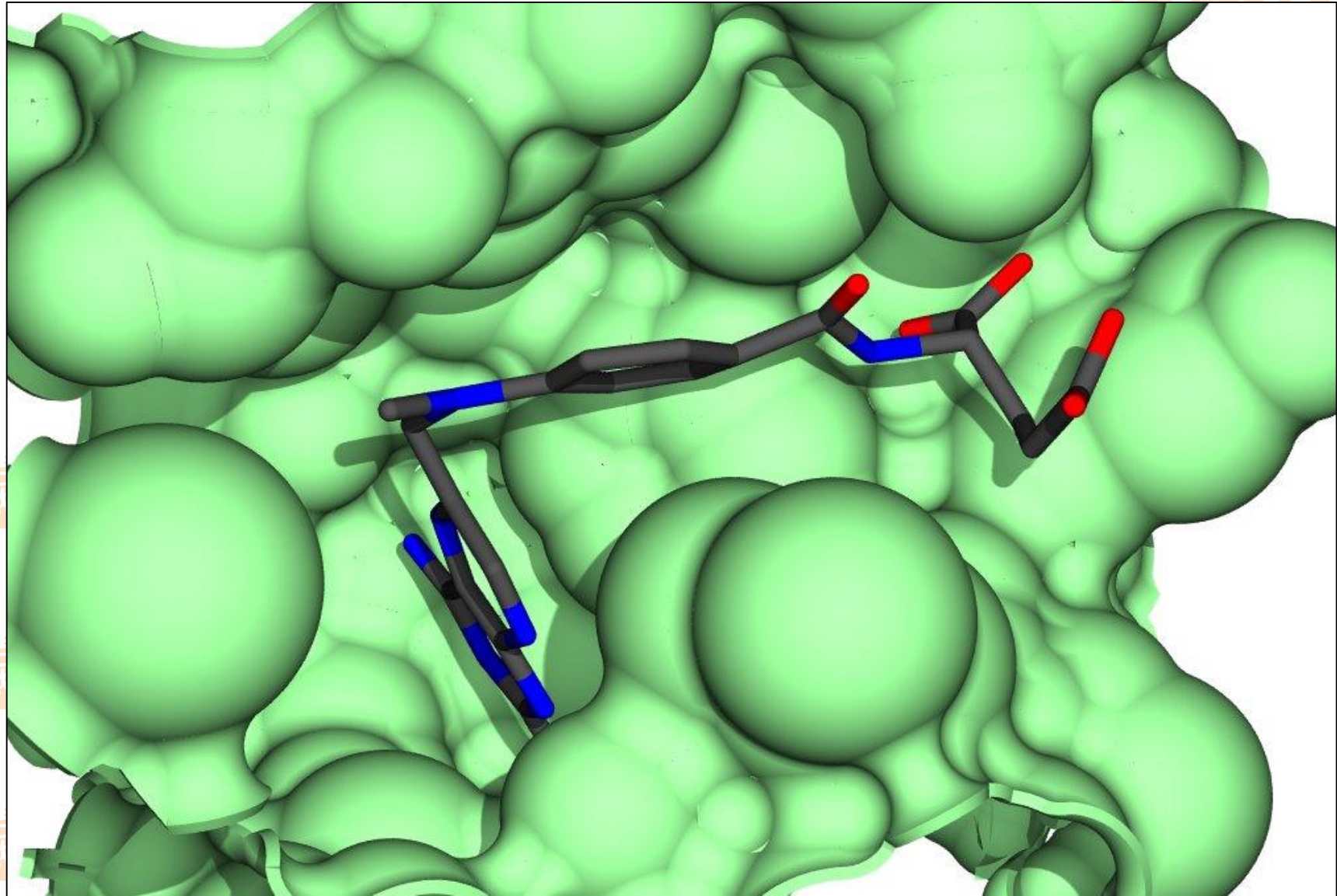


分子对接与虚拟筛选

Cao Yang
2015.5.25



Molecular Docking



分子对接的概念

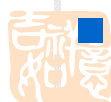


- 预测受体和配体分子形成的复合物结构

■ 分子对接分为两类

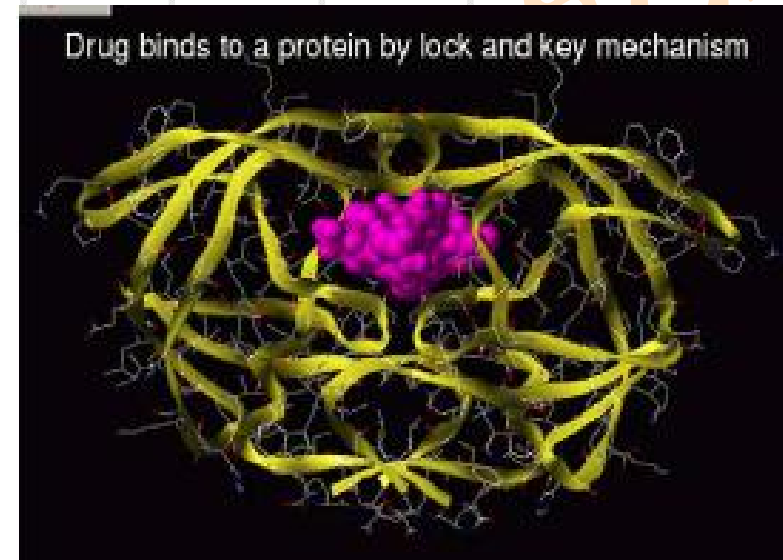
➤ 蛋白-蛋白分子对接

➤ 蛋白-小分子的对接



分子对接的原理

- 理论基础:
- “锁和钥匙模型”
- “诱导契合模型”



重要原则:

- 互补性: 决定识别过程的选择性
- 预组织性: 决定识别过程的结合能力

分子对接的基本原理

配体与受体的结合强度取决于结合的自由能变化

$$\Delta G_{\text{结合}} = \Delta H_{\text{结合}} - T\Delta S_{\text{结合}} = -RT \ln K_i$$

许多分子对接法忽略了熵效应，而在焓效应也只考虑配体与受体的相互作用能，即：

$$E_{\text{interaction}} = E_{\text{vdw}} + E_{\text{electrostatic}} + E_{\text{h-bond}}$$

分子对接的原理



- 搜索算法：如何找到最佳的结合位置

- 遗传算法
- 模拟退火

- 能量函数：如何评估结合强度

- 基于分子力场的方法
- 基于经验的方法
- 基于知识统计的打分函数



找出空穴，定出表面

□ Q-SiteFinder

□ www.modelling.leeds.ac.uk/qsitesfinder/



Q-SiteFinder

Ligand Binding Site Prediction

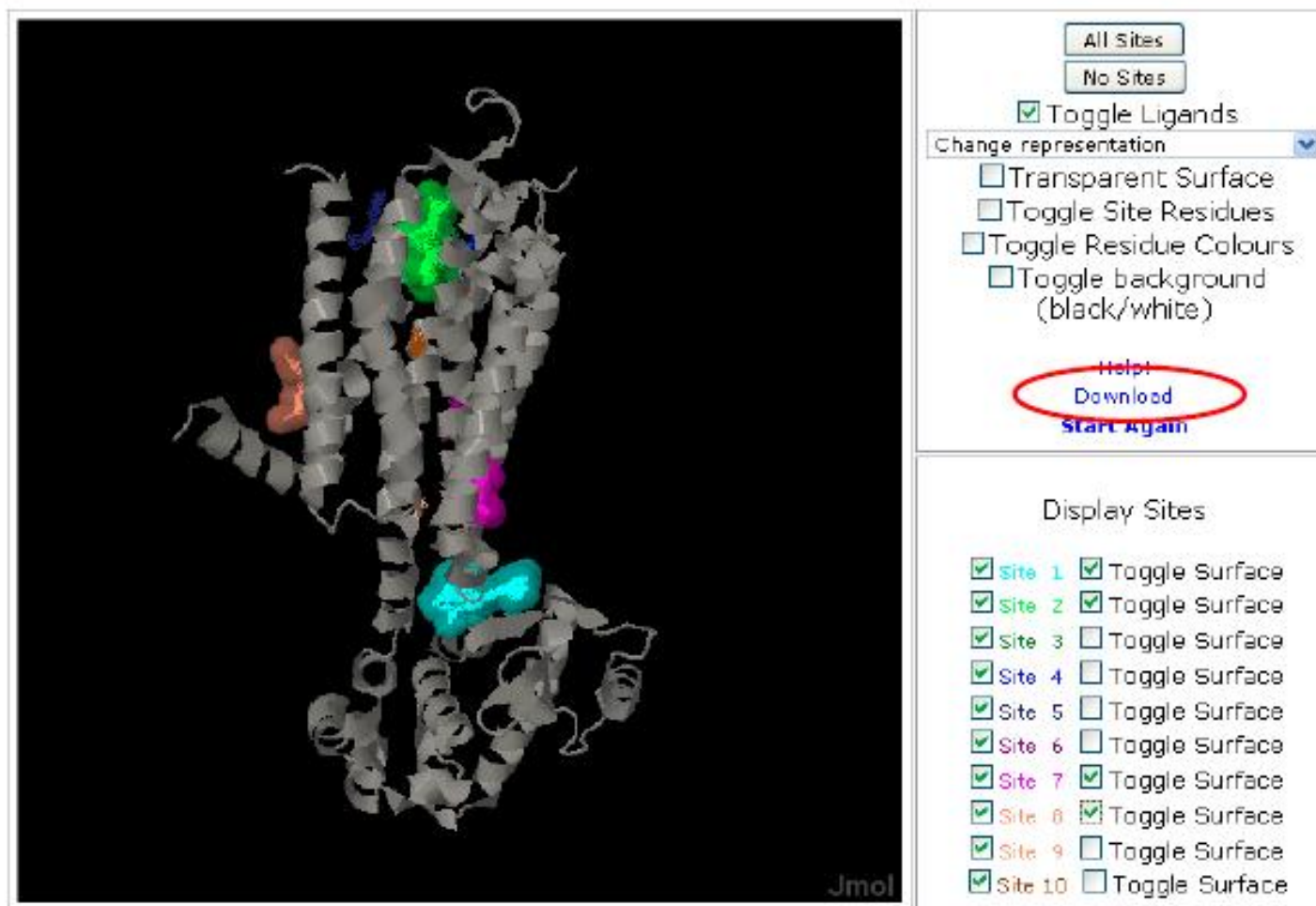
Submit to Q-SiteFinder*

Enter a PDB code:

OR Upload a PDB file to Q-SiteFinder:

*Interface uses Jmol viewer (Java required).

Q-SiteFinder



The image displays the Q-SiteFinder web interface. On the left, a 3D ribbon diagram of a protein structure is shown against a black background. Several binding sites are highlighted with colored surfaces: a green site at the top, an orange site on the left, a cyan site at the bottom, and a magenta site on the right. The protein backbone is shown in a grey ribbon representation.

On the right, there is a control panel with the following options:

-
-
- ☒ Toggle Ligands
- Change representation
- ☐ Transparent Surface
- ☐ Toggle Site Residues
- ☐ Toggle Residue Colours
- ☐ Toggle background (black/white)
- [Help!](#)
- [Download](#)
- [Start Again](#)

Below these controls is a section titled "Display Sites" with a list of 10 sites and their corresponding "Toggle Surface" options:

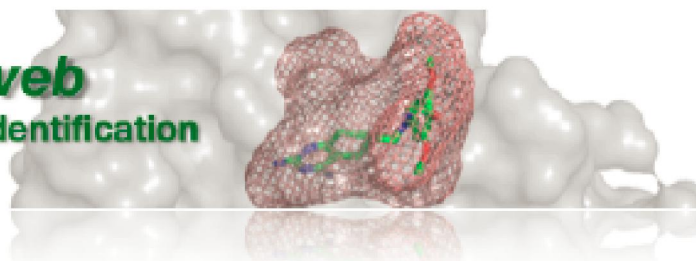
Site	Toggle Surface
<input checked="" type="checkbox"/> Site 1	<input checked="" type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 2	<input checked="" type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 3	<input type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 4	<input type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 5	<input type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 6	<input type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 7	<input checked="" type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 8	<input checked="" type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 9	<input type="checkbox"/> Toggle Surface
<input checked="" type="checkbox"/> Site 10	<input type="checkbox"/> Toggle Surface

找出空穴，定出表面



- SiteHound
- bsbbsinai.org/SHserver/SiteHound/Input.html

SITEHOUND-web Ligand Binding Site Identification in Protein Structures



SiteHound-web identifies ligand binding sites by computing interactions between a chemical probe and a protein structure. The input is a PDB file of a protein structure, the output is a list of "interaction energy clusters" corresponding to putative binding sites. Selecting a different chemical probe results in the identification of a different type of binding site.

For details see:
[Hernandez et al., NAR, 37:W413, 2009.](#)
[Gheri & Sanchez, Proteins, 74:417, 2009.](#)
[Gheri & Sanchez, Bioinformatics, 25:3185, 2009.](#)

A **standalone version** of SITEHOUND and instruction manual can be downloaded [here](#).

Please specify a PDB file or a PDB ID:

Probe ?

- ☒ Carbon (EasyMIFs) ☐ Carbon (Autogrid)
☐ Phosphate (EasyMIFs)

Clustering Algorithm ?

- ☐ Single
☒ Average

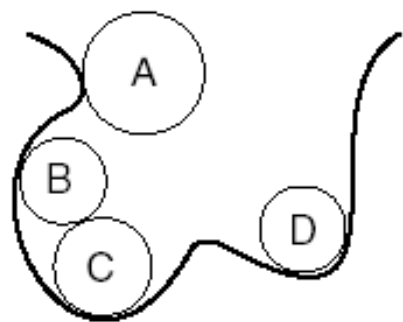


分子对接的分类

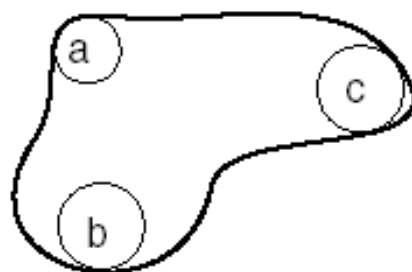
- 刚性对接：研究体系的构象不发生变化。
- 柔性对接：研究体系的构象是可以自由变化的。



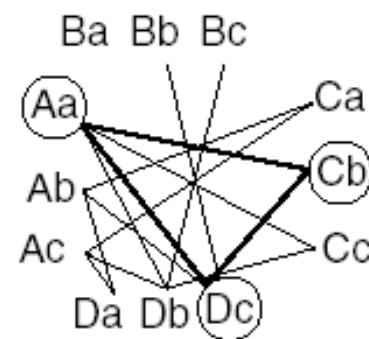
分子对接的算法举例



受体的活性位点



配体



代表基团的空间匹配



代表性对接软件

名称	构象搜索方法	结合评价方法	速度
Flex X	片段生长法	半经验自由能	快
LigandFit	蒙特卡罗模拟	半经验自由能	快
Glide	系统搜索	半经验自由能	一般
Gold	遗传算法	半经验自由能	快
AutoDock	遗传算法	半经验自由能	一般
Dock	片段生长法	分子力场	快
ICM-Dock	随机全局优化	半经验自由能	快
Fred (openeye)	系统搜索	半经验自由能	快

Autodock Vina



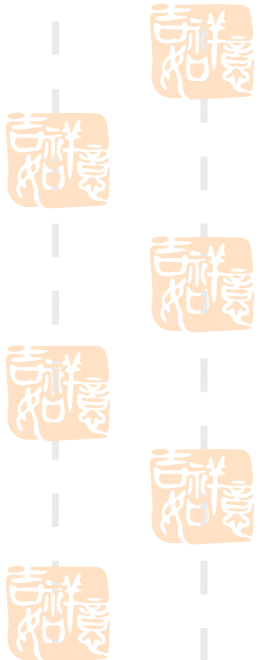
- <http://vina.scripps.edu/>
- <http://vina.scripps.edu/tutorial.html>
- <http://matrixspace.xicp.net:8002/webvs/webvina5.php>

Autodock Vina Web Interface

Designed by YC, 2015

center_x	<input type="text"/>
center_y	<input type="text"/>
center_z	<input type="text"/>
size_x	<input type="text"/>
size_y	<input type="text"/>
size_z	<input type="text"/>
Protein (pdbqt format)	<input type="text"/> <input data-bbox="1400 1157 1451 1177" type="button" value="浏览..."/>
Ligand (pdbqt format)	<input type="text"/> <input data-bbox="1400 1185 1451 1206" type="button" value="浏览..."/>
password	<input type="text"/>
<input data-bbox="1176 1241 1227 1262" type="button" value="Submit"/>	

© 2015 All Rights Reserved.



分子对接计算的注意点

■ 小分子问题

- 起始构象对接结果有一定影响
- 对分子进行加电荷和加氢处理

■ 蛋白质问题

- 如何选择合理的蛋白质活性位点

■ 对接问题

- 搜索结合模式的正确性、对接的效率、评分的正确性
- 采用多个软件进行评价，减少结合模式搜索误差
- 定量指标，需要结合分子动力学进一步评价



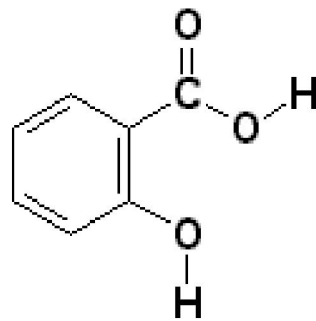
对接方法尚需解决的问题：

- 溶剂化效应
- 分子的柔性
- 打分函数

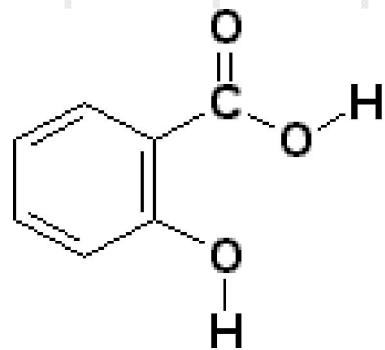


传统的药物开发方法

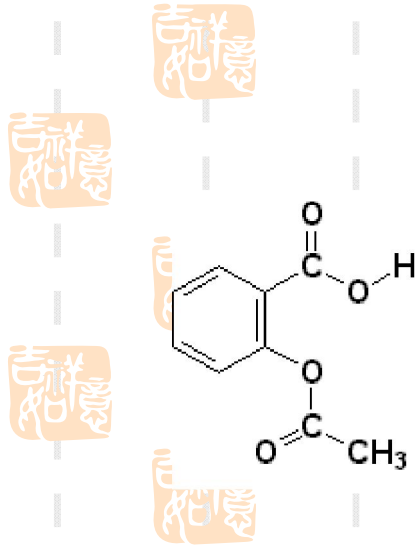
- 经验的有效天然产物



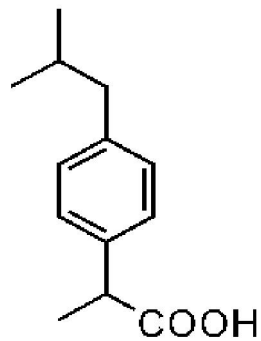
- 分离活性物质



• 合成化合物



Aspirin



Ibuprofen

布洛芬是公认的儿童首选抗炎药。



• 化合物的修饰

• (更有效,降低副作用)

药物设计与发现



- drug **targets** (usually proteins)
- binding of **ligands**



- “rational” drug design

- (benefits = saved time and **\$\$\$\$**)



现代的药物开发方法

NEW and IMPROVED!

- 区别在哪里？
- 药物开发从疾病入手（而不是已有的治疗方法）
- 应用疾病模型去寻找可能的药物靶点

现代的药物开发方法

Disease → genetic/biological target



High throughput screen (HTS)

- 发现 “hits” (compounds with binding in low nM to low μ M range)



Discovery of a “lead” molecule

- 先导化合物（亲和力高，有修改潜力）

现代的药物开发方法



- **manipulate structure to increase potency**

i.e. decrease K_i to low nM affinity



- **optimization of lead molecule into candidate drug**

fulfillment of required pharmacological properties:
potency, absorption, bioavailability, metabolism, safety



- **clinical trials**



现状

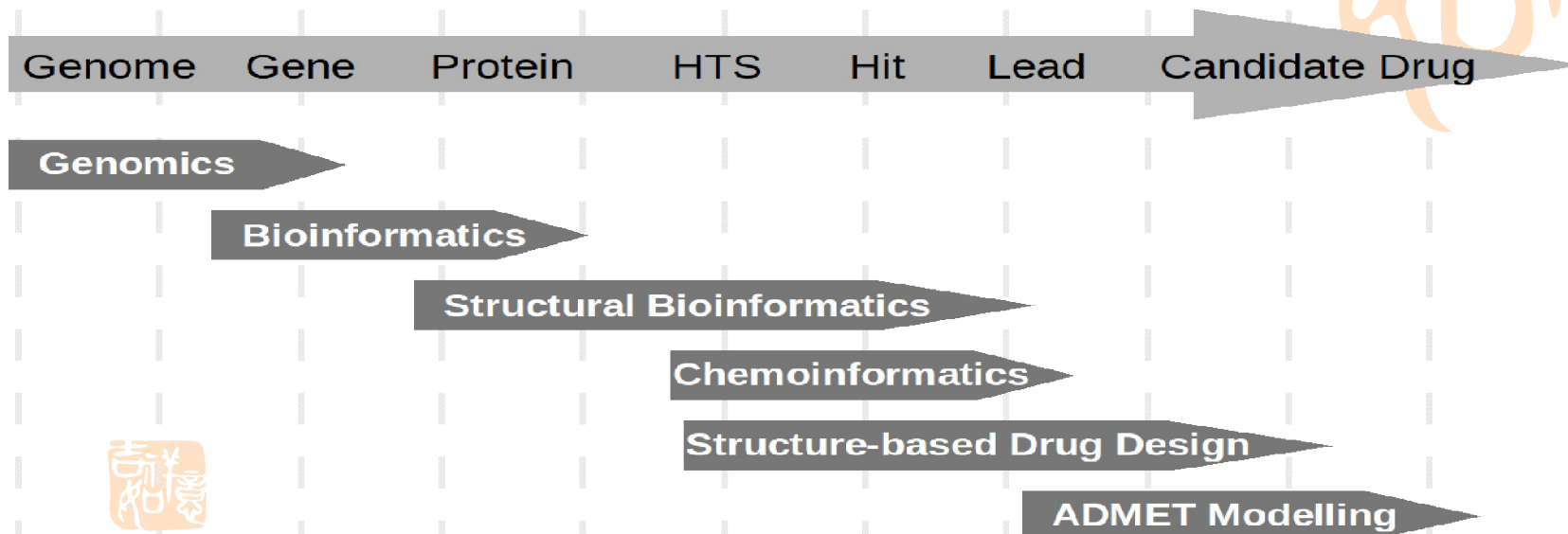
- 超过90% 的药物不能通过临床实验而被枪毙



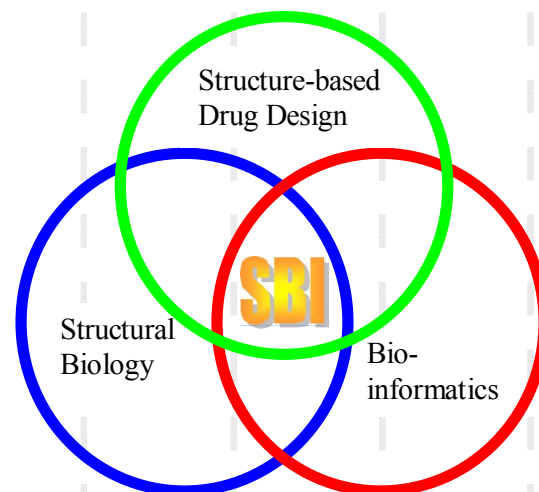
- 平均一款全新药物的研发耗资超过数亿美元，耗时10年



结构生物信息学对药物开发的影响

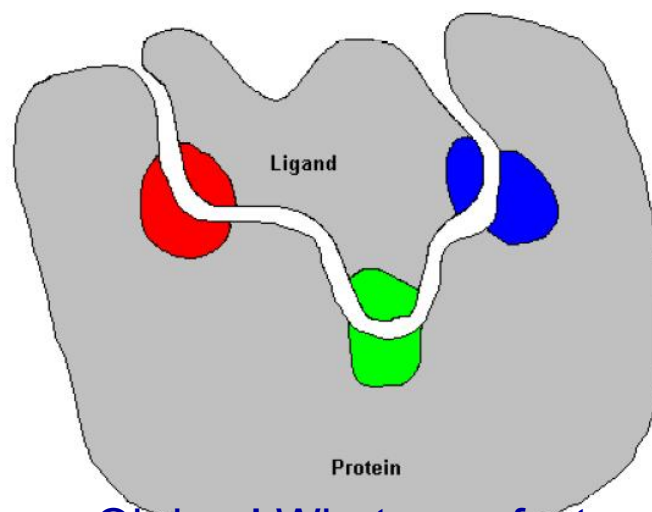


- Speeds up key steps in DD process by combining aspects of bioinformatics, structural biology, and structure-based drug design



虚拟筛选

首先要建立大量化合物（例如几十至上百万个化合物）的三维结构数据库，然后将库中的分子逐一与靶标分子进行“对接”（**docking**），通过不断优化小分子化合物的位置（取向）以及分子内部柔性键的二面角（构象），寻找小分子化合物与靶标大分子作用的最佳构象，计算其相互作用及结合能。在库中所有分子均完成了对接计算之后，即可从中找出与靶标分子结合的最佳分子（前**50**名或前**100**名）



Oh boy! What a perfect match

虚拟筛选



包括4个步骤：受体模型的建立；小分子库的产生；计算机筛选和命中化合物的后处理。

第一步，受体模型的建立：

1) 大分子结构获取

蛋白质结构的准备是虚拟筛选的重要一步。虚拟筛选的蛋白靶标的结构可以从PDB库中直接下载使用

也可以通过和家族中同源蛋白的序列、结构信息比较，同源模建而得



虚拟筛选



2) 接着是结合位点的描述，选择合适的配体结合口袋对分子对接至关重要

选择口袋有两种方式：

一种是直接从配体-受体复合物结构中抽出；

如果没有复合物结构，则需要根据生物功能如结合、突变等实验信息来手动选择结合部位



虚拟筛选



第二步，建立小分子数据库

二维结构用结构转换程序如CORINA、CONCORD实现三维结构的转化。



建好的三维结构加氢加电荷后，便可以用于对接程序。



虚拟筛选



第三步，对接和打分，这一步是虚拟筛选的核心步骤。

对接操作就是把每个小分子放到受体蛋白的配体结合位点，优化配体构象和位置，使之与受体有最佳的结合作用，给最佳结合构象打分，对所有化合物根据打分排序，然后从化合物库中挑出打分最高的小分子。



虚拟筛选



最后一步是命中化合物的后处理

通过计算分子的类药性质**ADME/T** (吸收**absorption**、器官分布**distribution**、体内代谢**metabolism**、排泄**excretion** 和毒性**toxicity**)性质的估算，排除那些不具有类药性质的分子。



可以利用一些经验规则如“五规则”等，快速排除那些不适合进一步药物开发的分子。

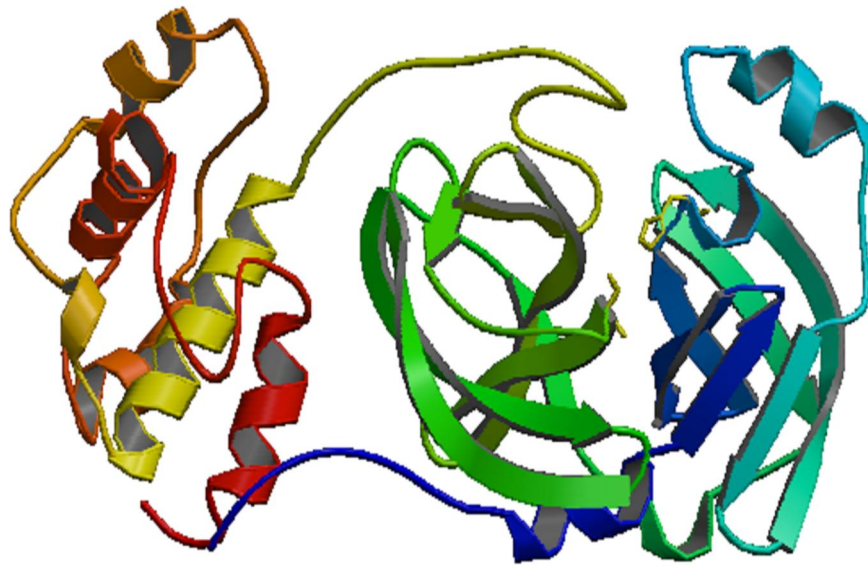


虚拟筛选



通过以上四步处理，大部分分子从化合物库中剔除，形成一个合理大小的化合物库，仅对这些适合成药的化合物或购买、或合成、或分离得到，然后再进行实际的生物测试。

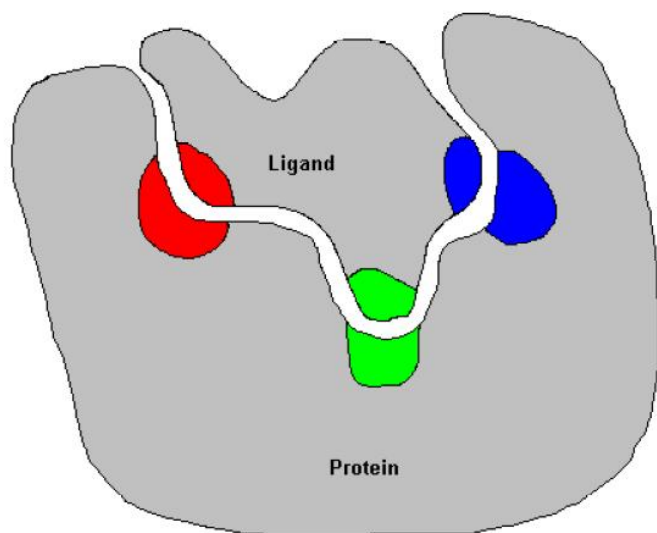




SARS冠状病毒**3C—like**蛋白酶在**SARS**病毒其他功能蛋白的形成过程中起重要作用，阻断**SARS**病毒**3C—like**蛋白酶的作用可以阻止**SARS**病毒的复制，并最终达到治疗**SARS**的目的

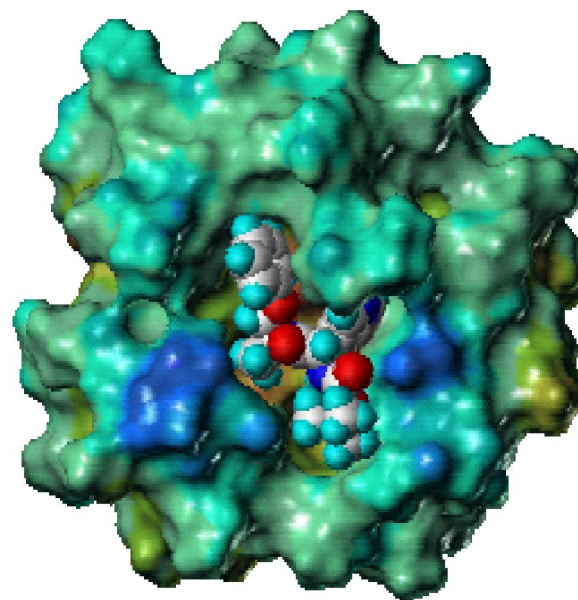
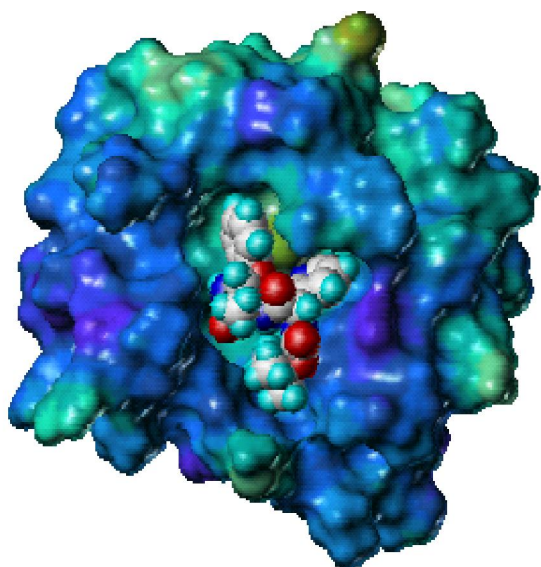
SARS病毒的**3C-like**蛋白酶的三维模建结构

在猪传染性胃肠炎病毒（**TGEV**）的主蛋白酶三维晶体结构的基础上，构建了**SARS**病毒的**3C—like**蛋白酶三维结构



用**DOCK**初筛，从每一个数据库筛选结果中选择**DOCK**打分前**1000**名的分子，进一步做类药性分析和多种评价函数包括**SYBYL**中的**Cscore**打分函数和**Autodock**的经验自由能评价方法进行打分，再根据药物化学家的经验来进行人工挑选，最终对每个数据库选择**100**个分子进行生物测试，这次虚拟筛选找到**300**个可能具有抗**SARS**冠状病毒的候选化合物，发现了**7**个具有高活性的化合物，进一步的细胞水平的实验，发现**5-HT**受体拮抗剂具有明显的抗**SARS**病毒感染和保护细胞的作用，其**EC50**值小于**10mg/L**

同一结构家族的组织蛋白酶抑制剂分别与**SARS**冠状病毒**3C-like**蛋白酶和**TGEV**的主蛋白酶的复合物模型如图，从图中可以看出两者的作用模式也是十分类似。



SARS病毒3C-like蛋白酶和TGEV主蛋白酶与抑制剂结合图

虚拟筛选的成功例子

靶 标	靶标分类	靶标结构	小分子库大小	所用方法	抑制剂活性 μM	实验数据
AmpC β -lactamase	Hydrolase	X-ray	200k	NWU DOCK	26	X-ray复合物
BCR—ABL	Kinase	X-ray	200k	DOCK	25	细胞的抑制活性实验
Anthrax EF	Adenylyl cyclase	X-ray	200k	NWU DOCK	20	酶动力学实验
IMPDH	Dehydrogrnase	X-ray	3500k	FlexX	30	酶动力学实验
Casein kinase II	Kinase	Homology	400k	DOCK	0.08	抑制活性、构效关系
K ⁺ 通道	Ion channel	Homology	50k	DOCK	10	细胞的抑制活性实验
Thyroid hormone receptor	Nuclear receptor	Homology	250k	ICM	0.75	抑制活性实验
CDK2	Kinase	X-ray	50k	LIDAEUS	2	X-ray复合物
TGF β RK	Kinase	X-ray	200k	Catalyst	0.005	X-ray复合物
cyclophilin	Immunophilin	X-ray		Unity/FlexX	6	细胞的抑制活性实验
tRNA-guanine transglycosylase		X-ray	800k	Unity/FlexX	0.25	酶动力学实验
PfDHFR	Reductase	Homology	230k	Catalyst/DOCK	0.9	酶动力学实验
α -Amylase	Hydrolase	X-ray	200k	Unity/FlexX		NMR,SPR,层析

吉祥如意

谢谢

吉祥如意

吉祥如意

吉祥如意

吉祥如意

吉祥如意

吉祥如意