

《计算机辅助药物设计》课程

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

基于结构的药物设计


Structure-based drug design, SBDD

主讲人：邹毅
Email: zouyi@cpu.edu.cn

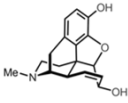
1

早期药物发现的途径


中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY



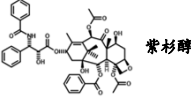
1803年德国药师萨顿从阿片中分离到黄色结晶



吗啡



1971年提取、分离、鉴定“卵巢癌和转移性乳腺癌治愈率达33%，总有效率75%以上”



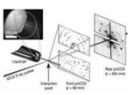
紫杉醇

一直到20世纪70年代，随机筛选/偶然发现仍然是最主要的先导化合物发现方式，至今仍不可替代


2

合理药物设计

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY



X-射线晶体衍射



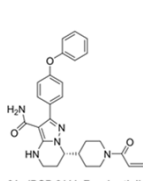
核磁共振

靶点三维结构

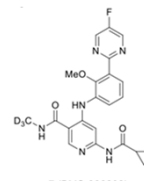
3

分子为什么这样设计，依据是什么？

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY



31a (BGB-3111, Zanubrutinib)



7 (BMS-986202)

4

课程回顾

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

理论计算方法

- 量子力学 → 量子化学
- 经典力学
- 分子力学 → 分子动力学

求解Schrödinger方程

求解经典牛顿力学方程

计算机辅助药物设计所涉及的理论计算方法

5

课程回顾

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

有无靶标蛋白的晶体结构？

有

同源建模

直接药物设计

- 分子对接
- 基于受体结构的药效团
- 全新药物设计
- 定量构效关系

无

间接药物设计

- 基于配体的药效团
- 相似性搜索

计算机辅助药物设计方法

6

基于结构的药物设计的基本概念



- **基于结构的药物设计 (Structure-based drug design, SBDD)**，广义上讲是基于配体结构和受体蛋白结构的药物设计的统称；狭义上讲就是基于受体结构的药物设计，即依据与药物作用的靶点（广义上的受体，如酶、受体、离子通道、抗原、核酸、多糖等）的三维结构，运用分子识别原理（互补性），设计对受体进行调控的先导物，或根据已有药物作用力大小和构效关系判断来推测新化合物的药效，达到发现活性分子的目的。
- 基于配体的药物设计也称为**间接药物设计**；基于受体结构的药物设计也称为**直接药物设计**。



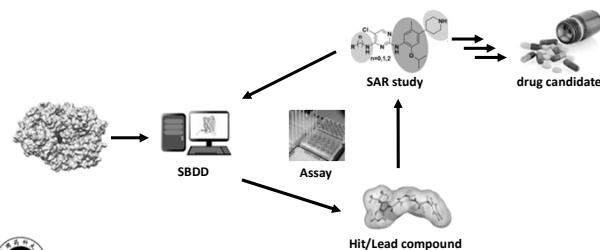
7

7

基于结构的药物设计的基本概念



基于结构的药物设计基本流程



8

8

Outline



- 蛋白质的结构与预测
- 基于结构的药物设计 (SBDD) 基本概念与理论
- 分子对接技术 (Molecular docking)
- 基于片段的药物设计 (Fragment-based drug design, FBDD)



9

9

第一节 Part One

蛋白质的结构与预测

10

10

蛋白质的结构与预测



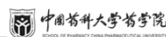
- 蛋白质的三维结构
- 蛋白质结构的获取方法
- 蛋白质三维结构的预测方法 ★
- 蛋白质结构的预处理 ★



11

11

蛋白质的三维结构



- 蛋白质是生物体中含量最丰富的生物大分子，参与了细胞中大部分的生命过程，是细胞最重要的组成物质。
- 当一个或多个蛋白的结构、功能或其参与相互作用的信号通路发生异常，将引起人类疾病的发生。经过严格的靶标验证实验，那么它就可能成为药物设计的靶标蛋白，通过设计药物分子与靶蛋白的结合便可达到治疗疾病的目的。
- **对靶蛋白三维结构的获取及分析则是SBDD的先决条件。**



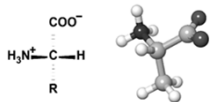
12

12

蛋白质的三维结构



蛋白质的基本结构单元：氨基酸



L-型氨基酸通式 (R表示不同侧链)
及其三维结构 (以丙氨酸为例)

组成蛋白质的种类繁多，结构各异，但元素组成基本相似，主要有碳 (C)、氢 (H)、氧 (O)、氮 (N) 和硫 (S)。除此以外，有些蛋白质发挥功能还需要少量金属离子的参与，如铁 (Fe^{2+})、镁 (Mg^{2+})、钙 (Ca^{2+}) 和锌 (Zn^{2+}) 等。



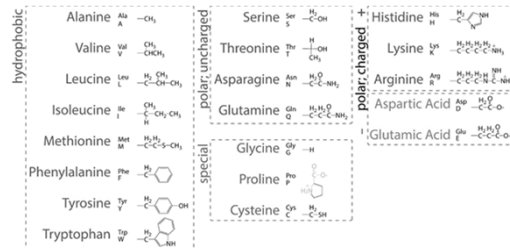
13

13

蛋白质的三维结构



氨基酸的分类



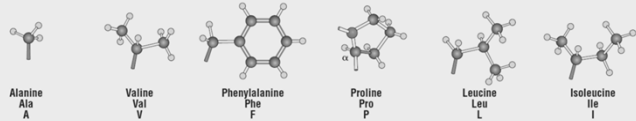
14

14

蛋白质的三维结构



疏水性氨基酸



- 这些疏水侧链参与了范德华 (van der Waals, vdW) 相互作用，它在维持蛋白质构象中起着主要的作用，是蛋白质折叠的主要驱动力。
- 苯丙氨酸Phe的苯基侧链有时能参与较弱的极性相互作用。



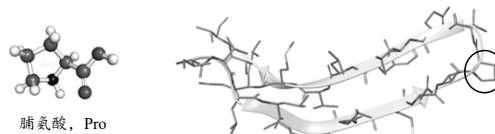
15

15

蛋白质的三维结构



Pro比其它任何氨基酸残基都具有更强的立体化学效应，它的侧链是和主链氮共价相连的，封闭的环常常形成一个转角，改变了主链的方向。



16

16

蛋白质的三维结构



带电氨基酸



带电荷的氨基酸都属于亲水性氨基酸，包括酸性氨基酸，如天冬氨酸和谷氨酸和碱性氨基酸，如精氨酸、赖氨酸和组氨酸。这些氨基酸主要参与了氢键相互作用和正负电荷的吸引作用 (即盐桥)。



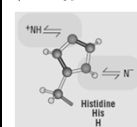
17

17

蛋白质的三维结构

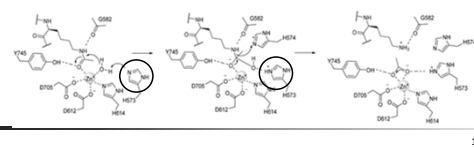


带电氨基酸



组氨酸, His

- 在中性状态下，其氮原子上的氢可在两个氮原子上互变 (τ 互变和 π 互变)，pKa均在6左右，具体以哪个异构体为主取决于其周围的环境；当两个氮均被质子化时，氨基酸整体带正电荷。
- 在pdb中，组氨酸的残基名都用“HIS”表示，而在Amber力场中， τ 互变、 π 互变及带正电的His则分别表示为“HIE、HID和HIP”。



18

18

蛋白质的三维结构

极性中性氨基酸

- Ser、Thr、Asn和Gln主要参与氢键相互作用，属于亲水性氨基酸。
- 芳香环类侧链如Tyr和Trp虽然是疏水的，但它们都能形成氢键，可以位于蛋白质的内部或外部。
- 具有完全疏水性侧链的Phe是蛋白质内部疏水核心的主要成分。

19

蛋白质的三维结构

极性中性氨基酸

31a (BGB-3111, Zanubrutinib) PDB ID: 6J6M

- 成对的Cys残基能形成二硫键，从而增强蛋白质结构的稳定
- Cys也与His一样，常见于酶活性位点
- 当侧链的巯基去质子化时具有较强的亲核性，常作为共价药物设计的靶向残基。

20

蛋白质的三维结构

蛋白中辅因子的作用

辅助蛋白质折叠 物质运输 辅助酶催化

21

蛋白质的三维结构

氨基酸通过肽键组成多肽

- 一个氨基酸的氨基与另一个氨基酸的羧基通过脱水形成的酰胺键称为**肽键**。
- 由两个氨基酸形成的肽为**二肽**，以此类推。不超过10个氨基酸形成的肽也被称为**寡肽**，超过10个氨基酸的肽也称为**多肽**。

主链 (backbone) 侧链、残基 (side chain, residue)

22

蛋白质的三维结构

氨基酸通过肽键组成多肽

- 位于肽键平面外的另外两个单键 (N-C_α和C_α-C) 在没有侧链位阻的情况下是可以自由旋转的。
- 绕肽键一边的N-C_α键旋转的角称为phi (φ)，绕着同一C_α原子的C_α-C键旋转的角被称为psi (ψ)。
- 其主链具有可旋转键和刚性平面相互交替的特征，自由度仅与两个构象角密切相关。

23

蛋白质的三维结构

拉氏图 (Ramachandran图)

将蛋白或多肽主链的φ和ψ值可视化而成的图形称为拉氏图，由 G. N. Ramachandran 等人提出的，以φ和ψ的角度为横纵坐标，规定φ和ψ角允许的构象区域的一个图形

- 主要的应用包括了辅助结构生物学中蛋白晶体结构的解析以及对同源建模后模型质量的检测。

在线服务: PROCHECK (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>)

24

蛋白质的三维结构

蛋白质的高级结构

- 在生理温度下，水溶液中蛋白质的多肽链在大多数情况下折叠成呈球状。蛋白质中不同氨基酸的序列是它的一级结构（primary structure），直接由编码它的基因序列决定，序列又决定了蛋白折叠成更高级别结构的方式，包括二级结构、三级结构和四级结构。

25

蛋白质的三维结构

蛋白质的高级结构：二级结构

α螺旋 (α-helix)
上: β转角 (β-turn)
下: β折叠 (β-sheet)
卷曲或环形区域 (coil or loop)

26

蛋白质的三维结构

快速识别出蛋白晶体结构中的二级结构

PDB ID: 4AGD
PDB ID: 2NNQ

27

蛋白质结构的获取方法

- X-射线晶体学 (X-ray crystallography)
- 冷冻电镜技术 (Cryo-EM)
- 核磁共振技术 (NMR)
- 蛋白质三维结构预测

28

蛋白质结构的获取方法

X-射线晶体学 (X-ray crystallography)

通过对衍射的位置、强度计算，读出原子坐标值（电子密度），解析结构得到晶体空间结构。

29

蛋白质结构的获取方法

在SBDD中，同一靶点的晶体结构要尽可能选择分辨率（Resolution）高（数值低）的结构作为靶标结构进行后续设计。

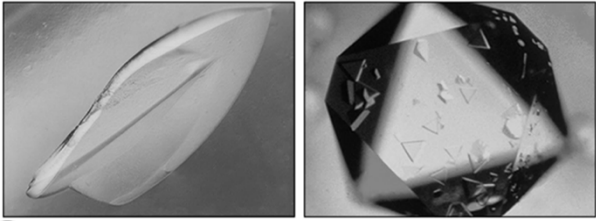
同一分子结构，不同分辨率的电子密度图

Resolution (Å)
3.00 2.70 2.00 1.50 1.20 1.00 0.65
Low Medium High Atomic

30

蛋白质结构的获取方法

各种蛋白质晶体



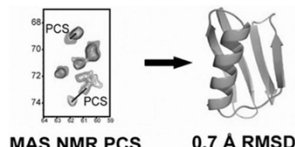
31

31

蛋白质结构的获取方法

核磁共振技术 (NMR)

核磁共振能够给出单个功能基团的信息,如离子化状态、pKa和氢键;提供溶液中主链和侧链的动态运动的位点特异信息。NMR的优势在于能在接近自然生理环境的条件下对生物分子进行研究。

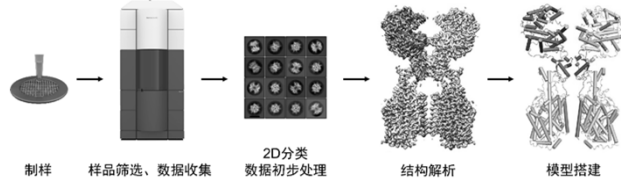


32

32

蛋白质结构的获取方法

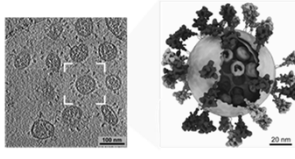
冷冻电镜, 是用于扫描电镜的超低温冷冻制样及传输技术(Cryo-SEM), 可实现直接观察液体、半液体及对电子束敏感的样品, 如生物、高分子材料等。



33

33

蛋白质结构的获取方法



采用冷冻电镜技术解析的新型冠状病毒的全病毒结构

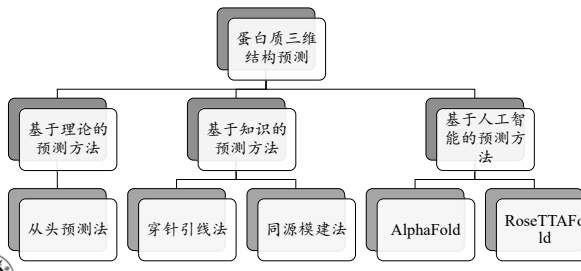
- 2020年10月份, 清华大学李赛团队与浙江大学李兰娟团队合作首次使用高分辨冷冻电镜断层成像技术解析出新冠病毒全病毒三维结构, 从原位状态下深度分析刺突蛋白的结构并开创性地揭示了核糖核蛋白复合物天然结构。这一真实的病毒三维图像让“看不见的敌人”清晰地展现在世人面前, 首次让世界看到新冠病毒全病毒结构。

34

34

蛋白质三维结构的预测方法

蛋白质三维结构预测



35

35

蛋白质三维结构的预测方法

基于理论的预测方法: 从头预测法 (ab initio prediction)

- 原理:** 采用理论计算 (分子力学、分子动力学、量子化学) 方法, 直接从分子和原子参数模拟肽段在三维空间中所有可能的姿态, 并计算出自由能最低的一个。
- 特点:** 计算量极大, 不常用。
- 代表工具:** QUARK <https://zhanglab.ccmb.med.umich.edu/QUARK/>

36

36

蛋白质三维结构的预测方法



基于知识的预测方法：穿针引线法，折叠识别（threading, fold recognition）

- **原理：**通过比较目标序列与已知的折叠库，找到目标序列的折叠模式，并使用相关的方法将各个部分的折叠连接起来。
- **特点：**折叠法是近年来发展起来的一种比较新的方法。它可以应用到没有同源结构的情况中，且不需要预测二级结构，即直接预测三级结构，从而可以绕过现阶段二级结构预测准确性不超过65%的限度，是一种有潜力的预测方法。
- **代表工具：**I-TASSER (Iterative Threading ASSEMBLY Refinement)
<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>



37

37

蛋白质三维结构的预测方法



基于知识的预测方法：同源建模法（homology modeling）

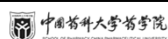
- **原理：**根据待建模序列（目标）与一个或多个已知蛋白结构的序列（模板）间的同源性（序列一致性），直接由目标序列的一级结构预测其三级结构
- **理论基础：**蛋白质三级结构的保守性远超过一级序列的保守性
- **特点：**模板蛋白和目标蛋白的序列一致性需要大于30%，越大建模准确性越有保障。
- **代表工具：**
(1) MODELLER (<https://salilab.org/modeller/>)
(2) SWISS-MODEL (<https://swissmodel.expasy.org/>)



38

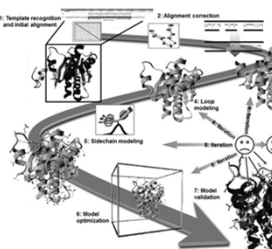
38

蛋白质三维结构的预测方法



同源建模的步骤（以DS MODELER为例）

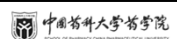
- 1、搜索并识别模板
- 2、序列比对
- 3、产生全原子模型（自动产生）
- 4、模型优化
- 5、模型评估



39

39

蛋白质三维结构的预测方法



同源建模的步骤（以DS MODELER为例）

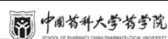
- 1、搜索并识别模板：查找目标蛋白序列（<https://www.uniprot.org/>）



40

40

蛋白质三维结构的预测方法



同源建模的步骤（以DS MODELER为例）

- 1、搜索并识别模板

- 载入序列
- 使用BLAST搜索模板

每行表示一条命中的氨基酸序列命中的序列按照E值（序列无错比对存在偶然性的可能性大小，表征了序列比对的可行度）进行排序。E值最低的序列，结果最可靠，排在第一行。

Query: P12345	Ref: P12345	Identity	Sequence Length	Alignment Length	Bit Score	E-value
1. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
2. P12345-123	1234_A	40	440	301	117.70	4.4440e-133
3. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
4. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
5. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
6. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
7. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
8. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
9. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
10. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
11. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
12. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
13. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
14. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
15. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
16. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
17. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
18. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
19. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133
20. ALPFA-MIT123	1234_A	40	440	301	117.70	4.4440e-133



41

41

蛋白质三维结构的预测方法



同源建模的步骤（以DS MODELER为例）

- 1、搜索并识别模板：模板的选取

- 挑选一个或多个合适的同源模板（templates）。一个理想的template需要涵盖整个target的长度，具有较高的序列一致性（Sequence identity），并且E值要够小（ $< 1 \times 10^{-5}$ ）。
- 一般而言，若有多条模板与target都具有相似的同源性，但模板之间相似性并不非常高，那么通常使用多模板来构建同源模型，使建模过程中模型的每个部分都采用最合适的模板。



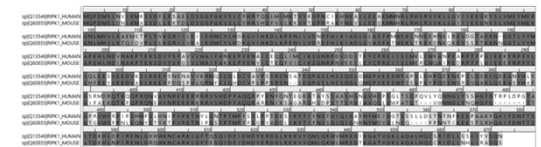
42

42

蛋白质三维结构的预测方法

同源建模的步骤（以DS MODELER为例）

2、序列比对



通过多序列比对将目标序列直接比至模板序列

43

43

蛋白质三维结构的预测方法

同源建模的步骤（以DS MODELER为例）

3、产生全原子模型（自动产生）

- 模型的PDF Total Energy越低，表明该模型在同源约束条件下优化的越好；模型同限定的同源约束条件偏差越小，该模型的可信度越大
- DOPE是一个基于原子统计势能的程序，主要用于模型评估。它的分数可以认为是衡量同一分子不同构象可信度的标准，能够帮助选择预测结构的最优模型。分数越低，模型质量越可靠

一般可以粗略地选取PDF Total Energy最低的模型作为最合理的初始模型；若其值相似，则可以利用DOPE score作为衡量模型质量的依据

44

44

蛋白质三维结构的预测方法

同源建模的步骤（以DS MODELER为例）

3、loop区和侧链的建模

loop区建模：主要是目标蛋白和模板蛋白的比对结果中存在缺口的部分如何处理的问题。第一种解决的方式是略去模板蛋白存在的残基，留下一个必须补上的缺口。另一种情况是将主链截断，插入缺少的残基。

侧链建模：构造各种可能的构象体，并利用基于能量的函数打分来实现侧链构象的选择的

45

45

蛋白质三维结构的预测方法

同源建模的步骤（以DS MODELER为例）

4、模型优化

模型产生初始坐标后，还要对分子结构进行进一步的优化，以此消除/缓解原子间的重叠或某些不合理构象（尤其是非保守区的构象）。优化一般采用分子力学或分子动力学

注意事项：

- 对柔性区域的构象进行构象分析可以采用分子动力学方法
- 在进行分子力学优化时，应尽量避免破坏模型的主链结构，一般要采取限制性优化方法
- 在修正过程中，应考虑溶剂效应的影响

46

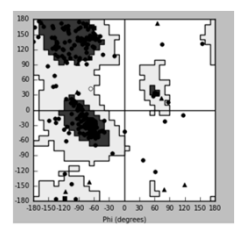
46

蛋白质三维结构的预测方法

同源建模的步骤（以DS MODELER为例）

5、模型评估：拉氏图

- 蛋白主链二面角的拉氏图（当落于核心区+允许区的氨基酸残基百分比>95%时，表明模型质量较高）



47

47

蛋白质三维结构的预测方法

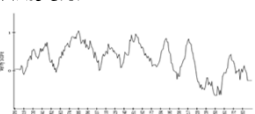
同源建模的步骤（以DS MODELER为例）

5、模型评估：Profile-3D

Profile-3D是UCLA的David Eisenberg教授开发的一种基于“穿线”（threading）法的模型评估程序。该方法采用3D-1D的打分函数来检测所构建模型与自身氨基酸序列的匹配度关系。分数越高，说明同源模型的可信度越大。

Protein	Verify Score	Expected High Score	Expected Low Score
P41131.M0002	179.12	195.312	87.6902

Verify Score越接近或高于Verify Expected High Score，模型的质量越好。





48

48

蛋白质结构的获取方法

基于人工智能的预测方法：AlphaFold

2020年12月1日，国际蛋白质结构预测竞赛CASP公布的数据显示，谷歌旗下人工智能技术公司DeepMind开发的深度学习算法AlphaFold的表现超过了大约100个其他团队，准确性达到了与实验室方法不分伯仲的水平，一举解决了困扰学界长达五十年之久的蛋白质折叠问题。

Experimental result
Computational prediction

49

蛋白质三维结构的预测方法

Article | Published: 15 July 2021

This is an unedited manuscript that has been accepted for publication. Nature Research are providing this early version of the manuscript as a service to our authors and readers. The manuscript will undergo copyediting, typesetting and a proof correction before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Highly accurate protein structure prediction with AlphaFold

John Jumper^{1,2}, Richard Evans¹, J. David Hasselis^{1,2}

nature

Explore content | Journal Information | Publish with us

Article | Published: 22 July 2021

This is an unedited manuscript that has been accepted for publication. Nature Research are providing this early version of the manuscript as a service to our authors and readers. The manuscript will undergo copyediting, typesetting and a proof correction before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Highly accurate protein structure prediction for the human proteome

Kathryn Tunyasuvunabool^{1,2}, Jonas Adler¹, J. David Hasselis^{1,2}

50

蛋白质三维结构的预测方法

<https://alphafold.ebi.ac.uk/>

Sequence: `MF033794.1` | 3 high priority | 5 A x 4

3D viewer

Model Confidence:

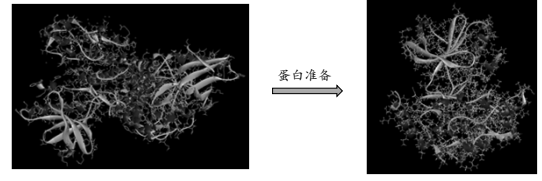
- Very High (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.



51

蛋白质结构的预处理



蛋白准备

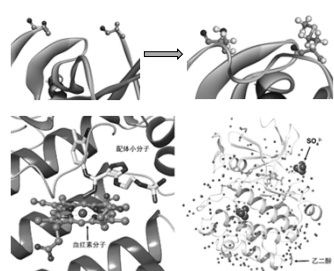
通过实验方法所测得的生物大分子三维结构或多或少存在一些缺陷，比如分辨率低导致模型质量不佳、晶体结构无H原子和电荷信息、在模型构建步骤中相对粗糙的力场带来的结构误差等，这就需要在做SBDD前对生物大分子（如蛋白）的三维结构进行准备和能量最小化。

52

蛋白质结构的预处理

蛋白结构准备的一般原则：

- 缺失氨基酸及其侧链的修复
- 正确设置键级、计算电荷和加氢
- 限制性优化
- 水的处理
- 辅因子的处理
- 其它小分子的处理

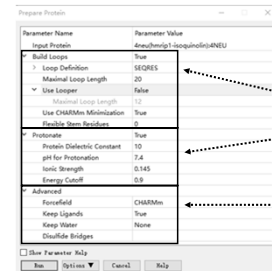


53

蛋白质结构的预处理

DS中蛋白结构准备的模块（Prepare Protein）：

- 缺失氨基酸及其侧链的修复
- 计算电荷、质子化和加氢
- 限制性优化
- 水的处理
- 辅因子的处理
- 其它小分子的处理



54

第二节 Part Two

基于结构的药物设计 (SBDD) 基本概念与理论

55

基本概念与理论

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

- 配体-受体相互作用理论 ★
- 水分子在药物设计中的作用
- 热力学理论在药物设计中的应用



56

配体-受体相互作用理论

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

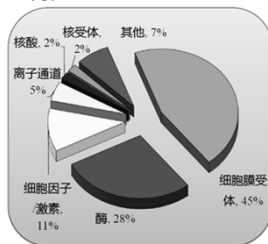
目前所发现的靶点的类型

靶点 (受体)：能识别和结合生物活性物质 (配体)，并产生生物效应的结构。

生物活性物质 (配体)：

内源性活性调节物：维持机体机能的基本生理机制，如多巴胺、激素等。

外源性药物：干预机体生理生化作用。



57

配体-受体相互作用理论

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

靶点学说 (受体理论)

一种有效的药物必须符合以下三个要求：

- 作用于靶点 (与机体内的某一种或多种靶点发生相互作用) (药效学要求)
- 暴露于靶点 (药物到达靶点、达到适宜的浓度 C_{max} 并维持足够的时间 AUC) (药代动力学要求)
- 不与其他无关靶点作用 (无 off target 效应) (安全性要求)

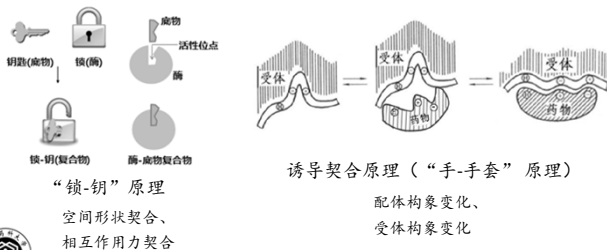


58

配体-受体相互作用理论

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

配体-受体相互作用方式



59

配体-受体相互作用理论

中国药科大学药学院
SCHOOL OF PHARMACY CHINA PHARMACEUTICAL UNIVERSITY

分子识别

药物分子与靶蛋白发生相互作用的驱动力是分子识别。

定义：药物在特定条件下通过分子间作用力 (可逆) 与靶点相结合的过程。其中，某些药物也可形成共价键 (不可逆)。

分子识别的两个决定性作用：互补性 (形状和电性互补) 和预组织作用 (去溶剂化和构象扰动)。



60

59

60

配体-受体相互作用理论			
配体-受体相互作用的类型			
作用类型	键能 (kJ/mol)	有效半径 (nm)	键长
共价键	140-800		
非键相互作用	范德华力	0.3-1.9	0.2-0.4
	疏水键	3.4	0.2-0.4
	π - π 堆积	8-12	0.5
	氢键	4-30	0.25-0.35
	阳离子- π	5-12	
静电作用	离子键	20-40	0.5-1.0
	离子-偶极键	~2	0.5-1.0
	偶极-偶极键	0.5	0.2-0.4

61

配体-受体相互作用理论

范德华力 (van der Waals Force, VDW)

- 相邻的电中性原子互相靠近, 会有短暂的微弱吸引力, 即范德华力
- 主要的分子间作用力
- 原子间距离为0.3~0.5 nm, 范德华力的能量为1.9 kJ/mol
- 主要分为: 色散力、诱导力、取向力

62

配体-受体相互作用理论

范德华力 (van der Waals Force, VDW)

色散力 (Dispersion force): 所有分子或原子之间都存在, 由于电子的运动产生的瞬时偶极键的作用力。

- 分子的变形性越大 (一般分子量越大, 变形性越大), 色散力越大
- 分子的电离势越低 (分子内所含的电子数越多), 色散力越大。

63

配体-受体相互作用理论

范德华力 (van der Waals Force, VDW)

诱导力 (Induction force): 在极性分子-非极性分子之间, 极性分子-极性分子之间, 都存在诱导力。

- 极性分子与非极性分子接近时, 极性分子的永久偶极产生的电场使非极性分子极化产生诱导偶极。永久偶极与诱导偶极间的吸引力称为诱导力。
- 诱导力同样存在于极性分子之间, 对极性分子而言, 诱导力是附加的取向力。

64

配体-受体相互作用理论

范德华力 (van der Waals Force, VDW)

取向力 (Orientation force): 取向力发生在极性分子-极性分子之间。

- 当极性分子互相接近时, 分子的永久偶极之间同极相斥、异极相吸, 使分子在空间按一定取向排列吸引, 而处于较稳定的状态。这种永久偶极间的吸引力称为取向力。
- 取向力与分子的偶极矩平方成正比, 即分子的极性越大, 取向力越大。

65

配体-受体相互作用理论

疏水键 (Hydrophobic bond)

疏水键是两个不溶于水的分子间的相互作用。这种因能量效应和熵效应等热力学作用使疏水基团在水中的相互结合作用称为疏水键。

- 疏水作用与分子中疏水基团和片段的数目成正比, 烷基越多, 疏水性越强。

66

配体-受体相互作用理论

疏水键的本质

疏水基团 + 有序水分子 + 表面暴露的疏水区域 + 亲水区域

水分子有序排列，熵值降低，系统能量增加，产生表面张力

为克服表面张力，疏水表面卷曲相互结合，疏水面积减少，水分子排挤成无序状态，熵值增加，系统能量降低。

药物分子与受体间疏水相互作用的示意图

67

配体-受体相互作用理论

氢键 (Hydrogen bond)

氢键: 氢原子与电负性大的杂原子共价结合，与另一具有未共用电子对的杂原子以氢为媒介，形成以X-H...Y形式的一种弱的静电引力。

- 氢键具有饱和性和方向性，在药物和受体形成氢键时，双方的原子和基团在空间的取向和距离决定了结合强度
- 氢键在化学和生物学的多个领域中扮演着重要的角色，是驱动生物分子间可逆相互作用的主要力量

68

配体-受体相互作用理论

氢键的方向性

3D mapping with hydrogen fluoride

Interaction energy (normalized)

$\theta = 90^\circ$

69

配体-受体相互作用理论

氢键供体能量强度

使用HCN作为受体探针计算34个氢键供体形成氢键的能量大小，排序结果如下：

Donor	Interaction Energy (kcal/mol)	Donor	Interaction Energy (kcal/mol)
1 methane	-0.5	18 dimethyl sulfoxide	-4.5
2 ethylene	-0.5	19 hydrogen fluoride	-4.7
3 ammonia	-1.8	20 difluoromethane	-4.8
4 methanol	-2.0	21 difluoromethane	-4.8
5 methylamine	-2.0	22 hydrogen cyanide	-5.1
6 methanethiol	-2.0	23 nitrobenzene	-5.5
7 diethylamine	-2.4	24 acetic acid	-5.5
8 acetone	-2.6	25 phenol	-5.8
9 benzene	-2.6	26 methanethiol	-6.4
10 acetone	-2.6	27 5-methylthiazole	-6.5
11 benzene	-3.0	28 hydrogen fluoride	-6.8
12 guanidine	-3.4	29 5-methylthiazole	-6.8
13 water	-3.8	30 benzamide	-7.7
14 aniline	-3.6	31 guanidine	-10.7
15 methanol	-4.1	32 indole	-12.7
16 acetamide	-4.5	33 dimethylammonium	-18.7
17 N-methylthiazole	-4.5	34 methylammonium	-21.2
18 dimethylamine	-4.7		
19 hydrogen fluoride	-4.8		
20 difluoromethane	-4.8		
21 difluoromethane	-4.8		
22 hydrogen cyanide	-5.1		
23 nitrobenzene	-5.5		
24 acetic acid	-5.5		
25 phenol	-5.8		
26 methanethiol	-6.4		
27 5-methylthiazole	-6.5		
28 hydrogen fluoride	-6.8		
29 5-methylthiazole	-6.8		
30 benzamide	-7.7		
31 guanidine	-10.7		
32 indole	-12.7		
33 dimethylammonium	-18.7		
34 methylammonium	-21.2		

Figures 1. Donor groups ranked by their interaction energy (in kcal/mol, below each molecule) with the HCN acceptor probe 61. Donor target atoms are colored in blue. Values annotated with "(*)" correspond to geometries that cannot be optimized and/or showing imaginary frequencies.

QM: B3LYP-D3BJ/cc-pVTZ

70

配体-受体相互作用理论

氢键受体能量强度

使用HCN作为受体探针计算34个氢键受体形成氢键的能量大小，排序结果如下：

Donor	Interaction Energy (kcal/mol)	Donor	Interaction Energy (kcal/mol)
80 pyridine	-7.8	91 imidazole	-9.8
81 urea	-8.2	92 N-methylacetamide	-9.0
82 acetamide	-8.3	93 N,N-dimethyl benzamide	-9.0
83 ammonia	-8.3	94 dimethyl sulfoxide	-10.2
84 pyridine-N-oxide	-8.3	95 dimethyl sulfoxide	-9.3
85 N,N-dimethyl acetamide	-8.4	96 guanidine	-10.0
86 trimethylamine	-8.4	97 trimethylamine-N-oxide	-13.3
87 N-methylacetamide	-8.5	98 5-methylthiazole	-12.0
88 methylamine	-8.6	99 5-methylthiazole	-12.0
89 dimethylamine	-8.6	100 acetate	-20.5

QM: B3LYP-D3BJ/cc-pVTZ

71

配体-受体相互作用理论

非典型氢键和分子内氢键

非典型氢键

分子内氢键

72

配体-受体相互作用理论



静电作用 (Electrostatic interaction)

- 离子键 (Ionic bond)
- 离子-偶极作用 (Ion-dipole interaction)
- 偶极-偶极作用 (Dipole-dipole interaction)



73

73

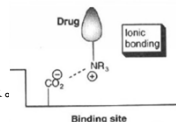
配体-受体相互作用理论



离子键 (Ionic bond)

离子键：两个具有相反电荷的离子间的静电引力，也叫离子离子相互作用。

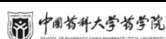
- 离子键无方向性，是随机转运的药物与受体的初始感受，对于两者间的趋近与识别有重要作用。
- 作用能较强，但在水中，去水合作用与静电引力对复合物形成的能量贡献相反，因而形成抵消效果。



74

74

配体-受体相互作用理论



离子-偶极作用 (Ion-dipole interaction)：带电荷的原子或基团与含有偶极的基团之间的静电引力。对稳定药物受体复合物起到重要作用，但弱于离子键。

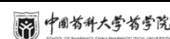
偶极-偶极作用 (Dipole-dipole interaction)：两个偶极分子或偶极键之间的静电引力，弱于离子-偶极作用。



75

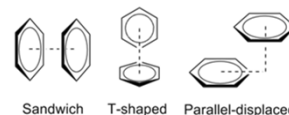
75

配体-受体相互作用理论

 π - π 相互作用 (π - π interaction)

π - π 相互作用：芳环含有环形 π 键，芳环之间的 π 键可发生 π - π 相互作用，又称为 π - π 堆积作用 (π - π stacking)。

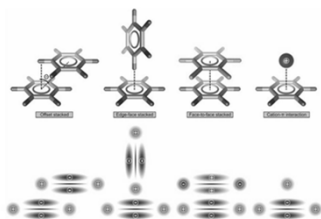
- 作用力较弱，但在生物体内普遍存在，作用能量大约8~12 kJ/mol。
- 芳环堆积的三种类型



76

76

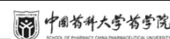
配体-受体相互作用理论

从静电吸引角度分析 π 体系的各种相互作用

77

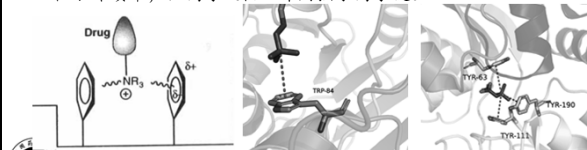
77

配体-受体相互作用理论

阳离子- π 相互作用 (Cation- π interaction)

阳离子- π 相互作用：由有机或无机阳离子与 π 电子体系的负电势之间的静电引力，能力强度与氢键相近。

- 在水介质中，阳离子- π 相互作用强于离子键。



78

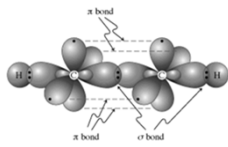
78

配体-受体相互作用理论



共价键 (Covalent bond)

键能高, 约140~800 KJ/mol, 结合牢, 驻留时间较长, 治疗指数高。
除了被特异性酶酶解而发生共价键断裂外, 不易被破坏。因此产生的药物作用持久, 为不可逆过程。



79

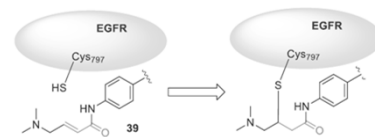
79

配体-受体相互作用理论



共价药物

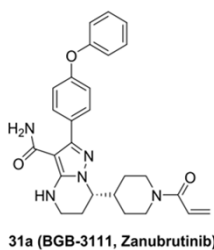
共价键键合类型常见于作用于病原体或肿瘤细胞的靶点。药物的共价基团往往具有较高的化学活性而缺乏特异选择性。



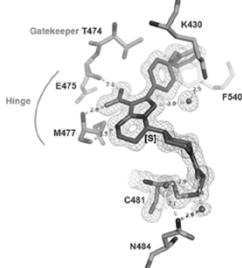
80

80

配体-受体相互作用理论



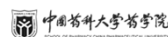
31a (BGB-3111, Zanubrutinib)



81

81

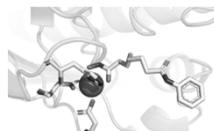
配体-受体相互作用理论



配位键 (Coordinate bond)

配位键: 又称配位共价键, 一种特殊的共价键, 其共用的电子对是由其中一原子独自供应, 另一原子提供空轨道。

- 过渡金属的核电荷高, 半径小, 有空的d轨道和自由的d电子, 它们容易接受配位体的电子对, 又容易将d电子反馈给配位体。因此, 它们都能形成稳定的配合物。

SAHA与HDAC8的晶体结构
(PDB ID: 1T69)

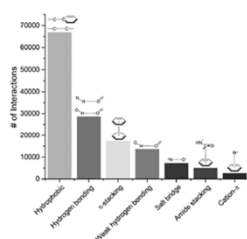
82

82

配体-受体相互作用理论



PDB中出现频率较高的非键相互作用



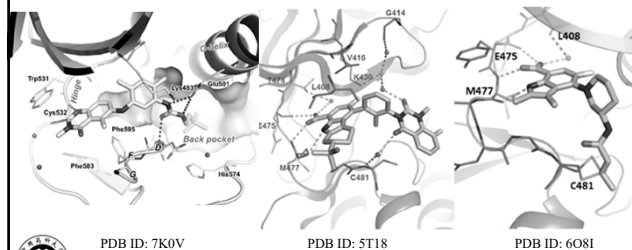
83

83

配体-受体相互作用理论



非键相互作用分析



PDB ID: 7K0V

PDB ID: 5T18

PDB ID: 6O81



J. Med. Chem. 2019, 62, 3228-3250

84

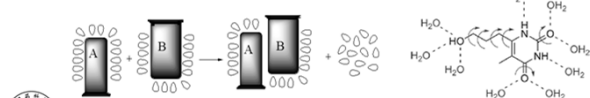
84

水分子在药物设计中的作用



药物设计时需要考虑的溶剂化效应

- 药物和受体在结合前都处于水的溶剂化状态，为了双方的结合需去溶剂化，这是耗能过程，能量由结合能得到补偿。
- 如果去溶剂化能大于结合能，药物难以与受体稳定结合，所以分子设计时要确定形成氢键的因素是否是多余。



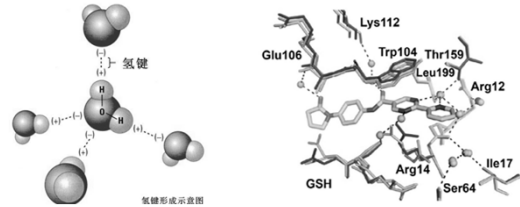
85

85

水分子在药物设计中的作用



水分子既是氢键给体也是氢键受体，可参与药物与受体的结合



86

86

水分子在药物设计中的作用



研究水分子的方法

- 基于经验：文献、蛋白中水分子的成键情况、比较同一蛋白中同一位置的水分子出现概率
- 基于计算：WaterMap、SZMAP、3D-RISM等
- 通过上述分析，我们可以知道：
 - (1) 哪些水分子在对接时应该保留并当成受体的一部分，哪些水应该是被删除掉的。
 - (2) 在水分子能量有利区设计分子或片段可以保留水或模拟水的作用，而在不利区可以替换或改善静电作用来实现化合物的优化。



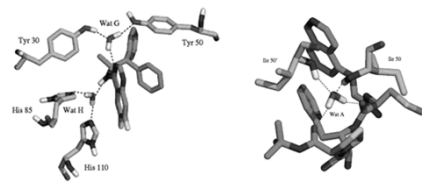
87

87

水分子在药物设计中的作用



基于经验：蛋白中水分子的成键情况



配体小分子与水分子相互作用

88

88

水分子在药物设计中的作用



基于计算：WaterMap

- 由Schrödinger公司开发的一种较为流行的基于分子动力学模拟（MD）的水分子位点预测工具，可预测蛋白质结合位点中水分子的位置和能量。
- WaterMap算法步骤：1、进行约2ns的约束性MD模拟（溶质约束），根据密度分布对水分子位置和方向进行聚类；2、使用非均相溶剂理论计算每个水合位点的热力学性质（相互作用能、焓增和自由能），以此来决定靶向哪个特定的水分子。



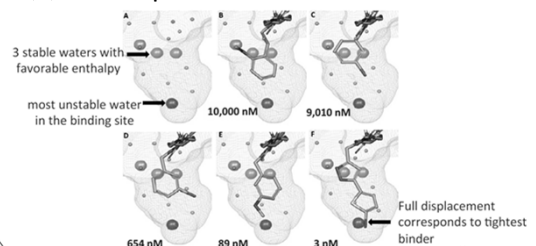
89

89

水分子在药物设计中的作用



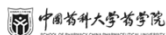
基于计算：WaterMap



90

90

水分子在药物设计中的作用



基于计算: SZMAP (solvent-zap-map)

- SZMAP是OpenEye应用软件包中的一个模块,是一款水分子位置与稳定性预测软件。无需分子动力学模拟采样。
- SZMAP采用显式水分子探针在高介电连续溶剂(high-dielectric continuum solvent)中快速计算分子表面附近溶剂化自由能的大小与分布。
- 它可以预测特定水分子对配体结合的稳定化或不稳定化效应,还可以识别结合水的位置和哪些位置的水是无序的。在结合位点,更好地理解这些效应有助于先导化合物的优化和药物设计其他方面问题的研究。

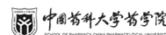


blog.molcalc.com.cn/2017/06/16/openeye-szmap-water.html

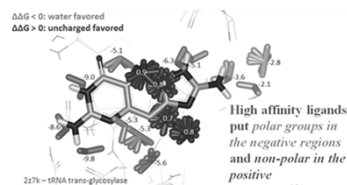
91

91

水分子在药物设计中的作用



基于计算: SZMAP (solvent-zap-map)



高结合亲和力的配体通常有极性基团处于负的自由能区,而非极性基团位于正自能能区。

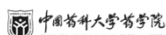


blog.molcalc.com.cn/2017/06/16/openeye-szmap-water.html

92

92

水分子在药物设计中的作用



基于计算: 3D-RISM (three-dimensional reference interaction site model)

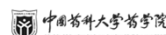
- 该程序集成在MOE、Flare和Amber中。
- 计算水H和O密度的时间平均分布,以及用于分析溶剂稳定性和溶剂化对结合自由能的贡献的自由能图。3D-RISM的独特之处在于它能够包含各种浓度的盐或疏水分子作为溶剂的一部分,从而生成离子,疏水基团以及水的密度图。
- 基于力场,无需分子动力学模拟采样。



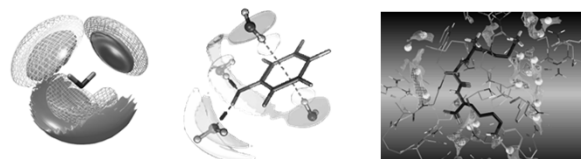
93

93

水分子在药物设计中的作用



基于计算: 3D-RISM (three-dimensional reference interaction site model)



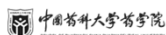
O (蓝)和H (白)原子的密度分布



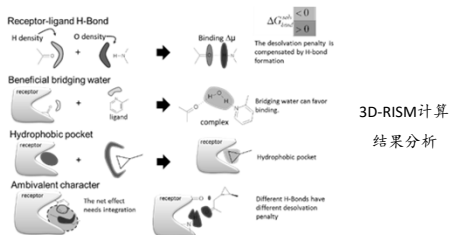
94

94

水分子在药物设计中的作用



基于计算: 3D-RISM (three-dimensional reference interaction site model)



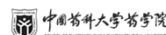
3D-RISM计算
结果分析



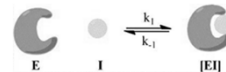
95

95

热力学理论在药物设计中的应用



多数药物与受体作用是以非共价键形式结合,形成的复合物与游离的配体和受体之间呈动态平衡。



这种结合作用是由热力学自由能(ΔG)变化所驱动,结合越强,配体-蛋白复合物的离解常数(K_d)越小, ΔG 负值越大。 ΔG 与 K_d 的关系由吉布斯-范特霍夫(Gibbs-Van't Hoff)方程表征:

$$\Delta G = RT \ln K_d$$

$R = 8.13 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$, T 为绝对温度。当 $T = 300 \text{ K}$ 时, $\Delta G = 5.7 \log K_d (\text{kJ/mol})$



96

96

热力学理论在药物设计中的应用



根据吉布斯-亥姆霍兹 (Gibbs-Helmholtz) 方程, ΔG 是由焓 (ΔH) 和熵 ($-T\Delta S$) 构成, 即:

$$\Delta G = \Delta H - T\Delta S$$

配体与受体结合产生自发过程的为负值, 所以从上式看, ΔH 为负值和 $-T\Delta S$ 为负值, 有利于结合过程。 ΔG 中 ΔH 和 $T\Delta S$ 有各自的贡献, 参与并维持着游离的配体和受体与复合物之间的平衡作用。



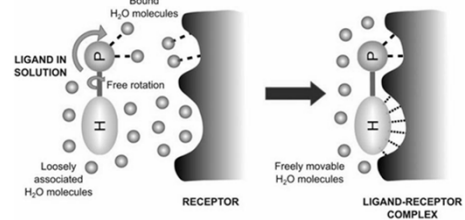
97

97

热力学理论在药物设计中的应用



蛋白-配体结合的过程



98

98

热力学理论在药物设计中的应用



配体与受体结合的焓-熵变化, 常常是有利的 ΔH 被不利的 $-T\Delta S$ 所抵消, 或者有利的 $-T\Delta S$ 以不利的 ΔH 作负性补偿, 这就是配体-受体结合作用的**焓-熵补偿 (enthalpy-entropy compensation)**。

受体-配体结合过程	焓 (ΔH)	熵 ($-T\Delta S$)
极性基团的去溶剂化	不利 (正值)	-
氢键的形成	有利 (负值)	-
静电相互作用	有利 (负值)	-
通过形状的互补与契合形成的范德华相互作用	有利 (负值)	-
疏水相互作用	-	有利 (负值)
配体和受体自由度的降低	-	不利 (正值)



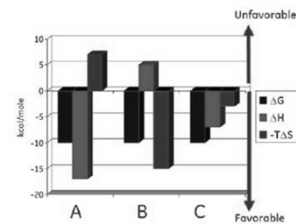
99

99

热力学理论在药物设计中的应用



如果 $|\Delta H| > |-T\Delta S|$, 则该结合为焓驱动结合, 反之, 则为熵驱动。



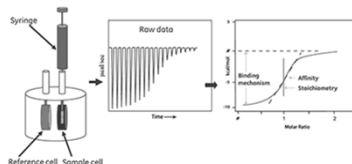
10

100

热力学理论在药物设计中的应用



• 热力学量的实验测定——等温滴定量热法 (isothermal titration calorimetry, ITC)



➢ 可直接测量生物分子结合过程中释放或吸收的热量。

➢ 准确地确定结合常数 (K_D)、反应化学量 (n)、焓 (ΔH) 和熵 ($-T\Delta S$)。



10

101

热力学理论在药物设计中的应用



• 热力学量的计算方法——结合自由能 (binding free energy) 的计算

在分子模拟领域, 准确计算结合自由能仍然是一个挑战。为此, 人们发展了许多方法。如:

➢ 自由能微扰 (FEP) 与热力学积分 (TI): 计算量很大, 不适用于生物大分子体系

➢ MM-PBSA (Molecular Mechanics-Poisson Boltzmann Surface Area, 分子力学泊松玻尔兹曼表面积): 准确度不如 FEP, 但计算量小, 在分子识别, 区分结合的强弱方面是一种有效的方法, 可得到 ΔG 、 ΔH 和 $-T\Delta S$ 的具体值。



102

102

热力学理论在药物设计中的应用

热力学量的计算方法——案例分析

脂蛋白相关磷脂酶A2(Lp-PLA2)抑制剂的空间结合模式与热力学特征比较

PDB ID: 5I9I PDB ID: 5I8P

J Med Chem 2016; 59 (10): 5115-5120

103

热力学理论在药物设计中的应用

热力学量的计算方法——案例分析

ITC测试可知, Dara与Lp-PLA2结合过程中产生的结合焓变化为 -53.67 KJ/mol, 比化合物1多了 -12.24 KJ/mol, 但两者与蛋白最终的结合自由能仅相差3.3 KJ/mol, 这主要是由于Dara化学结构柔性较大, 在与蛋白结合过程中结合熵损失明显大于化合物1, 最终导致两者与蛋白的结合自由能相差不大。

Inhibitors	K_d (M)	ΔG (KJ/mol)	ΔH (KJ/mol)	$-T\Delta S$ (KJ/mol)	IC_{50} (nM)
Darapladib	$(4.97 \pm 0.44) \times 10^{-8}$	-42.29 ± 0.22	-53.67 ± 1.12	11.38 ± 1.32	0.7
1	$(1.85 \pm 0.18) \times 10^{-7}$	-38.99 ± 0.24	-41.43 ± 0.04	2.45 ± 0.27	1.7

J Med Chem 2016; 59 (10): 5115-5120

104

第三节

Part Three

分子对接技术

Molecular docking

10

105

分子对接技术

- 分子对接的原理
- 分子对接的分类
- 常用的分子对接算法
- 分子对接的核心步骤
- 分子对接的一般流程
- 分子对接的应用
- Discovery Studio中的分子对接算法简介
- Discovery Studio中的非键相互作用分析
- 分子对接计算的注意点
- 分子对接尚未解决的问题

106

106

分子对接的原理

定义: 利用算法将配体小分子(如药物)放置到受体大分子(如靶标蛋白)的结合位点(Binding site), 预测小分子与受体结合构象(Pose)及作用能的过程。

- 分子对接是一种广泛应用的研究分子间相互作用的分子模拟方法, 其过程涉及小分子与受体之间的空间匹配和能量(性质, 电性等)匹配。
- 利用分子对接方法研究蛋白-配体作用模式在SBDD中的应用非常广泛。

107

107

分子对接的原理

分子对接的最初思想起源于1894年Fisher提出的受体学说。Fisher认为, 药物与体内的蛋白质大分子即受体会发生类似药物与锁的识别关系, 称为“锁钥模型”, 这种识别关系主要依赖两者的空间匹配。

Emil Fischer (1852 - 1919)

E. Fischer, Ber. Dtsch. Chem. Ges. 1894, 27, 2985.

"The specificity of an enzyme (the lock) for its substrate (the key) arises from their geometrically complementary shapes" (1894)

Substrate

Enzyme

108

108

分子对接的原理



真正算法上的实现是在1982年，美国加州大学旧金山校区（UCSF）的Irwin Kuntz教授率先提出了“分子对接”的概念，并发表了第一个分子对接软件“DOCK”，至今仍被广泛应用。



109

109

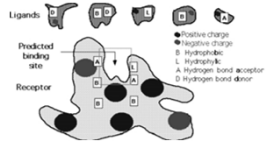
分子对接的原理



分子对接中的两大关键问题：空间匹配和能量匹配

- 空间匹配是分子间发生相互作用的前提。
- 能量匹配是分子间保持稳定结合的基础。

配体分子进入结合位点时，通过一定的程序计算它们之间的结合模式和结合能，并对结果进行评价，通过打分函数评判配体-受体的结合程度



110

110

分子对接的分类



根据对接过程中是否考虑研究体系的构象变化，将其分为

刚性对接：将配体和受体构象都当做刚性处理，以降低计算量

半柔性对接：受体保持刚性，配体则完全当做柔性分子进行处理，当前的大多数分子对接软件都采取这种模式

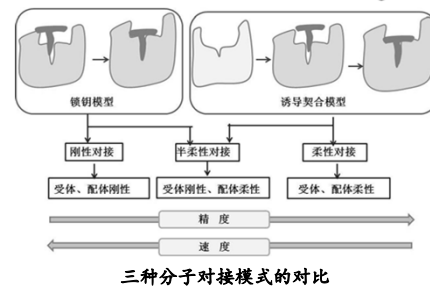
柔性对接：对接过程既考虑小分子柔性，又可考虑受体部分区域（如结合口袋中的残基）的柔性



111

111

分子对接的分类



三种分子对接模式的对比



112

112

常用的分子对接算法



刚性对接、半柔性对接：LibDock、CDOCKER、Glide、GOLD、AutoDock、DOCK、MOE_dock、FlexX、Surflex-dock

柔性对接：Flexible docking、Induced fit docking (IFD)

片段对接：MCSS

共价对接：Covdock、GOLD、MOE

蛋白-蛋白对接：ZDOCK、ICM-Docking



113

113

分子对接的核心步骤



搜索算法：搜索配体与受体的可能结合模式

- 配体的可能构象
- 配体在活性位点的可能取向

打分函数：对配体与受体的结合模式或进行打分，寻找最合理的结合

- 经验打分函数
- 能量打分函数
- 一致性打分 (Consensus Score)



11

114

分子对接的核心步骤：搜索算法



在分子对接中，配体小分子在受体结合口袋中会有不同的取向和构象，其中配体构象搜索方法起重要作用。对接问题的核心，就是在指定的构象空间内，采取合适的算法搜索合适的配体构象的过程。

- **系统性搜索 (Systematic search)**：穷举搜索、基于片段的搜索、构象系综搜索
- **启发式搜索 (Heuristic search)**：蒙特卡洛算法、遗传算法、模拟退火算法等
- **确定性搜索 (Deterministic search)**：分子动力学模拟



11

115

分子对接的核心步骤：搜索算法



系统搜索，就是在构象搜索时尽可能的去系统地遍历配体全部的自由度或结构参数的组合。但是，随着分子可旋转键数目的增多，配体构象空间会急剧膨胀，要做到完整采样并不现实。分子对接中常用的系统搜索算法有穷举搜索 (Exhaustive Search)、基于片段的搜索 (Fragment-based Search) 和构象系综搜索 (Conformational Ensemble Search) 等。



116

116

分子对接的核心步骤：搜索算法



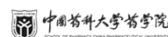
启发式搜索 (随机搜索法)，需对配体三维坐标进行随机改变产生新的构象，然后根据预先定义的评价函数对新构象进行取舍，通过不断迭代收敛于全局最优或近似最优构象。常用的启发式搜索算法包括蒙特卡洛 (Monte Carlo, MC) 算法、遗传算法 (Genetic Algorithm, GA)、模拟退火 (Simulated Annealing, SA) 算法和群智能 (Swarm intelligence, SI) 算法等。



117

117

分子对接的核心步骤：搜索算法



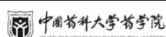
确定性搜索，初始状态可以确定下一状态的改变趋势，即向能量减小的方向进行。如果体系初始状态完全相同，当使用相同的参数进行构象搜索，就会到达相同的终态。这种方法在进行构象搜索时一般很难越过势垒，因此容易陷入局部最小值。分子对接中常用的确定性搜索算法主要是分子动力学模拟，代表性程序为基于CHARMM力场的CDOCKER。



118

118

分子对接的核心步骤：打分函数



打分函数是影响分子对接精度的另一个重要因素，其主要作用是评估配体和受体分子间的亲和力。用于评价同一分子的各种构象以及不同分子的最佳构象。

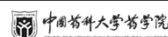
- 基于力场的打分函数 (Forcefield)
- 基于经验的打分函数 (Empirical)
- 基于知识的打分函数 (Knowledge-based)
- 一致性打分 (Consensus score)



119

119

分子对接的核心步骤：打分函数



基于力场的打分函数一般采用分子力场的非键作用能项来计算受体和配体之间的相互作用能。只考虑热焓对能量的贡献，不考虑熵的影响，一般情况下，采用标准力场的真空静电和范德华作用两项的加和来计算，有些打分函数还会考虑氢键项的贡献。

- 基于力场的打分函数通式： $\Delta G_{\text{binding}} = \Delta E_{\text{vdw}} + \Delta E_{\text{ele}} + [\Delta E_{\text{Hbond}}]$
- DOCK程序中采用AMBER的能量函数： $E = \sum_{i=1}^{uq} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right)$
- 其它常用的基于力场的打分函数：AutoDock的打分函数、Goldscore等



120

120

分子对接的核心步骤：打分函数



经验打分函数是受体和配体结合的自由能可以近似为不相关的不同能量项的贡献。通过计算并加和各个能量项求得总的结合自由能。

- 这些能量项包括：氢键作用能、疏水作用能、离子化、熵的贡献、金属离子作用等
- 常用的经验打分函数如PLP、ChemScore、GlideScore、LUDI等
- ChemScore采用如下函数形式：

$$\text{ChemScore} = S_{\text{Hbond}} + S_{\text{metal}} + S_{\text{lipophilic}} + P_{\text{rotor}} + P_{\text{strain}} + P_{\text{clash}} + [P_{\text{covalent}} + P_{\text{constraint}}]$$



121

121

分子对接的核心步骤：打分函数



基于知识的打分函数，通过加和受体和配体之间的原子对统计势（Pairwise Statistical Potential）来计算配体和受体之间的结合能力。

$$A = \sum_i \sum_j^{\text{pro}} \omega_{ij}(r)$$

- 距离依赖的原子对之间的作用势 $\omega_{ij}(r)$ 由下式计算得到：

$$\omega_{ij}(r) = -k_B T \ln[g_{ij}(r)] = -k_B T \ln\left[\frac{\rho_{ij}(r)}{\rho_{ij}^*}\right];$$

$\rho_{ij}(r)$ 是原子对*i-j*在距离*r*时的密度； ρ_{ij}^* 是原子对在参考态时的密度

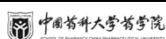
- 常用的基于知识的打分函数：PMF、IT-Score、DrugScore等



122

122

分子对接的核心步骤：打分函数



将几个不同打分函数的结果相结合，这种方法通常被称为“**一致性评分**”。根据这种方法，使用多个打分函数来评分结合构象，相当于对接结果进行二次打分（一致性打分）是非常有效的消除假阳性方法。

Models	Score1	Score2	Score3	Score4	Consensus
mol1	12 0	55 0	63 0	241 0	0
mol2	22 0	99 0	113 0	283 1	1
mol3	112 1	52 1	221 1	299 1	4
mol4	79 0	80 1	192 0	201 0	1
mol5	38 1	77 0	193 1	263 0	2

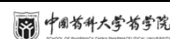
各打分函数的截断值设置为40%，即按每个打分函数排序，排名前40%的分子设为阳性分子



123

123

分子对接的一般流程



- 受体结构的准备
- 定义结合位点
- 配体结构准备
- 对接过程（参数设置）
- 对接结果分析



124

124

分子对接的一般流程



受体结构的准备

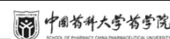
- 1、靶点检验：X-ray晶体结构要检查分辨率（分辨率越高越好）、种属、突变情况、结构的完整性等
- 2、受体结构处理（以晶体结构为例）：加氢、加电荷、带电残基的质子化状态、水分子的处理、助溶剂分子的删除、补充缺失的残基



125

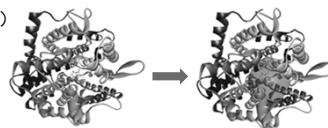
125

分子对接的一般流程



结合位点确定

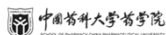
- 1、复合物中的配体去除（最简单）
- 2、同源蛋白模型的结合位点信息
- 3、定点突变实验
- 4、程序算法：Cavity（DS）、SiteMap（Schrödinger）、SiteFinder（MOE）



126

126

分子对接的一般流程



配体（小分子）结构准备

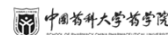
- 1、3D结构转化
- 2、加氢、加电荷、质子化状态 (pH: 7.4)、手性和互变状态确定
- 3、如果是小分子化合物库，还需进行类药性分析和假阳性子结构过滤



127

127

分子对接的一般流程



对接过程（参数设置）

1、对接算法的选择

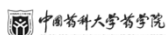
使用重对接（native docking），将复合物晶体结构中的小分子重新对接到该复合物中，比较对接构象与晶体构象的均方根偏差值（RMSD），该值越小（通常要小于2 Å）说明对接构象与晶体构象越接近，对接算法对小分子结合构象的重现性越好，对接算法越可靠。



128

128

分子对接的一般流程



重对接实验

PDB(complex)	CDOCKER	GOLD	Surflex-Dock	Glide (SP)
3GEN	0.2567	0.3374	0.5547	1.2907
3K54	1.0592	1.1577	2.5984	0.7963
3OCS	1.0082	0.4602	0.7332	1.2063
3PJ1	1.6826	11.0924	2.5355	1.9593
3PIX	0.9077	2.1161	1.4900	1.1274
3PIY	0.6866	5.9909	1.5996	0.4177
3PIZ	1.4018	0.8131	1.6543	1.7208
3PJ3	0.4107	0.5097	0.7705	0.8612
av ^a	0.9267	2.8100	1.4920	1.1720
std ^b	0.4772	3.8330	0.7861	0.4990

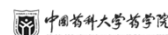


J Mol Graph Model 2015 Vol. 60 Pages 142-154

129

129

分子对接的一般流程



对接过程（参数设置）

2、打分函数的评价（同一个受体结构）

- 同一配体的不同结合模式（pose）：对于重对接而言，比较各pose的打分值与其RMSD值的相关性，相关性越高，则说明打分函数越能反映对接构象的准确性
- 不同活性的配体：比较各配体打分最高的结合构象的活性值与打分值的相关性，相关性越高，说明该打分函数能较为准确的反映配体的活性

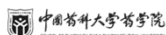
（注：选择不同的配体活性值之间要至少相差一个数量级。）



130

130

分子对接的一般流程



对接过程（参数设置）

3、其它参数设置

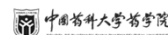
可先使用软件预设值，然后再根据初步对接结果（如重对接结果）进行调整



131

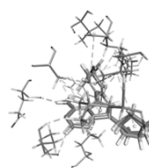
131

分子对接的一般流程



对接结果分析

1、RMSD值计算（重对接）



3GEN - RMSD Report		
Heavy Atom RMSD to 3GEN 11		
Name	Reference	RMSD (Å)
3GEN 1	3GEN 11	0.1173
3GEN 9	3GEN 11	0.6376
3GEN 38	3GEN 11	7.3890
3GEN 39	3GEN 11	0.9375
3GEN 40	3GEN 11	3.4741
3GEN 41	3GEN 11	0.9489
3GEN 42	3GEN 11	3.2732
3GEN 43	3GEN 11	2.7862
3GEN 60	3GEN 11	8.4332
3GEN 52	3GEN 11	3.4042

对接pose对应的RMSD值



132

132

分子对接的一般流程

中国药科大学药学院

对接结果分析

配体-受体非键相互作用分析 配体-受体相互作用2D图 受体结合口袋表面

133

133

分子对接的一般流程

中国药科大学药学院

对接结果分析

残基相互作用分析 多种打分函数分析 结合能分析
带溶剂化模型

134

134

分子对接的应用

中国药科大学药学院

- 确定配体和受体可能的结合构象，及其与靶标结合口袋内残基的相互作用细节，从而阐明活性化合物的作用机制，为分子的结构优化打下基础
- 对靶标进行基于结构的虚拟筛选，从而辅助发现对该靶标具有潜在活性的候选化合物
- 对同系列化合物进行分子对接，可以得到分子叠合构象，从而进行基于结构的3D-QSAR研究
- 通过反向对接找到小分子可能作用的靶标

135

135

Discovery Studio中的分子对接算法简介

中国药科大学药学院

锁-钥模型 诱导契合模型

刚性对接 半柔性对接 柔性对接

ZDOCK LibDock, CDOCKER Flexible Docking

精度 速度

136

136

Discovery Studio中的分子对接算法简介

中国药科大学药学院

	LibDock	CDOCKER	Flexible Docking
原理	基于热区匹配 半柔性	基于MD 半柔性	基于诱导契合 柔性
配体多构象	FAST、BEST、CAESAR	高温MD	FAST、BEST、CAESAR
受体多构象	\	\	ChiFlex
对接	High Quality、Fast Search for SASA.....	模拟退火	High Quality、Fast Search for SASA.....
能量最小化	局部最小化 SD、CG.....	全局最小化 CHARMm	全局最小化 CHARMm
相互作用能	\	CHARMm	CHARMm
评价函数	表面、化学环境匹配	分子力场	分子力场
速度	最快	慢	最慢
	虚拟筛选		机理解释/结构改造

137

137

Discovery Studio中的分子对接算法简介

中国药科大学药学院

LibDock

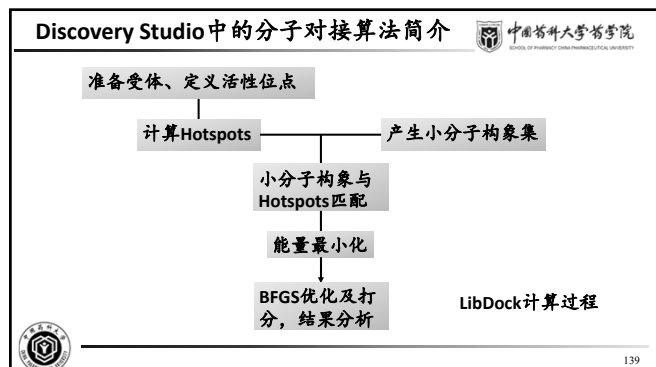
LibDock根据小分子构象与受体相互作用的热区（Hotspot）匹配的原理将这些构象刚性对接到受体的结合口袋中，其最大的优势在于速度快，可以并行运算，适合于对大规模数据库进行快速精确的虚拟筛选

极性格点 非极性格点

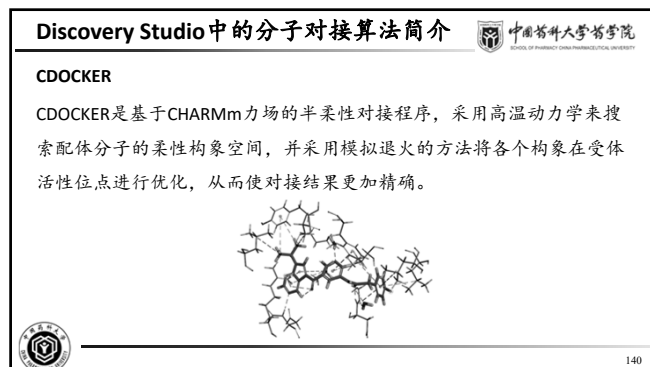
极性格点：受体中能与配体分子中可以形成氢键的原子相匹配的部分
非极性格点：受体中能与配体分子的非极性原子相匹配的部分

138

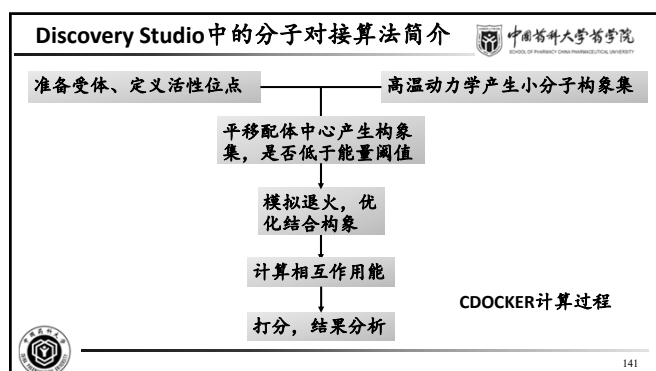
138



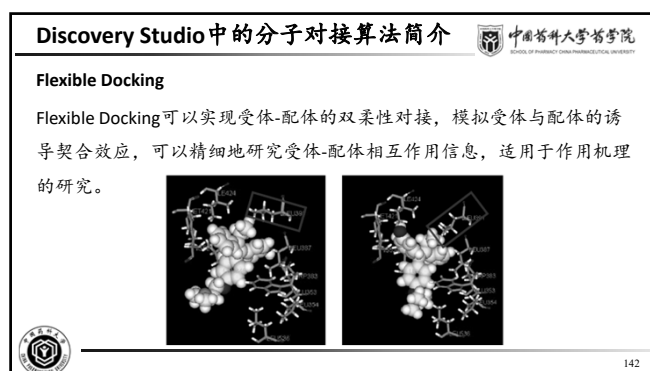
139



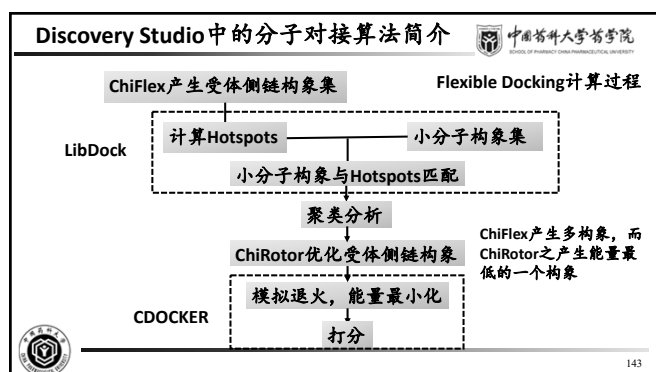
140



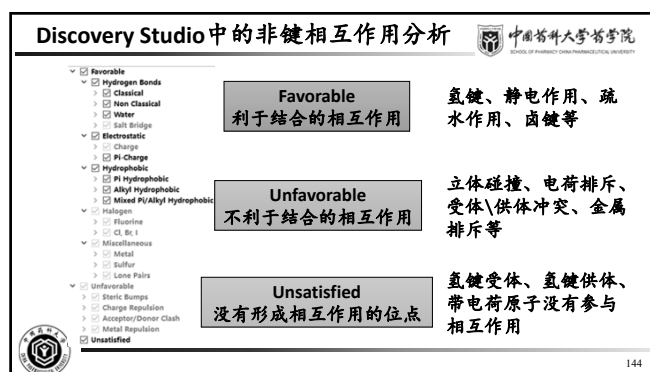
141



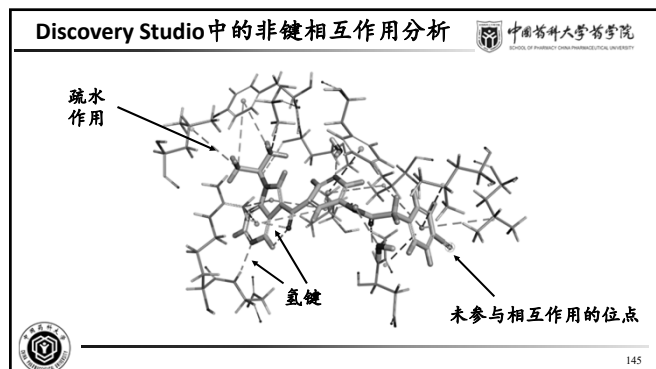
142



143



144



145

分子对接计算的注意点

小分子问题

- 小分子的起始构象、电荷和质子化状态对接结果有一定影响

蛋白受体问题

- 如何选择靶标晶体结构？定义活性位点的位置和大小？
- 活性位点氨基酸残基的结构和质子化状态是否正确？

对接问题

- 根据研究目的和研究体系，选择合适的对接算法

146

146

分子对接尚未解决的问题

- 1、溶剂化效应
- 2、熵效应
- 3、蛋白的柔性
- 4、打分函数的准确性

147

147

第四节

Part Four

基于片段的药物设计

Fragment-based drug design, FBDD

148

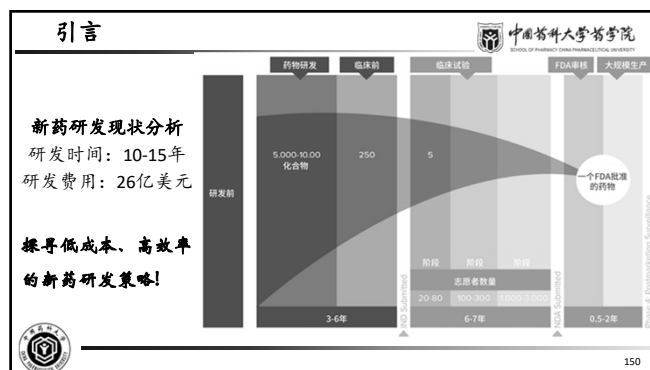
148

基于片段的药物设计

- 引言
- 原理
- 研究方法
- 从片段到先导物的药物设计策略
- 计算机辅助的片段药物设计策略

149

149



150

引言

中国药科大学药学院

高通量筛选 HTS

虚拟筛选 VS

微摩尔级的
苗头化合物
Hits

结构优化
如何着手?

151

151

引言

中国药科大学药学院

结构优化的困惑

- 许多药物靶标的活性位点是由多个口袋组成。高通量筛选的化合物过于“成熟”，通常不能与靶蛋白很好地结合，而对于其中单个片段的优化往往会影响整个分子，甚至导致结合位置的改变

HTS

高通量筛选或虚拟筛选得到的苗头化合物与靶标蛋白的结合模式

152

152

引言

中国药科大学药学院

结构优化的困惑

- 苗头化合物或先导化合物的优化，大多以加入或变换基团，以增加与靶标结合的机会和强度。一般“不敢”轻易除去基团或片段，以免丢失参与结合的原子或基团（即药效团）

结构如何优化?

153

153

引言

中国药科大学药学院

换一种思路：
从分子量较小的片段分子出发!

A

FP $IC_{50} \sim 1 \text{ mM}$
Starting fragment compound 1

B

FP $IC_{50} = 0.28 \mu\text{M}$
 $GI_{50} = 6 \mu\text{M}$
Lead from SAR by catalogue (2)

C

FP $IC_{50} = 0.009 \mu\text{M}$
 $GI_{50} = 0.014 \mu\text{M}$
Phase II candidate AUY922 (3)

154

154

引言

中国药科大学药学院

- 由HTS获得的苗头化合物与先导化合物在分子大小、亲脂性和极性表面积等理化参数上几乎一样，但是活性却比先导物差了很多，这就说明了HTS苗头化合物的成药质量差（药效和理化性质的不平衡）
- 相反，片段分子的活性虽弱，但可覆盖更全面的化学空间，可在优化过程中保持分子的成药质量。

Activity

Quality (balance of activity and property)

Good follow up

Bad modification

155

155

基于片段的药物设计原理

中国药科大学药学院

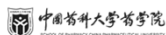
基于片段的药物设计 (Fragment-based drug design, FBDD) 是一种将高通量筛选和基于结构的药物设计相结合的药物发现新方法。

- 由于筛选的起始物是较小的分子，与受体结合的自由度与可能性较大，这是有利的一面，但因为只与受体部分位点结合，显示的活性强度要弱于较复杂的分子量较大的化合物，这是有待提高的一面。
- 低活性的片段分子，可借助与受体结合的结构特征，指导添加或连接有助于提高活性的基团或片段，同时控制分子的大小，以形成有成药前景的化合物

156

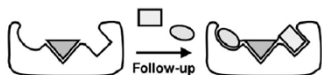
156

基于片段的药物设计原理



设计理念

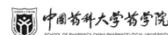
- 药物靶蛋白的活性位点是由多个口袋（子口袋）组成（前提假设）
- 寻找能与靶蛋白各个口袋特异性结合的片段（目标）
- 将上述各个片段用合适的连接器连接起来，组成高活性的化合物。片段连接后会引起结合自由能跳跃式下降，从而导致亲和力大幅度提高。



157

157

基于片段的药物设计原理



筛选得到的片段化合物通过结构优化成为先导物，以及最终成为候选药物，在提高活性的同时，尽量不要过多增加分子尺寸和亲脂性，为此需要在操作过程中用参数加以表征或“监视”活性与分子大小之间的关系。有如下参数：

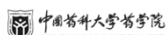
- 配体效率
- 契合质量



158

158

基于片段的药物设计原理



配体效率 (Ligand Efficiency, LE)

配体效率是指配体中每个原子对结合能的贡献：

$$\Delta G = 1.37pK_d, LE = \Delta G/N_{\text{非氢原子}}$$

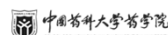
- LE是将配体和受体的结合能力同分子大小相结合的参数，用以衡量苗头物或先导物，以及优化的化合物的质量
- 如某化合物含41个非氢原子（MW: 540），当 $K_d = 10 \text{ nM}$ ， $LE = 0.27 \text{ kcal/mol}$ ；而含有30个非氢原子的化合物，若 $K_d = 10 \text{ nM}$ ， $LE = 0.37 \text{ kcal/mol}$



159

159

基于片段的药物设计原理

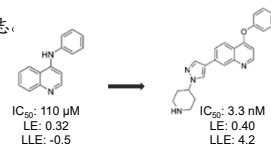


配体-亲脂性效率 (Ligand-Lipophilicity Efficiency, LLE)

LLE用来表征先导物和优化的质量：

$$LLE = pIC_{50} \text{ (或 } pK_i) - \log P \text{ (或 } \log D)$$

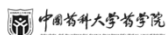
- 由于亲脂性对药物的溶解、溶出和代谢稳定性不利，所以高LLE值为结构优化的标志。



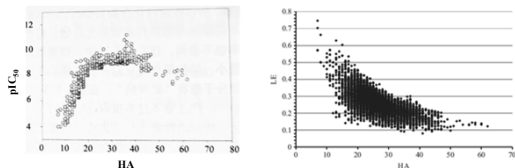
160

160

基于片段的药物设计原理



Reynolds等分析了8000个配体同20个靶蛋白的结合常数，计算了配体效率，发现相对分子量差别较大的化合物活性与（非氢）原子数之间并不呈线性关系



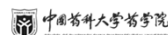
不能反映和揭示优化过程的实际情况



161

161

基于片段的药物设计原理



契合质量 (fit quality, FQ)

Reynolds等提出了配体的契合质量的概念 (fit quality, FQ)，以消除因分子量增大，LE的变化被掩饰和拉平的效应。

- FQ的计算方法是化合物的非氢原子数进行归一和标度化，得到相应的LE-Scale：

$$LE\text{-Scale} = -0.064 + 0.873 \times e^{(-0.026 \times HA)}$$

- 配体的契合质量FQ按照下式计算：

$$FQ = LE/LE\text{-Scale}$$



162

162

基于片段的药物设计研究方法

中国药科大学药学院

FBDD研究流程

- 片段库的构建
- 片段的筛选
- 片段的优化

163

163

基于片段的药物设计研究方法

中国药科大学药学院

片段库的构建

用于高通量筛选的化合物多为类药性分子 (drug-like)，大都遵循类药5规则。而用于FBDD筛选的化合物是分子小结构而简单的类先导物 (lead-like)，其一般特征为：

- 分子多样性
- 药化中的优势结构或骨架
- 有水溶性 ★
- 遵循类药3规则 (分子量低于300, clogP 低于3, 氢键供体、受体和可旋转键分别不多于3)
- 合成可修饰性

164

164

基于片段的药物设计研究方法

中国药科大学药学院

片段库举例：

	片段hit	高通量筛选hit	候选化合物
分子量	160 ~ 250	250 ~ 600	300 ~ 500
活性范围	mM ~ μM	μM	nM

165

165

基于片段的药物设计研究方法

中国药科大学药学院

片段筛选技术

- NMR技术
- 表面等离子共振技术 (SPR)
- X-射线单晶衍射
- 热漂移检测
- 高浓度筛选
- 恒温滴定量热法 (ITC)

2D-NMR Crystallization SPR Thermal Shift Assays

166

166

基于片段的药物设计研究方法

中国药科大学药学院

筛选技术用于检测化合物与大分子靶标结合的亲和力范围

167

167

从片段到先导物的药物设计策略

中国药科大学药学院

发现活性的片段分子仅是研究的第一步，将活性片段转化为先导化合物甚至是候选药物才是基于片段的药物设计的最终目的

从片段到先导物的设计方法：

- 片段生长 (fragment growth)
- 片段连接与融合 (fragment linking and fusion)
- 片段自组装 (fragment self-assembly)

168

168

从片段到先导物的药物设计策略



1、片段生长

以受体结合的第一个片段为核心，经过理性设计，在邻近处逐渐生长成活性强的较大分子



按片段逐个生长



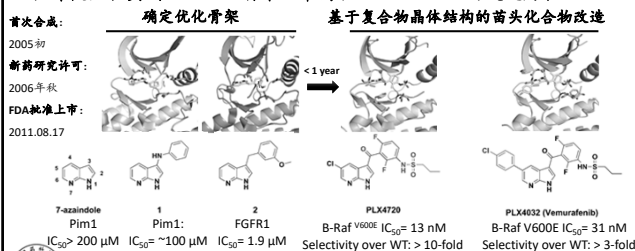
169

169

从片段到先导物的药物设计策略



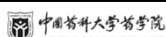
1、片段生长实例——FBDD首个上市药物 Vemurafenib 的发现历程



170

170

从片段到先导物的药物设计策略



2、片段连接与融合

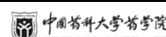
片段连接: 与受体结合的相邻的两个片段经连接基团连接成活性强的较大的分子



171

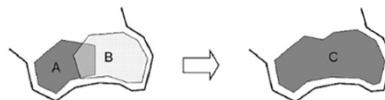
171

从片段到先导物的药物设计策略



2、片段连接与融合

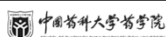
片段融合: 与受体结合的相互交叠或接近的两个片段合并成活性强的较大的分子



172

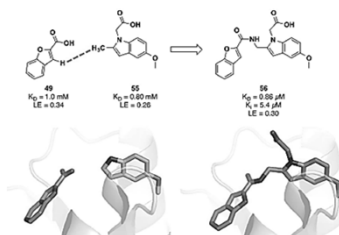
172

从片段到先导物的药物设计策略



2、片段连接与融合实例

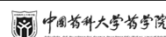
结核分枝杆菌泛酸合成酶 (MtPS) 命中片段的连接



173

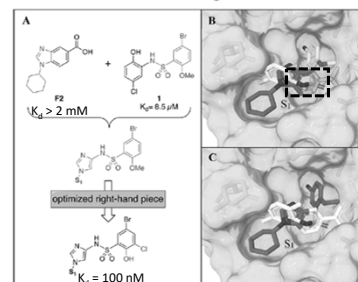
173

从片段到先导物的药物设计策略



2、片段连接与融合实例

通过片段融合设计 WDR5-MYC相互作用抑制剂



174

174

从片段到先导物的药物设计策略



3、片段自组装

分别结合在活性位点中相邻口袋的两个活性片段含有可相互反应的基团，这两个片段可自发地连接成为高活性的化合物



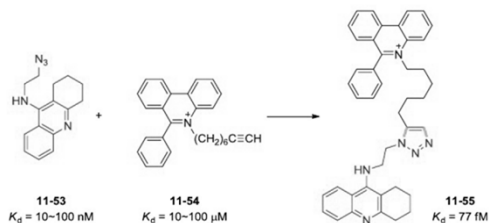
175

175

从片段到先导物的药物设计策略



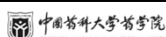
3、片段自组装实例



176

176

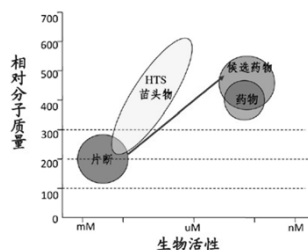
基于片段的药物设计优缺点



优点:

- 1、片段库仅需几百-几千个分子
- 2、化学结构新颖且多样，命中率高
- 3、从片段优化至候选药物，分子量和活性同步增长，更加符合新药发现的一般规律，可行性强

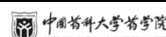
缺点: 晶体结构较难获得



177

177

计算机辅助的片段药物设计策略



适合于片段筛选的算法

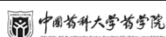
Software company or academic lab	Products	Type of calculation	Literature reference
Accelrys	LUDI	De novo design	Bolin (1992)
	MCS	Ensemble molecular mechanics	Miranker and Karplus (1991)
	LigandFit	Docking	Venkatesh et al. (2003)
BiosolveIT	FlexX	Docking	Ram et al. (1996)
	HexNovo	De novo design	Degen and Ravey (2006)
Cambridge Crystallographic Data Centre	GOLD	Docking	Jones et al. (1997), Verdoorn et al. (2003)
Chemical Computing Group, Inc.	MultiFragment search	Ensemble molecular mechanics	MOE (2009)
	Ligand-Receptor Docking	Docking	
OpenEye Scientific Software	FRED	Docking	McGee et al. (2003)
Schrodinger, LLC	GLIDE	Docking	Friesner et al. (2004), Halgren et al. (2004)
Tripos, L.P.	Surfex-Dock	Docking	Jain (2003)
UCSF	DOCK	Docking	Kuntz et al. (1982, 1990)



178

178

计算机辅助的片段药物设计策略



MCSS: Multiple Copy Simultaneous Search

MCSS是一种基于片段的分子对接技术，由Miranker与Karplus开发（1991）。片段先被随机分布在结合位点中，然后程序采用CHARMM对这些随机片段进行能量优化，通过采用独立的MCSS_Score来打分和排序，以找到最佳的片段位置。

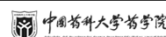
- $Score_{MCSS} = -(Energy_{Ligand} + Energy_{Ligand-Protein})$
- 计算效率高（多片段同时优化，片段之间互不影响）



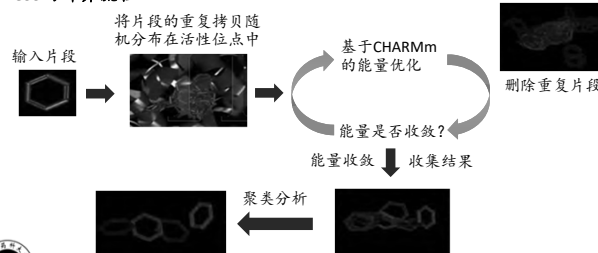
179

179

计算机辅助的片段药物设计策略



MCSS的计算流程



180

180

计算机辅助的片段药物设计策略



Ludi in DS

Ludi是应用最为广泛的基于片段的药物设计方法之一，它帮助我们找到新的分子骨架或者修改已有的分子骨架来提高小分子的活性。

Discovery Studio中的De Novo Protocol使用的就是Ludi算法，包括：

- De Novo Receptor
- De Novo Link
- De Novo Evolution



181

181

计算机辅助的片段药物设计策略



Ludi in DS

De Novo Receptor可以用来寻找与受体相互作用的全新片段，这种方法的优势是可以快速有效的对数据库中大量的片段进行筛选；

De Novo Link 可以将找到的活性片段连接或在已有骨架的基础上添加新的片段分子；

De Novo Evolution 也称为AutoLudi，主要用于进行Me better 药物设计工作，不仅可以搜索活性片段分子并且可以将片段和骨架自动连接直接产生新的分子。



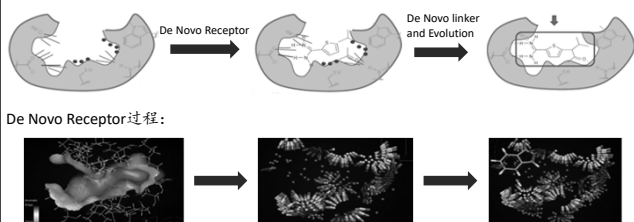
182

182

计算机辅助的片段药物设计策略



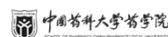
基于Ludi进行FBDD的基本流程



183

183

计算机辅助的片段药物设计优缺点



优点：

- 配体结构可能是全新的，不受现有知识的约束，也不受人的思维束缚
- 该方法既可用于先导化合物的发现，也可用于对先导化合物的结构优化

缺点：

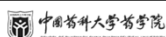
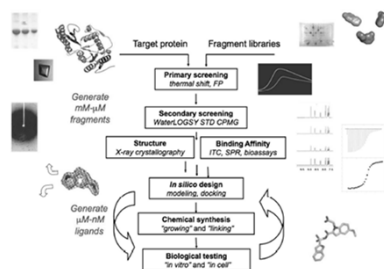
- 可能只是理论分子，仅对接有效，且可能不易合成。



184

184

基于片段的药物设计总结

FBDD筛选和
优化流程

185

185

谢谢大家



186

186