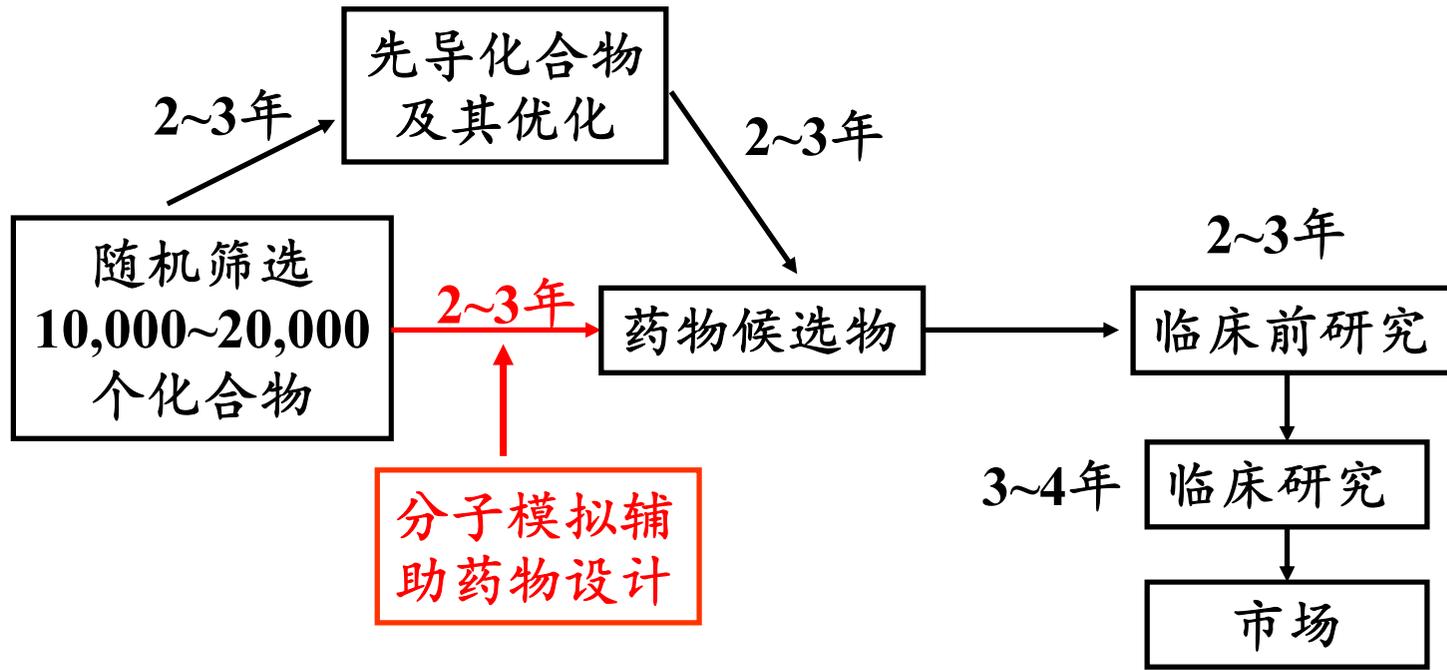




第八章 分子模拟在 药物设计中的应用简介



§ 8.1 药物研发过程与分子模拟的应用

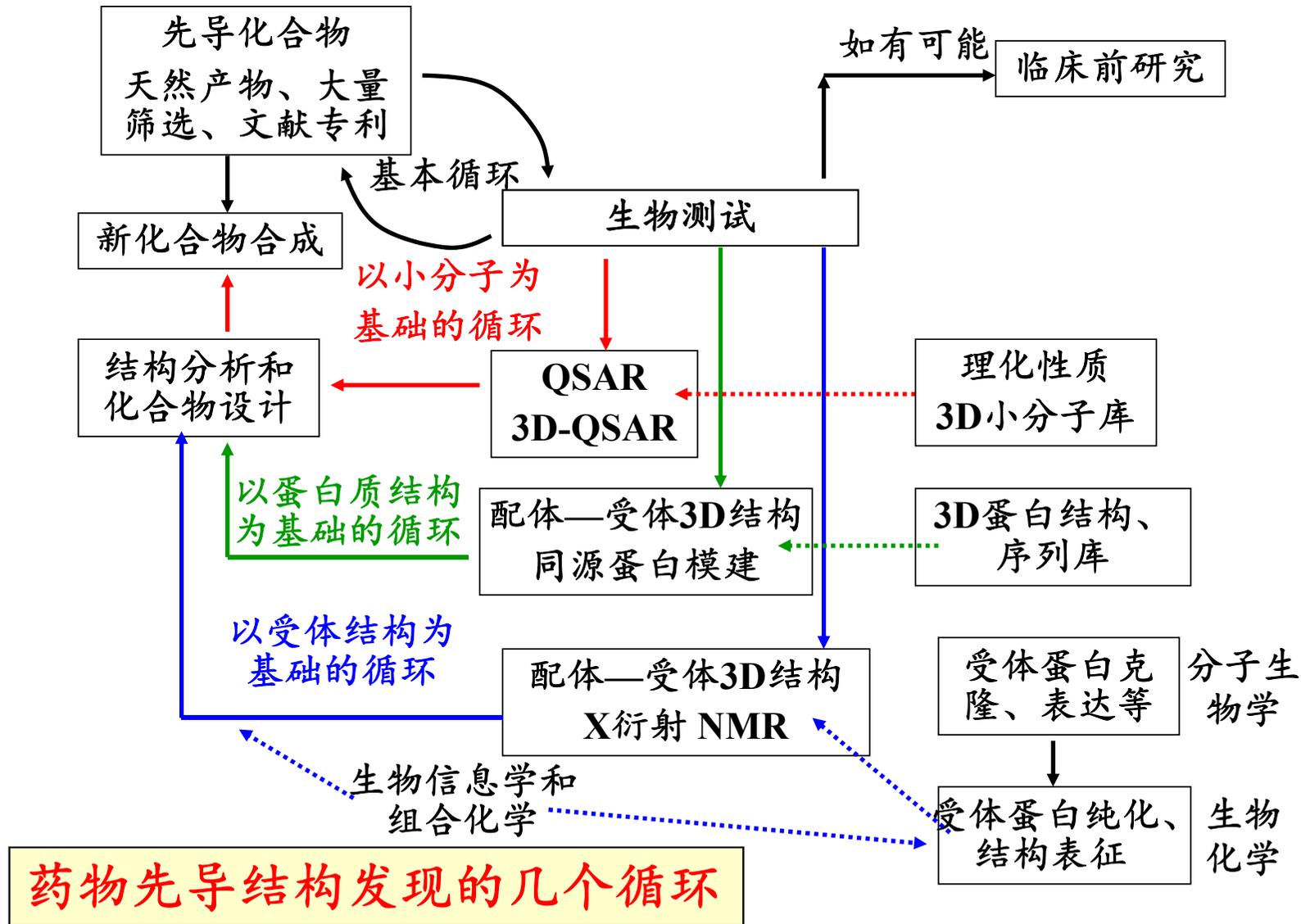


研制一种新药，平均花费~12年时间，成本26亿美元。2017年全球前**十**强药企研发投入平均占销售额**17.8%**



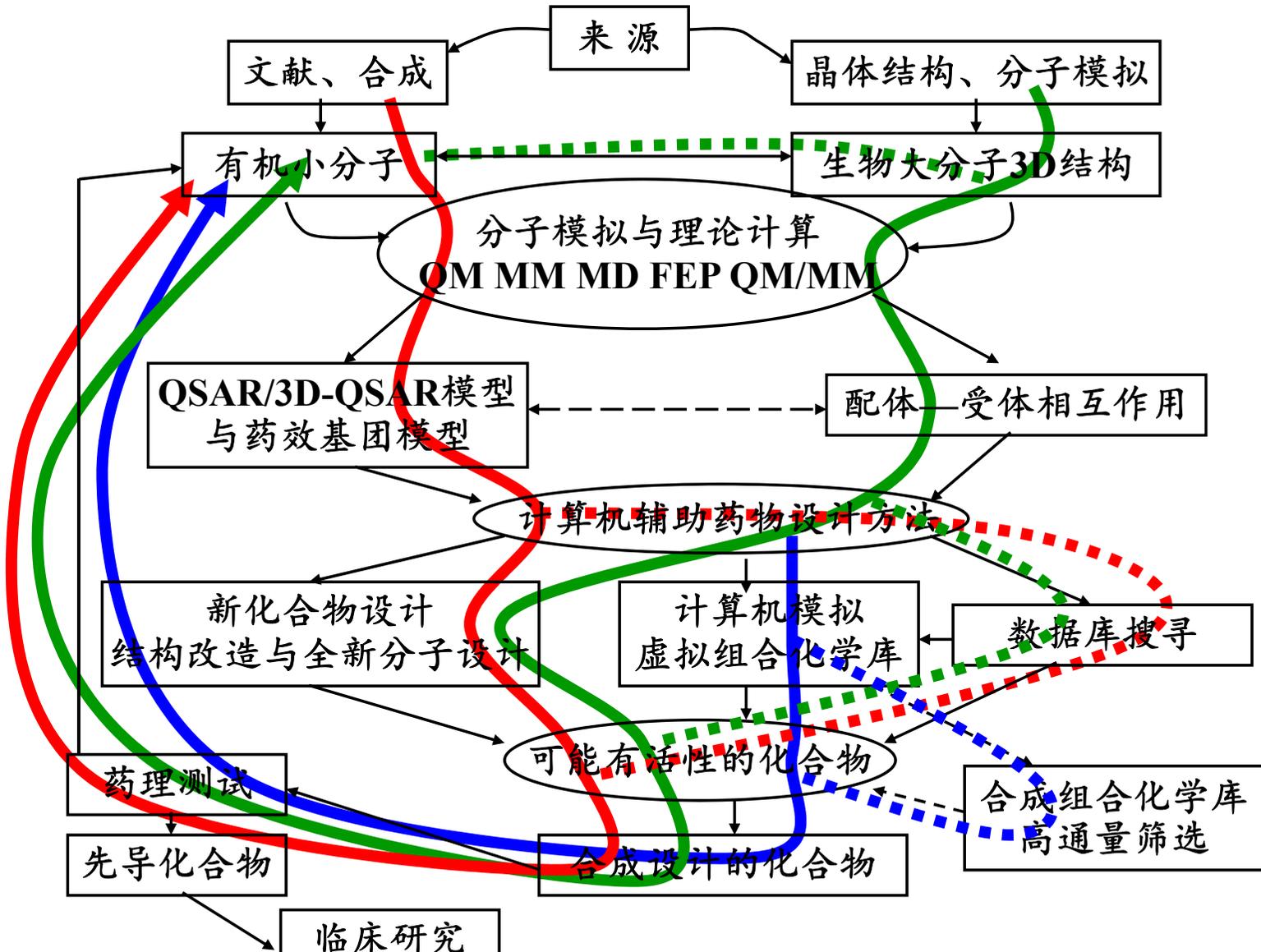
现代药物研究的四大支柱

- 分子生物学、基因组学及蛋白质组学
为药物研究提供新靶标，为疾病机理提供新的认识，受体是药物研究的核心
- 组合化学
用巧妙的战略和程序，从某些单体或基本化合物出发，快速合成大量化合物的方法，广泛用于先导化合物的发现和优化
- 高通量筛选(high through-put screening)
- 与药物研究相关的信息科学及技术
生物信息学、生物医学信息学、化学信息学





与组合化学相对应的计算机辅助药物设计
以受体的三维结构为基础的计算机辅助药物设计
药物小分子构效关系基础上的药物设计





§ 8.2 量化计算方法的应用

<100原子 *ab initio*/DFT 6-31G*

100~200 半经验

>200 MM MD 可以采用QM/MM模型

解决以下问题:

- 生物分子/药物分子的电子结构与活性关系
- 配体—受体相互作用机理
- 生物化学(酶催化)反应机理
- 生物体系中电子转移、质子转移、和能量传递
- 无机离子在生物过程中的作用
- 遗传、突变的量子理论
- 生物体系中的信号传递
- 量子药理学



生物大分子体系的量子化学计算方法

- 计算显微镜方法(computational microscope)

先用*ab initio*方法对组成大分子的各小分子碎片进行电子密度计算，再基于MEDLA(molecular electron density lego addember)原理运用计算机进行模糊电子密度碎片叠加得到大分子的电子密度图象(1994)

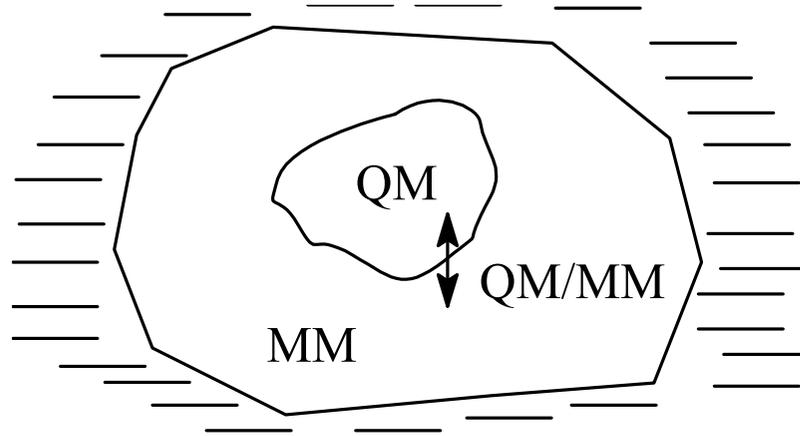
- 定域分子轨道法(Localized MO, LMO)

将分子轨道高度定域化，节省密度矩阵计算时间，并且可用来优化构型(1996)

- 线性标度半经验量子化学方法(linear scaling semiempirical quantum chemistry)

密度矩阵分解成若干个子体系的矩阵元的加合，DAC-divide and conquer

QM/MM模型



研究体系分成两部分，即QM(计算溶质或生物大分子的活性部分)和MM(溶剂或生物大分子的其它部分)，在QM与MM之间有一相互作用部分QM/MM。该模型广泛应用于酶催化机理的理论研究、溶剂效应的理论计算和其它的配体—受体的相互作用

Exploring 3rd ed. Chap 09 高级模拟技术 **ONIOM**



§ 8.3 定量构效关系

8.3.1 2D-QSAR定量模型

- **Hansch分析**—线性自由能相关途径(或超热力学途径)
活性与疏水性参数(π 或 $\lg P$)、电性参数(σ), 立体参数(E)等的线性关系

$$\lg(1/C) = a\pi + b\sigma + cE + d \quad C: \text{生物活性};$$

- **Free – Wilson模型**

$$\lg(1/C) = \sum \sigma_{ij} + \mu$$

σ_{ij} 为取代基 X_i 在位置 j 的基团贡献, μ 为参考分子的生物活性值



8.3.2 建立QSAR模型的统计方法

- **回归分析(LR)**

一元回归和多元线性回归分析(MLR)

$$Y = a + bX \qquad Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

相关系数 R , 标准偏差 S , F 检验来衡量效果

- **主成分分析(PCA)**

由于一些变量可能有相关，因此将原有变量重新组合成新的变量集合，组成时使变量之间正交，其对总方差的贡献按顺序排列，计算前 P 个主成分的积累贡献大于给定值(一般取0.85)，就取前 P 个主成分作新指标，用最小数主成分得最大总方差。这样既能使新指标保持原有指标的重要信息，又达到减少指标的目的



• 偏最小二乘(PLS)分析

对自变量矩阵 X 分解的同时,也将因变量矩阵 Y 进行分解。克服了基于最小二乘原理的多元线性回归(MLR)方法在自变量存在严重线性相关求逆矩阵的困难。同时弥补了PCA中未充分利用 Y 的信息的不足,结果更好

• 模式识别 (Pattern Recognition)

基本思想:根据实验数据得到一批训练点,参照模型或经验规律提出一批特征量(如将结构参数作为数量的模式向量成分)用以形成模式空间或特征空间,通过模式识别算法训练和分类,然后根据训练分类,所得判据对未知样本进行判断步骤:

1. 选择模式特征向量(结构参数): (无相关性, 标准化, 最小维数)

2. 模式分类:

主成分分析(作图分类); 聚类分析(对化合物归类分组); 判别分析(用判别函数加以分类)

3. 预报:

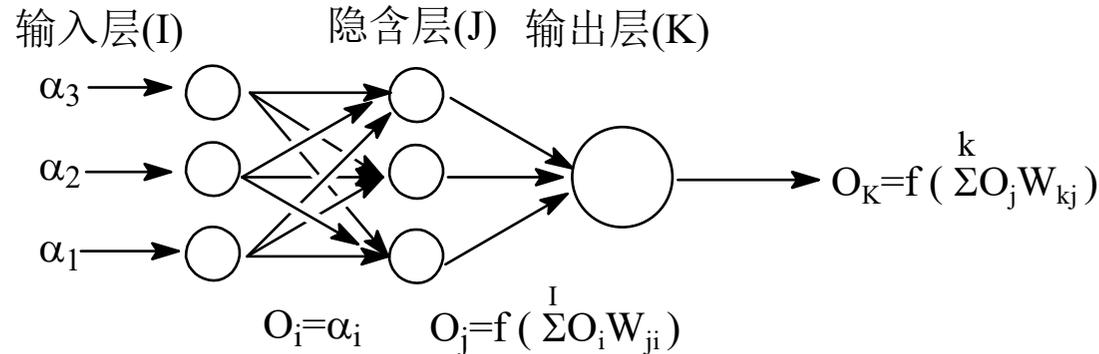
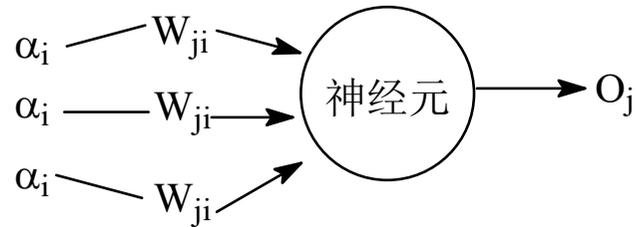
用主成分方程计算待预报物的值并在主成分平面图上找出其分类或用判别函数值判别分类



• 人工神经网络(artificial neural network)

模仿人的神经元处理信息的过程，由大量基本信息处理单元通过复杂连接而构成的非线性动力学系统，具有并行性、容错性、非线性和自学习性等。通常用误差反传(BP)人工神经网络

输入 权重 (节点) 输出





• 遗传算法(genetic algorithm)

首先利用某种编码方式把求解问题表达成二进制串(基因)产生初始的群体, 然后进行基因复制、杂交、变异三种操作产生新的群体。根据个体对环境的适应性, 不断淘汰适应性差的个体, 以得到最佳解

通过简单编码技术和繁殖机制, 遗传算法可以表现复杂现象, 解决组合优化过程

- 产生初始种群, 用评分函数来评价每个个体
- 选择操作—所有个体评分后, 根据个体得分结合随机方法选择被新种群保留的个体, 每个被淘汰的个体, 都将用新的个体替代
- 交叉操作—种群中选择两个被保留的个体为母体, 将这两个母体随机分为两段, 在不同的母体中选择一部分组成新的个体
- 突变操作—在种群中随机选择一个个体, 将此个体中的某一元素随机改变得到新的个体, 交叉和突变操作产生的新个体都要用得分函数予以评价, 根据得分情况进行选择, 产生新种群
- 部分替换操作—经过若干突变和交叉操作后, 对种群中得分最低的部分个体, 进行部分替换操作, 减少陷入局部极小的危险
- 比较操作—为将最好的若干个个体保存下来, 采用“精华”种群保存, 进行完交叉突变操作后, 逐一比较新种群中个体与精华种群的个体, 在新种群中发现更好的个体, 拷贝到精华种群去



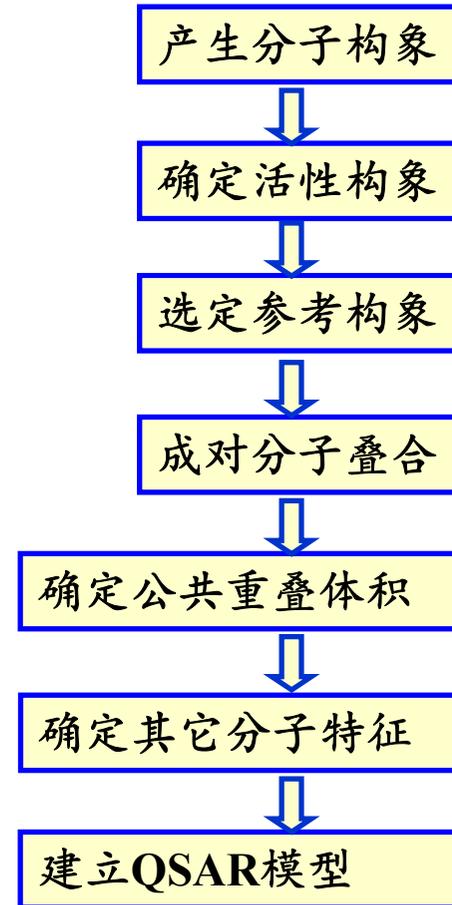
8.3.3 3D-QSAR 方法

- 距离几何3D-QSAR

通过选择合理的受体结合位点的模型以及和药物分子活性基团之间的结合模式，建立药物分子结合能力和活性之间的关系

- a. 定义药物分子中可能的作用位点
- b. 计算配体分子的距离矩阵
- c. 定义受体结合位点的分布
- d. 确定受体结合位点的分布

- 分子形状分析(MSA)

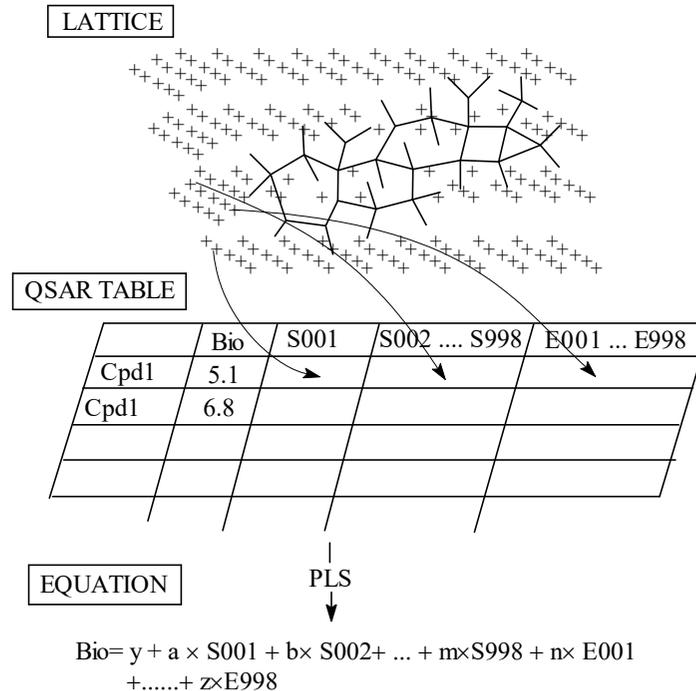




• 比较分子力场分析CoMFA

其基本出发点是，在分子水平上，影响生物活性的相互作用，通常是非共价键的作用力，如范德华力和静电相互作用。作用于同一受体的一系列药物分子的力场分布应相似，研究药物分子周围力场分布与其活性关联性，可以探测受体的某些性质，又可进行药物结构改进和新分子设计。

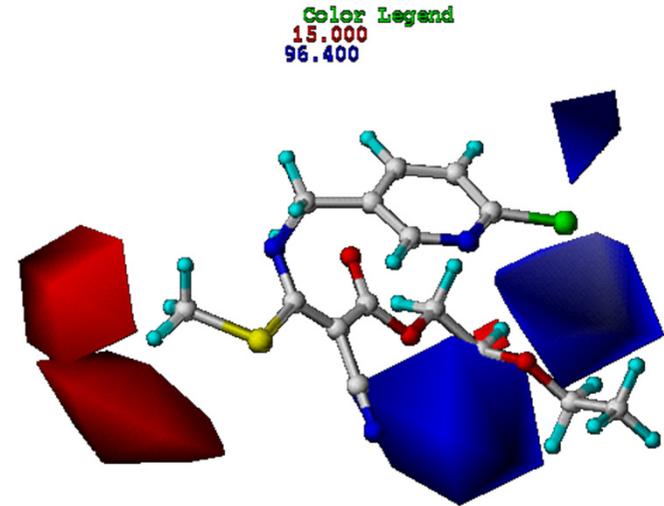
方法：将分子置于三维网络上(lattice)，按一定规则定位，然后由位于网格上的探针原子做立体场和静电场分析，计算出化合物间立体的和静电的相互作用之和。每个化合物均作如此分析并与各化合物的生物活性数据共同组成CoMFA的QSAR表，然后用偏最小二乘法(PLS)对数据进行回归分析，从而得到最佳的QSAR结果



CoMFA分析示意图

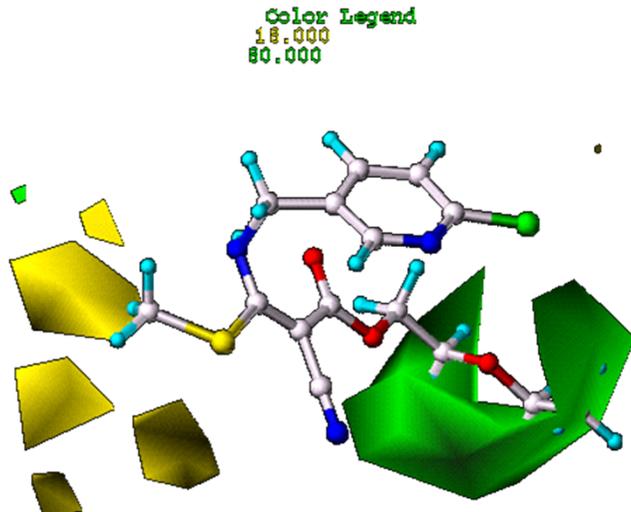
步骤:

- ① 确定化合物的药效构象
- ② 选择合适的定位规则及定位模板(分子迭合方式)
- ③ 计算CoMFA力场(通常选择 $C(sp^3)^+$ 轨道作为探针, 逐点计算作用能)
- ④ 用PLS分析数据并得QSAR方程
- ⑤ 以三维等值线显示QSAR方程
- ⑥ 在定位分子周围, 用不同颜色的网络线表示不同性质的等值区。



(在静电区域)

红色区是电荷密度区, 较高负电荷有利于受体结合, 蓝色区反之



(在立体场区域)

绿色区表明配体体积效应有利, 增大位阻将有利与受体结合, 黄色区反之



• 虚拟受体方法

采用多种探针原子和基团在药物分子表面建立一个虚拟的受体环境，虚拟受体与药物分子间相互作用的差别反映了药物分子之间的活性差别

Compass方法（基于形状的受体模型方法）

- ① Monte Carlo构象分析，产生多重构象
- ② 定义分子取向(Pose)
- ③ 分子叠合，质心出发162条射线，平均表面外一定距离定义“性质参考点”
- ④ 从性质参考点到某分子表面距离为矢量D的分量，162个分量表征分子形状
- ⑤ 三层神经网络训练

GERM方法（基于遗传算法）

RSM算法（受体表面模型法）



§ 8.4 药效团模型

药效团模型也是一种利用已有分子三维结构信息进行药物分子设计的方法，所谓药效团，是指药物分子中对活性起重要作用的“药效特征元素”的空间排列形式。“药效特征元素”是药物分子与受体相互作用时的活性部位，其可以是具体的原子或或基团，如氧原子、苯环等，也可以是某些具有特殊化学功能的结构，如疏水团、氢键给体、氢键受体等

药效团模型与QSAR方法都是以配体小分子为作为研究的起点，但QSAR主要用于指导先导化合物的改造，而药效团模型则可用于先导化合物的搜索



间接药物设计

- 基本假设：所有被研究的具有同类活性的一系列化合物都作用于同一受体的相同作用部位，且作用方式相同，化合物的形状和理化性质与受体活性部位的结构匹配。
- 构建药效基团步骤：
 - ❖ 选择化合物
 - ❖ 构象搜索
 - ❖ 分子叠合建立模型
 - ❖ 无/低活性用来“勘探”结合位点“空腔”的结构性质

AAA-active analogue approach

DISCO- distance comparison



§ 8.5 直接药物设计

SDBB(Structure-based drug design)

靶标受体3D结构已知，并且有配体-酶复合物的3D结构→配体-受体作用模型。

- 运用分子图形学和计算化学，计算有机小分子与受体的结合自由能→预测活性
- 运用分子对接(Dock)方法搜寻小分子数据库。
- 根据酶活性结构，设计结构新颖的化合物(全新药物设计方法 *de novo drug design*)

分子对接(Dock)

基于“锁与钥匙”模型，空间形状的匹配和能量匹配

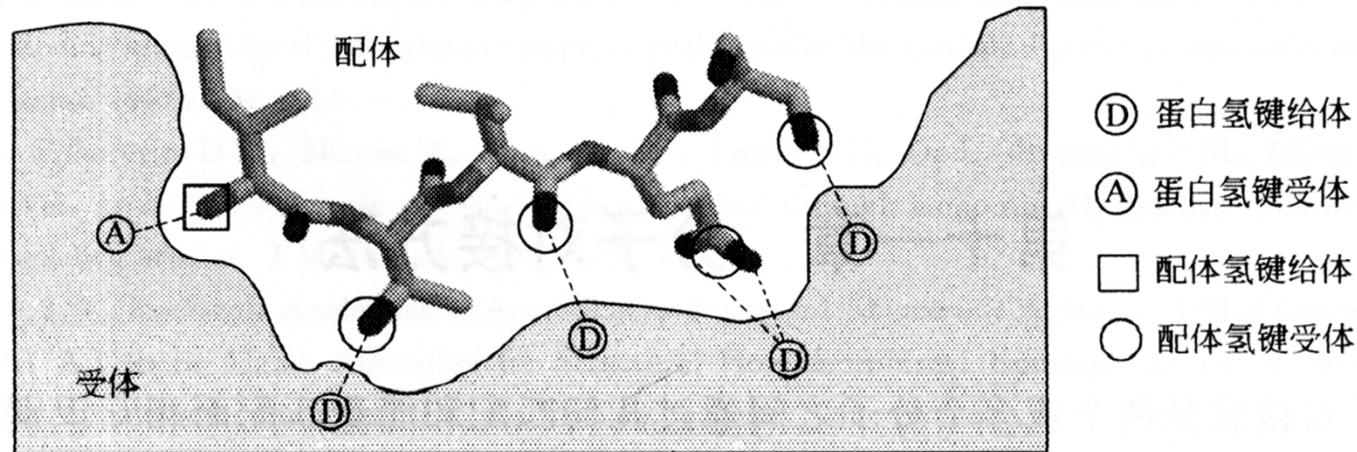
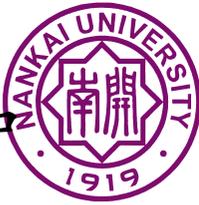
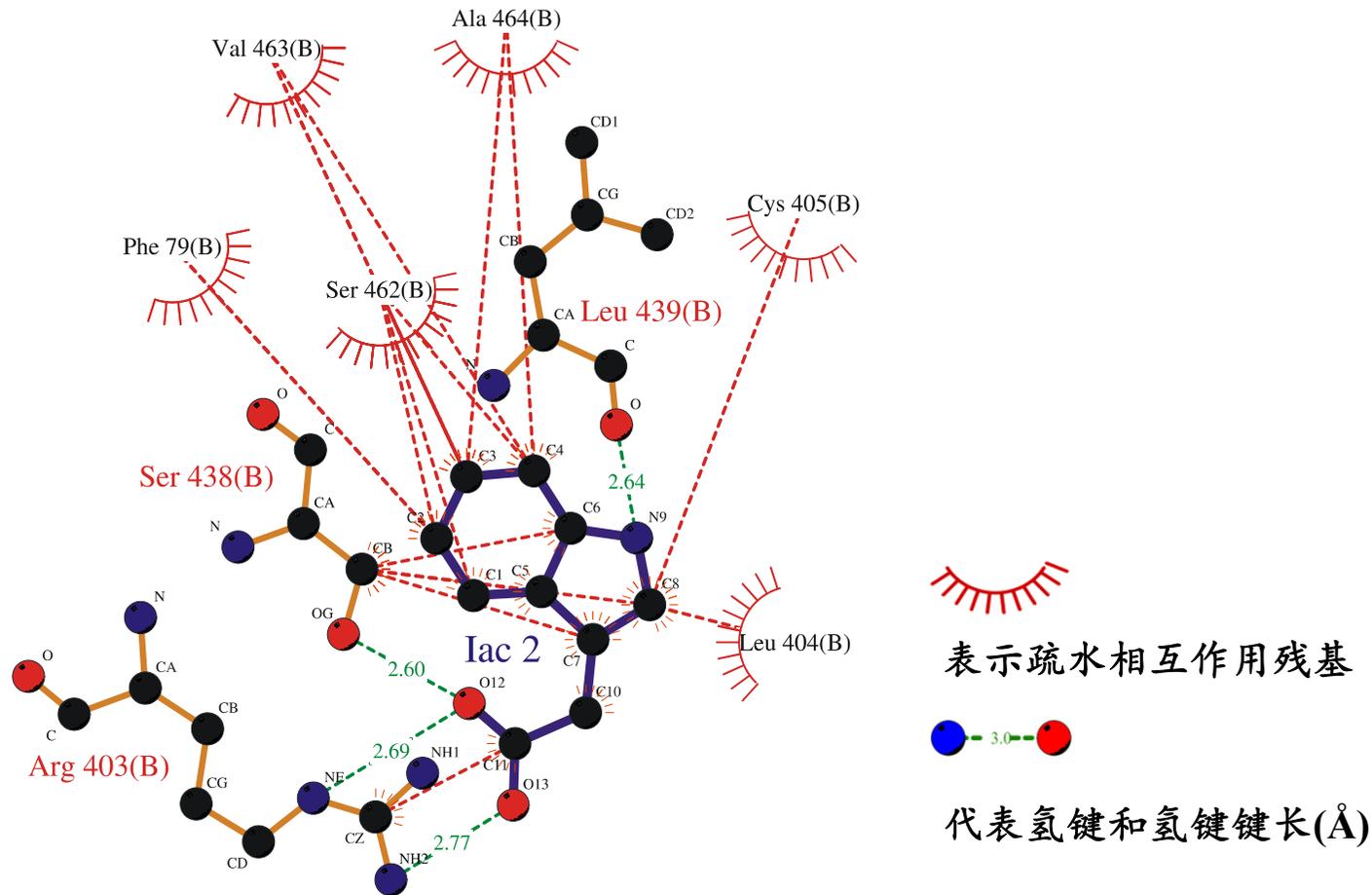


图 11.1 受体和配体之间相互匹配的简单示意



- 将小分子配体放置于受体的活性位点处，寻找其合理的取向和构象，使配体与受体的形状和相互作用的匹配最佳
- 根据不同的简化程度分为三类：
刚性对接、半柔性对接和柔性对接
- 药物设计中，分子对接方法主要用来从小分子数据库中搜寻与受体生物大分子有较好亲和力的小分子，进行药理测试，从中发现新的先导化合物。其优点是数据库来自现有化合物数据库，合成或订购容易，可加快先导化合物的发现过程

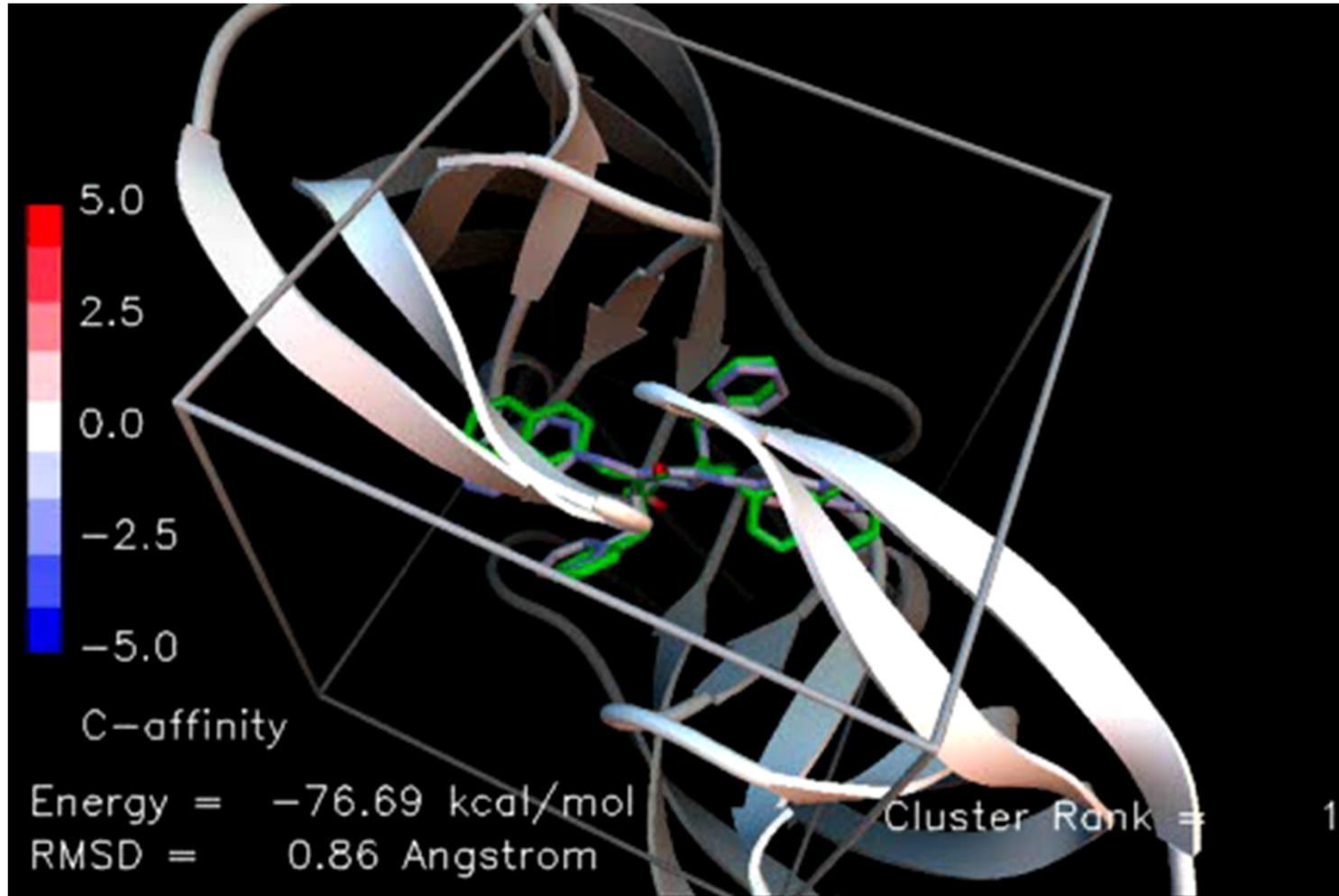


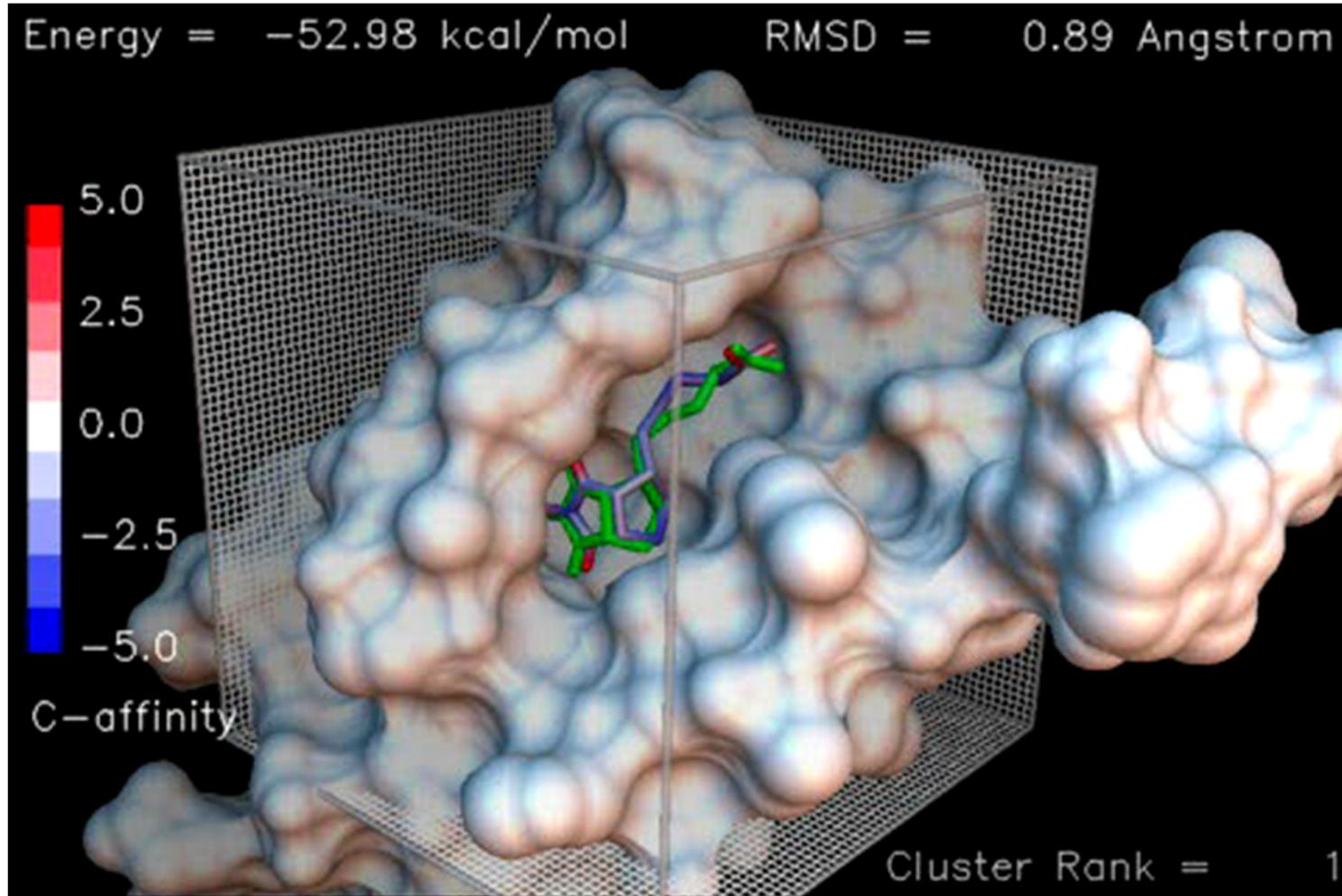
生长素IAA与受体间的相互作用

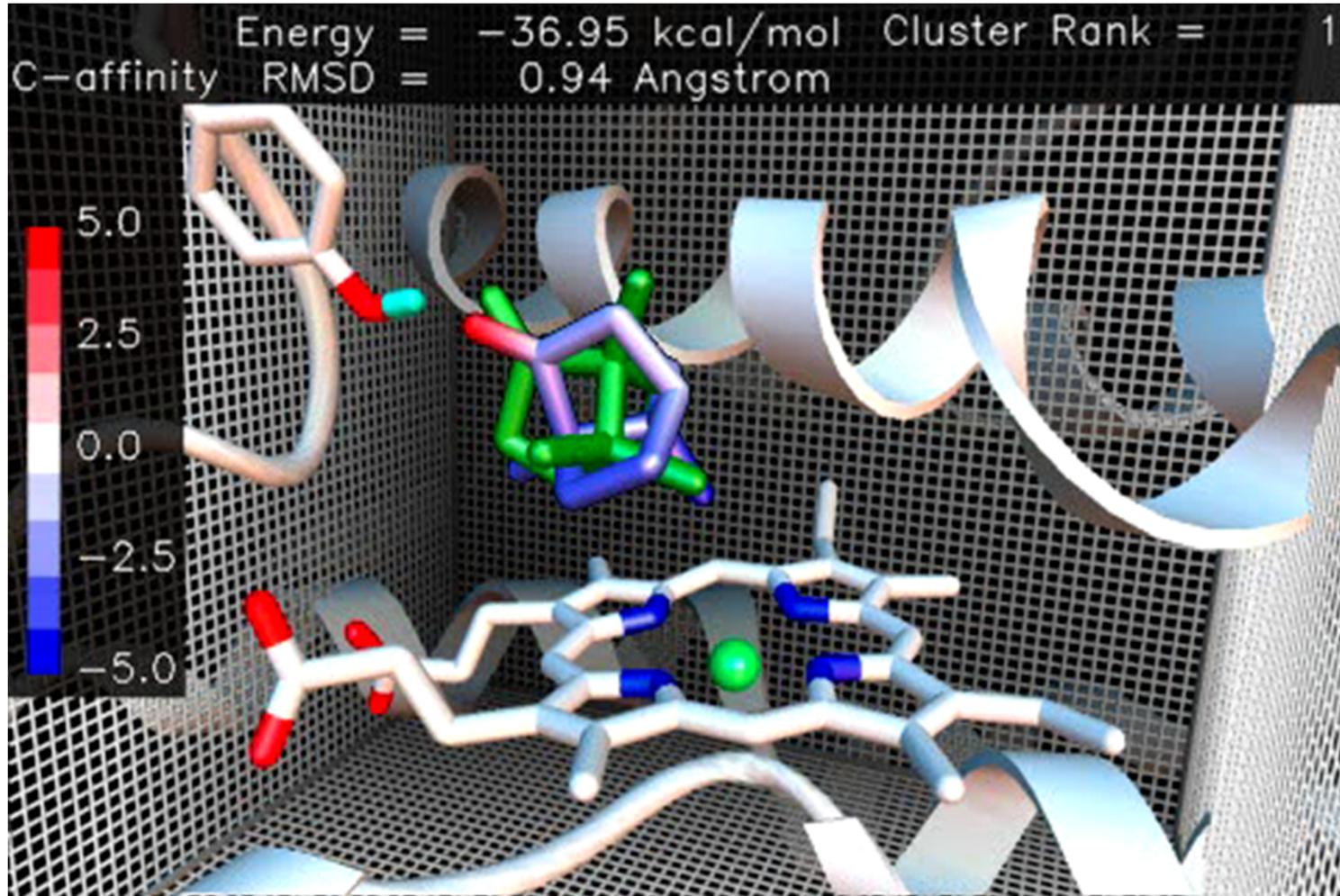


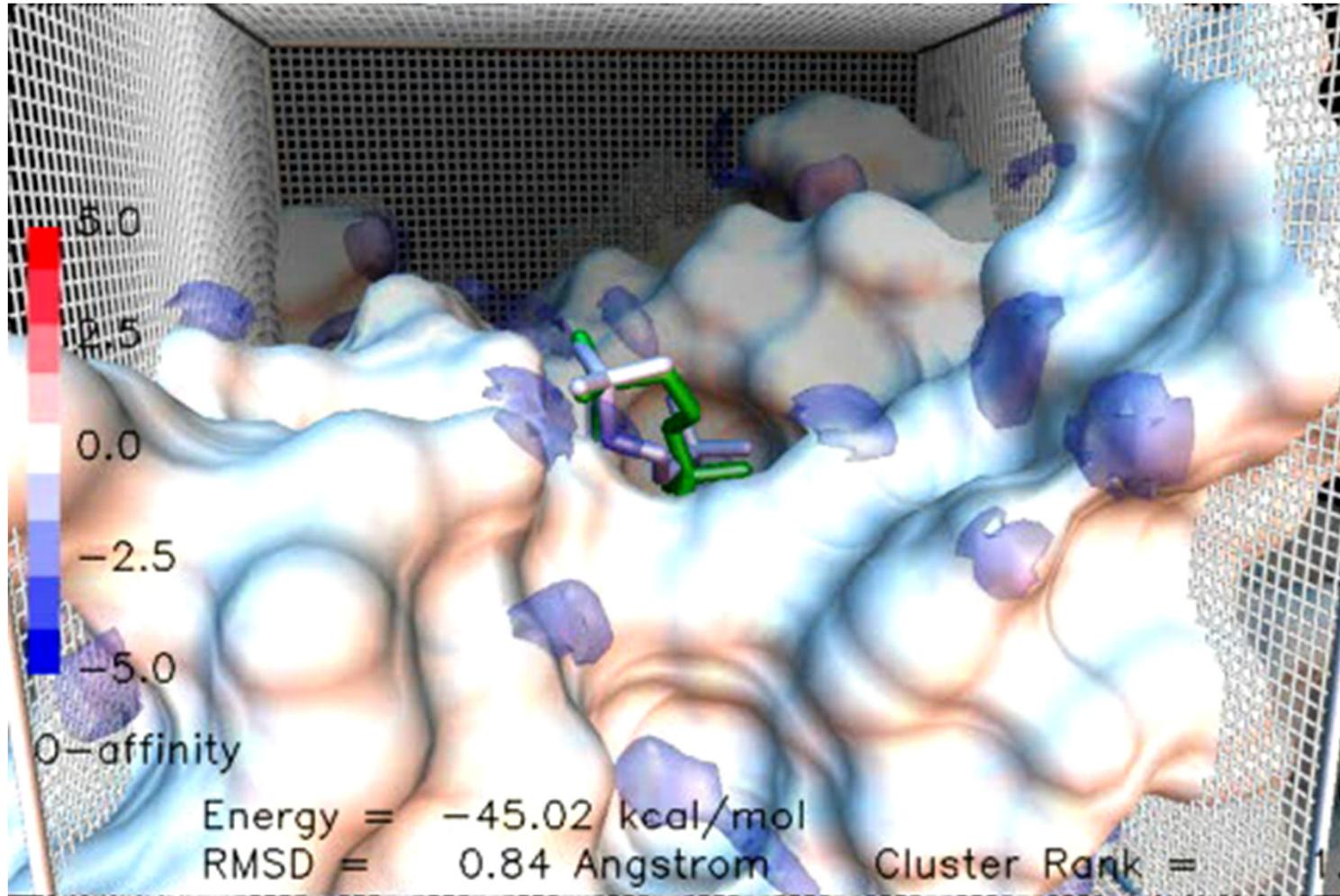
有代表性的分子对接方法

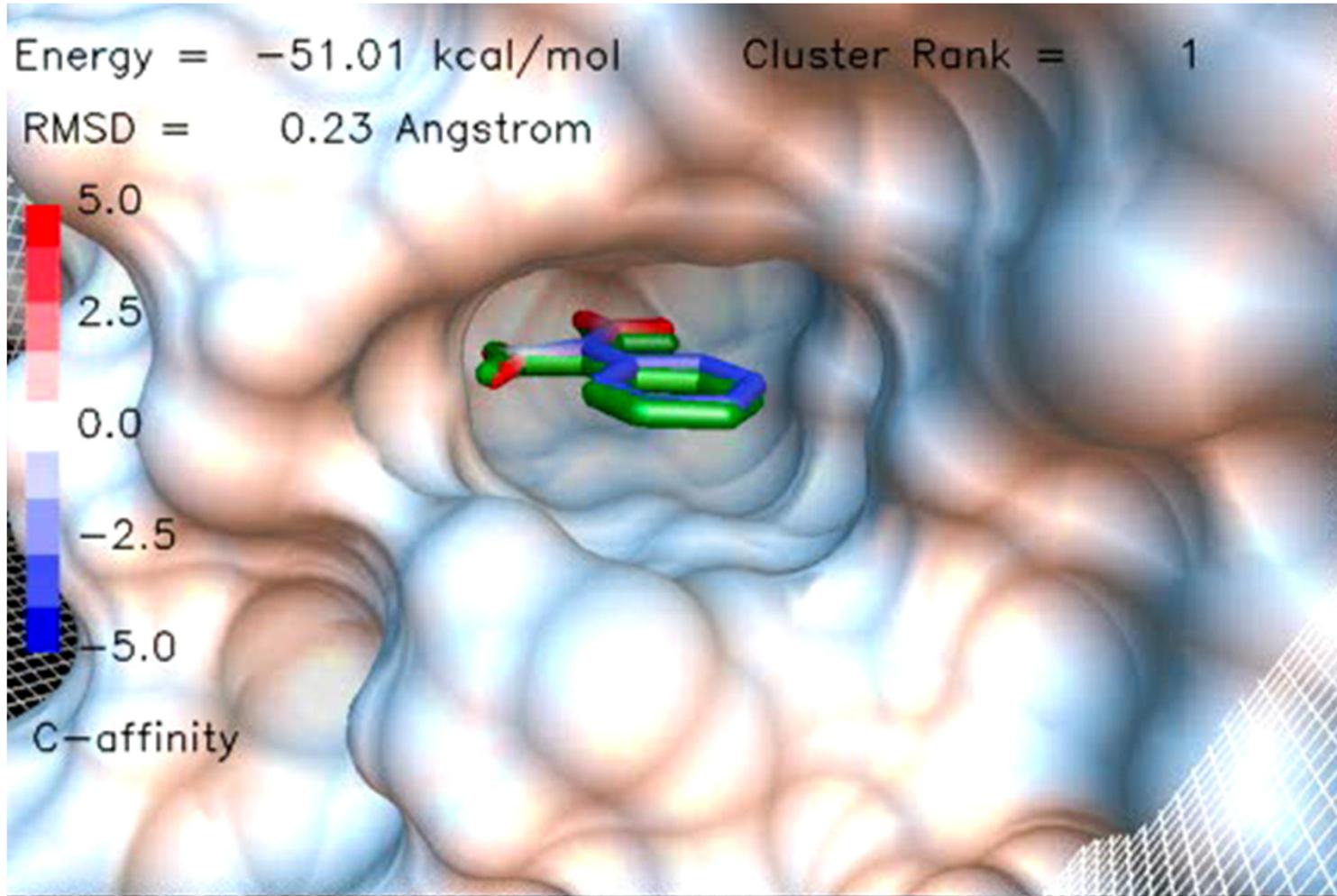
方法	优化方法	评价函数	速度
DOCK	片段生长	分子力场、表面匹配得分、化学环境匹配得分	快
eHiTS	系统搜索	半经验自由能评价函数	快
EUDOC	系统搜索	分子力场	快
FlexX	片段生长	半经验自由能评价函数	快
JCM-Dock	随机全局优化	半经验自由能评价函数	快
LigandFit	MC模拟	半经验自由能评价函数	快
FlexiDock	遗传算法	分子力场	慢
Fred	系统搜索	几何匹配、半经验自由能评价函数	快
Slide	系统搜索	半经验自由能评价函数	快
Affinity	MC/MM/MD	分子力场	慢
AutoDock	遗传算法	半经验自由能评价函数	一般
Glide	系统搜索	半经验自由能评价函数	一般
Gold	遗传算法	半经验自由能评价函数	快
QXP	MC模拟	分子力场	一般

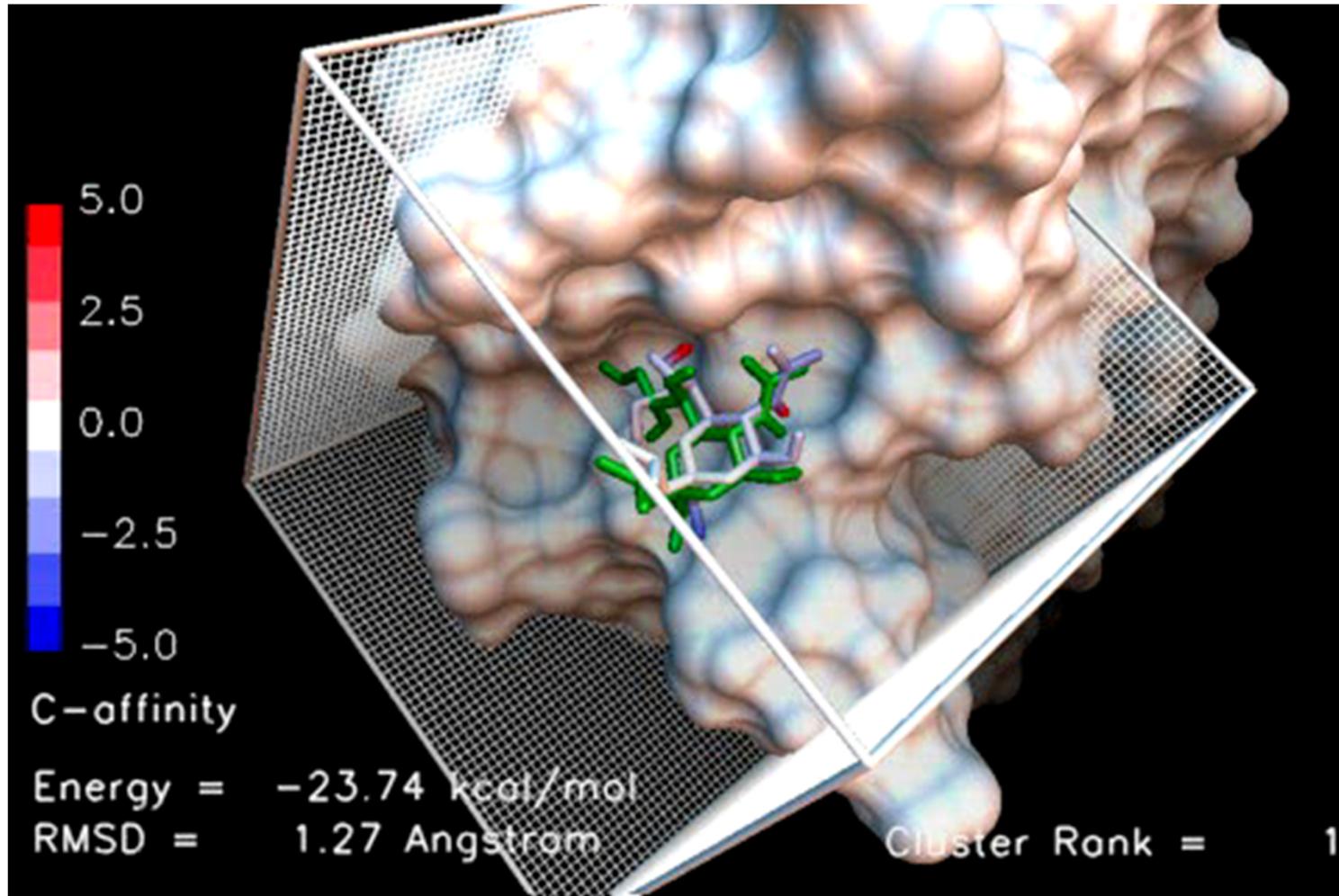










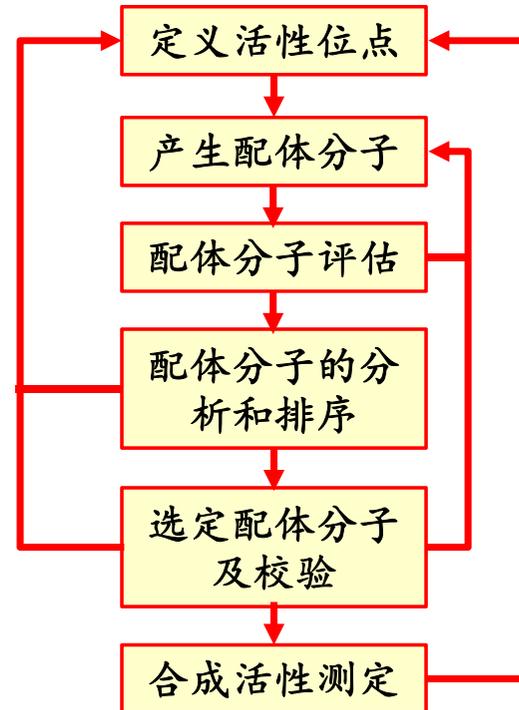




§ 8.6 全新药物设计

全新药物设计(*de novo drug design*)也称从头设计方法，是根据靶点的活性位点特征产生一系列的片段，通过这些片段的连接产生新的配体分子，其一般步骤如下：

- 根据结构特征对靶点进行分析定义活性位点，如结合部位的静电场分布、疏水场分布、氢键作用位点、空间约束特征等
- 根据活性位点的信息，产生相应的配体分子片段，通过连接这些片段产生新的配体
- 对新产生的配体进行评估和结构验证
- 活性分子的合成和活性测试





同源模建(homology modeling)

也称比较分子模拟(comparative molecular modeling)或基于知识的蛋白质结构模拟(knowledge based protein modeling), 其基本思想是: 同源膜蛋白氨基酸残基序列上的相似性表示它们应具有相似的三维空间结构

- 模板选择
- 序列联配
- 模型建立
- 模型优化



参考文献

1. 陈凯先、蒋华良、嵇汝运主编, 《计算机辅助药物设计—原理、方法及应用》上海科学技术出版社, 2000.10
2. 徐筱杰、侯廷军、乔学斌、章威编著, 《计算机辅助药物分子设计》, 化学工业出版社, 2004.4
3. 杨铭主编, 《结构生物学与药学研究》, 科学出版社, 2003.10