

## 计算化学方法与圆二色光谱结合 预测手性分子的性质

汇报人：李昆华

导师：张俊良 教授

日期：2024.11.8

1. 背景介绍
2. 计算化学方法与圆二色光谱结合的应用
  - 2.1 预测圆二色光谱和确定手性化合物的绝对构型
  - 2.2 预测手性化合物的ee值
  - 2.3 其它应用
3. 总结与展望

## 1. 背景介绍

## 2. 计算化学方法与圆二色光谱结合的应用

### 2.1 预测圆二色光谱和确定手性化合物的绝对构型

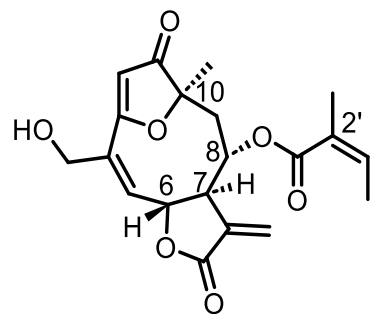
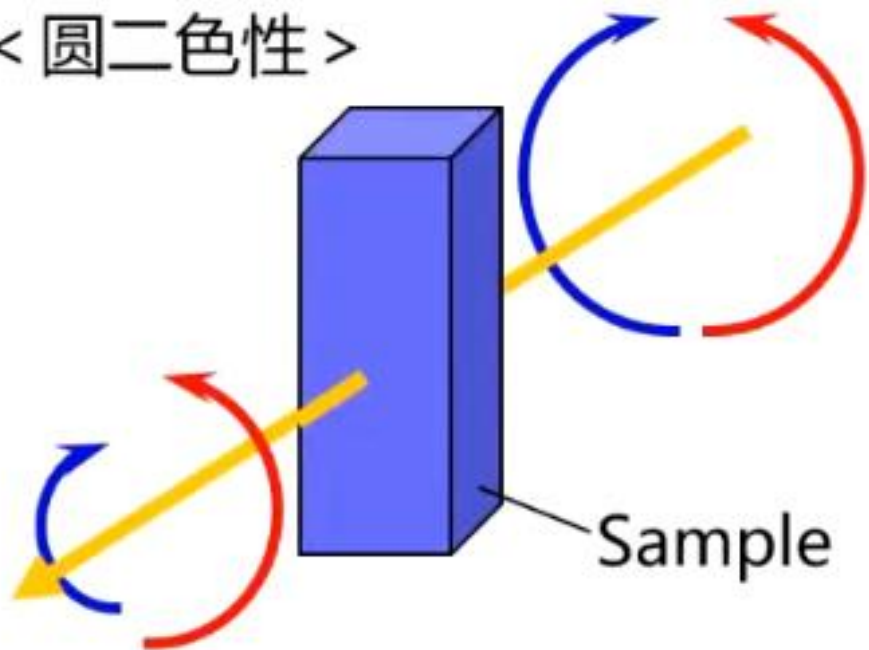
### 2.2 预测手性化合物的ee值

### 2.3 其它应用

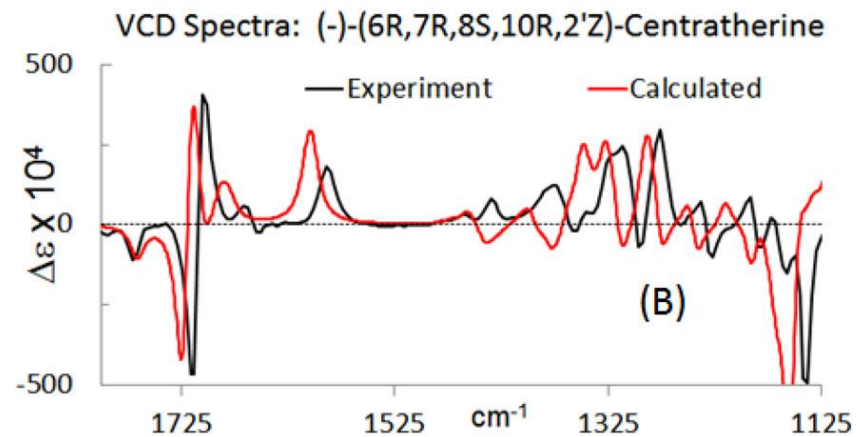
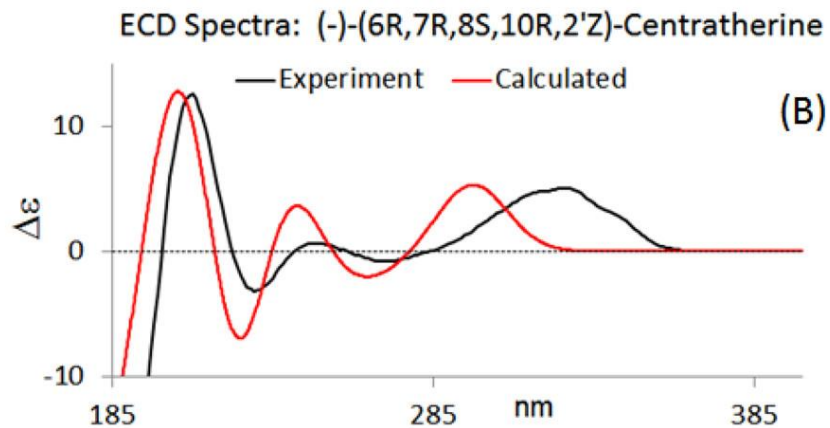
## 3. 总结与展望

# 背景：圆二色光谱简介

< 圆二色性 >



(-)-Centratherin



Grauso, L. et al. *Nat. Prod. Rep.* **2019**, *36*, 1005-1030.

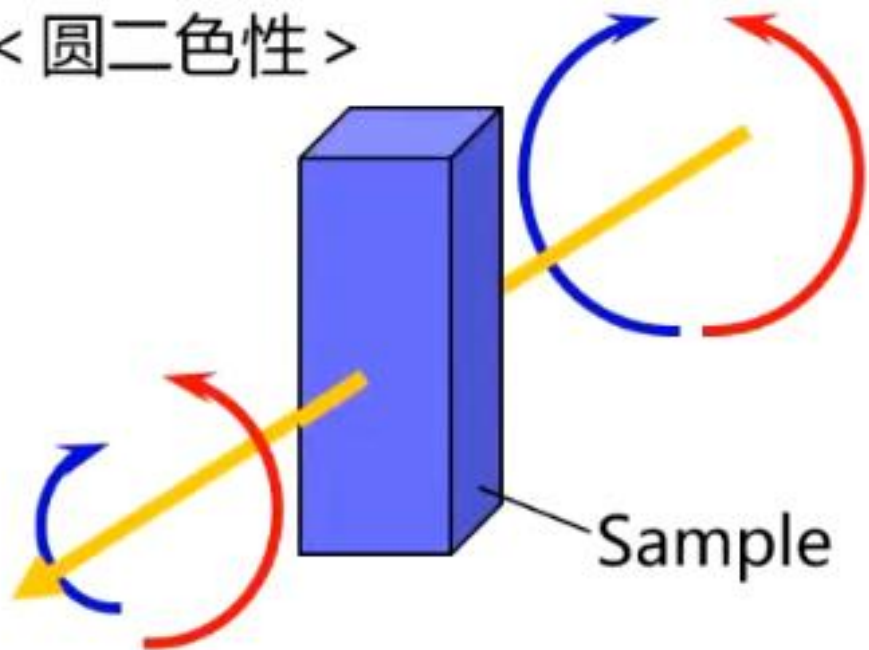
Merten, C. et al. *J. Org. Chem.* **2019**, *84*, 8797-8814.

Batista, A. N. L. et al. *Chem. Commun.* **2024**, *60*, 10439-10450.

Junior, F. M. et al. *J. Nat. Prod.* **2015**, *78*, 2617-2623.

# 背景：圆二色光谱简介

<圆二色性>



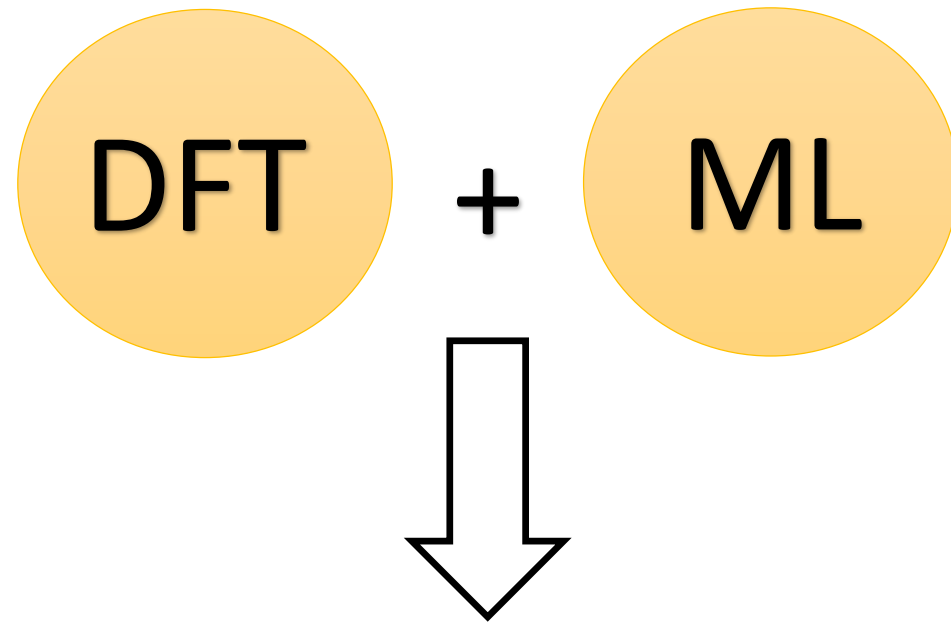
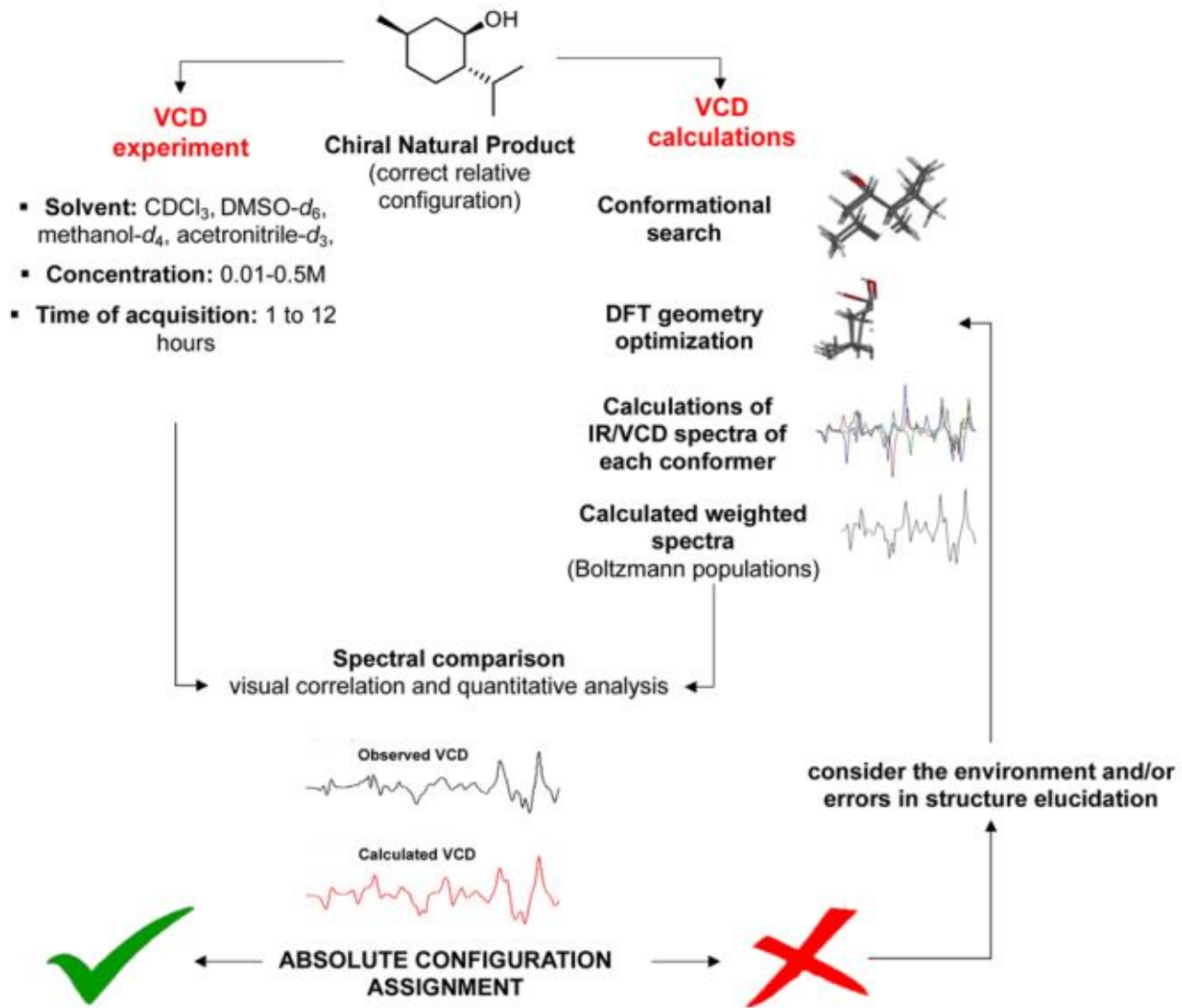
CD

- 灵敏度高
- 样品要求不高
- 可用经验规则
- 需要紫外生色基团

VCD

- 包含更多结构信息
- 不需要生色基团
- 对样品处理有一定要求
- 灵敏度较低

# 背景：计算方法



- DFT计算提供数据输入
- 机器学习建立模型
- 预测手性分子的多种性质

Grauso, L. et al. *Nat. Prod. Rep.* **2019**, *36*, 1005-1030.

Merten, C. et al. *J. Org. Chem.* **2019**, *84*, 8797-8814.

Batista, A. N. L. et al. *Chem. Commun.* **2024**, *60*, 10439-10450.

Szwarc, S. et al. *Org. Lett.* **2024**, *26*, 274-279.

# 背景：机器学习简介



*Reading*  
*Exploring*  
*Observing*  
...



*Brain*  
*Learning*



**Skills:**  
Translating  
Making decision  
Playing games



*Data*



*Machine*  
*Learning*



**Models:**  
Translating  
Making decision  
Playing games

# 背景：机器学习算法

多元线性回归 (MLR)  
线性判别分析 (LDA)

20世纪初

k近邻法 (kNN)  
逻辑回归  
贝叶斯分类器等

20世纪中期

前馈神经网络 (FNN)  
卷积神经网络 (CNN)  
支持向量机 (SVM)  
随机森林 (RF) 等

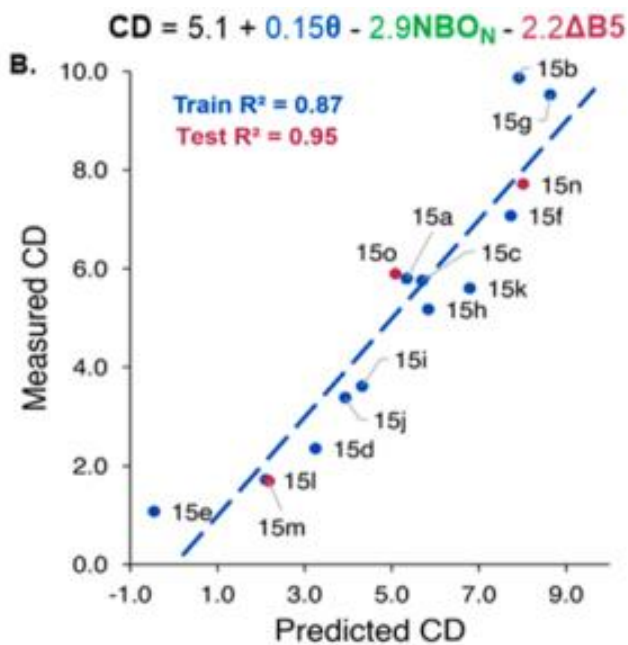
20世纪末

图神经网络 (GNN)  
扩散模型  
大语言模型等

21世纪后

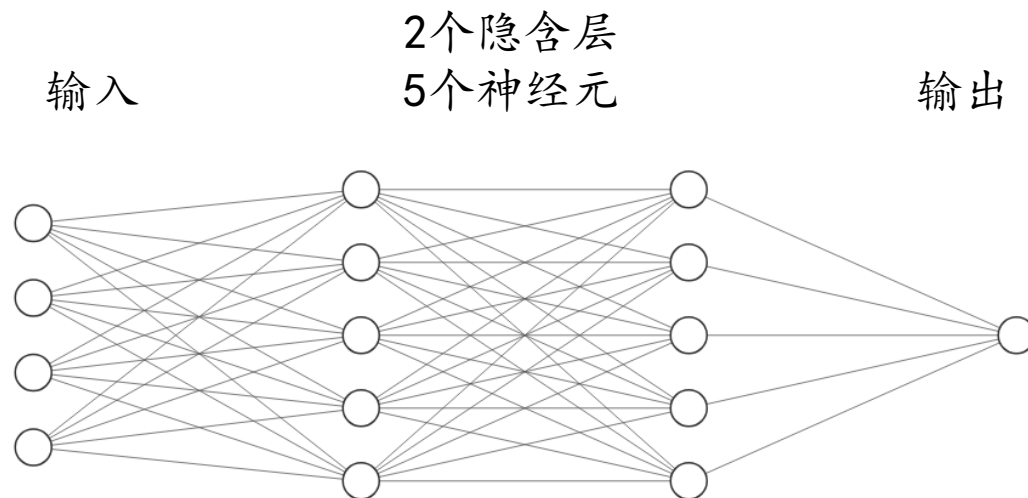
多元线性回归

- 线性关系模型
- 可解释
- 处理的问题相对简单



神经网络

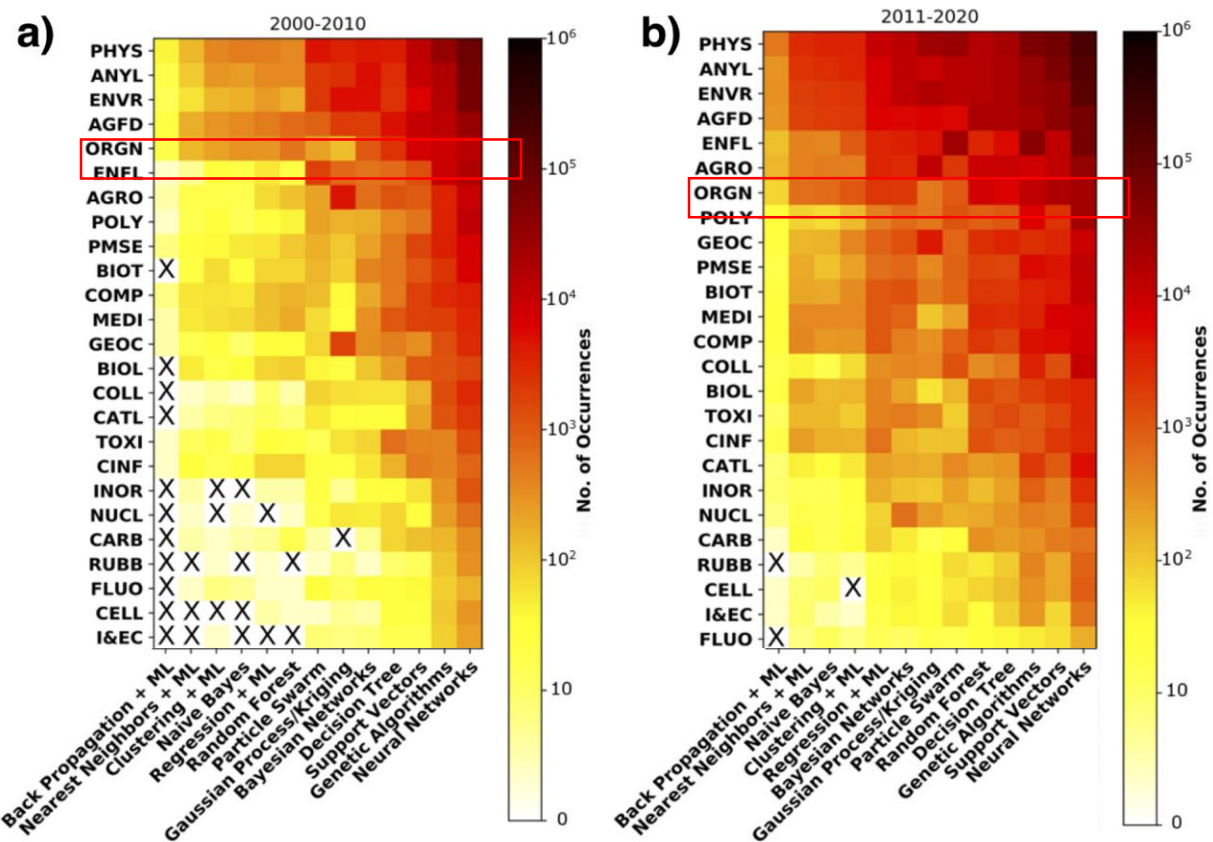
- 可用于处理复杂的问题
- 可用于大量数据
- 难以解释





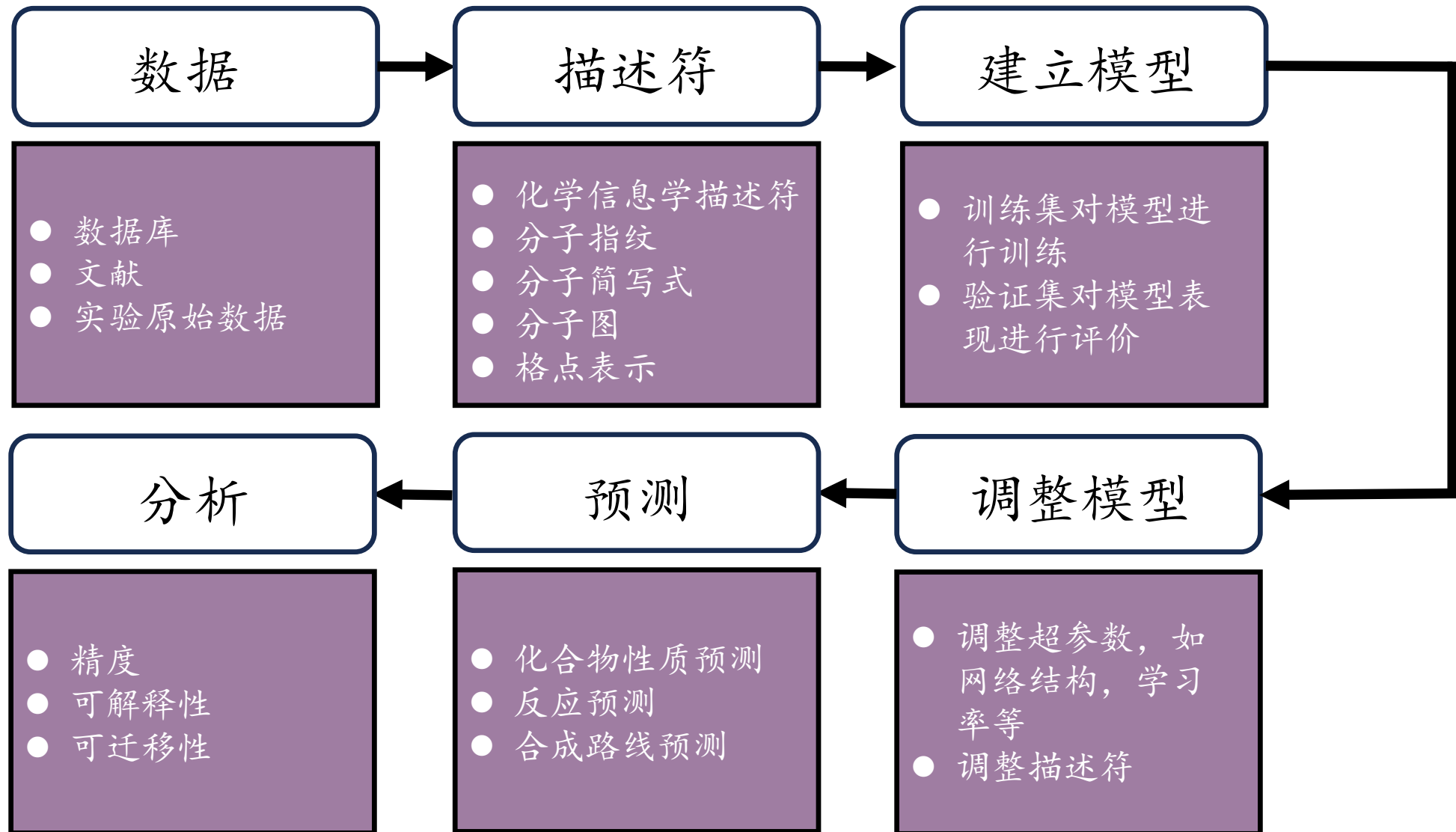
# 背景：机器学习的应用

- 应用最多的算法是神经网络
- 物理、分析、环境已经较多地应用机器学习
- 在有机化学领域中，应用最多的是蛋白质相关的研究



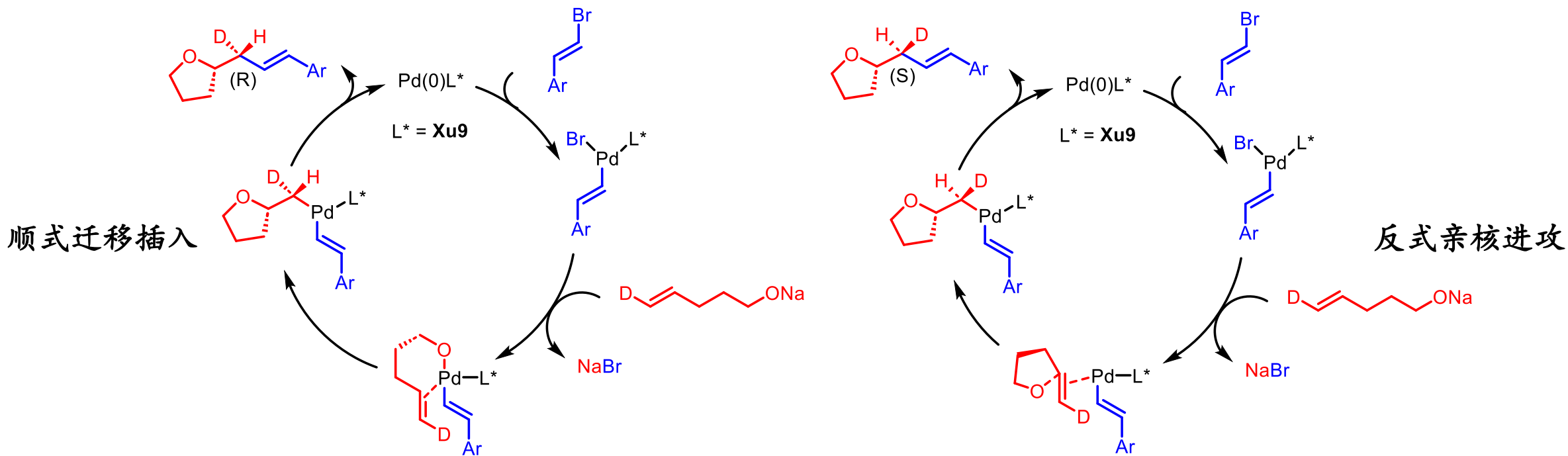
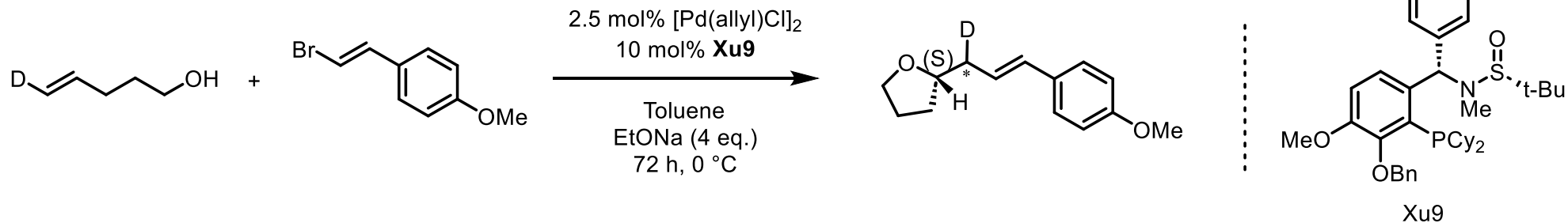
division	rank 1	rank 2	rank 3	rank 4
PHYS	electro* (56.3%)	spectroscopy (9.7%)	ion* (6.0%)	nano* (5.5%)
ANYL	sensor* (55.4%)	spectroscopy (13.1%)	characterization* (11.6%)	spectrometry (4.3%)
ENVR	*sensor* (60.9%)	soil* (14.5%)	water quality (4.2%)	environmental monitor* (3.0%)
AGFD	protein* (31.9%)	agricultur* (18.2%)	food (10.8%)	fruit* (5.7%)
ENFL	fuel* (19.2%)	petroleum (11.6%)	energy efficiency (11.1%)	batter* (10.7%)
AGRO	soil (43.3%)	crop* (25.4%)	groundwater (11.5%)	developing countr* (4.4%)
ORGN	protein* (64.6%)	amino acid* (19.7%)	peptide* (8.4%)	aromatic* (3.2%)

# 背景：机器学习的一般流程



1. 背景介绍
2. 计算化学方法与圆二色光谱结合的应用
  - 2.1 预测圆二色光谱和确定手性化合物的绝对构型
  - 2.2 预测手性化合物的ee值
  - 2.3 其它应用
3. 总结与展望

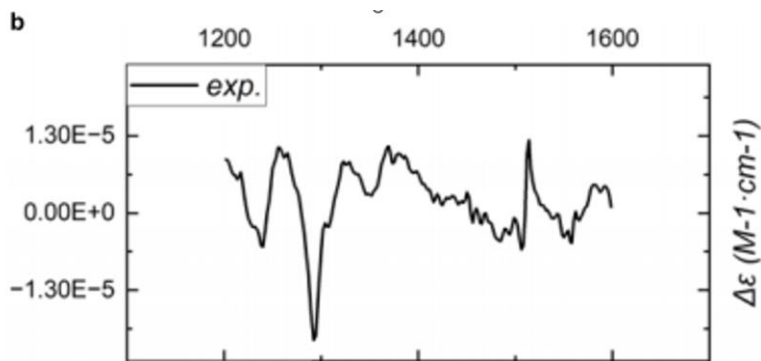
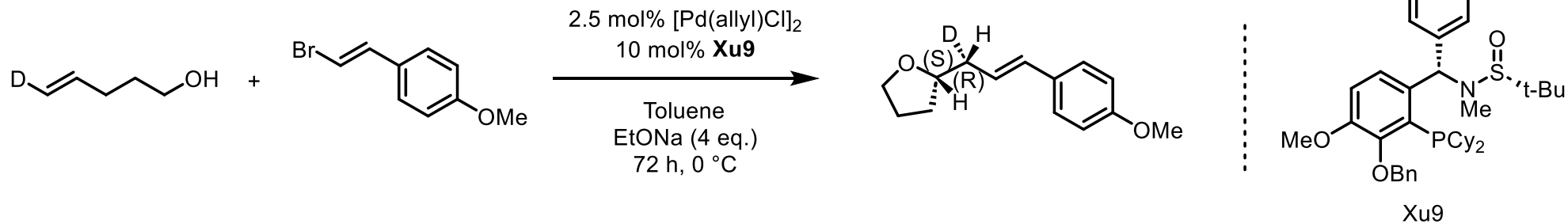
# DFT方法预测VCD光谱确定手性分子的绝对构型



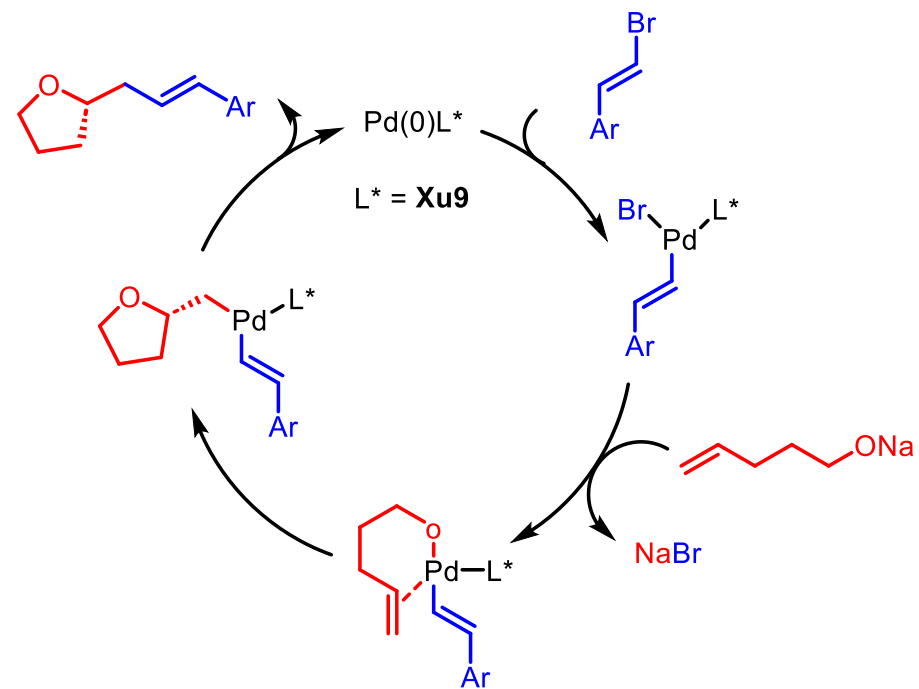
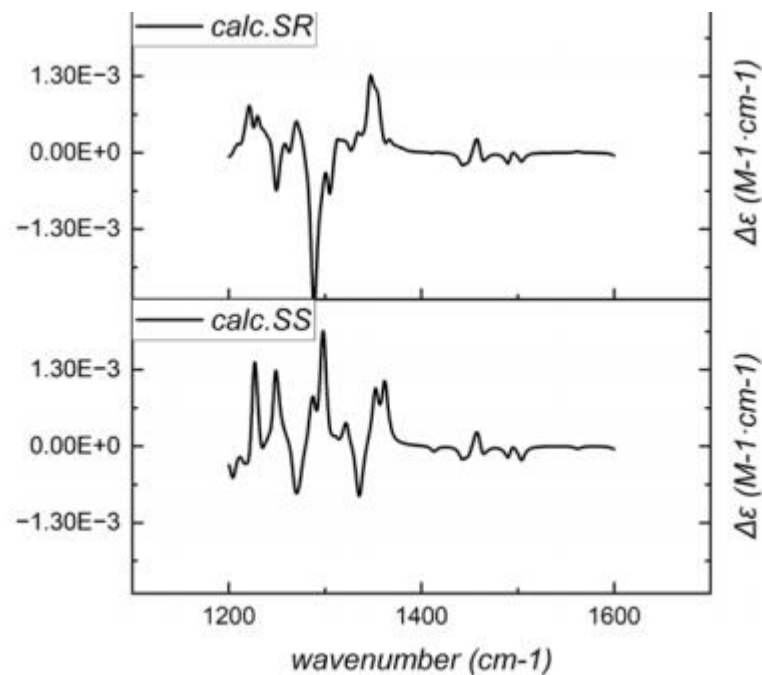
顺式迁移插入？反式亲核进攻？

需确定产物中氘代碳的绝对构型

# DFT方法预测VCD光谱确定手性分子的绝对构型

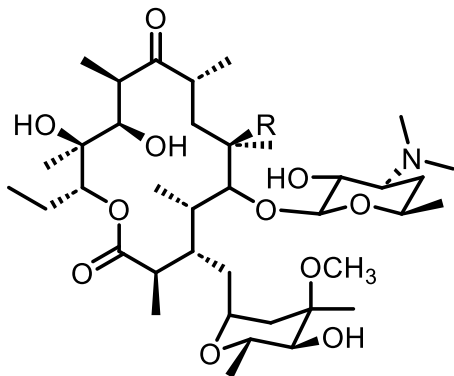


确定产物的构型为SR



顺式迁移插入！！

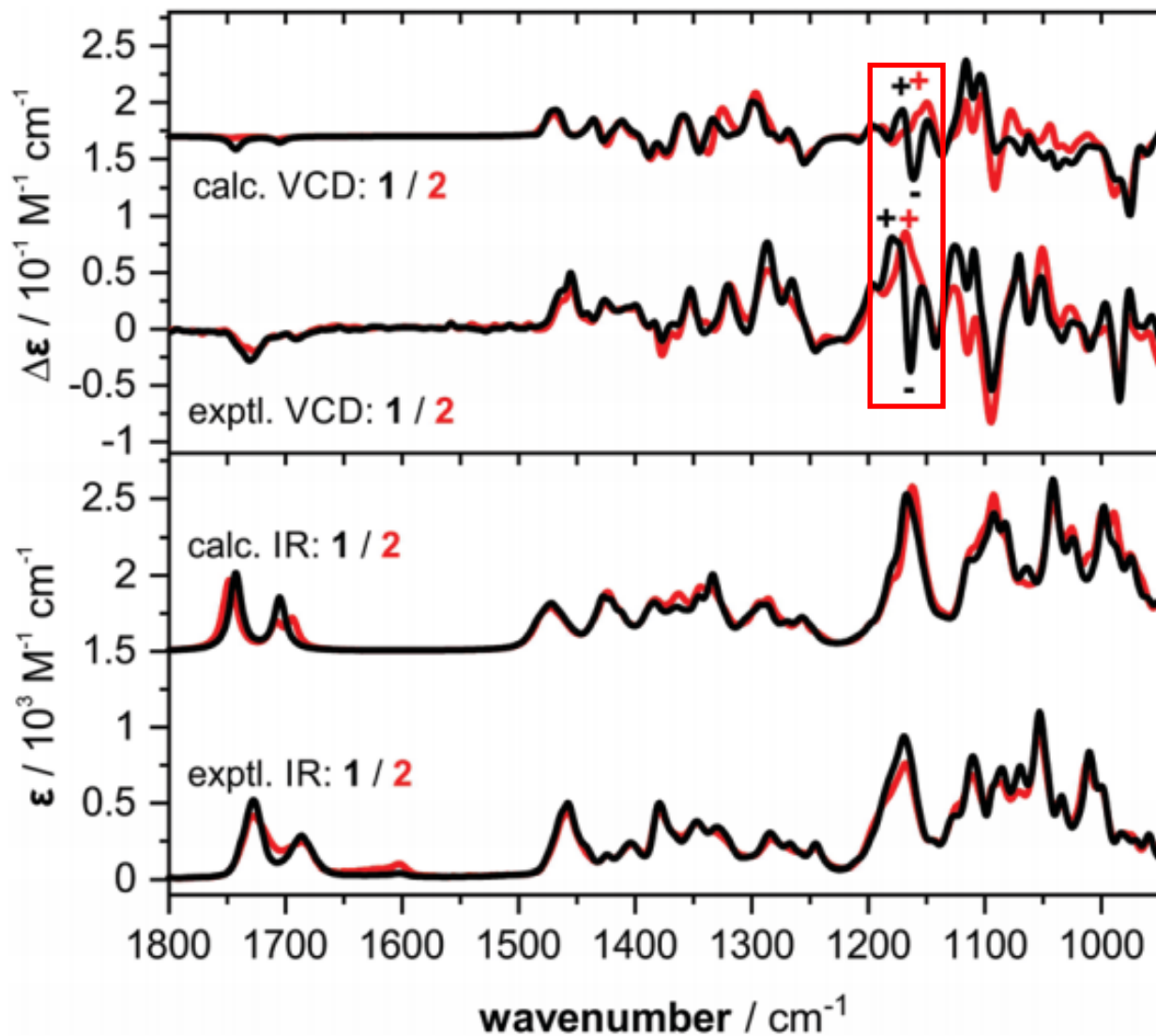
# DFT方法预测CD光谱用于区分手性大分子



clarithromycin (1) R = OCH<sub>3</sub>

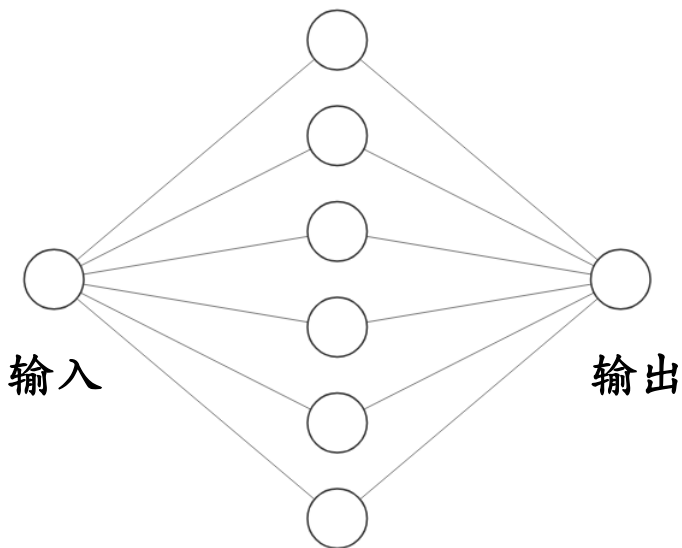
erythromycin (2) R = OH

- 确定了天然产物的绝对构型
- 区分了两种相似的大分子



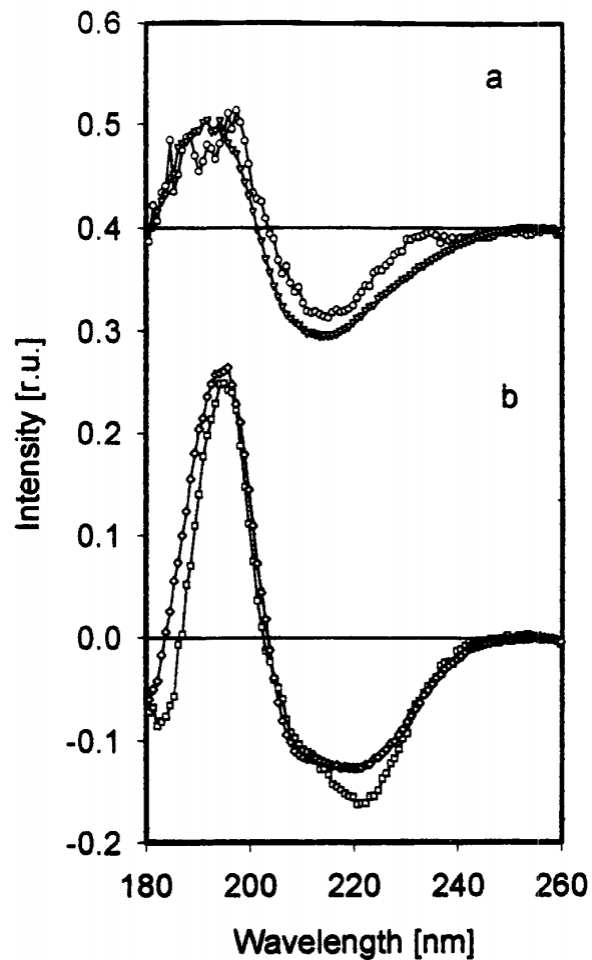
# 机器学习方法预测CD光谱

1层隐含层，6个神经元



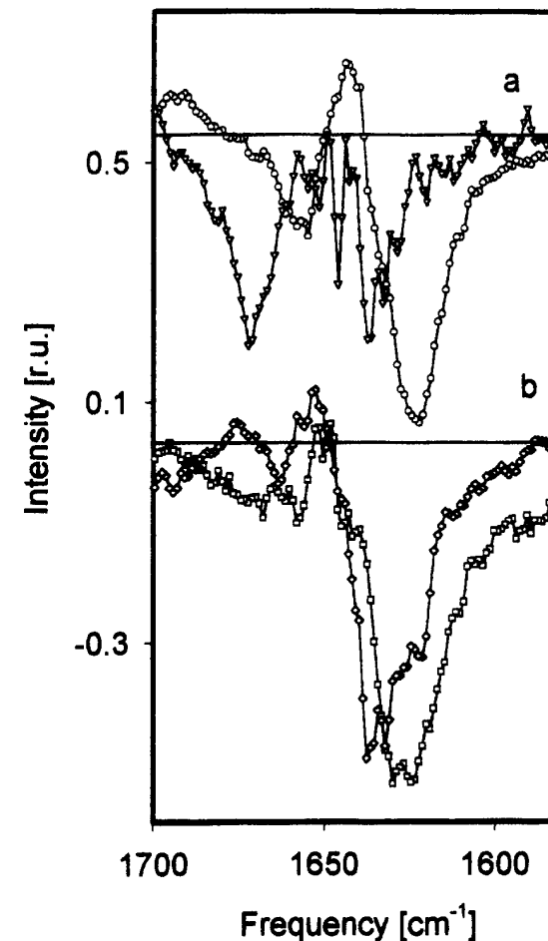
共28种蛋白质的实验CD和VCD光谱数据作为输入

由VCD预测得到的CD



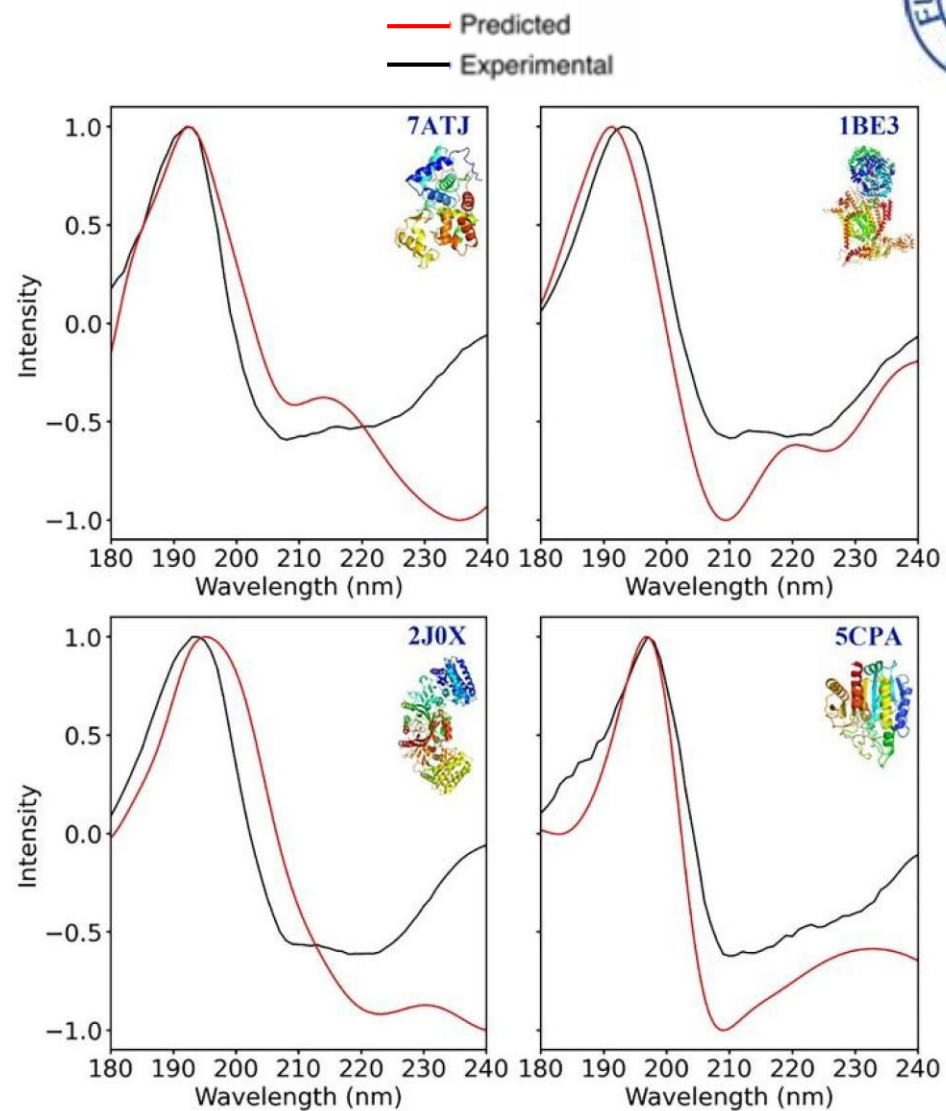
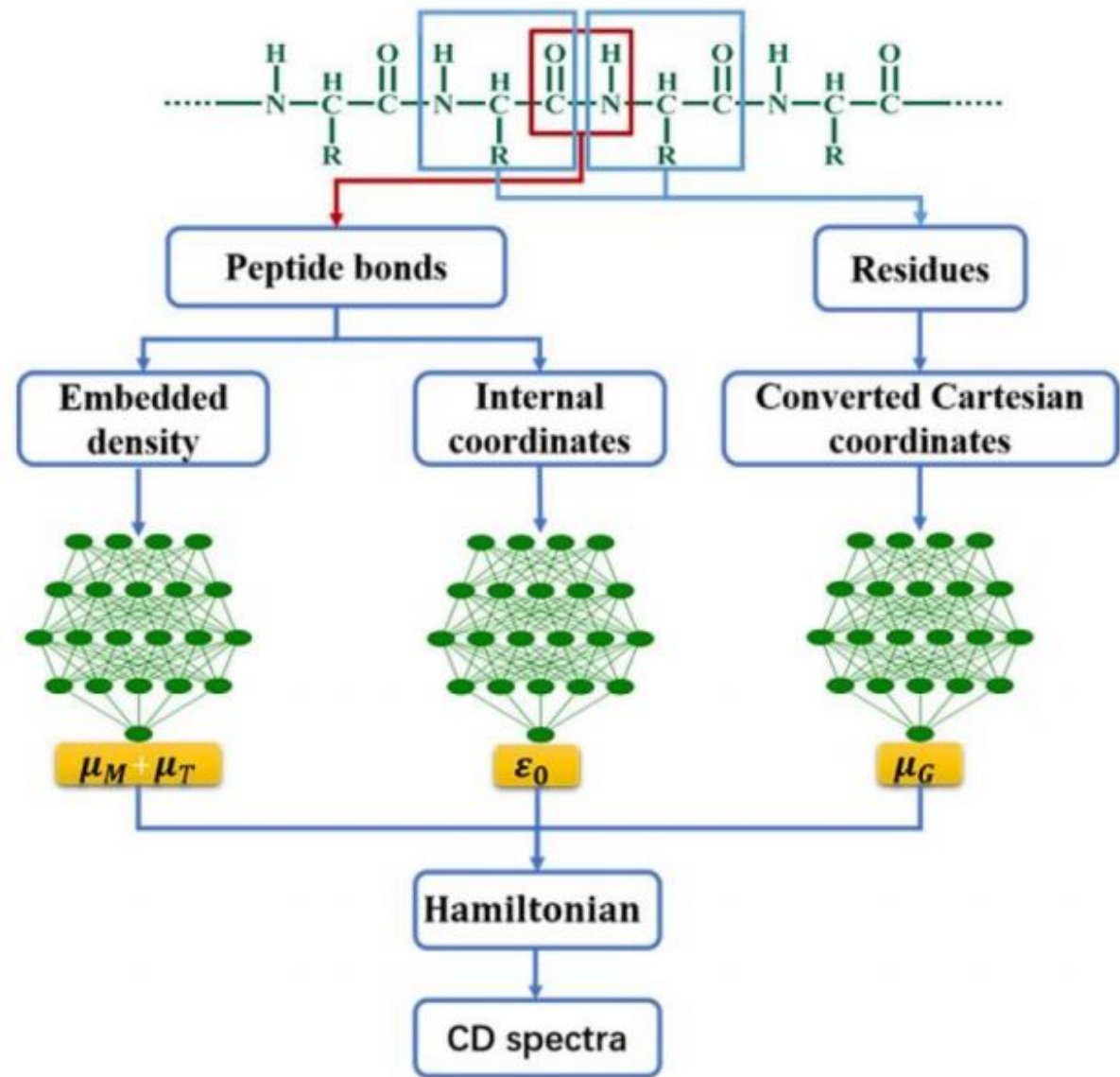
重合度比较高

由CD预测得到的VCD



重合度较低

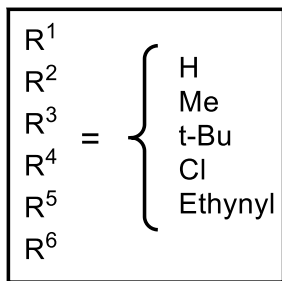
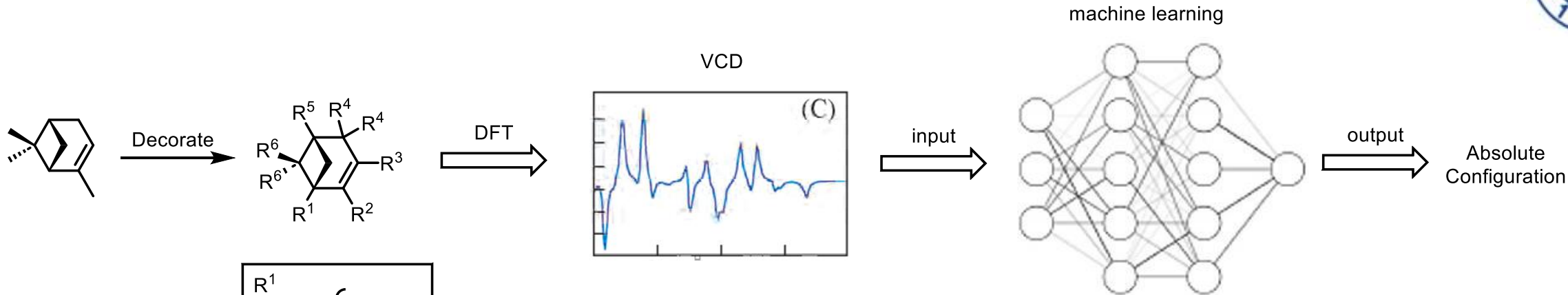
# 机器学习方法预测CD光谱



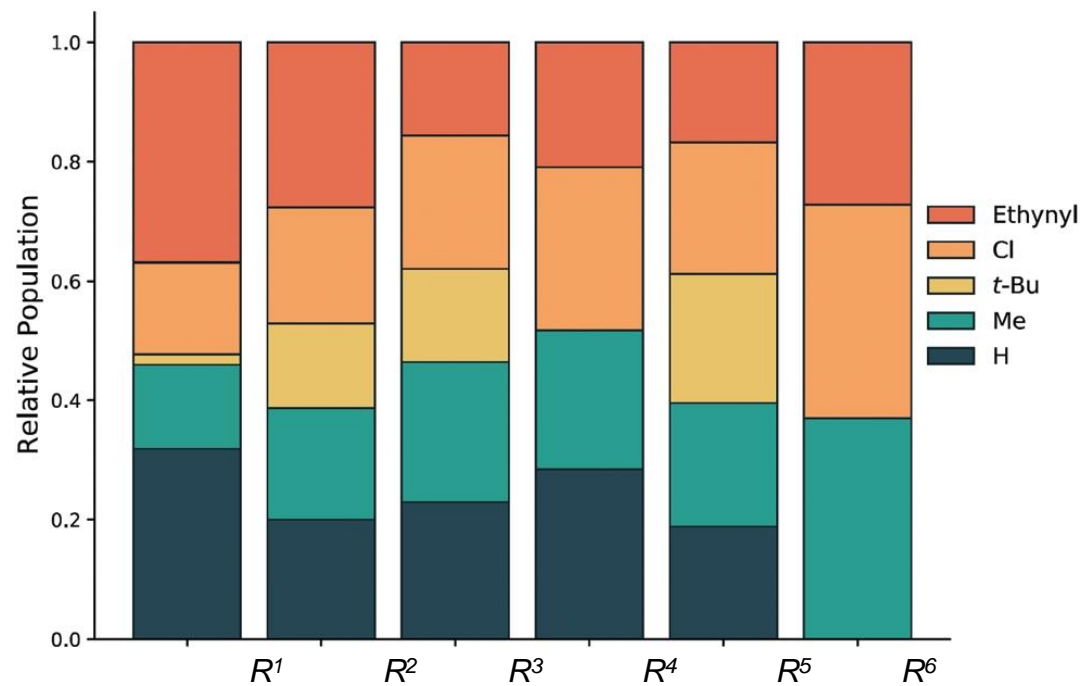
预测的CD谱与实验谱图重合度比较高



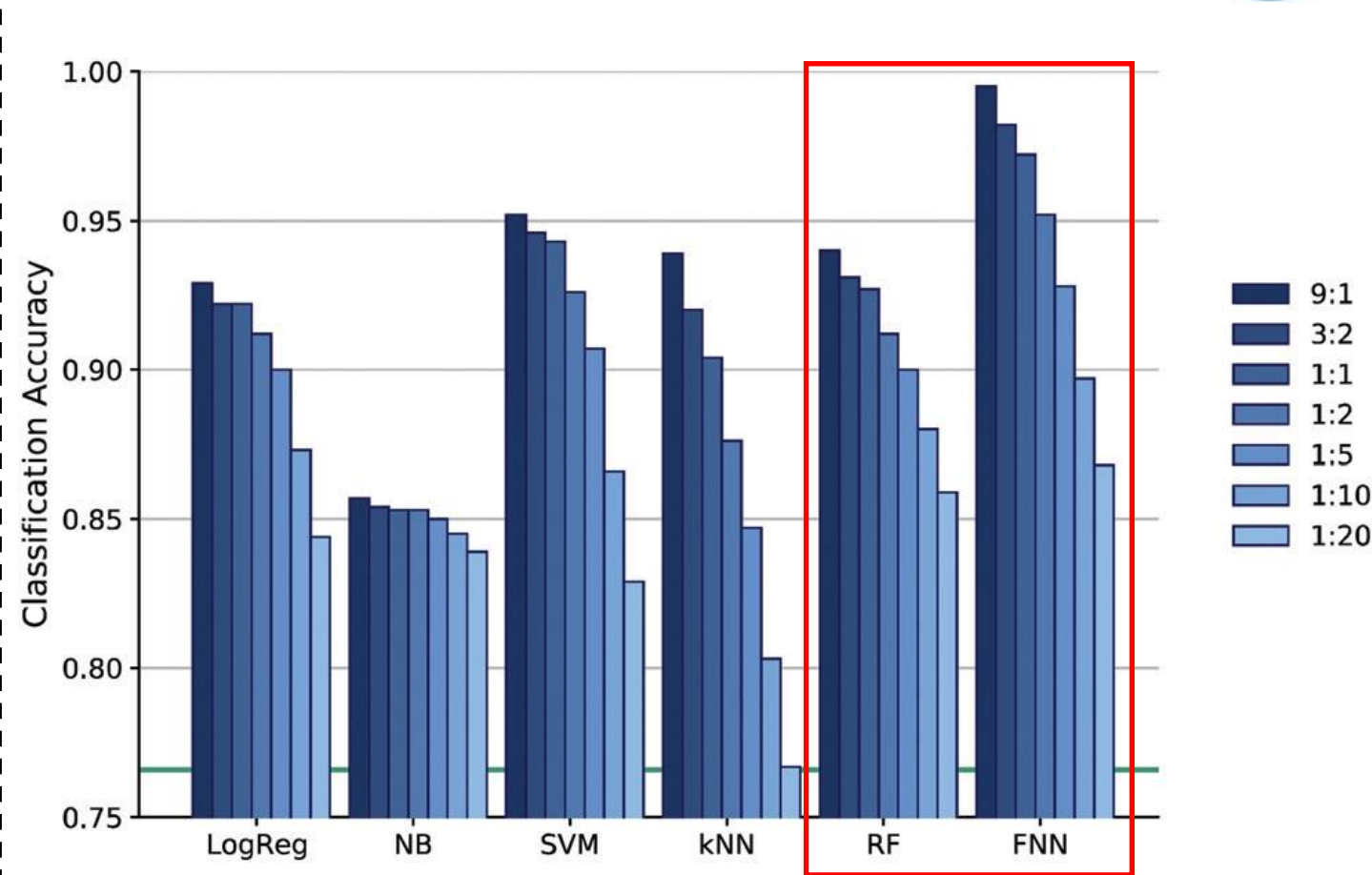
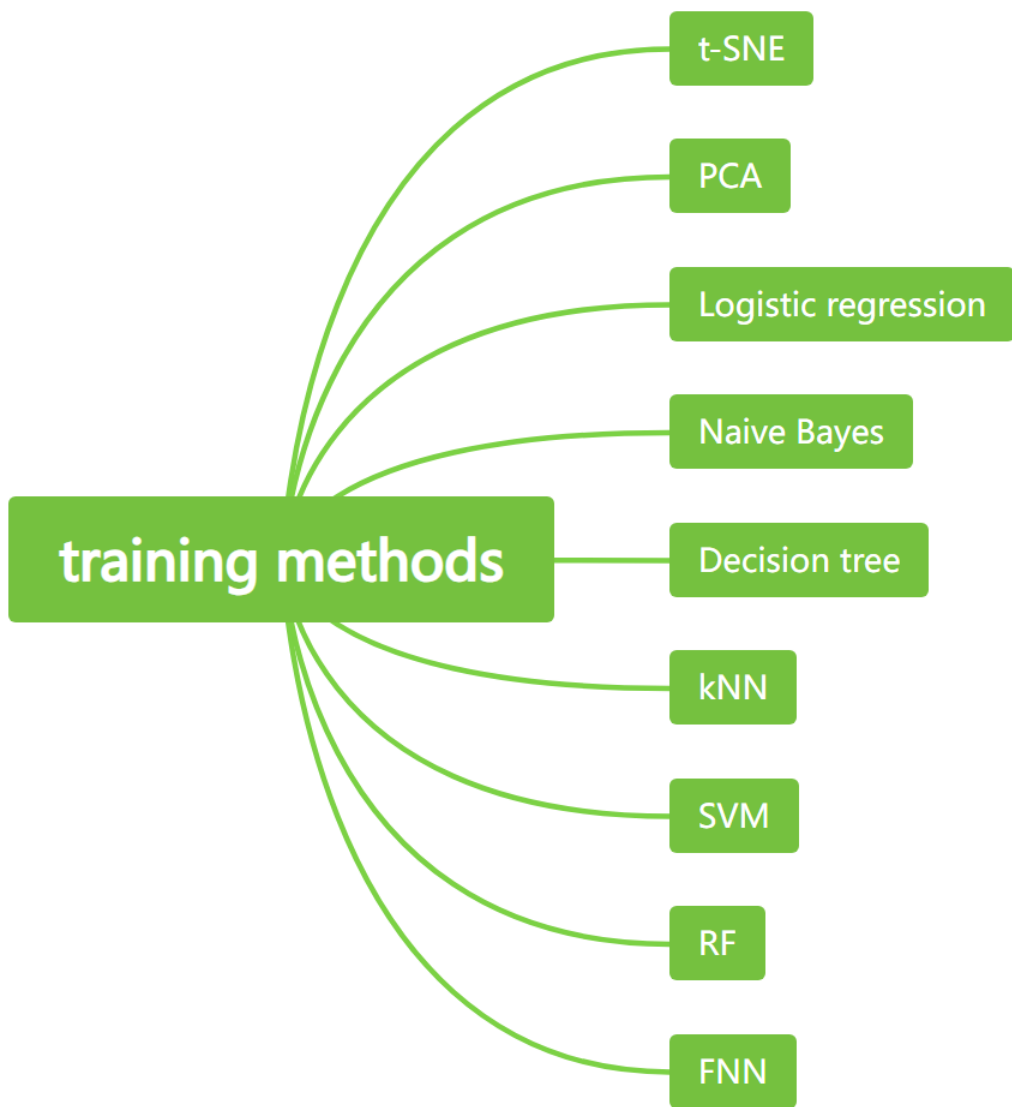
# 机器学习方法确定手性化合物的绝对构型



- 对骨架的取代基施加了一些限制
- 共3945对分子
- 计算的VCD谱作为输入



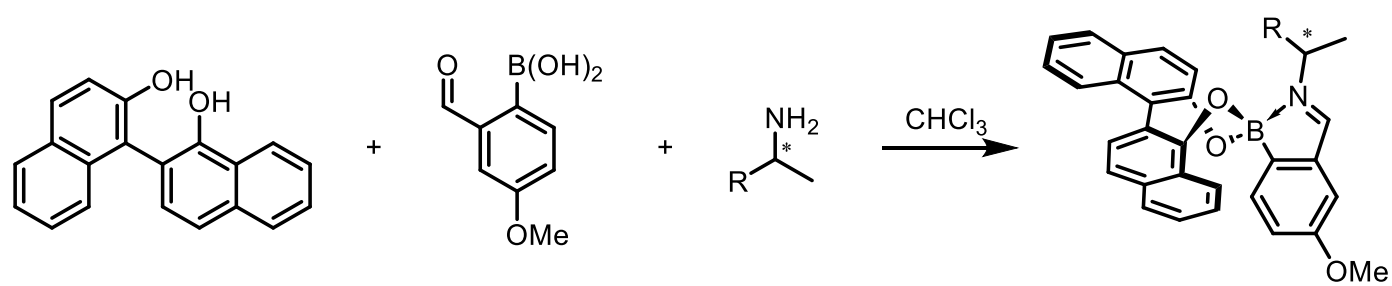
# 机器学习方法确定手性化合物的绝对构型



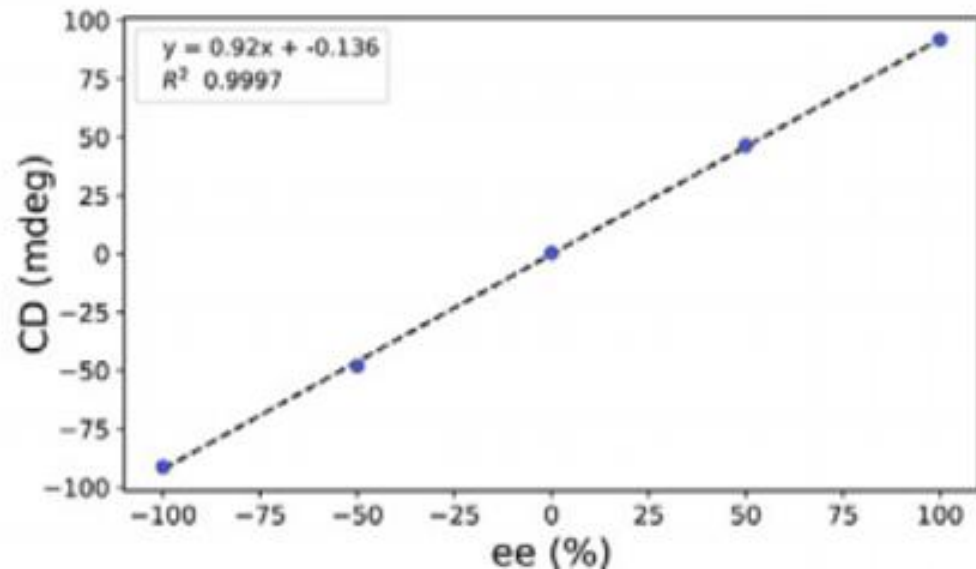
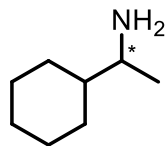
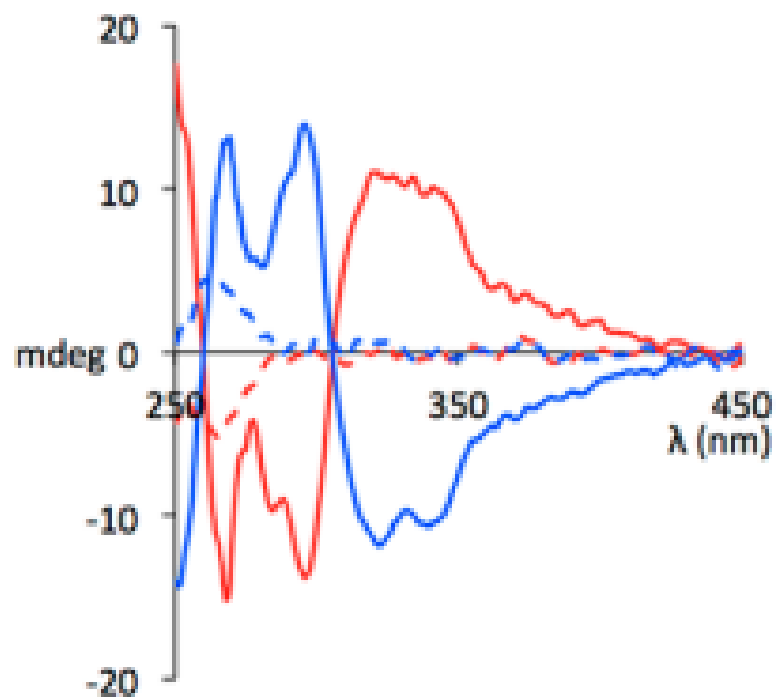
多种模型的结果很好

1. 背景介绍
2. 计算化学方法与圆二色光谱结合的应用
  - 2.1 预测圆二色光谱和确定手性化合物的绝对构型
  - 2.2 预测手性化合物的ee值
  - 2.3 其它应用
3. 总结与展望

# 机器学习方法预测手性化合物的ee值

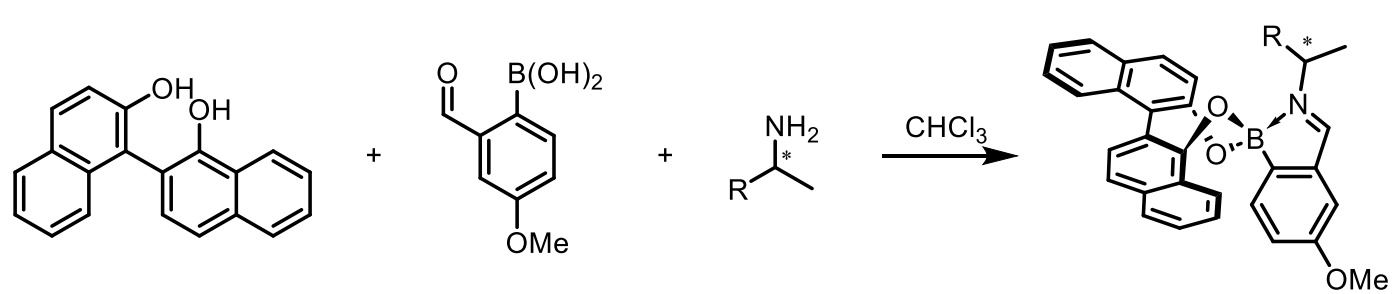


- 该转化可加大CD信号
- 共使用15种手性胺
- 输入对每种分子计算的一些量化参数和对应的最高CD强度

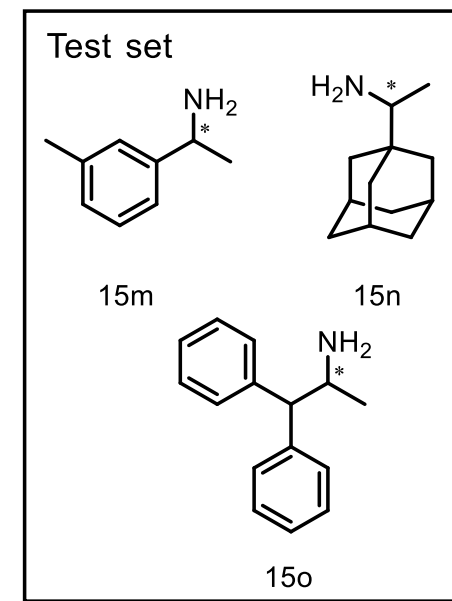
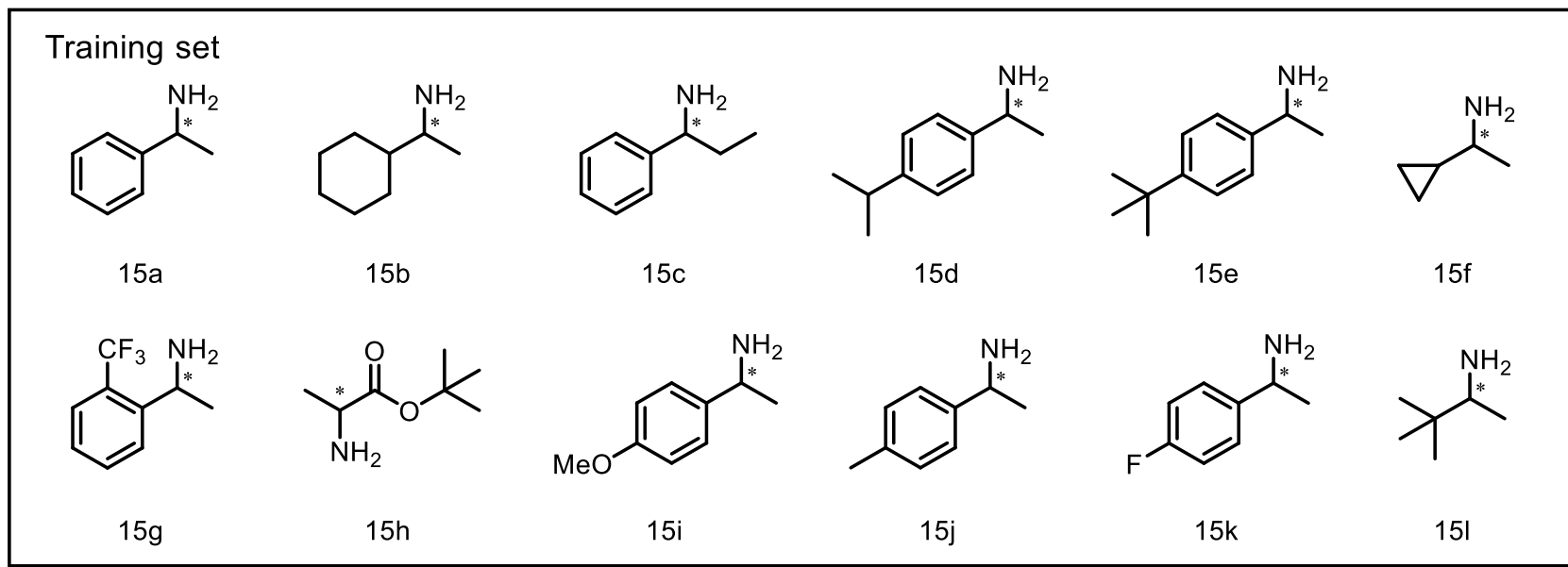


实验测定的一个校准曲线

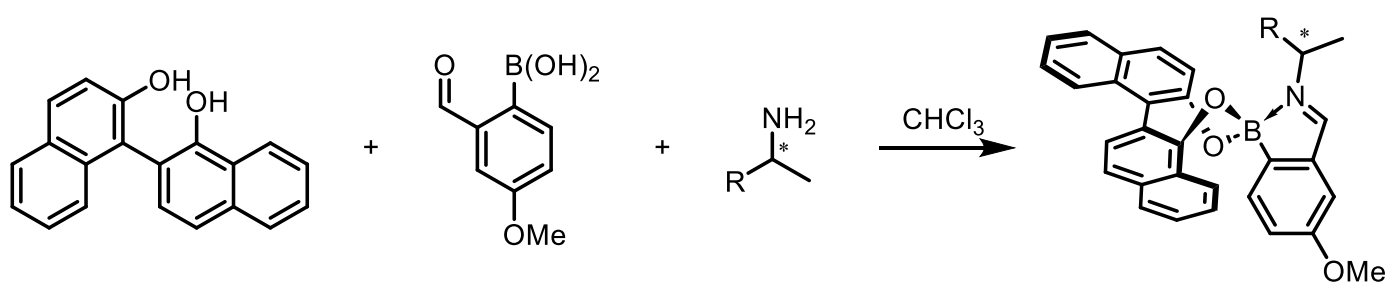
# 机器学习方法预测手性化合物的ee值



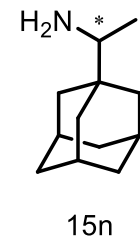
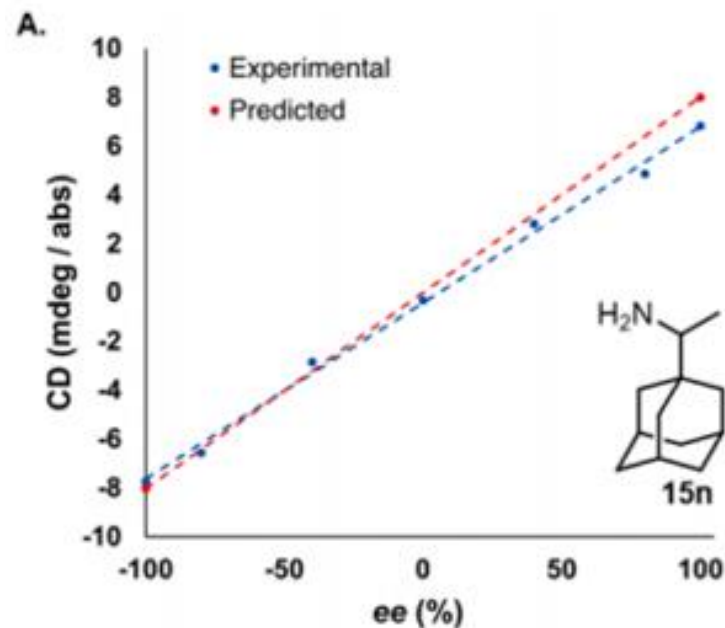
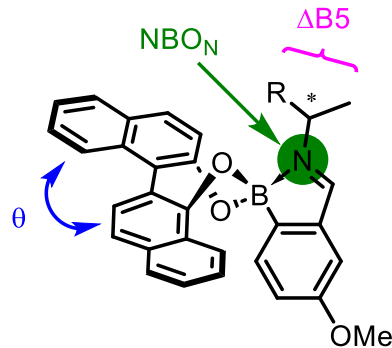
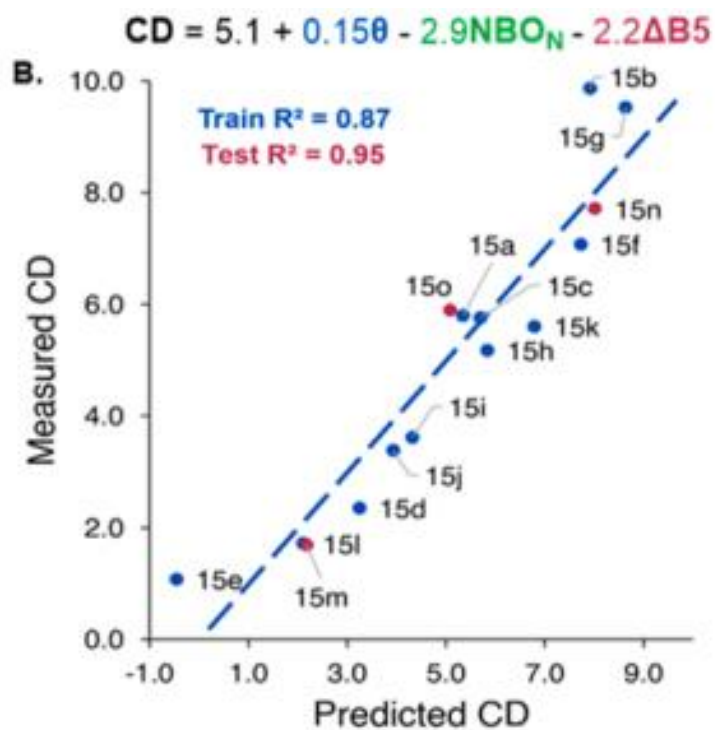
- 该转化可加大CD信号
- 共使用15种手性胺
- 输入对每种分子计算的一些量化参数和对应的最高CD强度



# 机器学习方法预测手性化合物的ee值



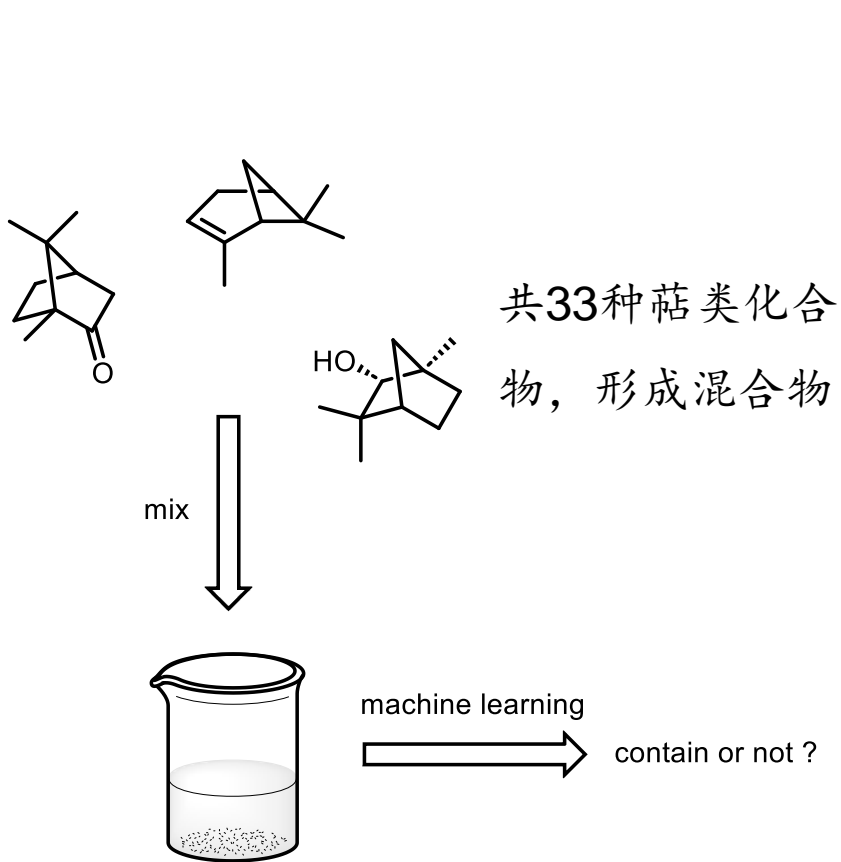
- 该转化可加大CD信号
- 共使用15种手性胺
- 输入对每种分子计算的一些量化参数和对应的最高CD强度



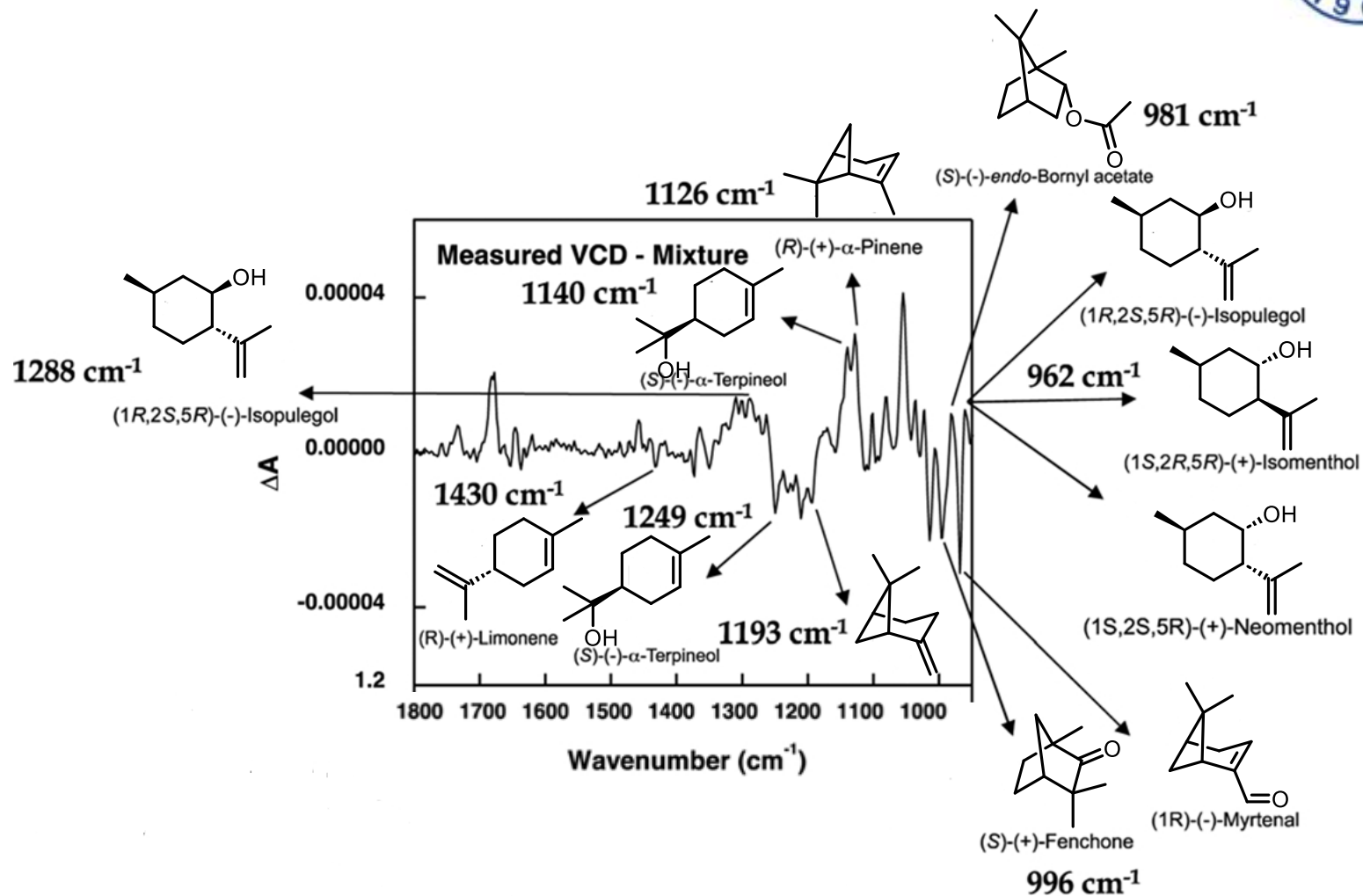
成功由VCD→ee值

1. 背景介绍
2. 计算化学方法与圆二色光谱结合的应用
  - 2.1 预测圆二色光谱和确定手性化合物的绝对构型
  - 2.2 预测手性化合物的ee值
  - 2.3 其它应用
3. 总结与展望

# 机器学习方法预测混合物的成分



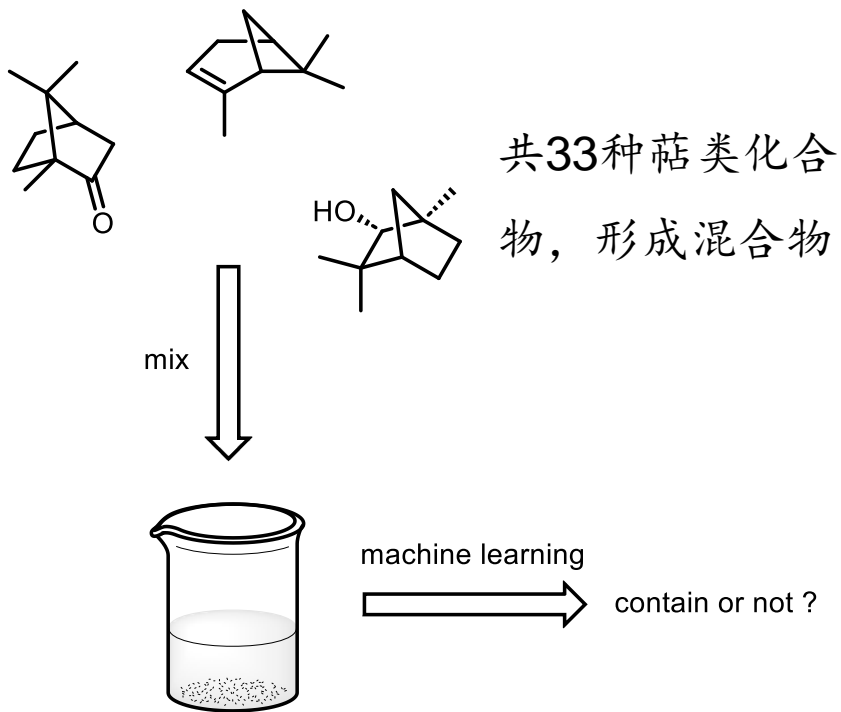
以单个萜和混合物的VCD光谱为输入，令机器学习混合物中每种萜的光谱特征



混合物中部分萜的光谱特征



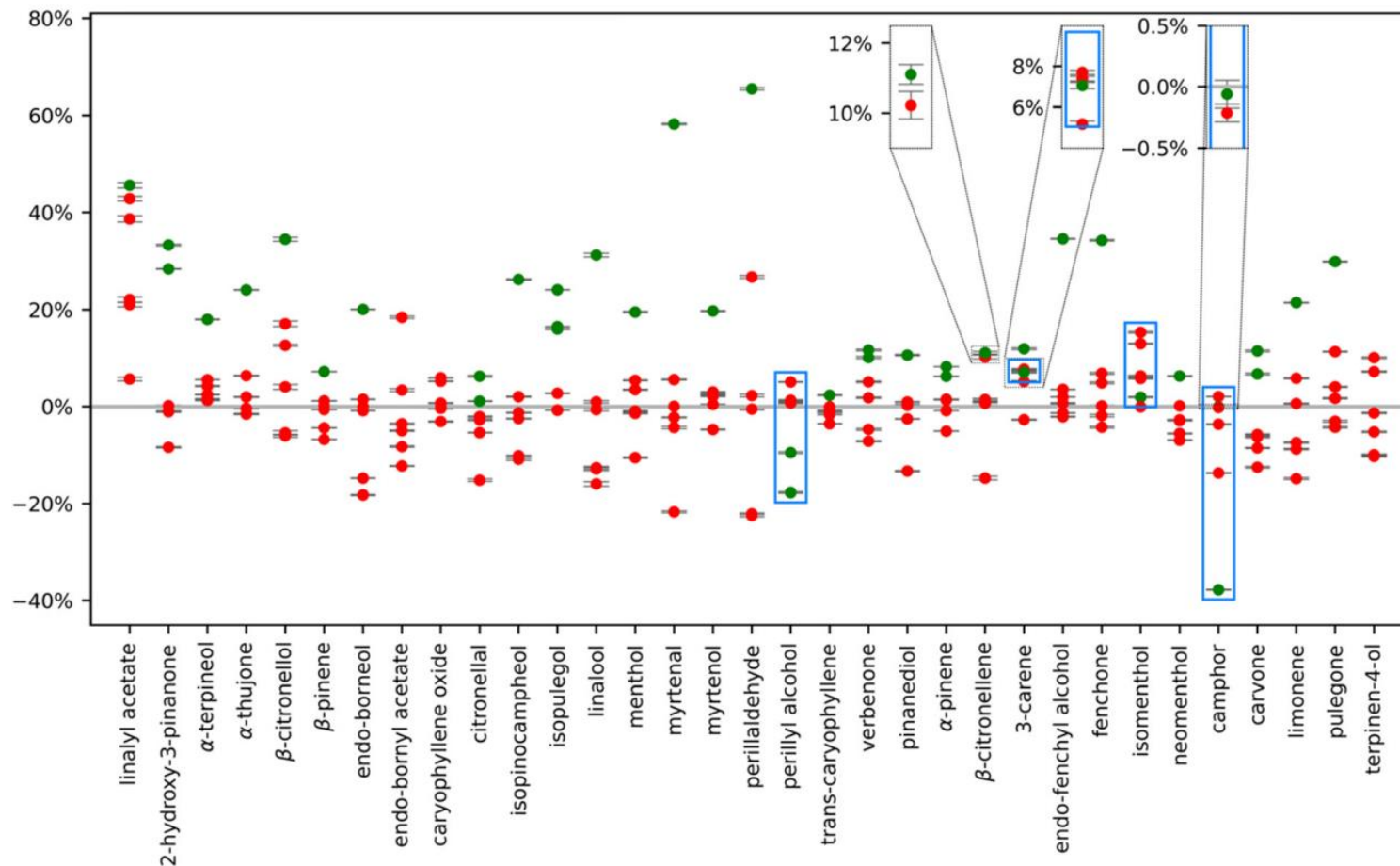
# 机器学习方法预测混合物的成分



以单个萜和混合物的VCD

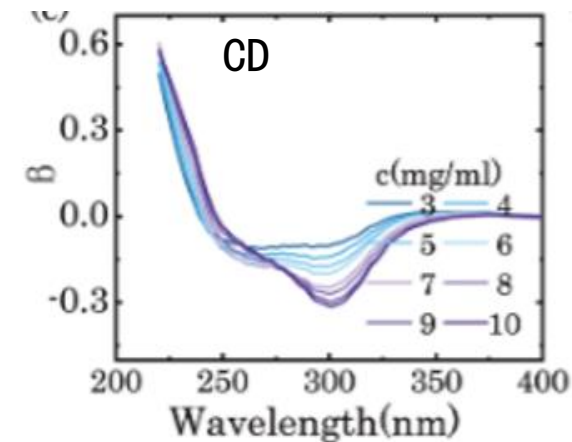
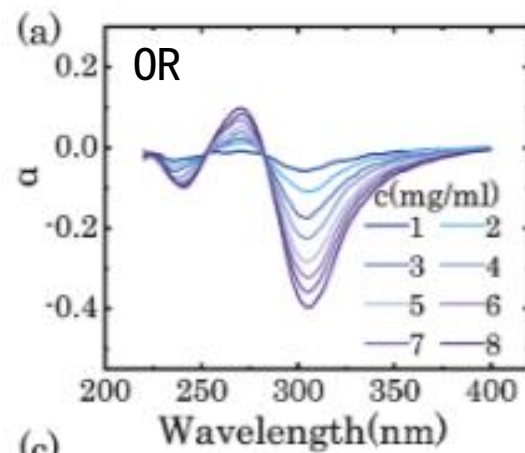
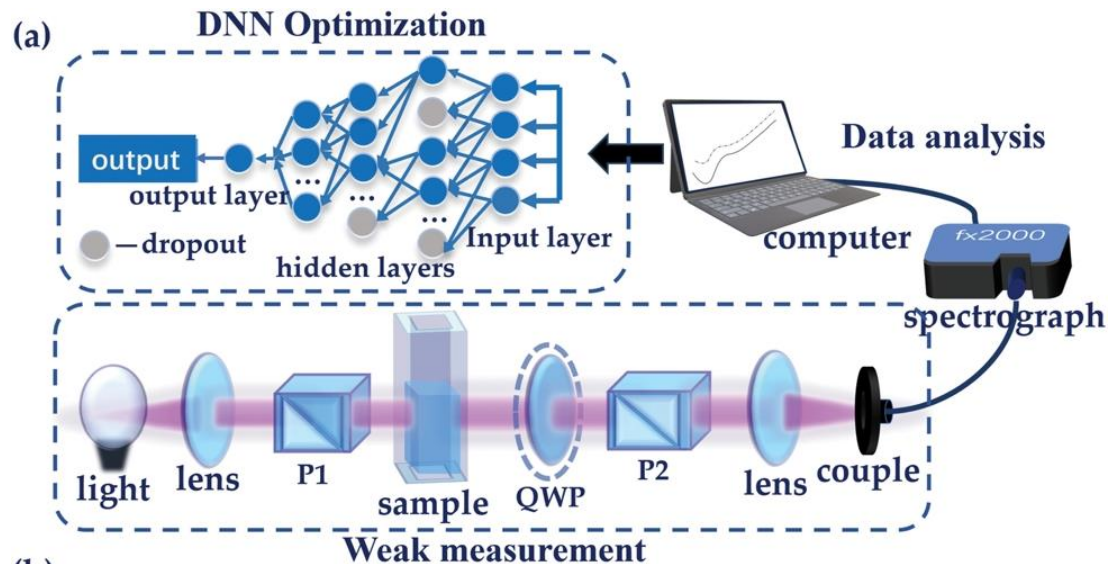
光谱为输入，令机器学习混

合物中每种萜的光谱特征

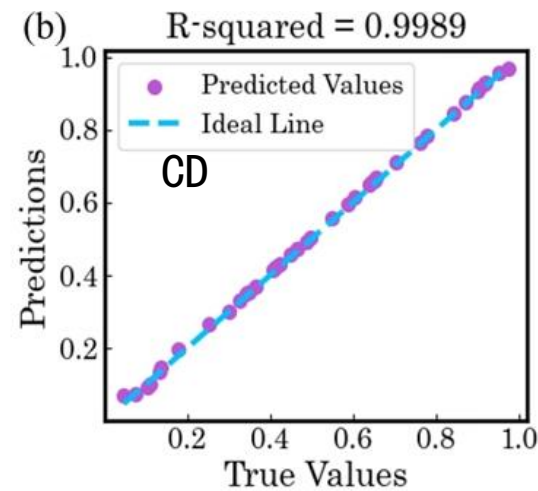
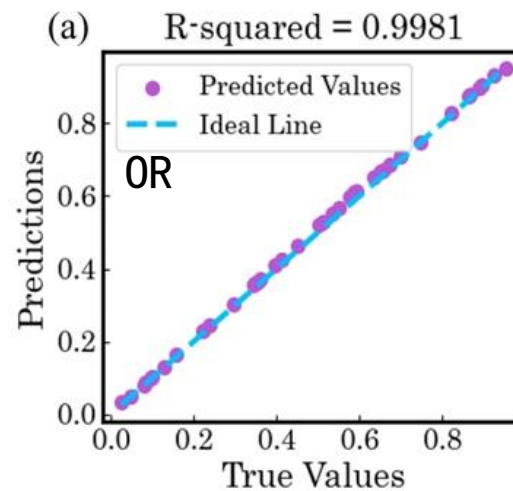


能够从不同混合物中成功识别出29种萜类化合物

# 机器学习方法预测手性化合物的浓度

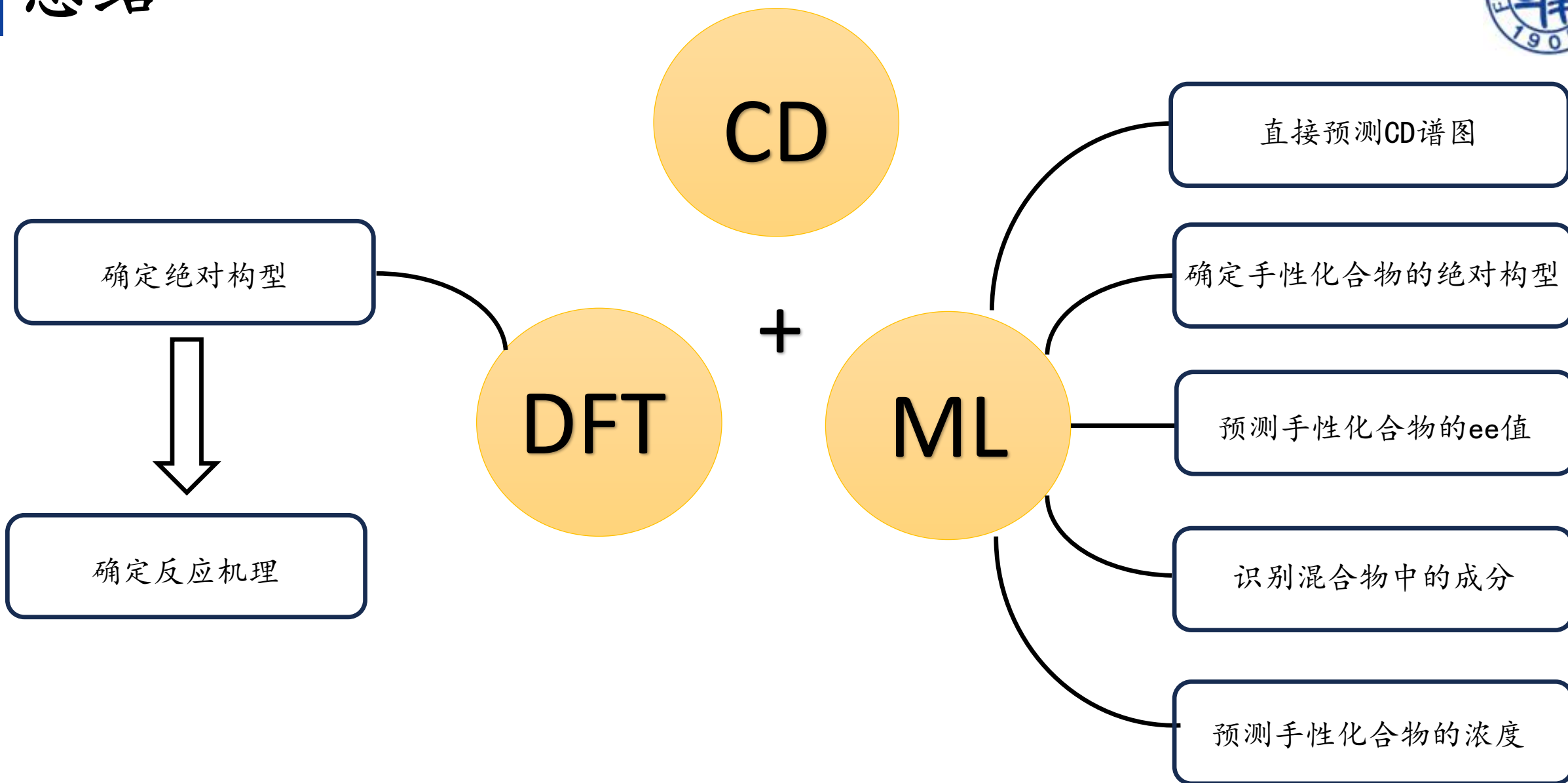


- 建立了实时监测系统
- 测量不同浓度的手性溶液的OR和CD光谱数据为输入，浓度作为输出

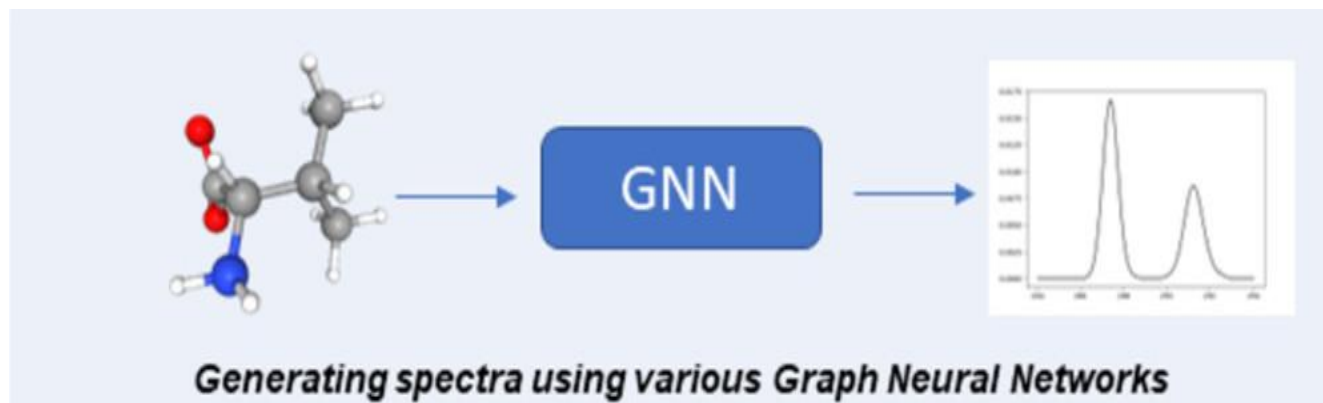


1. 背景介绍
2. 计算化学方法与圆二色光谱结合的应用
  - 2.1 预测圆二色光谱和确定手性化合物的绝对构型
  - 2.2 预测手性化合物的ee值
  - 2.3 其它应用
3. 总结与展望

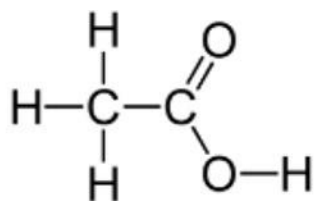
# 总结



# 展望：更先进的神经网络算法






1. Molecule








2. Molecular graph

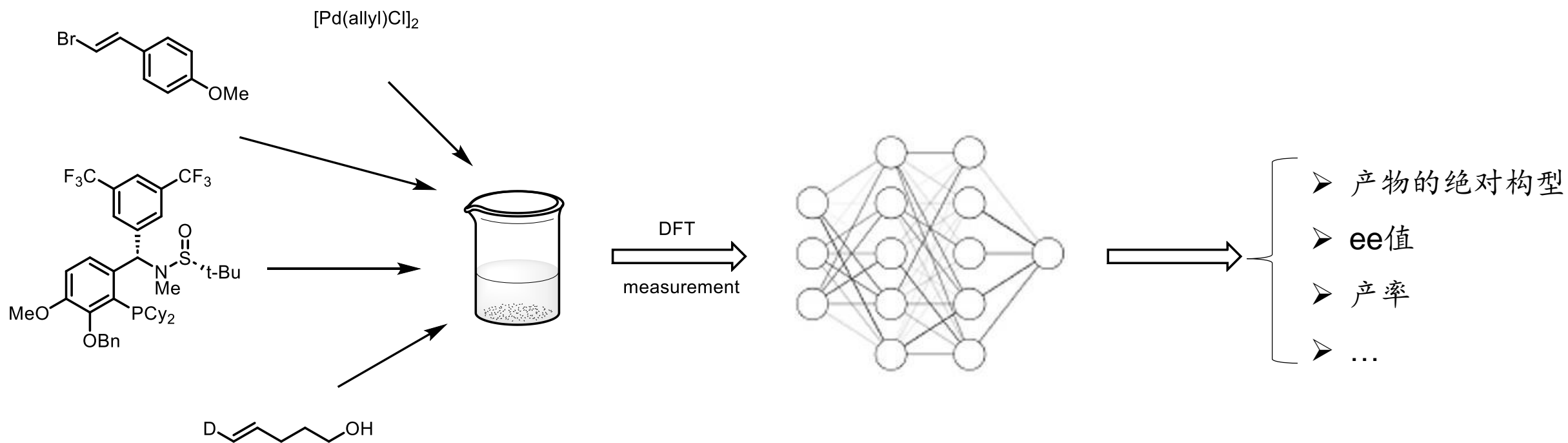
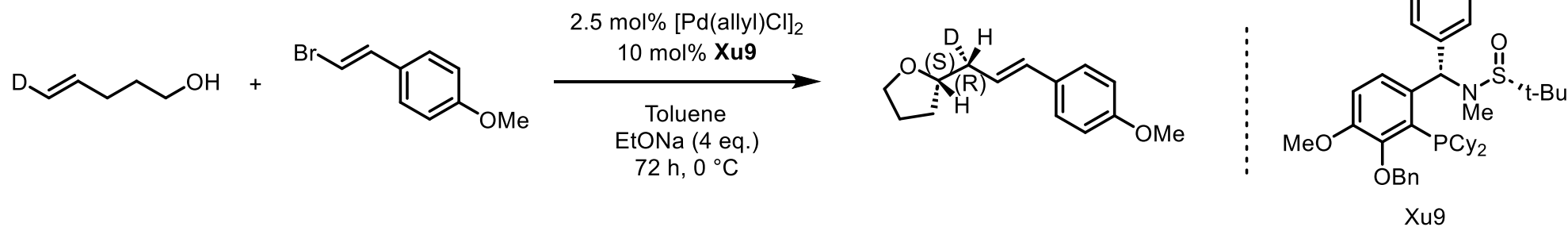


3. Atom and bond features

Atom (Node) features		
	Symbol	Atomic Number
	H	1
	O	8
	C	6

Bond (Edge) features		
	Type	Length
	Single	1.4
	Double	1.2
	Single	1.1
	Single	1.0
	Single	1.5

# 展望：应对更复杂更真实的体系



谢谢大家！  
敬请批评指正！