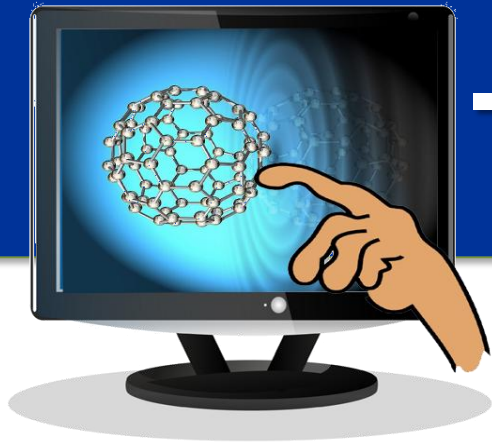


# 从AlphaGo 到材料基因 ——功能材料设计

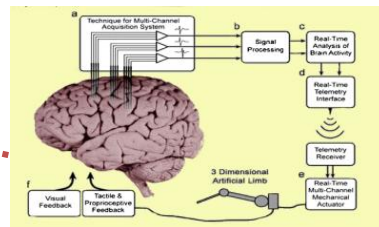
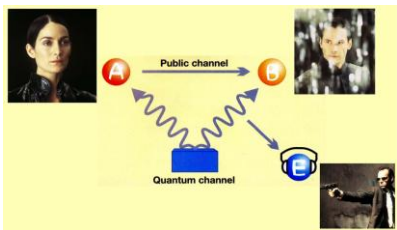


江俊 [jiangj1@ustc.edu.cn](mailto:jiangj1@ustc.edu.cn)

<http://staff.ustc.edu.cn/jiangj1>

微尺度物质科学国家研究中心

# 微尺度物质科学国家研究中心



重大关键技术

信息科技

量子比特

原子分子

纳米科技

交叉前沿与基础

基因蛋白

神经元

网络与认知

生物技术

量子信息、脑科学、人工智能、大数据

学科在融合



# 新材料——全球战略

## 北美

材料基因组计划  
21世纪国家纳米纲要  
未来工业材料计划  
先进汽车材料计划  
化石能材料计划  
建筑材料计划

## 欧洲

第六个框架计划  
COST计划  
尤里卡计划

## 俄罗斯

2030年前新材料与技术发展战略

## 日本

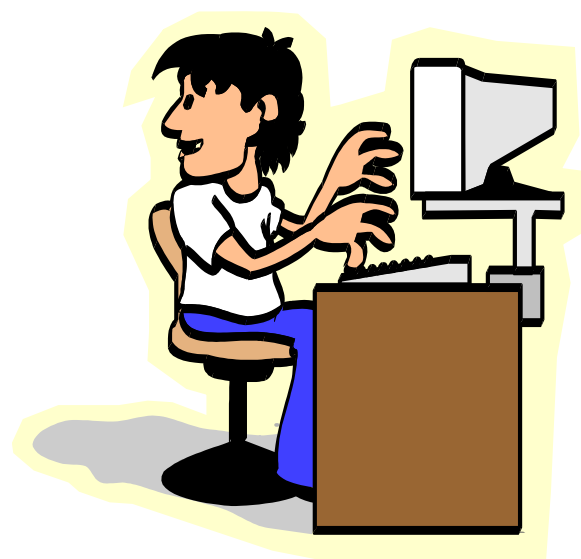
第五期科学技术基本计划

## 中国

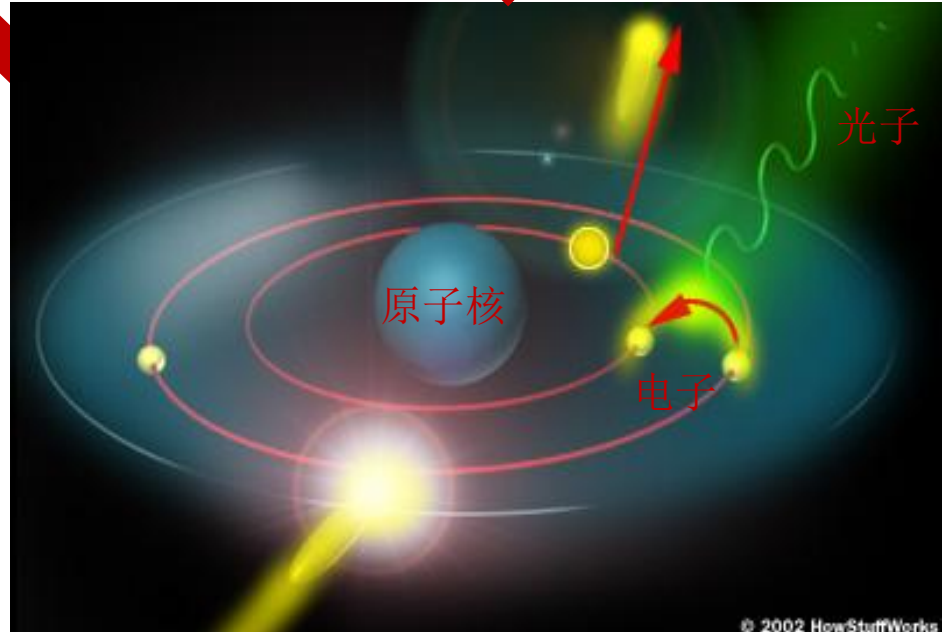
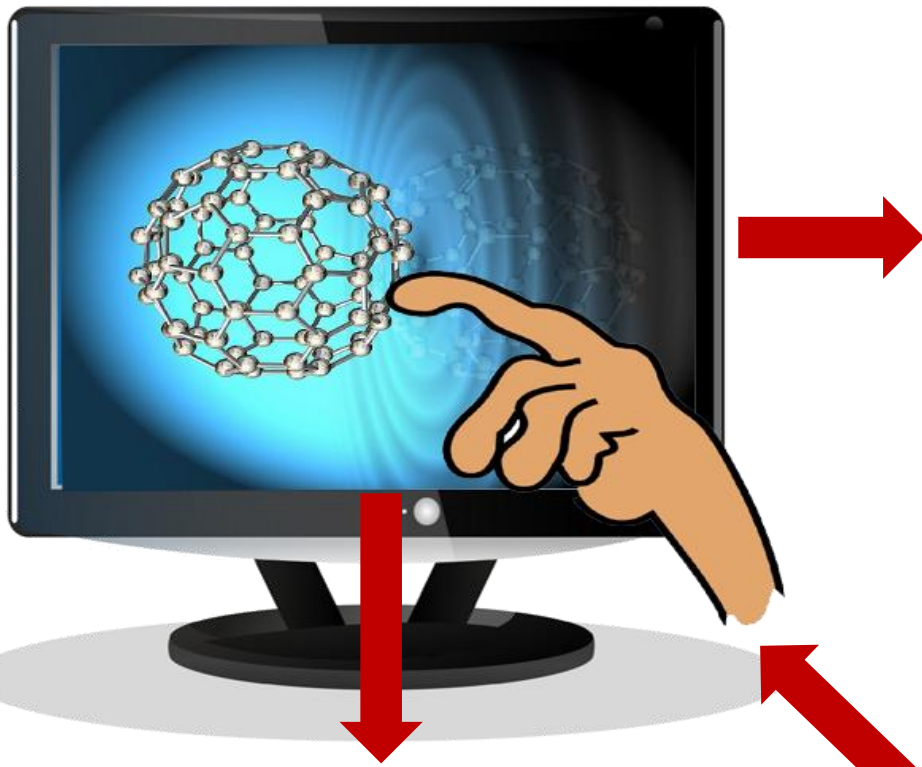
中国制造2025  
新材料产业发展指南  
七大战略新兴产业  
2015年中国新材料产  
值2万亿（工信部数据）



# 计算化学：化学与多个学科的交叉创新





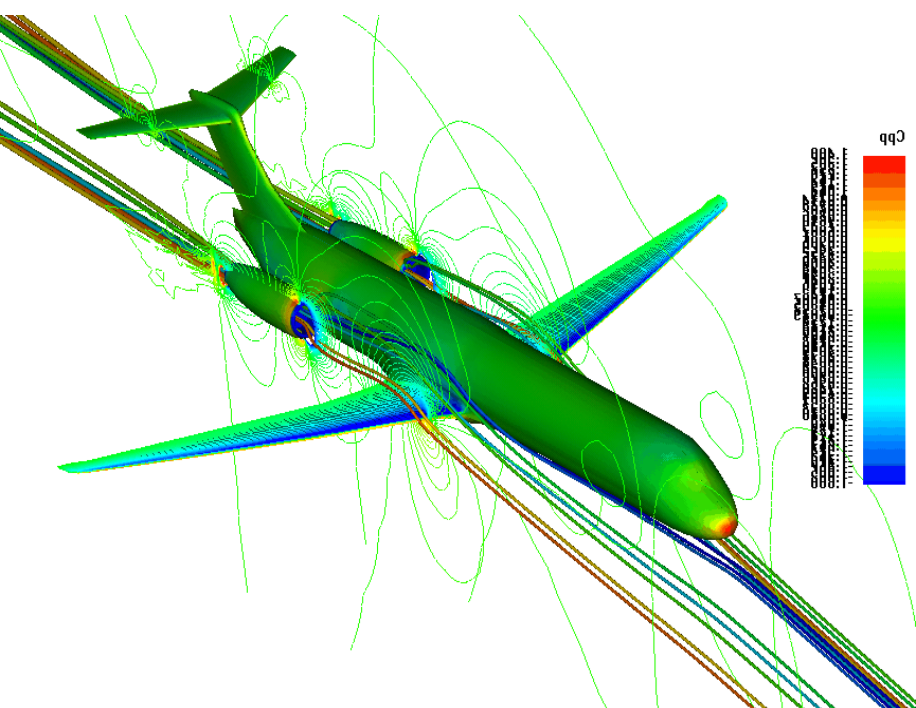


# 航空材料设计

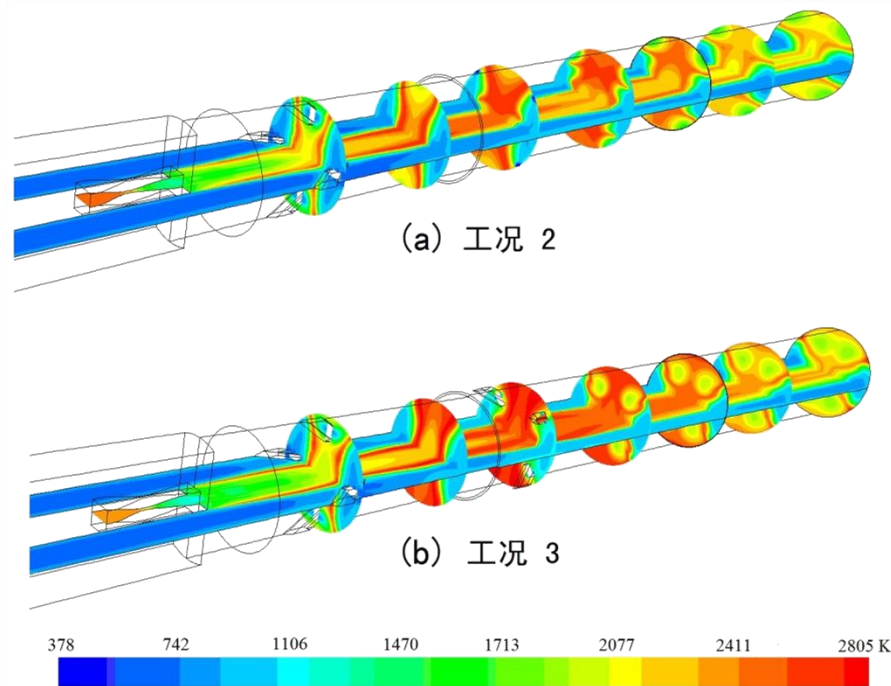
## 组合发动机燃烧大涡流场模拟

极端条件： 3000K, 10GPa, 70,000原子(Al, Fe, Cu), >10微米

西北工业大学/西安飞机工业集团

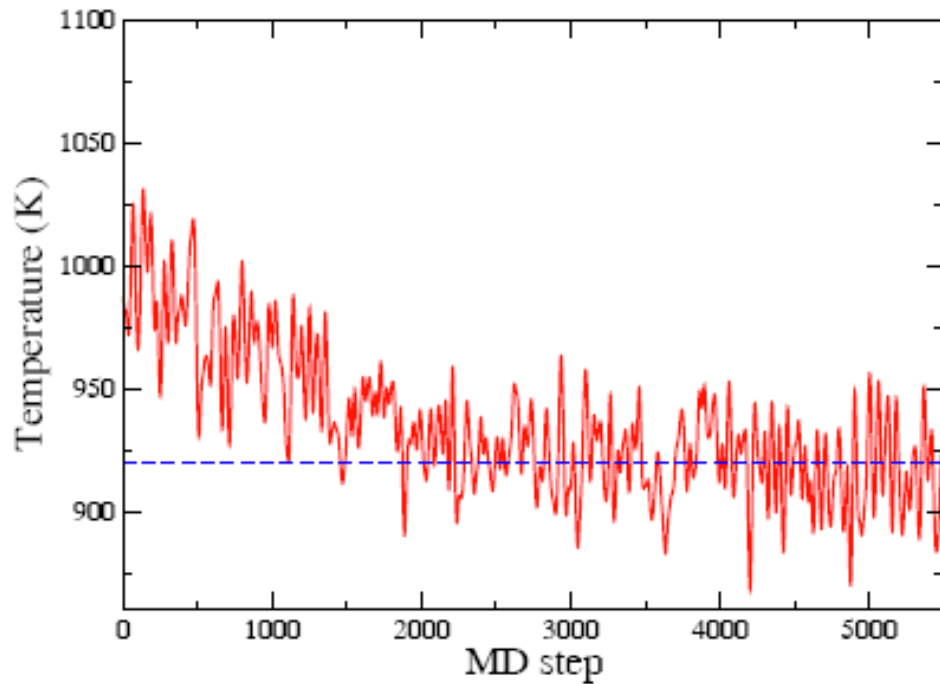
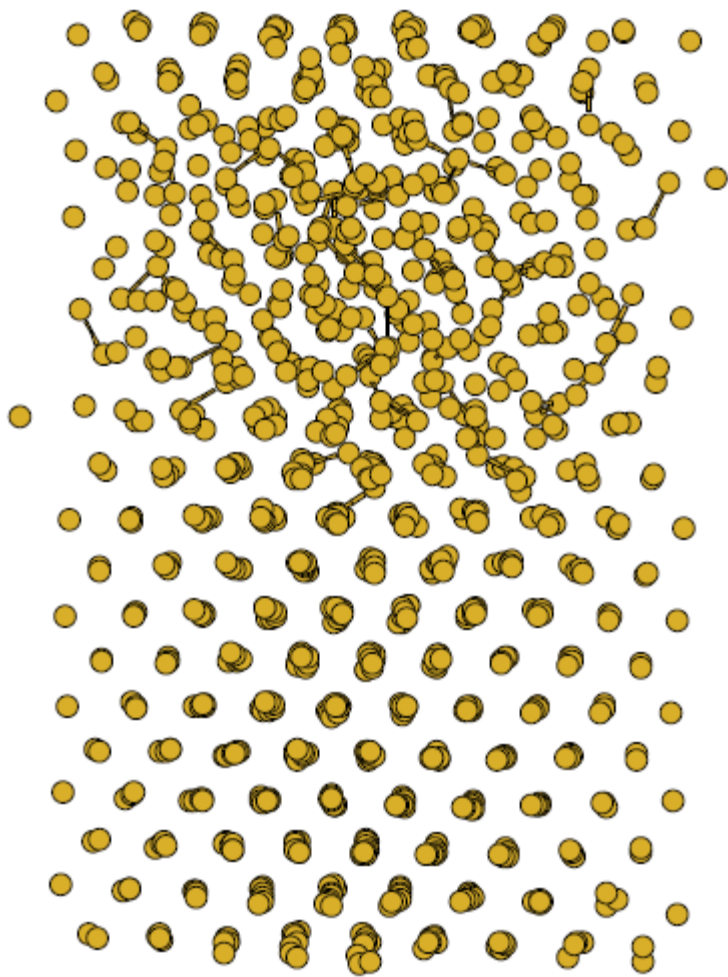


飞机降落过程中的流场  
模拟： 涡轮燃烧模拟



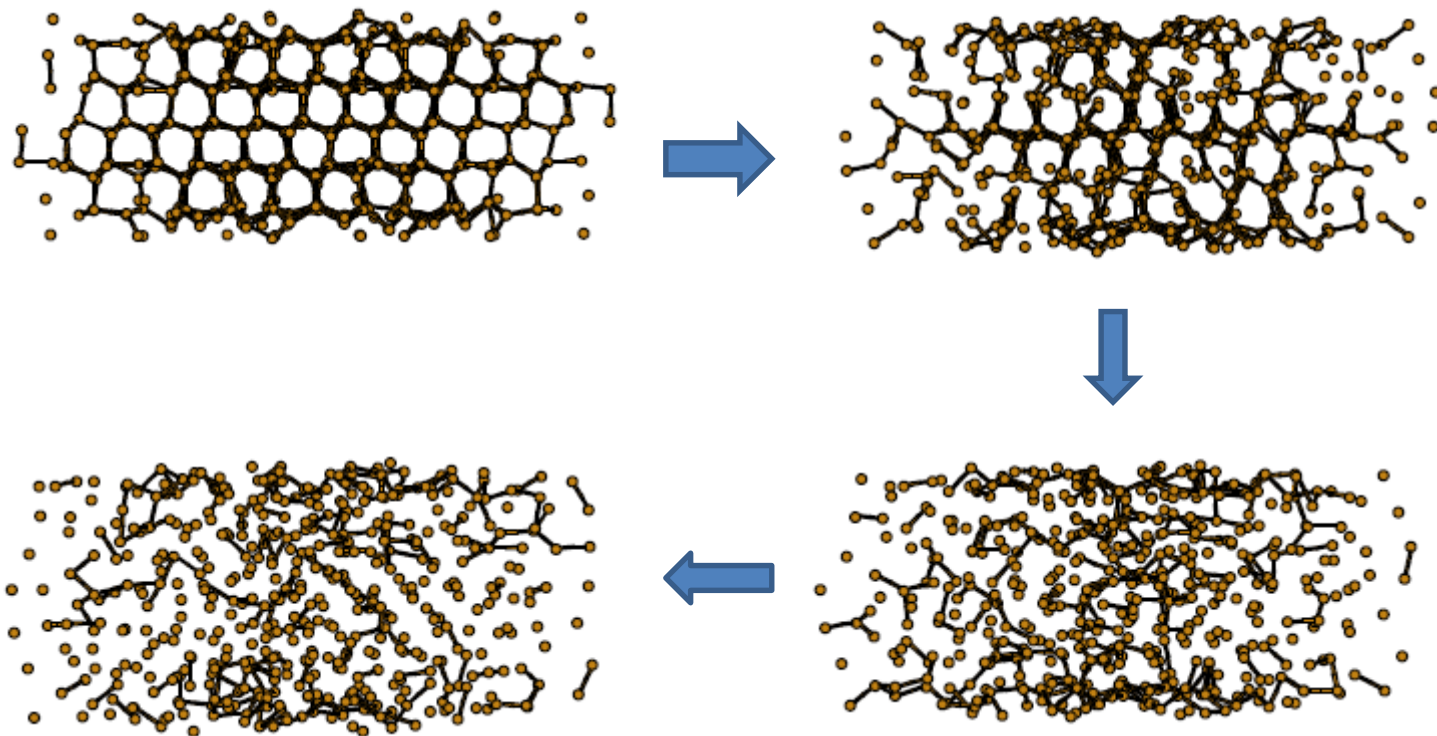
超燃模式下的小支流喷  
注温度分布

# 结构变化过程



AI的微结构及融化过程

# 结构变化过程





# 点石成金



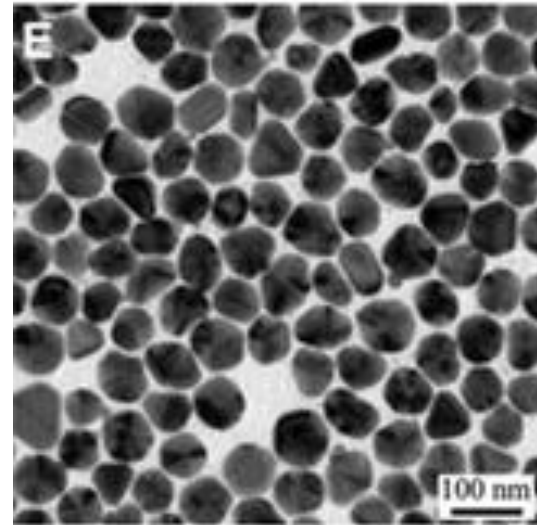
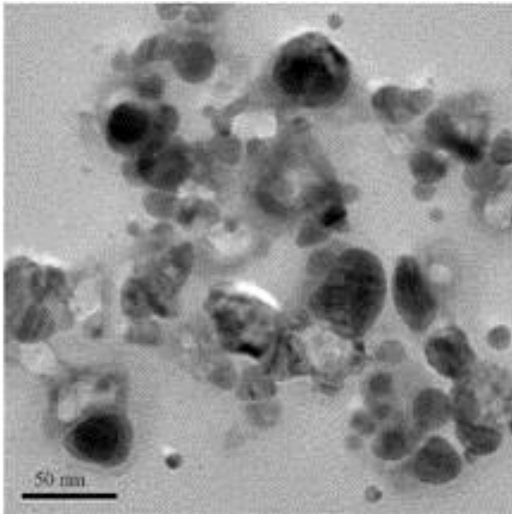
应用于西安飞机工业集团的航天部件设计

2012年完成课题12项，节约实验开发成本4320万

**理论模拟成本不到2万**

# 中国振华电子集团有限公司

- 成功研制基于纳米银的层状电阻
- 帮助其打破西方技术封锁
- 14年以来新增营业额2600万，税利达到680万



# 计算化学

## ——光电材料

# 光电材料设计

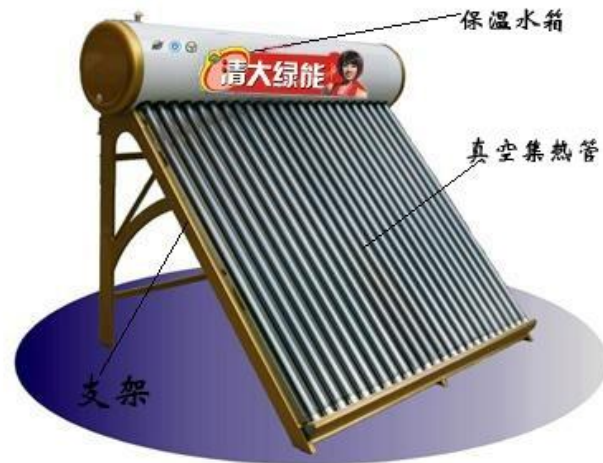
## 光的发射

照明、发光器件、荧光标记



## 光的吸收

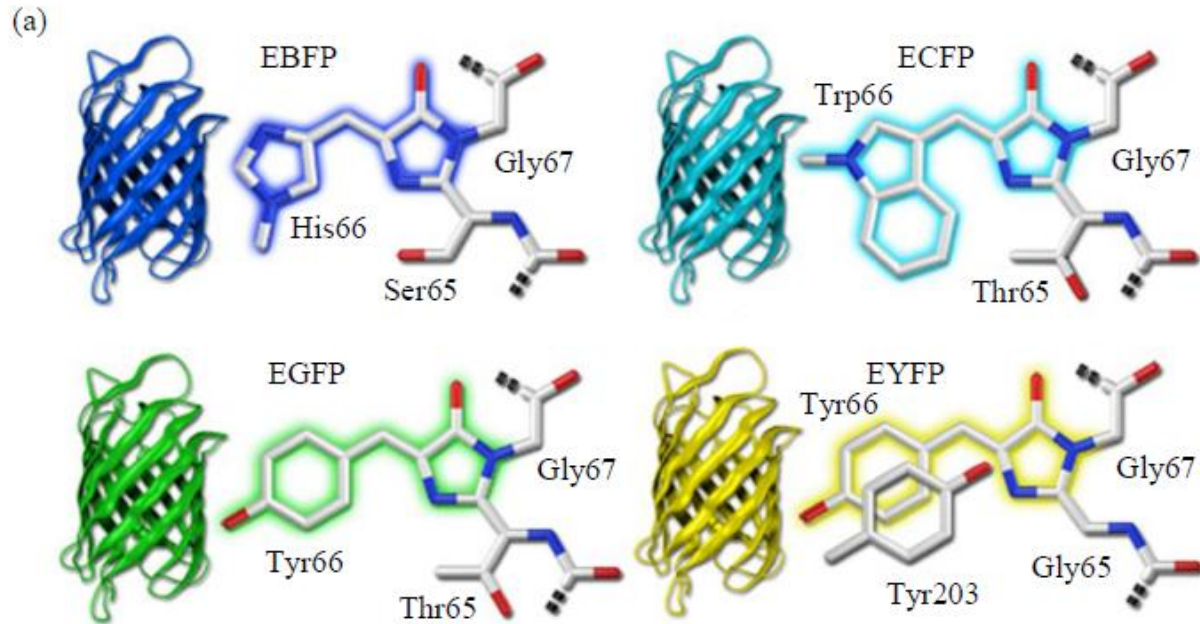
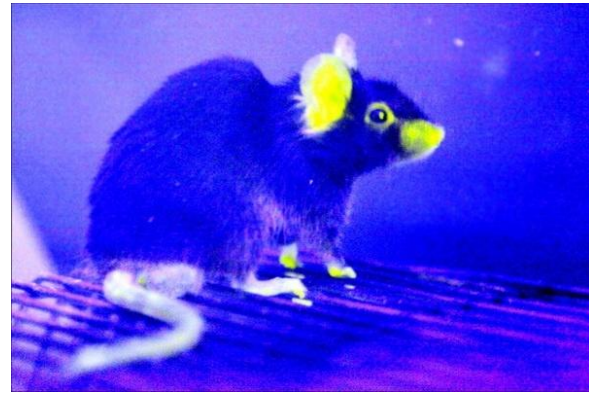
太阳能利用、污染治理



太阳能电动车 (09-B型)

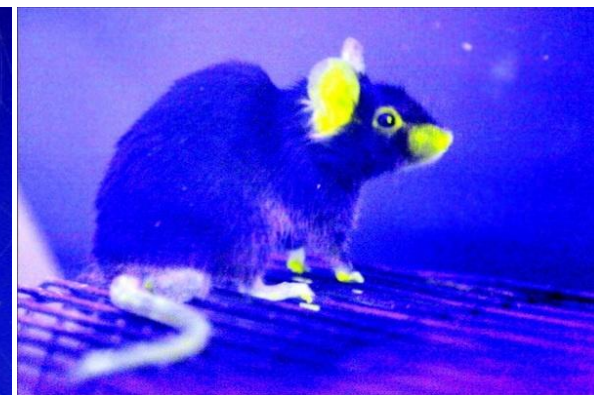
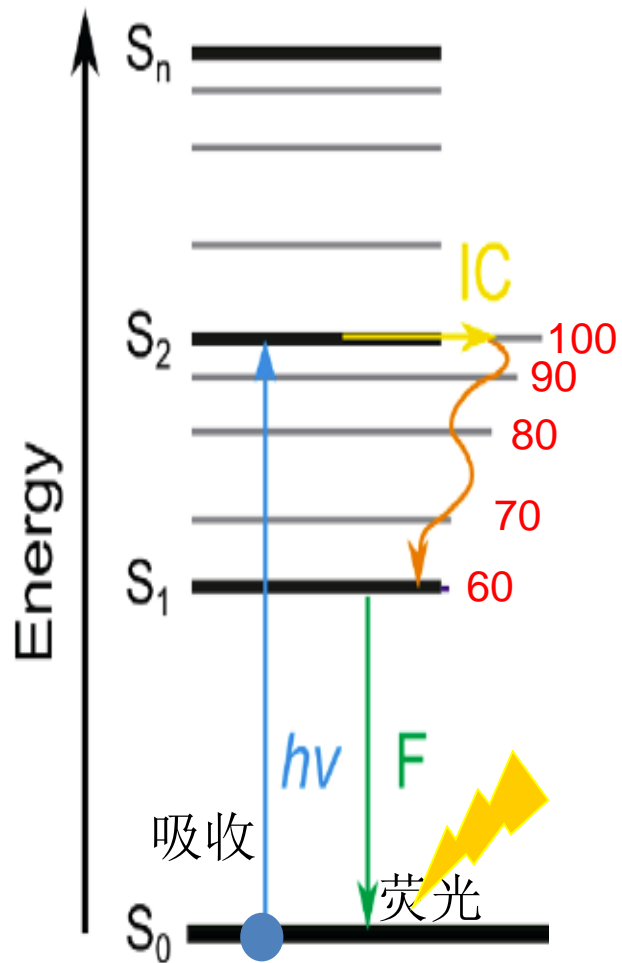


# 道法自然

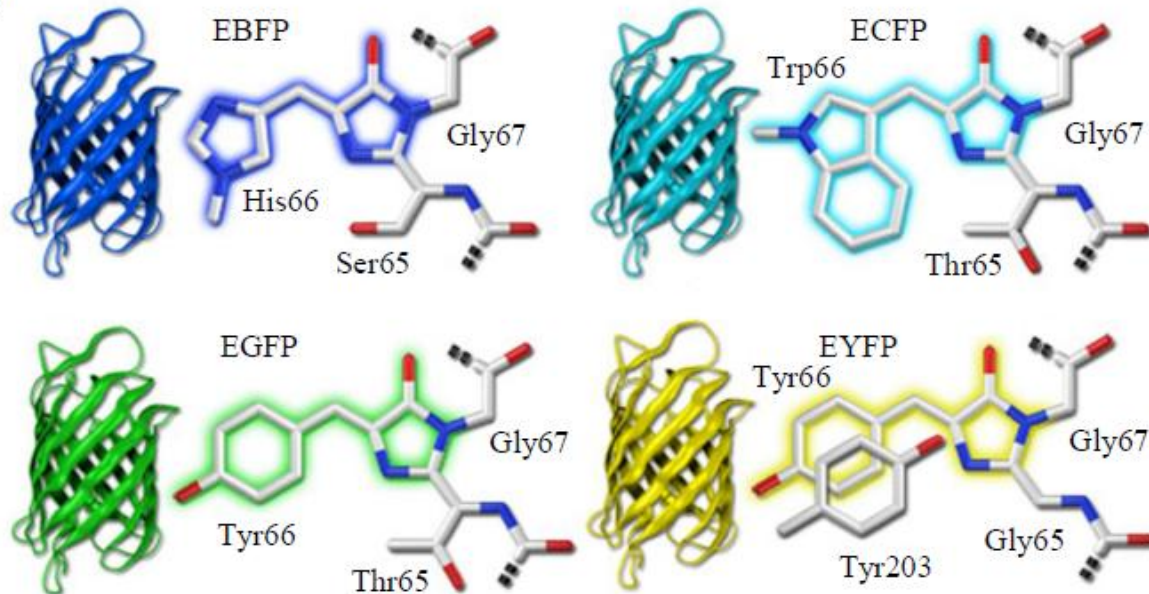


# 分子发光

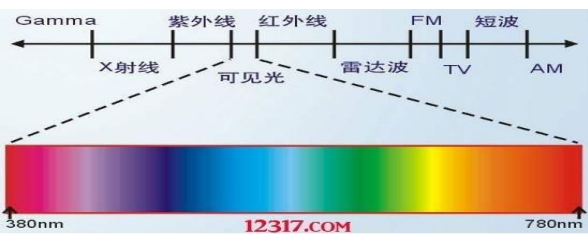
# 分子发光



(a)



基因（突变）改造—蛋白质表达—  
分子结构变化—荧光变化





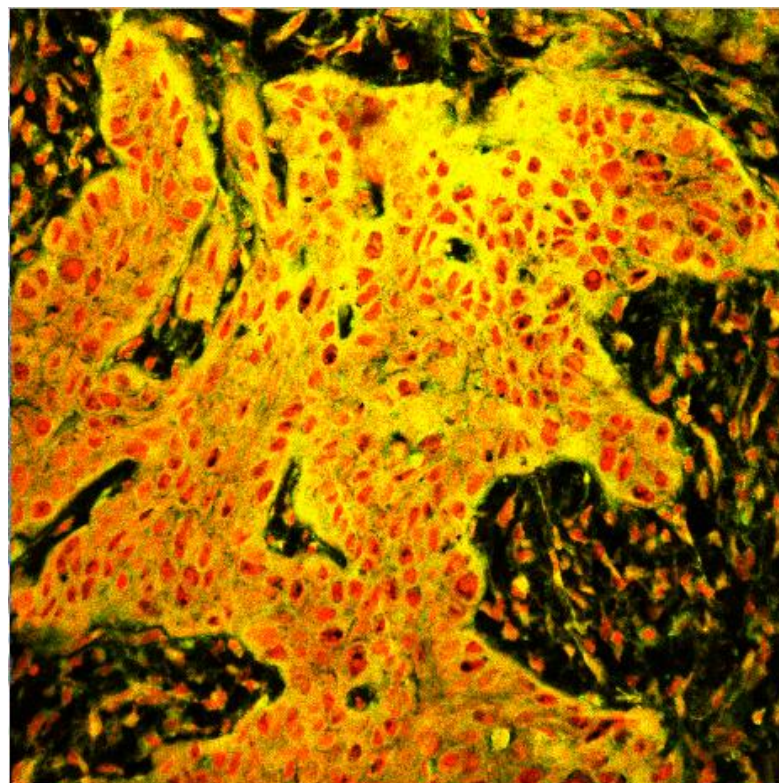
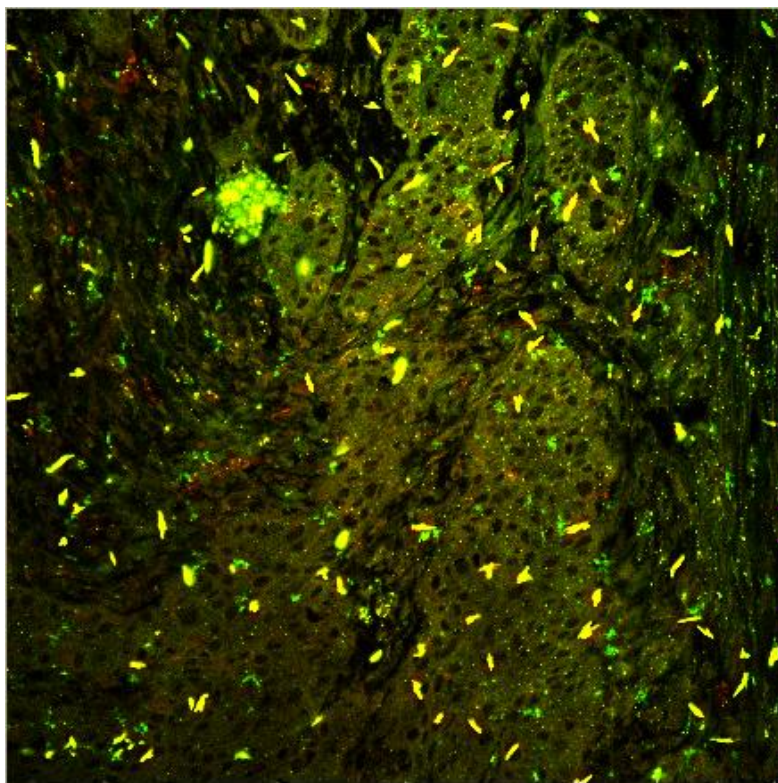
# 癌组织成像：食道

正常组织整体偏绿色  
5秒钟染色时间！

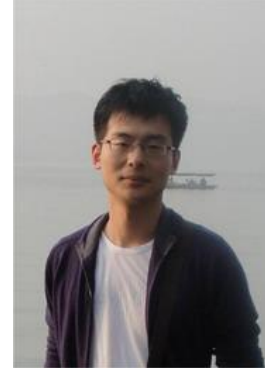
正常组织

癌组织

张国庆



千里马常有，而伯乐不常有





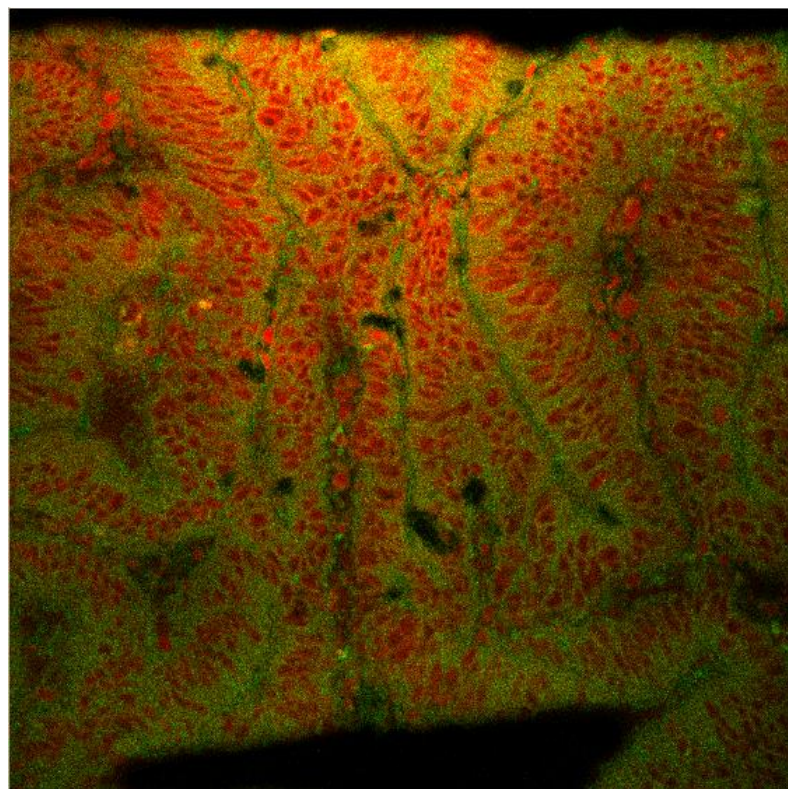
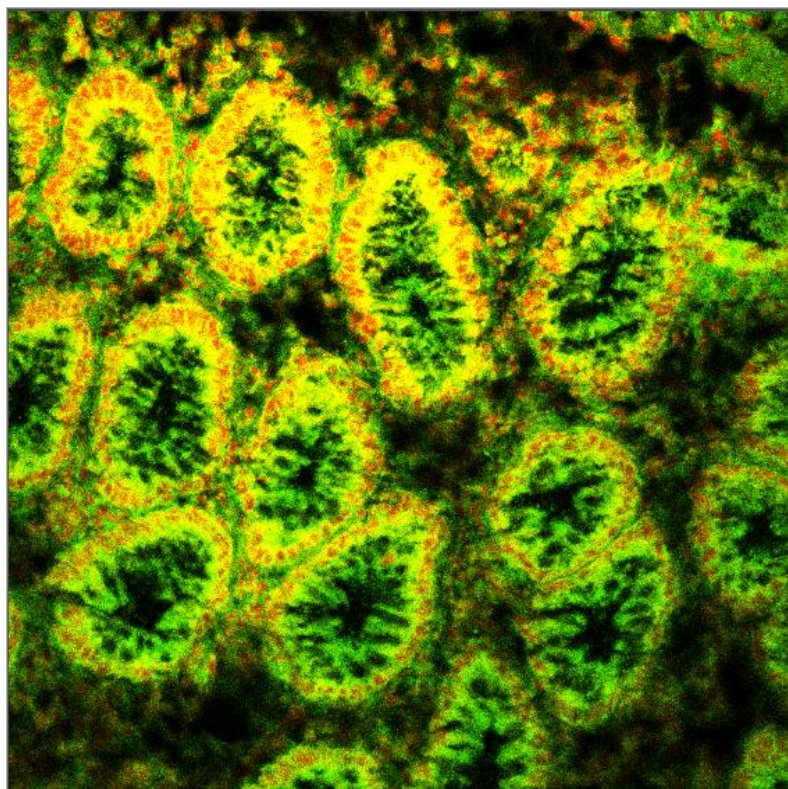
# 癌组织成像：肠

正常组织整体偏绿色  
5秒钟染色时间！

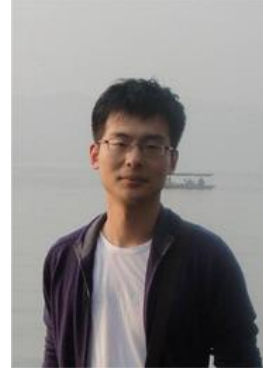
正常组织

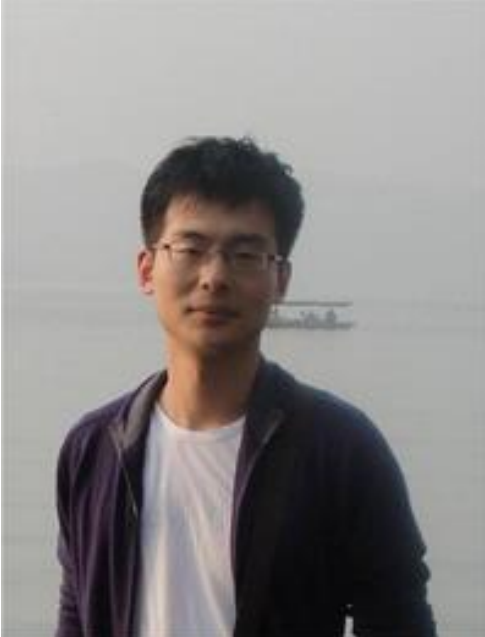
癌组织

张国庆



千里马常有，而伯乐不常有





张国庆



李军配







张国庆

科学与艺术、自然与社会

# 人见人爱的化学



科学与艺术、自然与社会

# 计算化学

## ——太阳能光催化



# 人类面临能源与环境问题

## □ 能源缺乏



## □ 环境问题形势严峻



二氧化碳排放  
将在2030达到峰值





**太阳能**

**取之不尽的能量来源**



**最伟大的太阳能反应**

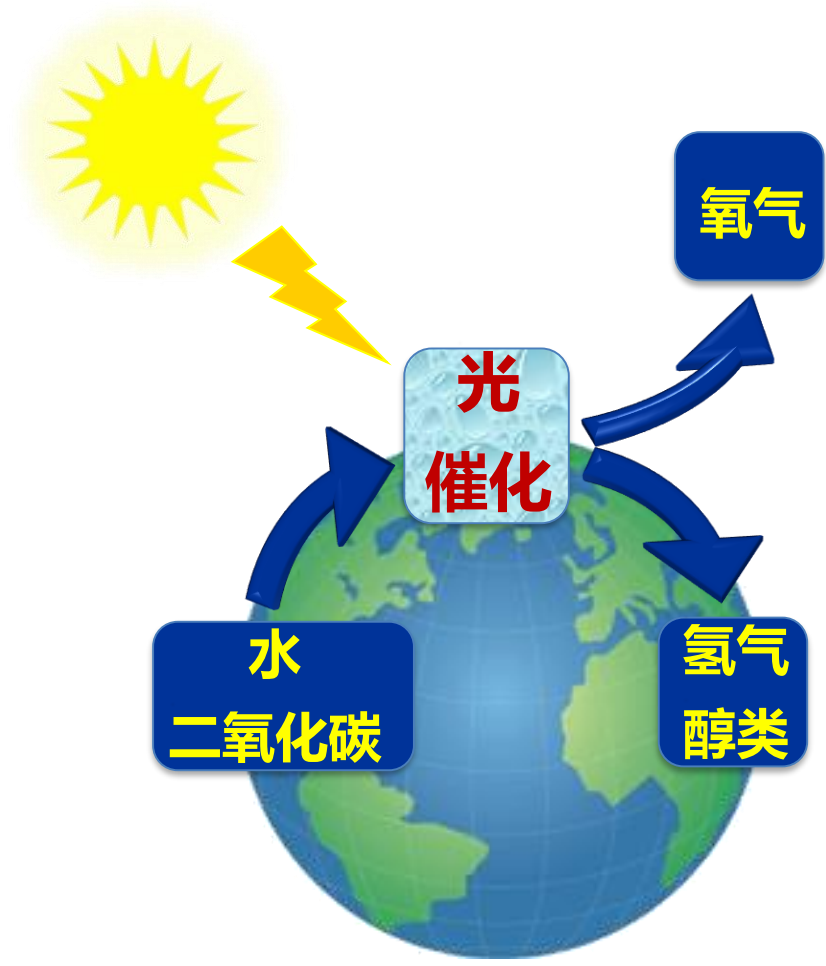
**光合作用**

**氧气  
糖类**

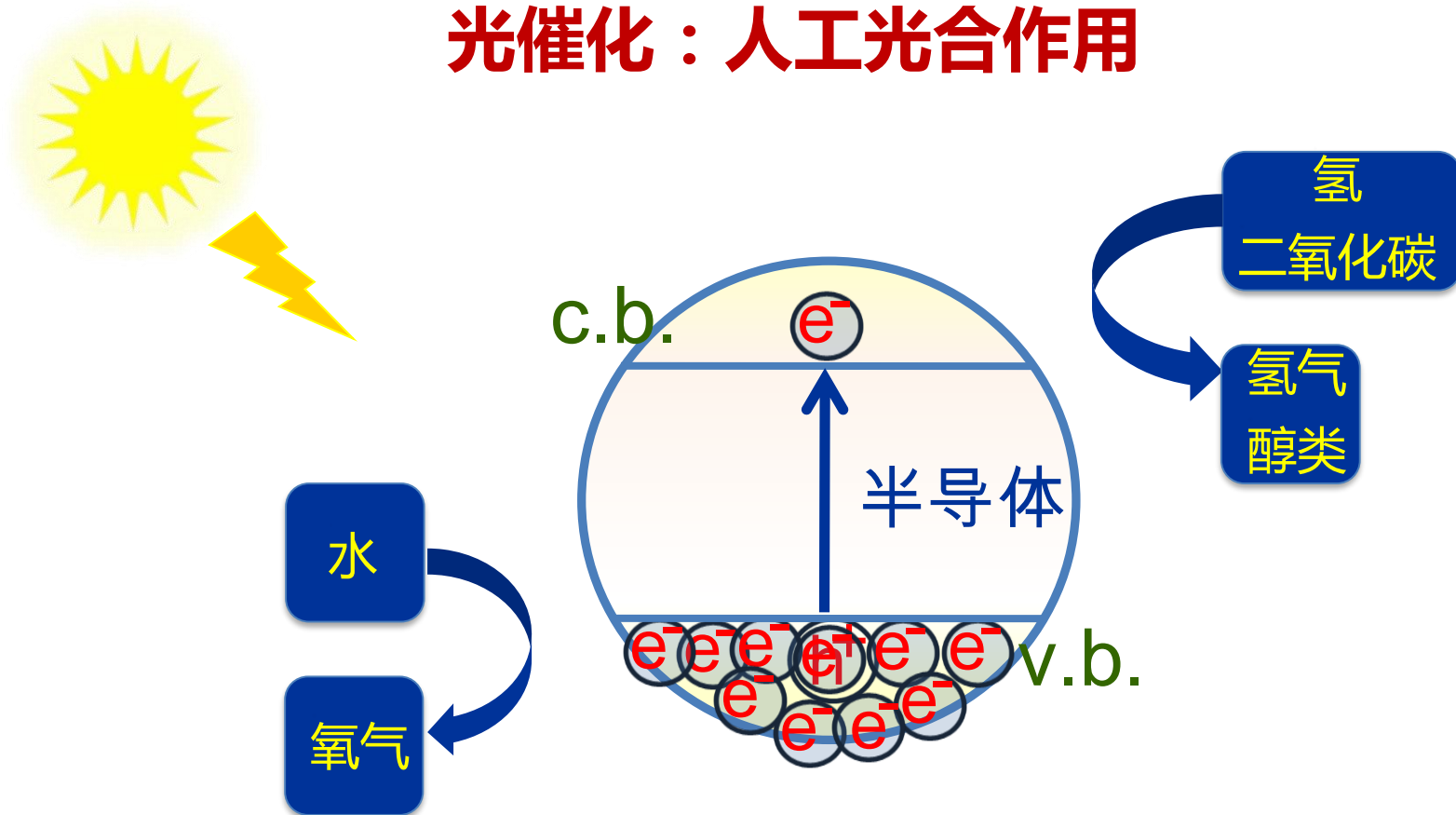


**丰富的物质**

**人工太阳能利用：光-化学能转化**



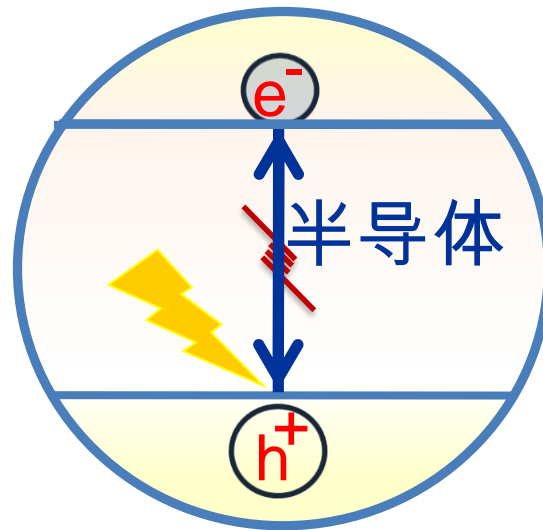
# 光催化：人工光合作用

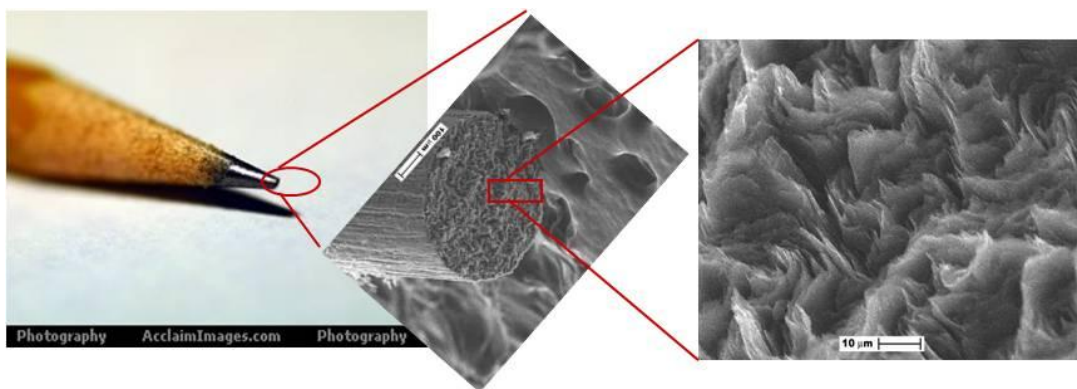
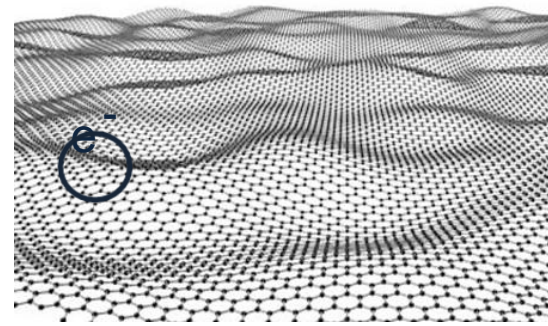
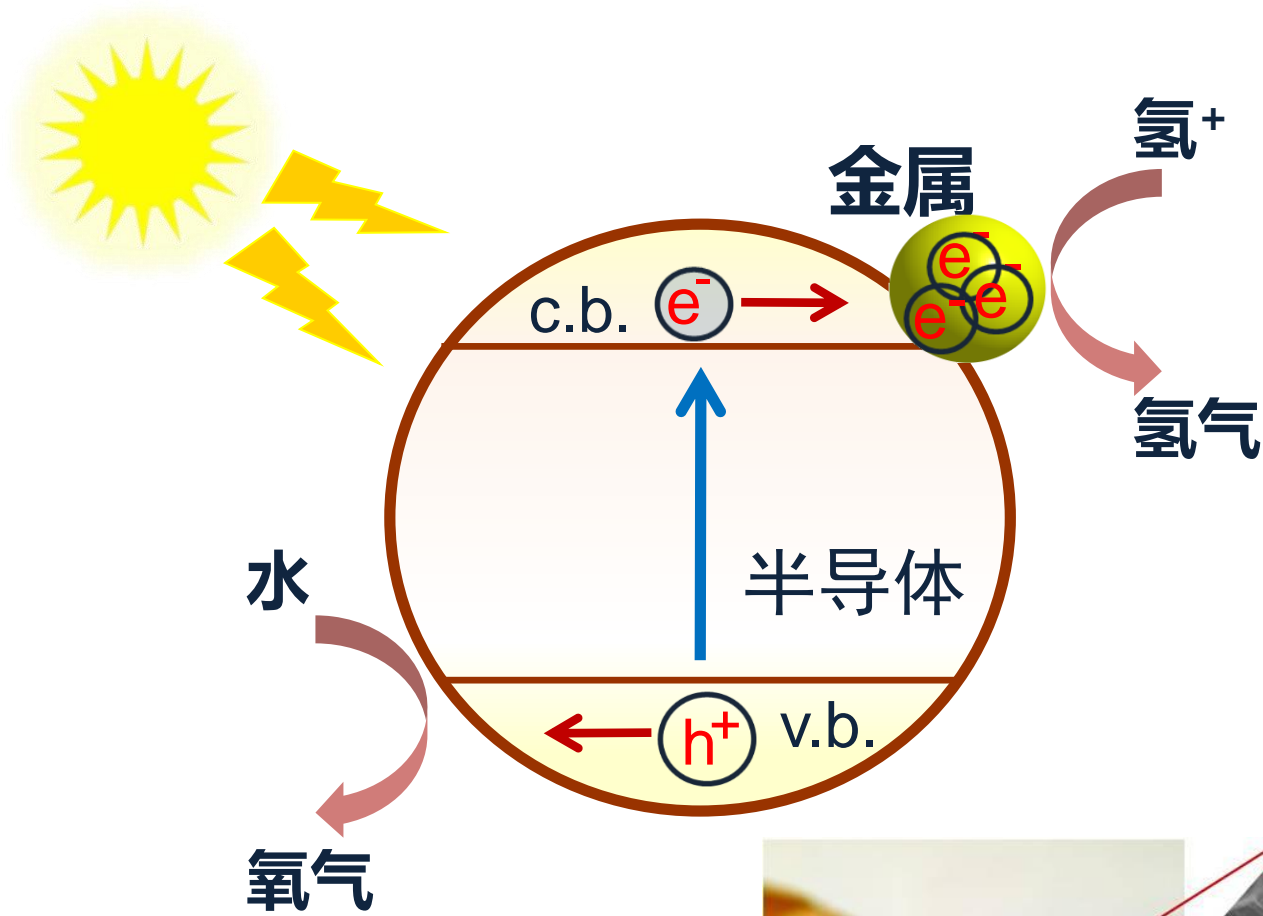


性能瓶颈

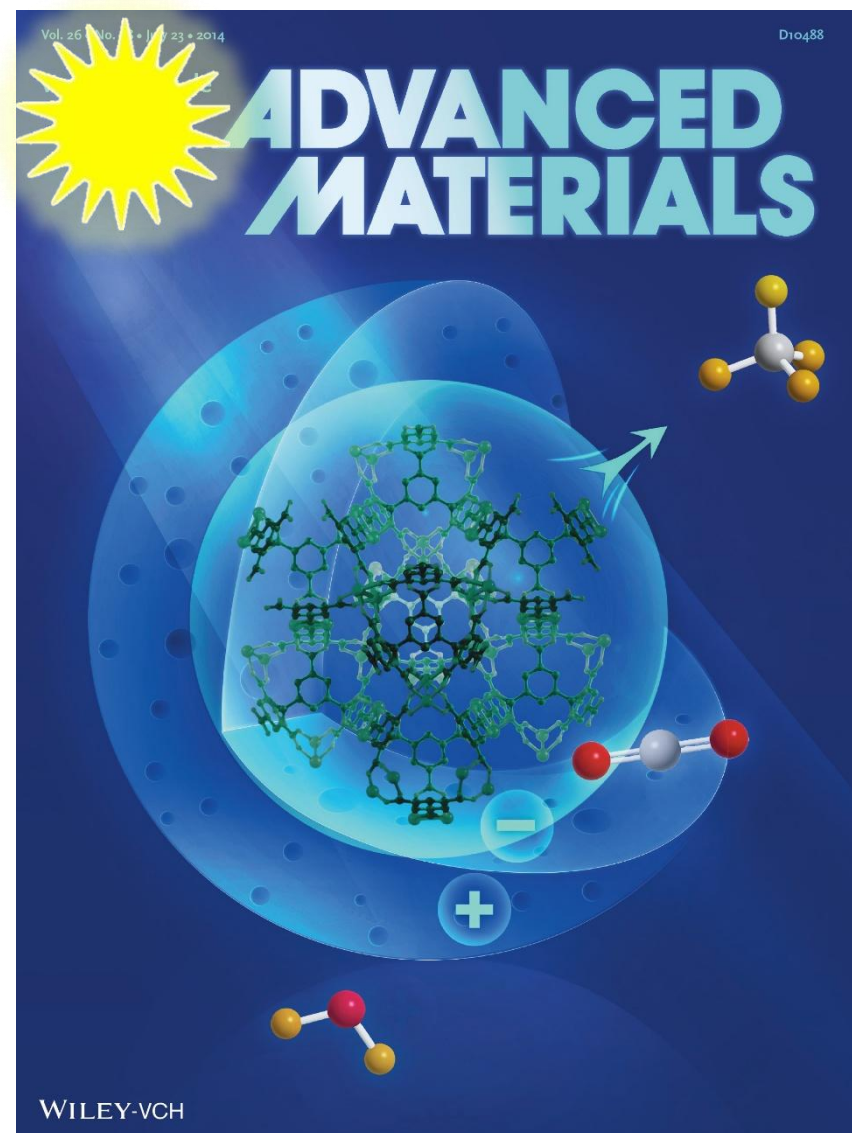
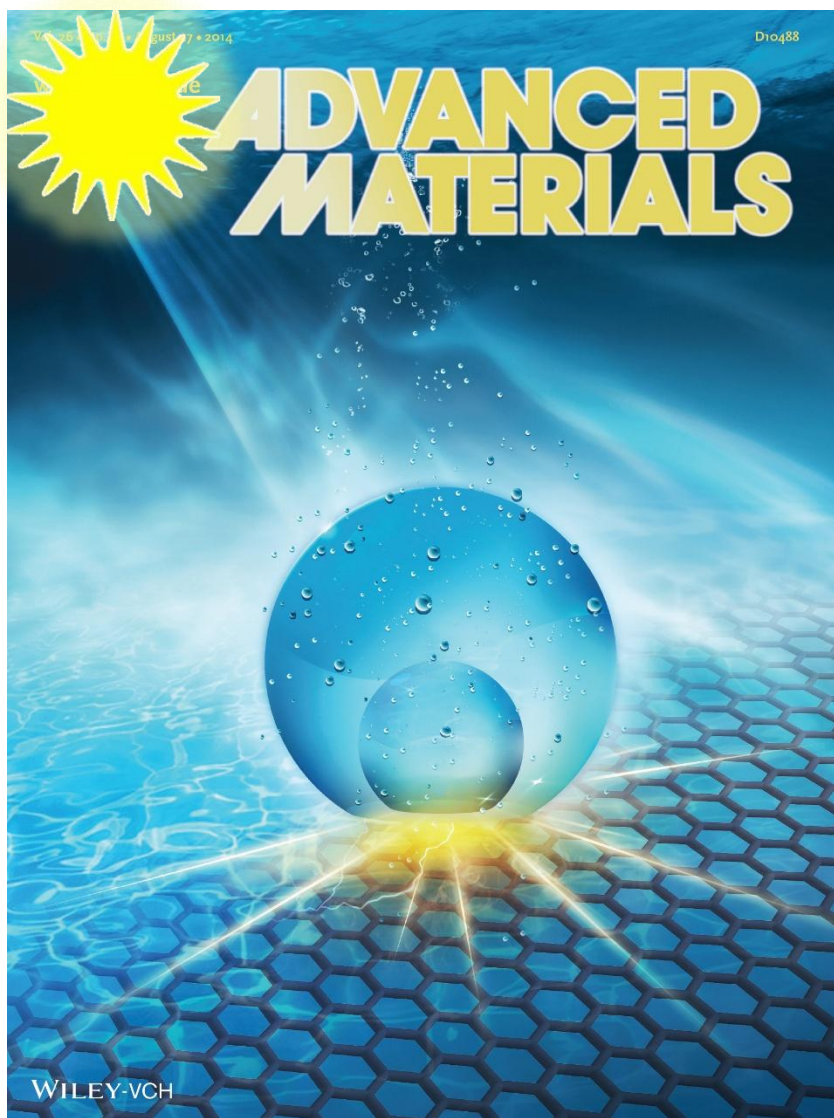


计算机设计





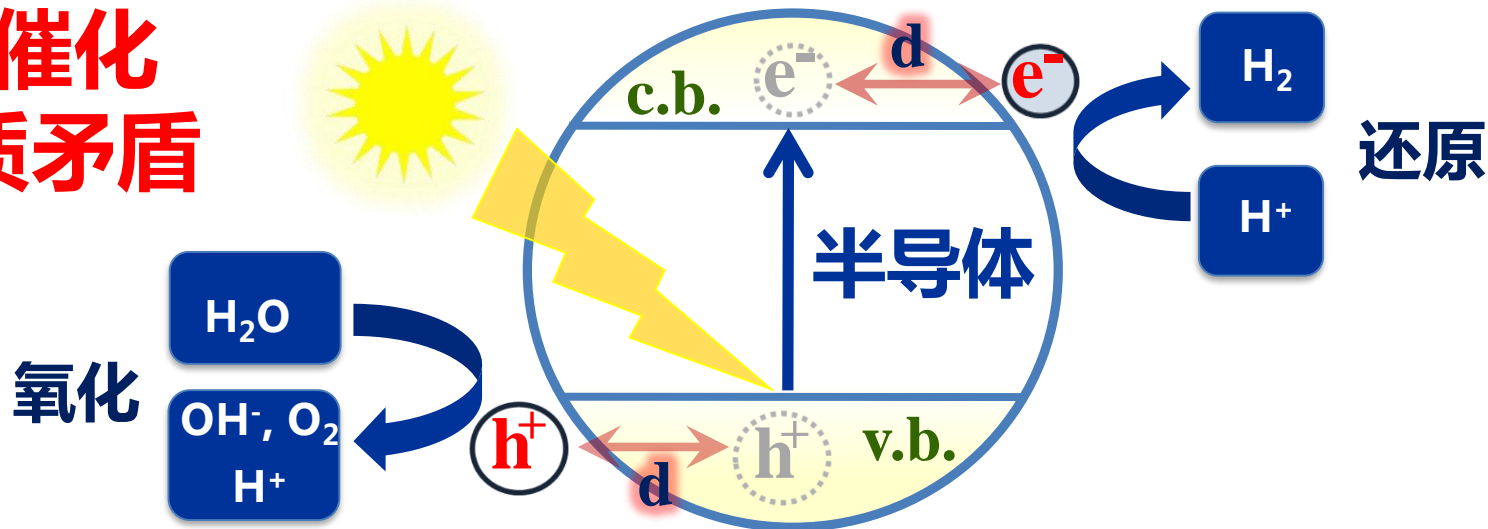




国家科技部973项目  
经费500万元

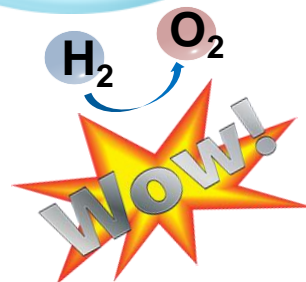


# 光催化 本质矛盾



$d <$  电子平均自由程  
(纳米尺度)

定域的光激发限制了  
离域的电荷转移



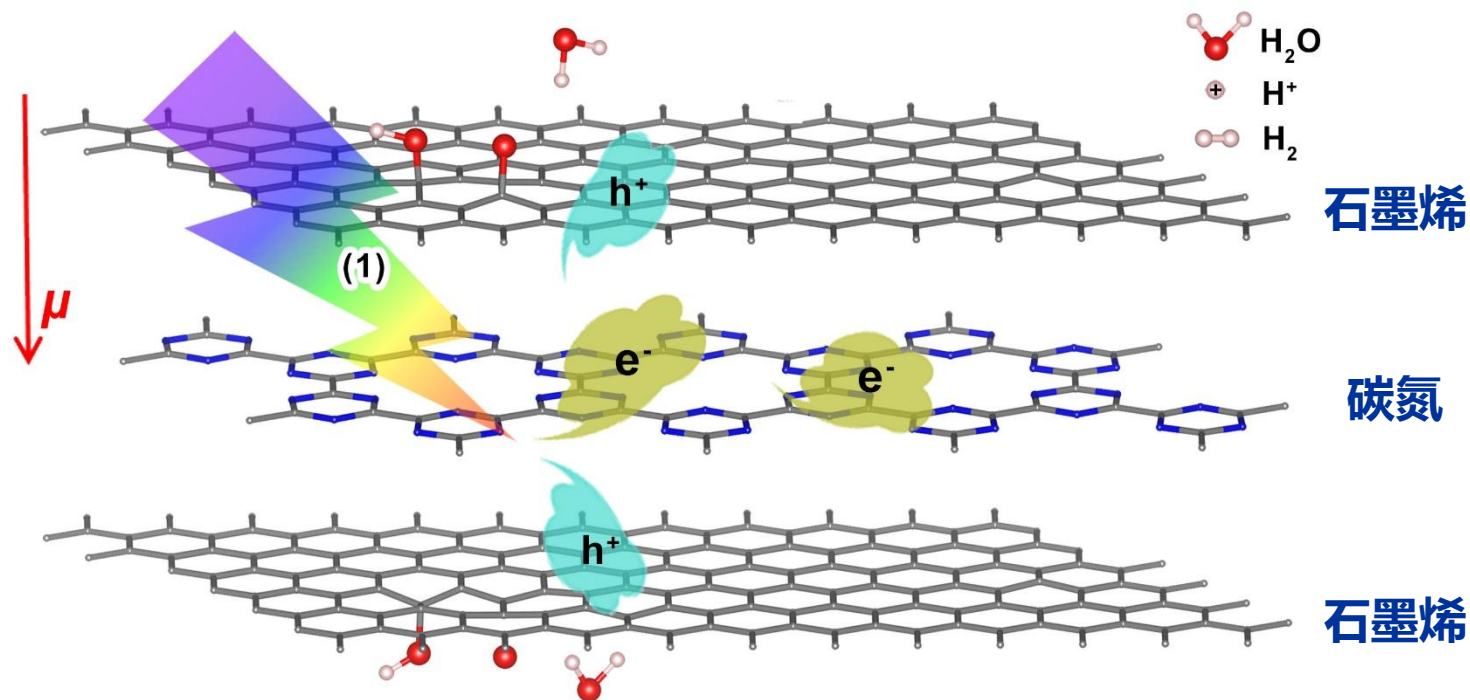
光生电子空穴的分离要求  
非常接近的  
氧化还原位点

**.vs.**

有效抑制逆反应则要求  
空间分离的  
氧化还原位点

# 电荷极化：驱动质子（正电荷电子态）运动

## 二维材料异质结相互作用诱导电荷极化

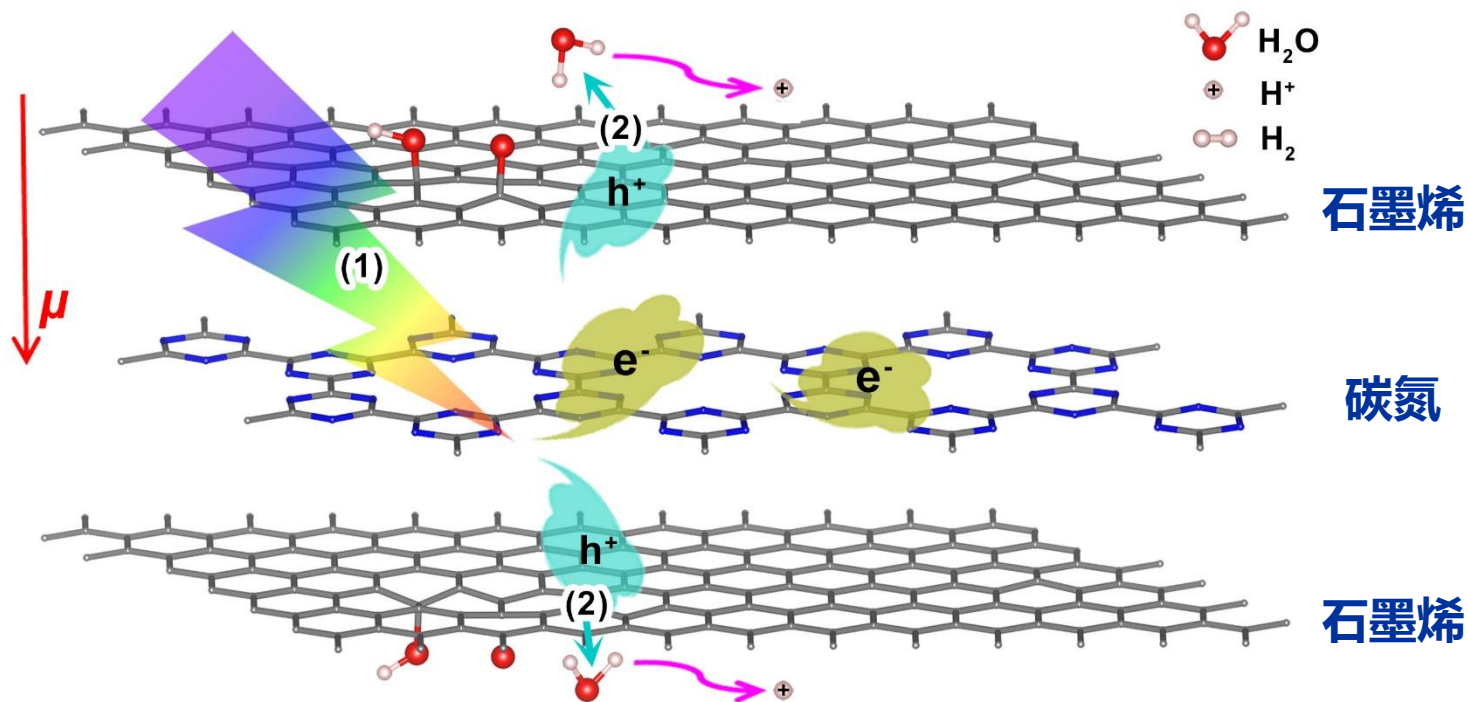


### Graphene-C<sub>x</sub>N<sub>y</sub>-Graphene 三明治结构

L. Yang, X. Y. Li, G. Z. Zhang, *et.al.*, Y. Luo, J. Jiang\*, *Nat. Commun.* 2017, 8, 16049

解决光催化本质矛盾：氧化/还原位点接近和分离

# 电荷极化：驱动质子（正电荷电子态）运动

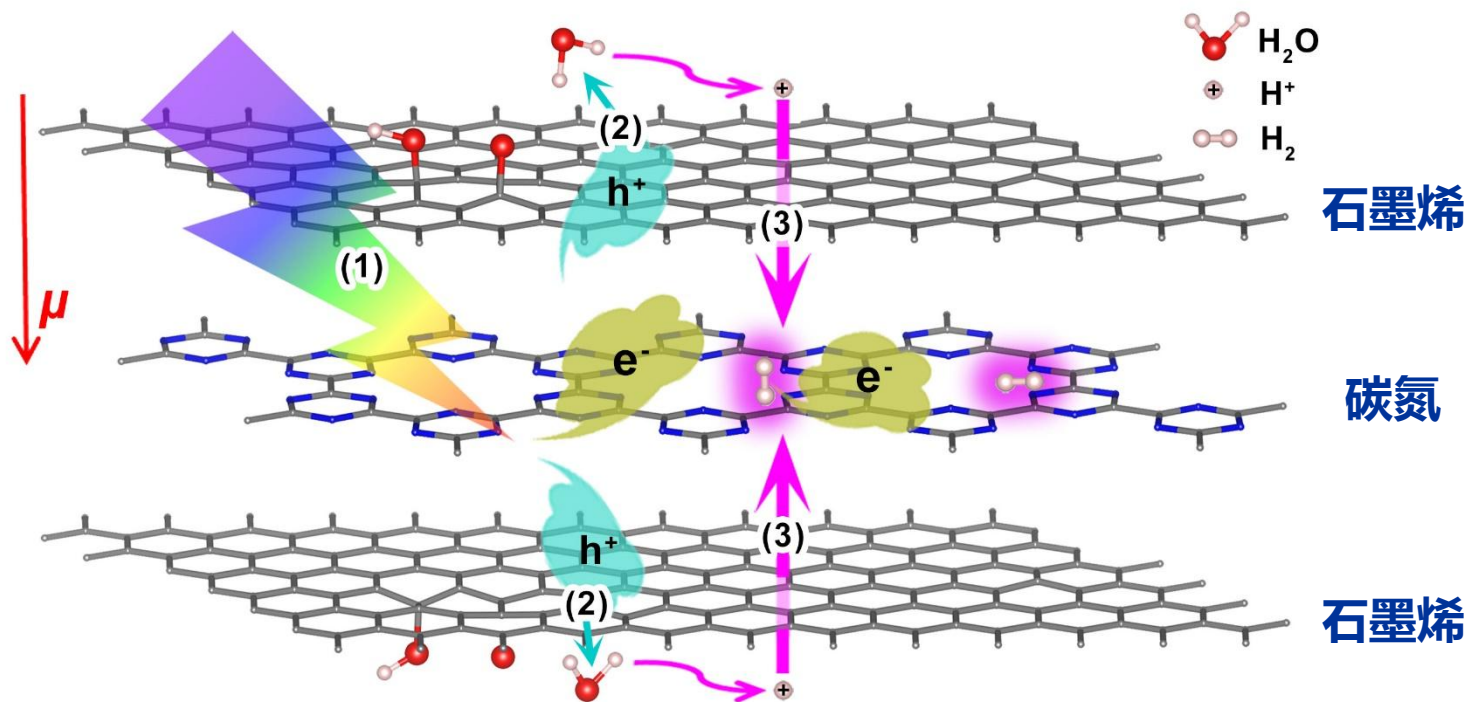


## Graphene-C<sub>x</sub>N<sub>y</sub>-Graphene 三明治结构

L. Yang, X. Y. Li, G. Z. Zhang, *et.al.*, Y. Luo, J. Jiang\*, *Nat. Commun.* 2017, 8, 16049

解决光催化本质矛盾：氧化/还原位点接近和分离

# 电荷极化：驱动质子（正电荷电子态）运动



## Graphene-C<sub>x</sub>N<sub>y</sub>-Graphene 三明治结构

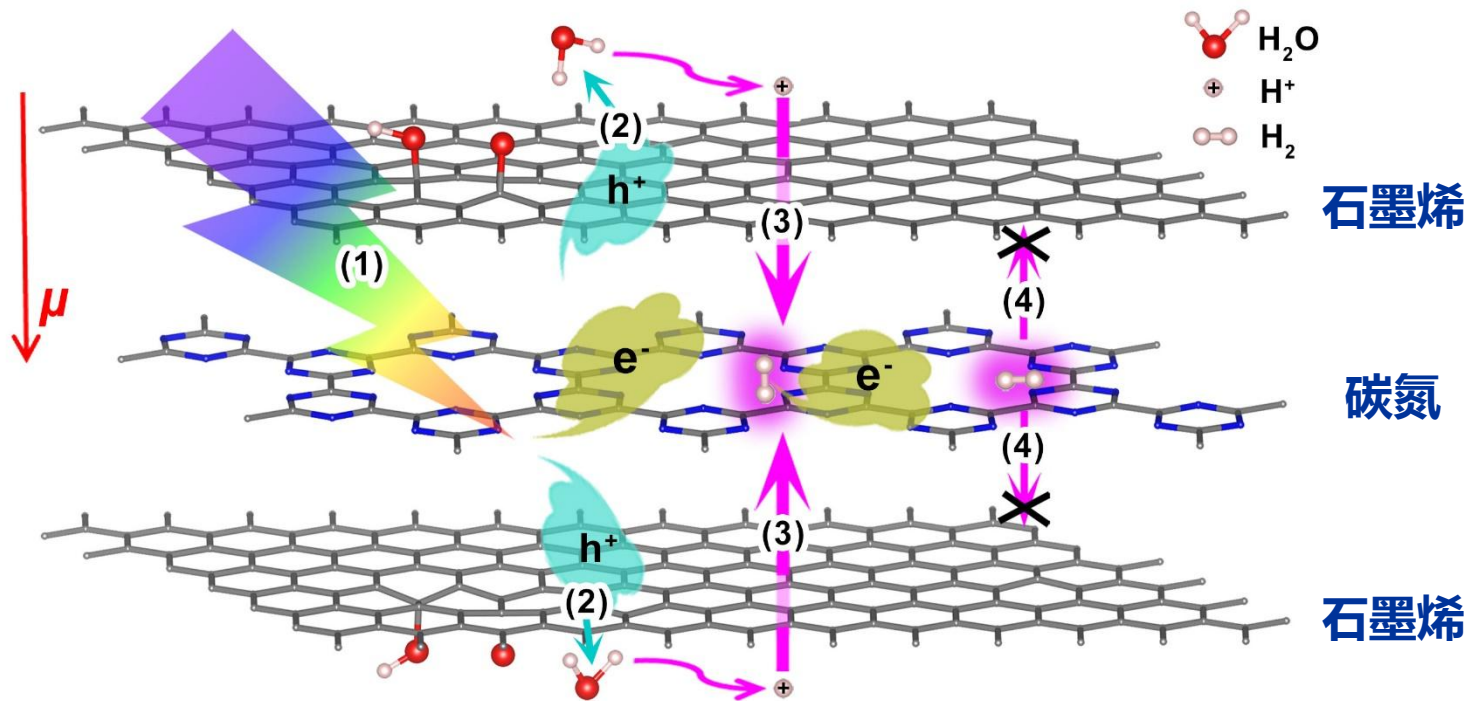
L. Yang, X. Y. Li, G. Z. Zhang, *et.al.*, Y. Luo, J. Jiang\*, *Nat. Commun.* 2017, 8, 16049

解决光催化本质矛盾：氧化/还原位点接近和分离

# 电荷极化：驱动质子（正电荷电子态）运动

协调电子和质子运动，抑制逆反应

在氧化/还原位点非常接近的前提下完全分离反应产物



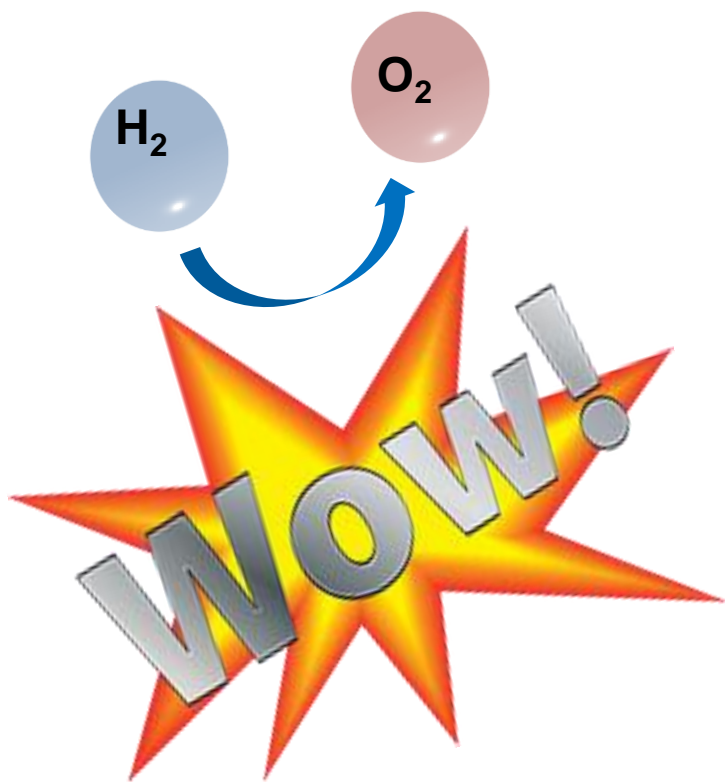
## Graphene-C<sub>x</sub>N<sub>y</sub>-Graphene 三明治结构

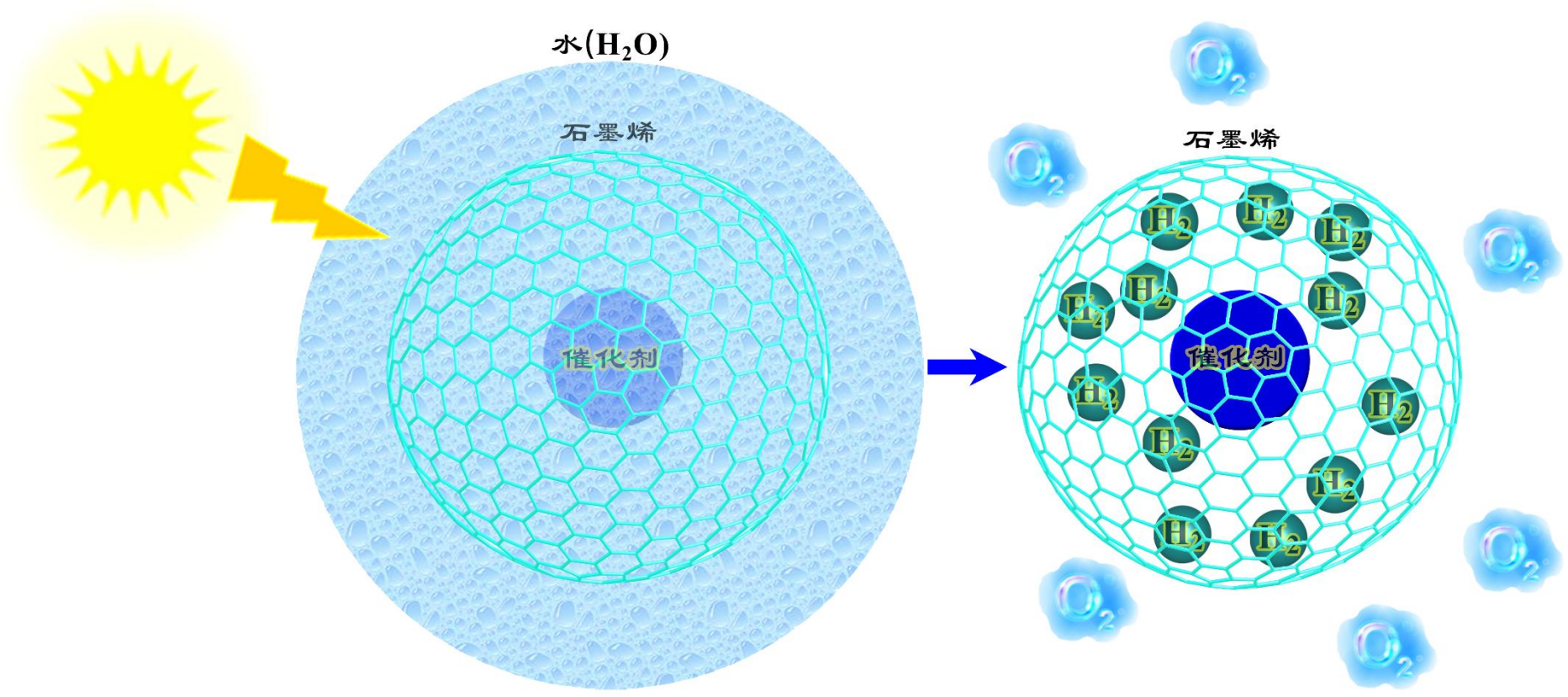
L. Yang, X. Y. Li, G. Z. Zhang, *et.al.*, Y. Luo, J. Jiang\*, *Nat. Commun.* 2017, 8, 16049

解决光催化本质矛盾：氧化/还原位点接近和分离



# 氢能利用的瓶颈：安全存储与运输



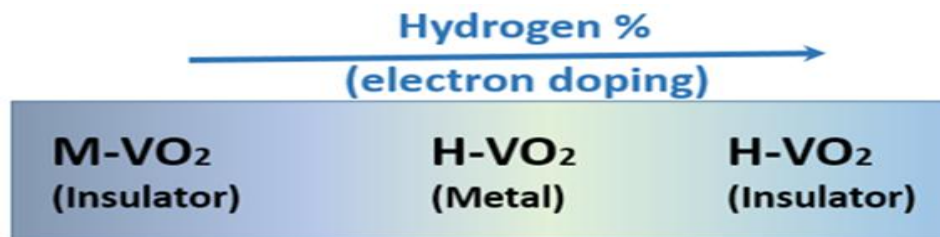


石墨烯三明治实现光解水制氢储氢一体化

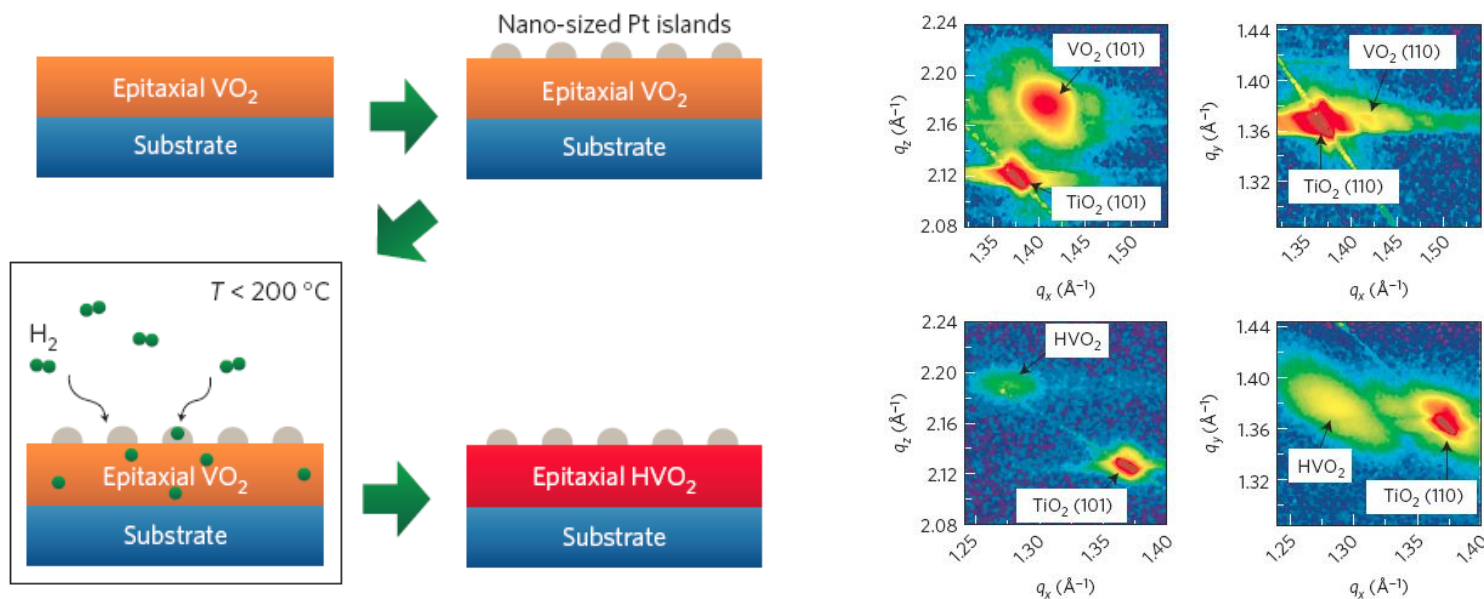




# ▶ 半导体材料加氢：智能窗、能源、雷达、通信



Hydrogen doping induce Metal-Insulator Transition (MIT):  
insulating monoclinic vanadium dioxide (M-VO<sub>2</sub>) to metallic H<sub>x</sub>VO<sub>2</sub>



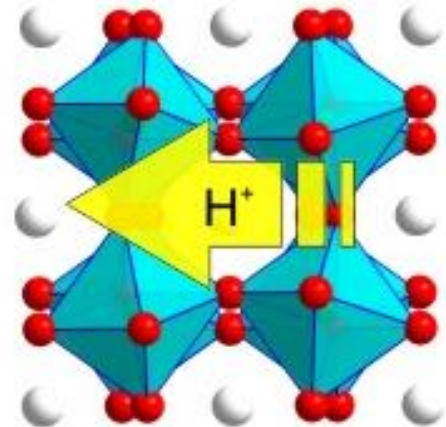
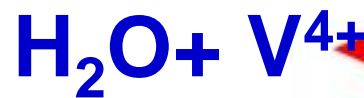
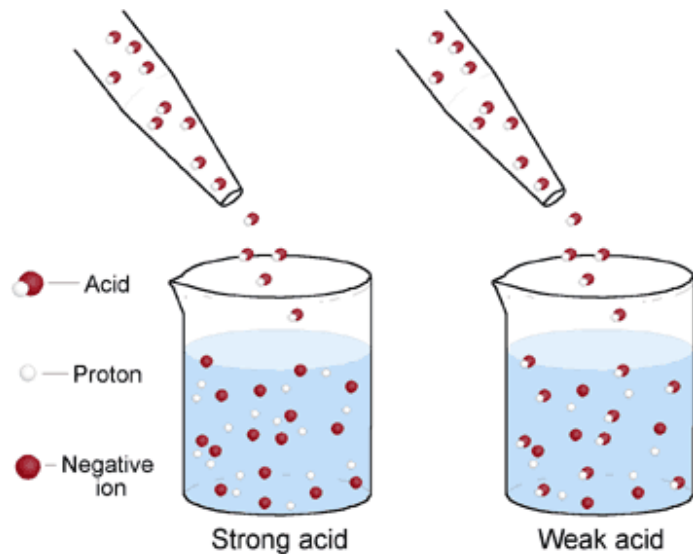
*Nature Mater.* 2016, 15, 113;  
*Nature Nanotech.* 2012, 7, 357;  
*Science.* 2011, 331, 746;

# Hydrogen penetration is never easy

1. Creating atomic H-source : Noble metal catalysis
2. Implantation : High energy driving force



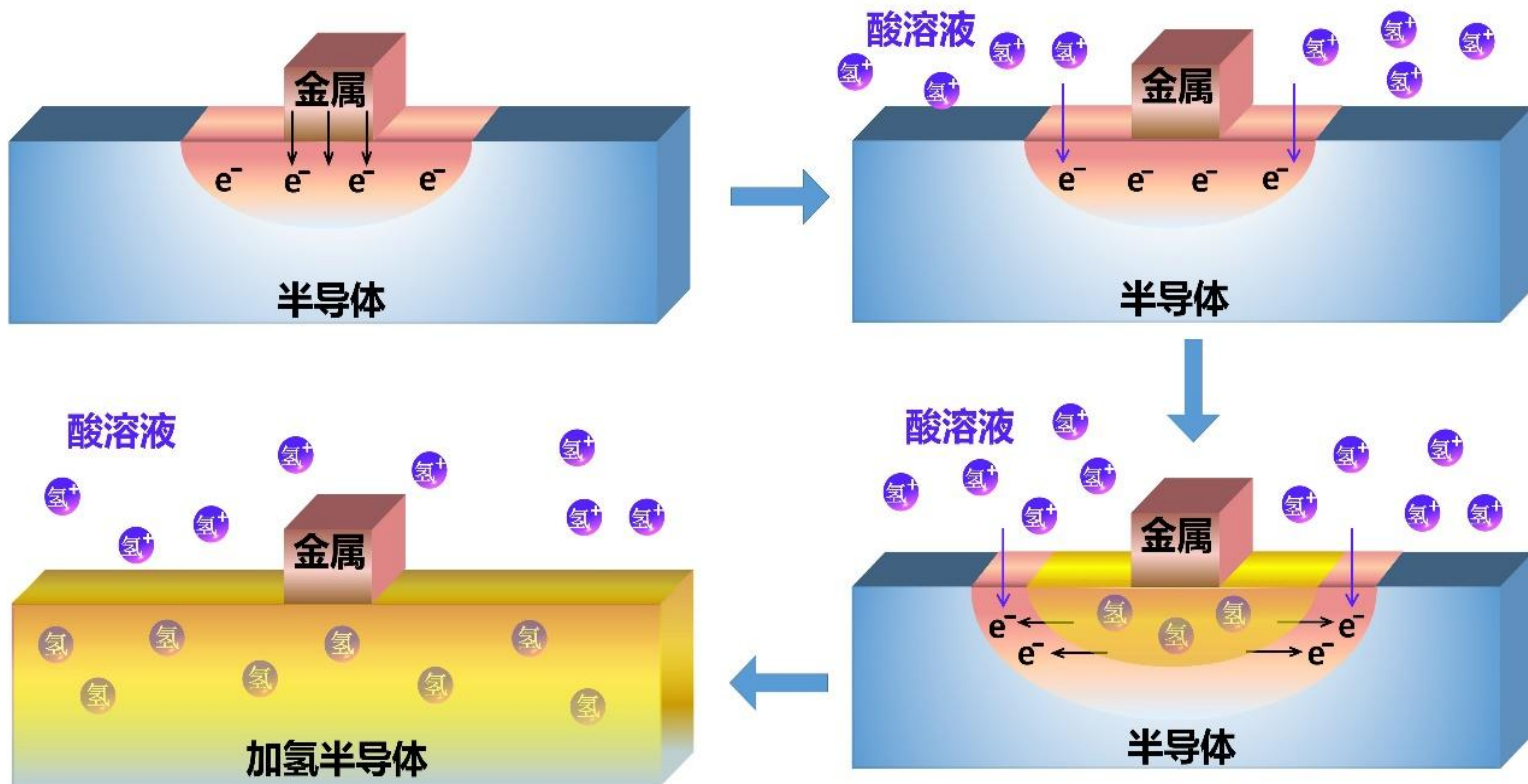
Convenient and cheap H-source ???  
Proton ( $H^+$ ) in acid solution



Positive charge of proton ( $H^+$ ) breaks lattice — Corrosion

# 极化电荷：驱动质子（正电荷）运动

调控电子运动和质子运动  
实现温和条件固体材料加氢改性

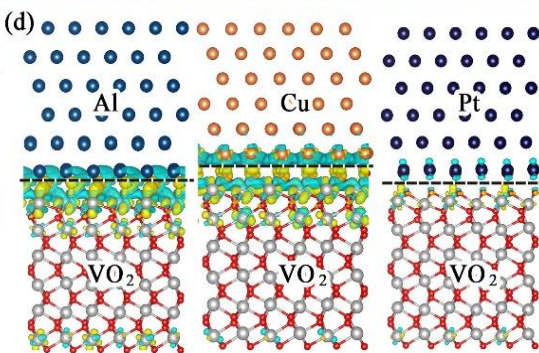
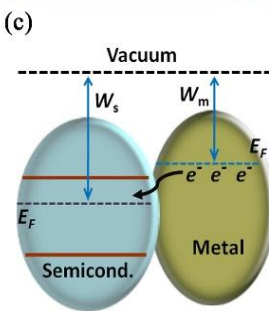
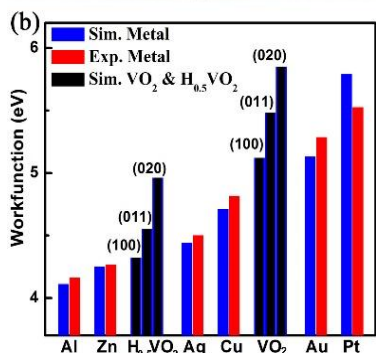
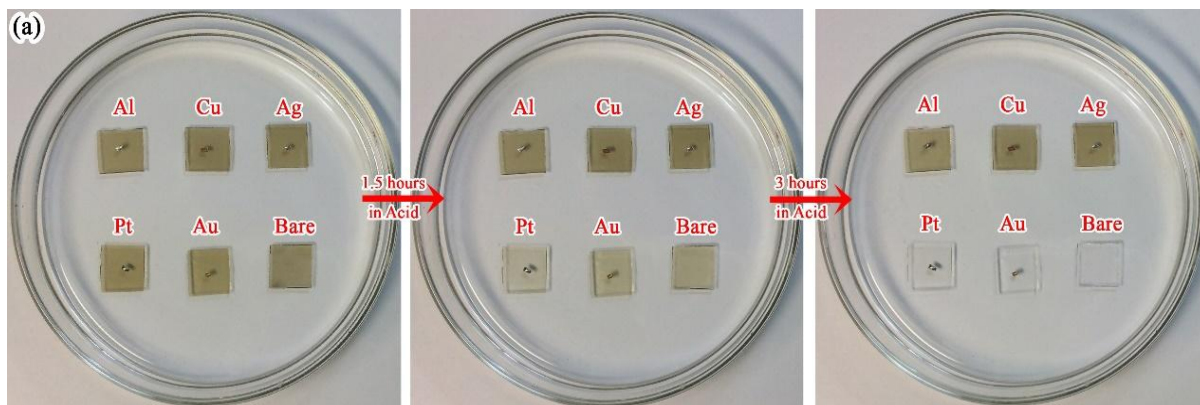


Y. Chen, Z. Wang *et al.*, J. Jiang\*, C. Zou\*, Y. Luo *Nat. Commun.* 2018, 9, 818

利用功函数差异驱动电子-质子共掺杂

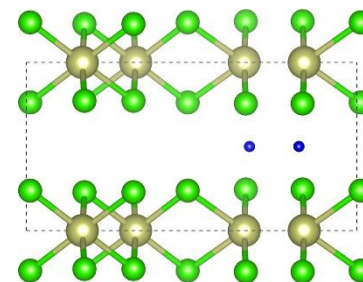
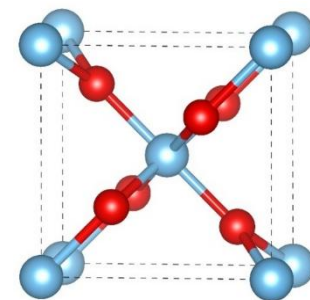
# 金属功函数筛选

Al、Cu、Ag、Pt、Au、Fe



# 半导体筛选

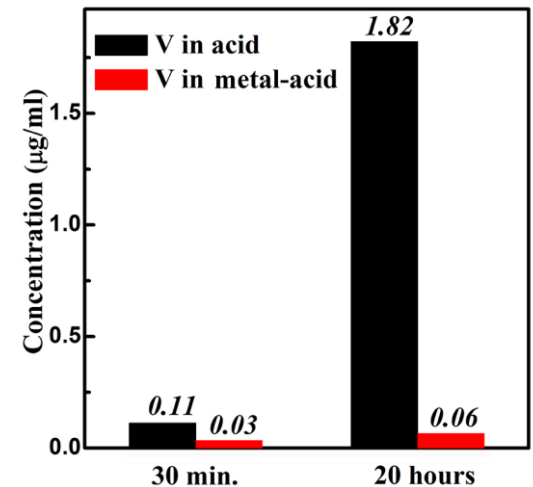
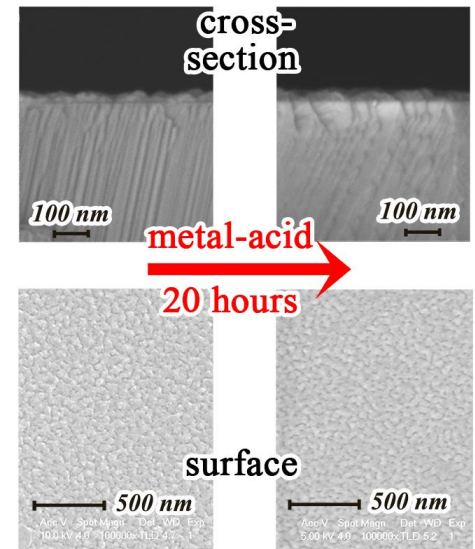
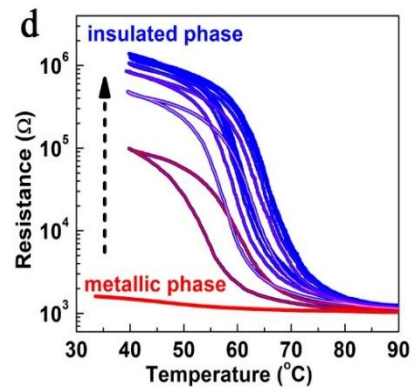
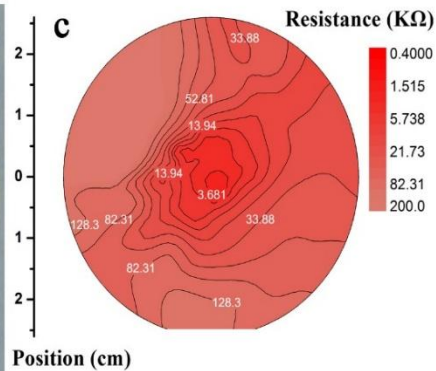
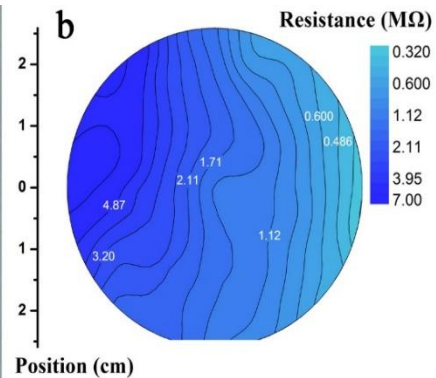
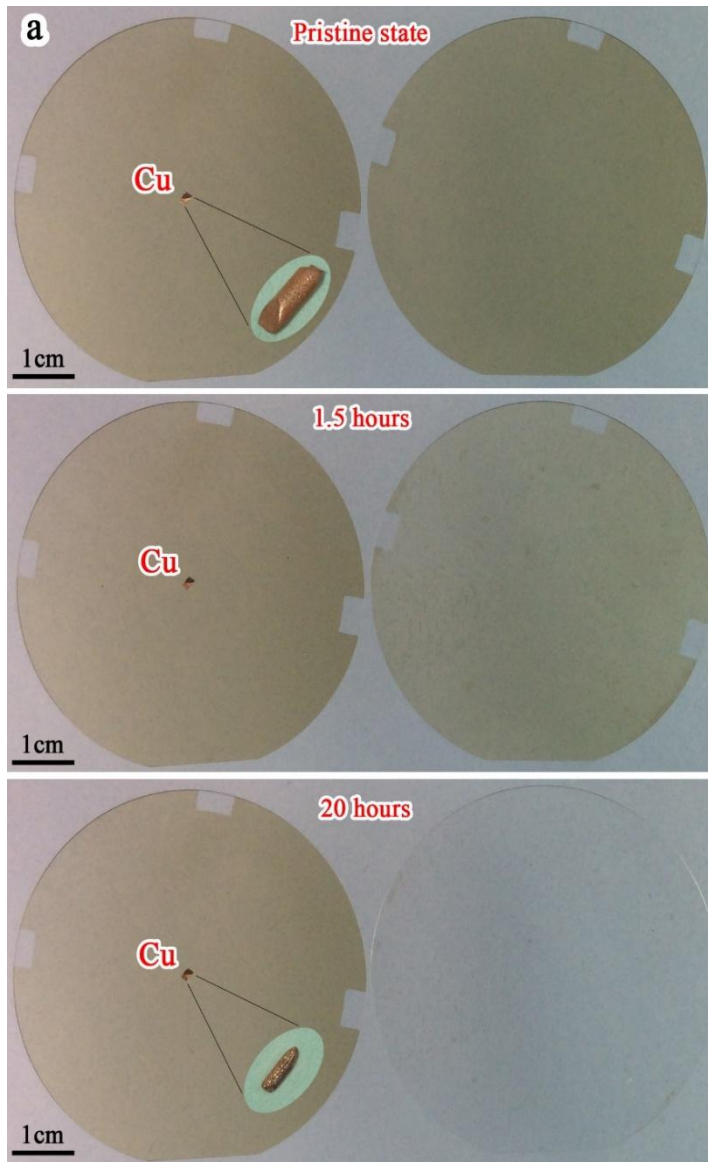
TiO<sub>2</sub>、SnO<sub>2</sub>、Cu<sub>2</sub>O、  
CeO<sub>2</sub>、ZnO、Mo<sub>2</sub>O<sub>3</sub>、  
IrCl<sub>3</sub>、RhCl<sub>3</sub>



Y. Chen, Z. Wang *et al.*, J. Jiang\*, C. Zou\*, Y. Luo *Nat. Commun.* 2018, 9, 818

电子-质子共掺杂——温和条件下改性材料增强光催化性能





Y. Chen, Z. Wang *et al.*, J. Jiang\*, C. Zou\*, Y. Luo *Nat. Commun.* 2018, 9, 818

加氢显著提高材料导电能力与稳定性

# 调控电子运动和原子运动 实现温和条件固体材料加氢改性

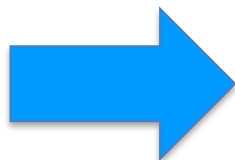


Y. Chen, Z. Wang *et al.*, J. Jiang\*, C. Zou\*, Y. Luo *Nat. Commun.* 2018, 9, 818

利用功函数差异驱动电子-质子共掺杂

# 功能材料的物理化学应用

光电器件  
(光)催化  
光物理/光化学  
生物医学  
发光材料  
能源材料  
分子光子学  
分子电子学



光电作用

能量转移

物质转化

电子态运动与演化

# 光能捕获

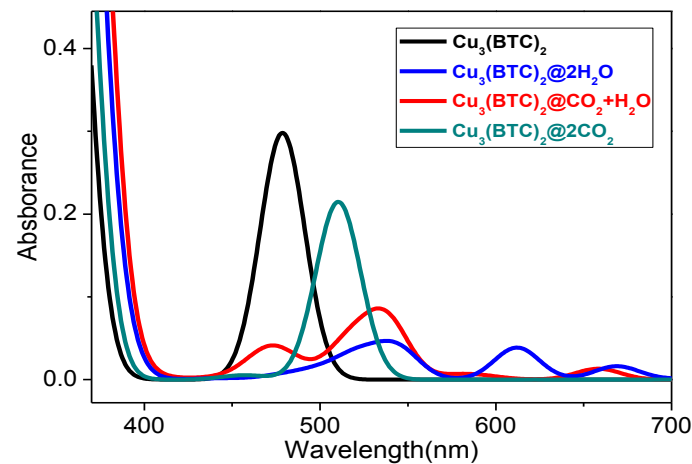
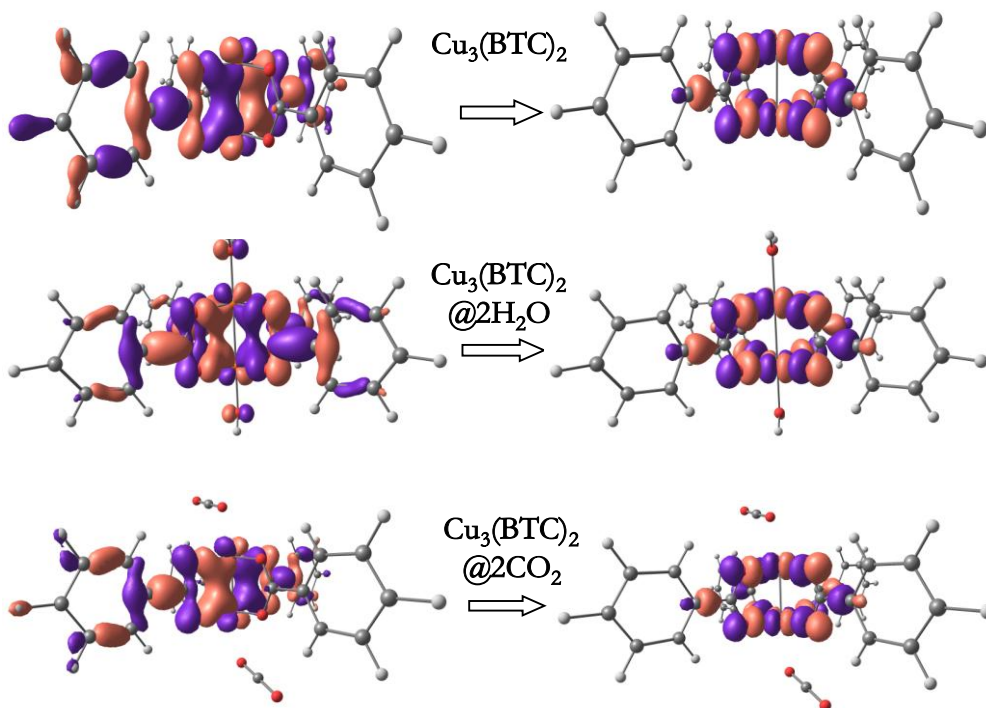
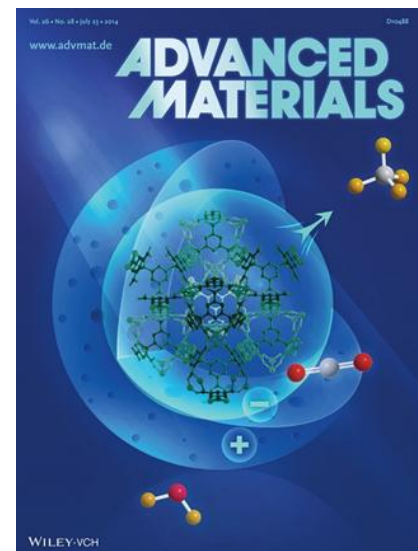
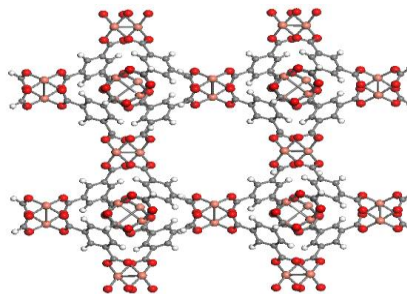
光电作用

能量转移

物质转化

金属有机骨架 (MOF)  $\text{Cu}_3(\text{BTC})_2$

可见光吸收差 .vs. 高效光催化 $\text{CO}_2$ 产生 $\text{CH}_4$



*Adv. Mater.* 2014, 26, 4788.

分子吸附调节电子激发跃迁偶极矩——增强可见光吸收

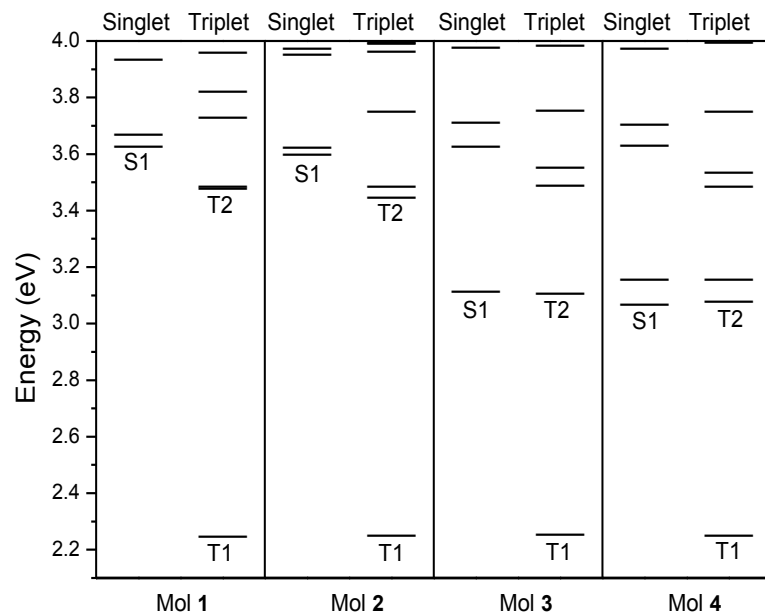
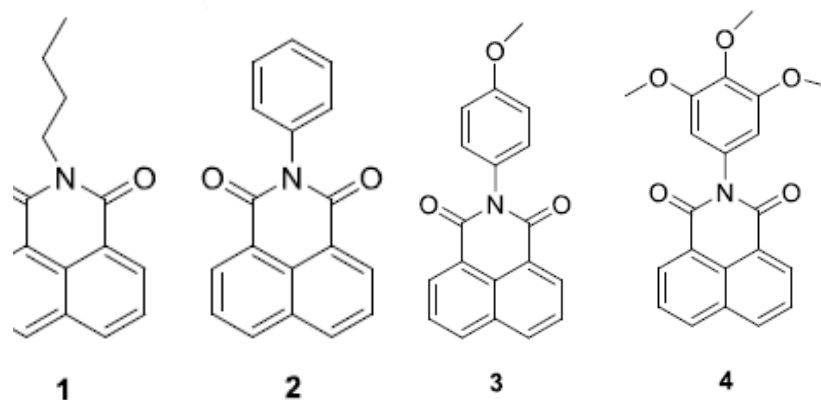
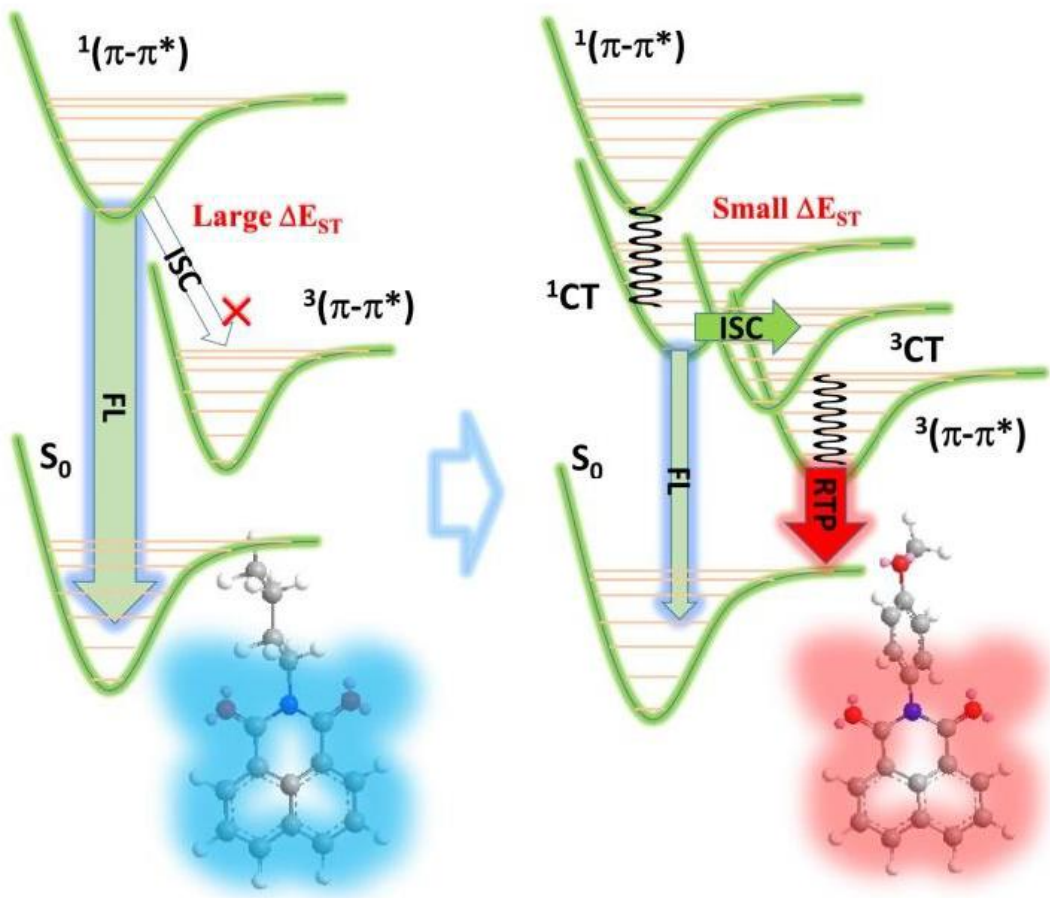


# 分子发光

光电作用

能量转移

物质转化



Angew. Chem. Int. Ed. 2016, 55, 9872

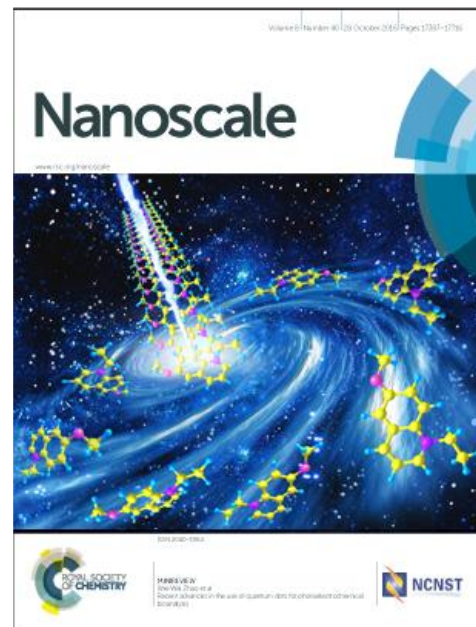
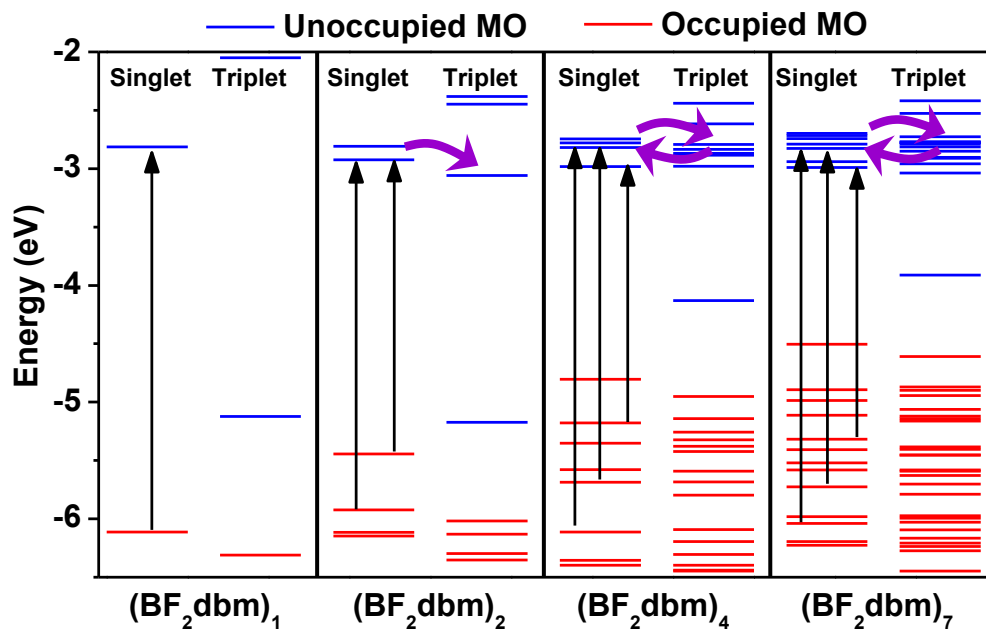
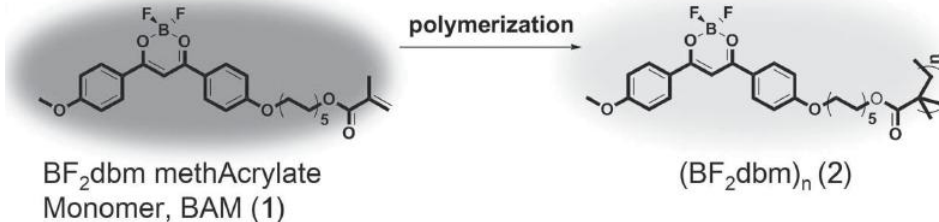
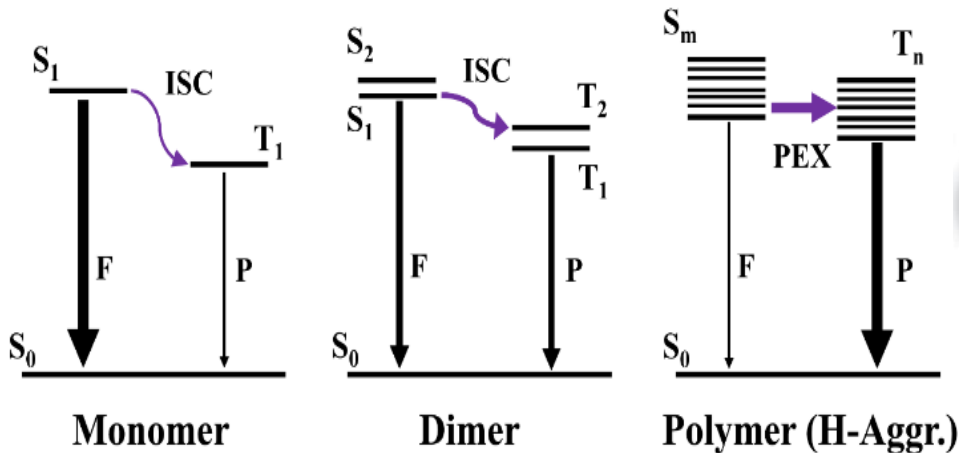
分子基团调节电子系间穿越通道——增强磷光发射

# 分子发光

光电作用

能量转移

物质转化



Macromol Rapid Comm. 2015, 36, 298  
Nanoscale 2016, 8, 17422

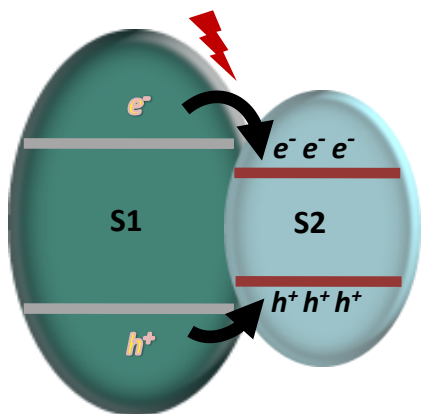
分子聚集调节电子系间穿越几率——增强磷光发射

# 电荷运动

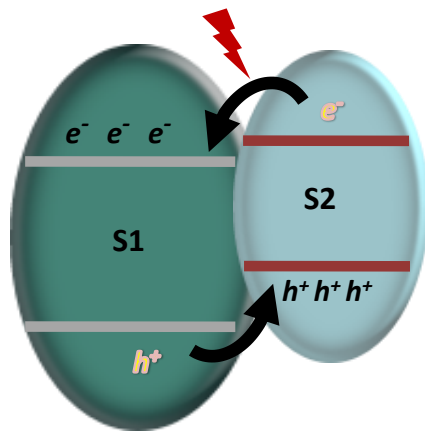
光电作用

能量转移

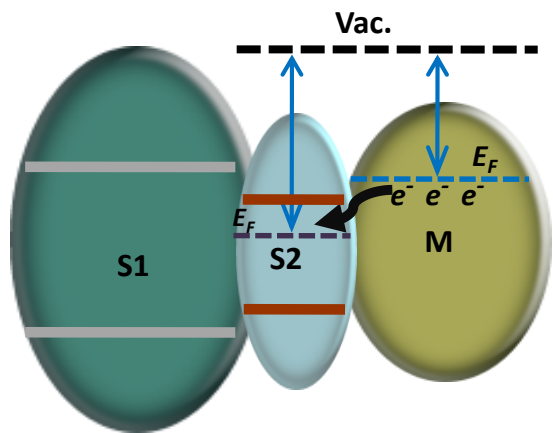
物质转化



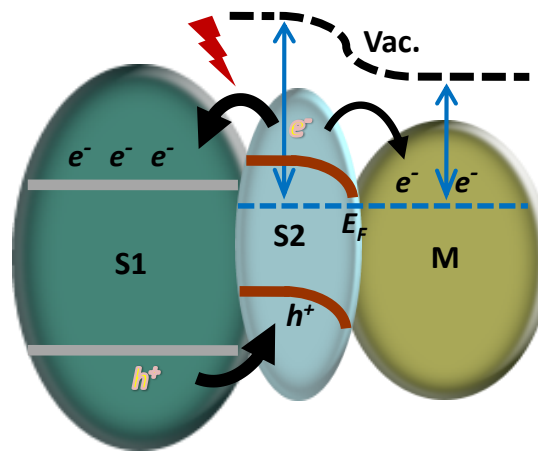
I型半导体异质结



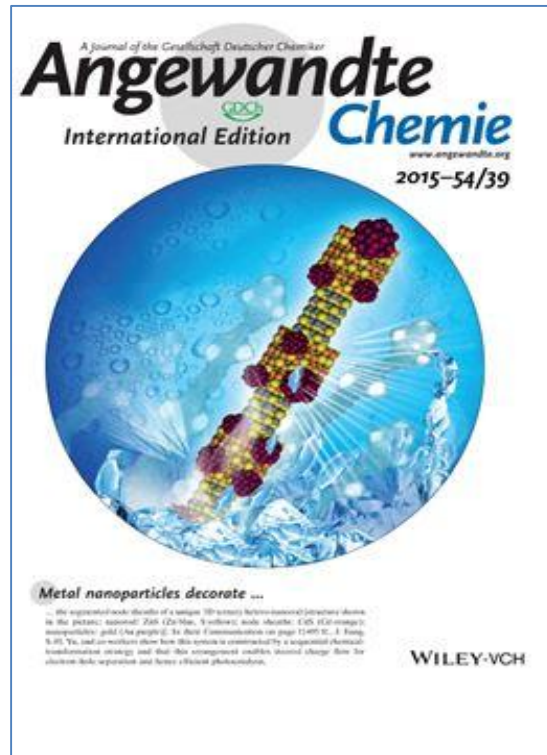
II型半导体异质结



金属团簇修饰I型半导体纳米异质结



II型半导体—(半导体/金属)纳米异质结



*Angew. Chem. Int. Ed.*  
2015, 54, 11495  
*Nano Res.* 2016, 9, 1590  
*Angew. Chem. Int. Ed.*  
2016, 39, 11495

金属修饰引导电子运动方向——实现高效光催化

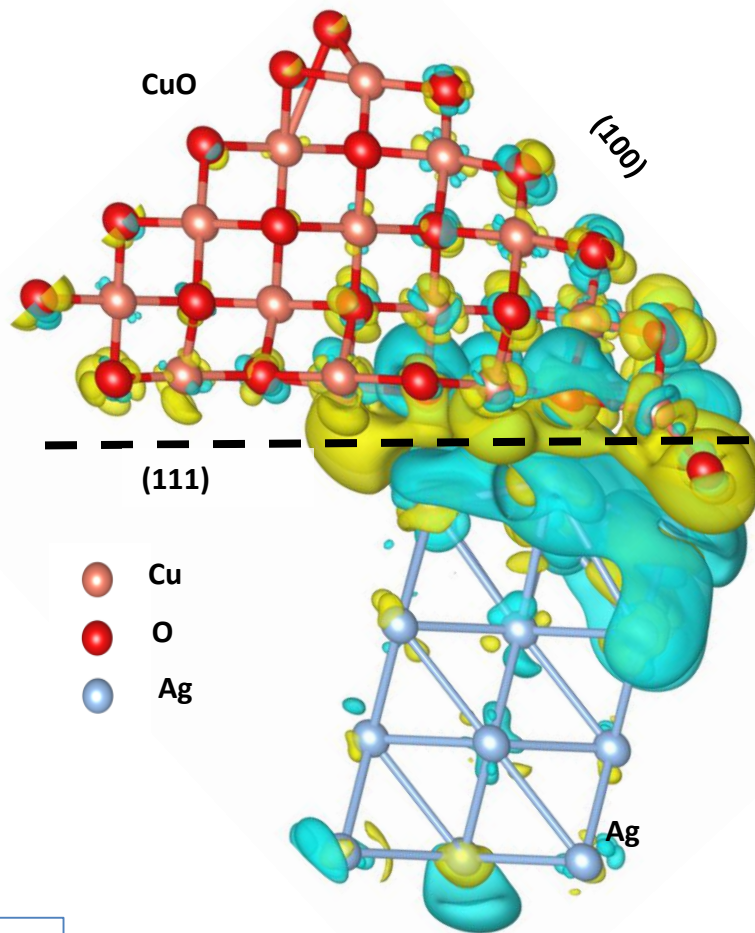
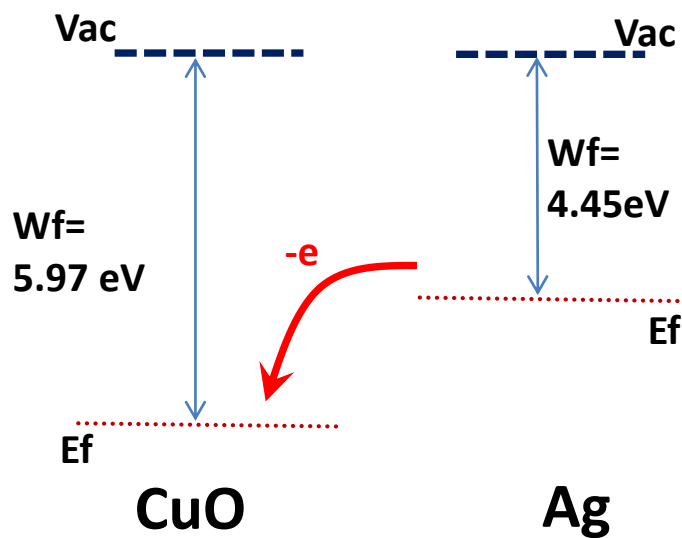
# 化学吸附与反应

光电作用

能量转移

物质转化

Ag提供极化电荷给CuO  
降低CO氧化势垒



*J. Am. Chem. Soc.* 2014, 136, 14650.

功函数差驱动复合材料界面电荷极化——增强反应活性



# 化学吸附与反应

光电  
作用

能量  
转移

物质  
转化

## Semicond-Metal Polarization :

Fe/VO<sub>2</sub> for H doping  
CuO/Ag oxidize CO  
Co<sub>2</sub>@C<sub>2</sub>N for O<sub>2</sub> activation  
Co@C<sub>3</sub>N<sub>4</sub> for HCOOH Dehydrogenation

*Nature. Commun.* 2018, 9, 818  
*J. Am. Chem. Soc.* 2014, 136, 14650.  
*J. Phys. Chem. Lett.* 2016, 7, 1750  
*J. Mater. Chem. A* 2018, DOI: 10.1039/C8TA02299B

## Metal-Metal Polarization :

Pt/Pd layers for water splitting  
Pt/Pd alloy for water splitting  
Fe/Co doping Pt for HER

*Angew. Chem. Int. Ed.* 2014, 53,12120.  
*Angew. Chem. Int. Ed.* 2015, 54,14810.  
*Adv. Mater.*, 2016, 28, 2077.

## Semicond Facet Polarization :

NiO nanosteps for Water splitting

*Sci. Rep.* 2015, 5, 8557.

## Semicond defect induced Polarization :

WO<sub>3</sub> defect for oxygen activation

*J. Am. Chem. Soc.* 2016, 138, 8928.

## Semicond-Chemical Group Polarization :

Si nanowires for water splitting

*Angew. Chem. Int. Ed.* 2015, 54, 2980.

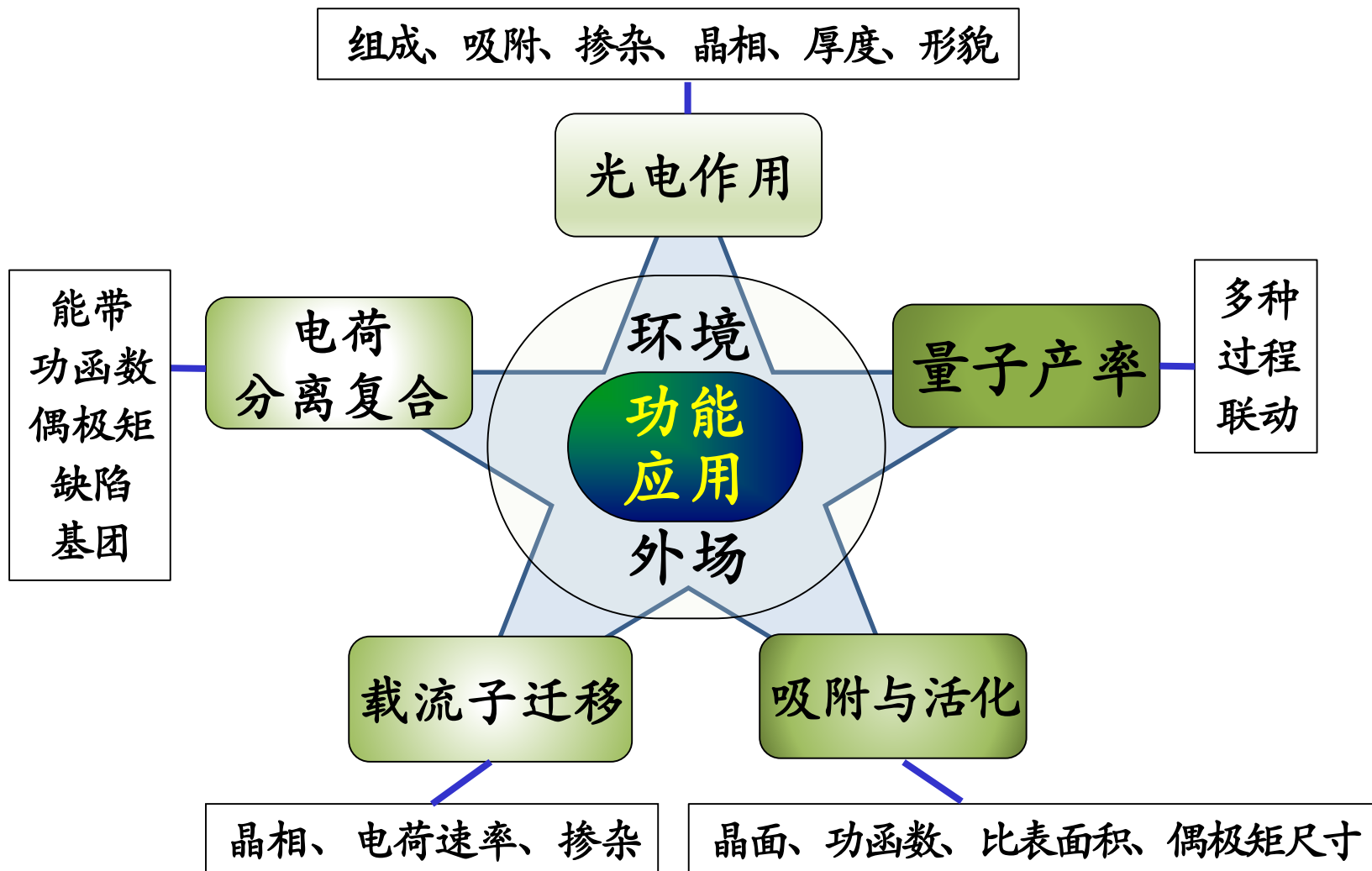
## Metal-Mol aggregates Polarization:

Pt catalyze HCOOH

*J. Phys. Chem. C.* 2015, 119, 19287.

功函数、晶面、缺陷、化学基团、聚集诱导极化

# 展望：实用的复合功能材料设计与仿真软件包



发掘可计算可测量的描述子，集成已有计算工具



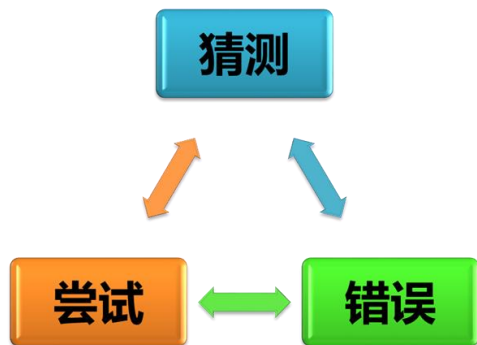
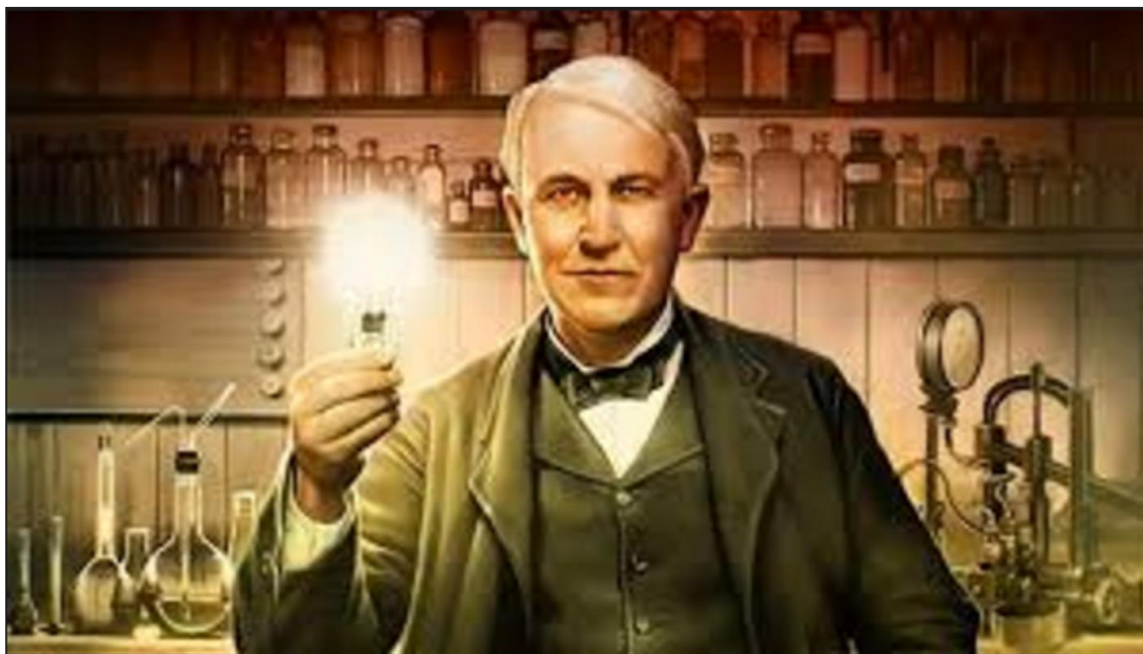
**理解材料构效关系是很艰难的事，  
我们常常看到的这样的景象**

# 新材料——研发瓶颈

## 低效试错模式大行其道

上帝说：要有光！

爱迪生说：要有灯丝！



10年试错：1600矿石、3000多种植物丝、3000多种金属丝！



15 years hard working



10 days happy time



Foldit screenshot illustrating tools and visualizations  
[David Baker](#) & 57000 Game Player



# AlphaGo





# AlphaGo

## 阿法狗

### 人工智能(AI)巅峰成就

棋盘361个交叉点  
黑、白、空3种状态  
可能性是3的361次方

天文数字！  
所有的可能性相加起来，  
比宇宙中的原子数还要多

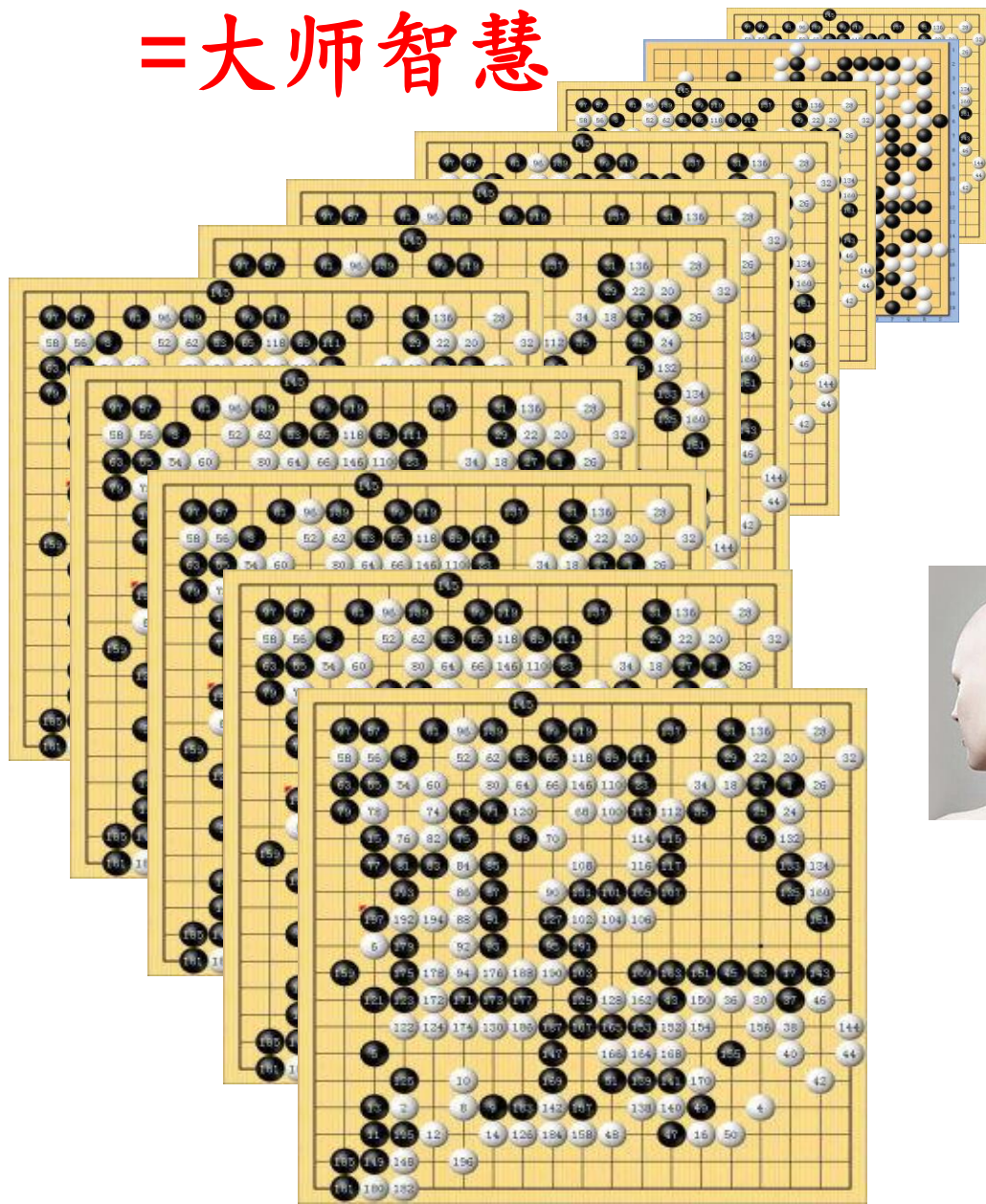




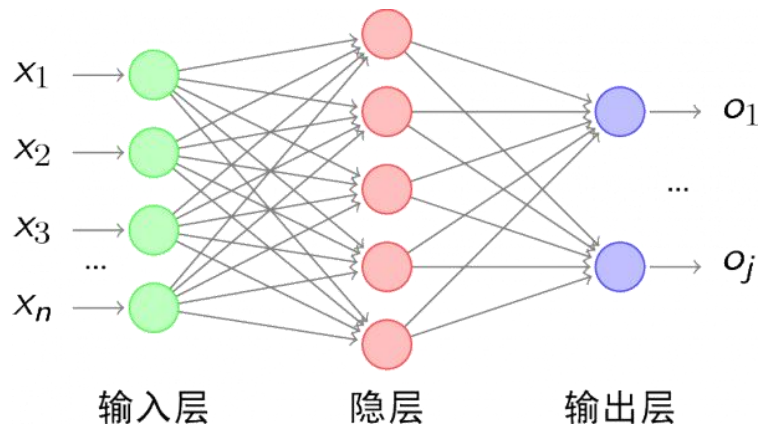
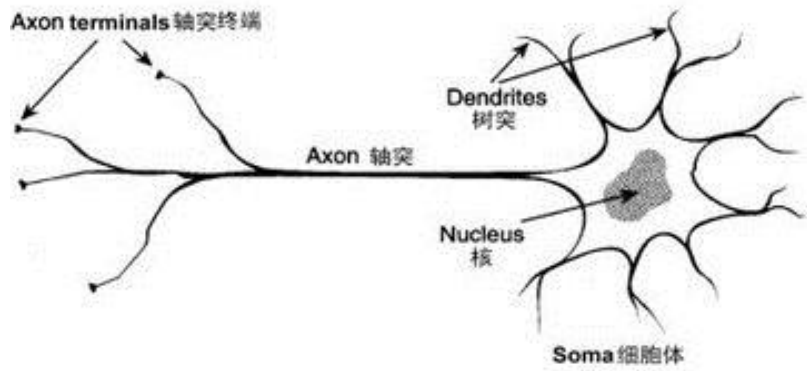
人工智能+大数据  
=大师智慧



AlphaGo







# 功能材料基因图谱：构效关系

应用需求与实验室研发  
光电信息 化学转化 能源环境 生命 药物

应用鸿沟：分离的性能参数－集成后的协同作用

电学

光学

力学

磁学

热学

动态

活性

稳定

外观

尺度鸿沟：空间、时间、能量跨尺度耦合

结构：组成、键合、尺寸、缺陷...；特征：能量排列、电荷分布、耦合作用、演化...

材料鸿沟：复杂的体系构造－简化的理想模型

量子力学的“第一原理”

元素周期表、5个基本物理常数

依赖原点的  
传统模式

# 材料基因与大数据

应用需求与实验室研发  
光电信息 化学转化 能源环境 生命 药物

理论规律与大数据关联

电学

光学

力学

磁学

热学

动态

活性

稳定

外观

理论规律与大数据关联

结构：组成、键合、尺寸、缺陷...；特征：能量排列、电荷分布、耦合作用、演化...

理论规律与大数据关联

量子力学的“第一原理”

元素周期表、5个基本物理常数

跳离原点的  
关联模式

宏观  
性能

微观  
基因

# 材料基因与大数据

应用需求与实验室研发  
光电信息 化学转化 能源环境 生命 药物

基于第一性原理的材料基因大数据

电学 — 光学 — 力学 — 磁学 — 热学

复合 — 活性 — 稳定 — 外观

能量排列 — 电荷分布 — 耦合作用 — 演化

组成 — 键合 — 尺寸 — 晶相 — 缺陷

元素周期表、5个基本物理常数

基因大数据  
预测模式



# 材料大数据基础

- 科学文献：“原始创新” & “垃圾paper”
- 企业与研究人员积累：“冗余” & “无效” 数据
- 网络资源：“共享” & “拥有”

**RCSB PDB** PROTEIN DATA BANK  
An Information Portal to 105,499 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Advanced Search | Browse by Annotations

Welcome

Deposit

Search

Visualize

Analyze

RCSB PDB

Deposit

Search

Visualize

Analyze

Download

Learn

More

MyPDB Login

### A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data. The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

### January Molecule of the Month

Cascade and CRISPR

### Latest Entries

As of Tuesday, Jan 06

**4NY1** PDB Entry

X-ray structure of the unliganded uridine phosphorylase from *Yersinia pseudotuberculosis* at 1.7 Å resolution

View in 3D

### New Features

December 2014 Release

Improved Representation of Large Structures

Updates to PDB Archive Visualization. New Queries for Ribosomes and Viruses

November 2014 Release

Redesigned Home Page

Simple, Clean, Usable. Tools & Functions More Visible.

Reference Management

Quickly save references to Mendeley Desktop

### News

Publications

New Calendar Celebrates 15th Anniversary of Molecule of the Month

A new online calendar highlights many of the structures featured in this ongoing column.

01/06/15

Video Challenge for High School Students in 2014

12/10/14

Integration of Large Structures with the Main PDB Archive

09/18/14

Inclusion of Large Structures in the Main PDB Archive

PDB at a Glance

34981 Distinct Protein Sequences | 27406 Structures of Human Sequences | 7454 Nucleic Acid Containing Structures | More Statistics

About About Us | Citing Us | Publications | Team | Careers | Usage & Privacy

Help Contact Us | Help System | Website FAQ | Browser Test | Glossary

RCSB Partners Nucleic Acid Database | Structural Biology Knowledgebase

The RCSB PDB (citation) is managed by two members of the Research Collaboratory for Structural Bioinformatics.

RUTGERS UC San Diego

RCSB PDB is a member of the




12万生物大分子数据

分子库：2000万

material genome big data

Molecules ▾  🔍

< Prior 1 / 3 Next >

1,5-Hexadiyne Formula $C_6H_6$	
phenyl-t Formula $C_6H_6$	
vinyl vinylene vinylidene Formula $C_6H_6$	

晶体库：30万

material genome big data

Crystals ▾ +element:S +element:Pb -element:Pd +system:Cubic 🔍

<Prior 1 / 1 Next>

Pb4S4

Formula  $Pb_4S_4$  (PbS)

System Cubic

a 592.37pm      b 592.37pm      c 592.37pm

$\alpha$  90°       $\beta$  90°       $\gamma$  90°

Pb8S100Hf28Ti72

Formula  $Pb_8S_{100}Hf_{28}Ti_{72}$  ( $Pb_2S_{25}Hf_7Ti_{18}$ )

System Cubic

a 1744.89pm      b 1744.89pm      c 1744.89pm

$\alpha$  90°       $\beta$  90°       $\gamma$  90°

Pb8S100Ti72Ti28

Formula  $Pb_8S_{100}Ti_{72}Ti_{28}$  ( $Pb_2S_{25}Ti_{18}Ti_7$ )

System Cubic

a 1709.52pm      b 1709.52pm      c 1709.52pm

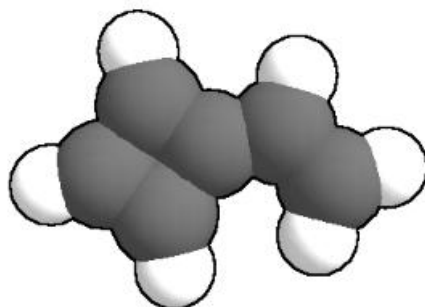
$\alpha$  90°       $\beta$  90°       $\gamma$  90°

量子化学计算  
30万分子

## 1,3-Cyclobutadiene, 1-ethenyl-

space fill ball stick

Download Structure (SDF)



### 识别

IUPAC 名称

ethenylcyclobutadiene

InChI Key

KZDURZABPFPZHJ-UHFFFAOYSA-N

InChI

InChI=1S/C6H6/c1-2-6-4-3-5-6/h2-5H,1H2

标准 SMILES

C=CC1=CC=C1

### 计算性质

分子式

C6H6

分子量

78.11184g/mol

Exact Mass

78.04695g/mol

分子量 (单同位素)

78.04695g/mol

XLogP3

1.9

氢键给体数

0

氢键受体数

0

可旋转键数

1

拓扑极性表面积

0A<sup>2</sup>

重原子数

6

形式电荷

0

复杂度

104

同位素原子数

0

已定义手性中心数

0

未明示手性中心数

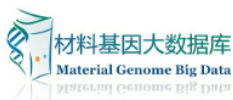
0

已定义手性轴数

0

未明示手性轴数





请输入需要搜索的内容 搜索

数据概览 : 10,000,000 分子 ; 100,000 晶体 ; 200,000 谱

## 数据上传

欢迎

材料查询

数据录入

基因服务

问答平台

材料类型
-----请选择-----

结构数据录入
浏览...

谱学数据录入
-----请选择-----

性质录入

+ 输入关键字

- 外观描述	输入关键字
- 味道	输入关键字
- 沸点	输入关键字

计算输出文件录入
浏览...

打开

计算机 > work (D:) > 360cloud > 材料基因 > 关键字挖掘

组织	名称	修改日期	类型
迅雷下载	adsorption energy	2016/6/13 21:10	文件夹
音乐	bonding energy	2016/6/15 17:32	文件夹
家庭组	energy gap	2016/6/15 17:23	文件夹
计算机	fermi energy	2016/6/15 17:28	文件夹
win7 (C:)	fluorescence-absorption peak	2016/6/15 17:34	文件夹
work (D:)	transition barrier = activation energy	2016/6/13 21:10	文件夹
life (E:)	workfuntion	2016/6/15 17:26	文件夹
网络	Word-List.txt	2016/6/15 17:34	文本文档

文件名(N): 答辩-5-18.pptx 所有文件

打开(O)

**材料更新** more

**低维碳材料库正式建立** »2015-03-26

2015年03月：低维碳材料库正式建立，首批录入富勒烯分子约5万

**用户指南** more

**材料数据查询** (加速研发, 激发灵感)

包括功能分子结构与性质、固体材料晶体结构与性能、药物分

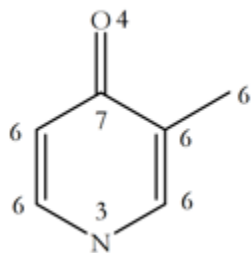
**新闻动态** more

**第一届材料计算设计与模拟国际会议召开** »2015-10-14

会议现场大会主席李依依院士主持开幕式为了回顾材料科学给材

# 材料结构的数字化识别

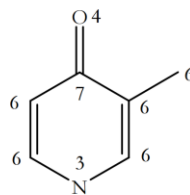
分子结构  $\longrightarrow$  拓扑分析  $\longrightarrow$  向量矩阵

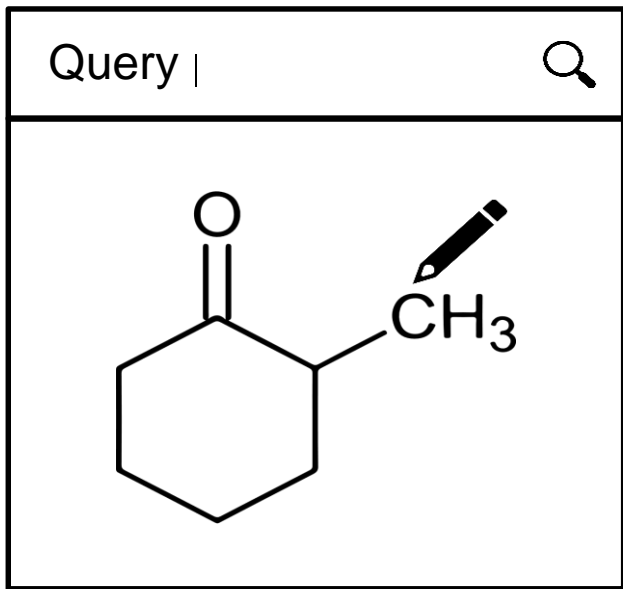



	unique ap	frequency
1	c31c2102	3
2	c31c2101	2
3	o11c2103	2
4	n20c2101	2
5	c31c2103	1
6	c21c2102	1
7	c21c2101	1
8	c21c2103	1
9	c31c3101	1
10	n20c3103	1
11	n20c3102	1
12	o11c3102	1
13	o11n2004	1
14	n20c2102	1
15	o11c3101	1
16	o11c2102	1
17	c31c1001	1
18	c21c1004	1
19	c31c1002	1
20	o11c1003	1
21	c21c1003	1
22	c21c1002	1
23	n20c1003	1
	28 total	

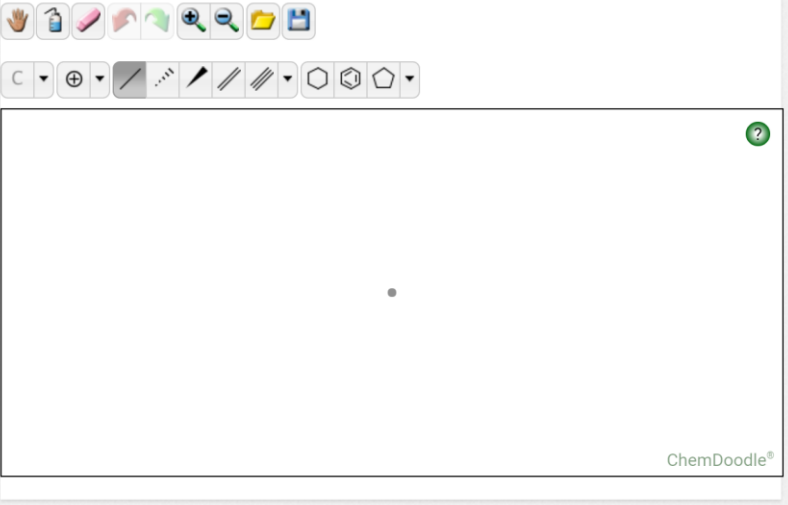
	unique tt	frequency
1	o11c31c31c21	1
2	o11c31c31c10	1
3	c31c31c21c21	1
4	c21c31c31c21	1
5	c21c31c31c10	1
6	n20c21c31c31	1
7	o11c31c21c21	1
8	c31c21n20c21	1
9	n20c21c21c31	1
10	n20c21c31c10	1
11	c21n20c21c21	1
	11 total	

AP/TT  
Descriptors





Molecules  



ChemDoodle®

AP:  
· 011-C20-4  
· .....  
TT:  
· C10-C30-C20-C30  
· .....

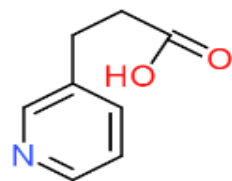
```
{bool: {  
  must: [  
    API项, TT项, .....  
  ],  
  .....  
}}
```

返回结果

```
{  
  "CAS-123-234-X": [...],  
  "CAS-987-654-Y": [...]  
  .....  
}
```

索引服务器  
**ElasticSearch**

# 材库1.01 [www.dcaiku.com](http://www.dcaiku.com)



ChemDoodle<sup>®</sup>

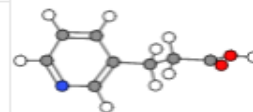
< Prior

1 / 141177

Next >

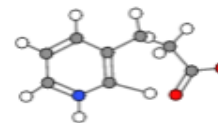
3-Pyridinepropionic acid

Formula  $C_8H_9NO_2$



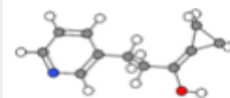
AGN-PC-01GO1N

Formula  $C_8H_9NO_2$



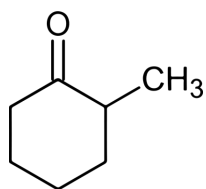
AGN-PC-0HSZNV

Formula  $C_{11}H_{13}NO$



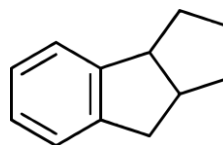
AGN-PC-002FDF





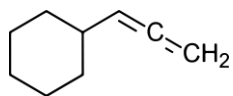
521ms

5匹配



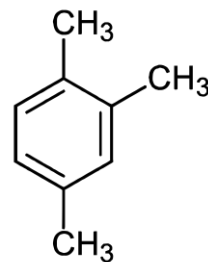
2504ms

2匹配



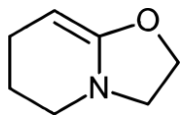
740ms

1匹配



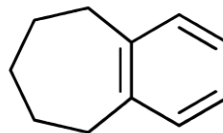
1546ms

2匹配



894ms

1近似匹配

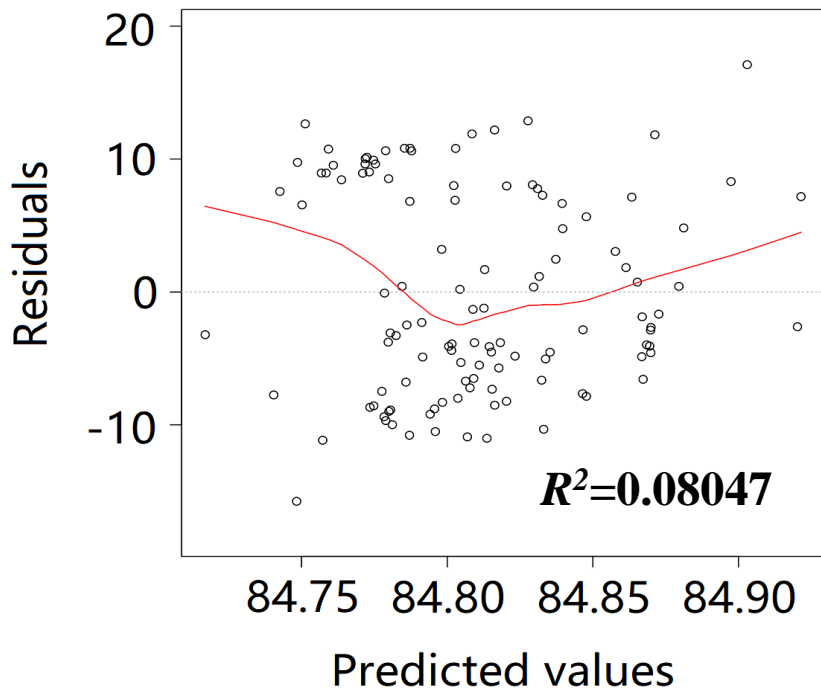
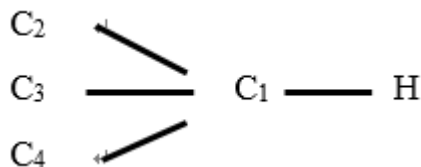


2473ms

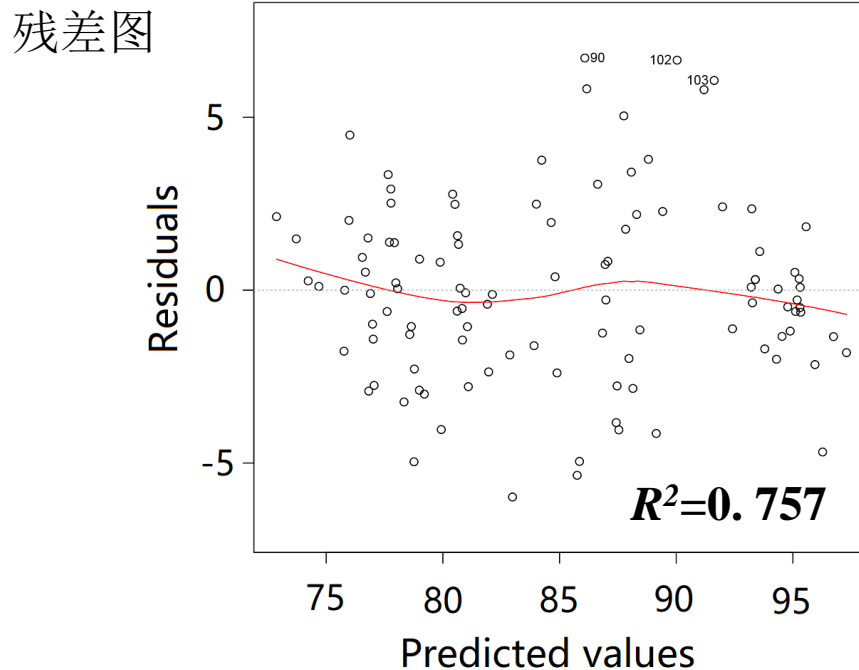
1匹配

# 构效关系函数化

C-H 键能：



$$E = A/d^2 + B$$



$$E = 85.54125 - 0.10373N_1^{-1} + 266.29556N_1 - 1.29568N_1^{-2} \\
 * (N_2^2 + N_3^2 + N_4^2) - 66.71595N_1^{-1} * (N_2^2 + N_3^2 + N_4^2)$$

$$N_1 = Q_{C1} * Q_H / d_{C1H}^2$$

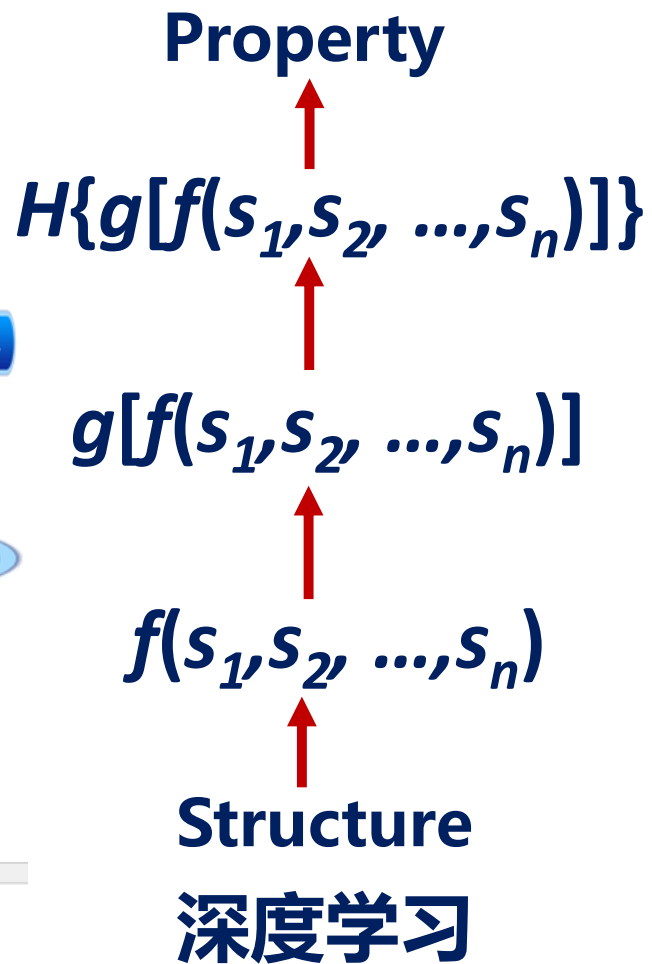
$$N_2 = Q_{C1} * Q_{C2} / d_{C1C2}^2$$

$$N_3 = Q_{C1} * Q_{C3} / d_{C1C3}^2$$

$$N_4 = Q_{C1} * Q_{C4} / d_{C1C4}^2$$

# 构效关系函数化



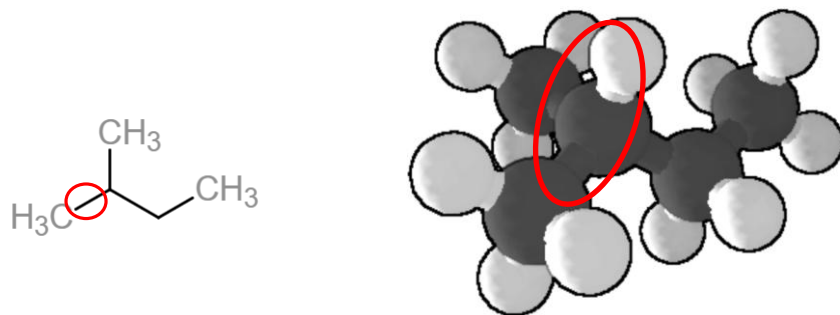




# 卷积神经网络(cnn)分析键能

C-H键的活化反应依赖于C-H键键能。

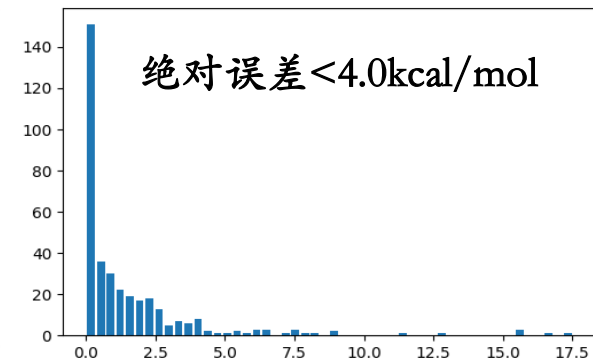
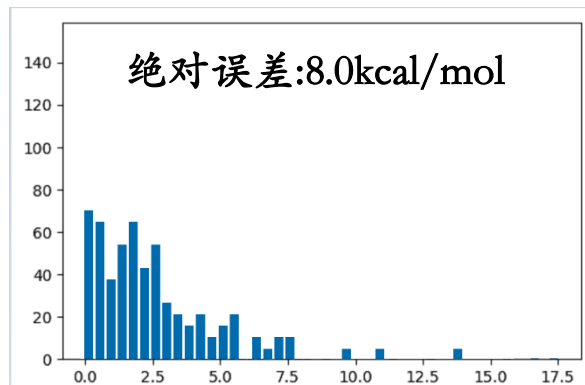
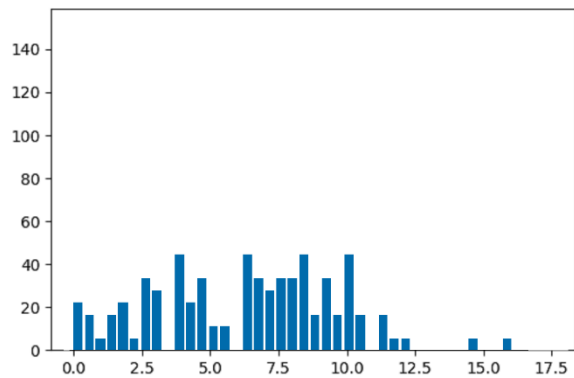
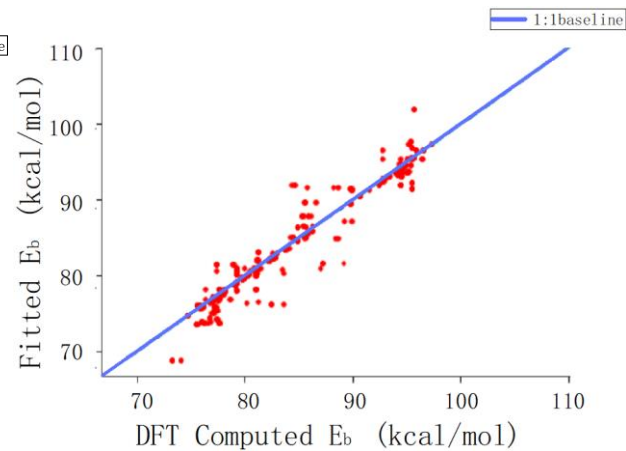
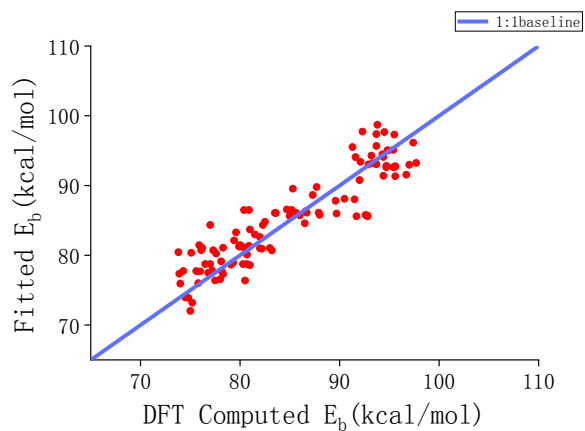
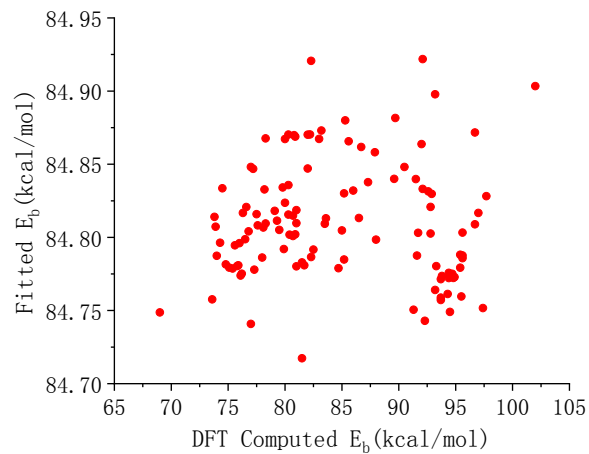
Gaussian B2LYP|6-31G.DP计算获得大量键能数据。



利用tensorflow框架

随机选取900组数据进行训练，200组数据进行测试

[C<sub>0</sub>-H键能, C<sub>0</sub>Mulliken电荷, H电荷, C<sub>0</sub>-C<sub>1</sub>键长, C<sub>1</sub>电荷, C<sub>0</sub>-C<sub>2</sub>键长, C<sub>2</sub>电荷, C<sub>0</sub>-C<sub>3</sub>键长, C<sub>3</sub>电荷]



文献\*中提出的模型 $E \sim d^{-2}$

引入库仑模型的统计分析

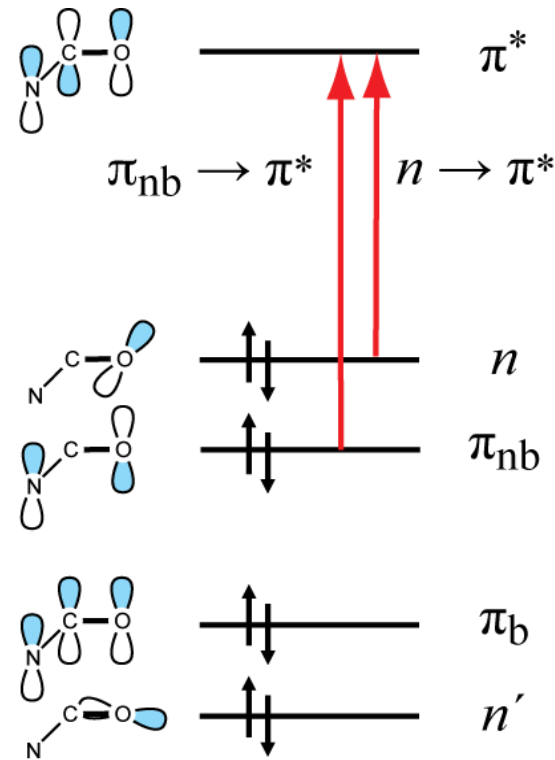
cnn训练模型

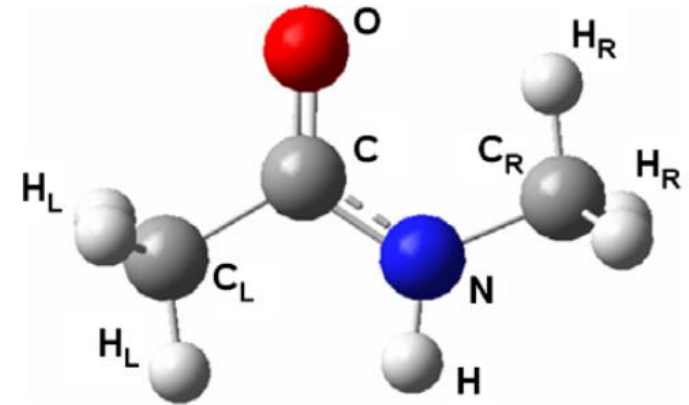
Pearson coefficients=0.89

训练结果的相关性强

# 卷积神经网络(cnn)分析蛋白质紫外吸收

- ◆ N-methylacetamide (NMA) is widely used as a model to mimic the **peptide bond**.
- ◆ The first two bands of the backbone electronic excitations are  $n\pi^*$  and  $\pi\pi^*$  amide bands.





MD Simulation



Extract the structure(10000)



Calculate The Transition energy by TDDFT



Statistical molecular geometry parameters



Neural Network Machine Learning

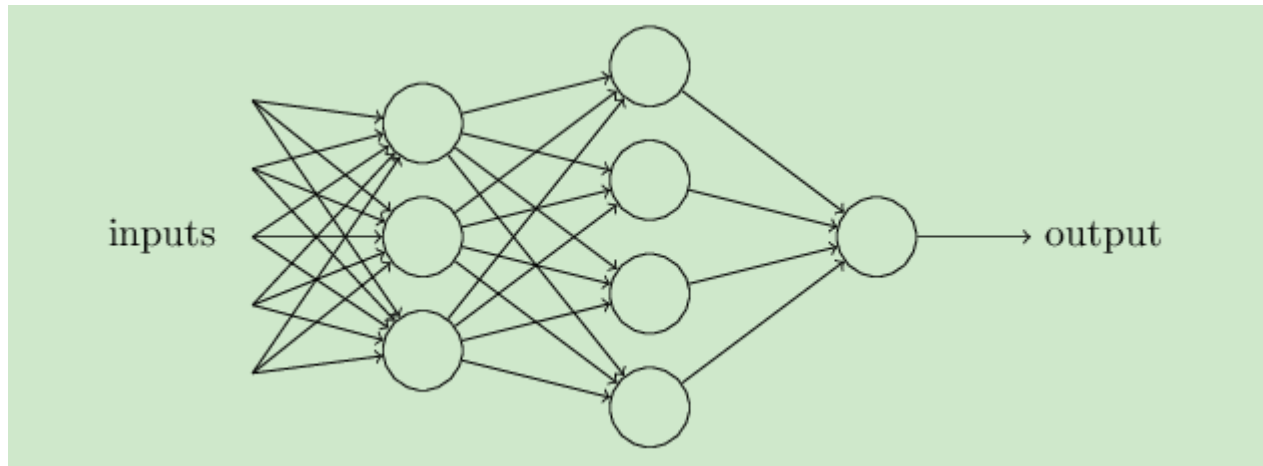


## Neural Networks



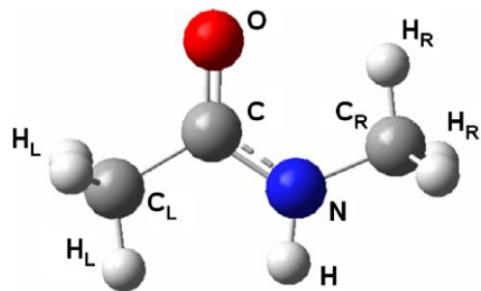
## Model : DNN(Deep Neural Networks)

- ◆ We use 9 structure parameters as inputs (x)
- ◆ Use three layers of neural network and one output layer (y)
- ◆ Train set : 9000; Test set :1000
- ◆ The neurons in each neural network are 16, 32, 128
- ◆ Use three layers of neural network and one output layer (y)
- ◆ using an activation function to remove linearization



# Predict dipole with neural network analysis

**NMA molecule: peptide bond**



**Dipole Moments from  
10000 MD snapshots**

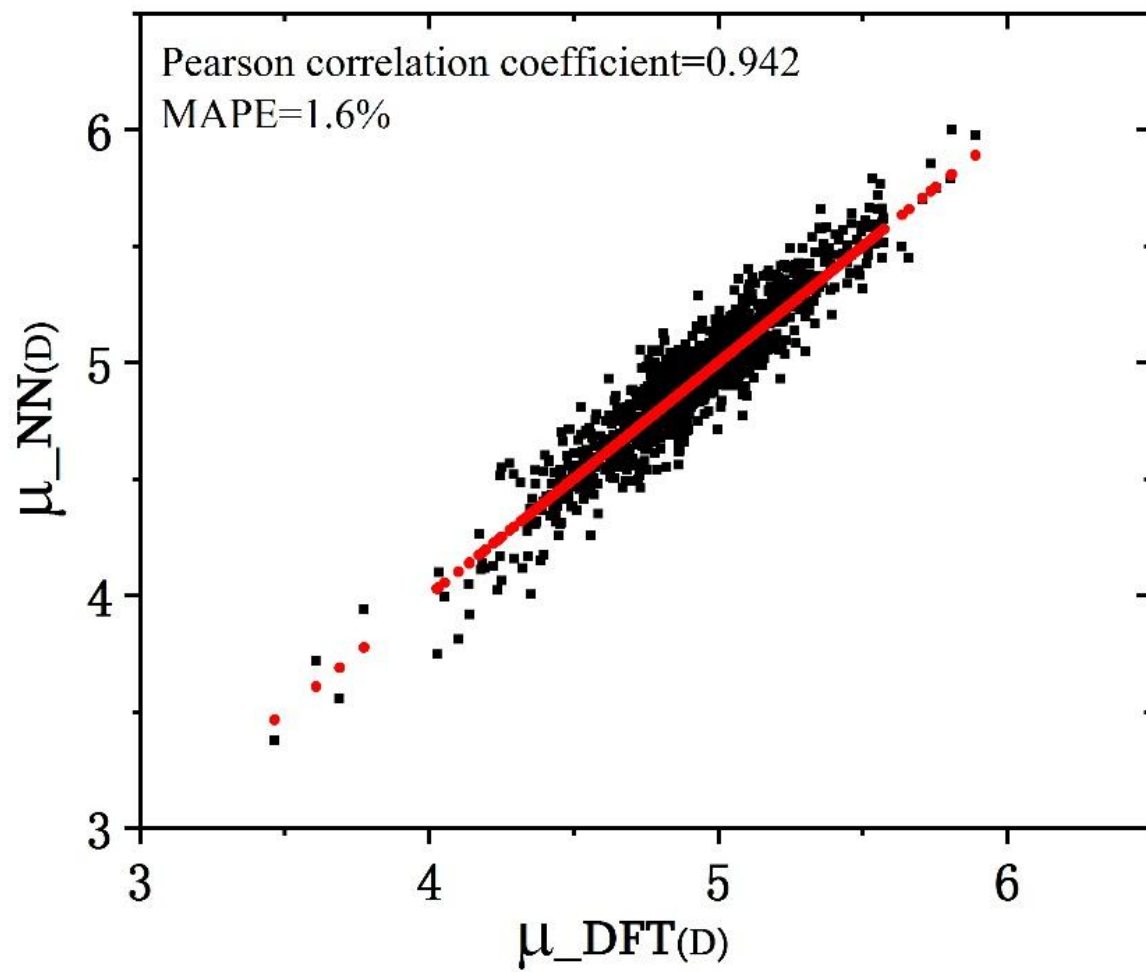
**Train set : 9000**

**Test set :1000**

PBE1PBE/cc-pVDZ

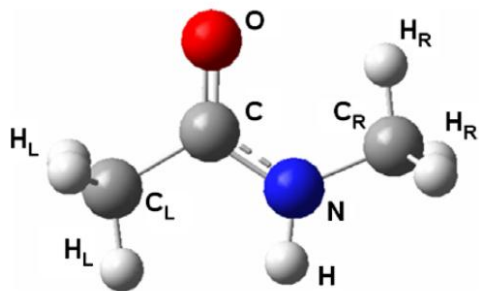
Average error: 1.8%

Pearson correlation coefficients: 0.94



# Predict dipole with neural network analysis

**NMA molecule: peptide bond**



B3LYP/6-311++G\*\*

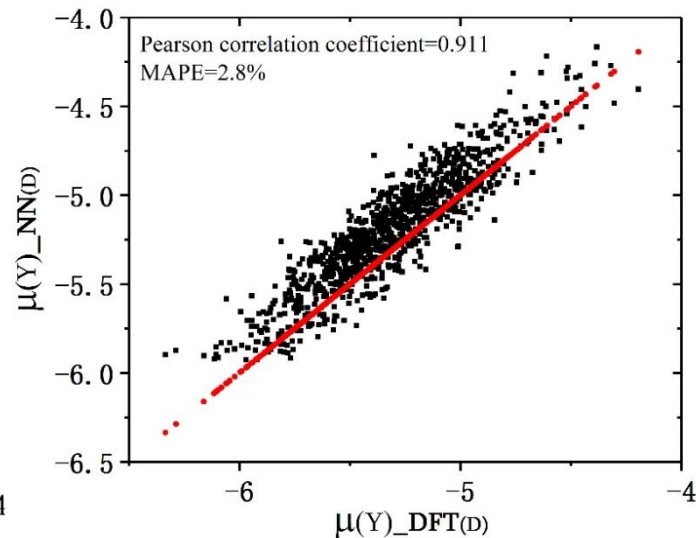
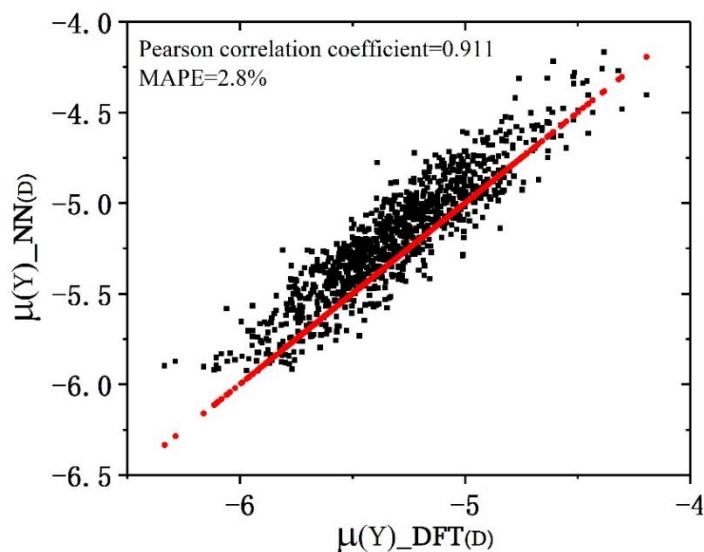
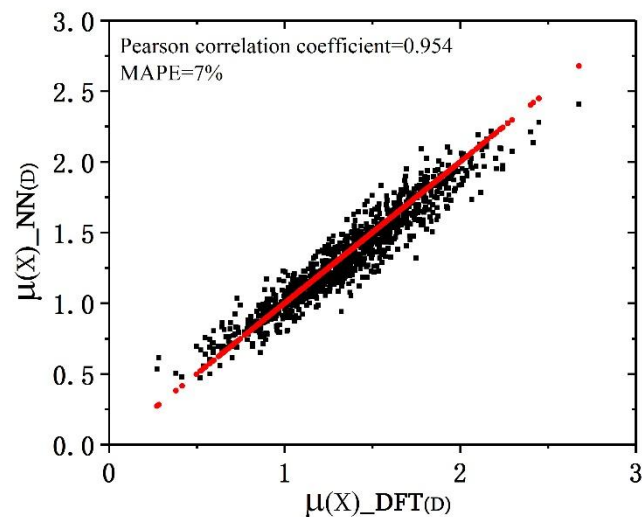
Average error: 2.8~7.0%

Pearson correlation coefficients: > 0.91

**Dipole Moments from  
10000 MD snapshots**

**Train set : 9000**

**Test set :1000**



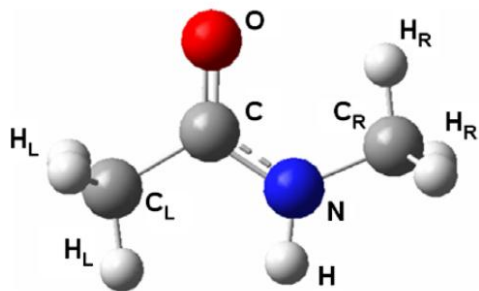
# Predict absorption spectra with neural network analysis

NMA molecule: peptide bond

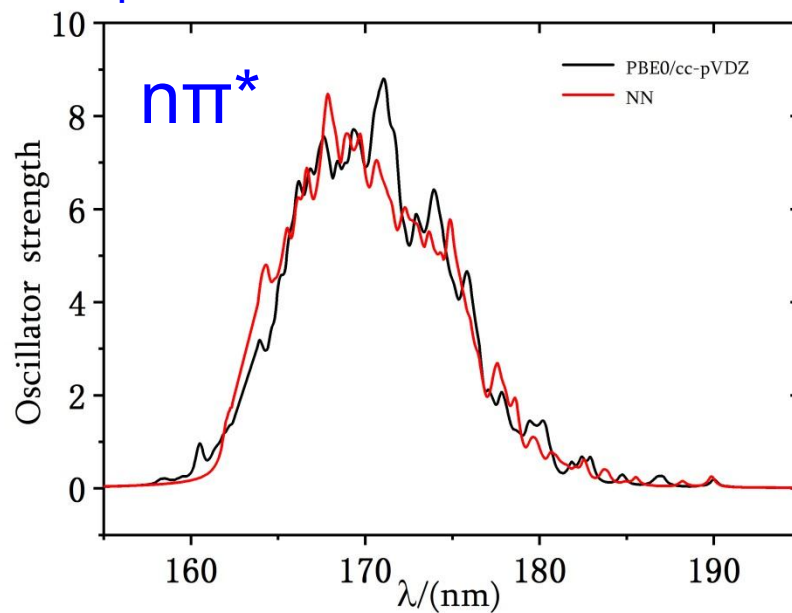
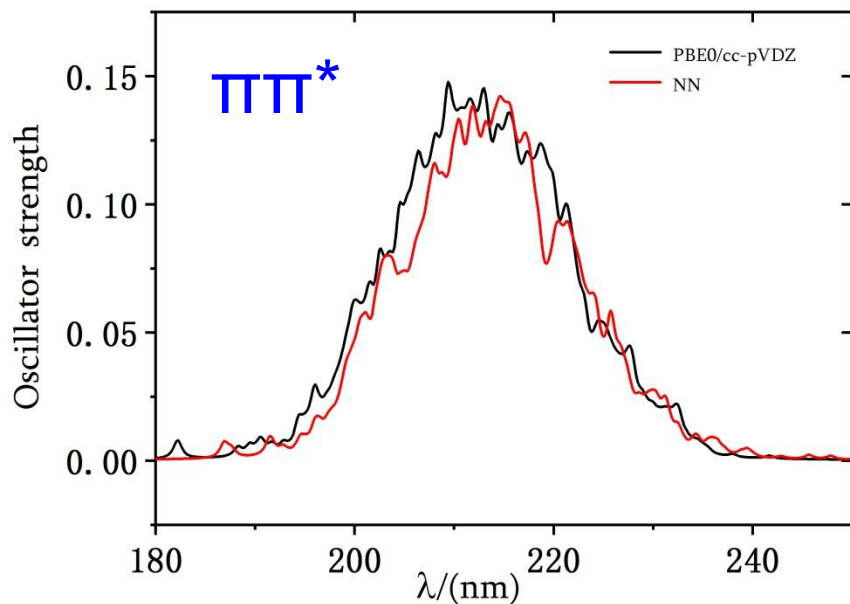
Spectra from 10000 MD snapshots

Train set : 9000

Test set :1000

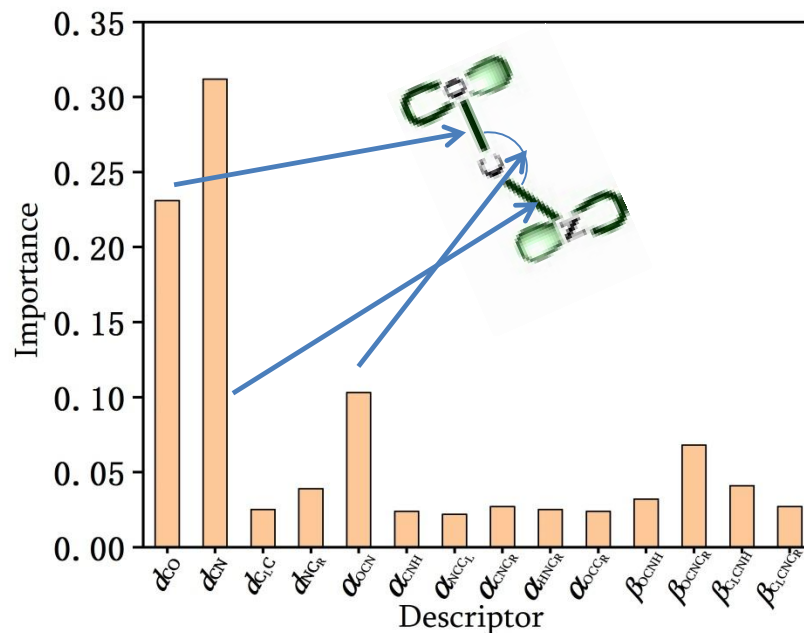
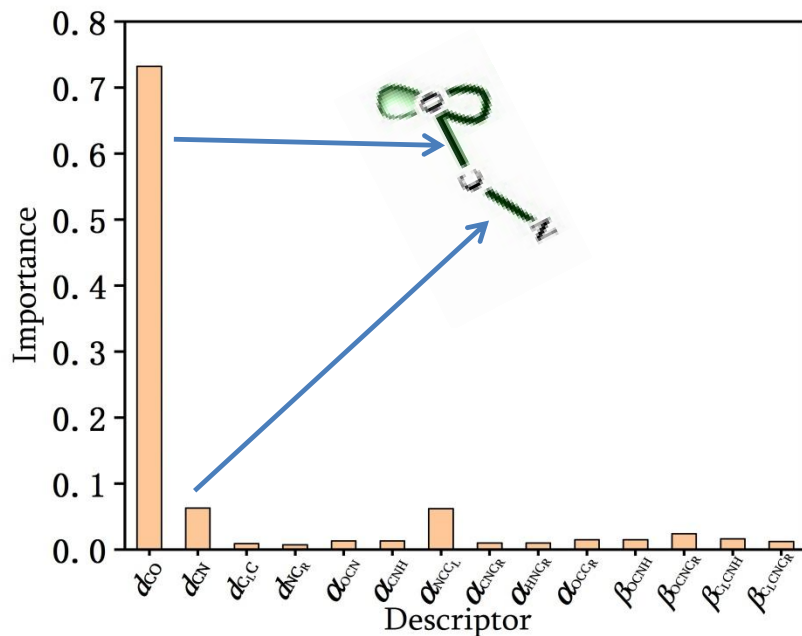
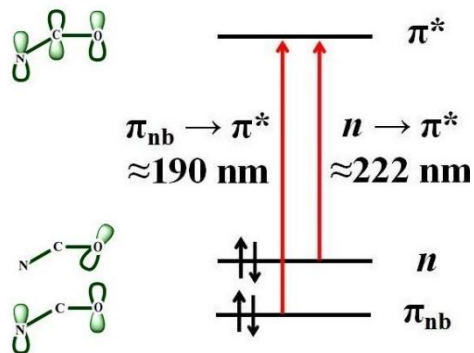


PBE1PBE/cc-pVDZ





# Importance analysis on transitions

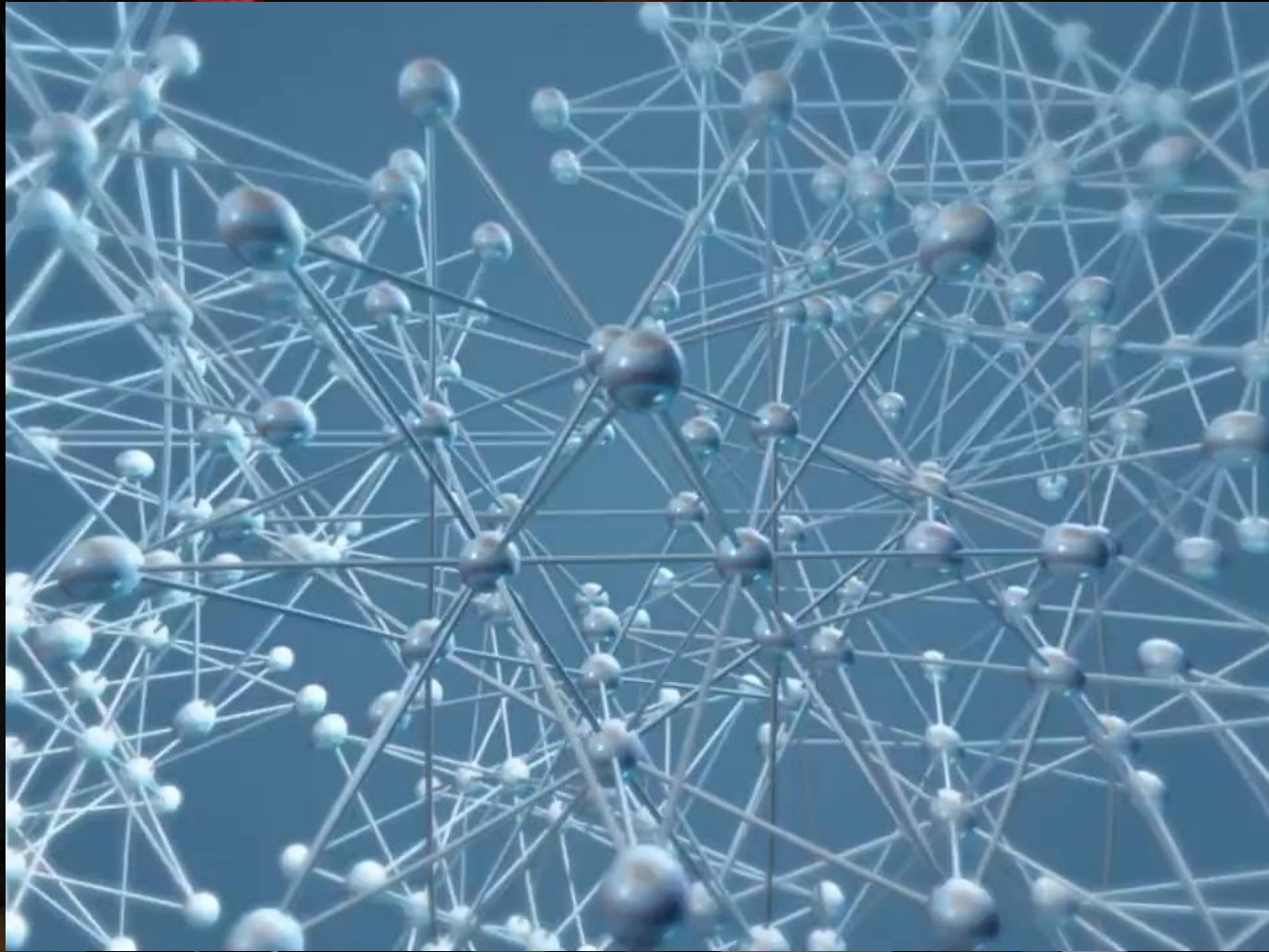


Dipole Moment



A descriptor easy to **compute/predict, measure, and manipulate**

# 创造的快乐



材料基因

CCTV 1

综合

新闻联播  
XINWEN LIANBO

中科大微尺度物质科学国家实验室教授 江俊  
材料数据是涉及到国家战略安全的

# 总结

---

- **新材料之突破，或将以今天陌生的方式到来.....**
- **物理、化学、材料大数据、人工智能的融合不可避免**



# 致谢

中国科学技术大学

罗毅

俞书宏

熊宇杰

张国庆

邹崇文

李任之

冯超

霍姚远



材库1.01

检索：[www.dcaiku.com](http://www.dcaiku.com)

上传：[www.mgbig.com](http://www.mgbig.com)

基金：中央组织部、国家科技部、国家基金委、中国科学院



**谢谢！**

**敬请批评指正！**