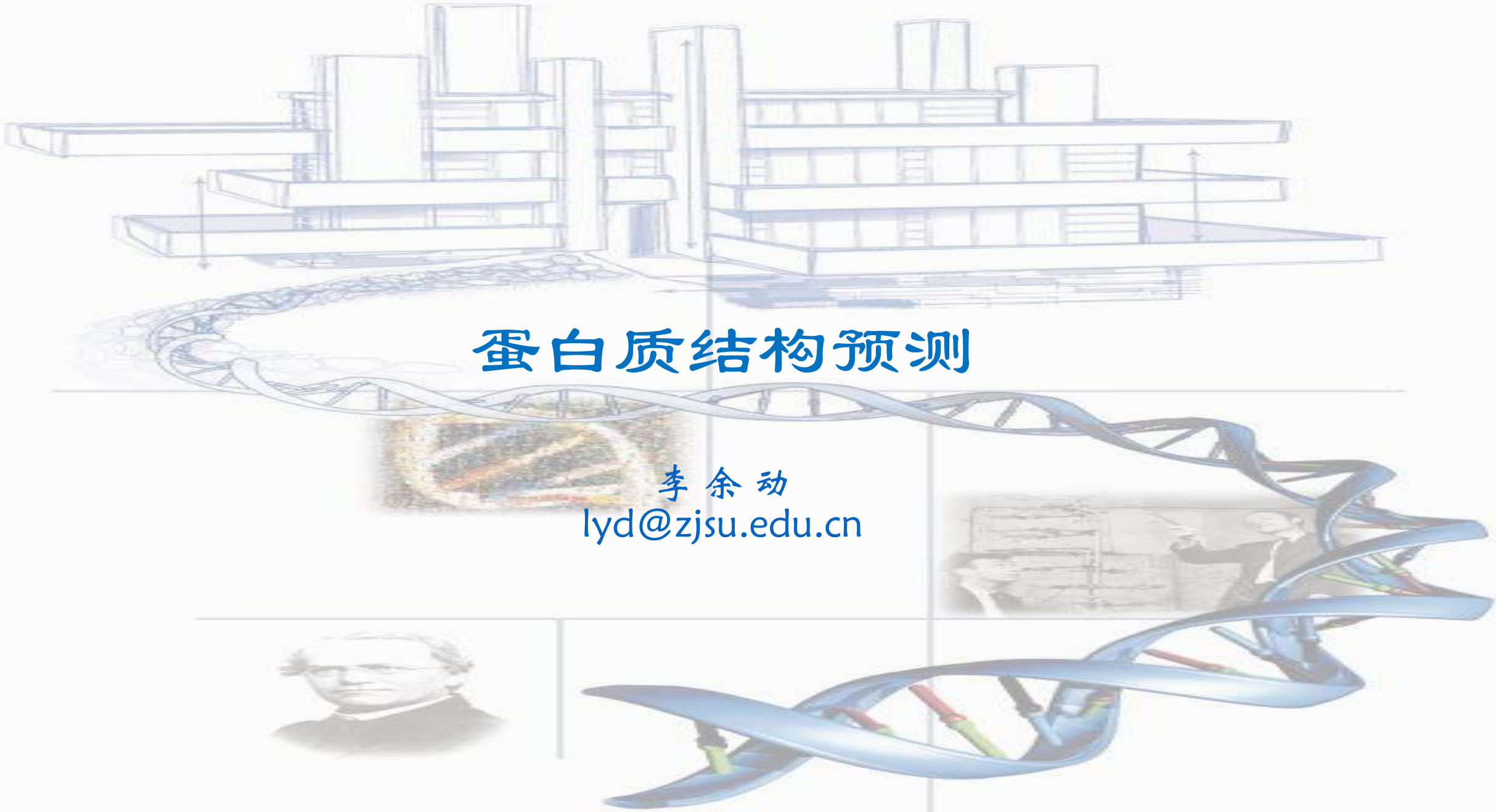


蛋白质结构预测

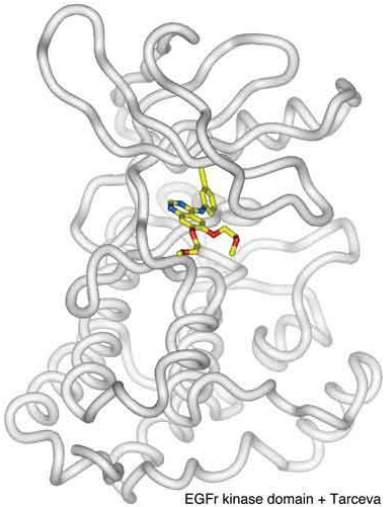
李 余 动
lyd@zjsu.edu.cn

目录

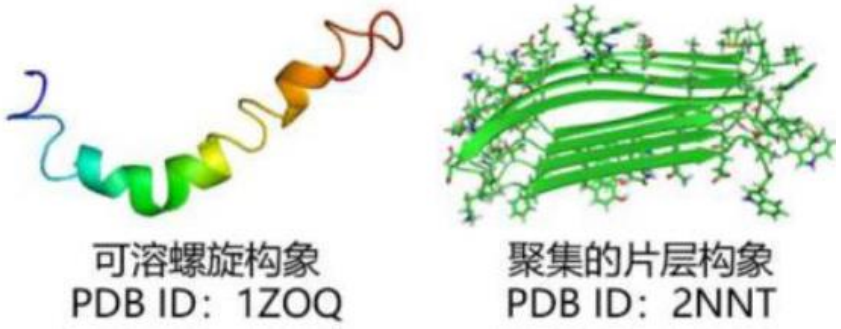
CONTENTS

1. 蛋白质的结构与功能

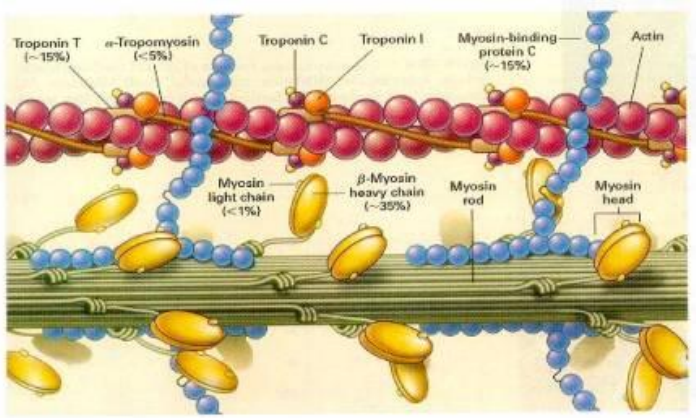
- 蛋白质是最重要的一类生物大分子，主要功能：
  - 酶(Enzyme)：各种代谢酶都是蛋白质。
  - 细胞结构(Structure)：胶原蛋白、角蛋白、细菌外壳蛋白等。
  - 运输(Transport): 血红蛋白（氧气）、各跨膜蛋白等。
  - 其它，如nutrition(蛋清蛋白), hormones、defense(免疫蛋白)等。



健康个体脑组织　　AD患者脑组织

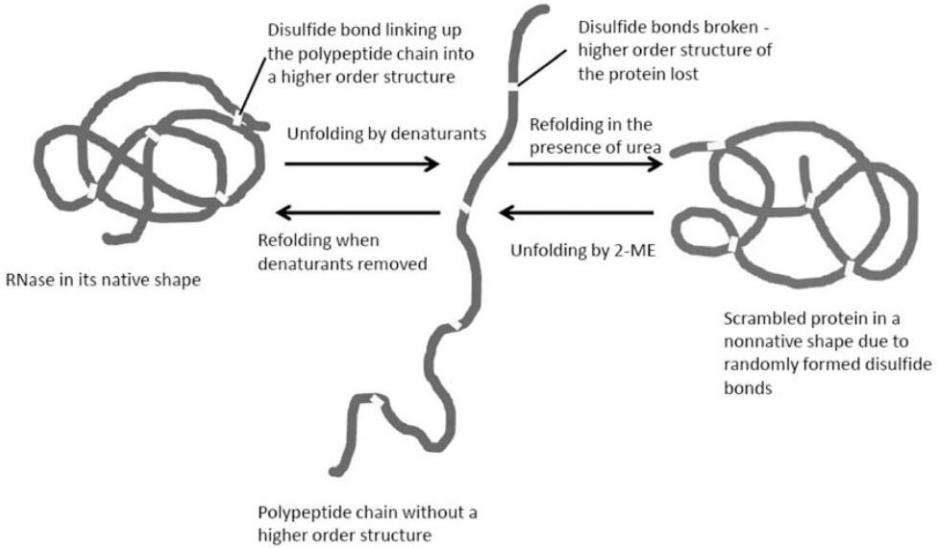


阿尔茨海默症(AD)的错误折叠淀粉样肽(beta-amyloid, Aβ)

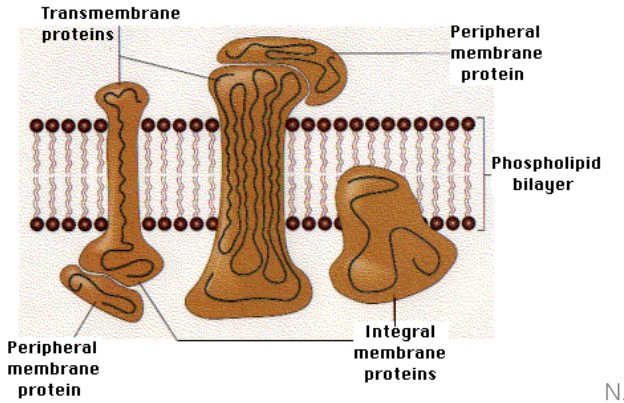


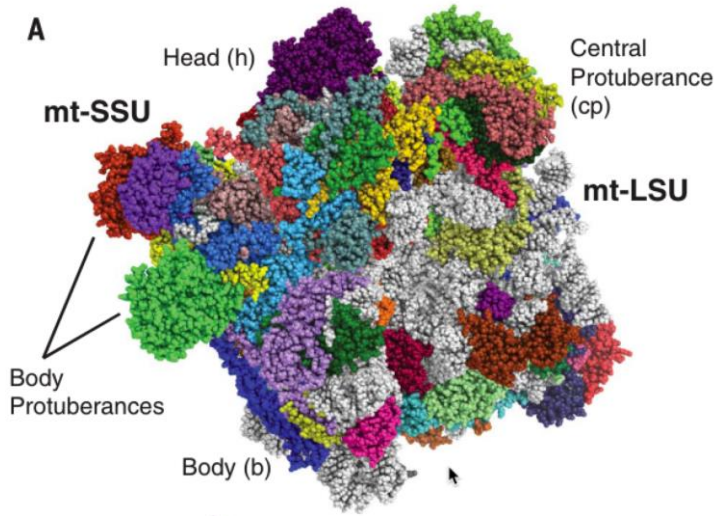
蛋白质结构决定功能

- 蛋白质的功能主要由三级结构所决定，蛋白质的三级结构主要由一级序列所决定
- Anfinsen原理认为氨基酸序列包含了形成热力学上最稳定的天然空间结构的全部信息。
  - 球蛋白 (Globular proteins): 疏水的内核 & 亲水的表面
  - 膜蛋白 (Membrane proteins): 特定的疏水氨基酸跨过膜内疏水区。
  - 无序性 (Intrinsically disordered): 许多蛋白质必须与其他蛋白质结合后才能够获得稳定的结构。



Anfinsen牛胰核糖核酸酶的变性实验

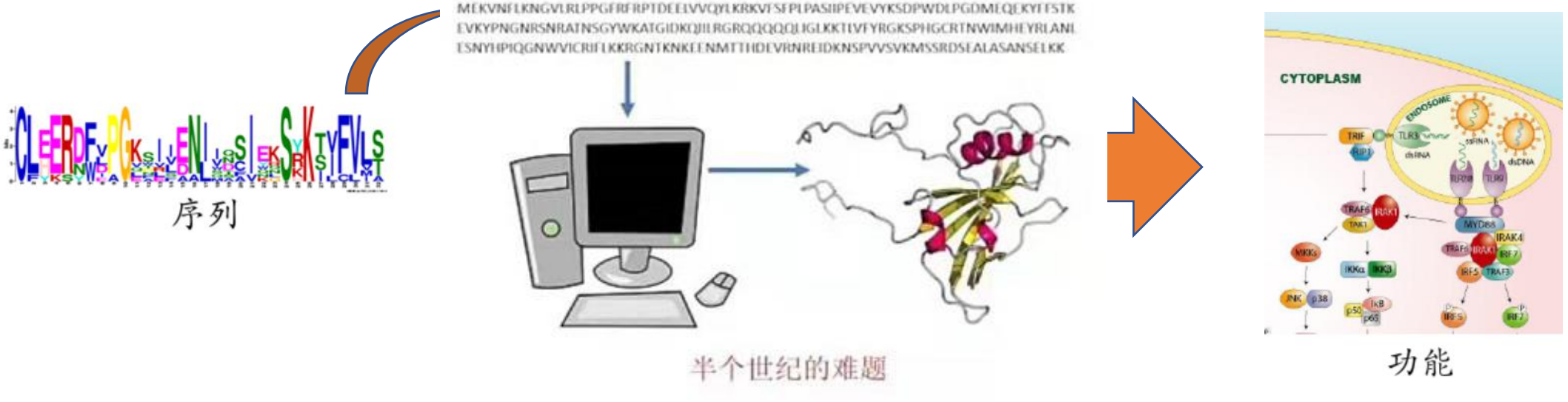




N. Desai, A. Brown, A. Amunts, and V. Ramakrishnan, "The structure of the yeast mitochondrial ribosome," *Science*, vol. 355, no. 6324, pp. 528–531, Feb. 2017.

1. 蛋白质的结构与功能

- 发掘蛋白质的结构特征是理解蛋白质生物功能的基础，对于生物医药领域的研究非常重要
- 获取精确的蛋白质三维空间结构是研究药物和靶标之间的相互作用进行药物设计的基础



预测蛋白质的结构和功能非常的困难!

蛋白质结构的四个基本层面

一级结构 Primary structure

氨基酸序列

二级结构 Secondary structure

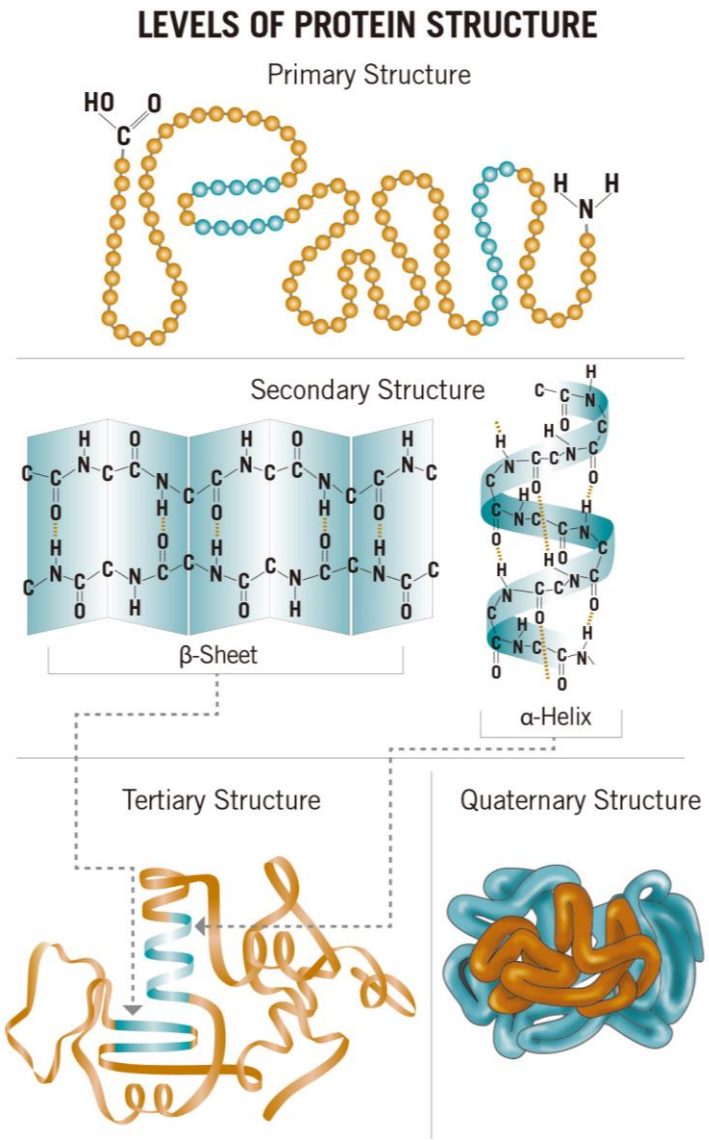
周期性的结构构象，如α螺旋和β折叠

三级结构 Tertiary structure

整条多肽链的三维空间结构

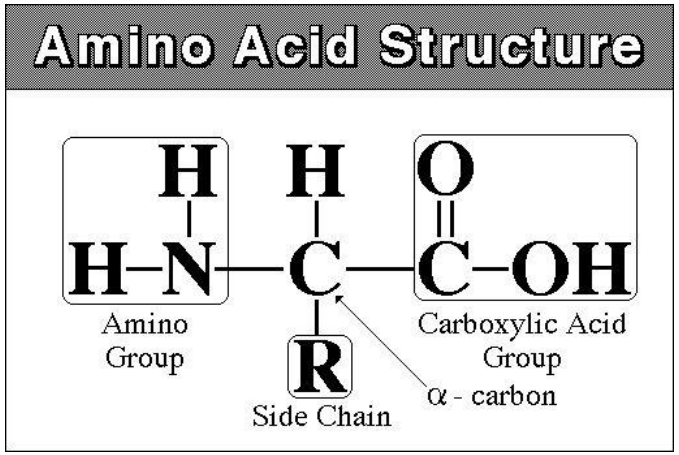
四级结构 Quaternary structure

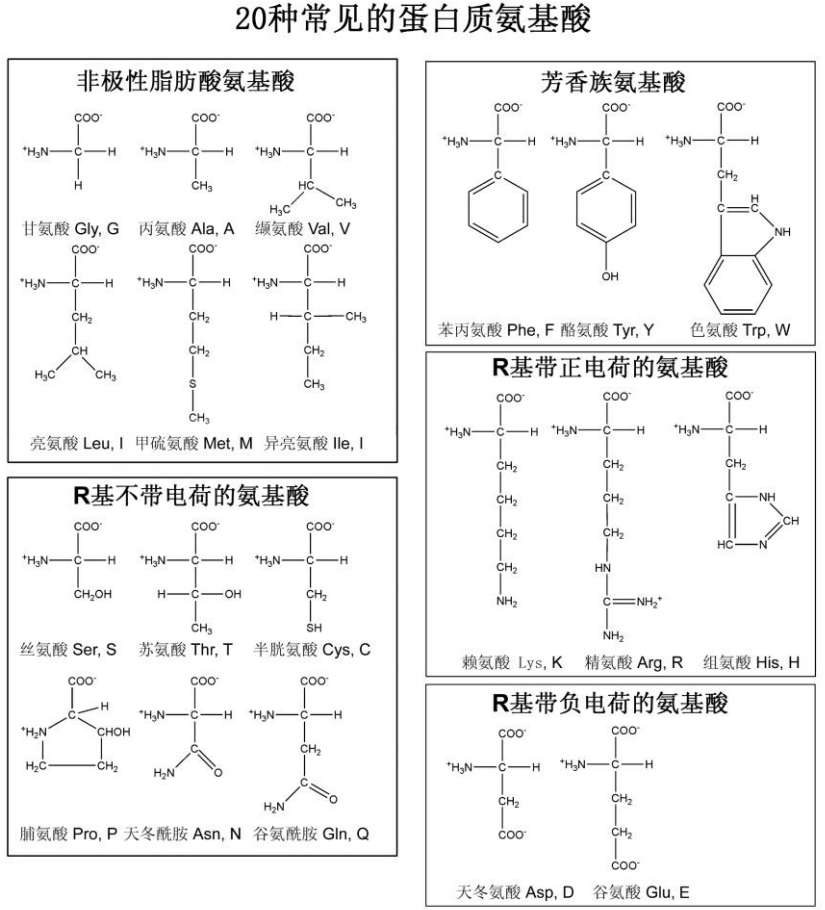
几个蛋白质分子（亚基）形成的复合体，如四聚体



氨基酸(amino acid)

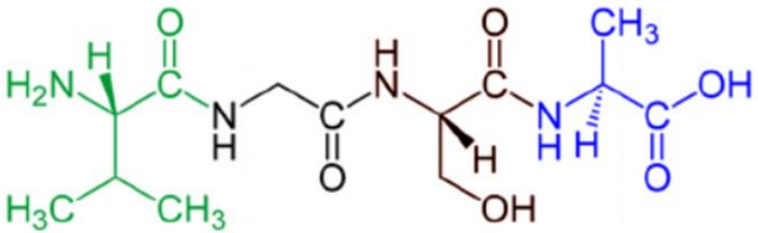
- 氨基酸是蛋白质的基本组成单位，天然蛋白质常用20种氨基酸
  - 一些细菌可会编码合成 selenocysteine (硒代半胱氨酸) 和 pyrolysine (吡咯赖氨酸)
- 蛋白质的性质由氨基酸R基团所决定



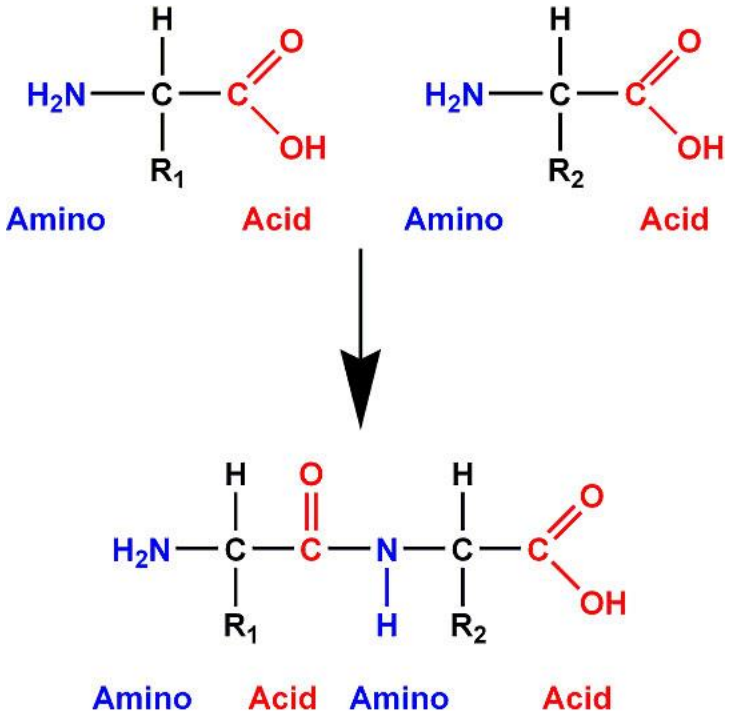


一级结构

- 一级结构(primary structure)指多肽链的氨基酸残基的排列顺序。
  - 氨基酸的线性序列，氨基酸残基之间以共价键连接
  - 肽键是指一个氨基酸的羧基与另一个氨基酸的氨基脱水缩合形成的酰胺键，其化学式为-CO-NH-。
  - 多肽中的氨基酸失水部分称氨基酸残基

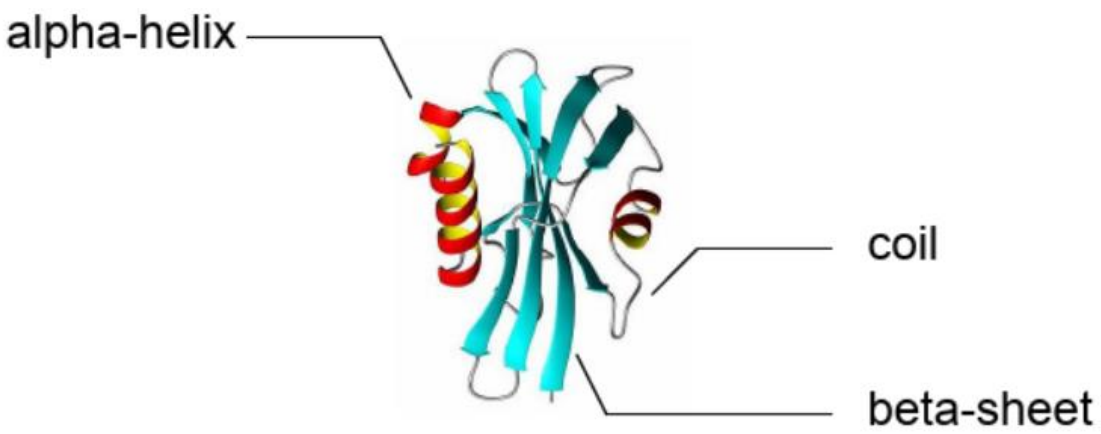


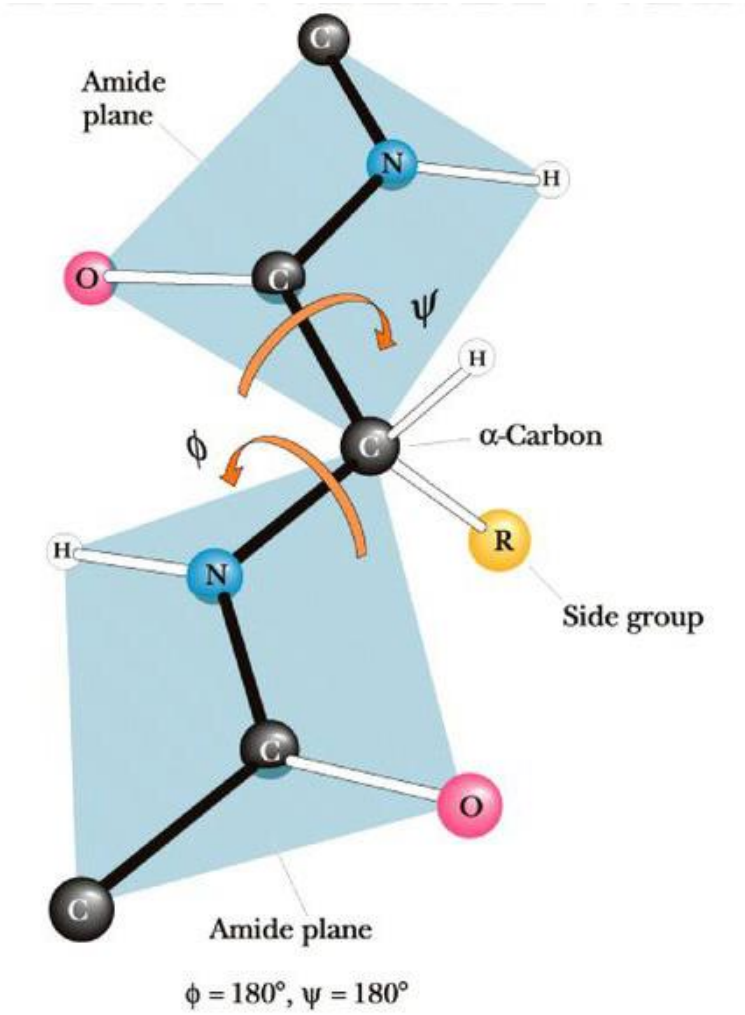
V G S A



二级结构

- 二级结构(secondary structure)指多肽链主链原子借助氢键沿一维方向排列成具有周期性的结构构象。
  - 氨基酸残基局部空间内的排列
  - 短程的、非共价的相互作用
  - 周期性的结构模式： α-helix, β-sheet, loops, coils



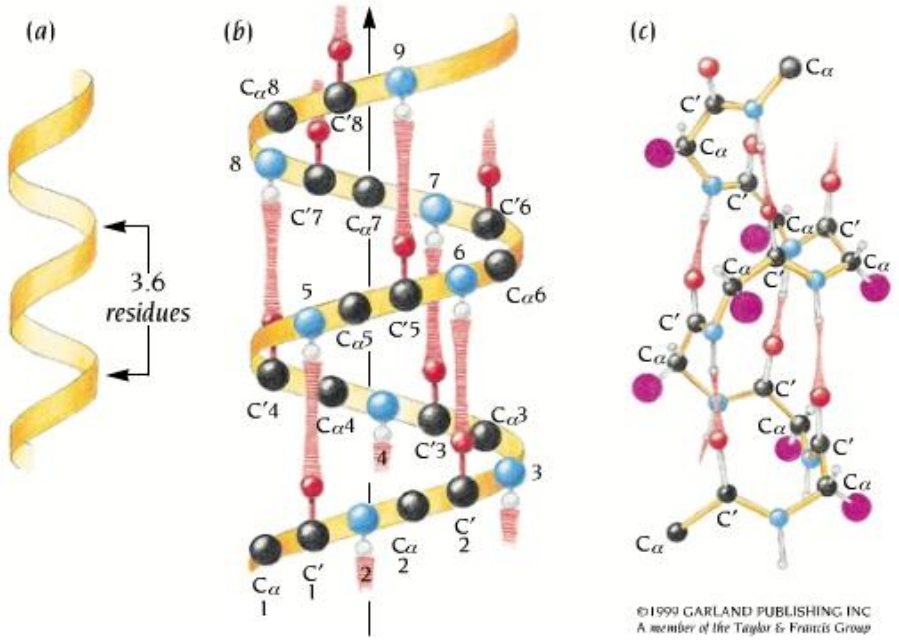


决定骨架的各区域会形成何种二级结构

α - helix（α螺旋）

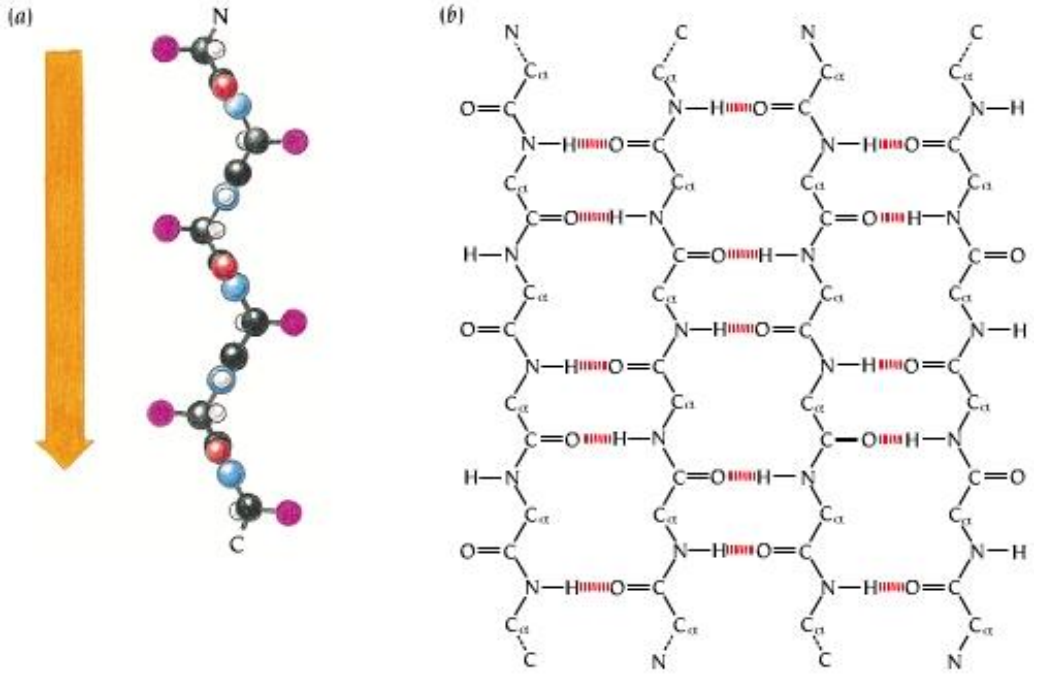
- 蛋白质中最多的二级结构，多肽主链围绕中心轴呈有规律的螺旋式上升。
- 平均长度：10个氨基酸残基 (10 A^0)
  - 长度范围：5-40aa
  - 每一圈：3.6个aa
  - 通过氢键 (~per 4aa) 稳定结构，氢键的方向与螺旋长轴基本平行
  - 通常在内核的表面，疏水残基向内，亲水残基向外

C = black
O = red
N = blue



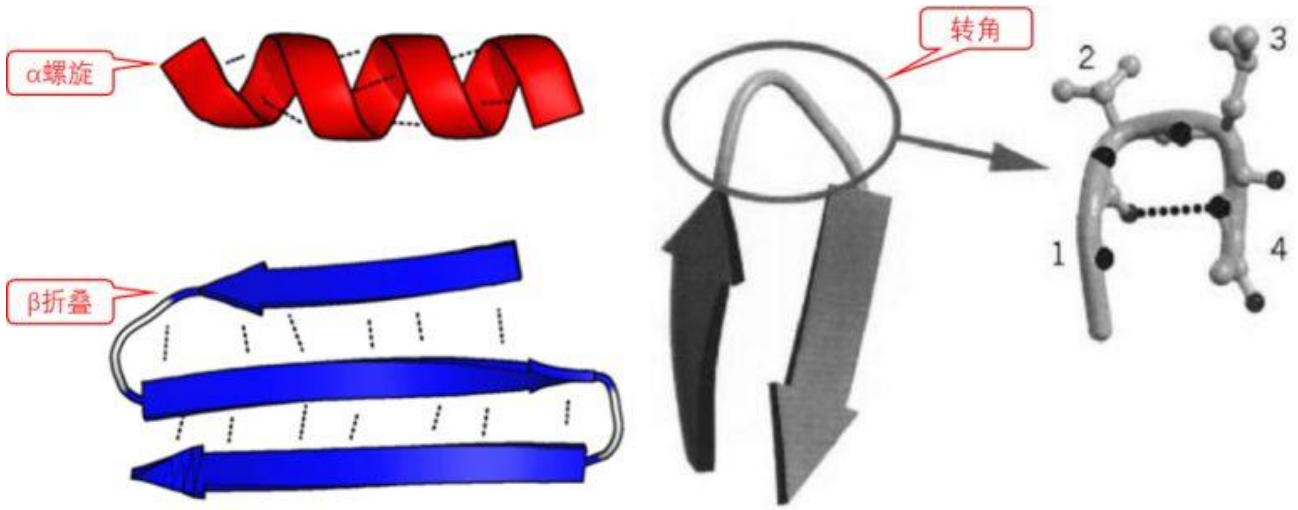
β-Sheets（ β折叠）

- 肽平面折叠成锯齿状，一般不单独出现，成对或多个出现
- β链通过氢键连接，相邻肽链主链的N-H键和C=O键之间形成有规则的氢键，稳定结构
- 相互作用的部分通过短的/长的loop连接
- 平行或反平行的β–sheet



Loops & β-turns （β转角）

- 连接α-helix或β-sheet，柔性好，构象变化余地大
- 通常由4个氨基酸残基构成，借1，4残基之间形成氢键，形成一个紧密的环，结构稳定
- 受点突变的影响小，带电荷、极性的氨基酸比例高
- 倾向成为活性位点，已经发现蛋白质的抗体识别、磷酸化、糖基化等位点经常出现在转角或紧靠转角处

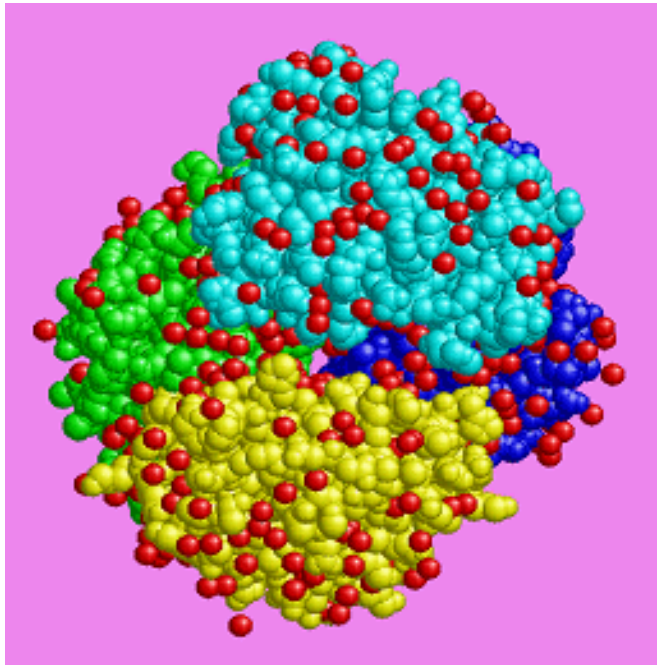


Random Coils(无规则卷曲)

- 主链骨架无规律盘绕的部分，泛指无法归入明确的二级结构
- 无序性 (Intrinsically disordered): 介导蛋白质-蛋白质之间的相互作用
- 酶的功能部位常常处于这种构象区域

三级结构和四级结构

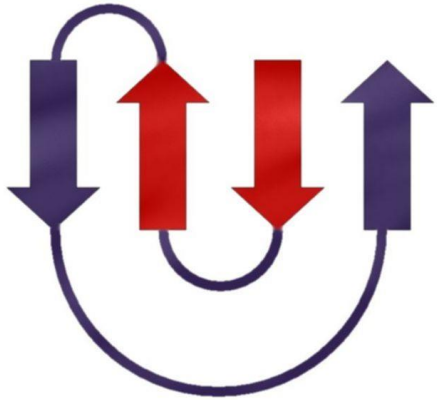
- 三级结构 (tertiary structure)
  - 肽链折叠成三维的空间结构
  - 二级结构在空间上的排布
  - 长程的、共价与非共价的相互作用
- 四级结构(quaternary structure)
  - 多个肽链在空间上的排布
  - 每一条多肽链都有其完整的三级结构，称为亚基（subunit）
    - ✓血红蛋白由4个亚基构成
  - 单体多肽不具备四级结构



血红蛋白的四级结构

超二级结构(supersecondary structure)

- 超二级结构是指相邻的二级结构单元组合形成的排列规律，结构可辨认的二级结构组合体，同时充当三级结构的构件(building block)
  - 基本形式有αα、ββ、βαβ
- Motifs（模体或基序）：超二级结构或二级结构的组合
  - the Greek key motif, composed of 4 beta sheet strands;
  - The TIM barrel, composed of 8 alpha helices and 8 beta sheet strands.

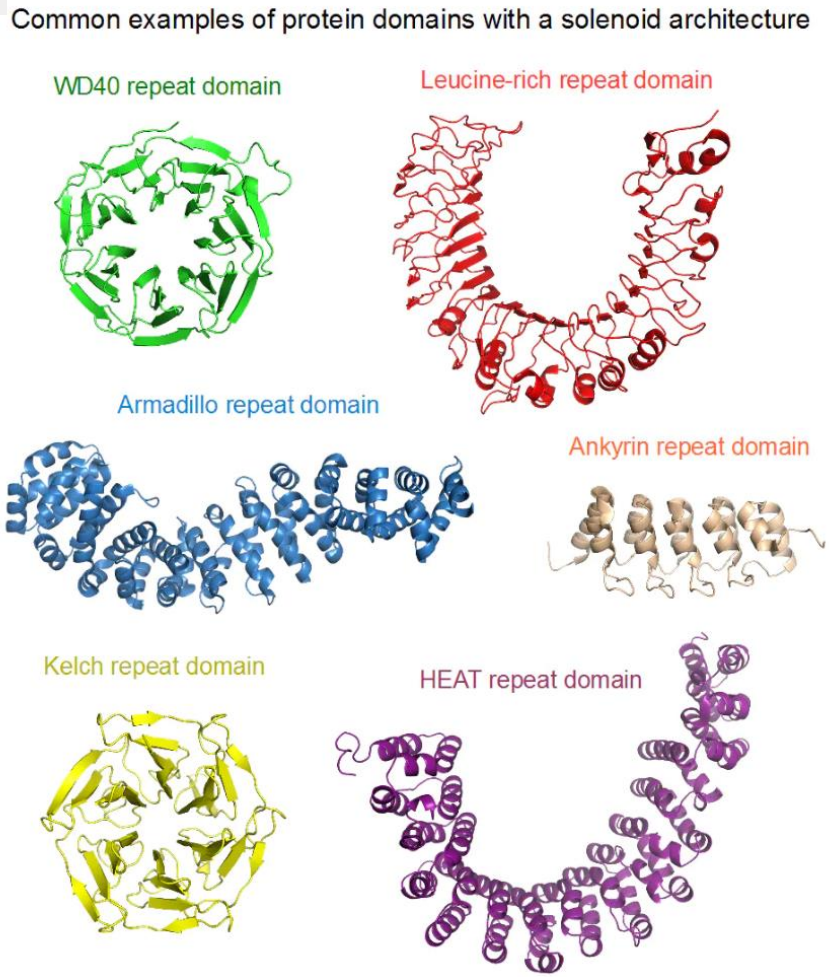


Motifs指DNA、蛋白质等生物大分子中的保守序列，介于二级和三级结构之间的另一种结构层次。

结构域(Domain)

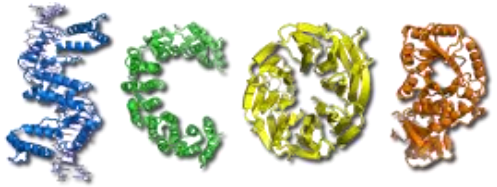
结构域:生物大分子中具有特异结构和独立功能的区域

- 蛋白质通常被认为是由一个或多个结构域组成的。
- 每个结构域都是一个相对独立的单元，通常由50~30个氨基酸残基组成，其间以柔性的铰链(hinge)相连，以便相对运动。
- 蛋白质分子中不同的结构域常由基因的不同外显子所编码，内含子通常位于DNA的结构域之间。



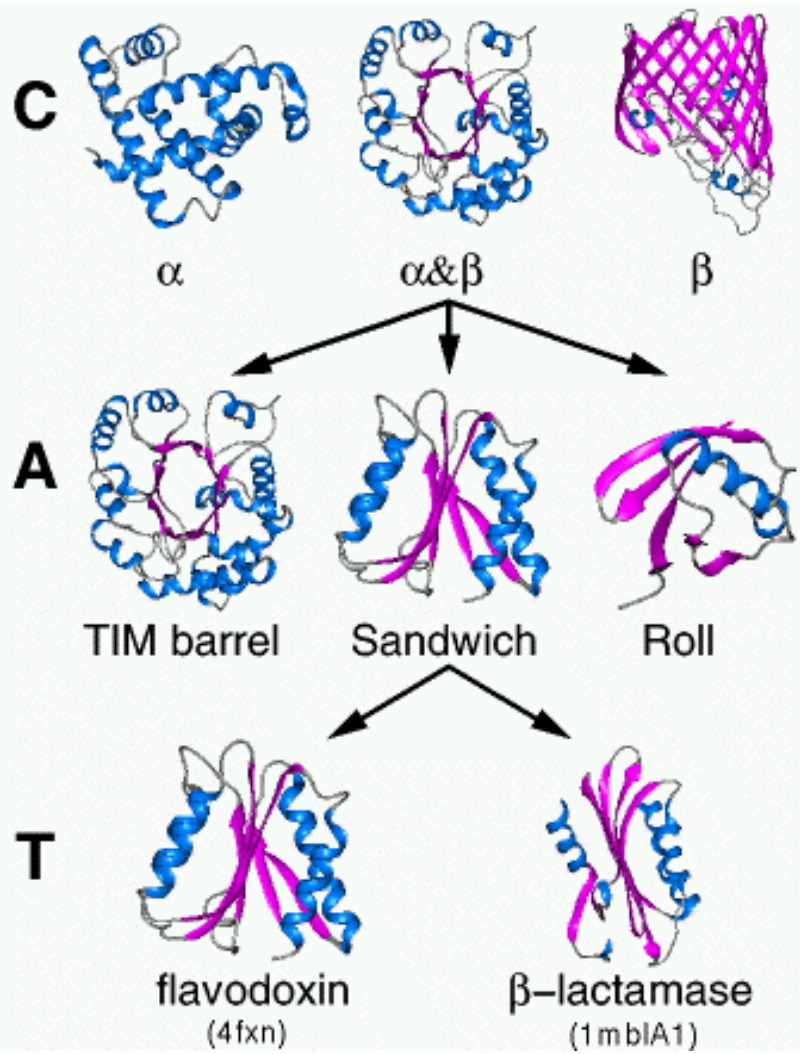
蛋白质结构分类数据库SCOP

- SCOP(Structural Classification of Proteins)数据库由英国医学研究委员会开发和维护；
  - http://scop2.mrc-lmb.cam.ac.uk/
- 提供关于PDB中已知结构的蛋白质之间结构和进化关系的详细描述。
- 可以按结构和进化关系对蛋白质分类，分类结果是一个具有层次结构的树，其主要的层次是家族、超家族和折叠：
  - 家族(family)：具有明显的进化关系，通常将序列同一性在30%以上的蛋白质归于同一家族；
  - 超家族(superfamily)：序列相似度低，但具有远源进化关系，具有共同的进化起源；
  - 折叠(fold)：无论是否有共同的进化起源，只要主要二级结构相似；
  - 结构类型(class): 结构类型相似，包括α螺旋、 β 折叠、α/ β 结构域、 α+ β 结构域等。
    - ✓ α / β 类蛋白质：由α螺旋和 β 折叠交替排列
    - ✓ α + β 类蛋白质：由分开的α螺旋和 β 折叠组成，其中 β 折叠一般为平行结构



蛋白质结构分类数据库CATH

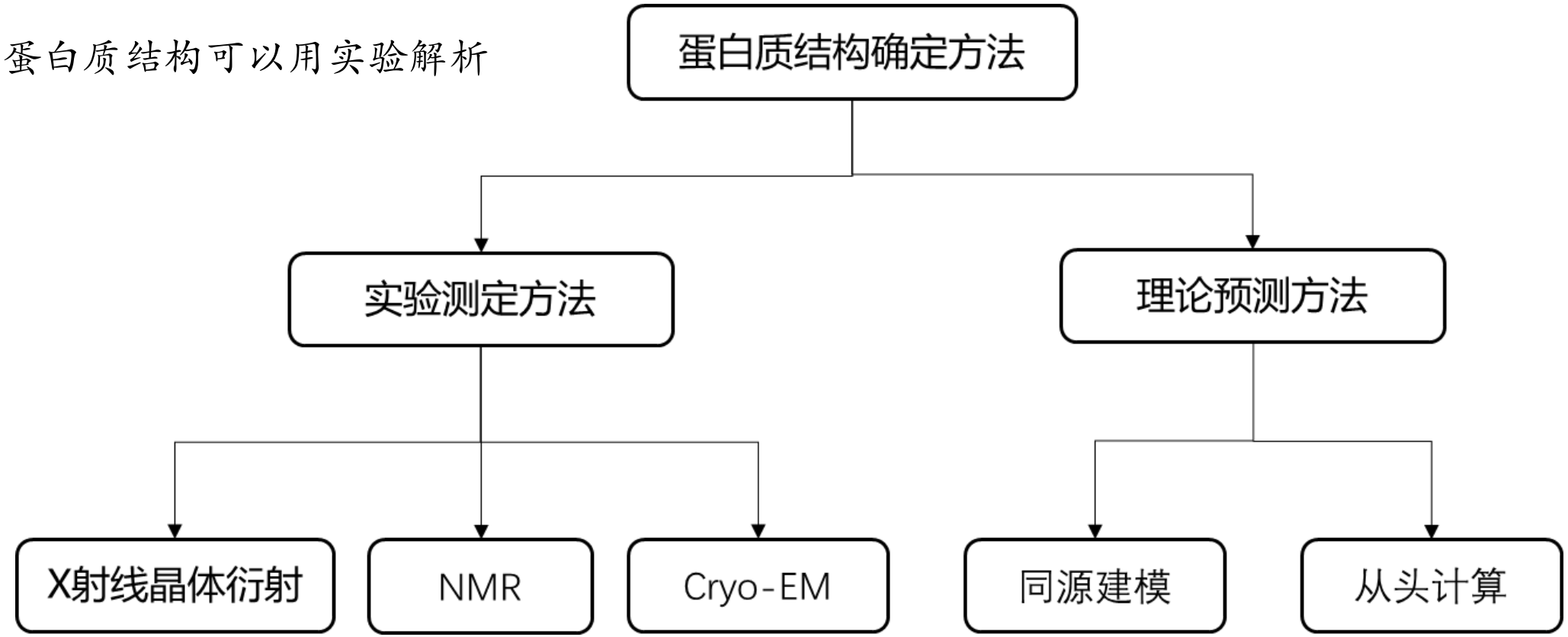
- CATH (Classification by Class, Architecture, Topology & Homology)数据库由英国伦敦大学UCL开发与维护；
  - http://cathdb.info/
- 不同于SCOP注重从蛋白质进化的角度来对蛋白质分类，CATH偏重于从结构监督对蛋白质分类，可分为：
  - Class（类型）：α、β、α&β
  - Architecture（构架）：超二级结构
  - Topology（拓扑结构）：二级结构形状与结构的关系
  - Homology(序列同源)：序列相似性



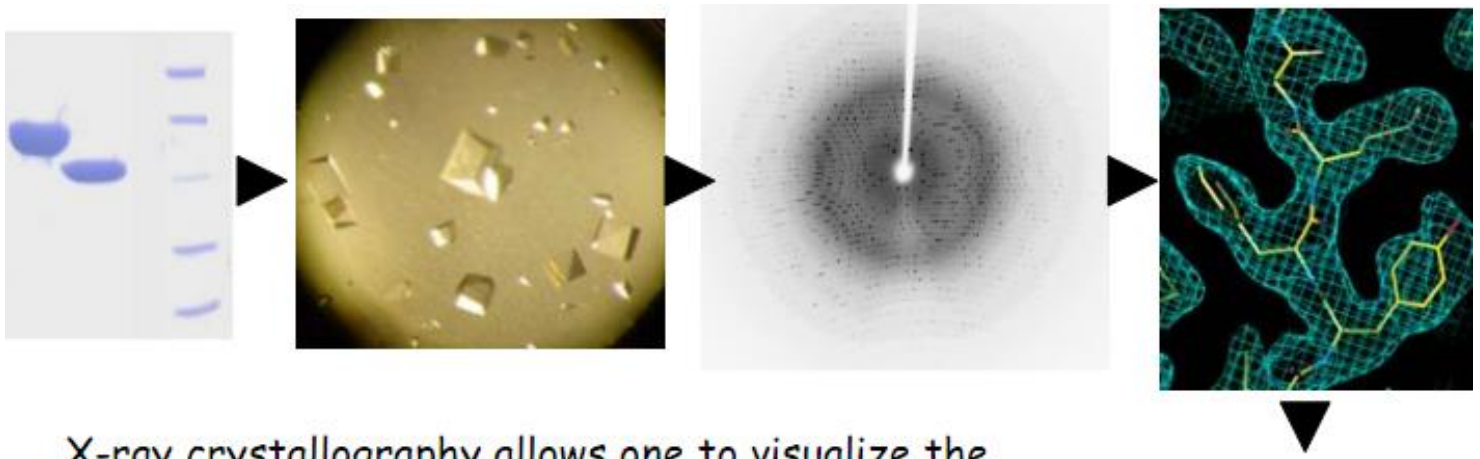
3.蛋白质三维结构测定方法

实验方法的缺点：

- 耗时长，几个月到几年
- 费用高
- 不是所有的蛋白质结构可以用实验解析



蛋白质三级结构测定方法

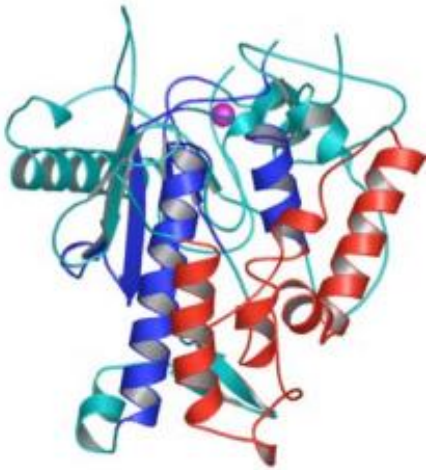


X-ray crystallography allows one to visualize the 3-Dimensional structure of a macromolecule.

Why would anyone want to do this?

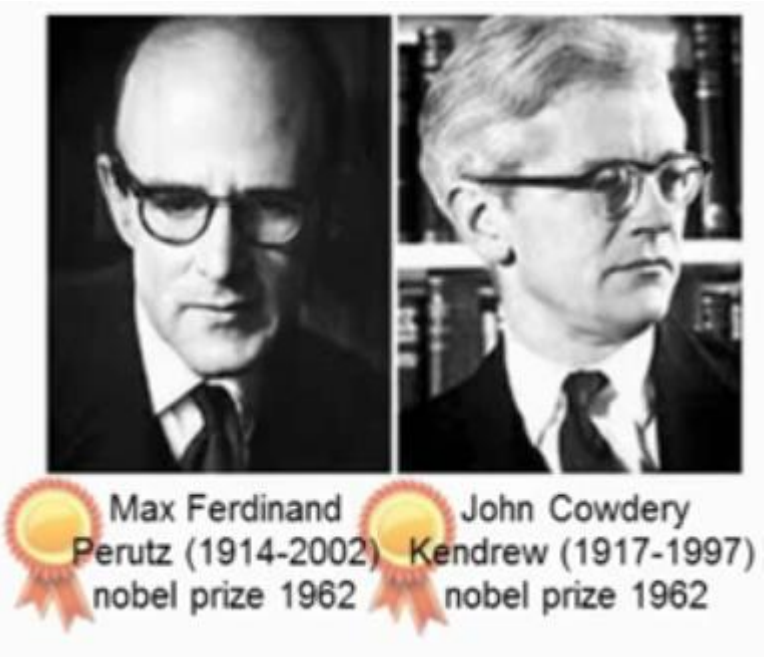
Understand the biochemical properties and the biological function

i.e. a reaction mechanism (if an enzyme)
or how a protein interacts in a complex.



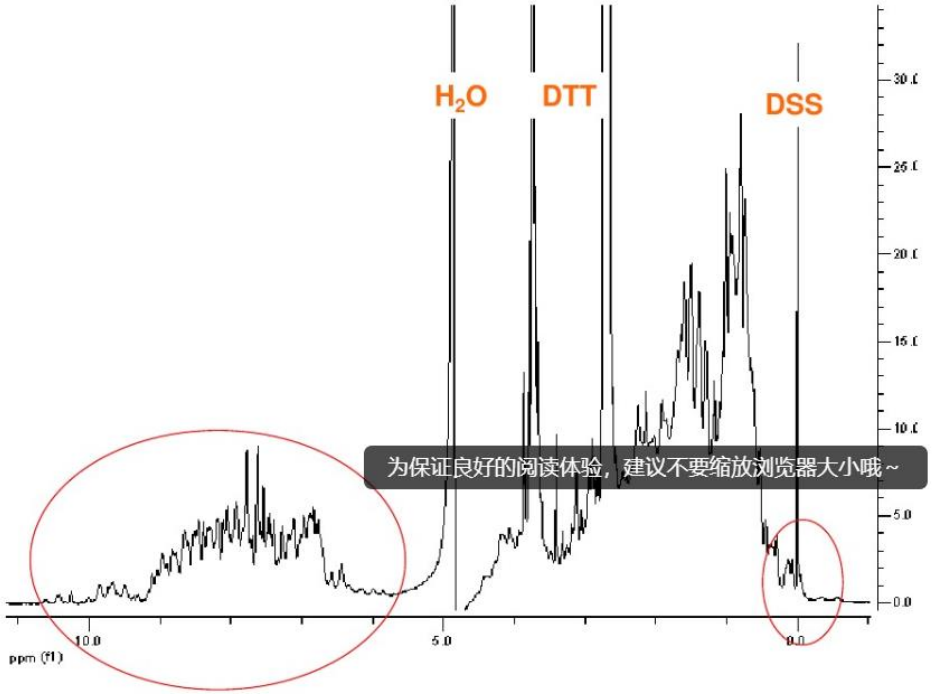
X射线晶体衍射图谱法

- 第一个蛋白质的三维空间结构是1958由Kendrew和Perutz博士用X-射线晶体衍射法测定



蛋白质三级结构测定方法

核磁共振法NMR



- 采集用于解析一个蛋白质所需要的全部图谱时间约1~2个月
- 限于分析长度不超过150个氨基酸残基的小蛋白质

蛋白质三级结构测定方法

X射线衍射法
X-ray Crystallography

核磁共振法

冷冻电子显微镜

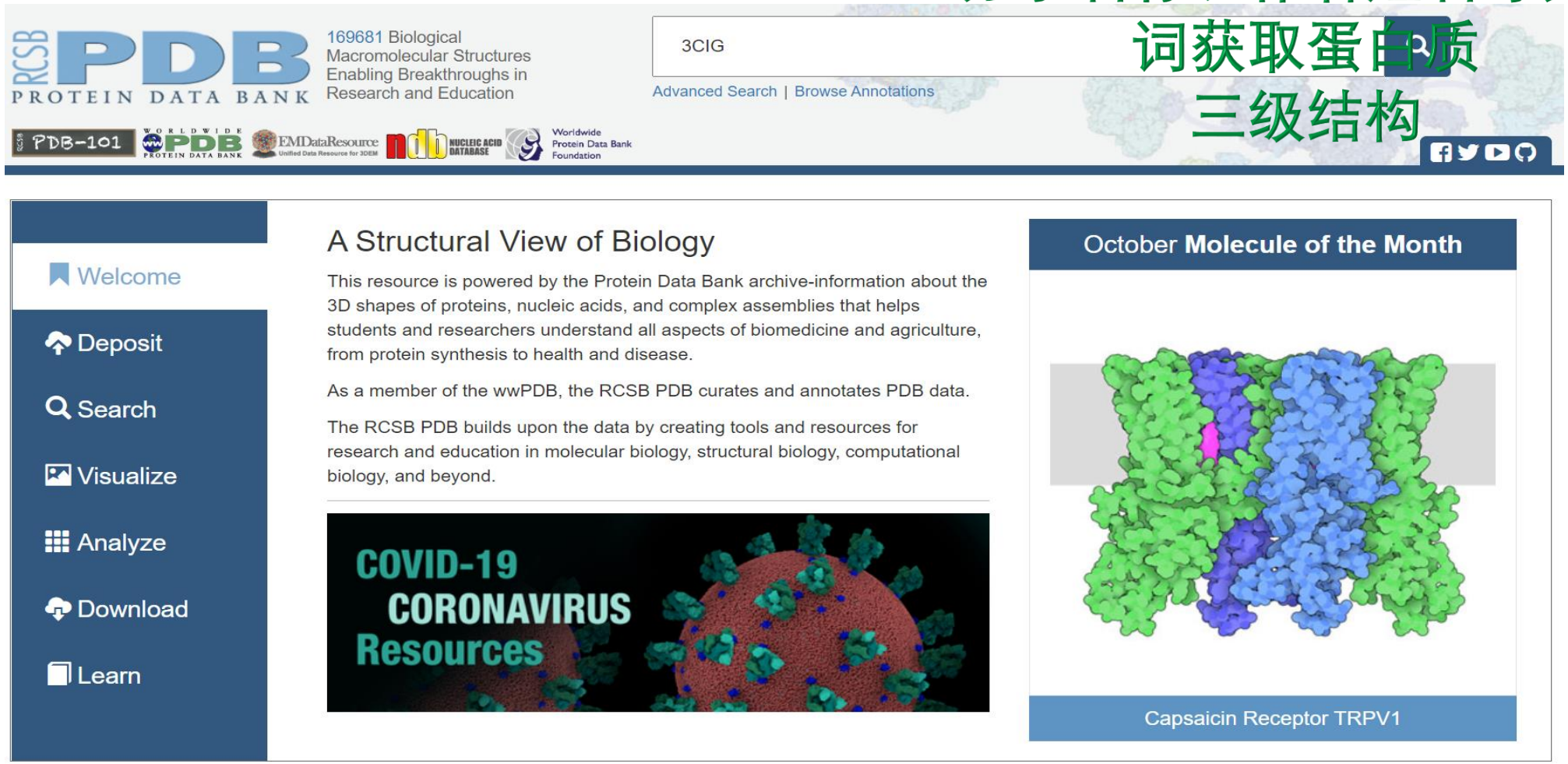
RCSB PDB
PROTEIN DATA BANK

- 无论用哪种方法测定的蛋白质三级结构，都要提交到PDB数据库中。
- 获取蛋白质三级结构最直接的方法——从PDB数据库中搜索

蛋白质结构数据库（Protein Data Bank, PDB）

PDB: https://www.rcsb.org/

搜索PDB ID、
分子名称、作者姓名等关键
词获取蛋白质
三级结构



- 目前最主要的收集生物大分子（包括蛋白质、核酸和糖）结构的数据库。

RCSB PDB PROTEIN DATA BANK

169963 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Enter search term(s)

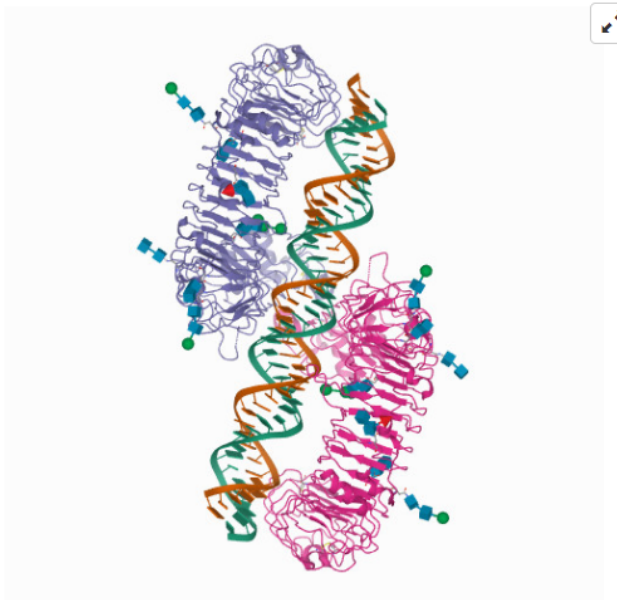
Advanced Search | Browse Annotations

查询结果

PDB-101 | wwPDB | EMDataResource | NDB NUCLEIC ACID DATABASE | Worldwide Protein Data Bank Foundation

Structure Summary | 3D View | Annotations | Experiment | Sequence

Biological Assembly 1



3D View: Structure | Electron Density | Ligand Interaction

Global Symmetry: Cyclic - C2 (3D View)
Global Stoichiometry: Homo 2-mer - A2

Find Similar Assemblies

由此可查询蛋白的详细信息

3CIY

Mouse Toll-like receptor 3 ectodomain complexed with doubl

DOI: 10.2210/pdb3CIY/pdb NDB: PR0335

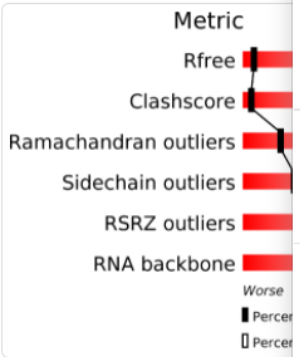
Classification: IMMUNE SYSTEM/RNA
Organism(s): Mus musculus
Expression System: Trichoplusia ni
Mutation(s): No

Deposited: 2008-03-12 Released: 2008-05-06
Deposition Author(s): Liu, L., Botos, I., Wang, Y., Leonard, J.N., Shiloach, J.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 3.41 Å
R-Value Free: 0.333
R-Value Work: 0.288
R-Value Observed: 0.288

wwPDB Validation



Display Files | Download Files

FASTA Sequence

PDB Format
PDB Format (gz)

PDBx/mmCIF Format
PDBx/mmCIF Format (gz)

PDBML/XML Format (gz)

Biological Assembly 1

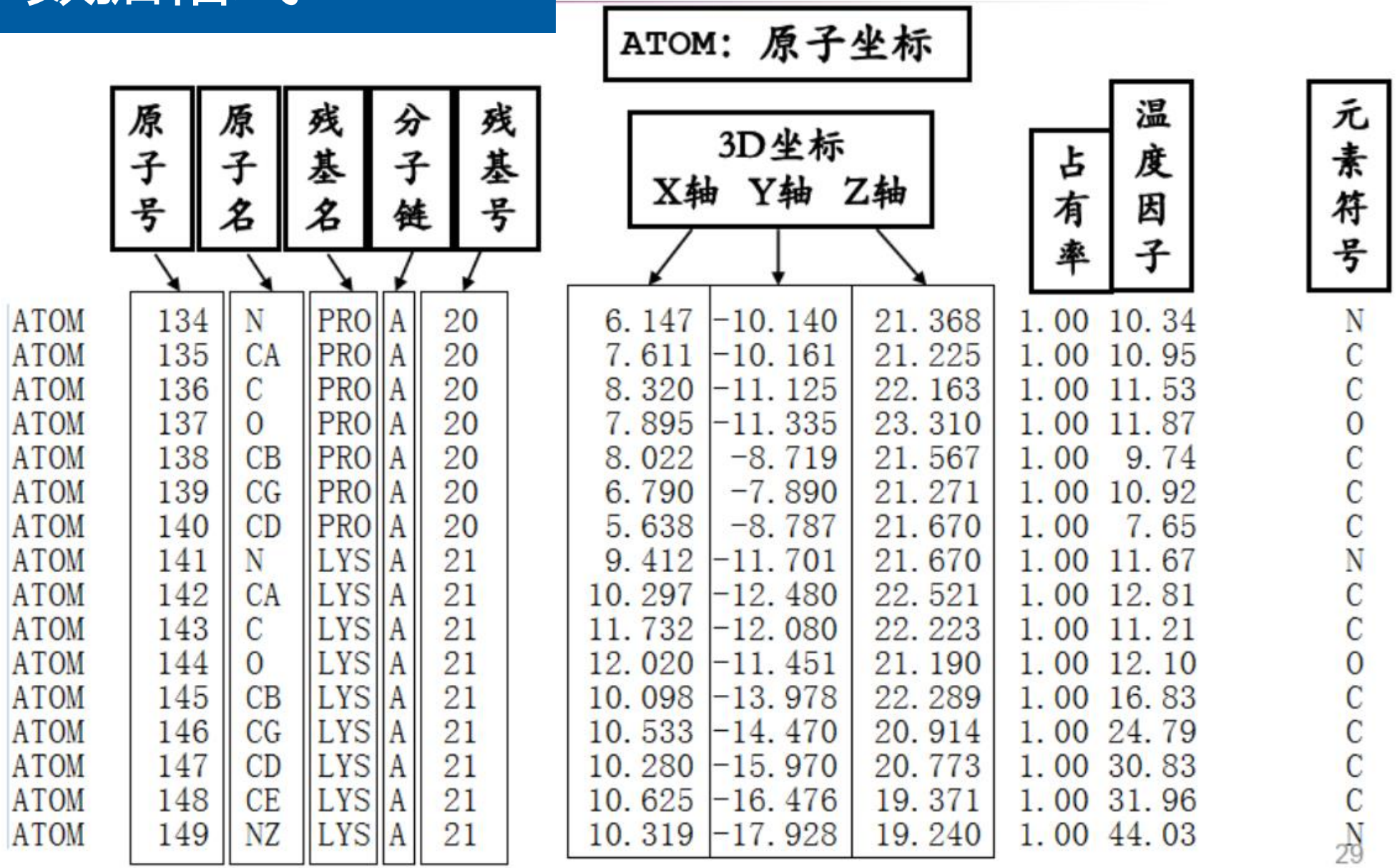
Structure Factors (CIF)
Structure Factors (CIF - gz)

Validation Full PDF
Validation XML

fo-fc Map (DSN6)
2fo-fc Map (DSN6)
Map Coefficients (MTZ format)

点此下载PDB文件

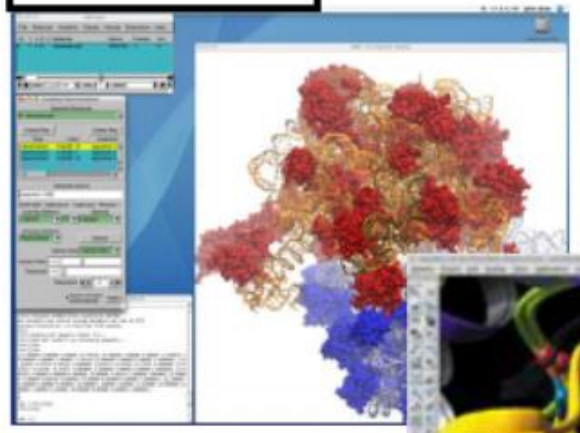
3.2 PDB数据格式



PDB文件存储蛋白质中每个原子的3D坐标

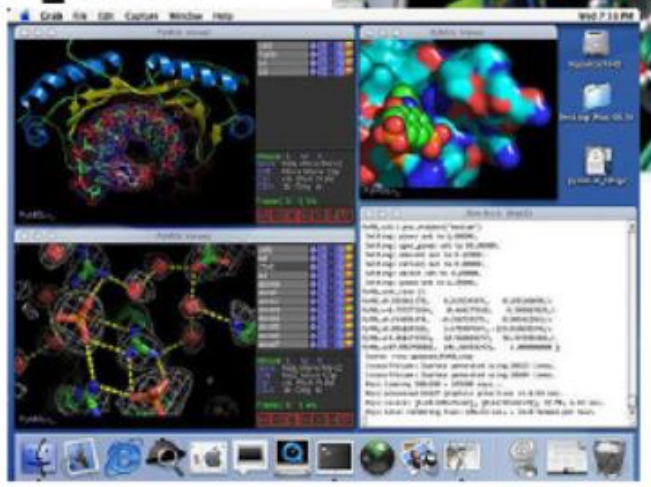
图形软件

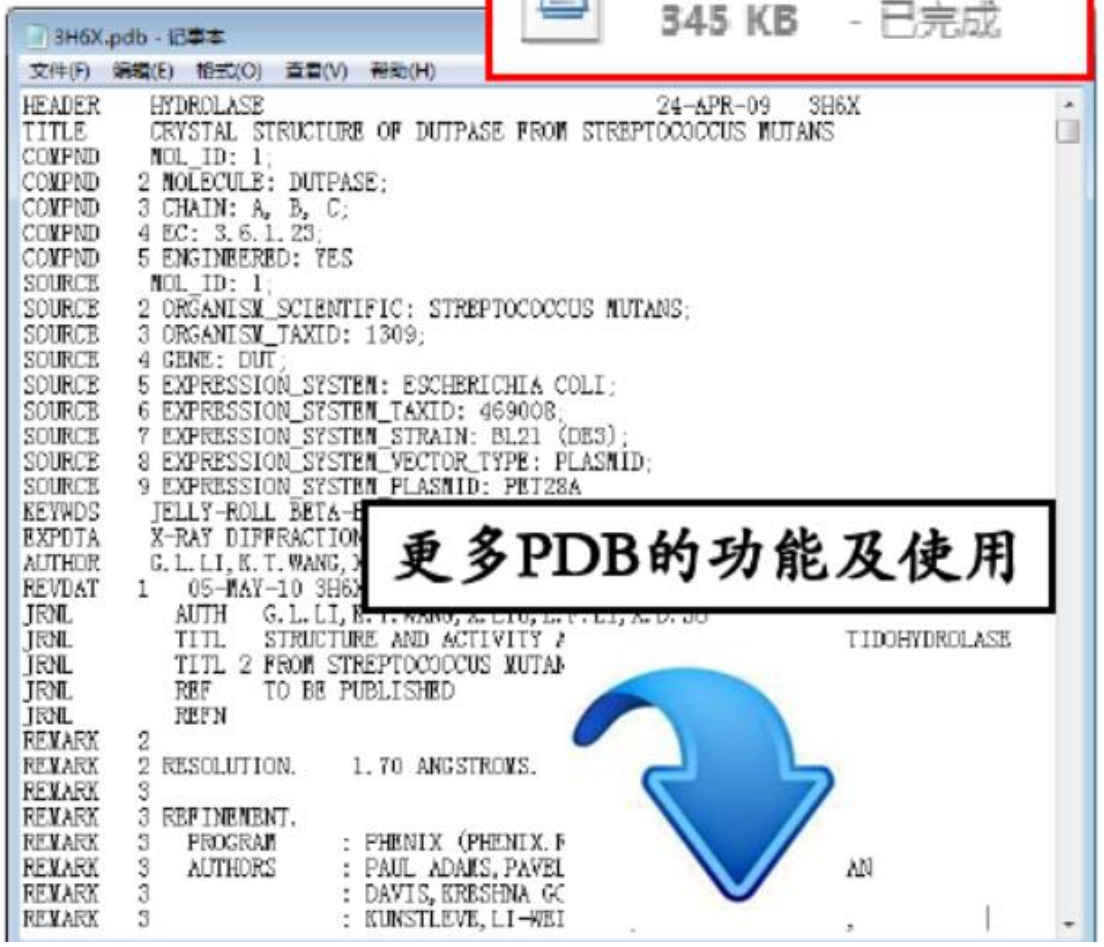
VMD



Maestro

Pymol



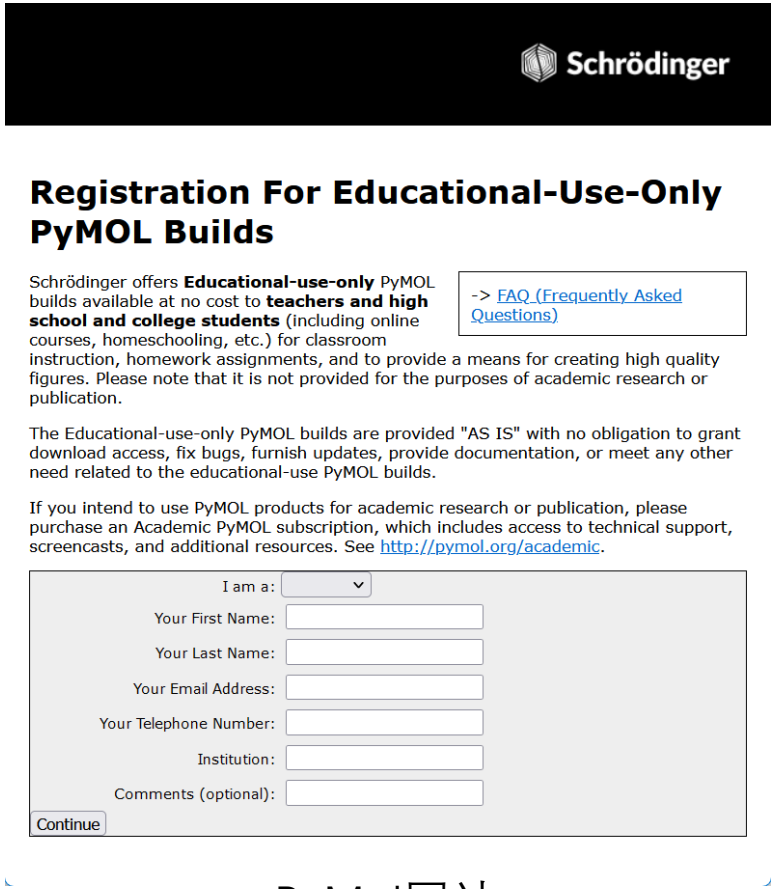


更多PDB的功能及使用

使用图形软件进行更多蛋白质的结构分析

4. 利用PyMol查看蛋白质三维结构

- **PyMOL软件下载**
  - PyMOL教育版软件下载地址：https://pymol.org/educational
  - **按要求注册后下载使用**



PyMol网站

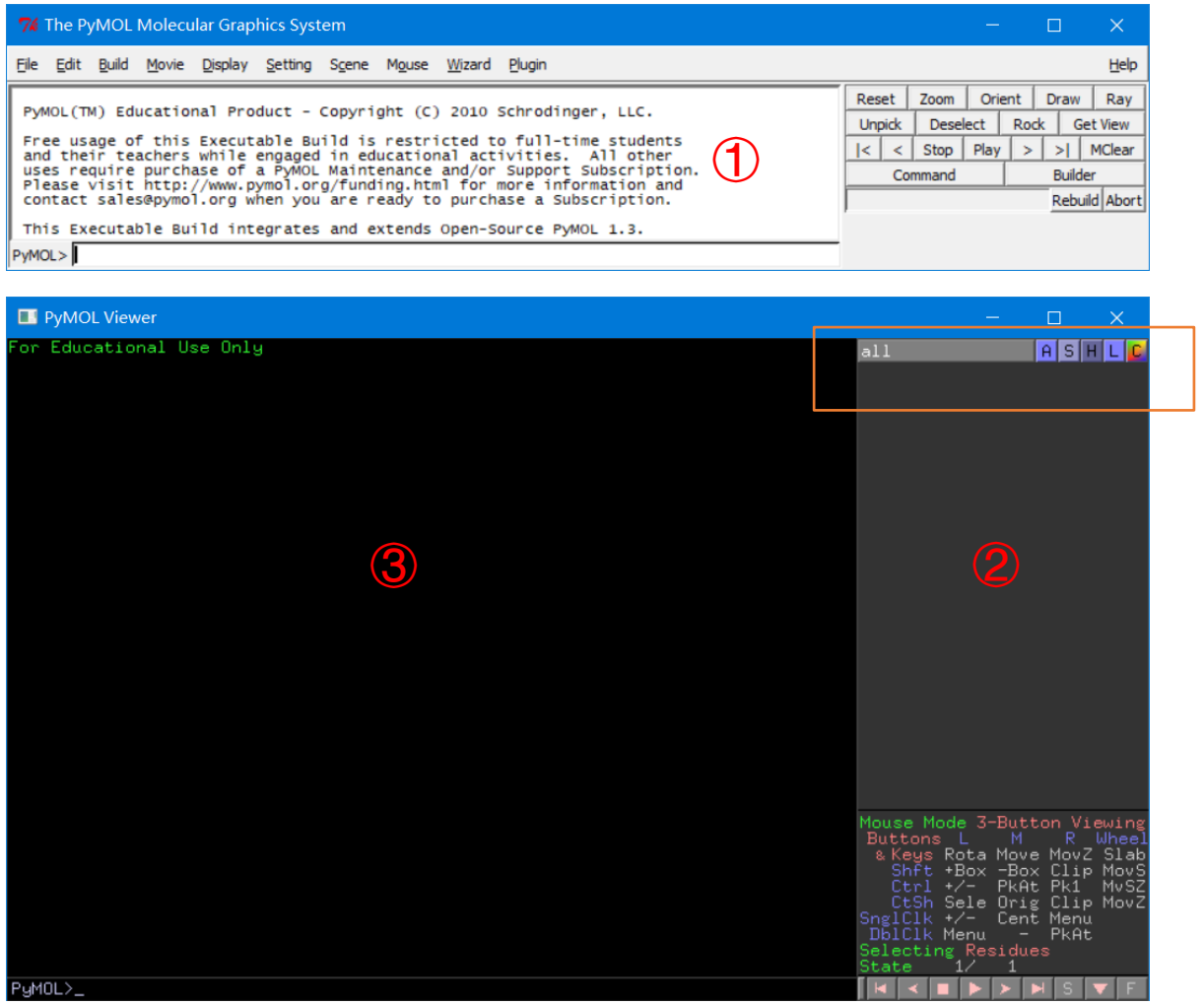
4. 利用PyMol查看蛋白质三维结构

- **PyMOL软件界面由三个模块组成：**

①外部窗口(external GUI)，包括菜单栏、信息显示框和命令行。

②内部窗口(internal GUI)，可以选定对象并进行操作，对象名称旁边有5个操作按键，分别为A(Actions), S(Show), H(Hide), L(Label), C(Color)，可实现各种操作。

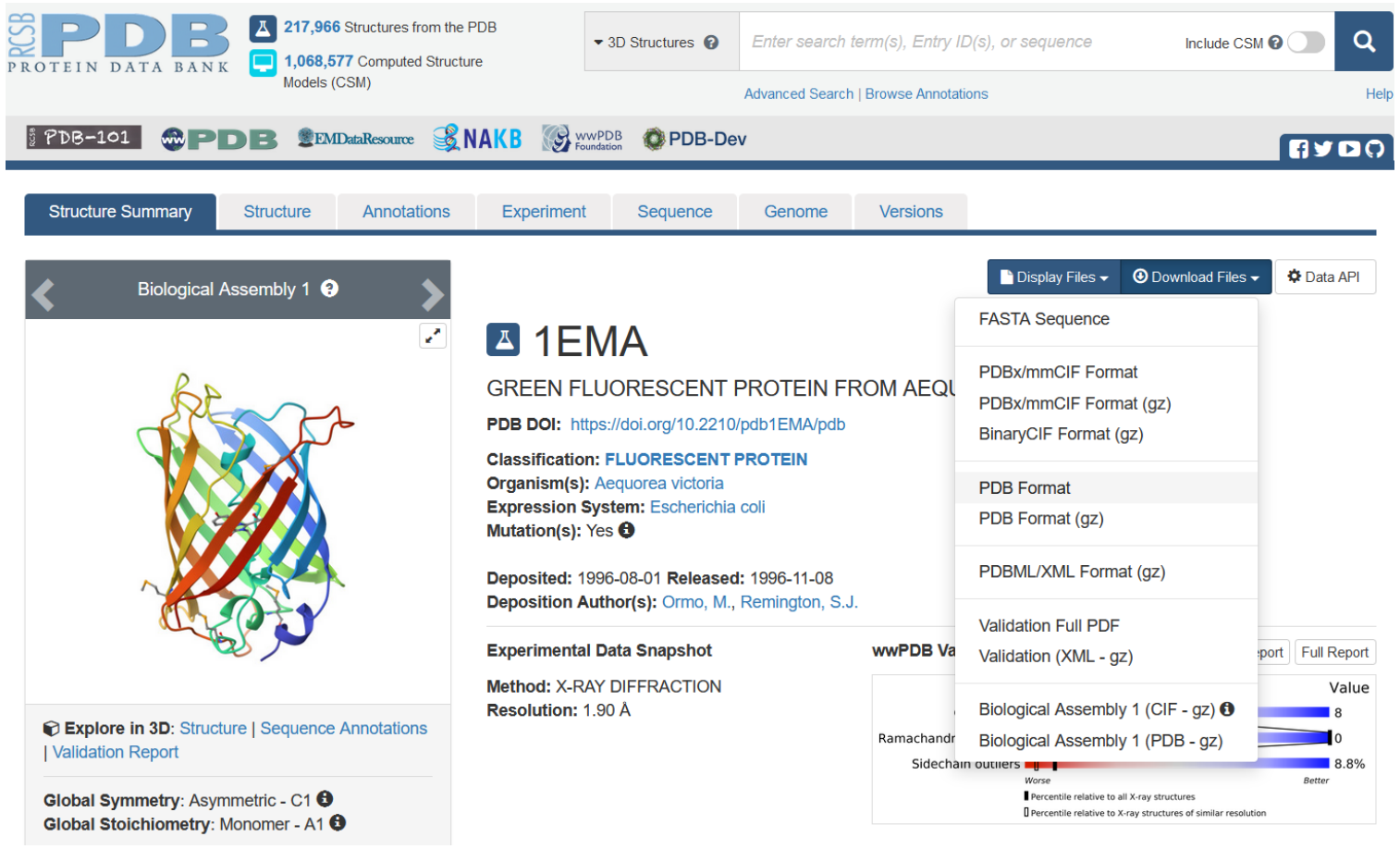
③图像浏览区，可显示结构图像。



PyMol图形界面

4. 利用PyMol查看蛋白质三维结构

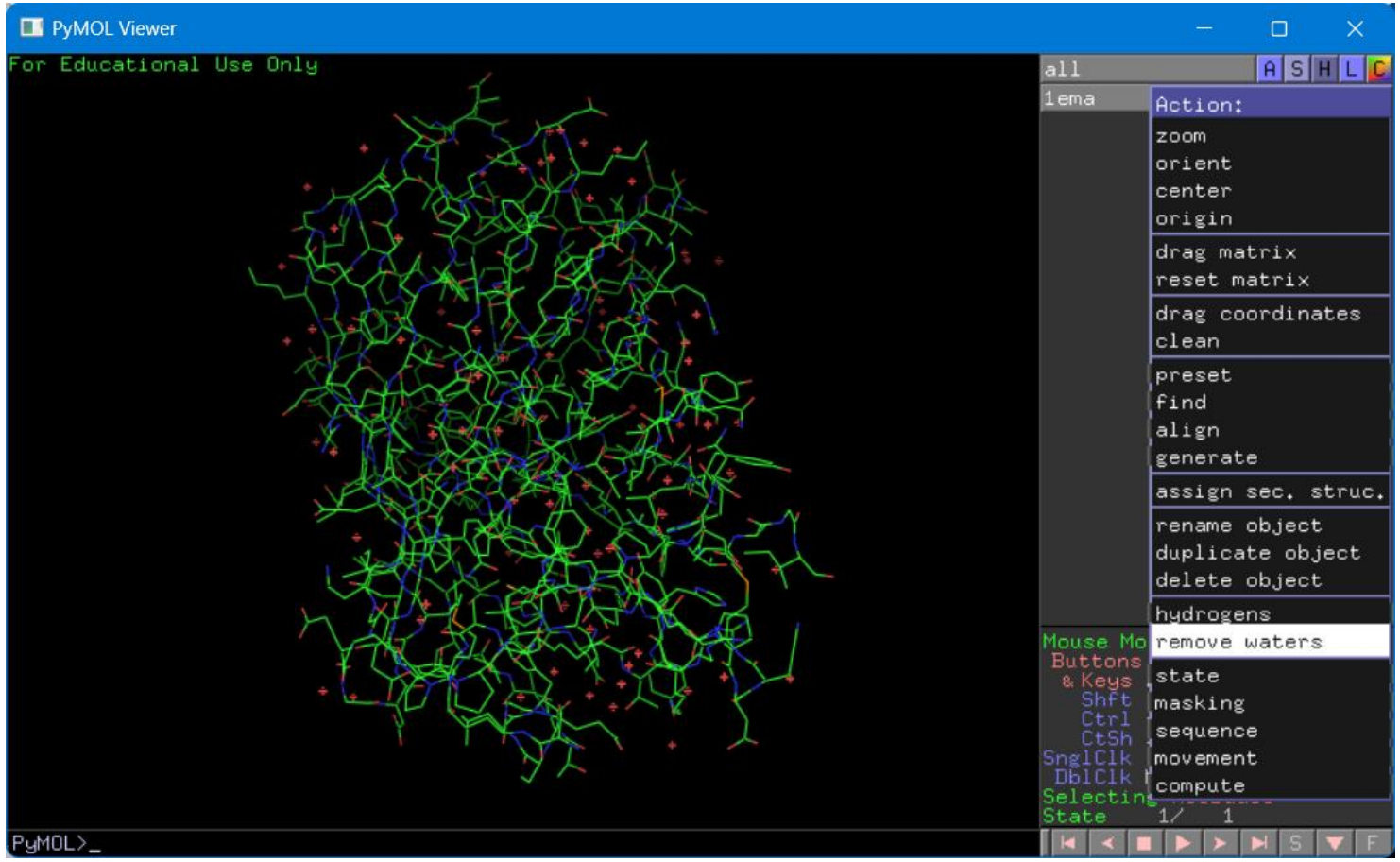
- **获得蛋白质结构数据文件**
  - 在PDB网站下载GFP蛋白(ID:1EMA)的pdb文件(1ema.pdb)；



PyMol图形界面

4. 利用PyMol查看蛋白质三维结构

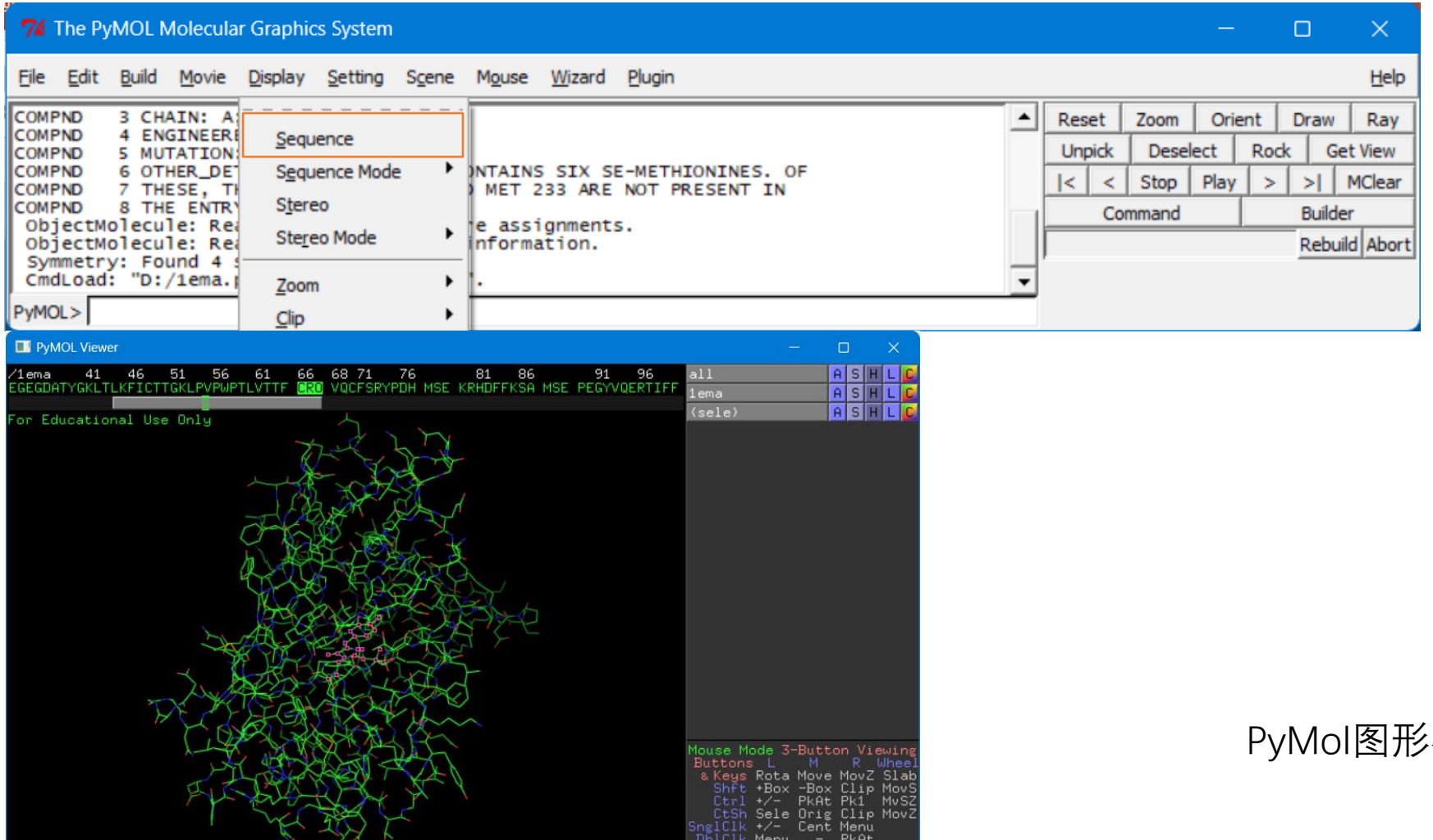
- **图形界面**
  - File → open, 选择“1ema.pdb”，打开后的GFP蛋白以线框模型(wireframe)显示;
  - 单击“all”图层的右边菜单A→"remove waters"，可去除水分子。



环绕蛋白质周围的水分子是蛋白质晶体结构解析时必然存在的分子，但它们并不是蛋白质本身的一部分，需要去掉水分子，以便更清晰地观察蛋白质的结构。

4. 利用PyMol查看蛋白质三维结构

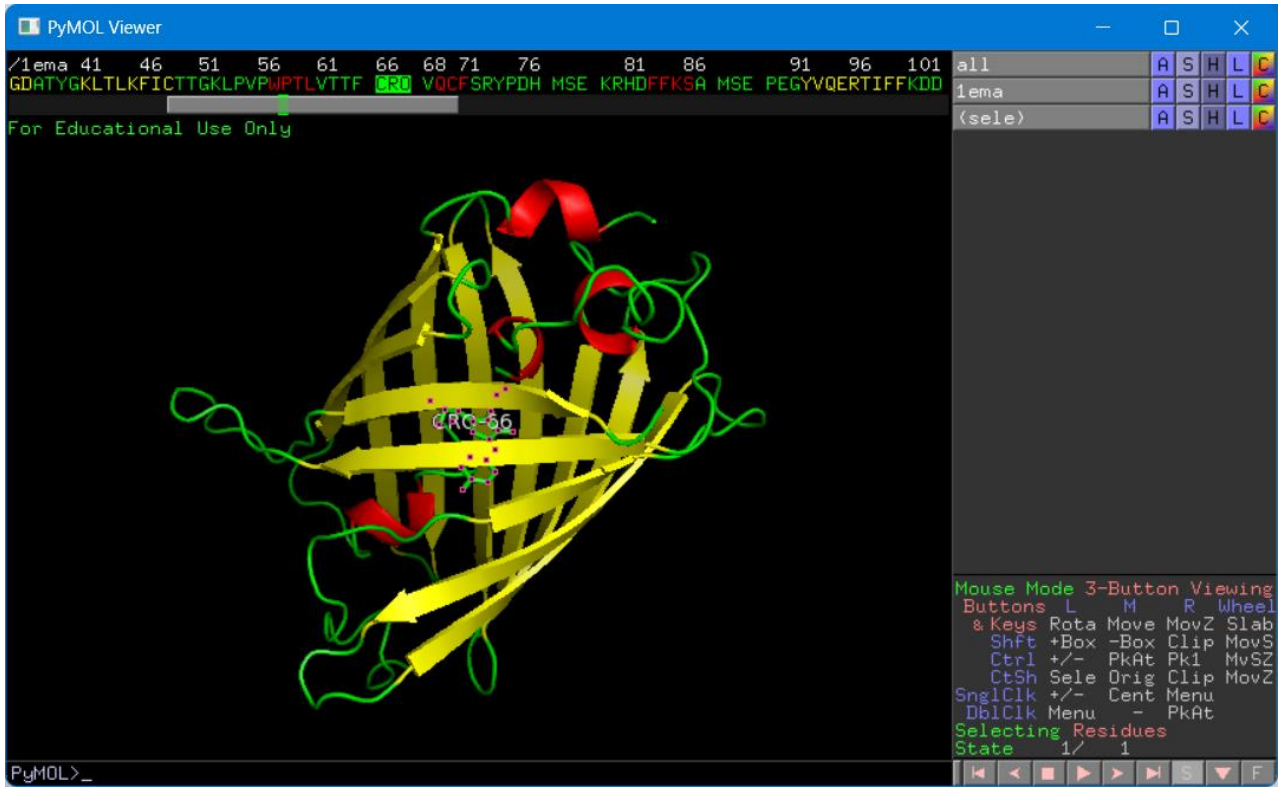
- 点击菜单Display →Sequence，可显示蛋白质序列，方便定位特定残基。
- 在序列上选择某个残基或一段序列，会出现图层(sele)。这里用鼠标点击序列66位的CRO，即荧光基团序列Ser65-Tyr66-Gly67。



PyMol图形界面

4. 利用PyMol查看蛋白质三维结构

- 点击1ema图层旁边的S→as→cartoon，以卡通模型显示，其他模型即被替换。
- 同时鼠标单击(sele)图层的S→sticks，以棍状模型显示选中的序列。
- 点击C→by ss→Helix Sheet Loop，以二级结构标记颜色。



PyMol图形界面

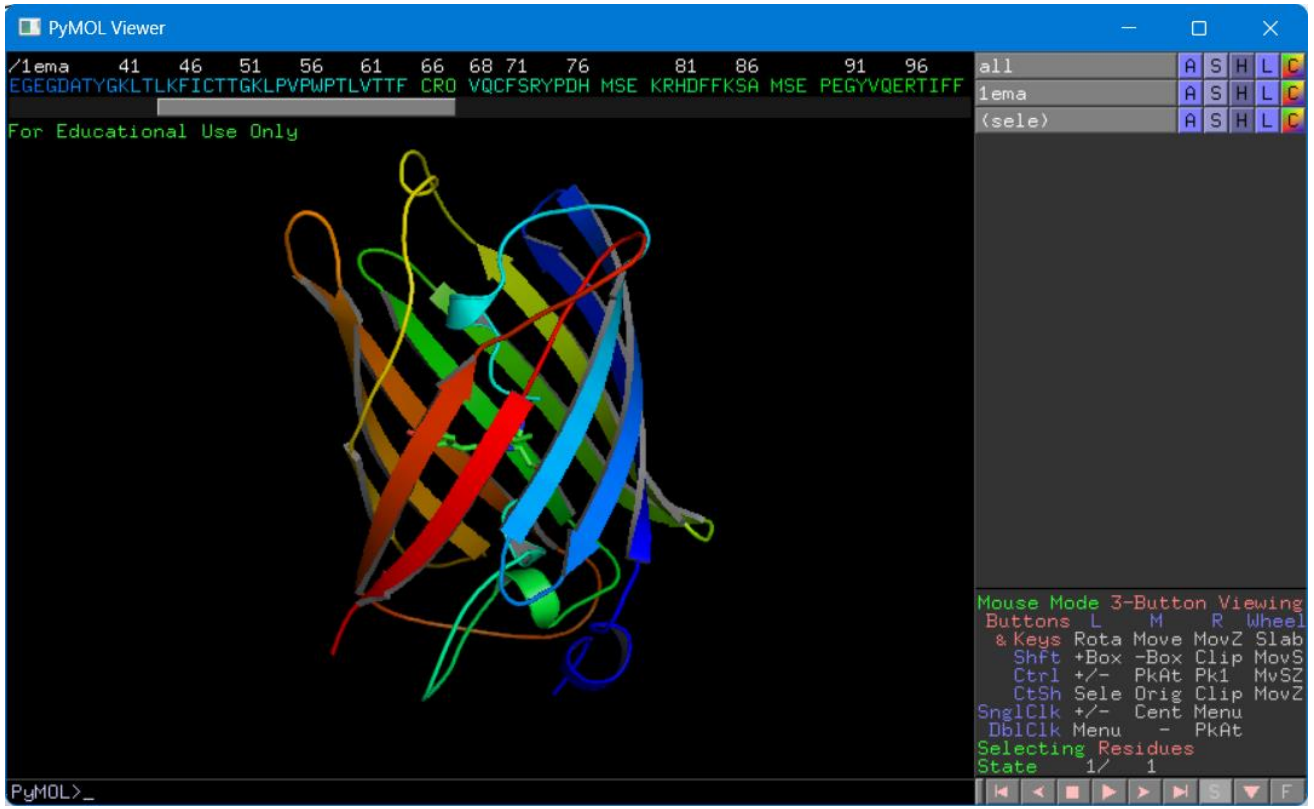
4. 利用PyMol查看蛋白质三维结构

- **PyMol显示窗口中鼠标操作**

  - 旋转图像：按住鼠标左键拖动，分子绕中心旋转；
  - 变换旋转中心：Ctrl/Shift/Alt+鼠标左键点击并拖动光标。
  - 缩放图像：按住鼠标右键向上、向下移动，可放大或缩小图像。
  - 移动剪切平面：按住“Shift”键，按住鼠标右键拖动，分子在X-Y平面内平移。

4. 利用PyMol查看蛋白质三维结构

- 鼠标单击1ema图层旁边的A → preset→ publication, 显示文献中常见的卡通结构图，二级结构以不同颜色显示，小分子以棍状结构显示。
- 再用鼠标中键点击小分子某个原子就可以使小分子居中，以左键拖动就可以看到三级结构的构象变化。



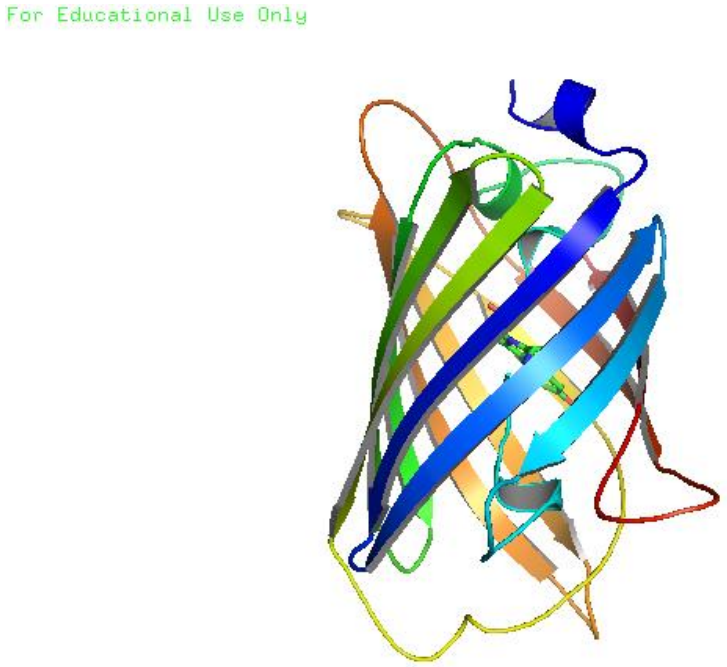
PyMol图形界面

4. 利用PyMol查看蛋白质三维结构

- **分析图像输出**

  - 点击File菜单→ Save Iamge，可以把图像存为PNG格式文件，用于WORD、PPT文档等。
  - 保存图片前，点击Display → Background →White，可将背景调为白色。

- 点击File菜单→ Save Session，保存PyMol文件。



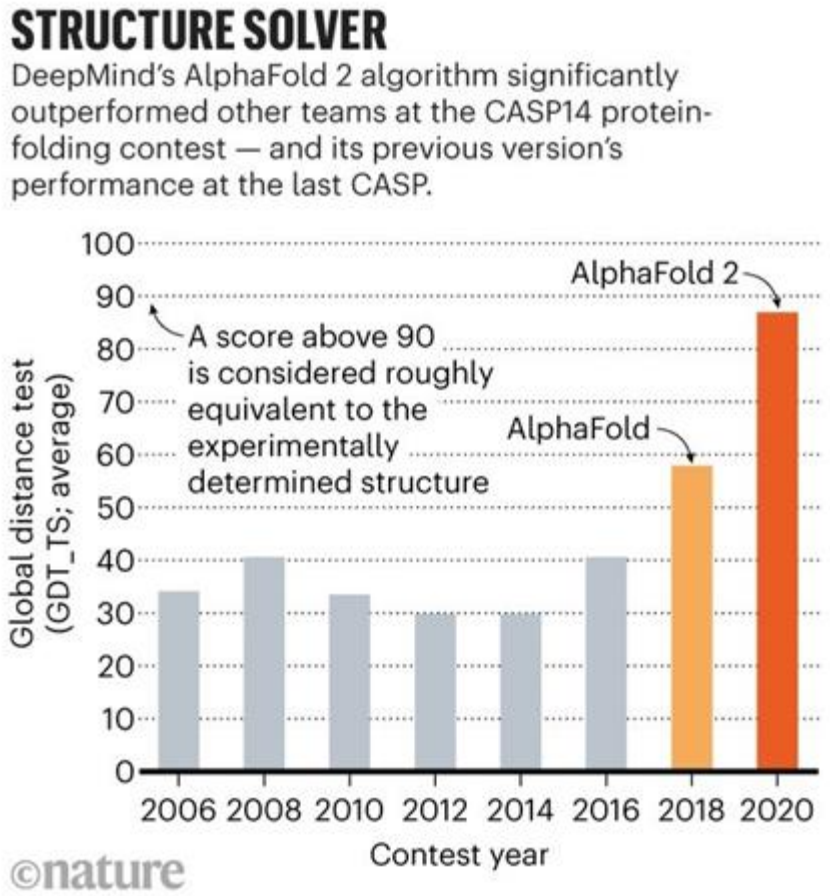
调整显示选项

5. 蛋白质三级结构预测

- 理论预测方法:

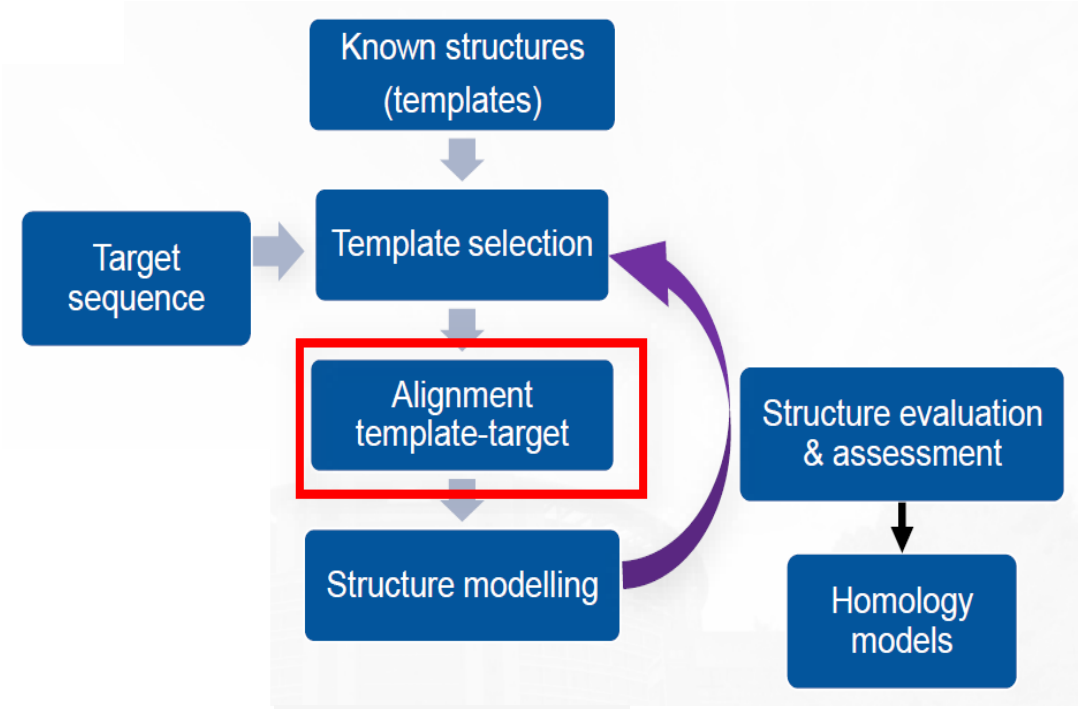
  通过已经建立的各种理论研究计算、推导、预测蛋白质的空间结构

  - **同源建模法(Homology Modeling)**
  - **穿线法(Threading)**
  - **从头计算法(Ab Initio Prediction)**
    - **AlphaFold - 深度学习算法**



Homology Modeling

- 搜索已知三级结构的同源蛋白质序列(模板)
  - PSI-BLAST
  - multiple sequence alignment (MSA)
- 选取与给定序列相似性最高的结构作为模板
- 将氨基酸残基替换到结构模板中对应的位置上，降低自由能
- 准确性好
  - 序列相似性高⇒ 模型可靠性高
  - >30% sequence identity
- 常用工具：MODELLER，Swiss-Model



同源建模流程

利用SWISS-MODEL进行同源建模

BIOZENTRUM University of Basel The Center for Molecular Life Sciences SWISS-MODEL

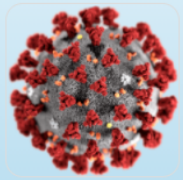
Modelling Repository Tools Documentation Log in Create Account

Welcome to SWISS-MODEL

SWISS-MODEL is a fully automated protein structure homology-modelling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

点击“Starting Modelling”，进入工作页面

Start Modelling



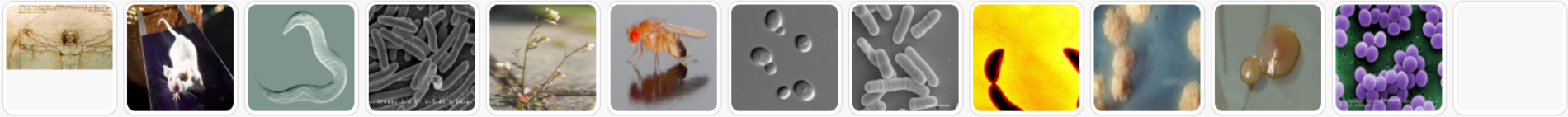
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes coronavirus disease 2019 (COVID-19).

We modelled the full SARS-CoV-2 proteome based on the NCBI reference sequence NC_045512 and annotations from UniProt.

The results are available here.

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in SWISS-MODEL Repository?

Search SWISS-MODEL Repository

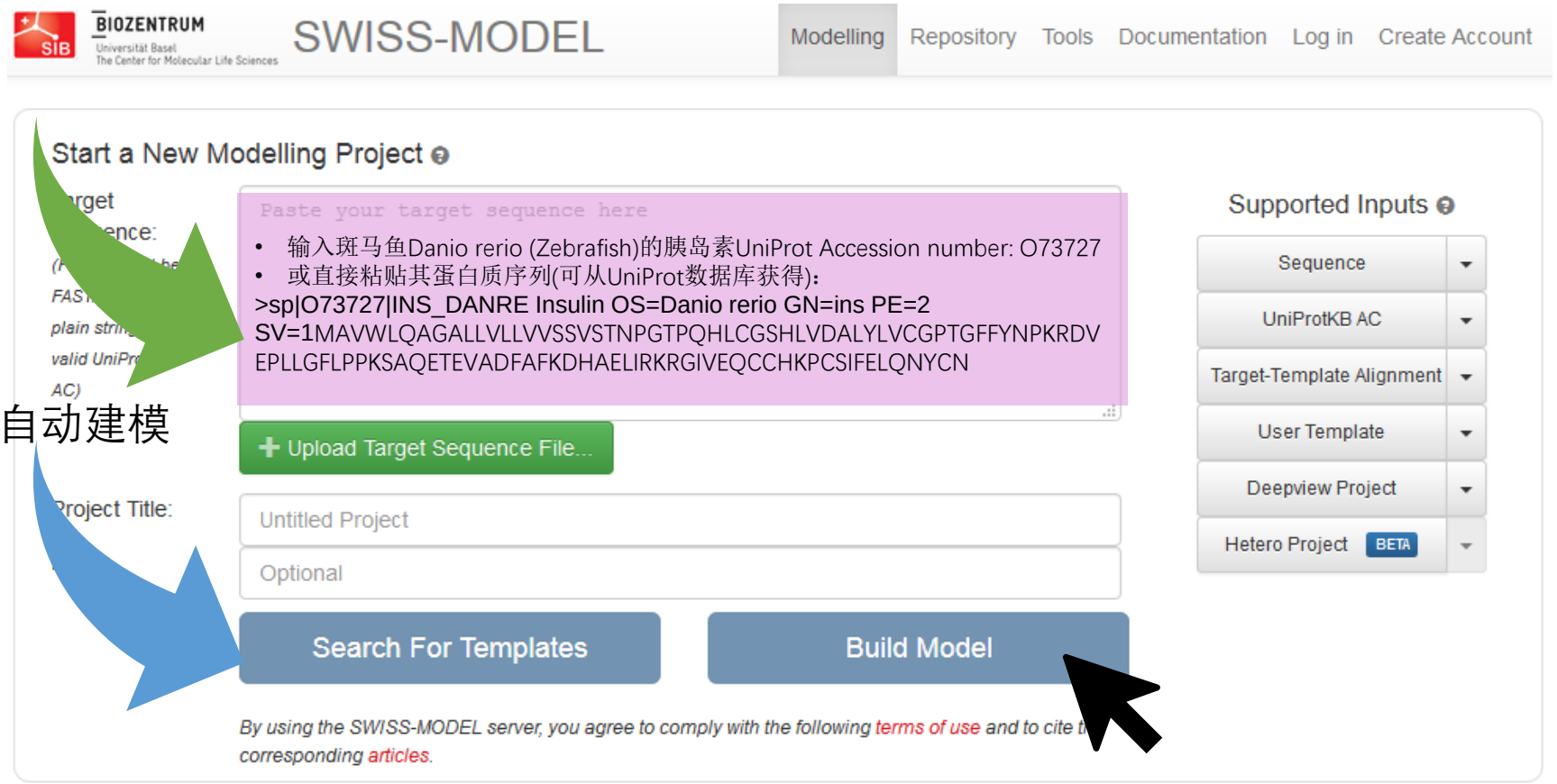


https://swissmodel.expasy.org/

利用SWISS-MODEL进行同源建模

1.输入蛋白质序列信息

2.选择蛋白模版或直接提交自动建模



SWISS-MODEL输入页面

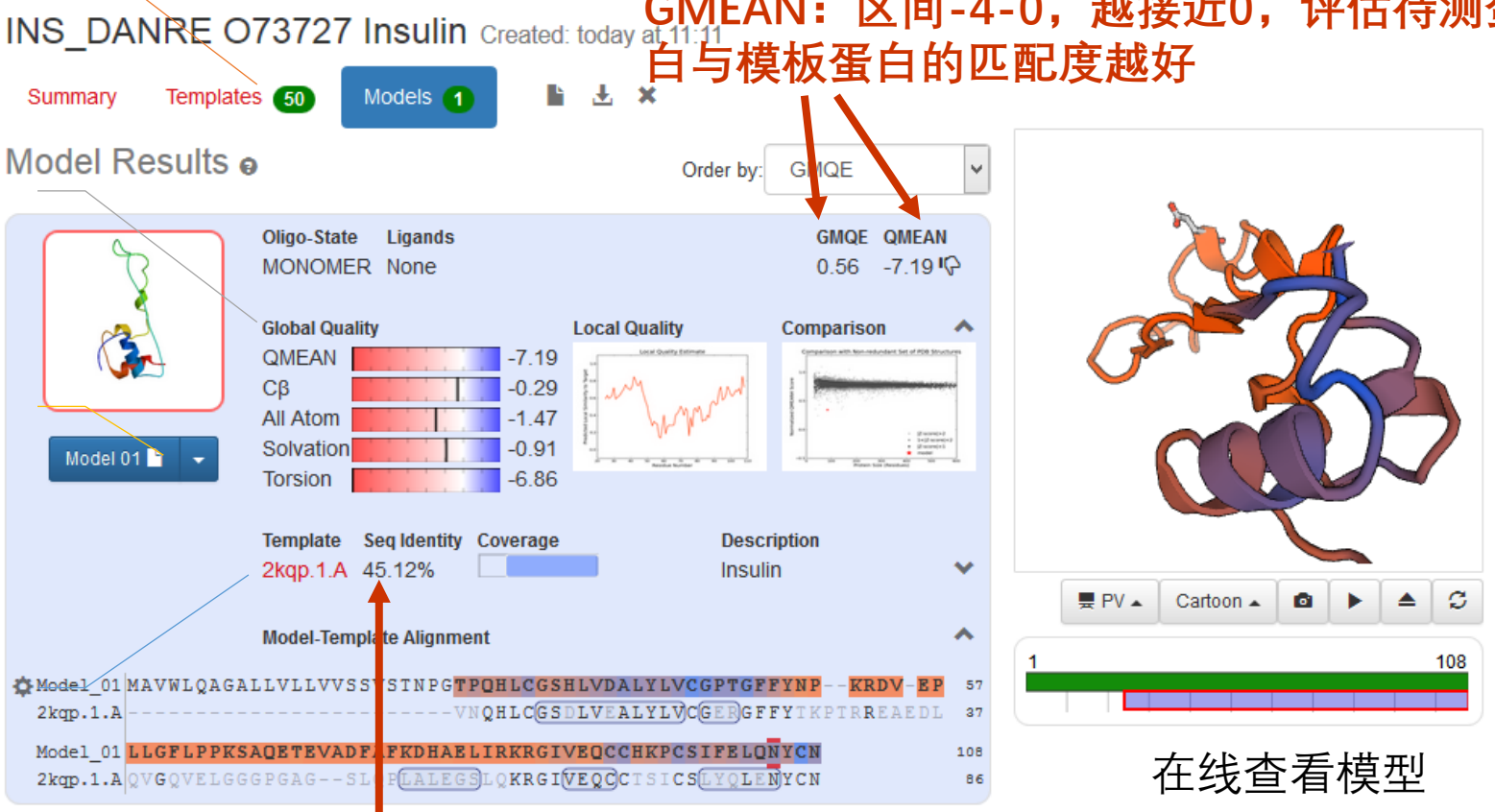
利用SWISS-MODEL进行同源建模

查看数据库中的结构信息以及本次采用的模版

建模的简要评估

点击PDB format链接下载PDB文件到本地，用PyMol查看结构

点击链接跳转新页面，从SMTL.id-coordinates-download coordinate下载模版PDB文件



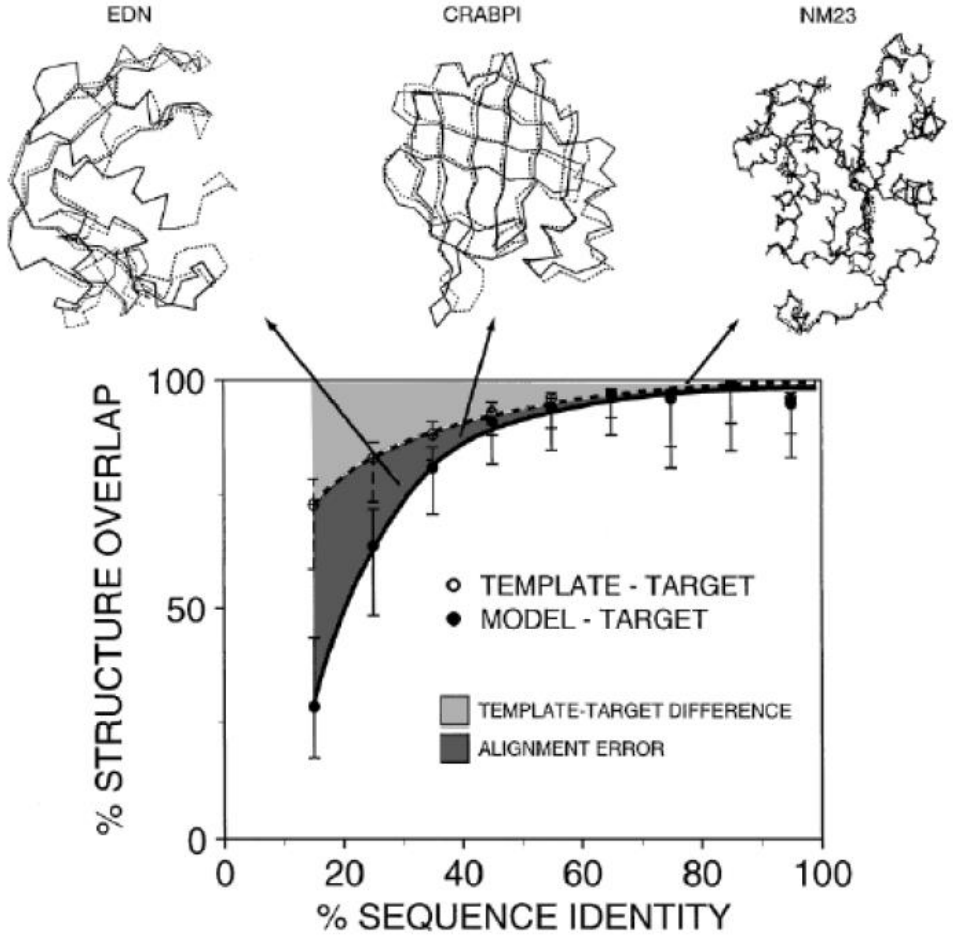
1. 序列一致性>30%，预测结果可用

2.若结果可用，查看评分
GMQE：置信度为0-1，值越大表明质量越好；
GMEAN：区间-4-0，越接近0，评估待测蛋白与模板蛋白的匹配度越好

在线查看模型

SWISS-MODEL建模结果页面

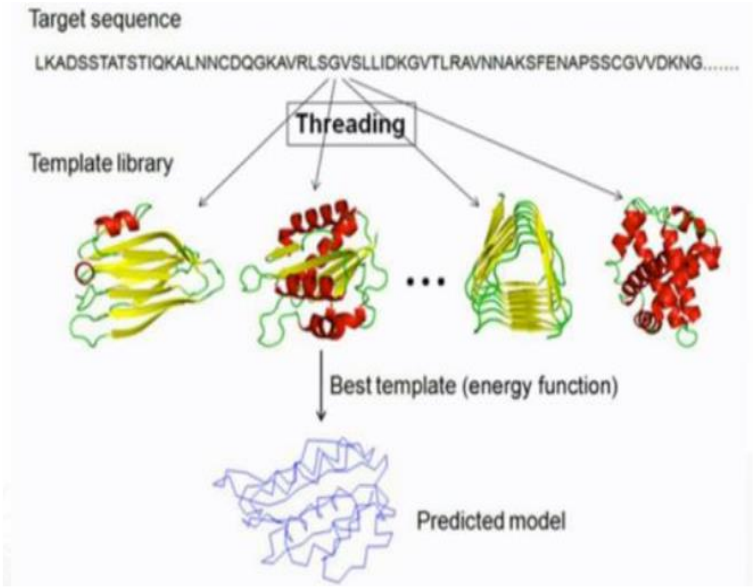
如果目标序列与模板序列一致度极高，那么同源建模法是最准确的方法。



Marti-Renom et al. *Annu. Rev. Biophys. Biomol. Struct.* 29 (2000): 291-325.

Threading - Fold Recognition(折叠识别)

- 实验发现： 蛋白质折叠的类型有限 (~1,000)
  - 问题： 能否根据不同的模版，预测给定蛋白质的折叠类型，并进一步拼装成三级结构?
- 计算要求:
  - 能量函数
  - 模版库(template library)
- 计算方法
  - 将给定序列与每一个模板的序列匹配，打分
  - 将模板连接起来，氨基酸残基替代
  - 优化模型：能量函数
- 计算性能： 不定
  - 序列相似性高⇒ 模型可靠性高
  - 常用工具：I-TASSER



Online Services
- I-TASSER
- QUARK
- LOMETS
- COACH
- COFACTOR
- MetaGO
- MUSTER
- CEthreader
- SEGMER
- FG-MD
- ModRefiner
- REMO
- DEMO
- SPRING
- COTH
- BSpred
- ANGLOR
- EDock
- BSP-SLIM
- SAXSTER
- FUpred
- ThreaDom
- ThreaDomEx
- EvoDesign
- GPCR-I-TASSER
- MAGELLAN
- BindProf

I-TASSER

Protein Structure & Function Predictions

(The server completed predictions for 575136 proteins submitted by 137484 users from 149 countries or regions)
(The template library was updated on 2020/09/22)

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure prediction and structure-based function annotation. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database BioLiP. I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments. It was also ranked the best for function prediction in CASP9. The server is in active development with the goal to provide the most accurate protein structure and function predictions using state-of-the-art algorithms. Please report problems and questions at I-TASSER message board and our developers will study and answer the questions accordingly. (>> More about the server ...)

Structure models for the 2019-nCov Coronavirus genome by C-I-TASSER NEW

[Queue] [Forum] [Download] [Search] [Registration] [Statistics] [Remove] [Potential] [Decoys] [News] [Annotation] [About] [FAQ]

I-TASSER On-line Server (View an example of I-TASSER output):

Copy and paste your sequence within [10, 1500] residues in FASTA format. Click here for a sample input:

Or upload the sequence from your local computer:

选择文件 未选择文件

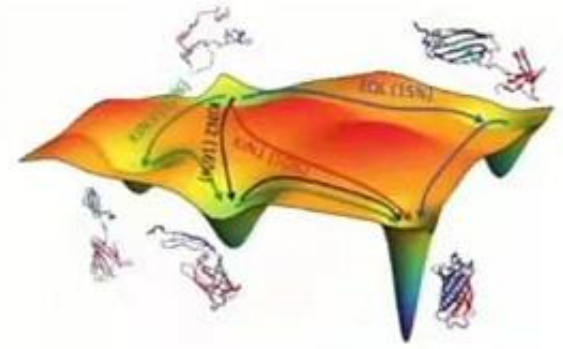
Email: (mandatory, where results will be sent to)

Password: (mandatory, please click here if you do not have a password)

https://zhanglab.ccmb.med.umich.edu/I-TASSER/

Ab Initio Prediction(从头预测方法)

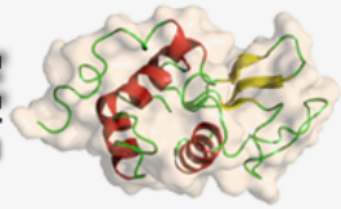
- 蛋白质折叠到最小能量状态：
  - 蛋白质的三维结构取决于自身的氨基酸序列，并且处于最低自由能状态。(Anfinsen, Science 1973)
- 从头预测根据能量函数计算结构的最小自由能：
  - 一个蛋白质有成千上万个原子，巨大的搜索空间，计算量大
  - 能量函数本身不光滑，非常难优化
- 能量函数
  - 键能 (bond energy)
  - 键的转角能 (bond angle energy)
  - 二面角能 (dihedral angle energy)
  - 范德华力 (van der Waals energy)
  - 静电力 (electrostatic energy)
- 常用工具：QUARK
  - 适用于没有同源模板的蛋白质，且氨基酸序列长度200以内



https://www.researchgate.net/figure/Example-of-a-folding-free-energy-landscape-FEL-of-the-green-fl-uores-cent-protein_fig1_290324826

QUARK ONLINE

Ab Initio Protein Structure Prediction



QUARK is a computer algorithm for ab initio protein structure prediction and protein peptide folding, which aims to construct the correct protein 3D model from amino acid sequence only. QUARK models are built from small fragments (1-20 residues long) by replica-exchange Monte Carlo simulation under the guide of an atomic-level knowledge-based force field. QUARK was ranked as the No 1 server in Free-modeling (FM) in CASP9 and CASP10 experiments. Since no global template information is used in QUARK simulation, the server is suitable for proteins that do not have homologous templates in the PDB library. Go to example to view an example of QUARK output. The server is only for non-commercial use. Questions about the QUARK server can be posted at the Service System Discussion Board.

Cut and paste your sequence (in FASTA format, less than 200 AA. Example input

```
DDFSFQWLKYLEYLNMDDNNIPSTKSNTFTGLVSLKYLSLSKTFTSLQTLTNETFVSLAHSPLLTLNLTKNHISKIANGT
FSWLGQLRILDLGLNEIEQKLSGQEWRGLRNIFEIYLSYNKYLQLSTSSFALVPSLQRLMLRRVALKNVDISPSPFRPLR
NLTILDLSNNNIANINEDLLEGLENLEILDFQHNNLARLWKRANPGGPVNFLKGLSHLHILNLESNGLDEIPVGVFKNLF
ELKSINLGLNNLNKLEPFIFDDQTSLRSLNLQKNLITSVEKDVFGPPFQNLNSLDMRFNPFDCTCESISWFVNWINQTHT
NISELSTHYLCNTPHHYYGFPLKLFDTSSCKDSAPFENLYFQGHHHHHHWSHPQFEK
```

Or upload sequence from your computer:

选择文件 未选择文件

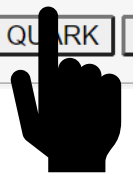
Email: (mandatory, where results will be sent to.)

Password: (mandatory, click here if you do not have a QUARK password)

ID: (optional, name of the protein)

Advanced options [?]

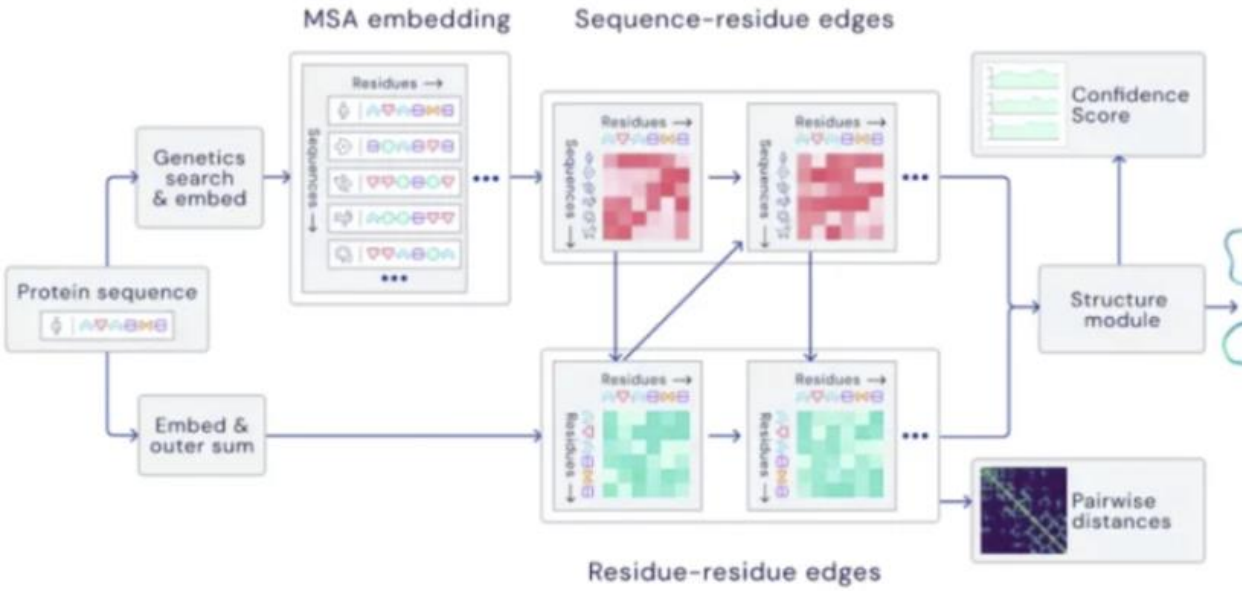
Run QUARK | Clear Form

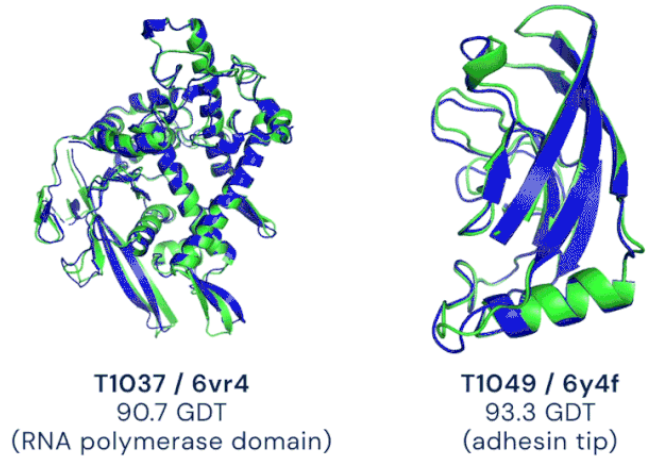


https://zhanglab.ccmb.med.umich.edu/QUARK/

人工智能方法：AlphaFold2

- 基于深度学习（多层神经网络）
- 基于同一家族蛋白质的序列比对
- 不做能量优化，而是预测原子之间的相互作用关系





AlphaFold2预测的蛋白结构与实验结果几乎一致

AlphaFold蛋白质结构数据库

- AlphaFoldDB: https://alphafold.ebi.ac.uk/

模型质量评估软件

- 模型质量评估软件(Model Quality Assessment Programs, MQAPs)
  - 并不比较预测模型与真实结构的差别大小，而是从空间几何学、立体化学和能量分布等方面评估一个模型的自身合理性。
  - PROCHECK、Verify3D、SAVES等

SAVES: The Structure Analysis and Verification Server (v5.0)

作业

1. 在最新版的PDB中检索绿色荧光蛋质GFP的三维结构，下载其PDB文件，并利用PyMOL显示GFP三维结构及其荧光基团位点。

2. 利用SWISS-MODEL预测新冠病毒刺突(Spike)蛋白的三级结构(Spike蛋白的NCB索引号为YP_009724390.1)，注明预测结果评价信息，并利用PyMOL显示其与ACE2相互作用的关键位点。